Ann. Inst. Statist. Math.
Vol. 53, No. 2, 289–306 (2001)

ASYMPTOTICALLY LOCAL MINIMAX ESTIMATION OF INFINITELY SMOOTH DENSITY WITH CENSORED DATA

EDUARD BELITSER[1]* AND BORIS LEVIT[2]

[1]*Department of Mathematics, Vrije Universiteit, de Boelelaan 1081a, 1081 HV Amsterdam, the Netherlands*
[2]*Mathematical Institute, Utrecht University, Budapestlaan 6, 3584 CD Utrecht, the Netherlands*

(Received February 9, 1998; revised September 28, 1999)

Abstract. The problem of the nonparametric minimax estimation of an infinitely smooth density at a given point, under random censorship, is considered. We establish the exact asymptotics of the local minimax risk and propose the efficient kernel-type estimator based on the well known Kaplan-Meier estimator.

Key words and phrases: Efficient estimator, local minimax risk, Kaplan-Meier estimator, kernel, random censorship.

1. Introduction and some definitions

Suppose $T_1, \dots, T_n$ are iid random variables (lifetimes) with common distribution function F and density f and suppose $C_1, \dots, C_n$ are iid random variables (censoring times) with common distribution function G. Assume that the lifetimes and censoring times are independent. According to the classical *random censorship* model, one observes the bivariate sample $(Z_1, \delta_1), \dots, (Z_n, \delta_n)$, where $Z_i = \min(T_i, C_i)$ and $\delta_i = I\{T_i \le C_i\}$. Estimation problems with censored observations arise often in lifetime research, and in medical and biological applications the random censorship may be a realistic model. We suppose F and G are unknown and our goal is to estimate the density $f(x)$ at a given point x, using the observed data.

The problem of density estimation under random censorship is not new and has long been treated in the literature (see Mielniczuk (1986), Lo *et al.* (1989), Kulasekera (1995), Huang and Wellner (1995) and further references therein). Many interesting aspects of the problem were investigated in those papers and all these studies led to a better understanding of risk computations for Kaplan-Meier based estimators. However, in the context of the well established minimax estimation framework, the issue of optimality of considered estimators remained open. Typically, the notion of asymptotic optimality is associated with the so called optimal rate of convergence of the minimax risk. It is an interesting and challenging task to derive the exact asymptotic behaviour of the minimax risk in the density estimation problem and to find an estimator achieving this asymptotics. Results of such kind have only been obtained in a limited number of studies for models with independent identically distributed observations.

We now elucidate all the above mentioned notions. To make the problem of minimax

*Now at Mathematical Institute, Leiden University, Niels Bohrweg 1, 2333 CA Leiden, the Netherlands.

density estimation feasible, one typically restrict oneself to a certain nonparametric class of densities $\mathcal{F}$ described usually by some smoothness assumptions. As a measure of quality of an estimator $\tilde{f}_n$ introduce the *maximal risk* of the estimator and *minimax risk*:

$$R_n(\tilde{f}_n, \mathcal{F}) = R_n(\tilde{f}_n, \mathcal{F}, x) = \sup_{f \in \mathcal{F}} E_f(\tilde{f}_n(x) - f(x))^2,$$

$$r_n(\mathcal{F}) = r_n(\mathcal{F}, x) = \inf_{\tilde{f}_n} R_n(\tilde{f}_n, \mathcal{F}, x), \tag{1.1}$$

where the infimum is taken over all estimators $\tilde{f}_n$.

The minimax risk expresses the least possible mean loss when the worst case happens and, in a way, reflects the complexity of the estimation problem over the class $\mathcal{F}$. In practice, however, only one density is in the background. This raises the question how to characterize the difficulty of estimation problem contributed solely by this particular curve. A natural way to do this is to localize the risk function. To be more precise, let $\mathcal{T}$ be a topology on the class $\mathcal{F}$. For each neighbourhood $V \in \mathcal{T}$ define the *local maximal risk* of an estimator $\tilde{f}_n(x)$ and *local minimax risk* as $R_n(\tilde{f}_n, V)$ and $r_n(V)$ respectively. The estimator $\tilde{f}_n$ is called *locally asymptotically minimax* (or just *efficient*) at f if there exists a neighbourhood $V_0 \ni f$ such that for any neighbourhood V, $f \in V \subseteq V_0$ (from now on we will just say: for any sufficiently small neighbourhood of f)

$$\lim_{n\to\infty} \frac{R_n(\tilde{f}_n, V)}{r_n(V)} = 1,$$

with the convention $0/0 = 1$.

Ibragimov and Hasminskii (1982) derived the exact asymptotics of the global minimax risk over a class of multivariate densities with compactly supported characteristic functions. The exact asymptotic behaviour of the local minimax risk for the class of analytic densities was studied in Golubev and Levit (1996) and in Belitser (1998), in noncensored case and under censorship, respectively. The major difficulty for the class of analytic densities is that there are no efficient kernel estimators with compactly supported kernels due to the phenomenon of the "long-range reciprocal memory" contained in two separated sets of values of an analytic function. Therefore, the censoring may cause problems in estimation of analytic densities. However, in Belitser (1998) it has been shown that one can choose a kernel with exponentially decreasing tails under condition that censoring is not of severe influence. The proof of efficiency of the estimator is based on the martingale technique and is rather involved.

In this paper we consider a class of infinitely smooth functions, somewhat broader class than the class of analytic functions. In this case, as we show, there are efficient estimators with compactly supported kernels. This facilitates also the use of the result of Lo *et al.* (1989) about strong representation of the Kaplan-Meier estimator by a sum of independent random variables in the proof of the upper bound. The treatment of the lower bound is based on the van Trees inequality and is in essence similar to that in Belitser (1998). A certain useful tool, the Approximation Lemma, plays an important role in the proofs of the main results. It reflects the fact that any density from the considered nonparametric class can be approximated by a sequence of smooth functionals with a negligible approximation error, thus linking our problem with a regular estimation problem. We establish the exact limiting behavior of the local minimax risk and propose an efficient kernel-type estimator. The constructed estimator is basically a modified Fourier integral estimator introduced by Konakov (1972) and Davis (1975). The modification consists in trimming the so called sinc kernel $K(u) = \sin u/(\pi u)$ by a suitably chosen compactly supported infinitely smooth function.

2. Main results

Denote from now on the Fourier transformation of an absolutely integrable function f by $\hat{f} : \hat{f}(t) = \int e^{ity} f(y)dy$. In the proofs we shall use notations $C_1, C_2, \ldots$ for generic positive constants which are assumed to be different in different proofs. Unless otherwise specified, all asymptotic relations refer to $n \to \infty$.

Define now the nonparametric class $\mathcal{F}_\delta$ of underlying densities and the topology $\mathcal{T}_\delta$ on it.

DEFINITION 2.1. For given $P, \delta > 0$, $0 < r < 1$ denote

$$\mathcal{F}_\delta = \mathcal{F}_\delta(P,r) = \left\{ f(\cdot) : (2\pi)^{-1} \int \exp(2\delta|t|^r)|\hat{f}(t)|^2 dt < P \right\}. \tag{2.1}$$

One can easily see that functions from this class are infinitely differentiable. Note also that the class $\mathcal{F}_\delta$ is quite broad: the Gauss, student and Cauchy distributions are, among many others, for appropriate δ, in this class, as well as their mixtures.

DEFINITION 2.2. Let $\mathcal{S}_\delta$ and $\mathcal{U}_\delta = \mathcal{U}_\delta(r)$ be the topologies on $\mathcal{F}_\delta(P,r)$ induced by the distances

$$\rho_s(f,g) = \sup_y |f(y) - g(y)| + \sup_y |f'(y) - g'(y)| + \int |f(y) - g(y)|\, dy,$$

$$\rho_u(f,g) = \left(\int \exp(2\delta|t|^r)|\hat{f}(t) - \hat{g}(t)|^2 dt \right)^{1/2} + \int |f(y) - g(y)| dy,$$

respectively. $\mathcal{U}_\delta$ is a strong topology—closeness with respect to ρ_u implies, by the formula for the inverse Fourier transform, closeness of all derivatives in the uniform topology: for $g, h \in \mathcal{F}_\delta$,

$$\sup_y |g^{(m)}(y) - h^{(m)}(y)| \le C_1 \left(\int t^{2m} |\hat{g}(t) - \hat{h}(t)|^2 dt \right)^{1/2} \le C_2 \rho_u(g,h).$$

Therefore, $\mathcal{S}_\delta \subseteq \mathcal{U}_\delta$. Let $\mathcal{T}_\delta$ be any topology on $\mathcal{F}_\delta(P,r)$ such that $\mathcal{S}_\delta \subseteq \mathcal{T}_\delta \subseteq \mathcal{U}_\delta$.

Remark 2.1. $\mathcal{S}_\delta$ and $\mathcal{U}_\delta$ are possible choices of weak and strong topologies respectively, for which the properties stated in the assertions below hold locally uniformly, i.e. for each $f \in \mathcal{F}_\delta$ there exists a neighbourhood $V(f)$ such that these properties hold uniformly over this neighbourhood. In fact, in assertions concerning the upper bound for the local minimax risk, one need to prove the local uniformity only for the topology $\mathcal{S}_\delta$, and in assertions concerning the lower bound only for the topology $\mathcal{U}_\delta$.

Now we describe a class of kernels to be used in the construction of the estimator.

Denote, for some positive b, β, A, $m \geq 0$,

$$q_r(y) = q_r(y,b,\beta) = \begin{cases} A\exp\left(-\dfrac{1}{(b^2-y^2)^\beta}\right), & -b<y<b \\ 0, & y \notin (-b,b), \end{cases} \tag{2.2}$$

$$a_n = a_n(m,\delta,r) = \left(\frac{\log n + m\log\log n}{2\delta}\right)^{1/r}, \tag{2.3}$$

$$s_n(y) = s_n(y,m,\delta,r) = \frac{\sin(a_n y)}{\pi y} \quad \text{for} \quad y \neq 0 \text{ and } s_n(0) = a_n/\pi, \tag{2.4}$$

where the constants δ and r appear in the definition of the class $\mathcal{F}_\delta$ and the constant A is defined by the requirement:

$$q_r(0) = 1. \tag{2.5}$$

Note that $s_n(y) = a_n K(a_n y)$ with $K(y) = \sin y/(\pi y)$ (cf. Konakov (1972) and Davis (1975)). The other constants are chosen according to the following conditions:

(i) the constant b is any fixed number such that $b + x < \tau_G$, where $\tau_G = \inf\{y : G(y) = 1\}$ and x is the point at which we want to estimate the density f;

(ii) the constant β is any fixed number such that $\beta/(\beta+1) > r$, where r is the parameter in the definition of the class $\mathcal{F}_\delta$;

(iii) the constant m is any fixed number such that $m/2 > 1/r - 1$.

Note that if $x > \tau_G$, then even consistent estimation of $f(x)$ is not possible. Next introduce the kernel

$$\phi_n(y) = \phi_n(y,\tau,\delta,m) = q_r(y)s_n(y), \tag{2.6}$$

and define the following estimator

$$\tilde{f}_n = \tilde{f}_n(x) = \int \phi_n(x-y)d\tilde{F}_n(y), \tag{2.7}$$

where $\tilde{F}_n(y)$ is the Kaplan-Meier estimator, a well known nonparametric efficient estimator of the distribution function $F(y)$:

$$\tilde{F}_n(y) = 1 - \prod_{i:Z_{(i)}<y} \left(\frac{n-i}{n-i+1}\right)^{\Delta_{(i)}}, \tag{2.8}$$

with the convention $0^0 = 1$. Here the $Z_{(i)}$ denote the ordered sequence of Z_i's and the $\Delta_{(i)}$'s are correspondent indicators. A rich literature is devoted to this estimator and its properties (see Andersen *et al.* (1993) and further references therein).

Remark 2.2. As is shown in Weits (1993), in case of Hölder-type class, the Kaplan-Meier estimator is not optimal with respect to the convergence rate of second order minimax risk. The problem of second order efficiency of a smoothed version of the Kaplan-Meier estimator for the infinitely smooth class will be treated elsewhere.

Remark 2.3. Since, by the standard formula for the Fourier transform of the product of two functions,

$$\hat{\phi}_n(t) = \frac{1}{2\pi}(\hat{q}_r * I_{[-a_n,a_n]})(t), \tag{2.9}$$

$\hat{\phi}_n(t)$ is nothing else but a smoothed indicator of $[-a_n, a_n]$. Here $*$ is the convolution operation and I_S denotes the indicator function of set S. In words, convolution of a function with the kernel ϕ_n in the time domain corresponds to "smooth thresholding" the Fourier transform of the function in the frequency domain.

Note also that the function $\hat{q}_r(t)$ is even. The asymptotic behaviour of $\hat{q}_r(t)$ for $0 < r < 1$, as $|t| \to \infty$, is described in Fedoruk ((1977), p. 229). We adapt this result in a simplified form, suitable for our purposes: for some positive A_1 and A_2,

$$|\hat{q}_r(t)| \le A_1 \exp\{-A_2|t|^{\beta/(\beta+1)}\}, \qquad 0 < r < 1. \tag{2.10}$$

The constants A_1, A_2 depend in general on b, β.

Denote $a \wedge b = \min\{a, b\}$ and $\bar{F}(u) = 1 - F(u)$. In the next theorem the local asymptotic performance of the estimator $\tilde{f}_n$ with respect to the topology $\mathcal{T}_\delta$ is established. The proofs of the theorems are given in the last section.

THEOREM 2.1. *Let $f_0 \in \mathcal{F}_\delta$ be such that $x < \tau_G \wedge \tau_{F_0}$ and distribution function G is continuous at point x. Then, for any sufficiently small neighbourhood $V(f_0)$, the relation*

$$\limsup_{n\to\infty} n(\log n)^{-1/r} E_f(\tilde{f}_n(x) - f(x))^2 \le \sigma^2(f)$$

holds uniformly over $f \in V(f_0)$, where

$$\sigma^2(f) = \sigma^2(f, x) = \frac{f(x)}{(2\delta)^{1/r}\pi\bar{G}(x)} \tag{2.11}$$

and the estimator $\tilde{f}_n(x)$ is defined by (2.7).

Theorem 2.1 gives an upper bound for the local minimax risk $r_n(V)$: for a sufficiently small neighbourhood $V(f_0)$

$$\limsup_{n\to\infty} n(\log n)^{-1/r} r_n(V) \le \limsup_{n\to\infty} n(\log n)^{-1/r} R_n(\tilde{f}_n, V) \le \sup_{f\in V} \sigma^2(f).$$

If we can provide a lower bound for the local minimax risk, coinciding asymptotically with the upper one, then we clearly determine the asymptotic behaviour of the local minimax risk. The next theorem describes the lower bound for the local minimax risk.

THEOREM 2.2. *For any neighbourhood $V \subseteq \mathcal{F}$,*

$$\liminf_{n\to\infty} n(\log n)^{-1/r} r_n(V) \ge \sup_{f\in V} \sigma^2(f),$$

where the local minimax risk $r_n(V)$ and $\sigma^2(f)$ are defined by (1.1) and (2.11) respectively.

In view of Theorems 2.1 and 2.2, the estimator $\tilde{f}_n$ is efficient. Indeed, for each $f \in \mathcal{F}_\delta$ and for any sufficiently small neighbourhood $V(f)$,

$$\lim_{n\to\infty} \sup_{f\in V} E_f(\tilde{f}_n(x) - f(x))^2 / r_n(V) = 1.$$

Remark 2.4. A distinguishing feature of the estimator (2.7) is that it does not depend on a specific neighbourhood while being efficient.

Moreover, as an immediate consequence of Theorems 2.1 and 2.2, we obtain the asymptotic behaviour of the local minimax risk.

COROLLARY 2.1. *Let $f_0 \in \mathcal{F}_\delta$ be such that $x < \tau_G \wedge \tau_{F_0}$ and suppose the distribution function G is continuous at the point x. Then for any sufficiently small neighbourhood $V(f_0)$*

$$\lim_{n\to\infty} n(\log n)^{-1/r} r_n(V) = \sup_{f\in V} \sigma^2(f).$$

Remark 2.5. Since $\sigma^2(\cdot)$ is a continuous functional, this implies also that

$$\lim_{V\downarrow f_0} \lim_{n\to\infty} n(\log n)^{-1/r} r_n(V) = \lim_{V\downarrow f_0} \sup_{f\in V} \sigma^2(f) = \sigma^2(f_0).$$

Remark 2.6. It should be mentioned that the results on local minimaxity are stronger than the global results. Indeed, the global results can be easily derived from Theorems 2.1 and 2.2 provided the condition of uniform boundness $f(x) \le M$ is included in the definition of the nonparametric class to ensure that the minimax risk is not infinite.

Remark 2.7. Using the result of Lo *et al.* (1989), one can derive a central limit theorem and a law of iterated logarithm for the estimator $\tilde{f}_n(x)$:

$$\sqrt{n}(\log n)^{-1/(2r)}(\tilde{f}_n(x) - f(x)) \xrightarrow{d} \mathcal{N}(0, \sigma^2(f, x)) \qquad n \to \infty,$$
$$\limsup_{n\to\infty} \sqrt{n}(\log n)^{-1/(2r)}(2\log\log n)^{-1/2}|\tilde{f}_n(x) - f(x)| = \sigma(f, x) \qquad \text{almost surely.}$$

Here $\to^d$ means convergence in distribution and $\sigma^2(f, x)$ is defined by (2.11).

Remark 2.8. In the proof of Theorem 2.1 we have to assume that the constant b from (2.2) is chosen in such a way that $x + b \le \tau_{F_0}$. Although this seems to be rather restrictive at first sight because we do not know the density f_0, we can assume this without loss of generality. The point is that we can let the constant b depend on n so that $b_n \to 0$ as $n \to \infty$. Provided b_n converges to zero slowly enough, one needs to modify only slightly the proof of the Approximation Lemma 4.1 in the following manner. By the most detailed analysis of Fedoruk's asymptotics for $\hat{q}_r(t)$ as $t \to \infty$ (see Fedoruk (1977), p. 229), one finds that $A_1 = A_1(b_n)$ and $A_2 = A_2(b_n)$ (see (2.10)), which leads to $C_7(b_n) \to \infty$ as $b_n \to 0$ in (4.4). Choose b_n in such a way that $b_n \to 0$ and $C_7(b_n) = o(\log n)$ as $n \to \infty$. All the other proofs remain unchanged.

3. Auxiliary results

In this section we provide technical results which we will need below.

LEMMA 3.1. *Let $q_r(y)$ and $\phi_n(y)$ be defined by (2.2) and (2.6) respectively, and let the function $h(y)$ be continuous at x and satisfy $\int q_r^2(x-y)|h(y)|dy < \infty$. Then, as $n \to \infty$, the relations*

$$\begin{aligned}\int \phi_n^2(x-y)h(y)dF(y) &= h(x)f(x)\int \phi_n^2(y)dy + o((\log n)^{1/r}) \\ &= \frac{h(x)f(x)}{\pi}\left(\frac{\log n}{2\delta}\right)^{1/r} + o((\log n)^{1/r})\end{aligned}$$

hold locally uniformly in $f \in \mathcal{F}_\delta$.

PROOF. To prove the second equality, by (2.4) and (2.5), write

$$\begin{aligned}\int \phi_n^2(y)dy &= \int_{|y|\le a_n^{-1/2}} \phi_n^2(y)dy + \int_{|y|>a_n^{-1/2}} \phi_n^2(y)dy \\ &= (1+o(1))\int_{|y|\le a_n^{-1/2}} s_n^2(y)dy + O(1)\int_{|y|>a_n^{-1/2}} s_n^2(y)dy \\ &= (1+o(1))a_n\pi^{-2}\int_{|y|\le a_n^{1/2}} \frac{\sin^2(y)}{y^2}dy + O(1)a_n\int_{|y|>a_n^{1/2}} \frac{\sin^2(y)}{y^2}dy \\ &= \frac{a_n}{\pi}(1+o(1)) = \frac{1}{\pi}\left(\frac{\log n}{2\delta}\right)^{1/r}(1+o(1)),\end{aligned}$$

where a_n is defined by (2.3).

Let us prove the first relation. Let $O_\epsilon(x) = \{y : |x-y| < \epsilon\}$ be the open interval around x of radius $\epsilon = \epsilon_n$ such that $\epsilon_n \to 0$ and $\epsilon_n^{-2}(\log n)^{-1/r} = o(1)$ as $n \to \infty$. We obviously have

$$\begin{aligned}&\int \phi_n^2(x-y)h(y)dF(y) - h(x)f(x)\int \phi_n^2(y)dy \\ &\quad= \int_{O_\epsilon(x)} \phi_n^2(x-y)\left(h(y)f(y) - h(x)f(x)\right)dy \\ &\qquad+ \int_{(O_\epsilon(x))^C} \phi_n^2(x-y)\left(h(y)f(y) - h(x)f(x)\right)dy.\end{aligned}$$

So it is enough to prove that the right hand side of the last identity is of order $o\left((\log n)^{1/r}\right)$ locally uniformly. According to (2.4) and (2.6), one can bound the function $\phi_n^2(x-y)$ outside the interval $O_\epsilon(x)$ by $q_r^2(x-y)(\pi\epsilon_n)^{-2}$. Therefore, the inequality

$$\begin{aligned}&\int_{(O_\epsilon(x))^C} \phi_n^2(x-y)\left|h(y)f(y) - h(x)f(x)\right|dy \\ &\qquad\le \epsilon_n^{-2}C_1\int_{(O_\epsilon(x))^C} q_r^2(x-y)(|h(y)|+1)dy = o((\log n)^{1/r})\end{aligned}$$

holds locally uniformly due to the fact that our topology is stronger than the topology induced by the sup-norm (see Definition 2.2). Next, owing to this fact again, it is easy to see that

$$\int_{O_\epsilon(x)} \phi_n^2(x-y)\left(h(y)f(y) - h(x)f(x)\right)dy = o(1)\int \phi_n^2(y)dy$$

locally uniformly and the first equality is proved. □

LEMMA 3.2. *The functional* $\int_y^\infty \phi_n(x-u)dF(u)$ *is bounded uniformly in* y *and locally uniformly in* $F \in \mathcal{F}_\delta$.

PROOF. Fix some $\epsilon > 0$. Denote $D_1(y) = O_\epsilon(x) \cap (y, +\infty) = (b_1, b_2)$, $D_2(y) = (O_\epsilon(x))^C \cap (y, +\infty)$, where $O_\epsilon(x)$ is the open interval around x of radius ϵ. Then

$$\psi_n(y) = \int_{D_1(y)} \phi_n(x-u)dF(u) + \int_{D_2(y)} \phi_n(x-u)dF(u) = I_1 + I_2,$$

say. The second term I_2, the integral over $D_2(y)$, is clearly bounded locally uniformly. For the term I_1, we first note that for any $u \in O_\epsilon(x)$

$$|f(u) - f(x)| \leq \sup_{y \in O_\epsilon(x)} |f'(y)||u-x| \leq C_2|u-x|$$

locally uniformly since $\sup_{y \in O_\epsilon(x)} |f'(y)|$ is bounded locally uniformly (see Definition 2.2). Consequently,

$$\begin{aligned} |I_1| &\leq f(x) \left| \int_{D_1(y)} \phi_n(x-u)du \right| + C_2 \left| \int_{D_1(y)} |x-u|\phi_n(x-u)du \right| \\ &\leq C_1 + C_2\pi^{-1} \int_{D_1(y)} q_r(x-u)du \leq C_3 \end{aligned}$$

locally uniformly because $f(x)$ is also bounded locally uniformly and

$$\left| \int_{D_1(y)} \phi_n(x-u)du \right| = \left| \int_{a_n(b_1-x)}^{a_n(b_2-x)} \frac{q_r(a_n^{-1}u)\sin u}{\pi u} du \right| \leq C_4. \qquad \square$$

LEMMA 3.3. *Let the function* $h_1(u)$ *be an integrable function, let the function* $h_2(u)$ *be of bounded variation such that* $h_2(-\infty) = H_2$. *Then*

$$\iint h_1(u)h_1(v)h_2(u \wedge v)dudv = \int \left(\int_v^\infty h_1(u)du \right)^2 dh_2(v) + H_2 \left(\int h_1(u)du \right)^2,$$

provided the left hand side or the right hand side of this equality is well defined.

PROOF. Denote $H_1(u) = \int_u^\infty h_1(v)dv$. Integrating by parts twice, we obtain

$$\begin{aligned} &\iint h_1(u)h_1(v)h_2(u \wedge v)dudv \\ &= \int h_1(u) \left(\int_{-\infty}^u h_1(v)h_2(v)dv \right) du + \int h_1(u) \left(\int_u^\infty h_1(v)h_2(u)dv \right) du \\ &= \int h_1(u) \left(-\int_{-\infty}^u h_2(v)dH_1(v) \right) du + \int h_1(u)h_2(u)H_1(u)du \\ &= \int h_1(u)\left(-h_2(u)H_1(u)\right)du + H_2 \left(\int h_1(u)du \right)^2 \end{aligned}$$

$$+\int h_1(u)\left(\int_{-\infty}^{u} H_1(v)dh_2(v)\right)du+\int h_1(u)h_2(u)H_1(u)du$$
$$=\int h_1(u)\left(\int_{-\infty}^{u} H_1(v)dh_2(v)\right)du+H_2\left(\int h_1(u)du\right)^2$$
$$=-\int\left(\int_{-\infty}^{u} H_1(v)dh_2(v)\right)dH_1(u)+H_2\left(\int h_1(u)du\right)^2$$
$$=\int (H_1(u))^2dh_2(u)+H_2\left(\int h_1(u)du\right)^2. \qquad \square$$

The following result which is due to Lo *et al.* (1989) gives a representation of the Kaplan-Meier estimator as an average of independent random variables plus a remainder term. First introduce some notations:

$$\begin{aligned} &g(y)=\int_{-\infty}^{y}\frac{dF(u)}{(\bar F(u))^2\bar G(u-)},\\ &\xi_i(t)=\xi(Z_i,\Delta_i,t)=-\bar F(t)g(Z_i\wedge t)+\frac{\bar F(t)}{\bar H(t)}I\{Z_i\le t,\Delta_i=1\}, \end{aligned} \tag{3.1}$$

where H is the distribution function of Z_1.

LEMMA 3.4. (Lo *et al.* (1989)). *Let F be continuous. Then*

$$\tilde F_n(y)=F(y)+\frac{1}{n}\sum_{i=1}^{n}\xi_i(y)+R_n(y),$$

where for any $T<\tau_F\wedge\tau_G$ and any $\alpha\ge 1$

$$\sup_{y\le T}E|R_n(y)|^\alpha=O((\log n/n)^\alpha)\qquad as\quad n\to\infty.$$

Remark 3.1. Actually, the result of Lo *et al.* (1989) concerns the case of nonnegative "lifetimes" $X_1,\dots,X_n$. It is however a straightforward matter to extend this to any continuous distribution function F.

Remark 3.2. Tracing the proof of this lemma, one can show that this representation holds locally uniformly over a sufficiently small neighbourhood of any F such that $T<\tau_G\wedge\tau_F$, in the topology generated by the distance in variation.

Remark 3.3. Note that the random variables $\xi_i(y)$, $i=1,\dots,n$, are bounded uniformly in $y\le T$, independent and, by routine calculations,

$$E\xi_i(y)=0,\qquad E\left(\xi_i(y)\xi_i(u)\right)=\bar F(y)\bar F(u)g(y\wedge u). \tag{3.2}$$

4. Approximation lemma

The following lemma is of particular importance. It reflects the fact that each function from the class $\mathcal{F}_\delta$ can be approximated with a negligible error by a sequence of "smooth functionals", which exhibits a close resemblance of our density estimation problem with the problem of estimating a smooth functional.

LEMMA 4.1. (APPROXIMATION LEMMA) *As $n \to \infty$, the relation*

$$\left(\int \phi_n(x-y)dF(y) - f(x)\right)^2 = O\left(n^{-1}\right)$$

holds uniformly over $\mathcal{F}_\delta$.

PROOF. Recalling Definition 2.1, we obtain the following uniform bound:

$$\begin{aligned}
&\left(\int \phi_n(x-y)dF(y) - f(x)\right)^2 \\
&= \left(\frac{1}{2\pi}\int e^{-itx}(\hat{\phi}_n(t)-1)\hat{f}(t)dt\right)^2 \\
&\le \frac{1}{2\pi}\int \exp\{2\delta|t|^r\}|\hat{f}(t)|^2 dt \cdot \frac{1}{2\pi}\int \exp\{-2\delta|t|^r\}|\hat{\phi}_n(t)-1|^2 dt \\
&\le C_1 \int |\hat{\phi}_n(t)-1|^2 \exp\{-2\delta|t|^r\}dt \\
&\le C_1 . \int_{-a_n}^{a_n} |\hat{\phi}_n(t)-1|^2 \exp\{-2\delta|t|^r\}dt + C_2\int_{|t|\ge a_n} \exp\{-2\delta|t|^r\}dt \\
&= 2C_1 \int_0^{a_n} |\hat{\phi}_n(t)-1|^2 \exp\{-2\delta|t|^r\}dt + O\left(n^{-1}\right)
\end{aligned}$$

since, according to Gradshtein and Ryzhik (1980), equation 2.621, and condition (iii),

$$\begin{aligned}
&\int_{|t|\ge a_n} \exp(-2\delta|t|^r)dt \\
&= 2\int_{a_n}^{\infty} \exp(-2\delta t^r)dt = \frac{1}{\delta r}\int_{\exp\{2\delta a_n^r\}}^{\infty} \left(\frac{\log u}{2\delta}\right)^{(1-r)/r} \frac{du}{u^2} \\
&= \frac{\exp\{-2\delta a_n^r\}}{r\delta a_n^{r-1}}(1+o(1)) = \frac{2(\log n)^{(1-r)/r}}{r(2\delta)^{1/r} n(\log n)^m}(1+o(1)) = O\left(n^{-1}\right).
\end{aligned}$$

Hence it suffices to prove that

$$\int_0^{a_n} |\hat{\phi}_n(t)-1|^2 \exp\{-2\delta|t|^r\} = O(n^{-1}). \tag{4.1}$$

Since the function $\hat{q}_r(u)$ is even,

$$\int_{u>t+a_n} |\hat{q}_r(u)|du \le \int_{u<t-a_n} |\hat{q}_r(u)|du = \int_{u>a_n-t} |\hat{q}_r(u)|du$$

for $t \in [0, a_n]$. By (2.5), we have also that $\int \hat{q}_r(u)du = q_r(0)2\pi = 2\pi$. By using the last two relations, (2.9) and (2.10), we obtain that, for $t \in [0, a_n]$,

$$\begin{aligned}
2\pi|\hat{\phi}_n(t)-1| &= |(\hat{q}_r * I_{(-a_n,a_n)})(t) - 1| \\
&= \left|\int (I_{(-a_n,a_n)}(t-u)-1)\hat{q}_r(u)du\right| = \left|\int_{|t-u|>a_n} \hat{q}_r(u)du\right| \\
&\le 2\int_{u>a_n-t} |\hat{q}_r(u)|du \le C_3\int_{a_n-t}^{\infty} \exp\{-A_2 u^{\beta/(\beta+1)}\}du \\
&\le C_4 \exp\{-C_5(a_n-t)^{\beta/(\beta+1)}\}.
\end{aligned} \tag{4.2}$$

The last inequality follows from the following relations: for $t \in [a_n - 1, a_n]$

$$\int_{a_n-t}^{\infty} \exp\{-A_2 u^{\beta/(\beta+1)}\}du \le B_1 \le B_1 e^{A_2} \exp\{-A_2(a_n - t)^{\beta/(\beta+1)}\},$$

and for $t \in [0, a_n - 1)$

$$\int_{a_n-t}^{\infty} \exp\{-A_2 u^{\beta/(\beta+1)}\}du \le B_2 \int_{a_n-t}^{\infty} u^{-1/(\beta+1)} \exp\left\{-\frac{A_2}{2} u^{\beta/(\beta+1)}\right\} du$$
$$= B_3 \exp\left\{-\frac{A_2}{2}(a_n - t)^{\beta/(\beta+1)}\right\}.$$

Recall now the c_r-inequality (see Loève (1963), p. 155):

$$|h_1 + h_2|^r \le c_r|h_1|^r + c_r|h_2|^r, \quad r > 0, \tag{4.3}$$

where $c_r = 1$ or $c_r = 2^{r-1}$ according as $r \le 1$ or $r > 1$. So, we prove (4.1) by combining (4.2) with the c_r-inequality and the fact that $\beta/(\beta + 1) > r$ (see condition (ii)):

$$\begin{aligned}
&\int_0^{a_n} |\hat{\phi}_n(t) - 1|^2 \exp\{-2\delta|t|^r\}dt \\
&\le C_6 \int_0^{a_n} \exp\{-2C_5(a_n - t)^{\beta/(\beta+1)} - 2\delta t^r\}dt \\
&= C_6 \int_0^{a_n} \exp\{-2C_5 t^{\beta/(\beta+1)} - 2\delta(a_n - t)^r\}dt \\
&\le C_6 \int_0^{a_n} \exp\{-2C_5 t^{\beta/(\beta+1)} - 2\delta(a_n^r - t^r)\}dt \\
&\le C_6 e^{-2\delta a_n^r} \int_0^{\infty} \exp\{-2C_5 t^{\beta/(\beta+1)} + 2\delta t^r\}dt = \frac{C_7}{n(\log n)^m}
\end{aligned} \tag{4.4}$$

uniformly over $\mathcal{F}_\delta$. □

Remark 4.1. As one can see from the proof of this lemma, a stronger bound on the approximation error is in fact valid. Namely, the relation

$$\left(\int \phi_n(x - y)dF(y) - f(x)\right)^2 = O\left(\frac{1}{n(\log n)^{m+1-1/r}}\right)$$

holds uniformly over $\mathcal{F}_\delta$. Thus, we can make the error of approximation smaller by choosing a larger m in (2.3).

Remark 4.2. Certainly, the proof of this approximation property is almost trivial if $\phi_n(y) = s_n(y)$, where the function s_n is defined by (2.4) (cf. also Ibragimov and Hasminskii (1982)). Let us clarify why this is a bad choice of the kernel function for the estimator (2.7). The risk of the estimator is bounded from above by a sum of two terms (see the proof of Theorem 2.1) which we call the approximation term and the stochastic term. The first term is analogous to the bias term in the noncensored case and comes from the approximation error. The second term has a stochastic origin and is analogous to the variance of an estimator in the noncensored case. So, choosing $\phi_n(y) = s_n(y)$ provides a small approximation error, while leading to a bigger stochastic term since this function is badly localized in the time domain. The idea is to find a proper localizing factor, the function $q_r(y)$, such that both the stochastic term becomes smaller and the approximation property remains valid.

5. Proofs of theorems

PROOF OF THEOREM 2.1. Without loss of generality we suppose that the constant b in (2.2) is chosen in such a way that $x + b < \tau_{F_0} \wedge \tau_G$; see Remark 2.8. Then $x + b < \tau_F \wedge \tau_G$ uniformly over a sufficiently small neighbourhood of f_0 since our topology is stronger than the topology induced by the distance in variation; see Definition 2.2. Now using integration by parts, Lemma 3.4 (see also Remark 3.2 and (3.2)) and the elementary inequality

$$(a+b)^2 \le (1+\gamma)a^2 + (1+\gamma^{-1})b^2, \qquad 0 < \gamma \le 1, \tag{5.1}$$

we have that, uniformly over a sufficiently small neighbourhood of f_0,

$$\begin{aligned} E_f \left(\int \phi_n(x-y) d(\tilde{F}_n(y) - F(y)) \right)^2 & \\ = E_f \left(\int (\tilde{F}_n(y) - F(y)) d\phi_n(x-y) \right)^2 & \\ \le \frac{(1+\gamma_n)}{n} \iint \bar{F}(t)\bar{F}(u) g(u \wedge t) d\phi_n(x-t) d\phi_n(x-u) & \\ +(1+\gamma_n^{-1}) E_f \int \left(R_n(y) \phi_n'(x-y) \right)^2 dy & \\ \le \frac{(1+\gamma_n)}{n} \iint \bar{F}(t)\bar{F}(u) g(u \wedge t) d\phi_n(x-t) d\phi_n(x-u) & \\ +(1+\gamma_n^{-1}) C_1 n^{-2} (\log n)^{2+2/r}, & \end{aligned} \tag{5.2}$$

where g is defined by (3.1) and γ_n is to be chosen later. We can apply Lemma 3.4 because the kernel $\phi_n(x-y)$ has finite support $[x-b, x+b]$ such that $x + b < \tau_G \wedge \tau_F$ uniformly in a neighbourhood of f_0.

Tedious but straightforward calculations lead to

$$\begin{aligned} &\iint \bar{F}(t)\bar{F}(u) g(u \wedge t) d\phi_n(x-t) d\phi_n(x-u) \\ &\quad = \int \frac{\phi_n^2(x-t) dF(t)}{1 - G(t-)} + \iint \phi_n(x-t)\phi_n(x-u) h(t \wedge u) dF(u) dF(t), \end{aligned}$$

where

$$h(y) = \int_{-\infty}^{y} \frac{dF(u)}{(\bar{F}(u))^2 \bar{G}(u-)} - \frac{1}{\bar{F}(y)\bar{G}(y-)}.$$

By Lemma 3.3, we have

$$\begin{aligned} &\iint \phi_n(x-t)\phi_n(x-u) h(t \wedge u) dF(u) dF(t) \\ &\quad = \int \left(\int_t^{\infty} \phi_n(x-u) dF(u) \right)^2 dh(t) - \left(\int \phi_n(x-u) dF(u) \right)^2 \\ &\quad = -\int \left(\int_t^{\infty} \phi_n(x-u) dF(u) \right)^2 \frac{dG(t-)}{\bar{F}(t)(\bar{G}(t-))^2} - \left(\int \phi_n(x-u) dF(u) \right)^2. \end{aligned}$$

Therefore,

$$\iint \bar{F}(t)\bar{F}(u)g(u \wedge t)d\phi_n(x-t)d\phi_n(x-u) \le \int \frac{\phi_n^2(x-t)dF(t)}{1-G(t-)}.$$

Now we evaluate the risk of the estimator (2.7). From the last relation, (5.2) and again the elementary inequality (5.1) it follows that

$$\begin{aligned} &E_f(\tilde{f}_n(x) - f(x))^2 \\ &= E_f\left(\int \phi_n(x-y)d(\tilde{F}_n(y) - F(y)) + \int \phi_n(x-y)dF(y) - f(x)\right)^2 \\ &\le \frac{(1+\gamma_n)^2}{n}\int \frac{\phi_n^2(x-t)dF(t)}{1-G(t-)} + \frac{(\gamma_n^{-1}+2+\gamma_n)C_1(\log n)^{2+2/r}}{n^2} \\ &+(1+\gamma_n^{-1})\left(\int \phi_n(x-y)dF(y) - f(x)\right)^2 \end{aligned}$$

uniformly in a neighbourhood of f_0. Next, choose γ_n in such a way that $\gamma_n \to 0$ and $\gamma_n(\log n)^{1/r} \to \infty$ as $n \to \infty$. Using the last relation, Lemma 3.1 and Approximation Lemma 4.1, we obtain that

$$\limsup_{n\to\infty} n(\log n)^{-1/r}E_f(\tilde{f}_n(x) - f(x))^2 \le \frac{f(x)}{(2\delta)^{1/r}\pi\bar{G}(x)}$$

uniformly over a sufficiently small neighbourhood of f_0. □

PROOF OF THEOREM 2.2. Let $f_0(y)$ be an arbitrary density from the neighbourhood V and let F_0 be the corresponding distribution function. Consider the following family of densities:

$$f_\theta(y) = f_\theta(y, x, \phi_n, f_0) = f_0(y)(1 + \theta[\phi_n(x-y) - \bar{\phi}_n(x)]),$$

where $|\theta| \le \theta_n$,

$$\bar{\phi}_n(x) = \int \phi_n(x-y)f_0(y)dy$$

and $\phi_n(y)$ is defined by (2.6) with

$$a_n = a_n(m,\delta,r) = \left(\frac{\log n - m\log\log n}{2\delta}\right)^{1/r} \tag{5.3}$$

instead of a_n defined by (2.3). Let θ_n be such that $\epsilon_n \le \theta_n \le \rho_n$, where the positive sequences ϵ_n and ρ_n satisfy

$$\frac{1}{\epsilon_n^2 n(\log n)^{1/r}} = o(1), \qquad \rho_n^2 n = o(1).$$

One can choose for example $\theta_n = n^{-1/2}(\log n)^{-1/(4r)}$.

The proof of the theorem will proceed via the following two claims.

PROPOSITION 5.1. *For sufficiently large n, $f_\theta \in V$.*

PROOF. Take $\epsilon > 0$ such that $O_\epsilon(f_0) \subseteq V$, where $O_\epsilon(f_0) = \{f \in \mathcal{F}_\delta : \rho(f, f_0) < \epsilon\}$. We prove now that $f_\theta \in O_\epsilon(f_0)$ for sufficiently large n where $\rho = \rho_u$ generates the strong topology $\mathcal{U}_\delta$.

Denote $\psi(y) = \psi(y,x) = f_0(y)\phi_n(x-y)$. First, by the Minkowski inequality, we have

$$\begin{aligned}
\rho_u(f_\theta, f_0) &= |\theta| \left(\int \exp\{2\delta|t|^r\} |\hat{\psi}(t,x) - \bar{\phi}_n(x)\hat{f}_0(t)|^2 dt \right)^{1/2} \\
&\quad + |\theta| \int |f_0(y)(\phi_n(x-y) - \bar{\phi}_n(x))| dy \\
&\le 2\theta_n \left(\int \exp\{2\delta|t|^r\} |\hat{\psi}(t,x)|^2 dt \right)^{1/2} \\
&\quad + 2\theta_n |\bar{\phi}_n(x)| \left(\int \exp\{2\delta|t|^r\} |\hat{f}_0(t)|^2 dt \right)^{1/2} \\
&\quad + \theta_n \int |f_0(y)(\phi_n(x-y) - \bar{\phi}_n(x))| dy \\
&\le 2\theta_n \left(\int \exp\{2\delta|t|^r\} |\hat{\psi}(t,x)|^2 dt \right)^{1/2} + C_1 \theta_n (\log n)^{1/r}.
\end{aligned}$$

Since $\theta_n = o(n^{-1/2})$, it suffices to show that the first term on the right hand side of the last inequality converges to zero as $n \to \infty$.

Note that
$$\hat{\psi}(t,x) = (2\pi)^{-1} \int e^{ixu} \hat{f}_0(t+u)\hat{\phi}_n(u) du.$$

Using the generalized Minkowski inequality (see Nikol'skii (1975), p. 20), Definition 2.1, property (2.9), and the c_r-inequality (4.3), we obtain that

$$\begin{aligned}
&\left(\int \exp\{2\delta|t|^r\} |\hat{\psi}(t,x)|^2 dt \right)^{1/2} \\
&\le C_2 \left(\int \left| \int \exp\{\delta|t|^r\} e^{ixu} \hat{f}_0(t+u)\hat{\phi}_n(u) du \right|^2 dt \right)^{1/2} \\
&\le C_2 \int \left(\int \left| \exp\{\delta|t|^r\} e^{ixu} \hat{f}_0(t+u)\hat{\phi}_n(u) \right|^2 dt \right)^{1/2} du \\
&\le C_2 \int \left(\int \left| \exp\{\delta|t+u|^r\} \hat{f}_0(t+u) \exp\{\delta|u|^r\} \hat{\phi}_n(u) \right|^2 dt \right)^{1/2} du \\
&\le C_3 \left(\int \exp\{2\delta|t|^r\} |\hat{f}_0(t)|^2 dt \right)^{1/2} \int \exp\{\delta|u|^r\} |\hat{\phi}_n(u)| du \\
&\le C_4 \int \exp\{\delta|u|^r\} |\hat{\phi}_n(u)| du \\
&\le C_4 \Big(\int \exp\{2\delta|u|^r\} \left| \left(\hat{q}_r * I_{(-a_n,a_n)} \right)(u) \right|^2 du \Big)^{1/2} \\
&= C_4 \Big(\int \left| \int \exp\{\delta|u|^r\} \hat{q}_r(u-t) I_{(-a_n,a_n)}(t) dt \right|^2 du \Big)^{1/2}
\end{aligned}$$

$$\le C_4 \int (\int \left|\exp\{\delta|u|^r\}\hat{q}_r(u-t)I_{(-a_n,a_n)}(t)\right|^2 du)^{1/2}dt$$
$$\le C_5(\int \exp\{2\delta|t|^r\}|\hat{q}_r(t)|^2dt)^{1/2}\int e^{\delta|u|^r}I_{(-a_n,a_n)}(u)du$$
$$\le C_6 n^{1/2}(\int \exp\{2\delta|t|^r\}|\hat{q}_r(t)|^2dt)^{1/2}$$

since, by equation 2.621 from Gradshtein and Ryzhik (1980), (5.3) and condition (iii),

$$\int e^{\delta|u|^r}I_{(-a_n,a_n)}(u)du = \frac{2}{\delta r}\int_1^{\exp\{\delta a_n^r\}}\left(\frac{\log u}{\delta}\right)^{(1-r)/r}du = \frac{\exp\{\delta a_n^r\}a_n^{1-r}}{\delta r}(1+o(1))$$
$$= \frac{2n^{1/2}(\log n)^{1/r-1}}{r(2\delta)^{1/r}(\log n)^{m/2}}(1+o(1)) \le C_7\sqrt{n}.$$

Now evaluate, by (2.10),

$$\int e^{2\delta|t|^r}|\hat{q}_r(t)|^2dt \le A_1\int \exp\{2\delta|t|^r - 2A_2|t|^{\beta/(\beta+1)}\}dt \le C_8.$$

Recalling the condition on the θ_n, we finally obtain that

$$\rho_u(f_\theta, f_0) \le C_9\theta_n n^{1/2} + C_1\theta_n(\log n)^{1/r}$$
$$\le C_9\rho_n n^{1/2} + C_1\rho_n(\log n)^{1/r} = o(1)$$

as $n \to \infty$. □

If X_i is distributed with density $f_\theta(y)$, then the corresponding observation (Z_i, Δ_i) has the density

$$f_\theta(y,\tau) = (f_\theta(y)(1-G(y)))^\tau (g(y)(1-F_\theta(y)))^{1-\tau}, \qquad \tau \in \{0,1\}.$$

Let $I(\theta)$ be the Fisher information about θ contained in the observation (Z, Δ), i.e.

$$I(\theta) = E_f\left[\frac{\partial \log f_\theta(Z,\Delta)}{\partial\theta}\right]^2.$$

PROPOSITION 5.2. *As $n \to \infty$,*

$$\sup_{|\theta|<\theta_n} I(\theta) = \frac{f_0(x)\bar{G}(x)}{\pi}\left(\frac{\log n}{2\delta}\right)^{1/r}(1+o(1)).$$

PROOF. By straightforward calculations,

$$I(\theta) = \int \frac{(\phi_n(x-y)-\bar{\phi}_n(x))^2 f_0(y)(1-G(y))dy}{1+\theta(\phi_n(x-y)-\bar{\phi}_n(x))}$$
$$+\int \frac{\{\int_{-\infty}^y f_0(u)(\phi_n(x-u)-\bar{\phi}_n(x))du\}^2 dG(y)}{1-F_\theta(y)} = I_1 + I_2,$$

say. Split the second term in the right hand side of the last inequality into two parts: the integral over $(-\infty, x+\epsilon]$ and the integral over $(x+\epsilon, \infty)$:

$$\int \frac{\{\int_{-\infty}^{y} f_0(u)(\phi_n(x-u) - \bar{\phi}_n(x))du\}^2 dG(y)}{1 - F_\theta(y)} = \int_{-\infty}^{x+\epsilon} + \int_{x+\epsilon}^{\infty} = I_{21} + I_{22},$$

say. The integral $\int_{-\infty}^{y} f_0(u)(\phi_n(x-u) - \bar{\phi}_n(x))du$ is bounded for $y \in (-\infty, x+\epsilon]$ by Lemma 3.2 and Approximation Lemma 4.1. Obviously,

$$1 - F_\theta(y) = (1 - F_0(y))(1 + o(1)) \geq (1 - F_0(x+\epsilon))(1 + o(1))$$

for $y \in (-\infty, x+\epsilon]$. Therefore, the integral I_{21} is bounded.

Further note that

$$\int_{-\infty}^{y} f_0(u)(\phi_n(x-u) - \bar{\phi}_n(x))du = -\int_{y}^{\infty} f_0(u)(\phi_n(x-u) - \bar{\phi}_n(x))du$$

and the function $\phi_n(x-y)$ is bounded for $y \in (x+\epsilon, \infty)$. Therefore, for $y \in (x+\epsilon, \infty)$ and sufficiently large n,

$$\begin{aligned}
&\frac{\{\int_{-\infty}^{y} f_0(u)(\phi_n(x-u) - \bar{\phi}_n(x))du\}^2}{1 - F_\theta(y)} \\
&= \frac{\{\int_{y}^{\infty} f_0(u)(\phi_n(x-u) - \bar{\phi}_n(x))du\}^2}{1 - F_\theta(y)} \leq \frac{C_1\{\int_{y}^{\infty} f_0(u)du\}^2}{1 - F_\theta(y)} \\
&= \frac{C_1(1 - F_0(y))^2}{(1 - F_0(y))(1 + o(1))} \leq C_2(1 - F_0(y)) \leq C_2.
\end{aligned}$$

Thus, we obtained that I_{22} is also bounded and consequently I_2 is bounded uniformly in θ, $|\theta| < \theta_n$, for sufficiently large n.

According to Lemmas 3.1 and 3.2, it is not difficult to see that

$$I_1 = \int \phi_n^2(x-y) f_0(y) \bar{G}(y) dy (1 + o(1)) = \frac{f_0(x)\bar{G}(x)}{\pi} \left(\frac{\log n}{2\delta}\right)^{1/r} (1 + o(1))$$

uniformly in θ, $|\theta| < \theta_n$. The proposition is proved. □

Now we proceed to prove the theorem. Introduce $\nu(x) = \theta_n^{-1}\nu_0(\theta_n^{-1}x)$, where $\nu_0(x)$ is a probability density on the interval $[-1, 1]$ such that $\nu_0(-1) = \nu_0(1) = 0$, $\nu_0(x)$ is continuously differentiable for $|x| < 1$ and it has finite Fisher information

$$I_0 = \int_{-1}^{1} (\nu_0'(x))^2 \nu_0^{-1}(x) dx.$$

The function $\nu(x)$ is a probability density with support $[-\theta_n, \theta_n]$. It is easy to calculate the Fisher information of the distribution defined by density $\nu(x) : I(\nu) = I_0\theta_n^{-2}$. Under these conditions, one can apply the van Trees inequality for the Bayes risk below (see Gill and Levit (1995)): for any estimator $\tilde{f}_n$,

$$\int E_{f_\theta}(\tilde{f}_n - f_\theta(x))^2 \nu(\theta) d\theta \geq \frac{\left(\int (\partial f_\theta(x)/\partial\theta)\, \nu(\theta) d\theta\right)^2}{n \int I(\theta)\nu(\theta)d\theta + I(\nu)}.$$

Using this, Propositions 5.1 and 5.2, we obtain that

$$\begin{aligned} r_n(V) &= \inf_{\tilde{f}_n} \sup_{f \in V} E_f(\tilde{f}_n - f(x))^2 \geq \inf_{\tilde{f}_n} \sup_{|\theta| \leq \theta_n} E_{f_\theta}(\tilde{f}_n - f_\theta(x))^2 \\ &\geq \inf_{\tilde{f}_n} \int E_{f_\theta}(\tilde{f}_n - f_\theta(x))^2 \nu(\theta) d\theta \geq \frac{(\int (\partial f_\theta(x)/\partial\theta)\nu(\theta)d\theta)^2}{n \int I(\theta)\nu(\theta)d\theta + I(\nu)} \\ &= \frac{(f_0(x)(\phi_n(0) - \bar{\phi}_n(x)))^2}{n \int I(\theta)\nu(\theta)d\theta + I_0\theta_n^{-2}} \\ &\geq \frac{(f_0(x)/\pi)^2(\log n/(2\delta))^{2/r}(1+o(1))}{nf_0(x)\bar{G}(x)(\log n/(2\delta))^{1/r}\pi^{-1}(1+o(1)) + I_0\epsilon_n^{-2}} \\ &\geq \frac{f_0(x)}{n\pi\bar{G}(x)} \left(\frac{\log n}{2\delta}\right)^{1/r} (1+o(1)) \end{aligned}$$

as $n \to \infty$. This implies that

$$\liminf_{n\to\infty} n(\log n)^{-1/r} r_n(V) \geq \frac{f_0(x)}{(2\delta)^{1/r}\pi\bar{G}(x)} = \sigma^2(f_0).$$

The function f_0 was chosen arbitrarily from the neighbourhood V and hence, by the same reasoning, this relation is valid for any function $f \in V$:

$$\liminf_{n\to\infty} n(\log n)^{-1/r} r_n(V) \geq \sigma^2(f).$$

Therefore

$$\liminf_{n\to\infty} n(\log n)^{-1/r} r_n(V) \geq \sup_{f\in V} \sigma^2(f),$$

which proves the theorem. □

References

Andersen, P. K., Borgan, O. Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer, New York.

Belitser, E. (1998). Efficient estimation of analytic density under random censorship, *Bernoulli*, **4**, 519–543.

Davis, K. B. (1975). Mean square error properties of density estimates, *Ann. Statist.*, **3**, 1025–1030.

Fedoruk, M. V. (1977). *Metod Perevala*, Nauka, Moscow (in Russian).

Gill, R. D. and Levit, B. Y. (1995). Applications of the van Trees inequality: A Bayesian Cramér-Rao bound, *Bernoulli*, **1**, 59–79.

Golubev, G. K. and Levit, B. Y. (1996). Asymptotically efficient estimation for analytic distributions, *Math. Methods Statist.*, **5**(3), 357–368.

Gradshtein, I. S. and Ryzhik, I. M. (1980). *Table of Integrals, Series, and Products*, Academic Press, New York.

Huang, J. and Wellner, J. A. (1995). Estimation of a monotone density and monotone hazard under random censoring, *Scand. J. Statist.*, **22**, 3–33.

Ibragimov, I. A. and Hasminskii, R. Z. (1982). Estimation of distribution density belonging to a class of entire functions, *Theory Probab. Appl.*, **27**, 551–562.

Konakov, V. D. (1972). Non parametric estimation of density functions, *Theory Probab. Appl.*, **17**, 377–379.

Kulasekera, K. B. (1995). A bound on the L_1-error of a nonparametric density estimator with censored data, *Statist. Probab. Lett.*, **23**, 233–238.

Lo, S. H., Mack, Y. P. and Wang, J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator, *Probab. Theory Related. Fields*, **80**, 461–473.

Loève, M. (1963). *Probability Theory*, 3rd ed., Van Nostrand Reinhold, New York.

Mielniczuk, J. (1986). Some asymptotic properties of kernel estimators of a density function in case of censored data, *Ann. Statist.*, **14**, 766–773.

Nikol'skii, S. M. (1975). *Approximation of Functions of Several Variables and Imbedding Theorems*, Springer, Berlin–Heidelberg–New York.

Weits, E. (1993). The second order optimality of a smoothed Kaplan-Meier estimator, *Scand. J. Statist.* **20**, 111–132.