Asymptotically local minimax estimation of infinitely smooth density with censored data

E.N. Belitser, B.Y. Levit

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

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 limiting behavior of the local minimax risk and propose the efficient kernel -type estimator based on the Kaplan-Meier estimator.

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

1 Introduction

Suppose $T_1, \ldots, T_n$ are iid random variables (lifetimes) with common distribution function F and suppose $C_1, \ldots, C_n$ 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 $(Z_1, \delta_1), \ldots, (Z_n, \delta_n)$, 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, using the observed data, to estimate the density $f(x)$ at a given point x.

The problem of density estimation under random censorship is not new and has long been treated in the literature (see Mielniczuk (1986), Lo, Mack and Wang (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 a Kaplan-Meier based estimators.

We propose estimator of the form

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

where $\tilde{F}_n$ is the Kaplan-Meier estimator of the distribution function and $\phi_n(y)$ is some sequence of functions (the exact definitions is given in the next section) which we will call *kernel.*

To make the problem of minimax density estimation feasible, one typically restrict oneself to a certain nonparametric class of densities described usually by some smoothness assumptions. Belitser (1996), motivated by study of noncensored case in Golubev and Levit (1996), considered the class of analytic functions. In this paper the major difficulty was that there are no efficient (i.e. asymptotically locally minimax) kernel estimators with finitely supported kernels – to estimate efficiently one has to use all observations. So, the censoring may cause problems in estimation of analytic densities. However, it has been shown that, under condition that censoring is not of severe influence, one can choose a kernel with exponentially decreasing tails. The proof of efficiency of the estimator was based on the martingale technique and remains still rather involved.

In this note we consider a class of infinitely smooth functions – the definition is given in terms of Fourier trasformations in the next section. In this case, as we show, there are efficient estimators with finitely supported kernels. This facilitates also the use of the result of Lo, Mack and Wang (1989) about strong representation of the Kaplan-Meier estimator by a sum of independent random variables. We establish the exact limiting behavior of the local minimax risk and propose an efficient estimator of the form (**??**), with a finitely supported kernel.

2 Definitions and main result

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

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

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

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

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

One can easily see that functions from this class are infinitely differentiable.

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

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

Remark 1. This is one of the possible choices of topology, when the properties stated in the assertions hold locally uniformly, i.e. for each $f \in \mathcal{F}_\delta$ there exists a vicinity $V(f)$ such that these properties hold uniformly over this vicinity.

For each vicinity $V \in \mathcal{T}_\delta$ define the *local maximal risk* of an estimator $\tilde{f}_n(x)$ and *local minimax risk*:

$$R_n(\tilde{f}_n, V) = R_n(\tilde{f}_n, V, x) = \sup_{f\in V} \mathbf{E}_f(\tilde{f}_n((x) - f(x))^2, \tag{3}$$

$$r_n(V) = r_n(V,x) = \inf_{\tilde{f}_n} R_n(\hat{f}_n, V, x), \tag{4}$$

where the infimum is taken over all estimators $\tilde{f}_n$. The estimator $\tilde{f}_n$ is called *locally asymptotically minimax* (or just *efficient*) at f if for any sufficiently small vicinity V of f

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

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

$$\phi_n(y) = r(y)s_n(y), \tag{5}$$

where for some $a, \alpha > 0$, $m \geq 0$,

$$\begin{aligned} s_n(y) &= s_n(y,\delta,m,r) = \frac{\sin(a_n y)}{\pi y}, \qquad a_n = \left(\frac{\log n + m \log\log n}{2\delta}\right)^{1/r}, \\ r(y) &= r(y,a,\alpha) = \begin{cases} \exp\left(\frac{1}{a^{2\alpha}} - \frac{1}{(a^2-y^2)^\alpha}\right), & -a<y<a \\ 0, & y \notin (-a,a). \end{cases} \end{aligned} \tag{6}$$

Constant α is chosen in such a way that $\alpha/(\alpha+1) > r$, where r is the parameter from the definition of class (**??**). Constant a will be chosen later. In the sequel we will suppress the dependence on constants in notations.

Remark 2. Since

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

$\hat{\phi}_n(t)$ is nothing else but a smoothed indicator of $[-a_n, a_n]$. In words, convolution of a function with kernel ϕ_n in time domain corresponds to thresholding the Fourier transformation of the function in frequency domain.

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, p. 229). We adapt this result in simplified and suitable for our purposes form: for some $A_1, A_2 > 0$,

$$|\hat{r}(t)| \leq A_1 \exp\left\{-A_2 |t|^{\frac{\alpha}{\alpha+1}}\right\}. \tag{8}$$

Constants A_1, A_2 depend in general on a and α.

Define now the following estimator

$$\tilde{f}_n = \tilde{f}_n(x) = \int \phi_n(x-y) d\tilde{F}_n(y), \tag{9}$$

where $\phi_n(y)$ is defined by (**??**)-(**??**) and $\tilde{F}_n(y)$ is the well known Kaplan-Meier estimator:

$$\tilde{F}_n(y) = 1 - \prod_{i:\, Z_{(i)} < y} \left(\frac{n-i}{n-i+1}\right)^{\Delta_{(i)}}, \tag{10}$$

with the conventions $0^0 = 1$. Here $Z_{(i)}$'s denote the ordered sequence of Z_i's and $\Delta_{(i)}$'s are correspondent indicators.

Denote $a \wedge b = \min\{a, b\}$ and $\tau_F = \inf\{y : \ F(y) = 1\}$. In the next Theorem the efficiency of the estimator $\tilde{f}_n$ and the asymptotic behavior of the local minimax risk are established.

Theorem 1. *Let $f \in \mathcal{F}_\delta$ be such that $x + \delta \leq \tau_G \wedge \tau_F$ for some $\delta > 0$ and distribution function G is continuous at point x. Then for any sufficiently small vicinity $V(f)$*

$$\lim_{n\to\infty} \frac{n}{(\log n)^{1/r}} r_n(V) = \sup_{f \in V} \frac{f(x)}{(2\delta)^{1/r}\pi(1-G(x))}$$

*and the estimator $\tilde{f}_n$ defined by (**??**) is efficient.*

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

Theorem 1 implies also that

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

3 Auxiliary results

In this section we provide some technical results which we will need in the proof of Theorem 1.

Lemma 1. *Let $\phi_n(y)$ be defined by (**??**) and $h(y)$ be a locally integrable, continuous at x function. Then, as $n \to \infty$, the relation*

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

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

Proof. The first equality can be proved by standard arguments. To prove the second equality, write

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

□

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

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

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

Proof. By (**??**), we obtain the following uniform bound:

$$\begin{aligned}&\int \phi_n(x-y)dF(y) - f(x) \\ &= \left(\frac{1}{2\pi}\int e^{-itx}(\hat{\phi}_n(t)-1)\hat{f}(t)dt\right)^2\end{aligned}$$

$$
\begin{aligned}
&\le \frac{1}{2\pi}\int \exp(2\delta|t|^r)|\hat{f}(t)|^2 dt \cdot \frac{1}{2\pi}\int \exp(-2\delta|t|^r)|\hat{\phi}_n(t)-1|^2 dt \\
&\le C_1 \int |\hat{\phi}_n(t)-1|^2 \exp(-2\delta|t|^r)dt \\
&\le C_1 \int_{-a_n}^{a_n} |\hat{\phi}_n(t)-1|^2 \exp(-2\delta|t|^r)dt + C_2\int_{|t|\ge a_n} \exp(-2\delta|t|^r)dt \\
&= 2C_1 \int_0^{a_n} |\hat{\phi}_n(t)-1|^2 \exp(-2\delta|t|^r)dt + O\left(\frac{1}{n(\log n)^m}\right). \qquad (11)
\end{aligned}
$$

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

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

for $t \in [0, a_n]$. Now using the last inequality, (**??**), (**??**) and fact that $\int \hat{r}(u)du = r(0)2\pi = 2\pi$, we have that, for $t \in [0, a_n]$,

$$
\begin{aligned}
|\hat{\phi}_n(t)-1| &= \left|(2\pi)^{-1}(\hat{r} * I_{[-a_n,a_n]})(t) - 1\right| \\
&= (2\pi)^{-1}\left|\int (I_{[-a_n,a_n]}(t-u)-1)\hat{r}(u)du\right| \\
&= (2\pi)^{-1}\left|\int_{|t-u|>a_n} \hat{r}(u)du\right| \le 2(2\pi)^{-1}\int_{u>a_n-t}|\hat{r}(u)|du \\
&\le C_3 \int_{a_n-t}^{\infty} \exp\left\{-A_2 u^{\frac{\alpha}{\alpha+1}}\right\}du \\
&\le C_4 \exp\left\{-C_5(a_n-t)^{\frac{\alpha}{\alpha+1}}\right\}. \qquad (12)
\end{aligned}
$$

In view of (**??**), combining (**??**) with c_r-inequality (see Loève (1963), p. 155) and fact that $\alpha/(\alpha+1) > r$ completes the proof:

$$
\begin{aligned}
&\int_0^{a_n} |\hat{\phi}_n(t)-1|^2 \exp(-2\delta|t|^r)dt \\
&\le C_6 \int_0^{a_n} \exp\left\{-2C_5(a_n-t)^{\frac{\alpha}{\alpha+1}} - 2\delta t^r\right\}dt \\
&= C_6 \int_0^{a_n} \exp\left\{-2C_5 y^{\frac{\alpha}{\alpha+1}} - 2\delta(a_n-y)^r\right\}dy \\
&\le C_6 \int_0^{a_n} \exp\left\{-2C_5 y^{\frac{\alpha}{\alpha+1}} - 2\delta(a_n^r-y^r)\right\}dy \\
&\le C_6 e^{-2\delta a_n^r}\int_0^{\infty} \exp\left\{-2C_5 y^{\frac{\alpha}{\alpha+1}} + 2\delta y^r\right\}dy = \frac{C_7}{n(\log n)^m}.
\end{aligned}
$$

□

The following result which is due to Lo, Mack and Wang (1989) gives a representation of the Kaplan-Meier estimator by an average of independent random variables. For brevity sake, introduce some notations:

$$\bar{F}(y) = 1 - F(y), \qquad H(y) = \mathbf{P}\{Z_1 \le y\} = 1 - \bar{F}(y)\bar{G}(y),$$

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

$$\xi_i(t) = \xi(Z_i, \Delta_i, t) = -\bar{F}(t)g(Z_i \wedge t) + \frac{\bar{F}(t)}{\bar{H}(t)} I\{Z_i \le t, \Delta_i = 1\}.$$

Lemma 3 (Lo, Mack and Wang (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 $\alpha \ge 1$

$$\sup_{0 \le y \le T} \mathbf{E}|R_n(y)|^{\alpha} = O\left((\log n/n)^{\alpha}\right) \quad \text{as} \quad n \to \infty.$$

Remark 3. Tracing the proof of this lemma, one can show that this representation holds locally uniformly in the topology generated by the distance in variation.

We are going to use this result in the proof of the theorem. Note that uniformly in $0 \le y \le T$ the random variables $\xi_i(y)$'s are bounded, independent and by straightforward calculations,

$$\mathbf{E}\xi_i(y) = 0, \qquad \mathbf{Cov}(\xi_i(y), \xi_i(u)) = \bar{F}(y)\bar{F}(u)g(y \wedge u). \tag{13}$$

4 Proof of Theorem

Upper bound. Choose constant a in the definition of the estimator in such a way that $0 < a < \delta$. Now using the integration by parts, Lemma **??** (see also Remark 3 and (**??**)) and the elementary inequality

$$(a+b)^2 \le (1+\gamma)\,a^2 + (1+\gamma^{-1})\,b^2, \qquad 0 < \gamma \le 1, \tag{14}$$

we have that, uniformly over sufficiently small vicinity of f,

$$\begin{aligned} &\mathbf{E}\left(\int \phi_n(x-y)d(\tilde{F}_n(y) - F(y))\right)^2 \\ &\le \frac{(1+\gamma_n)}{n} \iint \bar{F}(t)\bar{F}(u)g(u \wedge t)d\phi_n(x-t)d\phi_n(x-u) \end{aligned}$$

$$
\begin{aligned}
&+ (1+\gamma_n^{-1})\,\mathbf{E}\int (R_n(y))^2\, d\phi_n(x-y) \\
\leq\ & \frac{(1+\gamma_n)}{n}\iint \bar F(t)\bar F(u) g(u\wedge t) d\phi_n(x-t)d\phi_n(x-u) \\
&+ \frac{(1+\gamma_n^{-1})C_1(\log n)^2}{n^2},
\end{aligned}
$$

where γ_n is to be chosen later. We are in a position to apply Lemma **??** because kernel $\phi(x-y)$ has finite support $[x-a, x+a]$ such that $x+a < \tau_G \wedge \tau_F$ uniformly in a vicinity of f.

Tedious but straightforward calculations lead to

$$
\begin{aligned}
&\iint \bar F(t)\bar F(u) g(u\wedge t) d\phi_n(x-t)d\phi_n(x-u) \\
&= \int \frac{\phi_n^2(x-t)dF(t)}{1-G(t)} + \iint \phi_n(x-t)\phi_n(x-u)h(t\wedge u)dF(u)dF(t) \\
&= \int \frac{\phi_n^2(x-t)dF(t)}{1-G(t)} + \int \left(\int_t^\infty \phi_n(x-u)dF(u)\right)^2 dh(t) \\
&\leq \int \frac{\phi_n^2(x-t)dF(t)}{1-G(t)} - \int \left(\int_t^\infty \phi_n(x-u)dF(u)\right)^2 \frac{dG(t)}{\bar F(t)(\bar G(t))^2} \\
&\leq \int \frac{\phi_n^2(x-t)dF(t)}{1-G(t)}.
\end{aligned}
$$

Now we evaluate the risk of the estimator (**??**). From the last two relations and again elementary inequality (**??**) it follows that, uniformly in a vicinity of f,

$$
\begin{aligned}
&\mathbf{E}_f\left(\tilde f_n(x) - f(x)\right)^2 \\
&= \mathbf{E}_f\left(\int \phi_n(x-y)d(\tilde F_n(y)-F(y)) + \int \phi_n(x-y)dF(y) - f(x)\right)^2 \\
&\leq \frac{(1+\gamma_n)^2}{n}\int \frac{\phi_n^2(x-t)dF(t)}{1-G(t)} + \frac{(\gamma_n^{-1}+2+\gamma_n)C_1(\log n)^2}{n^2} \\
&\quad + (1+\gamma_n^{-1})\left(\int \phi_n(x-y)dF(y) - f(x)\right)^2.
\end{aligned}
$$

We choose now γ_n such that $\gamma_n \to 0$ and $(\gamma_n(\log n)^{1/r})^{-1} = o(1)$ as $n\to\infty$. Using the last relation, Lemmas **??** and **??**, we obtain that

$$
\limsup_{n\to\infty} \frac{n}{(\log n)^{1/r}} \mathbf{E}_f(\tilde f_n(x) - f(x))^2 \leq \frac{f(x)}{(2\delta)^{1/r}\pi(1-G(x))} \qquad (15)
$$

uniformly over sufficiently small vicinity of f.

Lower bound. Let $f_0(y)$ be an arbitrary density from vicinity V. Consider the following family of functions:

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

where $|\theta| \leq \theta_n$, $\theta_n = n^{-1/2}(\log n)^{-1/(4r)}$, $\phi_n(y)$ is defined by (**??**) and $\bar{\phi}_n(x) = \int \phi_n(x-y) f_0(y)dy$.

If X_i is distributed according to density $f_\theta(y)$, then the correspondent 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 Lemma describes the Fisher information $I(\theta)$ about θ contained in the observation (Z, Δ). The proof of this Lemma is almost identical to the proof of Proposition 2 in Belitser (1996) and is omitted.

Lemma 4. *As $n \to \infty$, the relation*

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

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

Now we proceed proving 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 a 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 $[-\theta_n, \theta_n]$ and the Fisher information $I(\nu) = I_n(\nu) = I_0\theta_n^{-2}$.

Obviously, $f_\theta \in V$ for sufficiently large n. Applying now the van Trees inequality for the Bayes risk below (see Gill and Levit (1995)) and Lemma **??**, we obtain that for sufficiently large n

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

or

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

Function f_0 was chosen arbitrarily from the vicinity V and hence, by the same reasoning, this relation is valid for any function $f \in V$. Therefore

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

and combining this with (**??**) proves Theorem. □

References

Belitser, E. (1996), Efficient estimation of analytic density under random censorship. Technical Report 953, Department of Mathematics, Utrecht University.

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

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

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

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

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

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.

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

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