Efficient Estimation of Analytic Density under Random Censorship

Eduard Belitser

Mathematical Institute, University of Utrecht, Budapestlaan 6, 3584 CD Utrecht, Netherlands

The nonparametric minimax estimation of an analytic density at a given point, under random censorship, is considered. Although the problem of estimating density is known to be irregular in a certain sense, we make some connections relating this problem to the problem of estimating smooth functionals. Under condition that the censoring is not too severe, we establish the exact limiting behavior of the local minimax risk and propose the efficient (locally asymptotically minimax) estimator – an integral of some kernel with respect to the Kaplan-Meier estimator.

Keywords: asymptotic local minimax risk, density estimation, Kaplan-Meier estimator, kernel, random censorship.

1 Introduction

Let $(X_1, Y_1), \ldots (X_n, Y_n)$ be independent identically distributed pairs of random variables where X_1 and Y_1 are independent and have the distribution functions F and G respectively. We assume also that the distribution of X_1 is absolutely continuous with density f. The following model is known as *random censorship model*. We observe only the pairs (Z_i, Δ_i), $i = 1, 2, \ldots, n$, with $Z_i = \min(X_i, Y_i)$ and $\Delta_i = I\{X_i \leq Y_i\}$. In survival analysis the X_i's are called survival times and the Y_i's censoring times. Estimation problems with censored observations arise often in lifetime research. In survival analysis lifetime data are typically subject to censoring. We suppose F and G are unknown and our goal is, using the observed data, to estimate the density $f(x)$ at a given point x.

The problem of density estimation under random censorship has been treated by a number of authors (see for example Diehl and Stute (1988), Mielniczuk (1986), Lo, Mack and Wang (1989), Hentzschel (1992), Kulasekera (1995), Huang and Wellner (1995), Liu (1996)). In Hentzschel (1992) the estimator based on the orthonormal system of the Laguerre series on the positive line is investigated and under some assumptions the rates of the mean integrated square error and the mean square error are obtained. In Kulasekera (1995) upper $\mathbf{L}_1$-bounds for the kernel-type estimator are given for two classes of densities: monotonically decreasing densities on $[0, \infty)$ and densities which are of bounded variation on $[0, 1]$. For a decreasing density function, Huang and Wellner (1995) showed that the nonparametric maximum likelihood estimator of the density is asymptotically equivalent to the estimator obtained by differentiating the least concave majorant of the Kaplan-Meier estimator and established the asymptotic distributions of the different estimators at a fixed point.

However, in all the above mentioned papers the question of optimality has not been touched on. In a recent paper Liu (1996) considered the general problem of estimating functionals of a distribution F for some nonparametric classes defined in terms of the Hellinger modulus of continuity. With regard to density estimation, minimax Kaplan-Meier based kernel procedures were shown, under some conditions, to be of optimal rate and, moreover, within certain lower and upper bounds.

In the nonparametric minimax estimation context the notion of asymptotic optimality is usually associated with the optimal rate of convergence of the minimax risk. In order to derive the exact asymptotics of the minimax risk and to be able to compare the estimators with the optimal rate of convergence, one may strengthen the optimal rate results by finding optimal constants when they exist. Results about the optimal constants in minimax density estimation have only been obtained in a limited number of works for models with independent identically distributed observations. The majority of authors has considered the global minimax risk. However, studying the so called local minimax risk yields more exact results. We mention the work of Golubev and Levit (1996) whose results motivated the present study. In the problem of estimation of an analytic density at a given point, with independent identically distributed observations, they derived the exact limiting behavior of the local minimax risk and proposed an efficient estimator.

To elucidate the ideas of the results, we give here some heuristic arguments. The unknown underlying density f is assumed to belong to the class of densities with exponentially decreasing Fourier transforms (analytic densities). This nonparametric class has the advantage that one can treat the problem of estimating $f(x)$ as if a smooth functional was to be estimated. In particular, it turns out that any density from this nonparametric class can be represented in the following asymptotic form (see Lemma 5 below):

$$f(x) = \int \phi_n(x-y)f(y)dy + O(n^{-1/2}), \qquad \text{as} \quad n \to \infty,$$

locally uniformly in f in a proper topology, where $\phi_n(y)$ is some sequence of functions (see the exact definitions in the next section) which we will call *kernel* treating this notion in a broader sense than is usual in the literature. The local minimax quadratic risk proves to be of order $\log n/n$ and therefore the remainder term can be neglected, while the first term resembles a "smooth" functional (it is a sequence of functionals because of its dependence on n) to which one can apply well developed methods for deriving an optimal estimator and its asymptotic behavior. So, in case there is no censoring one can expect the estimator $\tilde{f}_n(x) = \int \phi_n(x-y)dF_n(y)$, with the empirical distribution function F_n, to be optimal in some sense. If for independent identically distributed observations the estimator of the density is some functional of the empirical distribution function $T(F_n)$, then in the case of censoring one tries usually to use the estimator $T(\tilde{F}_n)$, with the well known Kaplan-Meier estimator $\tilde{F}_n(y)$ (see below) instead of the empirical distribution function F_n. Thus, it is natural to propose the estimator

$$\tilde{f}_n = \tilde{f}_n(x) = \int \phi_n(x-y)d\tilde{F}_n(y). \tag{1}$$

In this paper we establish, under condition that the censoring is not too severe, the exact limiting behavior of the local minimax risk up to a constant and show that the estimator of the form (1), with a properly chosen kernel, is locally asymptotically efficient. We emphasize here that the choice of nonparametric class (analytic densities) has made it possible. We propose a wide class of kernels on which the estimator can be based, which turns out to be important in the estimation problem with censored observations. Using the martingale approach enables us to derive the exact upper bound for the local minimax risk. The lower bound for the local minimax risk is based on the elementary van Trees inequality (Gill and Levit (1995)).

2 Definitions and main results

In this section we summarize the main results. First we define the notion of efficiency.

Prior information about an unknown density f is usually formalized by assuming $f \in \mathcal{F}$, for some class of densities $\mathcal{F}$. Suppose now that we have some topology on $\mathcal{F}$. For each neighbourhood V define the *local minimax risk*:

$$r_n(V) = r_n(V, x) = \inf_{\tilde{f}_n} \sup_{f \in V} \mathbf{E}_f(\tilde{f}_n((x) - f(x))^2, \tag{2}$$

where the infimum is taken over all estimators $\tilde{f}_n$. The most convenient and natural way to introduce the notion of efficiency is the following: the estimator $\tilde{f}_n$ is called *asymptotically efficient* (or just *efficient*) if for each density $f \in \mathcal{F}$ 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), for some positive sequence β_n,

$$\limsup_{n\to\infty} \beta_n^{-2} \left(\sup_{f\in V} \mathbf{E}_f(\tilde{f}_n(x) - f(x))^2 - r_n(V) \right) = 0$$

while

$$\liminf_{n\to\infty} \beta_n^{-2} r_n(V) > 0.$$

The sequence β_n is called the *minimax rate of convergence*. Note also that one can write lim instead of limsup. If the limit

$$\lim_{V\downarrow f_0} \lim_{n\to\infty} \beta_n^{-2} r_n(V) = \sigma^2(f_0)$$

exists, then, $\sigma(f_0)$, together with β_n, describes the exact behavior of the local minimax risk and represents in a way the difficulty of the estimation problem at the point f_0.

Denote from now on the Fourier transform of an absolutely integrable function f by $\hat{f}$:

$$\hat{f}(t) = \int e^{ity} f(y)dy.$$

Define now the nonparametric class $\mathcal{F}_\delta$ of underlying densities.

Definition. For given $P, \delta > 0$, denote

$$\mathcal{F}_\delta = \mathcal{F}_\delta(P) = \left\{ f(\cdot) : \ (2\pi)^{-1} \int \cosh^2(\delta t) |\hat{f}(t)|^2 dt \le P \right\}. \tag{3}$$

Remark 1. This is a class of analytic functions. Below we describe it more precisely. Let the class $\mathcal{A}_\delta = \mathcal{A}_\delta(P)$ consist of functions admitting bounded analytic continuation into the strip $\{y + iu, \ |u| \le \delta\}$ and $\int |f(y + i\delta)|^2 dy \le P < \infty$. In case there is no censoring the nonparametric classes of the type $\mathcal{A}_\delta$ were considered first in Ibragimov and Hasminski (1983), where the minimax rates of convergence in $\mathbf{L}_p$ were derived. There is a close relationship between the classes $\mathcal{F}_\delta$ and $\mathcal{A}_\delta$: for any β, $0 < \beta < \delta$, there exists $Q > 0$ such that

$$\mathcal{A}_\delta(P) \subseteq \mathcal{F}_\delta(P) \subseteq \mathcal{A}_{\delta-\beta}(Q).$$

Indeed, if a density $f \in \mathcal{A}_\delta$, then, according to Timan (1963, p. 137), the limit

$$\lim_{u \to \delta} \mathrm{Re} f(y + iu) = g_f(y)$$

exists for almost all y and $f(y)$ can be represented as a convolution:

$$f(y) = \frac{1}{2\delta} \int \left(\cosh\left(\frac{\pi(y-u)}{2\delta} \right) \right)^{-1} g_f(u) du.$$

Furthermore, because of the relation (see for example Gradshtein and Ryzhik (1980), 3.983.1)

$$\frac{1}{2\delta} \int e^{itu} \left(\cosh\left(\frac{\pi u}{2\delta} \right) \right)^{-1} du = (\cosh(\delta t))^{-1},$$

$$\hat{f}(t) = (\cosh(\delta t))^{-1} \hat{g}(t).$$

By the Parseval formula,

$$\frac{1}{2\pi} \int \cosh^2(\delta t) |\hat{f}(t)|^2 dt = \int g_f^2(y) dy \le P$$

and hence the first inclusion holds. The second inclusion follows immediately from the Paley-Wiener theorem (see for example Katznelson (1976, p. 174)).

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. Let $\mathcal{T}_\delta$ be the topology on $\mathcal{F}_\delta$ induced by the distance

$$\rho(f,g)=\Big(\int \cosh^2(\delta t)|\hat{f}(t)-\hat{g}(t)|^2 dt\Big)^{1/2}+\int |f(y)-g(y)|dy.$$

Remark 2. This is a strong topology – closeness with respect to ρ implies, by the formula for the inverse Fourier transform, closeness of all derivatives in the uniform topology: for $g,h\in\mathcal{F}_\delta$,

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

In almost every estimation problem with censored data one faces the well known unstable behavior of the Kaplan-Meier process $\sqrt{n}(\tilde{F}_n(y)-F(y))$ (here $\tilde{F}_n$ is the Kaplan-Meier estimator, see definition below) in the right tails of F and G. Therefore, the lighter are the tails of the kernel, the less restrictive conditions on the censoring mechanism are needed. On the other hand, it turns out that when constructing an efficient estimator, one has to use observations distant from x as well as those close to x. Roughly speaking, this corresponds to the fact that even for y's distant from x the values $f(y)$ still carry some information about $f(x)$ – analytic functions have "long memory". This is formalized by imposing the following restriction on the nonparametric class $\mathcal{F}_\delta$.

Definition. For given $\alpha,\tau_0>0$, $m\ge 1$, denote

$$\tilde{\mathcal{F}}_\delta=\tilde{\mathcal{F}}_\delta(m)=\mathcal{F}_\delta(P)\cap\Big\{f:\ \inf_y\Big\{e^{\tau_0 y^{2m}}(1-F(y))(1-G(y))\Big\}>\alpha\Big\}\,, \tag{4}$$

where $\mathcal{F}_\delta(P)$ is defined by (3) and F is the distribution function corresponding to the density f.

Remark 3. The restriction on the original class $\mathcal{F}_\delta$ defined by (3) expresses the requirement for the censoring mechanism to allow distant observations with positive probability as the number of observations tends to infinity. Indeed, for some $0<p<1$ let

$$y_n=\left(\frac{1}{\tau_0}\log\left(\frac{-\alpha n}{\log(1-p)}\right)\right)^{1/2m},$$

then

$$\begin{aligned}\mathbf{P}\Big\{Z_{(n)}>y_n\Big\} &= 1-(1-(1-F(y_n))(1-G(y_n)))^n\\ &\ge 1-(1-\alpha e^{-\tau_0 y_n^{2m}})^n\\ &= 1-\left(1+\frac{\log(1-p)}{n}\right)^n\ge p\,,\end{aligned}$$

where $Z_{(n)}=\max_{1\le i\le n} Z_i$.

Remark 4. Without loss of generality we suppose that m is integer. Indeed, we will see later that both the upper bound and the lower bound for the local minimax risk do not depend on m, nor on α, P and τ_0.

Definition. Let $\tilde{\mathcal{T}}_\delta$ be the topology induced by $\mathcal{T}_\delta$ on $\tilde{\mathcal{F}}_\delta$.

Let us establish several conventions: sometimes we will write $F \in \tilde{\mathcal{F}}_\delta$ meaning actually that the corresponding density $f \in \tilde{\mathcal{F}}_\delta$; in the proofs we will denote generic positive constants by $C_1, C_2, C_3, \dots$ and they are assumed to be different in the proofs of different assertions; all symbols O and o correspond to the asymptotics $n \to \infty$ unless otherwise specified; if we say that a particular property holds locally uniformly, this means that for each $f \in \tilde{\mathcal{F}}_\delta$ there exists a neighbourhood V of f such that this property holds uniformly over V.

We propose the following class of kernels to be used in the construction of the estimator:

$$\phi_n(y) = \phi_n(y, \tau, \delta, m) = r(y)s(y), \tag{5}$$

where

$$r(y) = r(\tau, m, y) = e^{-\tau y^{2m}}, \qquad s(y) = s(\delta, n, y) = \frac{\sin(a_n y)}{\pi y}, \qquad a_n = \frac{\log n}{2\delta} \tag{6}$$

and τ is any fixed number such that $\tau > \frac{3}{2}\tau_0$, where constant τ_0 appears in the definition of the class $\tilde{\mathcal{F}}_\delta$.

Let us state some properties of kernels of the form (5)-(6):

$$\hat{\phi}_n(t) = \frac{1}{2\pi}(\hat{r} * I_{(-a_n, a_n)})(t), \tag{7}$$

$$\int \phi_n^2(y)dy = \frac{a_n}{\pi}(1 + o(1)) = \frac{\log n}{2\pi\delta}(1 + o(1)), \tag{8}$$

where $*$ is the convolution operation and I_S denotes the indicator function of set S. The first property is merely application of the standard formula for the Fourier transform of the product of two functions. To get the second relation, split the integral into two terms: integral over small neighbourhood of 0 and integral over its complement; further notice that the first term is asymptotically equivalent to

$$a_n \pi^{-2} \int \left(\frac{\sin y}{y}\right)^2 dy = a_n/\pi$$

and the second is of order $O(1)$.

Note also that the function $\hat{r}(t)$ is even. The asymptotic behavior of $\hat{r}(t)$, as $|t| \to \infty$, is described in Fedoruk (1977, pp. 213-214, 220). We adapt this result in simplified form, suitable for our purposes: for some $A_1, A_2 > 0$,

$$|\hat{r}(t)| \le A_1 \exp\left\{-A_2 |t|^{\frac{2m}{2m-1}}\right\}. \tag{9}$$

The constants A_1, A_2 depend in general on m and τ.

Define now the following estimator

$$\tilde{f}_n = \tilde{f}_n(x) = \int \phi_n(x-y)d\tilde{F}_n(y), \tag{10}$$

where $\phi_n(y)$ is defined by (5)-(6) and $\tilde{F}_n(y)$ is the Kaplan-Meier estimator, a well known nonparametric efficient estimator of the distribution function $F(y)$ (see for example Andersen et al. (1993)):

$$\tilde{F}_n(y) = 1 - \prod_{i:\,Z_{(i)}<y} \left(\frac{n-i}{n-i+1}\right)^{\Delta_{(i)}}, \tag{11}$$

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

In the next Theorem the local asymptotic performance of the estimator $\tilde{f}_n$ with respect to the topology $\tilde{\mathcal{T}}_\delta$ is established. The proofs of the theorems are given in the last section.

Theorem 1. *Let the distribution function G be continuous at point x. Then the relation*

$$\limsup_{n\to\infty} \frac{n}{\log n} \mathbf{E}_f(\tilde{f}_n(x) - f(x))^2 \le \sigma^2(f)$$

holds locally uniformly in $f \in \tilde{\mathcal{F}}_\delta$, where

$$\sigma^2(f) = \sigma^2(f,x) = \frac{f(x)}{2\pi\delta(1-G(x))} \tag{12}$$

and the estimator $\tilde{f}_n(x)$ is defined by (10).

Theorem 1 gives an upper bound for the local minimax risk (2): for a sufficiently small neighbourhood $V(f)$

$$\limsup_{n\to\infty} \frac{n}{\log n} r_n(V) \le \sup_{f\in V} \sigma(f).$$

If we can provide a lower bound for the local minimax risk, coinciding asymptotically with the upper one, then we clearly determine the asymptotic behavior of the local minimax risk. The treatment of the lower bound is similar to that in Golubev and Levit (1996), with the difference that one has to take into account the censoring mechanism. The next Theorem describes the lower bound for the local minimax risk.

Theorem 2. *Let the distribution function G be continuous at point x. Then, for each neighbourhood $V \subseteq \tilde{\mathcal{F}}_\delta$,*

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

where the local minimax risk $r_n(V)$ and $\sigma^2(f)$ are defined by (2) and (12) respectively.

In view of Theorems 1 and 2, the estimator $\tilde{f}_n$ is efficient. Indeed, for each $f \in \tilde{\mathcal{F}}_\delta$ and for any sufficiently small neighbourhood $V(f)$,

$$\lim_{n\to\infty} \frac{n}{\log n}\left(\sup_{f\in V} \mathbf{E}_f(\tilde{f}_n(x) - f(x))^2 - r_n(V)\right) = 0\,.$$

Moreover, as an immediate consequence of Theorem 1 and Theorem 2, we obtain the asymptotic behavior of the local minimax risk.

Corollary 1. *Let $f \in \tilde{\mathcal{F}}_\delta$. Then for any sufficiently small neighbourhood $V(f)$*

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

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

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

Remark 5. Note that the smaller is α and the bigger are P, m, τ_0 in the definition of the class $\tilde{\mathcal{F}}_\delta$, the less restrictive is this class, while the asymptotic behavior of the local minimax risk in no way depends on α, P, m and τ_0.

Remark 6. Compared to the result of Golubev and Levit (1996), we see that the fact of censorship does not influence the convergence rate, but it does influence the optimal constant.

Remark 7. Since the Kaplan-Meier estimator is asymptotically normal, it seems plausible that a central limit theorem for the estimator $\tilde{f}_n(x)$ can be given (cf. Yang (1994)):

$$\sqrt{\frac{n}{\log n}}(\tilde{f}_n(x) - f(x)) \xrightarrow{D} \mathcal{N}(0, \sigma^2(f,x)) \quad \text{as} \quad n \to \infty,$$

where $\sigma^2(f,x)$ is defined by (12). For related result, see Yang (1994) where a central limit theorem for the functional $\int \psi dF$ is established, however one can not apply the methods of Yang (1994) to our functional $\tilde{f}_n(x)$ directly because of dependence on n. This problem will not be treated here.

3 Preliminaries: the Kaplan-Meier estimator

Our treatment of the upper bound for the minimax risk relies heavily on the martingale approach to the Kaplan-Meier estimator (see Gill (1980)). Below we present necessary preliminaries, beginning with a suitable adaptation from Gill (1980).

Let N_n be the process counting observed X_i's and Y be the process giving the number at risk:

$$N_n(u) = \#\{i : \ Z_i \leq u, \ \Delta_i = 1\},$$
$$Y_n(u) = \#\{i : \ Z_i \geq u\},$$
$$J_n(u) = I\{Y_n(u) > 0\},$$

where symbol # denotes the number of elements in a set. Let $X(u-)$ denote left hand limit of X at point u. It is known (see for example Gill (1980)) that for y such that $F(y) < 1$ and $Y_n(y) > 0$

$$\begin{aligned}
\tilde{F}_n(y) - F(y) &= (1 - F(y)) \int_{-\infty}^{y} \frac{(1 - \tilde{F}_n(u-))}{(1 - F(u))} \frac{J_n(u)}{Y_n(u)} \left(dN_n(u) - Y_n(u) \frac{dF(u)}{1 - F(u-)} \right) \\
&= (1 - F(y)) \int_{-\infty}^{y} \frac{(1 - \tilde{F}_n(u-))}{(1 - F(u))} \frac{J_n(u)}{Y_n(u)} dM_n(u), \qquad (13)
\end{aligned}$$

where $M_n(u)$ is a square integrable martingale with the predictable variation process

$$\langle M, M \rangle (y) = \int_{-\infty}^{y} Y_n(u) \frac{1 - F(u)}{(1 - F(u-))^2} dF(u), \qquad (14)$$

while $J_n(u)$, $\tilde{F}_n(u-)$, $Y_n(u)$ are left continuous adapted processes.

We close this section with two technical lemmas which will be needed in the following section.

For the following result we refer to Weits (1993).

Lemma 1. *Let $\mathcal{B} = \mathcal{B}(A, \tau) = \{F : \ (1 - F(A))(1 - G(A)) \geq \tau > 0\}$. Then, as $n \to \infty$, the relation*

$$\mathbf{E} \frac{(1 - \tilde{F}_n(y-))^2}{(1 - F(y-))^2} \frac{J_n(y)}{Y_n(y)} = \frac{1}{n(1 - F(y-))(1 - G(y-))} + O(n^{-2})$$

holds uniformly over $\mathcal{B}$ and y, $y \leq A$.

The proof of this Lemma is essentially contained in the proof of Lemma 4 in Weits (1993). In the paper of Weits Y_n/n corresponds to our Y_n.

Lemma 2. *For all $n \geq 2$,*

$$\mathbf{E} \left(\frac{J_n(y)}{Y_n(y)} \right) \leq p(y) (1 - p(y))^{n-1} + \frac{(1 - p(y))^{n/2}}{p(y)} + \frac{2}{np(y)},$$

where $p = p(y) = (1 - F(y-))(1 - G(y-))$.

Proof. Denote

$$\mu(n) = \mu(n,y) = n\mathbf{E}(J_n(y)/Y_n(y)) = \sum_{l=1}^{n} \frac{1}{l/n} \binom{n}{l} p^l (1-p)^{n-l}.$$

Reasoning as in Weits (1993), we obtain the following recursive equation for $\mu(n)$:

$$\begin{aligned}
\mu(n) &= n \sum_{l=1}^{n-1} \frac{1}{l} \left(\binom{n-1}{l-1} + \binom{n-1}{l} \right) p^l (1-p)^{n-l} + p^n \\
&= n \left(\frac{1}{n} \sum_{l=1}^{n} \frac{n!}{(n-l)!l!} p^l (1-p)^{n-l} + (1-p) \sum_{l=1}^{n-1} \frac{1}{l} \binom{n-1}{l} p^l (1-p)^{n-l-1} \right) \\
&= n \left(\frac{1}{n} (1-(1-p)^n) + (1-p) \frac{1}{n-1} \mu(n-1) \right) \\
&= 1 + (1-p) \frac{n}{n-1} \mu(n-1) - (1-p)^n .
\end{aligned}$$

Certainly $\mu(n) \leq \lambda(n)$, where $\lambda(n)$ satisfies the recursive equation:

$$\lambda(n) = 1 + (1-p)\frac{n}{n-1}\lambda(n-1), \qquad \lambda(1) = \mu(1) = p, \qquad n \geq 2.$$

Let $C(n)$ be a solution of the corresponding homogeneous equation:

$$C(n) = (1-p)\frac{n}{n-1}C(n-1) \qquad C(1) = 1, \qquad n \geq 2.$$

Thus,

$$C(n) = (1-p)^{n-1} n.$$

Let $B(n)$ be such that $\lambda(n) = C(n)B(n)$. Then $B(n)$ satisfies

$$B(n) = B(n-1) + \frac{n-1}{n(1-p)}(C(n-1))^{-1}, \qquad B(1) = p, \qquad n \geq 2.$$

It is easy to see that

$$B(n) = p + \sum_{k=2}^{n} (1-p)^{1-k} k^{-1}$$

and

$$\lambda(n) = p(1-p)^{n-1} n + n \sum_{k=2}^{n} (1-p)^{n-k} k^{-1}.$$

Now we bound the second term in the last relation

$$\begin{aligned}
\sum_{k=2}^{n} (1-p)^{n-k} k^{-1} &\leq \sum_{k=1}^{[n/2]} (1-p)^{n-k} + \frac{1}{[n/2]+1} \sum_{k=[n/2]+1}^{n} (1-p)^{n-k} \\
&\leq \frac{(1-p)^{n/2}}{p} + \frac{2}{np}
\end{aligned}$$

and hence the claim follows

$$\mathbf{E}\left(\frac{J_n(y)}{Y_n(y)}\right) = \frac{\mu(n)}{n} \le \frac{\lambda(n)}{n} \le p\,(1-p)^{n-1} + \frac{(1-p)^{n/2}}{p} + \frac{2}{np}\,.$$

$\square$

4 Auxiliary results

In this section we provide further technical results which we will need in the proof of theorems.

Lemma 3. *Let $\phi_n(y)$ be defined by (5) and $h(y)$ be a bounded and continuous at x function. Then, as $n \to \infty$, the relation*

$$\int \phi_n^2(x-y)h(y)dF(y) = h(x)f(x)\int \phi_n^2(y)dy(1+o(1)) = \frac{h(x)f(x)\log n}{2\pi\delta}(1+o(1)).$$

holds locally uniformly in $F \in \tilde{\mathcal{F}}_\delta$,

Proof. The second relation follows immediately from (8). It remains to prove only the first identity. Let $O_\epsilon(x) = \{y : \ |x-y| < \epsilon\}$ be the open interval around x radius $\epsilon = \epsilon_n$, $\epsilon_n \to 0$, $(\epsilon_n^2 \log n)^{-1} = o(1)$ as $n \to \infty$. We have obviously

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

So it is enough to prove that the right-hand side of the last identity is of order $o(\log n)$ locally uniformly.

One can bound the function $\phi_n(x-y)$ outside the interval $O_\epsilon(x)$ as follows:

$$\phi_n^2(x-y) \le \frac{e^{-2\tau(x-y)^{2m}}}{\pi^2\epsilon_n^2}.$$

Therefore, the inequality

$$\begin{aligned}
&\int_{(O_\epsilon(x))^C} \phi_n^2(x-y)\,|h(y)f(y) - h(x)f(x)|\,dy \\
&\le C_1\epsilon_n^{-2}\int_{(O_\epsilon(x))^C} e^{-2\tau(x-y)^{2m}}dy \\
&= o(\log n)
\end{aligned}$$

holds locally uniformly because our topology is stronger than the topology induced by the sup-norm (see Remark 2). Next, by (8), it is easy to see that

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

locally uniformly and Lemma follows. □

Lemma 4. *The functional*

$$\psi_n(y) = \psi_n(y,x,F) = \int_y^\infty \phi_n(x-u)dF(u) \tag{15}$$

is bounded locally uniformly in $F \in \tilde{\mathcal{F}}_\delta$ and uniformly in y.

Proof. 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

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

The second term (integral over $D_2(y)$) is clearly bounded. For the first term, we have that

$$\begin{aligned}\left|\int_{D_1(y)} \phi_n(x-u)dF(u)\right| &\le \left|f(x)\int_{D_1(y)} \phi_n(x-u)du\right| \\ &\quad + \left|\int_{D_1(y)} f'(u^*)(u-x)\phi_n(x-u)du\right| \\ &\le C_1 + C_2\int_{D_1(y)} e^{-\tau(x-u)^{2m}}du \le C_3\end{aligned}$$

locally uniformly because $\sup_{u\in O_\epsilon(x)}|f'(u)|$ is bounded locally uniformly (see Remark 2) and

$$\left|\int_{D_1(y)} \phi_n(x-u)du\right| = \left|\int_{a_n(b_1-x)}^{a_n(b_2-x)} \frac{e^{-\tau(u/a_n)^{2m}}\sin u}{\pi u}du\right| \le C_4.$$

□

Lemma 5. *As $n\to\infty$, the relation*

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

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

Proof. Recalling the definition of the class (4), we obtain the following uniform bound:

$$\begin{aligned}
\left(\int \phi_n(x-y)dF(y) - f(x)\right)^2 &= \left(\frac{1}{2\pi}\int e^{-itx}(\hat{\phi}_n(t)-1)\hat{f}(t)dt\right)^2 \\
&\leq \frac{1}{2\pi}\int \cosh^2(\delta t)|\hat{f}(t)|^2 dt \cdot \frac{1}{2\pi}\int \frac{|\hat{\phi}_n(t)-1|^2}{(\cosh(\delta t))^2}dt \\
&\leq C_1 \int \frac{|\hat{\phi}_n(t)-1|^2}{(\cosh(\delta t))^2}dt \\
&\leq C_1 \int_{-a_n}^{a_n} \frac{|\hat{\phi}_n(t)-1|^2}{(\cosh(\delta t))^2}dt + C_2 \int_{|t|\geq a_n} \frac{dt}{(\cosh(\delta t))^2} \\
&= 2C_1 \int_0^{a_n} \frac{|\hat{\phi}_n(t)-1|^2}{(\cosh(\delta t))^2}dt + O(n^{-1}). \qquad (16)
\end{aligned}$$

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

$$\int_{u>t+a_n} |\hat{r}(u)|du \leq \int_{u<t-a_n} |\hat{r}(u)|du = \int_{u>a_n-t} |\hat{r}(u)|du$$

for $t \in [0, a_n]$. Now using the last inequality, (7), (9) and fact that

$$\int \hat{r}(u)du = r(0)2\pi = 2\pi,$$

we have that, for $t \in [0, a_n]$,

$$\begin{aligned}
|\hat{\phi}_n(t) - 1| &= \left|(2\pi)^{-1}(\hat{r} * I_{(-a_n,a_n)})(t) - 1\right| \\
&= \left|(2\pi)^{-1}\int (I_{(-a_n,a_n)})(t-u) - 1)\hat{r}(u)du\right| \\
&= (2\pi)^{-1}\left|\int_{|t-u|>a_n} \hat{r}(u)du\right| \\
&\leq 2(2\pi)^{-1}\int_{u>a_n-t} |\hat{r}(u)|du \\
&\leq C_3 \int_{u>a_n-t} \exp\left\{-A_2 u^{\frac{2m}{2m-1}}\right\} du \\
&\leq C_4 \exp\left\{-A_2(a_n-t)^{\frac{2m}{2m-1}}\right\}. \qquad (17)
\end{aligned}$$

Combining this with (16) completes the proof:

$$\begin{aligned}
\int_0^{a_n} \frac{|\hat{\phi}_n(t)-1|^2}{(\cosh(\delta t))^2}dt &\leq C_5 \int_0^{a_n} \exp\left\{-2A_2(a_n-t)^{\frac{2m}{2m-1}} - 2\delta t\right\} dt \\
&= C_5 \int_0^{a_n} \exp\left\{-2A_2 y^{\frac{2m}{2m-1}} - 2\delta(a_n - y)\right\} dy \\
&\leq C_5 e^{-2\delta a_n} \int_0^{\infty} \exp\left\{-2A_2 y^{\frac{2m}{2m-1}} + 2\delta y\right\} dy \\
&= \frac{C_6}{n}.
\end{aligned}$$

□

Lemma 6. *As $n \to \infty$, the relations*

$$\begin{aligned} \mathbf{E}\left(\frac{\psi_n(Z_{(n)})}{1-F(Z_{(n)})}\right)^2 &= O(n^{-1}), \\ \mathbf{E}\left(\psi_n(Z_{(n)})\right)^2 &= O(n^{-1}) \end{aligned}$$

hold locally uniformly in $F \in \tilde{\mathcal{F}}_\delta$, where $\psi_n(y)$ is defined by (15).

Proof. From the definition of the nonparametric class (4) it follows that for each $F \in \tilde{\mathcal{F}}_\delta$

$$H(y) \stackrel{\text{def}}{=} \mathbf{P}\{Z_1 \le y\} = 1-(1-F(y))(1-G(y)) \le 1-\alpha e^{-\tau_0 y^{2m}}. \tag{18}$$

Fix some $\epsilon > 0$ and notice that

$$H(y) \le H(x+\epsilon) \le 1-\alpha e^{-\tau_0 (x+\epsilon)^{2m}} = q < 1 \qquad \text{for all} \quad y \le x+\epsilon\,.$$

Further, by the definition of the kernel function (5), it is easy to see that, with some constant $C_1 > 0$,

$$\phi_n^2(x-y) \le (\pi\epsilon)^{-2} e^{-2\tau(x-y)^{2m}} \le C_1 e^{-3\tau_0 y^{2m}} \qquad \text{for all} \quad y \ge x+\epsilon\,.$$

Besides, we have obviously that

$$\phi_n^2(y) \le a_n^2 \pi^{-2}\,.$$

Now, using the Hölder inequality and all the inequalities above, we obtain the second relation:

$$\begin{aligned} \mathbf{E}\left(\int_{Z_{(n)}}^{\infty} \phi_n(x-y)dF(y)\right)^2 &\le \mathbf{E}\int_{-\infty}^{\infty} I\{Z_{(n)} \le y\}\phi_n^2(x-y)dF(y) \\ &= \int_{-\infty}^{\infty} H^n(y)\phi_n^2(x-y)dF(y) \\ &= \int_{-\infty}^{x+\epsilon} H^n(y)\phi_n^2(x-y)dF(y) + \int_{x+\epsilon}^{\infty} H^n(y)\phi_n^2(x-y)dF(y) \\ &\le q^n \pi^{-2} a_n^2 F(x+\epsilon) + C_2 \int_{x+\epsilon}^{\infty} \left(1-\alpha e^{-\tau_0 y^{2m}}\right)^n \phi_n^2(x-y)dy \\ &\le q^n \pi^{-2} a_n^2 + C_3 \int_{x+\epsilon}^{\infty} e^{-3\tau_0 y^{2m}} \left(1-\alpha e^{-\tau_0 y^{2m}}\right)^n dy \\ &\le C_4 e^{-C_5 n}(\log n)^2 + C_6 \int_0^{C_7} (1-u)^n du = O(n^{-1}) \end{aligned}$$

locally uniformly because $f(y)$ is bounded locally uniformly and $0 < C_7 \le 1$.

To prove the first relation, note that from (4)

$$
\begin{aligned}
&\mathbf{E}\left(\frac{\psi_n(Z_{(n)})}{1-F(Z_{(n)})}\right)^2 \\
&= \mathbf{E}\left(\frac{\psi_n(Z_{(n)})}{1-F(Z_{(n)})}\left(I\{Z_{(n)}\le x+\epsilon\}+I\{Z_{(n)}>x+\epsilon\}\right)\right)^2 \\
&\le C_8\mathbf{E}\left(\psi_n(Z_{(n)})\right)^2 + 2\alpha^{-2}\mathbf{E}\left(\exp\{2\tau_0(Z_{(n)})^{2m}\}I\{Z_{(n)}>x+\epsilon\}\psi_n^2(Z_{(n)})\right).
\end{aligned}
$$

Therefore, it remains only to show that the second term in the right hand side of the last inequality is of order $O(n^{-1})$ locally uniformly:

$$
\begin{aligned}
&\mathbf{E}\left(\exp\{2\tau_0(Z_{(n)})^{2m}\}I\{Z_{(n)}>x+\epsilon\}\psi_n^2(Z_{(n)})\right) \\
&= \mathbf{E}\left(\exp\{2\tau_0(Z_{(n)})^{2m}\}I\{Z_{(n)}>x+\epsilon\}\left(\int_{Z_{(n)}}^{\infty}\phi_n(x-y)dF(y)\right)^2\right) \\
&\le \mathbf{E}\left(\exp\{2\tau_0(Z_{(n)})^{2m}\}I\{Z_{(n)}>x+\epsilon\}\int_{Z_{(n)}}^{\infty}\phi_n^2(x-y)dF(y)\right) \\
&\le C_9\mathbf{E}\left(\exp\{2\tau_0(Z_{(n)})^{2m}\}I\{Z_{(n)}>x+\epsilon\}\int_{Z_{(n)}}^{\infty}e^{-3\tau_0 y^{2m}}dy\right) \\
&\le C_9\mathbf{E}\int_{Z_{(n)}}^{\infty}e^{-\tau_0 y^{2m}}dy = C_9\int H^n(y)e^{-\tau_0 y^{2m}}dy \\
&\le C_9\int\left(1-\alpha e^{-\tau_0 y^{2m}}\right)^n e^{-\tau_0 y^{2m}}dy \le C_{10}\int_0^1(1-u)^n du = O(n^{-1})
\end{aligned}
$$

locally uniformly. □

For brevity sake, denote from now on

$$\bar{F}(y) = 1 - F(y).$$

Lemma 7. *Let the distribution function G be continuous at the point x. Then, as $n \to \infty$, the relation*

$$
\begin{aligned}
&\frac{n}{\log n}\mathbf{E}\int_{-\infty}^{Z_{(n)}}\left(\phi_n(x-y)\bar{F}(y)-\psi_n(y)\right)^2\frac{(1-\tilde{F}_n(y-))^2}{(1-F(y-))^4}\frac{J_n(y)}{Y_n(y)}\bar{F}(y)dF(y) \\
&\qquad = \sigma^2(f)(1+o(1)) \qquad (19)
\end{aligned}
$$

holds locally uniformly in $F \in \tilde{\mathcal{F}}_\delta$, where $\sigma^2(f)$ and ψ_n are defined by (12) and (15) respectively.

Proof. By continuity of $F(y)$, we write the the left-hand side of (19), for some positive ϵ, as a sum of two terms

$$I_1 = \mathbf{E} \int_{-\infty}^{Z_{(n)}} \left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right)^2 \frac{(1-\tilde{F}_n(y-))^2}{(1-F(y-))^3} \frac{J_n(y)}{Y_n(y)} I\{Z_{(n)} \le x+\epsilon\} dF(y)$$

and

$$I_2 = \mathbf{E} \int_{-\infty}^{Z_{(n)}} \left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right)^2 \frac{(1-\tilde{F}_n(y-))^2}{(1-F(y-))^3} \frac{J_n(y)}{Y_n(y)} I\{Z_{(n)} > x+\epsilon\} dF(y) .$$

To evaluate the first term, observe first that by (18)

$$\mathbf{P}\{Z_1 \le x+\epsilon\} = H(x+\epsilon) \le 1 - \alpha e^{-\tau_0 (x+\epsilon)^{2m}} = q < 1 ,$$

$$\begin{aligned} \left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right)^2 &\le 2\phi_n^2(x-y)\bar{F}^2(y) + 2\psi_n^2(y) \\ &\le 4\pi^{-2} a_n^2 e^{-2\tau(x-y)^{2m}} \bar{F}^2(y) \\ &= C_1(\log n)^2 e^{-2\tau(x-y)^{2m}} \bar{F}^2(y) \\ &\le C_2(\log n)^2 e^{-3\tau_0 y^{2m}} \bar{F}^2(y) \end{aligned} \tag{20}$$

and $J_n(y)/Y_n(y) \le 1$. Thus, recalling (4), we bound the first term as follows:

$$\begin{aligned} I_1 &\le C_2(\log n)^2 H^n(x+\epsilon) \int_{-\infty}^{\infty} \frac{e^{-3\tau_0 y^{2m}} dF(y)}{(1-F(y))}. \\ &\le C_2(\log n)^2 q^n \alpha^{-1} \int_{-\infty}^{\infty} e^{-3\tau_0 y^{2m} + \tau_0 y^{2m}} dF(y) \\ &\le C_3 e^{-C_4 n} . \end{aligned} \tag{21}$$

For the second term, we split the integral in I_2 into two parts: the integral over $(-\infty, x+\epsilon]$ and the integral over $(x+\epsilon, Z_{(n)}]$. Since for $y \ge x+\epsilon$

$$\begin{aligned} \left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right)^2 &\le 2\phi_n^2(x-y)\bar{F}^2(y) + 2\psi_n^2(y) \\ &\le 4(\pi\epsilon)^{-2} e^{-2\tau(x-y)^{2m}} \bar{F}^2(y) \\ &\le C_5 e^{-3\tau_0 y^{2m}} \bar{F}^2(y), \end{aligned}$$

we bound the expectation of the integral over $(x+\epsilon, Z_{(n)}]$ merely by

$$C_5 \int_{x+\epsilon}^{\infty} \mathbf{E}\left(\frac{J_n(y)}{Y_n(y)}\right) \frac{e^{-3\tau_0 y^{2m}} dF(y)}{(1-F(y))}.$$

Thus,

$$I_2 \le \mathbf{E}\left(\int_{-\infty}^{x+\epsilon} \left(\phi_n(x-y)\bar{F}(y) - \psi_n(y)\right)^2 \frac{(1-\tilde{F}_n(y-))^2}{(1-F(y-))^3} \frac{J_n(y)}{Y_n(y)} dF(y)\right)$$

$$+ C_5 \int_{x+\epsilon}^{\infty} \mathbf{E} \left(\frac{J_n(y)}{Y_n(y)} \right) \frac{e^{-3\tau_0 y^{2m}} dF(y)}{(1-F(y))} = S_1 + S_2 \,,$$

say.

To evaluate S_2, we make use of Lemma 2 and (4):

$$\begin{aligned} S_2 &= C_5 \int_{x+\epsilon}^{\infty} e^{-3\tau_0 y^{2m}} \mathbf{E} \left(\frac{J_n(y)}{Y_n(y)} \right) \frac{dF(y)}{(1-F(y))} \\ &\le 2C_5 \int_{x+\epsilon}^{\infty} \frac{e^{-3\tau_0 y^{2m}} (H^{n/2}(y) + n^{-1}) dF(y)}{(1-F(y))^2 (1-G(y))} \\ &\le C_6 \int_{x+\epsilon}^{\infty} e^{-3\tau_0 y^{2m} + 2\tau_0 y^{2m}} \left(\left(1 - \alpha e^{-\tau_0 y^{2m}} \right)^{n/2} + n^{-1} \right) f(y) dy \\ &\le C_7 \int_0^{C_8} (1-u)^{n/2} du + C_6 n^{-1} \\ &= O(n^{-1}) \end{aligned}$$

locally uniformly.

Therefore, to complete the proof, it remains only to prove that the relation

$$\frac{n}{\log n} S_1 = \sigma^2(f)(1+o(1)) \tag{22}$$

holds locally uniformly.

Since

$$(1-F(x+\epsilon))(1-G(x+\epsilon)) \ge \alpha e^{-\tau_0 (x+\epsilon)^{2m}} = \gamma > 0$$

uniformly over $\tilde{\mathcal{F}}_\delta$, by (20) and Lemma 1 we have that

$$\begin{aligned} S_1 &= \mathbf{E} \left(\int_{-\infty}^{x+\epsilon} \left(\phi_n(x-y) \bar{F}(y) - \psi_n(y) \right)^2 \frac{(1-\tilde{F}_n(y-))^2}{(1-F(y-))^3} \frac{J_n(y)}{Y_n(y)} dF(y) \right) \\ &= \int_{-\infty}^{x+\epsilon} \frac{\left(\phi_n(x-y) \bar{F}(y) - \psi_n(y) \right)^2 dF(y)}{n(1-F(y))^2 (1-G(y-))} + O\left((\log n)^2 n^{-2} \right) \end{aligned}$$

uniformly. Finally, we obtain by Lemma 3 and 4 that

$$S_1 = \int_{-\infty}^{x+\epsilon} \frac{\phi_n^2(x-y) dF(y)}{n(1-G(y-))} + O(n^{-1}) = \frac{\log n}{n} \sigma^2(f)(1+o(1))$$

locally uniformly. Thus (22) holds and the proof of Lemma is complete. □

5 Proofs of Theorems

Upper bound: proof of Theorem 1. First we provide necessary preliminaries. By (13) and integration by parts, we have

$$\begin{aligned}
&\int_{-\infty}^{Z_{(n)}} \phi_n(x-y)d(\tilde{F}_n(y)-F(y)) \\
&= \int_{-\infty}^{Z_{(n)}} \left(\phi_n(x-y)\bar{F}(y)-\psi_n(y)\right) \frac{(1-\tilde{F}_n(y-))}{(1-F(y-))}\frac{J_n(y)}{Y_n(y)}dM_n(y) \\
&+ \frac{\tilde{F}_n(Z_{(n)})-F(Z_{(n)})}{1-F(Z_{(n)})}\psi_n(Z_{(n)}). \qquad (23)
\end{aligned}$$

Since the first term of (23) is the integral of a predictable locally bounded process (almost all its sample paths are locally bounded) with respect to a square integrable martingale with the predictable variation process (14) (see for example Gill (1980)), one can represent its second moment as follows:

$$\begin{aligned}
&\mathbf{E}\left(\int_{-\infty}^{Z_{(n)}} \left(\phi_n(x-y)\bar{F}(y)-\psi_n(y)\right) \frac{(1-\tilde{F}_n(y-))}{(1-F(y-))}\frac{J_n(y)}{Y_n(y)}dM_n(y)\right)^2 \\
&= \mathbf{E}\int_{-\infty}^{Z_{(n)}} \left(\phi_n(x-y)\bar{F}(y)-\psi_n(y)\right)^2 \frac{(1-\tilde{F}_n(y-))^2}{(1-F(y-))^4}\frac{J_n(y)}{Y_n(y)}\bar{F}(y)dF(y). \qquad (24)
\end{aligned}$$

Recall that $\tilde{F}_n(y)$ is constant on $[Z_{(n)},\infty)$. So, using (23), (24) and (twice) the elementary inequality

$$(a+b)^2 \le (1+\gamma)\,a^2 + (1+\gamma^{-1})\,b^2, \qquad 0<\gamma\le 1, \qquad (25)$$

we obtain

$$\begin{aligned}
&\mathbf{E}\left(\int_{-\infty}^{\infty} \phi_n(x-y)d(\tilde{F}_n(y)-F(y))\right)^2 \\
&\le (1+\gamma_n)\,\mathbf{E}\left(\int_{-\infty}^{Z_{(n)}} \phi_n(x-y)d(\tilde{F}_n(y)-F(y))\right)^2 + (1+\gamma_n^{-1})\,\mathbf{E}\left(\psi_n(Z_{(n)})\right)^2 \\
&\le (1+\gamma_n)^2\,\mathbf{E}\left(\int_{-\infty}^{Z_{(n)}} \left(\phi_n(x-y)\bar{F}(y)-\psi_n(y)\right)^2 \frac{(1-\tilde{F}_n(y-))^2}{(1-F(y-))^4}\frac{J_n(y)}{Y_n(y)}\bar{F}(y)dF(y)\right) \\
&+ (\gamma_n^{-1}+2+\gamma_n)\,\mathbf{E}\left(\frac{\psi_n(Z_{(n)})}{1-F(Z_{(n)})}\right)^2 + (1+\gamma_n^{-1})\,\mathbf{E}\left(\psi_n(Z_{(n)})\right)^2, \qquad (26)
\end{aligned}$$

where γ_n is to be chosen later.

Now we evaluate the risk of the estimator (10). From the last inequality and again the elementary inequality (25) it follows that

$$\begin{aligned}
&\mathbf{E}_f(\tilde{f}_n(x)-f(x))^2 = \mathbf{E}_f\left(\tilde{f}_n(x)-\int\phi_n(x-y)dF(y)+\int\phi_n(x-y)dF(y)-f(x)\right)^2 \\
&\le (1+\gamma_n)\,\mathbf{E}\left(\int\phi_n(x-y)d(\tilde{F}_n(y)-F(y))\right)^2
\end{aligned}$$

$$
\begin{aligned}
&+(1+\gamma_n^{-1})\left(\int \phi_n(x-y)dF(y)-f(x)\right)^2 \\
\le\ &(1+\gamma_n)^3\,\mathbf{E}\int_{-\infty}^{Z_{(n)}}\left(\phi_n(x-y)\bar F(y)-\psi_n(y)\right)^2\frac{(1-\tilde F_n(y-))^2}{(1-F(y-))^4}\frac{J_n(y)}{Y_n(y)}\bar F(y)dF(y) \\
&+(\gamma_n^{-1}+3+3\gamma_n+\gamma_n^2)\,\mathbf{E}\left(\frac{\psi_n(Z_{(n)})}{1-F(Z_{(n)})}\right)^2+(\gamma_n^{-1}+2+\gamma_n)\mathbf{E}\left(\psi_n(Z_{(n)})\right)^2 \\
&+(1+\gamma_n^{-1})\left(\int \phi_n(x-y)dF(y)-f(x)\right)^2.
\end{aligned}
$$

We choose now γ_n such that $\gamma_n \to 0$ and $(\gamma_n \log n)^{-1} = o(1)$ as $n \to \infty$. Combining the last relation with Lemmas 5, 6 and 7 proves the theorem. □

Lower bound: proof of Theorem 2. Let $f_0(y)$ be an arbitrary density from neighbourhood V and let F_0 be the corresponding distribution function. Consider the following family of functions (cf. Golubev and Levit (1996)):

$$f_\theta(y) = f_\theta(y, x, \phi_n, f_0) = f_0(y)(1 + \theta(\phi_n(x-y) - \bar\phi_n(x)),$$

where $|\theta| \le \theta_n$, $\phi_n(y)$ is defined by (5) and $\bar\phi_n(x) = \int \phi_n(x-y)f_0(y)dy$. Let θ_n be such that $\epsilon_n \le \theta_n \le \rho_n$, where the positive sequences ϵ_n and ρ_n satisfy

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

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

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

Proposition 1. *For sufficiently large n, $f_\theta \in V$.*

Proof. Take $\epsilon > 0$ such that $O_\epsilon(f_0) \subseteq V$, where

$$O_\epsilon(f_0) = \left\{ f \in \tilde{\mathcal{F}}_\delta : \ \rho(f, f_0) < \epsilon \right\}.$$

We prove now that $f_\theta \in O_\epsilon(f_0)$ for sufficiently large n.

It is easy to check the condition on the nonparametric class (see (4)): for sufficiently large n,

$$\inf_y \left\{ e^{\tau_0 y^{2m}}(1-F_\theta(y))(1-G(y)) \right\} = \inf_y \left\{ e^{\tau_0 y^{2m}}(1-F_0(y))(1-G(y)) \right\}(1+o(1)) > \alpha$$

since $F_0 \in \tilde{\mathcal{F}}_\delta$.

Denote

$$\psi(t) = \psi(t, x) = f_0(y)\phi_n(x-y).$$

First, by Minkowski inequality and the definition (5)-(6) of the kernel ϕ_n,

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

So it suffices to show that the the first term on right-hand side of the last inequality converges to zero as $n \to \infty$. Since

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

by the generalized Minkowski inequality (Nikol'skii (1975, p. 20)), (4), (7) and (9) it follows that

$$\begin{aligned}
&\left(\int \cosh^2(\delta t) |\hat{\psi}(t,x)|^2 dt \right)^{1/2} \\
&\leq C_2 \left(\int \left| \int e^{\delta|t|} e^{ixu} \hat{f}_0(t+u) \hat{\phi}_n(u) du \right|^2 dt \right)^{1/2} \\
&\leq C_2 \int \left(\int \left| e^{\delta|t|} e^{ixu} \hat{f}_0(t+u) \hat{\phi}_n(u) \right|^2 dt \right)^{1/2} du \\
&\leq C_2 \int \left(\int \left| e^{\delta|t+u|} \hat{f}_0(t+u) e^{\delta|u|} \hat{\phi}_n(u) \right|^2 dt \right)^{1/2} du \\
&\leq C_3 \left(\int \cosh^2(\delta t) |\hat{f}_0(t)|^2 dt \right)^{1/2} \int e^{\delta|u|} |\hat{\phi}_n(u)| du \\
&\leq C_4 \int e^{\delta|u|} |\hat{\phi}_n(u)| du \\
&\leq C_4 \left(\int e^{2\delta|u|} \left| \left(\hat{r} * I_{(-a_n,a_n)} \right)(u) \right|^2 du \right)^{1/2} \\
&= C_4 \left(\int \left| \int e^{\delta|u|} \hat{r}(u-t) I_{(-a_n,a_n)}(t) dt \right|^2 du \right)^{1/2} \\
&\leq C_4 \int \left(\int \left| e^{\delta|u|} \hat{r}(u-t) I_{(-a_n,a_n)}(t) \right|^2 du \right)^{1/2} dt \\
&\leq C_5 \left(\int e^{2\delta|t|} |\hat{r}(t)|^2 dt \right)^{1/2} \int e^{\delta|u|} I_{(-a_n,a_n)}(u) du \\
&\leq C_6 \left(\int \exp\left\{ 2\delta|t| - 2A_2 |t|^{\frac{2m}{2m-1}} \right\} dt \right)^{1/2} \int e^{\delta|u|} I_{(-a_n,a_n)}(u) du
\end{aligned}$$

$$
\begin{aligned}
&\leq C_7 \int e^{\delta|u|} I_{(-a_n,a_n)}(u)du \\
&\leq C_8 n^{1/2}.
\end{aligned}
$$

Recalling the condition on the θ_n, we obtain finally that

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

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

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

The following Proposition describes the Fisher information $I(\theta)$ about θ contained in the observation (Z, Δ).

Proposition 2. *As $n \to \infty$, the relation*

$$I(\theta) \stackrel{\text{def}}{=} \mathbf{E}\left[\frac{\partial \log f_\theta(Z,\Delta)}{\partial\theta}\right]^2 = \frac{f_0(x)(1 - G(x))\log n}{2\pi\delta}(1 + o(1)). \tag{27}$$

holds uniformly in θ, $|\theta| < \theta_n$.

Proof. By straightforward calculations,

$$
\begin{aligned}
I(\theta) &= \mathbf{E}\left[\frac{\partial \log f_\theta(Z,\Delta)}{\partial\theta}\right]^2 \\
&= \int \frac{(\phi_n(x-y) - \bar\phi_n(x))^2 f_0(y)(1 - G(y))dy}{1 + \theta(\phi_n(x-y) - \bar\phi_n(x))} \\
&\quad + \int \frac{\left(\int_{-\infty}^y f_0(u)(\phi_n(x-u) - \bar\phi_n(x))du\right)^2 dG(y)}{1 - F_\theta(y)}.
\end{aligned}
$$

We 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 $(x+\epsilon, \infty)$. If we bound the integrands of both integrals, then we obtain in fact that the second term in the last inequality is of order $O(1)$. Since, according to Lemmas 4 and 5, $\int_{-\infty}^y f_0(z)(\phi_n(x-z) - \bar\phi_n(x))dz$ is bounded and

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

for $y \in (-\infty, x+\epsilon]$, the integral over $(-\infty, x+\epsilon]$ is bounded. Note that

$$\int_{-\infty}^y f_0(z)(\phi_n(x-z) - \bar\phi_n(x))dz = -\int_y^\infty f_0(z)(\phi_n(x-z) - \bar\phi_n(x))dz$$

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

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

for $y \in (x+\epsilon, \infty)$. Thus, we obtained that

$$\int \frac{\left(\int_{-\infty}^{y} f_0(z)(\phi_n(x-z) - \bar{\phi}_n(x))dz\right)^2 dG(y)}{1-F_\theta(y)} = O(1)$$

uniformly in θ, $|\theta| < \theta_n$. According to Lemmas 3 and 4, it is not difficult to see that

$$\int \frac{(\phi_n(x-y) - \bar{\phi}_n(x))^2 f_0(y)(1-G(y))dy}{1+\theta(\phi_n(x-y) - \bar{\phi}_n(x))} = \frac{f_0(x)(1-G(x))\log n}{2\pi\delta}(1+o(1))$$

uniformly in θ, $|\theta| < \theta_n$. Relation (27) is proved. □

Now we proceed to prove the theorem. Introduce

$$\nu(x) = \nu_n(x) = \theta_n^{-1}\nu_0(\theta_n^{-1}x),$$

where $\nu_0(x)$ is a probability density on the interval $[-1,1]$ with finite Fisher information

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

such that $\nu_0(-1) = \nu_0(1) = 0$ and $\nu_0(x)$ is continuously differentiable for $|x| < 1$. The function $\nu(x)$ is the probability density with support $[-\theta_n, \theta_n]$. It is easy to calculate the Fisher information of the distribution defined by density $\nu(x)$:

$$I(\nu) = I_n(\nu) = I_0\theta_n^{-2}.$$

Applying now the van Trees inequality for the Bayes risk below (see Gill and Levit (1995), van Trees (1968), cf. Golubev and Levit (1996)) and Propositions 1 and 2, we obtain that for sufficiently large n

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

or

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

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

$$\liminf_{n\to\infty}\frac{n}{\log n}r_n(V) \geq \sigma^2(f).$$

Therefore

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

which proves Theorem. □

Acknowledgement. The author is grateful to B. Levit for helpful conversations and suggestions.

References

Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (1993). *Statistical models based on counting processes.* Springer Verlag, New York.

Diehl, S. and W. Stute (1988). Kernel density and hazard function estimation in the presence of censoring. *J. Mult. Anal. 25*, 299–310.

Fedoruk, M. V. (1977). *Metod perevala.* Nauka, Moscow. In Russian.

Gill, R. D. (1980). *Censoring and Stochastic Integrals*, Volume 124. Mathematisch Centrum, Amsterdam.

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

Golubev, G. K. and B. Y. Levit (1996). Asymptotically Efficient Estimation for Analytic Distributions. *Math. Methods Statist. 5*(3), 357–368.

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

Hentzschel, J. (1992). Density estimation with Laguerre series and censored samples. *Statistics 23*, 49–61.

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

Ibragimov, I. A. and R. Z. Hasminskii (1983). Estimation of distribution density. *Journ. Sov. Math. 25*, 40–57. Originally published in Russian in 1980.

Katznelson, Y. (1976). *An introduction to harmonic analysis.* Dover publications, New York. Second corrected edition.

Kulasekera, K. B. (1995). A bound on the $\mathbf{L}_1$-error of a nonparametric density estimator with censored data. *Statist. Probab. Lett. 23*, 233–238.

Liu, R. C. (1996). Optimal rates, minimax estimations, and K-M based procedures for estimating functionals of a distribution under censoring. Technical report, Department of Mathematics, Cornell University.

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

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 Verlag, Berlin–Heidelberg–New York.

Timan, A. F. (1963). *Theory of Approximation of Functions of a Real Variable.* Pergamon Press, Oxford–London–New York–Paris.

van Trees, H. L. (1968). *Detection, Estimation and Modulation Theory, Part 1.* Wiley, New York.

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

Yang, S. (1994). A central limit theorem for functionals of the Kaplan-Meier estimator. *Statist. Probab. Lett. 21*, 337–345.