Bandwidth selection for kernel density estimation with length-biased data

M.I. Borrajo[a]*, W. González-Manteiga[a] and M.D. Martínez-Miranda[b]

[a]Departmento de Estatística, Análise Matemática e Optimización, Universidade de Santiago de Compostela, Spain;

[b]Cass Business Shool, City University of London, 106 Buhill Row, EC1Y 8TZ London (UK)

Abstract

Length-biased data are a particular case of weighted data, which arise in many situations: biomedicine, quality control or epidemiology among others. In this paper we study the theoretical properties of kernel density estimation in the context of length-biased data, proposing two consistent bootstrap methods that we use for bandwidth selection. Apart from the bootstrap bandwidth selectors we suggest a rule-of-thumb. These bandwidth selection proposals are compared with a least-squares cross-validation method. A simulation study is accomplished to understand the behaviour of the procedures in finite samples.

Keywords: Bootstrap, Rule-of-thumb, Cross-validation, Non-parametric, Weighted data.
AMS Subject Classification: 62G07; 62G09; 62G20.

1 Introduction

In general a sample is supposed to have the same basic characteristics as the population it represents. However, in practice it is usual that the sample may not be completely representative of the population, and bias is introduced in the sampling scheme, we known them as weighted data. This type of samples is produced when the probability of choosing an observation depends on its value and/or other covariates of interest. Weighted data arise in many sampling processes, see Patil and Rao (1977), and also in a wide variety of fields such as biomedicine, Chakraborty and Rao (2000), epidemiology, Simon (1980), textile fibres, Cox (2005), as well as social sciences, economics, Heckman (1990), or quality control.

*Corresponding author. Email: mariaisabel.borrajo@usc.es

Some specific examples are the visibility bias problem that arises when using aerial survey techniques to estimate, for instance, wildlife population density; or a damage model where an observation may be damaged by a process depending on the variable and then the observed data are clearly biased. Also the textile fibres problem is a classical motivating example.

Let us denote by f the density function of an unobserved random variable X, and sassume that the available information refers to a closely related random variable Y with weighted or biased distribution determined by the density function:

$$f_{Y,\omega}(y) = \frac{\omega(y)f(y)}{\mu_\omega} \qquad y > 0,$$

where ω is a known function and $\mu_\omega = \int \omega(x)f(x)dx < \infty$.

A particular case of weighted data is the length-biased data, where the probability of an observation to be sampled is directly proportional to its value in a simple linear way. In this case the weight function that determines the bias is the identity function, i.e., $\omega(y) = y$. This sort of data are quite common in problems related to renewal processes, epidemiological cohort studies or screening programs for the study of chronic diseases, see Zelen and Feinleib (1969).

Cox (2005) proposed an estimator for the mean and another for the distribution function in the context of weighted data. Vardi (1982, 1985) showed that this last estimator was the maximum likelihood estimator of the distribution function under weighted sampling and that the estimation of the mean is $\sqrt{n}-$consistent. Density estimation for this type of data started in the 80's when Bhattacharyya and Richardson (1988) defined the first density estimator for length-biased data based on the problem of fibres, which was continued with theoretical developments in Richardson et al. (1991). Furthermore, Jones (1991) proposed a modification of the common kernel density estimator adapted to length-biased data which is widely used. In the same paper he showed that this proposal has some advantages over the previous one, and better asymptotic properties. Ahmad (1995) extended to the multivariate case these two kernel density estimators. Another extensions using Fourier series have been proposed in Jones and Karunamuni (1997). Later a third non-parametric estimator has been considered in Guillamón et al. (1998).

Density estimation for weighted data has also been studied from other points of view, Barmi and Simonoff (2000) proposed a simple transformation-based approach motivated by the form of the non-parametric maximun likelihood estimator of the density. Efromovich (2004) presented asymptotic results on sharp minimax density estimation. Projection methods are developed in Brunel et al. (2009). Asgharian et al. (2002) and de Uña-Álvarez (2004) studied the problem under the common settings of surviva analysis. Also wavelet theory has been used in this context, see Chesneau (2010) which constructed an adaptative estimator based on the BlockShrink algorithm and Ramírez and Vidakovic (2010) which applied dyadic wavelet density estimation. Cutillo et al. (2014) proposed linear and non linear wavelet density estimators and recently Comte and Rebafka (2016) defined the estimation through out the distribution function and using a known link function.

The use of non-parametric methods implies to choose a bandwidth parameter, which determines the degree of smoothness to be considered in the estimation. The choice of the bandwidth parameter is crucial and it has motivated several papers in the literature in the recent decades. Marron (1988), Scott (1992) and Silverman (1986) provide a full description of the problem as well as a review of several bandwidth selection methods. Later methods such as plug-in or bootstrap methods, have been defined in Hall and Marron (1987), Sheather and Jones (1991) and Marron (1992). Fourier transforms have also been used in this context, see Chiu (1992). To explore the most relevant bandwidth selection methods in density estimation for complete data see the reviews of Turlach (1993), Cao et al. (1994), Jones et al. (1996) or Heidenreich et al. (2013), and the recent work on local linear density estimation by Mammen et al. (2011, 2014).

This paper is organised as follows. In Section 2 we develop asymptotic theory for the kernel density estimator of Jones (1991) for length-biased data, and we also define two different consistent bootstrap procedures. In Section 3 we propose new data-driven bandwidth selection methods: a rule-of-thumb based on the Normal distribution and two bootstrap bandwidth selectors based on the procedures presented in the previous section. These proposals are competitors of a cross-validation method which, to the extent of our knowledge, is the only existent data-driven bandwidth selector in this context. In Section 4, we carry out an extensive simulation study to evaluate the performance of the presented bandwidth selectors for finite samples. We draw some conclusions in Section 5. Final remarks are given in Section 6 as well as a discussion of how the methodology developed in this paper can be generalised to a widespread weight function. Finally we add in the appendix the proofs of the theoretical results.

2 Theoretical developments

Hereafter we will work under the scenario of the length-biased data even though all the results can be generalised to the weighted data case under appropriate assumptions, see final remarks in Section 6.

Hence, let us write the density function of the observed variable Y as

$$f_Y(y) = \frac{yf(y)}{\mu}, \qquad y > 0,$$

with $\mu = \int yf(y)dy$.

Let $Y_1, \ldots, Y_n$ be an i.i.d. (independent identically distributed) sample from f_Y, Jones (1991) defined the following kernel density estimator based on the structure of the one proposed in Parzen (1962) and Rosenblatt (1956):

$$\hat{f}_h(y) = \frac{1}{n}\hat{\mu}\sum_{i=1}^{n}\frac{1}{Y_i}K_h(y - Y_i), \tag{1}$$

where $\hat{\mu} = \left(\frac{1}{n}\sum_{i=n}^{n}\frac{1}{Y_i}\right)^{-1}$, see Cox (2005) to find out this estimation, and $K_h(\cdot) = \frac{1}{h}K(\frac{\cdot}{h})$, with K being a symmetric kernel function.

In the following result we obtain the value of the pointwise mean and variance of $\hat{f}_h$ with the corresponding error rates, as well as its mean squared error (MSE), which is defined as:

$$\mathrm{MSE}(h,y) = E\left[(\hat{f}_h(y) - f(y))^2\right].$$

We need to introduce the following hypotheses:

(A.1) $E\left[\frac{1}{X}\right] < +\infty$, $E\left[\frac{1}{Y^{2\nu}}\right] < +\infty$ where $\nu \in \mathbb{N}$, $\nu \geq 3$,

(A.2) $\int K(u)du = 1$, $\int uK(u)du = 0$ and $\mu_2(K) < +\infty$,

(A.3) $\lim_{n\to\infty} nh = +\infty$,

(A.4) y a continuity point of f,

(A.5) f has two continuous derivates,

(A.6) K is twice differentiable.

Theorem 2.1. *Under conditions (A.1) to (A.4) we have:*

$$E\left[\hat{f}_h(y)\right] = (K_h \circ f)(y) + O\left(\frac{1}{n}\right) \quad \textit{and}$$

$$Var\left[\hat{f}_h(y)\right] = n^{-1}\left[\left(K_h^2 \circ \gamma\right)(y) - \left(K_h \circ f\right)^2(y)\right] + O\left(\frac{1}{n}\right),$$

where $\circ$ denotes the convolution between two functions and $\gamma(y) = \mu f(y)/y$. Moreover, adding condition (A.5), we have:

$$\mathrm{MSE}(h,y) = \frac{1}{4}h^4\left(f''(y)\right)^2\mu_2^2(K) + \frac{\gamma(y)}{nh}R(K) + o\left(h^4 + \frac{1}{nh}\right), \tag{2}$$

where $\mu_2(K) = \int u^2K(u)du$ and $R(K) = \int K^2(u)du$.

Now, defining the mean integrated squared error (MISE) as

$$\mathrm{MISE}(h) = E\int\left(\hat{f}_h(y) - f(y)\right)^2 dy, \tag{3}$$

and denoting by AMISE its asymptotic version, the following result is a consequence of Theorem 2.1.

Corollary 2.2. *Under conditions (A.1), (A.2), (A.3) and (A.5),*

$$\mathrm{MISE}(h) = \frac{1}{4}h^4\mu_2^2(K)R(f'') + \frac{R(K)\mu c}{nh} + o\left(h^4 + \frac{1}{nh}\right),$$

$$\mathrm{AMISE}(h) = \frac{1}{4}h^4\mu_2^2(K)R(f'') + \frac{R(K)\mu c}{nh}, \quad \textit{with} \quad c = \int \frac{1}{y}f(y)dy.$$

As a consequence, the optimal bandwidth value which minimises AMISE(h) *is:*

$$h_{\mathrm{AMISE}} = \left(\frac{R(K)\mu c}{n\mu_2^2(K)R(f'')}\right)^{1/5}. \tag{4}$$

2.1 Resampling bootstrap methods

In this section we develop two different bootstrap procedures that can be applied in the context of length-biased data. Both of them are consistent in the way it is shown below and they conform the basis to define different data-driven bandwidth selection methods.

2.1.1 Bootstrapping using Jones' estimator

In this first mehtod we follow the work by Cao (1990, 1993) using the so-called smooth bootstrap to develop a bandwidth selector for the kernel density estimator of Jones (1991), given in (1). It is remarkable that one bootstrap bandwidth selector can be implemented in practice without requiring resampling and any Monte Carlo approximation.

Given an i.i.d. sample, $Y_1, \ldots, Y_n$ from f_Y, and $\hat{f}_g$ the density estimator introduced in (1) with pilot bandwidth g, the smooth bootstrap samples, $Y_1^*, \ldots, Y_n^*$, are generated by sampling randomly with replacement n times from the estimated density $\hat{f}_{Y,g}(y) = y\hat{f}_g(y)/\hat{\mu}$.

Let Y^* denote the random variable generated by the bootstrap method presented above. From the bootstrap sample let define the bootstrap density estimator of Y^* as

$$\hat{f}_h^*(y) = \frac{1}{n}\hat{\mu}^* \sum_{i=1}^{n} \frac{1}{Y_i^*} L_h\left(y - Y_i^*\right), \tag{5}$$

where $\hat{\mu}^* = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i^*}\right)^{-1}$, and $L_h(\cdot) = \frac{1}{h}L(\frac{\cdot}{h})$, with L being a symmetric kernel function like K.

The following result provides the expression of the mean, the variance and the mean squared error of $\hat{f}_h^*(y)$ under the bootstrap distribution. We use the notation E^*, Var^* and MSE* to refer to the bootstrap distribution.

Theorem 2.3. *Under conditions (A.1) to (A.4)*

$$E^*\left[\hat{f}_h^*(y)\right] = \left(L_h \circ \hat{f}_g\right)(y) + O_P\left(\frac{1}{n}\right) \quad \textit{and}$$
$$Var^*\left[\hat{f}_h^*(y)\right] = n^{-1}\left[\left(L_h^2 \circ \hat{\gamma}_g\right)(y) - \left(L_h \circ \hat{f}_g\right)^2(y)\right] + O_P\left(\frac{1}{n}\right).$$

Moreover, adding condition (A.6), we obtain

$$\mathrm{MSE}^*(h,y) = \frac{1}{4}h^4\left(\hat{f}_g''(y)\right)^2\mu_2^2(L) + \frac{\hat{\gamma}_g(y)}{nh}R(L) + o_P\left(h^4 + \frac{1}{nh}\right),$$

where $\hat{\gamma}_g(y) = \hat{\mu}\hat{f}_g(y)/y$, $\mu_2(L) = \int u^2 L(u)du$ *and* $R(L) = \int L^2(u)du$.

The same way we have done in Corollary 2.2, the integrated versions of the MSE*, MISE* and its asymptotic version are easily deduced from the theorem above.

Corollary 2.4. *Under hypothesis (A.2), (A.3), (A.4) and (A.6)*

$$\text{MISE}^*\left(h\right) = \frac{1}{4}h^4\mu_2^2(L)R(\hat{f}_g'') + \frac{R(L)\hat{\mu}\hat{c}}{nh} + o_P\left(h^4 + \frac{1}{nh}\right),$$

$$\text{AMISE}^*(h) = \frac{1}{4}h^4\mu_2^2(L)R(\hat{f}_g'') + \frac{R(L)\hat{\mu}\hat{c}}{nh}, \qquad \text{with } \hat{c} = \hat{\mu}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i^2}.$$

Therefore, the asymptotic expression of the optimal bootstrap bandwidth is:

$$h_{\text{AMISE}^*} = \left(\frac{R(L)\hat{\mu}\hat{c}}{n\mu_2^2(L)R(\hat{f}_g'')}\right)^{1/5},$$

which is a plug-in version of (4).

The following corollary is a consequence of the previous results.

Corollary 2.5. *Under assumptions (A.1) to (A.4),* MISE* *and* AMISE* *are consistent estimators of* MISE *and* AMISE*, respectively.*

2.1.2 Bootstrapping using a common kernel density estimator

In this second method, we are also using a smooth bootstrap procedure in which we will use a common kernel density estimator instead of Jones'. The idea of defining this second procedure is to be able to use the methodology developed in Bose and Dutta (2013) for bandwidth selection in kernel density estimation.

Given an i.i.d. sample, $Y_1, \ldots, Y_n$ from f_Y, and denote by $\tilde{f}_{K,g}$ the common kernel density estimator with pilot bandwidth g and a kernel function K, the smooth bootstrap samples, $Y_1^*, \ldots, Y_n^*$, are generated by sampling randomly with replacement n times from $\tilde{f}_{K,g}$. Let Y^* denote again the random variable generated by the bootstrap method presented above. From the bootstrap sample let us define the bootstrap density estimator of Y^* as the one presented in (5), taking into account that the bootstrap sample is generated differently.

Now we provide the expression for the point wise mean and variance of $\hat{f}_h^*(y)$ under this bootstrap distribution.

Theorem 2.6. *Under conditions (A.1) to (A.4)*

$$E^*\left[\hat{f}_h^*(y)\right] = \frac{1}{\int \frac{1}{z}\tilde{f}_{K,g}(z)dz}\int \frac{1}{z}L_h(y-z)\tilde{f}_{K,g}(z)dz + O_P\left(\frac{1}{n}\right) \quad \textit{and}$$

$$Var^*\left[\hat{f}_h^*(y)\right] = \frac{1}{n\left(\int \frac{1}{z}\tilde{f}_{K,g}(z)dz\right)^2}\left[\int \frac{1}{z^2}L_h^2(y-z)\tilde{f}_{K,g}(z)dz - \int \frac{1}{z}L_h(y-z)\tilde{f}_{K,g}(z)dz^2\right] + \\ + O_P\left(\frac{1}{n}\right).$$

Moreover,

$$\begin{aligned}\mathrm{MSE}^*(h,y) = & \left(\frac{\int \frac{1}{z} L_h(y-z)\tilde{f}_{K,g}(z)dz}{\int \frac{1}{z}\tilde{f}_{K,g}(z)dz} - \hat{f}_h(y)\right)^2 + \\ & + \frac{1}{n\left(\int \frac{1}{z}\tilde{f}_{K,g}(z)dz\right)^2}\left[\int \frac{1}{z^2} L_h^2(y-z)\tilde{f}_{K,g}(z)dz - \int \frac{1}{z} L_h(y-z)\tilde{f}_{K,g}(z)dz^2\right] + \\ & + O_P\left(\frac{1}{n}\right).\end{aligned}$$

Remark that for this bootstrap method we do not get manageable explicit expressions of the error criteria as we got in the previous one; and the way to obtain MISE$^*(h)$ *is integrating the expression above, but we neither obtain an explicit expression.*

3 Bandwidth selection

In this section we describe bandwidth selection methods for the density estimator defined in (1). These methods consist of adaptations of common automatic selectors for kernel density estimation with complete data to the context of length-biased data. We propose a Normal scale rule and two bootstrap selectors derived from the consistent resampling procedures given in the previous section. These proposals are defined as competitors of the cross-validation method proposed in Guillamón et al. (1998).

Two new methods are based on estimating the infeasible optimal expression (4), in which the unknown elements are $R(f'')$, c and μ. However, we have previously shown that these last two terms can be easily estimated, and then the only term that still needs to be estimated is $R(f'')$. The last bootstrap bandwidth selection procedure is based on the minimisation of the $MISE^*(h)$ and does not require those estimations.

3.1 Rule-of-thumb for bandwidth selection

This method is based on the rule-of-thumb, Silverman (1986), for complete data. The idea is to assume that the underlying distribution is Normal, $N(\mu,\sigma)$, and in this situation

$$R(f'') = \frac{3}{8}\pi^{-1/2}\sigma^{-5}.$$

To get a suitable estimator of σ in the context of length-biased data is not trivial. We suggest to estimate it as follows. Cox (2005) states that $E_Y\left[X^r\right] = \frac{\mu_{r+1}}{\mu}$, where μ_{r+1} denotes the (r+1)-th order moment of the original and not observable variable X. So,

$$\mu_2 = \mu E_Y\left[X\right] \Rightarrow \hat{\mu}_2 = \widehat{\hat{\mu} E_Y\left[X\right]} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}\right)^{-1}\left(\frac{1}{n}\sum_{j=1}^{n} Y_j\right),$$

thus,

$$\hat{\sigma}^2 = \hat{\mu}_2 - \hat{\mu}^2 = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}\right)^{-1}\left[\left(\frac{1}{n}\sum_{i=1}^{n}Y_i\right) - \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}\right)^{-1}\right],$$

and finally

$$\hat{h}_{\mathrm{RT}} = \left(\frac{R(K)\hat{\mu}\hat{c}8\sqrt{\pi}}{n\mu_2^2(K)3}\right)^{1/5}\hat{\sigma}.$$

Another possible estimator for σ could be obtained using a robust method such as the interquartile range (IQR)

$$\hat{\sigma}_{\mathrm{IQR}} = \frac{\mathrm{IQR}}{\Phi^{-1}(0.75) - \Phi^{-1}(0.25)},$$

where Φ is the Normal distribution function.

3.2 Cross-validation

The method previously defined is based on minimising estimations of the MISE, more precisely of the AMISE. This procedure relies on the minimisation of the ISE (integrated squared error), the methodology is the same as in Rudemo (1982) and Bowman (1984) applied to (1), and it was developed in Guillamón et al. (1998).

Let write:

$$\mathrm{ISE}(h) = \int\left(\hat{f}_h(z) - f(z)\right)^2 dz = \int \hat{f}_h^2(z)dz - 2\int \hat{f}_h(z)f(z)dz + \int f^2(z)dz. \qquad (6)$$

Note that $\int f^2(z)dz$ does not depend on h, so the minimisation of the ISE is equivalent to minimise the following function:

$$\int \hat{f}_h^2(z)dz - 2\int \hat{f}_h(z)f(z)dz = \int \hat{f}_h^2(z)dz - 2E[\hat{f}_h],$$

which can be estimated by

$$\mathrm{CV}(h) = \int \hat{f}_h^2(z)dz - 2\widehat{E[\hat{f}_h]}.$$

The addends of this estimation may be expressed as follows:

$$\int \hat{f}_h^2(z)dz = \int\left(\frac{1}{nh}\hat{\mu}\sum_{i=1}^{n}\frac{1}{Y_i}K\left(\frac{z-Y_i}{h}\right)\right)\left(\frac{1}{nh}\hat{\mu}\sum_{j=1}^{n}\frac{1}{Y_j}K\left(\frac{z-Y_j}{h}\right)\right)$$

$$= n^{-2}h^{-1}\hat{\mu}^2\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{Y_i}\frac{1}{Y_j}(K\circ K)\left(\frac{Y_i-Y_j}{h}\right)$$

$$\widehat{E[\hat{f}_h]} = \hat{\mu} n^{-1} \sum_{i=1}^{n} \frac{\hat{f}_{-i}(Y_i)}{Y_i} = \hat{\mu} n^{-1} \sum_{i=1}^{n} Y_i^{-1} \left(\sum_{j \neq i} \frac{1}{Y_j} \right)^{-1} \left(\sum_{j \neq i} \frac{1}{Y_j} K_h(Y_i - Y_j) \right),$$

realising that $E[\hat{f}_h] = \int \hat{f}_h(z) f(z) dz = \int \hat{f}_h(z) \frac{\mu f_Y(z)}{z} dz$, and define $\hat{f}_{-i}$ as the estimator in (1) calculated with all the data points except Y_i.

The cross-validation bandwidth is obtained following the proposal in Dutta (2016), where the CV function is minimized in a compact interval of the form $[c_1 \mathrm{IQR} n^{-1/5}, c_2 \mathrm{IQR}(\log(n)/n)^{1/5}]$, where IQR is the inter-quartile range and c_1 and c_2 are positive constants (see Dutta (2016) for the choice of these values). We will be denoted hereafter this bandwidth value as $\hat{h}_{\mathrm{CV}}$.

3.3 Bootstrap for bandwidth selection

3.3.1 Using Jones' estimator

The asymptotic expression of the optimal bootstrap bandwidth can be considered to derive a consistent bandwidth estimate. Cao (1993) suggested such approach for kernel density estimation with complete data. Since all the quantities involved in the expression are known, the result will be a bandwidth estimate which can be computed in practice without involving any resampling and Monte Carlo approximations. The only issue is to determine the pilot bandwidth g involved in the estimation of $R(f'')$. To this goal we first obtain the asymptotical (infeasible) optimal pilot bandwidth and then we propose two feasible estimations.

We define the optimal pilot bandwidth by optimising the MSE of $R(\hat{f}_g'') = \frac{1}{n} \hat{\mu} \sum_{i=1}^{n} \frac{1}{Y_i} \frac{1}{h^3} L'' \left(\frac{y - Y_i}{h} \right)$ as an estimator of $R(f'')$. Let $\hat{f}_g$ be the estimator in (1) with L a symmetric kernel function and assume the following conditions:

(A.7) $\int |u|^3 L(u) du < \infty$,

(A.8) L is twice differentiable, with bounded second derivative and verifies that $\lim_{u \to \pm\infty} u^3 L(u) = 0$, $\lim_{u \to \pm\infty} u^4 L'(u) du = 0$; $\int |u|^4 |L''(u)| du < \infty$, $\int L''^2(u) du < \infty$,

(A.9) $\int u^4 L(u) du < \infty$,

(A.10) f is six times differentiable with f, f'', f''', $f^{(4)} \in L_2(\mathbb{R})$ and verifies the limit condition $\lim_{y \to \pm\infty} f''(y) f'''(y) = 0$.

The result below provides us with the exact value of the optimal pilot bandwidth, in terms of the AMSE (asymptotic mean squared error) of the curvature of Jones' estimator.

Theorem 3.1. *Under hypothesis (A.7) to (A.10) we have that:*

$$\text{AMSE}\left(\int (\hat{f}_g''(y))^2 dy\right) = n^{-2} g^{-10} \left(\int L''(y)^2 dy\right) c^2\mu^2 + g^4 \mu_2^2(L) \left(\int f'''(y)^2 dy\right)^2 + \\ + 2n^{-1} g^{-3} \mu_2(L) c\mu \int L''(y)^2 dy$$

and

$$g_0 = \arg\min_g \text{AMSE}\left(R(\hat{f}_g''(y))\right) = d_0 n^{-1/7},$$

with

$$d_0 = \left[\frac{5}{2}\mu_2(L)^{-1} c\mu \int (L''(y))^2 dy \left(\int (f'''(y))^2 dy\right)^{-1}\right]^{1/7}.$$

From the expression of the optimal pilot bandwidth we can get an estimator, $\hat{g}_0$, just by plugging-in estimates of the unknown quantities. A simpler proposal could be to estimate the pilot by rescaling the rule-of-thumb for bandwidth selection with the corresponding order of the pilot, this is multiplying that value by the factor $\frac{n^{-1/7}}{n^{-1/5}}$.

Hence, we define two possibilities for the bootstrap bandwidth estimate:

$$\hat{h}_{\text{Bopt}} = \left(\frac{R(L)\hat{\mu}\hat{c}}{n\mu_2^2(L) R(\hat{f}_{\hat{g}_0}'')}\right)^{1/5} \quad \text{and}$$

$$\hat{h}_{\text{B}_{\text{RT}}} = \left(\frac{R(L)\hat{\mu}\hat{c}}{n\mu_2^2(L) R(\hat{f}_{\hat{g}_1}'')}\right)^{1/5}, \text{ with } \hat{g}_1 = \frac{n^{-1/7}}{n^{-1/5}} \hat{h}_{\text{RT}}.$$

Remark. *The asymptotic expression of the MSE in* (2) *given by*

$$\text{AMSE}\left(\hat{f}_h(y)\right) = \frac{1}{4} h^4 \left(f''(y)\right)^2 \mu_2^2(K) + \frac{\gamma(y)}{nh} R(K),$$

can be used to obtain the expression of an optimal local bandwidth, following similar steps as for the global one, but from the expression:

$$h_{\text{AMSE}}(y) = \left(\frac{\gamma(y) R(K)}{n(f''(y))^2 \mu_2^2(K)}\right)^{1/5}.$$

Then, a similar method as the described by González-Manteiga et al. (2004) for local linear regression, could be proposed in the context of density estimation with length-biased data.

3.3.2 Using a common kernel density estimator

Bose and Dutta (2013) proposed a new bootstrap bandwidth selector for complete data arguing that they do not need to assume a shape for the unknown density at any stage, and moreover they only require f to be four times differentiable instead of the six times needed in the method presented above.

Following their methodology we propose to obtain a smooth bootstrap bandwidth minimising the $\mathrm{MISE}^*(h)$ in a compact interval I, and assuming that the pilot bandwidth g can be set as $\frac{1}{8}n^{-1/(2p+2s+1)}$, where p and s are the orders of the kernels K and L respectively. This fixed value for the pilot has been set in Bose and Dutta (2013) after extensive simulation studies using different mixtures of normals on f.

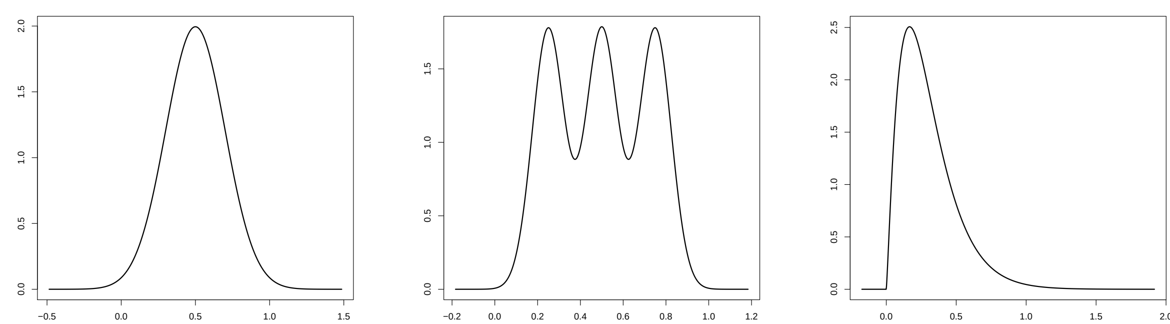
Hence, we define this bootstrap bandwidth as follows:

$$\hat{h}_{\mathrm{B}} = arg\min_{h\in I} \mathrm{MISE}^*(h).$$

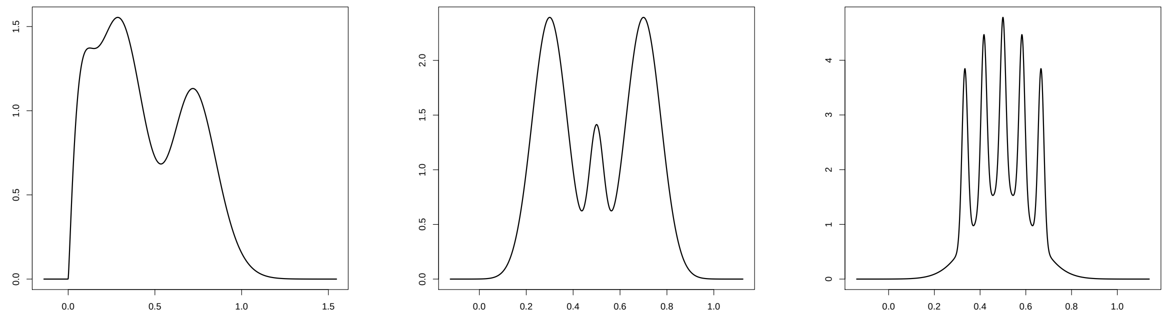
4 Finite sample study

In this section we evaluate the performance of the bandwidth selection procedures presented in Section 3. To this goal we have carried out a simulation study including rule-of-thumb ($\hat{h}_{\mathrm{RT}}$), cross-validation bandwidth ($\hat{h}_{\mathrm{CV}}$),the bootstrap bandwidths ($\hat{h}_{\mathrm{Bopt}}$) and ($\hat{h}_{\mathrm{B_{RT}}}$) with the two possible pilots and ($\hat{h}_{\mathrm{B}}$). We have considered as benchmarks the infeasible optimal bandwidth values h_{MISE} and h_{ISE} which correspond, respectively, to the optimal bandwidths obtained from MISE and ISE criterion defined in (3) and (6).

We have simulated six models with densities shown in Figure 1, some of them have been taken from Mammen et al. (2011) and others from Marron and Wand (1992) but rescaled to the interval $[0, 1]$. We have chosen these models to cover a wide range of densities with different complexity levels, including different number of modes and asymmetry.



(a) Model 1 (b) Model 2 (c) Model 3



(d) Model 4 (e) Model 5 (f) Model 6

Figure 1: The six simulated densities in the finite sample study.

These six models are:

- Model 1: a Normal distribution $N(0.5, 0.2^2)$.
- Model 2: a trimodal mixture of three Normal distributions, $N(0.25, 0.075^2)$, $N(0.5, 0.075^2)$ and $N(0.75, 0.075^2)$ with coefficients $\frac{1}{3}$.
- Model 3: a gamma distribution, $Gamma(a, b)$, with $a = b^2$ and $b = 1.5$ applied on $5x$ with $x \in \mathbb{R}^+$.
- Model 4: a mixture of three gamma distributions, $Gamma(a_i, b_i)$, $i = 1, \ldots, 3$ with $a_i = b_i^2$, $b_1 = 1.5$, $b_2 = 3$ and $b_3 = 6$ applied on $8x$ and $x \in \mathbb{R}^+$, with coefficients $\frac{1}{3}$.
- Model 5: a mixture of three Normal distributions, $N(0.3, \frac{3}{40}^2)$, $N(0.7, \frac{3}{40}^2)$ and $N(0.5, \frac{1}{32}^2)$ with coefficients $\frac{9}{20}$, $\frac{9}{20}$ and $\frac{1}{10}$ respectively.
- Model 6: a mixture of six Normal distributions, $N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, 6$ with $\mu_1 = 0.5$, $\mu_2 = \frac{1}{3}$, $\mu_3 = \frac{5}{12}$, $\mu_4 = \frac{1}{2}$, $\mu_5 = \frac{7}{12}$, $\mu_6 = \frac{2}{3}$, $\sigma_1 = \frac{1}{8}$ and $\sigma_i = \frac{1}{80}$, $i = 2, \ldots, 6$ with coefficients $c_1 = \frac{1}{2}$ and $c_i = \frac{1}{10}$, $i = 2, \ldots, 6$.

We have simulated 1000 length-biased samples from each model considering sample sizes $n = 50$, 100, 200 and 500, using the Epanechnikov kernel. From these samples we have evaluated the performance of the bandwidth selectors through the following measures:

$$m_1 = \text{mean}(\text{ISE}(\hat{h})), \quad m_2 = \text{std}(\text{ISE}(\hat{h})),$$

$$m_3 = \text{mean}(\hat{h} - h_{\text{ISE}}), \quad m_4 = \text{std}(\hat{h} - h_{\text{ISE}}).$$

The first two measures, m_1 and m_2 are referred to the error of the estimation, so they provide us with information about the overall performance and variability of the different methods. Meanwhile, m_3 and m_4 measure respectively, the bias and variability of the difference between the theoretical benchmark and the value selected by the proposals. This provides information about the way the methods are choosing the bandwidth parameter.

Table 1: Mean and standard deviations of the ISE and of the difference between the benchmark and the bandwidths selectors(criteria m_1 to m_4) for Model 1 multiplied by 10^2.

	Model 1						
	h_{ISE}	$\hat{h}_{\text{RT}}$	$\hat{h}_{\text{CV}}$	$\hat{h}_{\text{Bopt}}$	$\hat{h}_{\text{B}_{\text{RT}}}$	$\hat{h}_{\text{B}}$	h_{MISE}
				n=50			
m_1	4.61	7.95	7.52	8.72	8.93	9.05	5.20
m_2	5.31	14.39	7.38	8.82	11.23	8.33	8.41
m_3	—	-8.19	3.51	-9.65	-9.86	-9.27	1.15
m_4	—	6.25	10.75	5.83	6.03	5.78	5.32
				n=100			
m_1	3.11	4.92	4.52	5.65	5.79	5.64	3.55
m_2	4.21	10.83	4.83	6.14	8.12	5.06	7.85
m_3	—	-6.62	2.25	-8.63	-8.82	-8.02	0.45
m_4	—	5.48	8.22	5.06	5.23	4.67	4.70
				n=200			
m_1	1.94	2.98	2.73	3.61	3.70	3.29	2.23
m_2	2.98	7.45	3.29	4.08	5.49	3.25	6.15
m_3	—	-5.77	1.28	-8.19	-8.33	-7.16	0.37
m_4	—	4.53	6.87	4.11	4.34	3.79	4.00
				n=500			
m_1	1.00	1.51	1.35	1.95	1.99	1.16	1.53
m_2	1.95	3.99	2.09	2.45	3.02	3.79	2.07
m_3	—	-5.01	0.80	-7.53	-7.60	0.49	-5.56
m_4	—	3.78	5.14	3.32	3.55	3.52	3.14

Table 2: Mean and standard deviations of the ISE and of the difference between the benchmark and the bandwidths selectors(criteria m_1 to m_4) for Model 2 multiplied by 10^2.

	Model 2						
	h_{ISE}	$\hat{h}_{\text{RT}}$	$\hat{h}_{\text{CV}}$	$\hat{h}_{\text{Bopt}}$	$\hat{h}_{\text{B}_{\text{RT}}}$	$\hat{h}_{\text{B}}$	h_{MISE}
				n=50			
m_1	11.91	15.15	17.99	14.27	14.27	14.28	13.42
m_2	5.24	5.70	6.82	6.33	6.39	7.57	7.21
m_3	—	1.59	14.13	-0.24	-0.31	-3.95	-3.18
m_4	—	10.14	19.332	10.01	10.07	9.02	9.39
				n=100			
m_1	7.89	10.60	12.25	9.01	9.00	8.80	8.23
m_2	3.89	3.17	5.90	3.77	3.77	4.32	4.27
m_3	—	4.83	8.94	2.26	2.22	0.23	-0.54
m_4	—	4.50	14.16	4.41	4.43	4.55	4.15
				n=200			
m_1	4.90	7.30	6.85	5.50	5.51	5.78	5.06
m_2	2.49	2.14	4.55	2.46	2.45	2.48	2.57
m_3	—	5.11	2.99	2.14	2.18	2.23	0.33
m_4	—	1.35	8.68	1.34	1.32	1.98	1.20
				n=500			
m_1	2.47	4.21	3.06	2.74	2.74	4.28	2.55
m_2	1.45	1.45	2.45	1.52	1.55	1.44	1.68
m_3	—	4.52	0.68	1.52	1.50	4.48	0.03
m_4	—	1.25	4.29	0.98	1.07	1.31	1.27

Table 3: Mean and standard deviations of the ISE and of the difference between the benchmark and the bandwidths selectors(criteria m_1 to m_4) for Model 3 multiplied by 10^2.

Model 3							
	h_{ISE}	$\hat{h}_{\text{RT}}$	$\hat{h}_{\text{CV}}$	$\hat{h}_{\text{Bopt}}$	$\hat{h}_{\text{B}_{\text{RT}}}$	$\hat{h}_{\text{B}}$	h_{MISE}
n=50							
m_1	7.66	9.14	25.01	9.51	9.47	15.24	8.75
m_2	5.99	7.65	10.41	8.41	8.34	11.40	6.95
m_3	—	-0.12	24.39	-1.62	-1.55	-6.87	0.53
m_4	—	4.84	11.82	4.83	4.82	9.24	4.27
n=100							
m_1	5.18	6.00	22.71	6.32	6.29	8.13	5.86
m_2	3.67	4.33	7.37	4.79	4.80	6.25	4.16
m_3	—	0.43	25.88	-1.80	-1.78	-5.08	0.50
m_4	—	3.83	8.52	3.78	3.76	4.39	3.49
n=200							
m_1	3.52	4.00	21.60	4.27	4.25	4.39	3.94
m_2	2.54	2.81	5.56	3.20	3.19	3.41	2.77
m_3	—	0.75	27.58	-1.90	-1.87	-2.16	0.70
m_4	—	3.09	6.41	3.07	3.05	3.30	2.89
n=500							
m_1	1.99	2.21	21.24	2.35	2.35	2.39	2.17
m_2	1.35	1.42	3.45	1.59	1.61	1.44	1.47
m_3	—	1.13	30.01	-1.76	-1.75	2.10	0.54
m_4	—	2.15	3.80	2.10	2.11	2.69	2.06

Table 4: Mean and standard deviations of the ISE and of the difference between the benchmark and the bandwidths selectors(criteria m_1 to m_4) for Model 4 multiplied by 10^2.

Model 4							
	h_{ISE}	$\hat{h}_{\mathrm{RT}}$	$\hat{h}_{\mathrm{CV}}$	$\hat{h}_{\mathrm{Bopt}}$	$\hat{h}_{\mathrm{B_{RT}}}$	$\hat{h}_{\mathrm{B}}$	h_{MISE}
n=50							
m_1	7.75	9.27	17.23	9.61	9.57	17.69	8.92
m_2	4.93	7.20	4.65	7.58	7.59	9.13	6.29
m_3	—	-4.12	37.86	-5.33	-5.44	-15.51	-1.42
m_4	—	13.53	26.01	13.71	13.66	12.12	12.47
n=100							
m_1	5.47	6.12	15.71	6.39	6.35	10.32	6.03
m_2	3.56	4.07	2.96	4.44	4.44	5.81	3.87
m_3	—	-0.06	43.64	-2.32	-2.41	-10.33	0.49
m_4	—	7.63	18.89	7.72	7.65	7.38	7.22
n=200							
m_1	3.68	4.13	15.22	4.16	4.16	5.63	4.09
m_2	2.29	2.75	2.48	2.67	2.76	3.46	2.93
m_3	—	1.65	48.79	-1.20	-1.33	-6.50	0.57
m_4	—	4.61	13.04	4.53	4.50	5.08	4.54
n=500							
m_1	2.10	2.40	14.83	2.30	2.30	2.53	2.30
m_2	1.24	1.27	2.62	1.42	1.43	1.70	1.46
m_3	—	2.56	52.42	-0.62	-0.70	-2.32	0.27
m_4	—	3.06	11.46	2.99	3.00	3.74	3.08

Table 5: Mean and standard deviations of the ISE and of the difference between the benchmark and the bandwidths selectors(criteria m_1 to m_4) for Model 5 multiplied by 10^2.

Model 5							
	h_{ISE}	$\hat{h}_{\text{RT}}$	$\hat{h}_{\text{CV}}$	$\hat{h}_{\text{Bopt}}$	$\hat{h}_{\text{B}_{\text{RT}}}$	$\hat{h}_{\text{B}}$	h_{MISE}
n=50							
m_1	5.05	20.39	25.64	18.38	18.28	18.84	15.89
m_2	8.12	6.82	15.53	7.41	7.44	8.17	8.64
m_3	—	5.86	10.47	4.03	3.92	3.60	-0.36
m_4	—	3.58	18.54	3.53	3.56	4.31	3.21
n=100							
m_1	9.26	14.44	13.91	11.59	11.51	13.76	9.70
m_2	4.63	4.16	10.80	4.47	4.48	4.79	4.89
m_3	—	6.05	3.39	3.55	3.47	5.12	-0.16
m_4	—	1.67	11.47	1.63	1.64	2.86	1.50
n=200							
m_1	5.73	10.21	11.27	7.19	7.26	11.52	5.92
m_2	2.69	2.71	13.25	2.73	2.72	3.26	2.74
m_3	—	5.65	5.36	2.83	2.91	6.44	-0.08
m_4	—	1.16	14.83	1.13	1.15	2.13	1.02
n=500							
m_1	2.9	6.25	3.44	3.67	3.68	11.21	2.99
m_2	1.2	1.44	2.48	1.28	1.29	1.93	1.22
m_3	—	5.03	0.12	2.15	2.17	8.58	0.07
m_4	—	0.87	2.66	0.83	0.84	1.36	0.77

Table 6: Mean and standard deviations of the ISE and of the difference between the benchmark and the bandwidths selectors(criteria m_1 to m_4) for Model 6 multiplied by 10^2.

	Model 6						
	h_{ISE}	$\hat{h}_{\mathrm{RT}}$	$\hat{h}_{\mathrm{CV}}$	$\hat{h}_{\mathrm{Bopt}}$	$\hat{h}_{\mathrm{B_{RT}}}$	$\hat{h}_{\mathrm{B}}$	h_{MISE}
				n=50			
m_1	46.43	67.75	63.66	70.76	71.31	68.98	49.66
m_2	13.19	8.09	11.56	8.14	8.25	4.84	18.32
m_3	—	3.84	8.73	2.86	2.56	17.72	-2.58
m_4	—	5.37	8.08	5.34	5.35	6.35	5.09
				n=100			
m_1	30.11	65.97	55.52	64.74	64.50	66.83	30.99
m_2	10.14	4.25	8.73	5.08	5.22	4.35	10.99
m_3	—	5.56	8.69	4.22	4.15	19.51	-0.38
m_4	—	1.51	4.54	1.47	1.47	4.62	1.32
				n=200			
m_1	18.27	62.74	47.91	52.12	51.45	65.21	18.49
m_2	5.58	2.47	14.92	4.51	4.03	4.17	5.58
m_3	—	5.07	7.57	3.51	3.44	19.12	0.14
m_4	—	0.52	4.17	0.46	0.42	4.39	0.30
				n=500			
m_1	9.2	53.41	41.59	32.98	32.71	64.64	9.21
m_2	2.7	3.37	20.43	3.58	3.24	3.17	2.71
m_3	—	4.38	7.12	2.70	2.68	19.81	-0.03
m_4	—	0.36	4.60	0.24	0.22	3.07	0.14

An overview of these numbers indicates that the performance of the methods depends on the complexity of the underlying model. Let classify the models in “easy” (Model 1, Model 3 and Model 4), “intermediate” (Model 2 and Model 5) and “hard” (Model 6) estimation problems.

Regarding to the measure m_1, the rule-of-thumb performs better in smoother densities, such as Model 1, Model 3 and Model 4, however the bootstrap bandwidths are also really competitive for these models, while cross-validation has in general a poorer performance. We have to remark that in Model 4, the bootstrap bandwidth $\hat{h}_{\mathrm{B}}$ needs a large sample size in order to be competitive. Increasing the complexity of the densities, Model 2 and Model 5, the performance of the rule-of-thumb decreases considerably and the bootstrap procedures $\hat{h}_{\mathrm{B_{RT}}}$ and $\hat{h}_{\mathrm{B_{NS}}}$ seems to be more accurate; however $\hat{h}_{\mathrm{B}}$ has a worse performance and the gain with the increasing of the sample size is slower. Note also that depending on the design and the sample size, cross-validation can also produce competitive results. In hard estimation problems as Model 6, bootstrap bandwidths $\hat{h}_{\mathrm{B_{RT}}}$ and $\hat{h}_{\mathrm{B_{NS}}}$ are still valuable competitors.

In terms of variability, which is measured by criterion m_2 and m_4, the cross-validation method exhibits the highest values. The variability of the rule-of-thumb and the bootstrap bandwidths is in general moderate, with the only exception of $\hat{h}_{\mathrm{B}}$ in Model 6 where it exhibits higher values than the other bootstrap rules and the rule-of-thumb.

The bias in bandwidth selection is measured through m_3. Rule-of-thumb and bootstrap bandwidths with pilots generally show bias in the same direction and amount, except for Model 3 where they do not follow this pattern, even though the overall result is good. In smoother models both, tend to oversmooth and the opposite happens with cross-validation. The bias of the other bootstrap bandwidth selector, $\hat{h}_{\mathrm{B}}$, tends to be higher except for very large sample sizes of Model 1.

Moreover, the performance of all the methods proposed may also be compared graphically through the box plots of the errors (ISE's) computed in the 1000 Monte Carlo samples (see Figures 2 - 7). These plots confirm that the behaviour of the selectors depends on the complexity of the underlying model. The bootstrap proposals seems to have in general a good behaviour outperformed only by cross-validation in Model 6, due to its complexity.

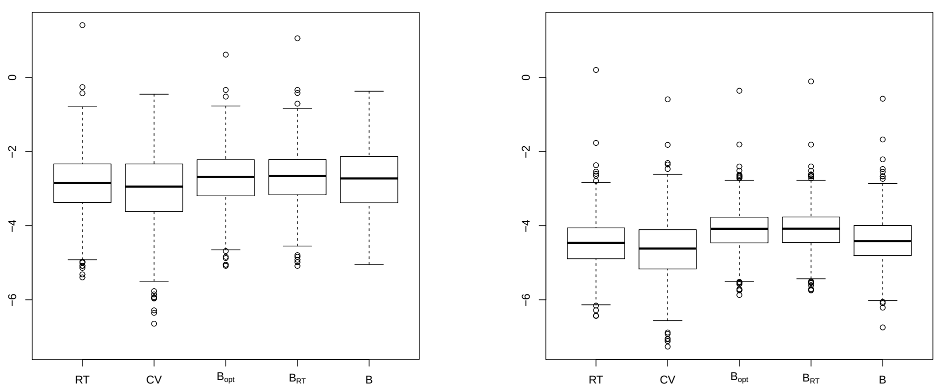


Figure 2: The plot of log(ISE) for the density estimation using the different bandwidth selectors for Model 1 and sizes n=50, 500 (from left to right).

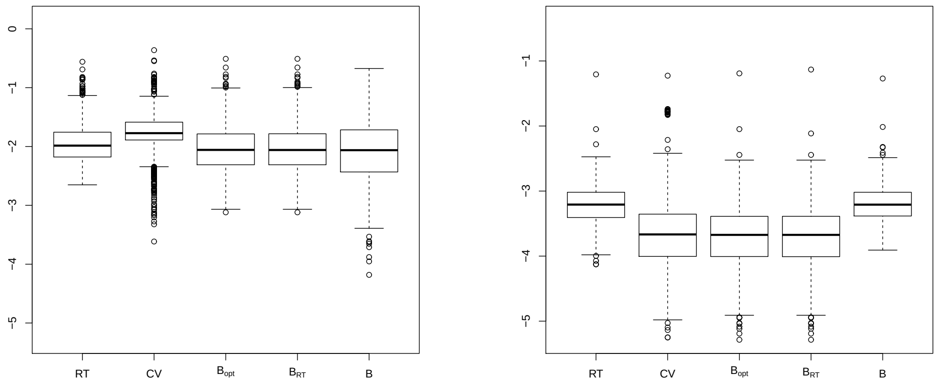


Figure 3: The plot of log(ISE) for the density estimation using the different bandwidth selectors for Model 2 and sizes n=50, 500 (from left to right).

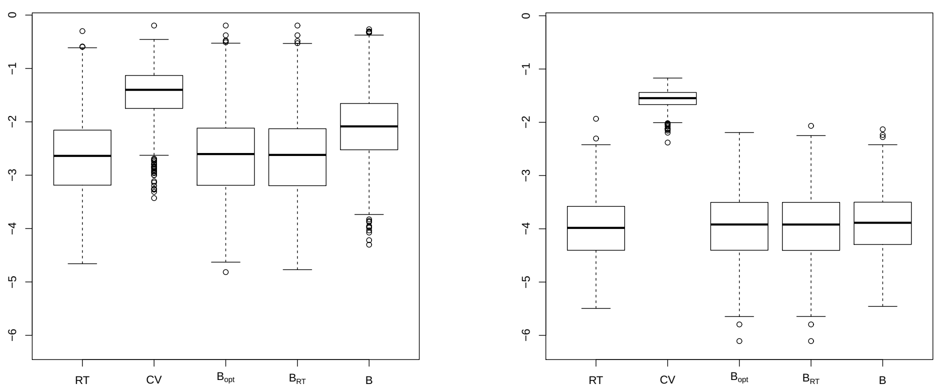


Figure 4: The plot of log(ISE) for the density estimation using the different bandwidth selectors for Model 3 and sizes n=50, 500 (from left to right).

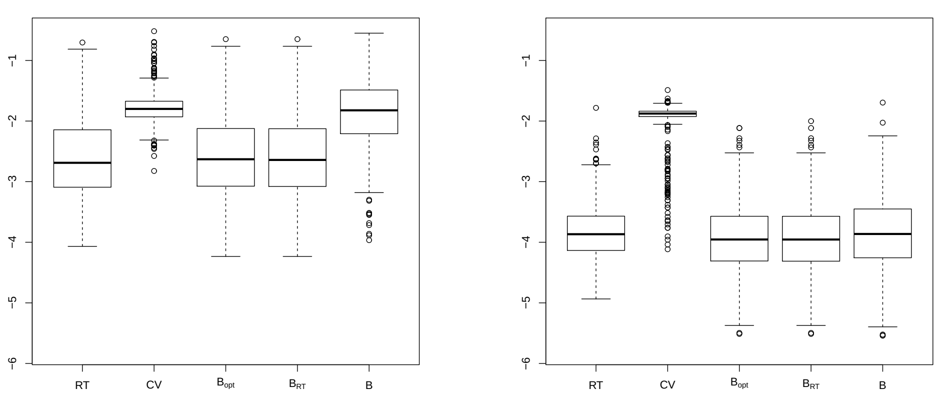


Figure 5: The plot of log(ISE) for the density estimation using the different bandwidth selectors for Model 4 and sizes n=50, 500 (from left to right).

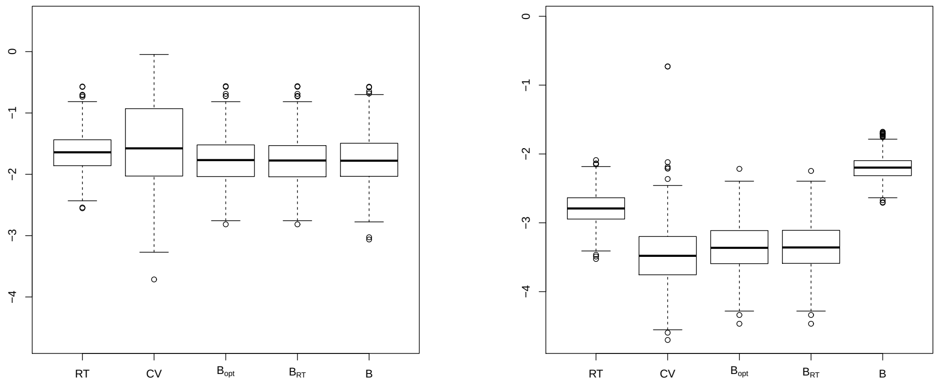


Figure 6: The plot of log(ISE) for the density estimation using the different bandwidth selectors for Model 5 and sizes n=50, 500 (from left to right).

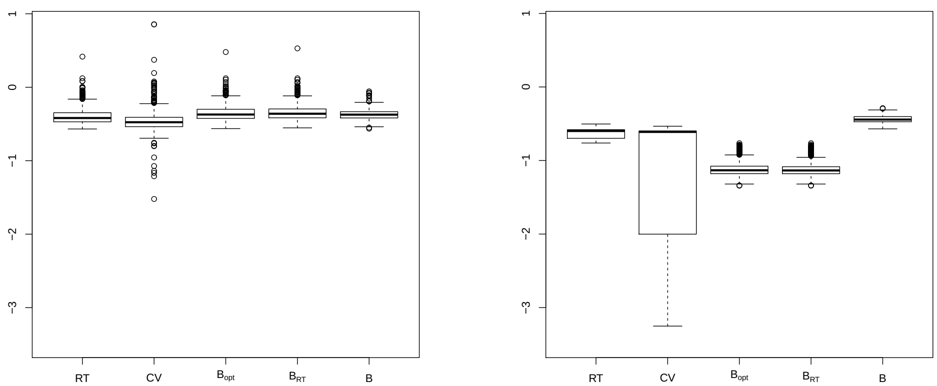


Figure 7: The plot of log(ISE) for the density estimation using the different bandwidth selectors for Model 6 and sizes n=50, 500 (from left to right).

5 Conclusions

We have considered density estimation in the context of length-biased data, specifically we have focused on the kernel estimator introduced by Jones (1991). We have developed with great detail asymptotic expansions of the MSE, MISE and AMISE of this estimator. Furthermore, we have proposed new bandwidth selection methods and we have studied their behaviour in finite samples through an extensive simulation study. As a general comment, some methods outperforms the others depending on the complexity of the underlying model. Nevertheless, our bandwidth selection proposals have shown to perform quite well and in general, better than the current available cross-validation method. The only exception is the case of very complex densities, with several features and peaks, where cross-validation exhibits the best results, but even in this case our proposals are still competitive.

6 Further extensions

As we have remarked in Section 2, the methods presented in this paper can be easily generalised for a general known weight function ω, where the particular case of length-biased data is that of $\omega(y) = y$. First, an appropriate modification of the estimator in (1) must be defined, as it has already been presented in Jones (1991):

$$\hat{f}_{h,\omega}(y) = \frac{1}{n}\hat{\mu}_\omega \sum_{i=1}^{n} \omega(Y_i)^{-1} K_h(y - Y_i),$$

with $\hat{\mu}_\omega = \left(\frac{1}{n} \sum_{i=1}^{n} \omega(Y_i)^{-1} \right)^{-1}$.

Theorem 2.1 and Corollary 2.2 can be generalised assuming the following conditions:

(B.1) $E\left[\frac{1}{\omega(X)}\right] < \infty,\ E\left[\frac{1}{\omega(Y)^l}\right] < \infty \quad l = 1, \ldots, 2\nu,$

(B.2) $\int K(u)du = 1,\ \int uK(u)du = 0$ and $\mu_2(K) < +\infty$

(B.3) $\lim_{n\to\infty} nh = +\infty,$

(B.4) y a continuity point of f,

(B.5) f and ω are two times differentiable in y.

We immediately get the error measures as and their optimal bandwidth parameters for the length-biased data:

$$\text{MSE}(\hat{f}_{h,\omega}(y)) = \frac{1}{4}h^4 \left(f''(y)\right)^2 \mu_2^2(K) + \frac{\gamma_\omega(y)}{nh} R(K) + o\left(h^4 + \frac{1}{nh}\right),$$

with $\gamma_\omega(y) = \frac{f(y)\mu_\omega}{\omega(y)}$ and

$$h_{\text{AMSE},\omega}(y) = \left(\frac{\gamma_\omega(y) R(K)}{n(f''(y))^2 \mu_2^2(K)} \right)^{1/5}.$$

We also obtain:

$$\text{MISE}(\hat{f}_{h,\omega}) = \frac{1}{4}h^4 \mu_2^2(K) R(f'') + \frac{R(K)\mu_\omega c_\omega}{nh} + o\left(h^4 + \frac{1}{nh}\right)$$

where $c_\omega = \int \frac{1}{\omega(y)} f(y)dy$ and then

$$h_{\text{AMISE},\omega} = \left(\frac{R(K)\mu_\omega c_\omega}{n\mu_2^2(K) R(f'')} \right)^{1/5}.$$

The bootstrap methods can be also modified in the same way. Then, the smooth bootstrap samples, $Y_1^*, \ldots, Y_n^*$, can be generated by sampling randomly with replacement n times from the estimated density $\hat{f}_{Y,g,\omega}(y) = \omega(y)\hat{f}_{g,\omega}(y)/\hat{\mu}_\omega$. Here g is again a pilot bandwidth.

For the extension of the bandwidth selectors we need to take into account not only the above modification of the density estimator but also

$$\hat{\sigma}^2_\omega = \left(\frac{1}{n}\sum_{i=1}^n \frac{1}{\omega(Y_i)}\right)^{-1}\left[\left(\frac{1}{n}\sum_{i=1}^n \omega(Y_i)\right) - \left(\frac{1}{n}\sum_{i=1}^n \frac{1}{\omega(Y_i)}\right)^{-1}\right].$$

Apart from these considerations the procedures can be obtained in the same way as the length-biased case.

Acknowledgements

The authors are grateful for constructive comments from the associate editor and two reviewers which helped to improve the paper. They also acknowledge the support from the Spanish Ministry of Economy and Competitiveness, through grant number MTM2013-41383P, which includes support from the European Regional Development Fund (ERDF). Support from the IAP network StUDyS from Belgian Science Policy, is also acknowledged. M.I. Borrajo has been supported by FPU (FPU2013/00473) from the Spanish Ministry of Education.

References

Ahmad, I.A. (1995), 'On multivariate kernel estimation for samples from weighted distributions', *Statistics & Probability Letters*, 22, pp. 121–129.

Asgharian, M., M'Lan, C.E., and Wolfson, D.B. (2002), 'Length-biased sampling with right censoring: an unconditional approach', *Journal of the American Statistical Association*, 97, pp. 201–209.

Barmi, H.E., and Simonoff, J.S. (2000), 'Transformation-based density estimation for weighted distributions', *Journal of Nonparametric Statistics*, 12, pp. 861–878.

Bhattacharyya, F.L.A., B, and Richardson, G.D. (1988), 'A comparioson of nonparametric unweighted and length-biased density estimation of fibres', *Communications in Statistics-Theory and Methods*, 17, pp. 3629–3644.

Bose, A., and Dutta, S. (2013), 'Density estimation using bootstrap bandwidth selector', *Statistics & Probability Letters*, 83, 245–256.

Bowman, A.W. (1984), ‘An alternative method of cross-validation for the smoothing of density estimates’, *Biometrika*, 71, 353–360.

Brunel, E., Comte, F., and Guilloux, A. (2009), ‘Nonparametric density estimation in presence of bias and censoring’, *Test*, 18, pp. 166–194.

Cacoullos, T. (1966), ‘Estimation of a multivariate density’, *Annals of the Institute of Statistical Mathematics*, 18, pp. 179–189.

Cao, R. (1990), ‘Aplicaciones y nuevos resultados del Método Bootstrap en la estimación no paramétrica de curvas’, Ph.D. dissertation, Universidade de Santiago de Compostela.

Cao, R. (1993), ‘Bootstrapping the mean integrated squared error’, *Journal of Multivariate Analysis*, 45, pp. 137–160.

Cao, R., Cuevas, A., and González-Manteiga, W. (1994), ‘A comparative study of several smoothing methods in density estimation’, *Computational Statistics and Data Analysis*, 17, pp. 153–176.

Chakraborty, R., and Rao, C.R. (2000), ‘23 Selection biases of samples and their resolutions’, *Handbook of Statistics*, 18, pp. 675–712.

Chesneau, C. (2010), ‘Wavelet block thresholding for density estimation in the presence of bias’, *Journal of the Korean Statistical Society*, 39, pp. 43–53.

Chiu, S.T. (1992), ‘An automatic bandwidth selector for kernel density estimation’, *Biometrika*, 79, pp. 771–782.

Collomb, G. (1976), ‘Estimation non parametrique de la regression par la méthode du noyau’, Ph.D. dissertation, Universit Paul Sabatier de Toulouse.

Comte, F., and Rebafka, T. (2016), ‘Nonparametric weighted estimators for biased data’, *Journal of Statistical Planning and Inference*, 174, pp. 104–128.

Cox, D. (2005), ‘Some sampling problems in technology’, in *Selected Statistical Papers of Sir David Cox*, Vol. 1, eds. D. Hand and A. Herzberg, Cambridge University Press, pp. pp. 81–92.

Cutillo, L., De Feis, I., Nikolaidou, C., and Sapatinas, T. (2014), ‘Wavelet density estimation for weighted data’, *Journal of Statistical Planning and Inference*, 146, pp. 1–19.

de Uña-Álvarez, J. (2004), ‘Nonparametric estimation under length-biased sampling and type I censoring: a moment based approach’, *Annals of the Institute of Statistical Mathematics*, 56, pp. 667–681.

Dutta, S. (2016), ‘Cross-validation revisited’, *Communications in Statistics-Simulation and Computation*, 45, pp. 472–490

Efromovich, S. (2004), 'Density estimation for biased data', *Annals of Statistics*, 32, pp. 1137–1161.

González-Manteiga, W., Martínez-Miranda, M.D., and Pérez-González, A. (2004), 'The choice of smoothing parameter in nonparametric regression through Wild Bootstrap', *Computational Statistics and Data Analysis*, 47, pp. 487–515.

Guillamón, A., Navarro, J., and Ruiz, J.M. (1998), 'Kernel density estimation using weighted data', *Communications in Statistics-Theory and Methods*, 27, pp. 2123–2135.

Hall, P., and Marron, J.S. (1987), 'Estimation of integrated squared density derivatives', *Statistics & Probability Letters*, 6, pp. 109–115.

Heckman, J.J. (1990), 'Selection bias and self-selection', in *Econometrics*, Springer, pp. pp. 201–224.

Heidenreich, N.B., Schindler, A., and Sperlich, S. (2013), 'Bandwidth selection for kernel density estimation: a review of fully automatic selectors', *Advances in Statistical Analysis*, 97, pp. 403–433.

Jones, M.C. (1991), 'Kernel density estimation for length biased data', *Biometrika*, 78, pp. 511–519.

Jones, M.C., and Karunamuni, R.J. (1997), 'Fourier series estimation for length biased data', *Australian Journal of Statistics*, 39, pp. 57–68.

Jones, M.C., Marron, J.S., and Sheather, S.J. (1996), 'A brief survey of bandwidth selection for density estimation', *Journal of the American Statistical Association*, 91, pp. 401–407.

Mammen, E., Martínez-Miranda, M.D., Nielsen, J.P., and Sperlich, S. (2011), 'Do-validation for kernel density estimation', *Journal of the American Statistical Association*, 106, pp. 651–660.

Mammen, E., Martínez-Miranda, M.D., Nielsen, J.P., and Sperlich, S. (2014), 'Further theoretical and practical insight to the do-validated bandwidth selector', *Journal of the Korean Statistical Society*, 43, pp. 355–365.

Marron, J.S. (1988), 'Automatic smoothing parameter selection: a survey', *Empirical Economics*, 13, pp. 187–208.

Marron, J.S. (1992), 'Bootstrap bandwidth selection', in *Exploring the limits of bootstrap*, eds. R. Lepage and L. Billard, Wiley, pp. pp. 249–262.

Marron, J.S., and Wand, M.P. (1992), 'Exact mean integrated squared error', *The Annals of Statistics*, 20, pp. 712–736.

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

Patil, G.P., and Rao, C.R. (1977), ‘The weighted distributions: A survey of their applications’, *Applications of Statistics*, 383, pp. 383–405.

Ramírez, P., and Vidakovic, B. (2010), ‘Wavelet density estimation for stratified size-biased sample’, *Journal of Statistical Planning and Inference*, 140, pp. 419–432.

Richardson, G.D., Kazempour, M.K., and Bhattacharyya, B. (1991), ‘Length biased density estimation of fibres’, *Journal of Nonparametric Statistics*, 1, pp. 127–141.

Rosenblatt, M. (1956), ‘Remarks on some nonparametric estimates of a density function’, *The Annals of Mathematical Statistics*, 27, pp. 832–837.

Rudemo, M. (1982), ‘Empirical choice of histograms and kernel density estimators’, *Scandinavian Journal of Statistics*, 9, pp. 65–78.

Scott, D.W. (1992), *Multivariate density estimation: Theory, practice and visualisation*, Wiley.

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, pp. 683–690.

Silverman, B.W. (1986), *Density estimation for statistics and data analysis*, CRC press.

Simon, R. (1980), ‘Length biased sampling in etiologic studies’, *American Journal of Epidemiology*, 111, pp. 444–452.

Turlach, B.A. (1993), ‘Bandwidth selection in kernel density estimation: A review’, Technical report, Université catholique de Louvain.

Vardi, Y. (1982), ‘Nonparametric estimation in the presence of length bias’, *The Annals of Statistics*, 10, pp. 616–620.

Vardi, Y. (1985), ‘Empirical distributions in selection bias models’, *The Annals of Statistics*, 13, pp. 178–203.

Zelen, M., and Feinleib, M. (1969), ‘On the theory of screening for chronic diseases’, *Biometrika*, 56, 601–614.

Appendices

A Proof of Theorem 2.1

First of all we rewrite the estimator in (1) as follows:

$$\hat{f}_h(y) = \frac{1}{n}\frac{1}{\frac{1}{n}\sum_{i=n}^{n}\frac{1}{Y_i}}\sum_{i=1}^{n}\frac{1}{Y_i}K_h(y-Y_i) = \frac{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}K_h(y-Y_i)}{\frac{1}{n}\sum_{i=n}^{n}\frac{1}{Y_i}} = \frac{\phi_n(y)}{\xi_n}. \tag{7}$$

We start by calculating the punctual mean of (1) for which we need the mean of the numerator and denominator in (7), so:

$$\overline{\phi_n(y)} := E[\phi_n(y)] = E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}K_h(y-Y_i)\right] = \int \frac{1}{z}K_h(y-z)f_Y(z)dz = \frac{1}{\mu}(K_h \circ f)(y).$$

$$\overline{\xi_n} := E[\xi_n] = E\left[\frac{1}{n}\sum_{j=1}^{n}\frac{1}{Y_j}\right] = \int \frac{1}{z}f_Y(z)dz = \frac{1}{\mu}.$$

We divide this proof in two separated but linked paragraphs, detailing all the results involving mean and variance calculations respectively.

Mean
Applying the linearisation technique used in Collomb (1976) with $\nu \geq 2$ and taking into account that $\xi_n \neq 0\ \forall\, n$, we can write down

$$\begin{aligned}\hat{f}_h(y) &= \frac{\phi_n(y)}{\xi_n} = \frac{\phi_n(y)}{\overline{\xi_n}}\cdot\frac{\overline{\xi_n}}{\xi_n} = \frac{\phi_n(y)}{\overline{\xi_n}}\left[1+\sum_{k=1}^{\nu-1}(-1)^k\left(\frac{\xi_n-\overline{\xi_n}}{\overline{\xi_n}}\right)^k + (-1)^\nu\frac{\overline{\xi_n}}{\xi_n}\left(\frac{\xi_n-\overline{\xi_n}}{\overline{\xi_n}}\right)^\nu\right]\\
&= \frac{\phi_n(y)}{\overline{\xi_n}}\left[1+\sum_{k=1}^{\nu-1}(-1)^k\left(\frac{\xi_n-\overline{\xi_n}}{\overline{\xi_n}}\right)^k\right] + (-1)^\nu\hat{f}_h(y)\left(\frac{\xi_n-\overline{\xi_n}}{\overline{\xi_n}}\right)^\nu.\end{aligned}$$

Using the notation $S_n^{a,b}(y) := E[\phi_n^a(y)(\xi_n-\overline{\xi_n})^b]$, $s_n^{a,b}(y) := E[(\phi_n(y)-\overline{\phi_n(y)})^a(\xi_n-\overline{\xi_n})^b]$, $\sigma_n^{a,b}(y) = E[\hat{f}_h(y)^a(\xi_n-\overline{\xi_n})^b]$ and knowing that $\overline{\phi_n}S_n^{0,k}(y)+s_n^{1,k}(y) = S_n^{1,k}(y)$ we can write:

$$
\begin{aligned}
E[\hat{f}_h(y)] &= E\left[\frac{\phi_n(y)}{\overline{\xi_n}} + \frac{\phi_n(y)}{\overline{\xi_n}}\sum_{k=1}^{v-1}(-1)^k\left(\frac{\xi_n-\overline{\xi_n}}{\overline{\xi_n}}\right)^k + (-1)^\nu \hat{f}_h(y)\left(\frac{\xi_n-\overline{\xi_n}}{\overline{\xi_n}}\right)^\nu\right] \\
&= \frac{\overline{\phi_n(y)}}{\overline{\xi_n}} + \sum_{k=1}^{\nu-1}(-1)^k\frac{E[\phi_n(y)(\xi_n-\overline{\xi_n})^k]}{\overline{\xi_n}^{k+1}} + \frac{(-1)^\nu}{\overline{\xi_n}^\nu}E[\hat{f}_h(y)(\xi_n-\overline{\xi_n})^\nu] \\
&= \frac{\overline{\phi_n(y)}}{\overline{\xi_n}} + \sum_{k=1}^{\nu-1}(-1)^k\frac{S_n^{1,k}(y)}{\overline{\xi_n}^{k+1}} + (-1)^\nu\frac{\sigma_n^{1,\nu}(y)}{\overline{\xi_n}^\nu} \\
&= \frac{\overline{\phi_n(y)}}{\overline{\xi_n}} + \frac{\overline{\phi_n(y)}S_n^{0,2}(y) - \overline{\xi_n}s_n^{1,2}(y) - \overline{\xi_n}S_n^{1,1}(y)}{\overline{\xi_n}^3} + \sum_{k=1}^{\nu-1}(-1)^k\frac{S_n^{1,k}(y)}{\overline{\xi_n}^{k+1}} + (-1)^\nu\frac{\sigma_n^{1,\nu}(y)}{\overline{\xi_n}^\nu}
\end{aligned}
$$

Then, we have proved that

$$
E[\hat{f}_h(y)] = \frac{\overline{\phi_n(y)}}{\overline{\xi_n}} + c_n(y) + c_n^{(\nu)}(y) + \frac{(-1)^\nu\sigma_n^{1,\nu}(y)}{\overline{\xi_n}^\nu},
$$

where
$c_n(y) = \dfrac{\overline{\phi_n(y)}S_n^{0,2}(y) - \overline{\xi_n}S_n^{1,1}(y)}{\overline{\xi_n}^3} = \dfrac{\overline{\phi_n(y)}E[(\xi_n-\overline{\xi_n})^2] - \overline{\xi_n}E[\phi_n(y)(\xi_n-\overline{\xi_n})]}{\overline{\xi_n}^3}$ and
$c_n^{(\nu)}(y) = \dfrac{s_n^{1,2}(y)}{\overline{\xi_n}^3} + \displaystyle\sum_{k=1}^{\nu-1}(-1)^k\frac{S_n^{1,k}(y)}{\overline{\xi_n}^{k+1}} = \dfrac{E[(\phi_n(y)-\overline{\phi_n(y)})(\xi_n-\overline{\xi_n})^2]}{\xi_n^3} +$
$+\sum_{k=1}^{\nu-1}(-1)^k\frac{E[\phi_n(y)(\xi_n-\overline{\xi_n})^k]}{\overline{\xi_n}^{k+1}}$.

The first addend corresponds to the asymptotic expression of the mean obtained by Jones (1991). We want now to expand each of the other terms and study the rate of convergence. To this aim we use some basic statistical properties and we proceed as follows:

$$
\begin{aligned}
E\left[(\xi_n-\overline{\xi_n})^2\right] &= Var\left[\xi_n\right] = E\left[\xi_n^2\right] - \overline{\xi_n}^2 \\
&= E\left[\left(\frac{1}{n}\sum_{j=1}^n\frac{1}{Y_j}\right)^2\right] - \frac{1}{\mu^2} = \frac{1}{n^2}E\left[\sum_{i=1}^n\frac{1}{Y_j^2} + \sum_{i\neq j}\frac{1}{Y_i}\frac{1}{Y_j}\right] - \frac{1}{\mu^2} \\
&= \frac{1}{n^2}nE\left[\frac{1}{Y_1^2}\right] + \frac{(n^2-n)}{n^2}E\left[\frac{1}{Y_1}\frac{1}{Y_2}-\right]\frac{1}{\mu^2} = \frac{1}{n}\int\frac{1}{z^2}g(z)dz + \frac{n-1}{n}E\left[\frac{1}{Y_1}\right]^2 - \frac{1}{\mu^2} \\
&= \frac{1}{n}E\left[\frac{1}{X}\right] + \frac{n-1}{n\mu^2} - \frac{1}{\mu^2} = \frac{1}{n\mu}\left(E\left[\frac{1}{X}\right] - \frac{1}{\mu}\right).
\end{aligned}
$$

$$
\begin{aligned}
E\left[\phi_n(y)(\xi_n-\overline{\xi_n})\right] &= E\left[\phi_n(y)\xi_n\right]-\overline{\xi_n}\,\overline{\phi_n(y)} \\
&= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}K_h(y-Y_i)\right)\left(\frac{1}{n}\sum_{j=1}^{n}\frac{1}{Y_j}\right)\right]-\frac{1}{\mu^2}\left(K_h\circ f\right)(y) \\
&= \frac{1}{n^2}E\left[\sum_{i=1}^{n}\frac{1}{Y_i^2}K_h(y-Y_i)+\sum_{i\neq j}\frac{1}{Y_i}\frac{1}{Y_j}K_h(y-Y_i)\right]-\frac{1}{\mu^2}\left(K_h*f\right)(y) \\
&= \frac{n}{n^2}E\left[\frac{1}{Y_1^2}K_h(y-Y_1)\right]+\frac{n(n-1)}{n^2}E\left[\frac{1}{Y_1}\frac{1}{Y_2}K_h(y-Y_1)\right]-\frac{1}{\mu^2}\left(K_h\circ f\right)(y) \\
&= \frac{1}{n\mu^2}\left(K_h\circ\gamma\right)(y)-\frac{1}{n\mu^2}\left(K_h\circ f\right)(y).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
c_n(y) &= \mu^3\left[\frac{1}{\mu}\left(K_h\circ f\right)(y)\frac{1}{n\mu}\left(E\left[\frac{1}{X}\right]-\frac{1}{\mu}\right)-\frac{1}{\mu}\left(\frac{1}{n\mu^2}\left(K_h\circ\gamma\right)(y)-\frac{1}{n\mu^2}\left(K_h\circ f\right)(y)\right)\right] \\
&= \underbrace{\frac{\mu}{n}\left(K_h\circ f\right)(y)\left(E\left[\frac{1}{X}\right]-\frac{1}{\mu}\right)}_{(a)}-\underbrace{\frac{1}{n}\left(K_h\circ\gamma\right)(y)}_{(b)}+\underbrace{\frac{1}{n}\left(K_h\circ f\right)(y)}_{(c)}.
\end{aligned}
$$

Applying Theorem 2.1 of Cacoullos (1966) with $g(z)=f(z)$ for (a) and (c), and $g(z)=\frac{f(z)}{z}$ for (b), we easily obtain that every of these addends is $O(1/n)$. Therefore, $c_n(y)=O(1/n)$.

To expand the next two terms we use the Hölder inequality and, taking into account that K is bounded, we only require the finiteness of the second order moment of $\frac{1}{Y}$.

$$
E\left[\phi_n(y)\left(\xi_n-\overline{\xi_n}\right)^2\right]\leq E\left[\phi_n^2(y)\right]^{1/2}E\left[\left(\xi_n-\overline{\xi_n}\right)^4\right]^{1/2}=O(1/n)^{1/2}O(1/n^2)^{1/2}=O(1/n^{3/2})
$$

$$
E\left[\left(\xi_n-\overline{\xi_n}\right)^2\right]=O(1/n).
$$

Therefore,

$$
c_n^{(\nu)}(y)=O(1/n).
$$

Finally,

$$
\sigma_n^{1,\nu}(y)=E\left[\hat{f}_h(y)\left(\xi_n-\overline{\xi_n}\right)^{\nu}\right]\leq E\left[\hat{f}_h^{\,2}(y)\right]^{1/2}E\left[\left(\xi_n-\overline{\xi_n}\right)^{2\nu}\right]^{1/2}=O(1)O(1/n^{\nu})^{1/2}=O(1/n^{\nu/2}).
$$

Variance

To get the variance, we compute the expected value of the squared estimator. We follow the same techniques as in the previous operations replacing $\hat{f}_h(y)$ by $\hat{f}_h^{\,2}(y)$. Applying again the linearisation method of Collomb (1976) with $\nu\geq 2$,

$$
\begin{aligned}
\hat{f_h}^2(y) &= \frac{\phi_n^2(y)}{\xi_n^2} = \frac{\phi_n^2(y)}{\overline{\xi}_n^2} \cdot \frac{\overline{\xi}_n^2}{\xi_n^2(y)} \\
&= \frac{\phi_n^2(y)}{\overline{\xi}_n^2}\left[1 + \sum_{k=1}^{\nu-1}(-1)^k \left(\frac{\xi_n^2(y) - \overline{\xi}_n^2}{\overline{\xi}_n^2}\right)^k\right] + (-1)^\nu \hat{f_h}^2(y)\left(\frac{\xi_n^2(y) - \overline{\xi}_n^2}{\overline{\xi}_n^2}\right)^\nu \\
&= \frac{\phi_n^2(y)}{\overline{\xi}_n^2}\left[1 + \sum_{k=1}^{\nu-1}(-1)^k \sum_{j=0}^{k} \frac{k!}{j!(k-j)!} 2^{k-j} \left(\frac{\xi_n - \overline{\xi}_n}{\overline{\xi}_n}\right)^{k+j}\right] + \\
&+ (-1)^\nu \hat{f_h}^2(y) \sum_{j=0}^{\nu} \frac{\nu!}{j!(\nu-j)!} 2^{\nu-j} \left(\frac{\xi_n - \overline{\xi}_n}{\overline{\xi}_n}\right)^{\nu+j}.
\end{aligned}
$$

To obtain the mean of $\hat{f_h}^2(y)$ we need to work on $\dfrac{S_n^{2,l}(y)}{\overline{\xi}_n^2} = \dfrac{\overline{\phi}_n^2 s_n^{0,l}(y) + 2\overline{\phi}_n^2 s_n^{1,l}(y) + s_n^{2,l}(y)}{\overline{\xi}_n^2}$ as follows:

$$
\begin{aligned}
&E\left[\hat{f_h}^2(y)\right] = \frac{1}{\overline{\xi}_n^2} E\left[\phi_n(y)^2\right] + E\left[\frac{\phi_n(y)^2(y)}{\overline{\xi}_n^2} \sum_{k=1}^{\nu-1}(-1)^k \sum_{j=0}^{k} \frac{k!}{j!(k-j)!} 2^{k-j} \left(\frac{\xi_n - \overline{\xi}_n}{\overline{\xi}_n}\right)^{k+j}\right] + \\
&+ (-1)^\nu \sum_{j=0}^{\nu} \frac{\nu! 2^{\nu-j}}{j!(\nu-j)!} \frac{E\left[\hat{f_h}^2(y)\left(\xi_n - \overline{\xi}_n\right)^{\nu+j}\right]}{\overline{\xi}_n^{\nu+j}} \\
&= \frac{E\left[\phi_n^2(y)\right]}{\overline{\xi}_n^2} + \sum_{k=1}^{\nu-1}(-1)^k \sum_{j=0}^{k} \frac{k!}{j!(k-j)!} 2^{k-j} \frac{S_n^{2,k+j}(y)}{\overline{\xi}_n^{k+j+2}} + (-1)^\nu \sum_{j=0}^{k} \frac{\nu!}{j!(\nu-j)!} 2^{\nu-j} \frac{\sigma_n^{2,\nu+j}(y)}{\overline{\xi}_n^{\nu+j}} \\
&= \frac{\overline{\phi}_n^2}{\overline{\xi}_n^2} + \frac{s_n^{0,2}(y)}{\overline{\xi}_n^2} + \sum_{k=1}^{\nu-1}(-1)^k \sum_{j=0}^{k} \frac{k!}{j!(k-j)!} 2^{k-j} \frac{\overline{\phi}_n^2 s_n^{0,k+j}(y) + 2\overline{\phi}_n s_n^{1,k+j}(y) + s_n^{2,k+j}(y)}{\overline{\xi}_n^{k+j+2}} +
\end{aligned}
$$

$$
\begin{aligned}
&+(-1)^{\nu}\sum_{j=0}^{\nu}\frac{\nu!}{j!(\nu-j)!}2^{\nu+j}\frac{\sigma_n^{2,\nu+j}(y)}{\overline{\xi}_n^{\nu+j}} \\
&=\frac{\overline{\phi}_n^2}{\overline{\xi}_n^2}+\frac{s_n^{2,0}(y)}{\overline{\xi}_n^2}-2\frac{2\overline{\phi}_n s_n^{1,1}(y)+s_n^{2,1}(y)}{\overline{\xi}_n^3}-\frac{\overline{\phi}_n^2 s_n^{0,2}(y)+2\overline{\phi}_n s_n^{1,2}(y)+s_n^{2,2}(y)}{\overline{\xi}_n^4}+ \\
&+4\frac{\overline{\phi}_n^2 s_n^{0,2}(y)+2\overline{\phi}_n s_m^{1,2}(y)+s_n^{2,2}(y)}{\overline{\xi}_n^4}+4\frac{S_n^{2,3}(y)}{\overline{\xi}_n^5}+\frac{S_n^{2,4}(y)}{\overline{\xi}_n^6}+ \\
&+\sum_{k=3}^{\nu-1}(-1)^k\sum_{j=0}^{k}\frac{k!}{j!(k-j)!}2^{k-j}\frac{S_n^{2,k+j}(y)}{\overline{\xi}_n^{k+j+2}}+(-1)^{\nu}\sum_{j=0}^{\nu}\frac{\nu!}{j!(\nu-j)!}2^{\nu+j}\frac{\sigma_n^{2,\nu+j}(y)}{\overline{\xi}_n^{\nu+j}} \\
&=\frac{\overline{\phi}_n^2}{\overline{\xi}_n^2}+\frac{s_n^{0,2}(y)}{\overline{\xi}_n^2}+3\frac{\overline{\phi}_n s_n^{0,2}(y)}{\overline{\xi}_n^4}-4\frac{\overline{\phi}_n s_n^{1,1}(y)}{\overline{\xi}_n^3}-2\frac{s_n^{2,1}(y)}{\overline{\xi}_n^3}+ \\
&+3\left(\frac{s_n^{2,2}(y)}{\overline{\xi}_n^4}+\frac{2\overline{\phi}_n s_n^{1,2}(y)}{\overline{\xi}_n^4}\right)+4\frac{S_n^{2,3}(y)}{\overline{\xi}_n^5}+\frac{S_n^{2,4}(y)}{\overline{\xi}_n^6}+ \\
&+\sum_{k=3}^{\nu-1}(-1)^k\sum_{j=0}^{k}\frac{k!}{j!(k-j)!}2^{k-j}\frac{S_n^{2,k+j}(y)}{\overline{\xi}_n^{k+j+2}}+(-1)^{\nu}\sum_{j=0}^{\nu}\frac{\nu!}{j!(\nu-j)!}2^{\nu+j}\frac{\sigma_n^{2,\nu+j}(y)}{\overline{\xi}_n^{\nu+j}}.
\end{aligned}
$$

Hence,

$$
E\left[\hat{f}_h^{\,2}(y)\right]=\frac{\overline{\phi}_n^2(y)}{\overline{\xi}_n^2}+\varphi_n(y)+\Gamma_n^{(\nu)}(y)+(-1)^{\nu}\Delta^{(\nu)}(y),
$$

where

$$
\begin{aligned}
&\varphi_n(y)=\frac{s_n^{0,2}(y)}{\overline{\xi}_n^2}+3\frac{\overline{\phi}_n s_n^{0,2}(y)}{\overline{\xi}_n^4}-4\frac{\overline{\phi}_n s_n^{1,1}(y)}{\overline{\xi}_n^3}, \\
&\Gamma_n^{(\nu)}(y)=-2\frac{s_n^{2,1}(y)}{\overline{\xi}_n^3}+3\left(\frac{s_n^{2,2}(y)}{\overline{\xi}_n^4}+\frac{2\overline{\phi}_n s_n^{1,2}(y)}{\overline{\xi}_n^4}\right)+4\frac{S_n^{2,3}(y)}{\overline{\xi}_n^5}+\frac{S_n^{2,4}(y)}{\overline{\xi}_n^6}+ \\
&+\sum_{k=3}^{\nu-1}(-1)^k\sum_{j=0}^{k}\frac{k!}{j!(k-j)!}2^{k-j}\frac{S_n^{2,k+j}(y)}{\overline{\xi}_n^{k+j+2}}, \\
&\Delta^{(\nu)}(y)=\sum_{j=0}^{\nu}\frac{\nu!}{j!(\nu-j)!}2^{\nu+j}\frac{\sigma_n^{2,\nu+j}(y)}{\overline{\xi}_n^{\nu+j}}.
\end{aligned}
$$

As we have done before for the mean, we must study the convergence order of these terms.

$$
\begin{aligned}
s_n^{2,0}(y)&=E\left[\left(\phi_n(y)-\overline{\phi}_n(y)\right)^2\right]=Var\left[\phi_n(y)\right]=E\left[\phi_n(y)^2\right]-\overline{\phi}_n(y)^2 \\
&=\frac{1}{n\mu^2}\left(K_h^2\circ\gamma\right)(y)+\frac{n-1}{n\mu^2}\left(K_h\circ f\right)^2(y)-\frac{1}{\mu^2}\left(K_h\circ f\right)^2(y) \\
&=\frac{1}{n\mu^2}\left(K_h^2\circ\gamma\right)(y)-\frac{1}{n\mu^2}\left(K_h\circ f\right)^2(y),
\end{aligned}
$$

$$
\begin{aligned}
&\overline{\phi}_n^2(y) s_n^{0,2}(y) = \overline{\phi}_n^2(y) E\left[\left(\xi_n - \overline{\xi}_n\right)^2\right] = \overline{\phi}_n^2(y) \frac{1}{n\mu}\left(E\left[\frac{1}{X}\right] - \frac{1}{\mu}\right) \\
&= \frac{1}{n\mu^3}\left(K_h \circ f\right)^2(y)\left(E\left[\frac{1}{X}\right] - \frac{1}{\mu}\right),
\end{aligned}
$$

$$
\begin{aligned}
&\overline{\phi}_n(y) s_n^{1,1}(y) = \overline{\phi}_n(y) E\left[\left(\phi_n(y) - \overline{\phi}_n(y)\right)\left(\xi_n - \overline{\xi}_n\right)\right] \\
&= \overline{\phi}_n(y) E\left[\phi_n(y)\left(\xi_n - \overline{\xi}_n\right)\right] - \overline{\phi}_n^2(y) E\left[\left(\xi_n - \overline{\xi}_n\right)\right] \\
&= \frac{1}{\mu}\left(K_h \circ f\right)(y)\left[\frac{1}{n\mu^2}\left(K_h \circ \gamma\right)(y) - \frac{1}{n\mu^2}\left(K_h \circ f\right)(y)\right] \\
&= \frac{1}{n\mu^3}\left(K_h \circ f\right)(y)\left(K_h \circ \gamma\right)(y) - \frac{1}{n\mu^3}\left(K_h \circ f\right)^2(y),
\end{aligned}
$$

$$
s_n^{2,1}(y) = E\left[\left(\phi_n(y) - \overline{\phi}_n(y)\right)^2\left(\xi_n - \overline{\xi}_n\right)\right] = O(1/n) - \frac{1}{n\mu^3}\left(K_h \circ \gamma\right)(y) - \frac{1}{n\mu^3}\left(K_h \circ f\right)^2(y),
$$

$$
\begin{aligned}
&s_n^{2,2}(y) = E\left[\left(\phi_n(y) - \overline{\phi}_n(y)\right)^2\left(\xi_n - \overline{\xi}_n\right)^2\right] \leq E\left[\left(\phi_n(y) - \overline{\phi}_n(y)\right)^4\right]^{1/2} E\left[\left(\xi_n - \overline{\xi}_n\right)^2\right]^{1/2} \\
&= O(1/n) O(1/n^{1/2}) = O(1/n^{3/2}).
\end{aligned}
$$

In the same way as with this last term and assuming that the l-th order centred moment of the variable $\frac{1}{Y} < +\infty$ with $l = 1, \ldots, 2\nu$, we obtain

$\overline{\phi}_n(y) s_n^{1,2}(y) = O(1/n^{3/2})$, $\quad S_n^{2,3}(y) = O(1/n^{5/2})$, $\quad S_n^{2,4}(y) = O(1/n^3)$, $\quad S_n^{2,k+j}(y) = O(1/n^{\frac{k+j}{2}+1})$ and $\sigma_n^{2,\nu+j}(y) = O(1/n^{\nu+j})$.

Finally, gathering all the addends properly, we get

$$
\begin{aligned}
E\left[\hat{f_h}^2(y)\right] &= \frac{\overline{\phi}_n^2(y)}{\overline{\xi}_n^2} + \varphi_n(y) + \Gamma_n^{(\nu)}(y) + (-1)^\nu \Delta^{(\nu)}(y) \\
&= \left(K_h \circ f\right)^2(y) + \frac{1}{n}\left(K_h^2 \circ \gamma\right)(y) - \frac{1}{n}\left(K_h \circ f\right)^2(y).
\end{aligned}
$$

Then,

$$
\begin{aligned}
Var\left[\hat{f}_h(y)\right] &= \left(K_h \circ f\right)^2(y) + \frac{1}{n}\left(K_h^2 \circ \gamma\right)(y) - \frac{1}{n}\left(K_h \circ f\right)^2(y) - \left(K_h \circ f\right)^2(y) + O(1/n) \\
&= \frac{1}{n}\left[\left(K_h^2 \circ \gamma\right)(y) - \left(K_h \circ f\right)^2(y)\right] + O(1/n)
\end{aligned}
$$

To get the MSE it is enough to realise that

$$
\mathrm{MSE}\left(\hat{f}_h(y)\right) = \mathrm{Bias}^2\left(\hat{f}_h(y)\right) + \mathrm{Var}\left(\hat{f}_h(y)\right),
$$

and apply a Taylor expansion as it is done with the kernel density estimator with complete data, then:

$$
\mathrm{MSE}\left(\hat{f}_h(y)\right) = \frac{1}{4} h^4\left(f''(y)\right)^2 \mu_2^2(K) + \frac{\gamma(y)}{nh} R(K) + o\left(h^4 + \frac{1}{nh}\right).
$$

B Proof of Theorem 2.3

We now obtain the MSE of the bootstrap estimator under the bootstrap distribution. To this aim we follow similar steps as in Appendix A. Remind that now, the estimator is given by (5), and it can be rewritten as follows:

$$\hat{f_h}^*(y) = \frac{\frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i^*} L_h\left(y - Y_i^*\right)}{\frac{1}{n}\sum_{j=1}^n \frac{1}{Y_j^*}} = \frac{\phi_n^*(y)}{\xi_n^*}.$$

From the expression above we compute the mean of the numerator and the denominator as follows:

$$\overline{\phi}_n^*(y) := E^*\left[\phi_n^*(y)\right] = E^*\left[\frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i^*} L_h\left(y - Y_i^*\right)\right] = \int \frac{1}{z} L_h(y-z)\hat{f}_{Y,g}(z)dz = \frac{1}{\hat{\mu}}\left(L_h \circ \hat{f}_g\right)(z).$$

$$\overline{\xi}_n^* := E^*[\xi_n^*] = E^*\left[\frac{1}{n}\sum_{j=1}^n \frac{1}{Y_j^*}\right] = \int \frac{1}{z}\hat{f}_{Y,g}(z)dz = \frac{1}{\hat{\mu}}.$$

Using the linearisation procedure in Collomb (1976) with $\nu \geq 2$ we have that $E^*[\hat{f_h}^*(y)] = \frac{\overline{\phi}_n^*(y)}{\overline{\xi}_n^*} + c_n^*(y) + c_n^{*(\nu)}(y) + \frac{(-1)^\nu \sigma_n^{*1,\nu}(y)}{\overline{\xi}_n^\nu}$ where

$$c_n^*(y) = \frac{\overline{\phi}_n^*(y) S_n^{*0,2} - \overline{\xi}_n^* S_n^{*1,1}}{{\overline{\xi}_n^*}^3} = \frac{\overline{\phi}_n^*(y) E^*\left[\left(\xi_n^* - \overline{\xi}_n^*\right)^2\right] - \overline{\xi}_n^* E^*\left[\phi_n^*(y)\left(\xi_n^* - \overline{\xi}_n^*\right)\right]}{{\overline{\xi}_n^*}^3},$$

$$c_n^{*(\nu)}(y) = \frac{s_n^{*1,2}(y)}{{\overline{\xi}_n^*}^3} + \sum_{k=3}^{\nu-1}(-1)^k \frac{S_n^{*1,k}(y)}{{\overline{\xi}_n^*}^{k+1}} = \frac{E^*\left[\left(\phi_n^*(y) - \overline{\phi}_n^*(y)\right)\left(\xi_n^* - \overline{\xi}_n^*\right)^2\right]}{{\overline{\xi}_n^*}^3} +$$

$+\sum_{k=3}^{\nu-1}(-1)^k \frac{E^*\left[\phi_n^*(y)\left(\xi_n^* - \overline{\xi}_n^*\right)^k\right]}{{\overline{\xi}_n^*}^{k+1}}$ and

$\sigma_n^{*1,\nu}(y) = E^*\left[\hat{f_h}^*(y)\left(\xi_n^* - \overline{\xi}_n^*\right)^\nu\right]$

To obtain the variance of the bootstrap estimator we compute

$$E^*[\hat{f_h}^{*2}(y)] = \frac{\overline{\phi}_n^{*2}(y)}{{\overline{\xi}_n^*}^2} + \varphi_n^*(y) + \Gamma_n^{*(\nu)}(y) + (-1)^\nu \Delta^{*(\nu)}(y),$$

with

$$\varphi_n^*(y) = \frac{s_n^{*2,0}(y)}{{\overline{\xi}_n^*}^2} + 3\frac{{\overline{\phi}_n}^{*2}(y) s_n^{*0,2}}{{\overline{\xi}_n^*}^4} - 4\frac{\overline{\phi}_n^*(y) s_n^{*1,1}}{{\overline{\xi}_n^*}^3},$$

$$\Gamma_n^{*(\nu)}(y) = -2\frac{s_n^{*2,1}(y)}{\overline{\xi}_n^{*3}} + 3\left(\frac{s_n^{*2,2}(y)}{\overline{\xi}_n^{*4}} + \frac{2\overline{\phi}_n^*(y)s_n^{*1,2}}{\overline{\xi}_n^{*4}}\right) + 4\frac{S_n^{*2,3}(y)}{\overline{\xi}_n^{*5}} + \frac{S_n^{*2,4}(y)}{\overline{\xi}_n^{*6}} +$$
$+\sum_{k=3}^{\nu-1}(-1)^k\sum_{j=0}^k \frac{k!2^{k-j}}{j!(k-j)!}\frac{S_n^{*2,k+j}(y)}{\overline{\xi}_n^{*k+j+2}}$ and

$$\Delta^{*(\nu)}(y) = \sum_{j=0}^{\nu}\frac{\nu!}{j!(\nu-j)!}2^{\nu-j}\frac{\sigma_n^{*2,\nu+j}(y)}{\overline{\xi}_n^{*\nu+j}}.$$

Here the dominant terms are $\frac{\overline{\phi}_n^{*2}(y)}{\overline{\xi}_n^{*2}} + \frac{s_n^{*2,0}(y)}{\overline{\xi}_n^{*2}}$, which are the ones we need to study.

$$\frac{\overline{\phi}_n^{*2}(y)}{\overline{\xi}_n^{*2}} = \frac{\left(\frac{1}{\hat{\mu}}(L_h\circ\hat{f}_g)(y)\right)^2}{\left(\frac{1}{\hat{\mu}}\right)^2}$$

$$\begin{aligned}
s_n^{*2,0}(y) &= E^*\left[\left(\phi_n^*(y)-\overline{\phi}_n^*(y)\right)^2\right] = Var^*\left[\phi_n^*(y)\right] = E^*\left[\phi_n^{*2}(y)\right] - E^*\left[\phi_n^*(y)\right]^2 \\
&= E^*\left[\phi_n^{*2}(y)\right] - \overline{\phi}_n^{*2}(y) = \frac{1}{n\hat{\mu}^2}(L_h^2\circ\hat{\gamma}_g)(y) + \frac{n-1}{n\hat{\mu}^2}(L_h\circ\hat{f}_g)^2(y) - \frac{1}{\hat{\mu}^2}(L_h\circ\hat{f}_g)^2(y) \\
&= \frac{1}{n\hat{\mu}^2}(L_h^2\circ\hat{\gamma}_g)(y) - \frac{1}{n\hat{\mu}^2}(L_h\circ\hat{f}_g)^2(y)
\end{aligned}$$
$$\Rightarrow \frac{s_n^{*2,0}(y)}{\overline{\xi}_n^{*2}} = \hat{\mu}^2 s_n^{*2,0}(y) = \frac{1}{n}(L_h^2\circ\hat{\gamma}_g)(y) - \frac{1}{n}(L_h\circ\hat{f}_g)^2(y),$$

taking into account that $E^*[\phi_n^{*2}(y)] = \frac{1}{n\hat{\mu}}(L_h^2\circ\hat{\gamma}_g)(y) + \frac{n-1}{n\hat{\mu}^2}(L_h\circ\hat{f}_g)^2(y)$.

Then, as the other terms are negligible, we get

$$Var^*\left[\hat{f}_h^*(y)\right] = \frac{1}{n}\left[(L_h^2\circ\hat{\gamma}_g)(y) - (L_h\circ\hat{f}_h)^2(y)\right].$$

Under the regularity conditions previously established we get that

$$\text{MSE}^*(\hat{f}_h^*(y)) = \frac{1}{4}h^4\left(\hat{f}_g''(y)\right)^2\mu_2^2(K) + \frac{\hat{\gamma}_g(y)}{nh}R(L) + o_P\left(h^4+\frac{1}{nh}\right).$$

C Proof of Theorem 2.6

This proof has basically the same aim as the one of Theorem 2.3 (Appendix B), with the particularity that the generation of the bootstrap sample is made differently.

Let us remind the expression of the bootstrap estimator

$$\hat{f}_h^*(y) = \frac{\frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i^*}L_h\left(y-Y_i^*\right)}{\frac{1}{n}\sum_{j=1}^n\frac{1}{Y_j^*}} = \frac{\phi_n^*(y)}{\xi_n^*}.$$

Firstly we will obtain the mean of the bootstrap estimator. Following again the linearisation procedure in Collomb (1976) and the previous proof, we only need the mean of the numerator and the denominator, so:

$$\overline{\phi}_n^*(y) := E^*\left[\phi_n^*(y)\right] = E^*\left[\frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i^*} L_h\left(y - Y_i^*\right)\right] = \int \frac{1}{z} L_h(y-z)\tilde{f}_{K,g}(z)dz.$$

$$\overline{\xi}_n^* := E^*[\xi_n^*] = E^*\left[\frac{1}{n}\sum_{j=1}^n \frac{1}{Y_j^*}\right] = \int \frac{1}{z}\tilde{f}_{K,g}(z)dz.$$

Remind that in this context, $\tilde{f}_{K,g}$ denote the common kernel density estimator with kernel K and bandwidth g.

Hence,

$$E^*\left[\hat{f}_h^*(y)\right] = \frac{\overline{\phi}_n^*(y)}{\overline{\xi}_n^*} + O_P\left(1/n\right) = \frac{\int \frac{1}{z} L_h(y-z)\tilde{f}_{K,g}(z)dz}{\int \frac{1}{z}\tilde{f}_{K,g}(z)dz} + O_P\left(1/n\right).$$

To compute the variance we follow again the previous proof, and taking only into account the dominant terms we get

$$E^*[\hat{f}_h^{*^2}(y)] = \frac{\overline{\phi}_n^{*^2}(y)}{\overline{\xi}_n^{*\,2}} + \frac{s_n^{*^{2,0}}(y)}{\overline{\xi}_n^{*\,2}} + O_P(1/n), \quad \text{where}$$

$$\frac{\overline{\phi}_n^{*^2}(y)}{\overline{\xi}_n^{*\,2}} = \frac{\left(\int \frac{1}{z} L_h(y-z)\tilde{f}_{K,g}(z)dz\right)^2}{\left(\int \frac{1}{z}\tilde{f}_{K,g}(z)dz\right)^2} \quad \text{and}$$

$$\begin{aligned} s_n^{*^{2,0}}(y) &= E^*\left[\left(\phi_n^*(y) - \overline{\phi}_n^*(y)\right)^2\right] = Var^*\left[\phi_n^*(y)\right] = E^*\left[\phi_n^{*^2}(y)\right] - E^*\left[\phi_n^*(y)\right]^2 \\ &= E^*\left[\phi_n^{*^2}(y)\right] - \overline{\phi}_n^{*^2}(y) = \frac{1}{n}\int \frac{1}{z^2} L_h^2(y-z)\tilde{f}_{K,g}(z)dz - \frac{1}{n}\left(\int \frac{1}{z} L_h(y-z)\tilde{f}_{K,g}(z)dz\right)^2, \end{aligned}$$

considering that $E^*[\phi_n^{*2}(y)] = \dfrac{1}{n}\displaystyle\int \frac{1}{z^2} L_h^2(y-z)\tilde{f}_{K,g}(z)dz - \frac{n-1}{n}\left(\int \frac{1}{z} L_h(y-z)\tilde{f}_{K,g}(z)dz\right)^2$.

Then we get

$$Var^*\left[\hat{f}_h^{\,*}(y)\right] = \frac{1}{n}\int \frac{1}{z^2} L_h^2(y-z)\tilde{f}_{K,g}(z)dz - \frac{1}{n}\left(\int \frac{1}{z} L_h(y-z)\tilde{f}_{K,g}(z)dz\right)^2.$$

Finally, just noting that MSE* can be computed as the sum of the squared bias and the variance we obtain the final equation.

D Proof of Theorem 3.1

Along this proof we will calculate the mean and the variance of $R(\hat{f}_g'')$ as an estimator of $R(f'')$in order to determine its MSE and an expression of the pilot bandwidth g. To this purpose we use the U-statistics theory and its projections, as it has been done for complete data in Cao (1990), as well as some common statistical results.

We will start by calculating its mean. First of all rewrite

$$R(\hat{f}_g'') = \int \left(\hat{f}_g''(y)\right)^2 dy = n^{-1}g^{-5}\int \left(L''(u)\right)^2 du \left(\frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i^2}\right)\hat{\mu}^2 +$$
$$+ n^{-2}g^{-6}\hat{\mu}^2 \sum_{i\neq j}\frac{1}{Y_i}\frac{1}{Y_j}\int L''\left(\frac{y-Y_i}{g}\right)L''\left(\frac{y-Y_j}{g}\right)dy.$$

Hence,

$$E\left[\int \hat{f}_g''^2(y)dy\right]$$
$$= n^{-1}g^{-5}\int \left(L''(u)\right)^2 du E\left[\frac{\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i^2}\right)}{\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i}\right)^2}\right] + n^{-2}g^{-6}E\left[\frac{\sum_{i\neq j}\frac{1}{Y_i}\frac{1}{Y_j}\int L''\left(\frac{y-Y_i}{g}\right)L''\left(\frac{y-Y_j}{g}\right)dy}{\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i}\right)^2}\right]$$
$$= n^{-1}g^{-5}\int \left(L''(u)\right)^2 duc\mu + n^{-1}(n-1)g^{-6}\mu^2\int \mu_g^2(y)dy + o(n^{-2}g^{-5}), \tag{8}$$

where $\mu_g(y) := E\left[\frac{1}{Y}L''\left(\frac{y-Y}{g}\right)\right]$ and we use that

$$E\left[\left(\frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i}\right)^2\right] = \frac{1}{\mu^2} + O(n^{-1}) \quad \text{and } E\left[\frac{1}{n}\sum_{i=1}^n \frac{1}{Y_i^2}\right] = E\left[\frac{1}{Y^2}\right] = \frac{c}{\mu},$$

that have been obtained using Taylor expansions.

Moreover, using again Taylor expansion and the regularity conditions imposed on L and f, we can rewrite

$$\mu_g(y) = \frac{1}{\mu}\left(g^3 f''(y) + \frac{1}{6}g^5\int u^4 L''(u)\int (1-t)^3 f^{(iv}(y-gut)dtdu\right),$$

and applying Fubini's theorem and Cauchy-Schwarz inequality, we obtain

$$\int \mu_g^2(y)dy = \frac{1}{\mu^2}\left(g^6\int f''^2(y)dy + g^8\mu_2(L)\int (f'''(y))^2dy + o(g^9)\right). \tag{9}$$

Then, replacing this value in (8), we get the mean of the estimator as follows:

$$E\left[\int \hat{f}_g''^2(y)dy\right] = n^{-1}g^{-5}c\mu\int \left(L''(u)\right)^2 du +$$
$$+ n^{-1}(n-1)\int (f''(y))^2dy + n^{-1}(n-1)g^2\mu_2(L)\int (f'''(y))^2dy + o(g^3). \tag{10}$$

Once having the mean, the bias can be immediately obtained:

$$Bias\left[R\left(\hat{f}_h''\right)\right] = n^{-1}g^{-5}c\mu\int\left(L''(u)\right)^2 du + g^2\mu_2(L)\int (f'''(y))^2dy + o(n^{-1}) + o(g^3). \quad (11)$$

Next step is calculating the variance. For this aim we need to rewrite the expression of the estimator in an appropriate and different way.

$$\begin{aligned}
R\left(\hat{f}_g''\right) &= n^{-1}g^{-5}\int (L''(u))^2du\hat{\mu}^2\left(\frac{1}{n}\sum_{i=1}\frac{1}{Y_i^2}\right) + n^{-2}g^{-6}\hat{\mu}^2\sum_{i\neq j}\frac{1}{Y_i}\frac{1}{Y_j}\int L''\left(\frac{y-Y_1}{g}\right)L''\left(\frac{y-Y_j}{g}\right)dy,\\
&= n^{-1}g^{-5}\int (L''(u))^2du\hat{\mu}^2\left(\frac{1}{n}\sum_{i=1}\frac{1}{Y_i^2}\right) + n^{-2}g^{-6}\hat{\mu}^2\sum_{i\neq j}\left[H_n(Y_i,Y_j) + \int\left(\frac{1}{Y_i}L''\left(\frac{y-Y_i}{g}\right) - \mu_g(y)\right)\mu_g(y) + \right.\\
&+ \int\left(\frac{1}{Y_j}L''\left(\frac{y-Y_j}{g}\right) - \mu_g(y)\right)\mu_g(y) + \int\mu_g^2(y)\\
&= n^{-1}g^{-5}\int (L''(u))^2du\hat{\mu}^2\left(\frac{1}{n}\sum_{i=1}\frac{1}{Y_i^2}\right) + n^{-1}(n-1)g^{-6}\hat{\mu}^2\int\mu_g^2(y)dy +\\
&+ n^{-2}g^{-6}\hat{\mu}^2\sum_{i\neq j}H_n(Y_i,Y_j) + n^{-2}g^{-6}\hat{\mu}^2 2(n-1)\sum_{i=1}^{n}W_i,
\end{aligned} \quad (12)$$

where $W_i = \int\left(\frac{1}{Y_i}L''\left(\frac{y-Y_i}{g}\right) - \mu_g(y)\right)\mu_g(y)dy \quad i = 1,\ldots,n$,
$H_n(Y_i,Y_j) := \int\left(\frac{1}{Y_i}L''\left(\frac{y-Y_i}{g}\right) - \mu_g(y)\right)\left(\frac{1}{Y_j}L''\left(\frac{y-Y_j}{g}\right) - \mu_g(y)\right)$, and we use that

$$\begin{aligned}
\int L''\left(\frac{y-Y_1}{g}\right)L''\left(\frac{y-Y_j}{g}\right)dy &= H_n(Y_i,Y_j) + \int\left(\frac{1}{Y_i}L''\left(\frac{y-Y_i}{g} - \mu_g(y)\right)\mu_g(y)dy\right) +\\
&+ \int\left(\frac{1}{Y_j}L''\left(\frac{y-Y_j}{g} - \mu_g(y)\right)\mu_g(y)dy\right) + \int\mu_g^2(y)dy \quad \text{and}\\
\int\frac{1}{Y_i}L''\left(\frac{y-Y_i}{g}\right)\frac{1}{Y_j}L''\left(\frac{y-Y_j}{g}\right) &= H_n(Y_i,Y_j) + \int\left(\frac{1}{Y_i}L''\left(\frac{y-Y_i}{g}\right) - \mu_g(y)\right)\mu_g(y)dy +\\
&+ \int\left(\frac{1}{Y_j}L''\left(\frac{y-Y_j}{g}\right) - \mu_g(y)\right)\mu_g(y)dy + \int\mu_g^{"}(y)dy
\end{aligned}$$

Now we calculate the variance of each term; for the first two we need to use common statistical techniques while for the others we will need to do a more complex expansion.

$$Var\left[n^{-1}g^{-5}\int (L''(u))^2du\hat{\mu}^2\left(\frac{1}{n}\sum_{i=1}\frac{1}{Y_i^2}\right)\right] = n^{-2}g^{-10}\left(\int (L''(u))^2du\right)^2 Var\left[\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i^2}}{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}\right)}\right]$$
$$= o(n^{-3}g^{-10}).$$

$$Var\left[n^{-1}(n-1)g^{-6}\hat{\mu}^2\int\mu_g^2(y)dy\right] = n^{-2}(n-1)^2g^{-12}\int\mu_g(y)dyVar\left[\frac{1}{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{Y_i}\right)^2}\right]$$
$$= o(n^{-1}g^{-12}).$$

Before getting involved in the variance of the third term we need some previous developments related to $H_n(Y_i, Y_j)$.

Firstly, we obtain the expression of $\int\int \mu_g(x)\mu_g(y)\int L^{''}\left(\frac{x-z}{g}\right)L^{''}\left(\frac{y-z}{g}\right)\gamma(z)dzdxdy$ using Taylor expansions on f and γ:

$$\int\int \mu_g(x)\mu_g(y)\int L^{''}\left(\frac{x-z}{g}\right)L^{''}\left(\frac{y-z}{g}\right)\gamma(z)dzdxdy = \frac{g^8}{\mu^2}\int\int\int\int\int\int_0^1\int_0^1 u_1^2L^{''}(u_1)u_2^2L^{''}(u_2)$$
$$(1-t_1)(1-t_2)f^{''}(x-gu_1t_1)f^{''}(x+gw-gu_2t_2)L^{''}(v)L^{''}(v+w)\gamma(x-gv)dt_1dt_2du_1du_2dvdwdx$$
$$= I_0 - I_1 + I_2 - I_3 + I_4, \tag{13}$$

with

$$I_j = \frac{1}{\mu^2}\frac{g^{8+j}}{j!}\int u_1^2L^{''}(u_1)\int u_2^2L^{''}(u_2)\int_0^1(1-t_1)\int_0^1(1-t_2)\int f^{''}(x-gu_1t_1)\gamma^{(j)}(x)\cdot$$
$$\cdot\int f^{''}(x+gw-gu_2t_2)N_j(w)dxdwdt_1dt_2du_1du_2 \quad j=0,\ldots,3$$

$$I_4 = \frac{1}{\mu^2}\frac{g^{12}}{6}\int u_1^2L^{''}(u_1)\int u_2^2L^{''}(u_2)\int_0^1(1-t_1)\int_0^1(1-t_2)\int f^{''}(x-gu_1t_1)\cdot$$
$$\cdot\int f^{''}(x+gw-gu_2t_2)\int v^4L^{''}(v)L^{''}(v+w)\int_0^1(1-z^3)\gamma^{(iv}(x-gvz)dzdvdwdxdt_2dt_1du_2du_1.$$

From these expressions, we obtain:

$$I_0 = \frac{g^{12}}{\mu^2}\int \gamma(y)f^{''}(y)f^{(vi}(y)dy + o(g^{12}),$$
$$I_1 = \frac{-2g^{12}}{\mu^2}\int \gamma^{'}(y)f^{''}(y)f^{(v}(y)dy + o(g^{12}),$$
$$I_2 = \frac{g^{12}}{\mu^2}\int\left(\gamma^{''}(y)\right)^2 f^{(iv}(y)dy + o(g^{12}),$$
$$I_3 = o(g^{12}) \quad \text{and} \quad I_4 = o(g^{12}).$$

Hence, getting back on (13), we have

$$\int\int \mu_g(x)\mu_g(y)\int L^{''}\left(\frac{x-z}{g}\right)L^{''}\left(\frac{y-z}{g}\right)\gamma(z)dzdxdy = \frac{g^{12}}{\mu^2}\int\gamma(y)\left(f^{(iv}(y)\right)^2dy + o(g^{12}).$$

Secondly, we need to remark that

$$\int\int\left[\int L^{''}\left(\frac{x-z}{g}\right)L^{''}\left(\frac{y-z}{g}\right)dz\right]^2dxdy = g^3\left(\int\gamma^2(z)dz\right)\left(\int(L^{''}\circ L^{''})^2(v)dv\right) + o(g^3).$$

And now, we are in a position to obtain the expression of $E\left[H_n(Y_i, Y_j)^2\right]$:

$$
\begin{aligned}
E\left[H_n(Y_i,Y_j)^2\right] &= \int\int\left(\frac{1}{\mu^2}\int L''\left(\frac{x-z}{g}\right)L''\left(\frac{y-z}{g}\right)\gamma(z)dz-\mu_g(x)\mu_g(y)\right)^2 dxdy\\
&= \frac{1}{\mu^4}\int\int\left(\int L''\left(\frac{x-z}{g}\right)L''\left(\frac{y-z}{g}\right)\gamma(z)dz\right)^2 dxdy-\\
&-\frac{2}{\mu^4}\int\int\mu_g(x)\mu_g(y)\int L''\left(\frac{x-z}{g}\right)L''\left(\frac{y-z}{g}\right)\gamma(z)dzdxdy+\frac{1}{\mu^4}\left(\int\mu_g^2(y)dy\right)^2\\
&= \frac{1}{\mu^4}\left(g^3\left(\int\gamma^2(z)dz\right)\left(\int(L''\circ L'')^2(v)dv\right)\right)-\frac{2g^{12}}{\mu^6}\int\gamma(y)(f^{(iv}(y))^2dy+\\
&+\frac{g^{12}}{\mu^8}\int f(y)(f^{(iv}(y))^2dy-\frac{g^{12}}{\mu^8}\left(\int f(y)f^{(iv}(y)dy\right)^2+o(g^{12}).
\end{aligned}\tag{14}
$$

Remember, that we were doing all these calculations in order to be able to obtain the variance of the third term of (12), so

$$
\begin{aligned}
Var\left[n^{-2}g^{-6}\hat{\mu}^2\sum_{i\neq j}H_n(Y_i,Y_j)\right] &= 2n^{-3}g^{-12}\mu^4E\left[H_n^2(Y_i,Y_j)\right]+o(n^{-5}g^{-12})\\
&= 2n^{-2}g^{-9}\int\gamma^2(z)dz\int(L''\circ L'')^2(v)dv+o(n^{-3}g^{-9}).
\end{aligned}
$$

The variance of the fourth term of (12) is calculated as follows:

$$
\begin{aligned}
&Var\left[n^{-2}g^{-6}\hat{\mu}^2 2(n-1)\sum_{i=1}^{n}W_i\right] = 4n^{-4}(n-1)^2g^{-12}\left(Var\left[\sum_{i=1}^{n}W_i\right]\mu^4+o(n^{-1})\right)\\
&= 4n^{-3}(n-1)^2g^{-12}\mu^4\left[\frac{g^{12}}{\mu^4}\int\gamma(y)\left(f^{(iv}(y)\right)^2dy-\frac{g^{-12}}{\mu^4}\left(\left(\int(f'''(y))^2dy\right)^2+o(g^{12})\right)\right],
\end{aligned}
$$

where we have used that W_i are centred and independent variables, as well as the expression of its second order moment.

Lastly, it can be seen, using Cauchy-Schwarz inequality, that the covariate terms between the addends of (12) are negligible.

Then, we finally obtain

$$
\begin{aligned}
Var\left[R\left(\hat{f}_f''\right)\right] &= 2n^{-2}g^{-9}\int\gamma^2(y)dy\int(L''\circ L'')^2(v)dv+o(n^{-2}g^{-9})+\\
&+O(n^{-1})+O(n^{-3/2}n^{-9/2}).
\end{aligned}\tag{15}
$$

Gathering (11) and (15), we have

$$
\begin{aligned}
\mathrm{MSE}\left(R(\hat{f}_g'')\right) &= 2n^{-2}g^{-9}\int\gamma^2(y)dy\int(L''\circ L'')^2(v)dv+n^{-2}g^{-10}\left(\int(L''(y))^2dy\right)^2c^2\mu^2+\\
&+g^4\mu_2^2(L)\left(\int(f'''(y))^2dy\right)^2+2n^{-1}g^{-3}\mu_2(L)c\mu\int(L''(y))^2dy\int(f'''(y))^2dy+\\
&+o(n^{-3}g^{-9})+o(n^{-2}g^{-9})+O(n^{-1})+O(n^{-3/2}g^{-9/2})\\
&= A^2n^{-2}g^{-10}+B^2g^4+2ABn^{-1}g^{-3}+o(n^{-2}g^{-9})+O(n^{-1})+\\
&+O(n^{-3/2}g^{-9/2}),
\end{aligned}
$$

where $A := c\mu \int (L''(u))^2 du$ and $B := \mu_2(L) \int (f'''(y))^2 dy$.

Then, its asymptotic version is

$$\text{AMSE}\left(R(\hat{f}_h'')\right) = (An^{-1}g^{-5} + Bg^2)^2,$$

and the value of the bandwidth g minimising the quantity above is

$$g_0 = \arg\min_g \text{AMSE} = d_o n^{-1/7},$$

with $d_0 = \left(\frac{5}{2}AB\right)^{1/7}$.