Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Statistical Planning and Inference 132 (2005) 3–19

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Mean residual life estimation

Belkacem Abdous, Alexandre Berred*

Université Laval, Département de médecine sociale et préventive, Pavillon de l'Est, local 1138A, Ste-Foy, Québec, G1K 7P4, Canada

Received 12 November 2002; received in revised form 8 March 2004
Available online 12 August 2004

Abstract

The mean residual life function is of interest in many fields such as reliability, survival analysis, actuarial studies, etc. Given a sample from an unknown distribution function, we use the local linear fitting technique to estimate the corresponding mean residual life function. The limit behaviour of the obtained estimator is presented.

MSC: Primary 62G05; Secondary 62G20, 62G30

Keywords: Mean residual life function; Expected remaining life function; Mean excess function; Kernel smoothing; Non parametric estimation; Survival function

1. Introduction

Let X be a random variable with distribution function F, survival function $\bar{F}(x)=1-F(x)$ and such that $E(X)<\infty$. The mean residual life (MRL) function $e(x)$ of X is defined by (see, e.g., Kotz and Shanbhag (1980), Hall and Wellner (1981), Guess and Proschan (1988))

$$e(x) = E(X-x|X>x) = \begin{cases} \frac{\int_x^\infty \bar{F}(y)\mathrm{d}y}{\bar{F}(x)} & \text{if } \bar{F}(x)>0 \\ 0 & \text{otherwise.} \end{cases}$$

The MRL function, also called expected remaining life function or mean excess function, has been widely studied in life time random variables context. It plays an important role in

* Corresponding author. Lab. de Mathématiques Appliquées, BP 540, 25 rue Philippe Lebon, Le Havre Cedex 76058, France. Tel.: +33-232-744322; fax: +33-232 -744314.

E-mail address: berred@univ-lehavre.fr (A. Berred).

0378-3758/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.06.012

many fields such as industrial reliability, biomedical science, life insurance and demography among others. For a detailed discussion and statistical applications of the MRL function, we refer to Embrechts et al. (1997). Given a sample $X_1, \ldots, X_n$ from F, a natural estimate of $e(x)$ is given by its empirical version

$$e_n(x) = \frac{\int_x^\infty \bar{F}_n(y)\mathrm{d}y}{\bar{F}_n(x)} \mathbb{1}_{[X_{(n)}>x]}, \tag{1.1}$$

where $X_{(n)} = \max_{1 \leqslant i \leqslant n} X_i$ and $\bar{F}_n(x)$ is the empirical survival function. Several authors studied the properties of $e_n(x)$, see e.g. Csörgő and Zitikis (1996) and the references therein. Despite its good asymptotic behavior, the empirical MRL function $e_n(x)$ is a rough estimate of $e(x)$ and it does not take into account the smoothness of $e(x)$. Some nonparametric estimates of $e(x)$ have been proposed in the literature. For instance, Ruiz and Guillamòn (1996) estimated the numerator in $e(x)$ by a recursive kernel estimate and used the empirical survival function to estimate the denominator, while Chaubey and Sen (1999) used the so called Hille's theorem (Hille, 1948) to smooth both the numerator and denominator in $e(x)$. To smooth $e_n(x)$, one can view $e(x)$ either as a ratio of two functions or as a function on its own. There are many smoothing techniques which can be applied to this problem. In this work, we adopt the classical kernel smoothing method for its simplicity. In the sequel, $K(\cdot)$ will stand for an arbitrary probability density and its corresponding survival function will be denoted by $\mathbb{K}(t) = \int_t^\infty K(u)\,\mathrm{d}u$. We will also need a sequence of smoothing parameters (or bandwidths) $h = h_n > 0$. If we view $e(x)$ as a ratio of two functions, we can use the following kernel estimator:

$$\hat{e}_n(x) = \frac{\int_x^\infty \sum_{i=1}^n \mathbb{K}\left(\frac{u-X_i}{h}\right)\mathrm{d}u}{\sum_{i=1}^n \mathbb{K}\left(\frac{x-X_i}{h}\right)}. \tag{1.2}$$

While if we view $e(x)$ as a function on its own, we can convolve $e_n(x)$ with $K_h(\cdot) = K(\cdot/h)/h$ and obtain the estimator

$$\widetilde{e}_n(x) = \int_{-\infty}^{+\infty} K_h(x-u)e_n(u)\mathrm{d}u. \tag{1.3}$$

For a discussion of ratio functions estimates see Patil et al. (1994). Note that if the right end of F's support is finite, i.e. $U_F < \infty$, with $U_F = \sup\{x : F(x) < 1\}$, then the MRL will be right truncated. In this case, both $\hat{e}_n$ and $\widetilde{e}_n$ will suffer from boundary effects. While, if the support of F has a finite left end $L_F = \inf\{x : F(x) > 0\}$, then $e(x) = E(X) - x$ for $x \leqslant L_F$, consequently, the boundary effect problem does not occur for L_F. There are many ways to bypass the problem of boundary effects in nonparametric estimation (see, e.g. Wand and Jones, 1995). If one is willing to accept negative values of the kernel estimates, a higher order kernel can be used. This means that K is allowed to take negative values in order to cancel some of its higher order moments. A successful and popular method which reduces the bias and alleviates the problem of boundary effects is the so-called local polynomial fitting (see Fan and Gijbels (1996) for an introduction and references). For ease of presentation, we will limit ourselves to the local linear fitting only. The idea behind this procedure relies on the fact that if the MRL function $e(\cdot)$ is continuously differentiable at x, then by Taylor's Theorem

$$e(y) \approx e(x) + (y-x)e'(x).$$

In other words, $e(y)$ is approximately linear in the neighborhood of x. For this reason, we seek for a linear polynomial which minimizes the following weighted least squares problem

$$\int_{-\infty}^{U_F} K_h(y-x)[e_n(y) - a_0 - a_1(y-x)]^2 \mathrm{d}y, \tag{1.4}$$

where U_F is possibly infinite and is assumed to be known. Without loss of generality, the kernel K will be taken as a symmetric probability density function with support $[-1, 1]$. Then, standard calculation shows that the coefficients $\hat{a}_{0n}(x)$ and $\hat{a}_{1n}(x)$ of the unique polynomial which minimizes (1.4) are given by

$$\hat{a}_{in}(x) = \begin{cases} \frac{1}{h^i}\int_{-1}^{1} W_i(v,1)e_n(x+hv)\mathrm{d}v & \text{if } U_F = +\infty \text{ or if } U_F < \infty \\ & \text{and } x < U_F - h \\ \frac{1}{h^i}\int_{-1}^{\vartheta} W_i(v,\vartheta)e_n(x+hv)\mathrm{d}v & \text{if } (U_F < \infty \text{ and } x = U_F - \vartheta h, \\ 0 & \text{if } U_F < \infty \text{ and } x \geqslant U_F + h. \end{cases}$$

where $i = 0, 1$, the constant ϑ is fixed in $]-1, 1]$ and the weight functions W_i are defined by

$$W_0(v,\eta) = \frac{K(v)\int_{-1}^{\eta}(u^2 - uv)K(u)\,\mathrm{d}u}{\int_{-1}^{\eta}\int_{-1}^{\eta}(u^2-uv)K(u)K(v)\mathrm{d}u\,\mathrm{d}v} 1_{[-1,\eta]}(v)$$

$$W_1(v,\eta) = \frac{K(v)\int_{-1}^{\eta}(v-u)K(u)\mathrm{d}u}{\int_{-1}^{\eta}\int_{-1}^{\eta}(u^2-uv)K(u)K(v)\mathrm{d}u\,\mathrm{d}v} 1_{[-1,\eta]}(v),$$

where η stands for either 1 or ϑ. By analogy with Taylor's Theorem, $\hat{a}_{0n}(x)$ and $\hat{a}_{1n}(x)$ are expected to estimate $e(x)$ and $e'(x)$ respectively. In the sequel, we will put emphasis on the estimate $\hat{a}_{0n}(x)$ only even if the asymptotic properties of both estimates can be obtained. A brief description of the main results of this work is as follows. In Section 2, we investigate some asymptotic properties of $\hat{a}_{0n}(x)$. Section 3 gives some guidance on the smoothing parameter selection problem. Section 4 is devoted to some simulation results. Proofs are postponed until Section 5.

2. Asymptotic properties of $\hat{a}_{0n}(x)$

In this section, we mainly focus on the asymptotic behavior of the bias and variance of $\hat{a}_{0n}(x)$ and establish its asymptotic normality. First, we show in the following Theorem that under mild regularity conditions, $\hat{a}_{0n}(x)$ is asymptotically unbiased for $e(x)$ and that its variance goes to 0 as $n \to \infty$.

Theorem 2.1. *Suppose that X has a continuous distribution function F with a finite second moment. Assume that the kernel K is a symmetric probability density with support* $[-1, 1]$ *and that the bandwidth satisfies* $h \to 0$ *as* $n \to \infty$. *Then, for any x*

$$\lim_{n\to\infty} \left[E(\hat{a}_{0n}(x)) - e(x)\right] = 0. \tag{2.1}$$

$$\lim_{n\to\infty} \operatorname{Var}(\hat{a}_{0n}(x)) = 0. \tag{2.2}$$

The proof of these results relies on the following lemma which gives closed expressions of the expectation and the variance of the empirical MRL $e_n(x)$.

Lemma 2.1. *Let X be a random variable having a finite second moment and a c.d.f. F with possibly infinite right end of support* U_F. Let $u \leqslant v < U_F$ *be fixed, then*

$$E(e_n(u)) = e(u)(1 - F^n(u)), \tag{2.3}$$

$$\begin{aligned}\operatorname{Cov}(e_n(u), e_n(v)) &= S_n(u, v)\left(2\frac{\mu_1(v)}{\bar{F}(v)} - e^2(v)\right) + e(u)e(v)(1 - F^n(u))F^n(v) \\ &\quad + \frac{F^n(v) - F^n(u)}{F(v) - F(u)}\mu_0(v)\left(v - u + \frac{\mu_0(v) - \mu_0(u)}{\bar{F}(v)}\right),\end{aligned} \tag{2.4}$$

where $\bar{F}(\cdot)$ *stands for the survival function and*

$$S_n(y, z) = \sum_{i=1}^{n} \frac{F^{n-i}(y)(1 - F^i(z))}{i}, \quad \mu_i(y) = \int_y^\infty (t - y)^i \bar{F}(t)\mathrm{d}t,$$

$$for \quad i = 0, 1.$$

Besides, upon imposing more restrictive conditions on F, one can exhibit rates of convergence in (2.1)–(2.2) and also show that the estimator $\hat{a}_{1n}(x)$ is asymptotically unbiased and convergent. Details are omitted because they are technical, at the most.

Theorem 2.2. *Under assumptions of Theorem* 2.1 *and* $E|X|^3 < \infty$, *the standardized version of* $\hat{a}_{0n}(x)$ *is asymptotically normal* $N(0, 1)$.

3. Automatic selection of the smoothing parameter

Given the crucial importance of the bandwidth h, we give some guidance to select this parameter appropriately. A common criterion to assess the performance of a nonparametric estimate is the MSE. By using Lemma 2.1, we obtain

$$
\begin{aligned}
\mathrm{MSE}(x,h) &= E(\hat{a}_{0n}(x)-e(x))^2 \\
&= \left(\int_{-1}^{\eta} W_0(u,\eta)e(x+hu)(1-F^n(x+hu))\mathrm{d}u - e(x)\right)^2 \\
&\quad + 2\int_{\Omega} W_0(u,\eta)W_0(v,\eta)S_n(x+hu,x+hv) \\
&\quad \times \left(2\frac{\mu_1(x+hv)}{\bar{F}(x+hv)} - e^2(x+hv)\right)\mathrm{d}u\,\mathrm{d}v \\
&\quad + 2\int_{\Omega} W_0(u,\eta)W_0(v,\eta)e(x+hu)e(x+hv)(1-F^n(x+hu)) \\
&\quad \times F^n(x+hv)\mathrm{d}u\,\mathrm{d}v \\
&\quad + 2\int_{\Omega} W_0(u,\eta)W_0(v,\eta)\frac{F^n(x+hv)-F^n(x+hu)}{F(x+hv)-F(x+hu)}\mu_0(x+hv) \\
&\quad \times \left\{h(v-u) + \frac{\mu_0(x+hv)-\mu_0(x+hu)}{\bar{F}(x+hv)}\right\}\mathrm{d}u\,\mathrm{d}v
\end{aligned}
$$

where

$$
\eta = \min\left(\frac{U_F - x}{h}, 1\right) = \begin{cases} 1 & \text{if } U_F = +\infty \text{ or if } U_F < \infty \text{ with } x < U_F - h \\ \vartheta & \text{if } U_F < \infty \text{ with } x = U_F - \vartheta h \text{ and } \vartheta \in]-1,1] \end{cases},
$$

and $\Omega = \{-1 \leqslant u \leqslant v \leqslant \eta\}$. Needless to say that the minimization of $\mathrm{MSE}(x,h)$ with respect to h is a difficult task. Even its asymptotic expansion depends on too many unknowns in addition to be difficult to handle for x in the boundary region of $F's$ support. For instance, when x belongs to the interior region, i.e. $\eta = 1$, then under some appropriate regularity assumptions, the asymptotic mean square error $\mathrm{AMSE}(x,h)$ can be expressed as

$$
\begin{aligned}
\mathrm{AMSE}(x,h) &= \frac{h^4}{4}e''^2(x)\left(\int_{-1}^{\eta} v^2 W_0(v,\eta)\mathrm{d}v\right)^2 \\
&\quad + 2\frac{h}{n}\frac{e^2(x)f(x)}{\bar{F}^2(x)}\int_{\Omega} vW_0(u,\eta)W_0(v,\eta)\mathrm{d}u\,\mathrm{d}v \\
&\quad + \frac{2}{n\bar{F}(x)}\left(2\frac{\mu_1(x)}{\bar{F}(x)} - e^2(x)\right)\int_{\Omega} W_0(u,\eta)W_0(v,\eta)\mathrm{d}u\,\mathrm{d}v.
\end{aligned}
$$

Minimization of $\mathrm{AMSE}(x,h)$ with respect to h gives a rather cumbersome expression

$$
h^*(x) = n^{-1/3}C(F,K),
$$

where

$$
C(F,K) = \left[\frac{-2e^2(x)f(x)\int_{\Omega} vW_0(u,\eta)W_0(v,\eta)\mathrm{d}u\,\mathrm{d}v}{\bar{F}^2(x)e''^2(x)\left(\int_{-1}^{\eta} v^2 W_0(v,\eta)\mathrm{d}v\right)^2}\right]^{1/3}.
$$

Many classical bandwidth selection methods such as "*rules of thumb*", "*solve-the-equation*", "*plug-in*", "*double kernel*" and "*smoothed bootstrap*" can be adapted to estimate a local or a global bandwidth. Due to their simplicity, hereafter, we shall adopt the "*rules of thumb*" and "*double kernel*" techniques. More precisely, let us switch to the global criterion MISE given by

$$\mathrm{MISE}(h) = \int_{-\infty}^{\infty} \mathrm{MSE}(x, h)\mathrm{d}x.$$

Then, as a first approach, we arbitrary set $F(\cdot)$ to a parametric family. Namely, we adopt the generalized Pareto distribution (GPD). This distribution plays an important role in the statistical analysis of extremes (see, e.g., Embrechts et al., 1997). As we will see later on, the GPD is an appropriate approximation of the excess distribution function $F_u(x) = P(X - u > x | X > u)$ for large u. For $v \in \mathbb{R}$ and $\beta > 0$, the GPD is defined by (see, Embrechts et al., 1997, p. 162)

$$G_{\xi,v,\beta}(x) = \begin{cases} 1 - (1 + \xi(x-v)/\beta)^{-1/\xi}, & \text{if } \xi \neq 0 \\ 1 - \exp(-(x-v)/\beta), & \text{if } \xi = 0, \end{cases}$$

where

$$\begin{aligned} (x-v)/\beta \geqslant 0 \quad & \text{if} \quad \xi \geqslant 0, \\ 0 \leqslant (x-v)/\beta \leqslant -1/\xi \quad & \text{if} \quad \xi < 0. \end{aligned}$$

In this case, the rule of thumb technique consists in minimizing the $\mathrm{MISE}(h)$ associated to the GPD whose parameters v, β and ξ are estimated from the sample at hand. Estimation of these parameters can be based, for instance, on maximum likelihood procedure. For details and other estimation techniques, see Embrecht et al. (1997, Section 6.5). The obtained smoothing parameter will be denoted h_{Ref}. Of course, h_{Ref} is of a limited value, many other more sophisticated techniques can be adapted to this context (see, e.g., Wand and Jones (1995) for a review). In this work, we limit ourselves to the L_1-double kernel method. This technique is fully automatic and has proven to be powerful in density estimation context (see Berlinet and Devroye, 1994). In order to outline this technique and sketch its adaptation to MRL estimation context, let $r \geqslant 1$ be an arbitrary integer and let $\{\hat{a}_{in}^{[r]}(x), i = 0, \ldots, r\}$ stand for the coefficients of the polynomial which minimizes the following analogous of (1.4)

$$\int_{-\infty}^{U_F} K_h(y - x)\left[e_n(y) - \sum_{i=0}^{r} a_i (y - x)^i\right]^2 \mathrm{d}y.$$

After some standard algebra, one can show that $\hat{a}_{0n}^{[r]}(x)$, the estimate of $e(x)$, can be written

$$\hat{a}_{0n}^{[r]}(x) = \int_{-1}^{\eta} W_0^{[r]}(v, \eta) e_n(x + hv)\mathrm{d}v, \tag{3.1}$$

where $\eta = \min\left(\frac{U_F - x}{h}, 1\right)$ and the function $W_0^{[r]}(\cdot, \eta)$ is a kernel of order $(r+1)$, i.e.

$$\int v^i W_0^{[r]}(v, \eta)\mathrm{d}v = \begin{cases} 1 & \text{if } i = 0, \\ 0 & \text{if } i = 1, \ldots, r, \\ C \neq 0 & \text{if } i = r + 1. \end{cases}$$

Now, assume that $p > q \geqslant 1$ are two given integers and note that the kernel $W_0^{[p]}(\cdot,\cdot)$ being of order $p+1 > q+1$, the bias of the estimate $\hat{a}_{0n}^{[p]}(\cdot)$ should be negligible in comparison to that of $\hat{a}_{0n}^{[q]}(x)$. Thus, whenever the L^1 error $\int \left|\hat{a}_{0n}^{[q]}(x) - e(x)\right| \mathrm{d}x$ exists, one could estimate it by

$$
\begin{aligned}
DK(p,q,h) &= \int \left|\hat{a}_{0n}^{[q]}(x) - \hat{a}_{0n}^{[p]}(x)\right| \mathrm{d}x \\
&= \int \left|\int_{-1}^{\eta} \left[W_0^{[q]}(v,\eta) - W_0^{[p]}(v,\eta)\right] e_n(x+hv)\mathrm{d}v\right| \mathrm{d}x. \qquad (3.2)
\end{aligned}
$$

Hence, the bandwidth h_{DK} which minimizes $DK(p,q,h)$ should approximate h_{L^1} the L^1 optimal bandwidth. This smoothing parameter selection technique is known as the double kernel method. It has been mainly used in density estimation context. Its L^2 version has been studied by Abdous (1999) and Jones (1998). Investigation of the asymptotic properties of h_{DK} are beyond the scope of the present paper, we will only evaluate its performance by a simulation study.

4. Simulations

To assess the performance of the local linear estimate defined above and the L^1 double kernel approach, we present a simulation study in which we considered the following models (Table 1): where $\mathbb{1}_A(x) = 1$ if x belongs to A and $=0$ elsewhere, while $\phi(\cdot)$ and $\Phi(\cdot)$ stand for the pdf and cdf of a standard normal distribution respectively.

As a starting kernel, we took Epanechnikov kernel, i.e.

$$K(x) = \frac{3}{4}(1-x^2)\mathbb{1}_{[-1,1]}(x).$$

Closed expressions for both $W_0^{[1]}$ and $W_0^{[2]}$ can be derived by putting

$$\mu_i = \mu_i(\eta) = \int_{-1}^{\eta} v^i K(v)\,\mathrm{d}v, \qquad \text{for } i \geqslant 0.$$

Table 1
Simulated models

Probability density function	MRL function
Uniform: $\mathbb{1}_{(0,1)}(x)$	$\frac{1-x}{2}\mathbb{1}_{(0,1)}(x)$
Maxwell: $x\exp(-\frac{x^2}{2})\mathbb{1}_{(0,\infty)}(x)$	$\sqrt{2\pi}\exp(x^2/2)(1-\Phi(x))\mathbb{1}_{(0,\infty)}(x)$
Pareto: $\frac{2}{x^3}\mathbb{1}_{(1,\infty)}(x)$	$(2-x)\mathbb{1}_{(0,1]}(x) + x\mathbb{1}_{(1,\infty)}(x)$
Truncated normal: $2\phi(x)\mathbb{1}_{(0,\infty)}(x)$	$(-x + \frac{\phi(x)}{1-\Phi(x)})\mathbb{1}_{(0,\infty)}(x)$
Lognormal: $\frac{1}{x\sqrt{2\pi}}\exp\left\{\frac{-(\log x)^2}{2}\right\}\mathbb{1}_{(0,\infty)}(x)$	$(-x + \sqrt{e}\frac{(1-\Phi(\log x - 1))}{(1-\Phi(\log x))})\mathbb{1}_{(0,\infty)}(x)$

Then, routine computations enable us to see that the empirical survival function can be written as

$$\overline{F}_n(u) = \begin{cases} 1, & \text{if } u < X_{(1)}, \\ \frac{n-i}{n}, & \text{if } X_{(i)} \leqslant u < X_{(i+1)}, \\ 0, & \text{if } u \geqslant X_{(n)}, \end{cases}$$

where $X_{(1)} \leqslant \cdots \leqslant X_{(n)}$ are the order statistics associated to the sample $X_1, \ldots, X_n$. Upon substituting the above expression for the empirical survival function in the definition of the empirical MRL function (1.1), we obtain

$$e_n(x) = \sum_{i=0}^{n-1} \{\overline{X}_{in} - x\} \mathbb{1}_{[X_{(i)}, X_{(i+1)})}(x),$$

where $\overline{X}_{in} = \frac{1}{n-i} \sum_{j=i+1}^{n} X_{(j)}$ for $i = 0, \ldots, n-1$ and $X_{(0)} = -\infty$ by convention. By (3.1), the estimators $\hat{a}_{0n}^{[1]}(x)$ and $\hat{a}_{0n}^{[2]}(x)$ can be written as

$$\hat{a}_{0n}^{[1]}(x) = \begin{cases} \int_{-1}^{\eta_n} W_0^{[1]}(v, \eta_n) e_n(x + hv) \mathrm{d}v, & \text{if } x < X_{(n)} + h, \\ 0, & \text{if } x \geqslant X_{(n)} + h, \end{cases}$$

$$\hat{a}_{0n}^{[2]}(x) = \begin{cases} \int_{-1}^{\eta_n} W_0^{[2]}(v, \eta_n) e_n(x + hv) \mathrm{d}v, & \text{if } x < X_{(n)} + h, \\ 0, & \text{if } x \geqslant X_{(n)} + h, \end{cases}$$

where $\eta_n := \eta = \min\left(1, (X_{(n)} - x)/h\right)$, and

$$W_0^{[1]}(v, \eta) = \frac{\mu_2 - v\mu_1}{\mu_0\mu_2 - \mu_1^2} K(v) \mathbb{1}_{[-1,\eta]}(v),$$

$$W_0^{[2]}(v, \eta) = \frac{(\mu_2\mu_4 - \mu_3^2) - v(\mu_1\mu_4 - \mu_2\mu_3) + v^2(\mu_3\mu_1 - \mu_2^2)}{\mu_0(\mu_2\mu_4 - \mu_3^2) - \mu_1(\mu_1\mu_4 - \mu_2\mu_3) + \mu_2(\mu_3\mu_1 - \mu_2^2)} K(v) \mathbb{1}_{[-1,\eta]}(v).$$

Since $W_0^{[1]}$ and $W_0^{[1]}$ are higher order kernels, a simplification is given by

$$\hat{a}_{0n}^{[1]}(x) = \begin{cases} \overline{X}_n - x, & \text{if } x < X_{(1)} - h, \\ \int_{-1}^{\eta_n} W_0^{[1]}(v, \eta_n) e_n(x + hv) \mathrm{d}v, & \text{if } X_{(1)} - h \leqslant x < X_{(n)} + h, \\ 0, & \text{if } x \geqslant X_{(n)} + h, \end{cases} \tag{4.1}$$

$$\hat{a}_{0n}^{[2]}(x) = \begin{cases} \overline{X}_n - x & \text{if } x < X_{(1)} - h, \\ \int_{-1}^{\eta_n} W_0^{[2]}(v, \eta_n) e_n(x + hv) \mathrm{d}v & \text{if } X_{(1)} - h \leqslant x < X_{(n)} + h, \\ 0 & \text{if } x \geqslant X_{(n)} + h. \end{cases}$$

According to these results, the difference $\hat{a}_{0n}^{[1]}(x) - \hat{a}_{0n}^{[2]}(x)$ is null if $x < X_{(1)} - h$ or $x \geqslant X_{(n)} + h$. Consequently the double kernel criteria can be written as

$$\begin{aligned} DK(1, 2, h) &= \int_{X_{(1)}-h}^{X_{(n)}+h} \left| \hat{a}_{0n}^{[2]}(x) - \hat{a}_{0n}^{[1]}(x) \right| \mathrm{d}x, \\ &= \int_{X_{(1)}-h}^{X_{(n)}+h} \left| \int_{-1}^{\eta} \left[W_0^{[2]}(v, \eta) - W_0^{[1]}(v, \eta) \right] e_n(x + hv) \mathrm{d}v \right| \mathrm{d}x, \end{aligned}$$

Table 2
Mean and standard deviation of the bandwidths h_{DK} and h_{ref}

Distribution	$N = 100$ $n = 50$		$N = 500$ $n = 100$	
	h_{DK}	h_{ref}	h_{DK}	h_{ref}
Uniform	0.368(0.263)	0.112(0.101)	0.312(0.232)	0.071(0.070)
Maxwell	0.588(0.315)	0.659(0.263)	0.519(0.322)	0.536(0.271)
Exponential	0.665(0.310)	0.809(0.257)	0.583(0.314)	0.751(0.270)
Pareto	0.603(0.365)	0.408(0.414)	0.528(0.372)	0.420(0.427)
Truncated normal	0.329(0.286)	0.730(0.310)	0.164(0.197)	0.739(0.324)
Lognormal	0.815(0.239)	0.884(0.225)	0.801(0.271)	0.808(0.317)

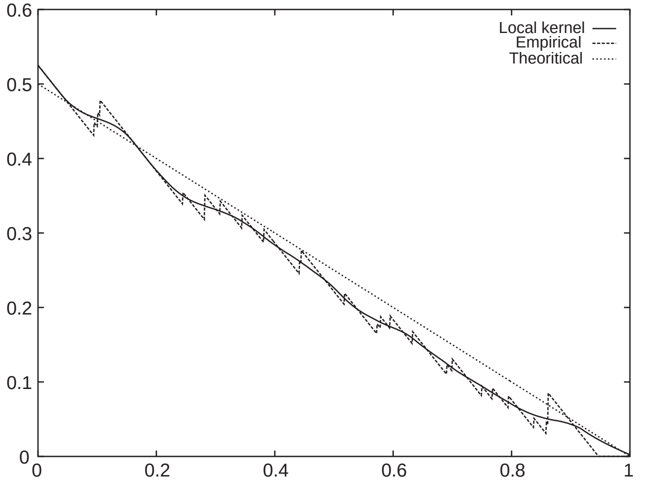

Fig. 1. Uniform, $n = 25$, $h = 0.36$.

For every model listed in Table 1, we generated $N = 100, 500$ samples of sizes $n = 50, 100$. For each sample, we computed the associated bandwidths h_{ref} and h_{DK}. A summary of these simulations is provided in Table 2. For each model and combination of N and n, we supplied the mean and the standard deviation (specified between brackets) of the obtained bandwidths.

Next, the expression (4.1) of $\hat{a}_{0n}^{[1]}(x)$ shows that this estimator vanishes for any $x \geqslant X_{(n)} + h$. This fact has prevented us from evaluating the global errors L^1 or L^2 of $\hat{a}_{0n}^{[1]}(x)$, since the theoretical MRL functions are not necessarily Lebesgue integrable. For this reason, instead of giving a global assessment, we will content ourselves by providing some graphics to illustrate the behavior of $\hat{a}_{0n}^{[1]}(x)$ under various MRL shapes. Furthermore, it is well known in the literature (see e.g. Csörgő and Zitikis, 1996) that the estimation of the MRL function $e(x)$ for large values of x is a challenging problem. Unfortunately, as shown by Figs. 1–6, this problem still holds for our estimators. A possible solution to this problem, among others, could be the use of semi-parametric approaches.

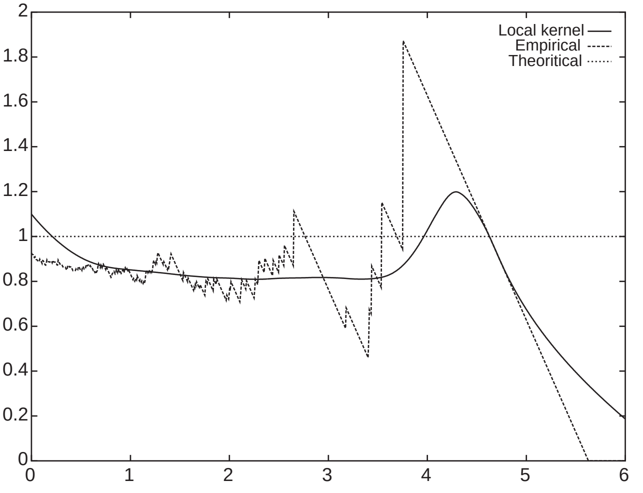

Fig. 2. Exponential, $n = 200$, $h = 0.65$.

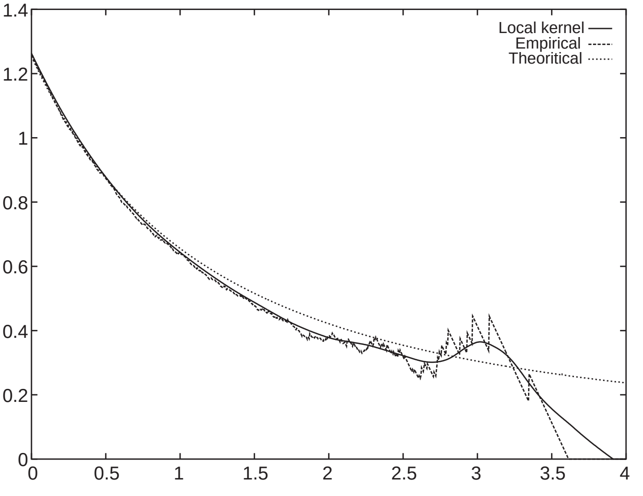

Fig. 3. Maxwell, $n = 700$, $h = 0.63$.

5. Proofs

Proof of Lemma 2.1. First, note that $e_n(\cdot)$ can be written as

$$e_n(u) = \int_u^\infty \frac{\bar{F}_n(z)}{\bar{F}_n(u)} \mathbb{1}_{[X_{(n)}>u]} \mathrm{d}z.$$

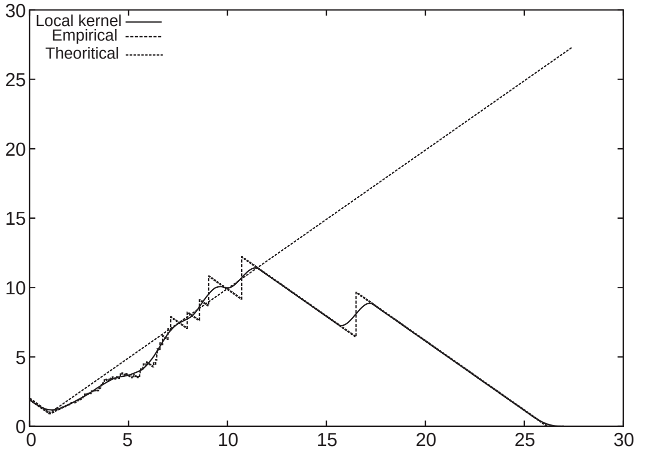

Fig. 4. Pareto, $n = 700$, $h = 0.8$.

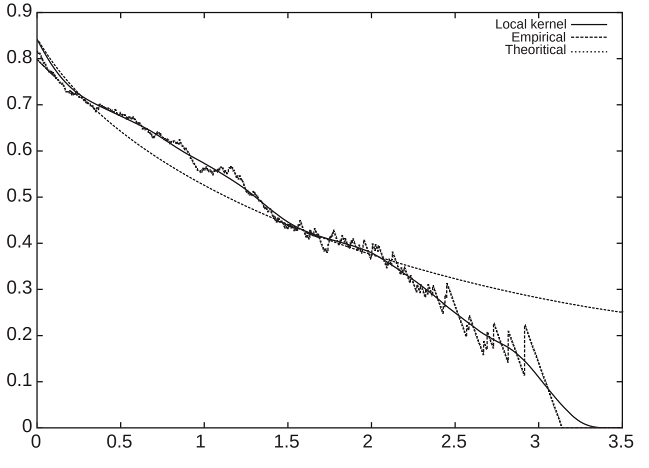

Fig. 5. Trun. normal, $n = 500$, $h = 0.28$.

Since $\bar{F}_n(\cdot)$ is a sum of i.i.d. Bernoulli random variables, we have for any $z \geqslant u$,

$$
\begin{aligned}
E\left(\frac{\bar{F}_n(z)}{\bar{F}_n(u)} \mathbb{1}_{[X_{(n)}>u]}\right) &= \sum_{j=1}^{n} \sum_{k=0}^{n} \frac{k}{j} P[n\bar{F}_n(z) = k, n\bar{F}_n(u) = j] \\
&= \sum_{j=1}^{n} \sum_{k=0}^{j} \frac{k}{j} \frac{n!}{(n-j)!(j-k)!k!} F^{n-j}(u) \\
&\quad \times (F(z) - F(u))^{j-k} (1 - F(z))^k \\
&= \frac{\bar{F}(z)}{\bar{F}(u)} (1 - F^n(u)),
\end{aligned}
$$

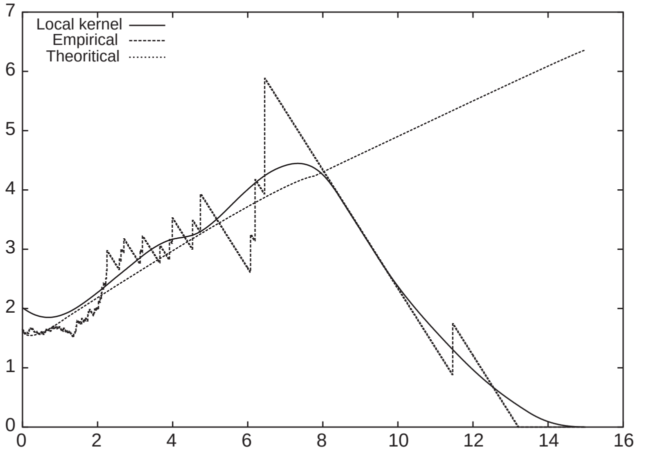

Fig. 6. Lognormal, $n = 100$, $h = 0.9$.

which gives (2.3). As for the covariance term, by taking $u \leqslant v$, we get

$$E(e_n(u)e_n(v)) = \int_v^\infty \int_u^\infty E\left(\frac{\bar{F}_n(s)\bar{F}_n(t)}{\bar{F}_n(u)\bar{F}_n(v)} \mathbb{1}_{[X_{(n)}>v]}\right) \mathrm{d}s\, \mathrm{d}t.$$

A closed form for this expectation can be obtained upon splitting the integral on the following domains:

$$\{u \leqslant s \leqslant v \leqslant t\}, \quad \{u \leqslant v \leqslant s \leqslant t\}, \quad \{u \leqslant v \leqslant t \leqslant s\}.$$

Then, the verification of (2.4) is similar to that of (2.3), though it involves longer calculation. □

Proof of Theorem 2.1. An application of Lemma 2.1 yields

$$E(\hat{a}_{0n}(x)) = \begin{cases} \int_{-1}^{1} W_0(v,1)e(x+hv) & \text{if } U_F = +\infty \text{ or if } U_F < \infty \\ (1 - F^n(x+hv))\mathrm{d}v & \text{and } x < U_F - h \\ \int_{-1}^{\vartheta} W_0(v,\vartheta)e(x+hv) & \text{if } U_F < \infty \text{ and } x = U_F - \vartheta h \\ (1 - F^n(x+hv))\,\mathrm{d}v & \\ 0 & \text{if } U_F < \infty \text{ and } x \geqslant U_F + h. \end{cases}$$

If either $U_F = +\infty$ or $U_F < \infty$ with $x < U_F - h$, then we can write

$$|E(\hat{a}_{0n}(x)) - e(x)| \leqslant |T_1| + |T_2| + |T_3|,$$

where

$$T_1 = \int_{-1}^{1} W_0(v,1)\frac{\mu_0(x+hv)-\mu_0(x)}{\bar F(x+hv)}\mathrm{d}v$$
$$T_2 = \int_{-1}^{1} W_0(v,1)\mu_0(x)\left[\frac{1}{\bar F(x+hv)}-\frac{1}{\bar F(x)}\right]\mathrm{d}v$$
$$T_3 = \int_{-1}^{1} W_0(v,1)e(x+hv)F^n(x+hv)\mathrm{d}v,$$

and $\mu_0(y)=\int_y^\infty \bar F(z)\mathrm{d}z$. The first term T_1 goes to 0 as $n\to\infty$ because

$$|T_1|\leqslant \frac{h}{\bar F(x+h)}\int_{-1}^{1}|vW_0(v,1)|\mathrm{d}v.$$

The second term T_2 satisfies

$$|T_2|\leqslant \mu_0(x)\max\left(\left|\frac{1}{\bar F(x+h)}-\frac{1}{\bar F(x)}\right|,\left|\frac{1}{\bar F(x-h)}-\frac{1}{\bar F(x)}\right|\right)\int_{-1}^{1}|W_0(v,1)|\mathrm{d}v.$$

The d.f. F being continuous, the right-hand side of the above inequality converges to 0 as $n\to\infty$. Similarly, the last term T_3 tends also to 0 because $F(x+h)<1$ and

$$|T_3|\leqslant \frac{\mu_0(x-h)}{\bar F(x+h)}F^n(x+h)\int_{-1}^{1}|W_0(v,1)|\mathrm{d}v.$$

This completes the proof of (2.1) for $U_F=+\infty$ or $U_F<\infty$ with $x<U_F-h$. Next, if $U_F<\infty$ and $x=U_F-\vartheta h$ with ϑ being fixed in $]-1,1]$, then

$$\begin{aligned}|E(\hat e_{n0}(x))| &= \left|\int_{-1}^{\vartheta} W_0(v,\vartheta)\frac{\mu_0(U_F+(v-\vartheta)h)}{\bar F(U_F+(v-\vartheta)h)}(1-F^n(U_F+(v-\vartheta)h))\mathrm{d}v\right|\\ &\leqslant h\int_{-1}^{\vartheta}|(v-\vartheta)W_0(v,\vartheta)|\mathrm{d}v.\end{aligned}$$

The above inequality ensues because $\bar F(\cdot)$ is a decreasing function and $\mu_0(\cdot)$ satisfies $0\leqslant\mu_0(U_F+(v-\vartheta)h)\leqslant(\vartheta-v)h\bar F(U_F+(v-\vartheta)h)$ for $v\in]-1,\vartheta[$. Similarly, we have

$$e(x)=\begin{cases} e(U_F-\vartheta h)\leqslant h\vartheta, & \text{if } \vartheta>0,\\ 0, & \text{if } \vartheta\leqslant 0.\end{cases}$$

By letting $h\to 0$, it follows that both $e(x)$ and $E(\hat e_{n0}(x))$ converge to 0. This concludes the proof of (2.1). Next, to establish (2.2), note that in accordance to (2.4), one has

$$\begin{aligned}\mathrm{Var}(\hat a_{0n}(x)) &= \int_{[-1,\eta]^2} W_0(u,\eta)W_0(v,\eta)\mathrm{Cov}(e_n(x+hu),e_n(x+hv))\mathrm{d}u\,\mathrm{d}v\\ &:= V_1+V_2+V_3,\end{aligned} \tag{5.1}$$

where

$$\eta = \begin{cases} 1 & \text{if } U_F = +\infty \text{ or if } U_F < \infty \text{ with } x < U_F - h \\ \vartheta & \text{if } U_F < \infty \text{ with } x = U_F - \vartheta h \text{ and } \vartheta \in]-1, 1] \end{cases}$$

and

$$\begin{aligned}
V_1 &= 2\int_\Omega W_0(u,\eta)W_0(v,\eta)S_n(x+hu, x+hv) \\
&\quad \times \left(2\frac{\mu_1(x+hv)}{\bar F(x+hv)} - e^2(x+hv)\right)\mathrm{d}u\,\mathrm{d}v \\
V_2 &= 2\int_\Omega W_0(u,\eta)W_0(v,\eta)e(x+hu)e(x+hv) \\
&\quad \times (1-F^n(x+hu))F^n(x+hv)\mathrm{d}u\,\mathrm{d}v \\
V_3 &= 2\int_\Omega W_0(u,\eta)W_0(v,\eta)\frac{F^n(x+hv)-F^n(x+hu)}{F(x+hv)-F(x+hu)}\mu_0(x+hv) \\
&\quad \times \left\{h(v-u) + \frac{\mu_0(x+hv)-\mu_0(x+hu)}{\bar F(x+hv)}\right\}\mathrm{d}u\,\mathrm{d}v.
\end{aligned}$$

with $\Omega = \{-1 \leqslant u \leqslant v \leqslant \eta\}$. First, we have to distinguish between the two cases $\eta = 1$ and $\eta = \theta$. If $\eta = 1$, then the functions $\bar F(\cdot)$, $\mu_0(\cdot)$ and $\mu_1(\cdot)$ being non-negative and decreasing, we easily see that

$$\begin{aligned}
|V_1| &\leqslant \left(4\frac{\mu_1(x-h)}{\bar F(x+h)} + 2\frac{\mu_0^2(x-h)}{\bar F^2(x+h)}\right) S_n(x+h, x-h)\left(\int_{-1}^1 |W_0(u,1)|\mathrm{d}u\right)^2 \\
|V_2| &\leqslant 2F^n(x+h)\frac{\mu_0^2(x-h)}{\bar F^2(x+h)}\left(\int_{-1}^1 |W_0(u,1)|\mathrm{d}u\right)^2 \\
|V_3| &\leqslant 4nhF^{n-1}(x+h)\mu_0(x-h) \\
&\quad \times \left[1 + \frac{\bar F(x-h)}{\bar F(x+h)}\right]\int_{-1}^1\int_{-1}^1 |uW_0(u,1)W_0(v,1)|\mathrm{d}u\,\mathrm{d}v.
\end{aligned}$$

The continuity of $\bar F(\cdot)$, $\mu_0(\cdot)$ and $\mu_1(\cdot)$ together with the fact that $F(x+h) < 1$ entail that both V_2 and V_3 tend to 0 as $n \to \infty$. The first term V_1 also goes to 0 because

$$\lim_{n\to\infty} nS_n(x+h, x-h) = \frac{1}{\bar F(x)}.$$

Indeed,

$$\begin{aligned}
&\left|nS_n(x+h, x-h) - \frac{1}{\bar F(x)}\right| \\
&\quad \leqslant \left|nS_n(x+h, x-h) - \frac{1}{\bar F(x+h)}\right| + \left|\frac{1}{\bar F(x+h)} - \frac{1}{\bar F(x)}\right|.
\end{aligned}$$

Clearly, the second term on the right-hand side of this inequality converges to 0 as $n \to \infty$, as for the first one, we have

$$\begin{aligned}
&\left| nS_n(x+h, x-h) - \frac{1}{\bar F(x+h)} \right| \\
&\quad \leqslant \left| n \sum_{i=1}^{n} \frac{F^{n-i}(x+h)}{i} - \frac{1}{\bar F(x+h)} \right| + n \sum_{i=1}^{n} \frac{F^{n-i}(x+h) F^i(x-h)}{i} \\
&\quad \leqslant \sum_{i=0}^{n-1} \frac{i}{n-i} F^i(x+h) + \frac{F^n(x+h)}{\bar F(x+h)} + nF^n(x+h) \sum_{i=1}^{n} \frac{1}{i}.
\end{aligned}$$

Since $F(x+h) < 1$, the last two terms tend to zero and an application of Tannery's theorem (see Apostol, 1957, p. 458) shows that the first term also tends to 0 as $n \to \infty$. To complete the proof of (2.2), we have to consider the case when x belongs to the boundary region of F's support, i.e. $\eta = \theta$. Indeed, since $x = U_F - \vartheta h$ and $\bar F(\cdot)$ is a decreasing function, we have

$$\begin{aligned}
&0 \leqslant \mu_i(x+hv) \leqslant h^{i+1}(\vartheta - v)^{i+1} \bar F(x+hv)/(i+1) \\
&\quad \text{for } i = 0, 1 \text{ and } v \in [-1, \vartheta].
\end{aligned}$$

Therefore, for any u and v such that $-1 \leqslant u \leqslant v \leqslant \vartheta$, we have

$$\left| 2\frac{\mu_1(x+hv)}{\bar F(x+hv)} - e^2(x+hv) \right| \leqslant h^2(\vartheta - v)^2,$$

and

$$h(v-u)\left(1 - \frac{\bar F(x+hu)}{\bar F(x+hv)}\right) \leqslant h(v-u) + \frac{\mu_0(x+hv) - \mu_0(x+hu)}{\bar F(x+hv)} \leqslant 0.$$

Then, by using the notations in (5.1) we get

$$\begin{aligned}
|V_1| &\leqslant 2h^2 S_n(U_F, U_F - (1+\vartheta)h) \int_{-1 \leqslant u \leqslant v \leqslant \vartheta} (\vartheta - v)^2 |W_0(u,\vartheta) W_0(v,\vartheta)| \mathrm{d}u\, \mathrm{d}v \\
&\leqslant 2nh^2(1 - F^n(U_F - (1+\vartheta)h)) \\
&\quad \times \int_{-1 \leqslant u \leqslant v \leqslant \vartheta} (\vartheta - v)^2 |W_0(u,\vartheta) W_0(v,\vartheta)| \mathrm{d}u\, \mathrm{d}v \\
|V_2| &\leqslant 2h^2 \int_{-1 \leqslant u \leqslant v \leqslant \vartheta} (\vartheta - v)(\vartheta - u) |W_0(u,\vartheta) W_0(v,\vartheta)| \mathrm{d}u\, \mathrm{d}v \\
|V_3| &\leqslant 4h^2 \int_{-1 \leqslant u \leqslant v \leqslant \vartheta} (\vartheta - v)(v - u) |W_0(u,\vartheta) W_0(v,\vartheta)| \mathrm{d}u\, \mathrm{d}v. \quad \square
\end{aligned}$$

Proof of Theorem 2.2. The estimator $\hat a_{0n}(x)$ being a sum of non i.i.d. random variables, we use the projection method of Hajek (1968) to approximate $\hat a_{0n}(x)$ by an asymptotically equivalent sum of i.i.d. random variables and then we apply the classical central limit theorem. The approximating sum is defined by

$$\widetilde{e}_n(x) = \sum_{i=1}^{n} E\left(\hat{a}_{0n}(x)|X_i\right) - (n-1)E(\hat{a}_{0n}(x)).$$

An explicit expression of $\widetilde{e}_n(\cdot)$ is easily obtained once we note that for any $u < v$ and $i \in \{1, \ldots, n\}$,

$$\begin{aligned}\frac{\bar{F}_n(v)}{\bar{F}_n(u)}\mathbb{1}_{[X_{(n)}>u]} = \mathbb{1}_{[X_i \leqslant u]}\frac{T_{n,i}(v)}{T_{n,i}(u)}\mathbb{1}_{[X_{(n,i)}>u]} + \mathbb{1}_{[u<X_i\leqslant v]}\frac{T_{n,i}(v)}{1+T_{n,i}(u)} \\ + \mathbb{1}_{[X_i>v]}\frac{1+T_{n,i}(v)}{1+T_{n,i}(u)},\end{aligned}$$

where $T_{n,i}(z) = \sum_{\substack{j=1\\ j\neq i}}^{n} \mathbb{1}_{[X_j>z]}$ for any z, and $X_{(n,i)} = \max_{\substack{1\leqslant j\leqslant n\\ j\neq i}} X_j$. Now, it suffices to use (2.3) and the fact that $E(\hat{a}_{0n}(x)) = E(\widetilde{e}_n(x))$ to end up with

$$\widetilde{e}_n(x) - E(\widetilde{e}_n(x)) = \frac{1}{n}\sum_{i=1}^{n} U_n(X_i),$$

where

$$\begin{aligned}U_n(X_i) = \int K_h(x-u)\Bigg\{\frac{1-F^n(u)}{\bar{F}(u)}(X_i - u - e(u))\mathbb{1}_{[X_i>u]} \\ + ne(u)F^{n-1}(u)\bar{F}(u)\left(\frac{F(u)}{\bar{F}(u)}\mathbb{1}_{[X_i>u]} - \mathbb{1}_{[X_i\leqslant u]}\right)\Bigg\}\,\mathrm{d}u.\end{aligned}$$

Hence, similar computations to those used in Lemma 2.1 show that

$$\begin{aligned}\mathrm{Var}(\widetilde{e}_n(x)) = \frac{2}{n}\int_{u<v} K_h(x-u)K_h(x-v) \\ \times\Bigg\{\frac{1-F^n(u)}{\bar{F}(u)}\frac{1-F^n(v)}{\bar{F}(v)}\left(2\mu_1(v) - e^2(v)\bar{F}(v)\right) \\ + nF^{n-1}(v)\bar{F}(v)e(v)\frac{1-F^n(u)}{\bar{F}(u)}(v - u + e(v) - e(u)) \\ + n^2 e(u)e(v)F^n(u)F^{n-1}(v)\bar{F}(v)\Bigg\}\,\mathrm{d}u\,\mathrm{d}v.\end{aligned}$$

Obviously, this expression and (2.1) ensure that $\mathrm{Var}(\hat{a}_{0n}(x))/\mathrm{Var}(\widetilde{e}_n(x))$ converges to 1 as $n \to \infty$. Consequently, by using the fact that

$$E(\hat{a}_{0n}(x) - \widetilde{e}_n(x))^2 = \mathrm{Var}(\hat{a}_{0n}(x)) - \mathrm{Var}(\widetilde{e}_n(x)),$$

we see that the standardized version of $\hat{a}_{0n}(x)$ and $\widetilde{e}_n(x)$ have the same asymptotic distribution. Thus, to complete the proof, it suffices to show that the standardized version of $\widetilde{e}_n(x)$ is asymptotically normal. By Lyapounov's theorem, this holds if $(n\mathrm{Var}(\widetilde{e}_n(x)))^{-3/2}n^{-1/2}E|U_n(X_1)|^3$ converges to zero. However, since

$$n\mathrm{Var}(\widetilde{e}_n(x)) \to \frac{2}{\bar{F}(x)}\left(2\frac{\mu_1(x)}{\bar{F}(x)} - e^2(x)\right)\int_{\Omega} K(u)K(v)\mathrm{d}u\,\mathrm{d}v, \quad \text{as } n \to \infty,$$

we only need to show that $n^{-1/2}E|U_n(X)|^3 \to 0$. This follows after observing that $U_n(X)$ satisfies

$$|U_n(X)| \leqslant \frac{1}{\bar{F}(x+h)}\left(|X-x| + h + \frac{\mu_0(x-h)}{\bar{F}(x+h)}\right) + n\mu_0(x-h)F^{n-1}(x+h)\left(\frac{F(x+h)}{\bar{F}(x+h)} + 1\right). \quad \square$$

Acknowledgements

The authors are deeply indebted to the referees for their careful reading of the manuscript. Their comments and suggestions have contributed to a greatly improved presentation of the results of this paper.

References

Abdous, B., 1999. L^2 version of the double kernel method. Statistics 32, 249–266.

Apostol, T.M., 1957. Mathematical Analysis, 2nd Edition. Addison-Wesley, Reading, MA.

Berlinet, A., Devroye, L., 1994. A comparison of kernel density estimates. Publications de l'Institut de Statistique de l'Université de Paris, vol. 38, pp. 3–59.

Chaubey, Y.P., Sen, P.K., 1999. On smooth estimation of mean residual life. J. Statist. Plann. Inference 75, 223–236.

Csörgő, M., Zitikis, R., 1996. Mean residual life processes. Ann. Statist. 24 (4), 1717–1739.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events. Karatzas, I, Yor, M. (Eds.), Springer, Berlin

Fan, J., Gijbels, I., 1996. Local polynomial modeling and its applications. In: Monographs on Statistics and Applied Probability, vol. 66. Chapman & Hall, London.

Guess, F., Proschan, F., 1988. Mean residual life. In: Rao Krishnaiah, P.R., Rao, C.R. (Eds.), Handbook of Statistics, vol. 7. North-Holland, Amsterdam, pp. 215–224.

Hajek, J., 1968. Asymptotic normality of simple linear statistics under alternatives. Ann. Math. Statist. 39, 325–346.

Hall, W.J., Wellner, J.A., 1981. Mean residual life. In: Csörgő, M., Dawson, D.A., Rao, J.N.K., Saleh, A.K.Md.E. (Eds.), Statistics and Related Topics. North-Holland, Amsterdam, pp. 169–184.

Hille, E., 1948. Functional Analysis and Semigroups. Am. Math. Soc. Colloq. Pub., vol. 31, American Mathematical Society.

Jones, M.C., 1998. On some kernel density estimation bandwidth selectors related to the double kernel method. Sankhya Ser. A 60, 249–264.

Kotz, S., Shanbhag, D.N., 1980. Some new approaches to probability distributions. Adv. in Appl. Probab. 12, 903–921.

Patil, P.N., Wells, M.T., Marron, J.S., 1994. Some heuristics of kernel based estimators of ratio functions. Nonparametric Statist. 4, 203–209.

Ruiz, J.M., Guillamòn, A., 1996. Nonparametric recursive estimator of residual life and vitality functions under mixing dependence conditions. Commun. Statist.- Theory and Methods 4, 1999–2011.

Wand, M.P., Jones, M.C., 1995. Kernel Smoothing, Chapman and Hall, London.