UNIVERSIDADE DA CORUÑA

DEPARTAMENTO DE MATEMÁTICAS

NONPARAMETRIC STATISTICAL INFERENCE FOR RELATIVE CURVES IN TWO-SAMPLE PROBLEMS

Elisa Mª Molanes López

Marzo 2007

Tesis Doctoral

DEPARTAMENTO DE MATEMÁTICAS

NONPARAMETRIC STATISTICAL INFERENCE FOR RELATIVE CURVES IN TWO-SAMPLE PROBLEMS

Elisa Mª Molanes López
Marzo 2007

Tesis Doctoral

Realizado el acto público de defensa y mantenimiento de esta Tesis Doctoral el día 30 de marzo de 2007 en la Facultad de Informática de la Universidad de La Coruña, ante el tribunal formado por:

Presidente:	Dr. D. Wenceslao González Manteiga
Vocales:	Dr. D. Noël Veraverbeke
	Dra. Dña. Ingrid Van Keilegom
	Dr. D. Kenneth R. Hess
Secretario:	Dr. D. José Antonio Vilar Fernández

obtuvo la máxima calificación de SOBRESALIENTE CUM LAUDE, siendo director de la misma el Dr. D. Ricardo Cao Abad.

To my family and friends

Acknowledgements

First of all, I would like to express how lucky and grateful I feel for having the chance to get a national grant that supported me financially during my PhD and allowed me to write this thesis under the supervision of people such as Ricardo Cao, Noël Veraverbeke, Paul Janssen, Ingrid Van Keilegom and Kenneth Hess. All of them supported me a lot during my research and I could find in them not only good researchers but also very good people. I will never forget their kindness to me and specially their hospitality when I arrived for the first time to their universities and they did not know too much about me. Without them, this would not be possible at all. Besides, the fact that I could travel to Belgium and the United States of America and I could visit the Center for Statistics, the Institut de Statistique and the MD Anderson Medical Center changed my life in a positive way. It was really a good idea of my supervisor to send me there! Thanks as well to my colleagues of the Mathematics Department at the University of A Coruña who always treated me well and created a good work environment.

I would like to thank all the people who supported me along my life and made it better in several ways. Of course, a special mention is given to my sister, Mari Carmen, and my parents, Carmen and Emilio, who were always next to me in an unconditional way, loving me and supporting me; and to Ronald, that even when he is now a bit far from here, he loves me and supports me every day. I will never forget what they are to me, what I am to them and how loved I am.

Of course I would also like to thank all my friends for the good moments.

Finally, I would also like to mention that this research has been supported by Grants BES-2003-1170 (EU ESF support included), BFM2002-00265 and MTM2005-00429 (EU ERDF support included). Thanks are also due to Sonia Pértega Díaz, Francisco Gómez Veiga, Salvador Pita Fernández and Emilio Casariego Vale from the "Hospital Juan Canalejo" in A Coruña for providing the prostate cancer and gastric cancer data sets.

Elisa María Molanes López
A Coruña, Dec 7, 2006

Table of Contents

List of Tables

List of Figures

Preface

Analyzing the duration of life is an interesting issue in many fields. Survival analysis refers to all the statistical methodology developed for analyzing lifetimes or times till the occurrence of an event of interest. There are many fields where these methods have been widely used such as medicine, biology, engineering, and social and economic sciences. In all of them the common and essential element is the presence of a nonnegative variable of interest that sometimes may not be observed completely due to different phenomena such as censoring and truncation. Censoring is a well-known setup in the literature that has received considerable attention for a long time. Under right censoring it may happen that the lifetime is only partially observed due to the previous occurrence of censoring. On the other hand, truncation is another situation that may appear jointly with censorship in survival applications. However, it has been appealed interesting more recently, mainly because of the AIDS epidemic. Under left truncation, it may happen that the time origin of the lifetime precedes the time origin of the study which makes impossible at all to observe the case.

The analysis of survival data includes different methods, such as life tables, regression models and two sample problems, among others. The oldest technique used to describe the survival in a sample is to compute its corresponding life table, the earliest nonparametric hazard rate estimate which is based on grouped lifetimes. When some covariates are observed it may be interesting to study if they are correlated with the survival time. If this is the case, then regression models appropriately defined, such as the Cox proportional hazard model or a frailty model, are useful techniques to describe such effects. The main difference between a frailty model and the Cox proportional hazard model is that the former is specifically designed to deal with clustered survival data, where the units or events of observation are grouped in clusters of fixed or variable size and sometimes show a certain ordering within the cluster. The novelty of a frailty model is that it takes into account the correlation structure observed between the events within a cluster by introducing a frailty term for each cluster. Sometimes, the interest is to compare the survival times of two populations. For example, in clinical trials when the objective is to study the effectiveness of a new treatment for a generally terminal disease. Comparing the survival times of both groups (the treated group against the placebo group) can address this question. ROC curves and PP plots are graphical procedures that provide a graphical idea of possible differences between two lifetime distributions. More recent approaches are the relative density, the relative distribution and the relative hazard function of a lifetime with respect to another one.

In this monograph, kernel type estimators of the relative density and the relative distribution functions are presented and different global bandwidth selectors are designed to appropriately select the smoothing parameter of the relative density kernel type estimators.

In Chapter 1 a more detailed introduction to survival analysis, the bootstrap, nonparametric curve estimation, two sample problems and relative curves is given.

The simplest case when the data are completely observed is studied in Chapter 2. Several bandwidth selectors are designed for two kernel type estimators of the relative density, based on plug-in ideas and the bootstrap technique. A simulation study presents some results where the behavior of these and a classical selector are compared.

Chapter 3 deals with the problem of estimating the relative density and relative distribution with right censored and left truncated data. Three bandwidth selectors are proposed for the relative density kernel type estimator considered for this scenario, and their performance, under different percentages of censoring and truncation, is checked through a simulation study.

In Chapter 4, a test of the null hypothesis of equal populations is designed using the relative distribution function via the empirical likelihood approach.

Chapter 5 includes two applications to real data where the techniques studied in previous chapters and extensions of them are applied. One of the data sets is related to prostate cancer (PC). The second data set contains information regarding patients who suffer from gastric cancer (GC). While the PC data set contains complete information regarding the variables of interest, this is not the case in the GC data set, which is a clear example of left truncated and right censored data, two very common characteristics of the data coming from survival analysis. After a descriptive study of both data sets, we apply the methodology developed in this monograph. For example, estimates of the relative density for different variables of interest registered in two groups that we wish to compare, are computed using the kernel type estimators and the bandwidth selectors introduced in Chapters 1–3.

Finally, in Chapter 6, we introduce some future research lines. The Appendix includes some definitions, theorems and inequalities used along the thesis.

It is interesting to mention here that all the simulation studies and practical applications carried out along this research have been implemented in Matlab 7.0.

Chapter 1

Introduction

— Todas aquellas cosas del sótano estaban ordenadas,
tenían sentido, eran parte del pasado,
pero completaban la historia del presente.
Paulo Coelho

1.1 Survival analysis

The field of survival analysis emerged in the 20th century and experienced a big growth during the second half of the century, specially due to the Kaplan-Meier method (1958) to estimate the survival function of a sample with censored data, the Cox proportional hazards model (1972) to quantify the effects of covariates on the lifetime or the martingale-based approach to survival analysis introduced by Aalen (1975). Next, we concentrate on a small part of the very rich field of survival analysis, the left truncated and right censored model.

1.1.1 The LTRC model

Let X_0, C_0 and T_0 denote respectively, the variable of interest with cdf F_0, the variable of censoring with cdf L_0 and the variable of truncation with cdf G_0.

When the data are subject to both left truncation and right censoring (LTRC), the statistician observes, for every individual in the sample, the random vector (T_0, Y_0, δ_0), where Y_0 denotes the minimum between the variable of interest and the variable of censoring, with cdf W_0, and $\delta_0 = 1_{\{X_0 \leq C_0\}}$ indicates if the variable of interest is censored ($\delta_0 = 0$) or not ($\delta_0 = 1$). Besides, the data are observable only when the condition $T_0 \leq Y_0$ holds. Otherwise, when $T_0 > Y_0$, nothing is observed. Therefore, under the LTRC model, the statistician is able to observe a sample of n vectors, $\{(T_{01}, Y_{01}, \delta_{01}),\ldots,(T_{0n}, Y_{0n}, \delta_{0n})\}$, where the observed sample size, n, is random as a consequence of the truncation mechanism. Denoting by N the real sample size, that is fixed but unknown, and defining α_0 as the probability of absence of truncation in the population, $\alpha_0 = P(T_0 \leq Y_0)$, it follows that n is a $\mathrm{Bin}(N, \alpha_0)$ random variable. By the Strong Law of Large Numbers (SLLN), it is easy to prove that $\frac{n}{N} \to \alpha_0$ almost surely as $N \to \infty$.

Considering that $T_0 \leq Y_0$ holds, if the observation is uncensored ($\delta_0 = 1$), the variable of interest is completely observed ($Y_0 = X_0$). However, when the observation is censored ($\delta_0 = 0$), Y_0 coincides with the censoring value C_0 and one only knows that the variable of interest, X_0, is larger than the observed censoring value C_0.

It is assumed that the random variables, X_0, C_0 and T_0, are mutually independent and that their cdf's are continuous. Therefore, under these conditions, it is easy to prove that W_0 can be expressed in terms of F_0 and L_0 as follows:

$$1 - W_0 = (1 - F_0)(1 - L_0).$$

Without loss of generality we will assume that the random variables are positive and we will use, for any cdf, say F, the following notation to denote respectively, its left-continuous inverse or quantile function and the left and right endpoints of its support:

$$F^{-1}(t) = \inf\{x \in \mathbb{R} : F(x) \geq t\}, \tag{1.1}$$

$$a_F = \inf\{x \in \mathbb{R} : F(x) > 0\},$$

$$b_F = \sup\{x \in \mathbb{R} : F(x) < 1\}.$$

Conditional on the value of n, $(T_{0i}, Y_{0i}, \delta_{0i}), i = 1, \ldots, n$, are still iid and the joint distribution of (Y_0, T_0) is given by

$$H_0(y,t) = P\{Y_0 \leq y, T_0 \leq t | T_0 \leq Y_0\} = \alpha_0^{-1} \int_0^y G_0(t \wedge z) \mathrm{d}W_0(z) \text{ for } y,t > 0,$$

where $\wedge$ denotes the minimum of two values.

Related to this model, we introduce below the following definitions that will be needed later on for further discussion:

$$\begin{aligned} B_0(t) &= P(T_0 \leq t \leq Y_0 / T_0 \leq Y_0) = \alpha_0^{-1} P(T_0 \leq t \leq Y_0) \\ &= \alpha_0^{-1} G_0(t)(1 - F_0(t))(1 - L_0(t)) = \alpha_0^{-1} G_0(t)(1 - W_0(t)), \end{aligned}$$

$$\begin{aligned} W_{01}(t) &= P(Y_0 \leq t, \delta_0 = 1 / T_0 \leq Y_0) = \alpha_0^{-1} P(X_0 \leq t, T_0 \leq X_0 \leq C_0) \\ &= \int_{a_{F_0}}^t \alpha_0^{-1} P(T_0 \leq y \leq C_0) \mathrm{d}F_0(y) \\ &= \int_{a_{F_0}}^t \alpha_0^{-1} G_0(y)(1 - L_0(y)) \mathrm{d}F_0(y), \end{aligned}$$

and

$$\xi_{0i}\left(Y_{0i}, T_{0i}, \delta_{0i}, z\right) = \frac{1_{\{Y_{0i} \leq z, \delta_{0i}=1\}}}{B_0(Y_{0i})} - \int_{a_{W_0}}^{z} \frac{1_{\{T_{0i} \leq u \leq Y_{0i}\}}}{B_0^2(u)} \mathrm{d}W_{01}(u), \tag{1.2}$$

where B_0 denotes the difference between the marginal distribution functions of Y_0 and T_0, given $T_0 \leq Y_0$, W_{01} denotes the conditional subdistribution of the uncensored data given $T_0 \leq Y_0$, and $\xi_{0i}\left(Y_{0i}, T_{0i}, \delta_{0i}, z\right), i = 1, \ldots, n$, are iid processes with mean zero and covariance structure given by:

$$\begin{aligned} \Gamma(z_1, z_2) &= Cov\left[\xi_{0i}\left(Y_{0i}, T_{0i}, \delta_{0i}, z_1\right), \xi_{0i}\left(Y_{0i}, T_{0i}, \delta_{0i}, z_2\right)\right] \qquad (1.3) \\ &= q_0(z_1 \wedge z_2), \end{aligned}$$

where

$$q_0(z) = \int_{a_{W_0}}^{z} \frac{\mathrm{d}W_{01}(t)}{B_0^2(t)}. \tag{1.4}$$

1.1.2 Distribution function estimation under LTRC data

For randomly left truncated and right censored data, the most widely used estimator of the survival function is that proposed by Tsai, Jewell and Wang (1987). It is known in the literature as the TJW product limit estimator and generalizes that proposed by Kaplan and Meier (1958) for right censored data and the one by Lynden-Bell (1971) for left truncated data. Assuming no ties in the data, the TJW product limit estimator is defined as follows

$$1 - \hat{F}_{0n}(x) = \prod_{Y_{0i} \leq x} \left[1 - \left(nB_{0n}(Y_{0i})\right)^{-1}\right]^{\delta_{0i}}, \tag{1.5}$$

where $B_{0n}(z)$ denotes the empirical estimate of the function $B_0(z)$, introduced in Subsection 1.1.1:

$$B_{0n}(z) = n^{-1} \sum_{i=1}^{n} 1_{\{T_{0i} \leq z \leq Y_{0i}\}} = n^{-1} \sum_{i=1}^{n} 1_{\{T_{0i} \leq z\}} 1_{\{z \leq Y_{0i}\}}.$$

There is a simple way to generalize (1.5) for addressing the presence of ties in the data. Let assume that $t_{01}, \ldots, t_{0s}$ denote the s distinct failure times observed in the data. Then, a modified version of (1.5) that takes into account the presence of ties is given by

$$1 - \tilde{F}_{0n}(x) = \prod_{t_{0i} \leq x} \left[1 - \frac{D_{0n}(t_{0i})}{nB_{0n}(t_{0i})}\right], \tag{1.6}$$

where, for a given time point t, $nB_{0n}(t)$ represents the number of individuals at risk and $D_{0n}(t)$ represents the number of individuals that fail at t, i.e.

$$D_{0n}(t) = \sum_{j=1}^{n} 1_{\{Y_{0j}=t, \delta_{0j}=1\}}.$$

Some care is recommended when applying directly (1.6). In practical applications, it may happen that at a given time point $t_0 \in \{t_{01}, \ldots, t_{0s}\}$, the number of individuals at risk and failures are equal. In situations like this, the straightforward application of (1.6) leads to $1 - \tilde{F}_{0n}(t) = 0$ for all $t \geq t_0$, even when survivors and deaths are observed beyond that point. One possibility to overcome this problem is to estimate the survival function conditional on survival to a time for which it is guaranteed that this will never happen.

As it was pointed out by several authors, under the LTRC model, F_0 is identifiable only if some conditions on the support of F_0 and G_0 are satisfied. For right censored data it can be difficult to handle the upper tail of F_0. The presence of a random mechanism of left truncation creates further complications in the lower tail that are propagated throughout the entire observable range. Therefore, under the LTRC model, both the upper and the lower tails of F_0 are affected. Due to this fact, the product limit estimators (1.5) and (1.6) show an unstable behaviour when the size of the risk set is small. A slightly modified version of them, say $\check{F}_{0n}$, was proposed by Lai and Ying (1991) to solve this problem. Introducing a weight function that discards those factors of the product in (1.6) that correspond to small risk set sizes, $\check{F}_{0n}$ is defined as follows:

$$1 - \check{F}_{0n}(x) = \prod_{t_{0i} \leq x} \left[1 - \frac{D_{0n}(t_{0i}) 1_{\{nB_{0n}(t_{0i}) \geq cn^{\alpha}\}}}{nB_{0n}(t_{0i})} \right], \tag{1.7}$$

where c and α are two previously specified values satisfying that $c > 0$ and $0 < \alpha < 1$. Like (1.6) does, $\check{F}_{0n}$ takes into account the presence of ties in the data.

Table 1.1: Comparing the behaviour of $\tilde{F}_{0n}(t)$ vs $\check{F}_{0n}(t)$.

t	$nB_{0n}(t)$	$F_0(t)$	$\tilde{F}_{0n}(t)$	$\check{F}_{0n}(t)$
0.1000	76	0.0328	0.0485	0.0485
0.3053	72	0.0967	0.0989	0.0989
0.5105	59	0.1565	0.1387	0.1387
0.7158	50	0.2123	0.1719	0.1719
0.9211	40	0.2644	0.2425	0.2425
1.1263	33	0.3130	0.3025	0.3025
1.3316	28	0.3584	0.3475	0.3475
1.5368	24	0.4009	0.3736	0.3736
1.7421	19	0.4405	0.4035	0.4035
1.9474	13	0.4775	0.4718	0.4718
2.1526	5	0.5121	0.5925	0.5925
2.3579	4	0.5443	0.6944	0.6944
2.5632	3	0.5745	0.6944	0.6944
2.7684	3	0.6026	0.6944	0.6944
2.9737	2	0.6289	0.8472	0.6944
3.1789	1	0.6534	0.8472	0.6944

Next, an example is introduced with the objective of illustrating the behaviour of (1.6) and (1.7) when the risk sets become small. Let us consider $\{X_{01}, \ldots, X_{0n}\}$ a sample of n iid exponential random variables with mean 3 that are subject to both right censoring and left truncation. Let $\{T_{01}, \ldots, T_{0n}\}$ be a sample of n iid truncation variables with distribution function G_0 given by

$$G_0(x) = (1 + \exp(-x))^{-1}, \quad x \in (-\infty, \infty)$$

and let $\{C_{01}, \ldots, C_{0n}\}$ be a sample of n iid censoring variables with distribution function L_0 given by

$$L_0(x) = 1 - \left(1 - \frac{x}{5}\right)^3, \quad x \in (0, 5).$$

Denoting by $Y_{0i} = \min\{X_{0i}, C_{0i}\}$, only those values that satisfy $T_{0i} \leq Y_{0i}$ are considered. We use the index $i = 1, \ldots, n$ for the first n values satisfying this condition.

First, in Table 1.1 the values of $\tilde{F}_{0n}(t)$ and $\check{F}_{0n}(t)$ (with $c = 1$ and $\alpha = \log_{10}(3)/2$) based on a simulated sample of $n = 100$ observed data from this LTRC model are tabulated for different values of t and compared with the true values. Also, the number of individuals at risk is given. Note that based on the selected values for c and α, $\check{F}_{0n}(t)$ discards risk sets of less than 3 individuals. Table 1.2 shows the mean squared error (MSE) obtained as the average of the squared deviations between the true value of $F_0(t)$ and the estimates (either $\tilde{F}_{0n}(t)$ or $\check{F}_{0n}(t)$) obtained over 1000 simulations for different values of t. Also the minimum, maximum, mean, median and standard deviation of the estimates are given in this table.

Table 1.2: Results obtained from a simulation study performed to compare the behaviour of $\tilde{F}_{0n}(t)$ vs $\check{F}_{0n}(t)$ when the risk sets associated to t are small.

	min	mean	median	max	std	MSE
$\tilde{F}_{0n}(3)$	0.2918	0.6305	0.6248	1.0000	0.1074	0.0115
$\check{F}_{0n}(3)$	0.2918	0.6212	0.6210	0.8651	0.0944	0.0090
$\tilde{F}_{0n}(4)$	0.3351	0.7067	0.6930	1.0000	0.1425	0.0212
$\check{F}_{0n}(4)$	0.3161	0.6488	0.6565	0.8618	0.0952	0.0167
$\tilde{F}_{0n}(5)$	0.3552	0.7159	0.6998	1.0000	0.1499	0.0315
$\check{F}_{0n}(5)$	0.3552	0.6484	0.6535	0.8701	0.0991	0.0363
$\tilde{F}_{0n}(6)$	0.3688	0.7206	0.6985	1.0000	0.1494	0.0430
$\check{F}_{0n}(6)$	0.3130	0.6505	0.6532	0.8691	0.0962	0.0551

After looking at Table 1.1 and the standard deviations shown in Table 1.2, it is evident the unstable behaviour that (1.6) presents when the risk sets become small. However, looking at the MSE of the estimates, which is a measure of the deviation between the real value and the estimates, it is observed that (1.7) does not always outperform (1.6). The reason is that even when the estimates given by $\check{F}_{0n}(t)$ present less variability, they tend to present larger biases. Since the MSE takes into account the variability and the bias of the estimates jointly, the final MSE turns out to be larger for $\check{F}_{0n}(t)$ than for $\tilde{F}_{0n}(t)$, for some values of t.

In the literature, several results have been proved for the TJW product limit estimator, $\hat{F}_{0n}$. They are collected below in the following theorems.

Theorem 1.1.1. *(Strong representation: Theorem 1 (c) in Gijbels and Wang (1993))*
For $a_{G_0} < a_0 \leq z \leq b_0 < b_{W_0}$, it follows that $B_0(z) \geq \epsilon$ for some $\epsilon > 0$, and if $a_{G_0} < a_{W_0}$, then

$$\hat{F}_{0n}(z) - F_0(z) = (1 - F_0(z))n^{-1} \sum_{i=1}^{n} \xi_{0i} \left(Y_{0i}, T_{0i}, \delta_{0i}, z\right) + s_{0n}(z),$$

where the ξ_{0i}'s have been defined in (1.2),

$$P\left(\sup_{0\leq z\leq b_0} n \left|s_{0n}(z)\right| > x + 4\epsilon^{-2}\right) \leq K\left[e^{-\lambda x} + (x/50)^{-2n} + e^{-\lambda x^3}\right]$$

with some $\lambda > 0$, and this implies that

$$\sup_{0\leq z\leq b_0} \left|s_{0n}(z)\right| = O\left(n^{-1}\ln n\right) \quad \text{a.s.}$$

and

$$E\left(\sup_{0\leq z\leq b_0} \left|s_{0n}(z)\right|^{\alpha}\right) = O\left(n^{-\alpha}\right), \text{ for any } \alpha > 0.$$

Theorem 1.1.2. *(Weak convergence: Corollary 1 (d) in Gijbels and Wang (1993))*
Assume that $a_{G_0} < a_{W_0}$ and $b_0 < b_{W_0}$. Then,

(a) For $0 < z < b_{W_0}$, $\hat{F}_{0n}(z) \to F_0(z)$ a.s.

(b) $\sup_{0\leq z\leq b_0} \left|\hat{F}_{0n}(z) - F_0(z)\right| = O\left(\left(n^{-1}\ln\ln n\right)^{1/2}\right)$ a.s.

(c) The PL-process $\alpha_{0n}(z) = n^{1/2}\left(\hat{F}_{0n}(z) - F_0(z)\right)$ converges weakly on $D[0, b_0]$ to a Gaussian process with mean zero and covariance structure given by

$$\left[1 - F_0(z_1)\right]\left[1 - F_0(z_1)\right]\Gamma(z_1, z_2),$$

where $\Gamma(z_1, z_2)$ is given in (1.3).

Theorem 1.1.3. *(Strong Gaussian approximation when $a_{G_0} = a_{W_0}$: Theorem 2 in Zhou (1996))*
Assuming that $a_{G_0} = a_{W_0}$ and the integral condition, $\int_{a_{W_0}}^{\infty} \frac{\mathrm{d}F_0(z)}{G_0^2(z)} < \infty$, holds. Then,

$$\hat{F}_{0n}(z) - F_0(z) = (1 - F_0(z))n^{-1} \sum_{i=1}^{n} \xi_{0i} \left(Y_{0i}, T_{0i}, \delta_{0i}, z\right) + s_{0n}(z)$$

uniformly in $a_{W_0} \leq z \leq b_0 < b_{W_0}$ *with*

$$\sup_{a_{W_0} \leq z \leq b_0} |s_{0n}(z)| = O\left(n^{-1} \ln^{1+\epsilon} n\right) \quad \text{a.s. } for\ \epsilon > 0.$$

Theorem 1.1.4. *(Strong representation: Theorem 2.2 in Zhou and Yip (1999)) Suppose that* $a_{G_0} \leq a_{W_0}$ *and the integral condition,* $\int_{a_{W_0}}^{b_0} \frac{\mathrm{d}W_{01}(z)}{B_0^3(z)} < \infty$, *is satisfied for some* $b_0 < b_{W_0}$. *Then, uniformly in* $a_{W_0} \leq z \leq b_0 < b_{W_0}$, *we have*

$$\hat{F}_{0n}(z) - F_0(z) = (1 - F_0(z))n^{-1} \sum_{i=1}^{n} \xi_{0i}\left(Y_{0i}, T_{0i}, \delta_{0i}, z\right) + s_{0n}(z)$$

with

$$\sup_{a_{W_0} \leq z \leq b_0} |s_{0n}(z)| = O\left(n^{-1} \ln \ln n\right) \quad \text{a.s.}$$

It is worth mentioning here that Li (1995) obtained an almost sure representation of the TJW product limit estimator analogously to the result given previously by Gijbels and Wang (1993) and that we summarized in Theorem 1.1.1 above. While Gijbels and Wang (1993) proved their result using the classical approach considered by Major and Rejto (1988) for the censored case, Li (1995) used in his proof results of empirical U-statistics processes. Apart from this subtle difference, the result given by Li (1995) covers the case $a_{G_0} = a_{W_0}$ which was not included in the previous result by Gijbels and Wang (1993). Besides, Li (1995) requires that the integral condition, $\int_{a_{W_0}}^{\infty} \frac{\mathrm{d}F_0(s)}{P(T_0 \leq s \leq C_0)^2} < \infty$, is satisfied and states that the remainder term is of order $O\left(n^{-1} \ln^3 n\right)$ a.s. instead of order $O\left(n^{-1} \ln n\right)$ a.s. It is later on, in 1996 and 1999, when other authors study the case $a_{G_0} = a_{W_0}$, obtaining similar results (see Theorems 1.1.3 and 1.1.4 above) to the previously obtained by Li in 1995.

Theorem 1.1.5. *(Exponential bound: Theorem 1 in Zhu (1996))*
a) Let consider $a_{G_0} < a_{W_0}$ *and* $b_0 < b_{W_0}$.
Then, for $\epsilon > 8\alpha_0^2 \left(nG_0(a_{W_0})(1 - W_0(b_0))\right)^{-1}$, *where* α_0 *was defined in Subsection 1.1.1, it follows that*

$$P\left\{\sup_{a_{W_0} \leq z \leq b_0} \left|\hat{F}_{0n}(z) - F_0(z)\right| > \epsilon\right\} \leq D_1 \exp\left\{-D_2 \epsilon^2 n\right\},$$

where D_1 *and* D_2 *are absolute constants.*
b) Let consider $a_{G_0} = a_{W_0}$, $a_{W_0} < a_0 < b_0 < b_{W_0}$ *and*

$$\int_{a_{W_0}}^{\infty} \frac{\mathrm{d}F_0(z)}{G_0^2(z)} < \infty.$$

Then, for $\epsilon > 8\alpha_0^2 \left(nG_0(a_0)(1-W_0(b_0))\right)^{-1}$ *it follows that*

$$P\left\{\sup_{a_0\le z\le b_0}\left|\hat{F}_{0n}(z)-F_0(z)\right|>\epsilon\right\}\le \tilde{D}_1\exp\left\{-\tilde{D}_2\epsilon^2 n\right\},$$

where $\tilde{D}_1$ *and* $\tilde{D}_2$ *are absolute constants.*

Theorem 1.1.6. *(Functional law of the iterated logarithm: Corollary 2.2 in Zhou and Yip (1999) and Theorem 2.3 in Tse (2003))*
Suppose that $a_{G_0}\le a_{W_0}$ *and the integral condition,* $\int_{a_{W_0}}^{b_0}\frac{\mathrm{d}W_{01}(z)}{B_0^3(z)}<\infty$*, holds for some* $b_0<b_{W_0}$*. Then, it follows that*

$$\limsup_{n\to\infty}\left(\frac{n}{2\ln\ln n}\right)^{1/2}\sup_{a_{W_0}<z\le b_0}\left|\hat{F}_{0n}(z)-F_0(z)\right|=\sup_{a_{W_0}<z\le b_0}\sigma_0(z)\quad \text{a.s.}$$

and

$$\liminf_{n\to\infty}(n\ln\ln n)^{1/2}\sup_{a_{W_0}<z\le b_0}\frac{\left|\hat{F}_{0n}(z)-F_0(z)\right|}{1-F_0(z)}=\frac{\pi}{8^{1/2}}(q_0(b_0))^{1/2}\quad \text{a.s.,}$$

where $\sigma_0^2(z)=(1-F_0(z))^2q_0(z)$ *for* $a_{W_0}<z<b_{W_0}$ *and* $q_0(z)$ *has been defined in (1.4).*

Theorem 1.1.6 above was first proved by Zhou and Yip (1999) invoking the approximation theorems of Komlós, Major and Tusnády (1975) for the univariate empirical process. They were able to find a strong approximation of the PL-process $\alpha_{0n}(z)=n^{1/2}\left(\hat{F}_{0n}(z)-F_0(z)\right)$, by a two parameter Gaussian process at the almost sure rate of $O\left(\frac{\ln\ln n}{\sqrt{n}}\right)$. Based on this fact, they established the functional law of the iterated logarithm for the PL-process as shown in Theorem 1.1.6. However, more recently, Tse (2003) remarked that the KMT theorems are not directly applied for the LTRC model and that the claim given by Zhou and Yip (1999) cannot be true without severe restrictions. Therefore, under more general conditions, Tse (2003) proves the strong approximation of the PL-process by a two parameter Kiefer type process at the almost sure rate of $O\left((\ln n)^{3/2}/n^{1/8}\right)$, that even when it is not as fast as that claimed by Zhou and Yip (1999), it still allows to prove Theorem 1.1.6.

Let $\mu_{0n}(x,y)=\alpha_{0n}(x)-\alpha_{0n}(y)$ denote the oscillation modulus of the PL-process. Consider three sequences of non-increasing positive constants, $\{a_n\}$ and $\{c_n\}$, such that $na_n\uparrow$ and $nc_n\uparrow$.

Convergence results concerning the Lipschitz-$\frac{1}{2}$ oscillation modulus of the PL-process are given in the following theorems.

Theorem 1.1.7. *(Remark 2.1 in Zhou et al (2003))*
Assume that $\int_{a_{W_0}}^{b_0}\frac{\mathrm{d}W_{01}(x)}{B_0^3(x)}<\infty$*,* F_0 *has continuous derivative* f_0 *and* B_0 *is a* γ*-Hölder continuous function with* $\frac{1}{2}<\gamma<1$ *in a neighborhood of a fixed point* $x_0\in(a_{W_0},b_{W_0})$*.*

If a_n and c_n satisfy

(i) $\frac{na_n}{\ln\ln n} \to \infty$,

(ii) $\frac{\ln \frac{c_n}{a_n}}{\ln\ln n} \to k, 0 \leq k < \infty$,

then, it follows that

$$\limsup_{n\to\infty} \sup_{0\leq t\leq a_n} \sup_{0\leq x\leq c_n} \frac{|\mu_{0n}(x_0+x, x_0+x+t)|}{\sqrt{2a_n \ln\ln n}} = (k+1)^{\frac{1}{2}} \sqrt{\frac{f_0(x_0)(1-F_0(x_0))}{B_0(x_0)}} \quad \text{a.s.}$$

and

$$\limsup_{n\to\infty} \sup_{0\leq x\leq c_n} \frac{|\mu_{0n}(x_0+x, x_0+x+a_n)|}{\sqrt{2a_n \ln\ln n}} = (k+1)^{\frac{1}{2}} \sqrt{\frac{f_0(x_0)(1-F_0(x_0))}{B_0(x_0)}} \quad \text{a.s.}$$

Theorem 1.1.8. *(Remark 2.2 in Zhou et al (2003))*
Assume that $\int_{a_{W_0}}^{b_0} \frac{\mathrm{d}W_{01}(x)}{B_0^3(x)} < \infty$, *$F_0$ has continuous derivative f_0 and B_0 is a γ-Hölder continuous function with $\frac{1}{2} < \gamma < 1$ in a neighborhood of a fixed point $x_0 \in (a_{W_0}, b_{W_0})$. If a_n satisfies $\frac{na_n}{\ln\ln n} \to \infty$, then, it follows that*

$$\begin{aligned}\limsup_{n\to\infty} \sup_{0\leq t\leq a_n} \frac{|\mu_{0n}(x_0, x_0+t)|}{\sqrt{2a_n \ln\ln n}} &= \limsup_{n\to\infty} \frac{|\mu_{0n}(x_0, x_0+a_n)|}{\sqrt{2a_n \ln\ln n}} \\ &= \sqrt{\frac{f_0(x_0)(1-F_0(x_0))}{B_0(x_0)}} \quad \text{a.s.}\end{aligned}$$

In some biomedical studies or clinical trials, it may be interesting to predict the survival time of a patient given a vector of covariables such as age, cholesterol or glucose level in blood, etc. In this context of conditional survival analysis, different product limit estimators have been considered for the setting in which the lifetime is subject to left truncation and/or right censoring mechanisms. Beran (1981), González-Manteiga and Cadarso-Suárez (1994), Van Keilegom and Veraverbeke (1997), Dabrowska (1989) and Akritas (1994) considered a conditional product limit estimator for RC data and different contexts such as fixed designs with weights of Gasser-Müller type and random designs with Nadaraya-Watson or k-nearest neighbours weights. In a similar way, La Valley and Akritas (1994) considered a conditional product limit estimator for LT data and more recently, Iglesias-Pérez and González-Manteiga (1999) defined a product limit estimator to estimate the survival function conditional on some covariates when the lifetimes are subject to both censoring and truncation.

1.2 The bootstrap

The name 'bootstrap' refers to the analogy with pulling oneself up by one's own bootstraps. Efron (1979) introduces this statistical term to refer to a computer-intensive resampling method for estimating the variability of statistical quantities and obtaining confidence regions. Efron's bootstrap lies in to resample the data. More specifically, let us consider a sample of independent and identically distributed data with common distribution F_0, $\{X_{01}, \ldots, X_{0n}\}$, and let $\theta(F_0)$ and $\theta(F_{0n})$ be respectively an unknown population parameter and an estimate based on the observed sample. When the objective is to estimate the probability distribution of $\sqrt{n}\,(\theta(F_{0n}) - \theta(F_0))$, let say $G_n(t) = P\,(\sqrt{n}\,(\theta(F_{0n}) - \theta(F_0)) \leq t)$, Efron's nonparametric method approximates $G_n(t)$ by the probability distribution function, $G_n^*(t)$, of its bootstrap analogues,

$$\sqrt{n}\,(\theta(F_{0n}^*) - \theta(F_{0n}))\,,$$

where now F_{0n}^* denotes the empirical distribution function of the 'bootstrap resample', an artificial random sample, $\{X_{01}^*, \ldots, X_{0n}^*\}$, drawn from F_{0n} and

$$G_n^*(t) = P_n^*\left(\sqrt{n}\,(\theta(F_{0n}^*) - \theta(F_{0n})) \leq t\right),$$

where P_n^* refers to probability corresponding to F_{0n}, i.e. probability conditionally on the observations $\{X_{01}, \ldots, X_{0n}\}$.

Hall (1992) describes a physical analogue of the main principle of the bootstrap, considering a Russian nesting doll (also known as matryoshka), which is a nest of wooden figures with slightly different features painted on each. Specifically, he considers the problem of estimating the number of freckles on the face of the outer doll, let say doll 0, using the information given by the number of freckles appearing in the other dolls, let say, doll 1, doll 2 and so on, in decreasing order, from larger to smaller.

Let n_i be the number of freckles painted on doll i. Since the doll 1 is smaller in size than the doll 0, it is expected that n_0 will be larger than n_1, and therefore, n_1 would be an underestimate of n_0. However, it seems reasonable to think that the relationship between n_0 and n_1 will resemble the relationship existing between n_1 and n_2, and so $n_0/n_1 \approx n_1/n_2$ and therefore $\tilde{n}_0 = n_1^2/n_2$ could be a reasonable estimate of n_0.

If we use this same idea and go deeply into the dolls, the estimate of n_0 could be further refined.

Formally, in a mathematical formulation, what we wish to determine is the value of t, t_0, that solves what we will call 'the population equation'

$$E\left[f_t(F_0, F_{0n}) \diagup F_0\right] = 0, \tag{1.8}$$

where F_0 and F_{0n} denote, respectively, the population distribution function and the empirical distribution of the sample $\{X_{01}, \ldots, X_{0n}\}$ drawn from F_0. For example, let assume that the true parameter value is

$$\theta_0 = \theta(F_0) = \left\{ \int x \mathrm{d}F_0(x) \right\}^3,$$

i.e, the third power of the mean of F_0. The bootstrap estimate $\hat{\theta}$ of θ_0 is given by

$$\hat{\theta} = \theta(F_{0n}) = \left\{ \int x \mathrm{d}F_{0n}(x) \right\}^3.$$

If we now wish to correct the bootstrap estimate of θ for bias, we first need to solve the population equation (1.8), with $f_t(F_0, F_{0n}) = \theta(F_{0n}) - \theta(F_0) + t$ and to define the bias-corrected estimate of θ by $\hat{\theta} + t_0$.

In the case of the Russian nesting doll, (1.8) is given by

$$n_0 - tn_1 = 0.$$

As it was explained before, the approach used to estimate the root of this equation, was firstly to replace the pair (n_0, n_1) by (n_1, n_2), secondly to consider the solution $\hat{t}_0$ of the equation $n_1 - tn_2 = 0$, the sample analogous of the population equation, and finally to obtain the approximate of n_0 by $\hat{n}_0 = \hat{t}_0 n_1$ with $\hat{t}_0 = n_1/n_2$.

To obtain an approximate of the root of (1.8), the same argument used in the Russian nesting doll example, can be considered now. Therefore, the population equation is replaced by what we will call 'the sample equation'

$$E\left[f_t(F_{0n}, F_{0n}^*) \diagup F_{0n}\right] = 0,$$

where F_{0n}^* denotes the distribution function of a sample drawn from F_{0n}. Note that the sample equation is always completely known and therefore it may be solved either exactly or via Monte Carlo approximation.

The bootstrap principle is based on the idea that the solution of the population equation can be well approximated by the solution of the sample equation.

Coming back to the example of the bias-corrected estimate of the third power of the mean of F_0, it happens that the sample equation is given by

$$E\left[\theta(F_{0n}^*) - \theta(F_{0n}) + t \diagup F_{0n}\right] = 0,$$

whose solution is

$$t = \hat{t}_0 = \theta(F_{0n}) - E\left[\theta(F_{0n}^*) \diagup F_{0n}\right].$$

If μ denotes the mean of F_0, i.e., $\mu = \int x \mathrm{d}F_0(x)$, then $\theta(F_0) = \mu^3$. Let $\bar{X}_0 = n^{-1} \sum_{i=1}^n X_{0i}$ denote the sample mean of the sample drawn from F_0 and consider the nonparametric estimate of θ given by $\theta(F_{0n}) = \bar{X}^3$.

Simple algebra yields an explicit expression for $E\left[\theta(F_{0n})\diagup F_0\right]$ in terms of the mean, μ, the variance, $\sigma^2 = E\left[(X_{01}-\mu)^2\right]$, and the skewness, $\gamma = E\left[(X_{01}-\mu)^3\right]$ as given below

$$E\left[\theta(F_{0n})\diagup F_0\right] = \mu^3 + n^{-1}3\mu\sigma^2 + n^{-2}\gamma.$$

Consequently,

$$\hat{t}_0 = \bar{X_0}^3 - \bar{X_0}^3 + n^{-1}3\bar{X}_0\hat{\sigma}^2 + n^{-2}\hat{\gamma},$$

where

$$\begin{aligned}\hat{\sigma}^2 &= n^{-1}\sum_{i=1}^n \left(X_{0i}-\bar{X}_0\right)^2,\\ \hat{\gamma} &= n^{-1}\sum_{i=1}^n \left(X_{0i}-\bar{X}_0\right)^3,\end{aligned}$$

are, respectively, the sample variance and the sample skewness, and the bootstrap bias-reduced estimate of μ^3 is given by

$$\hat{\theta}_1 = \hat{\theta} + \hat{t}_0 = \bar{X_0}^3 + n^{-1}3\bar{X}_0\hat{\sigma}^2 + n^{-2}\hat{\gamma}.$$

Sometimes, things may be more complicated than in this case and the value of $\hat{t}_0$ may need to be obtained via Monte Carlo simulation. To exemplify how we should proceed in those cases, we will consider again the last example analysed. In that case, the value of $E\left[\theta(F_{0n}^*)\diagup F_{0n}\right]$ could be approximated by first taking B resamples independently from F_{0n}, let say $\left\{X_{01}^{*,b},\ldots,X_{0n}^{*,b}\right\}$, with $1 \le b \le B$. Then, for each resample b, we compute $\hat{\theta}_b^* = \theta(F_{0n}^{*b})$, where F_{0n}^{*b} denotes the distribution function of $X_0^{*,b}$ and we approximate $E\left[\theta(F_{0n}^*)\diagup F_{0n}\right]$ by the sample mean of $\hat{\theta}_b^*$.

Taking into account the fact that the larger the number of resamples B is taken, the more accurate the approximation will be, this procedure will be suitable for a large number of resamples B.

1.3 Curve estimation

Curve estimation refers to all the statistical methods that have been proposed in the literature to estimate different curves of interest such as density functions, hazard rates, survival functions or regression curves. This problem has been faced using different methodologies. The classical approach consists in fitting a parametric model based on the observed data. This approach, although straightforward, presents an important disadvantage when the parametric assumptions are wrong. A more flexible approach allows the data to speak for themselves in the sense that no parametric assumptions are considered. This approach is known in the literature with the name of nonparametric curve estimation.

1.3.1 The Parzen-Rosenblatt estimator

From here on, the interest will be pointed out in the kernel method, one of the most widely used nonparametric methods for estimating a density function. Its origins go back to the papers of Rosenblatt (1956) and Parzen (1962) and the basic idea of this method is to place locally some bumps at the observations and sum them to get an estimate of the density. The shape of these bumps is given by a so-called kernel function (K) while the length of the interval around the observations where the bump is located is given by a parameter (h), called smoothing parameter or simply bandwidth. Let $\{X_{01}, X_{02}, \ldots, X_{0n}\}$ be a sample of n iid observations coming from a population X_0 with density function f_0. The kernel estimator of f_0 is given by

$$\begin{aligned} \tilde{f}_{0h}(x) &= \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_{0i}}{h}\right) = \frac{1}{n}\sum_{i=1}^{n} K_h\left(x - X_{0i}\right) \\ &= \int K_h(x-v)\mathrm{d}F_{0n}(v) = (K_h * F_{0n})(x), \end{aligned} \tag{1.9}$$

where $K_h(x) = h^{-1}K\left(\frac{x}{h}\right)$, F_{0n} denotes the empirical distribution function of F_0 based on the sample $\{X_{01}, X_{02}, \ldots, X_{0n}\}$ and $*$ denotes convolution.

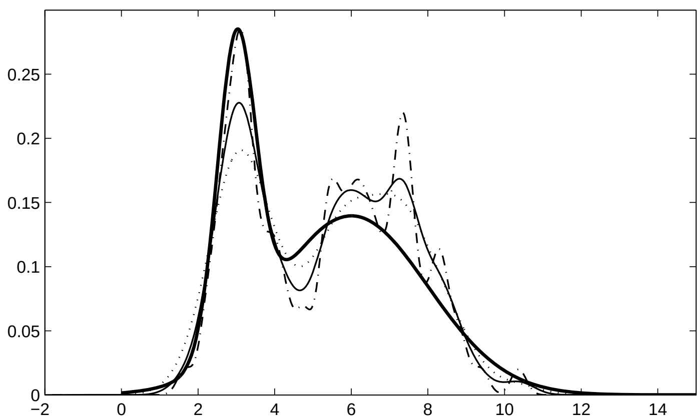

Figure 1.1: True density (thick solid line) and kernel type density estimates for model (1.10) using bandwidths of sizes $h = 0.6$ (dotted line), $h = 0.4$ (solid line) and $h = 0.2$ (dashed-dotted line).

It is always assumed that K integrates to one and usually that it is a symmetric probability density function. Table 1.3 collects a list of possible kernels K of this type. While the selection of K is not a crucial issue, however, the selection of the smoothing parameter must be appropriately addressed. The results can vary considerably depending on how much the data are smoothed. As h tends to zero, the bumps placed at the observations in order to construct $\tilde{f}_{0h}(x)$ will tend to Dirac delta functions that take value 1 at the

Table 1.3: Kernel functions.

Name	Density function $K(x)$
Gaussian	$(2\pi)^{-1/2}e^{-\frac{x^2}{2}}$
Cauchy	$(\pi(1+x^2))^{-1}$
Gamma(p)	$\Gamma(p)^{-1}x^{p-1}e^{-x}1_{\{x>0\}}$
Laplace	$\frac{1}{2}e^{-\lvert x\rvert}$
Uniform	$\frac{1}{2}1_{\{\lvert x\rvert<1\}}$
Epanechnikov	$\frac{3}{4}(1-x^2)1_{\{\lvert x\rvert<1\}}$
Biweight	$\frac{15}{16}(1-x^2)^2 1_{\{\lvert x\rvert<1\}}$
Triweight	$\frac{35}{32}(1-x^2)^3 1_{\{\lvert x\rvert<1\}}$
Triangular	$(1-\lvert x\rvert)1_{\{\lvert x\rvert<1\}}$
Logistic	$e^x/(e^x+1)^2$
Extrem value	$e^x e^{-e^x}$

observations. Consequently, the estimation of f_0 becomes very spiked and simply reproduces the observed data. However, the larger the value of h chosen, the smoother the curve becomes in the sense that all the information contained in the data is more and more obscured as h tends to infinity. This is illustrated in Figure 1.1 where, based on 100 data generated from a mixture X_0 of two Gaussian distributions, with density function given by

$$f_0(x) = 0.3N(3, 0.5^2) + 0.7N(6, 2^2), \tag{1.10}$$

three kernel type estimates, $\tilde{f}_{0h}(x)$, are plotted using different sizes of bandwidths ($h = 0.2, 0.4, 0.6$).

Below, some theoretical properties of $\tilde{f}_{0h}(x)$ are collected.

Theorem 1.3.1. *It follows that*

$$E\left[\tilde{f}_{0h}(x)\right] = \frac{1}{h}\int K\left(\frac{x-y}{h}\right) f_0(y)\mathrm{d}y, \tag{1.11}$$

$$Var\left[\tilde{f}_{0h}(x)\right] = \frac{1}{nh^2}\int K^2\left(\frac{x-y}{h}\right) f_0(y)\mathrm{d}y - \frac{1}{n}\left\{\frac{1}{h}\int K\left(\frac{x-y}{h}\right) f_0(y)\mathrm{d}y\right\}^2. \tag{1.12}$$

It seems reasonable to think that the discrepancy between the kernel estimate $\tilde{f}_{0h}$ based on a bandwidth h and the theoretical density f_0, will depend on h. Therefore, having a way to measure the discrepancy between both curves can help to address the question of how to select the bandwidth h. For example, trying to select the h minimizing the discrepancy between both curves. Different discrepancy measures have been proposed in the literature, either locally at a point x, such as the mean squared error,

$$MSE(\tilde{f}_{0h}(x)) = E\left[\left(\tilde{f}_{0h}(x) - f_0(x)\right)^2\right] = Bias^2\left[\tilde{f}_{0h}(x)\right] + Var\left[\tilde{f}_{0h}(x)\right], \tag{1.13}$$

or globally, such as the mean integrated squared error,

$$MISE(\tilde{f}_{0h}) = E\left[\int \left(\tilde{f}_{0h}(x) - f_0(x)\right)^2 \mathrm{d}x\right] = ISB(h) + IV(h), \qquad (1.14)$$

where

$$ISB(h) = \int Bias^2\left[\tilde{f}_{0h}(x)\right]\mathrm{d}x,$$
$$IV(h) = \int Var\left[\tilde{f}_{0h}(x)\right]\mathrm{d}x.$$

Using Theorem 1.3.1, it is easy to get exact expressions for the MSE and MISE. For example, when $f_0(x) = \phi_\sigma(x-\mu)$, a $N(\mu,\sigma^2)$ density, and $K(x) = \phi(x)$, the standard normal, it follows that:

$$Bias\left[\tilde{f}_{0h}(x)\right] = E\left[\tilde{f}_{0h}(x)\right] - f_0(x) = \phi_{\sqrt{\sigma^2+h^2}}(x-\mu) - \phi_\sigma(x-\mu)$$

and

$$Var\left[\tilde{f}_{0h}(x)\right] = \frac{1}{n}\left\{\frac{1}{2\sqrt{\pi}h}\phi_{\sqrt{\frac{1}{2}h^2+\sigma^2}}(x-\mu) - \phi^2_{\sqrt{\sigma^2+h^2}}(x-\mu)\right\}.$$

However, excepting for special selections of the kernel function and parametric assumptions of the true density f_0, the resulting expressions do not have any intuitive meaning. Therefore, the most practical way to proceed is to get asymptotic expressions for (1.11) and (1.12) under certain conditions and then to plug-in them in (1.13) and (1.14) to finally get asymptotic expressions for the MSE and MISE depending on h.

Let define $R(g) = \int g^2(x)\mathrm{d}x$, for every integrable squared function g and consider that f_0, K and h satisfy respectively condition (D1), (K1) and (B1) below:

(D1) f_0 has continuous derivatives up to order three, $f_0^{(2)}$ is square integrable and $f_0^{(3)}$ is bounded.

(K1) K is a symmetric and square integrable function, $\int K(t)\mathrm{d}t = 1$ and $d_K = \int t^2K(t)\mathrm{d}t$ is not null.

(B1) $\{h = h_n\}$ is a sequence of bandwidths satisfying that $h \to 0$ and $nh \to \infty$ as the sample size n tends to ∞.

Theorem 1.3.2. *Assuming conditions (B1), (D1) and (K1), it follows that*

$$Bias\left[\tilde{f}_{0h}(x)\right] = \frac{1}{2}h^2 f_0^{(2)}(x)d_K + o\left(h^2\right),$$

$$Var\left[\tilde{f}_{0h}(x)\right] = \frac{1}{nh}f_0(x)R(K) + o\left((nh)^{-1}\right).$$

Theorem 1.3.3. *Assuming conditions (B1), (D1) and (K1), it follows that*

$$ISB(h) \approx \frac{1}{4}h^4 {d_K}^2 R\left(f_0^{(2)}\right),$$

$$IV(h) \approx \frac{1}{nh}R(K).$$

Based on Theorems 1.3.2 and 1.3.3, it is straightforward to obtain asymptotic expressions for respectively, the MSE and MISE. Besides, they provide a theoretical justification of the practical behaviour shown previously in Figure 1.1. The AMSE and AMISE tell us that, when the smoothing parameter is selected in such a way that the bias is reduced, a marked increase in the random variation (or variance component) is added. On the other hand, when h is selected large enough to vanish the variance, then the absolute value of the bias increases. Therefore, the optimal local bandwidth to estimate f_0 at a given value x should be found as a compromise between the bias and the variance of $\tilde{f}_{0h}(x)$. Selecting h by minimizing the asymptotic MSE meets this compromise. If the objective, however, is to find an optimal global bandwidth to estimate f_0 at whatever point x, then, using a similar argument, a way to find that compromise is by selecting the value of h which minimizes, in this case, the asymptotic MISE. Using straightforward calculations it is easy to get expressions for the asymptotic optimal bandwidths, either locally or globally. They are shown in the following theorem.

Theorem 1.3.4. *Assuming conditions (B1), (D1) and (K1), it follows that the values of h minimizing AMSE and AMISE are respectively given by*

$$h_{AMSE} = \left(\frac{f_0(x)R(K)}{d_K^2 {f_0^{(2)}(x)}^2}\right)^{\frac{1}{5}} n^{-\frac{1}{5}}$$

and

$$h_{AMISE} = \left(\frac{R(K)}{d_K^2 R\left(f_0^{(2)}\right)}\right)^{\frac{1}{5}} n^{-\frac{1}{5}}. \tag{1.15}$$

So far we have seen that there are at least two ways of choosing the smoothing parameter, either locally, based on a local measure of discrepancy such as the MSE, or globally, based on a global error criterion such as the MISE. However, a straightforward computation of the two optimal bandwidths obtained in Theorem 1.3.4 can not be used in practical applications because they involve unknown values depending on the true density function f_0.

An overview of different global bandwidth selectors proposed in the literature is given below.

Cross validation selectors: There exist different methods that fall in this category. The pseudolikelihood crossvalidation method was proposed by Habbema *et al* (1974) and Duin (1976). Its basic idea consists in selecting h as the value that maximizes the pseudo-likelihood given by

$$\prod_{i=1}^{n} \tilde{f}_{0h}(X_{0i}). \tag{1.16}$$

However, since this function has a trivial maximum at $h = 0$, the kernel estimator $\tilde{f}_{0h}(x)$ in (1.16) is replaced by its leave-one-out modified version, $\tilde{f}_{0h}^{i}(x)$, when evaluated at X_{0i}

$$\tilde{f}_{0h}^{i}(x) = \frac{1}{(n-1)h} \sum_{j=1, j\neq i}^{n} K\left(\frac{x - X_{0j}}{h}\right). \tag{1.17}$$

This selector presents some nice properties. For example, it was shown that it minimizes the Kullback-Leibler distance between the true density and $\tilde{f}_{0h}$, and that it has a nice behaviour in L^1. Even so, it presents an important disadvantage: the lack of consistency when estimating heavy-tailed densities.

The biased crossvalidation method was proposed by Scott and Terrell (1987). Considering the asymptotic expression obtained for the MISE, they estimate the unknown quantity $R\left(f_0^{(2)}\right)$ by $R\left(\tilde{f}_{0h}^{(2)}\right)$ and derive a score function $BCV(h)$ to be minimized with respect to h. Based on simple algebra, it follows that

$$R\left(\tilde{f}_{0h}^{(2)}\right) = \frac{1}{nh^5} R\left(K^{(2)}\right) + \frac{1}{n^2h^5} \sum \sum\nolimits_{i\neq j} K^{(2)} * K^{(2)}(X_{0i} - X_{0j}).$$

Besides, if the optimal bandwidth is of order $n^{-\frac{1}{5}}$, Scott and Terrell (1987) show that $R\left(\tilde{f}_{0h}^{(2)}\right)$ yields a biased estimate of $R\left(f_0^{(2)}\right)$ since it holds that

$$E\left[R\left(\tilde{f}_{0h}^{2}\right)\right] = R\left(f_0^{(2)}\right) + n^{-1}h^{-5}R\left(K^{(2)}\right) + O\left(h^2\right).$$

Consequently, based on this fact, these authors propose to estimate $R\left(f_0^{(2)}\right)$ by $R\left(\tilde{f}_{0h}^{(2)}\right) - n^{-1}h^{-5}R\left(K^{(2)}\right)$ and find the value of h which minimizes the resulting bias crossvalidation function, BCV:

$$BCV(h) = \frac{R(K)}{nh} + h^4 \frac{d_K^2}{4n^2} \sum \sum_{i\neq j} K_h^{(2)} * K_h^{(2)}(X_{0i} - X_{0j}).$$

The main disadvantage of this method is the fact that, with commonly used kernels, it may happen that $\lim_{h\to 0^+} BCV(h) = \infty$ and $\lim_{h\to\infty} BCV(h) = 0$ and when $4R(K) - d_K^2 K(0) > 0$ (which holds for the Gaussian kernel), $BCV(h) > 0$ for all $h > 0$. Consequently, in situations like this, no global minimum exists (see Cao *et al* (1994)).

The smoothed crossvalidation method was proposed by Hall *et al* (1992). It is based on the exact expression of the MISE:

$$MISE(h) = IV(h) + ISB(h), \tag{1.18}$$

where

$$\begin{aligned} IV(h) &= \frac{R(K)}{nh} - \frac{1}{n}\int (K_h * f_0)^2(x)\mathrm{d}x, \\ ISB(h) &= \int (K_h * f_0 - f_0)^2(x)\mathrm{d}x. \end{aligned}$$

Marron and Wand (1992) showed that $\frac{R(K)}{nh}$ yields a good approximation of the first term in (1.18). Considering $\tilde{f}_{0g}$, another kernel estimate of f_0 based on a bandwidth g and a kernel L, and replacing f_0 in $ISB(h)$ by $\tilde{f}_{0g}$, an estimate of the second term appearing in (1.18), is obtained. Adding both estimates, $\widehat{IV}(h) + \widehat{ISB}(h)$, yields an estimate of $MISE(h)$, where

$$\widehat{IV}(h) = \frac{R(K)}{nh}$$

and

$$\widehat{ISB}(h) = \int \left(K_h * \tilde{f}_{0g} - \tilde{f}_{0g}\right)^2(x)\mathrm{d}x.$$

After doing some algebra and defining Δ_0 as the Dirac delta function for zero, it is possible to rewrite $\widehat{ISB}(h)$ as follows:

$$\widehat{ISB}(h) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i=1}^{n}(K_h * K_h - 2K_h + \Delta_0) * L_g * L_g(X_{0i} - X_{0j}).$$

Hall *et al* (1992) deleted in $\widehat{ISB}(h)$ the terms in the double sum for which $i = j$ and considered the approximation $n \approx n - 1$ to define the following score function that approximates $MISE(h)$:

$$\begin{aligned} SCV(h) &= \frac{R(K)}{nh} + \frac{1}{n(n-1)}\sum\sum_{i \neq j}(K_h * K_h - 2K_h + \Delta_0) \\ &\quad * L_g * L_g(X_{0i} - X_{0j}). \end{aligned} \tag{1.19}$$

The smoothed crossvalidation bandwidth is the value of h that minimizes (1.19). As Cao *et al* (1994) showed, this method can be motivated using bootstrap ideas.

Plug-in selectors: These selectors are all based on the asymptotic expression obtained for the optimal bandwidth that minimizes the MISE. As it was mentioned previously, the direct application of (1.15) is not plausible because the unknown quantity, $R\left(f_0^{(2)}\right)$, depends on the true density f_0.

Direct plug-in methods first assume a parametric model for f_0 such a Gaussian density and based on it an estimation of the unknown quantity is obtained and plugged in (1.15). Bandwidth selectors of this type are called in the literature rules of thumb.

More sophisticated plug in selectors consist in estimating the unknown functional using kernel-type estimators such as, for example, $R\left(\tilde{f}_{0g}^{(2)}\right)$. Under sufficiently smoothness conditions on f_0, it follows that $R\left(f_0^{(\ell)}\right) = (-1)^\ell \Psi_{2\ell}(f_0)$, where

$$\Psi_\ell(f_0) = \int f_0^{(\ell)}(x) f_0(x) \mathrm{d}x = E\left[f_0^{(\ell)}(X_0)\right]. \tag{1.20}$$

Due to the fact that $\Psi_\ell(f_0)$ can be expressed in terms of an expectation, a natural kernel-type estimator of it can be defined as follows:

$$\tilde{\Psi}_\ell\left(g; L\right) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n L_g^{(\ell)}\left(X_{0i} - X_{0j}\right).$$

Therefore, under sufficiently smoothness conditions on f_0, $R\left(f_0^{(2)}\right)$ can be estimated by $\tilde{\Psi}_4\left(g; L\right)$. Proceeding in this way, another bandwidth g, so called pilot bandwidth, has to be selected to estimate the unknown functional. In fact, after studying the asymptotic $MSE\left(\tilde{\Psi}_4\left(g; L\right)\right)$ it can be shown, for a second order kernel L satisfying some smoothness conditions, that the asymptotic optimal bandwidth has the following expression

$$g_{AMSE,\Psi_4(f_0)} = \left(\frac{2L^{(4)}(0)}{-d_L \Psi_6(f_0)}\right)^{\frac{1}{7}} n^{-\frac{1}{7}}.$$

However, once again, this expression can not be directly applied due to the unknown quantity depending on f_0, $\Psi_6(f_0)$, that appears in it, which needs to be estimated previously. There are at least two ways to proceed from here: using a parametric estimation for this unknown quantity or a kernel type estimation. The second one leads to another selection problem of a so called prepilot bandwidth. Different plug in selectors based on this strategy will differ in how many stages or successive kernel functional estimations are required till considering a parametric assumption on f_0 and on the way in which the needed (pre-)pilot bandwidths are selected as well.

Another strategy consists in choosing h as the root of the following equation

$$h = \left(\frac{R(K)}{d_K^2 \tilde{\Psi}_4\left(\gamma(h); L\right)}\right)^{\frac{1}{5}} n^{-\frac{1}{5}},$$

where the pilot bandwidth to estimate the functional $\Psi_4(f_0)$ is considered to be a function γ of h:

$$\gamma(h) = \left(\frac{2L^{(4)}(0)d_K^2}{R(K)d_L}\right)^{\frac{1}{7}} \left(-\frac{\Psi_4(f_0)}{\Psi_6(f_0)}\right)^{\frac{1}{7}} h^{\frac{5}{7}}. \tag{1.21}$$

Note that $\gamma(h)$ is naturally motivated by the relationship existing between the asymptotic optimal bandwidths h_{AMISE} and $g_{AMSE,\Psi_4(f_0)}$:

$$g_{AMSE,\Psi_4(f_0)} = \left(\frac{2L^{(4)}(0)d_K^2}{R(K)d_L}\right)^{\frac{1}{7}} \left(-\frac{\Psi_4(f_0)}{\Psi_6(f_0)}\right)^{\frac{1}{7}} h_{AMISE}^{\frac{5}{7}}.$$

As it is obvious, in order to apply (1.21), estimates of the functionals $\Psi_4(f_0)$ and $\Psi_6(f_0)$ are required. Either parametric or kernel type estimations are plausible here resulting in different solve-the-equation (STE) rules for selecting h. Likewise in most of sophisticated plug in selectors, plugging kernel type estimations of these unknown quantities in (1.21) leads to another stage selection problem. The novelty now is that, in the last step of this procedure aimed to select h, an equation depending on h must be solved using numerical methods.

Bootstrap selectors: Bootstrap-based choices of the bandwidth for kernel density estimators have been proposed by several authors (see, for instance, Taylor (1989), Faraway and Jhun (1990) and Cao (1993)). The basic idea behind all these proposals consists of estimating $MISE(h)$ using the bootstrap technique and then minimizing over h.

As Faraway and Jhun (1990) explain, the direct implementation would be to resample $\left\{X_{01}^{*,j}, \ldots, X_{0n}^{*,j}\right\}$ from the empirical distribution F_{0n} and then construct a large number, B, of bootstrap estimates as follows

$$\tilde{f}_{0h}^{*j}(x) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_{0i}^{*,j}}{h}\right) \quad \text{for } j = 1, \ldots, B.$$

Defining the bootstrapped approximations of the variance and the bias as follows,

$$Var^*\left[\tilde{f}_{0h}(x)\right] = B^{-1} \sum_{j=1}^{B} \left(\tilde{f}_{0h}^{*j}(x) - \bar{f}_{0h}^{*}(x)\right)^2$$

and

$$Bias^*\left[\tilde{f}_{0h}(x)\right] = \bar{f}_{0h}^{*}(x) - \tilde{f}_{0h}(x),$$

where

$$\bar{f}_{0h}^{*}(x) = B^{-1} \sum_{j=1}^{B} \tilde{f}_{0h}^{*j}(x),$$

it turns out that the estimate of the bias vanishes for every h. Since the bias increases with h, its contribution to the MISE can be substantial for some values of h. Therefore, this naive bootstrap selector fails in providing a good estimate of $MISE(h)$.

Faraway and Jhun (1990), being aware of this problem, consider an initial estimate of the density $\tilde{f}_{0g}(x)$ and resample $\left\{X_{01}^{*,j},\ldots,X_{0n}^{*,j}\right\}$ from it. Then, estimating $\tilde{f}_{0h}^{*j}(x)$ as before and defining the bootstrapped bias by $\bar{f}_{0h}^{*}(x)-\tilde{f}_{0g}(x)$, the modified bootstrap MISE is defined as follows:

$$\begin{aligned} MISE^*(h,g) &= B^{-1}\sum_{j=1}^{B}\int\left(\tilde{f}_{0h}^{*j}(x)-\bar{f}_{0h}^{*}(x)\right)^2\mathrm{d}x+\int\left(\bar{f}_{0h}^{*}(x)-\tilde{f}_{0g}(x)\right)^2\mathrm{d}x \\ &= B^{-1}\sum_{j=1}^{B}\int\left(\tilde{f}_{0h}^{*j}(x)-\tilde{f}_{0g}(x)\right)^2\mathrm{d}x. \end{aligned}$$

Taylor (1989) considers a similar strategy but with $h = g$. Apart from avoiding the selection problem of the pilot bandwidth g, Taylor (1989) realizes that the bootstrap resampling can be avoided when using a Gaussian kernel because there exists a closed expression for the bootstrap MISE

$$\begin{aligned} MISE^*(h) &= \frac{1}{2n^2h(2\pi)^{1/2}}\left[\sum_{i,j}\exp\left\{-\frac{(X_{0j}-X_{0i})^2}{8h^2}\right\}\right. \\ &\quad -\frac{4}{3^{1/2}}\sum_{i,j}\exp\left\{-\frac{(X_{0j}-X_{0i})^2}{6h^2}\right\} \\ &\quad \left.+2^{1/2}\sum_{i,j}\exp\left\{-\frac{(X_{0j}-X_{0i})^2}{4h^2}\right\}+n2^{1/2}\right]. \end{aligned}$$

However, as Cao *et al* (1994) point out, $MISE^*(h)$ tends to zero as h tends to infinity and consequently, it is an unsuitable estimator of $MISE(h)$ for smoothing purposes because it has no finite minimum. A different approach is given by Cao (1993). Like Faraway and Jhun (1990), he considers a pilot bandwidth g, different from h and like Taylor (1989) does, he finds a closed expression of the bootstrap MISE, which does not depend on the resampled values. The main results obtained in Cao (1993) can be summarized as follows:

(a) Based on Fubini's theorem the exact expression of the bootstrap MISE is

$$MISE^*(h,g) = IV^*(h,g) + ISB^*(h,g),$$

where

$$IV^*(h,g) = (nh)^{-1}d_K + n^{-1}\int\left(\int K(u)\tilde{f}_{0g}(x-hu)\mathrm{d}u\right)^2\mathrm{d}x$$

and

$$ISB^*(h,g) = \int \left(\int K(u)(\tilde{f}_{0g}(x-hu) - \tilde{f}_{0g}(x))\mathrm{d}u \right)^2 \mathrm{d}x.$$

(b) Asymptotic expressions obtained for the MISE and the bootstrap MISE under regularity conditions suggest that the pilot bandwidth g should be selected in order to minimize the MSE of $R\left(\tilde{f}_{0g}^{(2)}\right)$. It can be shown that the dominant part of the value g minimizing this MSE turns out to be of exact order $n^{-\frac{1}{7}}$. This selection of g leads to the following asymptotic bootstrap MISE:

$$\begin{aligned} MISE^*(h,g) &= 4^{-1}d_K^2 h^4 R\left(\tilde{f}_{0g}^{(2)}\right) + (nh)^{-1}R(K) - n^{-1}R\left(\tilde{f}_{0g}\right) \\ &\quad + O_P(n^{-1}h^2) + O_P(h^6) + O_P(n^{\frac{2}{7}}h^8). \end{aligned}$$

(c) Denoting by h_{MISE} and h_{MISE^*} the minimizers of $MISE(\tilde{f}_{0h})$ and $MISE^*(h,g)$, respectively, it holds that

$$\frac{h_{MISE^*} - h_{MISE}}{h_{MISE}} = O\left(n^{-\frac{5}{14}}\right)$$

and

$$n^{\frac{6}{5}} g^{\frac{9}{2}} c_1 (h_{MISE^*} - h_{MISE}) \to N(0,1) \quad \text{(weakly)},$$

where c_1 is a constant that depends on f_0 and K.

Since there are some terms inflating artificially the bias, Cao (1993) also defines a modified version of the bootstrap MISE, $MMISE^*(h,g)$, where the terms in the ISB^* with $i=j$ are removed. However, the minimizer of this modified bootstrap MISE, let say h_{MMISE^*}, exhibits a worse behaviour than the previously proposed, h_{MISE^*}, even when it is based on an improved estimator of the curvature of f_0. For example, it holds that

$$\frac{h_{MMISE^*} - h_{MISE}}{h_{MISE}} = O\left(n^{-\frac{4}{13}}\right)$$

and

$$n^{\frac{6}{5}} g^{\frac{9}{2}} c_1 (h_{MMISE^*} - h_{MISE}) \to N\left(\frac{3}{2}, 1\right) \quad \text{(weakly)}.$$

This surprising behaviour can be explained by the fact that the positive bias that arises when estimating the curvature of f_0, is compensated by a negative bias appearing when the asymptotic distribution of the selector is derived.

1.3.2 Smoothed empirical distribution function

The empirical distribution function, F_{0n}, is not a smooth estimation of F_0. Even when the steps of F_{0n} are small, it is known that a second order improvement can be achieved using the following kernel type estimate

$$\tilde{F}_{0h}(x) = \int \mathbb{K}\left(\frac{x-t}{h}\right) \mathrm{d}F_{0n}(t). \tag{1.22}$$

It was Nadaraya (1964) who first proposed this estimator and proved under mild conditions that $F_{0n}(x)$ and $\tilde{F}_{0h}(x)$ has asymptotically the same mean and variance. Although this estimate can be obtained by simply integrating $\tilde{f}_{0h}(x)$ (see 1.9), the bandwidth parameters that optimize global measures of the accuracy of $\tilde{F}_{0h}(x)$, such as the MISE, differ from those of $\tilde{f}_{0h}(x)$. Therefore, in the literature there exist several papers that study the optimal selection of h to estimate F_0 via $\tilde{F}_{0h}$. For instance, Polansky and Baker (2000) designed a multistage plug in bandwidth selector in this setting using the asymptotic expression of the MISE given by Azzalini (1981) and Jones (1990), under several conditions on the bandwidth sequence, $h = h_n$, the density f_0 and the kernel K. In the following, we first introduce the required conditions and then we summarize their result in Theorem 1.3.5.

Let consider the following condition:

(D2) The density f_0 is a continuous and differentiable function with a finite mean and with a continuous, bounded and square integrable first derivative, $f_0^{(1)}$.

Theorem 1.3.5. *Assume conditions (B1), (K1) and (D2). Then,*

$$MISE(\tilde{F}_{0h}) = n^{-1}\nu(F_0) - n^{-1}hD_K + \frac{1}{4}h^4 d_K^2 R(f_0^{(1)}) + o\left(n^{-1}h + h^4\right),$$

where

$$\nu(F_0) = \int_{-\infty}^{\infty} F_0(x)(1-F_0(x))\mathrm{d}x, \tag{1.23}$$

$$D_K = 2\int_{-\infty}^{\infty} xK(x)\mathbb{K}(x)\mathrm{d}x.$$

Remark 1.3.1. As in the setting of density estimation, the selection of K has little effect on $MISE(\tilde{F}_{0h})$ and, as Jones (1990) proved, a slight improvement can be achieved using a uniform density on $[-\sqrt{3}, \sqrt{3}]$.

Remark 1.3.2. From the asymptotic expression obtained for $MISE(\tilde{F}_{0h})$ in Theorem 1.3.5, it follows that the global bandwidth that minimizes $AMISE(\tilde{F}_{0h})$ is given by

$$h_{AMISE,F_0} = \left(\frac{D_K}{d_K^2 R\left(f_0^{(1)}\right)}\right)^{1/3} n^{-1/3}.$$

Based on the expression obtained for h_{AMISE,F_0}, Polansky and Baker (2000) proposed different plug-in bandwidth selectors that we detail below. The simplest bandwidth selector proposed by these authors relies on considering a parametric reference for f_0 and estimating the unknown functional, $R\left(f_0^{(1)}\right)$, appearing in h_{AMISE,F_0}, using that parametric reference. Using a Gaussian reference, it is easy to obtain that

$$R(f_0^{(1)}) = \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{2\pi\sigma^6} \exp\left\{-(x-\mu)^2/\sigma^2\right\} \mathrm{d}x = \frac{1}{4\sigma^3\pi^{1/2}}.$$

Therefore,

$$h_{AMISE,F_0} = \left(\frac{4\sigma^3 D_K \pi^{1/2}}{d_K^2}\right)^{1/3} n^{-1/3},$$

which suggests the following simple bandwidth selector, known in the literature as the rule of thumb

$$h_N = \left(\frac{4\hat{\sigma}^3 D_K \pi^{1/2}}{d_K^2}\right)^{1/3} n^{-1/3},$$

where $\hat{\sigma}$ is an estimator of the standard deviation, for example, the sample standard deviation, S, or as Silverman (1986) suggested, the minimum between S and $\widehat{IQR}/1.349$, where $\widehat{IQR}$ denotes the sample interquartile range, i.e. $\widehat{IQR} = F_{0n}^{-1}(0.75) - F_{0n}^{-1}(0.25)$, where F_{0n}^{-1} is the quantile function defined in (1.1). This last estimator is more suitable for non-normal densities.

In the special case $K = \phi$, where ϕ denotes the standard Gaussian density, it follows that $h_{AMISE,F_0} = 1.587\sigma n^{-1/3}$ and consequently $h_N = 1.587\hat{\sigma} n^{-1/3}$.

More sophisticated plug-in bandwidth selectors rely on the same idea presented previously in Subsection 1.3.1, when different plug-in bandwidth selectors for $\tilde{f}_{0h}(x)$ (see (1.9)) were introduced. Rather than using a parametric model for f_0 to estimate the unknown functional $R(f_0^{(1)})$, a nonparametric estimate is used. Since, under sufficiently smoothness conditions, it happens that $R(f_0^{(1)}) = \Psi_2(f_0)$ (see equation (1.20)), $R(f_0^{(1)})$ can be estimated by $\tilde{\Psi}_2(g_{AMSE,\Psi_2(f_0)}; L)$, where

$$g_{AMSE,\Psi_2(f_0)} = \left(\frac{2L^{(2)}(0)}{-d_L \Psi_4(f_0)}\right)^{1/5} n^{-1/5}$$

and depends on the unknown density function, f_0, through the functional $\Psi_4(f_0)$.

If we wish to use $\tilde{\Psi}_4(g_{AMSE,\Psi_4(f_0)}; L)$ as an estimate of $\Psi_4(f_0)$, we then need to propose a way to estimate the unknown quantity appearing in the expression of $g_{AMSE,\Psi_4(f_0)}$. If we again use the same strategy as for $g_{AMSE,\Psi_2(f_0)}$, we arrive to a never-ending process. Therefore, as Sheather and Jones (1991) proposed and it was already mentioned in Subsection 1.3.1, this iterative and never-ending process can be stopped at some stage by considering a parametric model for f_0 and by estimating the unknown functional required in that stage using that parametric model.

In case that a Gaussian reference with mean μ and variance σ^2 is considered, there exist simple expressions for the functionals $\Psi_\ell(f_0)$. Specifically, it follows that

$$\Psi_\ell(f_0) = \frac{(-1)^{\ell/2}\ell!}{(2\sigma)^{\ell+1}(\ell/2)!\pi^{1/2}},$$

which suggests the following parametric estimate for $\Psi_\ell(f_0)$:

$$\tilde{\Psi}_\ell^{GR} = \frac{(-1)^{\ell/2}\ell!}{(2\hat{\sigma})^{\ell+1}(\ell/2)!\pi^{1/2}}, \tag{1.24}$$

where $\hat{\sigma}$ was introduced above.

As Polansky and Baker (2000) suggest, this provides a b-stage plug in selector of h which algorithm follows the steps given below:

Step 1. Calculate $\tilde{\Psi}_{2b+2}^{GR}$ using (1.24).

Step 2. For $\ell = b, b-1, \ldots, 1$, calculate $\tilde{\Psi}_{2\ell}(g_{2\ell}; L)$ where

$$g_{2\ell} = \left(\frac{2L^{2\ell}(0)}{-d_L\tilde{\Psi}_{2\ell+2}(g_{2\ell+2}; L)}\right)^{1/(2\ell+3)} n^{-1/(2\ell+3)}$$

if $\ell < b$, and

$$g_{2\ell} = \left(\frac{2L^{2\ell}(0)}{-d_L\tilde{\Psi}_{2b+2}^{GR}}\right)^{1/(2\ell+3)} n^{-1/(2\ell+3)}$$

if $\ell = b$.

Step 3. Calculate the b-stage plug-in bandwidth selector by

$$\tilde{h}_b = \left(\frac{D_K}{d_K^2\tilde{\Psi}_2(g_2; L)}\right)^{1/3} n^{-1/3}.$$

1.3.3 Density function estimation under LTRC data

A natural extension of (1.9) to the case of LTRC data, it is given by

$$\hat{f}_{0h,n}(x) = \int K_h(x-t)\,\mathrm{d}\hat{F}_{0n}(t), \tag{1.25}$$

where the empirical distribution function in (1.9) is replaced by the TJW product limit estimator introduced in (1.5).

Using Theorem 1, Gijbels and Wang (1993) study the asymptotic behaviour of (1.25) under condition (B1) introduced previously and condition (K2) below:

(K2) For a positive integer $p \geq 1$, K is a p-order kernel function in $L^2[-1,1]$ of bounded variation with support in $[-1,1]$.

Later on, Sánchez-Sellero *et al* (1999) obtain an asymptotic expansion for the MISE of $\hat{f}_{0h,n}$ under (B1) and the following conditions:

(D3) The density function f_0 is six times differentiable and its sixth derivative is bounded. The first, second, third and fourth derivatives of f_0 are integrable, and the limits of f_0 and any of its first five derivatives at $-\infty$ or ∞ are zero.

(K3) K is a symmetric probability density, three times continuously differentiable, with first derivative integrable and satisfying that

$$\lim_{|x|\to\infty} x^j K^{(j)}(x) = 0, \quad j = 0,1,2,3.$$

(Q1) The function q_0 is twice continuously differentiable.

(W1) w is a weight function compactly supported by a set of points x satisfying $B_0(x) \geq \epsilon$ for some constant $\epsilon > 0$ and is three times continuously differentiable.

The results given by Gijbels and Wang (1993) and Sánchez-Sellero *et al* (1999) are collected below in Theorems 1.3.6 and 1.3.7.

Theorem 1.3.6. *(Consistency and asymptotic normality of (1.25))*
Assume (B1) and (K2). If $a_{G_0} < a_{W_0}$ and f_0 is a p times continuously differentiable function at z with $f_0(z) > 0$, for $a_{G_0} < z < b_{W_0}$, then,

(i) $\hat{f}_{0h,n}(z) = f_0(z) + \beta_{0n}(z) + \sigma_{0n}(z) + e_{0n}(z)$, *where*

$$\beta_{0n}(z) = h^{-1} \int F_0(z-hx)\mathrm{d}K(x) - f_0(z),$$

$$\sigma_{0n}(z) = (nh)^{-1} \sum_{i=1}^{n} \int \left[1 - F_0(z-hx)\right] \xi_{0i}\left(Y_{0i}, T_{0i}, \delta_{0i}, z-hx\right) \mathrm{d}K(x)$$

and $e_{0n}(z)$ *satisfies*

$$\sup_{0\leq z\leq b_0} |e_{0n}(z)| = O\left((\ln n)(nh)^{-1}\right) \quad \text{a.s.}$$

and

$$E \sup_{0\leq z\leq b_0} |e_{0n}(z)| = O\left((nh)^{-\alpha}\right), \ \textit{for any } \alpha > 0.$$

(ii)

$$Bias\left[\hat{f}_{0h,n}(z)\right] = h^p f_0^{(p)}(z) M_p + o\left(h^p\right) + O\left((nh)^{-1}\right)$$

and

$$Var\left[\hat{f}_{0h,n}(z)\right] = (nh)^{-1} f_0(z)\left[1 - F_0(z)\right]\left[B_0(z)\right]^{-1} R(K),$$

where $M_p = \frac{(-1)^p}{p!}\int x^p K(x)\mathrm{d}x$.

(iii) $\hat{f}_{0h,n}(z) \to f_0(z)$ a.s.

(iv) $(nh)^{\frac{1}{2}}\left\{\hat{f}_{0h,n}(z) - E\left[\hat{f}_{0h,n}(z)\right]\right\} \xrightarrow{d} N\left(0, \sigma_0^2(z)\right)$, *where*

$$\sigma_0^2(z) = f_0(z)[1 - F_0(z)][B_0(z)]^{-1} R(K).$$

(v) Let $d = \lim_{n\to\infty} nh^{2p+1}$. *If* $d < \infty$, *then,*

$$(nh)^{\frac{1}{2}}\left\{\hat{f}_{0h,n}(z) - f_0(z)\right\} \xrightarrow{d} N\left(d^{\frac{1}{2}} f_0^{(p)}(z) M_p, \sigma_0^2(z)\right),$$

where $\sigma_0^2(z)$ *has been defined in (iv) and* M_p *in (ii).*

Theorem 1.3.7. *(Asymptotic expression of* $MISE_w\left(\hat{f}_{0h,n}\right)$*: Theorem 2.1 in Sánchez-Sellero et al (1999))*
Assume conditions (B1), (D3), (K3), (Q1) and (W1). Then,

$$\begin{aligned} MISE_w\left(\hat{f}_{0h,n}\right) &= \frac{1}{4} d_K^2 h^4 \int {f_0^{(2)}(x)}^2 w(x)\mathrm{d}x + n^{-1}h^{-1}R(K) \\ & \int (1 - F_0(x))^2 B_0(x)^{-2} w(x)\mathrm{d}W_{01}(x) \\ & + n^{-1}\int (q_0(x) - B_0(x)^{-1}) f_0(x)^2 w(x)\mathrm{d}x \\ & + O(h^6) + O(n^{-1}h) + O((nh)^{\frac{-3}{2}}), \end{aligned}$$

where

$$MISE_w\left(\hat{f}_{0h,n}\right) = \int E\left[\left(\hat{f}_{0h,n}(t) - f_0(t)\right)^2\right] w(t)\mathrm{d}t.$$

Based on the asymptotic expression obtained for $MISE_w(\hat{f}_{0h,n})$, Sánchez-Sellero *et al* (1999) design two data driven bandwidth selectors based on plug-in ideas and the bootstrap technique.

When densities are used for classification purposes, the possibility of obtaining tight confidence bands can lead to a considerable reduction of the chance of misclassification. With this objective in mind, Sun and Zhou (1998) design a fully sequential procedure for constructing a fixed-width confidence band for f_0 in a given interval $[a, b]$ and such that it has approximately a coverage probability of $1 - \alpha$ and determines each $f_0(x)$ to within $\pm\epsilon$, where ϵ denotes the desired precision. After introducing the required conditions, we next collect in Theorems 1.3.8 and 1.3.9 the results given by Sun and Zhou (1998) .

Let assume $a_{G_0} \leq a_{W_0}$, $b_{G_0} \leq b_{W_0}$ and $\int_{a_{W_0}}^{\infty} \frac{\mathrm{d}F_0(x)}{G_0^2(x)} < \infty$. Besides, let us consider the following conditions on K, f_0 and the cdf's of T_0 and C_0.

(K4) K is a symmetric and continuously probability kernel with support $[-1,1]$.

(D4) The cdf's G_0 and L_0 have bounded first derivatives on $[a-\theta, b+\theta]$ for some $\theta > 0$. The density f_0 is continuous, bounded and bounded away from zero on $[a-\theta, b+\theta]$ with bounded second derivative on the same interval.

Theorem 1.3.8. *(Asymptotic distribution for the maximal deviation between $\hat{f}_{0h,n}$ and f_0: Theorem 1 in Sun and Zhou (1998))*
Assume conditions (K4) and (D4) and consider that $h = n^{-\gamma}$ for $\frac{1}{5} < \gamma < \frac{1}{2}$. Then, it follows that

$$P\left\{(2\gamma \ln n)^{\frac{1}{2}}\left[R(K)^{-\frac{1}{2}}E_n - e_n\right] < x\right\} \to \exp\left(-2\exp\left(-x\right)\right),$$

where

$$\begin{aligned} E_n &= (nh)^{\frac{1}{2}} \sup_{a\leq x\leq b}\left\{\left(\frac{B_0(x)}{(1-F_0(x))f_0(x)}\right)^{\frac{1}{2}}\left|\hat{f}_{0h,n}(x) - f_0(x)\right|\right\}, \\ e_n &= \left[2\ln\left(\frac{b-a}{n^{-\gamma}}\right)\right]^{\frac{1}{2}} + \left[2\ln\left(\frac{b-a}{n^{-\gamma}}\right)\right]^{-\frac{1}{2}}\left\{\ln\left(\frac{R\left(K^{(1)}\right)}{4\pi^2 R(K)}\right)^{\frac{1}{2}}\right\}. \end{aligned}$$

Based on Theorem 1.3.8, Sun and Zhou (1998) consider the following variable-width confidence band for $\hat{f}_{0h}$:

$$\hat{f}_{0h,n}(x) \pm \left(\frac{(1-\hat{F}_{0n}(x))\hat{f}_{0h,n}(x)R(K)}{n^{1-\gamma}B_{0n}(x)}\right)^{\frac{1}{2}}\left[\frac{z_\alpha}{(2\gamma\ln n)^{\frac{1}{2}}} + e_n\right],$$

where z_α satisfies $\exp\left(-2\exp(-z_\alpha)\right) = 1-\alpha$.

Using this confidence band, which asymptotically presents a coverage probability of $1-\alpha$, they consider N_ϵ below as the stopping rule,

$$N_\epsilon = \inf\left\{n \geq 1, \left[\sup_{a\leq x\leq b}\left(\frac{(1-\hat{F}_{0n}(x))\hat{f}_{0h,n}(x)}{n^{1-\gamma}B_{0n}(x)}\right)R(K)\right]^{\frac{1}{2}}\left(\frac{z_\alpha}{(2\gamma\ln n)^{\frac{1}{2}}} + e_n\right) \leq \epsilon\right\}$$

and define the following confidence bands

$$\hat{f}_{0h,N_\epsilon}(x) \pm \left(\frac{(1-\hat{F}_{0N_\epsilon}(x))\hat{f}_{0h,N_\epsilon}(x)R(K)}{N_\epsilon^{1-\gamma}B_{0N_\epsilon}(x)}\right)^{\frac{1}{2}}\left[\frac{z_\alpha}{(2\gamma\ln N_\epsilon)^{\frac{1}{2}}} + e_{N_\epsilon}\right] \qquad (1.26)$$

and

$$\hat{f}_{0h,N_\epsilon}(x) \pm \epsilon. \qquad (1.27)$$

The asymptotic properties of (1.26) and (1.27) are collected in Theorem 1.3.9 below.

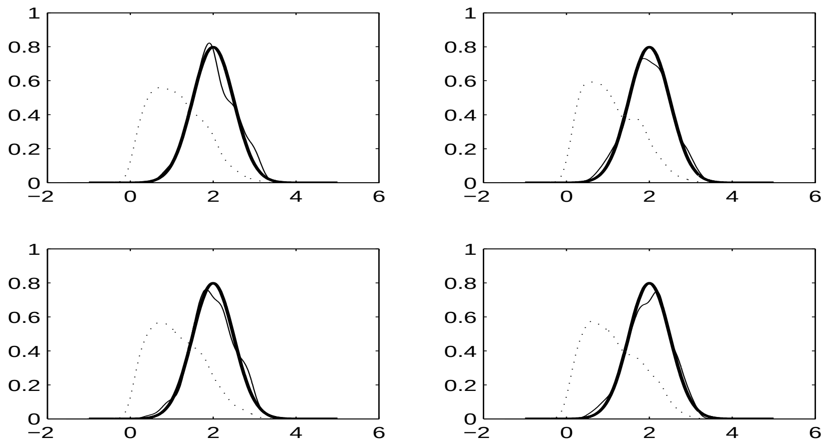

Figure 1.2: True density (thick solid line) and kernel type density estimates based on (1.9) (dotted line) and (1.25) (thin solid line) using a bandwidth of size $h = 0.15$ and 4 samples of size 1000 generated from scenario (i).

Theorem 1.3.9. *(Behaviour of confidence bands for $\tilde{f}_{0h}$: Theorem 2 in Sun and Zhou (1998))*
Assume conditions (K4) and (D4) and consider that $h = n^{-\gamma}$ for $\frac{1}{5} < \gamma < \frac{1}{2}$. Then, it follows that

$$P\left(f_0(x) \in \hat{f}_{0h,N_\epsilon}(x) \pm \left[\frac{\left(1 - \hat{F}_{0N_\epsilon}(x)\right) \hat{f}_{0h,N_\epsilon}(x) R(K)}{N_\epsilon^{1-\gamma} B_{0N_\epsilon}(x)} \right]^{\frac{1}{2}} \left(\frac{z_\alpha}{(2\gamma \ln N_\epsilon)^{\frac{1}{2}}} + e_{N_\epsilon} \right), \forall a \leq x \leq b \right) \to 1 - \alpha$$

as $\epsilon \to 0$ and

$$\liminf_{\epsilon \to 0} P\left\{ f_0(x) \in \hat{f}_{0h,N_\epsilon}(x) \pm \epsilon, \forall a \leq x \leq b \right\} \geq 1 - \alpha.$$

Next we illustrate through two examples the importance of taking into account the presence of censoring and truncation in the generation mechanism of the data when the objective is to estimate their density function. Let consider the following two settings:

(i) $f_0(t) = \frac{1}{0.5\sqrt{2\pi}} \exp\left\{-\frac{0.5(t-2)^2}{0.25}\right\}$, $L_0(t) = 1 - \exp\{-t\}$ and $G_0(t) = 1 - \exp\{-2t\}$,

(ii)$f_0(t) = \frac{1}{2.5\sqrt{2\pi}} \exp\left\{-\frac{0.5(t-3)^2}{6.25}\right\}$, $L_0(t) = 1-\exp\left\{-\frac{3t}{7}\right\}$ and $G_0(t) = 1-\exp\left\{-\frac{30t}{7}\right\}$.

For scenario (i), the objective is to estimate the normal density with mean 2 and standard deviation 0.5, denoted by f_0. However, for scenario (ii) the objective is to estimate the normal density with mean 3 and standard deviation 2.5, also denoted by f_0. While the first scenario presents a 78 percent of censoring and 34 percent of truncation, the second one only presents percentages of 68 and 21, respectively.

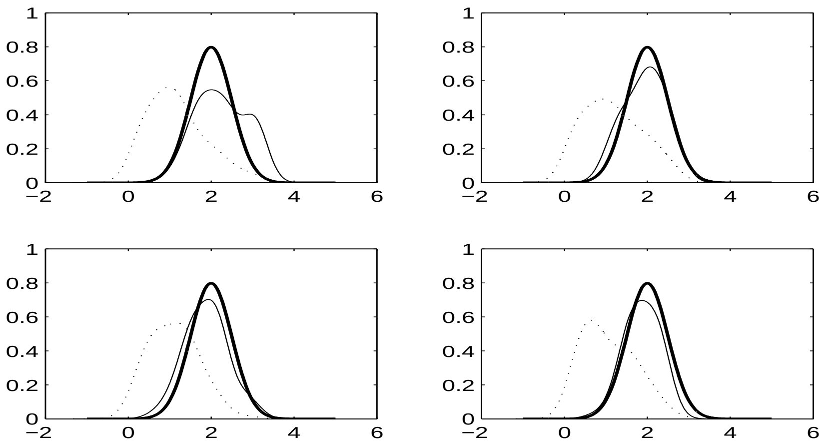

Figure 1.3: True density (thick solid line) and kernel type density estimates based on (1.9) (dotted line) and (1.25) (thin solid line) using a bandwidth of size $h = 0.30$ and 4 samples of size 100 generated from scenario (i).

Samples of 1000 and 100 data were obtained separately from both scenarios satisfying the condition $T_{0i} \leq Y_{0i}$. Using (1.9) and (1.25), kernel density estimates of f_0 were computed and plotted in Figures 1.2–1.5. From these figures, it is clearly observed the improvement that (1.25) yields over (1.9) when dealing with LTRC data.

1.3.4 Smoothed distribution function estimation under LTRC data

From the definition of $\hat{F}_{0n}(x)$, see (1.5), it is clear that it provides a stepwise estimate of F_0, and therefore it does not take into account the smoothness of F_0 when it is absolutely continuous. As in the case of complete data, in situations like this, it worths considering the following kernel type estimate of F_0

$$\hat{F}_{0h}(x) = \int \mathbb{K}\left(\frac{x-t}{h}\right) \mathrm{d}\hat{F}_{0n}(t). \tag{1.28}$$

It is interesting to note here that, as in the case of complete data, $\hat{F}_{0h}$ can now be obtained via integrating an appropriate kernel type estimate of f_0 (see 1.25) in the setting of LTRC data. The estimator $\hat{F}_{0h}$ was first proposed by Zhou and Yip (1999) who proved the law of iterated logarithm for it. However, little detail was given regarding their asymptotic properties. Briefly, it was stated that the asymptotic normality of $\hat{F}_{0h}$ could be proved in a similar way as in the case of the PL-estimator (see Corollary 2.1 in Zhou and Yip (1999)).

It was later on, when Chen and Wang (2006) studied the asymptotic properties of $\hat{F}_{0h}$ in more detail. More precisely, they obtained an asymptotic expression of the MISE,

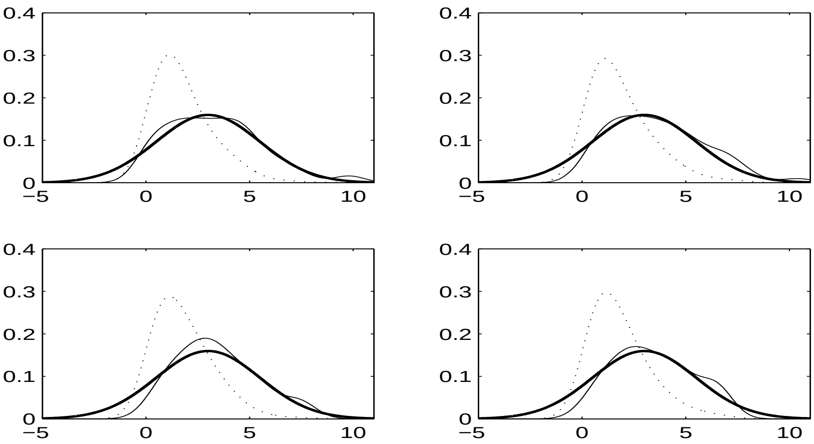

Figure 1.4: True density (thick solid line) and kernel type density estimates based on (1.9) (dotted line) and (1.25) (thin solid line) using a bandwidth of size $h = 0.75$ and 4 samples of size 1000 generated from scenario (ii).

proved the asymptotic normality of $\hat{F}_{0h}$ and proposed a plug-in bandwidth selector based on the asymptotically optimal smoothing parameter minimizing the AMISE.

Below, we first introduce some conditions that will be required for further discussion and we next collect the results proved by Chen and Wang (2006).

(K5) The kernel K is a symmetric probability density of bounded variation and with compact support $[-c, c]$.

(D5) $F_0^{-1}(p)$ has continuous derivatives at least to order $q + \ell$ at some $p = F_0(t)$, where $q = \min\left\{i : F_0^{-1(i)}(F_0(t)) \neq 0, i \geq 1\right\}$, $\ell = \min\left\{i : F_0^{-1(q+i)}(F_0(t)) \neq 0, i \geq 1\right\}$ and for a positive integer k, $F_0^{-1(k)}$ denotes the k^{th} derivative of the quantile function of X_0, F_0^{-1}.

Theorem 1.3.10. *(Theorem 2.1 in Chen and Wang (2006))*
Assuming conditions (B1), (K5) and (D5). If $a_{G_0} < a_{W_0}$, for $0 \leq x < b_{W_0}$, we have

$$\begin{aligned}
Bias(\hat{F}_{0h}(t)) &= -A_{q\ell}(t)h^{\frac{\ell+1}{q}} \int u^{\frac{\ell+1}{q}} K(u)\mathrm{d}u + o\left(h^{\frac{\ell+1}{q}}\right) + O\left(n^{-1}\right), \\
Var(\hat{F}_{0h}(t)) &= \frac{1}{n}\left(1 - F_0(t)\right)^2 q_0(t) - \frac{A_1(t)}{n}h \\
&\quad + \frac{2}{n}(1 - F_0(t))q_0(t)A_{q\ell}(t)h^{\frac{\ell+1}{q}} \int\int_{u\geq v} v^{\frac{\ell+1}{q}} K(v)\mathrm{d}v \\
&\quad + o\left(hn^{-1} + h^{\frac{\ell+1}{q}} n^{-1}\right),
\end{aligned}$$

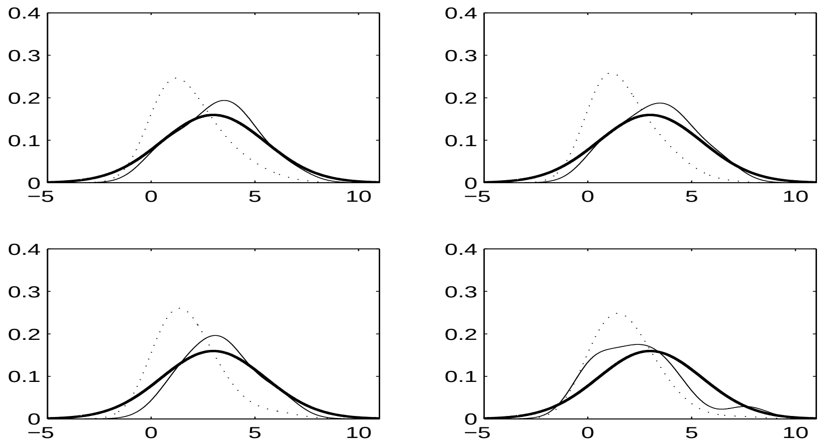

Figure 1.5: True density (thick solid line) and kernel type density estimates based on (1.9) (dotted line) and (1.25) (thin solid line) using a bandwidth of size $h = 1$ and 4 samples of size 100 generated from scenario (ii).

where $q_0(t)$ was introduced in (1.4) and

$$\begin{aligned} A_1(t) &= -2\frac{(1-F_0(t))f_0(t)}{B_0(t)}\int\int_{u\geq v} K(u)vK(v)\mathrm{d}u\mathrm{d}v \\ &= \frac{(1-F_0(t))f_0(t)}{B_0(t)} D_K > 0, \\ A_{q\ell}(t) &= \frac{q!^{\frac{q+\ell+1}{q}} F_0^{-1(q+\ell)}(F_0(t))}{q(q+\ell)!\left[F_0^{-1(q)}(F_0(t))\right]^{\frac{q+\ell+1}{q}}}. \end{aligned}$$

Corollary 1.3.11. *(Corollary 2.1 in Chen and Wang (2006))*
Under the assumptions of Theorem 1.3.10, it follows that

$$\begin{aligned} MSE(\hat{F}_{0h}(t)) &= \frac{1}{n}(1-F(t))^2 q_0(t) - \frac{A_1(t)h}{n} + \frac{B_{q\ell}(t)h^{\frac{\ell+1}{q}}}{n} \\ &\quad + D_{q\ell}(t)h^{\frac{2\ell+2}{q}} + o\left(hn^{-1} + h^{\frac{\ell+1}{q}}n^{-1} + h^{\frac{2\ell+2}{q}}\right), \end{aligned} \tag{1.29}$$

where

$$\begin{aligned} B_{q\ell}(t) &= 2(1-F_0(t))q_0(t)A_{q\ell}(t)\int u^{\frac{\ell+1}{q}}K(u)\mathrm{d}u, \\ D_{q\ell}(t) &= A_{q\ell}^2(t)\left(\int u^{\frac{\ell+1}{q}}K(u)\mathrm{d}u\right)^2. \end{aligned}$$

If $\ell+1$ is larger than q and $\int u^{\frac{\ell+1}{q}}K(u)\mathrm{d}u$ is not null, there exists an expression for the asymptotically optimal bandwidth, $h_{AMSE,F_0(t)}(t)$, which minimizes the sum of the second *and fourth terms in the right handside of (1.29):*

$$h_{AMSE,F_0(t)}(t) = \left(\frac{A_1(t)}{2(\ell+1)D_{q\ell}(t)}\right)^{\frac{q}{2(\ell+1)-q}} n^{-\frac{q}{2(\ell+1)-q}}.$$

Remark 1.3.3. When $f_0(t) \neq \infty$ and $f_0^{(1)}(t) \neq 0$, it is satisfied that

$$\begin{aligned} MSE(\hat{F}_{0h}(t)) &= \frac{1}{n}(1-F_0(t))^2 q_0(t) - \frac{A_1(t)h}{n} + A_{11}^2(t)d_K^2 h^4 + o\left(hn^{-1}+h^4\right) \\ &= \frac{1}{n}(1-F_0(t))^2 q_0(t) - \frac{(1-F_0(t))f_0(t)D_K h}{B_0(t)n} + \frac{1}{4}f_0^{(1)}(t)^2 d_K^2 h^4 \\ &\quad + o\left(hn^{-1}+h^4\right) \end{aligned}$$

and

$$h_{AMSE,F_0(t)}(t) = \left(\frac{(1-F_0(t))f_0(t)D_K}{B_0(t)f_0^{(1)}(t)^2 d_K^2}\right)^{1/3} n^{-1/3}.$$

Now, the consistency and the asymptotic normality of $\hat{F}_{0h}(t)$ as stated in Corollary 1.3.12 below, follows from Theorem 1.3.10 and Corollary 1.3.11.

Corollary 1.3.12. *(Corollary 2.2 in Chen and Wang (2006))*
Under the assumptions of Theorem 1.3.10. If $a_{G_0} < a_{W_0}$, for $0 \leq t < b_{W_0}$ it follows that

$$\hat{F}_{0h}(t) - F_0(t) \to 0 \quad \text{a.s. } as\ n \to \infty,$$

and, if $nh^{\frac{2\ell+2}{q}}$ goes to zero, then

$$n^{1/2}\left(\hat{F}_{0h}(t) - F_0(t)\right) \xrightarrow{d} N\left(0, \sigma_0^2(t)\right),$$

where $\sigma_0^2(t)$ was previously introduced in Theorem 1.1.6.

1.4 Two-sample problems

1.4.1 Nonparametric statistical tests

The study of differences among groups or changes over time is a goal in fields such as medical research and social science research. The traditional method for this purpose is the usual parametric location and scale analysis. However, this is a very restrictive tool, since a lot of the information available in the data is unaccessible. In order to make a better use of this information it is convenient to focus on distributional analysis, i.e., on the general two-sample problem of comparing the cumulative distribution functions (cdf), F_0 and F_1, of two random variables, X_0 and X_1 (see Gibbons and Chakraborti (1991) for a detailed review of the general two-sample problem).

Consider two independent random samples drawn independently from X_0 and X_1:

$$\{X_{01}, \ldots, X_{0n_0}\} \text{ and } \{X_{11}, \ldots, X_{1n_1}\}$$

and consider that the hypothesis of interest, H_0, is that they are drawn from identical populations, i.e. $H_0 : F_0(x) = F_1(x)$ for all x.

If this common population is Gaussian and if it is assumed that in case a difference between the populations exists, this is only due to a difference between means or only between variances, then the two-sample Student's t test for equality of means and the F test for equality of variances are respectively the best tests to use in these cases. Unfortunately, it is not always the case that these conditions are satisfied. Sometimes, for example, there is not enough information in the data to check the validity of these assumptions. In situations like this, when the validity of these assumptions is controversial, the best option is to use nonparametric procedures to test the null hypothesis H_0 for unspecified continuous distributions.

Under H_0, a single random sample of size $n = n_0+n_1$ can be considered as drawn from this common distribution. Therefore, the combined ordered configuration of the n_0 X_0's and n_1 X_1's random variables in the pooled sample, which is one of the $\binom{n_0+n_1}{n_0}$ possible equally likely arrangements, can provide information about the differences that may exist between the populations. In fact, most of the nonparametric tests designed in the literature in this setting are based on some function of the pooled ordered configuration where the type of function considered depends on the type of alternative the test is designed to detect.

There are different alternatives to H_0 that may be considered. The easiest to analyze using distribution-free techniques are those which state some functional relationship between the two distributions:

- A general one-sided alternative, which specifies that one of the random variables is stochastically larger than the other

$$\begin{aligned} H_A^1 : F_0(x) &\geq F_1(x) \text{ for all } x \\ F_0(x) &> F_1(x) \text{ for some } x. \end{aligned}$$

- A general two-sided alternative

$$H_A^2 : F_0(x) \neq F_1(x) \text{ for some } x.$$

- The location alternative, interesting to detect a difference in location

$$H_A^L : F_1(x) = F_0(x-\theta) \text{ for all } x \text{ and some } \theta \neq 0.$$

- The scale alternative, useful to detect a difference in scale

$$H_A^S : F_1(x) \quad = \quad F_0(\theta x) \text{ for all } x \text{ and some } \theta \neq 1.$$

- The Lehmann-type alternative which specifies that X_1 is distributed as the largest of k X_0 variables

$$H_A^L : F_1(x) \quad = \quad F_0^k(x) \text{ for all } x \text{ and some positive integer } k \neq 1.$$

There exists a large variety of nonparametric tests in the literature. Below we give some details on three of them: the Wald-Wolfowitz runs test, the Kolmogorov-Smirnov two-sample test and the Mann-Whitney U test.

The Wald-Wolfowitz runs test (Wald and Wolfowitz (1940)) is specially useful in a preliminary analysis when no particular form of alternative is yet formulated. Then, if the null hypothesis is rejected, further studies should be carried out with other tests with the objective of detecting which type of difference there exists between the two populations. Under H_0, it is expected that X_0 and X_1 will be well mixed in the pooled ordered configuration. Defining a run as a sequence of identical letters preceded and followed by a different letter or no letter, the total number of runs in the ordered pooled sample is indicative of the degree of mixing. If few runs are present, this fact would suggest that the pooled sample is made of two different samples drawn from distinguishable populations.

Defining R as the total number of runs, the test rejects H_0 when $R \leq c_\alpha$, such that c_α is the largest integer satisfying for an α level of significance that $P_{H_0}(R \leq c_\alpha) \leq \alpha$. There exists a simple expression for the null distribution of R,

$$f_R(r) = \begin{cases} \dfrac{\binom{n_0-1}{\frac{r}{2}-1}\binom{n_1-1}{\frac{r}{2}-1}}{\binom{n_0+n_1}{n_0}} & \text{if} \quad r \text{ is even}, \\ \dfrac{\binom{n_0-1}{\frac{r-1}{2}}\binom{n_1-1}{\frac{r-3}{2}}+\binom{n_0-1}{\frac{r-3}{2}}\binom{n_1-1}{\frac{r-1}{2}}}{\binom{n_0+n_1}{n_0}} & \text{if} \quad r \text{ is odd}, \end{cases} \tag{1.30}$$

for $r = 2, 3, \ldots, n$. However, when the n_0 and n_1 are large values, the computation of $f_R(r)$ becomes very laborious. So, for n_0 and n_1 both larger than 10 and assuming $\frac{n_0}{n} \to \kappa^2$ and $\frac{n_1}{n} \to 1 - \kappa^2$ where κ^2 is fixed $(0 < \kappa^2 < 1)$, it follows that

$$E\left(\frac{R}{n}\right) \quad \to \quad 2\kappa^2(1-\kappa^2)$$

$$Var\left(\frac{R}{n^{1/2}}\right) \quad \to \quad 4(\kappa^2)^2(1-\kappa^2)^2.$$

Defining the following standardized random variable

$$Z = \frac{R - 2n\kappa^2(1-\kappa^2)}{2n^{1/2}\kappa^2(1-\kappa^2)}$$

and denoting by f_Z its density function, Wald and Wolfowitz (1940) showed that

$$\ln f_Z(z) \to -\ln\sqrt{2\pi} - \frac{1}{2}z^2,$$

which proves that Z approaches a standard normal density as n tends to infinity. They first obtained the density of Z using (1.30) and then evaluated the factorials appearing in the resulting expression using the Stirling's formula (see Appendix A).

The Kolmogorov-Smirnov test (Smirnov (1939)) is in fact a one-sample test that can be adapted to the two-sample problem. While in the one sample case this statistic compares the empirical distribution function of a random sample with a hypothesized cumulative distribution function, in the two-sample case the comparison is carried out between the empirical distributions of the two samples, F_{0n_0} and F_{1n_1}.

The two-sided Kolmogorov-Smirnov two-sample test, D_{n_0,n_1}, is defined as follows:

$$D_{n_0,n_1} = \max_x \left| F_{0n_0}(x) - F_{1n_1}(x) \right|.$$

Since the order is preserved under a monotone transformation, this statistic is distribution-free for any continuous common distribution. There are several methods to compute the exact distribution of D_{n_0,n_1} under H_0. However, they are only appropriate when the sample sizes are small. When n_0 and n_1 approach infinity in such a way that n_0/n_1 remains constant, Smirnov (1939) proved that

$$\lim_{n_0,n_1\to\infty} P\left(\sqrt{\frac{n_0 n_1}{n}} D_{n_0,n_1} \le d\right) = L(d)$$

where

$$L(d) = 1 - 2\sum_{i=1}^{\infty} (-1)^{i-1} \exp\left(-2i^2 d^2\right).$$

As the Wald-Wolfowitz test, D_{n_0,n_1} should be used in preliminary studies because it is sensitive to all types of differences between the cumulative distributions functions.

The Mann-Whitney U test (Mann and Whitney (1947)) is based on the position of the X_{1j}'s in the combined ordered sequence. When most of the X_{1j}'s are larger than most of the X_{0i}'s or vice versa, this would be an evidence against a random mixing and then against the null hypothesis of identical distributions. Assuming that the possibility $X_{0i} = X_{1j}$ for some (i, j) does not need to be considered, since continuous distributions are assumed, the Mann-Whitney U test is defined as the number of times an X_{1j} precedes an X_{0i} in the combined ordered arrangement of the two samples in increasing order of magnitude.

Defining

$$D_{ij} = \begin{cases} 1, & \text{if} \quad X_{1j} < X_{0i}, \; i = 1, \ldots, n_0, \\ 0, & \text{if} \quad X_{1j} > X_{0i}, \; j = 1, \ldots, n_1, \end{cases}$$

the Mann-Whitney U statistic can be rewritten in terms of these indicator variables, D_{ij}, as follows:

$$U = \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} D_{ij}.$$

The null hypothesis can be reparameterized to $H_0 : p = 1/2$ where $p = P(X_1 < X_0) = \int_{-\infty}^{\infty} F_1(x) \mathrm{d}F_0(x)$. Based on the expressions of the mean and the variance of U,

$$E\left[\frac{U}{n_0 n_1}\right] = p,$$

$$Var\left(\frac{U}{n_0 n_1}\right) = \frac{1}{n_0 n_1}\left\{p - p^2(n-1) + (n_1 - 1)p_1 + (n_0 - 1)p_2\right\},$$

where

$$\begin{aligned} p_1 &= P(X_{1j} < X_{0i}, X_{1k} < X_{0i} \text{ with } j \neq k), \\ &= \int_{-\infty}^{\infty} F_1(x)^2 \mathrm{d}F_0(x), \\ p_2 &= P(X_{0i} > X_{1j}, X_{0h} > X_{1j} \text{ with } i \neq h), \\ &= \int_{-\infty}^{\infty} (1 - F_0(y))^2 \mathrm{d}F_1(y), \end{aligned}$$

it is straightforward to show that $\frac{U}{n_0 n_1}$ is a consistent estimator for p, since it is an unbiased estimator of p and $Var\left(\frac{U}{n_0 n_1}\right) \to 0$ as the sample sizes increase.

When the objective is to determine rejection regions of size α corresponding to the Mann-Whitmey test, we first need to find the null distribution of U. Under H_0, it follows that

$$P_{n_0,n_1}(u) = P(U = u) = \frac{r_{n_0,n_1}(u)}{\binom{n}{n_0}},$$

where $P_{n_0,n_1}(\cdot)$ denotes the probability mass function of U, $r_{n_0,n_1}(u)$ counts the number of distinguishable arrangements of the n_0 X_{0i}'s and the n_1 X_{1j}'s such that the number of times that an X_{1j} precedes an X_{0i} is u. A systematic method to generate critical values, defining the rejection region of U, is based on the following recurrence relationship:

$$r_{n_0,n_1}(u) = r_{n_0,n_1-1}(u) + r_{n_0-1,n_1}(u - n_1),$$

which implies that

$$(n_0 + n_1) P_{n_0,n_1}(u) = n_0 P_{n_0,n_1-1}(u) + n_1 P_{n_0-1,n_1}(u - n_1)$$

if the following initial and boundary conditions are defined:

$$\begin{aligned} r_{i,j}(u) &= 0 \text{ for all } u < 0 \\ r_{i,0}(0) &= r_{0,i}(0) = 1 \\ r_{i,0}(u) &= 0 \text{ for all } u \neq 0 \\ r_{0,i}(u) &= 0 \text{ for all } u \neq 0. \end{aligned}$$

When n_0 and n_1 are too large, the asymptotic probability distribution can be used instead of the exact null distribution of the test.

Based on a generalization of the Central Limit Theorem, it follows that U approaches to the standard normal as $n_0, n_1 \to \infty$ with $n_1/n_0 \to \kappa^2$. When H_0 is true, it follows that $p_1 = 1/3, p_2 = 1/3$ and $p = 1/2$, and therefore, it is satisfied that:

$$E\left[U|H_0\right] = n_0 n_1/2 \quad \text{and} \quad Var\left(U|H_0\right) = n_0 n_1 (n+1)/12.$$

Consequently, it follows that

$$Z = \frac{U - n_0 n_1/2}{\sqrt{n_0 n_1 (n+1)/12}}$$

can be approximated by a standard normal which gives reasonably accurate for $n_0 = n_1$ as small as 6.

Finally, it is worth mentioning here that the Mann Whitney test can be adapted to the case in which $X_{0i} = X_{1j}$ for some i and j. Besides, the Mann Whitney test is considered as the best nonparametric test for location by many statisticians.

1.4.2 ROC curves

A basic classification tool in medicine is the discrete classifier or the binary test, which yields discrete results, positive or negative, to infer whether a disease is present or absent. It is common practice to assess the accuracy of these tests using measures of sensitivity (SN) and specificity (SP), where

$$SN = \frac{TP}{TP + FN}$$

$$SP = \frac{TN}{TN + FP}$$

and TN, TP, FN and FP represent, respectively, the number of true negatives, true positives, false negatives and false positives, when the binary test is applied to a large population.

In contrast to binary tests, continuous tests or classifiers produce numeric values in a continuous scale. Considering that large values are indicative of a higher likelihood of disease, one individual can be classified as ill (positive) or healthy (negative) depending on if the value given by the continuous test exceeds or not the value given by a selected threshold. It is obvious that choosing a high threshold produces a low likelihood of a false positive result and a high likelihood of a false negative result, which is translated respectively in a high specificity and a low sensitivity. Therefore, while there are two values of specificity and sensitivity that summarize the accuracy of a binary classifier, this is not the case for continuous classifiers since a pair of values of sensitivity and specificity exists for every threshold considered in the continuous scale.

The receiver operating characteristic (ROC) curve captures in a unique graph the tradeoff between test's sensitivity and specificity along the whole range of a continuous test. Specifically, it plots SN vs $(1-SP)$ as the threshold varies. The ROC curve of a test with perfect accuracy, would run vertically from (0,0) to (0,1) and then horizontally from (0,1) to (1,1). A continuous test that does not perform better than a random guessing would run diagonally from (0,0) to (1,1). In practice, ROC curves lie between these two extremes. If the ROC curve of a test lies below the diagonal (0,0)-(1,1), in the lower right of the unit square, this means that the test is incorrect more often than correct and it could be improved reversing the labels of positive and negative.

Let X denote the diagnostic variable associated to a continuous test and let F_0 and F_1 denote the distribution of the diagnostic variable conditional on membership of two groups, P_0 and P_1. In the discussion above, P_0 and P_1 would be the healthy and the disease group, respectively. When F_0 is absolutely continuous, a formal definition of the corresponding ROC curve is a plot of $ROC(p) = 1 - F_1(F_0^{-1}(1-p))$ against p, where F_0^{-1} denotes the quantile function of F_0 introduced in (1.1). It is interesting to note here that $ROC(p)$ is nothing else than the distribution function of the random variable $1 - F_0(X_1)$, which is known in the literature as 'placement value' (see Cai (2004) and Pepe and Cai (2004)). In statistical terms, the ROC curve represents the non-null distribution of the p-value $(1 - F_0(X_1))$ for testing the null hypothesis that an individual comes from P_0.

Suppose that the diagnostic variable X is fully observed and that two independent samples $\{X_{01}, \ldots, X_{0n_0}\}$ and $\{X_{11}, \ldots, X_{1n_1}\}$ are drawn from respectively, P_0 and P_1, then a natural and simple way to estimate the ROC curve is obtained by replacing F_0 and F_1 in $ROC(p)$ by their corresponding empirical estimates, F_{0n_0} and F_{1n_1}, i.e.

$$ROC_{n_0,n_1}(p) = 1 - F_{1n_1}(F_{0n_0}^{-1}(1-p)).$$

In fact, this is equivalent to plot the empirical proportions $\#\{X_{1j} > \theta\}/n_1$ against $\#\{X_{0i} > \theta\}/n_0$ for different values of the threshold θ. Before going on, it is interesting to mention here the connection existing between this stepwise estimate of the ROC curve and the Mann-Whitney U statistic, introduced previously in Subsection 1.4.1. The area under the curve $ROC_{n_0,n_1}(p)$ is precisely $1 - \frac{U}{n_0 n_1}$, according to the notation used in Subsection 1.4.1.

A kernel type estimator of $ROC(p)$, first proposed by Lloyd (1998) in a heuristic way, is given by

$$\widehat{ROC}_{h_0,h_1}(p) = 1 - \tilde{F}_{1h_1}(\tilde{F}_{0h_0}^{-1}(1-p)),$$

where $\tilde{F}_{0h_0}$ and $\tilde{F}_{1h_1}$ denote kernel type estimates of respectively, F_0 and F_1 as defined in equation (1.22), with bandwidths h_0 and h_1.

Lloyd and Yong (1999) studied the asymptotic behaviour of these two estimators of $ROC(p)$ and showed that for any reasonable choice of the bandwidths, the kernel estimator performs better than the empirical one for moderate to large sample sizes.

Below we collect in Theorems 1.4.1 and 1.4.2 the results proved by Lloyd and Yong (1999). Let consider the following conditions:

(D6) The density functions, f_0 and f_1 are continuous at $x = F_0^{-1}(1-p)$.

(D7) The density f_0 is differentiable at $x = F_0^{-1}(1-p)$ and $f_0^{(1)}$ is continuous at $x = F_0^{-1}(1-p)$.

(D8) The density f_1 is differentiable at $x = F_0^{-1}(1-p)$ and $f_1^{(1)}$ is continuous at $x = F_0^{-1}(1-p)$.

(D9) The second moment $\int x^2 f_0(x)\mathrm{d}x$ is finite.

(K6) The kernel function $K(t)$ is symmetric about zero, twice differentiable with $K^{(2)}(t)$ continuous and bounded and it satisfies that $d_K = 1$.

Theorem 1.4.1. *(Theorem 1 in Lloyd and Yong (Lloyd and Yong) for a symmetric kernel)*
Assume conditions (D6)-(D8) and (K6). If $n_0, n_1 \to \infty$ and $h_0, h_1 \to 0$, we have that

$$\begin{aligned} Var\left(\widehat{ROC}_{h_0,h_1}(p)\right) &= \frac{ROC(p)(1-ROC(p))}{n_1} + \frac{(ROC^{(1)}(p))^2 p(1-p)}{n_0} \\ &\quad -D_K ROC^{(1)}(p)\left[\frac{h_0 f_1(F_0^{-1}(1-p))}{n_0} + \frac{h_1 f_0(F_0^{-1}(1-p))}{n_1}\right] \\ &\quad +O\left(\frac{h_0^2+h_1^2}{n_0}\right) + O\left(h_1^4 + n_0^{-2}\right), \end{aligned}$$

$$\begin{aligned} E\left[\widehat{ROC}_{h_0,h_1}(p)\right] &= ROC(p) + (h_0^2 - h_1^2) f_1^{(1)}(F_0^{-1}(1-p)) + ROC^{(2)}(p) \\ &\quad \left\{h_0^2 f_0^2(F_0^{-1}(1-p)) + \frac{p(1-p)}{n_0}\right\} \\ &\quad -\frac{1}{2} ROC^{(2)}(p) D_K \frac{f_0(F_0^{-1}(1-p))h_0}{n_0} + o\left(n_0^{-1} + h_0^2\right). \end{aligned}$$

Theorem 1.4.2. *(Theorem 2 in Lloyd and Yong (1999))*
Assume conditions (D6), (D8) and (D9). If $n_0, n_1 \to \infty$, we have that

$$\begin{aligned} Var\left(ROC_{n_0,n_1}(p)\right) &= \frac{ROC(p)(1-ROC(p))}{n_1} + (ROC^{(1)}(p))^2 \frac{p(1-p)}{n_0} + O\left(n_0^{-3/2}\right), \\ E\left[ROC_{n_0,n_1}(p)\right] &= ROC(p) + O\left(n_0^{-3/4}\right). \end{aligned}$$

For further discussion, we need to introduce new definitions. Let n be $n = \sqrt{n_0 n_1}$ and let κ_n^2 be the ratio between the two sample sizes n_1 and n_0, i.e. $\kappa_n^2 = n_1/n_0$. Similarly, let define $h = \sqrt{h_1 h_0}$ and $s_n^2 = h_1/h_0$.

Let k_n be the extra number of observations required for the empirical estimator to perform as well as the kernel one in terms of the MSE criterion, i.e. $MSE(ROC_{n_0,n_1}(p)) \leq MSE(\widehat{ROC}_{h_0,h_1}(p))$.

Furthermore, let us consider the following condition:

(S1) It is satisfied that $\kappa_n^2 = \frac{n_1}{n_0} \to \kappa^2 > 0$ and $s_n^2 = \frac{h_1}{h_0} \to s^2 > 0$.

Based on Theorem 1.4.3 below, it can be stated that the deficiency of the empirical estimator $ROC_{n_0,n_1}(p)$ is of order nh which diverges by assumption and includes as well the case of the asymptotically optimal bandwidth which is of order $O\left(n^{-1/3}\right)$.

Theorem 1.4.3. *(Theorem 3 in Lloyd and Yong (1999))*
Assume conditions (D6)-(D9), (K6) and (S1). Then, a necessary and sufficient condition for $\lim_{n\to\infty} k_n/n = 1$, *is that* $nh \to \infty$ *and* $nh^4 \to 0$. *In this case,*

$$\lim_{n\to\infty}\left\{\frac{k_n - n}{nh}\right\} = \frac{D_K ROC^{(1)}(p)(\kappa f_1\left(F_0^{-1}(1-p)\right)/s + s f_0\left(F_0^{-1}(1-p)\right)/\kappa)}{ROC(p)(1-ROC(p))/\kappa + \kappa(ROC^{(1)}(p))^2 p(1-p)} > 0.$$

Lloyd and Yong (1999) were the first who proposed an empirical method to select the bandwidths h_0 and h_1. Specifically, they considered the problem of selecting these bandwidths separately and they used for each one, the 2-stage plug-in selector detailed in Subsection 1.3.2.

Later on, Hall and Hyndman (2003) considered the problem of estimating the bandwidths h_0 and h_1 jointly rather than separately. When F_0 has much lighter tails than F_1, they realized that the error of an estimator of F_0 in its tail can yield a relatively large contribution to the error of the corresponding estimator of $ROC(p)$. Therefore, with the purpose of avoiding this problem, they considered a weighted global error criterion as follows

$$MISE_w(\widehat{ROC}_{h_0,h_1}) = \int E\left[(\widehat{ROC}_{h_0,h_1}(p) - ROC(p))^2\right] f_0(F_0^{-1}(p))\mathrm{d}p.$$

Since it can be proved that

$$MISE_w(\widehat{ROC}_{h_0,h_1}) \approx \int E\left[\tilde{F}_{0h_0}(t) - F_0(t)\right]^2 f_1^2(t)\mathrm{d}t + \int E\left[\tilde{F}_{1h_1}(t) - F_1(t)\right]^2 f_0^2(t)\mathrm{d}t,$$

the bandwidths h_0 and h_1 can indeed be obtained separately but using, respectively the error criteria:

$$\begin{aligned} MISE_{w_0}(\tilde{F}_{0h_0}) &= \int E\left[\tilde{F}_{0h_0}(t) - F_0(t)\right]^2 w_0(t)\mathrm{d}t, \\ MISE_{w_1}(\tilde{F}_{1h_1}) &= \int E\left[\tilde{F}_{1h_1}(t) - F_1(t)\right]^2 w_1(t)\mathrm{d}t, \end{aligned}$$

where $w_0(t) = f_1^2(t)$ and $w_1(t) = f_0^2(t)$. Based on asymptotic representations of these two error criteria, Hall and Hyndman (2003) proved that the optimal bandwidths have the following expressions:

$$h_0 = n_0^{-1/3} c(f_0, f_1) \text{ and } h_1 = n_1^{-1/3} c(f_1, f_0),$$

where

$$c(f_0, f_1)^3 = \left\{ \nu(\mathbb{K}) \int f_0(u) f_1^2(u) \mathrm{d}u \right\} / \left\{ d_K^2 \int \left[f_0^{(1)}(u) f_1(u) \right]^2 \mathrm{d}u \right\}$$

and $\nu(\mathbb{K})$ is the functional introduced previously in (1.23) but with F_0 replaced by $\mathbb{K}$. Note that, even when the constants differ, the order of these bandwidths coincides with that of the bandwidths previously used by Lloyd and Yong (1999) in a more heuristic way.

Based on the expressions obtained for the optimal bandwidths h_0 and h_1, Hall and Hyndman (2003) proposed two bandwidth selectors in this scenario. The first and simplest one is the classical rule of thumb, where f_0 and f_1 in expressions $c(f_0, f_1)$ and $c(f_1, f_0)$ are replaced by Gaussian densities. Their mean and scale are given respectively by the sample mean and by the minimum between the sample interquartile range divided by 1.349 and the sample standard deviation of respectively, the comparison sample $\{X_{01}, \ldots, X_{0n_0}\}$ and the reference sample $\{X_{11}, \ldots, X_{1n_1}\}$. Simple algebra gives

$$\begin{aligned} c(f_0, f_1) &= \left(\frac{4\sqrt{\pi}\nu(\mathbb{K})}{d_K^2} \frac{\sigma_0^3 (\sigma_0^2 + \sigma_1^2)^{5/2}}{(\sigma_1^2 + 2\sigma_0^2)^{1/2} \left[\sigma_1^4 + \sigma_1^2 \sigma_0^2 + 2\sigma_0^2 (\mu_0 - \mu_1)^2 \right]} \right)^{1/3} \\ &\quad \cdot \left(\exp\left[\frac{(\mu_0 - \mu_1)^2 \sigma_0^2}{(\sigma_0^2 + \sigma_1^2)(2\sigma_0^2 + \sigma_1^2)} \right] \right)^{1/3}. \end{aligned}$$

Consequently, the bandwidth selectors of h_0 and h_1 based on Gaussian densities are given by:

$$\hat{h}_0^{GR} = n_0^{-1/3} \hat{c}^{GR}(f_0, f_1) \text{ and } \hat{h}_1^{GR} = n_1^{-1/3} \hat{c}^{GR}(f_1, f_0),$$

where

$$\begin{aligned} \hat{c}^{GR}(f_0, f_1) &= \left(\frac{4\sqrt{\pi}\nu(\mathbb{K})}{d_K^2} \frac{\hat{\sigma}_0^3 (\hat{\sigma}_0^2 + \hat{\sigma}_1^2)^{5/2}}{(\hat{\sigma}_1^2 + 2\hat{\sigma}_0^2)^{1/2} \left[\hat{\sigma}_1^4 + \hat{\sigma}_1^2 \hat{\sigma}_0^2 + 2\hat{\sigma}_0^2 (\hat{\mu}_0 - \hat{\mu}_1)^2 \right]} \right)^{1/3} \\ &\quad \cdot \left(\exp\left[\frac{(\hat{\mu}_0 - \hat{\mu}_1)^2 \hat{\sigma}_0^2}{(\hat{\sigma}_0^2 + \hat{\sigma}_1^2)(2\hat{\sigma}_0^2 + \hat{\sigma}_1^2)} \right] \right)^{1/3}, \end{aligned}$$

with $\hat{\sigma}_k = \min\left\{ \widehat{IQR}_k / 1.349, S_k \right\}$, $S_k^2 = \frac{1}{n_k - 1} \sum_{\ell=1}^{n_k} (X_{k\ell} - \hat{\mu}_k)^2$ and $\hat{\mu}_k = n_k^{-1} \sum_{\ell=1}^{n_k} X_{k\ell}$ for $k = 0, 1$.

The second bandwidth selector proposed by Hall and Hyndman (2003) for h_0 and h_1 uses plug-in ideas in more steps and its algorithm is summarized below:

Step 1. Compute the following quantities:

$$\begin{aligned}
\hat{d} &= \frac{n_0^{1/3}\hat{\sigma}_1}{n_1^{1/3}\hat{\sigma}_0},\\
\hat{J}(f_0^{GR}, f_1^{GR}) &= 64\int (f_0^{GR(1)}(u))^2 f_0^{GR}(u)(f_1^{GR}(u))^3 \mathrm{d}u,\\
\hat{I}(f_0^{GR(1)}, f_1^{GR}) &= \int f_0^{GR(1)}(u) f_0^{GR(3)}(u)(f_1^{GR}(u))^2 \mathrm{d}u,\\
\hat{I}(f_1^{GR}, f_0^{GR(1)}) &= \int f_1^{GR}(u) f_1^{GR(2)}(u)(f_0^{GR(1)}(u))^2 \mathrm{d}u,
\end{aligned}$$

and define the bandwidths g_0 and g_2 as follows:

$$\begin{aligned}
g_0 &= \left(\frac{3\hat{J}(f_0^{GR}, f_1^{GR})}{\sqrt{2\pi}}(1+2\hat{d}^2)^{3/2}\left[2\hat{d}^2\hat{I}(f_0^{GR(1)}, f_1^{GR}) + \hat{I}(f_1^{GR}, f_0^{GR(1)})\right]^2\right)^{1/7}\\
&\quad \cdot (n_0 n_1)^{-1/7},\\
g_2 &= \hat{d} g_0.
\end{aligned}$$

Step 2. Estimate $I_1 = \int f_0(u) f_1^3(u)\mathrm{d}u$, $I_2 = \int f_0^2(u) f_1^2(u)\mathrm{d}u$ and $I_3 = \int f_0(u) f_1(u) f_1^{(2)}(u)\mathrm{d}u$ using Gaussian references as follows:

$$\begin{aligned}
\hat{I}_1 &= \int f_0^{GR}(u)(f_1^{GR}(u))^3 \mathrm{d}u,\\
\hat{I}_2 &= \int (f_0^{GR}(u))^2 (f_1^{GR}(u))^2 \mathrm{d}u,\\
\hat{I}_3 &= \int f_0^{GR}(u) f_1^{GR}(u)(f_1^{GR}(u))^{(2)} \mathrm{d}u,
\end{aligned}$$

and compute:

$$g_1 = n_1^{-2/5}\left(\frac{2\frac{n_1}{n_0}R(K)\hat{I}_1 + \rho(K)\hat{I}_2}{2d_K^2(\hat{I}_3)^2}\right)^{1/5},$$

where

$$\rho(K) = \int\left\{\int K(u)K(u+v)\mathrm{d}u\right\}^2 \mathrm{d}v.$$

Step 3. Compute $\hat{c}(f_0, f_1)^3$ using the formula:

$$\hat{c}(f_0, f_1)^3 = \left(\frac{\nu(\mathbb{K})}{d_K^2}\right)\frac{n_0^{-1}\sum_{1\le i\le n_0}\tilde{f}^2_{1g_1}(X_{0i})}{n_1^{-1}\left|\sum_{1\le j\le n_1}(\tilde{f}^{(1)}_{0g_2}(X_{1j}))^2\tilde{f}^j_{1g_0}(X_{1j})\right|},$$

where $\tilde{f}^j_{1g_0}$ denotes the leave-one-out estimator of f_1 introduced in (1.17) and

$$\begin{aligned}
\tilde{f}^2_{1g_1}(y) &= \frac{2}{n_1(n_1-1)g_1^2}\sum\sum_{1\le j_1<j_2\le n_1} K\left(\frac{y-X_{1j_1}}{g_1}\right)K\left(\frac{y-X_{1j_2}}{g_1}\right),\\
(\tilde{f}^{(1)}_{0g_2}(y))^2 &= \frac{2}{n_0(n_0-1)g_2^4}\sum\sum_{1\le i_1<i_2\le n_0} K^{(1)}\left(\frac{y-X_{0i_1}}{g_2}\right)K^{(1)}\left(\frac{y-X_{0i_2}}{g_2}\right).
\end{aligned}$$

Step 4. Analogously, $\hat{c}(f_1, f_0)^3$ is obtained following the previous Steps 1-3 but interchanging the roles of f_0 and f_1, $\{X_{01}, \ldots, X_{0n_0}\}$ and $\{X_{11}, \ldots, X_{1n_1}\}$, n_0 and n_1 and $\hat{\sigma}_0$ and $\hat{\sigma}_1$ in all the formulas.

Step 5. Finally, compute the bandwidths $\hat{h}_0$ and $\hat{h}_1$ by

$$\hat{h}_0 = n_0^{-1/3}\hat{c}(f_0, f_1),$$
$$\hat{h}_1 = n_1^{-1/3}\hat{c}(f_1, f_0).$$

Although it was not explicitly mentioned above, it is clear that in the previous steps we have used the notation $f_k^{GR}(x)$ (with $k = 1, 2$) to refer a Gaussian density with mean $\hat{\mu}_k$ and variance $\hat{\sigma}_k^2$ and for any positive integer ℓ, $f_k^{GR(\ell)}$ denotes the ℓ^{th} derivative of f_k^{GR}.

1.4.3 Relative curves

Useful tools for performing distributional comparisons are the relative distribution function, also called the two-sample vertical quantile comparison function, $R(t)$, and the relative density function or grade density, $r(t)$, of X_1 with respect to (wrt) X_0:

$$R(t) = P(F_0(X_1) \leq t) = F_1\left(F_0^{-1}(t)\right), \qquad 0 < t < 1,$$

where

$$r(t) = R^{(1)}(t) = \frac{f_1\left(F_0^{-1}(t)\right)}{f_0\left(F_0^{-1}(t)\right)}, \qquad 0 < t < 1.$$

These two curves, as well as estimators for them, have been studied by Gastwirth (1968), Ćwik and Mielniczuk (1993), Hsieh (1995), Handcock and Janssen (1996, 2002), Hsieh and Turnbull (1996) and Cao *et al* (2000, 2001) under different scenarios. Assuming that F_0 is continuous and F_1 is absolutely continuous with respect to F_0, the relative density, r, exists and satisfies the equation above.

These functions are closely related to other statistical methods. For instance, the ROC curve, used in the evaluation of the performance of medical tests for separating two groups and that was introduced in Subsection 1.4.2, is related to R through the relationship $ROC(t) = 1 - R(1-t)$ (see, for instance, Holmgren (1995) and Li *et al* (1996) for details) and the density ratio $\lambda(x) = \frac{f_1(x)}{f_0(x)}, x \in \mathbb{R}$, used by Silverman (1978) is linked to r through $\lambda(x) = r(F_0(x))$. The Kullback-Leibler divergence, D_{KL}, is defined in terms of the density ratio as follows:

$$D_{KL}(F_1, F_0) = \int_{-\infty}^{\infty} \ln\left(\frac{f_1(x)}{f_0(x)}\right) \mathrm{d}F_1(x).$$

Using a change of variable, it can be easily rewritten in terms of the Shannon entropy of the relative density:

$$D_{KL}(F_1, F_0) = \int_0^1 \ln(r(t))r(t)\mathrm{d}t.$$

Given the close relationship existing between the ROC curve and the relative distribution, all the estimators introduced for $ROC(p)$ in Subsection 1.4.2 are now applicable to the relative distribution with minor modifications, when both samples are completely observed. Specifically, $ROC_{n_0,n_1}(p)$ suggests to estimate $R(t)$ empirically by:

$$R_{n_0,n_1}(t) = 1 - ROC_{n_0,n_1}(1-t) = F_{1n_1}\left(F_{0n_0}^{-1}(t)\right), \tag{1.31}$$

and $\widehat{ROC}_{h_0,h_1}(p)$ suggests to estimate $R(t)$ by the smooth estimator:

$$\hat{R}_{h_0,h_1}(t) = 1 - \widehat{ROC}_{h_0,h_1}(1-t) = \tilde{F}_{1h_1}\left(\tilde{F}_{0h_0}^{-1}(t)\right).$$

Relative density estimates can provide more detailed information about the performance of a diagnostic test which can be useful not only in comparing different tests but also in designing an improved one. To illustrate this idea, we consider here two normal populations, X_0, with mean 0 and variance 1, and X_1, with mean 0.5 and variance 0.0625. The densities of these two populations are displayed in the left-top panel of Figure 1.6. Instead of considering a threshold that moves along the values of the interest variable, X, as in the case of a ROC curve, we consider now, for every value ℓ for the relative density, the two regions $\{t \in [0,1] : r(t) \geq \ell\}$ and $\{t \in [0,1] : r(t) < \ell\}$, which need not be two intervals now. Let us consider for this example, $\ell = 1$, then there exist two values, $c_1 = F_0^{-1}(t_1) = 0.0842$ and $c_2 = F_0^{-1}(t_2) = 0.9815$, with $t_1 = 0.5335$ and $t_2 = 0.8368$, for which $r(t_i) = \ell = 1$, for $i = 1, 2$. Considering that X_0 measures the value of interest, X, in the healthy population and X_1 in the ill population, we now can define a diagnostic test that classifies an individual in the healthy population if the value of X measured for this person does not fall inside the interval (c_1, c_2). Otherwise, the individual is classified in the ill population. Proceeding in this way, we have computed the curve of sensitivity and specificity that result as the level ℓ moves between the maximum and the minimum values attained by $r(t)$. Analogously to the construction of the ROC curve, we have plotted the obtained pairs $((1-SP)$,$SN)$ in the left-bottom panel of Figure 1.6. Comparing this curve, let us say $S(t)$ with the ROC curve, $ROC(t)$ (as plotted in the left-bottom panel of Figure 1.6), it is clear that the new diagnostic test performs better than the classical one, because $ROC(t)$ is closer to the diagonal in the unit square in $(0, 1)$.

There are different global measures that are used in the literature to compare medical tests for separating two groups. The area under the ROC curve (AUC) or the proportion of similar responses (PSR), also known as the overlapping coefficient, are two of them.

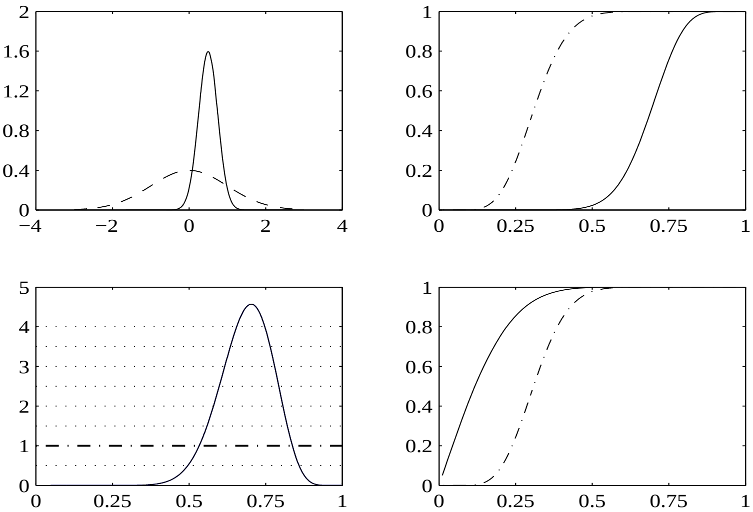

Figure 1.6: Plot of a diagnostic test based on the relative density. The density of X_0 (dashed line) and X_1 (solid line) are plotted in the left-top panel. The relative distribution of X_1 wrt X_0 (solid line) and the corresponding ROC curve (dashed-dotted line) are plotted in the right-top panel. The relative density of X_1 wrt X_0 (solid line) is plotted in the left-bottom panel and the ROC curve (dashed-dotted line) is jointly plotted in the right-bottom panel with the new curve $S(t)$ (solid line).

The overlapping coefficient is defined as:

$$PSR(f_0, f_1) = \int \min\left(f_0(x), f_1(x)\right)\mathrm{d}x,$$

where f_0 and f_1 denote the density functions of two populations, X_0 and X_1. A graphical representation of $PSR(f_0, f_1)$ is displayed in Figure 1.7.

There exists a close relationship between $PSR(f_0, f_1)$ and the dissimilarity index or area between the curves, $ABC(f_0, f_1)$:

$$ABC(f_0, f_1) = \int |f_0(x) - f_1(x)|\, \mathrm{d}x.$$

In fact, it can be easily proved that:

$$PSR(f_0, f_1) = 1 - \frac{1}{2}ABC(f_0, f_1).$$

As Stine and Heyse (2001) mention, one of the important properties of $PSR(f_0, f_1)$ is that it is invariant under transformations of the populations.

Nonparametric estimates of the proportion of similar responses have been used in the literature by replacing f_0 and f_1 by smoothed estimates. It is interesting to mention here that $PSR(f_0, f_1)$ can be rewritten in terms of the relative density of X_1 wrt X_0, $r_0^1(t)$,

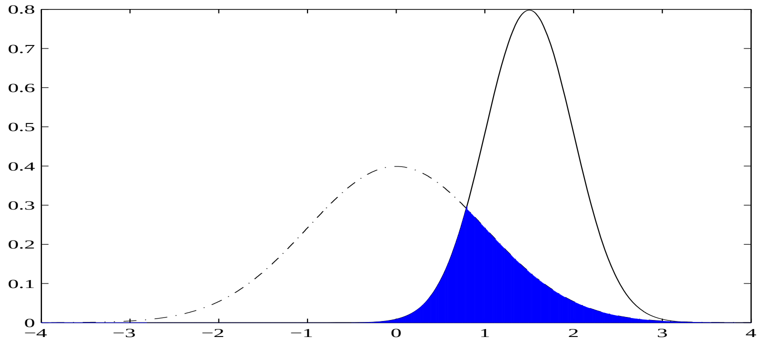

Figure 1.7: Graphical representation of $PSR(f_0, f_1)$, where f_0 is a N(0,1) density (dashed-dotted line) and f_1 is a N(1.5,0.25) density (solid line). The dark area represents the $PSR(f_0, f_1)$.

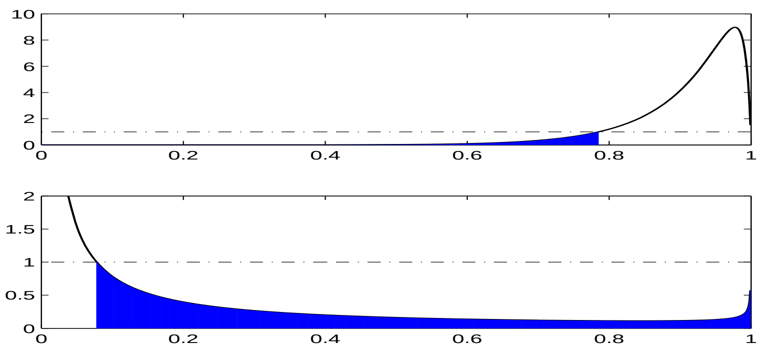

Figure 1.8: Graphical representation of $PSR(f_0, f_1)$ based on $r_0^1(t)$, the relative density of X_1 wrt X_0 (top panel) and $r_1^0(t)$, the relative density of X_0 wrt X_1 (bottom panel). The sum of the dark areas gives the value of $PSR(f_0, f_1)$.

and the relative density of X_0 wrt X_1, $r_1^0(t)$. From a simple reasoning, it is seen that:

$$PSR(f_0, f_1) = \int_{t:r_0^1(t)<1} r_0^1(t)\mathrm{d}t + \int_{t:r_1^0(t)<1} r_1^0(t)\mathrm{d}t.$$

Figures 1.7 and 1.8 correspond to two populations: X_0, a normal distributed random variable with mean 0 and variance 1, and X_1, a normal distributed random variable with mean 1.5 and variance 0.25. An alternative graphical representation of $PSR(f_0, f_1)$, in terms of the relative densities $r_0^1(t)$ and $r_1^0(t)$, can be found in Figure 1.8.

As Handcock and Morris (1999) point out, the differences between two distributions can be basically divided into two components: differences in location and differences in shape, a concept that comprises scale, skew and other distributional aspects. When F_1 is simply a location-shifted version of F_0, then $F_1(t) = F_0(t - c)$ or $F_1(t) = F_0(ct)$ for some constant c. If this is not the case, then $F_0(t)$ can be always location-adjusted to force it to have the same location as $F_1(t)$. If the differences existing between these two distributions, $F_0(t)$ and $F_1(t)$, are not only due to a shift in location, then all the differences remaining after a location-adjustment will be due to changes in shape. With this idea in mind, Handcock and Morris (1999) decompose the relative distribution into two components, one of them, $r_L(t)$, representing differences in location and the other one, $r_S(t)$, representing differences in shape:

$$r(t) = r_L(t) \cdot r_S(R_L(t)),$$

where

$$\begin{aligned} R_L(t) &= P(F_0(X_{0L}) \leq t) = F_{0L}(F_0^{-1}(t)), \\ R_S(t) &= P(F_{0L}(X_1) \leq t) = F_1(F_{0L}^{-1}(t)), \\ r_L(t) &= \frac{f_{0L}(F_0^{-1}(t))}{f_0(F_0^{-1}(t))}, \\ r_S(t) &= \frac{f_1(F_{0L}^{-1}(t))}{f_{0L}(F_{0L}^{-1}(t))}, \end{aligned}$$

and F_{0L} and f_{0L} denote the distribution and density function of a new random variable, X_{0L}, that has the shape of X_0 but the location of X_1.

This relationship is easily obtained from the fact that:

$$R(t) = R_S(R_L(t)).$$

Note that in the notation above, $R_L(t)$ and $r_L(t)$ refer respectively to the relative distribution and relative density of X_{0L} wrt X_0. Analogously, $R_S(t)$ and $r_S(t)$ denote respectively the relative distribution and relative density of X_1 wrt X_{0L}. While $R_L(t)$ and $r_L(t)$ isolate the differences in location between X_0 and X_1, $R_S(t)$ and $r_S(t)$ collect the remainder distributional differences, in scale, skew, etc.

Additionally, the shape component, $r_S(R_L(t))$, can be further decomposed. For instance, we could wish to isolate the difference in scale existing between the two populations. In that case, it follows that:

$$r(t) = r_L(t) \cdot r_{Sc}(R_L(t)) \cdot r_{rSh}(R_{LSc}(t)),$$

where

$$\begin{aligned} R_{Sc}(t) &= P(F_{0L}(X_{0LSc}) \leq t) = F_{0LSc}(F_{0L}^{-1}(t)), \\ R_{rSh}(t) &= P(F_{0LSc}(X_1) \leq t) = F_1(F_{0LSc}^{-1}(t)), \\ R_{LSc}(t) &= P(F_0(X_{0LSc}) \leq t) = F_{0LSc}(F_0^{-1}(t)), \end{aligned}$$

$$
\begin{aligned}
r_{Sc}(t) &= \frac{f_{0LSc}(F_{0L}^{-1}(t))}{f_{0L}(F_{0L}^{-1}(t))}, \\
r_{rSh}(t) &= \frac{f_1(F_{0LSc}^{-1}(t))}{f_{0LSc}(F_{0LSc}^{-1}(t))}, \\
r_{LSc}(t) &= \frac{f_{0LSC}(F_0^{-1}(t))}{f_0(F_0^{-1}(t))},
\end{aligned}
$$

and F_{0LSc} and f_{0LSc} denote the distribution and density function of a new random variable, X_{0LSc}, that has the residual shape of X_0 but the location and scale of X_1.

In the notation above, $R_{Sc}(t)$ and $r_{Sc}(t)$ refer respectively to the relative distribution and relative density of X_{0LSc} wrt X_{0L}. Analogously, $R_{rSh}(t)$ and $r_{rSh}(t)$ denote respectively the relative distribution and relative density of X_1 wrt X_{0LSc} and $R_{LSc}(t)$ and $r_{LSc}(t)$ refer to respectively the relative distribution and relative density of X_{0LSc} wrt X_0. While $R_{Sc}(t)$ and $r_{Sc}(t)$ isolate the differences in scale between X_0 and X_1, $R_{rSh}(t)$ and $r_{rSh}(t)$ collect what we call the residual shape, i.e., the remainder distributional differences excluding those in scale.

This approach of decomposing the relative density could be further extended to add parametric effects and isolate the part of the residual shape following such parametric model.

Several authors have considered the following smooth estimator of the relative density:

$$
\begin{aligned}
\hat{r}_h(t) &= \int_{-\infty}^{\infty} K_h\left(t - F_{0n_0}(v)\right) \mathrm{d}F_{1n_1}(v) = \frac{1}{n_1}\sum_{j=1}^{n_1} K_h\left(t - F_{0n_0}(X_{1j})\right) \qquad (1.32) \\
&= \left(K_h * F_{1n_1}F_{0n_0}^{-1}\right)(t)
\end{aligned}
$$

which can be derived in a natural way from the fact that

$$
r(t) \approx \frac{1}{h}\int_0^1 K\left(\frac{t-z}{h}\right) \mathrm{d}R\,(z) = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{t-F_0\,(z)}{h}\right) \mathrm{d}F_1\,(z)\,,
$$

under sufficiently smooth conditions.

Ćwik and Mielniczuk (1993) and Handcock and Janssen (2002) studied the asymptotic normality of $\hat{r}_h(t)$ or a slightly modified version of it that corrects for possible boundary effects. The result given by Handcock and Janssen (2002) is summarized in the following theorem.

Theorem 1.4.4. *(Asymptotic normality of $\hat{r}_h(t)$: Theorem 1 in Handcock and Janssen (2002))*
Assume condition (K6) and that K has compact support. If $t \in (0,1)$ and the densities $f_0(x)$ and $f_1(x)$ are smooth enough so that $r(t)$ is uniformly continuous, then for each

bandwidth sequence $\{h = h_{n_1}\}$ *such that* $h \to 0$, $n_1 h^3 \to \infty$, $n_1 h^5 \to 0$ *and* $n_1/n_0 \to \kappa^2 < \infty$, *we have*

$$\sqrt{n_1 h}\left(\hat{r}_h(t) - r(t)\right) \xrightarrow{d} N\left(0, r(t)R(K) + \kappa^2 r^2(t) R(K)\right).$$

It is worth mentioning here that Ćwik and Mielniczuk (1993) studied the MISE of only the dominant part of $\hat{r}_h(t)$ and proposed from it, a plug-in STE bandwidth selector. On the other hand, Handcock and Janssen (2002) obtained the asymptotic bias and variance of $\hat{r}_h(t)$ for large sample sizes and also a more accurate expression of the variance for small sample sizes. However, since these authors did not carry out a detailed study of the MISE of $\hat{r}_h(t)$, this will be done in the following chapter.

Motivated by the kernel relative density estimator (1.32) one can define a smooth relative distribution function estimator:

$$\begin{aligned} \hat{R}_h(t) &= \int_{-\infty}^{t} \hat{r}_h(x)\mathrm{d}x = \int \mathbb{K}\left(\frac{t - F_{0n_0}(x)}{h}\right) \mathrm{d}F_{1n_1}(x) \\ &= \frac{1}{n_1}\sum_{j=1}^{n_1} \mathbb{K}\left(\frac{t - F_{0n_0}(X_{1j})}{h}\right). \end{aligned} \qquad (1.33)$$

This is a smoothed version of the estimator given in (1.31).

Chapter 2

Bandwidth selection for the relative density with complete data

— Don't say, "Yes!"
Just take my hand and dance with me.
Oriah Mountain Dreamer

2.1 Kernel-type relative density estimators

Consider the two-sample problem with completely observed data:

$$\{X_{01}, \ldots, X_{0n_0}\},\ \{X_{11}, \ldots, X_{1n_1}\},$$

where the X_{0i}'s are independent and identically distributed as X_0 and the X_{1j}'s are independent and identically distributed as X_1. These two sequences are independent each other.

Throughout the whole thesis all the asymptotic results are obtained under the following assumption:

(S2) Both sample sizes n_0 and n_1 tend to infinity in such a way that, for some constant $0 < \kappa^2 < \infty$, $\lim_{n_1 \to \infty} \frac{n_1}{n_0} = \kappa^2$.

Besides, we assume the following conditions on the underlying distributions, the kernels K and M and the bandwidths h and h_0 to be used in the estimators (see (1.32) above and (2.1) and (2.2) below):

(B2) $h \to 0$ and $n_1 h^3 \to \infty$.

(B3) $h_0 \to 0$ and $n_0 h_0^4 \to 0$.

(D10) F_0 and F_1 have continuous density functions, f_0 and f_1, respectively.

(D11) f_0 is a three times differentiable density function with $f_0^{(3)}$ bounded.

(K7) K is a symmetric four times differentiable density function with compact support $[-1, 1]$ and $K^{(4)}$ is bounded.

(K8) M is a symmetric density and continuous function except at a finite set of points.

(R1) r is a twice continuously differentiable density with compact support contained in $[0, 1]$.

As it was mentioned at the end of Chapter 1, since $\frac{1}{h}\int_0^1 K\left(\frac{t-z}{h}\right) \mathrm{d}R(z)$ is close to $r(t)$ and for smooth distributions it is satisfied that:

$$\frac{1}{h}\int_0^1 K\left(\frac{t-z}{h}\right) \mathrm{d}R(z) = \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{t-F_0(z)}{h}\right) \mathrm{d}F_1(z),$$

a natural way to define a kernel-type estimator of $r(t)$ is by replacing the unknown functions F_0 and F_1 by some appropriate estimators. We consider two proposals, $\hat{r}_h(t)$, as introduced previously in (1.32), and the following one

$$\begin{aligned}\hat{r}_{h,h_0}(t) &= \int_{-\infty}^{\infty} K_h\left(t - \tilde{F}_{0h_0}(v)\right) \mathrm{d}F_{1n_1}(v) = \frac{1}{n_1}\sum_{j=1}^{n_1} K_h\left(t - \tilde{F}_{0h_0}(X_{1j})\right) \\ &= \left(K_h * F_{1n_1}\tilde{F}_{0h_0}^{-1}\right)(t),\end{aligned} \tag{2.1}$$

where $K_h(t) = \frac{1}{h}K\left(\frac{t}{h}\right)$, K is a kernel function, h is the bandwidth used to estimate r, F_{0n_0} and F_{1n_1} are the empirical distribution functions based on X_{0i}'s and X_{1j}'s, respectively, and $\tilde{F}_{0h_0}$ is a kernel-type estimate of F_0 given by:

$$\tilde{F}_{0h_0}(x) = n_0^{-1}\sum_{i=1}^{n_0} \mathbb{M}\left(\frac{x - X_{0i}}{h_0}\right), \tag{2.2}$$

where $\mathbb{M}$ denotes the cdf of the kernel M and h_0 is the bandwidth used to estimate F_0.

2.2 Selection criterion based on the mean integrated squared error

Using a Taylor expansion, $\hat{r}_h(t)$ can be written as follows:

$$\begin{aligned}\hat{r}_h(t) &= \int_{-\infty}^{\infty} K_h\left(t - F_0(z)\right) \mathrm{d}F_{1n_1}(z) \\ &\quad + \int_{-\infty}^{\infty} K_h^{(1)}\left(t - F_0(z)\right)\left(F_0(z) - F_{0n_0}(z)\right) \mathrm{d}F_{1n_1}(z) \\ &\quad + \int_{-\infty}^{\infty} \left(F_0(z) - F_{0n_0}(z)\right)^2 \\ &\quad \int_0^1 (1-s) K_h^{(2)}\left(t - F_0(z) - s\left(F_{0n_0}(z) - F_0(z)\right)\right) \mathrm{d}s \mathrm{d}F_{1n_1}(z).\end{aligned}$$

Let us define $\tilde{U}_{n_0} = F_{0n_0} \circ F_0^{-1}$ and $\tilde{R}_{n_1} = F_{1n_1} \circ F_0^{-1}$. Then, $\hat{r}_h(t)$ can be rewritten in a useful way for the study of its mean integrated squared error (MISE):

$$\hat{r}_h(t) = \tilde{r}_h(t) + A_1(t) + A_2(t) + B(t), \text{ where}$$

$$\tilde{r}_h(t) = \int_{-\infty}^{\infty} K_h\left(t - F_0(z)\right) \mathrm{d}F_{1n_1}(z) = \frac{1}{n_1}\sum_{j=1}^{n_1} K_h\left(t - F_0(X_{1j})\right),$$

$$A_1(t) = \int_0^1 \left(v - \tilde{U}_{n_0}(v)\right) K_h^{(1)}(t-v)\, \mathrm{d}\left(\tilde{R}_{n_1} - R\right)(v),$$

$$A_2(t) = \frac{1}{n_0}\sum_{i=1}^{n_0} \int_{-\infty}^{\infty} (F_0(w) - 1_{\{X_{0i} \le w\}}) K_h^{(1)}\left(t - F_0(w)\right) \mathrm{d}F_1(w),$$

$$B(t) = \int_{-\infty}^{\infty} (F_0(z) - F_{0n_0}(z))^2 \int_0^1 (1-s) K_h^{(2)}\left(t - F_0(z) - s\left(F_{0n_0}(z) - F_0(z)\right)\right) \mathrm{d}s \mathrm{d}F_{1n_1}(z).$$

Proceeding in a similar way, we can rewrite $\hat{r}_{h,h_0}(t)$ as follows:

$$\hat{r}_{h,h_0}(t) = \tilde{r}_h(t) + A_1(t) + A_2(t) + \hat{A}(t) + \hat{B}(t), \text{ where}$$

$$\hat{A}(t) = \int \left(F_{0n_0}(w) - \tilde{F}_{0h_0}(w)\right) K_h^{(1)}\left(t - F_0(w)\right) \mathrm{d}F_{1n_1}(w),$$

$$\hat{B}(t) = \int_{-\infty}^{\infty} (F_0(z) - \tilde{F}_{0h_0}(z))^2 \int_0^1 (1-s) K_h^{(2)}\left(t - F_0(z) - s\left(\tilde{F}_{0h_0}(z) - F_0(z)\right)\right) \mathrm{d}s \mathrm{d}F_{1n_1}(z).$$

Our main result is an asymptotic representation for the MISE of $\hat{r}_h(t)$ and $\hat{r}_{h,h_0}(t)$. Before presenting the main result, we include some useful lemmas.

Lemma 2.2.1. *Assume conditions (S2), (B2), (B3), (D10), (D11), (K7), (K8) and (R1). Then*

$$\int_0^1 E[(\tilde{r}_h(t) - r(t))^2] \mathrm{d}t = \frac{1}{n_1 h} R(K) + \frac{1}{4} h^4 d_K^2 R\left(r^{(2)}\right) + o\left(\frac{1}{n_1 h} + h^4\right).$$

The proof of Lemma 2.2.1 is not included here because it is a classical result in the setting of ordinary density estimation in a one-sample problem (see Theorem 1.3.3 in Subsection 1.3.1 or Wand and Jones (1995) for more details).

Lemma 2.2.2. *Assume the conditions in Lemma 2.2.1. Then*

$$\int_0^1 E[A_2^2(t)]\mathrm{d}t = \frac{1}{n_0 h} R(r) R(K) + o\left(\frac{1}{n_0 h}\right) = O\left(\frac{1}{n_0 h}\right).$$

Proof of Lemma 2.2.2. Standard algebra gives

$$\begin{aligned} E[A_2^2(t)] &= \frac{1}{n_0^2 h^4} \sum_{i=1}^{n_0} \sum_{j=1}^{n_0} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E\left[(F_0(w_1) - 1_{\{X_{0i} \le w_1\}})(F_0(w_2) - 1_{\{X_{0j} \le w_2\}})\right] \\ &\quad K^{(1)}\left(\frac{t - F_0(w_1)}{h}\right) K^{(1)}\left(\frac{t - F_0(w_2)}{h}\right) \mathrm{d}F_1(w_1)\mathrm{d}F_1(w_2). \end{aligned}$$

Due to the independence between X_{0i} and X_{0j} for $i \neq j$, and using the fact that

$$Cov(1_{\{F_0(X_{0i}) \le u_1\}}, 1_{\{F_0(X_{0i}) \le u_2\}}) = (1 - u_1)(1 - u_2)\omega(u_1 \wedge u_2),$$

where $\omega(t) = \frac{t}{1-t}$, the previous expression can be rewritten as follows

$$\begin{aligned} E[A_2^2(t)] &= \frac{2}{n_0 h^4} \int_0^1 \int_{u_2}^1 (1 - u_1)(1 - u_2)\omega(u_2) \\ &\quad K^{(1)}\left(\frac{t - u_1}{h}\right) K^{(1)}\left(\frac{t - u_2}{h}\right) r(u_1) r(u_2) \mathrm{d}u_1 \mathrm{d}u_2 \\ &= -\frac{1}{n_0 h^4} \int_0^1 \omega(u_2) \mathrm{d}\left[\int_{u_2}^1 (1 - u_1) K^{(1)}\left(\frac{t - u_1}{h}\right) r(u_1) \mathrm{d}u_1\right]^2. \end{aligned}$$

Now, using integration by parts, it follows that

$$\begin{aligned} E[A_2^2(t)] &= \frac{-1}{n_0 h^4} \lim_{u_2 \to 1^-} \omega(u_2)\Omega^2(u_2) + \frac{1}{n_0 h^4} \lim_{u_2 \to 0^+} \omega(u_2)\Omega^2(u_2) \\ &\quad + \frac{1}{n_0 h^4} \int_0^1 \Omega^2(u_2)\omega^{(1)}(u_2)\mathrm{d}u_2, \end{aligned} \tag{2.3}$$

where

$$\Omega(u_2) = \int_{u_2}^1 (1 - u_1) K^{(1)}\left(\frac{t - u_1}{h}\right) r(u_1)\mathrm{d}u_1.$$

Since Ω is a bounded function and $\omega(0) = 0$, the second term in the right hand side of (2.3) vanishes. On the other hand, due to the boundedness of $K^{(1)}$ and r, it follows that $|\Omega(u_2)| \le \left\|K^{(1)}\right\|_\infty \|r\|_\infty \frac{(1-u_2)^2}{2}$, which let us conclude that the first term in (2.3) is zero as well. Therefore,

$$E[A_2^2(t)] = \frac{1}{n_0 h^4} \int_0^1 \Omega(u_2)^2 \omega^{(1)}(u_2)\mathrm{d}u_2. \tag{2.4}$$

Now, using integration by parts, it follows that

$$\begin{aligned}
\Omega(u_2) &= \left[-(1-u_1)r(u_1)hK\left(\frac{t-u_1}{h}\right)\right]_{u_2}^{1} \\
&\quad + \int_{u_2}^{1} hK\left(\frac{t-u_1}{h}\right)[-r(u_1)+(1-u_1)r^{(1)}(u_1)]\mathrm{d}u_1 \\
&= h(1-u_2)r(u_2)K\left(\frac{t-u_2}{h}\right) \\
&\quad + h\int_{u_2}^{1} K\left(\frac{t-u_1}{h}\right)[(1-u_1)r^{(1)}(u_1)-r(u_1)]\mathrm{d}u_1,
\end{aligned}$$

and plugging this last expression in (2.4), it is concluded that

$$E[A_2^2(t)] = \frac{1}{n_0 h^2}(I_{21}(t)+2I_{22}(t)+I_{23}(t)),$$

where

$$I_{21}(t) = \int_0^1 r^2(u_2)K^2\left(\frac{t-u_2}{h}\right)\mathrm{d}u_2,$$

$$\begin{aligned}
I_{22}(t) &= \int_0^1 \frac{1}{1-u_2}r(u_2)K\left(\frac{t-u_2}{h}\right) \\
&\quad \int_{u_2}^{1} K\left(\frac{t-u_1}{h}\right)[(1-u_1)r^{(1)}(u_1)-r(u_1)]\mathrm{d}u_1\mathrm{d}u_2,
\end{aligned}$$

$$\begin{aligned}
I_{23}(t) &= \int_0^1 \frac{1}{(1-u_2)^2}\int_{u_2}^{1}\int_{u_2}^{1} K\left(\frac{t-u_1}{h}\right)[(1-u_1)r^{(1)}(u_1)-r(u_1)] \\
&\quad K\left(\frac{t-u_1^*}{h}\right)[(1-u_1^*)r^{(1)}(u_1^*)-r(u_1^*)]\mathrm{d}u_1\mathrm{d}u_1^*\mathrm{d}u_2.
\end{aligned}$$

Therefore,

$$\int_0^1 E[A_2^2(t)]\mathrm{d}t = \int_0^1 \frac{1}{n_0h^2}I_{21}(t)\mathrm{d}t + 2\int_0^1 \frac{1}{n_0h^2}I_{22}(t)\mathrm{d}t + \int_0^1 \frac{1}{n_0h^2}I_{23}(t)\mathrm{d}t. \tag{2.5}$$

Next, we will study each summand in (2.5) separately. The first term can be handled by using changes of variable and a Taylor expansion:

$$\int_0^1 \frac{1}{n_0h^2}I_{21}(t)\mathrm{d}t = \frac{1}{n_0h}\int_0^1 r^2(u_2)\left(\int_{-\frac{u_2}{h}}^{\frac{1-u_2}{h}} K^2(s)\mathrm{d}s\right)\mathrm{d}u_2.$$

Let us define $\mathbb{K}_2(x) = \int_{-\infty}^{x} K^2(s)\,\mathrm{d}s$ and rewrite the previous term as follows

$$\int_0^1 \frac{1}{n_0h^2}I_{21}(t)\mathrm{d}t = \frac{1}{n_0h}\int_0^1 r^2(u_2)\left(\mathbb{K}_2\left(\frac{1-u_2}{h}\right) - \mathbb{K}_2\left(-\frac{u_2}{h}\right)\right)\mathrm{d}u_2.$$

Now, by splitting the integration interval into three subintervals: $[0,h]$, $[h,1-h]$ and $[1-h,1]$, using changes of variable and the fact that

$$\mathbb{K}_2(x) = \begin{cases} R(K) & \forall x \geq 1, \\ 0 & \forall x \leq -1, \end{cases}$$

it is easy to show that

$$\int_0^1 \frac{1}{n_0 h^2} I_{21}(t)\mathrm{d}t = \frac{1}{n_0 h} R(K)R(r) + O\left(\frac{1}{n_0}\right).$$

To study the second term in the right hand side of (2.5), we first take into account the fact that

$$\begin{aligned} \left|\int_0^1 I_{22}(t)\mathrm{d}t\right| \leq & \int_0^1\int_0^1 \frac{r(u_2)}{1-u_2} K\left(\frac{t-u_2}{h}\right) \int_{u_2}^1 K\left(\frac{t-u_1}{h}\right)(1-u_1)|r^{(1)}(u_1)|\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}t \\ & + \int_0^1\int_0^1 \frac{r(u_2)}{1-u_2} K\left(\frac{t-u_2}{h}\right) \int_{u_2}^1 K\left(\frac{t-u_1}{h}\right) r(u_1)\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}t. \end{aligned}$$

Now, simple algebra, changes of variable and the Cauchy-Schwarz inequality lead to:

$$\begin{aligned} \left|\int_0^1 I_{22}(t)\mathrm{d}t\right| \leq & \int_0^1\int_0^1 r(u_2) K\left(\frac{t-u_2}{h}\right) h\|r^{(1)}\|_\infty \int_{-1}^1 K(s)\mathrm{d}s\mathrm{d}u_2\mathrm{d}t \\ & + \int_0^1\int_0^1 \frac{r(u_2)}{1-u_2} K\left(\frac{t-u_2}{h}\right) \\ & \left[\left(\int_{u_2}^1 K^2\left(\frac{t-u_1}{h}\right)\mathrm{d}u_1\right)^{\frac{1}{2}} \left(\int_{u_2}^1 r^2(u_1)\mathrm{d}u_1\right)^{\frac{1}{2}}\right] \mathrm{d}u_2\mathrm{d}t \\ \leq & \; h^2\|r^{(1)}\|_\infty + h^{\frac{3}{2}} R(K)^{\frac{1}{2}} \|r\|_\infty^2 \int_0^1 \frac{1}{(1-u_2)^{\frac{1}{2}}}\mathrm{d}u_2 \\ = & \; h^2\|r^{(1)}\|_\infty + 2h^{\frac{3}{2}} R(K)^{\frac{1}{2}} \|r\|_\infty^2. \end{aligned}$$

Then, using conditions $\|r\|_\infty < \infty$, $\left\|r^{(1)}\right\|_\infty < \infty$ and $R(K) < \infty$, we can conclude that

$$\int_0^1 \frac{1}{n_0 h^2} I_{22}(t)\mathrm{d}t = O\left(\frac{1}{n_0} + \frac{1}{n_0 h^{\frac{1}{2}}}\right) = O\left(\frac{1}{n_0 h^{\frac{1}{2}}}\right).$$

Below, we will study the third term in the right hand of (2.5). First, we take into account the fact that

$$\left|\int_0^1 I_{23}(t)\mathrm{d}t\right| \leq I_{23}^{(1)} + 2I_{23}^{(2)} + I_{23}^{(3)}, \text{ where} \tag{2.6}$$

$$
\begin{aligned}
I_{23}^{(1)} &= \int_0^1\int_0^1 \frac{1}{(1-u_2)^2}\int_{u_2}^1 K\left(\frac{t-u_1}{h}\right)(1-u_1)|r^{(1)}(u_1)| \\
&\quad \int_{u_2}^1 K\left(\frac{t-u_1^*}{h}\right)(1-u_1^*)|r^{(1)}(u_1^*)|\mathrm{d}u_1^*\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}t, \\
I_{23}^{(2)} &= \int_0^1\int_0^1 \frac{1}{(1-u_2)^2}\int_{u_2}^1 K\left(\frac{t-u_1}{h}\right)(1-u_1)|r^{(1)}(u_1)| \\
&\quad \int_{u_2}^1 K\left(\frac{t-u_1^*}{h}\right) r(u_1^*)\mathrm{d}u_1^*\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}t
\end{aligned}
$$

and

$$
\begin{aligned}
I_{23}^{(3)} &= \int_0^1\int_0^1 \frac{1}{(1-u_2)^2}\int_{u_2}^1 K\left(\frac{t-u_1}{h}\right) r(u_1) \\
&\quad \int_{u_2}^1 K\left(\frac{t-u_1^*}{h}\right) r(u_1^*)\mathrm{d}u_1^*\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}t.
\end{aligned}
$$

Then, using changes of variable and the Cauchy Schwarz inequality it is easy to show that, after simple algebra, the terms in the right hand side of (2.6) can be bounded as follows:

$$
I_{23}^{(1)} \leq \|r^{(1)}\|_\infty^2 h^2 \int_0^1\int_0^1 \frac{(1-u_2)^2}{(1-u_2)^2}\mathrm{d}u_2\mathrm{d}t = O\left(h^2\right) = O\left(h^{\frac{3}{2}}\right),
$$

$$
\begin{aligned}
I_{23}^{(2)} &\leq \int_0^1\int_0^1 \frac{1}{(1-u_2)^2}\int_{u_2}^1 K\left(\frac{t-u_1}{h}\right)(1-u_1)|r^{(1)}(u_1)| \\
&\quad \left[\left(\int_{u_2}^1 K\left(\frac{t-u_1^*}{h}\right)^2 \mathrm{d}u_1^*\right)^{\frac{1}{2}}\left(\int_{u_2}^1 r^2(u_1^*)\mathrm{d}u_1^*\right)^{\frac{1}{2}}\right]\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}t \\
&= h^{\frac{3}{2}} R(K)^{\frac{1}{2}}\|r\|_\infty\|r^{(1)}\|_\infty \int_0^1\int_0^1 (1-u_2)^{-\frac{1}{2}}\mathrm{d}u_2\mathrm{d}t \\
&= 2h^{\frac{3}{2}} R(K)^{\frac{1}{2}}\|r\|_\infty\|r^{(1)}\|_\infty = O\left(h^{\frac{3}{2}}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
I_{23}^{(3)} &\leq \int_0^1\int_0^1 \frac{1}{(1-u_2)^2}\int_{u_2}^1 K\left(\frac{t-u_1}{h}\right) r(u_1) \\
&\quad \left[\left(\int_{u_2}^1 K\left(\frac{t-u_1^*}{h}\right)\mathrm{d}u_1^*\right)^{\frac{1}{2}}\left(\int_{u_2}^1 K\left(\frac{t-u_1^*}{h}\right) r^2(u_1^*)\mathrm{d}u_1^*\right)^{\frac{1}{2}}\right] \\
&\quad \mathrm{d}u_1\mathrm{d}u_2\mathrm{d}t \\
&\leq \int_0^1\int_0^1 \frac{1}{(1-u_2)^2}\int_{u_2}^1 K\left(\frac{t-u_1}{h}\right) r(u_1) \\
&\quad \left[h^{\frac{1}{2}}\left(\int_{u_2}^1 K^2\left(\frac{t-u_1^*}{h}\right)\mathrm{d}u_1^*\right)^{\frac{1}{4}}\left(\int_{u_2}^1 r^4(u_1^*)\mathrm{d}u_1^*\right)^{\frac{1}{4}}\right]\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}t \\
&= h^{\frac{3}{2}}\|K\|_\infty^{\frac{1}{2}}\|r\|_\infty^2 \int_0^1 (1-u_2)^{-\frac{1}{2}}\mathrm{d}u_2 = 2h^{\frac{3}{2}}\|K\|_\infty^{\frac{1}{2}}\|r\|_\infty^2 = O\left(h^{\frac{3}{2}}\right).
\end{aligned}
$$

Consequently, we can conclude that

$$\int_0^1 \frac{1}{n_0 h^2} I_{23}(t)\mathrm{d}t = O\left(\frac{1}{n_0 h^{\frac{1}{2}}}\right).$$

Therefore, it has been shown that

$$\int_0^1 E[A_2^2(t)]\mathrm{d}t = \frac{1}{n_0 h} R(r)R(K) + O\left(\frac{1}{n_0}\right) + O\left(\frac{1}{n_0 h^{\frac{1}{2}}}\right).$$

Finally the proof concludes using condition (B2). □

Lemma 2.2.3. *Assume the conditions in Lemma 2.2.1. Then*

$$\int_0^1 E[A_1^2(t)]\mathrm{d}t = o\left(\frac{1}{n_0 h}\right).$$

Proof of Lemma 2.2.3. Direct calculations lead to

$$E[A_1^2(t)] = \frac{1}{h^4} E\left[I_1(t)\right], \tag{2.7}$$

where

$$\begin{aligned} I_1(t) &= E\left[\int_0^1\int_0^1 (v_1 - \tilde{U}_{n_0}(v_1))(v_2 - \tilde{U}_{n_0}(v_2))K^{(1)}\left(\frac{t-v_1}{h}\right)\right. \\ &\left. K^{(1)}\left(\frac{t-v_2}{h}\right)\mathrm{d}(\tilde{R}_{n_1} - R)(v_1)\mathrm{d}(\tilde{R}_{n_1} - R)(v_2)/X_{01},\ldots,X_{0n_0}\right]. \end{aligned} \tag{2.8}$$

To tackle with (2.7) we first study the conditional expectation (2.8). It is easy to see that

$$I_1(t) = Var[V(t)/X_{01},\ldots,X_{0n_0}],$$

where

$$V(t) = \frac{1}{n_1}\sum_{j=1}^{n_1}\left(X_{1j} - \tilde{U}_{n_0}(X_{1j})\right)K^{(1)}\left(\frac{t - X_{1j}}{h}\right).$$

Thus

$$\begin{aligned} I_1(t) &= \frac{1}{n_1}\left\{\int_0^1\left[(v - \tilde{U}_{n_0}(v))K^{(1)}\left(\frac{t-v}{h}\right)\right]^2\mathrm{d}R(v)\right. \\ &\left. -\left[\int_0^1 (v - \tilde{U}_{n_0}(v))K^{(1)}\left(\frac{t-v}{h}\right)\mathrm{d}R(v)\right]^2\right\} \end{aligned}$$

and

$$\begin{aligned} E[A_1^2(t)] &= \frac{1}{n_1 h^4}\int_0^1 E\left\{\left[(v - \tilde{U}_{n_0}(v))\right]^2\right\}\left[K^{(1)}\left(\frac{t-v}{h}\right)\right]^2\mathrm{d}R(v) \\ &-\frac{1}{n_1 h^4}\int_0^1\int_0^1 E\left[(v_1 - \tilde{U}_{n_0}(v_1))(v_2 - \tilde{U}_{n_0}(v_2))\right] \\ &K^{(1)}\left(\frac{t-v_1}{h}\right)K^{(1)}\left(\frac{t-v_2}{h}\right)\mathrm{d}R(v_1)\mathrm{d}R(v_2). \end{aligned}$$

Taking into account that

$$E\left[\sup_v |(\tilde{U}_{n_0}(v)-v)|^2\right] = \int_0^\infty P\left(\sup_v |(\tilde{U}_{n_0}(v)-v)|^2 > c\right) \mathrm{d}c,$$

we can use the Dvoretzky-Kiefer-Wolfowitz inequality (Dvoretzky *et al* (1956)), to conclude that

$$E\left[\sup_v |(\tilde{U}_{n_0}(v)-v)|^2\right] \leq \int_0^\infty 2e^{-(2n_0 c)}\mathrm{d}c = \frac{2}{n_0}\int_0^\infty ye^{-y^2}\mathrm{d}y = O\left(\frac{1}{n_0}\right). \qquad (2.9)$$

Consequently, using (2.9) and the conditions $\|r\|_\infty < \infty$ and $\left\|K^{(1)}\right\|_\infty < \infty$ we obtain that $E[A_1^2(t)] = O\left(\frac{1}{n_1 n_0 h^4}\right)$. The proof is concluded using condition (B2). □

Lemma 2.2.4. *Assume the conditions in Lemma (2.2.1). Then*

$$\int_0^1 E[B^2(t)]\mathrm{d}t = o\left(\frac{1}{n_0 h}\right).$$

Proof of Lemma 2.2.4. By using a Taylor expansion, we can rewrite $E\left[B^2(t)\right]$ as it is detailed below:

$$\begin{aligned} E\left[B^2(t)\right] &= E\left[T_1^2(t) + T_2^2(t) + T_3^2(t) + 2T_1(t)T_2(t) + 2T_1(t)T_3(t) + 2T_2(t)T_3(t)\right] \\ &= E\left[T_1^2(t)\right] + E\left[T_2^2(t)\right] + E\left[T_3^2(t)\right] \\ &\quad +2E\left[T_1(t)T_2(t)\right] + 2E\left[T_1(t)T_3(t)\right] + 2\left[T_2(t)T_3(t)\right], \end{aligned}$$

where

$$\begin{aligned} T_1(t) &= \frac{1}{2h^3}\int |F_0(z_1) - F_{0n_0}(z_1)|^2 \int_0^1 (1-s_1)\left|K^{(2)}\left(\frac{t-F_0(z_1)}{h}\right)\right| \\ &\quad \mathrm{d}s_1 \mathrm{d}F_{1n_1}(z_1), \\ T_2(t) &= \frac{1}{2h^3}\int |F_0(z_1) - F_{0n_0}(z_1)|^2 \int_0^1 (1-s_1)\, s_1 \frac{|F_{0n_0}(z_1) - F_0(z_1)|}{h} \\ &\quad \left|K^{(3)}\left(\frac{t-F_0(z_1)}{h}\right)\right| \mathrm{d}s_1 \mathrm{d}F_{1n_1}(z_1), \\ T_3(t) &= \frac{1}{2h^3}\int |F_0(z_1) - F_{0n_0}(z_1)|^2 \int_0^1 (1-s_1)\frac{s_1^2}{2}\frac{(F_{0n_0}(z_1) - F_0(z_1))^2}{h^2} \\ &\quad \left|K^{(4)}(\xi_n)\right| \mathrm{d}s_1 \mathrm{d}F_{1n_1}(z_1) \end{aligned}$$

with ξ_{n_0} a value between $\frac{t-F_0(z_1)}{h}$ and $\frac{t-F_{0n_0}(z_1)}{h}$.

It will be enough to study only the terms $E\left[T_i^2(t)\right]$ for $i = 1, 2, 3$. Using the bounds obtained for these terms and the Cauchy-Schwarz inequality it is easy to bound the other terms in $E\left[B^2(t)\right]$. To deal with $E\left[T_i^2(t)\right]$, we proceed as follows. The first step is to consider the equation $E[T_i^2(t)] = E\left[E\left[T_i^2(t)/X_{01},...,X_{0n_0}\right]\right]$. Next, by using the independence between the two samples, the Dvoretzky-Kiefer-Wolfowitz inequality, changes of variable and the conditions $\|r\|_\infty < \infty$, $\left\|K^{(1+i)}\right\|_\infty < \infty$, it is straightforward to show that $E[T_1^2(t)] = O\left(\frac{1}{n_0^2h^4}\right)$, $E[T_2^2(t)] = O\left(\frac{1}{n_0^3h^6}\right)$ and $E\left[T_3^2(t)\right] = O\left(\frac{1}{n_0^4h^{10}}\right)$. Consequently, (B2) implies that $E[T_i^2(t)] = o\left(\frac{1}{n_0h}\right)$. A direct application of the Cauchy-Schwarz inequality as well as the bounds obtained for the terms $E\left[T_i^2(t)\right]$ and condition (B2) imply $E\left[T_i(t)T_j(t)\right] = o\left(\frac{1}{n_0h}\right)$ with $i \neq j$. □

Lemma 2.2.5. *Assume the conditions in Lemma 2.2.1. Then*

$$\int_0^1 E[2A_1(t)(\tilde{r}_h - r)(t)]\mathrm{d}t = 0$$

$$\int_0^1 E[2A_2(t)(\tilde{r}_h - r)(t)]\mathrm{d}t = 0$$

$$\int_0^1 E[2B(t)(\tilde{r}_h - r)(t)]\mathrm{d}t = o\left(\frac{1}{n_1h} + h^4\right)$$

$$\int_0^1 E[2A_1(t)A_2(t)]\mathrm{d}t = o\left(\frac{1}{n_1h} + h^4\right)$$

$$\int_0^1 E[2A_1(t)B(t)]\mathrm{d}t = o\left(\frac{1}{n_1h} + h^4\right)$$

$$\int_0^1 E[2A_2(t)B(t)]\mathrm{d}t = o\left(\frac{1}{n_0h} + h^4\right).$$

Proof of Lemma 2.2.5. In order to show that $\int_0^1 E[2A_1(t)(\widetilde{r}_h - r)(t)]\mathrm{d}t = 0$ we start studying $E\left[A_1(t)(\widetilde{r}_h - r)(t)\right]$:

$$E\left[A_1(t)(\widetilde{r}_h - r)(t)\right] = E\left[(\widetilde{r}_h - r)(t)E\left[\frac{1}{h^2}\int(v - \widetilde{U}_{n_0}(v))K^{(1)}\left(\frac{t-v}{h}\right)\right.\right.$$
$$\left.\left.\mathrm{d}(\widetilde{R}_{n_1} - R)(v)/X_{11},\ldots,X_{1n_1}\right]\right].$$

Below, we will show that

$$E\left[\frac{1}{h^2}\int(v - \widetilde{U}_{n_0}(v))K^{(1)}\left(\frac{t-v}{h}\right)\mathrm{d}(\widetilde{R}_{n_1} - R)(v)/X_{11},\ldots,X_{1n_1}\right] = 0.$$

In fact,

$$E\left[\frac{1}{h^2}\int(v - \widetilde{U}_{n_0}(v))K^{(1)}\left(\frac{t-v}{h}\right)\mathrm{d}(\widetilde{R}_{n_1} - R)(v)/X_{11},\ldots,X_{1n_1}\right]$$
$$= \frac{1}{h^2}\int E\left(v - \frac{\sum_{i=1}^{n_0} 1_{\{X_{0i}\leq F_0^{-1}(v)\}}}{n_0}\right)K^{(1)}\left(\frac{t-v}{h}\right)\mathrm{d}(F_{1n_1}F_0^{-1} - F_1F_0^{-1})(v)$$
$$= \frac{1}{h^2}\int\left(v - E(1_{\{X_{0i}\leq F_0^{-1}(v)\}})\right)K^{(1)}\left(\frac{t-v}{h}\right)\mathrm{d}(F_{1n_1}F_0^{-1} - F_1F_0^{-1})(v)$$
$$= \frac{1}{h^2}\int(v - v)K^{(1)}\left(\frac{t-v}{h}\right)\mathrm{d}(F_{1n_1}F_0^{-1} - F_1F_0^{-1})(v) = 0.$$

In order to show that $\int_0^1 E[2A_2(t)(\widetilde{r}_h-r)(t)]\mathrm{d}t = 0$ we start studying $E[A_2(t)(\widetilde{r}_h-r)(t)]$. It follows that

$$\begin{aligned}
&E[A_2(t)(\widetilde{r}_h - r)(t)]\\
&= \frac{1}{n_0h^2}E\left[\sum_{i=1}^{n_0}\int(F_0(w)-1_{\{X_{0i}\leq w\}})K^{(1)}\left(\frac{t-F_0(w)}{h}\right)\mathrm{d}F_1(w)(\widetilde{r}_h-r)(t)\right]\\
&=\frac{1}{h^2}E\left[E\left[\int(F_0(w)-1_{\{X_{0i}\leq w\}})K^{(1)}\left(\frac{t-F_0(w)}{h}\right)\mathrm{d}F_1(w)(\widetilde{r}_h-r)(t)/X_{11},\ldots,X_{1n_1}\right]\right]\\
&= \frac{1}{h^2}E\left[(\widetilde{r}_h-r)(t)\int E\left[(F_0(w)-1_{\{X_{0i}\leq w\}})K^{(1)}\left(\frac{t-F_0(w)}{h}\right)/X_{11},\ldots,X_{1n_1}\right]\mathrm{d}F_1(w)\right]\\
&= \frac{1}{h^2}E\left[(\widetilde{r}_h-r)(t)\int E\left[(F_0(w)-1_{\{X_{0i}\leq w\}})K^{(1)}\left(\frac{t-F_0(w)}{h}\right)\right]\mathrm{d}F_1(w)\right]\\
&= \frac{1}{h^2}E\left[(\widetilde{r}_h-r)(t)\int(F_0(w)-F_0(w))K^{(1)}\left(\frac{t-F_0(w)}{h}\right)\mathrm{d}F_1(w)\right] = 0.
\end{aligned}$$

The proof of the other results stated in Lemma 2.2.5 are omitted here. They can be obtained as straightforward consequences from the Cauchy-Schwarz inequality and the bounds obtained in the previous lemmas. □

Lemma 2.2.6. *Assume the conditions in Lemma 2.2.1. Then*
(i) $\int_0^1 E[\hat{A}^2(t)]\mathrm{d}t = o\left(\frac{1}{n_0h}\right)$.
(ii) $\int_0^1 E[\hat{B}^2(t)]\mathrm{d}t = o\left(\frac{1}{n_0h}\right)$.

Proof of Lemma 2.2.6. We start proving (i). Let us define $D_{n_0}(w) = \tilde{F}_{0h_0}(w) - F_{0n_0}(w)$, then

$$\begin{aligned}
&E[\hat{A}^2(t)] = E[E[\hat{A}^2(t)/X_{11},\ldots,X_{1n_1}]]\\
&= E\left[\iint E[D_{n_0}(w_1)D_n(w_2)]K_h^{(1)}(t-F_0(w_1))K_h^{(1)}(t-F_0(w_2))\mathrm{d}F_{1n_1}(v_1)\mathrm{d}F_{1n_1}(v_2)\right].
\end{aligned}$$

Based on the results set for $D_{n_0}(w)$ in Hjort and Walker (2001), the conditions (D11) and (K8) and since $E[D_{n_0}(w_1)D_{n_0}(w_2)] = Cov(D_{n_0}(w_1), D_{n_0}(w_2)) + E[D_{n_0}(w_1)]E[D_{n_0}(w_2)]$, it follows that $E[D_{n_0}(w_1)D_{n_0}(w_2)] = O\left(\frac{h_0^4}{n_0}\right) + O(h_0^4)$.

Therefore, for any $t \in [0,1]$, we can bound $E[\hat{A}^2(t)]$, using suitable constants C_2 and C_3 as follows

$$\begin{aligned}
E[\hat{A}^2(t)] \;=\;& C_2\frac{h_0^4}{h^4}\frac{1}{n_1}\int\left(K^{(1)}\left(\frac{t-F_0(z)}{h}\right)\right)^2 f_1(z)\mathrm{d}z + C_3\frac{h_0^4}{h^4}\frac{(n_1-1)}{n_1}\\
&\iint\left|K^{(1)}\left(\frac{t-F_0(z_1)}{h}\right)\right|\left|K^{(1)}\left(\frac{t-F_0(z_2)}{h}\right)\right| f_1(z_1)f_1(z_2)\mathrm{d}z_1\mathrm{d}z_2.
\end{aligned}$$

Besides, condition (R1) allows us to conclude that $\int\left(K^{(1)}\left(\frac{t-F_0(z)}{h}\right)\right)^2 f_1(z)\mathrm{d}z = O(h)$ and $\iint\left|K^{(1)}\left(\frac{t-F_0(z_1)}{h}\right)\right|\left|K^{(1)}\left(\frac{t-F_0(z_2)}{h}\right)\right| f_1(z_1)f_1(z_2)\mathrm{d}z_1\mathrm{d}z_2 = O(h^2)$ uniformly in $t \in [0,1]$.

Therefore, $\int_0^1 E\left[\hat{A}^2(t)\right] \mathrm{d}t = O\left(\frac{h_0^4}{n_0 h^3}\right) + O\left(\frac{h_0^4}{h^2}\right)$, which, taking into account conditions (B2) and (B3), implies (i).

We next prove (ii). The proof is parallel to that of Lemma 2.2.4. The only difference now is that instead of requiring $E[\sup |F_{0n_0}(x) - F_0(x)|^p] = O\left(n_0^{-\frac{p}{2}}\right)$, where p is an integer larger than 1, it is required that

$$E\left[\sup\left|\tilde{F}_{0h_0}(x) - F_0(x)\right|^p\right] = O\left(n_0^{-\frac{p}{2}}\right). \tag{2.10}$$

To conclude the proof, below we show that (2.10) is satisfied. Define $H_{n_0} = \sup\left|\tilde{F}_{0h_0}(x) - F_0(x)\right|$, then, as it is stated in Ahmad (2002), it follows that $H_{n_0} \leq E_{n_0} + W_{n_0}$ where $E_{n_0} = \sup |F_{0n_0}(x) - F_0(x)|$ and $W_{n_0} = \sup\left|E\tilde{F}_{0h_0}(x) - F_0(x)\right| = O\left(h_0^2\right)$. Using the binomial formula it is easy to obtain that, for any integer $p \geq 1$, $H_{n_0}^p \leq \sum_{j=0}^p C_j^p W_{n_0}^{p-j} E_{n_0}^j$, where the constants C_j^p's (with $j \in \{0, 1, \ldots, p-1, p\}$) are the binomial coefficients. Therefore, since $E[E_{n_0}^j] = O\left(n_0^{-\frac{j}{2}}\right)$ and $W_{n_0}^{p-j} = O\left(h_0^{2(p-j)}\right)$, condition (B3) leads to $W_{n_0}^{p-j} E\left[E_{n_0}^j\right] = O\left(n_0^{-\frac{p}{2}}\right)$.
As a straightforward consequence, (2.10) holds and the proof of (ii) is concluded. □

Theorem 2.2.7 (AMISE)**.** *Assume conditions (S2), (D10), (R1), (K7) and (B2). Then*

$$MISE(\hat{r}_h) = AMISE(h) + o\left(\frac{1}{n_1 h} + h^4\right) + o\left(\frac{1}{n_0 h}\right)$$

with

$$AMISE(h) = \frac{1}{n_1 h} R(K) + \frac{1}{4} h^4 d_K^2 R(r^{(2)}) + \frac{1}{n_0 h} R(r) R(K).$$

If conditions (D11), (K8), and (B3) are assumed as well, then the same result is satisfied for the $MISE(\hat{r}_{h,h_0})$.

The proof of Theorem 2.2.7 is a direct consequence of the previous lemmas where each one of the terms that result from expanding the expression for the MISE are studied. As it was proved above, some of them produce dominant parts in the final expression for the MISE while others yield negligible terms.

Remark 2.2.1. From Theorem 2.2.7 it follows that the optimal bandwidth, minimizing the asymptotic mean integrated squared error of any of the estimators considered for r, is given by

$$h_{AMISE} = \mathbf{C} n_1^{-\frac{1}{5}}, \text{ where } \mathbf{C} = \left(\frac{R(K)(R(r)\kappa^2 + 1)}{d_K^2 R(r^{(2)})}\right)^{\frac{1}{5}}. \tag{2.11}$$

Remark 2.2.2. Note that $AMISE(h)$ derived from Theorem 2.2.7 does not depend on the bandwidth h_0. A higher-order analysis should be considered to address simultaneously the bandwidth selection problem of h and h_0.

2.3 Plug-in and STE selectors

2.3.1 Estimation of density functionals

It is very simple to show that, under sufficiently smooth conditions on r ($r \in C^{(2\ell)}(\mathbb{R})$), the functionals

$$R\left(r^{(\ell)}\right) = \int_0^1 \left(r^{(\ell)}(x)\right)^2 \mathrm{d}x \tag{2.12}$$

appearing in (2.11), are related to other general functionals of r, denoted by $\Psi_{2\ell}(r)$:

$$R\left(r^{(\ell)}\right) = (-1)^\ell \int_0^1 r^{(2\ell)}(x)\, r(x)\, \mathrm{d}x = (-1)^\ell\, \Psi_{2\ell}(r), \tag{2.13}$$

where

$$\Psi_\ell(r) = \int_0^1 r^{(\ell)}(x)\, r(x)\, \mathrm{d}x = E\left[r^{(\ell)}\left(F_0(X_1)\right)\right].$$

The equation above suggests a natural kernel-type estimator for $\Psi_\ell(r)$ as follows

$$\hat{\Psi}_\ell(g; L) = \frac{1}{n_1} \sum_{j=1}^{n_1} \left[\sum_{k=1}^{n_1} \frac{1}{n_1} L_g^{(\ell)}\left(F_{0n_0}(X_{1j}) - F_{0n_0}(X_{1k})\right)\right], \tag{2.14}$$

where L is a kernel function and g is a smoothing parameter called pilot bandwidth. Likewise in the previous section, this is not the only possibility and we could consider another estimator of $\Psi_\ell(r)$,

$$\tilde{\Psi}_\ell(g, h_0; L) = \frac{1}{n_1} \sum_{j=1}^{n_1} \left[\sum_{k=1}^{n_1} \frac{1}{n_1} L_g^{(\ell)}\left(\tilde{F}_{0h_0}(X_{1j}) - \tilde{F}_{0h_0}(X_{1k})\right)\right], \tag{2.15}$$

where F_{0n_0} in (2.14) is replaced by $\tilde{F}_{0h_0}$. Since the difference between both estimators decreases as h_0 tends to zero, it is expected to obtain the same theoretical results for both estimators. Therefore, we will only show theoretical results for $\hat{\Psi}_\ell(g; L)$.

We will obtain the asymptotic mean squared error of $\hat{\Psi}_\ell(g; L)$ under the following assumptions.

(R2) The relative density $r \in C^{(\ell+6)}(\mathbb{R})$.

(K9) The kernel L is a symmetric function of order 2, $L \in C^{(\ell+7)}(\mathbb{R})$ and satisfies that $(-1)^{\frac{\ell}{2}+2} L^{(\ell)}(0)\, d_L > 0$, $L^{(\ell)}(1) = L^{(\ell+1)}(1) = 0$, with $d_L = \int_{-\infty}^{\infty} x^2 L(x)\mathrm{d}x$.

(B4) $g = g_{n_1}$ is a positive-valued sequence of bandwidths satisfying

$$\lim_{n_1\to\infty} g = 0 \text{ and } \lim_{n_1\to\infty} n_1 g^{\max\{\alpha,\beta\}} = \infty,$$

where

$$\alpha = \frac{2(\ell+7)}{5}, \quad \beta = \frac{1}{2}(\ell+1) + 2.$$

Condition (R2) implies a smooth behaviour of r in the boundary of its support, contained in $[0,1]$. If this smoothness fails, the quantity $R\left(r^{(\ell)}\right)$ could be still estimated through its definition, using a kernel estimation for $r^{(\ell)}$ (see Hall and Marron (1987) for the one-sample problem setting). Condition (K9) can only hold for even ℓ. In fact, since $(-1)^{\left(\frac{\ell}{2}+2\right)}$ would be a complex number for odd ℓ, condition (K9) does not make sense for odd ℓ. Observe that in condition (B4) for even ℓ, $\max\{\alpha,\beta\}=\alpha$ for $\ell=0,2$ and $\max\{\alpha,\beta\}=\beta$ for $\ell=4,6,\ldots$

Theorem 2.3.1. *Assume conditions (S2), (B4), (D10), (K9) and (R2). Then it follows that*

$$\begin{aligned} MSE\left(\hat{\Psi}_\ell(g;L)\right) &= \left[\frac{1}{n_1 g^{\ell+1}} L^{(\ell)}(0)\left(1+\kappa^2\Psi_0(r)\right)+\frac{1}{2}d_L\Psi_{\ell+2}(r)g^2\right. \\ &\left.+O\left(g^4\right)+o\left(\left(n_0 g^{\ell+1}\right)^{-1}\right)\right]^2+\frac{2}{n_1^2 g^{2\ell+1}}\Psi_0(r)R\left(L^{(\ell)}\right) \\ &+o\left(\left(n_1^2 g^{2\ell+1}\right)^{-1}\right)+O\left(n_0^{-1}\right). \end{aligned} \tag{2.16}$$

Proof of Theorem 2.3.1. Below, we will briefly detail the steps followed to study the asymptotic behaviour of the mean squared error of $\hat{\Psi}_\ell(g;L)$ defined in (2.14). First of all, let us observe that

$$\hat{\Psi}_\ell(g;L)=\frac{1}{n_1}L_g^{(\ell)}(0)+\frac{1}{n_1^2}\sum_{j=1}^{n_1}\sum_{k=1,j\neq k}^{n_1}L_g^{(\ell)}\left(F_{0n_0}(X_{1j})-F_{0n_0}(X_{1k})\right),$$

which implies:

$$E\left[\hat{\Psi}_\ell(g;L)\right]=\frac{1}{n_1 g^{\ell+1}}L^{(\ell)}(0)+\left(1-\frac{1}{n_1}\right)E\left[L_g^{(\ell)}\left(F_{0n_0}(X_{11})-F_{0n_0}(X_{12})\right)\right].$$

Starting from the equation

$$\begin{aligned} &E\left[L_g^{(\ell)}\left(F_{0n_0}(X_{11})-F_{0n_0}(X_{12})\right)\right] \\ &=E\left[E\left[L_g^{(\ell)}\left(F_{0n_0}(X_{11})-F_{0n_0}(X_{12})\right)/X_{01},\ldots,X_{0n_0}\right]\right] \\ &=E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}L_g^{(\ell)}\left(F_{0n}(x_1)-F_{0n_0}(x_2)\right)f_1(x_1)f_1(x_2)\,\mathrm{d}x_1\mathrm{d}x_2\right] \\ &=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}E\left[L_g^{(\ell)}\left(F_{0n}(x_1)-F_{0n_0}(x_2)\right)\right]f_1(x_1)f_1(x_2)\,\mathrm{d}x_1\mathrm{d}x_2 \end{aligned}$$

and using a Taylor expansion, we have

$$E\left[L_g^{(\ell)}\left(F_{0n_0}(X_{11})-F_{0n_0}(X_{12})\right)\right]=\sum_{i=0}^{7}I_i, \tag{2.17}$$

where

$$I_0 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{g^{\ell+1}} L^{(\ell)} \left(\frac{F_0(x_1) - F_0(x_2)}{g} \right) f_1(x_1) f_1(x_2) \, \mathrm{d}x_1 \mathrm{d}x_2,$$

$$I_i = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{i! g^{\ell+i+1}} L^{(\ell+i)} \left(\frac{F_0(x_1) - F_0(x_2)}{g} \right) E\left[\left(F_{0n_0}(x_1) - F_0(x_1) - F_{0n_0}(x_2) + F_0(x_2) \right)^i \right] f_1(x_1) f_1(x_2) \, \mathrm{d}x_1 \mathrm{d}x_2,$$

$i = 1, \ldots, 6$,

$$I_7 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{7! g^{\ell+7+1}} E\left[L^{(\ell+7)}(\xi_n) \left(F_{0n_0}(x_1) - F_0(x_1) - F_{0n_0}(x_2) + F_0(x_2) \right)^7 \right] f_1(x_1) f_1(x_2) \, \mathrm{d}x_1 \mathrm{d}x_2$$

and ξ_{n_0} is a value between $\frac{F_0(x_1)-F_0(x_2)}{g}$ and $\frac{F_{0n_0}(x_1)-F_{0n_0}(x_2)}{g}$.

Now, consider the first term, I_0, in (2.17). Using changes of variable and a Taylor expansion, it is easy to show that $I_0 = \Psi_\ell(r) + \frac{1}{2} d_L \Psi_{\ell+2}(r) g^2 + O(g^4)$. Assume $x_1 > x_2$ and define $Z = \sum_{i=1}^{n_0} 1_{\{x_2 < X_{0i} \leq x_1\}}$. Then, the random variable Z has a $Bi(n_0, p)$ distribution with $p = F_0(x_1) - F_0(x_2)$ and mean equal to $\mu = n_0 p$. It is easy to show that, for $i = 1, \ldots, 6$,

$$I_i = 2 \int_{-\infty}^{\infty}\int_{x_2}^{\infty} \frac{1}{i! g^{\ell+i+1}} L^{(\ell+i)} \left(\frac{F_0(x_1) - F_0(x_2)}{g} \right) f_1(x_1) f_1(x_2) \frac{1}{n_0^i} \mu_i(Z) \, \mathrm{d}x_1 \mathrm{d}x_2, \quad (2.18)$$

where

$$\mu_r(Z) = E\left[(Z - E[Z])^r \right] = \sum_{j=0}^{r} (-1)^j \binom{r}{j} m_{r-j} \mu^j \quad (2.19)$$

and m_k denotes the kth non-central moment of Z, $m_k = E\left[Z^k\right]$.

Let m_k^{df} be the k-th descending factorial moment of a discrete distribution, let say Z,

$$m_k^{df} = E\left[Z!/(Z-k)! \right].$$

Since $Z!/(Z-k)! = \sum_{j=0}^{k} s(k,j) Z^j$ with $s(k,j)$'s the Stirling numbers of first kind (see Appendix A), it follows that

$$m_k^{df} = \sum_{j=0}^{k} s(k,j) m_j.$$

Similarly, since $Z^k = \sum_{j=0}^{k} \frac{S(k,j) Z!}{(Z-j)!}$ where the $S(k,j)$'s denote the Stirling numbers of second kind (see Appendix A), it follows that

$$m_k = \sum_{j=0}^{k} S(k,j) m_j^{df}. \quad (2.20)$$

Now, let $FMG(t)$ be the factorial moment generating function of Z, $FMG(t) = E\left[t^Z\right]$. Since it is satisfied that

$$FMG(1+t) = 1 + \sum_{k\leq 1} \frac{m_k^{df} t^k}{k!}, \tag{2.21}$$

the factorials moments of Z can be obtained straightforwardly from the corresponding factorial moment generating function associated to Z.

In particular, for Z, a binomial distribution with parameters n_0 and p, there exists a closed expression for $FMG(t)$ given in terms of n_0 and p:

$$FMG(t) = (1+pt)^{n_0}.$$

Therefore, based on this expression and the relationship existing between $FMG(1+t)$ and the descending factorial moments of Z (see (2.21)), it is easily proved that

$$m_k^{df} = \frac{n_0! p^k}{(n_0-k)!}.$$

Consequently, replacing m_j^{df} in (2.20) by the equation above, the non central moments of Z can be rewritten as follows

$$m_k = \sum_{j=0}^{k} \frac{S(k,j)\, n_0! p^j}{(n_0-j)!}, \tag{2.22}$$

where

$$S(k,j) = \frac{\sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^\ell (k-\ell)^j}{k!}.$$

Therefore, based on the previous formulas (2.18), (2.19) and (2.22), and after straightforward calculations, it is easy to show that:

$$\begin{aligned}
I_1 &= 0, \\
I_2 &= \frac{1}{n_0 g^{\ell+1}} \Psi_0(r) L^{(\ell)}(0) + O\left(\frac{1}{n_0 g^\ell}\right), \qquad (2.23) \\
I_i &= O\left(\frac{1}{n_0^2 g^{\ell+2}}\right), \text{ for } i = 3, 4, \\
I_i &= O\left(\frac{1}{n_0^3 g^{\ell+3}}\right), \text{ for } i = 5, 6.
\end{aligned}$$

Coming back to the last term in (2.17) and using the Dvoretzky-Kiefer-Wolfowitz inequality and condition (K9), it is easy to show that $I_7 = O\left(\frac{1}{n_0^{\frac{7}{2}} g^{\ell+8}}\right)$. Therefore, using condition (B4), it follows that

$$\begin{aligned}
E\left[\hat{\Psi}_\ell(g;L)\right] &= \Psi_\ell(r) + \frac{1}{2} d_L \Psi_{\ell+2}(r) g^2 + \frac{1}{n_1 g^{\ell+1}} L^{(\ell)}(0) + \frac{1}{n_0 g^{\ell+1}} L^{(\ell)}(0)\, \Psi_0(r) \\
&\quad + O\left(g^4\right) + o\left(\frac{1}{n_0 g^{\ell+1}}\right).
\end{aligned}$$

For the sake of brevity we will only prove (2.23). Similar arguments can be used to handle the terms I_3, I_4, I_5 and I_6.

First of all, since the integrand is a symmetric function in (x_1, x_2),

$$I_2 = \int_{-\infty}^{\infty} \int_{x_2}^{\infty} \frac{1}{g^{\ell+1+2}} L^{(\ell+2)} \left(\frac{F_0(x_1) - F_0(x_2)}{g} \right) f_1(x_1) f_1(x_2) \frac{1}{n^2} \mu_2(Z) \, \mathrm{d}x_1 \mathrm{d}x_2,$$

where $\mu_2(Z) = n_0 (F_0(x_1) - F_0(x_2))(1 - F_0(x_1) + F_0(x_2))$, which implies that

$$\begin{aligned} I_2 &= \frac{1}{n_0 g^{\ell+1+2}} \int_{-\infty}^{\infty} \int_{x_2}^{\infty} L^{(\ell+2)} \left(\frac{F_0(x_1) - F_0(x_2)}{g} \right) f_1(x_1) f_1(x_2) \\ &\quad (F_0(x_1) - F_0(x_2))(1 - F_0(x_1) + F_0(x_2)) \, \mathrm{d}x_1 \mathrm{d}x_2. \end{aligned}$$

Now, standard changes of variable give

$$\begin{aligned} I_2 &= \frac{1}{n_0 g^{\ell+1+2}} \int_0^1 \int_v^1 L^{(\ell+2)} \left(\frac{u-v}{g} \right) (u-v)(1-u+v) r(u) r(v) \, \mathrm{d}u \mathrm{d}v \\ &= \frac{1}{n_0 g^{\ell+1+2}} \int_0^1 \int_0^{\frac{1-v}{g}} L^{(\ell+2)}(x)(gx)(1-gx) r(v+gx) r(v) \, g \mathrm{d}x \mathrm{d}v. \end{aligned}$$

A Taylor expansion for $r(v+gx)$ and Fubini's theorem imply

$$I_2 = \frac{1}{n_0 g^{\ell+1}} \int_0^{\frac{1}{g}} \int_0^{1-gx} L^{(\ell+2)}(x) x r^2(v) \, \mathrm{d}v \mathrm{d}x + O\left(\frac{1}{n_0 g^{\ell}} \right).$$

Now, choosing n_0 such that $g = g_{n_0} < 1$ and using condition (K9),

$$I_2 = \frac{1}{n_0 g^{\ell+1}} \int_0^1 \int_0^1 L^{(\ell+2)}(x) x r^2(v) \, \mathrm{d}v \mathrm{d}x + O\left(\frac{1}{n_0 g^{\ell}} \right).$$

On the other hand, condition (K9) implies that

$$\begin{aligned} \int_0^1 x L^{(\ell+2)}(x) \, \mathrm{d}x &= \left[x L^{(\ell+1)}(x) \right]_0^1 - \int_0^1 L^{(\ell+1)}(x) \, \mathrm{d}x \\ &= L^{(\ell+1)}(1) - \left[L^{(\ell)}(1) - L^{(\ell)}(0) \right] = L^{(\ell)}(0). \end{aligned}$$

Thus,

$$I_2 = \frac{1}{n_0 g^{\ell+1}} \Psi_0(r) L^{(\ell)}(0) + O\left(\frac{1}{n_0 g^{\ell}} \right),$$

which proves (2.23).

In order to study the variance of $\hat{\Psi}_\ell(g; L)$, note that

$$Var\left[\hat{\Psi}_\ell(g; L) \right] = \sum_{i=1}^{3} c_{n_1, i} V_{\ell, i}, \tag{2.24}$$

where

$$\begin{aligned} c_{n_1,1} &= \frac{2(n_1-1)}{n_1^3}, \\ c_{n_1,2} &= \frac{4(n_1-1)(n_1-2)}{n_1^3}, \\ c_{n_1,3} &= \frac{(n_1-1)(n_1-2)(n_1-3)}{n_1^3}, \end{aligned}$$

$$V_{\ell,1} = Var\left[L_g^{(\ell)}\left(F_{0n_0}(X_{11})-F_{0n_0}(X_{12})\right)\right], \tag{2.25}$$

$$V_{\ell,2} = Cov\left[L_g^{(\ell)}\left(F_{0n_0}(X_{11})-F_{0n_0}(X_{12})\right), L_g^{(\ell)}\left(F_{0n_0}(X_{12})-F_{0n_0}(X_{13})\right)\right], \tag{2.26}$$

$$V_{\ell,3} = Cov\left[L_g^{(\ell)}\left(F_{0n_0}(X_{11})-F_{0n_0}(X_{12})\right), L_g^{(\ell)}\left(F_{0n_0}(X_{13})-F_{0n_0}(X_{14})\right)\right]. \tag{2.27}$$

Therefore, in order to get an asymptotic expression for the variance of $\hat{\Psi}_\ell(g;L)$, we will start getting asymptotic expressions for the terms (2.25), (2.26) and (2.27) in (2.24). To deal with the term (2.25), we will use

$$\begin{aligned} V_{\ell,1} &= E\left[\left(L_g^{(\ell)}\left(F_{0n_0}(X_{11})-F_{0n_0}(X_{12})\right)\right)^2\right] \\ &\quad -E^2\left[L_g^{(\ell)}\left(F_{0n_0}(X_{11})-F_{0n_0}(X_{12})\right)\right] \end{aligned} \tag{2.28}$$

and study separately each term in the right hand side of (2.28). Note that the expectation of $L_g^{(\ell)}\left(F_{0n_0}(X_{11})-F_{0n_0}(X_{12})\right)$ has been already studied when dealing with the expectation of $\hat{\Psi}_\ell(g;L)$. Next we study the first term in the right hand side of (2.28). Using a Taylor expansion around $\frac{F_0(x)-F_0(y)}{h}$,

$$\begin{aligned} L^{(\ell)}\left(\frac{F_{0n_0}(x)-F_{0n_0}(y)}{h}\right) &= L^{(\ell)}\left(\frac{F_0(x)-F_0(y)}{h}\right) + L^{(\ell+1)}\left(\frac{F_0(x)-F_0(y)}{h}\right) \\ &\quad \left(\frac{F_{0n_0}(x)-F_0(x)-(F_{0n_0}(y)-F_0(y))}{h}\right) \\ &\quad +\frac{1}{2}L^{(\ell+2)}(\xi_{n_0})\left(\frac{F_{0n_0}(x)-F_0(x)-(F_{0n_0}(y)-F_0(y))}{h}\right)^2, \end{aligned}$$

where ξ_{n_0} is a value between $\frac{F_{0n_0}(x)-F_{0n_0}(y)}{h}$ and $\frac{F_0(x)-F_0(y)}{h}$. Then, the term:

$$\begin{aligned} &E\left[\left(L_g^{(\ell)}\left(F_{0n_0}(X_{11})-F_{0n_0}(X_{12})\right)\right)^2\right] \\ &= E\left[E\left[\left(L_g^{(\ell)}\left(F_{0n_0}(X_{11})-F_{0n_0}(X_{12})\right)\right)^2 / X_{01},...,X_{0n_0}\right]\right] \\ &= E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} L_g^{(\ell)^2}\left(F_{0n_0}(x)-F_{0n_0}(y)\right) f_1(x) f_1(y)\, dx dy\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\left[L_g^{(\ell)^2}\left(F_{0n_0}(x)-F_{0n_0}(y)\right)\right] f_1(x) f_1(y)\, dx dy \end{aligned}$$

can be decomposed in a sum of six terms that can be bounded easily. The first term in that decomposition can be rewritten as $\frac{1}{g^{2\ell+1}}\Psi_0(r)R\left(L^{(\ell)}\right)+o\left(\frac{1}{g^{2\ell+1}}\right)$ after applying some changes of variable and a Taylor expansion. The other terms can be easily bounded using the Dvoretzky-Kiefer-Wolfowitz inequality and standard changes of variable. These bounds and condition (B4) prove that the order of these terms is $o\left(\frac{1}{g^{2\ell+1}}\right)$. Consequently,

$$\begin{aligned} V_{\ell,1} &= \frac{1}{g^{2\ell+1}}\Psi_0(r)R\left(L^{(\ell)}\right)+o\left(\frac{1}{g^{2\ell+1}}\right)-\left(\Psi_\ell(r)+o\left(1\right)\right)^2 \\ &= \frac{1}{g^{2\ell+1}}\Psi_0(r)R\left(L^{(\ell)}\right)-\Psi_\ell^2(r)+o\left(\frac{1}{g^{2\ell+1}}\right)+o\left(1\right). \end{aligned}$$

The term (2.26) can be handled using

$$\begin{aligned} V_{\ell,2} &= E\left[L_g^{(\ell)}\left(F_{0n_0}\left(X_{11}\right)-F_{0n_0}\left(X_{12}\right)\right)L_g^{(\ell)}\left(F_{0n_0}\left(X_{12}\right)-F_{0n_0}\left(X_{13}\right)\right)\right] \\ &\quad -E^2\left[L_g^{(\ell)}\left(F_{0n_0}\left(X_{11}\right)-F_{0n_0}\left(X_{12}\right)\right)\right]. \end{aligned} \tag{2.29}$$

As for (2.28), it is only needed to study the first term in the right hand side of (2.29).

Note that

$$\begin{aligned} & E\left[L_g^{(\ell)}\left(F_{0n_0}\left(X_{11}\right)-F_{0n_0}\left(X_{12}\right)\right)L_g^{(\ell)}\left(F_{0n_0}\left(X_{12}\right)-F_{0n_0}\left(X_{13}\right)\right)\right] \\ &= E\left[E\left[L_g^{(\ell)}\left(F_{0n_0}\left(X_{11}\right)-F_{0n_0}\left(X_{12}\right)\right)L_g^{(\ell)}\left(F_{0n_0}\left(X_{12}\right)-F_{0n_0}\left(X_{13}\right)\right)/X_{01,\ldots,}X_{0n_0}\right]\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\left[L_g^{(\ell)}\left(F_{0n_0}\left(y\right)-F_{0n_0}\left(z\right)\right)L_g^{(\ell)}\left(F_{0n_0}\left(z\right)-F_{0n_0}\left(t\right)\right)\right] \\ & f_1\left(y\right)f_1\left(z\right)f_1\left(t\right)\mathrm{d}y\mathrm{d}z\mathrm{d}t. \end{aligned}$$

Taylor expansions, changes of variable, the Cauchy-Schwarz inequality and the Dvoretzky-Kiefer-Wolfowitz inequality, give:

$$\begin{aligned} & E\left[L_g^{(\ell)}\left(F_{0n_0}\left(X_{11}\right)-F_{0n_0}\left(X_{12}\right)\right)L_g^{(\ell)}\left(F_{0n_0}\left(X_{12}\right)-F_{0n_0}\left(X_{13}\right)\right)\right] \\ &= \int_0^1 r^{(\ell)^2}\left(z\right)r\left(z\right)\mathrm{d}z+O\left(\frac{1}{n_0}\right)+O\left(\frac{1}{n_0^2}\right)+O\left(\frac{1}{n_0^3}\right)+O\left(\frac{1}{n_0^4 g^{2((\ell+1)+4)}}\right). \end{aligned}$$

Now, condition (B4) implies

$$\begin{aligned} & E\left[L_g^{(\ell)}\left(F_{0n_0}\left(X_{11}\right)-F_{0n_0}\left(X_{12}\right)\right)L_g^{(\ell)}\left(F_{0n_0}\left(X_{12}\right)-F_{0n_0}\left(X_{13}\right)\right)\right] \\ &= \int_0^1 r^{(\ell)^2}\left(z\right)r\left(z\right)\mathrm{d}z+o\left(1\right). \end{aligned}$$

Consequently, using (2.29), $V_{\ell,2}=O(1)$.

To study the term $V_{\ell,3}$ in (2.27), let us define

$$A_\ell=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[L_g^{(\ell)}\left(F_{0n_0}\left(y\right)-F_{0n_0}\left(z\right)\right)-L_g^{(\ell)}\left(F_0\left(y\right)-F_0\left(z\right)\right)\right]f_1\left(y\right)f_1\left(z\right)\mathrm{d}y\mathrm{d}z.$$

It is easy to show that:

$$V_{\ell,3} = Var\left(A_\ell\right).$$

Now a Taylor expansion gives

$$Var\left(A_\ell\right) = \sum_{k=1}^{N} Var\left(T_k\right) + \sum_{k=1}^{N}\sum_{\substack{\ell=1\\ k\neq\ell}}^{N} Cov\left(T_k, T_\ell\right), \tag{2.30}$$

where

$$A_\ell = \sum_{k=1}^{N} T_k,$$

$$\begin{aligned} T_k &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{k! g^{\ell+1}} L^{(\ell+k)}\left(\frac{F_0\left(y\right)-F_0\left(z\right)}{g}\right) f_1\left(y\right) f_1\left(z\right) \\ &\quad \left(\frac{F_{0n_0}\left(y\right)-F_{0n_0}\left(z\right)-\left(F_0\left(y\right)-F_0\left(z\right)\right)}{g}\right)^k \mathrm{d}y\mathrm{d}z, \text{ for } k=1,\ldots,N-1, \end{aligned}$$

and

$$\begin{aligned} T_N &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{N! g^{\ell+1}} L^{(\ell+N)}\left(\xi_n\right) f_1\left(y\right) f_1\left(z\right) \\ &\quad \left(\frac{F_{0n_0}\left(y\right)-F_{0n_0}\left(z\right)-\left(F_0\left(y\right)-F_0\left(z\right)\right)}{g}\right)^N \mathrm{d}y\mathrm{d}z, \end{aligned}$$

for some positive integer N. We will only study in detail each one of the first N summands in (2.30). The rest of them will be easily bounded using the Cauchy-Schwarz inequality and the bounds obtained for the first N terms.

Now the variance of T_k is studied. First of all, note that

$$\begin{aligned} Var(T_k) \leq\; & E\left[{T_k}^2\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{1}{k! g^{\ell+k+1}}\right)^2 f_1(y_1)f_1(z_1)f_1(y_2)f_1(z_2) \\ & L^{(\ell+k)}\left(\frac{F_0(y_1)-F_0(z_1)}{g}\right) L^{(\ell+k)}\left(\frac{F_0(y_2)-F_0(z_2)}{g}\right) \\ & h_k(y_1,z_1,y_2,z_2)\mathrm{d}y_1\mathrm{d}z_1\mathrm{d}y_2\mathrm{d}z_2, \end{aligned}$$

where

$$\begin{aligned} h_k(y_1,z_1,y_2,z_2) &= E\Big\{\left[F_{0n_0}\left(y_1\right)-F_{0n_0}\left(z_1\right)-\left(F_0\left(y_1\right)-F_0\left(z_1\right)\right)\right]^k \\ &\quad \left[F_{0n_0}\left(y_2\right)-F_{0n_0}\left(z_2\right)-\left(F_0\left(y_2\right)-F_0\left(z_2\right)\right)\right]^k\Big\}. \end{aligned} \tag{2.31}$$

Using changes of variable we can rewrite $E\left[T_k^2\right]$ as follows:

$$\begin{aligned} E\left[{T_k}^2\right] &= \int_0^1\int_0^1\int_0^1\int_0^1 \left(\frac{1}{k!}\right)^2 r(s_1)r(t_1)r(s_2)r(t_2)L_g^{(\ell+k)}\left(s_1-t_1\right) \\ &\quad L_g^{(\ell+k)}\left(s_2-t_2\right) h_k(F_0^{-1}(s_1),F_0^{-1}(t_1),F_0^{-1}(s_2),F_0^{-1}(t_2))\mathrm{d}s_1\mathrm{d}t_1\mathrm{d}s_2\mathrm{d}t_2 \\ &= \int_0^1\int_{s_2-1}^{s_2}\int_0^1\int_{s_1-1}^{s_1} \left(\frac{1}{k!}\right)^2 r(s_1)r(s_1-u_1)r(s_2)r(s_2-u_2)L_g^{(\ell+k)}\left(u_1\right) \\ &\quad L_g^{(\ell+k)}\left(u_2\right) h_k(F_0^{-1}(s_1),F_0^{-1}(s_1-u_1),F_0^{-1}(s_2),F_0^{-1}(s_2-u_2))\mathrm{d}u_1\mathrm{d}s_1\mathrm{d}u_2\mathrm{d}s_2. \end{aligned}$$

Note that closed expressions for h_k can be obtained using the expressions for the moments of order $\mathbf{r}=(r_1,r_2,r_3,r_4,r_5)$ of $\mathbf{Z}$, a random variable with multinomial distribution with parameters $(n_0;p_1,p_2,p_3,p_4,p_5)$. Based on these expressions, condition (R2) and using integration by parts we can rewrite $E\left[{T_k}^2\right]$ as follows:

$$\begin{aligned} E\left[{T_k}^2\right] &= \int_0^1\int_{s_2-1}^{s_2}\int_0^1\int_{s_1-1}^{s_1} \left(\frac{1}{k!}\right)^2 L_g\left(u_1\right)L_g\left(u_2\right) \\ &\quad \frac{\partial^{2(\ell+k)}}{\partial u_1^{\ell+k}\partial u_2^{\ell+k}}(\tilde{h}_k(u_1,s_1,u_2,s_2))\mathrm{d}u_1\mathrm{d}s_1\mathrm{d}u_2\mathrm{d}s_2, \end{aligned}$$

where

$$\begin{aligned} \tilde{h}_k(u_1,s_1,u_2,s_2) &= r(s_1)r(s_1-u_1)r(s_2)r(s_2-u_2) \\ &\quad \cdot h_k(F_0^{-1}(s_1),F_0^{-1}(s_1-u_1),F_0^{-1}(s_2),F_0^{-1}(s_2-u_2)). \end{aligned}$$

On the other hand Lemma 2.3.2 implies that $\sup_{\mathbf{z}\in\mathbb{R}^4}|h_k(\mathbf{z})|=O\left(\frac{1}{n_0^k}\right)$. This result and condition (R2) allow us to conclude that $Var\left(T_k\right)\le E\left[T_k^2\right]=O\left(\frac{1}{n_0^k}\right)$, for $1\le k<N$, which implies that $Var\left(T_k\right)=o\left(\frac{1}{n_0}\right)$, for $2\le k<N$.

A Taylor expansion of order $N=6$, gives $Var\left(T_6\right)=O\left(\frac{1}{n_0^N g^{2(N+\ell+1)}}\right)$, which using condition (B4), proves $Var\left(T_6\right)=o\left(\frac{1}{n_0}\right)$. Consequently,

$$Var\left[\hat{\Psi}_\ell(g;L)\right] = \frac{2}{n_1^2 g^{2\ell+1}}\Psi_0(r)R\left(L^{(\ell)}\right)+O\left(\frac{1}{n_0}\right)+o\left(\frac{1}{n_1^2 g^{2\ell+1}}\right),$$

which concludes the proof. □

Lemma 2.3.2. *For $h_k(y_1,z_1,y_2,z_2)$ defined previously in (2.31), we have that:*

$$\sup_{\mathbf{z}\in\mathbb{R}^4}|h_k(\boldsymbol{z})|=O\left(\frac{1}{n_0^k}\right).$$

Proof of Lemma 2.3.2. We will show below how closed expressions for h_k can be obtained from the multinomial moments of order $\mathbf{r}=(r_1,r_2,r_3,r_4,r_5)$, $m_{\mathbf{r}}\left(\mathbf{Z}\right)=E\left[\prod_{i=1}^5 N_i^{r_i}\right]$, where the N_i's for $i = 1, \ldots, 5$ denote the five components of the multinomial variable $\mathbf{Z}$, i.e. $\mathbf{Z} = (N_1, N_2, N_3, N_4, N_5)$. Since the way of proceeding will be the same for all k, for the sake of brevity, we will next consider only some values of k, $k = 1, 2$.

First of all, note that the multinomial moments introduced above, $m_{\mathbf{r}}(\mathbf{Z})$, can be rewritten as follows:

$$m_{\mathbf{r}}(\mathbf{Z}) = \sum_{i_1=0}^{r_1} \cdots \sum_{i_5=0}^{r_5} S(r_1, i_1) \cdots S(r_5, i_5)\, m_{(\mathbf{i})}(\mathbf{Z}),$$

where

$$\begin{aligned}
m_{(\mathbf{i})}(\mathbf{Z}) &= m_{((i_1,i_2,i_3,i_4,i_5))}(\mathbf{Z}) = E\left[\prod_{i=1}^{5} N_i^{(r_i)}\right] = n_0^{\left(\sum_{i=1}^{5} r_i\right)} \prod_{i=1}^{5} p_i^{r_i}, \\
N_i^{(r_i)} &= N_i\,(N_i - 1) \cdots (N_i - r_i + 1), \\
n_0^{\left(\sum_{i=1}^{5} r_i\right)} &= n_0\,(n_0 - 1) \cdots \left(n_0 - \sum_{i=1}^{5} r_i + 1\right)
\end{aligned}$$

and

$$S(r_j, i_j) = \frac{\sum_{k=0}^{i_j} \binom{i_j}{k} (-1)^k (i_j - k)^{r_j}}{i_j!}.$$

Using the expressions above, we next find the value of $h_1(y_1, z_1, y_2, z_2)$ for the six possible arrangements of z_1, z_2, y_1 and y_2 given below:

(1) $z_1 < y_1 \leq z_2 < y_2$,
(2) $z_2 < z_1 \leq y_1 < y_2$,
(3) $z_1 < z_2 < y_1 < y_2$,
(4) $y_1 < z_1 \leq y_2 < z_2$,
(5) $y_2 < y_1 \leq z_1 < z_2$,
(6) $y_1 < y_2 < z_1 < z_2$.

Due to the symmetry property, it is enough to study the cases (1)-(3). Let start with scenario (1). In this case, it follows that:

$$h_1(y_1, z_1, y_2, z_2) = \frac{1}{n_0^2} E[(N_2 - \mu_2)(N_4 - \mu_4)],$$

where

$$\begin{aligned}
&N_2 = \sum_{j=1}^{n_0} 1_{\{z_1 \leq X_{0j} \leq y_1\}}, \\
&E[N_2] = \mu_2 = n_0 p_2^{(1)}, \text{ with } p_2^{(1)} = P(z_1 \leq X_{0j} \leq y_1), \\
&N_4 = \sum_{j=1}^{n_0} 1_{\{z_2 \leq X_{0j} \leq y_2\}}, \\
&E[N_4] = \mu_4 = n_0 p_4^{(1)}, \text{ with } p_4^{(1)} = P(z_2 \leq X_{0j} \leq y_2).
\end{aligned}$$

Consequently, when (1) holds, it follows that:

$$\begin{aligned} h_1(y_1, z_1, y_2, z_2) &= \frac{1}{n_0^2}\{E[N_2N_4] - \mu_2\mu_4\} \\ &= \frac{1}{n_0^2}\left\{n_0(n_0-1)p_2^{(1)}p_4^{(1)} - n_0^2p_2^{(1)}p_4^{(1)}\right\} \\ &= \frac{1}{n_0}\left\{(n_0-1)p_2^{(1)}p_4^{(1)} - n_0p_2^{(1)}p_4^{(1)}\right\} \\ &= -\frac{1}{n_0}p_2^{(1)}p_4^{(1)}. \end{aligned}$$

Proceeding in a similar way when (2) holds, it follows that

$$h_1(y_1, z_1, y_2, z_2) = \frac{1}{n_0^2}E[(N_3 - \mu_3)(N_4 - \mu_4 + N_3 - \mu_3 + N_2 - \mu_2)],$$

where

$$\begin{aligned} &N_2 = \sum_{j=1}^{n_0} 1_{\{z_2 \le X_{0j} \le z_1\}}, \\ &E[N_2] = \mu_2 = n_0p_2^{(2)}, \text{ with } p_2^{(2)} = P\left(z_2 \le X_{0j} \le z_1\right), \\ &N_3 = \sum_{j=1}^{n_0} 1_{\{z_1 \le X_{0j} \le y_1\}}, \\ &E[N_3] = \mu_2 = n_0p_3^{(2)}, \text{ with } p_3^{(2)} = P\left(z_1 \le X_{0j} \le y_1\right), \\ &N_4 = \sum_{j=1}^{n_0} 1_{\{y_1 \le X_{0j} \le y_2\}}, \\ &E[N_4] = \mu_4 = n_0p_4^{(2)}, \text{ with } p_4^{(2)} = P\left(y_1 \le X_{0j} \le y_2\right). \end{aligned}$$

Therefore, under scenario (2), it follows that

$$\begin{aligned} h_1(y_1, z_1, y_2, z_2) &= \frac{1}{n_0^2}\left\{E[N_3N_4] + E[N_3^2] + E[N_2N_3] - \mu_3\mu_4 - \mu_3^2 - \mu_2\mu_3\right\} \\ &= \frac{1}{n_0^2}\Big\{n_0(n_0-1)p_3^{(2)}p_4^{(2)} + n_0p_3^{(2)}((n_0-1)p_3^{(2)}+1) + n_0(n_0-1)p_2^{(2)}p_3^{(2)} \\ &\quad -n_0^2p_3^{(2)}p_4^{(2)} - n_0^2p_2^{(2)}p_3^{(2)} - n_0^2\left(p_3^{(2)}\right)^2\Big\} \\ &= -\frac{1}{n_0}p_3^{(2)}(1 - p_4^{(2)} - p_3^{(2)} - p_2^{(2)}). \end{aligned}$$

Finally, when scenario (3) holds, it follows that

$$h_1(y_1, z_1, y_2, z_2) = \frac{1}{n_0^2}E[(N_3 - \mu_3 + N_2 - \mu_2)(N_4 - \mu_4 + N_3 - \mu_3)],$$

where

$$\begin{aligned}
N_2 &= \sum_{j=1}^{n_0} 1_{\{z_1 \leq X_{0j} \leq z_2\}}, \\
E[N_2] &= \mu_2 = n_0 p_2^{(3)}, \text{ with } p_2^{(3)} = P\left(z_1 \leq X_{0j} \leq z_2\right), \\
N_3 &= \sum_{j=1}^{n_0} 1_{\{z_2 \leq X_{0j} \leq y_1\}}, \\
E[N_3] &= \mu_2 = n_0 p_3^{(3)}, \text{ with } p_3^{(3)} = P\left(z_2 \leq X_{0j} \leq y_1\right), \\
N_4 &= \sum_{j=1}^{n_0} 1_{\{y_1 \leq X_{0j} \leq y_2\}}, \\
E[N_4] &= \mu_4 = n_0 p_4^{(3)}, \text{ with } p_4^{(3)} = P\left(y_1 \leq X_{0j} \leq y_2\right).
\end{aligned}$$

After some simple algebra, it can be shown that

$$h_1(y_1, z_1, y_2, z_2) = \frac{1}{n_0}\left\{p_3^{(3)} - (p_3^{(3)} + p_2^{(3)})(p_4^{(3)} + p_3^{(3)})\right\}$$

when (3) holds. As a consequence, $\sup_{\mathbf{z} \in \Re^4} |h_1(\mathbf{z})| = O\left(\frac{1}{n_0}\right)$.

The value of $h_2(y_1, z_1, y_2, z_2)$ under the different arrangements of z_1, z_2, y_1 and y_2, can be obtained in a similar way. After simple algebra it follows that:

$$\begin{aligned}
h_2(y_1, z_1, y_2, z_2) = \frac{1}{n_0^3}&\left\{p_2^{(1)} p_4^{(1)}(n_0 - 1) + \left(p_2^{(1)}\right)^2 p_4^{(1)}(2 - n_0)\right. \\
&\left.+ p_2^{(1)}\left(p_4^{(1)}\right)^2(2 - n_0) + \left(p_2^{(1)}\right)^2\left(p_4^{(1)}\right)^2(-6 + 3n_0)\right\}
\end{aligned}$$

if (1) holds,

$$\begin{aligned}
h_2(y_1, z_1, y_2, z_2) = \frac{1}{n_0^3}&\left\{p_3^{(2)} + (n_0 - 3)p_2^{(2)} p_3^{(2)} + (n_0 - 3)p_3^{(2)} p_4^{(2)}\right. \\
&+ (3n_0 - 7)\left(p_3^{(2)}\right)^2 - (2n_0 - 4)p_2^{(2)} p_3^{(2)} p_4^{(2)} \\
&- (7n_0 - 14)p_2^{(2)}\left(p_3^{(2)}\right)^2 - (n_0 - 2)\left(p_2^{(2)}\right)^2 p_3^{(2)} \\
&- (n_0 - 2)p_3^{(2)}\left(p_4^{(2)}\right)^2 - (7n_0 - 14)\left(p_3^{(2)}\right)^2 p_4^{(2)} \\
&- (6n_0 - 12)\left(p_3^{(2)}\right)^3 + (6n_0 - 12)p_2^{(2)}\left(p_3^{(2)}\right)^3 \\
&+ (6n_0 - 12)\left(p_3^{(2)}\right)^3 p_4^{(2)} + (3n_0 - 6)\left(p_2^{(2)}\right)^2\left(p_3^{(2)}\right)^2 \\
&+ (3n_0 - 6)\left(p_3^{(2)}\right)^2\left(p_4^{(2)}\right)^2 + (6n_0 - 12)p_2^{(2)}\left(p_3^{(2)}\right)^2 p_4^{(2)} \\
&\left.+ (3n_0 - 6)\left(p_3^{(2)}\right)^4\right\}
\end{aligned}$$

if (2) holds and

$$
\begin{aligned}
h_2(y_1,z_1,y_2,z_2) \;=\; & \frac{1}{n_0^3}\Big\{p_3^{(3)}+(n_0-3)p_2^{(3)}p_3^{(3)}+(n_0-1)p_2^{(3)}p_4^{(3)} \\
& (n_0-3)p_3^{(3)}p_4^{(3)}-(8n_0-16)p_2^{(3)}p_3^{(3)}p_4^{(3)} \\
& +(3n_0-7)\left(p_3^{(3)}\right)^2-(6n_0-12)\left(p_3^{(3)}\right)^3+(3n_0-6)\left(p_3^{(3)}\right)^4 \\
& -(7n_0-14)p_2^{(3)}\left(p_3^{(3)}\right)^2-(n_0-2)\left(p_2^{(3)}\right)^2p_3^{(3)} \\
& +(6n_0-12)p_2^{(3)}\left(p_3^{(3)}\right)^3-(n_0-2)p_2^{(3)}\left(p_4^{(3)}\right)^2 \\
& -(n_0-2)\left(p_2^{(3)}\right)^2p_4^{(3)}-(n_0-2)p_3^{(3)}\left(p_4^{(3)}\right)^2 \\
& -(7n_0-14)\left(p_3^{(3)}\right)^2p_4^{(3)}+(6n_0-12)\left(p_3^{(3)}\right)^3p_4^{(3)} \\
& +(6n_0-12)p_2^{(3)}p_3^{(3)}\left(p_4^{(3)}\right)^2 \\
& +(12n_0-24)p_2^{(3)}\left(p_3^{(3)}\right)^2p_4^{(3)} \\
& +(6n_0-12)\left(p_2^{(3)}\right)^2p_3^{(3)}p_4^{(3)} \\
& +(3n_0-6)\left(p_2^{(3)}\right)^2\left(p_3^{(3)}\right)^2+(3n_0-6)\left(p_2^{(3)}\right)^2\left(p_4^{(3)}\right)^2 \\
& +(3n_0-6)\left(p_3^{(3)}\right)^2\left(p_4^{(3)}\right)^2\Big\}
\end{aligned}
$$

if (3) holds. Consequently, $\sup_{\mathbf{z}\in\Re^4}|h_2(\mathbf{z})|=O\left(\frac{1}{n_0^2}\right)$. □

Remark 2.3.1. If equation (2.17) is replaced by a three-term Taylor expansion $\sum_{i=1}^2 I_i+I_3^*$, where

$$
\begin{aligned}
I_3^* \;=\; & \frac{1}{3!g^{\ell+4}}\iint E\left[L^{(\ell+3)}(\zeta_n)\left(F_{0n_0}(x_1)-F_0(x_1)-F_{0n_0}(x_2)+F_0(x_2)\right)^3\right] \\
& f_1(x_1)f_1(x_2)\mathrm{d}x_1\mathrm{d}x_2
\end{aligned}
$$

and ζ_{n_0} is a value between $\frac{F_0(x_1)-F_0(x_2)}{g}$ and $\frac{F_{0n_0}(x_1)-F_{0n_0}(x_2)}{g}$, then $I_3^*=O\left(\frac{1}{n_0^{\frac{3}{2}}g^{\ell+4}}\right)$ and we would have to ask for the condition $n_0g^6\to\infty$ to conclude that $I_3^*=o\left(\frac{1}{n_0g^{\ell+1}}\right)$. However, this condition is very restrictive because it is not satisfied by the optimal bandwidth g_ℓ with $\ell=0,2$, which is $g_\ell\sim n_0^{-\frac{1}{\ell+3}}$. We could consider $\sum_{i=1}^3 I_i+I_4^*$ and then we would need to ask for the condition $n_0g^4\to\infty$. However, this condition is not satisfied by g_ℓ with $\ell=0$. In fact, it follows that $n_0g_\ell^4\to 0$ if $\ell=0$ and $n_0g_\ell^4\to\infty$ if $\ell=2,4,\ldots$ Something similar happens when we consider $\sum_{i=1}^4 I_i+I_5^*$ or $\sum_{i=1}^5 I_i+I_6^*$, i.e., the condition required in g, it is not satisfied by the optimal bandwidth when $\ell=0$. Only when we stop in I_7^*, the required condition, $n_0g^{\frac{14}{5}}\to\infty$, is satisfied for all even ℓ.

If equation (2.30) is reconsidered by the mean value theorem, and then we define $A_\ell = T_1^*$ with

$$T_1^* = \iint \frac{1}{g^{\ell+2}} L^{(\ell+1)}(\zeta_n)\left[F_{0n_0}(y) - F_{0n_0}(z) - (F_0(y) - F_0(z))\right] f_1(x_1) f_1(x_2)\,\mathrm{d}y\mathrm{d}z,$$

it follows that $Var(A_\ell) = O\left(\frac{1}{n_0 g^{2(\ell+2)}}\right)$. However, assuming that $g \to 0$, it is impossible to conclude from here that $Var(A_\ell) = o\left(\frac{1}{n_0}\right)$.

Remark 2.3.2. The first term in the right-hand side of (2.16) corresponds to the main squared bias term of MSE. Note that, using (K9) and (2.13), the bias term vanishes by choosing $g = g_\ell$

$$g_\ell = \left(\frac{2L^{(\ell)}(0)\left(1+\kappa^2\Psi_0(r)\right)}{-d_L\Psi_{\ell+2}(r)n_1}\right)^{\frac{1}{(\ell+3)}} = \left(\frac{2L^{(\ell)}(0)d_K^2\Psi_4(r)}{-d_L\Psi_{\ell+2}(r)R(K)}\right)^{\frac{1}{\ell+3}} h_{AMISE}^{\frac{5}{\ell+3}},$$

which is also the asymptotic MSE-optimal bandwidth.

Note that g_ℓ only has sense if the term in brackets is positive. Since

$$\Psi_{\ell+2}(r) = \Psi_{2(\frac{\ell}{2}+1)}(r) = (-1)^{\frac{\ell}{2}+1} R\left(r^{(\frac{\ell}{2}+1)}\right)$$

and $R\left(r^{(\frac{\ell}{2}+1)}\right)$ is positive by definition, the required condition to guarantee that

$$\frac{2L^{(\ell)}(0)(1+\kappa^2\Psi_0(r))}{-d_L\Psi_{\ell+2}(r)n_1} > 0$$

is that $(-1)^{\frac{\ell}{2}+2}L^{(\ell)}(0)d_L > 0$, which is exactly condition (K9) introduced previously.

2.3.2 STE rules based on Sheather and Jones ideas

As in the context of ordinary density estimation, the practical implementation of the kernel-type estimators proposed here (see (1.32) and (2.1)), requires the choice of the smoothing parameter h. Our two proposals, h_{SJ_1} and h_{SJ_2}, as well as the selector b_{3c} recommended by Ćwik and Mielniczuk (1993), are modifications of Sheather and Jones (1991). Since the Sheather & Jones selector is the solution of an equation in the bandwidth, it is also known as a solve-the-equation (STE) rule. Motivated by formula (2.11) for the AMISE-optimal bandwidth and the relation (2.13), solve-the-equation rules require that h is chosen to satisfy the relationship

$$h = \left(\frac{R(K)\left(\kappa^2\tilde{\Psi}_0(\gamma_1(h), h_0; L)+1\right)}{d_K^2\tilde{\Psi}_4(\gamma_2(h), h_0; L)\,n_1}\right)^{\frac{1}{5}},$$

where the pilot bandwidths for the estimation of $\Psi_0(r)$ and $\Psi_4(r)$ are functions of h ($\gamma_1(h)$ and $\gamma_2(h)$, respectively).

Motivated by Remark 2.3.2, we suggest taking

$$\gamma_1(h) = \left(\frac{2L(0)\, d_K^2 \tilde{\Psi}_4(g_4, h_0; L)}{-d_L \tilde{\Psi}_2(g_2, h_0; L)\, R(K)} \right)^{\frac{1}{3}} h^{\frac{5}{3}}$$

and

$$\gamma_2(h) = \left(\frac{2L^{(4)}(0)\, d_K^2 \tilde{\Psi}_4(g_4, h_0; L)}{-d_L \tilde{\Psi}_6(g_6, h_0; L)\, R(K)} \right)^{\frac{1}{7}} h^{\frac{5}{7}},$$

where $\tilde{\Psi}_j(\cdot)$, $(j = 0, 2, 4, 6)$ are kernel estimates (2.15). Note that this way of proceeding leads us to a never ending process in which a bandwidth selection problem must be solved at every stage. To make this iterative process feasible in practice one possibility is to propose a stopping stage in which the unknown quantities are estimated using a parametric scale for r. This strategy is known in the literature as the stage selection problem (see Wand and Jones (1995)). While the selector b_{3c} in Ćwik and Mielniczuk (1993) used a Gaussian scale, now for the implementation of h_{SJ_2}, we will use a mixture of betas based on the Weierstrass approximation theorem and Bernstein polynomials associated to any continuous function on $[0, 1]$ (see Kakizawa (2004) and references therein for the motivation of this method). Later on we will show the formula for computing the reference scale above-mentioned.

In the following we denote the Epanechnikov kernel by K, the uniform density in $[-1, 1]$ by M and we define L as a $\beta(9, 9)$ density function rescaled to the interval $[-1, 1]$:

$$L(x) = \frac{\Gamma(18)}{2\Gamma(9)\Gamma(9)} \left(\frac{x+1}{2} \right)^8 \left(1 - \frac{x+1}{2} \right)^8 1_{\{-1 \le x \le 1\}}.$$

Next, we detail the steps required in the implementation of h_{SJ_2}. Since the selector h_{SJ_1} is a modified version of h_{SJ_2}, a brief mention to it will be given later on after a more detailed presentation of h_{SJ_2}.

Step 1. Obtain $\hat{\Psi}_j^{PR}$ $(j = 0, 4, 6, 8)$, parametric estimates for $\Psi_j(r)$ $(j = 0, 4, 6, 8)$, with the replacement of $r(x)$ in $R\left(r^{(j/2)}\right)$ (see (2.12)), by a mixture of betas, $\tilde{b}(x; N, R)$, as it will be explained later on (see (2.33)).

Step 2. Compute kernel estimates for $\Psi_j(r)$ $(j = 2, 4, 6)$, by using $\tilde{\Psi}_j(g_j^{PR}, h_0; L)$ $(j = 2, 4, 6)$, with

$$g_j^{PR} = \left(\frac{2L^{(j)}(0) \left(\kappa^2 \hat{\Psi}_0^{PR} + 1 \right)}{-d_L \hat{\Psi}_{j+2}^{PR} n_1} \right)^{\frac{1}{j+3}}, j = 2, 4, 6.$$

Step 3. Estimate $\Psi_k(r)$ $(k = 0, 4)$, by means of $\tilde{\Psi}_0(\hat{\gamma}_1(h), h_0; L)$ and $\tilde{\Psi}_4(\hat{\gamma}_2(h), h_0; L)$, where

$$\hat{\gamma}_1(h) = \left(\frac{2L(0)\, d_K^2 \tilde{\Psi}_4\left(g_4^{PR}, h_0; L\right)}{-d_L \tilde{\Psi}_2\left(g_2^{PR}, h_0; L\right) R(K)} \right)^{\frac{1}{3}} h^{\frac{5}{3}}$$

and

$$\hat{\gamma}_2(h) = \left(\frac{2L^{(4)}(0)\, d_K^2 \tilde{\Psi}_4\left(g_4^{PR}, h_0; L\right)}{-d_L \tilde{\Psi}_6\left(g_6^{PR}, h_0; L\right) R(K)} \right)^{\frac{1}{7}} h^{\frac{5}{7}}.$$

Step 4. Select the bandwidth h_{SJ_2} as the one that solves the following equation in h:

$$h = \left(\frac{R(K)\left(\kappa^2 \tilde{\Psi}_0\left(\hat{\gamma}_1(h), h_0; L\right) + 1\right)}{d_K^2 \tilde{\Psi}_4\left(\hat{\gamma}_2(h), h_0; L\right) n_1} \right)^{\frac{1}{5}}.$$

In order to solve the equation above, it will be necessary to use a numerical algorithm. In the simulation study we will use the false-position method. The main reason is that the false-position algorithm does not require the computation of the derivatives, which simplifies considerably the implementation of the proposed bandwidth selectors. At the same time, this algorithm presents some advantages over others because it tries to combine the speed of methods such as the secant method with the security afforded by the bisection method. Figure 2.1 below exemplifies how the false-position method works.

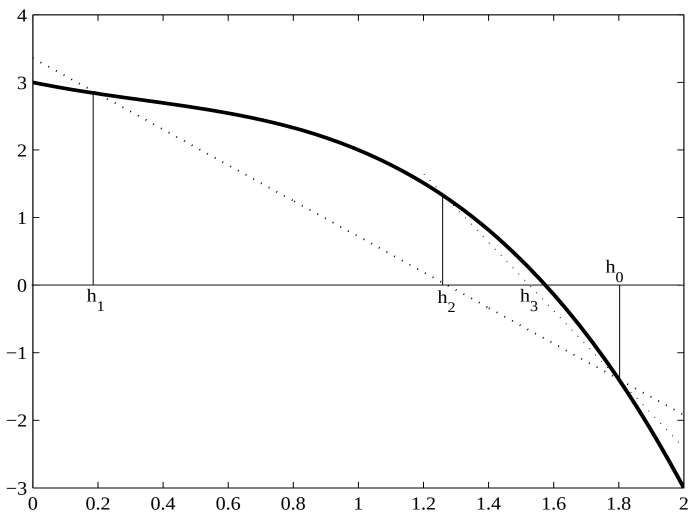

Figure 2.1: Graphical representation of the false-position algorithm.

Unlike the Gaussian parametric reference, used to obtain b_{3c}, the selector h_{SJ_2} uses in Step 1 a mixture of betas. Since we are trying to estimate a density with support in $[0,1]$ it seems more suitable to consider a parametric reference with this support. A mixture of betas is an appropriate option because it is flexible enough to model a large variety of relative densities, when derivatives of order 1, 3 and 4 are also required. This can be clearly observed in Figure 2.2 where the theoretical relative density, $r(x) = \beta(x, 4, 5)$, the function $\tilde{b}(x; N, R)$ with $N = 14, 74$ (for a pair of samples coming from that model), and the above-mentioned derivatives of both functions are shown.

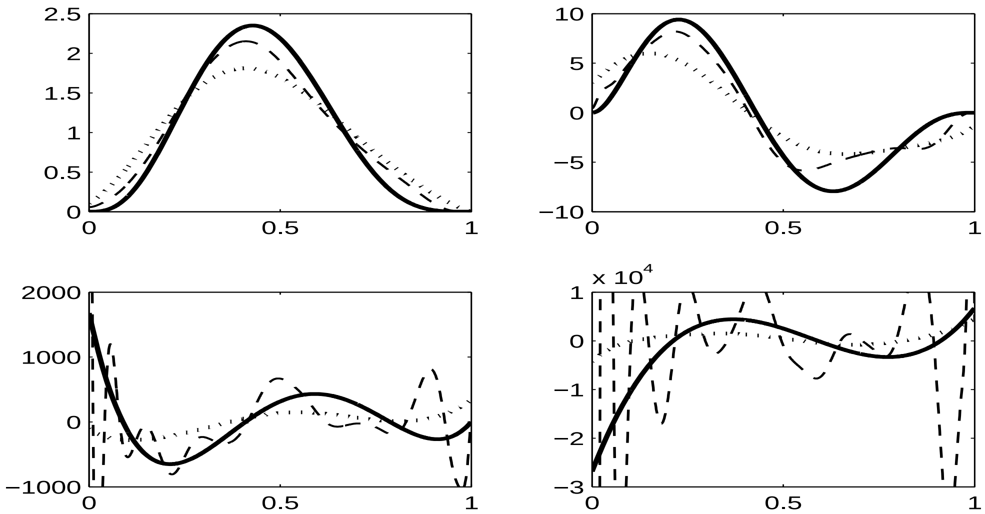

Figure 2.2: Relative density $r(x) = \beta(x, 4, 5)$ (solid lines) and $\tilde{b}(x, N, R)$ (with $N = 14$ dotted lines, with $N = 74$ dashed lines) for a pair of samples with sizes $n_0 = n_1 = 200$: $r(x)$ and $\tilde{b}(x, N, R)$ (left-top panel); $r^{(1)}(x)$ and $\tilde{b}^{(1)}(x, N, R)$ (right-top panel); $r^{(3)}(x)$ and $\tilde{b}^{(3)}(x, N, R)$ (left-bottom panel); $r^{(4)}(x)$ and $\tilde{b}^{(4)}(x, N, R)$ (right-bottom panel).

Let $\beta(x, a, b)$ be the beta density

$$\beta(x, a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad x \in [0,1],$$

and let $B(x; N, G)$ be the Bernstein polynomial associated to any continuous function G on the closed interval $[0,1]$:

$$\begin{aligned} B(x; N, G) &= (N+1)^{-1} \sum_{j=1}^{N+1} G\left(\frac{j-1}{N}\right) \beta(x, j, N-j+2) \\ &= \sum_{j=0}^{N} G\left(\frac{j}{N}\right) \frac{N!}{j!(N-j)!} x^j (1-x)^{N-j}. \end{aligned}$$

Applying Weierstrass's theorem it is known that $B(x;N,G)$ converges to $G(x)$ uniformly in $x \in [0,1]$ as N tends to ∞. For a distribution function, G, on $[0,1]$, it follows that $B(x;N,G)$ is a proper distribution function with density function $b(x;N,G) = B^{(1)}(x;N,G)$, i.e.

$$b(x;N,G) = \sum_{j=1}^{N} \left(G\left(\frac{j}{N}\right) - G\left(\frac{j-1}{N}\right) \right) \beta(x,j,N-j+1). \tag{2.32}$$

Based on this idea, we propose for $r(x)$ the following parametric fit $\tilde{b}(x;N,R)$ where the unknown relative distribution function R in (2.32) is replaced by a smooth estimate, $\tilde{R}_{g_{\tilde{R}}}$, as follows:

$$\tilde{b}(x;N,R) = \sum_{j=1}^{N} \left(\tilde{R}_{g_{\tilde{R}}}\left(\frac{j}{N}\right) - \tilde{R}_{g_{\tilde{R}}}\left(\frac{j-1}{N}\right) \right) \beta(x,j,N-j+1), \tag{2.33}$$

where

$$\tilde{R}_{g_{\tilde{R}}}(x) = n_1^{-1} \sum_{j=1}^{n_1} \mathbb{M}\left(\frac{x - \tilde{F}_{0h_0}(X_{1j})}{g_{\tilde{R}}} \right), \tag{2.34}$$

$$g_{\tilde{R}} = \left(\frac{2\int_{-\infty}^{\infty} xM(x)\,\mathbb{M}(x)\,\mathrm{d}x}{n_1 d_M^2 R\left(r^{PR(1)}\right)} \right)^{\frac{1}{3}}$$

and N is the number of betas in the mixture. We refer to the interested reader to Kakizawa (2004) for more details with respect to this choice.

Note that, for the sake of simplicity, we are using above the AMISE-optimal bandwidth ($g_{AMISE,R}$) for estimating a distribution function in the setting of a one-sample problem (see Polansky and Baker (2000) for more details in the kernel-type estimate of a distribution function). The use of this bandwidth requires the previous estimation of the unknown functional, $R\left(r^{(1)}\right)$. We will consider a quick and dirty method, the rule of thumb, that uses a parametric reference for r to estimate the above-mentioned unknown quantity. More specifically, our reference scale will be a beta with parameters (p,q) estimated from the smoothed relative sample $\left\{\tilde{F}_{0h_0}(X_{1j})\right\}_{j=1}^{n_1}$, using the method of moments.

Following the same ideas as for (2.34), the bandwidth selector used for the kernel-type estimator $\tilde{F}_{0h_0}$ introduced in (2.2) is based on the AMISE-optimal bandwidth in the one-sample problem:

$$h_0 = \left(\frac{2\int_{-\infty}^{\infty} xM(x)\,\mathbb{M}(x)\,\mathrm{d}x}{n_0 d_M^2 R\left(f_0^{PR(1)}\right)} \right)^{\frac{1}{3}}.$$

As it was already mentioned above, in most of the cases this methodology will be applied to survival analysis, so it is natural to assume that our samples come from distributions with support on the positive real line. Therefore, a gamma distribution, Gamma(α, β), has been considered as the parametric reference for f_0, let say f_0^{PR}, where the parameters (α, β) are estimated from the comparison sample $\{X_{0i}\}_{i=1}^{n_0}$, using the method of moments.

For the implementation of h_{SJ_1}, we proceed analogously to that of h_{SJ_2} above. The only difference now is that throughout the previous discussion, $\tilde{\Psi}_j(\cdot)$ and $\tilde{F}_{0h_0}(\cdot)$ are replaced by, respectively, $\hat{\Psi}_j(\cdot)$ and $F_{0n_0}(\cdot)$.

As a variant of the selector that Ćwik and Mielniczuk (1993) proposed, b_{3c} is obtained as the solution to the following equation:

$$b_{3c} = \left(\frac{R(K)\left(1+\kappa^2\hat{\Psi}_0(a)\right)}{d_K^2 \hat{\Psi}_4\left(\alpha_2\left(b_{3c}\right)\right) n_1} \right)^{\frac{1}{5}},$$

where $a = 1.781\hat{\sigma} n_1^{-\frac{1}{3}}$, $\hat{\sigma} = \min\left\{S_{n_1}, \widehat{IQR}/1.349\right\}$, S_{n_1} and $\widehat{IQR}$ denote, respectively, the empirical standard deviation and the sample interquartile range of the relative data, $\{F_{0n_0}(X_{1j})\}_{j=1}^{n_1}$,

$$\alpha_2\left(b_{3c}\right) = 0.7694 \left(\frac{\hat{\Psi}_4\left(g_4^{GS}\right)}{-\hat{\Psi}_6\left(g_6^{GS}\right)} \right)^{\frac{1}{7}} b_{3c}^{\frac{5}{7}},$$

where GS stands for standard Gaussian scale,

$$g_4^{GS} = 1.2407\hat{\sigma} n_1^{-\frac{1}{7}}, \quad g_6^{GS} = 1.2304\hat{\sigma} n_1^{-\frac{1}{9}}$$

and the estimates $\hat{\Psi}_j$ (with $j = 0, 4, 6$) were obtained using (2.14), with L replaced by the standard Gaussian kernel and with data driven bandwidth selectors derived from reducing the two-sample problem to a one-sample problem.

Remark 2.3.3. The explicit formula of b_{3c} detailed previously and used in the simulation study later on, is equation (3.7) appearing in Ćwik and Mielniczuk (1993) but generalized now to the case of different sample sizes. The constant appearing in $\alpha_2\left(b_{3c}\right)$ differs slightly from that in Section 5 of Sheather and Jones (1991) because different kernel functions were used. While we use the Epanechnikov kernel for estimating r and the Gaussian kernel for estimating the functionals $\Psi_j(r)$, Ćwik and Mielniczuk (1993) used the Gaussian kernel for both. The expression given by g_4^{GS} and g_6^{GS} correspond to, respectively, a and b given in Section 5 of Sheather and Jones (1991). The differences observed in the constants come from the fact that we estimate the dispersion of the relative data using the minimum of the empirical standard deviation and the ratio of the sample interquartile range ($\widehat{IQR}$) over 1.349. However, Sheather and Jones (1991) use $\widehat{IQR}/1.349$ to estimate the dispersion of the relative data. Therefore, replacing $\hat{\sigma}$ in g_4^{GS} and g_6^{GS} by $\widehat{IQR}$ /1.349 gives expressions a and b in Section 5 of Sheather and Jones (1991).

It is interesting to note that all the kernel-type estimators presented previously ($\hat{r}_h(t)$, $\hat{r}_{h,h_0}(t)$, $\tilde{R}_{g_{\tilde{R}}}(x)$ and $\tilde{F}_{0h_0}(x)$) were not corrected to take into account, respectively, the fact that r and R have support on $[0,1]$ instead of on the whole real line, and the fact that f_0 is supported only on the positive real line. Therefore, in order to correct the boundary effect in practical applications we will use the well known reflection method (as described in Silverman (1986)), to modify $\hat{r}_h(t)$, $\hat{r}_{h,h_0}(t)$, $\tilde{R}_{g_{\tilde{R}}}(x)$ and $\tilde{F}_{0h_0}(x)$, where needed.

2.3.3 A simulation study

We compare, through a simulation study, the performance of the bandwidth selectors h_{SJ_1} and h_{SJ_2}, proposed in Subsection 2.3.2, with the existing competitor b_{3c} recommended by Ćwik and Mielniczuk (1993). Although we are aware that the smoothing parameter N introduced in (2.33) should be selected by some optimal way based on the data, this issue goes beyond the scope of this research. Consequently, from here on, we will consider $N = 14$ components in the beta mixture reference scale model given by (2.33).

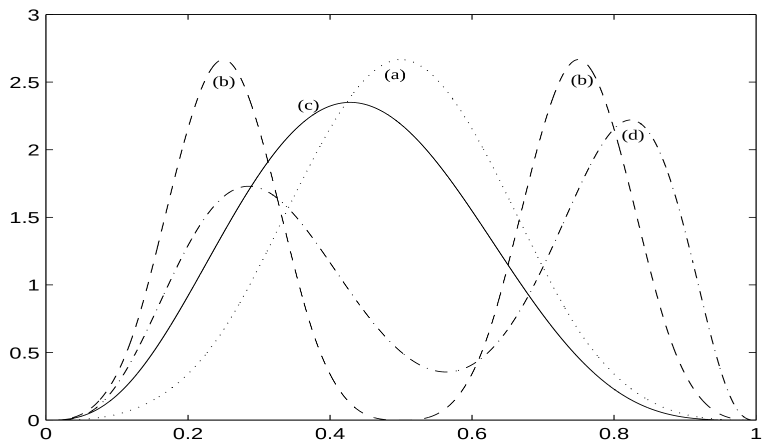

Figure 2.3: Plots of the relative densities (a)-(d).

We will consider the first sample coming from the random variate $X_0 = W^{-1}(U)$ and the second sample coming from the random variate $X_1 = W^{-1}(S)$, where U denotes a uniform distribution in the compact interval $[0,1]$, W is the distribution function of a Weibull distribution with parameters $(2,3)$ and S is a random variate from one of the following seven different populations (see Figures 2.3 and 2.4):

(a) $V = \frac{1}{4}(U_1 + U_2 + U_3 + U_4)$, where U_1, U_2, U_3, U_4 are iid $U[0,1]$.

(b) A mixture consisting of V_1 with probability $\frac{1}{2}$ and V_2 with probability $\frac{1}{2}$, where

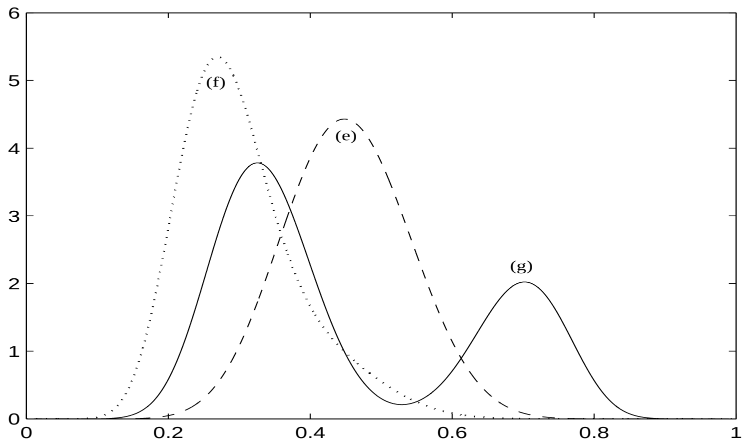

Figure 2.4: Plots of the relative densities (e)-(g).

$V_1 = \frac{V}{2}$, $V_2 = \frac{V+1}{2}$ and V as for model (a).

(c) A beta distribution with parameters 4 and 5 ($\beta(4,5)$).

(d) A mixture consisting of V_1 with probability $\frac{1}{2}$ and V_2 with probability $\frac{1}{2}$, where $V_1 = \beta(15,4)$ and $V_2 = \beta(5,11)$.

(e) A beta distribution with parameters 14 and 17 ($\beta(14,17)$).

(f) A mixture consisting of V_1 with probability $\frac{4}{5}$ and V_2 with probability $\frac{1}{5}$, where $V_1 = \beta(14,37)$ and $V_2 = \beta(14,20)$.

(g) A mixture consisting of V_1 with probability $\frac{1}{3}$ and V_2 with probability $\frac{2}{3}$, where $V_1 = \beta(34,15)$ and $V_2 = \beta(15,30)$.

Note that here we are considering that the relative distribution is the distribution of S and that the role of W is just to transform both samples for not restricting the study to the case that X_0 is $U[0,1]$.

Choosing different values for the pair of sample sizes n_0 and n_1 and under each of the models presented above, we start drawing 500 pairs of random samples and, according to every method, we select the bandwidths $\hat{h}$. Then, in order to check their performance we approximate by Monte Carlo the mean integrated squared error, EM, between the true relative density and the kernel-type estimate for r, given by (1.32) when $\hat{h} = b_{3c}, h_{SJ_1}$ or by (2.1) when $\hat{h} = h_{SJ_2}$.

The computation of the kernel-type estimations can be very time consuming when using a direct algorithm. Therefore, we will use linear binned approximations that, thanks to

Table 2.1: Values of EM for h_{SJ_1}, h_{SJ_2} and b_{3c} for models (e)-(g).

EM	Model (e)			Model (f)			Model (g)		
(n_0, n_1)	h_{SJ_1}	h_{SJ_2}	b_{3c}	h_{SJ_1}	h_{SJ_2}	b_{3c}	h_{SJ_1}	h_{SJ_2}	b_{3c}
$(50, 50)$	0.8437	0.5523	1.2082	1.1278	0.7702	1.5144	0.7663	0.5742	0.7718
$(100, 100)$	0.5321	0.3717	0.6654	0.6636	0.4542	0.7862	0.4849	0.3509	0.4771
$(200, 200)$	0.2789	0.2000	0.3311	0.4086	0.2977	0.4534	0.2877	0.2246	0.2830
$(100, 50)$	0.5487	0.3804	0.7162	0.6917	0.4833	0.8796	0.4981	0.3864	0.4982
$(200, 100)$	0.3260	0.2443	0.3949	0.4227	0.3275	0.4808	0.3298	0.2601	0.3252
$(400, 200)$	0.1739	0.1346	0.1958	0.2530	0.1924	0.2731	0.1830	0.1490	0.1811
$(50, 100)$	0.8237	0.5329	1.1189	1.1126	0.7356	1.4112	0.7360	0.5288	0.7135
$(100, 200)$	0.5280	0.3627	0.6340	0.6462	0.4288	0.7459	0.4568	0.3241	0.4449
$(200, 400)$	0.2738	0.1923	0.3192	0.3926	0.2810	0.4299	0.2782	0.2099	0.2710

Table 2.2: Values of EM for h_{SJ_2} and b_{3c} for models (a)-(d).

EM	Model (a)		Model (b)		Model (c)		Model (d)	
(n_0, n_1)	h_{SJ_2}	b_{3c}	h_{SJ_2}	b_{3c}	h_{SJ_2}	b_{3c}	h_{SJ_2}	b_{3c}
$(50, 50)$	0.1746	0.3493	0.4322	0.5446	0.1471	0.3110	0.2439	0.3110
$(100, 100)$	0.1208	0.2034	0.2831	0.3471	0.1031	0.1750	0.1474	0.1892
$(200, 200)$	0.0608	0.0964	0.1590	0.1938	0.0524	0.0827	0.0874	0.1068
$(100, 50)$	0.1241	0.2165	0.3319	0.3911	0.1139	0.2064	0.1897	0.2245
$(200, 100)$	0.0835	0.1227	0.1965	0.2286	0.0712	0.1029	0.1152	0.1351
$(400, 200)$	0.0462	0.0641	0.1089	0.1250	0.0399	0.0529	0.0661	0.0760
$(50, 100)$	0.1660	0.3308	0.3959	0.5117	0.1256	0.2615	0.2075	0.2826
$(100, 200)$	0.1073	0.1824	0.2577	0.3240	0.0898	0.1548	0.1281	0.1685
$(200, 400)$	0.0552	0.0897	0.1457	0.1820	0.0448	0.0716	0.0757	0.0953

their discrete convolution structures, can be efficiently computed by using the fast Fourier transform (FFT) (see Wand and Jones (1995) for more details).

The values of this criterion for the three bandwidth selectors, h_{SJ_1}, h_{SJ_2} and b_{3c}, can be found in Table 2.1 for models (e)-(g). Since from this table it is clear the outperformance of h_{SJ_2} over h_{SJ_1}, only the values for selectors h_{SJ_2} and b_{3c} were shown for models (a)-(d) (see Table 2.2).

A careful look at the tables points out that the new selectors, especially h_{SJ_2}, present a much better behaviour than the selector b_{3c}, particularly when the sample sizes are equal or when n_1 is larger than n_0. The improvement is even larger for unimodal relative densities (models (a), (c) and (e)). The ratio $\frac{n_1}{n_0}$ produces an important effect on the behaviour of any of the two selectors considered. For instance, it is clearly seen an asymmetric behaviour of the selectors in terms of the sample sizes.

In Figures 2.5 and 2.6 we plot the histograms of the 500 values of the bandwidths b_{3c} and h_{SJ_2} obtained in the simulation study for model (e) and sample sizes of, respectively, $(n_0, n_1) = (50, 50)$ and $(n_0, n_1) = (100, 100)$.

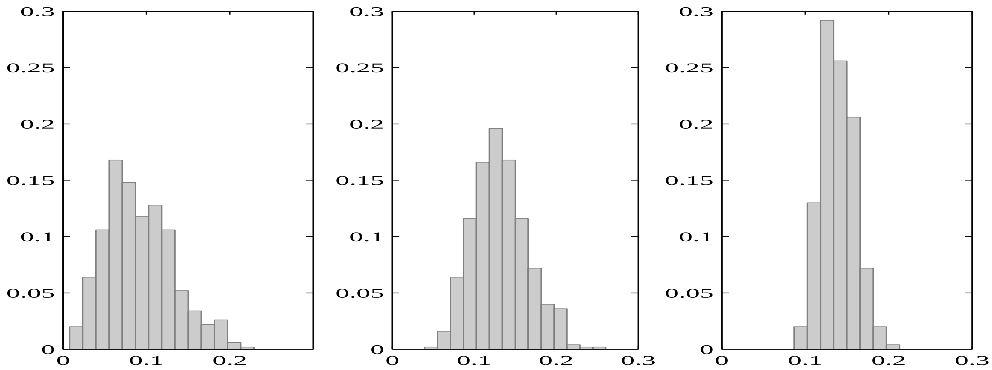

Figure 2.5: Histograms of the 500 values of b_{3c}, h_{SJ_1} and h_{SJ_2} (from left to right) obtained under model (e) and $(n_0, n_1) = (50, 50)$.

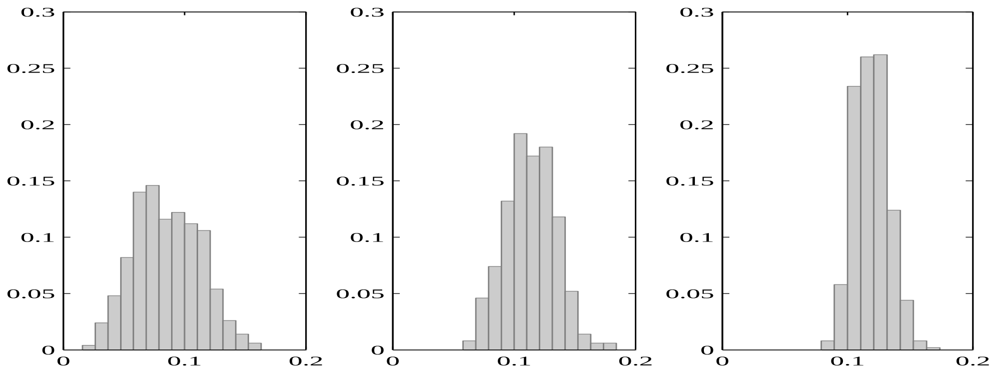

Figure 2.6: Histograms of the 500 values of b_{3c}, h_{SJ_1} and h_{SJ_2} (from left to right) obtained under model (e) and $(n_0, n_1) = (100, 100)$.

Table 2.3 includes, for models (a)–(d), the mean of the CPU times required per trial in the computation of b_{3c} and h_{SJ_2}. Since the CPU times required to compute h_{SJ_1} were very similar to the ones obtained for h_{SJ_2}, they were omitted in Table 2.3.

Other proposals for selecting h have been investigated. For instance, versions of h_{SJ_2} were considered in which either the unknown functionals Ψ_ℓ are estimated from the viewpoint of a one-sample problem or the STE rule is modified in such a way that only the function γ_2 is considered in the equation to be solved (see Step 4 in Subsection 2.3.2). In a simulation study similar to the one detailed here, but now carried out for these versions of h_{SJ_2}, it was observed a similar practical performance to that observed for h_{SJ_2}. However, a worse behaviour was observed when, in the implementation of h_{SJ_2}, the smooth estimate of F_0 is replaced by the empirical distribution function F_{0n_0}. Therefore, although h_{SJ_2} requires the selection of two smoothing parameters, a clear better practical behaviour is

Table 2.3: Mean of the CPU time (in seconds) required per trial in the computation of bandwidths h_{SJ_2} and b_{3c} for models (a)–(d).

CPU time	Model (a)		Model (b)		Model (c)		Model (d)	
(n_0, n_1)	h_{SJ_2}	b_{3c}	h_{SJ_2}	b_{3c}	h_{SJ_2}	b_{3c}	h_{SJ_2}	b_{3c}
$(50, 50)$	4.9024	0.0588	4.9252	0.1603	4.8988	0.0662	3.4335	0.0682
$(100, 100)$	4.7759	0.0879	4.5524	0.0812	4.7797	0.1075	3.3601	0.0637
$(200, 200)$	5.7760	0.1099	4.7692	0.1170	5.9758	0.1025	3.4739	0.0743
$(100, 50)$	4.8467	0.0536	4.9455	0.1328	4.8413	0.0702	3.4431	0.0564
$(200, 100)$	4.8545	0.0843	4.4674	0.0713	4.7913	0.0929	3.3451	0.0622
$(400, 200)$	5.6559	0.1079	4.8536	0.1140	5.8951	0.0993	3.3867	0.0839
$(50, 100)$	4.9909	0.0811	5.0604	0.1696	4.9540	0.0904	3.5827	0.0863
$(100, 200)$	3.1026	0.1078	4.8804	0.1151	5.1038	0.1116	3.6372	0.0785
$(200, 400)$	6.3867	0.1254	5.6517	0.1320	6.5351	0.1131	4.0607	0.1061

observed when considering the smoothed relative data instead of the non-smoothed ones.

A brief version of the contents of Sections 2.1–2.3 has been accepted for publication in the Annals of the Institute of Statistical Mathematics (see Molanes and Cao (2006a)).

2.4 Bootstrap selectors

2.4.1 Exact MISE calculations

As Cao (1993) points out, in the case of a one-sample problem, there exists a closed expression for the mean integrated squared error of the Parzen Rosenblatt estimator of an ordinary density function, f_0. Using Fubini's theorem and decomposing the mean integrated squared error into the integrated variance and the integrated squared bias, it follows that:

$$\begin{aligned} MISE(\tilde{f}_{0h}) &= \int \left(\int K(u) \left(f_0(x - hu) - f_0(x) \right) \mathrm{d}u \right)^2 \mathrm{d}x \\ &\quad + \frac{R(K)}{nh} + \frac{1}{n} \int \left(\int K(u) f_0(x - hu) \mathrm{d}u \right)^2 \mathrm{d}x. \end{aligned}$$

We next prove a similar result for the two-sample problem, when estimating the relative density function, $r(t)$.

Let us consider the following assumptions:

(D12) $F_0(X_1)$ is absolutely continuous.

(K10) L is a bounded density function in $(-\infty, \infty)$.

Theorem 2.4.1. *Assume conditions (D12) and (K10). Then, the MISE of the kernel relative density estimator in (1.32) can be written as follows:*

$$\begin{aligned} MISE\left(\hat{r}_h\right) &= \int r^2(t)\,\mathrm{d}t - 2\sum_{i=0}^{n_0} C_{n_0}^i a_{n_0,r(\cdot)}(i) \int L_h\left(t-\frac{i}{n_0}\right) r(t)\,\mathrm{d}t \\ &\quad + \frac{(L_h * L_h)(0)}{n_1} \\ &\quad + 2\frac{n_1-1}{n_1}\sum_{i=0}^{n_0}\sum_{j=i}^{n_0} P_{n_0}^{i,j-i,n_0-j} b_{n_0,r(\cdot)}(i,j)\,(L_h * L_h)\left(\frac{j-i}{n_0}\right), \end{aligned}$$

where

$$\begin{aligned} a_{n_0,r(\cdot)}(i) &= \int_0^1 s^i (1-s)^{n_0-i} r(s)\,\mathrm{d}s, \\ b_{n_0,r(\cdot)}(i,j) &= \int_0^1 (1-s_2)^{n_0-j} r(s_2)\, s_2^{j+1} \int_0^1 s_3^i (1-s_3)^{j-i} r(s_2 s_3)\,\mathrm{d}s_3 \mathrm{d}s_2, \ \text{for } j \geq i, \\ (L_h * L_h)(t) &= \int L_h(t-s) L_h(s)\,\mathrm{d}s, \end{aligned}$$

$C_{n_0}^i$ *and* $P_{n_0}^{i,j-i,n_0-j}$ *denote, respectively, the binomial coefficient* $\binom{n_0}{i}$ *and the multinomial coefficient* $\frac{n_0!}{i!(j-i)!(n_0-j)!}$.

Proof of Theorem 2.4.1. Standard bias-variance decomposition of MSE gives:

$$MISE\left(\hat{r}_h\right) = \int \left[E\left[\hat{r}_h(t)\right] - r(t)\right]^2 \mathrm{d}t + \int Var\left[\hat{r}_h(t)\right]\mathrm{d}t. \tag{2.35}$$

For the first term, it is easy to check that

$$\begin{aligned} E\left[\hat{r}_h(t)\right] &= E\left[L_h\left(t - F_{0n_0}(X_1)\right)\right] = E\left[E\left[L_h\left(t - F_{0n_0}(X_1)\right)/X_{01},...,X_{0n_0}\right]\right] \\ &= E\left[\int L_h\left(t - F_{0n_0}(y)\right) f_1(y)\,\mathrm{d}y\right] = \int E\left[L_h\left(t - F_{0n_0}(y)\right) f_1(y)\right]\mathrm{d}y \\ &= \sum_{i=0}^{n_0} L_h\left(t-\frac{i}{n_0}\right)\binom{n_0}{i}\int F_0(y)^i \left(1-F_0(y)\right)^{n_0-i} f_1(y)\,\mathrm{d}y \\ &= \sum_{i=0}^{n_0} L_h\left(t-\frac{i}{n_0}\right)\binom{n_0}{i}\int s^i (1-s)^{n_0-i} r(s)\,\mathrm{d}s. \end{aligned} \tag{2.36}$$

On the other hand it is straightforward to see that

$$
\begin{aligned}
Var\left[\hat{r}_h(t)\right] &= \frac{1}{n_1} Var\left[L_h\left(t - F_{0n_0}(X_1)\right)\right] \\
&+ \frac{n_1-1}{n_1} Cov\left[L_h\left(t - F_{0n_0}(X_{11})\right), L_h\left(t - F_{0n_0}(X_{12})\right)\right] \\
&= \frac{1}{n_1} E\left[L_h^2\left(t - F_{0n_0}(X_1)\right)\right] - \frac{1}{n_1} E^2\left[L_h\left(t - F_{0n_0}(X_1)\right)\right] \\
&+ \frac{n_1-1}{n_1} Var\left[\int L_h\left(t - F_{0n_0}(y)\right) f_1(y)\,\mathrm{d}y\right] \\
&= \frac{1}{n_1} E\left[L_h^2\left(t - F_{0n_0}(X_1)\right)\right] - E^2\left[L_h\left(t - F_{0n_0}(X_1)\right)\right] \\
&+ \frac{n_1-1}{n_1} E\left[\left(\int L_h\left(t - F_{0n_0}(y)\right) f_1(y)\,\mathrm{d}y\right)^2\right]. \qquad (2.37)
\end{aligned}
$$

In order to get a more explicit expression for the variance, we study the expectations in the right hand-side of the expression above.

The first expectation is

$$
\begin{aligned}
E\left[L_h^2\left(t - F_{0n_0}(X_1)\right)\right] &= E\left[E\left[L_h^2\left(t - F_{0n_0}(X_1)\right)/X_{01},\ldots,X_{0n_0}\right]\right] \\
&= \int E\left[L_h^2\left(t - F_{0n_0}(y)\right)\right] f_1(y)\,\mathrm{d}y \\
&= \sum_{i=0}^{n_0} L_h^2\left(t - \frac{i}{n_0}\right)\binom{n_0}{i}\int F_0(y)^i\left(1 - F_0(y)\right)^{n_0-i} f_1(y)\,\mathrm{d}y \\
&= \sum_{i=0}^{n_0} L_h^2\left(t - \frac{i}{n_0}\right)\binom{n_0}{i}\int s^i(1-s)^{n_0-i} r(s)\,\mathrm{d}s. \qquad (2.38)
\end{aligned}
$$

The last expectation can be written as

$$
\begin{aligned}
&E\left[\left(\int L_h\left(t - F_{0n_0}(y)\right) f_1(y)\,\mathrm{d}y\right)^2\right] \\
&= E\left[\int\int L_h\left(t - F_{0n_0}(y_1)\right) L_h\left(t - F_{0n_0}(y_2)\right) f_1(y_1) f_1(y_2)\,\mathrm{d}y_1\mathrm{d}y_2\right] = 2A,
\end{aligned}
$$

based on the symmetry of the integrand, where

$$\begin{aligned} A &= \int\int_{y_2>y_1} E\left[L_h\left(t-F_{0n_0}\left(y_1\right)\right)L_h\left(t-F_{0n_0}\left(y_2\right)\right)\right]f_1\left(y_2\right)f_1\left(y_1\right)\mathrm{d}y_2\mathrm{d}y_1 \\ &= \sum_{i=0}^{n_0}\sum_{\substack{j=0\\ i\le j}}^{n_0} L_h\left(t-\frac{i}{n_0}\right)L_h\left(t-\frac{j}{n_0}\right)\frac{n_0!}{i!\,(j-i)!\,(n_0-j)!} \\ &\quad \int\int_{y_2>y_1} F_0\left(y_1\right)^i\left(F_0\left(y_2\right)-F_0\left(y_1\right)\right)^{j-i}\left(1-F_0\left(y_2\right)\right)^{n_0-j}f_1\left(y_2\right)f_1\left(y_1\right)\mathrm{d}y_2\mathrm{d}y_1 \\ &= \sum_{i=0}^{n_0}\sum_{\substack{j=0\\ i\le j}}^{n_0} L_h\left(t-\frac{i}{n_0}\right)L_h\left(t-\frac{j}{n_0}\right)\frac{n_0!}{i!\,(j-i)!\,(n_0-j)!} \\ &\quad \int\int_{s_2>s_1} s_1^i\left(s_2-s_1\right)^{j-i}\left(1-s_2\right)^{n_0-j}r\left(s_2\right)r\left(s_1\right)\mathrm{d}s_2\mathrm{d}s_1. \end{aligned}$$

Therefore,

$$\begin{aligned} &E\left[\left(\int L_h\left(t-F_{0n_0}\left(y\right)\right)f_1\left(y\right)\mathrm{d}y\right)^2\right] \\ &= 2\sum_{i=0}^{n_0}\sum_{\substack{j=0\\ i\le j}}^{n_0} L_h\left(t-\frac{i}{n_0}\right)L_h\left(t-\frac{j}{n_0}\right)\frac{n_0!}{i!\,(j-i)!\,(n_0-j)!} \\ &\int\int_{s_2>s_1} s_1^i\left(s_2-s_1\right)^{j-i}\left(1-s_2\right)^{n_0-j}r\left(s_2\right)r\left(s_1\right)\mathrm{d}s_2\mathrm{d}s_1. \end{aligned} \tag{2.39}$$

Using (2.38), (2.36) and (2.39) in (2.37) and (2.36) and (2.37) in (2.35) gives

$$\begin{aligned} MISE\left(\hat{r}_h\right) &= \int\left(\sum_{i=0}^{n_0}L_h\left(t-\frac{i}{n_0}\right)\binom{n_0}{i}a_{n_0,r(\cdot)}\left(i\right)-r\left(t\right)\right)^2\mathrm{d}t \\ &+\frac{1}{n_1}\sum_{i=0}^{n_0}\binom{n_0}{i}a_{n_0,r(\cdot)}\left(i\right)\int L_h^2\left(t-\frac{i}{n_0}\right)\mathrm{d}t \\ &-\int\left(\sum_{i=0}^{n_0}L_h\left(t-\frac{i}{n_0}\right)\binom{n_0}{i}a_{n_0,r(\cdot)}\left(i\right)\right)^2\mathrm{d}t \\ &+2\frac{n_1-1}{n_1}\sum_{i=0}^{n_0}\sum_{j=i}^{n_0}\frac{n_0!}{i!\,(j-i)!\,(n_0-j)!}b_{n_0,r(\cdot)}\left(i,j\right) \\ &\int L_h\left(t-\frac{i}{n_0}\right)L_h\left(t-\frac{j}{n_0}\right)\mathrm{d}t \end{aligned}$$

and consequently we get that

$$\begin{aligned} MISE\left(\hat{r}_h\right) &= \int\left(\sum_{i=0}^{n_0} L_h\left(t-\frac{i}{n_0}\right) C_{n_0}^i a_{n_0,r(\cdot)}\left(i\right)-r\left(t\right)\right)^2 \mathrm{d}t \\ &+\frac{1}{n_1}\sum_{i=0}^{n_0} C_{n_0}^i a_{n_0,r(\cdot)}\left(i\right)\left(L_h * L_h\right)(0) \\ &-\int\left(\sum_{i=0}^{n_0} L_h\left(t-\frac{i}{n_0}\right) C_{n_0}^i a_{n_0,r(\cdot)}\left(i\right)\right)^2 \mathrm{d}t \\ &+2\frac{n_1-1}{n_1}\sum_{i=0}^{n_0}\sum_{j=i}^{n_0} P_{n_0}^{i,j-i,n_0-j} b_{n_0,r(\cdot)}\left(i,j\right)\left(L_h * L_h\right)\left(\frac{j-i}{n_0}\right). \end{aligned}$$

Some simple algebra concludes the proof. □

2.4.2 Resampling schemes

When using the bootstrap technique to estimate the MISE of, either $\hat{r}_h$ or $\hat{r}_{h,h_0}$, one possibility would be to approximate the distribution function of the ISE process and then compute its expectation. To this aim we first need to define a resampling scheme that imitates the procedure from which the two original samples were drawn. As pointed out in Cao (1993) for the setting of ordinary density estimation, this can be achieved by replacing the role of the true target density, in this case r, by some estimator of it. Since we are in a two-sample setting we need to draw a pair of resamples of n_0 and n_1 observations respectively, the first one coming from a population, say X_0^*, and the last one coming from another one, say X_1^*. Besides, the relative density of X_1^* wrt X_0^* should coincide with the kernel relative density estimator. Therefore, the ideas presented in Cao (1993) require some modifications to be adapted to this new setting. There exist at least two ways to proceed. Either replacing the roles of the densities, f_0 and f_1, by some appropriate estimators or considering a uniform distribution on $[0,1]$ for X_0^* and a distribution with density equal to the relative density estimator for X_1^*. The second possibility is justified by noting that the sampling distribution of $\hat{r}_h$ only depends on the two populations through their relative density, r.

We now present a bootstrap procedure to approximate $MISE(\hat{r}_h)$:

(a) Select a pilot bandwidth, g, and construct the relative density estimator $\hat{r}_g$ (see (1.32)).

(b) Draw bootstrap resamples $\left\{X_{01}^*,\ldots,X_{0n_0}^*\right\}$, from a uniform distribution on $[0,1]$, and $\left\{X_{11}^*,\ldots,X_{1n_1}^*\right\}$, with density function $\hat{r}_g$.

(c) Consider, for each $h > 0$, the bootstrap version of the kernel estimator (1.32):

$$\hat{r}_h^*(x) \quad = \quad n_1^{-1} \sum_{j=1}^{n_1} L_h(x - F_{0n_0}^*(X_{1j}^*)),$$

where $F_{0n_0}^*$ denotes the empirical distribution function of $\{X_{01}^*, \ldots, X_{0n_0}^*\}$.

(d) Define the bootstrap mean integrated squared error as a function of h:

$$MISE^*(\hat{r}_h^*) = E^* \left[\int (\hat{r}_h^*(x) - \hat{r}_g(x))^2 \, \mathrm{d}x \right]. \tag{2.40}$$

(e) Find the minimizer of (2.40). This value, denoted by $h_{MISE(\hat{r}_h)}^*$, is a bootstrap analogue of the MISE bandwidth for $\hat{r}_h$.

By definition, the bootstrap MISE function in (2.40) does not depend on the resamples. Therefore, in case that a closed expression could be found for it, Monte Carlo approximation could be avoided. In other words, there would be no need of drawing resamples (steps (b) and (c) in the bootstrap procedure sketched above) which always means an important computational load. In the one-sample problem this approach was plausible (see Cao *et al* (1994)) and yielded a considerable saving of computing time.

A bootstrap version for Theorem 2.4.1 can be proved using parallel arguments:

Theorem 2.4.2. *Assume condition (K10). Then,*

$$\begin{aligned} MISE^*\left(\hat{r}_h^*\right) \quad = \quad & \int \hat{r}_g^2\left(t\right) \mathrm{d}t - 2\sum_{i=0}^{n_0} C_{n_0}^i a_{n_0,\hat{r}_g(\cdot)}\left(i\right) \int L_h\left(t - \frac{i}{n_0}\right) \hat{r}_g\left(t\right) \mathrm{d}t \\ & + \frac{\left(L_h * L_h\right)\left(0\right)}{n_1} \\ & + 2\frac{n_1 - 1}{n_1} \sum_{i=0}^{n_0} \sum_{j=i}^{n_0} P_{n_0}^{i,j-i,n_0-j} b_{n_0,\hat{r}_g(\cdot)}\left(i,j\right) \left(L_h * L_h\right) \left(\frac{j-i}{n_0}\right), \end{aligned} \tag{2.41}$$

where

$$a_{n_0,\hat{r}_g(\cdot)}\left(i\right) = \int s^i \left(1-s\right)^{n_0-i} \hat{r}_g\left(s\right) \mathrm{d}s$$

and

$$b_{n_0,\hat{r}_g(\cdot)}\left(i,j\right) = \int_0^1 \left(1-s_2\right)^{n_0-j} \hat{r}_g\left(s_2\right) s_2^{j+1} \int_0^1 s_3^i \left(1-s_3\right)^{j-i} \hat{r}_g\left(s_2 s_3\right) \mathrm{d}s_3 \mathrm{d}s_2.$$

Based on the bootstrap scheme shown previously and the closed expression obtained for $MISE^*$, we propose two bootstrap bandwidth selectors. Both consist in approximating $MISE^*$ and finding its minimizer (which yields an approximation of $h_{MISE(\hat{r}_h)}^*$). While the first one, say h_{CE}^*, approximates (2.40) using the closed expression (2.41), the second proposal, say h_{MC}^*, estimates (2.40) by Monte Carlo taking a large number of resamples as described in steps (b) and (c).

Bandwidth selection is also an important issue when dealing with the estimator defined in (2.1). However, it is not easy to find a closed expression for the MISE of this estimator. Below we present two bootstrap procedures to approximate $MISE(\hat{r}_{h,h_0})$. The first proposal is as follows:

Smooth Uniform Monte Carlo Bootstrap resampling plan (SUMC)

(a) Select two pilot bandwidths, g and g_0, and construct the estimator $\hat{r}_{g,g_0}$ (see (2.1)) of the relative density r. Let H be the cdf of a uniform random variable on [0,1] and consider

$$\tilde{H}_b(x) = n_0^{-1} \sum_{i=1}^{n_0} \mathbb{U}\left(\frac{x - U_i}{b}\right),$$

a kernel-type estimate of H based on the uniform kernel U on $[-1,1]$ (with distribution function $\mathbb{U}$), the bandwidth parameter b and the sample $\{U_1, \ldots, U_{n_0}\}$ coming from H. Approximate the MISE function of $\tilde{H}_b$ by Monte Carlo and find its minimizer b_0.

(b) Draw bootstrap samples $\left\{X_{01}^*, \ldots, X_{0n_0}^*\right\}$ and $\left\{X_{11}^*, \ldots, X_{1n_1}^*\right\}$ from, respectively, a uniform distribution on $[0,1]$ and the density function $\hat{r}_{g,g_0}$.

(c) Consider, for each $h > 0$, the bootstrap version of the kernel estimator (2.1):

$$\hat{r}_{h,b_0}^*(x) = n_1^{-1} \sum_{j=1}^{n_1} L_h(x - \tilde{F}_{0b_0}^*(X_{1j}^*)),$$

where $\tilde{F}_{0b_0}^*$ denotes a kernel-type cdf estimate based on the bootstrap resample $\{X_{01}^*, \ldots, X_{0n_0}^*\}$, the uniform kernel on $[-1,1]$ and the bandwidth parameter b_0 computed previously in (a).

(d) Define the bootstrap mean integrated squared error as a function of h:

$$MISE^*(\hat{r}_{h,b_0}^*) = E_*\left[\int \left(\hat{r}_{h,b_0}^*(x) - \hat{r}_{g,g_0}(x)\right)^2 \mathrm{d}x\right]. \tag{2.42}$$

(e) Find a numerical approximation of the minimizer of (2.42). This value, denoted by h_{SUMC}^*, is a bootstrap version of the MISE bandwidth for $\hat{r}_{h,h_0}$.

Since we do not have a closed expression for $MISE^*(\hat{r}_{h,b_0}^*)$, this function is approximated by Monte Carlo.

The second proposal is sketched below.

Smooth Monte Carlo Bootstrap resampling plan (SMC)

(a) Select three pilot bandwidths, g, g_0 and g_1, and construct the estimators $\hat{r}_{g,g_0}$ and $\tilde{f}_{0g_1}$ of, respectively, the relative density r and the density f_0.

(b) Draw bootstrap resamples $\left\{X_{01}^*,\ldots,X_{0n_0}^*\right\}$ from $\tilde{f}_{0g_1}$ and $\left\{Z_1^*,\ldots,Z_{n_1}^*\right\}$ from $\hat{r}_{g,g_0}$. Define $X_{1j}^* = \tilde{F}_{0g_1}^{-1}(Z_j^*)$, $j=1,2,\ldots,n_1$.

(c) Consider, for each $h>0$, the bootstrap version of the kernel estimator (2.1):

$$\hat{r}_{h,h_0}^*(x) \;=\; n_1^{-1}\sum_{j=1}^{n_1} L_h(x-\tilde{F}_{0h_0}^*(X_{1j}^*))$$

with $\tilde{F}_{0h_0}^*$ a smooth estimate of F_0 based on the bootstrap resample $\left\{X_{01}^*,\ldots,X_{0n_0}^*\right\}$.

(d) Define the bootstrap mean integrated squared error as a function of h:

$$MISE^*(\hat{r}_{h,h_0}^*) = E^*\left[\int \left(\hat{r}_{h,h_0}^*(x)-\hat{r}_{g,g_0}(x)\right)^2 \mathrm{d}x\right]. \tag{2.43}$$

(e) Find the minimizer of (2.43), h_{SMC}^*, which is a bootstrap analogue of the MISE bandwidth for $\hat{r}_{h,h_0}$.

Once more, a Monte Carlo approach has to be used to approximate the function in (2.43).

2.4.3 A simulation study

Although in the previous subsection several bootstrap selectors have been proposed, some aspects of them remain unspecified such as, for example, how the required pilot bandwidths, g, g_0 or g_1, are chosen. In the following we denote by L the Gaussian kernel.

Let us start with the proposals h_{CE}^* and h_{MC}^*. The pilot bandwidth g is selected based on the AMSE-optimal bandwidth, $g_{AMSE,\Psi_4(r)}$, to estimate the value of $\Psi_4(r)$ through $\hat{\Psi}_4(g;L)$ (see (2.14)). Note that, under regularity assumptions on r, $\Psi_4(r)$ is equal to $R\left(r^{(2)}\right)$, the curvature of r. Based on the rule of thumb, the unknown quantities depending on r that appear in $g_{AMSE,\Psi_4(r)}$, are replaced by parametric estimates based on an appropriate fit for r. This procedure leads us to define g as follows

$$g \;=\; \left(\frac{-2L^{(4)}(0)(1+\kappa^2\hat{\Psi}_0^{PR})}{d_L\hat{\Psi}_6^{PR}}\right)^{\frac{1}{7}} n_1^{-\frac{1}{7}},$$

where $\hat{\Psi}_0^{PR}$ and $\hat{\Psi}_6^{PR}$ are parametric estimates of, respectively, $\Psi_0(r)$ and $\Psi_6(r)$, based on the parametric fit, $\hat{b}(x;N,R)$, considered for $r(x)$, similar to (2.33) but with $\tilde{F}_{0h_0}$ replaced by F_{0n_0} in (2.34).

In practice we have used a number of $N = 2n_1$ beta distributions in the mixture.

For implementing h^*_{SUMC} and h^*_{SMC}, g is obtained proceeding in the same way as explained above for selectors h^*_{CE} and h^*_{MC}. The only difference now is that $\tilde{F}_{0h_0}$ in (2.34), is replaced by the smooth estimate

$$\begin{aligned} \tilde{F}_{0g_0}(x) &= n_0^{-1}\sum_{i=1}^{n_0}\mathbb{U}\left(\frac{x-X_{0i}}{g_0}\right), \\ g_0 &= \left(\frac{2\int_{-\infty}^{\infty} xU(x)\,\mathbb{U}(x)\,\mathrm{d}x}{n_0 d_U^2 R\left(f_0^{PR(1)}\right)}\right)^{\frac{1}{3}}, \end{aligned}$$

where $R\left(f_0^{PR(1)}\right)$ denotes a parametric estimate of $R\left(f_0^{(1)}\right)$ based on a gamma fit for f_0, denoted by f_0^{PR}. Note that the definition of g_0 follows the same strategy as the definition of $g_{\tilde{R}}$ given above. The only difference now is that the target distribution is F_0 rather than R and therefore a parametric scale needs to be assumed for f_0 (and not for r).

The selector h^*_{SMC} entails a third pilot bandwidth, g_1. In this case, we consider the AMSE-optimal bandwidth, $g_{AMSE,\Psi_4(f_0)}$, to estimate $\Psi_4(f_0)$ in the case of a one-sample problem (see Wand and Jones (1995)). Note that, under regularity conditions on f_0, $\Psi_4(f_0)$ is equal to $R\left(f_0^{(2)}\right)$, the curvature of f_0. Using again the rule of thumb, the bandwidth g_1 is defined as follows

$$g_1 = \left(\frac{-2L^{(4)}(0)}{d_L\hat{\Psi}_6^{PR}}\right)^{\frac{1}{7}} n_1^{-\frac{1}{7}},$$

where $\hat{\Psi}_6^{PR}$ is a parametric estimate of $\Psi_6(f_0)$, based on considering a gamma fit for f_0.

It is worth mentioning here that all the kernel type estimates required in the computation of the pilot bandwidths are boundary corrected using the well-known reflection method. Likewise, all the kernel estimates required in steps (a) and (c) for implementing h^*_{SUMC} and h^*_{SMC}, are boundary corrected. However, for selectors h^*_{CE} and h^*_{MC} only $\hat{r}_g$ (in step (a)) is boundary corrected.

These bootstrap bandwidth selectors are compared with the plug-in STE bandwidth selector h_{SJ_2} proposed previously and the slightly modified version of the bandwidth selector given by Ćwik and Mielniczuk (1993), b_{3c}.

The simulations were carried out for different sample sizes and the performance of the different data-driven selectors was examined for the seven models considered for r in Subsection 2.3.3.

For each one of the seven relative populations listed in Subsection 2.3.3, a number of 250 pairs of samples were taken. For each pair of samples, the six bandwidth selectors were computed and, based on each one, the kernel-type relative density estimate, (1.32) or (2.1), was computed. For each bandwidth selector, let say $\hat{h}$, we obtained 250 estimations of r. Based on them, the following global error measure was approximated by Monte Carlo:

$$EM = E\left[\int (\hat{r}(t) - r(t))^2 \, \mathrm{d}t\right],$$

where $\hat{r}$ denotes $\hat{r}_{\hat{h}}$, for selectors h^*_{CE}, h^*_{MC} and b_{3c}, and the boundary corrected version of $\hat{r}_{\hat{h},h_0}$ for selectors h^*_{SUMC}, h^*_{SMC} and h_{SJ_2}. These results are collected in Tables 2.4 and 2.5. Since for sample sizes of $n_0 = n_1 = 50$ (see Table 2.4), the bandwidth selectors b_{3c}, h^*_{CE} and h^*_{MC} show a worse practical behaviour, we only summarize in Table 2.5 the results obtained for the three best bandwidth selectors, h_{SJ_2}, h^*_{SUMC} and h^*_{SMC}.

Table 2.4: Values of EM for b_{3c}, h_{SJ_2}, h^*_{CE}, h^*_{MC}, h^*_{SUMC} and h^*_{SMC} for models (a)-(g).

EM		Model						
(n_0, n_1)	Selector	(a)	(b)	(c)	(d)	(e)	(f)	(g)
$(50, 50)$	b_{3c}	0.3493	0.5446	0.3110	0.3110	1.2082	1.5144	0.7718
$(50, 50)$	h_{SJ_2}	0.1746	0.4322	0.1471	0.2439	0.5523	0.7702	0.5742
$(50, 50)$	h^*_{CE}	0.2791	0.5141	0.2095	0.2630	0.7905	1.0784	0.7675
$(50, 50)$	h^*_{MC}	0.2404	0.4839	0.1951	0.2911	0.8103	0.9545	0.7195
$(50, 50)$	h^*_{SUMC}	0.1719	0.3990	0.1473	0.2517	0.6392	0.7246	0.5917
$(50, 50)$	h^*_{SMC}	0.1734	0.3996	0.1486	0.2372	0.5984	0.7093	0.5918

Table 2.5: Values of EM for h_{SJ_2}, h^*_{SUMC} and h^*_{SMC} for models (a)-(g).

EM		Model						
(n_0, n_1)	Selector	(a)	(b)	(c)	(d)	(e)	(f)	(g)
$(50, 100)$	h_{SJ_2}	0.1660	0.3959	0.1256	0.2075	0.5329	0.7356	0.5288
$(50, 100)$	h^*_{SUMC}	0.1565	0.3733	0.1276	0.2076	0.5386	0.7199	0.5576
$(50, 100)$	h^*_{SMC}	0.1580	0.3716	0.1376	0.2015	0.5655	0.7296	0.5509
$(100, 50)$	h_{SJ_2}	0.1241	0.3319	0.1139	0.1897	0.3804	0.4833	0.3864
$(100, 50)$	h^*_{SUMC}	0.1291	0.3056	0.1020	0.1914	0.4403	0.5129	0.4310
$(100, 50)$	h^*_{SMC}	0.1297	0.3060	0.1021	0.1845	0.4324	0.5048	0.4393
$(100, 100)$	h_{SJ_2}	0.1208	0.2831	0.1031	0.1474	0.3717	0.4542	0.3509
$(100, 100)$	h^*_{SUMC}	0.1095	0.2718	0.0905	0.1467	0.3789	0.4894	0.3741
$(100, 100)$	h^*_{SMC}	0.1105	0.2712	0.0871	0.1499	0.3686	0.4727	0.3639
$(100, 150)$	h_{SJ_2}	0.1045	0.2528	0.0841	0.1318	0.3305	0.4648	0.3345
$(100, 150)$	h^*_{SUMC}	0.1136	0.2552	0.0951	0.1344	0.3719	0.4887	0.3528
$(100, 150)$	h^*_{SMC}	0.1147	0.2543	0.0863	0.1439	0.3814	0.4894	0.3477

When implementing h^*_{CE}, numerical estimates of $a_{n_0,\hat{r}_g(\cdot)}(i)$ and $b_{n_0,\hat{r}_g(\cdot)}(i,j)$, are required. Using the binomial formula, these integrals can be rewritten as follows:

$$a_{n_0,\hat{r}_g(\cdot)}(i) = \sum_{k=0}^{n_0-i} (-1)^k \binom{n_0 - i}{k} \int s^{i+k} \hat{r}_g(s) \mathrm{d}s, \tag{2.44}$$

$$b_{n_0,\hat{r}_g(\cdot)}(i,j) = \sum_{q=0}^{j-i}\sum_{p=0}^{n_0-j}(-1)^q\binom{j-i}{q}(-1)^p\binom{n_0-j}{p}\int\int_{s_1} s_1^{i+q}s_2^{j-i-q+p}\hat{r}_g(s_2)\hat{r}_g(s_1)\mathrm{d}s_2\mathrm{d}s_1. \tag{2.45}$$

Therefore a possible strategy to compute $a_{n_0,\hat{r}_g(\cdot)}(i)$ and $b_{n_0,\hat{r}_g(\cdot)}(i,j)$, could be based on numerical estimates of the terms in the right hand side of (2.44) and (2.45). However, from a practical point of view, this procedure presents the disadvantage of having to take into account more than one approximation error which finally leads to worse approximations of the quantities of interest, $a_{n_0,\hat{r}_g(\cdot)}(i)$ and $b_{n_0,\hat{r}_g(\cdot)}(i,j)$. Therefore, in the present simulation study, we estimated these quantities in a straightforward way.

While the practical implementation of h^*_{CE} is very time consuming, one can make the simulation study faster for a given pair of fixed sample sizes (n_0, n_1). The fact is that there are some computations that do not need to be carried out for every pair of simulated samples because they only depend on the sample sizes but not on the observations themselves. Therefore, carrying out these computations previously can lead to a considerable gain in computing time. In the simulation study carried out here, we took advantage of this fact.

From the simulation study carried out here, we conclude that all the proposed bootstrap selectors improve the one proposed by Ćwik and Mielniczuk (1993). However, only two of them, h^*_{SUMC} and h^*_{SMC}, show a similar behaviour to the plug-in selector of best performance studied in Subsection 2.3.2, h_{SJ_2}. Sometimes, it is even observed a slight improvement over h_{SJ_2}. However, this is not always the case. These facts and the intensive computing time required for any of the bootstrap selectors compared to the time required for any of the plug-in selectors make h_{SJ_2} be a good choice in this setting.

Chapter 3

Relative density and relative distribution with LTRC data

— Cuando alguien encuentra su camino, no puede tener miedo.
Tiene que tener el coraje suficiente para dar pasos errados.
Las decepciones, las derrotas, el desánimo
son herramientas que Dios utiliza para mostrar el camino.
Paulo Coelho

3.1 Kernel-type relative density estimator

Let (X_0, T_0, C_0) denote a random vector for the reference population, where X_0 is the variable of interest with cdf F_0, T_0 is a random left truncation variable with cdf G_0 and C_0 is a random right censoring variable with cdf L_0. Similar definitions apply for the random vector (X_1, T_1, C_1) for the target population.

We assume that the cdf's are continuous and that the three random variables in each vector are independent. In this setting, for the reference population, we observe (T_0, Y_0, δ_0) if $T_0 \leq Y_0$ where $Y_0 = \min\{X_0, C_0\}$ with cdf W_0, and $\delta_0 = 1_{\{X_0 \leq C_0\}}$ is the indicator of uncensoring. The following equation is satisfied $(1 - W_0) = (1 - F_0)(1 - L_0)$. When $T_0 > Y_0$ nothing is observed. A similar setting holds for the target population.

The data consist of two independent samples of LTRC data $\{(T_{11}, Y_{11}, \delta_{11}), \ldots, (T_{1n_1}, Y_{1n_1}, \delta_{1n_1})\}$ and $\{(T_{01}, Y_{01}, \delta_{01}), \ldots, (T_{0n_0}, Y_{0n_0}, \delta_{0n_0})\}$, where the observed sample sizes n_1 and n_0 are random and the real sample sizes (say N_1 and N_0) are unknown.

Let $\alpha_0 = P(T_0 \leq Y_0)$ and $\alpha_1 = P(T_1 \leq Y_1)$ be the probabilities of absence of truncation in the reference and target populations, respectively, a straightforward consequence of the SLLN (Strong Law of Large Numbers) is that $\frac{n_1}{N_1} \to \alpha_1$ and $\frac{n_0}{N_0} \to \alpha_0$ almost surely.

Consider the following definitions that will be needed later on

$$\begin{aligned} B_0(t) &= P(T_0 \le t \le Y_0 / T_0 \le Y_0) = \alpha_0^{-1} P(T_0 \le t \le Y_0) \\ &= \alpha_0^{-1} G_0(t)(1 - F_0(t))(1 - L_0(t)) = \alpha_0^{-1} G_0(t)(1 - W_0(t)), \end{aligned}$$

$$\begin{aligned} W_{01}(t) &= P(Y_0 \le t, \delta_0 = 1 / T_0 \le Y_0) = \alpha_0^{-1} P(X_0 \le t, T_0 \le X_0 \le C_0) \\ &= \int_{a_{F_0}}^{t} \alpha_0^{-1} P(T_0 \le y \le C_0) \, \mathrm{d}F_0(y) \\ &= \int_{a_{F_0}}^{t} \alpha_0^{-1} G_0(y)(1 - L_0(y)) \, \mathrm{d}F_0(y). \end{aligned}$$

Similar definitions for the target population are omitted here.

A natural kernel-type estimator for the relative density is as follows

$$\check{r}_h(t) = \frac{1}{h} \int K\left(\frac{t - \hat{F}_{0n_0}(y)}{h}\right) \mathrm{d}\hat{F}_{1n_1}(y), \tag{3.1}$$

where $\hat{F}_{0n_0}(y)$ denotes the TJW product limit estimator (see Tsai *et al* (1987)) for the reference distribution and $\hat{F}_{1n_1}(y)$ denotes the TJW product limit estimator for the target distribution.

Below we state a result by Zhou and Yip (1999) and two useful representations for two empirical processes related in a two-sample setup. The proofs of the last two are straightforward using the first result.

Lemma 3.1.1. *(Theorem 2.2 in Zhou and Yip (1999)) Assume $a_{G_1} \le a_{W_1}$ and that for some b_1 such that $a_{W_1} < b_1 < b_{W_1}$, it is satisfied that*

$$\int_{a_{W_1}}^{b_1} \frac{\mathrm{d}W_{11}(z)}{B_1^3(z)} < \infty.$$

Then, we have, uniformly in $a_{W_1} \le x \le b_1$

$$\hat{F}_{1n_1}(x) - F_1(x) = \hat{L}_{1n_1}(x) + \hat{s}_{1n_1}(x),$$

where

$$\hat{L}_{1n_1}(x) = \frac{1}{n_1} \sum_{i=1}^{n_1} \hat{\xi}_{1i}(Y_{1i}, T_{1i}, \delta_{1i}, x),$$

$$\hat{\xi}_{1i}(Y_{1i}, T_{1i}, \delta_{1i}, x) = (1 - F_1(x)) \left\{ \frac{1_{\{Y_{1i} \le x, \delta_{1i} = 1\}}}{B_1(Y_i)} - \int_{a_{W_1}}^{x} \frac{1_{\{T_{1i} \le v \le Y_{1i}\}}}{B_1^2(v)} \mathrm{d}W_{11}(v) \right\},$$

$$\sup_{a_{W_1} \le x \le b_1} |\hat{s}_{1n_1}(x)| = O\left(n_1^{-1} \ln \ln n_1\right) \quad \text{a.s.}$$

It is worth mentioning that the $\hat{\xi}_{1i}(Y_{1i}, T_{1i}, \delta_{1i}, x)$ are zero mean, iid processes with covariance structure given by

$$\begin{aligned} Cov\left(\hat{\xi}_{11}(Y_{11}, T_{11}, \delta_{11}, x_1), \hat{\xi}_{11}(Y_{11}, T_{11}, \delta_{11}, x_2)\right) &= (1 - F_1(x_1))(1 - F_1(x_2)) \\ &\cdot Q_1(F_0(x_1 \wedge x_2)), \end{aligned}$$

where $Q_1(s) = q_1\left(F_0^{-1}(s)\right)$, with

$$q_1(t) = \int_{a_{W_1}}^{t} \frac{\mathrm{d}W_{11}(u)}{B_1^2(u)}.$$

To obtain the appropriate approximation for the relative density estimator given above, we need asymptotic representations for the TJW product limit estimator based on each one of the pseudosamples:

$$\{(F_0(T_{11}), F_0(Y_{11}), \delta_{11}), \ldots, (F_0(T_{1n_1}), F_0(Y_{1n_1}), \delta_{1n_1})\}$$

and

$$\{(F_0(T_{01}), F_0(Y_{01}), \delta_{01}), \ldots, (F_0(T_{0n_0}), F_0(Y_{0n_0}), \delta_{0n_0})\}.$$

We denote these estimators by $\tilde{R}_{n_1}(t)$ and $\tilde{U}_{0n_0}(t)$. It is easy to check that

$$\tilde{R}_{n_1}(t) = \hat{F}_{1n_1}(F_0^{-1}(t)) \tag{3.2}$$

and

$$\tilde{U}_{0n_0}(t) = \hat{F}_{0n_0}(F_0^{-1}(t)). \tag{3.3}$$

First, we will show that, in fact, $\tilde{R}_{n_1}(t) = \hat{F}_{1n_1}\left(F_0^{-1}(t)\right)$. Since

$$\begin{aligned} 1 - \tilde{R}_{n_1}(t) &= \prod_{F_0(Y_{1j}) \le t} \left(1 - \{n_1 B_{n_1}(F_0(Y_{1j}))\}^{-1}\right)^{\delta_{1j}} \text{ with} \\ B_{n_1}(z) &= \frac{1}{n_1} \sum_{j=1}^{n_1} 1_{\{F_0(T_{1j}) \le z \le F_0(Y_{1j})\}} \end{aligned}$$

and

$$\begin{aligned} 1 - \hat{F}_{1n_1}(t) &= \prod_{Y_{1j} \le t} \left(1 - \{n_1 B_{1n_1}(Y_{1j})\}^{-1}\right)^{\delta_{1j}} \text{ with} \\ B_{1n_1}(z) &= \frac{1}{n_1} \sum_{j=1}^{n_1} 1_{\{T_{1j} \le z \le Y_{1j}\}}, \end{aligned}$$

it follows that $B_{n_1}(z) = B_{1n_1}(F_0^{-1}(z))$ or equivalently, $B_{n_1}(F_0(s)) = B_{1n_1}(s)$.

Consequently,

$$\begin{aligned}
1-\hat{F}_{1n_1}(F_0^{-1}(t)) &= \prod_{Y_{1j}\le F_0^{-1}(t)} \left(1-\{n_1 B_{n_1}(F_0(Y_{1j}))\}^{-1}\right)^{\delta_{1j}} \\
&= \prod_{F_0(Y_{1j})\le t} \left(1-\{n_1 B_{n_1}(F_0(Y_{1j}))\}^{-1}\right)^{\delta_{1j}} \\
&= 1-\tilde{R}_{n_1}(t).
\end{aligned}$$

A similar argument leads to conclude that: $\tilde{U}_{0n_0}(t)=\hat{F}_{0n_0}\left(F_0^{-1}(t)\right)$.

First of all let us define

$$\tilde{W}_1(t) = P\left(F_0(Y_{1i})\le t\right) = W_1\left(F_0^{-1}(t)\right),$$

$$\tilde{B}_1(t) = P\left(F_0(T_{1i})\le t\le F_0(Y_{1i}) / F_0(T_{1i})\le F_0(Y_{1i})\right) = B_1\left(F_0^{-1}(t)\right),$$

$$\tilde{W}_{11}(t) = P\left(F_0(Y_{1i})\le t, \delta_{1i}=1 / F_0(T_{1i})\le F_0(Y_{1i})\right) = W_{11}\left(F_0^{-1}(t)\right).$$

Based on the definitions it is easy to show that $a_{\tilde{W}_1}=a_{W_1\circ F_0^{-1}}$ and $b_{\tilde{W}_1}=b_{W_1\circ F_0^{-1}}$. In fact,

$$\begin{aligned}
a_{W_1\circ F_0^{-1}} &= \inf\left\{y : W_1\circ F_0^{-1}(y)>0\right\} \\
&= \inf\left\{y : W_1(F_0^{-1}(y))>0\right\} \\
&= \inf\left\{y : \tilde{W}_1(y)>0\right\} \\
&= a_{\tilde{W}_1}.
\end{aligned}$$

Let $\tilde{G}_1$ be the cdf of $F_0(T_{1i})$. The following two lemmas are a straightforward consequence of Lemma 3.1.1.

Lemma 3.1.2. *Assume $a_{\tilde{G}_1}\le a_{\tilde{W}_1}$ and that for some $\tilde{b}_1$, such that $a_{\tilde{W}_1}<\tilde{b}_1<b_{\tilde{W}_1}$, it is satisfied that*

$$\int_{a_{\tilde{W}_1}}^{\tilde{b}_1} \frac{d\tilde{W}_{11}(z)}{\tilde{B}_1^{\,3}(z)} \equiv \int_{a_{W_1}}^{F_0^{-1}(\tilde{b}_1)} \frac{dW_{11}(z)}{B_1^3(z)} < \infty.$$

Then, we have, uniformly in $a_{\tilde{W}_1}\le t\le \tilde{b}_1$

$$\tilde{R}_{n_1}(t)-R(t)=\tilde{L}_{n_1}(t)+\tilde{s}_{n_1}(t),$$

where

$$\tilde{L}_{n_1}(t)=\frac{1}{n_1}\sum_{i=1}^{n_1}\tilde{\xi}_{1i}\left(Y_{1i},T_{1i},\delta_{1i},t\right),$$

$$\tilde{\xi}_{1i}\left(Y_{1i},T_{1i},\delta_{1i},t\right)=\hat{\xi}_{1i}\left(Y_{1i},T_{1i},\delta_{1i},F_0^{-1}(t)\right),$$

$$\sup_{a_{\tilde{W}_1}\le t\le\tilde{b}_1}\left|\tilde{s}_{n_1}(t)\right| = O\left(n_1^{-1}\ln\ln n_1\right) \quad \text{a.s.}$$

Note that $\tilde{\xi}_{1i}(Y_{1i},T_{1i},\delta_{1i},x)$ are zero mean iid processes with covariance structure given by

$$\begin{aligned} &Cov\left(\tilde{\xi}_{11}(Y_{11},T_{11},\delta_{11},x_1),\tilde{\xi}_{11}(Y_{11},T_{11},\delta_{11},x_2)\right)\\ =\ &(1-R(x_1))(1-R(x_2))\int_{a_{\tilde{W}_1}}^{x_1\wedge x_2}\left[\tilde{B}_1(v)\right]^{-2}\mathrm{d}\tilde{W}_{11}(v)\\ =\ &(1-R(x_1))(1-R(x_2))\,Q_1(x_1\wedge x_2)\,. \end{aligned}$$

Let us define $\tilde{G}_0$ the cdf of $F_0(T_{0i})$, and

$$\tilde{W}_0(t)=P(F_0(Y_{0i})\le t)=W_0\left(F_0^{-1}(t)\right),$$

$$\tilde{B}_0(t)=P\left(F_0(T_{0i})\le t\le F_0(Y_{0i})\diagup F_0(T_{0i})\le F_0(Y_{0i})\right)=B_0\left(F_0^{-1}(t)\right),$$

$$\tilde{W}_{01}(t)=P\left(F_0(Y_{0i})\le t,\delta_{0i}=1\diagup F_0(T_{0i})\le F_0(Y_{0i})\right)=W_{01}\left(F_0^{-1}(t)\right).$$

Note that $a_{\tilde{W}_0}=a_{W_0\circ F_0^{-1}}$ and $b_{\tilde{W}_0}=b_{W_0\circ F_0^{-1}}$.

Lemma 3.1.3. *Assume $a_{\tilde{G}_0}\le a_{\tilde{W}_0}$ and that for some $\tilde{b}_0$ such that $a_{\tilde{W}_0}<\tilde{b}_0<b_{\tilde{W}_0}$, it is satisfied that*

$$\int_{a_{\tilde{W}_0}}^{\tilde{b}_0}\frac{\mathrm{d}\tilde{W}_{01}(z)}{\tilde{B}_0^3(z)}\equiv\int_{a_{W_0}}^{F_0^{-1}(\tilde{b}_0)}\frac{\mathrm{d}W_{01}(z)}{B_0^3(z)}<\infty.$$

Then, we have, uniformly in $a_{\tilde{W}_0}\le t\le\tilde{b}_0$

$$\tilde{U}_{0n_0}(t)-t=\tilde{L}_{0n_0}(t)+\tilde{s}_{0n_0}(t),$$

where

$$\tilde{L}_{0n_0}(t)=\frac{1}{n_0}\sum_{i=1}^{n_0}\tilde{\xi}_{0i}(Y_{0i},T_{0i},\delta_{0i},t),$$

$$\begin{aligned} \tilde{\xi}_{0i}(Y_{0i},T_{0i},\delta_{0i},t)\ =\ &(1-t)\left\{\frac{1_{\left\{Y_{0i}\le F_0^{-1}(t),\delta_{0i}=1\right\}}}{B_0(Y_{0i})}\right.\\ &\left.-\int_{a_{W_0}}^{F_0^{-1}(t)}\frac{1_{\{T_{0i}\le v\le Y_{0i}\}}}{B_0^2(v)}\mathrm{d}W_{01}(v)\right\}, \end{aligned}$$

$$\sup_{a_{\tilde{W}_0}\le t\le\tilde{b}_0}|\tilde{s}_{0n_0}(t)|=O\left(n_0^{-1}\ln\ln n_0\right)\quad\text{a.s.}$$

Once again $\tilde{\xi}_{0i}(Y_{0i},T_{0i},\delta_{0i},x)$ are zero mean iid processes with covariance structure given by

$$\begin{aligned} Cov\Big(\tilde{\xi}_{01}(Y_{01},T_{01},\delta_{01},x_1),\tilde{\xi}_{01}(Y_{01},T_{01},\delta_{01},x_2)\Big)&=(1-x_1)(1-x_2)\int_{a_{\tilde{W}_0}}^{x_1\wedge x_2}\left[\tilde{B}_0(v)\right]^{-2}\mathrm{d}\tilde{W}_{01}(v)\\ &=(1-x_1)(1-x_2)\,Q_0(x_1\wedge x_2)\,, \end{aligned}$$

where $Q_0(s) = q_0\left(F_0^{-1}(s)\right)$ with

$$q_0(t) = \int_{a_{W_0}}^{t} \frac{\mathrm{d}W_{01}(u)}{B_0^2(u)}.$$

3.1.1 Asymptotic properties

Consider a fixed value of t. Using a Taylor expansion the following expression for $\check{r}_h(t)$ is obtained:

$$\begin{aligned} \check{r}_h(t) &= \frac{1}{h}\int K\left(\frac{t-F_0(y)}{h}\right)\mathrm{d}\hat{F}_{1n_1}(y) \\ &+\frac{1}{h}\int \frac{F_0(y)-\hat{F}_{0n_0}(y)}{h}K^{(1)}\left(\frac{t-F_0(y)}{h}\right)\mathrm{d}\hat{F}_{1n_1}(y) \\ &+\frac{1}{h}\int \frac{(F_0(y)-\hat{F}_{0n_0}(y))^2}{2h^2}K^{(2)}(\Delta_{ty})\mathrm{d}\hat{F}_{1n_1}(y) \end{aligned} \quad (3.4)$$

with Δ_{ty} a value between $\frac{t-\hat{F}_{0n_0}(y)}{h}$ and $\frac{t-F_0(y)}{h}$.

For further discussion we now state some conditions that will be used along the proofs.

(B5) The bandwidth h satisfies $h \to 0$ and $\frac{n_1 h^3}{(\ln\ln n_1)^2} \to \infty$.

(E1) The endpoints of the supports satisfy $a_{G_1} < a_{W_0} < b_{W_1} < b_{W_0}$, $a_{\tilde{G}_1} < a_{\tilde{W}_1}$ and $a_{\tilde{G}_0} < a_{\tilde{W}_0}$.

(I1) There exist some $\tilde{b}_1 < b_{\tilde{W}_1}$ and $\tilde{b}_0 < b_{\tilde{W}_0}$ such that

$$\int_{a_{\tilde{W}_1}}^{\tilde{b}_1} \frac{\mathrm{d}\tilde{W}_{11}(s)}{\tilde{B}_1^{3}(s)} < \infty \quad \text{and} \quad \int_{a_{\tilde{W}_0}}^{\tilde{b}_0} \frac{\mathrm{d}\tilde{W}_{01}(s)}{\tilde{B}_0^3(s)} < \infty.$$

(K11) K is a twice differentiable density function on $[-1,1]$ with $K^{(2)}$ bounded.

(K12) K is a differentiable density function on $[-1,1]$ with $K^{(1)}$ bounded.

(Q2) Q_0 is differentiable in a neighborhood of t, with $Q_0^{(1)}(t)$ continuous at t.

(Q3) The functions Q_0 and Q_1 are twice continuously differentiable at t.

(R3) The relative distribution function, R, is Lipschitz continuous in a neighborhood of t with Lipschitz constant C_L.

(R4) The relative density, r, is twice continuously differentiable at t.

(R5) The relative density, r, is continuously differentiable at t.

Assumptions (K11), (K12), (Q2), (Q3), (R3), (R4) and (R5) are standard regularity conditions on the kernel and on some populational functions, while (B5) is a typical assumption about the sample size and the smoothing parameter. Condition (E1) is needed for identifiability and (I1) is needed to bound negligible terms.

Based on (3.4), it is easy to rewrite

$$\check{r}_h(t) = A_{n_1}(t) + B_{n_0,n_1}(t) + R_{n_0,n_1}(t)$$

with

$$A_{n_1}(t) = \frac{1}{h}\int K\left(\frac{t-v}{h}\right) \mathrm{d}\tilde{R}_{n_1}(v),$$
$$B_{n_0,n_1}(t) = \frac{1}{h^2}\int \left(v - \tilde{U}_{0n_0}(v)\right) K^{(1)}\left(\frac{t-v}{h}\right) \mathrm{d}\tilde{R}_{n_1}(v),$$
$$R_{n_0,n_1}(t) = \frac{1}{2h^3}\int \left(F_0(y) - \hat{F}_{0n_0}(y)\right)^2 K^{(2)}(\Delta_{ty})\, \mathrm{d}\hat{F}_{1n_1}(y),$$

where $\tilde{U}_{0n_0}$ and $\tilde{R}_{n_1}$ were introduced previously in (3.3) and (3.2).

Applying Lemma 3.1.2 we obtain

$$A_{n_1}(t) = A_{n_1}^{(1)}(t) + A_{n_1}^{(2)}(t) + A_{n_1}^{(3)}(t),$$

where

$$A_{n_1}^{(1)}(t) = \int K(u)\, r(t-hu)\, \mathrm{d}u,$$

$$A_{n_1}^{(2)}(t) = \frac{1}{n_1 h}\sum_{i=1}^{n_1} \int \tilde{\xi}_{1i}(Y_{1i}, T_{1i}, \delta_{1i}, t-hu)\, K^{(1)}(u)\, \mathrm{d}u,$$

$$A_{n_1}^{(3)}(t) = O\left(n_1^{-1}\ln\ln n_1\right) \quad \text{a.s.}$$

Consider the following decomposition for $B_{n_0,n_1}(t)$

$$B_{n_0,n_1}(t) = B_{n_0,n_1}^{(1)}(t) + B_{n_0,n_1}^{(2)}(t),$$

where

$$B_{n_0,n_1}^{(1)}(t) = \frac{1}{h^2}\int \left(v - \tilde{U}_{0n_0}(v)\right) K^{(1)}\left(\frac{t-v}{h}\right) \mathrm{d}R(v),$$
$$B_{n_0,n_1}^{(2)}(t) = \frac{1}{h^2}\int \left(v - \tilde{U}_{0n_0}(v)\right) K^{(1)}\left(\frac{t-v}{h}\right) \mathrm{d}\left(\tilde{R}_{n_1} - R\right)(v).$$

Applying Lemma 3.1.3, a further decomposition is obtained for both terms in $B_{n_0,n_1}(t)$:

$$B_{n_0,n_1}^{(1)}(t) = B_{n_0,n_1}^{(11)}(t) + B_{n_0,n_1}^{(12)}(t),$$

where

$$\begin{aligned} B^{(11)}_{n_0,n_1}(t) &= -\frac{1}{n_0h^2}\sum_{i=1}^{n_0}\int \tilde{\xi}_{0i}(Y_{0i},T_{0i},\delta_{0i},v)\,K^{(1)}\left(\frac{t-v}{h}\right)r(v)\,\mathrm{d}v \\ &= -\frac{1}{n_0h}\sum_{i=1}^{n_0}\int \tilde{\xi}_{0i}(Y_{0i},T_{0i},\delta_{0i},t-hu)\,K^{(1)}(u)\,r(t-hu)\,\mathrm{d}u, \\ B^{(12)}_{n_0,n_1}(t) &= -\frac{1}{h^2}\int \tilde{s}_{0n_0}(v)\,K^{(1)}\left(\frac{t-v}{h}\right)r(v)\,\mathrm{d}v \qquad (3.5) \end{aligned}$$

and

$$B^{(2)}_{n_0,n_1}(t) = B^{(21)}_{n_0,n_1}(t) + B^{(22)}_{n_0,n_1}(t),$$

where

$$B^{(21)}_{n_0,n_1}(t) = -\frac{1}{h^2}\int \tilde{L}_{0n_0}(v)\,K^{(1)}\left(\frac{t-v}{h}\right)\mathrm{d}\left(\tilde{R}_{n_1}-R\right)(v), \qquad (3.6)$$

$$B^{(22)}_{n_0,n_1}(t) = -\frac{1}{h^2}\int \tilde{s}_{0n_0}(v)\,K^{(1)}\left(\frac{t-v}{h}\right)\mathrm{d}\left(\tilde{R}_{n_1}-R\right)(v). \qquad (3.7)$$

Let us state the main result.

Theorem 3.1.4. *Assume conditions (S2), (B5), (E1), (I1), (K11), (Q2) and (R3) and let* $t < \min\left\{\tilde{b}_0, \tilde{b}_1\right\}$, *then*

$$\check{r}_h(t) = A^{(1)}_{n_1}(t) + A^{(2)}_{n_1}(t) + B^{(11)}_{n_0,n_1}(t) + C_{n_0,n_1}(t)$$

with $C_{n_0,n_1}(t) = o_P\left((n_1h)^{-\frac{1}{2}}\right)$ *and* $E\left[|C_{n_0,n_1}(t)|^d\right] = o\left((n_1h)^{-\frac{d}{2}}\right)$ *for* $d = 1, 2$.

Note that C_{n_0,n_1} is the sum of all the remainder terms, i.e.

$$C_{n_0,n_1}(t) = A^{(3)}_{n_1}(t) + B^{(12)}_{n_0,n_1}(t) + B^{(21)}_{n_0,n_1}(t) + B^{(22)}_{n_0,n_1}(t) + R_{n_0,n_1}(t).$$

Proof of Theorem 3.1.4. First we will study the term $B^{(12)}_{n_0,n_1}(t)$. We will show that $B^{(12)}_{n_0,n_1}(t) = O\left((n_0h)^{-1}\ln\ln n_0\right)$ a.s.

Recall (3.5). Since K has support on $[-1, 1]$,

$$\begin{aligned} \left|B^{(12)}_{n_0,n_1}(t)\right| &\le \frac{1}{h^2}\left\|K^{(1)}\right\|_\infty \sup_{t-h\le v\le t+h}\left|\tilde{s}_{0n_0}(v)\right| \int_{t-h}^{t+h} r(v)\,\mathrm{d}v \\ &= \frac{1}{h^2}\left\|K^{(1)}\right\|_\infty \sup_{t-h\le v\le t+h}\left|\tilde{s}_{0n_0}(v)\right| \left[R(t+h)-R(t-h)\right]. \end{aligned}$$

Based on the Lipschitz continuity of R and using Lemma 3.1.3

$$\begin{aligned} \left|B^{(12)}_{n_0,n_1}(t)\right| &\le \frac{2C_L}{h}\left\|K^{(1)}\right\|_\infty \sup_{t-h\le v\le t+h}\left|\tilde{s}_{0n_0}(v)\right| \\ &= O\left((n_0h)^{-1}\ln\ln n_0\right) \quad \text{a.s.} \end{aligned}$$

Below we will study the terms $B^{(21)}_{n_0,n_1}(t)$ and $B^{(22)}_{n_0,n_1}(t)$. We will show that

$$
\begin{aligned}
B^{(21)}_{n_0,n_1}(t) &= o_P\left((n_1 h)^{-\frac{1}{2}}\right), \\
B^{(22)}_{n_0,n_1}(t) &= O\left(\frac{\ln\ln n_0}{n_0 h^2}\left[h + \left\{\frac{(\ln n_1)^2}{n_1} \vee \left(\frac{h\ln n_1}{n_1}\right)^{\frac{1}{2}}\right\}\right]\right) \quad \text{a.s.}
\end{aligned}
$$

To deal with $B^{(21)}_{n_0,n_1}(t)$ in equation (3.6), first we will find a bound for the order of $E\left[B^{(21)}_{n_0,n_1}(t)^2\right]$. Using the covariance structure of $\tilde{\xi}_{0i}(Y_{0i}, T_{0i}, \delta_{0i}, t)$, taking into account the independence between the samples $\{(X_{0i}, T_{0i}, C_{0i})\}$ and $\{(X_{1j}, T_{1j}, C_{1j})\}$, and considering conditional expectations, we finally obtain that

$$
\begin{aligned}
E\left[\left(B^{(21)}_{n_0,n_1}(t)\right)^2\right] &= E\left[\frac{2}{n_0 h^4}\int_{t-h}^{t+h}\left[\int_u^{t+h}(1-v)K^{(1)}\left(\frac{t-v}{h}\right)\mathrm{d}\left(\tilde{R}_{n_1}-R\right)(v)\right]\right. \\
&\quad \left.(1-u)Q_0(u)K^{(1)}\left(\frac{t-u}{h}\right)\mathrm{d}\left(\tilde{R}_{n_1}-R\right)(u)\right].
\end{aligned}
$$

Using integration by parts we can rewrite the inner integral as follows:

$$
\begin{aligned}
&\int_u^{t+h}(1-v)K^{(1)}\left(\frac{t-v}{h}\right)\mathrm{d}\left(\tilde{R}_{n_1}-R\right)(v) \\
&= \frac{1}{h}\int_u^{t+h}\left(\tilde{R}_{n_1}-R\right)(v)(1-v)K^{(2)}\left(\frac{t-v}{h}\right)\mathrm{d}v \\
&- (1-u)K^{(1)}\left(\frac{t-u}{h}\right)\left(\tilde{R}_{n_1}-R\right)(u) \\
&+ \int_u^{t+h}\left(\tilde{R}_{n_1}-R\right)(v)K^{(1)}\left(\frac{t-v}{h}\right)\mathrm{d}v.
\end{aligned}
$$

Consequently

$$
E\left[\left(B^{(21)}_{n_0,n_1}(t)\right)^2\right] = \frac{2}{n_0 h^4}\left\{E\left[I_1(t)\right] - E\left[I_2(t)\right] + E\left[I_3(t)\right]\right\}
$$

with

$$
\begin{aligned}
I_1(t) &= \int_{t-h}^{t+h}\left\{\frac{1}{h}\int_u^{t+h}(1-v)K^{(2)}\left(\frac{t-v}{h}\right)\left(\tilde{R}_{n_1}-R\right)(v)\,\mathrm{d}v\right\} \\
&\quad Q_0(u)(1-u)K^{(1)}\left(\frac{t-u}{h}\right)\mathrm{d}\left(\tilde{R}_{n_1}-R\right)(u),
\end{aligned}
$$

$$
\begin{aligned}
I_2(t) &= \int_{t-h}^{t+h}Q_0(u)(1-u)^2K^{(1)}\left(\frac{t-u}{h}\right)^2 \\
&\quad \left(\tilde{R}_{n_1}-R\right)(u)\,\mathrm{d}\left(\tilde{R}_{n_1}-R\right)(u),
\end{aligned}
$$

$$\begin{aligned} I_3(t) &= \int_{t-h}^{t+h} \left\{ \int_{u}^{t+h} K^{(1)}\left(\frac{t-v}{h}\right) \left(\tilde{R}_{n_1} - R\right)(v)\,\mathrm{d}v \right\} \\ &\quad Q_0(u)(1-u) K^{(1)}\left(\frac{t-u}{h}\right) \mathrm{d}\left(\tilde{R}_{n_1} - R\right)(u). \end{aligned}$$

Rewriting $I_2(t)$ as follows

$$I_2(t) = \frac{1}{2} \int_{t-h}^{t+h} Q_0(u)(1-u)^2 K^{(1)}\left(\frac{t-u}{h}\right)^2 \mathrm{d}\left(\tilde{R}_{n_1} - R\right)^2(u)$$

and applying integration by parts, it is easy to obtain

$$I_2(t) = I_{21}(t) + I_{22}(t) + I_{23}(t),$$

where

$$\begin{aligned} I_{21}(t) &= \frac{1}{2} \int_{t-h}^{t+h} \left(\tilde{R}_{n_1} - R\right)^2(u)\, 2Q_0(u)(1-u)\; K^{(1)}\left(\frac{t-u}{h}\right)^2 \mathrm{d}u, \\ I_{22}(t) &= -\frac{1}{2} \int_{t-h}^{t+h} \left(\tilde{R}_{n_1} - R\right)^2(u)\, Q_0^{(1)}(u)(1-u)^2\; K^{(1)}\left(\frac{t-u}{h}\right)^2 \mathrm{d}u, \\ I_{23}(t) &= \frac{1}{h} \int_{t-h}^{t+h} \left(\tilde{R}_{n_1} - R\right)^2(u)\, Q_0(u)(1-u)^2 K^{(1)}\left(\frac{t-u}{h}\right) K^{(2)}\left(\frac{t-u}{h}\right) \mathrm{d}u. \end{aligned}$$

Standard algebra gives

$$\begin{aligned} E\left[|I_2(t)|\right] &\leq E\left[\sup_{a_{\tilde{W}} \leq x \leq \tilde{b}_1} \left|\left(\tilde{R}_{n_1} - R\right)(x)\right|^2\right] \cdot \left\{ \frac{h}{2} \left\|Q_0^{(1)}\right\|_\infty \int K^{(1)}(u)^2 \mathrm{d}u \right. \\ &\quad \left. + \left\|Q_0\right\|_\infty \left\|K^{(1)}\right\|_\infty \int \left|K^{(2)}(u)\right| \mathrm{d}u + h \left\|Q_0\right\|_\infty \int K^{(1)}(u)^2 \mathrm{d}u \right\}. \end{aligned}$$

Theorem 1 in Zhu (1996) can be applied for $a_{\tilde{G}_1} < a_{\tilde{W}_1} \leq \tilde{b}_1 < b_{\tilde{W}_1}$ and $\varepsilon > 0$, to obtain

$$P\left(\sup_{a_{\tilde{W}_1} \leq x \leq \tilde{b}_1} \left|\left(\tilde{R}_{n_1} - R\right)(x)\right| > \varepsilon\right) \leq C_1 \exp\left\{-n_1 D_1 \varepsilon^2\right\}, \text{ for some } C_1, D_1 > 0.$$

Consequently,

$$\begin{aligned} E\left[\sup_{a_{\tilde{W}} \leq x \leq \tilde{b}} \left|\left(\tilde{R}_{n_1} - R\right)(x)\right|^2\right] &= \int_0^\infty P\left(\sup_{a_{\tilde{W}} \leq x \leq \tilde{b}} \left|\left(\tilde{R}_{n_1} - R\right)(x)\right| > u^{\frac{1}{2}}\right) \mathrm{d}u \\ &\leq C_1 \int_0^\infty \exp\left\{-n_1 D_1 u\right\} \mathrm{d}u = O\left(\frac{1}{n_1}\right), \end{aligned}$$

which implies that $E\left[|I_2(t)|\right] = O\left(\frac{1}{n_1}\right)$.

Using integration by parts, the Cauchy-Schwarz inequality and conditions (K11) and (Q2), we can show that $E\left[|I_1(t) + I_3(t)|\right] = O\left(\frac{1}{n_1}\right)$.

These calculations are shown below in some more detail:

$$\begin{aligned}
E[I_1(t)] &= E\left[\int_{t-h}^{t+h} \frac{1}{h}\int_u^{t+h}(1-v)K^{(2)}\left(\frac{t-v}{h}\right)(\tilde{R}_{n_1}-R)(v)\mathrm{d}v\right.\\
&\quad \left. Q_0(u)(1-u)K^{(1)}\left(\frac{t-u}{h}\right)\mathrm{d}(\tilde{R}_{n_1}-R)(u)\right]\\
&= E\left[\frac{1}{h}\right]\int_u^{t+h}(1-v)K^{(2)}\left(\frac{t-v}{h}\right)(\tilde{R}_{n_1}-R)(v)\mathrm{d}v\\
&\quad Q_0(u)(1-u)K^{(1)}\left(\frac{t-u}{h}\right)(\tilde{R}_{n_1}-R)(u)\Big]_{u=t-h}^{u=t+h}\Big]\\
&\quad -E\left[\int_{t-h}^{t+h}(\tilde{R}_{n_1}-R)(u)\frac{\partial}{\partial u}\left[Q_0(u)(1-u)K^{(1)}\left(\frac{t-u}{h}\right)\right.\right.\\
&\quad \left.\left.\frac{1}{h}\int_u^{t+h}(1-v)K^{(2)}\left(\frac{t-v}{h}\right)(\tilde{R}_{n_1}-R)(v)\mathrm{d}v\right]\right].
\end{aligned}$$

Then $E[I_1(t)] = E[I_{11}(t)] - E[I_{12}(t)] + E[I_{13}(t)] + E[I_{14}(t)]$, where

$$\begin{aligned}
I_{11}(t) &= \int_{t-h}^{t+h}(\tilde{R}_{n_1}-R)(u)\left\{\frac{1}{h}(1-u)K^{(2)}\left(\frac{t-u}{h}\right)(\tilde{R}_{n_1}-R)(u)\right\}\\
&\quad Q_0(u)(1-u)K^{(1)}\left(\frac{t-u}{h}\right)\mathrm{d}u,
\end{aligned}$$

$$\begin{aligned}
I_{12}(t) &= \int_{t-h}^{t+h}(\tilde{R}_{n_1}-R)(u)\left\{\frac{1}{h}\int_u^{t+h}(1-v)K^{(2)}\left(\frac{t-v}{h}\right)(\tilde{R}_{n_1}-R)(v)\mathrm{d}v\right\}\\
&\quad Q_0^{(1)}(u)(1-u)K^{(1)}\left(\frac{t-u}{h}\right)\mathrm{d}u,
\end{aligned}$$

$$\begin{aligned}
I_{13}(t) &= \int_{t-h}^{t+h}(\tilde{R}_{n_1}-R)(u)\left\{\frac{1}{h}\int_u^{t+h}(1-v)K^{(2)}\left(\frac{t-v}{h}\right)(\tilde{R}_{n_1}-R)(v)\mathrm{d}v\right\}\\
&\quad Q_0(u)K^{(1)}\left(\frac{t-u}{h}\right)\mathrm{d}u,\\
I_{14}(t) &= \int_{t-h}^{t+h}(\tilde{R}_{n_1}-R)(u)\left\{\frac{1}{h}\int_u^{t+h}(1-v)K^{(2)}\left(\frac{t-v}{h}\right)(\tilde{R}_{n_1}-R)(v)\mathrm{d}v\right\}\\
&\quad Q_0(u)(1-u)\frac{1}{h}K^{(2)}\left(\frac{t-u}{h}\right)\mathrm{d}u.
\end{aligned}$$

Now we study the term $E[I_{11}(t)]$:

$$\begin{aligned}
E[|I_{11}(t)|] &\leq E\left[\int_{t-h}^{t+h}|(\tilde{R}_{n_1}-R)(u)|\frac{1}{h}|1-u|\left|K^{(2)}\left(\frac{t-u}{h}\right)\right||(\tilde{R}_{n_1}-R)(u)|\right|\\
&\quad \left. Q_0(u)||1-u|\left|K^{(1)}\left(\frac{t-u}{h}\right)\right|\mathrm{d}u\right]
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{h} E\left[\sup |(\tilde{R}_{n_1}-R)(u)|^2 \|Q_0\|_\infty \int_{t-h}^{t+h} \left|K^{(2)}\left(\frac{t-u}{h}\right)\right| \cdot \left|K^{(1)}\left(\frac{t-u}{h}\right)\right| \mathrm{d}u\right] \\
&= E\left[\sup |(\tilde{R}_{n_1}-R)(u)|^2 \|Q_0\|_\infty \int_{-1}^{1} \left|K^{(2)}(z)\right| \cdot \left|K^{(1)}(z)\right| \mathrm{d}z\right] \\
&= E\left[\sup |(\tilde{R}_{n_1}-R)(u)|^2\right] \|Q_0\|_\infty \int_{-1}^{1} \left|K^{(2)}(z)\right| \cdot \left|K^{(1)}(z)\right| \mathrm{d}z.
\end{aligned}
$$

Therefore $E[|I_{11}(t)|] = O\left(\frac{1}{n_1}\right)$.

In order to study $E[|I_{12}(t)|]$ we start from the inner integral in $I_{12}(t)$:

$$
\begin{aligned}
&\left|\int_u^{t+h} (1-v) K^{(2)}\left(\frac{t-v}{h}\right) (\tilde{R}_{n_1}-R)(v) \mathrm{d}v\right| \\
&\leq \int_u^{t+h} |1-v| \left|K^{(2)}\left(\frac{t-v}{h}\right)\right| \left|(\tilde{R}_{n_1}-R)(v)\right| \mathrm{d}v \\
&\leq \sup \left|(\tilde{R}_{n_1}-R)(v)\right| h \int \left|K^{(2)}(z)\right| \mathrm{d}z \\
&= h \sup \left|(\tilde{R}_{n_1}-R)(v)\right| \int \left|K^{(2)}(z)\right| \mathrm{d}z.
\end{aligned}
$$

Then

$$
\begin{aligned}
&E[|I_{12}(t)|] \\
&\leq E\left[\int_{t-h}^{t+h} \sup \left|(\tilde{R}_{n_1}-R)(u)\right| \right. \\
&\left. \int \left|K^{(2)}(z)\right| \mathrm{d}z \sup |(\tilde{R}_{n_1}-R)(u)| \left\|Q_0^{(1)}\right\|_\infty \left|K^{(1)}\left(\frac{t-u}{h}\right)\right| \mathrm{d}u\right] \\
&\leq hE\left[\sup |(\tilde{R}_{n_1}-R)(u)|^2 \left\|Q_0^{(1)}\right\|_\infty \left(\int \left|K^{(2)}(z)\right| \mathrm{d}z\right) \left(\int \left|K^{(1)}(z)\right| \mathrm{d}z\right)\right] \\
&\leq h \left\|Q_0^{(1)}\right\|_\infty \left(\int \left|K^{(2)}(z)\right| \mathrm{d}z\right) \left(\int \left|K^{(1)}(z)\right| \mathrm{d}z\right) E\left[\sup |(\tilde{R}_{n_1}-R)(u)|^2\right] = O\left(\frac{h}{n_1}\right),
\end{aligned}
$$

which implies that $E[|I_{12}(t)|] = O\left(\frac{h}{n_1}\right) = o\left(\frac{1}{n_1}\right)$.

In order to study $E[|I_{13}(t)|]$ we note that the inner integral in it is the same as the one appearing in $I_{12}(t)$, which has been previously studied. Then

$$
\begin{aligned}
&E[|I_{13}(t)|] \\
&\leq E\left[\int_{t-h}^{t+h} \sup |(\tilde{R}_{n_1}-R)(u)| \int \left|K^{(2)}(v)\right| \mathrm{d}v \sup |(\tilde{R}_{n_1}-R)(u)| \|Q_0\|_\infty \left|K^{(1)}\left(\frac{t-u}{h}\right)\right| \mathrm{d}u\right] \\
&\leq hE\left[\sup |(\tilde{R}_{n_1}-R)(u)|^2 \|Q_0\|_\infty \left(\int \left|K^{(2)}(z)\right| \mathrm{d}z\right) \left(\int \left|K^{(1)}(z)\right| \mathrm{d}z\right)\right] \\
&\leq hE\left[\sup |(\tilde{R}_{n_1}-R)(u)|^2\right] \|Q_0\|_\infty \left(\int \left|K^{(2)}(z)\right| \mathrm{d}z\right) \left(\int \left|K^{(1)}(z)\right| \mathrm{d}z\right) \\
&= O\left(\frac{h}{n_1}\right) = O\left(\frac{1}{n_1}\right).
\end{aligned}
$$

Proceeding in a similar way we can show that $E[|I_{14}(t)|] = O\left(\frac{1}{n_1}\right)$. Parallel algebra also shows that $E[|I_3(t)|] = O\left(\frac{1}{n_1}\right)$. Therefore, we can conclude that $E\left[B^{(21)}_{n_0,n_1}(t)^2\right] = O\left(\frac{1}{n_1^2 h^4}\right)$, which implies $(n_1 h)^{\frac{1}{2}} B^{(21)}_{n_0,n_1}(t) = o_P(1)$, using conditions (S2) and (B5).

The term $B^{(22)}_{n_0,n_1}(t)$ in (3.7) can be bounded as follows:

$$\begin{aligned}\left|B^{(22)}_{n_0,n_1}(t)\right| &\leq \frac{1}{h^2}\left\|K^{(1)}\right\|_\infty \sup_{t-h\leq v\leq t+h}\left|\tilde{s}_{0n_0}(v)\right| \\ &\quad \left\{\left(\tilde{R}_{n_1}(t+h)-\tilde{R}_{n_1}(t-h)\right)+\left(R(t+h)-R(t-h)\right)\right\} \\ &\leq B^{(221)}_{n_0,n_1}(t)+B^{(222)}_{n_0,n_1}(t),\end{aligned}$$

where

$$\begin{aligned}B^{(221)}_{n_0,n_1}(t) &= \frac{1}{h^2}\left\|K^{(1)}\right\|_\infty \sup_{t-h\leq v\leq t+h}\left|\tilde{s}_{0n_0}(v)\right|\left|2\left(R(t+h)-R(t-h)\right)\right|, \\ B^{(222)}_{n_0,n_1}(t) &= \frac{1}{h^2}\left\|K^{(1)}\right\|_\infty \sup_{t-h\leq v\leq t+h}\left|\tilde{s}_{0n_0}(v)\right| \\ &\quad \cdot\left|\tilde{R}_{n_1}(t+h)-\tilde{R}_{n_1}(t-h)-\left(R(t+h)-R(t-h)\right)\right|.\end{aligned}$$

Using the Lipschitz-continuity of R and Lemma 3.1.3 it follows that

$$B^{(221)}_{n_0,n_1}(t) = O\left(\frac{\ln\ln n_0}{n_0 h}\right) \quad \text{a.s.}$$

Lemma 3.1 in Zhou and Yip (1999) and Lemma 3.1.3 imply

$$\begin{aligned}\left|B^{(222)}_{n_0,n_1}(t)\right| &= O\left(\frac{\ln\ln n_0}{n_0 h^2}\right) 2\sup_{|s|\leq 2h}\left|\tilde{R}_{n_1}(t+s)-\tilde{R}_{n_1}(t)-(R(t+s)-R(t))\right| \\ &= O\left(\frac{\ln\ln n_0}{n_0 h^2}\right) O\left(\frac{n_1^{-\frac{1}{2}}(\ln n_1)^2}{n_1^{\frac{1}{2}}} \vee \frac{(h(-\ln h))^{\frac{1}{2}}}{n_1^{\frac{1}{2}}}\right) \\ &= O\left(\frac{\ln\ln n_0}{n_0 h^2}\right)\left\{\frac{(\ln n_1)^2}{n_1} \vee \frac{(h\ln n_1)^{\frac{1}{2}}}{n_1^{\frac{1}{2}}}\right\} \quad \text{a.s.,}\end{aligned}$$

where $\vee$ denotes maximum.

Consequently, we conclude from the previous results that

$$\left|B^{(22)}_{n_0,n_1}(t)\right| = O\left(\frac{\ln\ln n_0}{n_0 h^2}\left\{h+\left(\frac{(\ln n_1)^2}{n_1} \vee \left(\frac{h\ln n_1}{n_1}\right)^{\frac{1}{2}}\right)\right\}\right) \quad \text{a.s.}$$

Below we will show that $R_{n_0,n_1}(t) = o_P\left((n_1 h)^{-\frac{1}{2}}\right)$:

$$\begin{aligned}\left|R_{n_0,n_1}(t)\right| &\leq \frac{1}{2h^3}\left\|K^{(2)}\right\|_\infty \Delta^2_{0n_0}\int 1\left\{|\Delta_{ty}|\leq 1\right\} \mathrm{d}\hat{F}_{1n_1}(y), \\ \Delta_{0n_0} &= \sup_{a_{F_0}\leq y\leq F_0^{-1}(\tilde{b}_0)}\left|\hat{F}_{0n_0}(y)-F_0(y)\right|.\end{aligned}$$

Since $F_0^{-1}\left(\tilde{b}_0\right) < b_{W_0}$, $a_{G_1} \leq a_{W_0}$ and $\int_{a_{W_0}}^{b_W} \frac{\mathrm{d}W_{01}(z)}{C_0^3(z)} < \infty$, we have, by Lemma 2.5 in Zhou and Yip (1999) and Lemma 3.1.1, the following order for Δ_{0n_0}, $\Delta_{0n_0} = O\left(\left(\frac{\ln\ln n_0}{n_0}\right)^{\frac{1}{2}}\right)$ a.s. We further have that

$$1_{\{|\Delta_{ty}|\leq 1\}} \leq 1_{\left\{t-h-\Delta_{0n_0}\leq F_0(y)\leq t+h+\Delta_{0n_0}\right\}}.$$

Then

$$\begin{aligned}\left|R_{n_0,n_1}(t)\right| &\leq \frac{1}{2h^3}\left\|K^{(2)}\right\|_\infty \Delta_{0n_0}^2 \\ &\quad \left\{\hat{F}_{1n_1}\left(F_0^{-1}\left(t+h+\Delta_{0n_0}\right)\right) - \hat{F}_{1n_1}\left(F_0^{-1}\left(t-h-\Delta_{0n_0}\right)\right)\right\} \\ &= R_{n_0,n_1}^{(1)}(t) + R_{n_0,n_1}^{(2)}(t)\end{aligned}$$

with

$$\begin{aligned}R_{n_0,n_1}^{(1)}(t) &= \frac{1}{2h^3}\left\|K^{(2)}\right\|_\infty \Delta_{0n_0}^2 \\ &\quad \left\{\hat{F}_{1n_1}\left(F_0^{-1}\left(t+h+\Delta_{0n_0}\right)\right) - F_1\left(F_0^{-1}\left(t+h+\Delta_{0n_0}\right)\right)\right\} \\ &\quad -\frac{1}{2h^3}\left\|K^{(2)}\right\|_\infty \Delta_{0n_0}^2 \\ &\quad \left\{\hat{F}_{1n_1}\left(F_0^{-1}\left(t-h-\Delta_{0n_0}\right)\right) - F_1\left(F_0^{-1}\left(t-h-\Delta_{0n_0}\right)\right)\right\}\end{aligned}$$

$$\begin{aligned}R_{n_0,n_1}^{(2)}(t) &= \frac{1}{2h^3}\left\|K^{(2)}\right\|_\infty \Delta_{0n_0}^2 \\ &\quad \left\{F_1\left(F_0^{-1}\left(t+h+\Delta_{0n_0}\right)\right) - F_1\left(F_0^{-1}\left(t-h-\Delta_{0n_0}\right)\right)\right\}.\end{aligned}$$

The identity $R = F_1 \circ F_0^{-1}$ and condition (R3) imply

$$\left|R_{n_0,n_1}^{(2)}(t)\right| \leq \frac{1}{h^3}\left\|K^{(2)}\right\|_\infty \Delta_{0n_0}^2 C_L\left(h+\Delta_{0n_0}\right).$$

Using the order of Δ_{0n_0} and condition (B5), it follows that $(n_1 h)^{\frac{1}{2}} R_{n_0,n_1}^{(2)}(t) = o_P(1)$. On the other hand,

$$\begin{aligned}R_{n_0,n_1}^{(1)}(t) &\leq \frac{1}{2h^3}\left\|K^{(2)}\right\|_\infty \Delta_{0n_0}^2 \\ &\quad \sup_{|s|\leq h+\Delta_{0n_0}}\left|\tilde{R}_{n_1}(t+s) - \tilde{R}_{n_1}(t) - \left(R(t+s) - R(t)\right)\right| \\ &\leq \frac{1}{2h^3}\left\|K^{(2)}\right\|_\infty \Delta_{0n_0}^2 \\ &\quad \sup_{|s|\leq 2h}\left|\tilde{R}_{n_1}(t+s) - \tilde{R}_{n_1}(t) - \left(R(t+s) - R(t)\right) + 1\left\{\Delta_{0n_0} > h\right\}\right|.\end{aligned}$$

Applying Lemma 3.1 in Zhou and Yip (1999) and using (B5) we obtain

$$\sup_{|s|\leq 2h}\left|\left(\tilde{R}_{n_1}(t+s)-\tilde{R}_{n_1}(t)\right)-(R(t+s)-R(t))\right|$$
$$=O\left(\frac{\ln^2 n_0}{n_0}\vee\left(\frac{h\ln n_1}{n_0}\right)^{\frac{1}{2}}\right)\quad\text{a.s.}$$

Now, the exponential bound in Theorem 1 in Zhu (1996), gives

$$1_{\left\{\Delta_{0n_0}>h\right\}}=O_P\left(\exp\left\{-n_0h^2\right\}\right).$$

As a final conclusion we obtain $(n_1h)^{\frac{1}{2}}R^{(1)}_{n_0,n_1}(t)=o_P(1)$.

Consequently, under conditions (B5), (E1), (I1), (K11), (Q2), (R3) and (S2) and for $t<\min\left\{\tilde{b}_0,\tilde{b}_1\right\}$, we have the following expression for $\check{r}_h(t)$:

$$\check{r}_h(t)=A^{(1)}_{n_1}(t)+A^{(2)}_{n_1}(t)+B^{(11)}_{n_0,n_1}(t)+C_{n_0,n_1}(t)$$

with $C_{n_0,n_1}(t)=o_P\left((n_1h)^{-\frac{1}{2}}\right)$. It remains to prove that $E\left[\left|C_{n_0,n_1}(t)\right|^d\right]=o\left((n_1h)^{-\frac{d}{2}}\right)$, for $d=1,2$.

First of all, note that,

$$\begin{aligned}C_{n_0,n_1}(t)&=A^{(3)}_{n_1}(t)+B^{(12)}_{n_0,n_1}(t)+B^{(21)}_{n_0,n_1}(t)+B^{(22)}_{n_0,n_1}(t)+R_{n_0,n_1}(t)\\&=A^{(3)}_{n_1}(t)+B^{(12)}_{n_0,n_1}(t)+B^{(2)}_{n_0,n_1}(t)+R_{n_0,n_1}(t).\end{aligned}$$

Using Theorem 1 (c) in Gijbels and Wang (1993) it can be proved that

$$E\left[\sup_{t-h\leq v\leq t+h}\left|\tilde{s}_{0n_0}(v)\right|\right]=O\left(\frac{1}{n_0}\right).$$

Now, condition (R3) and

$$E\left[\left|B^{(12)}_{n_0,n_1}(t)\right|\right]\leq\frac{1}{h^2}\left\|K^{(1)}\right\|_\infty\left[R(t+h)-R(t-h)\right]E\left[\sup_{t-h\leq v\leq t+h}\left|\tilde{s}_{0n_0}(v)\right|\right]$$

imply $E\left[\left|B^{(12)}_{n_0,n_1}(t)\right|\right]=o\left(\frac{1}{\sqrt{n_0h}}\right)$. Similar arguments give $E\left[\left|A^{(3)}_{n_1}(t)\right|\right]=o\left(\frac{1}{\sqrt{n_0h}}\right)$.

Using Hölder inequality and condition (B5):

$$E\left[\left|B^{(21)}_{n_0,n_1}(t)\right|\right]\leq E\left[\left|B^{(21)}_{n_0,n_1}\right|^2\right]^{\frac{1}{2}}=o\left(\frac{1}{\sqrt{n_0h}}\right).$$

Secondly, we study the term $B^{(22)}_{n_0,n_1}(t)$. Using similar arguments as above, Theorem 1 (c) in Gijbels and Wang (1993), $E\left[\sup\left|\tilde{R}_{n_1}(v)-R(v)\right|\right]=O\left(\frac{1}{\sqrt{n_1}}\right)$ and condition (B5), it is easy to prove that:

$$E\left[\left|B^{(22)}_{n_0,n_1}(t)\right|\right] \leq \frac{2}{h^2}\left\|K^{(1)}\right\|_\infty E\left[\sup_{t-h\leq v\leq t+h}|\tilde{s}_{0n_0}(v)|\right]|(R(t+h)-R(t-h))|$$
$$+\frac{2}{h^2}\left\|K^{(1)}\right\|_\infty E\left[\sup_{t-h\leq v\leq t+h}|\tilde{s}_{0n_0}(v)|\right]E\left[\sup\left|\tilde{R}_{n_1}(v)-R(v)\right|\right] = o\left(\frac{1}{\sqrt{n_0h}}\right).$$

To deal with $R_{n_0,n_1}(t)$ we use some previous results:

$$\begin{aligned}
R^{(1)}_{n_0,n_1}(t) &\leq \frac{1}{2h^3}\left\|K^{(2)}\right\|_\infty \Delta^2_{0n_0} \\
&\quad \sup_{|s|\leq 2h}\left|\tilde{R}_{n_1}(t+s)-\tilde{R}_{n_1}(t)-(R(t+s)-R(t))+1_{\{\Delta_{0n_0}>h\}}\right|, \\
1_{\{\Delta_{0n_0}>h\}} &= O_P\left(\exp\left\{-n_0h^2\right\}\right), \\
\left|R^{(2)}_{n_0,n_1}(t)\right| &\leq \frac{1}{h^3}\left\|K^{(2)}\right\|_\infty \Delta^2_{0n_0}C_L(h+\Delta_{0n_0}).
\end{aligned}$$

By using similar arguments as above, it can be proved that $E\left[\Delta^2_{0n_0}\right] = O\left(\frac{1}{n_0}\right)$. Now the order $E\left[\sup\left|\tilde{R}_{n_1}(v)-R(v)\right|\right] = O\left(\frac{1}{\sqrt{n_1}}\right)$ can be used to conclude

$$\begin{aligned}
E\left[R^{(1)}_{n_0,n_1}(t)\right] &\leq \frac{1}{2h^3}\left\|K^{(2)}\right\|_\infty E\left[\Delta^2_{0n_0}\right]2E\left[\sup\left|\tilde{R}_{n_1}(v)-R(v)\right|\right] \\
&\quad +\frac{1}{2h^3}\left\|K^{(2)}\right\|_\infty \left[E\left(\Delta^4_{0n_0}\right)\right]^{\frac{1}{2}}\left[P(\Delta_{0n_0}>h)\right]^{\frac{1}{2}} \\
&= O\left(\frac{1}{n_0^{3/2}h^3}\right)+O\left(\frac{\exp\left(-\frac{n_0h^2}{2}\right)}{n_0h^3}\right) = o\left(\frac{1}{\sqrt{n_0h}}\right).
\end{aligned}$$

On the other hand,

$$E\left[R^{(2)}_{n_0,n_1}(t)\right] = O\left(\frac{1}{n_0h^2}\right)+O\left(\frac{1}{n_0^{\frac{3}{2}}h^3}\right) = o\left(\frac{1}{\sqrt{n_0h}}\right).$$

Collecting all the bounds, $E\left[|C_{n_0,n_1}(t)|\right] = O\left((n_1h)^{-\frac{1}{2}}\right)$.

To bound the second moment of $C_{n_0,n_1}(t)$,

$$\begin{aligned}
C_{n_0,n_1}(t)^2 &= A^{(3)}_{n_1}(t)^2 + B^{(12)}_{n_0,n_1}(t)^2 + B^{(2)}_{n_0,n_1}(t)^2 + R_{n_0,n_1}(t)^2 \\
&\quad +2A^{(3)}_{n_1}(t)B^{(12)}_{n_0,n_1}(t)+2A^{(3)}_{n_1}(t)B^{(2)}_{n_0,n_1}(t)+2A^{(3)}_{n_1}(t)R_{n_0,n_1}(t) \\
&\quad +2B^{(12)}_{n_0,n_1}(t)B^{(2)}_{n_0,n_1}(t)+2B^{(12)}_{n_0,n_1}(t)R_{n_0,n_1}(t)+2B^{(2)}_{n_0,n_1}(t)R_{n_0,n_1}(t).
\end{aligned}$$

Using the Cauchy-Schwarz inequality, to bound $E\left[C_{n_0,n_1}(t)^2\right]$, it will be enough to study the main terms $E\left[A^{(3)}_{n_1}(t)^2\right]$, $E\left[B^{(12)}_{n_0,n_1}(t)^2\right]$, $E\left[B^{(2)}_{n_0,n_1}(t)^2\right]$ and $E\left[R_{n_0,n_1}(t)^2\right]$.

First of all

$$\begin{aligned} E\left[|B_{n_0,n_1}^{(12)}(t)|^2\right] &\leq \frac{1}{h^4}\left\|K^{(1)}\right\|_\infty^2 |R(t+h)-R(t-h)|^2 E\left[\sup|\tilde{s}_{0n_0}(v)|^2\right] \\ &= O\left(\frac{1}{n_0^2h^2}\right) = o\left(\frac{1}{n_0h}\right). \end{aligned}$$

Similarly, it is not difficult to prove that $E\left[A_{n_1}^{(3)}(t)^2\right] = o\left(\frac{1}{n_0h}\right)$.

To handle $E\left[B_{n_0,n_1}^{(2)}(t)^2\right]$, it is enough to study: $E\left[B_{n_0,n_1}^{(21)}(t)^2\right]$ and $E\left[B_{n_0,n_1}^{(22)}(t)^2\right]$. We already know that $E\left[B_{n_0,n_1}^{(21)}(t)^2\right] = O\left(\frac{1}{n_1^2h^4}\right) = o\left(\frac{1}{n_0h}\right)$. On the other hand

$$B_{n_0,n_1}^{(22)}(t)^2 \leq \left(B_{n_0,n_1}^{(221)}(t) + B_{n_0,n_1}^{(222)}(t)\right)^2.$$

Standard arguments give:

$$\begin{aligned} E\left[B_{n_0,n_1}^{(221)}(t)^2\right] &= O\left(\frac{1}{n_0^2h^2}\right) = o\left(\frac{1}{n_0h}\right), \\ E\left[B_{n_0,n_1}^{(221)}(t)^2\right] &= O\left(\frac{1}{n_0^3h^4}\right) = o\left(\frac{1}{n_0h}\right). \end{aligned}$$

Finally, the second moment of $R_{n_0,n_1}(t)$ can be bounded by proving that

$$E\left[R_{n_0,n_1}^{(1)}(t)^2\right] = o\left(\frac{1}{n_0h}\right) \text{ and } E\left[R_{n_0,n_1}^{(2)}(t)^2\right] = o\left(\frac{1}{n_0h}\right),$$

using similar arguments as for the first moment. □

Corollary 3.1.5. *Assume conditions (S2), (B5), (E1), (I1), (K11), (Q2), (Q3), (R3), (R4) and let* $t < \min\left\{\tilde{b}_0, \tilde{b}_1\right\}$*, then*

$$E\left[\check{r}_h(t)\right] = r(t) + \frac{1}{2}r^{(2)}(t)\,d_K h^2 + o\left(h^2\right) + o\left((n_1h)^{-\frac{1}{2}}\right),$$

$$Var\left[\check{r}_h(t)\right] = \frac{R(K)\sigma^2(t)}{n_1h} + o\left((n_1h)^{-1}\right),$$

where

$$\begin{aligned} \sigma^2(t) &= Q_1^{(1)}(t)\,(1-R(t))^2 + \kappa^2 Q_0^{(1)}(t)\,r^2(t)\,(1-t)^2 \\ &= \frac{r(t)\,(1-R(t))}{B_1(F_0^{-1}(t))} + \kappa^2\frac{(1-t)r^2(t)}{B_0(F_0^{-1}(t))}. \end{aligned}$$

Note that this result generalizes those presented in Cao *et al* (2000) and Handcock and Janssen (2002) for, respectively, right censored data and complete data.

Proof of Corollary 3.1.5. To study the bias of $\check{r}_h(t)$, we use the order for the first moment of $C_{n_0,n_1}(t)$:

$$\begin{aligned} E[\check{r}_h(t)] &= \int K(u)\, r(t-hu)\,\mathrm{d}u + o\left((n_1h)^{-\frac{1}{2}}\right) \\ &= r(t) + \frac{1}{2} r^{(2)}(t)\, d_K h^2 + o\left(h^2\right) + o\left((n_1h)^{-\frac{1}{2}}\right). \end{aligned}$$

To handle the variance we use the bound for the second moment of $C_{n_0,n_1}(t)$ and the Cauchy-Schwarz inequality:

$$\begin{aligned} Var[\check{r}_h(t)] &= Var\left(A_{n_1}^{(2)}(t)\right) + Var\left(B_{n_0,n_1}^{(11)}(t)\right) \\ &\quad + o\left((n_1h)^{-1}\right) + o\left((n_1h)^{-\frac{1}{2}}\left(Var\left(A_{n_1}^{(2)}(t)\right)\right)^{\frac{1}{2}}\right) \\ &\quad + o\left((n_1h)^{-\frac{1}{2}}\left(Var\left(B_{n_0,n_1}^{(11)}\right)\right)^{\frac{1}{2}}\right). \end{aligned}$$

Then, we first obtain expressions for the variances of $A_{n_1}^{(2)}(t)$ and $B_{n_0,n_1}^{(11)}(t)$:

$$\begin{aligned} & Var\left(A_{n_1}^{(2)}(t)\right) = E\left[A_{n_1}^{(2)}(t)^2\right] \\ &= \frac{1}{n_1h^2}\int_{-1}^{1}\int_{-1}^{1} E\left[\hat{\xi}_{1i}\left(Y_{1i}, T_{1i}, \delta_{1i}, F_0^{-1}(t-hu)\right)\hat{\xi}_{1i}\left(Y_{1i}, T_{1i}, \delta_{1i}, F_0^{-1}(t-hv)\right)\right] \\ & K^{(1)}(u)\, K^{(1)}(v)\,\mathrm{d}u\mathrm{d}v \\ &= \frac{2}{n_1h^2}\int_{-1}^{1}\int_{-1}^{v} (1-R(t-hu))(1-R(t-hv))\, Q_1(t-hv)\, K^{(1)}(u)\, K^{(1)}(v)\,\mathrm{d}u\mathrm{d}v. \end{aligned}$$

Using Taylor expansions it is easy to derive:

$$\begin{aligned} Var\left(A_{n_1}^{(2)}(t)\right) &= \frac{2}{n_1h^2}\int_{-1}^{1}\int_{-1}^{v}\left(1-\left\{R(t) - hur(t) + h^2u^2r^{(1)}(\eta_1)\right\}\right) \\ &\quad \left(1-\left\{R(t) - hvr(t) + h^2v^2r^{(1)}(\eta_2)\right\}\right) \\ &\quad \left\{Q_1(t) - hvQ_1^{(1)}(t) + \frac{1}{2}h^2v^2Q_1^{(2)}(\eta)\right\} K^{(1)}(u)\, K^{(1)}(v)\,\mathrm{d}u\mathrm{d}v \\ &= A(t) + B(t) + C(t) + D(t) + O\left(\frac{1}{n_1}\right), \end{aligned}$$

where η_1 is a value between $t-hu$ and t, η and η_2 are some values between $t-hv$ and t and

$$\begin{aligned} A(t) &= \frac{2}{n_1h^2}(1-R(t))(1-R(t))\, Q_1(t)\int_{-1}^{1}\int_{-1}^{v} K^{(1)}(u)\, K^{(1)}(v)\,\mathrm{d}u\mathrm{d}v, \\ B(t) &= -\frac{2}{n_1h}(1-R(t))(1-R(t))\, Q_1^{(1)}(t)\int_{-1}^{1}\int_{-1}^{v} vK^{(1)}(u)\, K^{(1)}(v)\,\mathrm{d}u\mathrm{d}v, \\ C(t) &= \frac{2}{n_1h}(1-R(t))\, r(t)\, Q_1(t)\int_{-1}^{1}\int_{-1}^{v} K^{(1)}(u)\, vK^{(1)}(v)\,\mathrm{d}u\mathrm{d}v, \\ D(t) &= \frac{2}{n_1h}(1-R(t))\, r(t)\, Q_1(t)\int_{-1}^{1}\int_{-1}^{v} uK^{(1)}(u)\, K^{(1)}(v)\,\mathrm{d}u\mathrm{d}v. \end{aligned}$$

Straightforward calculations show that $A(t)=0$, $B(t)=\frac{R(K)(1-R(t))^2Q_1^{(1)}(t)}{n_1h}$ and $C(t)+D(t)=0$. These facts lead to:

$$Var\left(A_{n_1}^{(2)}(t)\right)=\frac{R(K)Q_1^{(1)}(t)}{n_1h}\left\{1-R(t)\right\}^2+O\left(\frac{1}{n_1}\right).$$

It is easy to check that $q_1\left(F_0^{-1}(t)\right)=\int_{a_{W_1}}^{F_0^{-1}(t)}\frac{\mathrm{d}F_1(u)}{(1-F_1(u))B_1(u)}$. Consequently, it follows that $Q_1^{(1)}(t)=\frac{r(t)}{(1-R(t))B_1\left(F_0^{-1}(t)\right)}$.

The expression shown above, for $Var\left(A_{n_1}^{(2)}(t)\right)$, was obtained after simple algebra by taking into account that:

$$\int_{-1}^{1}\int_{-1}^{v}K^{(1)}(u)K^{(1)}(v)\,\mathrm{d}u\mathrm{d}v = \int_{-1}^{1}K(v)K^{(1)}(v)\,\mathrm{d}v=0,$$

since K is symmetric. On the other hand,

$$\begin{aligned}\int_{-1}^{1}\int_{-1}^{v}vK^{(1)}(v)K^{(1)}(u)\,\mathrm{d}u\mathrm{d}v &= \int_{-1}^{1}vK^{(1)}(v)K(v)\,\mathrm{d}v\\ &= -\int_{-1}^{1}vK(v)K^{(1)}(v)\,\mathrm{d}v-\int_{-1}^{1}K^2(v)\,\mathrm{d}v\\ &= -\int_{-1}^{1}vK(v)K^{(1)}(v)\,\mathrm{d}v-R(K)\\ &\Rightarrow\ 2\int_{-1}^{1}vK^{(1)}(v)K(v)\mathrm{d}v=-R(K),\end{aligned}$$

which finally gives

$$\int_{-1}^{1}\int_{-1}^{v}K^{(1)}(v)K(u)\mathrm{d}u\mathrm{d}v=-\frac{R(K)}{2},$$

which proves the expression obtained for the term $B(t)$. To handle $C(t)+D(t)$,we have:

$$\begin{aligned}&\int_{-1}^{1}\int_{-1}^{v}K^{(1)}(u)vK^{(1)}(v)\mathrm{d}u\mathrm{d}v+\int_{-1}^{1}\int_{-1}^{v}uK^{(1)}(u)K^{(1)}(v)\mathrm{d}u\mathrm{d}v\\ &=\int_{-1}^{1}\int_{-1}^{v}K^{(1)}(u)vK^{(1)}(v)\mathrm{d}u\mathrm{d}v+\int_{-1}^{1}\int_{v}^{1}vK^{(1)}(v)K^{(1)}(u)\mathrm{d}u\mathrm{d}v\\ &=\int_{-1}^{1}\int_{-1}^{1}K^{(1)}(u)vK^{(1)}(v)\mathrm{d}u\mathrm{d}v=0,\end{aligned}$$

since K is symmetric. Besides, it holds that

$$\int_{-1}^{1}v^2K^{(1)}(v)\int_{-1}^{v}K^{(1)}(u)\,\mathrm{d}u\mathrm{d}v = \int_{-1}^{1}v^2K^{(1)}(v)K(v)\,\mathrm{d}v\le\left\|K^{(1)}\right\|_\infty$$

and

$$\left|\int_{-1}^{1}K^{(1)}(v)\int_{-1}^{v}u^2K^{(1)}(u)\,\mathrm{d}u\mathrm{d}v\right| \le \int_{-1}^{1}\left|K^{(1)}(v)\right|\int_{-1}^{v}\left|K^{(1)}(u)\right|\mathrm{d}u\mathrm{d}v\le 2\left\|K^{(1)}\right\|_\infty^2.$$

Based on the previous integrals it is easy to see that $A(t)=0$, $C(t)+D(t)=0$ and

$$\begin{aligned} B(t) &= -2h\left(1-R\left(t\right)\right)^2 Q_1^{(1)}\left(t\right)\left(-\frac{R(K)}{2}\right) \\ &= R(K)\left(1-R\left(t\right)\right)^2 Q_1^{(1)}\left(t\right)h. \end{aligned}$$

Below we study $Var\left(B_{n_0,n_1}^{(11)}\left(t\right)\right)$.

$$\begin{aligned} Var\left(B_{n_0,n_1}^{(11)}\left(t\right)\right) &= Var\left(-\frac{1}{n_0 h}\sum_{i=1}^{n_0}\int \tilde{\xi}_{0i}\left(Y_{0i},T_{0i},\delta_{0i},t-hu\right)K^{(1)}\left(u\right)r\left(t-hu\right)\mathrm{d}u\right) \\ &= \frac{1}{n_0 h^2}\int_{-1}^{1}\int_{-1}^{1}K^{(1)}\left(u\right)K^{(1)}\left(v\right)r\left(t-hu\right)r\left(t-hv\right) \\ &\quad E\left[\tilde{\xi}_{0i}\left(Y_{0i},T_{0i},\delta_{0i},t-hu\right)\tilde{\xi}_{0i}\left(Y_{0i},T_{0i},\delta_{0i},t-hv\right)\right]\mathrm{d}u\mathrm{d}v \\ &= \frac{2}{n_0 h^2}\int_{-1}^{1}\int_{-1}^{v}K^{(1)}\left(u\right)K^{(1)}\left(v\right)r\left(t-hu\right)r\left(t-hv\right) \\ &\quad E\left[\hat{\xi}_{0i}\left(Y_{0i},T_{0i},\delta_{0i},F_0^{-1}(t-hu)\right)\hat{\xi}_{0i}\left(Y_{0i},T_{0i},\delta_{0i},F_0^{-1}(t-hv)\right)\right]\mathrm{d}u\mathrm{d}v. \end{aligned}$$

Consider $u<v$, then

$$\begin{aligned} &E\left[\hat{\xi}_{0i}\left(Y_{0i},T_{0i},\delta_{0i},F_0^{-1}\left(t-hu\right)\right)\hat{\xi}_{0i}\left(Y_{0i},T_{0i},\delta_{0i},F_0^{-1}\left(t-hv\right)\right)\right] \\ &=\left(1-\left(t-hu\right)\right)\left(1-\left(t-hv\right)\right)Q_0\left(t-hv\right), \end{aligned}$$

where $Q_0\left(t-hv\right)=\int_{a_{W0}}^{F_0^{-1}(t-hv)}\frac{\mathrm{d}W_{01}(y)}{B_0^2(y)}$.

Consequently,

$$\begin{aligned} Var\left(B_{n_0,n_1}^{(11)}\left(t\right)\right) &= \frac{2}{n_0 h^2}\int_{-1}^{1}\int_{-1}^{v}\left(1-\left(t-hu\right)\right)\left(1-\left(t-hv\right)\right) \\ &\quad Q_0\left(t-hv\right)K^{(1)}\left(u\right)K^{(1)}\left(v\right)r\left(t-hu\right)r\left(t-hv\right)\mathrm{d}u\mathrm{d}v \\ &= \frac{2}{n_0 h^2}\left(1-t\right)^2 r^2\left(t\right)Q_0\left(t\right)\int_{-1}^{1}\int_{-1}^{v}K^{(1)}\left(u\right)K^{(1)}\left(v\right)\mathrm{d}u\mathrm{d}v \\ &\quad -\frac{2}{n_0 h}\left(1-t\right)^2 r^2\left(t\right)Q_0^{(1)}\left(t\right)\int_{-1}^{1}\int_{-1}^{v}K^{(1)}\left(u\right)vK^{(1)}\left(v\right)\mathrm{d}u\mathrm{d}v \\ &\quad +\frac{2}{n_0 h}\left(1-t\right)r^2\left(t\right)Q_0\left(t\right)\int_{-1}^{1}\int_{-1}^{v}K^{(1)}\left(u\right)vK^{(1)}\left(v\right)\mathrm{d}u\mathrm{d}v \\ &\quad +\frac{2}{n_0 h}\left(1-t\right)r^2\left(t\right)Q_0\left(t\right)\int_{-1}^{1}\int_{-1}^{v}uK^{(1)}\left(u\right)K^{(1)}\left(v\right)\mathrm{d}u\mathrm{d}v \\ &\quad -\frac{2}{n_0 h}(1-t)^2 r(t)r^{(1)}(t)Q_0(t)\int_{-1}^{1}\int_{-1}^{v}K^{(1)}(u)(u+v)K^{(1)}(v)\mathrm{d}u\mathrm{d}v \\ &\quad +O\left(\frac{1}{n_0}\right)=\frac{R(K)}{n_0 h}Q_0^{(1)}\left(t\right)r^2\left(t\right)\left(1-t\right)^2+O\left(\frac{1}{n_0}\right), \end{aligned}$$

where $Q_0^{(1)}\left(t\right)=\frac{1}{(1-t)B_0\left(F_0^{-1}(t)\right)}$.

Collecting all the previous results, we get $Var\left[\check{r}_h\left(t\right)\right]=\frac{R(K)\sigma^2(t)}{n_1 h}+o\left(\left(n_1 h\right)^{-1}\right)$. □

Remark 3.1.1. As a straightforward consequence of Corollary 3.1.5, an asymptotic formula for the mean squared error of the estimator is obtained

$$MSE\left(\check{r}_h\left(t\right)\right) = AMSE\left(\check{r}_h\left(t\right)\right) + o\left(h^4\right) + o\left(\left(n_1 h\right)^{-1}\right),$$

where

$$AMSE\left(\check{r}_h\left(t\right)\right) = \frac{1}{4} r^{(2)}\left(t\right)^2 d_K^2 h^4 + \frac{R(K)\sigma^2(t)}{n_1 h}.$$

Consequently, the smoothing parameter that minimizes this criterion is

$$h_{AMSE} = \left(\frac{R(K)\sigma^2(t)}{r^{(2)}\left(t\right)^2 d_K^2}\right)^{\frac{1}{5}} n_1^{-\frac{1}{5}}.$$

Corollary 3.1.6. *Assume conditions (S2), (B5), (E1), (I1), (K11), (Q2), (Q3), (R3), (R4) and let* $t < \min\left\{\tilde{b}_0, \tilde{b}_1\right\}$.
If $n_1 h^5 \to 0$, *then*

$$(n_1 h)^{\frac{1}{2}} \left\{\check{r}_h(t) - r(t)\right\} \xrightarrow{d} N\left(0, R(K)\sigma^2(t)\right).$$

If $n_1 h^5 \to c$, *for some* $c > 0$, *then*

$$(n_1 h)^{\frac{1}{2}} \left\{\check{r}_h(t) - r(t)\right\} \xrightarrow{d} N\left(\frac{1}{2} r^{(2)}(t) d_K c^{\frac{1}{2}}, R(K)\sigma^2(t)\right).$$

Proof of Corollary 3.1.6. It is a straightforward consequence of Theorem 3.1.4, Corollary 3.1.5 and the Lyapunov theorem for triangular arrays (see Theorem 4.9 in Petrov (1995)). Below, the main steps of the proof are given.

Let define $Z_{0i} = \int \tilde{\xi}_{0i}(Y_{0i}, T_{0i}, \delta_{0i}, t - hu) K^{(1)}(u) r(t - hu) \mathrm{d}u$ for $i = 1, \ldots, n_0$ and $Z_{1j} = \int \tilde{\xi}_{1j}(Y_{1j}, T_{1j}, \delta_{1j}, t - hu) K^{(1)}(u) \mathrm{d}u$ for $j = 1, \ldots, n_1$. From Corollary 3.1.5, it follows that

$$\begin{aligned} E\left[Z_{0i}\right] &= E\left[Z_{1j}\right] = 0, \\ Var[Z_{0i}] &= hR(K) Q_0^{(1)}(t) r^2(t) \left\{1 - t\right\}^2 + O\left(h^2\right) \text{ and} \\ Var[Z_{1j}] &= hR(K) Q_1^{(1)}(t) \left\{1 - R(t)\right\}^2 + O\left(h^2\right). \end{aligned}$$

Using Lyapunov condition for the sequences $\{Z_{0i}\}$ and $\{Z_{1j}\}$, we only need to prove that there exist some positive real numbers, c_1 and c_2, such as

$$\left(\sum_{i=1}^{n_0} Var(Z_{0i})\right)^{-1-\frac{c_1}{2}} \sum_{i=1}^{n_0} E\left|Z_{0i}\right|^{2+c_1} \to 0 \text{ and} \tag{3.8}$$

$$\left(\sum_{j=1}^{n_1} Var(Z_{1j})\right)^{-1-\frac{c_2}{2}} \sum_{j=1}^{n_1} E\left|Z_{1j}\right|^{2+c_2} \to 0. \tag{3.9}$$

Conditions (3.8) and (3.9) imply that

$$\begin{aligned}\frac{1}{(n_0h)^{\frac{1}{2}}}\sum_{i=1}^{n_0} Z_{0i} &\xrightarrow{d} N\left(0, R(K)Q_0^{(1)}(t)r^2(t)\left\{1-t\right\}^2\right) \text{ and}\\ \frac{1}{(n_1h)^{\frac{1}{2}}}\sum_{i=1}^{n_1} Z_{1j} &\xrightarrow{d} N\left(0, R(K)Q_1^{(1)}(t)\left\{1-R(t)\right\}^2\right).\end{aligned}$$

The proof of the corollary is finished using Theorem 3.1.4, Corollary 3.1.5 and conditions (B5) and (S2).

Let us prove Lyapunov condition. Using condition (K11) and the inequality $(a+b)^{2+c_2} \leq 2^{1+c_2}(a^{2+c_2}+b^{2+c_2})$, for any pair of positive real numbers a and b, it is easy to show that

$$\begin{aligned}E\left[|Z_{1j}|^{2+c_2}\right] &\leq \left\|K^{(1)}\right\|_\infty^{2+c_2} E\left[\left\{\frac{\delta_{1j}}{\tilde{B}_1(F_0(Y_{1j}))}+\int_{a_{\tilde{W}_1}}^{\tilde{b}_1}\frac{\mathrm{d}\tilde{W}_{11}(s)}{\tilde{B}_1^2}\right\}^{2+c_2}\right]\\ &\leq \left\|K^{(1)}\right\|_\infty^{2+c_2} 2^{1+c_2}\left\{\left(\int_{a_{\tilde{W}_1}}^{\tilde{b}_1}\frac{\mathrm{d}\tilde{W}_{11}(s)}{\tilde{B}_1^2}\right)^{2+c_2}+E\left[\left(\frac{\delta_{1j}}{\tilde{B}_1(F_0(Y_{1j}))}\right)^{2+c_2}\right]\right\}.\end{aligned}$$

Since

$$E\left[\left(\frac{\delta_{1j}}{\tilde{B}_1(F_0(Y_{1j}))}\right)^{2+c_2}\right] \leq \int_{a_{\tilde{W}_1}}^{\tilde{b}_1}\frac{\mathrm{d}\tilde{W}_{11}(s)}{\tilde{B}_1^{2+c_2}},$$

the proof of $E\left[|Z_{1j}|^{2+c_2}\right] < \infty$ follows from condition (I1).

Now, considering $c_2 = 1$, it is easy to check that

$$\left(\sum_{j=1}^{n_1} Var(Z_{1j})\right)^{-1-\frac{c_2}{2}}\sum_{j=1}^{n_1} E\left|Z_{1j}\right|^{2+c_2} = O\left(\frac{1}{(n_1h^3)^{1/2}}\right).$$

Consequently (3.9) holds for $c_2 = 1$, using condition (B5). Proceeding in a similar way it is easy to check that (3.8) is satisfied for $c_1 = 1$, which concludes the proof. □

Remark 3.1.2. Defining $MISE_w(\check{r}_h) = \int(\check{r}_h(t)-r(t))^2w(t)\mathrm{d}t$, it can be proved from Corollary 3.1.5 that

$$MISE_w(\check{r}_h) = AMISE_w(h) + o\left(\frac{1}{n_1h}+h^4\right)+o\left(\frac{1}{n_0h}\right), \tag{3.10}$$

where

$$\begin{aligned}AMISE_w(h) &= \frac{R(K)}{n_1h}\int\frac{r(t)(1-R(t))}{B_1(F_0^{-1}(t))}w(t)\mathrm{d}t+\frac{R(K)}{n_0h}\int\frac{(1-t)r^2(t)}{B_0(F_0^{-1}(t))}w(t)\mathrm{d}t\\ &+\frac{1}{4}h^4d_K^2\int(r^{(2)}(t))^2w(t)\mathrm{d}t\end{aligned}$$

and $w(t)$ denotes a weight function that is advisable to introduce in the MISE, specially when boundary effects may appear.

Therefore, the dominant part of the bandwidth that minimizes the mean integrated squared error in (3.10) is given by

$$h_{AMISE_{w},r} = \left(\frac{R(K) \int \sigma^2(t) w(t) \mathrm{d}t}{d_K^2 \int \left(r^{(2)}(t)\right)^2 w(t) \mathrm{d}t} \right)^{1/5} n_1^{-1/5}. \tag{3.11}$$

3.1.2 Plug-in selectors

Three plug-in global bandwidth selectors for $\check{r}_h(t)$ were designed in this setting of left truncated and right censored data. Generalizing the results given by Sánchez Sellero *et al* (1999) to the case of a two-sample problem with LTRC data, it can be proved that, if g goes to zero as the sample sizes increase, then

$$E\left[\left(\int \left(\check{r}_g^{(2)}(x)\right)^2 w(x)\mathrm{d}x - \int \left(r^{(2)}(x)\right)^2 w(x)\mathrm{d}x\right)^2\right] \tag{3.12}$$

$$= \frac{1}{n_1^2 g^{10}} R^2\left(K^{(2)}\right) \left(\int \sigma^2(t)w(t)\mathrm{d}t\right)^2 + g^4 d_K \left(\int r^{(2)}(t) r^{(4)}(t) w(t)\mathrm{d}t\right)^2$$
$$+ \frac{2d_K}{n_1 g^3} R^2\left(K^{(2)}\right) \int \sigma^2(t)w(t)\mathrm{d}t \int r^{(2)}(t) r^{(4)}(t) w(t)\mathrm{d}t$$
$$+ O\left(n_1^{-1} g^{-1}\right) + O\left(n_1^{-2} g^{-9}\right) + O\left(g^6\right) + O\left(n_1^{-3/2} g^{-9/2}\right).$$

Furthermore, if $n_1 g^3$ goes to infinity, the dominant part of the bandwidth that minimizes the mean squared error in (3.12), is given by

$$g_{AMSE,R(r^{(2)}w^{1/2})} = \left[\frac{R\left(K^{(2)}\right) \int \sigma^2(t) w(t)\mathrm{d}t}{-d_K \left(\int r^{(2)}(t) r^{(4)}(t) w(t)\mathrm{d}t\right)} \right]^{1/7} n_1^{-1/7}$$

if $\int r^{(2)}(t) r^{(4)}(t) w(t)\mathrm{d}t$ is negative, and by

$$g_{AMSE,R(r^{(2)}w^{1/2})} = \left[\frac{5R\left(K^{(2)}\right) \int \sigma^2(t) w(t)\mathrm{d}t}{2d_K \left(\int r^{(2)}(t) r^{(4)}(t) w(t)\mathrm{d}t\right)} \right]^{1/7} n_1^{-1/7}$$

if $\int r^{(2)}(t) r^{(4)}(t) w(t)\mathrm{d}t$ is positive.

Note that $h_{AMISE_{w},r}$ in (3.11) and $g_{AMSE,R(r^{(2)}w^{1/2})}$ satisfy the relationship below

$$g_{AMSE,R(r^{(2)}w^{1/2})} = \left(\frac{R(K^{(2)}) d_K \int \left(r^{(2)}(t)\right)^2 w(t)\mathrm{d}t}{-R(K) \left(\int r^{(2)}(t) r^{(4)}(t) w(t)\mathrm{d}t\right)} \right)^{1/7} h_{AMISE_{w},r}^{5/7}$$

if $\int r^{(2)}(t) r^{(4)}(t) w(t)\mathrm{d}t$ is negative, and

$$g_{AMSE,R(r^{(2)}w^{1/2})} = \left(\frac{5R(K^{(2)}) d_K \int \left(r^{(2)}(t)\right)^2 w(t)\mathrm{d}t}{2R(K) \left(\int r^{(2)}(t) r^{(4)}(t) w(t)\mathrm{d}t\right)} \right)^{1/7} h_{AMISE_{w},r}^{5/7}$$

if $\int r^{(2)}(t)r^{(4)}(t)w(t)\mathrm{d}t$ is positive.

The simplest way to propose a plug-in bandwidth selector for $\check{r}_h(t)$ would be to consider a parametric model for r and to replace the unknown quantities appearing in (3.11) by these parametric estimates. As it was previously mentioned in Chapter 2, these simple plug-in bandwidth selectors are known as rules of thumb. Therefore, from here on we will denote our first proposal, which is based on this idea, by h_{RT} and we will consider, as in Chapter 2, a mixture of betas (see (2.33)). The only difference now is that $\tilde{R}_{g_{\tilde{R}}}(\cdot)$ in $\tilde{b}(x;N,R)$ is replaced by

$$\tilde{R}_{n_1}(x) = 1 - \prod_{\hat{F}_{0n_0}(Y_{1j})\leq x} \left[1 - \left(n_1\tilde{B}_{1n_1}(\hat{F}_{0n_0}(Y_{1j}))\right)^{-1}\right]^{\delta_{1j}},$$

where $\hat{F}_{0n_0}(t)$ denotes the PL estimator associated to the sample $\{(T_{0i}, Y_{0i}, \delta_{0i})\}_{i=1}^{n_0}$ (see (1.5)) and $\tilde{B}_{1n_1}(t)$ is the empirical estimate of $\tilde{B}_1(t)$ introduced previously at the beginning of Chapter 3. To make clear the difference, we denote from here on, this new mixture of betas by $\hat{b}(x;N,R)$, which explicit expression is given by

$$\hat{b}(x;N,R) = \sum_{j=1}^{N}\left(\tilde{R}_{n_1}\left(\frac{j}{N}\right) - \tilde{R}_{n_1}\left(\frac{j-1}{N}\right)\right)\beta(x,j,N-j+1),$$

where, as in Chapter 2, N denotes the number of betas in the mixture. In practice we will set $N = 14$.

Our second bandwidth selector, h_{PI}, differs from h_{RT} in how the estimation of the functional $R\left(r^{(2)}w^{1/2}\right)$ is carried out. Rather than using a parametric estimate of this unknown quantity, based on a parametric model assumed for $r(t)$, we now consider a kernel type estimate of $r^{(2)}(t)$ with optimal bandwidth g given by the expression

$$g = \left[\frac{R\left(K^{(2)}\right)\int\hat{\sigma}^2(t)w(t)\mathrm{d}t}{-d_K\left(\int\hat{b}^{(2)}(t;N,R)\hat{b}^{(4)}(t;N,R)w(t)\mathrm{d}t\right)}\right]^{1/7} n_1^{-1/7}$$

if $\int\hat{b}^{(2)}(t;N,R)\hat{b}^{(4)}(t;N,R)w(t)\mathrm{d}t$ is negative, and by

$$g = \left[\frac{5R\left(K^{(2)}\right)\int\hat{\sigma}^2(t)w(t)\mathrm{d}t}{2d_K\left(\int\hat{b}^{(2)}(t;N,R)\hat{b}^{(4)}(t;N,R)w(t)\mathrm{d}t\right)}\right]^{1/7} n_1^{-1/7}$$

if $\int\hat{b}^{(2)}(t;N,R)\hat{b}^{(4)}(t;N,R)w(t)\mathrm{d}t$ is positive. Note that the expression of g is based on the asymptotically optimal bandwidth, $g_{AMSE,R(r^{(2)}w^{1/2})}$, to estimate the functional $R\left(r^{(2)}w^{1/2}\right)$. The difference is that now the unknown quantity in the denominator of

$g_{AMSE,R(r^{(2)}w^{1/2})}$ is replaced by a parametric estimate using the mixture of betas presented above, $\hat{b}(x; N, R)$, and the unknown function in the numerator, $\sigma^2(t)$, is estimated by

$$\hat{\sigma}^2(t) = \frac{\hat{b}(t; N, R)(1 - \tilde{R}_{n_1}(t))}{\tilde{B}_{1n_1}(t)} + \frac{n_1}{n_0}\frac{(1-t)\hat{b}^2(t; N, R)}{\tilde{B}_{0n_0}(t)}.$$

Finally, our last proposal is a more sophisticated plug-in bandwidth selector that requires to solve the following equation in h

$$h = \left(\frac{R(K) \int \hat{\sigma}^2(t) w(t) \mathrm{d}t}{d_K R\left(\check{r}^{(2)}_{\gamma(h)} w^{1/2}\right)} \right)^{1/5} n_1^{-1/5},$$

with

$$\gamma(h) = \left(\frac{R(K^{(2)}) d_K R\left(\hat{b}^{(2)} w^{1/2}\right)}{-R(K) \int \hat{b}^{(2)}(t; N, R) \hat{b}^{(4)}(t; N, R) w(t) \mathrm{d}t} \right)^{1/7} h^{5/7}$$

if $\int \hat{b}^{(2)}(t; N, R)\hat{b}^{(4)}(t; N, R)w(t)\mathrm{d}t$ is negative, and with

$$\gamma(h) = \left(\frac{5R(K^{(2)}) d_K R\left(\hat{b}^{(2)} w^{1/2}\right)}{2R(K) \int \hat{b}^{(2)}(t; N, R) \hat{b}^{(4)}(t; N, R) w(t) \mathrm{d}t} \right)^{1/7} h^{5/7}$$

if $\int \hat{b}^{(2)}(t; N, R)\hat{b}^{(4)}(t; N, R)w(t)\mathrm{d}t$ is positive. We denote this third proposal by h_{STE}.

3.1.3 A simulation study

A simulation study is carried out to check the practical behaviour of the plug-in bandwidth selectors proposed in the previous subsection.

Since we need to specify the model assumed for the relative density and the probability of censoring and truncation for both populations, we propose the following sampling scheme.

It is assumed that the random variable X_0 follows a Weibull distribution with parameters a and b, i.e.

$$F_0(x) = \left(1 - \exp\left\{-(ax)^b\right\}\right) 1_{\{x>0\}}.$$

The cdf of the censoring variable in the comparison population is chosen in such a way that satisfies the relationship

$$1 - L_0(x) = (1 - F_0(x))^{\mu_0},$$

where $\mu_0 > 0$ and L_0 denotes the cdf of the censoring variable C_0. Simple algebra yields that

$$L_0(x) = \left(1 - \exp\left\{-(a\mu_0^{1/b} x)^b\right\}\right) 1_{\{x>0\}},$$

or equivalently, that C_0 follows a Weibull distribution with parameters $a\mu_0^{1/b}$ and b. Consequently, it follows that the probability of censoring is $\beta_0 = P\left(C_0 < X_0\right) = \frac{\mu_0}{\mu_0+1}$. The truncation variable T_0 is selected in such a way that its cdf, G_0, satisfies

$$\begin{aligned} 1 - G_0(x) &= (1 - W_0(x))^{\nu_0} \\ &= (1 - F_0(x))^{\nu_0}(1 - L_0(x))^{\nu_0}, \end{aligned}$$

where $\nu_0 > 0$, and therefore, T_0 is a Weibull distributed random variable with parameters $a(\nu_0(1+\mu_0))^{1/b}$ and b and it is known that the probability of truncation is $1-\alpha_0 = P(Y_0 < T_0) = \frac{1}{\nu_0+1}$.

Since $R(t) = F_1(F_0^{-1}(t))$ it follows that $F_1(x) = R(F_0(x))$ and therefore we will consider that $X_1 = F_0^{-1}(Z)$ where Z denotes the random variable associated to the relative density, $r(t)$. The censoring variable in the second sample will be selected in such a way that its cdf, L_1, satisfies the relationship $1 - L_1(x) = (1 - F_1(x))^{\mu_1}$, where $\mu_1 > 0$, or equivalently, $1 - L_1(x) = (1 - R(F_0(x)))^{\mu_1}$. In this case, it is easy to prove that the probability of censoring in the second population is $\beta_1 = P(C_1 < X_1) = \frac{\mu_1}{\mu_1+1}$. Finally, the truncation variable in the second population, T_1, is selected in such a way that its cdf, G_1, satisfies the relationship

$$\begin{aligned} 1 - G_1(x) &= (1 - W_1(x))^{\nu_1} \\ &= (1 - R(F_0(x)))^{\nu_1}(1 - L_1(x))^{\nu_1}, \end{aligned}$$

where $\nu_1 > 0$, from where it can be easily proved that the probability of truncation in the second population is given by $1 - \alpha_1 = P(Y_1 < T_1) = \frac{1}{\nu_1+1}$.

The models assumed for r will be the seven models introduced previously in Chapter 2. For each one of these models, and each fixed pair of sample sizes, a large number of trials will be considered. For each trial, two samples subject to left truncation and right censoring are drawn. More specifically, a sample of n_0 iid random variables, $\{X_{01}, \ldots, X_{0n_0}\}$, and a sample of n_1 iid random variables, $\{X_{11}, \ldots, X_{1n_1}\}$, subject to left truncation and right censoring, are independently drawn from, respectively $X_0 \stackrel{d}{=} W(a, b)$ and the random variable $X_1 = F_0^{-1}(Z)$. Let $\{T_{01}, \ldots, T_{0n_0}\}$ be a sample of n_0 iid truncation variables with distribution function G_0, and let $\{C_{01}, \ldots, C_{0n_0}\}$ be a sample of iid censoring variables with distribution function L_0. Denoting by $Y_{0i} = \min\{X_{0i}, C_{0i}\}$, only those values that satisfy the condition $T_{0i} \leq Y_{0i}$ are considered. We use the index $i = 1, \ldots, n_0$ for the first n_0 values satisfying this condition. In a similar way, let $\{T_{11}, \ldots, T_{1n_1}\}$ be a sample of n_1 iid truncation variables with distribution function G_1, and let $\{C_{11}, \ldots, C_{1n_1}\}$ be a sample of iid censoring variables with distribution function L_1. Denoting by $Y_{1j} = \min\{X_{1j}, C_{1j}\}$,

Table 3.1: Values of EM_w for h_{RT}, h_{PI} and h_{STE} for models (a) and (b).

EM_w			Model (a)			Model (b)		
(n_0, n_1)	(β_0, β_1)	$(1-\alpha_0, 1-\alpha_1)$	h_{RT}	h_{PI}	h_{STE}	h_{RT}	h_{PI}	h_{STE}
$(50, 50)$	$(0.10, 0.10)$	$(0.10, 0.10)$	0.3355	0.3829	0.3831	0.5873	0.6182	0.6186
$(100, 100)$	$(0.10, 0.10)$	$(0.10, 0.10)$	0.2138	0.2363	0.2365	0.4278	0.4527	0.4522
$(200, 200)$	$(0.10, 0.10)$	$(0.10, 0.10)$	0.1115	0.1207	0.1207	0.2627	0.2611	0.2611
$(50, 50)$	$(0.20, 0.20)$	$(0.10, 0.10)$	0.3358	0.3856	0.3856	0.6141	0.6471	0.6494
$(100, 100)$	$(0.20, 0.20)$	$(0.10, 0.10)$	0.2068	0.2300	0.2300	0.4321	0.4554	0.4551
$(200, 200)$	$(0.20, 0.20)$	$(0.10, 0.10)$	0.1178	0.1285	0.1285	0.2857	0.2906	0.2905
$(50, 50)$	$(0.10, 0.10)$	$(0.20, 0.20)$	0.3582	0.4011	0.4015	0.6233	0.6567	0.6616
$(100, 100)$	$(0.10, 0.10)$	$(0.20, 0.20)$	0.2338	0.2603	0.2597	0.4847	0.5203	0.5201
$(200, 200)$	$(0.10, 0.10)$	$(0.20, 0.20)$	0.1413	0.1528	0.1528	0.3178	0.3275	0.3275

only those values that satisfy the condition $T_{1j} \leq Y_{1j}$ are considered. We use the index $j = 1, \ldots, n_1$ for the first n_1 values satisfying this condition.

In the simulation study carried out here, we consider $a = 3$, $b = 0.3$, K as the Gaussian kernel and the stepwise functions $\tilde{B}_{0n_0}(t)$ and $\tilde{B}_{1n_1}(t)$ in $\hat{\sigma}^2(t)$ are replaced by the following smoothed versions to avoid possible divisions by zero,

$$\tilde{B}_{0g_0}(t) = \frac{1}{n_0} \sum_{i=1}^{n_0} \mathbb{L}\left(\frac{t - \hat{F}_{0n_0}(T_{0i})}{g_0}\right) \mathbb{L}\left(\frac{\hat{F}_{0n_0}(Y_{0i}) - t}{g_0}\right)$$

and

$$\tilde{B}_{1g_1}(t) = \frac{1}{n_1} \sum_{j=1}^{n_1} \mathbb{L}\left(\frac{t - \hat{F}_{0n_0}(T_{1j})}{g_1}\right) \mathbb{L}\left(\frac{\hat{F}_{0n_0}(Y_{1j}) - t}{g_1}\right),$$

where $\mathbb{L}$ is the cdf of the biweight kernel, i.e.

$$\mathbb{L}(x) = \begin{cases} 0, & \text{if} \quad x < -1, \\ \frac{1}{2} + \frac{15}{16}x - \frac{5}{8}x^3 + \frac{3}{16}x^5, & \text{if} \quad x \in [-1, 1], \\ 1, & \text{if} \quad x > 1, \end{cases}$$

and $g_0 = g_1 = 0.10$.

The weight function used is defined as

$$w(t) = \mathbb{L}\left(\frac{t - q_{0.05}}{0.025}\right) \mathbb{L}\left(\frac{q_{0.95} - t}{0.025}\right),$$

where $q_{0.05}$ and $q_{0.95}$ denote respectively the 0.05 and 0.95 quantile of $r(t)$.

It is interesting to remember here the problem already mentioned in Chapter 1, regarding the PL estimator of a cdf with LTRC data and its bad performance from a failure time t_0 and beyond when the number of individuals at risk in t_0 is the same as the number of failures occurring in t_0. In order to avoid this problem we consider a modified version of

Table 3.2: Values of EM_w for h_{RT}, h_{PI} and h_{STE} for models (c) and (d).

EM_w			Model (c)			Model (d)		
(n_0, n_1)	(β_0, β_1)	$(1-\alpha_0, 1-\alpha_1)$	h_{RT}	h_{PI}	h_{STE}	h_{RT}	h_{PI}	h_{STE}
(50, 50)	(0.10, 0.10)	(0.10, 0.10)	0.2646	0.2950	0.2966	0.2938	0.3141	0.3140
(100, 100)	(0.10, 0.10)	(0.10, 0.10)	0.1625	0.1796	0.1796	0.2015	0.2169	0.2168
(200, 200)	(0.10, 0.10)	(0.10, 0.10)	0.1044	0.1135	0.1135	0.1373	0.1463	0.1463
(50, 50)	(0.20, 0.20)	(0.10, 0.10)	0.2604	0.2922	0.2925	0.3091	0.3363	0.3370
(100, 100)	(0.20, 0.20)	(0.10, 0.10)	0.1499	0.1644	0.1644	0.2130	0.2270	0.2269
(200, 200)	(0.20, 0.20)	(0.10, 0.10)	0.1014	0.1148	0.1148	0.1320	0.1347	0.1347
(50, 50)	(0.10, 0.10)	(0.20, 0.20)	0.2990	0.3330	0.3385	0.3390	0.3692	0.3682
(100, 100)	(0.10, 0.10)	(0.20, 0.20)	0.1953	0.2206	0.2202	0.2469	0.2701	0.2701
(200, 200)	(0.10, 0.10)	(0.20, 0.20)	0.1176	0.1277	0.1277	0.1675	0.1818	0.1818

Table 3.3: Values of EM_w for h_{RT} and h_{PI} for models (e)-(g).

EM_w			Model (e)		Model (f)		Model (g)	
(n_0, n_1)	(β_0, β_1)	$(1-\alpha_0, 1-\alpha_1)$	h_{RT}	h_{PI}	h_{RT}	h_{PI}	h_{RT}	h_{PI}
(50, 50)	(0.10, 0.10)	(0.10, 0.10)	0.9461	1.0485	1.1754	1.2521	0.7665	0.7714
(100, 100)	(0.10, 0.10)	(0.10, 0.10)	0.5815	0.6344	0.8346	0.8646	0.6102	0.5926
(200, 200)	(0.10, 0.10)	(0.10, 0.10)	0.3806	0.3964	0.5967	0.6065	0.4341	0.3860
(50, 50)	(0.20, 0.20)	(0.10, 0.10)	0.8293	0.9135	1.1123	1.1588	0.8166	0.8343
(100, 100)	(0.20, 0.20)	(0.10, 0.10)	0.5857	0.6384	0.8414	0.8650	0.6232	0.6051
(200, 200)	(0.20, 0.20)	(0.10, 0.10)	0.3783	0.3980	0.5366	0.5304	0.4388	0.3941
(50, 50)	(0.10, 0.10)	(0.20, 0.20)	0.9987	1.1050	1.3041	1.3919	0.9007	0.9411
(100, 100)	(0.10, 0.10)	(0.20, 0.20)	0.7075	0.7904	1.0579	1.1416	0.6911	0.6906
(200, 200)	(0.10, 0.10)	(0.20, 0.20)	0.4479	0.4871	0.7130	0.7523	0.4988	0.4800

the TJW product-limit estimator (see (1.5)) such as it discards those terms in the product for which the number of individuals at risk is the same as the number of failures.

For each one of the seven relative populations listed in Subsection 2.3.3, a number of 500 pairs of samples $\{(T_{01}, Y_{01}, \delta_{01}), \ldots, (T_{0n_0}, Y_{0n_0}, \delta_{0n_0})\}$ and $\{(T_{11}, Y_{11}, \delta_{11}), \ldots, (T_{1n_1}, Y_{1n_1}, \delta_{1n_1})\}$ were taken as explained before. For each pair of samples, the three bandwidth selectors proposed previously, h_{RT}, h_{PI} and h_{STE}, were computed and, based on each one, the kernel-type relative density estimate (3.1), was obtained. For each bandwidth selector, let say $\hat{h}$, we obtained 500 estimations of r. Based on them, the following global error criterion was approximated via Monte Carlo:

$$EM_w = E\left[\int (\check{r}_{\hat{h}}(t) - r(t))^2 w(t) \mathrm{d}t\right].$$

From Tables 3.1-3.3, it follows that the simple rule of thumb, h_{RT}, outperforms the plug-in selector, h_{PI}, and the solve-the-equation rule, h_{STE}. Besides, the effect of an increase in the percentages of truncation seems to produce a worse behaviour than the same increase in the percentages of censoring.

A brief version of the contents appearing in Section 3.1 is included in the paper Molanes and Cao (2006b).

3.2 Kernel-type relative distribution estimator

In this section we extend the results presented previously for the relative density to the relative distribution, $R(t)$. However, the problem of estimating $R(t)$ will be presented only briefly.

A natural kernel type estimator for the relative distribution is as follows

$$\check{R}_h(t) = \int \mathbb{K}\left(\frac{t - \hat{F}_{0n_0}(y)}{h}\right) \mathrm{d}\hat{F}_{1n_1}.$$

3.2.1 Asymptotic properties

Consider a fixed value of t. Using a Taylor expansion the following expression for $\check{R}_h(t)$ is obtained:

$$\begin{aligned}\check{R}_h(t) &= \int \mathbb{K}\left(\frac{t - F_0(y)}{h}\right) \mathrm{d}\hat{F}_{1n_1}(y) \\ &+ \int \frac{F_0(y) - \hat{F}_{0n_0}(y)}{h} K\left(\frac{t - F_0(y)}{h}\right) \mathrm{d}\hat{F}_{1n_1}(y) \\ &+ \int \frac{(F_0(y) - \hat{F}_{0n_0}(y))^2}{2h^2} K^{(1)}(\Delta_{ty}) \mathrm{d}\hat{F}_{1n_1}(y)\end{aligned}$$

with Δ_{ty} a value between $\frac{t-\hat{F}_{0n_0}(y)}{h}$ and $\frac{t-F_0(y)}{h}$. Based on this decomposition we can further obtain that

$$\check{R}_h(t) = A_{n_1}^R(t) + B_{n_0,n_1}^R(t) + R_{n_0,n_1}^R(t) \tag{3.13}$$

with

$$A_{n_1}^R(t) = \int \mathbb{K}\left(\frac{t - v}{h}\right) \mathrm{d}\tilde{R}_{n_1}(t),$$

$$B_{n_0,n_1}^R(t) = \frac{1}{h}\int \left(v - \tilde{U}_{0n_0}(v)\right) K\left(\frac{t - v}{h}\right) \mathrm{d}\tilde{R}_{n_1}(t),$$

$$R_{n_0,n_1}^R(t) = \frac{1}{2h^2}\int \left(F_0(y) - \hat{F}_0(y)\right)^2 K^{(1)}(\Delta_{ty}) \mathrm{d}\hat{F}_{1n_1}(y),$$

where $\tilde{U}_{0n_0}$ and $\tilde{R}_{n_1}$ were introduced previously in (3.3) and (3.2).

It is interesting to note here that a very similar decomposition was obtained in the previous section of this chapter when studying the asymptotic properties of $\check{r}_h(t)$. Besides, the first term in (3.13) was previously studied by Chen and Wang (2006) for the case

$a_{G_0} < a_{W_0}$. Therefore, under this assumption, and proceeding in a similar way as in the previous section, it is straightforward to prove the following results regarding the asymptotic behaviour of $\check{R}_h(t)$. For that reason their proofs are omitted here.

Theorem 3.2.1. *Assume conditions (S2), (B5), (E1), (I1), (K12), (Q2), (R3) and let* $t < \min\left\{\tilde{b}_0, \tilde{b}_1\right\}$, *then*

$$\check{R}_h(t) = A_{n_1}^{R,(1)}(t) + A_{n_1}^{R,(2)}(t) + B_{n_0,n_1}^{R,(11)}(t) + C_{n_0,n_1}^{R}(t),$$

where

$$A_{n_1}^{R,(1)}(t) = h \int \mathbb{K}(u) r(t-hu) \mathrm{d}u,$$

$$A_{n_1}^{R,(2)}(t) = \frac{1}{n_1} \sum_{i=1}^{n_1} \int \tilde{\xi}_{1i}\left(Y_{1i}, T_{1i}, \delta_{1i}, t-hu\right) K(u) \mathrm{d}u,$$

$$B_{n_0,n_1}^{R,(11)}(t) = -\frac{1}{n_0} \sum_{i=1}^{n_0} \int \tilde{\xi}_{0i}\left(Y_{0i}, T_{0i}, \delta_{0i}, t-hu\right) K(u) r(t-hu) \mathrm{d}u,$$

$$C_{n_0,n_1}^{R}(t) = o\left(\left(\frac{h}{n_1}\right)^{1/2}\right) \quad \text{a.s.}$$

and $E\left[\left|C_{n_0,n_1}^{R}(t)\right|^d\right] = o\left(\left(\frac{h}{n_1}\right)^{d/2}\right)$, *for* $d = 1, 2$.

Corollary 3.2.2. *Assume conditions (S2), (B5), (E1), (I1), (K12), (Q2), (Q3), (R3), (R5) and let* $t < \min\left\{\tilde{b}_0, \tilde{b}_1\right\}$, *then*

$$\begin{aligned}
E\left[\check{R}_h(t)\right] &= R(t) + \frac{1}{2} r^{(1)}(t) d_K h^2 + o\left(h^2\right) + o\left(\left(\frac{h}{n_1}\right)^{1/2}\right), \\
Var\left(\check{R}_h(t)\right) &= \frac{1}{n_1}\sigma_R^2(t) + \frac{D_K h}{n_1}\varsigma^2(t) + o\left(\frac{h}{n_1}\right),
\end{aligned}$$

where

$$\begin{aligned}
\sigma_R^2(t) &= Q_1(t)(1-R(t))^2 + \kappa^2 Q_0(t)(1-t)^2 r^2(t), \\
\varsigma^2(t) &= Q_1^{(1)}(t)(1-R(t))^2 + \kappa^2 Q_0^{(1)}(t)(1-t)^2 r^2(t) \\
&= \frac{(1-R(t))r(t)}{B_1(F_0^{-1}(t))} + \kappa^2 \frac{(1-t)r^2(t)}{B_0(F_0^{-1}(t))}
\end{aligned}$$

and

$$D_K = 2 \int \int_{u \geq v} K(u) v K(v) \mathrm{d}u \mathrm{d}v.$$

Remark 3.2.1. As a straightforward consequence of Corollary 3.2.2, an asymptotic formula for the mean squared error of the estimator is obtained

$$MSE\left(\check{R}_h\left(t\right)\right) = AMSE\left(\check{R}_h\left(t\right)\right) + o\left(h^4\right) + o\left(\frac{h}{n_1}\right),$$

where

$$AMSE\left(\check{R}_h\left(t\right)\right) = \frac{1}{4}r^{(1)}\left(t\right)^2 d_K^2 h^4 + \frac{1}{n_1}\sigma_R^2(t) + \frac{D_K h}{n_1}\varsigma^2(t).$$

Consequently, the smoothing parameter that minimizes this criterion is

$$h_{AMSE,R(t)} = \left(\frac{-D_K\varsigma^2(t)}{r^{(1)}\left(t\right)^2 d_K^2}\right)^{\frac{1}{3}} n_1^{-\frac{1}{3}}.$$

Unfortunately, under the conditions of Theorem 3.2.1, the optimal bandwidth is excluded. Note that condition (B5) is not satisfied for $h = h_{AMSE,R(t)}$. However, we believe that this condition could be replaced by a less restrictive one that holds as well for the interesting case $h = h_{AMSE,R(t)}$.

Corollary 3.2.3. *Assume conditions (S2), (B5), (E1), (I1), (K12), (Q2), (Q3), (R3), (R5) and let* $t < \min\left\{\tilde{b}_0, \tilde{b}_1\right\}$.
If $n_1 h^4 \to 0$, *then*

$$(n_1)^{\frac{1}{2}}\left\{\check{R}_h(t) - R(t)\right\} \xrightarrow{d} N\left(0, \sigma_R^2(t)\right).$$

If $n_1 h^4 \to c$, *for some* $c > 0$, *then*

$$(n_1)^{\frac{1}{2}}\left\{\check{R}_h(t) - R(t)\right\} \xrightarrow{d} N\left(\frac{1}{2}r^{(1)}(t) d_K c^{\frac{1}{2}}, \sigma_R^2(t)\right).$$

Chapter 4

Empirical likelihood approach

— *La vida siempre espera situaciones críticas*
para mostrar su lado brillante.
Paulo Coelho

4.1 Empirical likelihood

Likelihood methods can deal with incomplete data and correct for this problem. They can pool information from different sources or include information coming from the outside in the form of constraints that restrict the domain of the likelihood function or in the form of a prior distribution to be multiplied by the likelihood function. On the other hand, nonparametric inferences avoid the misspecification that parametric inferences can cause when the data of interest are wrongly assumed to follow one of the known and well studied parametric families.

Empirical likelihood methods can be defined as a combination of likelihood and nonparametric methods that allow the statisticians to use likelihood methods without having to assume that the data come from a known parametric family of distributions. It was first proposed by Thomas and Grunkemeier (1975) to set better confidence intervals for the Kaplan-Meier estimator. Later on, Owen (1988, 1990, 2001) and other authors showed the potential of this methodology. It is known, for example, that empirical likelihood is a desirable and natural inference procedure for deriving nonparametric and semiparametric confidence regions for mostly finite-dimensional parameters. Empirical likelihood confidence regions respect the range of the parameter space, are invariant under transformations, their shapes are determined by the data, in some cases there is no need to estimate the variance due to the internally studentizing and they are often Bartlett correctable.

When the interest is to determine if two samples come from the same population, it would be interesting to obtain confidence intervals and regions for an estimate of, for example, the relative density, $r(t)$, or the relative distribution, $R(t)$. Note that two identical populations have $R(t) = t$ and $r(t) = 1$, for $t \in (0, 1)$. Therefore, a simple visualization of the confidence region would allow to conclude if the two samples come from the same population or not. For example, when the testing is based on the relative distribution, one would reject the null hypothesis of equal populations if the identity function is outside the confidence region.

Claeskens *et al* (2003) developed an empirical likelihood procedure to set confidence intervals and regions for $R(t)$ in the case of a two-sample problem with complete data. Due to the nice properties shown by the empirical likelihood methods, we study in the following pages, the case of a two-sample problem with LTRC data via the relative distribution and the empirical likelihood methodology. We obtain in this setting a nonparametric generalization of Wilks theorem and confidence intervals for the value of $R(t)$, when t is a fixed point in the unit interval.

4.2 Two-sample test via empirical likelihood for LTRC data

Consider the two sample problem introduced in Chapter 3. We wish to construct a point and interval estimator for the relative distribution value $R(t) = \theta$ where t is a fixed point in $[0, 1]$. Note that this implies that there exists a value η such that $F_0(\eta) = t$ and $F_1(\eta) = \theta$.

Consider K a compactly supported kernel of order r and define $\mathbb{K}(x) = \int_{-\infty}^{x} K(u)\mathrm{d}u$. For $i = 1, \dots, n_0$ we define

$$U_{0i}(\eta) = \frac{\mathbb{K}\left(\frac{\eta - Y_{0i}}{h_0}\right)\delta_{0i}}{\alpha_0^{-1} G_0(Y_{0i})(1 - L_0(Y_{0i}))}$$

and analogously for $j = 1, \dots, n_1$, we consider

$$U_{1j}(\eta) = \frac{\mathbb{K}\left(\frac{\eta - Y_{1j}}{h_1}\right)\delta_{1j}}{\alpha_1^{-1} G_1(Y_{1j})(1 - L_1(Y_{1j}))}.$$

When the LTRC model was introduced in Subsection 1.1.1, it was proved that:

$$\mathrm{d}W_{01}(t) = \alpha_0^{-1} G_0(t)(1 - L_0(t))\mathrm{d}F_0(t)$$

and

$$\mathrm{d}W_{11}(t) = \alpha_1^{-1} G_1(t)(1 - L_1(t))\mathrm{d}F_1(t).$$

Based on these equations and under conditions (E2) and (E3) introduced later on, it is easy to check that

$$\mathbb{E}\left(U_{0i}(\eta)\right) = \int \frac{\mathbb{K}\left(\frac{\eta - u}{h_0}\right)\mathrm{d}W_{01}(u)}{\alpha_0^{-1} G_0(u)(1 - L_0(u))} = \int \mathbb{K}\left(\frac{\eta - u}{h_0}\right)\mathrm{d}F_0(u)$$

and

$$\mathbb{E}\left(U_{1j}(\eta)\right) = \int \frac{\mathbb{K}\left(\frac{\eta-u}{h_1}\right)\mathrm{d}W_{11}(u)}{\alpha_1^{-1}G_1(u)(1-L_1(u))} = \int \mathbb{K}\left(\frac{\eta-u}{h_1}\right)\mathrm{d}F_1(u),$$

which, under smooth conditions, approximates $F_0(\eta)$ and $F_1(\eta)$, respectively. Therefore, using Owen's (1988) idea, we can define the smoothed likelihood ratio function of θ by

$$\tilde{L}(\theta) = \sup_{(\tilde{p},\tilde{q},\eta)} \left(\prod_{i=1}^{n_0} n_0\tilde{p}_i\right)\left(\prod_{j=1}^{n_1} n_1\tilde{q}_j\right),$$

where $\tilde{p} = (\tilde{p}_1, \tilde{p}_2, \ldots, \tilde{p}_{n_0})$ and $\tilde{q} = (\tilde{q}_1, \tilde{q}_2, \ldots, \tilde{q}_{n_1})$ are two probability vectors consisting of nonnegative values and the supremum is taken over $(\tilde{p}, \tilde{q}, \eta)$ and subject to the following constraints:

$$t = \sum_{i=1}^{n_0} \tilde{p}_i U_{0i}(\eta) \text{ and } \theta = \sum_{j=1}^{n_1} \tilde{q}_j U_{1j}(\eta),$$
$$1 = \sum_{i=1}^{n_0} \tilde{p}_i \text{ and } 1 = \sum_{j=1}^{n_1} \tilde{q}_j.$$

However, since the definition of $U_{0i}(\eta)$ and $U_{1j}(\eta)$ involves some unknown quantities, we must replace them by some appropriate estimators. Noting that $U_{0i}(\eta)$ and $U_{1j}(\eta)$ can be rewritten in terms of other unknown quantities that can be easily estimated from the data:

$$U_{0i}(\eta) = \frac{\mathbb{K}\left(\frac{\eta-Y_{0i}}{h_0}\right)\delta_{0i}}{B_0(Y_{0i})(1-F_0(Y_{0i}^-))^{-1}}, \tag{4.1}$$
$$U_{1j}(\eta) = \frac{\mathbb{K}\left(\frac{\eta-Y_{1j}}{h_1}\right)\delta_{1j}}{B_1(Y_{1j})(1-F_1(Y_{1j}^-))^{-1}}, \tag{4.2}$$

we propose to define

$$V_{0i}(\eta) = \frac{\mathbb{K}\left(\frac{\eta-Y_{0i}}{h_0}\right)\delta_{0i}}{B_{0n_0}(Y_{0i})(1-\hat{F}_{0n_0}(Y_{0i}^-))^{-1}}, \tag{4.3}$$
$$V_{1j}(\eta) = \frac{\mathbb{K}\left(\frac{\eta-Y_{1j}}{h_1}\right)\delta_{1j}}{B_{1n_1}(Y_{1j})(1-\hat{F}_{1n_1}(Y_{1j}^-))^{-1}}, \tag{4.4}$$

where we replace in (4.1) and (4.2), B_0 and B_1 by their corresponding empirical estimates and F_0 and F_1 by their corresponding PL-estimates (see equation (1.5)). Before we go on in the discussion, it is worth noting here that the smoothed estimators of F_0 and F_1

introduced in Subsection 1.3.4 can be rewritten in terms of the V_{0i}'s and the V_{1j}'s (see (4.3) and (4.4) above). In fact, simple algebra allows to rewrite $\hat{F}_{0h_0}(\eta)$ and $\hat{F}_{1h_1}(\eta)$ as

$$\begin{aligned}\hat{F}_{0h_0}(\eta) &= \int \mathbb{K}\left(\frac{\eta - x}{h_0}\right) \mathrm{d}\hat{F}_{0n_0}(x) \\ &= \frac{1}{n_0}\sum_{i=1}^{n_0} \mathbb{K}\left(\frac{\eta - Y_{0i}}{h_0}\right) \frac{\delta_{0i}}{\left(1 - \hat{F}_{0n_0}(Y_{0i}^-)\right)^{-1} B_{0n_0}(Y_{0i})} \\ &= \frac{1}{n_0}\sum_{i=1}^{n_0} V_{0i}(\eta)\end{aligned}$$

and

$$\begin{aligned}\hat{F}_{1h_1}(\eta) &= \int \mathbb{K}\left(\frac{\eta - x}{h_1}\right) \mathrm{d}\hat{F}_{1n_1}(x) \\ &= \frac{1}{n_1}\sum_{j=1}^{n_1} \mathbb{K}\left(\frac{\eta - Y_{1j}}{h_1}\right) \frac{\delta_{1j}}{\left(1 - \hat{F}_{1n_1}(Y_{1j}^-)\right)^{-1} B_{1n_1}(Y_{1j})} \\ &= \frac{1}{n_1}\sum_{j=1}^{n_1} V_{1j}(\eta).\end{aligned}$$

We now define the estimated smoothed likelihood ratio function of θ by

$$L(\theta) = \sup_{(p,q,\eta)} \left(\prod_{i=1}^{n_0} n_0 p_i\right)\left(\prod_{j=1}^{n_1} n_1 q_j\right),$$

where $p = (p_1, p_2, \ldots, p_{n_0})$ and $q = (q_1, q_2, \ldots, q_{n_1})$ are two probability vectors consisting of nonnegative values and the supremum is taken over (p, q, η) and subject to the following constraints:

$$t = \sum_{i=1}^{n_0} p_i V_{0i}(\eta) \text{ and } \theta = \sum_{j=1}^{n_1} q_j V_{1j}(\eta), \tag{4.5}$$

$$1 = \sum_{i=1}^{n_0} p_i \text{ and } 1 = \sum_{j=1}^{n_1} q_j. \tag{4.6}$$

Based on $L(\theta)$, the estimated smoothed log-likelihood function (multiplied by minus two) is given by:

$$\ell(\theta) = -2\left\{\sum_{i=1}^{n_0} \ln(n_0 p_i) + \sum_{j=1}^{n_1} \ln(n_1 q_j)\right\},$$

where using the Lagrange multiplier method, it is straightforward to show that:

$$p_i = \frac{1}{n_0}\frac{1}{\lambda_0(V_{0i}(\eta) - t) + 1}, \tag{4.7}$$

$$q_j = \frac{1}{n_1}\frac{1}{\lambda_1(V_{1j}(\eta) - \theta) + 1}, \tag{4.8}$$

where $(\lambda_0, \lambda_1, \eta) = (\lambda_0(\theta), \lambda_1(\theta), \eta(\theta))$ are solutions to the following score equations:

$$\frac{\partial}{\partial\lambda_0}\ell(\theta) = \sum_{i=1}^{n_0} \frac{V_{0i}(\eta) - t}{1 + \lambda_0(V_{0i}(\eta) - t)} = 0, \tag{4.9}$$

$$\frac{\partial}{\partial\lambda_1}\ell(\theta) = \sum_{j=1}^{n_1} \frac{V_{1j}(\eta) - \theta}{1 + \lambda_1(V_{1j}(\eta) - \theta)} = 0, \tag{4.10}$$

$$\frac{\partial}{\partial\eta}\ell(\theta) = 2\lambda_0 \sum_{i=1}^{n_0} \frac{V_{0i}^{(1)}(\eta)}{1 + \lambda_0(V_{0i}(\eta) - t)} + 2\lambda_1 \sum_{j=1}^{n_1} \frac{V_{1j}^{(1)}(\eta)}{1 + \lambda_1(V_{1j}(\eta) - \theta)} = 0 \tag{4.11}$$

with

$$V_{0i}^{(1)}(\eta) = \frac{1}{h_0} K\left(\frac{\eta - Y_{0i}}{h_0}\right) \frac{\delta_{0i}}{B_{0n_0}(Y_{0i})(1 - \hat{F}_{0n_0}(Y_{0i}^-))^{-1}},$$

$$V_{1j}^{(1)}(\eta) = \frac{1}{h_1} K\left(\frac{\eta - Y_{1j}}{h_1}\right) \frac{\delta_{1j}}{B_{1n_1}(Y_{1j})(1 - \hat{F}_{1n_1}(Y_{1j}^-))^{-1}}.$$

Next, we show in detail the computations required above to obtain (4.7)-(4.11) by the Lagrange multiplier method. Note that the objective is to find the minimizer of $\ell(\theta)$ subject to the constraints (4.5) and (4.6). By the Lagrange multiplier method, this is equivalent to minimize $E(\theta)$ below:

$$\begin{aligned} E(\theta) &= \ell(\theta) - \mu_0 \left(t - \sum_{i=1}^{n_0} p_i V_{0i}(\eta)\right) - \mu_1 \left(\theta - \sum_{j=1}^{n_1} q_j V_{1j}(\eta)\right) \\ &\quad -\nu_0 \left(1 - \sum_{i=1}^{n_0} p_i\right) - \nu_1 \left(1 - \sum_{j=1}^{n_1} q_j\right), \end{aligned}$$

where $(\mu_0, \mu_1, \nu_0, \nu_1)$ is the Lagrange multiplier.

Therefore, we need to solve the following equations:

$$0 = \frac{\partial E(\theta)}{\partial\mu_0} = \sum_{i=1}^{n_0} p_i V_{0i}(\eta) - t, \tag{4.12}$$

$$0 = \frac{\partial E(\theta)}{\partial\nu_0} = \sum_{i=1}^{n_0} p_i - 1, \tag{4.13}$$

$$0 = \frac{\partial E(\theta)}{\partial p_i} = \frac{-2}{p_i} + \mu_0 V_{0i}(\eta) + \nu_0 \text{ with } i = 1, \ldots, n_0, \tag{4.14}$$

$$0 = \frac{\partial E(\theta)}{\partial \mu_1} = \sum_{j=1}^{n_1} q_j V_{1j}(\eta) - \theta, \tag{4.15}$$

$$0 = \frac{\partial E(\theta)}{\partial \nu_1} = \sum_{j=1}^{n_1} q_j - 1, \tag{4.16}$$

$$0 = \frac{\partial E(\theta)}{\partial q_j} = \frac{-2}{q_j} + \mu_1 V_{1j}(\eta) + \nu_1 \text{ with } j = 1, \ldots, n_1, \tag{4.17}$$

$$0 = \frac{\partial E(\theta)}{\partial \eta} = \mu_0 \sum_{i=1}^{n_0} p_i V_{0i}^{(1)}(\eta) + \mu_1 \sum_{j=1}^{n_1} q_j V_{1j}^{(1)}(\eta). \tag{4.18}$$

From equation (4.14), we have

$$(\mu_0 V_{0i}(\eta) + \nu_0)\, p_i = 2 \text{ for } i = 1, \ldots, n_0. \tag{4.19}$$

Summing over i and using equations (4.12) and (4.13), we obtain that

$$\nu_0 = 2n_0 - \mu_0 t. \tag{4.20}$$

Similarly, from equation (4.17), it follows that

$$(\mu_1 V_{1j}(\eta) + \nu_1)\, q_j = 2 \text{ for } j = 1, \ldots, n_1. \tag{4.21}$$

If we now sum over j, equations (4.15) and (4.16) allow to conclude that

$$\nu_1 = 2n_1 - \mu_1 \theta. \tag{4.22}$$

From equations (4.12), (4.13), (4.19) and (4.20), it follows that

$$\sum_{i=1}^{n_0} \frac{2\,(V_{0i}(\eta) - t)}{\mu_0 V_{0i}(\eta) + 2n_0 - \mu_0 t} = 0. \tag{4.23}$$

In a similar way, from equations (4.15), (4.16), (4.21) and (4.22), it is easily obtained that

$$\sum_{j=1}^{n_1} \frac{2\,(V_{1j}(\eta) - \theta)}{\mu_1 V_{1j}(\eta) + 2n_1 - \mu_1 \theta} = 0. \tag{4.24}$$

Defining $\lambda_0 = \frac{\mu_0}{2n_0}$ and $\lambda_1 = \frac{\mu_1}{2n_1}$, formulas (4.23) and (4.24) can now be rewritten as given above in (4.9) and (4.10). From equations (4.20) and (4.19), we now obtain

$$p_i = \frac{1}{n_0} \frac{1}{\lambda_0 V_{0i}(\eta) + \frac{\nu_0}{2n_0}} = \frac{1}{n_0} \frac{1}{\lambda_0\,(V_{0i}(\eta) - t) + 1},$$

i.e., equation (4.7). Similarly, equation (4.8) follows from equations (4.22) and (4.21) and finally, equation (4.11) is obtained by rewriting equation (4.18) in terms of λ_0 and λ_1.

Hence, the estimated smoothed loglikelihood function can now be rewritten as

$$\ell(\theta) = 2\sum_{i=1}^{n_0} \ln\left(\lambda_0(V_{0i}(\eta) - t) + 1\right) + 2\sum_{j=1}^{n_1} \ln\left(\lambda_1(V_{1j}(\eta) - \theta) + 1\right). \tag{4.25}$$

The value of θ minimizing $\ell(\theta)$ will give a point estimate, $\hat{\theta}$, of $R(t) = \theta_0$ based on empirical likelihood.

For practical matters, the score equations (4.9), (4.10) and (4.11) can be solved in two stages. In the first stage, we fix η and obtain $\lambda_0 = \lambda_0(\eta)$ and $\lambda_1 = \lambda_1(\eta)$ from the equations (4.9) and (4.10). In the second stage, $\tilde{\eta}$ is obtained as a solution to equation (4.11).

When the objective is the construction of a confidence interval for $R(t)$, all those values of θ for which the null hypothesis $H_0 : R(t) = \theta$ can not be rejected will be included in the confidence interval for $R(t)$. Since large values of $\ell(\theta)$ favors the (two-sided) alternative hypothesis, a formal definition of an asymptotic confidence interval for $R(t)$ based on empirical likelihood is given by:

$$I_{1-\alpha} = \{\theta : \ell(\theta) \leq c\},$$

where c is a value such that $P_{H_0}(I_{1-\alpha}) \to 1 - \alpha$.

Next, we introduce some conditions that will be needed later on:

(D13) There exists an integer r such that the densities f_k, $k = 0, 1$ satisfy that $f_k^{(r-1)}$ exists in a neighbourhood of η_0 and it is continuous at η_0. Additionally, $f_0(\eta_0)f_1(\eta_0) > 0$. Here η_0 denotes that value for which $F_0(\eta_0) = t$ and $F_1(\eta_0) = \theta_0$.

(K13) K denotes a twice differentiable kernel of order r ($r \geq 2$) with support on $[-1, 1]$, i.e.

$$\int x^k K(x)\mathrm{d}x = \begin{cases} 1, & k = 0, \\ 0, & 1 \leq k \leq r-1, \\ c \neq 0, & k = r. \end{cases}$$

(B6) The bandwidth sequences $h_0 = h_{0n_0}$ and $h_1 = h_{1n_1}$ tend to zero as n_0 and n_1 tend to infinity. Besides, as the sample sizes increase, $n_0 h_0^{4r} \to 0$, $n_1 h_1^{4r} \to 0$, $\frac{n_0 h_0^{2r}}{\ln(n_0)} \to \infty$ and $\frac{n_1 h_1^{2r}}{\ln(n_1)} \to \infty$.

(E2) For $k = 0, 1$, the value of $\alpha_k^{-1} G_k(a_{G_k})(1 - L_k(b_{L_k}))$ is positive.

(E3) For $k = 0, 1$, it is assumed that $a_{G_k} < a_{W_k}$ and that the integral condition

$$\int_{a_{W_k}}^{b} \frac{\mathrm{d}W_{k1}(t)}{B_k^3(t)} < \infty$$

is satisfied for $a_{W_k} < b < b_{W_k}$.

Before presenting the main results, we next introduce some previous lemmas. For the sake of simplicity we will denote from here on $h=\max\{h_0,h_1\}$, $\delta=O(h^r)$, $n=n_0+n_1$ and $\gamma_{0n_0}=\frac{n_0}{n_0+n_1}\to\gamma_0$ and $\gamma_{1n_1}=\frac{n_1}{n_0+n_1}\to\gamma_1$, where $\gamma_0=\frac{\kappa^2}{\kappa^2+1}$ and $\gamma_1=\frac{1}{\kappa^2+1}$.

Lemma 4.2.1. *Assume that conditions (D13), (S2), (K13), (B6), (E2) and (E3) hold.*

(i) For each fixed η, we have

$$\begin{aligned}
E[U_{01}(\eta)-t] &= F_0(\eta)-F_0(\eta_0)+O(h_0^r),\\
Var[U_{01}(\eta)-t] &= \int \mathbb{K}^2\left(\frac{\eta-x}{h_0}\right)M_0(x)\mathrm{d}F_0(x)-F_0^2(\eta)+O(h_0^r)\\
&= \int_{-\infty}^{\eta}\frac{\mathrm{d}F_0(u)}{\alpha_0^{-1}G_0(u)(1-L_0(u))}-F_0^2(\eta)+O(h_0),
\end{aligned}\tag{4.26}$$

$$\begin{aligned}
E[U_{11}(\eta)-\theta_0] &= F_1(\eta)-F_1(\eta_0)+o(h_1^r),\\
Var[U_{11}(\eta)-\theta_0] &= \int \mathbb{K}^2\left(\frac{\eta-x}{h_1}\right)M_1(x)\mathrm{d}F_1(x)-F_1^2(\eta)+O(h_1^r)\\
&= \int_{-\infty}^{\eta}\frac{\mathrm{d}F_1(u)}{\alpha_1^{-1}G_1(u)(1-L_1(u))}-F_1^2(\eta)+O(h_1),
\end{aligned}$$

where $M_0(x)=\frac{1-F_0(x)}{B_0(x)}$ and $M_1(x)=\frac{1-F_1(x)}{B_1(x)}$.

(ii) Uniformly for $\eta:|\eta-\eta_0|<\delta$ we have

$$n_0^{-1}\sum_{i=1}^{n_0}(U_{0i}(\eta)-t) = O(h_0^r)+O_P\left(\delta+n_0^{-1/2}\right),\tag{4.27}$$

$$n_0^{-1}\sum_{i=1}^{n_0}(U_{0i}(\eta)-t)^2 = \int \mathbb{K}^2\left(\frac{\eta-x}{h_0}\right)M_0(x)\mathrm{d}F_0(x)-F_0^2(\eta_0) +O_P\left(\delta+h_0^r+n_0^{-1/2}\right)\tag{4.28}$$

$$= \int_{-\infty}^{\eta_0}\frac{\mathrm{d}F_0(u)}{\alpha_0^{-1}G_0(u)(1-L_0(u))}-F_0^2(\eta_0) +O_P\left(\delta+h_0+n_0^{-1/2}\right),\tag{4.29}$$

$$n_1^{-1}\sum_{j=1}^{n_1}(U_{1j}(\eta)-\theta_0) = O(h_1^r)+O_P\left(\delta+n_1^{-1/2}\right),\tag{4.30}$$

$$n_1^{-1}\sum_{j=1}^{n_1}(U_{1j}(\eta)-\theta_0)^2 = \int \mathbb{K}^2\left(\frac{\eta-x}{h_1}\right)M_1(x)\mathrm{d}F_1(x)-F_1^2(\eta_0) +O_P\left(\delta+h_1^r+n_1^{-1/2}\right)\tag{4.31}$$

$$= \int_{-\infty}^{\eta_0}\frac{\mathrm{d}F_1(u)}{\alpha_1^{-1}G_1(u)(1-L_1(u))}-F_1^2(\eta_0) +O_P\left(\delta+h_1+n_1^{-1/2}\right).\tag{4.32}$$

(iii)

$$n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\eta_0+\delta)-t) = O\left(h_0^r\right)+O\left(\delta+n_0^{-1/2}(\ln n_0)^{1/2}\right) \quad \text{a.s.,} \tag{4.33}$$

$$\begin{aligned} n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\eta_0+\delta)-t)^2 &= \int \mathbb{K}^2\left(\frac{\eta_0-x}{h_0}\right)M_0(x)\mathrm{d}F_0(x)-F_0^2(\eta_0) \\ &\quad +O\left(\delta+h_0^r+n_0^{-1/2}(\ln n_0)^{1/2}\right) \quad \text{a.s.} \qquad (4.34) \\ &= \int_{-\infty}^{\eta_0}\frac{\mathrm{d}F_0(u)}{\alpha_0^{-1}G_0(u)(1-L_0(u))}-F_0^2(\eta_0) \\ &\quad +O\left(\delta+h_0+n_0^{-1/2}(\ln n_0)^{1/2}\right) \quad \text{a.s.,} \qquad (4.35) \end{aligned}$$

$$n_1^{-1}\sum_{j=1}^{n_1}(V_{1j}(\eta_0+\delta)-\theta_0) = O\left(h_1^r\right)+O\left(\delta+n_1^{-1/2}(\ln n_1)^{1/2}\right) \quad \text{a.s.,} \tag{4.36}$$

$$\begin{aligned} n_1^{-1}\sum_{j=1}^{n_1}(V_{1j}(\eta_0+\delta)-\theta_0)^2 &= \int \mathbb{K}^2\left(\frac{\eta_0-x}{h_1}\right)M_1(x)\mathrm{d}F_1(x)-F_1^2(\eta_0) \\ &\quad +O\left(\delta+h_1^r+n_1^{-1/2}(\ln n_1)^{1/2}\right) \quad \text{a.s.} \qquad (4.37) \\ &= \int_{-\infty}^{\eta_0}\frac{\mathrm{d}F_1(u)}{\alpha_1^{-1}G_1(u)(1-L_1(u))}-F_1^2(\eta_0) \\ &\quad +O\left(\delta+h_1+n_1^{-1/2}(\ln n_1)^{1/2}\right) \quad \text{a.s.} \qquad (4.38) \end{aligned}$$

(iv) Uniformly for $\eta : |\eta-\eta_0|<\delta$, we have

$$\lambda_0(\eta) = O\left(h_0^r\right)+O_P\left(\delta+n_0^{-1/2}(\ln n_0)^{1/2}\right), \tag{4.39}$$

$$\lambda_1(\eta) = O\left(h_1^r\right)+O_P\left(\delta+n_1^{-1/2}(\ln n_1)^{1/2}\right). \tag{4.40}$$

Furthermore, on the boundary points, it follows that

$$\lambda_0(\eta_0\pm\delta) = O\left(h_0^r\right)+O\left(\delta+n_0^{-1/2}(\ln n_0)^{1/2}\right) \quad \text{a.s.,} \tag{4.41}$$

$$\lambda_1(\eta_0\pm\delta) = O\left(h_1^r\right)+O\left(\delta+n_1^{-1/2}(\ln n_1)^{1/2}\right) \quad \text{a.s.} \tag{4.42}$$

Proof of Lemma 4.2.1. We start with proving (i). It is easy to see that

$$\begin{aligned} E\left[U_{01}(\eta)\right] &= E\left[\frac{\mathbb{K}\left(\frac{\eta-Y_{01}}{h_0}\right)\delta_{01}}{\alpha_0^{-1}G_0(Y_{01})(1-L_0(Y_{01}))}\right] = \int\frac{\mathbb{K}\left(\frac{\eta-x}{h_0}\right)\mathrm{d}W_{01}(x)}{\alpha_0^{-1}G_0(x)(1-L_0(x))} \\ &= \int \mathbb{K}\left(\frac{\eta-x}{h_0}\right)\mathrm{d}F_0(x). \end{aligned}$$

Using integration by parts, a change of variables and a Taylor expansion, it follows that

$$\int \mathbb{K}\left(\frac{\eta-x}{h_0}\right) \mathrm{d}F_0(x) = \int F_0(\eta - h_0 u) K(u) \mathrm{d}u =$$
$$\int \left\{ F_0(\eta) + f_0(\eta)(-h_0 u) + \frac{1}{2} f_0^{(1)}(\eta)(-h_0 u)^2 + \cdots + \frac{1}{r!} f_0^{(r-1)}(\eta)(-h_0 u)^r + o\left(h_0^r\right) \right\}$$
$$\times K(u) \mathrm{d}u = F_0(\eta) + O\left(h_0^r\right).$$

On the other hand,

$$\begin{aligned} E\left[U_{01}^2(\eta)\right] &= E\left[\left(\frac{\mathbb{K}\left(\frac{\eta-Y_{01}}{h_0}\right)\delta_{01}}{\alpha_0^{-1} G_0(Y_{01})(1-L_0(Y_{01}))}\right)^2\right] = \int \frac{\mathbb{K}^2\left(\frac{\eta-x}{h_0}\right) \mathrm{d}W_{01}(x)}{\alpha_0^{-2} G_0^2(x)(1-L_0(x))^2} \\ &= \int \frac{\mathbb{K}^2\left(\frac{\eta-x}{h_0}\right) \mathrm{d}F_0(x)}{\alpha_0^{-1} G_0(x)(1-L_0(x))} = \int \frac{\mathbb{K}^2\left(\frac{\eta-x}{h_0}\right)(1-F_0(x)) \mathrm{d}F_0(x)}{B_0(x)} \\ &= \int \mathbb{K}^2\left(\frac{\eta-x}{h_0}\right) M_0(x) \mathrm{d}F_0(x), \end{aligned}$$

where we define $\mathbb{M}_0(x) = \int_{-\infty}^{x} M_0(y) \mathrm{d}F_0(y)$. Using integration by parts and a Taylor expansion, it follows that

$$\begin{aligned} \int \mathbb{K}^2\left(\frac{\eta-x}{h_0}\right) M_0(x) \mathrm{d}F_0(x) &= \frac{2}{h_0} \int_{\eta-h_0}^{\eta+h_0} \mathbb{K}\left(\frac{\eta-x}{h_0}\right) K\left(\frac{\eta-x}{h_0}\right) \mathbb{M}_0(x) \mathrm{d}x \\ &= \frac{2}{h_0} \int_{\eta-h_0}^{\eta+h_0} \mathbb{K}\left(\frac{\eta-x}{h_0}\right) K\left(\frac{\eta-x}{h_0}\right) \\ &\quad \times \left\{\mathbb{M}_0(\eta) + M_0(x') f_0(x')(x-\eta)\right\} \mathrm{d}x, \end{aligned}$$

where x' is a value between x and η.

Therefore,

$$\begin{aligned} \int \mathbb{K}^2\left(\frac{\eta-x}{h_0}\right) M_0(x) \mathrm{d}F_0(x) &= -\mathbb{M}_0(\eta) \int_{\eta-h_0}^{\eta+h_0} \mathrm{d}\mathbb{K}^2\left(\frac{\eta-x}{h_0}\right) + O(h_0) \\ &= \mathbb{M}_0(\eta)\left\{\mathbb{K}^2(1) - \mathbb{K}^2(-1)\right\} + O(h_0) \\ &= \int_{-\infty}^{\eta} \frac{\mathrm{d}F_0(y)}{\alpha_0^{-1} G_0(y)(1-L_0(y))} + O(h_0). \end{aligned}$$

In a similar way, it can be proved that

$$\begin{aligned} E\left[U_{11}(\eta)\right] &= F_1(\eta) + O\left(h_1^r\right), \\ E\left[U_{11}^2(\eta)\right] &= \int \mathbb{K}^2\left(\frac{\eta-x}{h_0}\right) M_1(x) \mathrm{d}F_1(x) = \int_{-\infty}^{\eta} \frac{\mathrm{d}F_1(y)}{\alpha_1^{-1} G_1(y)(1-L_1(y))} + O(h_1) \end{aligned}$$

and the proof of (i) is finished.

We now prove (ii). Using a Taylor expansion, $n_0^{-1}\sum_{i=1}^{n_0}(U_{0i}(\eta)-t)$ can be rewritten as follows:

$$\begin{aligned} n_0^{-1}\sum_{i=1}^{n_0}(U_{0i}(\eta)-t) &= n_0^{-1}\sum_{i=1}^{n_0}(U_{0i}(\eta_0)-t)+n_0^{-1}\sum_{i=1}^{n_0}U_{0i}^{(1)}(\eta_0)(\eta-\eta_0) \\ &+\frac{1}{2}n_0^{-1}\sum_{i=1}^{n_0}U_{0i}^{(2)}(\eta')(\eta-\eta_0)^2, \end{aligned} \tag{4.43}$$

where η' is a value between η_0 and η. From the Central Limit Theorem and (i), it is easy to prove that

$$n_0^{-1}\sum_{i=1}^{n_0}(U_{0i}(\eta_0)-t)=E\left[U_{01}(\eta_0)-t)\right]+O_P\left(n_0^{-1/2}\right)=O\left(h_0^r\right)+O_P\left(n_0^{-1/2}\right). \tag{4.44}$$

By the SLLN (Strong Law of Large Numbers), it follows that

$$n_0^{-1}\sum_{i=1}^{n_0}U_{0i}^{(1)}(\eta_0)=f_0(\eta_0)+O\left(h_0^r\right). \tag{4.45}$$

Besides, from conditions (K13), (B6) and (E2), it is easy to prove the existence of a constant c such that

$$\left|\frac{K^{(1)}\left(\frac{\eta'-Y_{0i}}{h_0}\right)}{B_0(Y_{0i})\left(1-F_0\left(Y_{0i}^-\right)\right)^{-1}}\right| \leq c.$$

This result and the fact that $h_0^{-2}\delta=O(1)$, imply that

$$\left|n_0^{-1}\sum_{i=1}^{n_0}U_{0i}^{(2)}(\eta')(\eta-\eta_0)^2\right|\leq\frac{\delta^2}{h_0^2n_0}\sum_{i=1}^{n_0}\left|K^{(1)}\left(\frac{\eta'-Y_{0i}}{h_0}\right)\frac{1-F_0(Y_{0i}^-)}{B_0(Y_{0i})}\right|=O\left(\delta\right). \tag{4.46}$$

Now, (4.43), (4.44), (4.45) and (4.46) prove (4.27).
Similarly, it can be proved that equations (4.28)-(4.32) hold.

In order to prove (iii), we first note that

$$n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\eta)-t)^2 \quad = \quad I_1+I_2+I_3,$$

where

$$\begin{aligned} I_1 &= n_0^{-1}\sum_{i=1}^{n_0}\left(U_{0i}(\eta)-t\right)^2, \\ I_2 &= n_0^{-1}\sum_{i=1}^{n_0}\left\{\mathbb{K}\left(\frac{\eta-Y_{0i}}{h_0}\right)\delta_{0i}\left[\frac{1-\hat{F}_{0n_0}(Y_{0i}^-)}{B_{0n_0}(Y_{0i})}-\frac{1-F_0(Y_{0i}^-)}{B_0(Y_{0i})}\right]^2\right\}, \\ I_3 &= \frac{2}{n_0}\sum_{i=1}^{n_0}(U_{0i}(\eta)-t)\left\{\mathbb{K}\left(\frac{\eta-Y_{0i}}{h_0}\right)\delta_{0i}\left[\frac{1-\hat{F}_{0n_0}(Y_{0i}^-)}{B_{0n_0}(Y_{0i}^-)}-\frac{1-F_0(Y_{0i})}{B_0(Y_{0i})}\right]\right\}. \end{aligned}$$

Using conditions (E2) and (E3) and the fact that

$$\begin{aligned}
\sup_{i=1,\ldots,n_0} \left|\hat{F}_{0n_0}(Y_{0i}^-) - F_0(Y_{0i}^-)\right| &= O\left(\left(n_0^{-1}\ln n_0\right)^{1/2}\right) \quad \text{a.s.},\\
\sup_{i=1,\ldots,n_0} \left|B_0(Y_{0i}) - B_{0n_0}(Y_{0i})\right| &= O_P\left(n_0^{-1/2}\right),\\
\sup_{i=1,\ldots,n_0} \left|B_0(Y_{0i}) - B_{0n_0}(Y_{0i})\right| &= O\left(\left(n_0^{-1}\ln n_0\right)^{1/2}\right) \quad \text{a.s.},
\end{aligned}$$

it is easy to show that

$$\begin{aligned}
\sup_{i=1,\ldots,n_0} \left|B_{0n_0}^{-1}(Y_{0i})\right| &= O_P(1),\\
\sup_{i=1,\ldots,n_0} \left|\frac{1}{B_{0n_0}(Y_{0i})} - \frac{1}{B_0(Y_{0i})}\right| &= O\left(\left(n_0^{-1}\ln n_0\right)^{1/2}\right) \quad \text{a.s.}
\end{aligned}$$

Therefore, since

$$\frac{1-\hat{F}_{0n_0}(Y_{0i}^-)}{B_{0n_0}(Y_{0i})} - \frac{1-F_0(Y_{0i}^-)}{B_0(Y_{0i})} = \frac{F_0(Y_{0i}^-) - \hat{F}_{0n_0}(Y_{0i}^-)}{B_{0n_0}(Y_{0i})} + (1-F_0(Y_{0i}^-))\left\{\frac{1}{B_{0n_0}(Y_{0i})} - \frac{1}{B_0(Y_{0i})}\right\},$$

from conditions (K13) and (E2), it follows that $I_2 + I_3 = O\left(\left(n_0^{-1}\ln n_0\right)^{1/2}\right)$ a.s.

Consequently,

$$n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\eta)-t)^2 = n_0^{-1}\sum_{i=1}^{n_0}(U_{0i}(\eta)-t)^2 + O\left(\left(n_0^{-1}\ln n_0\right)^{1/2}\right) \quad \text{a.s.} \quad (4.47)$$

Now, using Bernstein's inequality and the Borel-Cantelli lemma, it can be proved that

$$\begin{aligned}
&\left|n_0^{-1}\sum_{i=1}^{n_0}(U_{0i}(\eta_0+\delta)-t)^2 - E\left[n_0^{-1}\sum_{i=1}^{n_0}(U_{0i}(\eta_0+\delta)-t)^2\right]\right|\\
&= O\left(\left(n_0^{-1}\ln(n_0)\right)^{1/2}\right) \quad \text{a.s.} \qquad (4.48)
\end{aligned}$$

Next, we show in more detail how (4.48) can be obtained. From conditions (K13) and (E2), there exists a positive constant, m, such that

$$P\left(\left|(U_{0i}(\eta_0+\delta)-t)^2 - E\left[(U_{0i}(\eta_0+\delta)-t)^2\right]\right| \le m\right) = 1.$$

Let define the event

$$A_{n_0} = \left\{\left|n_0^{-1}\sum_{i=1}^{n_0}(U_{0i}(\eta_0+\delta)-t)^2 - E\left[n_0^{-1}\sum_{i=1}^{n_0}(U_{0i}(\eta_0+\delta)-t)^2\right]\right| > \alpha^{1/2} n_0^{-\frac{1}{2}}(\ln n_0)^{\frac{1}{2}}\right\},$$

where α is a positive constant.

Since $(U_{0i}(\eta_0+\delta)-t)^2$, for $i=1,\ldots,n_0$, are independent and identically distributed, using Bernstein inequality, it follows that

$$P(A_{n_0}) \leq \exp\left\{-\frac{\alpha\ln(n_0)}{2\sigma^2+\frac{2}{3}m\alpha^{\frac{1}{2}}n_0^{-\frac{1}{2}}(\ln n_0)^{\frac{1}{2}}}\right\} = 2n_0^{-R_{n_0}}, \tag{4.49}$$

where σ^2 denotes the variance of $(U_{01}(\eta_0+\delta)-t)^2$ and $R_{n_0}=\frac{\alpha}{2\sigma^2+\frac{2}{3}m\alpha^{\frac{1}{2}}n_0^{-\frac{1}{2}}(\ln n_0)^{\frac{1}{2}}}$.

Since $n_0^{-\frac{1}{2}}(\ln n_0)^{\frac{1}{2}}$ tends to zero as n_0 tends to infinity and α can be chosen such that $\alpha-2\sigma^2>0$, we then can conclude that $R_{n_0}>1+\epsilon$ for n_0 large enough and some $\epsilon>0$. Hence, from (4.49) it follows that $\sum_{i=1}^{\infty}P(A_{n_0})<\infty$. Now, using Borel-Cantelli lemma one obtains that $P\left(A_{n_0}\text{ infinitely often }\right)=0$, which implies (4.48) above.

Consequently, combining (4.48) with (4.47) and (4.26), we obtain (4.35), i.e.

$$\begin{aligned} n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\eta_0+\delta)-t)^2 &= \int_{-\infty}^{\eta_0}\frac{\mathrm{d}F_0(u)}{\alpha_0^{-1}G_0(u)(1-L_0(u))}-F_0^2(\eta_0) \\ &\quad +O\left(\delta+h_0+\left(n_0^{-1}\ln(n_0)\right)^{1/2}\right) \quad \text{a.s.} \end{aligned}$$

In a similar way, (4.33), (4.34) and (4.36)-(4.38) can be proved.

Next, we prove (iv). To this end, we first note that

$$\begin{aligned} \sum_{i=1}^{n_0}\frac{V_{0i}(\eta)-t}{1+\lambda_0(\eta)(V_{0i}(\eta)-t)} &= \sum_{i=1}^{n_0}\frac{(V_{0i}(\eta)-t)\left(1+\lambda_0(V_{0i}(\eta)-t)-\lambda_0(\eta)(V_{0i}(\eta)-t)\right)}{1+\lambda_0(\eta)(V_{0i}(\eta)-t)} \\ &= \sum_{i=1}^{n_0}(V_{0i}(\eta)-t)-\lambda_0(\eta)\sum_{i=1}^{n_0}\frac{(V_{0i}(\eta)-t)^2}{1+\lambda_0(\eta)(V_{0i}(\eta)-t)}. \end{aligned} \tag{4.50}$$

From equations (4.9) and (4.50), it follows that

$$n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\eta)-t)-\lambda_0(\eta)n_0^{-1}\sum_{i=1}^{n_0}\frac{(V_{0i}(\eta)-t)^2}{1+\lambda_0(\eta)(V_{0i}(\eta)-t)}=0. \tag{4.51}$$

Define $S_{n_0}=\max_{1\leq i\leq n_0}|V_{0i}(\eta)-t|$. Using conditions (K13) and (E2), it is easy to prove that $\max_{1\leq i\leq n_0}|U_{0i}(\eta)-t|\leq c_1$ a.s., where c_1 denotes a positive constant. Since

$$V_{0i}(\eta)-t=U_{0i}(\eta)-t+\mathbb{K}\left(\frac{\eta-Y_{0i}}{h_0}\right)\delta_{0i}\left[\frac{1-\hat{F}_{0n_0}(Y_{0i}^-)}{B_{0n_0}(Y_{0i})}-\frac{1-F_0(Y_{0i}^-)}{B_0(Y_{0i})}\right]$$

and

$$\sup\left|\frac{1-\hat{F}_{0n_0}(Y_{0i}^-)}{B_{0n_0}(Y_{0i})}-\frac{1-F_0(Y_{0i}^-)}{B_0(Y_{0i})}\right|=O\left(n_0^{-\frac{1}{2}}(\ln n_0)^{\frac{1}{2}}\right) \quad \text{a.s.},$$

it follows that $S_{n_0}\leq c$ a.s. for a positive constant c.

Besides, it follows that

$$n_0^{-1} \sum_{i=1}^{n_0} (V_{0i}(\eta) - t) \quad = \quad J_1 + J_2,$$

where

$$\begin{aligned} J_1 &= n_0^{-1} \sum_{i=1}^{n_0} (U_{0i}(\eta) - t) \,, \\ J_2 &= n_0^{-1} \sum_{i=1}^{n_0} \mathbb{K}\left(\frac{\eta - Y_{0i}}{h_0}\right) \delta_{0i} \left[\frac{1 - \hat{F}_{0n_0}(Y_{0i}^-)}{B_{0n_0}(Y_{0i})} - \frac{1 - F_0(Y_{0i}^-)}{B_0(Y_{0i})}\right] \\ &= O\left(\left(n_0^{-1} \ln n_0\right)^{1/2}\right) \quad \text{a.s.} \end{aligned}$$

Therefore, using this result and equations (4.27) and (4.51) it follows that

$$|\lambda_0(\eta)|\, n_0^{-1} \sum_{i=1}^{n_0} \frac{(V_{0i}(\eta) - t)^2}{1 + \lambda_0(\eta)(V_{0i}(\eta) - t)} \quad = \quad O\left(h_0^r\right) + O_P\left(\delta + n_0^{-1/2} + \left(n_0^{-1} \ln n_0\right)^{1/2}\right).$$

Finally, using (4.47) and (4.29), we can conclude that

$$\begin{aligned} |\lambda_0(\eta)|\, n_0^{-1} \sum_{i=1}^{n_0} \frac{(V_{0i}(\eta) - t)^2}{1 + \lambda_0(\eta)(V_{0i}(\eta) - t)} &\geq \frac{|\lambda_0(\eta)|}{1 + |\lambda_0(\eta)|\, S_{n_0}} n_0^{-1} \sum_{i=1}^{n_0} (V_{0i}(\eta) - t)^2 \qquad (4.52) \\ &\geq \frac{|\lambda_0(\eta)|}{1 + c\,|\lambda_0(\eta)|} \left\{\int_{-\infty}^{\eta_0} \frac{\mathrm{d}F_0(u)}{\alpha_0^{-1} G_0(u)(1 - L_0(u))} \right. \\ &\qquad \left. - F_0^2(\eta_0) + o_P(1)\right\} \end{aligned}$$

and therefore $\lambda_0(\eta) = O\left(h_0^2\right) + O_P\left(\delta + n_0^{-1/2} + \left(n_0^{-1} \ln n_0\right)^{1/2}\right)$, which proves (4.39). In a similar way (4.40) can be obtained.

Now, using (iii) in Lemma 4.2.1, we can proceed in a similar way as that used to obtain (4.52) and conclude that (4.41) and (4.42) are satisfied. □

Lemma 4.2.2. *Assume that conditions (D13), (S2), (K13), (B6), (E2) and (E3) hold and that* $\delta^{-1} = O\left(\frac{\sqrt{n}h}{\sqrt{\ln \ln n}}\right)$. *Then, for sufficiently large sample sizes* n_0 *and* n_1*, there exists, with probability one, a solution* $\tilde{\eta}$ *of equation (4.11), such that* $|\tilde{\eta} - \eta_0| < \delta$ a.s.

Proof of Lemma 4.2.2. Before starting with the proof we introduce some notation. Define

$$\Delta_\delta = \delta + h_0^r + n_0^{-1/2} (\ln n_0)^{\frac{1}{2}}$$

and

$$H(\theta_0, \eta) = n_0 H_0(\theta_0, \eta) + n_1 H_1(\theta_0, \eta),$$

where

$$\begin{aligned} H_0(\theta_0,\eta) &= n_0^{-1}\sum_{i=1}^{n_0}\ln\left\{1+\lambda_0(\eta)\left(V_{0i}(\eta)-t\right)\right\}, \\ H_1(\theta_0,\eta) &= n_1^{-1}\sum_{j=1}^{n_1}\ln\left\{1+\lambda_1(\eta)\left(V_{1j}(\eta)-\theta_0\right)\right\}. \end{aligned}$$

Using a Taylor expansion of $\ln(x)$ around 1, it follows that:

$$\begin{aligned} H_0(\theta_0,\eta_0+\delta) &= n_0^{-1}\sum_{i=1}^{n_0}\ln\left\{1+\lambda_0(\eta_0+\delta)\left(V_{0i}(\eta_0+\delta)-t\right)\right\} \\ &= n_0^{-1}\sum_{i=1}^{n_0}\Big\{\ln(1)+\lambda_0(\eta_0+\delta)\left(V_{0i}(\eta_0+\delta)-t\right) \\ &\quad -\frac{1}{2}\left[\lambda_0(\eta_0+\delta)\left(V_{0i}(\eta_0+\delta)-t\right)\right]^2 \\ &\quad +\frac{1}{3}(x')^{-3}\left[\lambda_0(\eta_0+\delta)\left(V_{0i}(\eta_0+\delta)-t\right)\right]^3\Big\}, \end{aligned}$$

where x' is a value between 1 and $\lambda_0(\eta_0+\delta)\left(V_{0i}(\eta_0+\delta)-t\right)$. Now, using (4.41) it is straightforward to show that

$$\begin{aligned} H_0(\theta_0,\eta_0+\delta) &= \lambda_0(\eta_0+\delta)n_0^{-1}\sum_{i=1}^{n_0}\left(V_{0i}(\eta_0+\delta)-t\right) \\ &\quad -\frac{1}{2}\lambda_0^2(\eta_0+\delta)n_0^{-1}\sum_{i=1}^{n_0}\left(\left(V_{0i}(\eta_0+\delta)-t\right)\right)^2+O\left(\Delta_\delta^3\right) \quad \text{a.s.} \end{aligned} \tag{4.53}$$

From (4.51), it follows that

$$\begin{aligned} 0 &= n_0^{-1}\sum_{i=1}^{n_0}\left(V_{0i}(\eta_0+\delta)-t\right)-\lambda_0(\eta_0+\delta)n_0^{-1}\sum_{i=1}^{n_0}\frac{\left(V_{0i}(\eta_0+\delta)-t\right)^2}{1+\lambda_0(\eta_0+\delta)(V_{0i}(\eta_0+\delta)-t)} \\ &= n_0^{-1}\sum_{i=1}^{n_0}\left(V_{0i}(\eta_0+\delta)-t\right)-\lambda_0(\eta_0+\delta)n_0^{-1} \\ &\quad \sum_{i=1}^{n_0}\frac{\left(V_{0i}(\eta_0+\delta)-t\right)^2\left[1+\lambda_0(\eta_0+\delta)(V_{0i}(\eta_0+\delta)-t)-\lambda_0(\eta_0+\delta)(V_{0i}(\eta_0+\delta)-t)\right]}{1+\lambda_0(\eta_0+\delta)(V_{0i}(\eta_0+\delta)-t)} \\ &= n_0^{-1}\sum_{i=1}^{n_0}\left(V_{0i}(\eta_0+\delta)-t\right) \\ &\quad -\lambda_0(\eta_0+\delta)n_0^{-1}\sum_{i=1}^{n_0}\left(V_{0i}(\eta_0+\delta)-t\right)^2 \\ &\quad +\lambda_0^2(\eta_0+\delta)n_0^{-1}\sum_{i=1}^{n_0}\frac{\left(V_{0i}(\eta_0+\delta)-t\right)^3}{1+\lambda_0(\eta_0+\delta)(V_{0i}(\eta_0+\delta)-t)} \end{aligned}$$

From (4.35) and (4.41) we next obtain that

$$\lambda_0(\eta_0+\delta) = n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\eta_0+\delta)-t)\left\{n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\eta_0+\delta)-t)^2\right\}^{-1} + O\left(\Delta_\delta^2\right) \quad \text{a.s.}$$

Replacing $\lambda_0(\eta_0+\delta)$ in (4.53) by the expression above, it follows that

$$H_0\left(\theta_0,\eta_0+\delta\right) = \frac{1}{2}\left\{n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\eta_0+\delta)-t)\right\}^2\left\{n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\eta_0+\delta)-t)^2\right\}^{-1} + O\left(\Delta_\delta^3\right) \quad \text{a.s.} \tag{4.54}$$

Now, a Taylor expansion yields

$$\begin{aligned} n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\eta_0+\delta)-t) &= n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\eta_0)-t) + \delta\left\{n_0^{-1}\sum_{i=1}^{n_0}V_{0i}^{(1)}(\eta')\right\} \\ &= n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\eta_0)-t) + \delta\left\{n_0^{-1}\sum_{i=1}^{n_0}V_{0i}^{(1)}(\eta_0)\right\} \\ &\quad +\delta\left\{n_0^{-1}\sum_{i=1}^{n_0}V_{0i}^{(1)}(\eta') - n_0^{-1}\sum_{i=1}^{n_0}V_{0i}^{(1)}(\eta_0)\right\}, \end{aligned} \tag{4.55}$$

where η' is between η_0 and $\eta_0+\delta$.

By a Taylor expansion, there exists an η^* between η_0 and η' for which it is satisfied that

$$\left|n_0^{-1}\sum_{i=1}^{n_0}V_{0i}^{(1)}(\eta') - n_0^{-1}\sum_{i=1}^{n_0}V_{0i}^{(1)}(\eta_0)\right| = \left|n_0^{-1}\sum_{i=1}^{n_0}V_{0i}^{(2)}(\eta^*)(\eta'-\eta_0)\right|.$$

Since

$$\begin{aligned} n_0^{-1}\sum_{i=1}^{n_0}V_{0i}^{(2)}(\eta^*) &= n_0^{-1}\sum_{i=1}^{n_0}U_{0i}^{(2)}(\eta^*) \\ &\quad + n_0^{-1}\frac{1}{h_0^2}\sum_{i=1}^{n_0}K^{(1)}\left(\frac{\eta^*-Y_{0i}}{h_0}\right)\delta_{0i}\left\{\frac{1-\hat{F}_{0n_0}(Y_{0i}^-)}{B_{0n_0}(Y_{0i})} - \frac{1-F_0(Y_{0i}^-)}{B_0(Y_{0i})}\right\} \\ &= n_0^{-1}\sum_{i=1}^{n_0}U_{0i}^{(2)}(\eta^*) \\ &\quad + O\left(h_0^{-2}n_0^{-\frac{1}{2}}(\ln n_0)^{\frac{1}{2}}\right) \quad \text{a.s.} \end{aligned}$$

and by the SLLN $n_0^{-1}\sum_{i=1}^{n_0}U_{0i}^{(2)}(\eta^*) = f_0^{(1)}(\eta^*) + O\left(h_0^r\right)$ a.s., condition (B6) gives

$$\left|n_0^{-1}\sum_{i=1}^{n_0}V_{0i}^{(2)}(\eta^*)\right| = \left|f_0^{(1)}(\eta^*)\right| + o(1) \quad \text{a.s.}$$

Consequently,

$$\left| n_0^{-1} \sum_{i=1}^{n_0} V_{0i}^{(1)}(\eta') - n_0^{-1} \sum_{i=1}^{n_0} V_{0i}^{(1)}(\eta_0) \right| = O(\delta) \quad \text{a.s.}$$

On the other hand, the first term in the right hand side of (4.55) is:

$$n_0^{-1} \sum_{i=1}^{n_0} (V_{0i}(\eta_0) - t) = n_0^{-1} \sum_{i=1}^{n_0} (U_{0i}(\eta_0) - t) + O\left(n_0^{-1/2} (\ln n_0)^{\frac{1}{2}}\right) \quad \text{a.s.} \tag{4.56}$$

and $n_0^{-1} \sum_{i=1}^{n_0} V_{0i}^{(1)}(\eta_0) - f_0(\eta_0) = O\left(\frac{\sqrt{\ln \ln n}}{\sqrt{nh}}\right)$ a.s. (see Gijbels and Wang (1993) and (iii) in Theorem 1.3.6). Since $\delta^{-1} = O\left(\frac{\sqrt{nh}}{\sqrt{\ln \ln n}}\right)$, it follows that $n_0^{-1} \sum_{i=1}^{n_0} V_{0i}^{(1)}(\eta_0) - f_0(\eta_0) = O(\delta)$ a.s. As a consequence, we have

$$n_0^{-1} \sum_{i=1}^{n_0} (V_{0i}(\eta_0 + \delta) - t) = n_0^{-1} \sum_{i=1}^{n_0} (V_{0i}(\eta_0) - t) + \delta f_0(\eta_0) + O\left(\delta^2\right) \quad \text{a.s.} \tag{4.57}$$

Besides, from equation (4.34), we have

$$n_0^{-1} \sum_{i=1}^{n_0} (V_{0i}(\eta) - t)^2 = \int \mathbb{K}^2 \left(\frac{\eta_0 - x}{h_0}\right) M_0(x) \mathrm{d}F_0(x) - F_0^2(\eta_0) + O\left(\delta + h_0^r + n_0^{-1/2} (\ln n_0)^{1/2}\right)$$

a.s. for any $\eta : |\eta - \eta_0| < \delta$. Hence, using (4.54), (4.56) and (4.57), we can conclude that

$$\begin{aligned} H_0(\theta_0, \eta_0 + \delta) = {} & \frac{1}{2 \int \mathbb{K}^2 \left(\frac{\eta_0 - x}{h_0}\right) M_0(x) \mathrm{d}F_0(x) - F_0^2(\eta_0) + O\left(\delta + h_0^r + n_0^{-1/2} (\ln n_0)^{1/2}\right)} \\ & \left[n_0^{-1} \sum_{i=1}^{n_0} (U_{0i}(\eta_0) - t) + O\left(n_0^{-1/2} (\ln n_0)^{\frac{1}{2}}\right) + \delta f_0(\eta_0) + O\left(\delta^2\right) \right]^2 \\ & + O\left(\Delta_\delta^3\right) \quad \text{a.s.} \end{aligned}$$

If we now consider $\delta = 0$, it follows that

$$\begin{aligned} H_0(\theta_0, \eta_0) = {} & \frac{1}{2 \int \mathbb{K}^2 \left(\frac{\eta_0 - x}{h_0}\right) M_0(x) \mathrm{d}F_0(x) - F_0^2(\eta_0) + O\left(h_0^r + n_0^{-1/2} (\ln n_0)^{1/2}\right)} \\ & \left[n_0^{-1} \sum_{i=1}^{n_0} (U_{0i}(\eta_0) - t) + O\left(n_0^{-1/2} (\ln n_0)^{\frac{1}{2}}\right) \right]^2 + O\left(\Delta_0^3\right) \quad \text{a.s.} \end{aligned}$$

Besides, for $\delta = O(h_0^r)$ satisfying $\delta^{-1} = o\left(n^{1/2} (\ln n)^{-1/2}\right)$ and under condition (B6), it follows that $H_0(\theta_0, \eta_0 + \delta) \geq H_0(\theta_0, \eta_0)$ a.s. In a similar way, it can be proved that $H_0(\theta_0, \eta_0 - \delta) \geq H_0(\theta_0, \eta_0)$ a.s. and that $H_1(\theta_0, \eta_0 \pm \delta) \geq H_1(\theta_0, \eta_0)$ a.s. Consequently, it follows that $H(\theta_0, \eta_0 \pm \delta) \geq H(\theta_0, \eta_0)$ a.s. and, with probability one, there exist a value $\tilde{\eta} \in (\eta_0 - \delta, \eta_0 + \delta)$ such that $H(\theta_0, \tilde{\eta}) \leq H(\theta_0, \eta)$, for all $\eta \in [\eta_0 - \delta, \eta_0 + \delta]$, i.e., with probability one, $H(\theta_0, \eta)$ achieves its minimum at a point $\tilde{\eta}$ in the interior of $A = \{\eta : |\eta - \eta_0| \leq \delta\}$. The proof concludes noting that this result is equivalent to say that $\tilde{\eta}$ is a root of the score equation (4.11), provided that (4.9) and (4.10) hold. □

Lemma 4.2.3. *Assume that conditions (D13), (S2), (K13), (B6), (E2) and (E3) hold. Then, for $\tilde{\eta}$ given in Lemma 4.2.2,*

$$\begin{aligned}\gamma_{1n_1} f_{1n_1}(\eta_0)\lambda_1(\tilde{\eta}) &= -\gamma_{0n_0} f_{0n_0}(\eta_0)\lambda_0(\tilde{\eta}) + O_P\left(h^{2r}\right) \\ &= -\gamma_{0n_0} f_{0n_0}(\eta_0)\lambda_0(\tilde{\eta}) + o_P\left(n^{-\frac{1}{2}}\right),\end{aligned}$$

$$\begin{aligned}\sqrt{n}\frac{\gamma_{1n_1}^{1/2} d_{2n}}{\gamma_{0n_0}^{1/2} d_{1n}^{1/2} f_{0n_0}(\eta_0)}\lambda_1(\tilde{\eta}) &\xrightarrow{d} N(0,1), \\ \sqrt{n}\frac{\gamma_{0n_0}^{1/2} d_{2n}}{\gamma_{1n_1}^{1/2} d_{1n}^{1/2} f_{1n_1}(\eta_0)}\lambda_0(\tilde{\eta}) &\xrightarrow{d} N(0,1),\end{aligned}$$

where

$$\begin{aligned}f_{0n_0}(\eta_0) &= n_0^{-1}\sum_{i=1}^{n_0} V_{0i}^{(1)}(\eta_0), \\ f_{1n_1}(\eta_0) &= n_1^{-1}\sum_{j=1}^{n_1} V_{1j}^{(1)}(\eta_0),\end{aligned}$$

$$\begin{aligned}d_{1n} &= \gamma_{1n_1} f_{1n_1}^2(\eta_0) v_0(\eta_0) + \gamma_{0n_0} f_{0n_0}^2(\eta_0) v_1(\eta_0), \\ d_{2n} &= \gamma_{1n_1} f_{1n_1}^2(\eta_0) u_{0n_0}(\eta_0) + \gamma_{0n_0} f_{0n_0}^2(\eta_0) u_{1n_1}(\eta_0)\end{aligned}$$

and

$$\begin{aligned}v_0(\eta_0) &= (1-F_0(\eta_0))^2\int_{-\infty}^{\eta_0}\frac{\mathrm{d}W_{01}(u)}{B_0^2(u)}, \\ v_1(\eta_0) &= (1-F_1(\eta_0))^2\int_{-\infty}^{\eta_0}\frac{\mathrm{d}W_{11}(u)}{B_1^2(u)}, \\ u_{0n_0}(\eta_0) &= \int_{-\infty}^{\infty}\mathbb{K}^2\left(\frac{\eta_0-x}{h_0}\right)M_0(x)\mathrm{d}F_0(x) - t^2, \\ u_{1n_1}(\eta_0) &= \int_{-\infty}^{\infty}\mathbb{K}^2\left(\frac{\eta_0-x}{h_1}\right)M_1(x)\mathrm{d}F_1(x) - \theta_0^2.\end{aligned}$$

Proof of Lemma 4.2.3. Before starting the proof we introduce some notation

$$\begin{aligned}Q_0(\eta,\lambda_0,\lambda_1) &= \frac{1}{n}\sum_{i=1}^{n_0}\frac{V_{0i}(\eta)-t}{1+\lambda_0\left(V_{0i}(\eta)-t\right)}, \\ Q_1(\eta,\lambda_0,\lambda_1) &= \frac{1}{n}\sum_{j=1}^{n_1}\frac{V_{1j}(\eta)-\theta_0}{1+\lambda_1\left(V_{1j}(\eta)-\theta_0\right)}, \\ Q_2(\eta,\lambda_0,\lambda_1) &= \frac{1}{n}\left\{\lambda_0\sum_{i=1}^{n_0}\frac{V_{0i}^{(1)}(\eta)}{1+\lambda_0\left(V_{0i}(\eta)-t\right)}\right. \\ &\quad \left. +\lambda_1\sum_{j=1}^{n_1}\frac{V_{1j}^{(1)}(\eta)}{1+\lambda_1\left(V_{1j}(\eta)-\theta_0\right)}\right\}.\end{aligned}$$

From Lemma 4.2.2, $Q_k(\tilde{\eta}, \lambda_0(\tilde{\eta}), \lambda_1(\tilde{\eta})) = 0$, for $k = 0, 1, 2$. By a Taylor expansion and equations (4.39) and (4.40), it follows that

$$\begin{aligned}0 = Q_k(\tilde{\eta}, \lambda_0(\tilde{\eta}), \lambda_1(\tilde{\eta})) &= Q_k(\eta_0, 0, 0) + \frac{\partial Q_k}{\partial \eta}(\eta_0, 0, 0)(\tilde{\eta} - \eta_0) + \frac{\partial Q_k}{\partial \lambda_0}(\eta_0, 0, 0)\lambda_0(\tilde{\eta}) \\ &\quad + \frac{\partial Q_k}{\partial \lambda_1}(\eta_0, 0, 0)\lambda_1(\tilde{\eta}) + O_P\left(h^{2r}\right).\end{aligned}$$

Besides, from condition (B6), it follows that $O_P\left(h^{2r}\right) = o_P(n^{-\frac{1}{2}})$.

On the other hand, using (4.34) and (4.37) in Lemma 4.2.1, it is easy to check that,

$$\begin{aligned}\frac{\partial Q_0}{\partial \eta}(\eta_0, 0, 0) &= \frac{\partial Q_2}{\partial \lambda_0}(\eta_0, 0, 0) = \gamma_{0n_0} f_{0n_0}(\eta_0) \\ \frac{\partial Q_1}{\partial \eta}(\eta_0, 0, 0) &= \frac{\partial Q_2}{\partial \lambda_1}(\eta_0, 0, 0) = \gamma_{1n_1} f_{1n_1}(\eta_0) \\ \frac{\partial Q_0}{\partial \lambda_0}(\eta_0, 0, 0) &= -\frac{n_0}{n} n_0^{-1} \sum_{i=1}^{n_0} (V_{0i}(\eta_0) - t)^2 \\ &= -\gamma_{0n_0} u_{0n_0}(\eta_0) + O\left(\delta + n^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}\right) \quad \text{a.s.} \\ \frac{\partial Q_1}{\partial \lambda_1}(\eta_0, 0, 0) &= -\frac{n_1}{n} n_1^{-1} \sum_{j=1}^{n_1} (V_{1j}(\eta_0) - \theta_0)^2 \\ &= -\gamma_{1n_1} u_{1n_1}(\eta_0) + O\left(\delta + n^{-\frac{1}{2}}(\ln n)^{\frac{1}{2}}\right) \quad \text{a.s.} \\ \frac{\partial Q_0}{\partial \lambda_1}(\eta_0, 0, 0) &= \frac{\partial Q_1}{\partial \lambda_0}(\eta_0, 0, 0) = \frac{\partial Q_2}{\partial \eta}(\eta_0, 0, 0) = 0.\end{aligned}$$

Therefore,

$$0 = Q_0(\eta_0, 0, 0) + \gamma_{0n_0} f_{0n_0}(\eta_0)(\tilde{\eta} - \eta_0) - \gamma_{0n_0} u_{0n_0}(\eta_0)\lambda_0(\tilde{\eta}) + O_P(h^{2r}),$$

$$0 = Q_1(\eta_0, 0, 0) + \gamma_{1n_1} f_{1n_1}(\eta_0)(\tilde{\eta} - \eta_0) - \gamma_{1n_1} u_{1n_1}(\eta_0)\lambda_1(\tilde{\eta}) + O_P(h^{2r})$$

and

$$0 = Q_2(\eta_0, 0, 0) + \gamma_{0n_0} f_{0n_0}(\eta_0)\lambda_0(\tilde{\eta}) + \gamma_{1n_1} f_{1n_1}(\eta_0)\lambda_1(\tilde{\eta}) + O_P(h^{2r}),$$

which can be expressed in vector form as given below

$$\begin{pmatrix} \lambda_0(\tilde{\eta}) \\ \lambda_1(\tilde{\eta}) \\ \tilde{\eta} - \eta_0 \end{pmatrix} = -S_n^{-1} \begin{pmatrix} Q_0(\eta_0, 0, 0) \\ Q_1(\eta_0, 0, 0) \\ 0 \end{pmatrix} + \begin{pmatrix} O_P\left(h^{2r}\right) \\ O_P\left(h^{2r}\right) \\ O_P\left(h^{2r}\right) \end{pmatrix},$$

where the Jacobian matrix S_n is as follows:

$$S_n = \begin{pmatrix} -\gamma_{0n_0} u_{0n_0}(\eta_0) & 0 & \gamma_{0n_0} f_{0n_0}(\eta_0) \\ 0 & -\gamma_{1n_1} u_{1n_1}(\eta_0) & \gamma_{1n_1} f_{1n_1}(\eta_0) \\ \gamma_{0n_0} f_{0n_0}(\eta_0) & \gamma_{1n_1} f_{1n_1}(\eta_0) & 0 \end{pmatrix},$$

and its inverse is obtained as

$$S_n^{-1} = \frac{1}{\det(S_n)} \begin{pmatrix} -\gamma_{1n_1}^2 f_{1n_1}^2(\eta_0) & \gamma_{0n_0}\gamma_{1n_1} f_{0n_0}(\eta_0) f_{1n_1}(\eta_0) & \gamma_{0n_0}\gamma_{1n_1} f_{0n_0}(\eta_0) u_{1n_1}(\eta_0) \\ \gamma_{0n_0}\gamma_{1n_1} f_{0n_0}(\eta_0) f_{1n_1}(\eta_0) & -\gamma_{0n_0}^2 f_{0n_0}^2(\eta_0) & \gamma_{0n_0}\gamma_{1n_1} f_{1n_1}(\eta_0) u_{0n_0}(\eta_0) \\ \gamma_{0n_0}\gamma_{1n_1} f_{0n_0}(\eta_0) u_{1n_1}(\eta_0) & \gamma_{0n_0}\gamma_{1n_1} f_{1n_1}(\eta_0) u_{0n_0}(\eta_0) & \gamma_{0n_0}\gamma_{1n_1} u_{0n_0}(\eta_0) u_{1n_1}(\eta_0) \end{pmatrix},$$

where

$$\det(S_n) = \gamma_{0n_0}\gamma_{1n_1} \left\{ \gamma_{0n_0} f_{0n_0}^2(\eta_0) u_{1n_1}(\eta_0) + \gamma_{1n_1} f_{1n_1}^2(\eta_0) u_{0n_0}(\eta_0) \right\}.$$

Therefore, $\lambda_0(\tilde{\eta})$ and $\lambda_1(\tilde{\eta})$ can now be rewritten as linear combinations of $Q_k(\eta_0, 0, 0)$ $(k = 0, 1)$, plus negligible remainder terms as given below:

$$\begin{aligned} \lambda_0(\tilde{\eta}) &= -\frac{\gamma_{1n_1} f_{1n_1}(\eta_0)}{\det(S_n)} \{\gamma_{0n_0} f_{0n_0}(\eta_0) Q_1(\eta_0, 0, 0) - \gamma_{1n_1} f_{1n_1}(\eta_0) Q_0(\eta_0, 0, 0)\} \\ &\quad + O_P\left(h^{2r}\right), \qquad (4.58) \\ \lambda_1(\tilde{\eta}) &= \frac{\gamma_{0n_0} f_{0n_0}(\eta_0)}{\det(S_n)} \{\gamma_{0n_0} f_{0n_0}(\eta_0) Q_1(\eta_0, 0, 0) - \gamma_{1n_1} f_{1n_1}(\eta_0) Q_0(\eta_0, 0, 0)\} \\ &\quad + O_P\left(h^{2r}\right), \qquad (4.59) \end{aligned}$$

which proves the first part of the lemma. Now, the Central Limit Theorem gives

$$\sqrt{n} \begin{pmatrix} \frac{Q_0(\eta_0,0,0)}{\sqrt{\gamma_{0n_0}}} \\ \frac{Q_1(\eta_0,0,0)}{\sqrt{\gamma_{1n_1}}} \end{pmatrix} \xrightarrow{d} N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} v_0(\eta_0) & 0 \\ 0 & v_1(\eta_0) \end{pmatrix} \right\}.$$

The proof concludes after combining this result with equations (4.58) and (4.59). □

Theorem 4.2.4. *Under conditions (D13), (S2), (K13), (B6), (E2) and (E3), it follows that* $\frac{d_2}{d_1}\ell(\theta) \xrightarrow{d} \chi_1^2$ *under* H_0*, where*

$$\begin{aligned} d_1 &= \gamma_1 f_1^2(\eta_0) v_0(\eta_0) + \gamma_0 f_0^2(\eta_0) v_1(\eta_0), \\ d_2 &= \gamma_1 f_1^2(\eta_0) u_0(\eta_0) + \gamma_0 f_0^2(\eta_0) u_1(\eta_0), \\ u_0(\eta_0) &= \int_{-\infty}^{\eta_0} \frac{dF_0(u)}{G_0(u)(1 - L_0(u))\alpha_0^{-1}} - t^2, \\ u_1(\eta_0) &= \int_{-\infty}^{\eta_0} \frac{dF_1(u)}{G_1(u)(1 - L_1(u))\alpha_1^{-1}} - \theta_0^2. \end{aligned}$$

Proof of Theorem 4.2.4. Using Lemma 4.2.3 and a Taylor expansion of $\ln(x)$ around 1, it follows that

$$\begin{aligned} \ell(\theta_0) &= 2n_0\lambda_0(\tilde{\eta}) n_0^{-1} \sum_{i=1}^{n_0} (V_{0i}(\tilde{\eta}) - t) - n_0\lambda_0^2(\tilde{\eta}) n_0^{-1} \sum_{i=1}^{n_0} (V_{0i}(\tilde{\eta}) - t)^2 \\ &\quad + 2n_1\lambda_1(\tilde{\eta}) n_1^{-1} \sum_{j=1}^{n_1} (V_{1j}(\tilde{\eta}) - \theta_0) - n_1\lambda_1^2(\tilde{\eta}) n_1^{-1} \sum_{j=1}^{n_1} (V_{1j}(\tilde{\eta}) - \theta_0)^2 + o_P(1). \end{aligned}$$

Besides, if we consider $\eta = \tilde{\eta}$ in (4.51), it follows that

$$n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\tilde{\eta})-t)-\lambda_0(\tilde{\eta})n_0^{-1}\sum_{i=1}^{n_0}\frac{(V_{0i}(\tilde{\eta})-t)^2}{1+\lambda_0(\tilde{\eta})(V_{0i}(\tilde{\eta})-t)}=0.$$

From this expression, straightforwardly it follows that

$$n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\tilde{\eta})-t) = \lambda_0(\tilde{\eta})n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\tilde{\eta})-t)^2-\lambda_0^2(\tilde{\eta})n_0^{-1}\sum_{i=1}^{n_0}\frac{(V_{0i}(\tilde{\eta})-t)^3}{1+\lambda_0(\tilde{\eta})(V_{0i}(\tilde{\eta})-t)}.$$

Now, using Lemma 4.2.3, we can conclude that

$$n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\tilde{\eta})-t) = \lambda_0(\tilde{\eta})n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\tilde{\eta})-t)^2+o_P\left(n^{-1}\right).$$

In a similar way, we can prove that

$$n_1^{-1}\sum_{j=1}^{n_1}(V_{1j}(\tilde{\eta})-\theta_0) = \lambda_1(\tilde{\eta})n_1^{-1}\sum_{j=1}^{n_1}(V_{1j}(\tilde{\eta})-\theta_0)^2+o_P\left(n^{-1}\right).$$

Besides, from equations (4.35) and (4.38) it follows that

$$\begin{aligned} n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}(\tilde{\eta})-t)^2 &= u_0(\eta_0)+o_P(1), \\ n_1^{-1}\sum_{j=1}^{n_1}(V_{1j}(\tilde{\eta})-\theta_0)^2 &= u_1(\eta_0)+o_P(1). \end{aligned}$$

Hence, using Lemmas 4.2.1 and 4.2.3, we can rewrite $\ell(\theta_0)$ as follows

$$\begin{aligned} \ell(\theta_0) &= 2n_0\lambda_0(\tilde{\eta})n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}-t)-n_0\lambda_0^2(\tilde{\eta})n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}-t)^2 \\ &\quad +2n_1\lambda_1(\tilde{\eta})n_1^{-1}\sum_{j=1}^{n_1}(V_{1j}-\theta_0)-n_1\lambda_1^2(\tilde{\eta})n_1^{-1}\sum_{j=1}^{n_1}(V_{1j}-\theta_0)^2+o_P(1) \\ &= n_0\lambda_0^2(\tilde{\eta})n_0^{-1}\sum_{i=1}^{n_0}(V_{0i}-t)^2+n_1\lambda_1^2(\tilde{\eta})n_1^{-1}\sum_{j=1}^{n_1}(V_{1j}-\theta_0)^2+o_P(1) \\ &= n_0\lambda_0^2(\tilde{\eta})u_{0n_0}(\eta_0)+n_1\lambda_1^2(\tilde{\eta})u_{1n_1}(\eta_0)+o_P(1) \\ &= n\gamma_{0n_0}\lambda_0^2(\tilde{\eta})u_{0n_0}(\eta_0)+n\gamma_{1n_1}\lambda_1^2(\tilde{\eta})u_{1n_1}(\eta_0)+o_P(1) \\ &= n\lambda_0^2(\tilde{\eta})\left\{\gamma_{0n_0}u_{0n_0}(\eta_0)+\frac{\gamma_{0n_0}^2f_{0n_0}^2(\eta_0)}{\gamma_{1n_1}f_{1n_1}^2(\eta_0)}u_{1n_1}(\eta_0)\right\}+o_P(1) \\ &= \frac{d_{1n}}{d_{2n}}\left\{\sqrt{n}\lambda_0(\tilde{\eta})\frac{\sqrt{\gamma_{0n_0}}d_{2n}}{\sqrt{\gamma_{1n_1}}f_{1n_1}(\eta_0)\sqrt{d_{1n}}}\right\}^2+o_P(1)\xrightarrow{d}\frac{d_1}{d_2}\chi_1^2, \end{aligned}$$

which concludes the proof. □

Theorem 4.2.4 is the nonparametric version of the Wilks theorem in the relative distribution context with LTRC data. Unlike in the case of complete data, the limit distribution of $\ell(\theta)$ under H_0 is not asymptotically pivotal because it depends on two unknown quantities, d_1 and d_2. Based on this theorem and using appropriate estimates of d_1 and d_2, let say $\tilde{d}_1$ and $\tilde{d}_2$ respectively, a possible approach to construct a $(1-\alpha)$-level confidence interval for $R(t) = \theta_0$ is:

$$I_{1-\alpha} = \left\{ \theta : \frac{\tilde{d}_2}{\tilde{d}_1} \ell(\theta) \leq c_{1-\alpha}(\chi_1^2) \right\},$$

where $c_{1-\alpha}(\chi_1^2)$ is the $(1-\alpha)$ quantile of a χ_1^2 distribution.

Chapter 5

Application to real data

— *Es justamente la posibilidad de realizar un sueño*
lo que hace que la vida sea interesante.
Paulo Coelho

5.1 Prostate cancer data

There exists an increasingly interest in the literature, in finding good diagnostic tests that help in early detection of prostate cancer (PC) and avoid the need of undergoing a prostate biopsy. There are several studies in which, through ROC curves, the performance of different diagnostic tests was investigated (see for example Okihara *et al* (2002), Lein *et al* (2003) and Partin *et al* (2003)). These tests are based on some analytic measurements such as the total prostate specific antigen (tPSA), the free PSA (fPSA) or the complex PSA (cPSA).

Table 5.1: Descriptive statistics of tPSA, fPSA and cPSA, overall and per group.

Desc. stat.	tPSA	tPSA–	tPSA+	fPSA	fPSA–	fPSA+	cPSA	cPSA–	cPSA+
n	1432	833	599	1432	833	599	1432	833	599
min	0.10	1.69	0.10	0.002	0.010	0.002	0.096	0.550	0.096
max	66	63	66	34.59	33.40	34.59	56.70	56.70	53.30
range	65.90	61.31	65.90	34.5880	33.3900	34.5880	56.6040	56.1500	53.2040
mean	8.3311	7.2686	9.8085	1.4692	1.4064	1.5565	6.8556	5.8507	8.2531
median	6.50	5.98	7.60	1.0645	1.0640	1.0700	5.35	4.85	6.45
std	6.4929	5.1082	7.7987	1.8350	1.6001	2.1171	5.4504	4.2252	6.5489
iqr	4.8700	3.7450	6.0250	1.0000	0.8703	1.1494	4.0434	3.0022	5.4975
iqr/1.349	3.6101	2.7761	4.4663	0.7413	0.6451	0.8520	2.9973	2.2255	4.0752

The data consist of 599 patients suffering from PC (+) and 835 patients PC-free (–).

Table 5.2: Quantiles of tPSA, fPSA and cPSA, overall and per group.

Quantiles	tPSA	tPSA–	tPSA+	fPSA	fPSA–	fPSA+	cPSA	cPSA–	cPSA+
0.05	3.30	3.20	3.40	0.30	0.34	0.22	2.70	2.50	2.81
0.10	3.80	3.70	3.98	0.44	0.46	0.39	3.09	2.95	3.27
0.15	4.22	4.18	4.44	0.52	0.54	0.48	3.45	3.35	3.70
0.20	4.50	4.39	4.90	0.60	0.64	0.55	3.68	3.56	4.15
0.25	4.83	4.60	5.32	0.70	0.73	0.65	3.90	3.71	4.45
0.30	5.12	4.83	5.85	0.77	0.80	0.73	4.17	3.87	4.88
0.35	5.50	5.12	6.20	0.83	0.85	0.79	4.42	4.09	5.19
0.40	5.80	5.40	6.68	0.90	0.91	0.89	4.79	4.33	5.55
0.45	6.11	5.72	7.10	0.99	0.99	0.97	5.03	4.60	6.05
0.50	6.50	5.98	7.60	1.06	1.06	1.07	5.35	4.85	6.45
0.55	6.90	6.21	8.37	1.15	1.14	1.20	5.75	5.07	7.05
0.60	7.48	6.60	9.10	1.25	1.23	1.33	6.14	5.42	7.56
0.65	8.10	7.13	9.88	1.40	1.32	1.45	6.65	5.80	8.11
0.70	8.97	7.70	10.53	1.51	1.48	1.59	7.30	6.17	8.88
0.75	9.70	8.34	11.34	1.70	1.60	1.80	7.94	6.71	9.95
0.80	10.68	9.34	12.80	1.92	1.85	2.00	8.74	7.44	10.82
0.85	12.06	10.23	14.30	2.22	2.12	2.32	10.15	8.26	12.05
0.90	14.17	12.00	16.88	2.61	2.50	2.75	11.74	9.57	14.32
0.95	17.79	15.00	20.66	3.60	3.47	4.18	14.62	11.84	18.03
0.96	18.90	15.62	24.91	3.94	3.62	4.86	15.50	12.61	21.52
0.97	20.11	16.36	31.09	4.76	3.90	5.38	17.33	13.22	24.82
0.98	26.30	18.59	37.22	5.67	5.01	7.56	23.32	14.76	30.78
0.99	38.18	24.11	49.53	8.42	6.33	9.82	31.61	19.36	42.72

For each patient the illness status has been determined through a prostate biopsy carried out for the first time in Hospital Juan Canalejo (Galicia, Spain) between January 2002 and September 2005. Values of tPSA, fPSA and cPSA were measured for each patient.

Two values of tPSA were extremely large compared to the rest of values that fell in the interval [0.10, 66]. These two outliers, 120 and 4002, registered in the PC-free group, were discarded from the analysis. After a more detailed look at the data, we observed that only 4.7 and 1.4 percent of the individuals in the PC+ group and PC– group, respectively, registered a value of tPSA larger than 22. In Table 5.1 we show some descriptive statistics of the data and in Table 5.2 we collect some quantiles.

Figures 5.1, 5.3 and 5.5 show the empirical distribution functions of the three variables of interest, overall and per group and Figures 5.2, 5.4 and 5.6 show the smooth estimates of their corresponding densities. For the computation of these estimates we have used the Parzen-Rosenblatt estimator introduced in (1.9) with Gaussian kernel K and bandwidths $\tilde{h}$ selected by the rule of thumb which uses a Gaussian parametric reference for computing the unknown quantities appearing in the expression of the AMISE optimal bandwidth in this setting, i.e.

$$\tilde{h} \;=\; \left(\frac{R(K)}{d_K^2 R(f^{(2)})}\right)^{\frac{1}{5}} n^{-\frac{1}{5}} = \left(\frac{(2\hat{\sigma})^5}{4!}\right)^{\frac{1}{5}} n^{-\frac{1}{5}},$$

where $\hat{\sigma}$ refers to the minimum between the standard deviation estimate and the interquartile range estimate divided by 1.349, shown in Table 5.1.

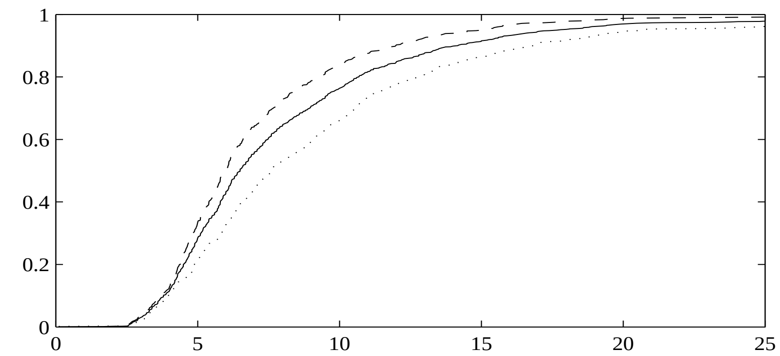

Figure 5.1: Empirical estimate of the distribution functions of tPSA for both groups, F, for PC– group, F_0, and for PC+ group, F_1: F_n (solid line), F_{0n_0} (dashed line) and F_{1n_1} (dotted line), respectively.

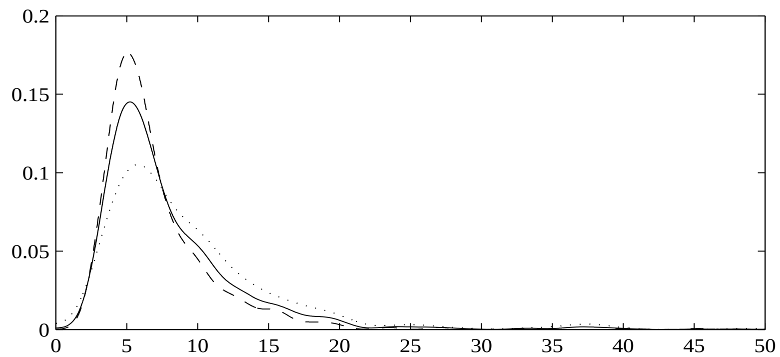

Figure 5.2: Smooth estimates of f (density of tPSA), f_0 (density of tPSA in the PC– group) and f_1 (density of tPSA in the PC+ group) by using, respectively, $\tilde{f}_h$ with $h = 0.8940$ (solid line), $\tilde{f}_{0h_0}$ with $h_0 = 0.7661$ (dashed line) and $\tilde{f}_{1h_1}$ with $h_1 = 1.3166$ (dotted line).

We compare, from a distributional point of view, the above mentioned measurements (tPSA, fPSA and cPSA) between the two groups in the data set (PC+ and PC–). To this end we compute and plot the empirical estimate of the relative distribution (see (1.31)) of each one of these measurements in the PC+ group wrt to its values in the PC– group (see Figure 5.7). We compute as well the appropriate bandwidths to estimate the corresponding relative densities using four of the data-driven bandwidth selectors proposed in Chapter

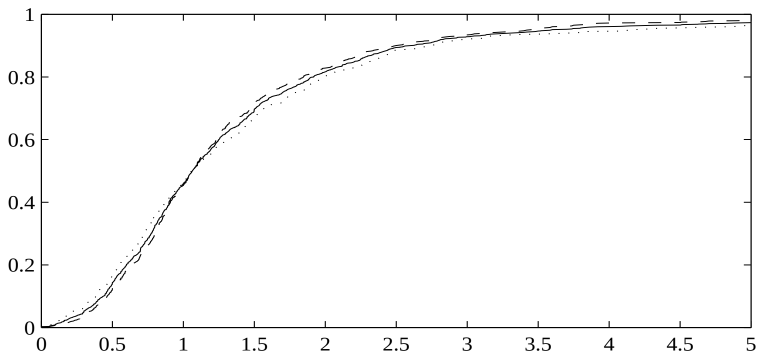

Figure 5.3: Empirical estimate of the distribution functions of fPSA for both groups, F, for PC– group, F_0 and for PC+ group F_1: F_n (solid line), F_{0n_0} (dashed line) and F_{1n_1} (dotted line), respectively.

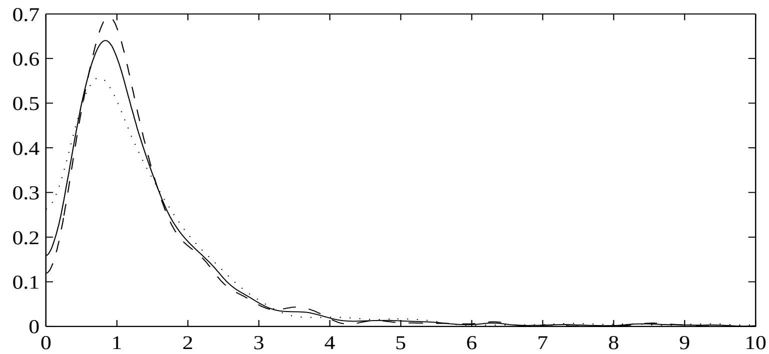

Figure 5.4: Smooth estimates of f (density of fPSA), f_0 (density of fPSA in the PC– group) and f_1 (density of fPSA in the PC+ group) by using, respectively, $\tilde{f}_h$ with $h = 0.1836$ (solid line), $\tilde{f}_{0h_0}$ with $h_0 = 0.1780$ (dashed line) and $\tilde{f}_{1h_1}$ with $h_1 = 0.2512$ (dotted line).

2, h_{SJ_1}, h_{SJ_2}, h^*_{SUMC}, h^*_{SMC}, and the classical one, proposed by Ćwik and Mielniczuk (1993) (see Table 5.3). While in the definition of h^*_{SUMC} and h^*_{SMC} a number of $N = 2n_1$ beta distributions is used in the parametric fit considered there, $\hat{b}(x; N, R)$, we now use a number of $N = 14$ betas. For each one of the measurements of interest, the value of h given by the selector of best performance, h_{SJ_2}, is used and the corresponding relative density estimate is computed using (2.1). These estimates are shown in Figure 5.8.

It is clear from Figure 5.8 that the relative density estimate is above one in the upper interval accounting for a probability of about 30% of the PC– distribution for the variables tPSA and cPSA. In the case of fPSA the 25% left tail of the PC– group and an interval in the upper tail that starts approximately at the quantile 0.7 of the PC– group, show as well that the relative density estimate is slightly above one. However, this effect is less remarkable that in the case of the variables tPSA and cPSA.

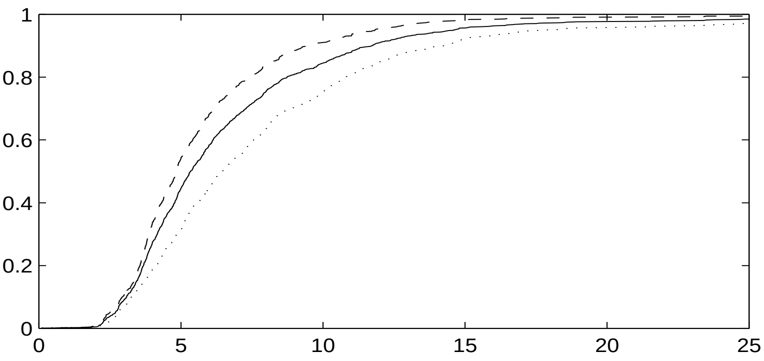

Figure 5.5: Empirical estimate of the distribution functions of cPSA for both groups, F, for PC– group, F_0 and for PC+ group F_1: F_n (solid line), F_{0n_0} (dashed line) and F_{1n_1} (dotted line), respectively.

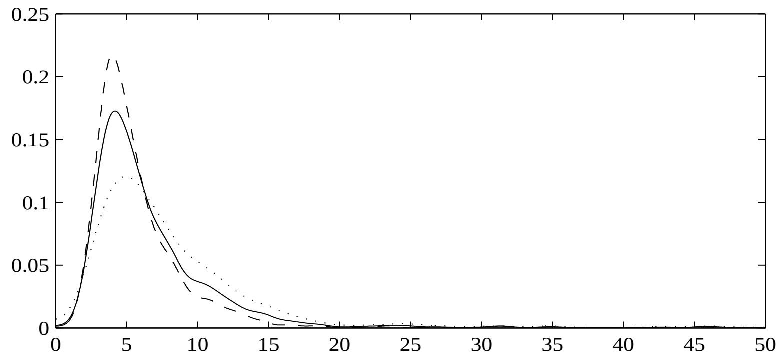

Figure 5.6: Smooth estimates of f (density of cPSA), f_0 (density of cPSA in the PC– group) and f_1 (density of cPSA in the PC+ group) by using, respectively, $\tilde{f}_h$ with $h = 0.7422$ (solid line), $\tilde{f}_{0h_0}$ with $h_0 = 0.6142$ (dashed line) and $\tilde{f}_{1h_1}$ with $h_1 = 1.2013$ (dotted line).

The two-sided Kolmogorov-Smirnov two-sample test, D_{n_0,n_1}, introduced in Subsection 1.4.1, has been computed for each one of the three variables of interest for testing the null hypothesis of equal distribution functions in the two groups (PC+ patients and PC– patients). For the variables tPSA, fPSA and cPSA, we obtained respectively, $D_{n_0,n_1} = 0.2132$ (p-value$= 2.3135 \cdot 10^{-14}$), $D_{n_0,n_1} = 0.0629$ (p-value$= 0.1219$) and $D_{n_0,n_1} = 0.2398$ (p-value$= 4.5503 \cdot 10^{-18}$). Based on the two-sided Kolmogorov-Smirnov two-sample tests, the only variable for which the null hypothesis cannot be rejected is fPSA. In fact, this was already expected from Figures 5.1, 5.3, 5.5, 5.7 and 5.8.

As it was mentioned in Chapter 3, Corollary 3.1.6 generalizes to LTRC data, the asymptotic result given by Handcock and Janssen (2002) for complete data. Based on this result, confidence intervals of $\hat{r}_h(t)$ can be obtained. Let c be a real number such that $n_1 h^5 \to c$. Then, it is satisfied that $\sqrt{n_1 h}\,(\hat{r}_h(t) - r(t))$ converges asymptotically to

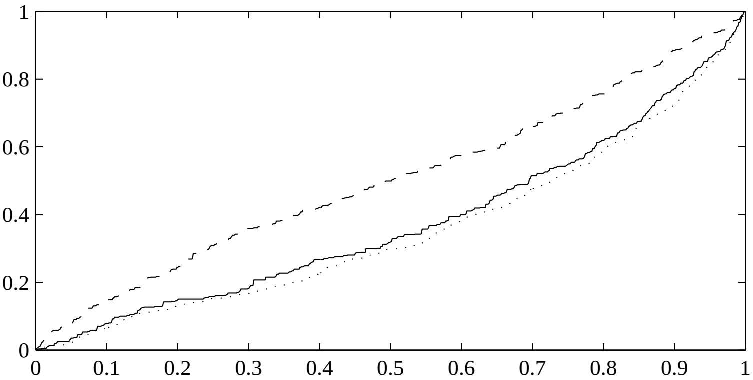

Figure 5.7: Relative distribution estimate, $R_{n_0,n_1}(t)$, of the PC+ group wrt the PC– group for the variables tPSA (solid line), fPSA (dashed line) and cPSA (dotted line).

Table 5.3: Bandwidths b_{3c}, h_{SJ_1}, h_{SJ_2}, h^*_{SUMC} and h^*_{SMC} selected for estimating the relative density of tPSA+ wrt tPSA–, the relative density of fPSA+ wrt fPSA– and the relative density of cPSA+ wrt cPSA–. The symbol (G) means Gaussian kernel and (E) Epanechnikov kernel.

Variables	b_{3c}	h_{SJ_1}	h_{SJ_2}	h^*_{SUMC}	h^*_{SMC}
tPSA+ wrt tPSA–	0.0923 (E)	0.0596 (E)	0.0645 (E)	0.0879 (G)	0.0837 (G)
fPSA+ wrt fPSA–	0.1259 (E)	0.0941 (E)	0.1067 (E)	0.1230 (G)	0.1203 (G)
cPSA+ wrt cPSA–	0.0822 (E)	0.0588 (E)	0.0625 (E)	0.0765 (G)	0.0739 (G)

a normal distribution with mean and variance given by, respectively,

$$\begin{aligned}\mu(t) &= \frac{1}{2}r^{(2)}(t)d_K c^{1/2},\\ \sigma^2(t) &= R(K)\left\{r(t)+\kappa^2 r^2(t)\right\}.\end{aligned}$$

Consequently, the following interval

$$\left(\hat{r}_h(t)-\frac{\mu(t)}{\sqrt{n_1h}}-\frac{q_{0.975}\sigma(t)}{\sqrt{n_1h}},\hat{r}_h(t)-\frac{\mu(t)}{\sqrt{n_1h}}-\frac{q_{0.025}\sigma(t)}{\sqrt{n_1h}}\right)$$

has an asymptotic coverage probability of 0.95. Here $q_{0.025}$ and $q_{0.975}$ are the 0.025 and 0.975 quantiles of a standard normal density. However, this interval can not be used directly because it depends on some unknown quantities. Therefore, we propose to use:

$$\left(\hat{r}_{h_{SJ_2}}(t)-\frac{\hat{\mu}(t)}{\sqrt{n_1h_{SJ_2}}}-\frac{q_{0.975}\hat{\sigma}(t)}{\sqrt{n_1h_{SJ_2}}},\hat{r}_{h_{SJ_2}}(t)-\frac{\hat{\mu}(t)}{\sqrt{n_1h_{SJ_2}}}-\frac{q_{0.025}\hat{\sigma}(t)}{\sqrt{n_1h_{SJ_2}}}\right),$$

where

$$\begin{aligned}\hat{\mu}(t) &= \frac{1}{2}\hat{r}^{(2)}_{h_1}(t)d_K(n_1h^5_{SJ_2})^{1/2},\\ \hat{\sigma}(t) &= R(K)\left\{\hat{r}_{h_{SJ_2}}(t)+\frac{n_1}{n_0}\hat{r}^2_{h_{SJ_2}}(t)\right\},\end{aligned}$$

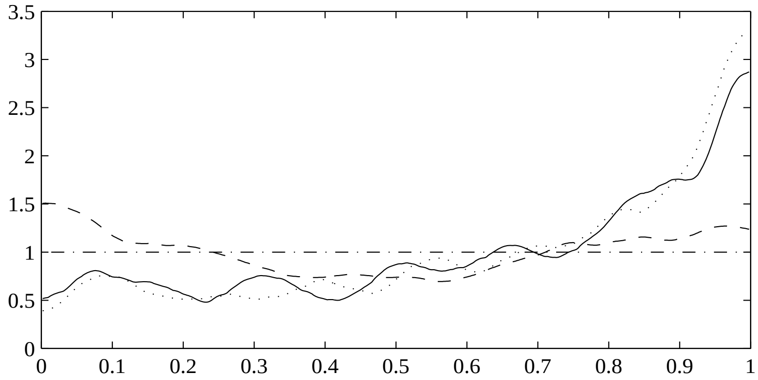

Figure 5.8: Relative density estimate of the PC+ group wrt the PC– group for the variables tPSA (solid line, $h_{SJ_2} = 0.0645$), cPSA (dotted line, $h_{SJ_2} = 0.0625$) and fPSA (dashed line, $h_{SJ_2} = 0.1067$).

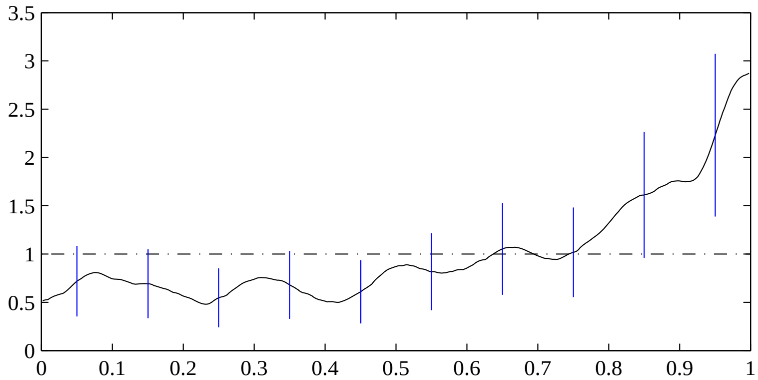

Figure 5.9: Two-sided simultaneous α confidence intervals (with $\alpha = 0.05$) for the relative density estimate of the tPSA in PC+ group wrt PC– group.

K is the Epanechnikov kernel and

$$\hat{r}_{h_1}^{(2)}(t) \quad = \quad \frac{1}{n_1 h_1^3} \sum_{j=1}^{n_1} L^{(2)} \left(\frac{t - \tilde{F}_{0h_0}(Y_{1j})}{h_1} \right),$$

where L denotes the Gaussian kernel, h_0 is selected as explained in Section 2.3.2 and $h_1 = 2 \cdot h_{SUMC}^*$. This subjective choice for h_1 is motivated by the fact that the optimal bandwidth for estimating $r^{(2)}$ is asymptotically larger than that for estimating r.

To test, at an α-level of 0.05, if the relative density is the function constantly one, we take ten equispaced points in $[0, 1]$, from 0.5 to 0.95, and compute, using Bonferroni correction, for each point a confidence interval with confidence level $1 - \frac{\alpha}{10}$. The relative density estimates previously computed for the interest variables, tPSA, fPSA and cPSA, and their corresponding confidence intervals are jointly plotted in Figures 5.9–5.11.

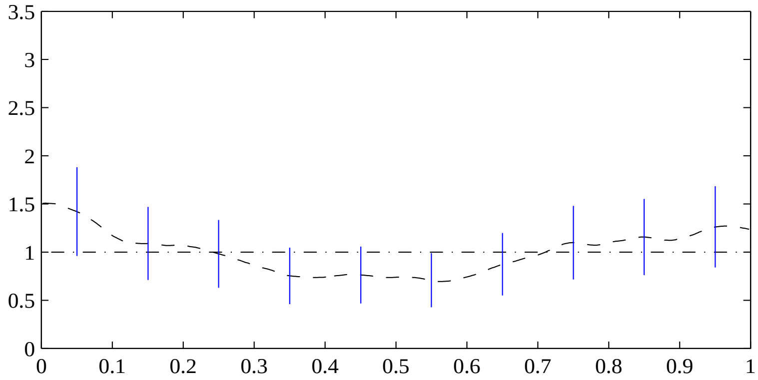

Figure 5.10: Two-sided simultaneous α confidence intervals (with $\alpha = 0.05$) for the relative density estimate of the fPSA in PC+ group wrt PC– group.

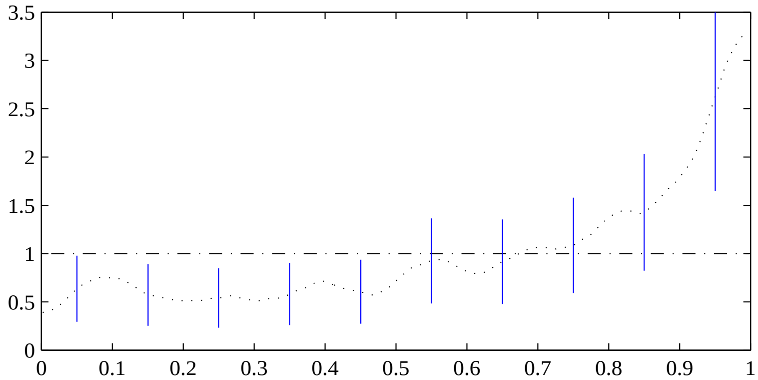

Figure 5.11: Two-sided simultaneous α confidence intervals (with $\alpha = 0.05$) for the relative density estimate of the cPSA in PC+ group wrt PC– group.

5.2 Gastric cancer data

A data set consisting of $n = 1956$ patients with gastric cancer has been collected in Hospital Juan Canalejo and Hospital Xeral-Calde (Galicia, Spain) during the period 1975-1993. A large number of variables have been registered for patients under study. Among them we will focus on age and metastasis status (+ or −) at diagnosis, sex, elapsed time from first symptoms to diagnosis (T) and time from first symptoms to death or loss of follow up (Y). This possible loss of follow up may cause censoring of the interest lifetime, time from first symptoms to death (X). On the other hand, truncation occurs when a patient is not diagnosed before death. In Figure 5.12 we have plotted the lifetimes and the truncation times registered for nine patients in the data set. It is interesting to note that three of these patients presented censored lifetimes. Since this implies that their real lifetimes

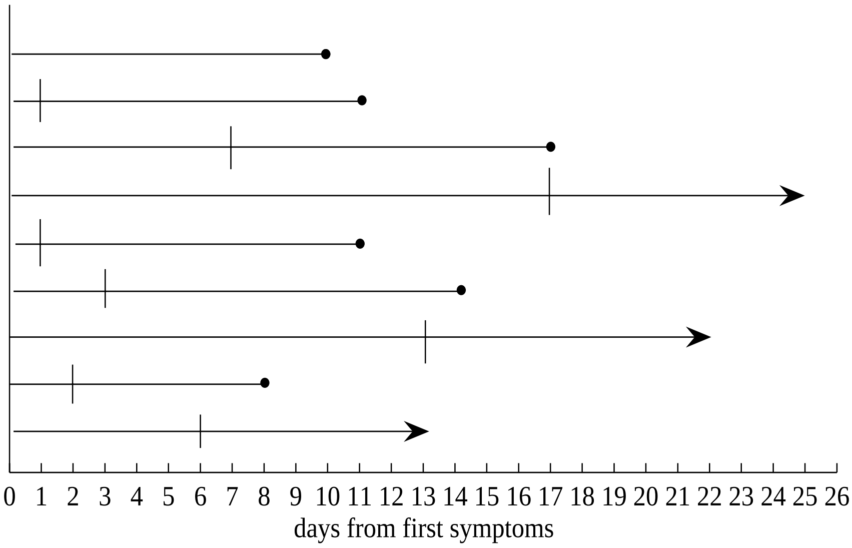

Figure 5.12: Truncation times (|) and lifetimes of 9 individuals in the gastric data set. Uncensored lifetimes are indicated by • while censored lifetimes are indicated by arrows.

would be larger than the observed ones, we use right arrows for the censored lifetimes.

Table 5.4: Descriptive statistics of survival overall, per sex, age and metastasis.

Desc. stat.	Y	men	women	age<70	age ≥ 70	met.–	met.+
n	1956	1228	728	1076	880	1443	513
cens (%)	22.90	21.66	25.00	25.74	19.43	25.78	14.81
min	1	8	1	10	1	1	10
max	14639	14639	8688	14639	8688	14639	6348
range	14638	14631	8687	14629	8687	14638	6338
mean	885.3269	835.7176	787.2606	1055.6350	690.4801	1171.5466	227.7661
median	208	209	202	281	139	286	111
std	1864.3241	1787.3590	1626.5015	2077.3549	1502.5734	2150.3750	522.1906
iqr	540	525	563	717	371	939	183
iqr/1.349	400.2965	389.1772	417.3462	531.5048	275.0185	696.0712	135.6560

Some descriptive statistics of the data are collected in Table 5.4. The means, medians, standard deviations and interquartile ranges, were computed using the TJW product-limit estimates of the variable X, based on the three-dimensional vectors, (T, Y, δ), registered for all the patients in the data set and for each one of the groups of interest, women versus men, metastasis– group versus metastasis+ group and patients older than 70 versus patients younger than 70. For example, consider the lifetime Y measured in all the patients and let $\hat{F}_n$ be the TJW estimate of its distribution function, F. Then, the mean, μ, the standard deviation, σ, the interquartile range, iqr, and the median, $q_{0.5}$, are estimated by means of:

$$\hat{\mu} = \sum_{k=1}^{n} Y(k) \left[\hat{F}_n(Y(k)) - \hat{F}_n(Y_k^-)\right],$$

$$\hat{\sigma} = \sqrt{\sum_{k=1}^{n} Y(k)^2 \left[\hat{F}_n(Y(k)) - \hat{F}_n(Y_k^-)\right] - \hat{\mu}^2},$$

$$\widehat{iqr} = \hat{F}_n^{-1}(0.75) - \hat{F}_n^{-1}(0.25),$$

$$\hat{q}_{0.5} = \hat{F}_n^{-1}(0.5).$$

The list of quantiles θ collected in Table 5.5 are computed in a similar way as $\hat{q}_{0.5}$, by means of $\hat{q}_\theta = \hat{F}_n^{-1}(\theta)$.

Table 5.5: Quantiles of survival overall, per sex, age and metastasis.

Quantiles	survival	men	women	age<70	age≥ 70	metastasis–	metastasis+
0.05	11	11	1	22	8	11	1
0.10	25	22	31	35	11	26	19
0.15	38	34	41	57	26	42	34
0.20	58	51	59	73	41	61	42
0.25	69	64	74	102	55	79	55
0.30	89	87	89	127	66	109	59
0.35	110	109	111	158	79	137	72
0.40	133	128	137	194	99	173	85
0.45	161	157	164	239	113	238	100
0.50	208	209	202	281	139	286	111
0.55	249	250	241	332	167	361	128
0.60	298	292	307	413	223	471	147
0.65	375	362	380	522	266	574	173
0.70	473	454	495	640	330	754	202
0.75	609	589	637	819	426	1018	238
0.80	848	810	920	1208	564	1851	279
0.85	1601	1430	2019	2810	960	3364	329
0.90	3957	3331	4585	5710	2059	6019	424
0.95	7379	6833	7487	9003	5233	9003	660
0.96	7487	7379	–	9003	6019	9003	766
0.97	9003	9003	–	–	6256	9003	997
0.98	9003	9003	–	–	6879	–	2137
0.99	–	–	–	–	7379	–	5011

Figures 5.13, 5.15 and 5.17 show density estimates of the lifetime registered in all the patients, per sex, age (< 70 vs ≥ 70) and presence of metastasis (yes (+) or no (–)). For the computation of these estimates we have used the kernel type estimator introduced in (1.25) with Gaussian kernel K and bandwidths $\tilde{h}$ selected by the rule of thumb

$$\tilde{h} = \left(\frac{R(K) \int \sigma^2(t) w(t) dt}{d_K^2 R(f^{G(2)} w^{1/2})} \right)^{\frac{1}{5}} n^{-\frac{1}{5}},$$

where

$$\begin{aligned}\sigma^2(t) &= \frac{f^G(t)(1-\hat{F}_n(t))}{\tilde{B}_g(t)},\\ \tilde{B}_g(t) &= n^{-1}\sum_{k=1}^{n}\mathbb{L}\left(\frac{t-T_k}{g}\right)\mathbb{L}\left(\frac{Y_k-t}{g}\right),\\ w(t) &= \mathbb{L}\left(\frac{t-q_{0.05}}{0.025}\right)\mathbb{L}\left(\frac{q_{0.95}-t}{0.025}\right),\end{aligned}$$

$g = 0.10$, $\mathbb{L}$ is the cdf of the biweight kernel, $\hat{F}_n(t)$ is the TJW estimate of F and f^G denotes a Gaussian density with mean and standard deviation given by the estimates shown in Table 5.4.

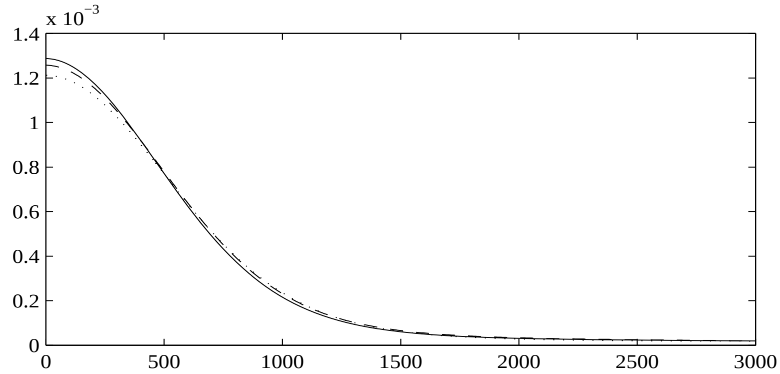

Figure 5.13: Smooth estimates of f (density of the lifetime), f_0 (density of lifetime in men) and f_1 (density of lifetime in women) by using, respectively, $\tilde{f}_h$ with $h = 413.3244$ (solid line), $\tilde{f}_{0h_0}$ with $h_0 = 431.9383$ (dashed line) and $\tilde{f}_{1h_1}$ with $h_1 = 437.8938$ (dotted line).

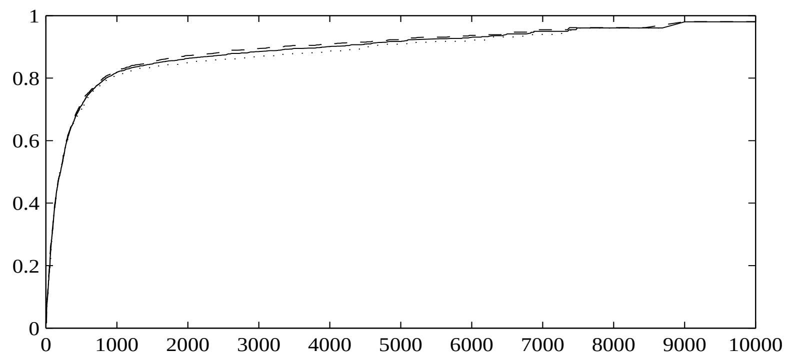

Figure 5.14: Product limit estimate of the distribution functions of survival overall, F, for men, F_0, and for women, F_1: $\hat{F}_n$ (solid line), $\hat{F}_{0n_0}$ (dashed line) and $\hat{F}_{1n_1}$ (dotted line), respectively.

In Figures 5.14, 5.16 and 5.18, the TJW estimates of the overall lifetime and the conditional lifetime, given the covariates, are plotted. Figure 5.19 show estimates of the relative

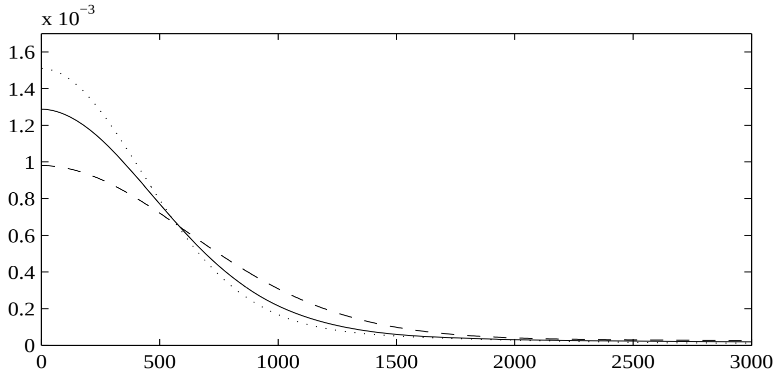

Figure 5.15: Smooth estimates of f (density of the lifetime), f_0 (density of lifetime for people aged$<$ 70) and f_1 (density of lifetime in people aged $\geq$ 70) by using, respectively, $\tilde{f}_h$ with $h = 413.3244$ (solid line), $\tilde{f}_{0h_0}$ with $h_0 = 533.1826$ (dashed line) and $\tilde{f}_{1h_1}$ with $h_1 = 379.0568$ (dotted line).

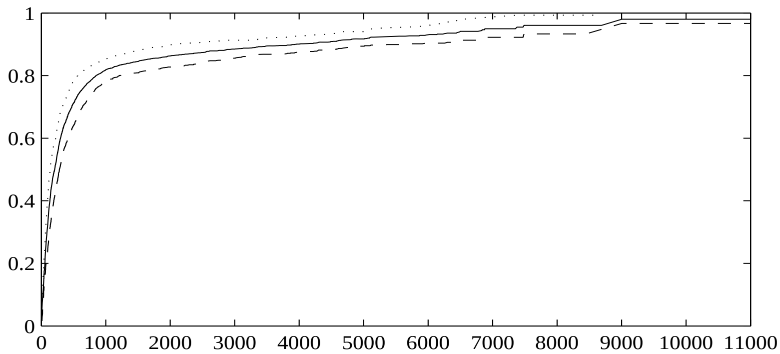

Figure 5.16: Product limit estimate of the distribution functions of survival for the whole sample, F, for people aged$<$ 70, F_0, and for people aged$\geq$ 70, F_1: $\hat{F}_n$ (solid line), $\hat{F}_{0n_0}$ (dashed line) and $\hat{F}_{1n_1}$ (dotted line), respectively.

lifetime distribution for women wrt men, for patients older than 70 (80) wrt younger patients, and for patients with presence of metastasis wrt patients without metastasis. For the computation of these estimates we have used a natural extension to LTRC data, $\check{R}_{n_0,n_1}$, of the estimator introduced in (1.31) for complete data:

$$\check{R}_{n_0,n_1}(t) = \hat{F}_{1n_1}\left(\hat{F}_{0n_0}^{-1}(t)\right),$$

where now, the role of the empirical estimates of F_0 and F_1 are replaced by the TJW estimates of F_0 and F_1, respectively.

The estimator presented in (3.1) has been applied to this problem. With the purpose of correcting its boundary effect, the well-known reflection method has been applied here. Kernel relative density estimations have been computed for the interest variable under

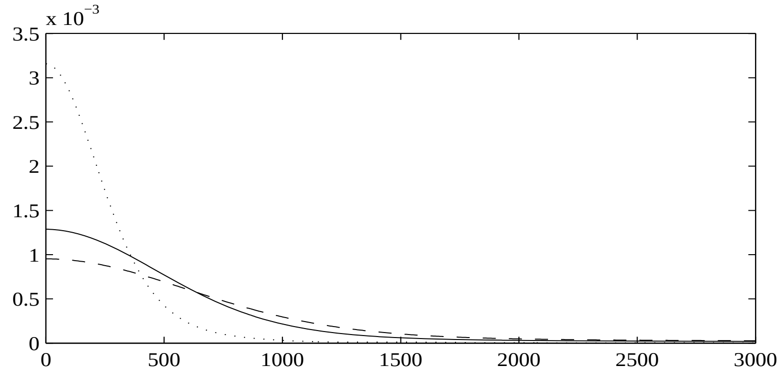

Figure 5.17: Smooth estimates of f (density of the lifetime), f_0 (density of lifetime in metastasis– group) and f_1 (density of lifetime in metastasis+ group) by using, respectively, $\tilde{f}_h$ with $h = 413.3244$ (solid line), $\tilde{f}_{0h_0}$ with $h_0 = 528.4267$ (dashed line) and $\tilde{f}_{1h_1}$ with $h_1 = 159.4171$ (dotted line).

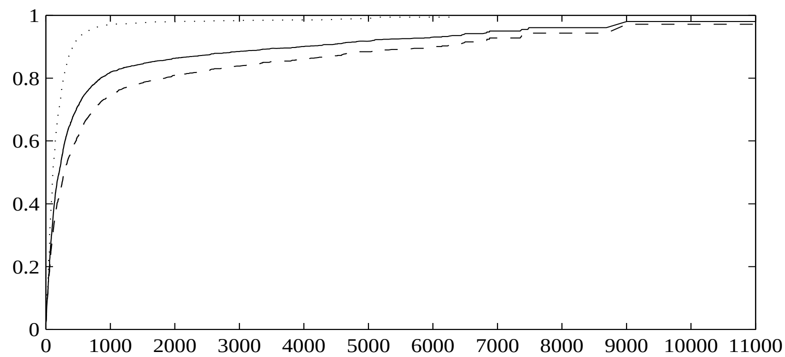

Figure 5.18: Product limit estimate of the distribution functions of survival for the whole sample, F, for metastasis– group, F_0 and for metastasis+ group F_1: $\hat{F}_n$ (solid line), $\hat{F}_{0n_0}$ (dashed line) and $\hat{F}_{1n_1}$ (dotted line), respectively.

the different choices for the two groups mentioned above: for women with respect to men, for the metastasis+ group with respect to the metastasis– group and for patients older than 80 (70) with respect to patients younger than 80 (70). Figure 5.20 collects the four estimations using the smoothing parameters obtained using the bandwidth selector h_{RT} introduced in Chapter 3. Table 5.6 collects these bandwidths as well as bandwidths h_{PI} and h_{STE}. As it is clear from Figure 5.20, sex does not affect the lifetime distribution, while the presence of metastasis and the age do. More specifically, those patients with metastasis tend to present smaller lifetimes than those without metastasis. Similarly, patients older than 80 have about double probability of presenting lifetimes around 10% of the lifetime distribution of those under 80 years old. A similar tendency is exhibited when using the cutoff age of 70. Now, the estimation is a little closer to a uniform density.

Table 5.6: Bandwidths h_{RT}, h_{PI} and h_{STE} selected for estimating the relative density of X in women wrt men, the relative density of X in patients older than 70 (80) wrt younger patients, the relative density of X in the metastasis+ group wrt the metastasis– group. The symbol (G) means Gaussian kernel.

Variables	h_{RT}	h_{PI}	h_{STE}
women wrt men	0.0934 (G)	0.0953 (G)	0.0952 (G)
aged<70 wrt aged≥ 70	0.0888 (G)	0.0860 (G)	0.0860 (G)
aged<80 wrt aged≥ 80	0.0879 (G)	0.0965 (G)	0.0965 (G)
metastasis+ wrt metastasis–	0.0966 (G)	0.0892 (G)	0.0892 (G)

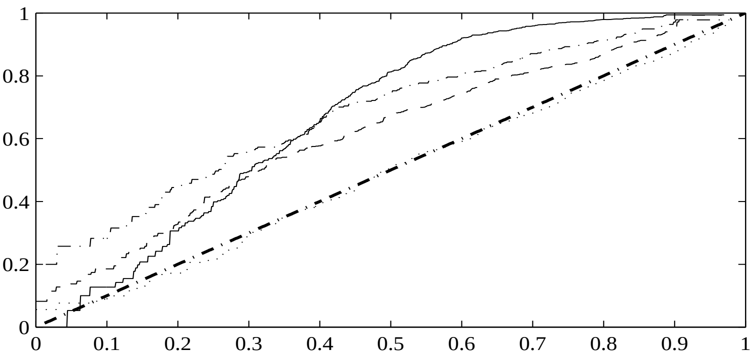

Figure 5.19: Relative distribution estimate, $\check{R}_{n_0,n_1}(t)$, of the the survival time for women wrt men (dotted line), for the metastasis+ group wrt the metastasis– group (solid line), for the group of patients aged≥ 80 wrt the group of patients aged<80 (dashed-dotted line) and for the group of patients aged≥ 70 wrt the group of patients aged<70 (dashed line).

Confidence intervals of $\check{r}_h(t)$ can be obtained using Corollary 3.1.6. Let c be a real number such that $n_1 h^5 \to c$. Then, it is satisfied that $\sqrt{n_1 h}\,(\check{r}_h(t) - r(t))$ converges asymptotically to a normal distribution with mean and variance given by, respectively,

$$\begin{aligned} \mu(t) &= \frac{1}{2} r^{(2)}(t) d_K c^{1/2}, \\ \sigma^2(t) &= R(K) \left\{ \frac{r(t)(1-R(t))}{B_1\left(F_0^{-1}(t)\right)} + \kappa^2 \frac{(1-t) r^2(t)}{B_0\left(F_0^{-1}(t)\right)} \right\}. \end{aligned}$$

Consequently, taking the 0.025 and 0.975 quantiles of a standard normal density, $q_{0.025}$ and $q_{0.975}$, the interval given below

$$\left(\check{r}_h(t) - \frac{\mu(t)}{\sqrt{n_1 h}} - \frac{q_{0.975}\sigma(t)}{\sqrt{n_1 h}}, \check{r}_h(t) - \frac{\mu(t)}{\sqrt{n_1 h}} - \frac{q_{0.025}\sigma(t)}{\sqrt{n_1 h}} \right)$$

has an asymptotic coverage probability of 0.95. However, this interval can not be used directly because it depends on some unknown quantities. Therefore, we propose to use:

$$\left(\check{r}_{h_{RT}}(t) - \frac{\check{\mu}(t)}{\sqrt{n_1 h_{RT}}} - \frac{q_{0.975}\check{\sigma}(t)}{\sqrt{n_1 h_{RT}}}, \check{r}_{h_{RT}}(t) - \frac{\check{\mu}(t)}{\sqrt{n_1 h_{RT}}} - \frac{q_{0.025}\check{\sigma}(t)}{\sqrt{n_1 h_{RT}}} \right),$$

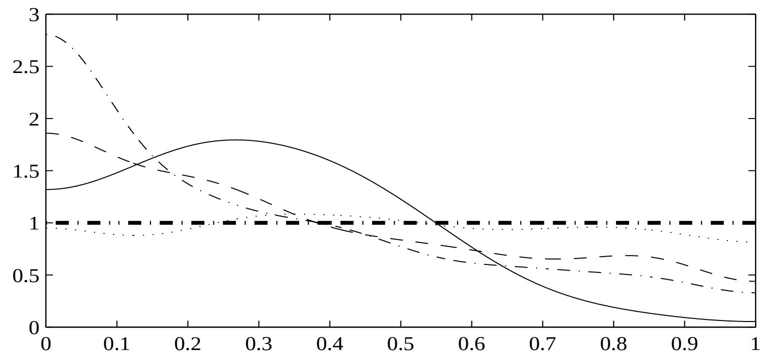

Figure 5.20: Relative density estimate of the survival time for women wrt men (dotted line, $h_{RT} = 0.0934$), for the metastasis+ group wrt the metastasis– group (solid line, $h_{RT} = 0.0966$), for the group of patients aged≥ 80 wrt the group of patients aged< 80 (dashed-dotted line, $h_{RT} = 0.0879$) and for the group of patients aged≥ 70 wrt the group of patients aged< 70 (dashed line, $h_{RT} = 0.0888$).

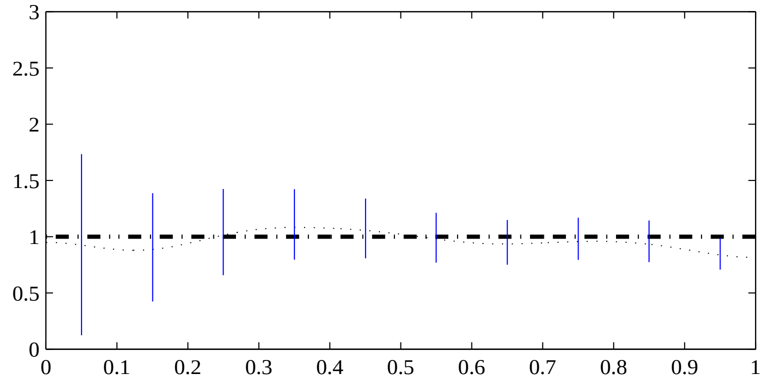

Figure 5.21: Two-sided simultaneous α confidence intervals (with $\alpha = 0.05$) for the relative density estimate of the lifetime in women wrt to men.

where

$$\check{\mu}(t) = \frac{1}{2}\check{r}_{h_1}^{(2)}(t) d_K (n_1 h_{RT}^5)^{1/2},$$

$$\check{\sigma}(t) = R(K)\left\{\frac{\check{r}_{h_{RT}}(t)(1-\check{R}_{n_0,n_1}(t))}{\tilde{B}_{1g_1}(\hat{F}_0^{-1}(t))} + \frac{n_1}{n_0}\frac{(1-t)\check{r}_{h_{RT}}^2(t)}{\tilde{B}_{0g_0}(\hat{F}_0^{-1}(t))}\right\}$$

with the subjective choices $g_0 = 0.05$ and $g_1 = 0.05$, K is the Gaussian kernel and

$$\check{r}_{h_1}^{(2)}(t) = \frac{1}{h_1^3}\sum_{j=1}^{n_1} K^{(2)}\left(\frac{t-\hat{F}_{0n_0}(Y_{1j})}{h_1}\right)\left\{\hat{F}_{1n_1}(Y_{1j}) - \hat{F}_{1n_1}(Y_{1j}^-)\right\},$$

where $h_1 = 2h_{RT}$. This subjective choice for h_1 is motivated by the fact that the optimal bandwidth for estimating $r^{(2)}$ is asymptotically larger than that for estimating r.

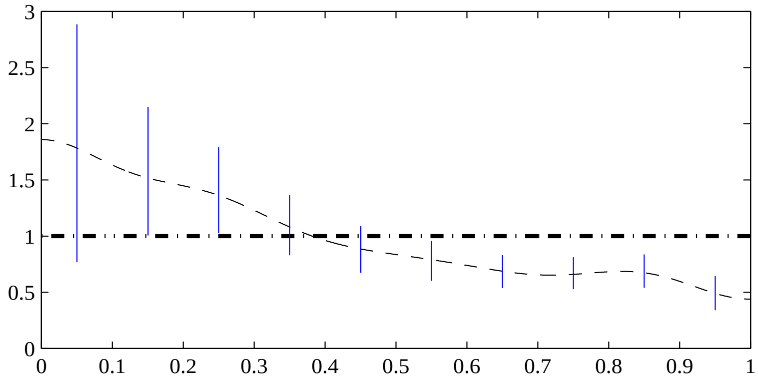

Figure 5.22: Two-sided simultaneous α confidence intervals (with $\alpha = 0.05$) for the relative density estimate of the lifetime in the patients aged≥ 70 wrt to the patients aged< 70.

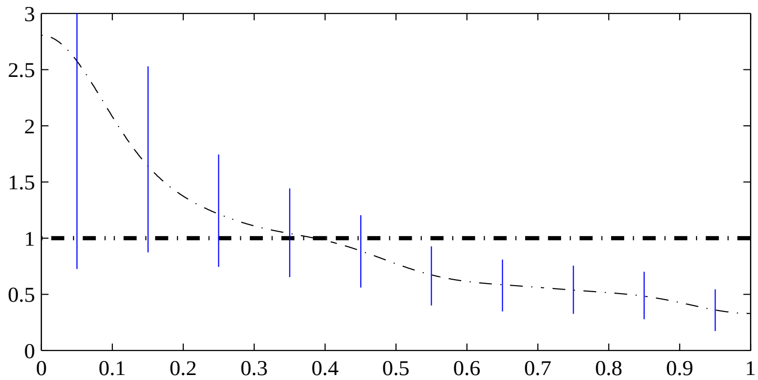

Figure 5.23: Two-sided simultaneous α confidence intervals (with $\alpha = 0.05$) for the relative density estimate of the lifetime in the patients aged≥ 80 wrt to the patients aged< 80.

To test, at an α-level of 0.05, if the relative density is the function constantly one, we take ten equispaced points in $[0, 1]$, from 0.5 to 0.95, and compute, using Bonferroni correction, for each point a confidence interval with confidence level $1 - \frac{\alpha}{10}$. The relative density estimates previously computed for the lifetime observed in the interest pair of groups and their corresponding confidence intervals are jointly plotted in Figures 5.21–5.24.

It is interesting to note here that we have assumed that truncation time, censoring time and survival time are independent of each other. As Tsai (1990) remarks, while for data subject to random censoring, it is well known that the assumption of independence between censoring time and survival time cannot be tested nonparametrically, for data subject to random truncation, however, it is possible to test independence for truncation time and survival time. We have computed the statistic K^* proposed by Tsai (1990) for

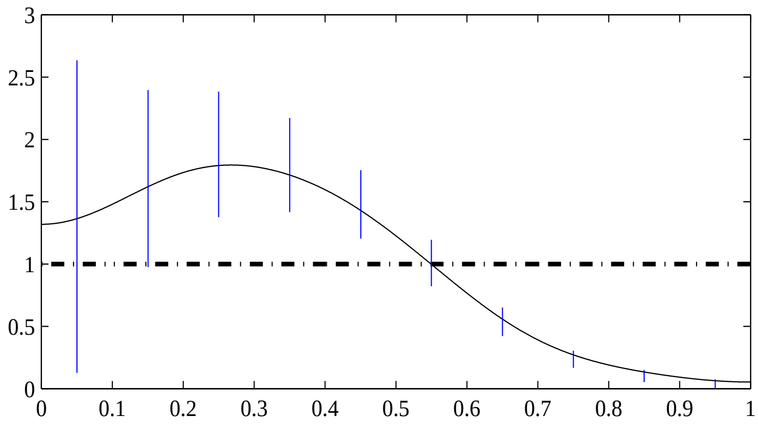

Figure 5.24: Two-sided simultaneous α confidence intervals (with $\alpha = 0.05$) for the relative density estimate of the lifetime in the metastasis+ group wrt the metastasis– group.

testing this independence assumption and which its an extension of Kendall's tau to LTRC data. As Tsai (1990) proves, under the null hypothesis of independence and the following assumption

$$P(Y \leq y, C \leq c/T = t) = P(Y \leq y/T = t) P(C \leq c/T = t),$$

it follows that

$$T^* = \frac{K^*}{\left\{\frac{1}{3}\sum_{k=1}^m (r_k^2 - 1)\right\}^{1/2}}$$

tends to the standard Gaussian distribution, where

$$\begin{aligned} K^* &= \sum_{k=1}^m S_{(k)}, \quad S_{(k)} = \sum_{j \in R(k)} sgn(T_j - T_{(k)}), \\ R(k) &= \left\{ j : T_j \leq Y_{(k)} \leq Y_j \right\}, \\ r_k &= card(R(k)), \quad card(R(k)) = \sum_{j=1}^m 1 \left\{ T_j \leq Y_{(k)} \leq Y_j \right\}, \end{aligned}$$

$Y_{(1)} < \cdots < Y_{(m)}$ denote the distinct observed survival times and $T_{(i)}$ is the concomitant of $Y_{(i)}$ for $i = 1, \ldots, m$. Therefore, a test of the null hypothesis can be carried out by comparing T^* with the standard Gaussian distribution.

We have computed the statistic T^* using all the survival times registered in the dataset and as well only the survival times observed in each one of the two groups of interest. Unfortunately, all the values obtained are indicative of non independence between the truncation variable and the survival time Y. For example, $T^* = -9.6031$ when computed for all the dataset, $T^* = -5.3213$ for women, $T^* = -7.0425$ for men, $T^* = -2.2930$ for the metastasis+ group, $T^* = -6.7123$ for the metastasis– group, $T^* = -4.9478$ for people aged< 70, $T^* = -7.6832$ for people aged≥ 70, $T^* = -8.4193$ for people aged< 80 and $T^* = -3.9893$ for people aged≥ 80. Therefore, the results obtained from the analysis carried out here should be taken carefully. Certainly, a substantive analysis of these data would require to consider the dependence existing between the truncation variable, T, and the time of interest, Y.

Chapter 6

Further research

— Escoger un camino significa abandonar otros.
Si pretendes recorrer todos los caminos posibles
acabarás no recorriendo ninguno.
Paulo Coelho

This thesis has also raised new and interesting questions that we detail below.

6.1 Reduced-bias techniques in two-sample problems

There exist in the literature many proposals to reduce the bias of the Parzen-Rosenblatt estimator, $\tilde{f}_{0h}(t)$, introduced in (1.9), and several proposals to improve the bias property of the smoothed empirical distribution function, $\tilde{F}_{0h}(t)$, introduced in (1.22). However, in the setting of a two-sample problem, we are not aware of any work that deals with the problem of reducing the bias of $\hat{r}_h(t)$ or $\hat{R}_h(t)$ introduced at the end of Chapter 1 for completely observed data (see (1.32) and (1.33) for more details).

Using the same ideas as Swanepoel and Van Graan (2005), we propose here the following version of $\hat{R}_h(t)$:

$$\hat{R}^T_{h,g}(t) = \frac{1}{n_1}\sum_{j=1}^{n_1}\mathbb{K}\left(\frac{\hat{R}_g(t)-\hat{R}_g(F_{0n_0}(X_{1j}))}{h}\right),$$

that is based on a nonparametric transformation of the data.

Following the same ideas as Janssen *et al* (2004), we propose as well the following variable bandwidth version of $\hat{r}_h(t)$:

$$\hat{r}^V_{h,g}(t) = \frac{1}{n_1 h}\sum_{j=1}^{n_1}K\left(\frac{t-F_{0n_0}(X_{1j})}{h}\hat{r}_g^{1/2}(F_{0n_0}(X_{1j}))\right)\hat{r}_g^{1/2}(F_{0n_0}(X_{1j})).$$

Similarly, when the data are subject to left truncation and right censoring, we define a version of $\check{R}_h(t)$ based on a nonparametric transformation of the data and a variable bandwidth version of $\check{r}_h(t)$ as given below:

$$\check{R}^T_{h,g}(t) = \frac{1}{n_1}\sum_{j=1}^{n_1}\mathbb{K}\left(\frac{\check{R}_g(t)-\check{R}_g(\hat{F}_{0n_0}(X_{1j}))}{h}\right)$$

and

$$\check{r}^V_{h,g}(t) = \frac{1}{n_1 h}\sum_{j=1}^{n_1} K\left(\frac{t-\hat{F}_{0n_0}(X_{1j})}{h}\check{r}_g^{1/2}(\hat{F}_{0n_0}(X_{1j}))\right)\check{r}_g^{1/2}(\hat{F}_{0n_0}(X_{1j})).$$

Note that the kernel-type estimator $\hat{r}_h(t)$ follows, in a natural way from the Parzen-Rosenblatt estimator, $\tilde{r}_h(t)$, associated to the random variable $Z = F_0(X_1)$ when it is assumed that F_0 is known, or in other words, when the two-sample problem of estimating the relative density is reduced to the one-sample problem of estimating the ordinary density function of Z. In fact, $\hat{r}_h(t)$ is obtained from $\tilde{r}_h(t)$ by replacing $F_0(X_{1j})$ by the empirical estimate $F_{0n_0}(X_{1j})$. If we now start from the following modification of $\tilde{r}_h(t)$:

$$\tilde{r}^V_h(t) = \frac{1}{n_1}\sum_{j=1}^{n_1} K\left(\frac{t-F_0(X_{1j})}{h}\phi(X_{1j})\right)\phi(X_{1j}),$$

where $\phi(x)$ is some function to be specified, it follows that:

$$\begin{aligned} E\left[\tilde{r}^V_h(t)\right] &= \frac{1}{h}\int K\left(\frac{t-F_0(y)}{h}\phi(y)\right)\phi(y)f_1(y)\mathrm{d}y \\ &= \int K(z\phi(F_0^{-1}(t-hz)))\phi(F_0^{-1}(t-hz))r(t-hz)\mathrm{d}z \\ &= \int K(z\tilde{\phi}(t-hz))\tilde{\phi}(t-hz)r(t-hz)\mathrm{d}z \\ &= r(t)+\frac{h^2}{2}\frac{d_K}{\tilde{\phi}^4(t)}\left\{6r(t)\tilde{\phi}^{(1)}(t)^2-2r(t)\tilde{\phi}(t)\tilde{\phi}^{(2)}(t)\right. \qquad (6.1) \\ &\quad \left.-4r^{(1)}(t)\tilde{\phi}(t)\tilde{\phi}^{(1)}(t)+r^{(2)}(t)\tilde{\phi}(t)^2\right\}+O\left(h^4\right) \end{aligned}$$

with $\tilde{\phi}(t) = \phi(F_0^{-1}(t))$.

If we choose $\tilde{\phi}(t) = r^{1/2}(t)$ or equivalently $\phi(t) = r^{1/2}(F_0(t))$, it can be proved that the second term in the right hand side of (6.1) vanishes. In fact, since

$$\begin{aligned} \tilde{\phi}^{(1)}(t) &= \frac{1}{2}r^{-1/2}(t)r^{(1)}(t), \\ \tilde{\phi}^{(2)}(t) &= \frac{1}{2}\left\{-\frac{1}{2}r^{-3/2}(t)r^{(1)}(t)^2+r^{-1/2}(t)r^{(2)}(t)\right\} \\ &= -\frac{1}{4}r^{-3/2}(t)r^{(1)}(t)^2+\frac{1}{2}r^{-1/2}(t)r^{(2)}(t), \end{aligned}$$

it follows that

$$\begin{aligned}
6r(t)\tilde{\phi}^{(1)}(t)^2 &= 6r(t)\frac{1}{4}r^{-1}(t)r^{(1)}(t)^2 = \frac{3}{2}r^{(1)}(t)^2, \\
-2r(t)\tilde{\phi}(t)\tilde{\phi}^{(2)}(t) &= -2r(t)r^{1/2}(t)\left\{-\frac{1}{4}r^{-3/2}(t)r^{(1)}(t)^2 + \frac{1}{2}r^{-1/2}(t)r^{(2)}(t)\right\} \\
&= \frac{1}{2}r^{(1)}(t)^2 - r(t)r^{(2)}(t), \\
-4r^{(1)}(t)\tilde{\phi}(t)\tilde{\phi}^{(1)}(t) &= -4r^{(1)}(t)r^{1/2}(t)\left(\frac{1}{2}r^{-1/2}(t)r^{(1)}(t)\right) = -2r^{(1)}(t)^2
\end{aligned}$$

and therefore

$$\begin{aligned}
&6r(t)\tilde{\phi}^{(1)}(t)^2 - 2r(t)\tilde{\phi}(t)\tilde{\phi}^{(2)}(t) - 4r^{(1)}(t)\tilde{\phi}(t)\tilde{\phi}^{(1)}(t) + r^{(2)}(t)\tilde{\phi}(t)^2 \\
&= \frac{3}{2}r^{(1)}(t)^2 + \frac{1}{2}r^{(1)}(t)^2 - r(t)r^{(2)}(t) - 2r^{(1)}(t)^2 + r^{(2)}(t)r(t) = 0.
\end{aligned}$$

Consequently, in the one-sample problem, it is expected that the estimator

$$\tilde{r}^V_{h,g}(t) = \frac{1}{n_1}\sum_{j=1}^{n_1} K\left(\frac{t - F_0(X_{1j})}{h}\tilde{r}_g^{1/2}(F_0(X_{1j}))\right)\tilde{r}_g^{1/2}(F_0(X_{1j}))$$

improves the bias properties of the Parzen-Rosenblatt estimator $\tilde{r}_h(t)$. More specifically, that the bias is reduced from order $O\left(h^2\right)$ to $O\left(h^4\right)$.

If we now come back to the original problem of estimating the relative density when F_0 is unknown, and we replace $F_0(x)$ and $\tilde{r}_g(t)$ in $\tilde{r}^V_{h,g}(t)$ by respectively $F_{0n_0}(x)$ and $\hat{r}_g(t)$, it is expected that this new estimator of $r(t)$, which was previously denoted by $\hat{r}^V_{h,g}(t)$, yields as well in the two-sample problem an improvement over $\hat{r}_h(t)$ in terms of bias reduction.

The motivation of the estimators $\hat{R}^T_{h,g}(t)$ and $\check{R}^T_{h,g}(t)$ introduced previously follows from the fact that the 'ideal estimator', $S_{n_1}(t)$ below, estimates $R(t)$ without bias. Let

$$S_{n_1}(t) = \frac{1}{n_1}\sum_{j=1}^{n_1}\mathbb{K}\left(\frac{R(t) - R(F_0(X_{1j}))}{h}\right)$$

and consider that $\mathbb{K}$ is the cdf of a density kernel K which is symmetric around zero and supported in $[-1, 1]$.

Next, we detail the computations required to conclude that $E\left[S_{n_1}(t)\right] = R(t)$.

Using changes of variable and integration by parts it is easy to see that:

$$\begin{aligned}
E\left[S_{n_1}(t)\right] &= E\left[\mathbb{K}\left(\frac{R(t)-R(F_0(X_{1j}))}{h}\right)\right] \\
&= \int_{-\infty}^{F_1^{-1}(R(t)+h)} \mathbb{K}\left(\frac{R(t)-R(F_0(y))}{h}\right) \mathrm{d}F_1(y) \\
&= \int_{-\infty}^{F_1^{-1}(R(t)-h)} \mathrm{d}F_1(y) + \int_{F_1^{-1}(R(t)-h)}^{F_1^{-1}(R(t)+h)} \mathbb{K}\left(\frac{R(t)-R(F_0(y))}{h}\right) \mathrm{d}F_1(y) \\
&= F_1(F_1^{-1}(R(t)-h)) + \int_{R^{-1}(R(t)-h)}^{R^{-1}(R(t)+h)} \mathbb{K}\left(\frac{R(t)-R(z)}{h}\right) \mathrm{d}R(z) \\
&= R(t)-h+h\int_{-1}^{1} \mathbb{K}(y)\mathrm{d}y \\
&= R(t)-h+h\left\{\lim_{y\to 1}\mathbb{K}(y)y - \lim_{y\to -1}\mathbb{K}(y)y - \int yK(y)\mathrm{d}y\right\} \\
&= R(t)-h+h(\mathbb{K}(1)+\mathbb{K}(-1)) = R(t).
\end{aligned}$$

It could be interesting to study in the future the asymptotic behaviour of these new kernel type estimators of the relative distribution and relative density, either for complete or LTRC data, and check if, in fact, they provide a bias reduction.

6.2 Testing the hypothesis of proportional hazard rates along time

For a single binary covariate, Cox proportional hazard model (PH) reduces to

$$\lambda(t,x) = \lambda_0(t)\exp\{x\beta\}. \tag{6.2}$$

Since the variable x only takes two values, let say 0 and 1, then the hazard function for group 0 is $\lambda_0(t)$ (i.e. the baseline hazard) and the hazard function for group 1 is $\lambda_1(t) = \lambda_0(t)\exp(\beta)$. Consequently, under model (6.2) one assumes that the ratio of the two hazards is constant along time. Since violations of the PH assumption can lead to incorrect inferences, it is important to check for PH violations. The objective for another future work is to design a goodness-of-fit test based on kernel relative hazard estimators for testing the general null hypothesis $H_0^G : \lambda_R(t) = \exp(\beta)$ for some $\beta \in \mathbb{R}$ against the alternative hypothesis $H_1 : \lambda_R(t) \neq \exp(\beta)$ for any $\beta \in \mathbb{R}$, where $\lambda_R(t) = \Lambda_R^{(1)}(t) = \frac{\lambda_1(\Lambda_0^{-1}(t))}{\lambda_0(\Lambda_0^{-1}(t))}$ denotes the relative hazard rate of X_1 wrt X_0 and $\Lambda_R(t) = -\ln(1-F_1(\Lambda_0^{-1}(t))) = \Lambda_1(\Lambda_0^{-1}(t))$ is the relative cumulative hazard function of X_1 wrt X_0. Here $\lambda_0(t)$ ($\lambda_1(t)$) and $\Lambda_0(t)$ ($\Lambda_1(t)$) refer to the hazard rate and the cumulative hazard function of X_0 (X_1).

In fact, some preliminary work was already done on this problem when the samples are subject to left truncation and right censoring. Starting with the simpler null hypothesis $H_0^S : \lambda_R(t) = \theta_0$ where θ_0 is now a fixed and known constant, we propose the following test statistic

$$I_{n_0,n_1} = \int \left(\hat{\lambda}_R(t) - \theta_0\right)^2 \mathrm{d}t,$$

where

$$\hat{\lambda}_R(t) = \frac{1}{h}\int K\left(\frac{t-\hat{\Lambda}_0(y)}{h}\right)\mathrm{d}\hat{\Lambda}_1(y)$$

is a kernel estimate of $\lambda_R(t)$, the relative hazard rate of X_1 wrt X_0,

$$\hat{\Lambda}_0(t) = \sum_{i=1}^{n_0}\frac{1_{\{Y_{0i}\le t,\delta_{0i}=1\}}}{n_0 B_{n_0}(Y_{0i})}$$

is a kernel estimate of the cumulative hazard function of X_0, $\Lambda_0(t) = -\ln(1-F_0(t))$, and

$$\hat{\Lambda}_1(t) = \sum_{j=1}^{n_1}\frac{1_{\{Y_{1j}\le t,\delta_{1j}=1\}}}{n_1 B_{n_1}(Y_{1j})}$$

denotes a kernel-type estimate of the cumulative hazard function of X_1, $\Lambda_1(t) = -\ln(1-F_1(t))$.

This statistic can be easily decomposed as follows:

$$I_{n_0,n_1} = J_{n_0,n_1}^{(1)} + 2J_{n_0,n_1}^{(2)} + J_{n_0,n_1}^{(3)},$$

where

$$\begin{aligned}
J_{n_0,n_1}^{(1)} &= \int \left(\hat{\lambda}_R(t) - E\left[\hat{\lambda}_R(t)\right]\right)^2 \mathrm{d}t,\\
J_{n_0,n_1}^{(2)} &= \int \left(\hat{\lambda}_R(t) - E\left[\hat{\lambda}_R(t)\right]\right)\left(E\left[\hat{\lambda}_R(t)\right] - \theta_0\right)\mathrm{d}t,\\
J_{n_0,n_1}^{(3)} &= \int \left(E\left[\hat{\lambda}_R(t)\right] - \theta_0\right)^2 \mathrm{d}t.
\end{aligned}$$

This is a typical decomposition where $J_{n_0,n_1}^{(3)}$ and the term $\left(E\left[\hat{\lambda}_R(t)\right] - C\right)$ in $J_{n_0,n_1}^{(2)}$ are not random. While the first term, $J_{n_0,n_1}^{(1)}$, can be handled using Central Limit Theorems for U-statistics, the second term, $J_{n_0,n_1}^{(2)}$, can be handled using classical forms of the Central Limit Theorem such as for example Lyapunov Theorem.

Based on Theorem 1 in Cao *et al* (2005), an asymptotic representation of $\hat{\lambda}_R(t)$ for LTRC data is given as follows

$$\hat{\lambda}_R(t) = A_{n_1}^{(1)}(t) + A_{n_1}^{(2)}(t) + B_{n_0,n_1}^{(11)}(t) + C_{n_0,n_1}(t)$$

with

$$\begin{aligned}
A_{n_1}^{(1)}(t) &= \int K(u)\lambda_R(t-hu)\mathrm{d}u, \\
A_{n_1}^{(2)}(t) &= \frac{1}{n_1 h}\sum_{j=1}^{n_1}\int \tilde{\xi}_{1j}^{H}(t-hu)K^{(1)}(u)\mathrm{d}u, \\
B_{n_0,n_1}^{(11)}(t) &= -\frac{1}{n_0 h^2}\sum_{i=1}^{n_0}\int \tilde{\xi}_{0i}^{H}(v)K^{(1)}\left(\frac{t-v}{h}\right)\lambda_R(v)\mathrm{d}v \\
&= -\frac{1}{n_0 h}\sum_{i=1}^{n_0}\int \tilde{\xi}_{0i}^{H}(t-uh)K^{(1)}(u)\lambda_R(t-hu)\mathrm{d}u, \\
\tilde{\xi}_{0i}^{H}(Y_{0i},T_{0i},\delta_{0i},t) &= \frac{1_{\left\{Y_{0i}\le\Lambda_0^{-1}(t),\delta_{0i}=1\right\}}}{B_0(Y_{0i})} - \int_{a_{W_0}}^{\Lambda_0^{-1}(t)}\frac{1_{\{T_{0i}\le v\le Y_{0i}\}}}{B_0^2(v)}\mathrm{d}W_{01}(v), \\
\tilde{\xi}_{1j}^{H}(Y_{1j},T_{1j},\delta_{1j},t) &= \frac{1_{\left\{Y_{1j}\le\Lambda_0^{-1}(t),\delta_{1j}=1\right\}}}{B_1(Y_{1j})} - \int_{a_{W_1}}^{\Lambda_0^{-1}(t)}\frac{1_{\{T_{1j}\le v\le Y_{1j}\}}}{B_1^2(v)}\mathrm{d}W_{11}(v)
\end{aligned}$$

and $C_{n_0,n_1}(t) = o\left((n_1 h)^{-\frac{1}{2}}\right)$.

Using this asymptotic representation, $\hat{\lambda}_R(t) - E\left[\hat{\lambda}_R(t)\right]$ can be rewritten as follows:

$$\hat{\lambda}_R(t) - E\left[\hat{\lambda}_R(t)\right] = A_{n_1}^{(2)}(t) + B_{n_0,n_1}^{(11)}(t) + C_{n_0,n_1}(t) - E\left[C_{n_0,n_1}(t)\right].$$

Consequently, it is satisfied that

$$J_{n_0,n_1}^{(1)} = J_{n_0,n_1}^{(11)} + J_{n_0,n_1}^{(12)} + J_{n_0,n_1}^{(13)} + J_{n_0,n_1}^{(14)} + J_{n_0,n_1}^{(15)} + J_{n_0,n_1}^{(16)},$$

where

$$\begin{aligned}
J_{n_0,n_1}^{(11)} &= \int A_{n_1}^{(2)}(t)^2\,\mathrm{d}t, \\
J_{n_0,n_1}^{(12)} &= \int B_{n_0,n_1}^{(11)}(t)^2\,\mathrm{d}t, \\
J_{n_0,n_1}^{(13)} &= \int \left(C_{n_0,n_1}(t) - E\left[C_{n_0,n_1}(t)\right]\right)^2\mathrm{d}t, \\
J_{n_0,n_1}^{(14)} &= 2\int A_{n_1}^{(2)}(t)\,B_{n_0,n_1}^{(11)}(t)\,\mathrm{d}t, \\
J_{n_0,n_1}^{(15)} &= 2\int A_{n_1}^{(2)}(t)\left(C_{n_0,n_1}(t) - E\left[C_{n_0,n_1}(t)\right]\right)\mathrm{d}t, \\
J_{n_0,n_1}^{(16)} &= 2\int B_{n_0,n_1}^{(11)}(t)\left(C_{n_0,n_1}(t) - E\left[C_{n_0,n_1}(t)\right]\right)\mathrm{d}t.
\end{aligned}$$

We can expect that $J_{n_0,n_1}^{(1k)} = o\left((n_1 h)^{-1}\right)$, for $k = 3,4,5,6$ and that the contributing terms in $J_{n_0,n_1}^{(1)}$, will be $J_{n_0,n_1}^{(11)}$ and $J_{n_0,n_1}^{(12)}$.

Let consider for example $J^{(11)}_{n_0,n_1}$. The other term, $J^{(12)}_{n_0,n_1}$, could be handled in a similar way. Rewriting $A^{(2)}_{n_1}(t)^2$ in the integrand of $J^{(11)}_{n_0,n_1}$ as a double sum:

$$A^{(2)}_{n_1}(t)^2 = \frac{1}{(n_1 h)^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_1}\int\int \tilde{\xi}^H_{1i}(t-hu)\,\tilde{\xi}^H_{1j}(t-hv)\,K^{(1)}(u)\,K^{(1)}(v)\,\mathrm{d}u\mathrm{d}v$$

and splitting it into two terms (sum over those terms with $i=j$ and sum over those terms with $i \neq j$), it follows that:

$$J^{(11)}_{n_0,n_1} = J^{(111)}_{n_0,n_1} + J^{(112)}_{n_0,n_1}$$

with

$$\begin{aligned} J^{(111)}_{n_0,n_1} &= \frac{1}{(n_1 h)^2}\sum_{i=1}^{n_1}\int\int\left(\int \tilde{\xi}^H_{1i}(t-hu)\,\tilde{\xi}^H_{1i}(t-hv)\,\mathrm{d}t\right)K^{(1)}(u)\,K^{(1)}(v)\,\mathrm{d}u\mathrm{d}v, \\ J^{(112)}_{n_0,n_1} &= \frac{1}{(n_1 h)^2}\sum_{i=1}^{n_1}\sum_{\substack{j=1\\ j\neq i}}^{n_1}\int\int\left(\int \tilde{\xi}^H_{1i}(t-hu)\,\tilde{\xi}^H_{1j}(t-hv)\,dt\right)K^{(1)}(u)\,K^{(1)}(v)\,\mathrm{d}u\mathrm{d}v. \end{aligned}$$

It is expected that $J^{(111)}_{n_0,n_1}$ can be handled using classical forms of the Central Limit Theorem such as Lyapunov Theorem. However, the asymptotic study of $J^{(112)}_{n_0,n_1}$, may be more problematic and Central Limit Theorems for U-statistics will be required here.

In fact, $J^{(112)}_{n_0,n_1}$ can be rewritten

$$J^{(112)}_{n_0,n_1} = \frac{2}{(n_1 h)^2}\sum_{1\le i<j\le n_1}^{n_1} H_{n_1}\left((T_{1i},Y_{1i},\delta_{1i}),(T_{1j},Y_{1j},\delta_{1j})\right),$$

where H_{n_1} denotes the kernel of the U-statistic:

$$\begin{aligned} H_{n_1}\left((T_{1i},Y_{1i},\delta_{1i}),(T_{1j},Y_{1j},\delta_{1j})\right) &= \iint\left(\int \tilde{\xi}^H_{1i}(t-hu)\,\tilde{\xi}^H_{1j}(t-hv)\,\mathrm{d}t\right)K^{(1)}(u)K^{(1)}(v)\,\mathrm{d}u\mathrm{d}v \\ &= \iiint A(t,u,(T_{1i},Y_{1i},\delta_{1i}))\,A(t,v,(T_{1j},Y_{1j},\delta_{1j}))\,\mathrm{d}t\mathrm{d}u\mathrm{d}v \end{aligned}$$

and

$$\begin{aligned} A(t,u,(T_{1i},Y_{1i},\delta_{1i})) &= K^{(1)}(u)\,\tilde{\xi}^H_{1i}(t-hu) \\ &= K^{(1)}(u)\,\frac{1_{\{Y_{1i}\le\Lambda_0^{-1}(t-hu),\delta_{1i}=1\}}}{B_1(z)} \\ &\quad -K^{(1)}(u)\int_{a_{W_1}}^{\Lambda_0^{-1}(t-hu)}\frac{1_{\{T_{1i}\le w\le Y_{1i}\}}}{B_1^2(w)}\mathrm{d}W_{11}(w). \end{aligned}$$

Note that H_{n_1} is a symmetric and degenerate kernel that depends on the sample size. Therefore, a complication is added here and we will not be able to use classical results for U-statistics. However, it is expected that, if the conditions of Theorem 1 in Hall (1984) are satisfied, the asymptotic limit distribution of $J^{(112)}_{n_0,n_1}$ could be established.

According to Theorem 1 in Hall (1984), we need to prove the following conditions:

(H1) H_{n_1} is a symmetric kernel.

(H2) $E\left[H_{n_1}((T_{11},Y_{11},\delta_{11}),(T_{12},Y_{12},\delta_{12}))\big/(T_{11},Y_{11},\delta_{11})\right]=0$ a.s.

(H3) $E\left[H_{n_1}^2((T_{11},Y_{11},\delta_{11}),(T_{12},Y_{12},\delta_{12}))\right]<\infty$.

(H4) As the sample size n_1 tends to infinity,

$$\frac{E\left[G_{n_1}((T_{11},Y_{11},\delta_{11}),(T_{12},Y_{12},\delta_{12}))\right]+n_1^{-1}E\left[H_{n_1}^4((T_{11},Y_{11},\delta_{11}),(T_{12},Y_{12},\delta_{12}))\right]}{\left\{E\left[H_{n_1}^2((T_{11},Y_{11},\delta_{11}),(T_{12},Y_{12},\delta_{12}))\right]\right\}^2}$$

tends to zero, where

$$\begin{aligned} G_{n_1}((\ell_1,y_1,d_1),(\ell_2,y_2,d_2)) &= E\left[H_{n_1}((T_1,Y_1,\delta_1),(\ell_1,z_1,d_1))\right. \\ &\quad \left. H_{n_1}((T_1,Y_1,\delta_1),(\ell_2,z_2,d_2))\right]. \end{aligned}$$

Note that in the discussion above we have considered the simpler problem of testing the null hypothesis $H_0^S:\lambda_R(t)=C$ where C is a fixed and known constant. However, as it was mentioned at the beginning of this section, the main interest in practical applications, is to test the more general null hypothesis: $H_0^G:\lambda_R(t)=\exp(\beta)$ for some $\beta\in\mathbb{R}$.

When the objective is to test the general null hypothesis H_0^G, a natural generalization of the test statistic I_{n_0,n_1} introduced previously for testing H_0^S, could be given by

$$I_{n_0,n_1}^G=\int\left(\hat{\lambda}_R(t)-\hat{\theta}\right)^2 \mathrm{d}t,$$

where $\theta=\exp(\beta)$ and $\hat{\theta}$ is an estimate of θ. On the one hand, it is interesting to note here that I_{n_0,n_1}^G can be decomposed in a sum of three terms where one of them is precisely the test statistic I_{n_0,n_1} introduced previously for the simpler null hypothesis H_0^S:

$$I_{n_0,n_1}^G=\int\left(\hat{\lambda}_R(t)-\theta_0+\theta_0-\hat{\theta}\right)^2 \mathrm{d}t=I_{n_0,n_1}+J_{n_0,n_1}+K_{n_0,n_1},$$

where

$$\begin{aligned} J_{n_0,n_1} &= \int\left(\theta_0-\hat{\theta}\right)^2 \mathrm{d}t, \\ K_{n_0,n_1} &= \int\left(\hat{\lambda}_R(t)-\theta_0\right)\left(\theta_0-\hat{\theta}\right) \mathrm{d}t. \end{aligned}$$

On the other hand, in this more general setting of testing H_0^G, we need to consider an estimate of the parameter β (or equivalently an estimate of θ) when the data are subject to left truncation and right censoring. Cox (1972) provides estimates for the regression parameters of a Cox model by maximizing the likelihood that does not depend on the baseline hazard function $\lambda_0(t)$. A more detailed justification was given later on by the same author (see Cox (1975)) under the term of partial likelihood. Even when there exists the possibility to extend this approach to work with LTRC data (see Alioum and Commenges (1996) for more details), the problem is that there does not exist a closed or explicit expression for the estimates of the parameter β based on the Cox partial likelihood. They are simply the solution of an equation which has to be solved numerically.

Even when the two-sample problem can be viewed as a special case of Cox proportional hazard model, it is also a very simple case of that more general model. Therefore, as Li (1995) suggests, there must be a more intuitive and simple approach to deal with the two-sample problem parameterized in terms of the general null hypothesis H_0^G. Next we give some detail on the approach given by Li (1995) regarding this problem and present the explicit formula of the estimator obtained for θ under the two-sample problem with LTRC data.

Under left truncation and right censoring, Li (1995) considers that one can only estimate the survival beyond a time t when t is larger than a_{G_0} for the first sample and larger than a_{G_1} for the second sample, i.e.:

$$1 - F_0^e(t) = \frac{1 - F_0(t)}{1 - F_0(a_{G_0})}, \quad \text{for } t \geq a_{G_0}, \tag{6.3}$$

$$1 - F_1^e(t) = \frac{1 - F_1(t)}{1 - F_1(a_{G_1})}, \quad \text{for } t \geq a_{G_1}. \tag{6.4}$$

Under the general hypothesis H_0^G it is assumed that:

$$\lambda_0(t) = \exp\{\beta\}\lambda_1(t) = \theta\lambda_1(t),$$

which implies that

$$1 - F_0(t) = (1 - F_1(t))^\theta. \tag{6.5}$$

In the particular case θ is a positive integer, this means that F_0 is distributed as the smallest of θ X_1 variables, which is also known as a Lehmann-type alternative although slightly different from the one defined in Section 1.4.1.

Since we can only estimate $F_0^e(t)$ and $F_1^e(t)$, equations (6.3), (6.4) and (6.5) lead to

$$(1 - F_0^e(t))(1 - F_0(a_{G_0})) = \{(1 - F_1^e(t))(1 - F_1(a_{G_1}))\}^\theta$$

and consequently

$$1 - F_0^e(t) = \frac{(1 - F_1(a_{G_1}))^\theta}{1 - F_0(a_{G_0})}(1 - F_1^e(t))^\theta = \gamma(1 - F_1^e(t))^\theta,$$

where $\gamma = \frac{(1-F_1(a_{G_1}))^\theta}{1-F_0(a_{G_0})}$.

Since $\Lambda_0^e(t) = -\ln(1 - F_0^e(t))$ and $\Lambda_1^e(t) = -\ln(1 - F_1^e(t))$, it follows that:

$$\begin{aligned}\Lambda_0^e(t) &= -\ln\left(\gamma\,(1 - F_1^e(t))^\theta\right)\\ &= -\ln(\gamma) - \ln\left((1 - F_1^e(t))^\theta\right)\\ &= \rho + \theta\Lambda_1^e(t),\end{aligned}$$

where $\rho = -\ln(\gamma)$.

In view of the relationship existing between $\Lambda_0^e(t)$ and $\Lambda_1^e(t)$, the problem we need to solve turns up to be more general than the one assumed by the Cox proportional hazard model when there is only a binary covariate that needs to be fit.

Li (1995) defines a consistent estimator of (θ, ρ) in this scenario as the pair $(\hat{\theta}_{n_0,n_1}, \hat{\rho}_{n_0,n_1})$ which minimizes the function $E_{n_0,n_1}(\theta, \rho)$ below:

$$E_{n_0,n_1}(\theta, \rho) = \int_a^b (\hat{\Lambda}_0(t) - \rho - \theta\hat{\Lambda}_1(t))^2 \mathrm{d}\hat{F}_{1n_1}(t)$$

and that is the sample analogous of

$$E(\theta, \rho) = \int_a^b (\Lambda_0^e(t) - \rho - \theta\Lambda_1^e(t))^2 \mathrm{d}F_1(t),$$

where the interval $[a, b]$ is chosen such that $\Lambda_0^e(t)$ and $\Lambda_1^e(t)$ can be consistently estimated for any $t \in [a, b]$.

From this definition it follows that:

$$\begin{aligned}\hat{\theta}_{n_0,n_1} &= \frac{\int_a^b \hat{\Lambda}_0(t)\hat{\Lambda}_1(t)\mathrm{d}\hat{F}_{1n_1}(t) - \int_a^b \hat{\Lambda}_0(t)\mathrm{d}\hat{F}_{1n_1}(t)\int_a^b \hat{\Lambda}_1(t)\mathrm{d}\hat{F}_{1n_1}(t)}{\int_a^b \hat{\Lambda}_1^2(t)\mathrm{d}\hat{F}_{1n_1}(t)\int_a^b \mathrm{d}\hat{F}_{1n_1}(t) - \left(\int_a^b \hat{\Lambda}_1(t)\mathrm{d}\hat{F}_{1n_1}(t)\right)^2},\\ \hat{\rho}_{n_0,n_1} &= \frac{\int_a^b \left(\hat{\Lambda}_1(t) - \hat{\theta}_{n_0,n_1}\hat{\Lambda}_1(t)\right)^2 \mathrm{d}\hat{F}_{1n_1}(t)}{\int_a^b \mathrm{d}\hat{F}_{1n_1}(t)}.\end{aligned}$$

In the special case $\rho = 0$, $\hat{\theta}_{n_0,n_1}$ simplifies to $\tilde{\theta}_{n_0,n_1}$ as given below:

$$\tilde{\theta}_{n_0,n_1} = \frac{\int_a^b \hat{\Lambda}_0(t)\hat{\Lambda}_1(t)\mathrm{d}\hat{F}_{1n_1}(t)}{\int_a^b \hat{\Lambda}_1^2(t)\mathrm{d}\hat{F}_{1n_1}(t)}.$$

Under the typical assumptions of identifiability of F_0 and F_1, i.e. $a_{G_0} \leq a_{W_0}$ and $a_{G_1} \leq a_{W_1}$, it follows that $F_0^e(t) = F_0(t)$ and $F_1^e(t) = F_1(t)$. Consequently, we propose to consider $\hat{\theta} = \tilde{\theta}_{n_0,n_1}$ in I_{n_0,n_1}^G introduced above since under the assumptions of identifiability of F_0 and F_1, $\rho = 0$.

Therefore, the future work we need to do involves, on the one hand, checking the conditions (H1)-(H4), what from Theorem 1 in Hall (1984) will allow to get the asymptotic behaviour of I_{n_0,n_1}. On the other hand, we need as well to study the asymptotic behaviour of the other terms, J_{n_0,n_1} and K_{n_0,n_1}, appearing in the decomposition of the test statistic $I^G_{n_0,n_1}$. Finally, a simulation study should be carried out to check the performance of this statistic with small to moderate sample sizes and under different percentages of censoring and truncation.

6.3 EL confidence bands for the relative distribution with LTRC data

As a continuation of the work developed in Chapter 4, it could be interesting to extend those results for testing the global hypothesis $H_0 : R(t) = t, \forall t \in [0, 1]$. Note that in Chapter 4 all the theoretical results were obtained for a fixed t and the objective was to make inference about $R(t) = \theta$. Cao and Van Keilegom (2006) designed a test statistic of the global hypothesis $H_0 : f_0(x) = f_1(x), \forall x$, in the setting of a two-sample problem with complete data. Here, we could either extend their approach based on EL and kernel density estimates to LTRC data or define a test statistic where the log likelihood function introduced in (4.25) is now integrated along t.

On the other hand, it is of interest to check, through a simulation study, the behaviour of the empirical likelihood confidence intervals briefly introduced at the end of Chapter 4 and to compare them with those obtained by means of the asymptotic normality of the kernel type estimator, $\check{R}_h(t)$, introduced at the end of Chapter 3 (see Theorem 3.2.3 for more details about its asymptotic normality). For this future project, we need to define estimators of the two unknown quantities, d_1 and d_2, appearing in the asymptotic limit distribution of $\ell(\theta)$, as well as an estimator of the asymptotic variance of $\check{R}_h(t)$. For this last problem, one possibility could be to use a bootstrap estimate of the variance. More specifically, we could generate a large number of pairs of bootstrap resamples, independently from each sample, using the 'simple' bootstrap, that was first designed by Efron (1981) for complete data and that, according to Gross and Lai (1996), seems to work as well with LTRC data. Then, after computing the bootstrap analogous of $\check{R}_h(t)$ for every pair of bootstrap samples, we could use the Monte Carlo method to obtain a bootstrap estimate of the variance of $\check{R}_h(t)$. Different bootstrap resampling plans could be studied as well.

6.4 Extensions for dealing with multi-state problems

Multistate models have been increasingly used to model the natural history of a chronic disease as well as to characterize the follow-up of patients under different clinical protocols. Multistate models may be considered as a generalization of the survival process where several events occur successively along time. Therefore, apart from registering the lifetime of each patient, there are some others times along their lifetime that are of interest in this setting, for example, the times at which the patients pass from one stage to another. Note that in the survival process there is only two possible stages, alive or dead, and one of them, dead, is a terminal event. However, this is not the case on multistate problems, where there are more than two stages and the patient could move from one to another before he reaches the terminal stage.

In this new setting, we could define two populations, for example, men versus women, and it could be interesting to compare as well their history of disease from a distributional point of view. The generalization of relative curves to this new setting is of interest, especially, to analyze a dataset of patients with HER2-overexpressing metastatic breast cancer treated with trastuzumab-based therapy at The University of Texas M.D. Anderson Cancer Center (Guarneri *et al* (2006)). Since the risk of a cardiac event increases with long term trastuzumab-based therapy, the patients may pass along their follow-up by periods of discontinuation of the treatment with trastuzumab.

Appendix A

Some useful material

Definition A.0.1. A stochastic process $\{X_t, t \in T\}$ is Gaussian if for all $k \in \mathbb{N}$ and all $t_1, t_2, \dots, t_k \in T$, the random vector $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ has a multivariate normal distribution, that is, all finite dimensional distributions of the process are Gaussian.

Definition A.0.2. With $\mathbb{Q}$ denoting the rational numbers (which is a countable set), we call a continuous-time stochastic process $\{X_t, t \in T\}$ separable, if for all $\alpha, \beta \in \mathbb{R}$, $\alpha < \beta$ and each open interval $T_0 \subset T \subset \mathbb{R}$,

$$P(\alpha \le X_t \le \beta, \forall t \in T_0) = P(\alpha \le X_t \le \beta, \forall t \in T_0 \cap \mathbb{Q}).$$

Definition A.0.3. $W(x_1, x_2)$ is said a two-dimensional Wiener-process if it is a separable Gaussian-process satisfying that

$$E[W(x_1, x_2)] = 0$$

and

$$E[W(x_{11}, x_{12})W(x_{21}, x_{22})] = \min\{x_{11}, x_{21}\} \min\{x_{12}, x_{22}\}.$$

Definition A.0.4. $K(x_1, x_2, y)$ is said a two-dimensional Kiefer-process if $K(x_1, x_2, y) = W(x_1, x_2, y) - x_1 x_2 W(1, 1, y)$, where $0 \le x_1 \le 1, 0 \le x_2 \le 1, 0 \le y < \infty$ and $W(x_1, x_2)$ is a two-dimensional Wiener process.

Definition A.0.5. A function $f(x)$ is γ-Hölder continuous if it satisfies

$$|f(x) - f(y)| \le L\,|x - y|^{\gamma}, \text{ for all } x, y \in \mathbb{R},$$

where $0 < \gamma \le 1$ and L is a positive constant.

Definition A.0.6. A function $f(x)$ is Lipschitz-continuous if there exists a positive constant L, such that:

$$|f(x) - f(y)| \leq L\,|x - y|\,.$$

The smallest such L is called the Lipschitz constant.

Definition A.0.7. Let a discrete distribution P have probability function p_k, and let a second discrete distribution Q have probability function q_k. Then, the relative entropy of P with respect to Q, also called the Kullback-Leibler distance, is defined by

$$D_{KL}(P,Q) = \sum_k p_k \ln\left(\frac{p_k}{q_k}\right).$$

Although $D_{KL}(P,Q) \neq D_{KL}(Q,P)$, so relative entropy is therefore not a true metric, it satisfies many important mathematical properties. For example, it is a convex function of p_k, it is always nonnegative, and it is equal to zero only if $p_k = q_k$.

When the two distributions P and Q are continuous, the summation is replaced by an integral, so that

$$D_{KL}(P,Q) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} \mathrm{d}x,$$

where p and q, denote the densities pertaining to P and Q, respectively.

Definition A.0.8. The Stirling formula or Stirling approximation, which is named in honor of James Stirling, is an approximation for large factorials. Formally, it is given by

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n} = 1.$$

Definition A.0.9. The Stirling numbers of first class, $s(n,m)$, are the integer coefficients of the falling factorial polynomials, $x(x-1)\cdots(x-n+1)$, i.e.:

$$x(x-1)\cdots(x-n+1) = \sum_{m=1}^{n} s(n,m)x^m.$$

The number of permutations of n elements which contain exactly m permutation cycles is counted by $(-1)^{n-m}s(n,m)$.

Since $x(x-1)\cdots(x-n) = x\left\{x(x-1)\cdots(x-n+1)\right\} - n\left\{x(x-1)\cdots(x-n+1)\right\}$, there exists a characterization of the Stirling numbers of first kind based on the following recurrence formula:

$$s(n+1,m) = s(n,m-1) - ns(n,m), \quad 1 \leq m < n,$$

subject to the following initial constraints: $s(n,0) = 0$ and $s(1,1) = 1$.

Definition A.0.10. The Stirling numbers of second class, $S(n,m)$, represent the number of ways a set of n elements can be partitioning in m nonempty sets.

There exist several generating functions of these numbers, for example:

$$x^n = \sum_{m=0}^{n} S(n,m)x(x-1)\cdots(x-m+1),$$

but they can also be computed from the following sum:

$$S(n,m) = \frac{1}{m!}\sum_{i=0}^{m}(-1)^i\binom{m}{i}(m-i)^n.$$

Theorem A.0.1. *(Fubini's Theorem)*
If $f(x,y)$ is continuous on the rectangular region $R = \{(x,y) : a \le x \le b, c \le y \le d\}$, then

$$\int_R f(x,y)\mathrm{d}(x,y) = \int_a^b \int_c^d f(x,y)\mathrm{d}y\mathrm{d}x = \int_c^d \int_a^b f(x,y)\mathrm{d}x\mathrm{d}y.$$

Theorem A.0.2. *(Hölder inequality)*
Let S be a measure space, let $1 \le p,q \le \infty$ with $1/p + 1/q = 1$, let f be in $L^p(S)$ and g be in $L^q(S)$. Then, fg is in $L^1(S)$ and

$$\|fg\|_1 \le \|f\|_p \|g\|_q .$$

The numbers p and q are known as the 'Hölder conjugates' of each other.

Theorem A.0.3. *(Cauchy-Schwarz inequality)*
This inequality states that if x and y are elements of inner product spaces then

$$|\langle x,y\rangle|^2 \le \langle x,x\rangle \langle y,y\rangle ,$$

where $\langle \cdot,\cdot \rangle$ denotes the inner product.

Theorem A.0.4. *(Weierstrass Approximation Theorem. Bernstein's Polynomials)*
Let $f : [0,1] \to \mathbb{R}$ be a continuous function and let ϵ be a positive number. There exists a number N such that for $n > N$

$$|f(x) - W_n(x)| < \epsilon, \text{ for } x \in [0,1],$$

where

$$W_n(x) = \sum_{k=0}^{n} f(k/n)\binom{n}{k}x^k(1-x)^{n-k}.$$

In other words, for every continuous function $f : [0,1] \to \mathbb{R}$ there exists a sequence $W_1, W_2, \ldots$ of polynomials uniformly approaching f in $[0,1]$.

Theorem A.0.5. *(Lyapunov Theorem)*
Let X_n be a sequence of independent random variables, at least one of which has a non-degenerate distribution. Let $E|X_n|^{2+\delta} < \infty$ for some $\delta > 0$ $(n = 1, 2, \ldots)$. If

$$B_n^{-1-\delta/2} \sum_{k=1}^{n} E|X_k - m_k|^{2+\delta} \to 0,$$

then

$$\sup_x |F_n(x) - \Phi(x)| \to 0,$$

where

$$m_n = E[X_n], \sigma_n^2 = Var(X_n), B_n = \sum_{k=1}^{n} \sigma_k^2$$

and Φ denotes the standard Gaussian distribution.

Theorem A.0.6. *(Bernstein inequality)*
Let $X_1, X_2, \ldots, X_n$ be independent random variables satisfying that

$$P(|X_i - E[X_i]| \leq m) = 1,$$

for each i, where $m < \infty$. Then, for $t > 0$,

$$P\left(\left|\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} E[X_i]\right| \geq nt\right) \leq 2\exp\left(-\frac{n^2t^2}{2\sum_{i=1}^{n} Var(X_i) + \frac{2}{3}mnt}\right),$$

for all $n = 1, 2, \ldots$

Theorem A.0.7. *(Borel-Cantelli Lemma)*
(i) For arbitrary events $\{B_n\}$, if $\sum_n P(B_n) < \infty$, then $P(B_n$ infinitely often $) = 0$.
(ii) For independent events $\{B_n\}$, if $\sum_n P(B_n) = \infty$, then $P(B_n$ infinitely often $) = 1$.

Theorem A.0.8. *(Dvoretzky-Kiefer-Wolfowitz inequality)*
Let F_{0n} denote the empirical distribution function for a sample of n iid random variables with distribution function F_0. There exists a positive constant C such that:

$$P\left(\sqrt{n}\sup_x (F_{0n}(x) - F_0(x)) \geq \lambda\right) \leq C\exp\left\{-2\lambda^2\right\}.$$

Provided that $\exp\left\{-2\lambda^2\right\} \leq \frac{1}{2}$, *$C$ can be chosen as* 1 *(see Massart (1990)). In particular, the two-sided inequality*

$$P\left(\sqrt{n}\sup_x |F_{0n}(x) - F_0(x)| \geq \lambda\right) \leq C\exp\left\{-2\lambda^2\right\}$$

holds without any restriction on λ.

Theorem A.0.9. *(Wilks Theorem)*
Consider the m*-dimensional parameter,* $\theta \in \Theta \subset \mathbb{R}^m$*, and the null hypothesis* $H_0 : \theta \in \Theta_0 \subset \mathbb{R}^k$ *vs the alternative* $H_1 : \theta \notin \Theta_0$*. Let* $\{X_{01}, \ldots, X_{0n}\}$ *be a sample of* n *iid random variables distributed as* F_0 *and let* $L(X, \theta)$ *denote the likelihood function associated to* $\{X_{01}, \ldots, X_{0n}\}$*. Then, the Wilks statistic,*

$$W(X) = -2 \ln \left\{ \frac{\sup_{\theta \in \Theta_0} L(X, \theta)}{\sup_{\theta \in \Theta} L(X, \theta)} \right\},$$

is, under H_0*, asymptotically* χ^2_{m-k} *distributed.*

Appendix B

Summary in Galician / Resumo en galego

Esta tese trata o problema de dúas mostras dentro da análise de supervivencia. No capítulo 1 preséntase unha breve introdución ao modelo de truncamento pola esquerda e censura pola dereita. Os fenómenos de censura e truncamento aparecen en problemas de supervivencia cando a variable de interese é un tempo de vida. Debido ao fenómeno de censura pola dereita só aqueles datos que levan asociado un valor de censura maior, son completamente observados. Con todo, cando o valor da censura se rexistra antes de que o dato de interese poida ser observado, entón a única información que se ten é que o dato, de poder observarse, sería maior que o valor da censura. En presenza de truncamento pola esquerda, soamente se observa o dato de interese cando este é maior que o valor da variable de truncamento. De non ser así, un non ten evidencia algunha da existencia dese dato.

Baixo censura pola dereita e truncamento pola esquerda un observa, para cada individuo na mostra, o vector aleatorio (T_0, Y_0, δ_0), onde Y_0 denota o mínimo entre a variable de interese, por exemplo, o tempo de vida X_0, e a variable de censura, que denotamos por C_0, e a variable δ_0 é a indicadora de non censura, é dicir, $\delta_0 = 1$ cando o dato Y_0 non está censurado (e por tanto $X_0 = Y_0$) e $\delta_0 = 0$ en caso contrario, cando $Y_0 = C_0 < X_0$. Ademais, os datos son observables só se T_0, a variable de truncamento, é menor ou igual que o valor de Y_0. No que segue, usaremos a notación F_0, L_0, G_0 e W_0 para referirnos respectivamente ás funcións de distribución das variables aleatorias X_0, C_0, T_0 e Y_0.

A presenza destes dous fenómenos, censura e truncamento, fai que os estimadores existentes na literatura para datos completos, perdan algunhas das súas boas propiedades. Por exemplo, o estimador empírico da distribución, que otorga o mesmo peso a cada un dos datos presentes na mostra, deixa de ser consistente cando os datos se ven afectados polos fenómenos de censura e truncamento. Tsai *et al* (1987) propuxeron o estimador límite produto de F_0, $\hat{F}_{0n}$, para unha mostra con datos independentes, identicamente distribuídos, truncados pola esquerda e censurados pola dereita

$$1 - \hat{F}_{0n}(t) = \prod_{Y_{0i} \leq t} \left[1 - (nB_{0n}(Y_{0i}))^{-1} \right]^{\delta_{0i}},$$

onde $B_{0n}(t)$ denota a estimación empírica da función $B_0(t)$, introducida na subsección 1.1.1:

$$B_{0n}(t) = n^{-1} \sum_{i=1}^{n} 1 \{T_{0i} \leq t \leq Y_{0i}\} = n^{-1} \sum_{i=1}^{n} 1 \{T_{0i} \leq t\} 1 \{t \leq Y_{0i}\}$$

e $nB_{0n}(t)$ representa o número de individuos na mostra que no tempo t están a risco de 'morrer'.

Na subsección 1.1.2 recompílanse resultados teóricos de interese sobre o comportamento asintótico do estimador límite produto $\hat{F}_{0n}$. Así, por exemplo, recóllense unha serie de teoremas encontrados na literatura que tratan da descomposición asintótica do estimador nunha suma de variables independentes e identicamente distribuídas, da converxencia débil do estimador, dunha cota exponencial, da lei funcional do logaritmo iterado e sobre a lipschitzianidade do módulo de oscilación do proceso límite produto, $\mu_{0n}(x, y)$.

Na sección 1.2 introdúcese o método bootstrap que será utilizado nos capítulos 1 e 2. Na sección 1.3 trátase a estimación non paramétrica de curvas e, en concreto, recompílanse diversas propiedades do estimador de Parzen-Rosenblatt, $\tilde{f}_{0h}$ (1.9), proposto por Parzen (1962) e Rosenblatt (1956) para a estimación dunha función de densidade f_0 a partir dunha mostra de datos independentes e identicamente distribuídos segundo f_0. Como é sabido, os métodos tipo núcleo, entre eles o estimador de Parzen-Rosenblatt, dependen dunha función chamada núcleo ou kernel, K, e dun parámetro de suavizado, h, tamén coñecido co nome de ventá. Mentres a elección do núcleo non é relevante, si o é a elección da ventá. Diversos selectores de h foron propostos na literatura para resolver o problema da elección deste parámetro de suavizado. Un breve repaso dos métodos de selección máis relevantes baseados en validación cruzada, ideas 'plug-in' ou a técnica 'bootstrap', preséntase na subsección 1.3.1. A continuación, na subsección 1.3.2 trátase o problema de estimar unha función de distribución suave no caso dunha mostra con datos completos. Preséntase o estimador empírico suavizado, $\tilde{F}_{0h}$, e así mesmo recóllense algúns dos resultados teóricos máis relevantes e un selector plug-in en varios estados proposto por Polansky and Baker (2000). As seguintes subseccións 1.3.3 e 1.3.4, estenden as dúas subseccións previas ao caso de datos censurados pola dereita e truncados pola esquerda e considéranse os análogos de $\tilde{f}_{0h}$ e $\tilde{F}_{0h}$ neste contexto, $\hat{f}_{0h,n}$ e $\hat{F}_{0h}$ ((1.25) e (1.28)). Na subsección 1.3.3 exemplifícase mediante unha pequena simulación a importancia que ten na práctica o feito de ter en conta a presenza destes dous fenómenos de censura e truncamento na mostra e o corrixir ou adaptar apropiadamente os métodos de estimación para este contexto.

Na sección 1.4 faise referencia ao problema de dúas mostras, e mais en concreto ao problema de contrastar a hipótese nula de que as dúas mostras proveñan da mesma distribución. Na subsección 1.4.1 faise un repaso dos principais estatísticos non paramétricos propostos no caso de observabilidade completa. Entre eles destacamos os estatísticos de Wald-Wolfowitz, Kolmogorov-Smirnov e Mann-Whitney. Na subsección 1.4.2 introdúcese o concepto de curva ROC, que é utilizada no campo da medicina para avaliar probas de diagnóstico que permiten clasificar os individuos en dous grupos, por exemplo, sans e enfermos. Dous estimadores, un empírico, $ROC_{n_0,n_1}(p)$, e outro baseado no método núcleo de estimación non-paramétrica, $\widehat{ROC}_{h_0,h_1}(p)$, preséntanse nesta subsección. Así mesmo, detállanse diversos selectores do parámetro de suavizado propostos na literatura para ese estimador tipo núcleo da curva ROC. Finalmente, na subsección 1.4.3 introdúcense os conceptos de curvas relativas. Mais en concreto comezamos presentando as definicións de distribución relativa e densidade relativa dunha poboación de comparación, X_1, con respecto a unha poboación de referencia, X_0. Tamén faise unha breve mención sobre a estreita relación existente entre o concepto de curva ROC e o concepto de distribución relativa así como entre a densidade relativa e a razón de densidades utilizada por Silverman (1978). Logo, inclúese unha descomposición da densidade relativa en compoñentes de localización, escala e forma residual e finalmente, este primeiro capítulo de introducción péchase coa presentación de dous estimadores tipo núcleo, un da densidade relativa, $\hat{r}_h(t)$, e outro da distribución relativa, $\hat{R}_h(t)$. Así mesmo, inclúese un teorema que proba a normalidade asintótica do estimador $\hat{r}_h(t)$.

No capítulo 2 tratamos en maior detalle a estimación non-paramétrica da densidade relativa con datos completos, é dicir, datos que non se ven afectados polo fenómeno de censura ou truncamento. Consideramos o problema de dúas mostras con datos completos $\{X_{01}, \ldots, X_{0n_0}\}$ e $\{X_{11}, \ldots, X_{1n_1}\}$, e definimos os seguintes estimadores tipo núcleo de $r(t)$:

$$\begin{aligned}\hat{r}_h(t) &= \int_{-\infty}^{\infty} K_h\left(t - F_{0n_0}(v)\right) \mathrm{d}F_{1n_1}(v) = \frac{1}{n_1}\sum_{j=1}^{n_1} K_h\left(t - F_{0n_0}(X_{1j})\right) \\ &= \left(K_h * F_{1n_1}F_{0n_0}^{-1}\right)(t)\end{aligned}$$

e

$$\begin{aligned}\hat{r}_{h,h_0}(t) &= \int_{-\infty}^{\infty} K_h\left(t - \tilde{F}_{0h_0}(v)\right) \mathrm{d}F_{1n_1}(v) = \frac{1}{n_1}\sum_{j=1}^{n_1} K_h\left(t - \tilde{F}_{0h_0}\left(X_{1j}\right)\right) \\ &= \left(K_h * F_{1n_1}\tilde{F}_{0h_0}^{-1}\right)(t).\end{aligned}$$

Baixo certas condicións faise na sección 2.2 un estudo minucioso do erro cadrático medio de $\hat{r}_h(t)$ e $\hat{r}_{h,h_0}(t)$ que nos permite obter a expresión da ventá asintoticamente óptima, h_{AMISE}. Como xa sucede noutros contextos de estimación non paramétrica de curvas, o valor de h_{AMISE} non pode ser usado directamente porque depende de funcionais descoñecidos, en concreto, funcionais dependentes da curva que desexamos estimar, neste caso $r(t)$. Porén, a expresión da ventá asintoticamente óptima permítenos propor dous selectores plug-in, h_{SJ_1} e h_{SJ_2}. A proposta de selectores plug-in de h para $\hat{r}(t)$ e $\hat{r}_{h,h_0}(t)$ é estudado en detalle na sección 2.3. Mais en concreto, na subsección 2.3.1 estudamos a estimación tipo núcleo dos funcionais descoñecidos que aparecen na expresión de h_{AMISE}. Despois de obter unha descomposición asintótica do erro cadrático medio dos estimadores tipo núcleo propostos para tales funcionais, obtéñense expresións para as ventás asintoticamente óptimas. Baseándonos nos resultados obtidos na sección e na subsección previas, a selección de h e en concreto os algoritmos de cálculo de h_{SJ_1} e h_{SJ_2} detállanse na subsección 2.3.2. O estudo de simulación da subsección 2.3.3 permite comparar e evaluar o comportamento práctico destes dous selectores e unha variante dun selector previamente proposto neste mesmo contexto por Ćwik and Mielniczuk (1993), b_{3c}. Os resultados obtidos na simulación permiten concluír que o comportamento de h_{SJ_1} mellora o comportamento de b_{3c} e que h_{SJ_2} destácase como o mellor dos tres.

Así mesmo, baseándonos na expresión cerrada que se obtén para o erro cadrático medio integrado do estimador de $r(t)$, $\hat{r}_h(t)$, propoñemos catro selectores da ventá baseados na técnica bootstrap, h^*_{CE}, h^*_{MC}, h^*_{SUMC} e h^*_{SMC}. Un estudo de simulación revela que os selectores h^*_{CE} e h^*_{MC} non presentan un bo comportamento práctico en mostras pequenas, ademais de requerir moita carga computacional. Os outros dous selectores bootstrap, h^*_{SUMC} e h^*_{SMC}, si que presentan un comportamento máis favorable, pero de novo, a súa elevada carga computacional, e a escasa mellora que só presentan nalgúns casos, cando os comparamos co selector plug-in h_{SJ_2}, fan deste último selector, a mellor elección.

No capítulo 3 tratamos a estimación non paramétrica da densidade e distribución relativas con datos truncados pola esquerda e censurados pola dereita. Máis en concreto, estudamos o erro cadrático medio e a normalidade asintótica dos estimadores de $r(t)$ e $R(t)$ considerados neste contexto, $\check{r}_h(t)$ e $\check{R}_h(t)$:

$$\check{r}_h(t) = \frac{1}{h}\int K\left(\frac{t-\hat{F}_{0n_0}(y)}{h}\right)\mathrm{d}\hat{F}_{1n_1}(y),$$

$$\check{R}_h(t) = \int \mathbb{K}\left(\frac{t-\hat{F}_{0n_0}(y)}{h}\right)\mathrm{d}\hat{F}_{1n_1}(y),$$

onde $\hat{F}_{0n_0}(y)$ e $\hat{F}_{n_1}(y)$ denotan o estimador límite produto proposto por Tsai *et al* (1987) de F_0 e F_1, respectivamente.

É importante sinalar que o caso da estimación non paramétrica de $r(t)$ é estudado en máis detalle que a de $R(t)$. Así, inclúense tres propostas, h_{RT}, h_{PI} e h_{STE}, para a selección da ventá de $\check{r}_h(t)$, todas elas baseadas na técnica plug-in. Mediante un estudo de simulación e baixo diferentes porcentaxes de censura e truncamento compróbase e compárase o comportamento dos tres selectores propostos. Deste estudo, a conclusión que se desprende, é que o selector h_{RT}, baseado na técnica do polgar, é o que mellor se comporta.

O capítulo 4 ten como obxeto a construción de intervalos de confianza para a distribución relativa no caso de dúas mostras independentes de datos suxeitos a censura pola dereita e truncamento pola esquerda. Utilizando a técnica de verosimilitude empírica, conséguese estender o teorema de Wilks a este contexto e o traballo de Claeskens *et al* (2003) que trata o mesmo problema pero con datos completamente observados.

O capítulo 5 inclúe dúas aplicacións a datos reais. A primeira delas recolle 1434 pacientes que foron sometidos a unha biopsia no Hospital Juan Canlejo entre xaneiro de 2002 e setembro de 2005, a cal permitiu determinar se o paciente presentaba cancro de próstata ou non. Para cada paciente, rexistráronse tamén os valores de tPSA, fPSA e cPSA. Segundo estudos realizados neste mesmo campo, parece que estas variables ofrecen boas propiedades diagnósticas. A segunda base de datos analizada recolle datos de 1956 pacientes diagnosticados con cancro de estómago. Entre outras variables medidas en cada individuo, destacamos o sexo, a idade e a presenza de metástase no momento do diagnóstico, o tempo de vida desde a data de primeiros síntomas e a demora diagnóstica. Mentres a primeira base de datos é un claro exemplo de datos completos, a segunda base recolle datos truncados pola esquerda e censurados pola dereita. Despois dunha análise distribucional de ambas bases de datos, as conclusións mais destacables son que as variables tPSA e cPSA presentan mellores propiedades diagnósticas que a variable fPSA e que o tempo de vida desde a data de primeiros síntomas de cancro de estómago é menor en grupos de idade maior e naqueles pacientes que mostran metástase no momento do diagnóstico.

No capítulo 6 faise unha exposición de varias futuras liñas de investigación que gardan estreita relación co tema principal desta tese, técnicas de redución do nesgo en problemas con dúas mostras, contraste da hipótese de riscos proporcionais ao longo do tempo de vida, bandas de confianza obtidas mediante verosimilitude empírica e a distribución relativa no caso de datos truncados pola dereita e censurados pola esquerda e, finalmente, a extensión do concepto de curvas relativas ó caso de problemas multiestado. No que se refire aos dous primeiros destes futuros traballos, inclúense a súa motivación, a estratexia que se vai seguir na demostración e algúns cálculos ao respecto, asi como a bibliografía consultada ata o momento.

Finalmente, incluimos un apéndice onde se recompilan definicións, teoremas e desigualdades que foron utilizadas ao longo da tese. Tamén se inclúe ao final deste documento, unha lista bibliográfica con todas as referencias consultadas durante a elaboración deste traballo de investigación.

Bibliography

Aalen, O.O. (1975). Statistical inference for a family of counting processes. PhD thesis, University of California, Berkeley.

Ahmad, I.A. (2002). On moment inequalities of the supremum of empirical processes with applications to kernel estimation, *Statistics & Probability Letters*, **57**, 215–220.

Akritas, M.G. (1994). Nearest neighbor estimation of a bivariate distribution under random censoring, *Annals of Statistics*, **22**, 1299–1327.

Alioum, A. and Commenges D. (1996). A proportional hazards model for arbitrarily censored and truncated data, *Biometrics*, **52**, 512–524.

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

Beran, R. (1981). Nonparametric regression with randomly censored data. Technical Report. University of California, Berkeley.

Cai, T. (2004). Semi-parametric ROC regression analysis with placement values, *Biostatistics*, **5**, 45–60.

Cao, R. (1993). Bootstrapping the mean integrated squared error, *Journal of Multivariate Analysis*, **45**, 137–160.

Cao, R., Cuevas, A. and González-Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation, *Computational Statistics & Data Analysis*, **17**, 153–176.

Cao, R., Janssen, P. and Veraverbeke, N. (2000). Relative density estimation with censored data, *The Canadian Journal of Statistics*, **28**, 97–111.

Cao, R., Janssen, P. and Veraverbeke, N. (2001). Relative density estimation and local bandwidth selection for censored data, *Computational Statistics & Data Analysis*, **36**, 497–510.

Cao, R., Janssen, P. and Veraverbeke, N. (2005). Relative hazard rate estimation for right censored and left truncated data, *Test*, **14**, 257–280.

Cao, R. and Van Keilegom, I. (2006). Empirical likelihood tests for two-sample problems via nonparametric density estimation, *Scandinavian Journal of Statistics*, **34**, 61–77

Chen, Q. and Wang, Q. (2006). On kernel estimation of lifetime distribution function with left truncated and right censored data. Technical Report. Chinese Academy of Science, Beijing (China).

Claeskens, G., Jing B.Y., Peng, L. and Zhou, W. (2003). Empirical likelihood confidence regions for comparison distributions and ROC curves, *The Canadian Journal of Statistics*, **31**, 173–190.

Cox, D.R. (1972). Regression models and life-tables, *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**, 187–220.

Cox, D.R. (1975). Partial likelihood, *Biometrika*, **62**, 269–276.

Ćwik, J. and Mielniczuk, J. (1993). Data-dependent bandwidth choice for a grade density kernel estimate, *Statistics & Probability Letters*, **16**, 397–405.

Dabrowska, D. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate, *Annals of Statistics*, **17**, 1157–1167.

Duin, R.P.W. (1976). On the choice of smoothing parameters of Parzen estimators of probability density functions, *IEEE Transactions on Computers*, **C-25**, 1175–1179.

Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator, *The Annals of Mathematical Statistics*, **27**, 642–669.

Efron, B. (1979). Bootstrap methods: another look at the jackknife, *The Annals of Statistics*, **7**, 1–26.

Efron, B. (1981). Censored data and the bootstrap, *Journal of the American Statistical Association*, **76**, 312–319.

Faraway, J.J. and Jhun, M. (1990). Bootstrap choice of bandwidth for density estimation, *Journal of the American Statistical Association*, **85**, 1119–1122.

Gastwirth, J.L. (1968). The first-median test: A two-sided version of the control median test, *Journal of the American Statistical Association*, **63**, 692–706.

Gibbons, J.D. and Chakraborti, S. (1991). Nonparametric statistical inference. Third edition, revised and expanded. Statistics: textbooks and monographs, v. 131, Dallas, Texas.

Gijbels, I. and Wang, J.L. (1993). Strong representations of the survival function estimator for truncated and censored data with applications, *Journal of Multivariate Analysis*, **47**, 210–229.

González-Manteiga, W. and Cadarso-Suárez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications, *Journal of Nonparametric Statistics*, **4**, 65–78.

Gross, S.T. and Lai T.L. (1996). Bootstrap methods for truncated and censored data, *Statistica Sinica*, **6**, 509–530.

Guarneri, V., Lehihan, D.J., Valero, V., Durand, J.B., Broglio, K., Hess, K.R., Boehnke, L., González-Angulo, A.M., Hortobagyi, G.N. and Esteva, F.J. (2006). Long-term cardiac tolerability of trastuzumab in metastatic breast cancer: the M.D. Anderson Cancer Center experience, *Journal of Clinical Oncology*, **24**, 4107–4115.

Habbema, J.D.F., Hermans, J. and van den Broek, K. (1974). A stepwise discrimination analysis program using density estimation. *Compstat 1974: Proceeding in Computational Statistics. Physica Verlag, Vienna.*

Hall, P.G. (1984). Central limit theorem for integrated squared error of multivariate nonparametric density estimators, *Journal of Multivariate Analysis*, **14**, 1–16.

Hall, P.G. (1992). The bootstrap and Edgeworth expansion. Springer Series in Statistics, New York.

Hall, P.G. and Hyndman, R.J. (2003). Improved methods for bandwidth selection when estimating ROC curves, *Statistics & Probability Letters*, **64**, 181–189.

Hall, P.G. and Marron, J.S. (1987). Estimation of integrated squared density derivatives, *Statistics & Probability Letters*, **6**, 109–115.

Hall, P.G., Marron, J.S. and Park, B.U. (1992). Smoothed cross-validation, *Probability Theory and Related Fields*, **92**, 1–20.

Handcock, M.S. and Janssen, P. (1996). Statistical inference for the relative distribution. Technical Report. Department of Statistics and Operations Research, New York University, New York.

Handcock, M.S. and Janssen, P. (2002). Statistical inference for the relative density, *Sociological Methods & Research*, **30**, 394–424.

Handcock, M.S. and Morris, M. (1999). Relative distribution methods in social sciences. Springer, New York.

Hjort, N.L. and Walker, S.G. (2001). A note on kernel density estimators with optimal bandwidths, *Statistics & Probability Letters*, **54**, 153–159.

Holmgren, E.B. (1995). The P-P plot as a method for comparing treatment effects, *Journal of the American Statistical Association*, **90**, 360–365.

Hsieh, F. (1995). The empirical process approach for semiparametric two-sample models with heterogeneous treatment effect, *Journal of the Royal Statistical Society. Series B*, **57**, 735–748.

Hsieh, F. and Turnbull, B.W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve, *The Annals of Statistics*, **24**, 25–40.

Iglesias-Pérez, M.C. and González-Manteiga, W. (1999). Strong representation of a generalized product-limit estimator for truncated and censored data with some applications, *Journal of Nonparametric Statistics*, **10**, 213–244.

Janssen, P., Swanepoel, J.W.H. and Veraverbeke N. (2004). Modifying the kernel distribution estimator towards reduced bias, *IAP statistics network. Technical Report*, **0458**.

Jones, M.C. (1990). The performance of kernel density functions in kernel distribution function estimation, *Statistics and Probability Letters*, **9**, 129–132.

Kakizawa, Y. (2004). Bernstein polynomial probability density estimation, *Journal of Nonparametric Statistics*, **16**, 709–729.

Kaplan, E.L. and Meier, P. (1958). Non-parametric estimation from incomplete observations, *Journal of the American Statistical Association*, **53**, 457–481.

Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent random variables, and the sample distribution function, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **32**, 111–131.

La Valley, M.P. and Akritas, M.G. (1994). Extensions of the Lynden-Bell-Woodroofe method for truncated data. Unpublished manuscript.

Lai, T.L. and Ying, Z. (1991). Estimating a distribution with truncated and censored data, *Annals of Statistics*, **19**, 417–442.

Lein, M., Kwiatkowski, M., Semjonow, A., Luboldt, H.J., Hammerer, P., Stephan, C., Klevecka, V., Taymoorian, K., Schnorr, D., Recker, F., Loening, S.A. and Jung, K. (2003). A multicenter clinical trial on the use of complexed prostate-specific antigen in low prostate specific antigen concentrations, *The Journal of Urology*, **170**, 1175–1179.

Li, L. (1995). Almost sure representations and two sample problems with left truncated and right censored data. PhD thesis. Columbia University.

Li, G., Tiwari, R.C. and Wells, M.T. (1996). Quantile comparison functions in two-sample problems with applications to comparisons of diagnostic markers, *Journal of the American Statistical Association*, **91**, 689–698.

Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars, *Monthly Notices of the Royal Astronomy Society*, **155**, 95–118.

Lloyd, C.J. (1998). Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems, *Journal of the American Statistical Association*, **93**, 1356–1364.

Lloyd, C.J. and Yong, Z. (1999). Kernel estimators are better than empirical, *Statistics & Probability Letters*, **44**, 221–228.

Major, P. and Rejto, L. (1988). Strong embedding of the estimator of the distribution function under random censorship, *Annals of Statistics*, **16**, 1113–1132.

Mann, H.B. and Whitney, D.R. (1947). On a test whether one of two random variables is stochastically larger than the other, *Annals of Mathematical Statistics*, **18**, 50–60.

Marron, J.S. and Wand, M.P. (1992). Exact mean integrated squared errors, *Annals of Statistics*, **20**, 712–736.

Massart P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality, *The Annals of Probability*, **18**, 1269–1283.

Molanes-López, E.M. and Cao, R. (2006a). Plug-in bandwidth selector for the kernel relative density estimator. To appear in *Annals of the Institute of Statistical Mathematics*.

Molanes-López, E.M. and Cao, R. (2006b). Relative density estimation for left truncated and right censored data. Unpublished manuscript.

Nadaraya, E.A. (1964). On estimating regression, *Theory of Probability and its Applications*, **9**, 141–142.

Okihara, K., Cheli, C.D., Partin, A.W., Fritche, H.A., Chan, D.W., Sokoll, L.J., Brawer, M.K., Schwartz, M.K., Vesella, R.L., Loughlin, K.R., Johnston, D.A. and Babaian, R.J. (2002). Comparative analysis of complexed prostate specific antigen, free prostate specific antigen and their ratio in detecting prostate cancer, *The Journal of Urology*, **167**, 2017–2024.

Owen, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, **75**, 237–249.

Owen, A.B. (1990). Empirical likelihood confidence regions, *The Annals of Statistics*, **18**, 90–120.

Owen, A.B. (2001). Empirical likelihood. Chapman & Hall, Boca Raton, FL.

Partin, A.W., Brawer, M.K, Bartsch, G., Horninger, W., Taneja, S.S., Lepor, H., Babaian, R., Childs S.J., Stamey, T., Fritsche, H.A., Sokoll, L., Chan, D.W., Thiel, R.P. and Chely, C.D. (2003). Complexed prostate specific antigen improves specificity for prostate cancer detection: results of a prospective multicenter clinical trial, *The Journal of Urology*, **170**, 1787–1791.

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

Pepe, M.S. and Cai, T. (2004). The analysis of placement values for evaluating discriminatory measures, *Biometrics*, **60**, 528–535.

Petrov, V. (1995). Limit theorems of Probability Theory. Sequences of independent random variables. Clarendon Press, Oxford.

Polansky, A.M. and Baker, E.R. (2000). Multistage plug-in bandwidth selection for kernel distribution function estimates, *Journal of Statistical Computation & Simulation*, **65**, 63–80.

Rosenblatt, M. (1956). Remarks on some non-parametric estimates of a density function, *Annals of Mathematical Statistics*, **27**, 642–669.

Sánchez-Sellero, C., González-Manteiga, W. and Cao, R. (1999). Bandwidth selection in density estimation with truncated and censored data, *Annals of the Institute of Statistical Mathematics*, **51**, 51–70.

Scott, D.W. and Terrell, G.R. (1987). Biased and unbiased cross-validation in density estimation, *Journal of the American Statistical Association*, **82**, 1131–1146.

Sheather, S.J. and Jones, M.C. (1991). A reliable data-based bandwidth selection method for kernel density estimation, *Journal of the Royal Statistical Society. Series B*, **53**, 683–690.

Silverman, B.W. (1978). Density ratios, empirical likelihood and cot death, *Applied Statistics*, **27**, 26–33.

Silverman, B.W. (1986). Density estimation. Chapman and Hall, London.

Smirnov, N.V. (1939). Estimate of deviation between empirical distribution functions in two independent samples (in Russian), *Bulletin of Moscow University*, **2**, 3–16.

Stine, R.A. and Heyse, J.F. (2001). Nonparametric estimates of overlap, *Statistics in Medicine*, **20**, 215–236.

Sun, L. and Zhou, Y. (1998). Sequential confidence bands for densities under truncated and censored data, *Statistics and Probability Letters*, **40**, 31–41.

Swanepoel, J.W.H. and Van Graan F.C. (2005). A new kernel distribution function estimator based on a non-parametric transformation of the data. *Scandinavian Journal of Statistics*, **32**, 551–562.

Taylor, C.C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation *Biometrika*, **76**, 705–712.

Thomas, D.R. and Grunkemeier, G.L. (1975). Confidence interval estimation of survival probabilities for censored data, *Journal of the American Statistical Association*, **70**, 865–871.

Tsai, W.Y. (1990). Testing the assumption of independence of truncation time and failure, *Biometrika*, **77**, 169–177.

Tsai, W.Y., Jewell, N.P. and Wang, M.C. (1987). A note on the product limit estimator under right censoring and left truncation, *Biometrika*, **74**, 883–886.

Tse, S. (2003). Strong gaussian approximations in the left truncated and right censored model, *Statistica Sinica*, **13**, 275–282.

Van Keilegom, I. and Veraverbeke N. (1997). Estimation and bootstrap with censored data in fixed design nonparametric regression, *Annals of the Institute of Statistical Mathematics*, **49**, 467–491.

Wald, A. and Wolfowitz, J. (1940). On a test whether two samples are from the same population, *Annals of Mathematical Statistics*, **11**, 147–162.

Wand, M.P. and Jones, M.C. (1995). Kernel Smoothing. Chapman and Hall, London.

Zhou, Y. (1996). A note on the TJW product limit estimator for truncated and censored data, *Statistics & Probability Letters*, **26**, 381–387.

Zhou, Y. and Yip, P.S.F. (1999). A strong representation of the product-limit estimator for left truncated and right censored data, *Journal of Multivariate Analysis*, **69**, 261–280.

Zhou, Y., Wu, G.F. and Jiang, X.L. (2003). Asymptotic behavior of the Lipschitz-1/2 modulus of the PL-process for truncated and censored data, *Acta Mathematica Sinica, English Series*, **19**, 729–738.

Zhu, Y. (1996). The exponential bound of the survival function estimator for randomly truncated and censored data, *Journal of Systems Science and Mathematical Sciences*, **9**, 175–181.