Fixed design local polynomial smoothing and bandwidth selection for right censored data.☆

Dimitrios Bagkavos[a,*], Dimitrios Ioannides[b]

[a] *Department of Mathematics, University of Ioannina, 45110, Greece.*
[b] *Department of Economics, University of Macedonia, 54006, Greece*

Abstract

The local polynomial smoothing of the Kaplan–Meier estimate for fixed designs is explored and analyzed. The first benefit, in comparison to classical convolution kernel smoothing, is the development of boundary aware estimates of the distribution function, its derivatives and integrated derivative products of any arbitrary order. The advancements proceed by developing asymptotic mean integrated square error optimal solve-the-equation plug–in bandwidth selectors for the estimates of the distribution function and its derivatives, and as a byproduct, a mean square error optimal bandwidth rule for integrated derivative products. The asymptotic properties of all methodological contributions are quantified analytically and discussed in detail. Three real data analyses illustrate the benefits of the proposed methodology in practice. Finally, numerical evidence is provided on the finite sample performance of the proposed technique with reference to benchmark estimates.

Keywords: Kaplan–Meier, local polynomial fitting, censoring, kernel smoothing, bandwidth selection.
2010 MSC: 62G10, 62N03

1. Introduction

Let T denote a continuous lifetime variable with cumulative distribution function (c.d.f.) $F_T(t) = P(T \leq t)$. Frequently the available data are beyond the experimenter's control and come in the form of scatterplot observations. For example, this is the case in lifetable analyses in the actuarial science, in data analyses in demography e.t.c., see Müller et al. (1997) and Wang et al. (1998). In such an occasion, the coordinates of the available data pairs consist of the response, which is usually an empirical estimate of the target curve, and the center of the associated time interval at which the curve is being estimated. Still, continuous estimates are desirable, especially when the analysis additionally depends on the estimate's derivatives. For this reason, the present research considers the local polynomial smoothing of the well-known Kaplan-Meier estimate (Kaplan and Meier, 1958), with first objective to provide continuous, boundary aware estimates for the distribution function, its derivatives of any arbitrary order and integrated c.d.f. derivative products for fixed designs under the random right censorship model. The reasoning for pursuing this approach becomes immediately obvious when observing that smoothing of scatterplot data intrinsically corresponds to formulating a reasonable nonparametric regression problem. The asymptotic unbiasedness property of the Kaplan-Meier estimate together with its

*Correspondence to: Department of Mathematics, University of Ioannina, 45100, Ioannina, Greece, tel: +302651008225.
Email address: dimitrios.bagkavos@gmail.com (Dimitrios Bagkavos)

strong representation as the underlying c.d.f. plus an asymptotically negligible error term, prompts its use as the response. At the same time, the center of each equidistant time interval in which the observed data range is split is used as the design. Hence, application of the local polynomial technique yields estimates of $F_T^{(\nu)}, \nu = 0, 1, \dots$ by matching the coefficients of polynomials fitted locally – through kernel weighted least squares – and the derivatives of F_T in a Taylor expansion of the regression function at a nearby point; precise formulation and details are provided in Section 2. This approach enables the development, also in Section 2, of local polynomial estimates for integrated c.d.f. derivative products of any arbitrary order. These are useful on their own right since they are necessary for the implementation of automatic bandwidth selectors, in estimation of population characteristics, statistical distance measures and in a variety of other settings.

Multiple benefits arise from the local polynomial smoothing of the Kaplan–Meier estimate. First, its definition does not involve a bandwidth and thus its use as the response in the aforementioned nonparametric regression problem greatly simplifies implementation of the resulting estimates which now depend on just one bandwidth; this is in contrast to the traditional approach which needs two bandwidths. In terms of performance, the Asymptotic Mean Integrated Square Error (AMISE) and central limit theorem for the estimates of $F_T^{(\nu)}$, quantified analytically in Section 3, are valid throughout the region of estimation and imply the absence of inflated bias at the endpoints. Further, the asymptotic properties of the integrated derivative product estimates, also quantified in Section 3, ensure efficient estimation of the functionals as opposed to using conventional kernel smoothers. A subsequent advantage thus results by their utilization in developing (in Section 4) a solve–the–equation AMISE–optimal plug–in bandwidth rule applicable to all estimates proposed here. The rule is built as a direct extension of the corresponding density estimation bandwidth selector for complete data proposed in Cheng (1997). The gain is its stable performance across the region of estimation; this is also reflected in its asymptotic properties, quantified analytically together with its convergence rate and asymptotic distribution in Section 4. It is worth noting here that the literature is rather thin on AMISE optimal bandwidth rules for convolution smoother estimates for right censored data. Since the plug-in rule proposed here is also applicable to classical kernel approach, it can also be thought as filling this important gap in the literature.

Section 5 investigates the finite sample performance of the proposed methodology. First, the analysis of three real world data sets illustrates how the proposed technique can help in capturing data patterns that remain undiscoverable either by the conventional kernel smoothing approach or by parametric estimates. Finally, distributional data are used to simulate and compare the finite sample MISE performance of the proposed estimates in comparison to frequently used estimates in the literature and in practice.

2. Local polynomial smoothing of the Kaplan–Meier estimate.

Let $T_1, T_2, \dots, T_n$ be a sample of i.i.d. survival times censored on the right by i.i.d. random variables $U_1, U_2, \dots, U_n$, which are independent from the T_i's. Let f_T be the common probability density function (p.d.f.) and F_T the c.d.f. of the T_i's. Denote with H the c.d.f. of the U_i's. Typically the observed right censored data are denoted by the pairs (X_i, δ_i), $i = 1, 2, \dots, n$ with $X_i = \min\{T_i,\ U_i\}$ and $\delta_i = \mathbf{1}_{\{T_i \le U_i\}}$ where $\mathbf{1}_{\{\cdot\}}$ is the indicator random variable of the event $\{\cdot\}$. The distribution function of the X_i's satisfies $1 - F = (1 - F_T)(1 - H)$. It is assumed that estimation happens in the interval $[0, M]$ where M satisfies the relationship

$$M = \sup\{x : 1 - F(x) > \varepsilon\} \text{ for a small } \varepsilon > 0.$$

We are interested in estimating $F_T^{(\nu)}(x), \nu = 0, 1, \ldots$. The Kaplan-Meier, introduced in Kaplan and Meier (1958), is the classical nonparametric estimate of $F_T \equiv F_T^{(0)}$ and is defined by

$$\hat{F}_T(x) = \begin{cases} 0, & 0 \le x \le X_{(1)}, \\ 1 - \prod_{i=1}^{k-1} \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}}, & X_{(k-1)} < x \le X_{(k)},\ k = 2, \ldots, n, \\ 1, & x > X_{(n)}, \end{cases} \tag{1}$$

where $(X_{(i)}, \delta_{(i)}), i = 1, \ldots, n$ are the ordered X_i's, along with their censoring indicators. According to the fixed design local polynomial principle, first partition the interval $[0, M]$ into g disjoint subintervals of equal length b; denote with $x_j = (j - \frac{1}{2})b$, $j = 1, \ldots, g$, the center of the jth interval. Denote with $\sigma^2(x_i)$ the variance of $\hat{F}_T(x_i)$ at x_i and let $\varepsilon_i, i = 1, \ldots, g$ be independent random vectors with mean 0 and variance 1. Also, set $m(x_i) = F_T(x_i)$. Since $\hat{F}_T(x)$ is an asymptotically unbiased estimate of $F_T(x)$ it can be used as the response to the local nonparametric regression problem

$$\hat{F}_T(x_i) = m(x_i) + \sigma(x_i)\varepsilon_i, \quad i = 1, \ldots, g.$$

Then, given a bandwidth h, the data $\{\hat{F}_T(x_i), x_i\}, i = 1, \ldots, g$ for which $|x_j - x| \le h$ are smoothed by locally fitting polynomials of fixed degree p. The polynomial coefficients are obtained by solving the optimization problem

$$\min_{\beta_k, k=0,\ldots,p} \sum_{j=1}^{g} \left\{ \hat{F}_T(x_j) - \sum_{k=0}^{p} \beta_k (x_j - x)^k \right\}^2 K\left(\frac{x_j - x}{h}\right). \tag{2}$$

Here K is a kernel function, usually a symmetric density, assumed to be supported on a symmetric and compact interval; however see also Funke and Hirukawa (2020) for an alternative approach based on asymmetric kernel functions in the closely related regression setting. Denote with $\hat{\beta}_k$ the estimates of β_k resulting by the solution of (2). A Taylor expansion of the regression function $m(x)$ in a nearby point x_0 such that $|x - x_0| \le \varepsilon$ for an arbitrarily small ε, yields that $\hat{F}_L^{(\nu)}(x) = \nu!\hat{\beta}_\nu$ is an estimate of $F_T^{(\nu)}(x), \nu = 0, \ldots, p$. According to Fan and Gijbels (1996), the optimal order of the local polynomial to use in (2) depends on the order of the derivative being estimated and is given by $p = \nu + 1$. This yields the solution

$$\hat{F}_L^{(\nu)}(x) = \nu! \sum_{i=1}^{g} K_\nu \left(\frac{x_i - x}{h}\right) \hat{F}_T(x_i),\ \nu = 0, 1, 2, \ldots, \tag{3}$$

where

$$K_\nu(u) = e_{\nu+1}^T S^{-1}(1, hu, \ldots, (hu)^\nu, (hu)^{\nu+1})^T K(u).$$

Here $e_{\nu+1}^T$ denotes a vector with $\nu + 2$ elements with 1 in the $(\nu + 1)$th position and zeros elsewhere and S is the $(\nu + 2) \times (\nu + 2)$ matrix $(S_{n,j+l})_{0 \le j,l \le \nu+1}$ with

$$S_{n,l}(x) = \sum_{i=1}^{g} K\left(\frac{x_i - x}{h}\right)(x_i - x)^l, \quad l = 0, 1, \ldots, 2\nu + 2.$$

Expression (3) shows that $\hat{F}_L^{(\nu)}(x)$ is very similar to a conventional kernel estimate with the difference that K_ν is defined as a function of the design points and locations. However there are some fundamental differences with the random design setting for censored data, explored in Bagkavos and Ioannides (2020). In the random design, smoothing is applied to the increments of the Kaplan–Meier estimate and the smoothing weights $S_{n,l}$ are random. In the fixed design the weights $S_{n,l}$ are deterministic and operate on the bin centers x_i. As a consequence, the quotient of the weights applied to the empirical estimate tend to 1 as $n \to \infty$; further, smoothing weights are identical irrespectively of whether the target is e.g. the distribution or the density function in both complete and censored data settings, see Cheng (1997). An equivalent representation for $\hat{F}_L^{(\nu)}(x)$ which sheds light on the inner mechanism of the technique can be defined as follows. Without loss of generality assume that K is supported on $[-1, 1]$. Let $0 < c < 1$ so that $x = ch \in [0, h)$ is a boundary point. Correspondingly, in the interior we have $x = ch, c > 1$, so that $x \in [h, M - h]$. Set $\hat{S}_c = (\mu_{i+j,c}(K))_{0 \le i,j \le \nu+1}$ where for any function g, for $i = 0, \ldots, 2\nu + 2$,

$$\mu_{i,c}(g) = \begin{cases} \int_{-\infty}^{c} u^i g(u)\, du, & \text{when } u \in [0, h), \\ \int u^i g(u)\, du \equiv \mu_i(g), & \text{when } u \in [h, M-h], \\ \int_{-c}^{\infty} u^i g(u)\, du, & \text{when } u \in (M-h, M]. \end{cases}$$

From the proof of Lemma 5 in Cheng (1994),

$$\frac{b}{h^{l+1}} S_{n,l} = \mu_l + o(1), l = 0, 1, \ldots, 2\nu + 2, \tag{4}$$

$$K_\nu(t) = \frac{b}{h^{\nu+1}} K^*_{\nu,c}(t) + o\left(bh^{-(\nu+1)}\right), \tag{5}$$

where for any two real valued deterministic sequences a_n and b_n, $a_n = o(b_n)$ as $n \to \infty$ if and only if $\lim_{n\to\infty} |a_n/b_n| = 0$; thus choosing $b \ll h$ as suggested by assumption A.3 below means that the asymptotic term in the right hand side of (5) is negligible. Let $K^*_{\nu,c}$ denote the so called *equivalent* kernel, defined by

$$K^*_{\nu,c}(u) = e^T_{\nu+1} \hat{S}_c^{-1} (1, u, \ldots, u^\nu, u^{\nu+1})^T K(u) I_{[-c,\infty)}(u).$$

An asymptotically equivalent representation for $\hat{F}_L^{(\nu)}(x)$ is given by

$$\hat{F}_L^{(\nu)}(x) = \frac{b\nu!}{h^{\nu+1}} \sum_{i=1}^{g} K^*_{\nu,c}\left(\frac{x_i - x}{h}\right) \hat{F}_T(x_i)(1 + o(1)).$$

The equivalent kernel satisfies throughout the region of estimation the moment conditions

$$\int u^q K^*_{\nu,c}(u)\, du = \delta_{\nu,q},\ 0 \le \nu, q \le \nu + 1, \tag{6}$$

where $\delta_{\nu,q}$ is Kronecker's delta, i.e. $\delta_{\nu,q} = 1$ for $\nu = q$ and 0 otherwise. It is immediately seen from (6) that $\hat{F}_L^{(\nu)}(x)$ automatically adjusts at the endpoints, without the extra modifications and without the undesirable side effects of boundary kernels such as negative estimate values.

Notice that the definition of $K^*_{\nu,c}$ for $x \in [h, M - h]$ does not depend on c. To see this first assume, without loss of generality, that the support of K is $[-1, 1]$. In the interior, i.e. for $c \to \infty$, $\hat{S}_c$ is equivalent to $\hat{S} = (\mu_{i+j}(K))_{0 \le i,j \le \nu+1}$ and $I_{(-\infty,c]}(u) = 1 = I_{[-c,\infty)}(u)$. Hence $K^*_{\nu,c} = K^*_\nu$ where

$$K^*_\nu(u) = e^T_{\nu+1} \hat{S}^{-1} (1, u, \ldots, u^\nu, u^{\nu+1})^T K(u),$$

with $\hat{S} = (\mu_{i+j}(K))_{0\le i,j\le \nu+1}$. Therefore in the interior $\hat{F}_L^{(\nu)}(x)$ can be written with slightly simpler notation as

$$\hat{F}_L^{(\nu)}(x) = \frac{b\nu!}{h^{\nu+1}} \sum_{i=1}^{g} K_\nu^* \left(\frac{x-x_i}{h}\right) \hat{F}_T(x_i)(1+o(1)).$$

It will be easier to study the statistical properties of $\hat{F}_L^{(\nu)}(x)$ by considering the following equivalent formulation

$$\begin{aligned}\hat{F}_L^{(\nu)}(x) &= \frac{b\nu!}{h^{\nu}} \sum_{i=1}^{g} W_\nu^* \left(\frac{x_i-x}{h}\right) \hat{f}_T(x_i)(1+o(1)) \\ &\equiv \frac{b\nu!}{h^{\nu}} \sum_{i=1}^{g} W_{\nu,c}^* \left(\frac{x_i-x}{h}\right) \hat{f}_T(x_i)(1+o(1)),\end{aligned}$$

where

$$W_\nu^*(u) = \int_{-\infty}^{u} K_\nu^*(t)\,dt \text{ and } W_{\nu,c}^*(u) = \int_{-\infty}^{u} K_{\nu,c}^*(t)\,dt.$$

To see the equivalence, for fixed j and for $k \in \{1,\dots,g\}$ set

$$c_{kj} = \mathbf{1}_{[x_k-\frac{b}{2},x_k+\frac{b}{2}]}(X_j, \delta_j = 1).$$

Since the $X_1, X_2, \dots, X_n$ are i.i.d., the strong law of large numbers yields

$$n^{-1}b^{-1}\sum_{j=1}^{n} c_{ij} \overset{a.s.}{\to} b^{-1}\int_{x_i-\frac{b}{2}}^{x_i+\frac{b}{2}} f_T(y)(1-H(y))\,dy \simeq b^{-1}bf_T(x_i)(1-H(x_i)) = f_T(x_i)(1-H(x_i)). \quad (7)$$

Dividing the empirical estimate of $f_T(x_i)(1-H(x_i))$ by an estimate of the survival function $1-H(x)$ of the censoring distribution yields an estimate of $f_T(x_i)$. Following Marron and Padgett (1987), by reversing the intuitive role played by T_i and U_i, $1-H(x)$ can be estimated by the (sightly modified) Kaplan–Meier estimator,

$$1-\hat{H}(x) = \begin{cases} 1, & 0 \le x \le X_{(1)}, \\ \prod_{i=1}^{k-1}\left(\frac{n-i+1}{n-i+2}\right)^{1-\delta_{(i)}}, & X_{(k-1)} < x \le X_{(k)}, k = 2,\dots,n, \\ \prod_{i=1}^{n}\left(\frac{n-i+1}{n-i+2}\right)^{1-\delta_{(i)}}, & X_{(n)} < x. \end{cases}$$

For $x \in [0, M]$, $\hat{H}$ converges strongly to H as according to Theorem 2.1 of Chen and Lo (1997), for $0 < p < 1/2$

$$\sup_{x\le M} |\hat{H}(x) - H(x)| = o(n^{-p}) \text{ a.s.}. \quad (8)$$

Hence, for fixed i and for $x_i \in [0, M]$ an empirical estimate of $f_T(x_i)$ at the ith bin center is obtained by

$$\hat{f}_T(x_i) = \frac{1}{n}\sum_{j=1}^{n}\frac{c_{ij}}{1-\hat{H}(x_i)} = \frac{1}{n}\sum_{j=1}^{n}\frac{c_{ij}}{1-H(x_i)}(1+o(n^{-p})) = bf_T(x_i)(1+o(n^{-p})).$$

By assumption A.3 below, $b = n^{-\lambda}$ with $1/2 < \lambda < 1$ and thus the term $o(bn^{-p})$ is asymptotically negligible. Also, for $X_i \in [x_j - b/2, x_j + b/2]$, $\hat{H}(X_i) = \hat{H}(x_j)(1+o(1))$. This, together with the asymptotic results in Satten and Datta (2001) page 209, allow writing the Kaplan–Meier as

$$\hat{F}_T(x_k) = n^{-1}\sum_{j=1}^{k}\sum_{i=1}^{n}\frac{\mathbf{1}_{[x_j-b/2,x_j+b/2]}(X_i,\delta_i=1)}{1-\hat{H}(X_i)} = \sum_{j=1}^{k} b\hat{f}_T(x_j)(1+o(1)),$$

from which the equivalence between the two formulations of $\hat{F}_L^{(\nu)}(x)$ immediately follows. Now, consider the functional

$$\theta_{\mu,\gamma} = \int_0^M F_T^{(\mu)}(x) F_T^{(\gamma)}(x)\, dx, \quad \mu, \gamma \geq 0,$$

where $\mu + \gamma$ is an even integer. Estimates of $\theta_{\mu,\gamma}$ are routinely employed in automatic (plug-in) bandwidth selectors. Using classical kernel smoothers in the place of $F_T^{(\mu)}(x)$ and $F_T^{(\gamma)}(x)$ will likely lead to inefficient endpoint estimation and diminish the global MSE rate of convergence of the resulting functional estimate. The absence of endpoint effects of $\hat{F}_L^{(\nu)}(x)$ motivates its use for effectively estimating $\theta_{\mu,\gamma}$ by

$$\begin{aligned} \hat{\theta}_{\mu,\gamma}(a) \equiv \hat{\theta}_{\mu,\gamma} \equiv \int_0^M \hat{F}_L^{(\mu)}(x) \hat{F}_L^{(\gamma)}(x)\, dx \simeq b \sum_{i=1}^{g} \hat{F}_L^{(\mu)}(x_i) \hat{F}_L^{(\gamma)}(x_i)\, dx \\ = b\mu!\gamma! \sum_{i=1}^{g} \sum_{j=1}^{g} \sum_{k=1}^{g} K_\mu \left(\frac{x_k - x_i}{a} \right) K_\gamma \left(\frac{x_k - x_j}{a} \right) \hat{F}_T(x_i) \hat{F}_T(x_j), \end{aligned} \tag{9}$$

where a denotes a bandwidth, typically different than h_ν. The asymptotic properties of $\hat{F}_L^{(\nu)}(x)$ and $\hat{\theta}_{\mu,\gamma}(a)$ are discussed next.

3. Asymptotic properties.

Denote the bias and variance of $\hat{F}_L^{(\nu)}(x)$ respectively by

$$b_{L(,c)}(x) = \begin{cases} b_{L,c}(x), & x \in [0,h) \cup (M-h, M] \\ b_L(x), & x \in [h, M-h] \end{cases},$$

$$\sigma^2_{L(,c)}(x) = \begin{cases} \sigma^2_{L,c}(x), & x \in [0,h) \cup (M-h, M] \\ \sigma^2_L(x), & x \in [h, M-h] \end{cases}.$$

Set

$$g(x) = f_T(x)(1 - H(x))^{-1}, \;\; G(x) = \int_0^x g(t)\, dt,$$

and define the constant

$$C_1 = \int_0^M g(x)\, dx. \tag{10}$$

Similarly to the definition of $b_{L(,c)}(x)$ and $\sigma^2_{L(,c)}(x)$, let $K^*_{\nu(,c)}$ and $W^*_{\nu(,c)}$ stand for $K^*_{\nu,c}$ and $W^*_{\nu,c}$ respectively in the boundary and K^*_ν and W^*_ν in the interior. In what follows, focus is given on the left boundary, i.e. $x = ch \in [0,h), 0 < c < 1$, since treatment of the right boundary, i.e. $x \in [M-h, M]$ is similar in an obvious manner. The case of estimation in the interior, i.e. $x = ch, c > 1$ so that $x \in [h, M-h)$ is obtained by letting $c \to \infty$. Let,

$$A_{i,j(,c)} = \begin{cases} A_{i,j,c} = \int_{-c}^{+\infty} x^i \{K^*_{\nu,c}(x)\}^j W^*_{\nu,c}(x)\, dx, \; i = 0,1,2\ldots, j = 1,2,\ldots, & 0 < c < 1, \\ A_{i,j} = \int_{-\infty}^{+\infty} x^i \{K^*_\nu(x)\}^j W^*_\nu(x)\, dx, \; i = 0,1,2\ldots, j = 1,2,\ldots, & c > 1, \end{cases}$$

that is, $A_{i,j(,c)}$ stands for $A_{i,j,c}$ for $x \in [0,h)$ and $A_{i,j}$ when x is in the interior. Similarly, for a positive integer l, let $\mu_{l(,c)}(K^*_{\nu(,c)})$ denote $\mu_l(K^*_\nu)$ in the interior and $\mu_{l,c}(K^*_{\nu,c})$ in the boundary. Let h_ν denote the bandwidth used when estimating the νth derivative of F_T. The following conditions are used throughout.

A.1 The kernel K is symmetric about the origin and satisfies

$$\int K^2 < +\infty, \int |u^2 K| < +\infty, \text{ and } \int uK = 0.$$

A.2 The kernel K has bounded support, vanishes at its endpoints and its first ν derivatives exist.

A.3 Assume $b = n^{-\lambda}$, where $1/2 < \lambda < 1$, g is such that $gb \to \infty$ and $g/n \to 0$. Also, for $\nu = 1, \ldots,$ $h_\nu \to 0$ and $b^{-1} h_\nu^{\nu+1} \to \infty$ as $n \to \infty$, i.e. as n grows, the bandwidth grows much faster than b.

A.4 As $n \to +\infty$, $h_\nu \to 0$ and for $\nu = 1, \ldots,$ $nh_\nu^\nu \to \infty$.

A.5 For fixed $\nu, F_T^{(\nu)}(x)$ is Lipschitz continuous and differentiable.

The asymptotic properties of $\hat{F}_L^{(\nu)}$ are summarized in the next theorem which is proved in Ioannides and Bagkavos (2019).

Theorem 1. *Assume that for $l = 0, \ldots, \nu + 1, K^{(l)}$ is bounded, absolutely integrable, with finite second moments and F_T is $l + 2$ times differentiable. Assume also that as $n \to +\infty$, $h_\nu \to 0, nh_\nu^{2\nu} \to +\infty$ and $b/h_\nu \to 0$. Then, the asymptotic bias and variance of $\hat{F}_L^{(\nu)}(x)$ are given by*

$$\begin{aligned} b_{L(,c)}(x) &= h_\nu^2 \frac{\nu!}{(\nu+2)!} \mu_{\nu+2(,c)}(K^*_{\nu(,c)}) F_T^{(\nu+2)}(x) + o(h_\nu^2), \\ \sigma^2_{L(,c)}(x) &= \frac{(\nu!)^2}{nh_\nu^{2\nu}} \Bigg[G(x) - 2h_\nu g(x) \int tK^*_{\nu(,c)}(s) W^*_{\nu(,c)}(s)\, ds \\ &\quad - \left\{ F_T^{(\nu)}(x) + h_\nu^2 \nu! ((\nu+2)!)^{-1} \mu_{\nu+2,c}(K^*_{\nu(,c)}) F_T^{(\nu+2)}(x) \right\}^2 \Bigg] + O(n^{-1} h_\nu^{2\nu}) + o(h_\nu^4), \end{aligned}$$

where

$$g(x) = f_T(x)(1 - H(x))^{-1}, \ \ G(x) = \int_0^x g(t)\, dt, \ \ W^*_{\nu(,c)}(s) = \int_{-\infty}^s K^*_{\nu(,c)}(u)\, du.$$

Further,

$$\hat{F}_L^{(\nu)}(x) \sim N\left(F_T^{(\nu)}(x) + b_{L(,c)}(x), \sigma^2_{L(,c)}(x) \right).$$

Remark 1. *The above results imply that $\hat{F}_L^{(\nu)}$ achieves the same rate of convergence in the boundary and in the interior and that the derivative order leaves the bias rate of convergence unaffected. However the second term on the right hand side of the variance expression is negative which implies that kernel smoothing improves the estimate variance by a second order effect.*

Remark 2. *Theorem 1 indicates two limitations that might be encountered in finite sample implementations of $\hat{F}_L^{(\nu)}$. One issue is the presence of $1 - H(x)$ in the denominator of the leading term in $\sigma^2_{L(,c)}(x)$. Even though this does not affect the variance rate of convergence it is expected to disproportionately inflate the estimate's variance in comparison to the uncensored case. Thus large amounts of censoring are expected to diminish the estimate's precision. Another point where caution is needed is that $\hat{F}_L^{(\nu)}$ might exhibit diminished finite sample performance at the right end point, as a consequence of the unreliable behavior of $\hat{F}_T$ for $x \in [M_F, M]$ where* *typically M_F denotes the largest uncensored observation, see for example Chen and Lo (1997). On the contrary, for $x \in [0, M_F]$, from (2.7) in Karunamuni and Yang (1991), $\hat{F}_T$ converges in probability to F_T with rate $n^{-1/2}$ implying a robust behavior there also for $\hat{F}_L^{(\nu)}$.*

Turning attention to the asymptotic properties of $\hat{\theta}_{\mu,\gamma}(a)$, these are given in the next theorem. Its proof is based on the strong convergence of $\hat{H}$ to H, repeated use of Lemma 1 of Ioannides and Bagkavos (2019) as well as repeated Riemannian approximations of integrals by sums using Lemma 2 in Bagkavos and Patil (2008). A full proof is available from the authors; see also the proof of Theorem 7 in Cheng (1994) (equiv. Theorem 2 in Cheng (1997)) for a very similar proof on the complete data density estimation setting.

Theorem 2. *Assume that F_T is $\mu+\gamma$ times differentiable. Assume also K is compactly supported and twice differentiable. Then, as $n \to \infty, a \to 0, na^{\mu+\gamma+1} \to \infty$ and $b/h_\nu \to 0$,*

$$\begin{aligned} \mathrm{E}\hat{\theta}_{\mu,\gamma}(a) - \theta_{\mu,\gamma} &= \frac{\mu!\gamma!}{na^{\mu+\gamma-1}} \left\{ \int \left(\frac{f_T(u)}{1-H(u)} \right) du \right\} \int W_\mu^* K_\gamma^* \\ &+ \frac{(1+\delta_{\mu\gamma})\gamma!}{(\gamma+2)!} a^2 \theta_{\mu,\gamma+2} \mu_{\gamma+2}(K_\gamma^*) + O(n^{-1}a^{-(\mu+\gamma)}) + o(a^2), \\ \mathrm{Var}\left\{ \hat{\theta}_{\mu,\gamma}(a) \right\} &= \frac{2(\mu!\gamma!)^2}{n^2 a^{2(\mu+\gamma)-1}} R(g) R(W_\mu^* K_\gamma^*) \\ &+ \frac{4}{n} \left\{ \int F_T \left(F_T^{(\mu+\gamma)} \right)^2 - \theta_{\mu,\gamma} \right\} + o(n^{-2}a^{-2(\mu+\gamma)+1}) + o(n^{-1}). \end{aligned}$$

Remark 3. *The immediate conclusion from Theorem 2 is that the leading squared bias term is of order $n^{-2}a^{-2(\mu+\gamma-1)}$ while the variance leading term is of order $n^{-2}a^{-2(\mu+\gamma)+1}$. Therefore bias dominates variance in the MSE expression of $\hat{\theta}_{\mu,\gamma}(a)$. This fact implies that in minimizing the functional's MSE expression with respect to a, it is enough to consider only the bias part.*

4. Plug in bandwidth selection.

By Theorem 1 and since the Lebesgue measure of $[0, h)$ tends to zero and therefore the corresponding integral is zero, the MISE of $\hat{F}_L^{(\nu)}(x)$ can be decomposed as

$$\begin{aligned} \mathrm{MISE}\left\{ \hat{F}_L^{(\nu)}(x) \right\} &= \int_0^{h_\nu} \mathrm{MSE}\left\{ \hat{F}_L^{(\nu)}(x) \right\} dx + \int_{h_\nu}^{M} \mathrm{MSE}\left\{ \hat{F}_L^{(\nu)}(x) \right\} dx \\ &\simeq \frac{h_\nu^4}{4} \mu_{\nu+2}^2(K_\nu^*) R\left(F_T^{(\nu+2)} \right) + \frac{(\nu!)^2}{nh_\nu^{2\nu}} \int_{h_\nu}^{M} G(x)\, dx - 2\frac{(\nu!)^2}{nh_\nu^{2\nu-1}} C_1 A_{1,1} \\ &\quad - h_\nu \int_{h_\nu}^{M} \left\{ F_T^{(\nu)}(x) + h_\nu^2 \nu!((\nu+2)!)^{-1} \mu_{\nu+2}(K_\nu^*) F_T^{(\nu+2)}(x) \right\}^2 dx \\ &\quad + O(n^{-1}h_\nu^{2\nu}) + o(h_\nu^4), \end{aligned} \quad (11)$$

where $a_n = O(b_n)$ if and only if $\limsup_{n\to\infty} |a_n/b_n| < \infty$. Write

$$\mathrm{MISE}\left\{ \hat{F}_L^{(\nu)} \right\} = \mathrm{AMISE}\left\{ \hat{F}_L^{(\nu)} \right\} + O(n^{-1}h_\nu^{-2\nu}) + o(h_\nu^4), \quad (12)$$

where

$$\mathrm{AMISE}\left\{ \hat{F}_L^{(\nu)} \right\} = \frac{h_\nu^4}{4} \mu_{\nu+2}^2(K_\nu^*) R\left(F_T^{(\nu+2)} \right) - 2\frac{(\nu!)^2}{nh_\nu^{2\nu-1}} C_1 A_{1,1}. \quad (13)$$

Note that under assumption A.4 the asymptotic terms in (12) vanish as the sample size increases. This means that the MISE of the local polynomial estimator can be effectively approximated by its AMISE. In turn this facilitates approximation of the MISE optimal bandwidth by the minimizer of (13), obtained by solving

$$\frac{\partial \text{AMISE}\left\{\hat{F}_L^{(\nu)}\right\}}{\partial h_\nu} = h_\nu^3 \mu_{\nu+2}^2(K_\nu^*) R(F_T^{(\nu+2)}) + 2(2\nu-1)\frac{(\nu!)^2}{n h_\nu^{2\nu}} C_1 A_{1,1} = 0,$$

for h_ν. This yields the optimal bandwidth rule

$$h_\nu = (-1)^{\nu+1}\left(\frac{2}{n}\right)^{\frac{1}{2\nu+3}} \left\{\frac{(2\nu-1)(\nu!)^2 C_1 A_{1,1}}{\mu_{\nu+2}^2(K_\nu^*) R(F_T^{(\nu+2)})}\right\}^{\frac{1}{2\nu+3}}. \tag{14}$$

Obviously h_ν cannot be used in practice as it depends on C_1 and $R(F_T^{(\nu+2)})$ which are unknown. $R(F_T^{(\nu+2)})$ is estimated by $\hat{\theta}_{\mu,\gamma}(a)$ for $\mu=\gamma=\nu+2$. Of course, $\hat{\theta}_{\nu+2,\nu+2}(a)$ requires an optimal bandwidth rule for a. Based on Remark 3, the MSE optimal bandwidth for $\hat{\theta}_{\mu,\gamma}$, denoted by $a_{\mu,\gamma}$, is given by

$$a_{\mu,\gamma} = \left\{\frac{\chi C_2 A_{0,1}}{n \mu_{\gamma+2}(K_\gamma^*)\theta_{\mu,\gamma+2}}\right\}^{\frac{1}{\mu+\gamma+1}}, \tag{15}$$

with

$$\chi = \begin{cases} -1 & \text{if } \theta_{\mu,\gamma+2} < 0 \\ \frac{(\mu+\gamma+1)\mu!(\gamma+2)!}{\gamma!(1+\delta_{\mu\gamma})} & \text{if } \theta_{\mu,\gamma+2} > 0 \end{cases} \quad \text{and } C_2 = \int_0^M g^{(\nu+2)}(u)\,du.$$

From (14) and (15) with $\mu=\gamma=\nu+2$, the optimal AMISE bandwidth for $\hat{F}_L^{(\nu)}(x)$ is the solution of

$$\hat{h}_\nu = (-1)^{\nu+1}\left(\frac{2}{n}\right)^{\frac{1}{2\nu+3}} \left\{\frac{(2\nu-1)(\nu!)^2 C_1 A_{1,1}}{\mu_{\nu+2}^2(K_\nu^*)\hat{\theta}_{\nu+2,\nu+2}(a(\hat{h}_\nu))}\right\}^{\frac{1}{2\nu+3}}, \tag{16}$$

with respect to $\hat{h}_\nu$, where

$$a(\hat{h}_\nu) = C(K) D(\theta) \hat{h}_\nu^{\frac{2\nu+1}{2\nu+3}}, \tag{17}$$

with

$$C(K) = \left\{\frac{\mu_{\nu+2}^2(K_\nu^*) A_{0,1}}{2^{\frac{1}{2\nu+3}}(2\nu-1)(\nu!)^2 C_1 A_{1,1}\mu_{\nu+4}(K_{\nu+2}^*)}\right\}^{\frac{1}{2\nu+3}}, \tag{18}$$

$$D(\theta) = \left(\frac{\chi C_2 \theta_{\nu+2,\nu+2}}{\theta_{\nu+2,\nu+4}}\right)^{\frac{1}{2\nu+3}}. \tag{19}$$

When no analytic solution to (16) is feasible, $\hat{h}_\nu$ is obtained by a numerical procedure such as the Newton-Raphson method. Of course, implementation of $D(\theta)$ depends on estimation of $\theta_{\nu+2,\nu+2}$ and $\theta_{\nu+2,\nu+4}$. According to the conventional solve-the-equation approach one would go a stage further and apply local polynomial fitting for estimation of both functionals before using a parametric reference model. However, such an approach is subject to inherit large amount of variability from the data which results in the bandwidth selector to become unstable. Moreover it requires computations of inverses of matrices of increasingly large dimensions and thus in a considerable decrease in computational speed. For these two reasons it is more effective to adopt a parametric reference at this stage; this has been also advocated by Cheng (1997). A suitable default parametric estimate of $\theta_{\mu,\gamma}$ for $\mu = \nu + 2$ and $\gamma = \nu + 4$, say $\tilde{\theta}_{\mu,\gamma}$, is the two parameter Weibull distribution given by

$$\tilde{\theta}_{\mu,\gamma} = \int_0^M \left(e^{-(\rho t)^\kappa}\right)^{(\mu)} \left(e^{-(\rho t)^\kappa}\right)^{(\gamma)} dt,$$

where κ, ρ are the scale and location parameters of the Weibull model estimated by maximum likelihood. The choice of this particular distribution is justified by its wide use in survival analysis and by its flexibility as it can mimic the behavior of other distributions such as the Rayleigh and the normal. It goes without saying that in presence of even partial information about the underlying density, $\tilde{\theta}_{\mu,\gamma}$ should be adjusted accordingly. Thus the suggested bandwidth $\hat{h}_\nu$ results by (16) after substituting $D(\theta)$ in (17). Now, let

$$\alpha_1 = \begin{cases} \frac{2(2\mu\gamma+1)+\mu+\gamma+1}{2(\mu+\gamma+1)(\mu+\gamma-1)}, & \text{if } \theta_{\mu,\gamma} < 0 \\ \frac{2\mu\gamma+1}{(\mu+\gamma+1)(\mu+\gamma-1)}, & \text{if } \theta_{\mu,\gamma} > 0 \end{cases},$$

and set

$$\begin{aligned} \mu_{DPI} &= n^{\alpha_1} \left\{ \frac{2}{n} \frac{(2\gamma-1)(\gamma!)^2 C_1 A_{1,1}}{\mu^2_{\gamma+2}(K^*_\gamma)\theta_{\gamma+2,\gamma+2}} \right\}^{\frac{1}{2\gamma+3}} \mu_\gamma^{\frac{\mu+\gamma}{\mu+\gamma-1}} \theta_{\mu,\gamma}^{\frac{1}{\mu+\gamma-1}} \\ &\quad + \frac{n^{\alpha_1}(1+\delta_{\mu\gamma})\gamma!}{(\gamma+2)!(\mu+\gamma+1)} \mu^{\frac{\mu+\gamma}{\mu+\gamma-1}} \left\{ \frac{2}{n} \frac{(2\gamma-1)(\gamma!)^2 C_1 A_{1,1}}{\mu^2_{\gamma+2}(K^*_\gamma)\theta_{\gamma+2,\gamma+2}} \right\}^{\frac{3}{2\gamma+3}} \\ &\quad \times \theta_{\mu,\gamma}^{\frac{2-(\mu+\gamma)}{\mu+\gamma-1}} \theta_{\mu,\gamma+2}\mu_{\gamma+2}(K^*_\gamma)\mathbf{1}_{[\theta_{\mu,\gamma}>0]}, \\ \sigma^2_{DPI} &= \frac{2(\mu!\gamma!)^2}{\mu+\gamma-1} n^{2\alpha_1-2} \mu_\gamma^{\frac{2(\mu+\gamma)}{\mu+\gamma-1}} \left\{ \frac{2}{n} \frac{(2\gamma-1)(\gamma!)^2 C_1 A_{1,1}}{\mu^2_{\gamma+2}(K^*_\nu)\theta_{\gamma+2,\gamma+2}} \right\}^{\frac{2(2\gamma+3)-(2\gamma+1)(2(\mu+\gamma)-1)}{(2\gamma+3)^2}} \\ &\quad \times \{C(K)D(\theta)\}^{-2(\mu+\gamma)+1} R(g)R(C^n_\mu * W^n_\gamma)\theta_{\mu,\gamma}^{\frac{4-2(\mu+\gamma)}{\mu+\gamma-1}}. \end{aligned}$$

The rate of convergence of $\hat{h}_\nu$ to the *ideal* bandwidth h_ν and its asymptotic distribution are quantified in the next theorem. First, let $a_n = O_p(b_n)$ denote that the sequence a_n is bounded in probability at the "rate" b_n, i.e. for each $\varepsilon > 0$ there exist M, N depending on ε such that $P(|a_n| \leq M|b_n|) > 1 - \varepsilon$, for all $n \geq N$.

Theorem 3. *Under conditions A.1–A.2, as $n \to \infty$,*

$$\frac{\hat{h}_\nu}{h_\nu} = 1 + O_p(n^{-\alpha}),$$

where

$$\alpha = \begin{cases} \frac{\mu+\gamma-1}{2(\mu+\gamma+1)}, & \textit{if } \theta_{\mu,\gamma} < 0, \\ \frac{2}{\mu+\gamma+1}, & \textit{if } \theta_{\mu,\gamma} > 0. \end{cases}$$

Further,

$$n^\alpha \left(\frac{\hat{h}_\nu}{h_\nu} - 1 \right) \xrightarrow{d} N(\mu_{DPI}, \sigma^2_{DPI}).$$

The conclusion from Theorem 3 is that the proposed bandwidth selector is expected to achieve optimal results faster (i.e. with smaller samples) compared to traditional approaches such as those based on cross validation or the Akaike Information Criterion. Its practical performance is exhibited by three applications to real world data sets and finite sample MISE simulations in the next section.

5. Numerical examples

Throughout this section, binning of each sample $(X_i, \delta_i), i = 1, \ldots, n$ is performed by splitting the observed data range into $g = [(X_{(n)} - X_{(1)})/b]$ disjoint intervals of equal length. In accordance to assumption A.3 the length is set to $b = x_2 - x_1 = n^{-3/4}$. Now, let $\hat{F}_L^{(0)} \equiv \hat{F}_L$ be the estimate of F_T and let $\hat{S}_L(x) = 1 - \hat{F}_L(x)$ be the corresponding survival function estimate. In all examples $\hat{F}_L$ and $\hat{S}_L$ are implemented with the MISE optimal bandwidth obtained by (16) for $\nu = 0$ after replacing the unknown quantities by suitable estimates i.e.

$$\hat{h}_0 = \left(\frac{2}{n}\right)^{\frac{1}{3}} \left\{ \frac{\hat{C}_1 A_{1,1}}{\mu_2^2(K_0^*)\hat{\theta}_{2,2}(a(\hat{h}_0))} \right\}^{\frac{1}{3}}. \tag{20}$$

In (20) and throughout the section $\hat{C}_1$ is the estimate of C_1 defined in (10), obtained by replacing the unknown quantities by consistent data driven estimates, i.e.

$$\hat{C}_1 = \int_0^M \hat{f}(x)(1 - \hat{H}(x))^{-1}\, dx, \tag{21}$$

where $M = X_{(n)}$ and $\hat{f}(x)$ is the density estimate of Marron and Padgett (1987) given by

$$\hat{f}(x) = \sum_{i=1}^{n} \frac{\delta_i}{n(1 - \hat{H}(x))s} K\left(\frac{x - X_i}{s}\right). \tag{22}$$

Estimator $\hat{f}(x)$ is implemented with the Integrated Square Error optimal bandwidth s resulting by the cross validation rule of Marron and Padgett (1987). Integration in (21) is performed by Simpson's rule. Both $A_{1,1}$ and $\mu_2(K_0^*)$ are calculated analytically based on the Epanechnikov kernel which is also used in all kernel implementations throughout this section. $\hat{\theta}_{2,2}(a(\hat{h}_0))$ is calculated by combining (9) with $\mu = \gamma = 2$ and (17) with $a(\hat{h}_0) = C(K)D(\theta)\hat{h}_0^{1/3}$. $C(K)$ is approximated by setting $\nu = 0$ in (18), estimating C_1 by $\hat{C}_1$ and calculating analytically the values of $\mu_2^2(K_0^*), A_{0,1}, A_{1,1}$ and $\mu_4(K_2^*)$. Similarly, $D(\theta)$ is approximated by replacing $\theta_{2,2}$ and $\theta_{2,4}$ in (19) with $\tilde{\theta}_{2,2}$ and $\tilde{\theta}_{2,4}$ respectively, obtained by either a Weibull reference model or, when available as it is the case in the first two real data examples below, by utilizing any existing maximum likelihood estimate. Throughout this section, integration in the definition of $\tilde{\theta}_{\mu,\gamma}$ is performed analytically (when feasible) or otherwise numerically by adaptive quadrature (function `integrate` in `R`). The constant C_2 in (19) is estimated by applying Simpson's rule on the second derivative of $\hat{f}(x)(1 - \hat{H}(x))^{-1}$, calculated by numerical differentiation. Using the estimates of $D(\theta)$ and $C(K)$ in $a(\hat{h}_0)$, substituting back to (20) and solving for $\hat{h}_0$ with Newton-Raphon yields the bandwidth used with $\hat{F}_L$.

The conventional kernel survival function estimate of Gulati and Padgett (1996) is also used throughout the section for comparison. The estimate is given by $\hat{S}(x) = 1 - \hat{F}(x)$ where

$$\hat{F}(x) = \hat{h}^{-1} \sum_{i=1}^{n} K\left(\frac{x - X_i}{\hat{h}}\right) \hat{F}_T(X_i). \tag{23}$$

In (23) $\hat{h} \equiv \hat{h}(x)$ is the default asymptotic MSE optimal rule of Gulati and Padgett (1996) given by

$$\hat{h} = \left\{ \frac{2\hat{f}(x)A_{1,1}}{n(1 - \hat{H}(x))(\hat{f}'(x))^2 \mu_2^2(K)} \right\}^{1/3},$$

where $\hat{f}'(x)$ is the first derivative of $\hat{f}(x)$ calculated by analytical differentiation of (22). The performance of the density estimate $\hat{F}_L^{(1)}(x) \equiv \hat{f}_L(x)$, resulting from (3) for $\nu = 1$, is also investigated in this section. Its bandwidth $\hat{h}_1$ is obtained by (16) for $\nu = 1$ as the solution (by Newton-Raphson) of

$$\hat{h}_1 = \left(\frac{2}{n}\right)^{\frac{1}{5}} \left\{ \frac{\hat{C}_1 A_{1,1}}{\mu_3^2(K_1^*)\hat{\theta}_{3,3}(a(\hat{h}_1))} \right\}^{\frac{1}{5}}. \tag{24}$$

In (24), $\hat{C}_1$ is again provided by (21), while $A_{1,1}$ and $\mu_3(K_1^*)$ are calculated analytically. $\hat{\theta}_{3,3}(a(\hat{h}_1))$ is calculated by combining (9) with $\mu = \gamma = 3$ and (17) with $a(\hat{h}_1) = C(K)D(\theta)\hat{h}_1^{1/5}$. $C(K)$ is approximated by setting $\nu = 1$ in (18), estimating C_1 by $\hat{C}_1$ and calculating $\mu_3^2(K_1^*), A_{0,1}, A_{1,1}$ and $\mu_5(K_3^*)$ analytically. Similarly, $D(\theta)$ is approximated by estimating $\theta_{3,3}$ and $\theta_{3,5}$ in (19) by $\tilde{\theta}_{3,3}$ and $\tilde{\theta}_{3,5}$ respectively, obtained by either the Weibull reference model or, if available, by utilizing any existing maximum likelihood estimates. The constant C_2 in (19) is estimated by applying Simpson's rule on the third derivative of $\hat{f}(x)(1 - \hat{H}(x))^{-1}$, calculated by numerical differentiation. Using the estimates of $D(\theta)$ and $C(K)$ in $a(\hat{h}_1)$, substituting back to (24) and solving for $\hat{h}_1$ yields the estimate of the AMISE optimal bandwidth h_1.

5.1. Danish fire loss data example

The first example is from the insurance practice and analyzes the Danish Fire Loss data, collected at Copenhagen Reinsurance. The data set comprises of 2167 fire losses over the period 1980 to 1990 and have been adjusted for inflation to reflect 1985 values. The observations are expressed in millions of Danish Krones. McNeil (1997) analyzed two subsets of the data, one consisting of values greater than 10 and the second with values of 20 million Krones respectively. Focusing on the first data set which consists of 109 observations, McNeil (1997) modeled the c.d.f. of the fire losses by testing three different distributions. These are the truncated lognormal, the Pareto and the Generalized Pareto distribution (GDP), with their parameters estimated by maximum likelihood. McNeil (1997) concluded that no single parametric model is totaly satisfactory, however the GDP with location parameter 10, scale parameter 6.98 and shape parameter 0.497 is perhaps the most suitable. Fig. 1 replicates the survival function estimate resulting from this model and compares it with $\hat{S}_L$ and $\hat{S}$.

The first outcome from Fig. 1 is that as expected, $\hat{S}_L$ corrects the boundary bias problem of $\hat{S}$. The second outcome is that $\hat{S}_L$ suggests a change in the fire loss data distribution between approximately 25 to 40 million Krones. Even though this is also expected based on the discussion in McNeil (1997), it is not captured by either the parametric nor the conventional kernel estimate. Even though the maximum likelihood estimate is based on the three parameter GDP distribution, still the shape restrictions imposed throughout the region of estimation dominate and mask the important features of the curve such as this shape change. The conventional kernel estimate $\hat{S}$ is somewhat more flexible and close to $\hat{S}_L$ in the interior. However the edge effects of $\hat{f}$, which are carried over in $\hat{f}'$, inherit excessive bias in the calculation of $\hat{h}$ resulting to a higher value than $\hat{h}_0$. In turn this results in an oversmoothed estimate which masks the true survival function shape. This can be seen also by the inflated estimation between approximately 15 to 20 million Krones as well as from the overestimation of the pattern change between $x = 25$ and $x = 40$. On the contrary $\hat{S}_L$ with the proposed bandwidth selector readily adjusts its estimation at the boundary and offers precise estimation of the true curve throughout the region of estimation, resulting in enhanced insights and inference.

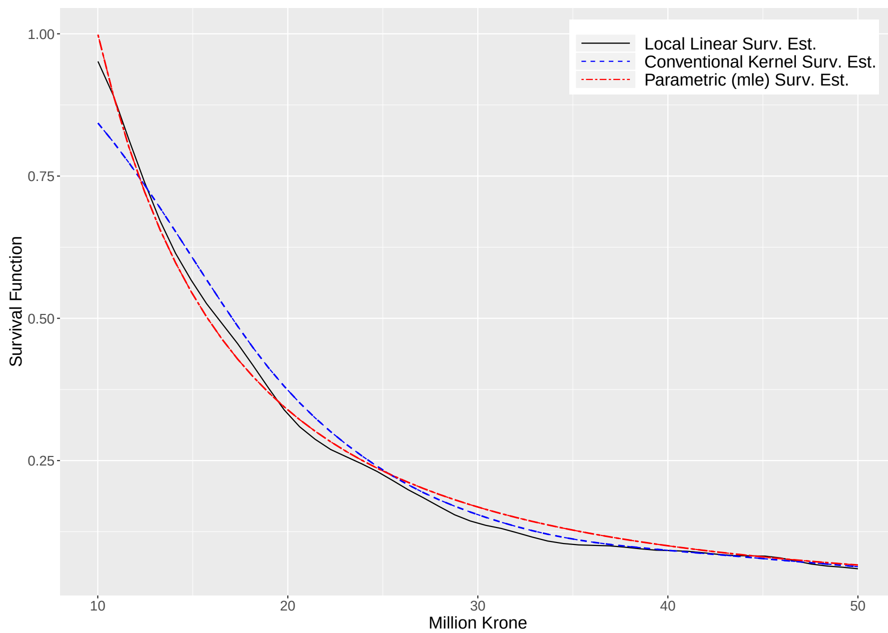

Figure 1: Parametric and local polynomial estimates for the Danish fire loss data.

5.2. Air conditioning unit failure data example

The second example is based on the well known air conditioning unit failure data set of Proschan (1963), available in Table 11.9, Lai and Xie (2006). The pooled data set consists of 213 time intervals (observations), in hours, between successive failures of the air conditioning system of each member of a fleet of 13 Boeing 720 jet airplanes. Multiple attempts in the literature are based on modeling the underlying survival function parametrically by maximum likelihood, assuming a specific distribution; see Kus (2007) and the references therein for an account of important contributions up to that point. However adopting a single parametric model at fleet level would imply that all failure times follow the same distribution irrespective of which plane they come from. This imposes a strong assumption which a practitioner would find rather unrealistic since it is expected that different planes are exposed to different conditions which affect failure occurrences. Even though using mixtures of distributions provides flexibility in capturing different data patterns in a single model, in practice this model would be uncertain as adding or deleting one plane from the sample would change the mixture.

Fig. 2 illustrates the survival function estimate suggested by $\hat{S}_L$. For comparison the survival function estimate proposed by Proschan (1963), given by $S_P(t) = \exp\{-t/93.14\}$ is also included. Even though S_P is not regarded as a realistic model since it is based on the exponential distribution and hence suggests that failures decline with time, its inclusion as a benchmark estimate (among many other parametric models) is justified because its goodness of fit is not rejected by the Kolmogorv–Smirnov test and hence corroborates with the overall data pattern. Fig. 2 exemplifies the versatility of $\hat{S}_L$. The probability of failure changes pattern as the time between successive failures increases. On the contrary S_P, even though it is close to $\hat{S}_L$, seems unable to capture this change proposing a strictly decreasing failure probability model. Consequently $\hat{S}_L$ would be more useful to a practitioner for reliability assessment and maintenance planning at fleet level.

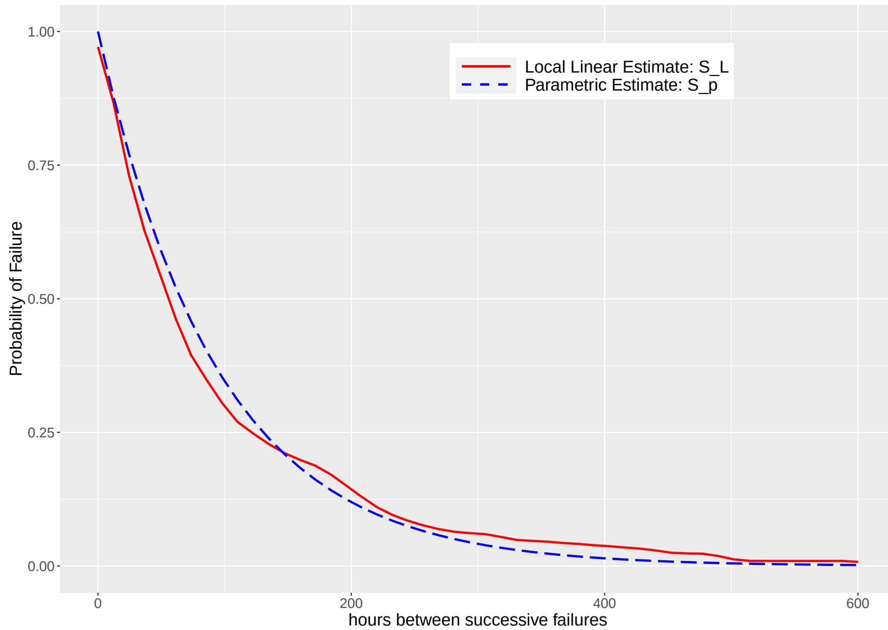

Figure 2: Parametric (S_P) and local polynomial $\hat{S}_L$ estimates for the air conditioning unit failure data.

5.3. Rear dump truck data example

The third application utilizes the rear dump truck data, analyzed among many others in Pulcini (2001) and Hua et al. (2017). The data set represents the time (in 1000's of hours) between failures of a 180-ton rear dump truck. The values in the data set indicate a *bathtub* failure model since the majority of observations occur either at the beginning of the data range (implying a faulty construction), or towards the end (implying ageing or wear out effects). In this and in similar occasions, practitioners are mostly interested on the shape of the failure pattern rather than estimating the probability of a failure occurring after a certain time point. Hence the most appropriate model for analyzing this data set would be an estimate of the hazard rate function which expresses the instantaneous probability of a failure in the next time instant, given that no failure has occurred up to that point. Now, the hazard rate function is defined as the ratio of the underlying density over the corresponding survival function. Thus, a sensible hazard rate estimate, say $\hat{\lambda}_L(x)$ will result by dividing $\hat{f}_L(x)$ with $\hat{S}_L(x)$.

$\hat{\lambda}_L(x)$ is implemented in Fig. 3. The estimate confirms the bathtub nature of the process, first identified in the histogram estimate of Pulcini (2001) and further explored in the semiparametric estimate of Hua et al. (2017). The discontinuity (i.e. the crude nature) of the histogram estimate in Pulcini (2001) prevents it from capturing the more subtle features of the failure pattern. This was achieved in Hua et al. (2017) where the change in failure was quantified in accordance to the increase of time in operation. However, the functional form of the parametric components in Hua et al. (2017) led to suggesting different patterns of failure on the same time frame. First $\hat{\lambda}_L(x)$ corrects the end point effects of the estimate in Hua et al. (2017) and second provides an unbiased point of view on the various shape changes free from the distributional assumptions. Specifically the local polynomial

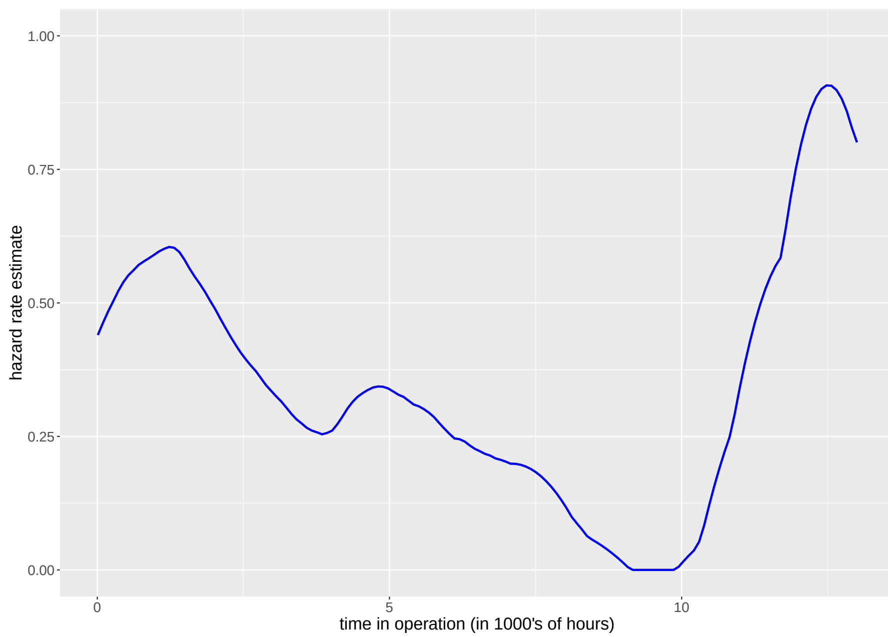

Figure 3: Local polynomial hazard rate estimate for the rear dump truck data.

estimate suggests that the first decline in failure starts at about 2,000 hours of operation up to approximately 4,000 hours. This is followed by a slight increase in failure up to approximately 5,000 hours and then declines until about 10,000 hours of operation where it shortly stabilizes, followed by a sharp increase afterwards which indicates the kick in of ageing and wear out effects.

5.4. MISE simulation examples

The last set of numerical examples uses distributional data to simulate the finite sample MISE performance of $\hat{F}_L$ and $\hat{f}_L$. Three distributions with different shape, routinely employed in modeling lifetime data, are used for this purpose. These are the positive truncated normal mixtures $\frac{2}{3}N^+(4, 0.4^2) + \frac{1}{3}N^+(3, 0.2^2)$ (NM1) and $0.6N^+(-3, 9) + 0.4N^+(10, 9)$ (NM2), where N^+ denotes a truncated normal distribution. The positive truncated normal density (Navarro and Hernandez, 2004) is defined by

$$f^+(t) = c \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\}, c > 0,\ c = c(\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{\Phi(\mu\sigma^{-1})},$$

where Φ denotes the standard normal c.d.f.. The third distribution is the one parameter Birnbaum–Saunders (BS, also known as fatigue life distribution) given by

$$f(x; \alpha) = \frac{\sqrt{x} + \sqrt{\frac{1}{x}}}{2\alpha x} \phi \left(\frac{\sqrt{x} + \sqrt{\frac{1}{x}}}{\alpha} \right). \tag{25}$$

In all examples (25) is implemented with the shape parameter $\alpha = 1.75$. For benchmarking purposes, the performance of $\hat{F}_L$ is compared to the kernel c.d.f. estimate $\hat{F}$ defined in (23) and to the corresponding parametric (maximum likelihood) estimate of each distribution, denoted by $\tilde{F}$. The comparison is performed on four different sample sizes, $n = 50, 100, 150$ and 250 and at four levels of censoring: 0% (no censoring), 15%, 30% and 50%.

Random censoring is implemented by independently generating n random censoring times from the uniform $U[0,k]$ distribution where k is selected so that the desired percentage of censoring is achieved on average across all iterations. Since $U[0,k]$ does not depend on the parameters of any of the three distributions considered here, the likelihood function is given by

$$L = \prod_{i=1}^{n} \{f(X_i;\boldsymbol{\theta})\}^{\delta_i} \{1 - F(X_i;\boldsymbol{\theta})\}^{1-\delta_i} . \tag{26}$$

In (26) $\boldsymbol{\theta}$ denotes the vector of unknown parameters and $F' = f$. In the case of the two normal mixture distributions

$$f(X_i;\boldsymbol{\theta}) \equiv f(X_i;\mu_1,\sigma_1,\mu_2,\sigma_2) = pf^{+}(X_i;\mu_1,\sigma_1) + (1-p)f^{+}(X_i;\mu_2,\sigma_2),$$

where for NM1 $(\mu_1,\sigma_1,\mu_2,\sigma_2) = (4, 0.4, 3, 0.2)$ and for NM2 $(\mu_1,\sigma_1,\mu_2,\sigma_2) = (-3, 9, 10, 9)$. For the Birnbaum–Saunders distribution $\boldsymbol{\theta} = \alpha$. Maximization of (26) with respect to the unknown parameters is performed with the `R` package `maxLik`. Similarly, the performance of $\hat{f}_L$ is benchmarked against $\hat{f}$, implemented as described in (22) and the corresponding maximum likelihood density estimates, denoted by $\tilde{f}$.

For each distribution, sample size and level of censoring the approximate Mean Integrated Squared Error of each estimate is calculated as follows. The differences $(\breve{F}^{(\nu)}(x_i) - F^{(\nu)}(x_i))^2, \nu = 0, 1$, with $\breve{F}^{(\nu)}$ being any of the three estimates considered here and $F^{(\nu)}(x_i)$ the true curve at x_i, are calculated for all equally spaced grid points $x_i, i = 1, \ldots, 50$. Simpson's extended rule is then applied to obtain the integrated square error approximation for each estimate. The averaged integrated differences across 10,000 iterations are reported on the tables. In every iteration the same sample values are used in calculating all estimators. Note that selection of the classical convolution kernel distribution and density estimates for benchmarking the MISE figures of the local polynomial estimates is sought so as to understand the gain in precision from the boundary correction and from the utilization of MISE optimal bandwidth selection rules. Similarly, inclusion of maximum likelihood estimates in the comparison is sought, not with purpose to identify which estimate is the best, but rather as a benchmark of the achieved improvements. For this reason, in cases where maximization of the likelihood function for a specific sample failed, calculation of the MISE for all estimates was repeated by drawing another sample until achieving convergence. In other words, the maximum likelihood MISE figures should be regarded as an *ideal* but nevertheless useful indication of the heights to which a very precise estimation procedure might achieve.

The results of the simulation in Tables 1 and 2 illustrate the benefits from both the boundary correction and the bandwidth selectors introduced with the local polynomial smoothing of the Kaplan-Meier estimate. The MISE comparison between $\hat{F}_L, \hat{f}_L$ and their corresponding convolution kernel counterparts confirms that the local polynomial smoothers introduced in (3), implemented with the MISE optimal bandwidth (16), improve the precision in estimation across all three example distributions, sample sizes and levels of censoring. Taking into account the 'ideal' nature of the parametric MISE figures, the results in Tables 1 and 2 suggest that the performance of $\hat{F}_L$ and $\hat{f}_L$ is closer to $\tilde{F}$ and $\tilde{f}$ rather than to the performance of $\hat{F}$ and $\hat{f}$. Another useful outcome from the simulation is drawn by comparing the MISE increase between samples of the same magnitude so as to understand the effect of censoring. Specifically, the effect of censoring on $\hat{F}_L$ and $\hat{f}_L$ starts becoming visible on the MISE figures of $\hat{F}_L, \hat{f}_L$ when the samples contain 30%-50% censored observations and more (i.e.

Table 1: Approximate MISE's of $\hat{F}_L, \hat{F}$ and $\tilde{F}$ for the truncated normal mixtures and the Birnbaum–Saunders distribution.

		NM1			NM2			BS		
Cens.	n	$\hat{F}_L$	$\hat{F}$	$\tilde{F}$	$\hat{F}_L$	$\hat{F}$	$\tilde{F}$	$\hat{F}_L$	$\hat{F}$	$\tilde{F}$
0%	50	0.318	0.615	0.25	0.194	0.206	0.137	0.281	0.299	0.228
	100	0.237	0.571	0.189	0.123	0.134	0.083	0.178	0.194	0.139
	150	0.212	0.544	0.169	0.09	0.115	0.068	0.131	0.167	0.113
	250	0.181	0.514	0.145	0.066	0.092	0.048	0.095	0.134	0.081
15%	50	0.341	0.616	0.271	0.206	0.299	0.145	0.298	0.433	0.242
	100	0.262	0.557	0.214	0.153	0.156	0.095	0.222	0.226	0.159
	150	0.226	0.525	0.183	0.121	0.125	0.072	0.175	0.181	0.121
	250	0.196	0.493	0.157	0.09	0.102	0.055	0.131	0.148	0.091
30%	50	0.434	0.703	0.334	0.219	0.563	0.128	0.317	0.816	0.213
	100	0.355	0.622	0.324	0.213	0.293	0.111	0.308	0.425	0.183
	150	0.324	0.561	0.299	0.193	0.237	0.095	0.28	0.344	0.158
	250	0.29	0.519	0.281	0.178	0.197	0.081	0.258	0.285	0.135
50%	50	0.506	0.758	0.378	0.289	0.533	0.262	0.448	0.773	0.436
	100	0.468	0.683	0.353	0.266	0.492	0.241	0.429	0.713	0.401
	150	0.395	0.617	0.321	0.243	0.328	0.204	0.381	0.476	0.339
	250	0.361	0.554	0.302	0.228	0.288	0.195	0.359	0.417	0.324

medium to heavy censoring). On the contrary lower levels of censoring seem to have negligible effect on the precision of the estimates. This increase is predominantly driven by the presence of the censoring distribution's survival function in the denominator of the estimate's variance leading terms and by the slow convergence rate of the Kaplan–Meier on the right tail beyond the last uncensored observation, see Remark 2. Even though not reported here, exactly the same numerical experiments were repeated by restricting the estimation range to $[0, M_F]$ where the Kaplan–Meier behaves quite robustly. Thus the experiments simulated solely the impact of censoring on the estimate's variance. The MISE figures exhibited significant increase only for the censoring level of 50%; in turn this verifies the suggestion at the end of Remark 2 for the robust performance of $\hat{F}_L^{(\nu)}$ in $[0, M_F]$. Finally, it should be noted that, even though to a lesser extent and for reasons related to maximization of (26) under censoring, medium to heavy censoring is also obvious on the performance of $\tilde{F}$.

6. Conclusions and future work

This research investigated the local polynomial smoothing of the Kaplan–Meier and showed that it leads to an effective and reliable way to estimate the c.d.f., its derivatives and auxiliary functionals for right censored data in the fixed design setting. The theoretical properties of all estimates and bandwidth selectors introduced herein

Table 2: Approximate MISE's of $\hat{f}_L, \hat{f}$ and $\tilde{f}$ for the truncated normal mixtures and the Birnbaum–Saunders distribution.

		NM1			NM2			BS		
Cens.	n	$\hat{f}_L$	$\hat{f}$	$\tilde{f}$	$\hat{f}_L$	$\hat{f}$	$\tilde{f}$	$\hat{f}_L$	$\hat{f}$	$\tilde{f}$
0%	50	0.098	0.135	0.077	0.083	0.078	0.101	0.256	0.511	0.191
	100	0.073	0.084	0.044	0.039	0.039	0.111	0.151	0.371	0.111
	150	0.039	0.068	0.035	0.027	0.029	0.029	0.113	0.311	0.082
	250	0.027	0.051	0.028	0.018	0.021	0.017	0.083	0.242	0.055
15%	50	0.116	0.165	0.102	0.096	0.096	0.112	0.328	0.513	0.252
	100	0.058	0.096	0.054	0.044	0.048	0.047	0.175	0.373	0.131
	150	0.043	0.074	0.042	0.031	0.035	0.033	0.125	0.301	0.101
	250	0.031	0.054	0.033	0.021	0.026	0.019	0.084	0.231	0.062
30%	50	0.156	0.232	0.152	0.128	0.135	0.131	0.256	0.602	0.247
	100	0.097	0.149	0.096	0.078	0.083	0.083	0.281	0.412	0.211
	150	0.066	0.106	0.071	0.053	0.059	0.053	0.176	0.312	0.131
	250	0.042	0.069	0.051	0.033	0.039	0.032	0.116	0.253	0.101
50%	50	0.451	0.512	0.488	0.433	0.461	0.384	0.442	0.487	0.436
	100	0.311	0.388	0.463	0.397	0.432	0.359	0.354	0.412	0.411
	150	0.251	0.351	0.453	0.382	0.418	0.316	0.316	0.385	0.385
	250	0.184	0.337	0.445	0.382	0.413	0.316	0.282	0.375	0.381

suggest a robust asymptotic behavior throughout the region of estimation. The MISE simulations indicate that this robust behavior is valid also for finite samples in the interval from the left endpoint up to at least the largest uncensored observation. The same simulations, in combination with Remark 2 indicate that two limitations are the impact of large amounts of censoring on the estimate's precision as well as possibly a diminished estimate performance at the right end point.

The methodological advances explored herein can be extended towards multiple directions. An natural step forward is calibration of the estimates and associated bandwidth rule towards incorporating covariate information. Another extension, especially useful for practitioners, is the development of statistical inference for goodness-of-fit hypothesis testing and confidence intervals for e.g. assessing the validity of parametric estimates. While research in both directions is already underway in ongoing work, multiple other topics arise by considering the adjustment and analytical study of this technique to different censoring schemes.

7. Appendix

First recall that for $0 < p < 1/2$, by (8),

$$\frac{1}{n}\sum_{j=1}^{g}\sum_{j=1}^{n}\frac{c_{ij}}{1-\hat{H}(x_j)}W_{\mu}^{*}\left(\frac{x_i - x}{a}\right) = \frac{1}{n}\sum_{j=1}^{g}\sum_{j=1}^{n}\frac{c_{ij}}{1-H(x_j)}W_{\mu}^{*}\left(\frac{x_i - x_j}{a}\right)(1+o(n^{-p})).$$

7.1. Auxiliary lemmas

Lemma 1. *Assume that F_T is twice differentiable, continuous and that $b = o(h)$. Then, as $n \to \infty$, for $i \neq j \neq k \neq l \in \{1, \ldots, g\}$*

$$\begin{aligned}
\mathrm{E}(c_i) &= b f_T(x_i)(1-H(x_i))(1+o(b)),\\
\mathrm{E}(c_j^2) &= \frac{b f_T(x_j)(1-H(x_j)) + (n-1)b^2 f_T^2(x_j)(1-H(x_j))^2}{n}(1+o(1)),\\
\mathrm{E}(c_j c_k) &= b^2 f_T(x_j) f_T(x_k)(1-H(x_j))(1-H(x_k))(1+o(b)),\\
\mathrm{E}(c_i^2 c_j^2) &= \frac{1}{n^4}\left\{n(n-1)b^2 + \frac{n!b^3}{(n-3)!}f_T(x_i)(1-H(x_i)) + \frac{n!b^3}{(n-3)!}f_T(x_j)(1-H(x_j))\right.\\
&\quad \left. + \frac{n!b^4}{(n-4)!}f_T(x_i)(1-H(x_i))f_T(x_j)(1-H(x_j))\right\}\\
&\quad \times f_T(x_i)(1-H(x_i))f_T(x_j)(1-H(x_j))(1+o(b)),\\
\mathrm{E}(c_i^2 c_j c_k) &= \frac{1}{n^4}\left\{\frac{n!b^3}{(n-3)!} + \frac{n!b^4}{(n-4)!}f_T(x_i)(1-H(x_i))\right\}\\
&\quad \times f_T(x_i)(1-H(x_i))f_T(x_j)(1-H(x_j))f_T(x_k)(1-H(x_k))(1+o(b)),\\
\mathrm{E}(c_i c_j c_k c_l) &= \frac{1}{n^4}\frac{n!}{(n-4)!}b^4 f_T(x_i)f_T(x_j)f_T(x_k)f_T(x_l)\\
&\quad \times (1-H(x_i))(1-H(x_j))(1-H(x_k))(1-H(x_l))(1+o(b)),\\
\mathrm{E}(c_i^4) &= \frac{1}{n^4}\left\{nbf_T(x_i)(1-H(x_i)) + 7n(n-1)b^2 f_T^2(x_i)(1-H(x_i))^2\right.\\
&\quad \left. + \frac{6n!b^3}{(n-3)!}(f_T(x_i)(1-H(x_i)))^3 + \frac{n!b^4}{(n-4)!}(f_T(x_i)(1-H(x_i)))^4\right\}(1+o(b)).
\end{aligned}$$

Proof. Only the last equation is proved here as the others are proved by straightforward calculus in an entirely similar manner. Using Lemma 1 in Ioannides and Bagkavos (2019) and noting that the same sample point cannot be in two distinct intervals say $\mathbf{1}_i$ and I_j which implies that $\mathrm{E}(c_{ir}c_{jr}) = 0$, then

$$\begin{aligned}
\mathrm{E}(c_i^4) = \mathrm{E}\left(\frac{1}{n}\sum_{r=1}^{n} c_{ir}\right)^4 &= \frac{1}{n^4}\sum_{r=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{s=1}^{n}\mathrm{E}(c_{ir}c_{il}c_{im}c_{is})\\
&= \frac{1}{n^4}\left\{\sum_{r=1}^{n}\mathrm{E}(c_{ir}^4) + 3\sum\sum_{r\neq l}\mathrm{E}(c_{ir}c_{il})^2 + 4\sum\sum_{r\neq l}\mathrm{E}(c_{ir}^3 c_{il})\right\}\\
&\overset{r \text{ fixed}}{=} \frac{1}{n^4}\left\{n\mathrm{E}(c_{ir}) + 7n(n-1)(\mathrm{E}c_{ir})^2 + 6n(n-1)(n-2)(\mathrm{E}c_{ir})^3 + \frac{n!}{(n-4)!}(\mathrm{E}c_{ir})^4\right\},
\end{aligned}$$

from which the last result immediately follows. □

An important consequence of Lemma 1, used throughout this section is

$$\mathrm{E}\left\{\frac{c_j^2}{(1-H(x_j))^2}\right\} = \left\{\frac{b}{n}\frac{f_T(x_j)}{1-H(x_j)} + \frac{(n-1)b^2 f_T^2(x_j)}{n}\right\}(1+o(1)), \tag{27}$$

$$\mathrm{E}\left\{\frac{c_j}{1-H(x_j)}\frac{c_k}{1-H(x_k)}\right\} = b^2 f_T(x_j) f_T(x_k)(1+o(b)), \tag{28}$$

$$\begin{aligned}\mathrm{E}\left\{\frac{c_i^2 c_j^2}{(1-H(x_i)^2(1-H(x_j)^2}\right\} &= \Big\{n(n-1)b^2\frac{f_T(x_i)f_T(x_j)}{(1-H(x_i))(1-H(x_j))} \\ &\quad + \frac{n!b^3}{(n-3)!}\frac{f_T^2(x_i)f_T(x_j)}{1-H(x_j)} + \frac{n!b^3}{(n-3)!}\frac{f_T(x_i)f_T^2(x_j)}{1-H(x_i)} \\ &\quad + \frac{n!b^4}{(n-4)!}f_T^2(x_i)f_T^2(x_j)\Big\}(1+o(b)),\end{aligned} \tag{29}$$

$$\begin{aligned}\mathrm{E}\left\{\frac{c_i^4}{(1-H(x_i))^4}\right\} &= \Big\{\frac{nbf_T(x_i)}{(1-H(x_i))^3} + \frac{7n(n-1)b^2 f_T^2(x_i)}{(1-H(x_i))^2} \\ &\quad + \frac{6n!b^3}{(n-3)!}\frac{f_T(x_i)}{1-H(x_i)} + \frac{n!b^4}{(n-4)!}f_T(x_i)\Big\}(1+o(b)).\end{aligned} \tag{30}$$

Further,

$$\begin{aligned}&\mathrm{E}\left\{\frac{c_i^2 c_j c_k}{(1-H(x_i))^2(1-H(x_j))(1-H(x_k))}\right\} = \\ &\left\{\frac{n!b^3}{(n-3)!}\frac{f_T(x_i)f_T(x_j)f_T(x_k)}{1-H(x_i)} + \frac{n!b^4}{(n-4)!}f_T^2(x_i)f_T(x_i)f_T(x_j)f_T(x_k)(1+o(b))\right\}(1+o(b)),\end{aligned} \tag{31}$$

$$\mathrm{E}\left\{\frac{c_i c_j c_k c_l}{(1-H(x_i))(1-H(x_j))(1-H(x_k))(1-H(x_l))}\right\} = \frac{n!b^4}{(n-4)!}f_T(x_i)f_T(x_j)f_T(x_k)f_T(x_l)(1+o(b)). \tag{32}$$

Define

$$\omega(t,u) = \int \frac{u-s}{a} K_\rho^n\left(\frac{u-s}{a}\right) W_\nu^n\left(\frac{t-s}{a}\right) ds.$$

Also let $C_\rho^*(r) = rK_\rho^n(r)$.

Lemma 2. *Under assumptions A.1–A.2,*

$$\int \omega(t,t)g(t)\,dt = ag(s)\int uK_\rho^n(u)W_\nu^n(u)\,du(1+O(a)), \tag{33}$$

$$\int \omega(t,u)g(t)\,dt = -\frac{a^{\nu+\rho-1}}{\nu!(\rho-1)!}G^{(\rho-1)}(u)(1+O(a^{\rho-1})), \tag{34}$$

$$\iint \omega(t,u)f_T(t)f_T(u)\,dt\,du = -\frac{a^{\nu+\rho}}{(\rho-1)!\nu!}\int F_T^{(\rho)}(z)F_T^{(\nu)}(z)\,dz(1+O(a^{\max(\rho,\nu)})), \tag{35}$$

$$\iint \frac{f_T(t)f_T(u)}{(1-H(t))(1-H(u))}\omega^2(t,u)\,dt\,du = a^3 R(g)R(C_\rho^n * W_\nu^n) + O(a^4), \tag{36}$$

$$\begin{aligned}\iint \frac{f_T(t)f_T(u)}{(1-H(t))(1-H(u))}\omega(t,t)\omega(t,u)\,dt\,du &= -\frac{a^{\nu+\rho}}{\nu!(\rho-1)!}\left\{\int g(u)g^{(\rho-2)}(u)\,du\right\} \\ &\quad\times\left\{\int rK_\rho^n(r)W_\nu^n(r)\,dr\right\}(1+O(a^\rho)),\end{aligned} \tag{37}$$

$$\iiint \frac{f_T(t)f_T(u)f_T(v)}{1-H(t)}\omega(t,u)\omega(t,v)\,dt\,du\,dv = \frac{a^{2(\nu+\rho)}}{(\nu!(\rho-1)!)^2}\int g(t)\left(f_T^{(\rho-2)}(t)\right)^2 dt(1+O(a^{2\nu})), \tag{38}$$

$$\int \frac{f_T(t)}{(1-H(t))^3}\omega^2(t,t)\,dt = a^2\left\{\int \frac{f_T(t)}{(1-H(t))^3}\,dt\right\}\left\{\int rK_\rho^n(r)W_\nu^n(r)\,dr\right\}^2(1+o(1)). \tag{39}$$

Proof. First, note that since W_ν^n is a distribution function $W_\nu^n(-\infty) = 0$ and $W_\nu^n(\infty) = 1$. Also by the moment conditions of K_ρ^n,

$$\int uK_\rho^n(u)\,du = 0 \text{ unless } \rho = 1.$$

Then,

$$\int \frac{u-s}{a} K_\rho^n\left(\frac{u-s}{a}\right)\left[W_\nu^n(v)G(av+s)\right]_{-\infty}^{+\infty} =$$
$$G(s)\int \frac{u-s}{a} K_\rho^n\left(\frac{u-s}{a}\right) W_\nu^n(\infty)\,ds - G(s)\int \frac{u-s}{a} K_\rho^n\left(\frac{u-s}{a}\right) W_\nu^n(-\infty)\,ds$$
$$= G(s)\int \frac{u-s}{a} K_\rho^n\left(\frac{u-s}{a}\right)\,ds = 0 \text{ unless } \rho = 1. \quad (40)$$

Starting with (33)

$$\begin{aligned}\int \omega(t,t)g(t)\,dt &= \iint \frac{t-s}{a} K_\rho^n\left(\frac{t-s}{a}\right) W_\nu^n\left(\frac{t-s}{a}\right) g(t)\,dt\,ds \\ &\overset{t-s=ua}{=} a\iint uK_\rho^n(u)W_\nu^n(u)g(s+ua)\,du\,ds \\ &= ag(s)\int uK_\rho^n(u)W_\nu^n(u)\,du(1+O(a)),\end{aligned}$$

after expanding $g(s+ua)$ in Taylor series around s and using the Lipschitz continuity of g. The proof of (34)–(39) is entirely similar (albeit longer) and therefore is omitted. Full details are available from the authors. □

Now, define

$$A_{\gamma,\mu}(a) = -3b^3\frac{\mu!\gamma!}{a^{\gamma+5}}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} K_\mu^n\left(\frac{x_k - x_i}{a}\right) K_\gamma^n\left(\frac{x_k - x_j}{a}\right)\hat{F}_T(x_i)\hat{F}_T(x_j),$$
$$B_{\gamma,\mu}(a) = -b^3\frac{\mu!\gamma!}{a^{\gamma+5}}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\left(\frac{x_k - x_i}{a}\right) K_\mu^{n\prime}\left(\frac{x_k - x_i}{a}\right) K_\gamma^n\left(\frac{x_k - x_j}{a}\right)\hat{F}_T(x_i)\hat{F}_T(x_j).$$

Lemma 3. *Assuming that K is $\max(\gamma,\mu)$ times differentiable. Provided $b = o(a)$ and that $ba^{-\max(\gamma,\mu)} \to 0$ as $n \to \infty$, then*

$$\frac{d}{da}\hat{\theta}_{\mu,\gamma}(a) = \left(A_{\gamma,\mu}(a) + B_{\gamma,\mu}(a)\right)(1 + o_p(1)).$$

Proof. First, recall that

$$\hat{\theta}_{\mu,\gamma}(a) = b\mu!\gamma!\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} K_\mu\left(\frac{x_k - x_i}{a}\right) K_\gamma\left(\frac{x_k - x_j}{a}\right)\hat{F}_T(x_i)\hat{F}_T(x_j)(1 + o(b)).$$

Thus,

$$\begin{aligned}\frac{d}{da}\hat{\theta}_{\mu,\gamma}(a) = b\mu!\gamma!\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\Bigg[&\left\{\frac{d}{da}K_\mu\left(\frac{x_k - x_i}{a}\right)\right\}\left\{K_\gamma\left(\frac{x_k - x_j}{a}\right)\right\} \\ &+ \left\{K_\mu\left(\frac{x_k - x_i}{a}\right)\right\}\left\{\frac{d}{da}K_\gamma\left(\frac{x_k - x_j}{a}\right)\right\}\Bigg]\hat{F}_T(x_i)\hat{F}_T(x_j)(1 + o_p(1)). \quad (41)\end{aligned}$$

The central concept of the proof is calculation of $\frac{d}{da}K_\nu(\cdot)$. Recall the definition of K_ν

$$K_\nu(u) = e_{\nu+1}^T S^{-1}(1, hu, \ldots, (hu)^\nu, (hu)^{\nu+1})^T K(u),$$

and S is the $(\nu+2)\times(\nu+2)$ matrix $(S_{n,j+l})_{0\le j,l\le \nu+1}$ with

$$S_{n,l}(x) = \sum_{i=1}^{g} K\left(\frac{x_i - x}{a}\right)(x_i - x)^l, \quad l = 0, 1, \ldots, 2\nu+2.$$

Since $S_{n,l}(x) \equiv S_{n,l}$ depends on a we need to also calculate the derivative of the matrix S^{-1}. For this, from standard linear algebra we know that for a $(\nu+2)\times(\nu+2)$ matrix A given by

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1\ \nu+2} \\ a_{21} & a_{22} & \ldots & a_{2\ \nu+2} \\ \ldots & \ldots & \ldots & \ldots \\ a_{\nu+2\ 1} & a_{n2} & \ldots & a_{\nu+2\ \nu+2} \end{pmatrix},$$

its determinant is given by

$$\det(A) = \sum_{\sigma\in s_{\nu+2}} \varepsilon(\sigma) a_{\sigma(1),1} a_{\sigma(2),2} \ldots a_{\sigma(\nu+1),\nu+1} a_{\sigma(\nu+2),\nu+2}, \tag{42}$$

where for every permutation $\sigma(1), \ldots, \sigma(\nu+2)$ of $s_{\nu+2} = \{1, 2, \ldots, \nu+2\}$, the products $\varepsilon(\sigma)\cdot a_{\sigma(1),1}\cdot a_{\sigma(2),2}\cdot \cdots \cdot a_{\sigma(\nu+2),\nu+2}$ across the set $s_{\nu+2}$ in (42) result by multiplying $a_{\sigma(1),1}$, which is in the $\sigma(1)$ row and 1st column of A, with $a_{\sigma(2),2}$, which is in the $\sigma(2)$ row and 2nd column of A, ..., with $a_{\sigma(\nu+2),\nu+2}$ in the $\sigma(\nu+2)$ row and $(\nu+2)$th column of A. $\varepsilon(\sigma)$ is -1 or $+1$ depending on whether σ is odd or even respectively. Let $l \in \{l_1, \ldots, l_{\nu+2}\}$. Applying this to the matrix $e_{\nu+1}^T S^{-1}$ yields

$$e_{\nu+1}^T S^{-1} = \frac{1}{\det(S)}\left(A_S^{(1)}, \ldots, A_S^{(\nu+1)}, A_S^{(\nu+2)}\right),$$

where for any matrix M,

$$A_M^{(i)} = \ldots, i = 1, \ldots, \nu+2.$$

Note that by (42),

$$\det(M) = \sum_{i=1}^{\nu+2} M_{i+1} A_M^{(i)}. \tag{43}$$

From (4) $bh^{-(l+1)} S_{n,l} = \mu_l + o(1), l = 0, 1, \ldots, 2\nu+2$ and thus

$$S^{-1} = \frac{b}{a^{\nu+1}} \hat{S}^{-1} + o(bh^{-(\nu+1)}),$$

or equivalently, by setting $K_\nu^n(u) = e_{\nu+1}^T \hat{S}^{-1}(1, hu, \ldots, (hu)^\nu, (hu)^{\nu+1})^T K(u)$,

$$K_\nu(u) = ba^{-(\nu+1)} K_\nu^n(u) + o(bh^{-(\nu+1)}).$$

From (4) and (5), for $b = o(a)$ (see also the proof of Lemma 5 in Cheng (1994))

$$\frac{d}{da}\frac{b}{a^l} S_{n,l} = \mu_l + o(1), l = 0, \ldots, 2\gamma+2. \tag{44}$$

By (43),

$$\frac{b^{\nu+2}}{a^{l_1+\cdots+l_{\nu+2}}}\det(S) = \frac{b^{\nu+2}}{a^{l_1+\cdots+l_{\nu+2}}} \sum_{l\in s_{\nu+2}} (-1)^{p(l)} S_{n,l_1} S_{n,l_2} \ldots S_{n,l_{\nu+2}}.$$

Thus,

$$\frac{d}{da}\left(\frac{b^{\nu+2}}{a^{l_1+\cdots+l_{\nu+2}}}\det(S)\right) = \sum_{l\in s_{\nu+2}}(-1)^{p(l)}\frac{d}{da}\left\{\left(\frac{b}{a^{l_1}}S_{n,l_1}\right)\left(\frac{b}{a^{l_2}}S_{n,l_2}\right)\cdots\left(\frac{b}{a^{l_{\gamma+2}}}S_{n,l_{\nu+2}}\right)\right\}$$

$$= \sum_{l\in s_{\nu+2}}(-1)^{p(l)}\left\{\left(\frac{d}{da}\frac{b}{a^{l_1}}S_{n,l_1}\right)\left(\frac{b}{a^{l_2}}S_{n,l_2}\right)\cdots\left(\frac{b}{a^{l_{\nu+2}}}S_{n,l_{\nu+2}}\right)\right.$$

$$+\left(\frac{b}{a^{l_1}}S_{n,l_1}\right)\left(\frac{d}{da}\frac{b}{a^{l_2}}S_{n,l_2}\right)\cdots\left(\frac{b}{a^{l_{\nu+2}}}S_{n,l_{\nu+2}}\right)+\cdots$$

$$\left.+\left(\frac{b}{a^{l_1}}S_{n,l_1}\right)\left(\frac{b}{a^{l_2}}S_{n,l_2}\right)\cdots\left(\frac{d}{da}\frac{b}{a^{l_{\nu+2}}}S_{n,l_{\nu+2}}\right)\right\}.$$

By (4), for any $k = 1,\dots,\nu+2$, $ba^{-l_k}S_{n.l_k} = a\mu_{l_k}(1+o(1))$ and thus

$$\frac{b^{\nu+2}}{a^{l_1+\cdots+l_{\nu+2}}}\det(S) = a^{\nu+2}\sum_{l\in s_{\nu+2}}(-1)^{p(l)}\mu_{l_1}\mu_{l_2}\dots\mu_{l_{\nu+2}}(1+o(1))$$

$$= a^{\nu+2}\det(\hat{S})(1+o(1)). \tag{45}$$

Differentiating (45) and noticing that $\det(\hat{S})$ no longer depends on a yields,

$$\frac{d}{da}\left(\frac{b^{\nu+2}}{a^{l_1+\cdots+l_{\nu+2}}}\det(S)\right) = (\nu+2)a^{\nu+1}\det(\hat{S})(1+o(1)). \tag{46}$$

Similarly, using (4) yields

$$\frac{b^{\nu+1}}{a^{l_1+\cdots+l_{\nu+2}-l}}A_S^{(l)} = a^{\nu+1}A_{\hat{S}}^{(l)}(1+o(1)), l = 1,\dots,\nu+2. \tag{47}$$

Differentiating (47) and noticing that $A_{\hat{S}}^{(l)}$ no longer depends on a yields,

$$\frac{d}{da}\left(\frac{b^{\nu+1}}{a^{l_1+\cdots+l_{\nu+2}-l}}A_S^{(l)}\right) = (\nu+1)a^{\nu}A_{\hat{S}}^{(l)}(1+o(1)). \tag{48}$$

Now, combine (45)–(48) to get

$$\frac{d}{da}\frac{a^{1+l}A_S^{(l)}}{b\det(S)} = \frac{d}{da}\frac{\frac{b^{\nu+1}}{a^{l_1+\cdots+l_{\nu+2}-1-l}}A_S^{(l)}}{\frac{b^{\nu+2}}{a^{l_1+\cdots+l_{\nu+2}}}\det(S)}$$

$$= \frac{\left(\frac{d}{da}\frac{b^{\nu+1}}{a^{l_1+\cdots+l_{\nu+2}-1-l}}A_S^{(l)}\right)\left(\frac{b^{\nu+2}}{a^{l_1+\cdots+l_{\nu+2}}}\det(S)\right)}{\left(\frac{b^{\nu+2}}{a^{l_1+\cdots+l_{\nu+2}}}\det(S)\right)^2}$$

$$= -\frac{A_{\hat{S}}^{(l)}}{a^2\det(\hat{S})}(1+o(1)), l = 1,2,\dots,\nu+2. \tag{49}$$

Also, from (45) and (47) we conclude that for $l = 1,2,\dots,\nu+2$

$$\frac{a^{1+l}A_S^{(l)}}{b\det(S)} = \frac{a^{1+l}}{b}\frac{\frac{a^{l_1+\cdots+l_{\nu+2}-1-l}}{b^{\nu+1}}(\nu+1)a^{\nu}A_{\hat{S}}^{(l)}(1+o(1))}{\frac{a^{l_1+\cdots+l_{\nu+2}}}{b^{\nu+2}}a^{\nu+2}\det(\hat{S})(1+o(1))} = \frac{A_{\hat{S}}^{(l)}}{a\det(\hat{S})}(1+o(1)). \tag{50}$$

Thus

$$\frac{d}{da}K_\nu\left(\frac{x_k-x_j}{a}\right) = \frac{d}{da}e_{\nu+1}^T S^{-1}(1,au,\dots,(au)^\nu,(au)^{\nu+1})^T K(u)$$

$$
\begin{aligned}
&\overset{(49),(50)}{=} \sum_{l=1}^{\nu+2} \Bigg[\left(-\frac{A_{\hat{S}}^{(l)}}{a^2 \det(\hat{S})} (1+o(1)) \right) \left\{ \frac{b}{a^2} K\left(\frac{x_k - x_i}{a} \right) \left(\frac{x_k - x_i}{a} \right)^{l-1} \right\} \\
&\quad + \left(\frac{A_{\hat{S}}^{(l)}}{a \det(\hat{S})} (1+o(1)) \right) \frac{d}{da} \left\{ \frac{b}{a^2} K\left(\frac{x_k - x_i}{a} \right) \left(\frac{x_k - x_i}{a} \right)^{l-1} \right\} \Bigg] \\
&= \sum_{l=1}^{\nu+2} \frac{A_{\hat{S}}^{(l)}}{\det(\hat{S})} \Bigg[-\frac{3b}{a^4} K\left(\frac{x_k - x_i}{a} \right) \left(\frac{x_k - x_i}{a} \right)^{l-1} \\
&\quad + \frac{b}{a^3} \frac{d}{da} K\left(\frac{x_k - x_i}{a} \right) \left(\frac{x_k - x_i}{a} \right)^{l-1} \Bigg] (1+o(1)). \qquad (51)
\end{aligned}
$$

Use (51) back to (41) in the second step below together with the assumption $ba^{-\max(\gamma,\mu)} \to 0$ as $n \to \infty$ to obtain

$$
\begin{aligned}
\frac{d}{da}\hat{\theta}_{\mu,\gamma}(a) &= b\mu!\gamma! \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} \Bigg[\left\{ \frac{d}{da} K_\mu\left(\frac{x_k - x_i}{a} \right) \right\} \left\{ K_\gamma\left(\frac{x_k - x_j}{a} \right) \right\} \\
&\quad + \left\{ K_\mu\left(\frac{x_k - x_i}{a} \right) \right\} \left\{ \frac{d}{da} K_\gamma\left(\frac{x_k - x_j}{a} \right) \right\} \Bigg] \hat{F}_T(x_i)\hat{F}_T(x_j)(1+o(b)) \\
&= b\mu!\gamma! \sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g} \Bigg\{ \Bigg(\sum_{l=1}^{\mu+2} \frac{A_{\hat{S}}^{(l)}}{\det(\hat{S})} \Bigg[-\frac{3b}{a^4} K\left(\frac{x_k - x_i}{a} \right) \left(\frac{x_k - x_i}{a} \right)^{l-1} \\
&\quad + \frac{b}{a^3} \frac{d}{da} K\left(\frac{x_k - x_i}{a} \right) \left(\frac{x_k - x_i}{a} \right)^{l-1} \Bigg] \Bigg) \frac{b}{a^{\gamma+1}} K_\gamma^n\left(\frac{x_k - x_j}{a} \right) \\
&\quad + \Bigg(\sum_{l=1}^{\gamma+2} \frac{A_{\hat{S}}^{(l)}}{\det(\hat{S})} \Bigg[-\frac{3b}{a^4} K\left(\frac{x_k - x_j}{a} \right) \left(\frac{x_k - x_j}{a} \right)^{l-1} \\
&\quad + \frac{b}{a^3} \frac{d}{da} K\left(\frac{x_k - x_j}{a} \right) \left(\frac{x_k - x_j}{a} \right)^{l-1} \Bigg] \Bigg) \frac{b}{a^{\mu+1}} K_\mu^n\left(\frac{x_k - x_i}{a} \right) \Bigg\} \\
&\quad \times \hat{F}_T(x_i)\hat{F}_T(x_j)(1+o(1)). \qquad (52)
\end{aligned}
$$

From the definition of the equivalent kernel (see also (5)),

$$
K_\nu^n\left(\frac{x_k - x_i}{a} \right) = \sum_{l=1}^{\nu+2} \frac{A_{\hat{S}}^{(l)}}{\det(\hat{S})} K\left(\frac{x_k - x_i}{a} \right) \left(\frac{x_k - x_i}{a} \right)^{l-1} (1+o(1)), \qquad (53)
$$

$$
\frac{1}{a}\left(\frac{x_k - x_i}{a} \right) K_\nu^{n\prime}\left(\frac{x_k - x_i}{a} \right) = \frac{d}{da} \sum_{l=1}^{\nu+2} \frac{A_{\hat{S}}^{(l)}}{\det(\hat{S})} K\left(\frac{x_k - x_i}{a} \right) \left(\frac{x_k - x_i}{a} \right)^{l-1} (1+o(1)). \qquad (54)
$$

Using (53) and (54) back to (52) and by straightforward calculations completes the proof. □

Lemma 4. *Assume that K has compact support, it vanishes at the endpoints, is symmetric about its origin and its first $\mu+2$ derivatives exist. Then, as $n \to \infty, h \to 0$ and $n^{\mu+\gamma+1} \to \infty$,*

$$
\begin{aligned}
\mathrm{E}\{B_{\nu,\rho}(a)\} &= \left\{ \rho a^{\rho-5} \int F_T^{(\rho)}(z) F_T^{(\nu)}(z)\, dz - \frac{1}{n}\frac{\nu!\rho!}{a^{\nu+4}} g(s) \int u K_\rho^n(u) W_\nu^n(u)\, du \right\} \left(1+o(n^{-p})\right), \\
\mathrm{Var}\{B_{\nu,\rho}(a)\} &= \frac{2(\nu!\rho!)^2}{n^2 a^{2(\nu+3)}} aR(g)R(C_\rho^n * W_\nu^n) + o(n^{-1}a^{-2}),
\end{aligned}
$$

with $C_\nu^*(x) = x(W_\nu^*)'(x)$.

Proof. By the definition of $B_{\nu,\rho}(a)$,

$$
\begin{aligned}
B_{\nu,\rho}(a) &= -\frac{b^3\nu!\rho!}{a^{\nu+5}}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\left(\frac{x_k-x_i}{a}\right)K_\rho^{n\prime}\left(\frac{x_k-x_i}{a}\right)K_\nu^{n}\left(\frac{x_k-x_j}{a}\right)\hat{F}_T(x_i)\hat{F}_T(x_j)\\
&= -\frac{b^3\nu!\rho!}{a^{\nu+5}}\Bigg\{\sum_{i=1}^{g}\sum_{k=1}^{g}\left(\frac{x_k-x_i}{a}\right)K_\rho^{n}\left(\frac{x_k-x_i}{a}\right)W_\nu^{n}\left(\frac{x_k-x_i}{a}\right)\frac{c_i^2}{(1-\hat{H}(x_i))^2}\\
&\quad+\sum_{\substack{i=1\\ i\neq j}}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\left(\frac{x_k-x_i}{a}\right)K_\rho^{*}\left(\frac{x_k-x_i}{a}\right)W_\nu^{*}\left(\frac{x_k-x_j}{a}\right)\frac{c_i}{1-\hat{H}(x_i)}\frac{c_j}{1-\hat{H}(x_j)}\Bigg\}(1+o(1)). \qquad (55)
\end{aligned}
$$

Now, using (8), together with (27) and (28) yields

$$
\begin{aligned}
\mathrm{E}(B_{\nu,\rho}(a)) &= -\frac{b^3\nu!\rho!}{a^{\nu+5}}\Bigg\{\sum_{i=1}^{g}\sum_{k=1}^{g}\left(\frac{x_k-x_i}{a}\right)K_\rho^{n}\left(\frac{x_k-x_i}{a}\right)W_\nu^{n}\left(\frac{x_k-x_j}{a}\right)\frac{1}{nb}\frac{f_T(x_i)}{1-H(x_i)}\\
&\quad+\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\left(\frac{x_k-x_i}{a}\right)K_\rho^{n}\left(\frac{x_k-x_i}{a}\right)W_\nu^{n}\left(\frac{x_k-x_j}{a}\right)f_T(x_i)f_T(x_j)\Bigg\}\left(1+o(n^{-p})\right). \qquad (56)
\end{aligned}
$$

Now, the two sums can be approximated as

$$
\begin{aligned}
&b^2\sum_{i=1}^{g}\sum_{k=1}^{g}\left(\frac{x_k-x_i}{a}\right)K_\rho^{n}\left(\frac{x_k-x_i}{a}\right)W_\nu^{n}\left(\frac{x_k-x_j}{a}\right)\frac{f_T(x_i)}{1-H(x_i)}\\
&\qquad=\iint\left(\frac{x-y}{a}\right)K_\rho^{*}\left(\frac{x-y}{a}\right)W_\nu^{*}\left(\frac{x-y}{a}\right)\frac{f_T(x)}{1-H(x)}\,dx\,dy+o(b), \qquad (57)
\end{aligned}
$$

and

$$
\begin{aligned}
&b^3\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\left(\frac{x_k-x_i}{a}\right)K_\rho^{n}\left(\frac{x_k-x_i}{a}\right)W_\nu^{n}\left(\frac{x_k-x_j}{a}\right)f_T(x_i)f_T(x_j)\\
&\qquad\simeq\iiint\left(\frac{x-z}{a}\right)K_\rho^{n}\left(\frac{x-z}{a}\right)W_\nu^{*}\left(\frac{y-z}{a}\right)f_T(x)f_T(y)\,dx\,dy\,dz+o(b^2). \qquad (58)
\end{aligned}
$$

Use (57) and (58) back to (56) to obtain

$$
\begin{aligned}
\mathrm{E}(B_{\nu,\rho}(a)) &= -\frac{\nu!\rho!}{a^{\nu+5}}\Bigg\{\iint\frac{x-y}{a}K_\rho^{n}\left(\frac{x-y}{a}\right)W_\nu^{n}\left(\frac{x-y}{a}\right)\frac{1}{n}\frac{f_T(x)}{1-H(x)}\,dx\,dy\\
&\quad+\iiint\frac{x-z}{a}K_\rho^{n}\left(\frac{x-z}{a}\right)W_\nu^{n}\left(\frac{y-z}{a}\right)f_T(x)f_T(y)\,dx\,dy\,dz\Bigg\}\left(1+o(n^{-p})\right)\\
&= -\frac{\nu!\rho!}{a^{\nu+5}}\underbrace{\left\{\frac{1}{n}\int\omega(x,x)g(x)\,dx+\iint\omega(x,y)f_T(x)f_T(y)\,dx\,dy\right\}}_{I}\left(1+o(n^{-p})\right)\\
&= -\frac{\nu!\rho!}{a^{\nu+5}}\left\{\frac{ag(s)}{n}\int uK_\rho^{n}(u)W_\nu^{n}(u)\,du-\frac{a^{\nu+\rho}}{(\rho-1)!\nu!}\int F_T^{(\rho)}(z)F_T^{(\nu)}(z)\,dz\right\}\left(1+o(n^{-p})\right),
\end{aligned}
$$

from which the result immediately follows. Regarding the variance, first set

$$
\pi(x_i,x_j)=\sum_{i=1}^{g}\frac{x_j-s}{a}K_\rho^{n}\left(\frac{x_j-s}{a}\right)W_\nu^{n}\left(\frac{x_j-s}{a}\right)b.
$$

Then,

$$
\mathrm{E}(B_{\nu,\rho}^2(a))=\mathrm{E}\Bigg\{\frac{b^3\nu!\rho!}{a^{\nu+5}}\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\left(\frac{x_k-x_i}{a}\right)K_\rho^{n}\left(\frac{x_k-x_i}{a}\right)W_\nu^{n}\left(\frac{x_k-x_j}{a}\right)\frac{c_i}{1-\hat{H}(x_i)}\frac{c_j}{1-\hat{H}(x_j)}\Bigg\}^2
$$

$$= \frac{b^2(\nu!\rho!)^2}{a^{2(\nu+5)}} \underbrace{\sum_{i=1}^{g}\sum_{j=1}^{g}\sum_{k=1}^{g}\sum_{l=1}^{g} \frac{\omega(x_i,x_j)\omega(x_k,x_l)\mathrm{E}\left(c_ic_jc_kc_l\right)}{(1-H(x_i))(1-H(x_j))(1-H(x_k))(1-H(x_l))}}_{II}(1+o(n^{-p})).$$

By (27)–(32) and the multinomial theorem,

$$\begin{aligned}
\mathrm{E}II = {} & \frac{n!}{(n-4)!}\left(\iint f_T(t)f_T(u)\omega(t,u)\,dt\,du\right)^2 \\
& + \frac{2n!}{(n-3)!}\left(\int g(t)\omega(t,t)\,dt\right)\left(\iint f_T(t)f_T(u)\omega(t,u)\,dt\,du\right) \\
& + \frac{4n!}{(n-3)!}\iiint \frac{f_T(t)f_T(u)f_T(v)}{1-H(t)}\omega(t,u)\omega(t,v)\,dt\,du\,dv \\
& + \frac{2n!}{(n-2)!}\iint \frac{f_T(t)f_T(u)}{(1-H(t))(1-H(u))}\left(\omega(t,u)^2+2\omega(t,t)\omega(t,u)\right)\,dt\,du \\
& + n\int \frac{f_T(t)}{(1-H(t))^3}\omega^2(t,t)\,dt.
\end{aligned}$$

Rearranging and using (33)–(39),

$$\begin{aligned}
\mathrm{E}II-(\mathrm{E}I)^2 = {} & \left[\frac{n!}{(n-4)!}-(n(n-1))^2\right]\frac{a^{2(\nu+\rho)}}{((\rho-1)!\nu!)^2}\left(\int F_T^{(\rho)}(z)F_T^{(\nu)}(z)\,dz\right)^2 \\
& - \left[\frac{2n!}{(n-3)!}-\frac{2n^2(n-1)}{n}\right]\frac{ag(s)a^{\nu+\rho}}{(\rho-1)!\nu!}\left(\int uK_\rho^n(u)W_\nu^n(u)\,du\right)\left(\int F_T^{(\rho)}(z)F_T^{(\nu)}(z)\,dz\right) \\
& + \frac{4n!}{(n-3)!}\frac{a^{2(\nu+\rho)}}{(\nu!(\rho-1)!)^2}\int g(t)\left(f_T^{(\rho-2)}(t)\right)^2\,dt + \frac{2n!}{(n-2)!}a^3R(g)R(C_\rho^n * W_\nu^n) \\
& - \frac{4n!}{(n-2)!}\frac{a^{\nu+\rho}}{\nu!(\rho-1)!}\left\{\int g(u)g^{(\rho-2)}(u)\,du\right\}\left\{\int rK_\rho^n(r)W_\nu^n(r)\,dr\right\} \\
& + na^2\left\{\int \frac{f_T(t)}{(1-H(t))^3}\,dt\right\}\left\{\int rK_\rho^n(r)W_\nu^n(r)\,dr\right\}^2 \\
& - n\left(\frac{ag(s)}{n}\int uK_\rho^n(u)W_\nu^n(u)\,du\right)^2 + O(n^2a^4)+o(n^{-1}a^{-2}).
\end{aligned}$$

Rearranging the above expression, multiplying by $(\nu!\rho!)^2a^{-2(\nu+5)}$ and noticing that the dominant term is $a^3R(g)R(C_\rho^n * W_\nu^n)$ yields the result. □

Lemma 5. *As $n\to\infty, h\to 0$ and $na^{\mu+\gamma+3}\to\infty$, and $b=o(a)$*

$$na^{\mu+\gamma-1}\left(\hat\theta_{\mu,\gamma}(a)-\theta_{\mu,\gamma}\right)\xrightarrow{d} N(\mu_*,\sigma_*^2),$$

where

$$\begin{aligned}
\mu_* &= \mu!\gamma!\left\{\int\left(\frac{f_T(u)}{1-H(u)}\right)du\right\}\int W_\mu^*K_\gamma^* + \frac{(1+\delta_{\mu\gamma})\gamma!}{(\gamma+2)!}na^{\mu+\gamma+1}\theta_{\mu,\gamma+2}\mu_{\gamma+2}(K_\gamma^*)+O(a), \\
\sigma_*^2 &= 2(\mu!\gamma!)^2a^{-1}R(g^{(\gamma)})R\left(W_\mu^*K_\gamma^*\right).
\end{aligned}$$

Proof. Set

$$\psi_{lk,\nu} = \sum_{i=1}^{g} W_\nu^n\left(\frac{x_i-x_k}{a}\right)\frac{c_{kl}}{1-\hat H(x_k)},$$

and

$$\mu_{k,\nu} = \mathrm{E}(\psi_{1k,\nu}).$$

Now, note that by Lemma 2 in Bagkavos and Patil (2008) in the third step below,

$$\mu_{k,\nu} = \mathrm{E}\sum_{i=1}^{g} W_\nu^n\left(\frac{x_i - x_k}{a}\right)\frac{c_{i1}}{1-\hat{H}(x_i)} = \sum_{i=1}^{g} W_\nu^n\left(\frac{x_i - x_k}{a}\right) b f_T(x_i)\, dt(1+o(bn^{-p}))$$
$$= \int W_\nu^n\left(\frac{t - x_k}{a}\right) f_T(t)\, dt(1+o(bn^{-p})). \quad (59)$$

By (6) of Bagkavos and Ioannides (2020), Lemma 2 of Bagkavos and Patil (2008) gives

$$\mathrm{E}(\psi_{1k,\mu}\psi_{1l,\gamma}) = \int W_\mu^n\left(\frac{t-x_k}{a}\right) W_\gamma^n\left(\frac{t-x_l}{a}\right)\frac{f(t)}{1-H(t)}\, dt(1+o(n^{-1})). \quad (60)$$

By the independence of X_1 and X_2 and using (7) of Bagkavos and Ioannides (2020),

$$\mathbb{E}(\psi_{1k,\mu}\psi_{2l,\gamma}) = \sum_{i=1}^{g} W_\mu^n\left(\frac{x_i-x_k}{a}\right) W_\gamma^n\left(\frac{x_i-x_l}{a}\right) b^2 f_T^2(x_i)(1+o(n^{-2p}))$$
$$+\sum\sum_{i\neq j} W_\mu^n\left(\frac{x_i-x_k}{a}\right) W_\gamma^n\left(\frac{x_j-x_l}{a}\right) b^2 f_T(x_i) f_T(x_j)(1+o(n^{-2p})). \quad (61)$$

Then,

$$\hat{\theta}_{\mu,\gamma}(a) = \frac{\mu!\gamma!}{(nb)^2}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{g} b\ (\psi_{lk,\mu}-\mu_{k,\gamma})(\psi_{mk,\gamma}-\mu_{k,\mu})$$
$$+\frac{\mu!\gamma!}{(nb)^2}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{g} b(\psi_{lk,\mu}\mu_{k,\mu}+\psi_{mk,\gamma}\mu_{k,\gamma}) - \frac{\mu!\gamma!}{(nb)^2}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{g} b\mu_{k,\mu}\mu_{k,\gamma}. \quad (62)$$

The next step is to prove

$$U_n = \sum\sum_{1\le l<m<n}\left\{b\sum_{k=1}^{g}(\psi_{lk,\mu}-\mu_{k,\gamma})(\psi_{mk,\gamma}-\mu_{k,\mu})\right\}\xrightarrow{d} N\left(0, \frac{n^2b^2}{2a^{2(\mu+\gamma)-1}}R(g)R(C_\mu^* * W_\gamma^*)\right), \quad (63)$$

$$\frac{\mu!\nu!}{(nb)^2}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{g} b\psi_{lk,\nu}\mu_{k,\nu}\to \mu!\nu!a^{-\nu}\times$$
$$\left\{\frac{a^\nu}{\nu!}R(F_T^{(\nu)}) + \frac{a^{\nu+2}}{(\nu+2)!}\int F_T^{(\nu)}F_T^{(\nu+2)}\int u^{\nu+2}K_\nu^*\right\}(1+o(n^{-p})), \quad (64)$$

and

$$\frac{\mu!\nu!}{(nb)^2}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{g} b\mu_{k,\mu}\mu_{k,\nu}\to\frac{1}{b^2}\left\{\theta_{\nu,\mu}+\frac{(1+\delta_{\nu\mu})a^2}{(\mu+1)(\mu+2)}\theta_{\nu,\mu+2}\mu_{\mu+2}(K_\mu^*)+o(a^2)\right\}. \quad (65)$$

Then, asymptotic normality of $\hat{\theta}_{\mu,\gamma}(a)$ will follow by using (63)–(65) in (62). Establishing (63)–(65) depends on repeated use of the following approximation. By (59),

$$\mu_{k,\nu}\mu_{l,\mu} = \left\{\int W_\nu^n\left(\frac{t-x_k}{a}\right) f_T(t)\, dt\right\}\left\{\int W_\mu^n\left(\frac{t-x_l}{a}\right) f_T(t)\, dt\right\}$$
$$= \frac{b^2}{a^{\nu+\mu}}\left\{\int W_\nu^*\left(\frac{t-x_k}{a}\right) f_T(t)\, dt\right\}\left\{\int W_\mu^*\left(\frac{t-x_l}{a}\right) f_T(t)\, dt\right\}(1+o(b^2a^{-(\nu+\mu)}))$$
$$= \frac{b^2}{a^{\nu+\mu}}\left\{\int K_\nu^*\left(\frac{t-x_k}{a}\right) F_T(t)\, dt\right\}\left\{\int K_\mu^*\left(\frac{t-x_l}{a}\right) F_T(t)\, dt\right\}(1+o(b^2a^{-(\nu+\mu)}))$$
$$= \frac{b^2}{a^{\nu+\mu}}a^2\left\{\int K_\nu^*(u)F_T(x_k+ua)\, du\right\}\left\{\int K_\mu^*(u)F_T(x_l+ua)\, du\right\}(1+o(b^2a^{-(\nu+\mu)}))$$

$$= \frac{b^2}{a^{\nu+\mu}} \left\{ \frac{a^{\nu+\mu}}{\mu!\nu!} F_T^{(\nu)}(x_k) F_T^{(\mu)}(x_l) + \frac{(1+\delta_{\nu\mu}) a^{\nu+\mu+2}}{\nu!(\mu+2)!} F_T^{(\nu)}(x_k) F_T^{(\mu+2)}(x_l) \mu_{\mu+2}(K_\mu^*) + o(a^{\nu+\mu}) \right\}$$

$$(1 + o(b^2 a^{-(\nu+\mu)})). \quad (66)$$

In showing (63), Theorem 1 of Hall (1984) is applied to U_n. Let

$$H_n(X_1, X_2) = b \sum_{k=1}^{g} (\psi_{1k,\mu} - \mu_{k,\gamma})(\psi_{2k,\gamma} - \mu_{k,\mu}),$$

$$H_n(X_1, x) = b \sum_{k=1}^{g} (\psi_{xk,\mu} - \mu_{k,\gamma})(\psi_{1k,\gamma} - \mu_{k,\mu}),$$

$$\psi_{xk,\nu} = (nb)^{-1} \sum_{i=1}^{g} W_\nu^n \left(\frac{x_i - x_k}{a} \right) \frac{I_{[x_k - \frac{b}{2}, x_k + \frac{b}{2}]}(x)}{1 - \hat{H}(x_k)}, \quad x \in \mathbb{R},$$

$$G_n(x, y) = \mathrm{E}\{H_n(X_1, x) H_n(X_1, y)\}$$

$$= b^2 \sum_{k=1}^{g} \sum_{l=1}^{g} (\psi_{xk,\mu} - \mu_{k,\gamma})(\psi_{yl,\mu} - \mu_{l,\gamma}) \mathrm{E}(\psi_{1k,\gamma} - \mu_{k,\mu})(\psi_{1l,\gamma} - \mu_{l,\mu}).$$

By definition, H_n is symmetric and $\mathrm{E}(H_n(X_1, X_2)|X_2) = 0$, thus

$$U_n = \sum\sum_{1 \le l < m \le n} H_n(X_l, X_m)$$

is a degenerate U-statistics. Proof of (63) will follow by application of Theorem 1 in Hall (1984), according to which $U_n \to N(0, \frac{1}{2} n^2 \mathrm{E} H_n^2(X_1, X_2))$. According to the theorem, (2.1) in Hall (1984) must be verified first. For this, first note that

$$\mathrm{E}[H_n^2(X_1, X_2)] = b^2 \mathrm{E} \left(\sum_{k=1}^{g} (\psi_{1k,\mu} - \mu_{k,\gamma})(\psi_{2k,\gamma} - \mu_{k,\mu}) \right)^2$$

$$= b^2 \mathrm{E} \sum_{k=1}^{g} \sum_{l=1}^{g} (\psi_{1k,\mu} - \mu_{k,\gamma})(\psi_{2k,\gamma} - \mu_{k,\mu})(\psi_{1l,\mu} - \mu_{l,\gamma})(\psi_{2l,\gamma} - \mu_{l,\mu})$$

$$= b^2 \mathrm{E} \sum_{k=1}^{g} \sum_{l=1}^{g} (\psi_{1k,\mu}\psi_{2k,\gamma} - \mu_{k,\mu}\psi_{1k,\mu} - \mu_{k,\gamma}\psi_{2k,\gamma} + \mu_{k,\gamma}\mu_{k,\mu})$$

$$\times (\psi_{1l,\mu}\psi_{2l,\gamma} - \mu_{l,\mu}\psi_{1l,\mu} - \mu_{l,\gamma}\psi_{2l,\gamma} + \mu_{l,\gamma}\mu_{l,\mu})$$

$$\simeq b^2 \sum_{k=1}^{g} \sum_{l=1}^{g} \{\mathrm{E}(\psi_{1k,\mu}\psi_{1k,\gamma} - \mu_{k,\mu}\psi_{1k,\mu} - \mu_{k,\gamma}\psi_{2k,\gamma} + \mu_{k,\gamma}\mu_{k,\mu})\}^2$$

$$= b^2 \sum_{k=1}^{g} \sum_{l=1}^{g} \{\mathrm{E}\psi_{1k,\mu}\psi_{1l,\gamma} - \mu_{k,\gamma}\mu_{l,\gamma} - \mu_{k,\mu}\mu_{l,\mu} + \mu_{k,\gamma}\mu_{l,\mu}\}^2 . \quad (67)$$

Using (60) and (66) in (67) yields

$$\mathrm{E}[H_n^2(X_1, X_2)] = b^2 \sum_{k=1}^{g} \sum_{l=1}^{g} \left\{ \mathrm{E}(\psi_{1k,\mu}\psi_{2l,\gamma}) - \frac{b^2}{(\gamma!)^2} F_T^{(\gamma)}(x_k) F_T^{(\gamma)}(x_l) \right.$$

$$\left. - \frac{b^2}{(\mu!)^2} F_T^{(\mu)}(x_k) F_T^{(\mu)}(x_l) + \frac{b^2}{\gamma!\mu!} F_T^{(\gamma)}(x_k) F_T^{(\mu)}(x_l) + O(a^2) \right\}^2 (1 + o(b^2 n^{-p}))$$

$$= \frac{b^4}{a^{2(\mu+\gamma)}} \iint \omega^2(t, u) g(t) g(u) \, dt \, du (1 + o(b^2 n^{-p}))$$

$$- \frac{2b^4}{a^{2(\mu+\gamma)}} \iiint \omega(t,u)\omega(t,v)g(t)f_T(u)f_T(v)\,dt\,du\,dv(1+o(b^2n^{-p}))$$

$$+ \frac{b^4}{a^{2(\mu+\gamma)}} \left\{ \iint \omega(t,u)f_T(u)f_T(u)\,dt\,du \right\}^2 (1+o(b^2n^{-p}))(1+o(b^2n^{-p})). \quad (68)$$

Using (35), (36) and (38) in (68),

$$\mathrm{E}[H_n^2(X_1,X_2)] = b^2 \sum_{k=1}^{g}\sum_{l=1}^{g} \Big\{ - \frac{a^{\mu+\gamma}}{(\gamma-1)!\mu!} \int F_T^{(\gamma)}(z)F_T^{(\mu)}(z)\,dz$$

$$+ \frac{a^{2(\mu+\gamma)}}{(\mu!(\gamma-1)!)^2} \int g(t) \left(f_T^{(\gamma-2)}(t) \right)^2 dt(1+O(a^{2\mu})) + aR(g)R(C_\rho^n * W_\nu^n) + O(a^4) \Big\}$$

$$= \frac{b^2}{a^{2(\mu+\gamma)-1}} R(g)R(C_\mu^* * W_\gamma^*)(1+o(1)). \quad (69)$$

Working in an entirely similar way

$$\mathbb{E}[H_n^4(X_1,X_2)] \le \frac{b^4}{a^{4(\mu+\gamma)-1}} R(K_\gamma^*)^4(1+o(1)),$$

$$\mathbb{E}[G_n^4(X_1,X_2)] = O\left(b^8 a^{-4(\mu+\gamma)}\right),$$

which concludes the proof of (63). Regarding (64), first note that

$$\frac{\mu!\nu!}{(nb)^2} \sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{g} b\psi_{1k,\nu}\mu_{k,\nu} = \frac{\mu!\nu!}{nb} \sum_{l=1}^{n}\sum_{k=1}^{g} \psi_{1k,\nu}\mu_{k,\nu}. \quad (70)$$

Now, using (70) in the sixth step below that $W_\nu^n(t) = ba^{-\nu}W_\nu^*(t)$, we have,

$$\frac{1}{nb}\sum_{l=1}^{n}\sum_{k=1}^{g} \psi_{lk,\nu}\mu_{k,\nu} = \frac{1}{b^2}\mathrm{E}\sum_{k=1}^{g} b\psi_{lk,\nu}\mu_{k,\nu} = \frac{1}{b^2}\sum_{k=1}^{g} b\mu_{k,\nu}(\mathrm{E}\psi_{1k,\nu}) = \frac{1}{b^2}\sum_{k=1}^{g} b\mu_{k,\nu}^2$$

$$= \frac{1}{b^2}\sum_{k=1}^{g} b \left(\sum_{i=1}^{g} W_\nu^n\left(\frac{x_i - x_k}{a}\right) bf_T(x_i)(1+O_p(bn^{-1/2})) \right)^2$$

$$= a^{-\nu}\left\{ \frac{a^\nu}{\nu!}R(F_T^{(\nu)}) + \frac{a^{\nu+2}}{(\nu+2)!}\int F_T^{(\nu)}F_T^{(\nu+2)} \int u^{\nu+2}K_\nu^* \right\}(1+o(n^{-p})).$$

Thus,

$$\frac{\mu!\gamma!}{(nb)^2}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{g} b\psi_{lk,\mu}\mu_{k,\mu} = \mu!\gamma!a^{-\nu}\left\{ \frac{a^\nu}{\nu!}R(F_T^{(\nu)}) + \frac{a^{\nu+2}}{(\nu+2)!}\int F_T^{(\nu)}F_T^{(\nu+2)} \int u^{\nu+2}K_\nu^* \right\}(1+o(n^{-p})),$$

from which (64) immediately follows. For (65), first note that

$$\frac{\mu!\nu!}{(nb)^2}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{g} b\mu_{k,\mu}\mu_{k,\nu} = \frac{\mu!\nu!}{b^2}\sum_{k=1}^{g} b\mu_{k,\mu}\mu_{k,\nu}. \quad (71)$$

Now,

$$\sum_{k=1}^{g} b\mu_{k,\mu}\mu_{k,\nu} = \sum_{k=1}^{g} b\left\{ \sum_{i=1}^{g} W_\nu^n\left(\frac{x_i-x_k}{a}\right)\frac{\mathrm{E}(c_{kl})}{1-\hat{H}(x_k)} \right\}\left\{ \sum_{i=1}^{g} W_\mu^n\left(\frac{x_i-x_k}{a}\right)\frac{\mathrm{E}(c_{kl})}{1-\hat{H}(x_k)} \right\}$$

$$= \frac{1}{a^{\nu+\mu}}\left\{ \frac{a^{\nu+\mu}}{\mu!\nu!}\theta_{\nu,\mu} + \frac{(1+\delta_{\nu\mu})a^{\nu+\mu+2}}{\nu!(\mu+2)!}\theta_{\nu,\mu+2}\mu_{\mu+2}(K_\mu^*) + o(a^{\nu+\mu}) \right\}. \quad (72)$$

By (71) and (72)

$$\frac{\mu!\nu!}{(nb)^2}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{k=1}^{g} b\mu_{k,\mu}\mu_{k,\nu} = \frac{1}{b^2 a^{\nu+\mu}}\left\{\frac{a^{\nu+\mu}}{\mu!\nu!}\theta_{\nu,\mu} + \frac{(1+\delta_{\nu\mu})a^{\nu+\mu+2}}{\nu!(\mu+2)!}\theta_{\nu,\mu+2}\mu_{\mu+2}(K_\mu^*) + o(a^{\nu+\mu})\right\},$$

from which (65) immediately follows. □

Lemma 6. *Assume that K has compact support, is Lipschitz continuous, is symmetric about its origin and its first $\mu+2$ derivatives exist. Then provided that $h \sim n^{-1/(\mu+\gamma+3)}$,*

$$\hat\theta_{\mu,\gamma}\left(a_{\hat\lambda}(h)\right) - \hat\theta_{\mu,\gamma}\left(a_\lambda(h)\right) = o(n^{-1/2}).$$

Proof. From the expressions for $\mathrm{E}\{B_{\mu,\gamma}(a)\}$ and $\mathrm{Var}\{B_{\mu,\gamma}(a)\}$ in Lemma 4,

$$B_{\mu,\gamma} = \gamma a^{\gamma-5}\int F_T^{(\gamma)}(z)F_T^{(\mu)}(z)\,dz + o_p(a^{-1}). \tag{73}$$

Moreover note that

$$A_{\mu,\gamma}(a) = \frac{-\gamma}{a^{\gamma-1}}\int F_T^{(\gamma)}(z)F_T^{(\mu)}(z)\,dz + o_p(a^{-1}). \tag{74}$$

Now, by the mean value theorem

$$\hat\theta_{\mu,\gamma}\left(a_{\hat\lambda}(h)\right) - \hat\theta_{\mu,\gamma}\left(a_\lambda(h)\right) = \frac{d}{da}\hat\theta_{\mu,\gamma}(a)\Big|_{a=a^*}\left(a_{\hat\lambda}(h) - a_\lambda(h)\right), \tag{75}$$

where a^* lies between $a_{\hat\lambda}(h)$ and $a_\lambda(h)$. By Lemma 3, (75) can be written as

$$\hat\theta_{\mu,\gamma}\left(a_{\hat\lambda}(h)\right) - \hat\theta_{\mu,\gamma}\left(a_\lambda(h)\right) = \left\{(A_\gamma(a^*) + B_\gamma(a^*))\left(a_{\hat\lambda}(h) - a_\lambda(h)\right) + (A_\mu(a^*) + B_\mu(a^*))\left(a_{\hat\lambda}(h) - a_\lambda(h)\right)\right\}(1 + o_p(1)). \tag{76}$$

Using (73) and (74) in (76) yields

$$\begin{aligned}
\hat\theta_{\mu,\gamma}\left(a_{\hat\lambda}(h)\right) - \hat\theta_{\mu,\gamma}\left(a_\lambda(h)\right) &= o_p(1/a^*)\left(a_{\hat\lambda}(h) - a_\lambda(h)\right)\\
&\overset{(17)}{=} o_p(h^{-\frac{\mu+\gamma+1}{\mu+\gamma-1}})C(K)D(\theta)h^{\frac{\mu+\gamma+1}{\mu+\gamma-1}}\left(\hat\lambda^{\frac{2}{\mu+\gamma-1}} - \lambda^{\frac{2}{\mu+\gamma-1}}\right)\\
&= o_p(1)\left\{\left(\lambda + O_p(n^{-1/2})\right)^{\frac{2}{\mu+\gamma-1}} - \lambda^{\frac{2}{\mu+\gamma-1}}\right\}\\
&= o_p(1)\left\{\lambda^{\frac{2}{\mu+\gamma-1}} + \binom{\frac{2}{\mu+\gamma-1}}{1}\lambda^{-\frac{\mu+\gamma+1}{\mu+\gamma-1}}O_p(n^{-1/2}) - \lambda^{\frac{2}{\mu+\gamma-1}}\right\} = o_p(n^{-1/2}).
\end{aligned}$$

□

7.2. Proof of Theorem 3

Let $\mu_\nu = \int u^\nu K$ and define the function L_λ as

$$L_\lambda(h) = h\left\{\mu_\mu^{\mu+\gamma}\hat\theta_{\mu,\gamma}(a_\lambda(h))\right\}^{\frac{1}{\mu+\gamma-1}} - n^{-\frac{1}{\mu+\gamma-1}}R(K)^{\frac{1}{\mu+\gamma-1}}.$$

Assume that K is positive only on $[-1,1]$. Then, for a fixed censored sample $X_1,\ldots,X_n$, $L_{\hat\lambda}(h)\to\infty$ as $h\to\infty$ and $L_{\hat\lambda}(h) < 0$ as $a_{\hat\lambda}(h)\downarrow b$ and $0 < b\downarrow 0$ (e.g. b $\downarrow 0$ means $b\to 0+$ (i.e. b goes to zero from above). This means that $L_{\hat\lambda}(h)$ has roots on the positive real line. Note that $\hat h$ is a root of $L_{\hat\lambda}(h)$ and $\hat h \sim n^{-\frac{1}{\mu+\gamma-1}}$. Then,

$$0 = L_{\hat\lambda}(\hat h) = h\left\{\mu_\mu^{\mu+\gamma}\hat\theta_{\mu,\gamma}(a_{\hat\lambda}(h))\right\}^{\frac{1}{\mu+\gamma-1}} - n^{-\frac{1}{\mu+\gamma-1}}R(K)^{\frac{1}{\mu+\gamma-1}}. \tag{77}$$

Using Lemma 6,

$$\left\{\hat{\theta}_{\mu,\gamma}(a_{\hat{\lambda}}(h))\right\}^{\frac{1}{\mu+\gamma-1}} = \left\{\hat{\theta}_{\mu,\gamma}(a_{\lambda}(h)) + o_p(n^{-1/2})\right\}^{\frac{1}{\mu+\gamma-1}}$$
$$= \left\{\hat{\theta}_{\mu,\gamma}(a_{\lambda}(h))\right\}^{\frac{1}{\mu+\gamma-1}} + \frac{1}{\mu+\gamma-1}\left\{\hat{\theta}_{\mu,\gamma}(a_{\lambda}(h))\right\}^{\frac{1}{\mu+\gamma-1}-1} o_p(n^{-1/2})$$
$$= \left\{\hat{\theta}_{\mu,\gamma}(a_{\lambda}(h))\right\}^{\frac{1}{\mu+\gamma-1}} + o_p(n^{-1/2}). \quad (78)$$

Using (78) back in (77) yields

$$L_{\hat{\lambda}}(\hat{h}) = h\left\{\mu_{\mu}^{\mu+\gamma}\right\}^{\frac{1}{\mu+\gamma-1}}\left\{\hat{\theta}_{\mu,\gamma}(a_{\lambda}(h))\right\}^{\frac{1}{\mu+\gamma-1}} - n^{-\frac{1}{\mu+\gamma-1}} R(K)^{\frac{1}{\mu+\gamma-1}},$$

and thus

$$L_{\lambda}(\hat{h}) = L_{\hat{\lambda}}(\hat{h}) + O_p(n^{-1/2}n^{-\frac{1}{\mu+\gamma-1}}) = L_{\hat{\lambda}}(\hat{h}) + O_p\left(n^{-\frac{\mu+\gamma+1}{2(\mu+\gamma-1)}}\right). \quad (79)$$

By Lemma 5 and the Delta method

$$n^{\alpha_1} L_{\lambda}(h_*) \xrightarrow{d} N(\mu_1, \sigma_1^2), \quad (80)$$

with

$$\mu_1 = \mathrm{E}\left\{n^{\alpha_1} h_* \mu_{\gamma}^{\frac{\mu+\gamma}{\mu+\gamma-1}} \hat{\theta}_{\mu,\gamma}(\alpha_{\lambda}(h))^{\frac{1}{\mu+\gamma-1}}\right\}$$
$$= n^{\alpha_1} h_* \mu_{\gamma}^{\frac{\mu+\gamma}{\mu+\gamma-1}} \mathrm{E}\left\{\hat{\theta}_{\mu,\gamma}(\alpha_{\lambda}(h))^{\frac{1}{\mu+\gamma-1}}\right\}$$
$$= n^{\alpha_1} h_* \mu_{\gamma}^{\frac{\mu+\gamma}{\mu+\gamma-1}} \Bigg[\theta_{\mu,\gamma} + \frac{\mu!\gamma!}{na^{\mu+\gamma-1}}\left\{\int\left(\frac{f_T(u)}{1-H(u)}\right)du\right\}\int W_{\mu}^* K_{\gamma}^*$$
$$+ \frac{(1+\delta_{\mu\gamma})\gamma!}{(\gamma+2)!} h_*^2 \theta_{\mu,\gamma+2}\mu_{\gamma+2}(K_{\gamma}^*)\Bigg]^{\frac{1}{\mu+\gamma-1}}$$
$$= n^{\alpha_1} h_* \mu_{\gamma}^{\frac{\mu+\gamma}{\mu+\gamma-1}} \theta_{\mu,\gamma}^{\frac{1}{\mu+\gamma-1}} + \frac{n^{\alpha_1}(1+\delta_{\mu\gamma})\gamma!}{(\gamma+2)!(\mu+\gamma+1)} h_*^3 \mu_{\gamma}^{\frac{\mu+\gamma}{\mu+\gamma-1}} \theta_{\mu,\gamma}^{\frac{2-(\mu+\gamma)}{\mu+\gamma+1}} \theta_{\mu,\gamma+2}\mu_{\gamma+2}(K_{\gamma}^*)$$
$$+ \frac{\mu!\gamma!}{na^{\mu+\gamma-1}}\int W_{\mu}^* K_{\gamma}^*, \quad (81)$$

after using in the last step above the expansion

$$\{\theta_{\mu,\gamma}(a) + g(x)\}^{\frac{1}{\mu+\gamma-1}} = \theta_{\mu,\gamma}(a)^{\frac{1}{\mu+\gamma-1}} + \frac{1}{\mu+\gamma-1}\theta_{\mu,\gamma}(a)^{\frac{1}{\mu+\gamma-1}-1} g(x),$$

where $g(x)$ is a generic function. Then, using the definition of h_* (see (14)) in (81) yields

$$\mu_1 = n^{\alpha_1}\left\{\frac{2}{n}\frac{(2\gamma-1)(\gamma!)^2 C_1 A_{1,1}}{\mu_{\gamma+2}^2(K_{\gamma}^*)\theta_{\gamma+2,\gamma+2}}\right\}^{\frac{1}{2\gamma+3}} \mu_{\gamma}^{\frac{\mu+\gamma}{\mu+\gamma-1}} \theta_{\mu,\gamma}^{\frac{1}{\mu+\gamma-1}}$$
$$+ \frac{n^{\alpha_1}(1+\delta_{\mu\gamma})\gamma!}{(\gamma+2)!(\mu+\gamma+1)} \mu^{\frac{\mu+\gamma}{\mu+\gamma-1}} \theta_{\mu,\gamma}^{\frac{2-(\mu+\gamma)}{\mu+\gamma-1}}$$
$$\times \left\{\frac{2}{n}\frac{(2\gamma-1)(\gamma!)^2 C_1 A_{1,1}}{\mu_{\gamma+2}^2(K_{\gamma}^*)\theta_{\gamma+2,\gamma+2}}\right\}^{\frac{3}{2\gamma+3}} \theta_{\mu,\gamma+2}\mu_{\gamma+2}(K_{\gamma}^*)(1+\chi^{-1}) + O\left(n^{-1}a^{-(\mu+\gamma)+1}\right).$$

Also, σ_1^2 is given by

$$\sigma_1^2 = \mathrm{Var}\left\{n^{\alpha_1} L_{\lambda}(h_*)\right\} = \mathrm{Var}\left\{n^{\alpha_1} h_* \mu_{\gamma}^{\frac{\mu+\gamma}{\mu+\gamma-1}} \hat{\theta}_{\mu,\gamma}(\alpha_{\lambda}(h))^{\frac{1}{\mu+\gamma-1}}\right\}$$

$$= n^{2\alpha_1} h_*^2 \mu_\gamma^{\frac{2(\mu+\gamma)}{\mu+\gamma-1}} \frac{1}{\mu+\gamma-1} \theta_{\mu,\gamma}^{2\frac{2-(\mu+\gamma)}{\mu+\gamma-1}} \frac{2(\mu!\gamma!)^2}{n^2 a^{2(\mu+\gamma)-1}} R(g) R(C_\mu^n * W_\gamma^n)$$
$$= n^{2\alpha_1-2} h_*^2 \mu_\gamma^{\frac{2(\mu+\gamma)}{\mu+\gamma-1}} \theta_{\mu,\gamma}^{\frac{4-2(\mu+\gamma)}{\mu+\gamma-1}} \frac{2(\mu!\gamma!)^2}{a^{2(\mu+\gamma)-1}(\mu+\gamma-1)} R(g) R(C_\mu^n * W_\gamma^n).$$

Use the fact that from (17), $a(\hat{h}_\nu) = C(K)D(\theta)\hat{h}_\nu^{\frac{2\nu+1}{2\nu+3}}$ as well as the definition of h_* in (14) to obtain,

$$\sigma_1^2 = n^{2\alpha_1-2} h_*^2 \mu_\gamma^{\frac{2(\mu+\gamma)}{\mu+\gamma-1}} \frac{2(\mu!\gamma!)^2 a R(g) R(C_\mu^n * W_\gamma^n)}{(\mu+\gamma-1)\left\{C(K)D(\theta)\hat{h}_*^{\frac{2\gamma+1}{2\gamma+3}}\right\}^{2(\mu+\gamma)-1}} \theta_{\mu,\gamma}^{\frac{4-2(\mu+\gamma)}{\mu+\gamma-1}}$$
$$= \frac{2(\mu!\gamma!)^2}{\mu+\gamma-1} n^{2\alpha_1-2} \mu_\gamma^{\frac{2(\mu+\gamma)}{\mu+\gamma-1}} \left\{\frac{2}{n}\frac{(2\gamma-1)(\gamma!)^2 C_1 A_{1,1}}{\mu_{\gamma+2}^2(K_\nu^*)\theta_{\gamma+2,\gamma+2}}\right\}^{\frac{2(2\gamma+3)-(2\gamma+1)(2(\mu+\gamma)-1)}{(2\gamma+3)^2}}$$
$$\times \{C(K)D(\theta)\}^{-2(\mu+\gamma)+1} R(g) R(C_\mu^n * W_\gamma^n) \theta_{\mu,\gamma}^{\frac{4-2(\mu+\gamma)}{\mu+\gamma-1}}.$$

Further,

$$\frac{d}{dh}L_\lambda(h) = \left\{\mu_\gamma^{\mu+\gamma}\hat{\theta}_{\mu,\gamma}(a_\lambda(h))\right\}^{\frac{1}{\mu+\gamma-1}}$$
$$+ h\mu_\gamma^{\frac{\mu+\gamma}{\mu+\gamma-1}} \frac{1}{\mu+\gamma-1}\{A_{n,\mu}(a_\lambda(h)) + B_{n,\mu}(a_\lambda(h) + A_{n,\gamma}(a_\lambda(h)) + B_{n,\gamma}(a_\lambda(h)\}$$
$$\times C(K)D(g_\lambda)h^{\frac{2-(\mu+\gamma)}{\mu+\gamma-1}}$$
$$\overset{(73),(74)}{=} \left\{\mu_\gamma^{\mu+\gamma}\frac{-\mu}{a^{\mu-1}}\int F_T^{(\mu)}(z)F_T^{(\gamma)}(z)\,dz + o_p(a^{-1}) + \mu a^{\mu-5}\int F_T^{(\mu)}(z)F_T^{(\gamma)}(z)\,dz\right.$$
$$\left. + \frac{-\gamma}{a^{\gamma-1}}\int F_T^{(\gamma)}(z)F_T^{(\mu)}(z)\,dz + \gamma a^{\gamma-5}\int F_T^{(\mu)}(z)F_T^{(\gamma)}(z)\,dz\right\}^{\frac{1}{\mu+\gamma-1}} + o_p(1). \tag{82}$$

Now,

$$L_\lambda(\hat{h}) = L_\lambda(h_*) + \frac{d}{dh}L_\lambda(h^{**})(\hat{h}-h_*), \tag{83}$$

where h^{**} is between $\hat{h}$ and h_*. By (79) and (83),

$$n^\alpha\left(\frac{\hat{h}-h_*}{h_*}\right) = n^\alpha\left(\frac{L_\lambda(\hat{h})-L_\lambda(h_*)}{h_*\frac{d}{dh}L_\lambda(h^{**})}\right) = n^\alpha\left(\frac{O_p\left(n^{-\frac{\mu+\gamma+1}{2(\mu+\gamma-1)}}\right)-L_\lambda(h_*)}{h_*\frac{d}{dh}L_\lambda(h^{**})}\right). \tag{84}$$

Using (82) in the denominator of (84) and subsequently applying (80) yields

$$n^\alpha\left(\frac{\hat{h}-h_*}{h_*}\right) = n^\alpha\left(\frac{n}{R(K)}\right)^{\frac{1}{\mu+\gamma-1}} L_\lambda(h_*)(-1+o_p(1)) \xrightarrow{d} N(\mu_{DPI}, \sigma_{DPI}^2),$$

which completes the proof.

Acknowledgments

The authors thank the associate editor and two anonymous reviewers for their helpful comments and suggestions which have greatly improved this research.

References

Alevizos, F., Bagkavos, D., Ioannides, D., 2019. Efficient estimation of a distribution based on censored data. Stat. Probab. Lett. 145, 359-364.

Bagkavos, D., Patil, P.N., 2008. Local polynomial fitting in failure rate estimation. IEEE Trans. Rel. 56, 126–163.

Ioannides, D. and Bagkavos, D., 2020. Smooth nonparametric survival analysis, *Proceedings of the 4th ISNPS conference, Salerno, Italy, June 2018.*

Bagkavos, D. and Ioannides, D., 2020. Local polynomial smoothing based on the Kaplan–Meier estimate, *under review.*

Cattaneo, M.D., Jansson, M., and Ma, X., 2019. Simple Local Polynomial Density Estimators. *J. Amer. Statist. Assoc.*, *to appear*, DOI: 10.1080/01621459.2019.1635480

Cao, R. Jacome, M.A., 2004. Presmoothed kernel density estimator for censored data. J. Nonparametr. Stat. 16, 289–309.

Chen, K., Lo, S.-H., 1997. On the rate of uniform convergence of the product-limit estimator: strong and weak laws. Ann. Statist. 25, pp. 1050–1087.

Cheng, M.-Y. Fan, J., Marron, J.S., 1997. On automatic boundary corrections. Ann. Statist. 25, pp. 1691–1708.

Cheng, M.-Y., 1997. Boundary aware estimators of integrated density derivative products. J. R. Stat. Soc. Ser. B Stat. Methodol. 59, 191–203.

Cheng, M.-Y., 1994. On boundary effects of smooth curve estimators. *Mimeo 2319*, Institute of Statistics, University of North Carolina, Chapel Hill.

Fan, J. and Gijbels, I., 1996. *Local polynomial modeling and its applications.* Chapman and Hall, London.

Fan, J., Gijbels, I., Brockmann, M., Engels, J., 1997. Local Polynomial regression: optimal kernels and asymptotic minimax efficiency. Ann. Inst. Stat. Math. 49, 79–99.

Funke, B., Hirukawa, M., 2020. Bias correction for local linear regression estimation using asymmetric kernels via the skewing method. Econ. Stat., In press. doi: https://doi.org/10.1016/j.ecosta.2020.01.004.

Hall, P., 1984. Central limit theorem for integrated square error of multivariate nonparametric density estimators. J. Multivariate Anal. 14, 1–16.

Hua, H., Patil, P.N., Bagkavos, D., 2017. Semiparametric smoothing approach to hazard rate estimation. J. Nonparametr. Stat. 29, 669–693.

Gulati, S., Padgett, W.J., 1996. Families of smooth confidence bands for the survival function under the general random censorship model. Lifetime Data Anal. 2, 349–362.

Jeganathan, P., Randrianampy, N.V., Paige, R.L., Trindade, A.A., 2019. An empirical saddlepoint approximation based method for smoothing survival functions under right censoring. Canad. J. Statist. 47, 238–261.

Kaplan, E.L., Meier, P., 1958. Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457–481.

Karunamuni, R. and Yang, S., 1991. Weak and strong uniform consistency rates of kernel density estimates for randomly censored data. Canad. J. Statist. 19, 349–359.

Kulasekera, K.B., Williams, C.L., Coffin, M., Manatuga, A., 2001. Smooth estimation of the reliability function. Lifetime Data Anal. 7, 415–433.

Kim, C., Oh, M., Yang, S.J., Choi, H., 2010. A local linear estimation of conditional hazard function in censored data. J. Korean Stat. Soc. 39, 347–355.

Kus, C., 2007. A new lifetime distribution. Comp. Statist. Data Analys. 51, 4497—4509.

Lai, C. D., Xie, M., 2006. *Stochastic Ageing and Dependence for Reliability.* Springer, New York.

Marron, J.S., Padgett, W.J., 1987. Asymptotically optimal bandwidth selection for kernel density estimators from randomly right censored samples. Ann. Statist. 15, 1520–1535.

McNeil, A., 1997. Estimating the tails of loss severity distributions using extreme value theory. Astin Bulletin 71, 117–137.

Müller, H.–G., Wang, J.L., 1994. Hazard rate estimation under random censoring with varying kernels and bandwidths. Biometrics 50, 61–76.

Mammen, E., Nielsen, J.P., 2007. A general approach to the predictability issue in survival analysis with applications. Biometrika 94, 873–892.

Müller, H.–G., Wang, J.L., Capra, W.B., 1997. From lifetables to hazard rates: the transformation approach. Biometrika 84, 881–892.

Navarro, J., Hernandez, P.J., 2004. How to obtain bathtub-shaped failure rate models from normal mixtures. Probab. Eng. Inform. Sc. 18, 511–531.

Nielsen, J.P., 2003. Smoothing and prediction with a view to actuarial science, biostatistics and finance. Scand. Actuar. J. 1, 51–74.

Nielsen, J.P., Tangaard, C., Jones, M.C., 2009. Local linear density estimation for filtered survival data, with bias correction. Statistics 43, 167–186.

Proschan, F., 1963. Theoretical explanation of observed decreasing failure rate. Technometrics 5, 375–383.

Pulcini, G., 2001. Modeling the failure data of a repairable equipment with bathtub type failure intensity. Reliab. Eng. Syst. Saf. 71, 209–218.

Satten, G.A. and Datta, S., 2015. The Kaplan–Meier estimator as an inverse-probability-of-censoring weighted average. Am. Stat. 55, 207–210.

Sheather, S.J., Jones, M.C., 1991. A reliable data-based bandwidth selection method for kernel density estimation. J. R. Stat. Soc. Ser. B Stat. Methodol. 53, 683–690.

Wang, J.L., Müller, H.–G., Capra, W.B., 1998. Analysis of oldest-old mortality: lifetables revisited. Ann. Statist. 26, 126–163.

Wang, Q., Liu, W., Liu, L., 2009. Probability density estimation for survival data with censoring indicators missing at random. J. Multivariate Anal. 100, 835–850.

Youndje, E., 2016. Some heuristics about bandwidth selection for the smooth Kaplan–Meier estimator. J. Korean Stat. Soc. 45, 568–580.

Zhang, B., 1998. A note on the integrated square errors of kernel density estimators under random censorship. Stoch. Proc. Appl. 75, 225–234.