On Aggregation of Uncensored and Censored Observations

Sam Efromovich[*]

Department of Mathematical Sciences, UTDallas, Richardson, TX 75080 USA

Abstract—In survival analysis a random right-censoring partitions data into uncensored and censored observations of the lifetime of interest. The dominance of uncensored observations is a familiar methodology in nonparametric estimation motivated by the classical Kaplan–Meier product-limit and Cox partial likelihood estimators. Nonetheless, for high rate censoring it is of interest to understand what, if anything, can be done by aggregating uncensored and censored observations for the staple nonparametric problems of density and regression estimation. The oracle, who knows distribution of the censoring lifetime, can use each subsample for consistent estimation and hence may shed light on the aggregation. The oracle's asymptotic theory reveals that density estimation, based on censored observations, is an ill-posed problem with slower rates of risk convergence, the ill-posedness occurs in frequency-domain, its severity increases with frequency, and accordingly a special aggregation on low frequencies may be beneficial. On the other hand, censored observations are not ill-posed for nonparametric regression and the aggregation is feasible. Based on these theoretical results, methodology of aggregation in frequency domain is developed and proposed estimators are tested on simulated and real examples.

1. INTRODUCTION

Consider a lifetime of interest T which is right-censored by another lifetime C. Then instead of a direct sample from T we get a sample from pair $(V, \Delta) := (\min(T, C), I(T \leq C))$ where $I(\cdot)$ is the indicator function. Accordingly, the random right-censoring partitions data into uncensored observations when we observe realizations of T and censored observations when we observe realizations of C. Assuming that T and C are independent continuous random variables, denote their densities and survival functions as f^T, S^T and f^C, S^C. Then the joint mixed density of (V, Δ) is

$$f^{V,\Delta}(t, \delta) = [f^T(t) S^C(t)]^\delta [f^C(t) S^T(t)]^{1-\delta}, \quad (t, \delta) \in [0, \infty) \times \{0, 1\}. \tag{1}$$

Note the symmetry in the formula with respect to T and C, it reflects the fact that C right-censors T and T right-censors C.

Statistical literature, devoted to analysis of the lifetime of interest T, treats the uncensored subsample as the dominate one. For instance, the classical Kaplan–Meier estimator of the survival function has jumps only at uncensored observations of T, and moreover Kaplan and Meier [30] refers to a censored T as a "loss". Further, a large portion of statistical literature treats censored observations as "missing data", and then uses the Buckley–James imputation of censored observations by statistics based on uncensored observations.

The aim of the paper is to understand how and when censored observations may be aggregated with uncensored for nonparametric estimation of the density f^T and nonparametric regression of T on a predictor. Of course, both these subsamples are needed for consistent estimation. Accordingly, the oracle's approach is used. The oracle knows distribution of the censoring variable C and may use uncensored and censored observations separately to answer the raised question. Then the recommended oracle's estimators are mimicked by corresponding data-driven nonparametric estimators based on an

---

[*]E-mail: efrom@utdalls.edu

estimated distribution of C. At the same time, if the distribution of C is known or may be estimated based on an extra sample from C, then the oracle's approach can be used directly. A practical example of the latter possibility will be presented in Section 3.

Let us stress that it is of a special interest to consider problems of density and regression estimation together because conclusions of the asymptotic theory for the problems are different. Namely, for density no aggregation is needed for asymptotically efficient estimation while for regression the aggregation is beneficial. At the same time, conclusions of numerical studies for small samples and high rates of censoring coalesce in terms of feasibility of aggregation for both density and regression estimation problems.

In the paper, with some obvious but not confusing abuse of the notions, a sample from (V, Δ) is called right-censored, its subsample with $\Delta = 1$ is called uncensored-data or uncensored observations because T is observed, and the complementary subsample with $\Delta = 0$ is called censored-data or censored observations because C is observed.

The context of the paper is as follows. Sections 2 and 3 are devoted to density and regression problems, respectively. Their structures are identical. The first subsection presents literature review, it is shown that the literature treats uncensored and censored subsamples differently with the former being the dominant source of information. The second subsection is the core mathematical statistics that contains the sharp minimax asymptotic theory. The theory provides both sharp constants and optimal rates of the MISE (mean integrated squared error) convergence. This section also explains the corresponding methodology of estimation. Further, it is of a special interest to show that this asymptotic theory is not just a complicated mathematical exercise, and it can be applied to simulated small samples and real-life practical examples with high rates of censoring. The latter is done in Subsections 3 and 4, respectively. The interested reader can even begin with Subsections 2.4 and 3.4.2 to check how the aggregation sheds a new light on longevity of a patient with small cell lung cancer. Proofs are in Section 4, and conclusions are in Section 5.

2. DENSITY ESTIMATION

The problem of estimation of the density f^T based on a right-censored sample of size n from the pair (V, Δ), defined in the Introduction, is considered. The structure of the section is outlined in the last paragraph of Section 1.

2.1. Literature Review

It is fair to say that the modern survival analysis, and distribution estimation in particular, are based on the pathbreaking product-limit methodology of Kaplan and Meier [30] for nonparametric estimation of survival function by a stepwise function with steps at uncensored lifetimes. The product-limit methodology is based on the understanding that censored observations are dominated by uncensored. Moreover, in that seminal paper censored observations are referred to as "losses." And sure enough, rigorous proof of the dominance was done later in [1, 4, 21, 38]. It took a bit longer to verify the dominance for density estimation. Efromovich [18] established that for large samples the oracle, who knows data and the distribution of censoring variable, can attain the sharp constant and rate of the MISE (mean integrated squared error) convergence using only uncensored observations. In other words, the oracle does not need censored observations for efficient estimation of the density, see also an interesting discussion in [6].

Despite all these results, there are still two unresolved issues. First, it is of interest to understand why the oracle does not use censored observations. Second, if the rate of censoring is high and the sample is small [3, 7, 9, 10, 13, 14, 50, 51], can the oracle use censored observations in an optimal way and aggregate them with uncensored ones? In what follows we present results that shed light on plausible answers. The theory, presented shorty in Subsection 2.2, shows that using censored observations yields consistent estimation, but rate of the MISE convergence is slower than for uncensored observations. In other words, estimation based on censored observations is ill-posed. This is the bad news. The good news is that the ill-posedness occurs in frequency-domain and low frequencies see only its onset. Accordingly, the oracle recommends to aggregate low-frequency Fourier coefficient estimates, based on uncensored and censored observations, and then use a corresponding series density estimate. In other words, the oracle states that the aggregation must be in frequency-domain and primarily on low frequencies. The interested reader will be able to check in Subsections 2.3 and 2.4 feasibility of these recommendation for simulated and real-life small samples, respectively.

2.2. Asymptotic Theory and Methodology

Let us recall the right-censored model. Estimation of f^T is based on a sample $(V_1, \Delta_1), \ldots, (V_n, \Delta_n)$ of n independent and identically distributed observations from the pair $(V, \Delta) := (\min(T, C), I(T \leq C))$. The right-censoring lifetime C partitions the sample into two subsamples where observations of either T or C are available. We refer to the two subsamples as uncensored and censored observations. Note that the number $N := \sum_{l=1}^{n} \Delta_l$ of uncensored observations has Binomial$(\mathbb{P}(\Delta = 1), n)$ distribution.

The main aim of this subsection is to explain what can and cannot be done by using uncensored and censored observations for estimating density f^T of a bounded lifetime of interest T. The oracle estimates the density over a finite interval, and without loss of generality it is assumed that the density is estimated over its support $[0, 1]$. It is also assumed that $S^C(1) > 0$, this allows consistent estimation. The oracle knows density f^C of the censoring variable, and accordingly can propose consistent estimators based on each of the two subsamples. In what follows $q_n := \lceil \ln(n + 20) \rceil$, $s_n := 3 + \lceil \ln(\ln(n + 3)) \rceil$, and $\lceil x \rceil$ is the smallest integer that larger or equal to x.

We begin with assumptions.

Assumption 1. *The lifetime T is independent of the censoring lifetime C.*

This is a standard assumption in the literature. Next, following [18], we introduce a class of estimated densities. Denote by $\{\varphi_0(t) := 1, \varphi_j(t) := 2^{1/2} \cos(\pi j t), j = 1, 2, \ldots\}$ the cosine basis on $[0, 1]$, and introduce a shrinking local Sobolev class of α-fold differentiable densities supported on $[0, 1]$,

$$\mathcal{F}_n := \mathcal{F}_n(f_0, \alpha, Q) := \Big\{ f : \ f(t) = f_0(t) + g(t) I(t \in [0, 1]),$$

$$g \in \mathcal{S}_1(\alpha, Q), \ |g(t)| \leq \min_{x \in [0,1]} f_0(t)/s_n, \ t \in [0, 1] \Big\}. \tag{2}$$

Here the anchor density f_0 is supported, continuous and positive on $[0, 1]$, and for $k \in \{0, 1\}$

$$\mathcal{S}_k(\alpha, Q) := \left\{ g : g(t) = \sum_{j=k}^{\infty} \theta_j \varphi_j(t), \ \sum_{j=k}^{\infty} (1 + (\pi j)^{2\alpha}) \theta_j^2 \leq Q < \infty, \ t \in [0, 1] \right\}. \tag{3}$$

The class $\mathcal{S}_0(\alpha, Q)$ is called the global Sobolev class, $\mathcal{S}_1(\alpha, Q) \subset \mathcal{S}_0(\alpha, Q)$, and $\mathcal{S}_1(\alpha, Q)$ is the class of Sobolev functions integrated to zero. As we will see shortly in Theorem 1, we need to use the local Sobolev class (2) because the Fisher information for right-censored observations depends on an underlying density. Let us also stress that f_0 is not necessarily the underlying density of interest, it simply anchors all underlying densities f^T in its vanishing in L_1-norm vicinity.

Theorem 1. *Consider a sample of size n from right-censored pair $(V, \Delta) := (\min(T, C), I(T \leq C))$, and the problem is to estimate density f^T of the lifetime of interest T under the MISE criterion. Let Assumption (1) hold, density f^C is positive and continuous on $[0, 1]$, $S^C(1) > 0$, and the oracle knows the right-censored data, density f^C and function class $\mathcal{F}_n$. Then*

$$\inf_{\tilde{f}^*} \sup_{f^T \in \mathcal{F}_n} \mathbb{E}_{f^T} \left\{ (n/d_u)^{2\alpha/(2\alpha+1)} \int_0^1 (\tilde{f}^*(t) - f^T(t))^2 dt \right\} \geq P_u (1 + o_n(1)), \tag{4}$$

where the infimum is over all possible oracle-estimators $\tilde{f}^*$. Furthermore, the lower bound is sharp and it is attainable by an oracle-estimator $\tilde{f}_u^*$ based solely on uncensored observations. If the oracle uses only censored observations and $\alpha > 1$, then

$$\inf_{\tilde{f}_c^*} \sup_{f^T \in \mathcal{F}_n} \mathbb{E}_{f^T} \left\{ (n/d_c)^{2\alpha/(2\alpha+3)} \int_0^1 (\tilde{f}_c^*(t) - f^T(t))^2 dt \right\} \geq P_c (1 + o_n(1)), \tag{5}$$

and the lower bound is sharp. Accordingly, censored observations are ill-posed, with respect to uncensored observations, and using only them slows down rate of the MISE convergence from $n^{-2\alpha/(2\alpha+1)}$ *to* $n^{-2\alpha/(2\alpha+3)}$. *In* (4) *and* (5)

$$P_u := \frac{Q^{1/(2\alpha+1)}\alpha^{2\alpha/(2\alpha+1)}(2\alpha+1)^{1/(2\alpha+1)}}{[\pi(\alpha+1)]^{2\alpha/(2\alpha+1)}}, \tag{6}$$

$$P_c := [Q(2\alpha+3)]^{3/(2\alpha+3)}(1/3)[\alpha/(\pi(\alpha+3))]^{2\alpha/(2\alpha+3)}, \tag{7}$$

$$d_u := \int_0^1 \frac{f^T(t)}{S^C(t)}dt, \; d_c := \int_0^1 \frac{S^T(t)}{f^C(t)}dt. \tag{8}$$

Now let us present oracle's estimators that attain the sharp lower bounds (4) and (5). We begin with the one based on uncensored observations, and note that we are using subscript u to highlight that. Set

$$\tilde{\theta}^*_{u0} := 1, \quad \tilde{\theta}^*_{uj} := n^{-1}\sum_{l=1}^{n}\Delta_l\frac{\varphi_j(V_l)}{S^C(V_l)}, \quad j \geq 1. \tag{9}$$

Then the oracle-estimator based on uncensored observations is

$$\hat{f}^*_u(t) := \sum_{j=0}^{J_n}\tilde{\theta}^*_{uj}I((\tilde{\theta}^*_{uj})^2 > c_{TH}d_un^{-1})\varphi_j(t) + \sum_{j=J_n+1}^{J^*_{un}}(1-(j/J^*_{un})^\alpha)\tilde{\theta}^*_{uj}\varphi_j(t), \tag{10}$$

where $J_n := 4\lceil\ln(n+3)\rceil$, c_{TH} is a positive constant, and

$$J^*_{un} := \lceil[(n/d_u)Q\pi^{-2\alpha}(\alpha+1)(2\alpha+1)/\alpha]^{1/(2\alpha+1)}\rceil. \tag{11}$$

Note that in (10) the classical hard thresholding is used on low frequencies and the shrinkage on high frequencies. For censored observations, consider the sine basis $\psi_j(t) := 2^{1/2}\sin(\pi jt)$, $j = 1, 2, \ldots$ Using subscript c to highlight that a statistic is based on censored observations, set

$$\tilde{\theta}^*_{c0} := 1, \quad \tilde{\theta}^*_{cj} := 2^{1/2} - n^{-1}(\pi j)\sum_{l=1}^{n}(1-\Delta_l)\frac{\psi_j(V_l)}{f^C(V_l)}, \quad j \geq 1. \tag{12}$$

Then the oracle-estimator based on censored observations is

$$\hat{f}^*_c(t) := \sum_{j=0}^{J_n}\tilde{\theta}^*_{cj}I((\tilde{\theta}^*_{cj})^2 > c_{TH}d_cn^{-1})\varphi_j(t) + \sum_{j=J_n+1}^{J^*_{cn}}(1-(j/J^*_{cn})^\alpha)\tilde{\theta}^*_{cj}\varphi_j(t), \tag{13}$$

where

$$J^*_{cn} := \lceil[(n/d_c)Q\pi^{-2\alpha-2}(\alpha+3)(2\alpha+3)/\alpha]^{1/(2\alpha+3)}\rceil. \tag{14}$$

Theorem 2. *Let Assumption* 1 *hold. Suppose that the anchor* $f_0 \in \mathcal{S}_0(\alpha+\beta, Q')$, $\beta > 0$, $Q' < \infty$ *where the global Sobolev class* $\mathcal{S}_0$ *is defined in* (3). *Consider* d_u *and* d_c *defined in* (8). *If* d_u *is finite, then the oracle-estimator* (10), *based on uncensored observations, attains the lower bound* (4). *If* d_c *is finite, then the oracle-estimator* (13), *based on censored observations, attains the lower bound* (5). *Further, the proposed Fourier coefficient estimates are unbiased and satisfy*

$$\mathbb{E}_f\{(\tilde{\theta}^*_{uj} - \theta_j)^2\} = n^{-1}\sigma^2_{uj}, \quad \sigma^2_{uj} = d_u(1+o_j(1)) \tag{15}$$

and

$$\mathbb{E}_f\{(\tilde{\theta}^*_{cj} - \theta_j)^2\} = n^{-1}(\pi j)^2\sigma^2_{cj}, \quad \sigma^2_{cj} = d_c(1+o_j(1)). \tag{16}$$

These properties yield the following unbiased aggregation of the two Fourier coefficient estimates based on uncensored-data and censored-data,

$$\tilde{\theta}^*_{aj} := \tilde{\theta}^*_{uj}\frac{(\pi j)^2\sigma^2_{cj}}{(\pi j)^2\sigma^2_{cj} + \sigma^2_{uj}} + \tilde{\theta}^*_{cj}\frac{\sigma^2_{uj}}{(\pi j)^2\sigma^2_{cj} + \sigma^2_{uj}}, \tag{17}$$

with the mean squared error satisfying

$$\mathbb{E}_f\{(\tilde{\theta}_{aj}^* - \theta_j)^2\} = \sigma_{uj}^2 \frac{(\pi j)^2 \sigma_{cj}^2}{(\pi j)^2 \sigma_{cj}^2 + \sigma_{uj}^2} =: \sigma_{uj}^2(1 - \nu_j), \quad 0 < \nu_j < (\pi j)^{-2}[d_u/d_c](1 + o_j(1)). \quad (18)$$

The assertions of Theorems 1 and 2 highlight and quantify ill-posedness of censored lifetimes with respect to uncensored. At the same time, formula (16) implies that for small j we see only the onset of ill-posedness. Accordingly, the frequency-domain aggregation (17) is feasible, line (18) explains its benefits, and the theory sheds light on the numerical study presented in the next subsections.

The nice feature of the proposed series oracle-estimator is that it implies the following algorithm of nonparametric estimation that will be referred to as E-estimator.

Algorithm of E-estimation. Let $f(x)$, $x \in [0, 1]$ be a square integrable function of interest. There are three steps that E-estimator makes for its estimation using the cosine basis $\{\varphi_j\}$.

Step 1. The function can be written as $f(x) = \sum_{j=0}^{\infty} \theta_j \varphi_j(x)$. Here $\theta_j := \int_0^1 f(x)\varphi_j(x)dx$ are Fourier coefficients of f. Suggest a sample mean estimator $\hat{\theta}_j$ of Fourier coefficients $\theta_j := \int_0^1 f(x)\varphi_j(x)dx$. Then calculate a corresponding sample variance estimator $\hat{v}_{jn}$ of the variance $v_{jn} := \mathrm{Var}(\hat{\theta}_j)$ of the sample mean estimator.

Step 2. The E-estimator is defined as $\hat{f}(x) := \sum_{j=0}^{\hat{J}} \hat{\theta}_j I(\hat{\theta}_j^2 > c_{TH}\hat{v}_{jn})\varphi_j(x)$. Here the empirical cutoff is $\hat{J} := \mathrm{argmin}_{0 \leq J \leq c_{J0} + c_{J1}\ln(n)}\{\sum_{j=0}^{J}[2\hat{v}_{jn} - \hat{\theta}_j^2]\}$, and c_{J0}, c_{J1}, and c_{TH} are parameters (non-negative constants).

Step 3. If there are bona fide restrictions on $f(x)$ (for instance, the probability density is nonnegative and integrated to one, or it is known that the function is monotonic) then a projection of $\hat{f}(x)$ on the bona fide function class is performed, see [17].

Note that Steps 2 and 3 in construction of the E-estimator are the same for all nonparametric statistical problems. As a result, as soon as a sample mean estimator of Fourier coefficients is proposed, this Fourier estimator yields the corresponding E-estimator. We will see shortly how the E-estimator performs.

Now let us explain a general methodology of how to adapt to unknown smoothness of f^T by using a a block-shrinkage estimator. To define the estimator, suppose that the oracle recommends to use a Fourier estimator $\bar{\theta}_j$ of θ_j satisfying $\mathbb{E}\{(\bar{\theta}_j - \theta_j)^2\} = dn^{-1}(1 + o_n(1) + o_j(1))$. Set $b_1 := J_n + 1$, $b_{k+1} := b_k + \lceil (1 + 1/s_n)^k \rceil$, $k = 1, 2, \ldots$, $B_k := \{j : b_k \leq j < b_{k+1}\}$, $L_k := b_{k+1} - b_k$, K_n is the smallest integer such that $b_{K_n} \geq n^{1/3}s_n$,

$$\bar{\Theta}_k := L_k^{-1} \sum_{j \in B_k} \bar{\theta}_j^2. \quad (19)$$

Also denote by $\bar{d}$ an estimate of d such that $\mathbb{E}\{(\bar{d} - d)^2\} = o_n(1)$. For instance, we can set $\bar{d} := n^{-1}\sum_{l=1}^{n}\Delta_l[S^C(V_l)]^{-2}$ for uncensored observations, and $\bar{d} := n^{-1}\sum_{l=1}^{n}(1 - \Delta_l)[f^C(V_l)]^{-2}$ for censored observations. Then the blockwise-shrinkage estimator that adapts to parameters (α, Q) and matches the MISE of a corresponding oracle's estimator is

$$\bar{f}(t) := \sum_{j=0}^{J_n} \bar{\theta}_j \varphi_j(t) + \sum_{k=1}^{K_n} \sum_{j \in B_k} \frac{\bar{\Theta}_k - \bar{d}n^{-1}}{\bar{\Theta}_k} I(\bar{\Theta}_k \geq (1 + 1/q_k)\bar{d}n^{-1}). \quad (20)$$

If the distribution of C is unknown, then the following method of moments estimator of the survival function S^C is used,

$$\hat{S}^C(t) := \exp\left\{-n^{-1}\sum_{l=1}^{n}(1 - \Delta_l)I(V_l \leq t)/\hat{S}^V(V_l)\right\}, \quad (21)$$

where

$$\hat{S}^V(t) := n^{-1} \sum_{l=1}^{n} I(V_l \geq t). \qquad (22)$$

Note that $\hat{S}^V(V_l) \geq n^{-1}$, and hence it can be used in the denominators of (21).

Lemma 1. *Consider estimation of $S^C(t)$ for $t \in [0, a]$. Suppose that Assumption* 1 *holds and $S^V(a) > 0$. Then there exist finite positive constants B_*, B and a sequence of finite constants B_k such that for any $l = 1, \ldots, n$, $z \in [0, a]$, positive ν and integer k,*

$$\mathbb{E}\{[\hat{S}^C(V_l) - S^C(V_l)]|V_l = z\} \leq B_* n^{-1}, \qquad (23)$$

$$\mathbb{P}(|\hat{S}^C(V_l) - S^C(V_l)| > \nu|V_l = z) \leq Bne^{-n\nu^2/B}, \qquad (24)$$

$$\mathbb{E}\{[\hat{S}^C(V_l) - S^C(V_l)]^{2k}|V_l = z\} \leq B_k n^{-k}. \qquad (25)$$

Lemma 1 will be proved using familiar properties of Bernoulli sums, and the interested reader can compare the assertion and simplicity of the proof with the beautiful and mathematically involved theory of product-limit survival estimators, see a nice exposition in [2]. One more remark about Lemma 1 is due. Its assumption is standard and requires that we are considering an interval $[0, a]$ such that $S^V(a) > 0$, and note that $S^V(a) = S^T(a)S^C(a)$.

Due to the symmetry between estimating distributions of C and T, the density f^C can be estimated by the same estimator as f^T only with Δ being replaced by $1 - \Delta$.

We have defined all estimates of the nuisance functions used by the oracle.

2.3. Numerical Study

Three rows of diagrams in Fig. 1 present results of different simulations, the simulations and diagrams are explained in the caption, and all estimates are data-driven. The top row presents the case when 23.5% of observations are censored (the theoretical $\mathbb{P}(\Delta = 0) = 0.25$), the underlying density f^T is the Bathtub (the solid line and see its discussion in [28]), and the censoring distribution is Uniform(0, 1.5). The short-dashed line is the hidden-data density estimate based on underlying (hidden) observations of T, this data-driven estimate is from R-package [20] and it is used as a benchmark. As we see, the hidden-data estimate is good and it indicates that the underlying sample is reasonable. All other estimates are based on the right-censored data. The circles show estimates of $f^C(V_l), l = 1, \ldots, n$. The dotted line is the uncensored-data estimate of f^T, it is based on uncensored observations, shown by the circles in the top-left diagram with $\Delta = 1$, and on estimated survival function of C. This estimate is also from R-package [20]. Visualization of the uncensored observations supports the estimate. Now let us look at the 47 censored observations shown in the left-top diagram by the circles with $\Delta = 0$. Visual analysis does not help us to see the underlying density because censored observations are ill-posed. Indeed, according to (1), censored observations allow us to evaluate S^T, and then its derivative yields the density. This is why it is difficult to visualize the underlying density in censored observations. Nonetheless, despite the small sample size and ill-posedness, let us look at the dot-dashed line which is the proposed density estimate based on the censored observations. The estimate is surprisingly good, and it will be explained shortly why such an outcome is possible. The long-dashed line is the proposed aggregated estimate. The middle row of diagrams shows a similar experiment with the same underlying density and the larger rate of censoring, here the theoretical $\mathbb{P}(\Delta = 0) = 0.38$. The particular simulation is chosen to show that despite the larger number of censored observations, the censored-data estimate (the dot-dashed line) is clearly worse than the others. This outcome is a real possibility due to the ill-posedness of censored observations. At the same time, the aggregated estimate is very good, and this is the main message of this simulation. The bottom row is devoted to estimation of the Bimodal density defined in ([20], p. 32). Here $\mathbb{P}(\Delta = 0) = 0.5$, and for the particular simulation the rate of censoring is 48%. Despite the larger number of censored observations, the estimate based on censored observations (the dot-dashed line) is dramatically worse than the estimate based on uncensored observations (the dotted line). Repeated simulations have indicated similar outcomes.
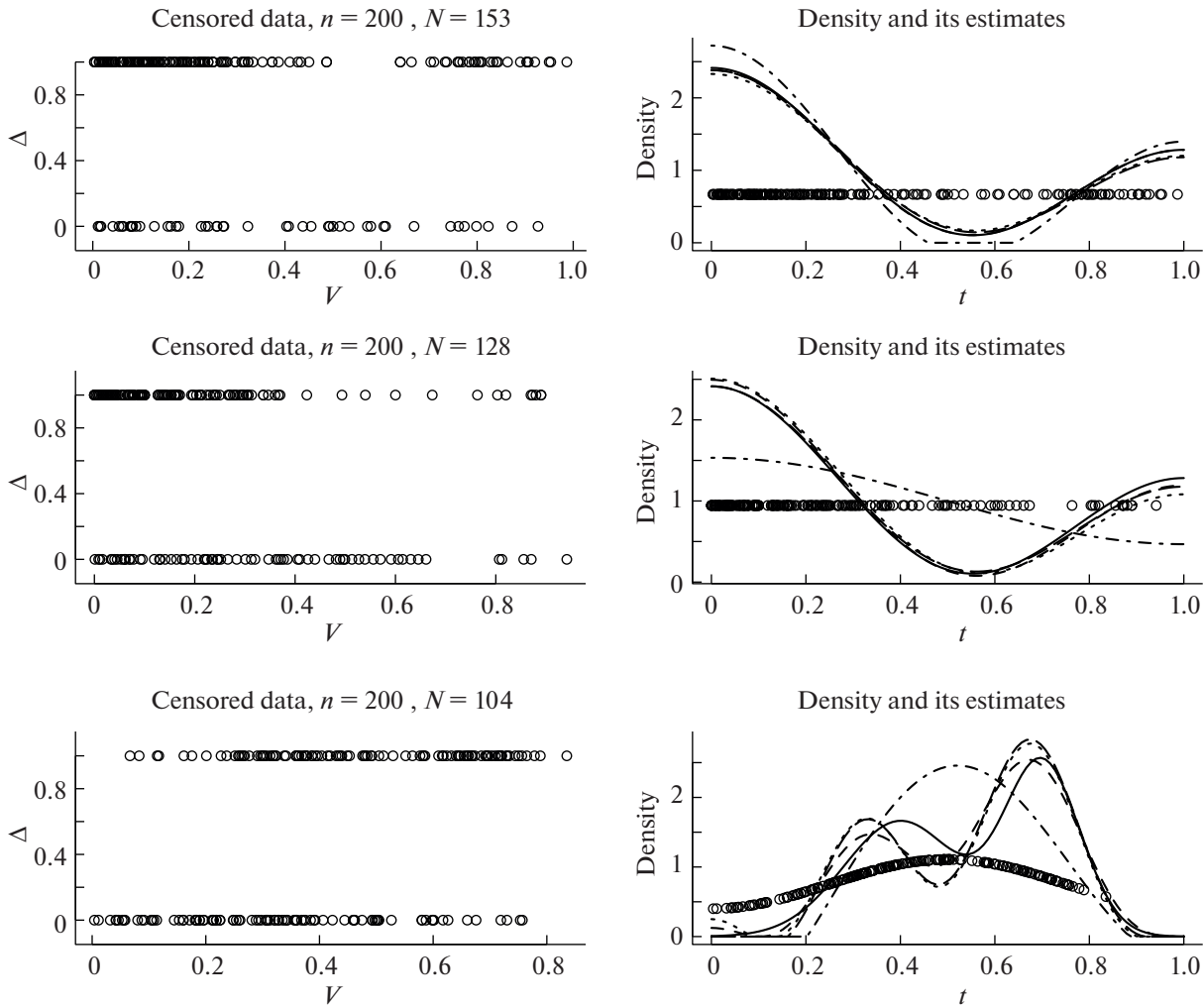
Fig. 1. Density estimation for three simulated examples shown in the corresponding rows. In each row the left diagram shows pairs (V_l, Δ_l), $l = 1, \ldots, n$ as well as the sample size n and the number of uncensored observations $N := \sum_{l=1}^{n} \Delta_l$. In these three experiments the censoring distributions are Uniform$(0, b)$ with b equal to 1.5, 1.01, and 1.1, respectively. A right diagram shows estimates of $f^C(V_l)$ by the circles. An underlying density f^T is shown by the solid line, it is the Bathtub in the two top diagrams and the bimodal in the bottom digram. The short-dashed, dotted, dot-dashed, and long-dashed lines are hidden-data, uncensored-data, censored-data, and aggregated estimates, respectively.

Let us explain the difference between the top two and the bottom experiments. The Bathtub density, considered in the two top experiments, is a low-frequency curve while the Bimodal is a high-frequency curve, see ([17], Sect. 3.3). This explains why estimates based on ill-posed censored observations may be visually appealing for the Bathtub density and not for the Bimodal. Now let us look at the aggregated estimates (the long-dashed lines). It is plain to realize that they are not from a class $\{\tilde{f}_A = \lambda \tilde{f}_1 + (1 - \lambda) \tilde{f}_2, \lambda \in [0, 1]\}$ of traditionally studied estimates aggregated in the time-domain, see a discussion and mathematically beautiful results in [40, 43, 47]. Instead, a special aggregation in frequency-domain is used, and this is why the aggregation may be beneficial even if a censored-data estimate is not good as in the case of the two bottom diagrams.

Are there situations when for small samples the ill-posed censored-data estimate may outperform the uncensored-data estimate? Figure 2 presents such examples. The left column of diagrams exhibits a particular outcome for an experiment identical to the top one in Fig. 1 only here the uniform C is replaced by $C = 0.01 + Z$, $f^Z(t) = [1 + 0.8 \cos(\pi t)]I(t \in [0, 1])$. Due to this monotonically decreasing density there are no observations of V larger 0.77 and just one uncensored observation larger 0.43. This creates a challenge for estimating right tail of the density, and this challenge is emphasized by the
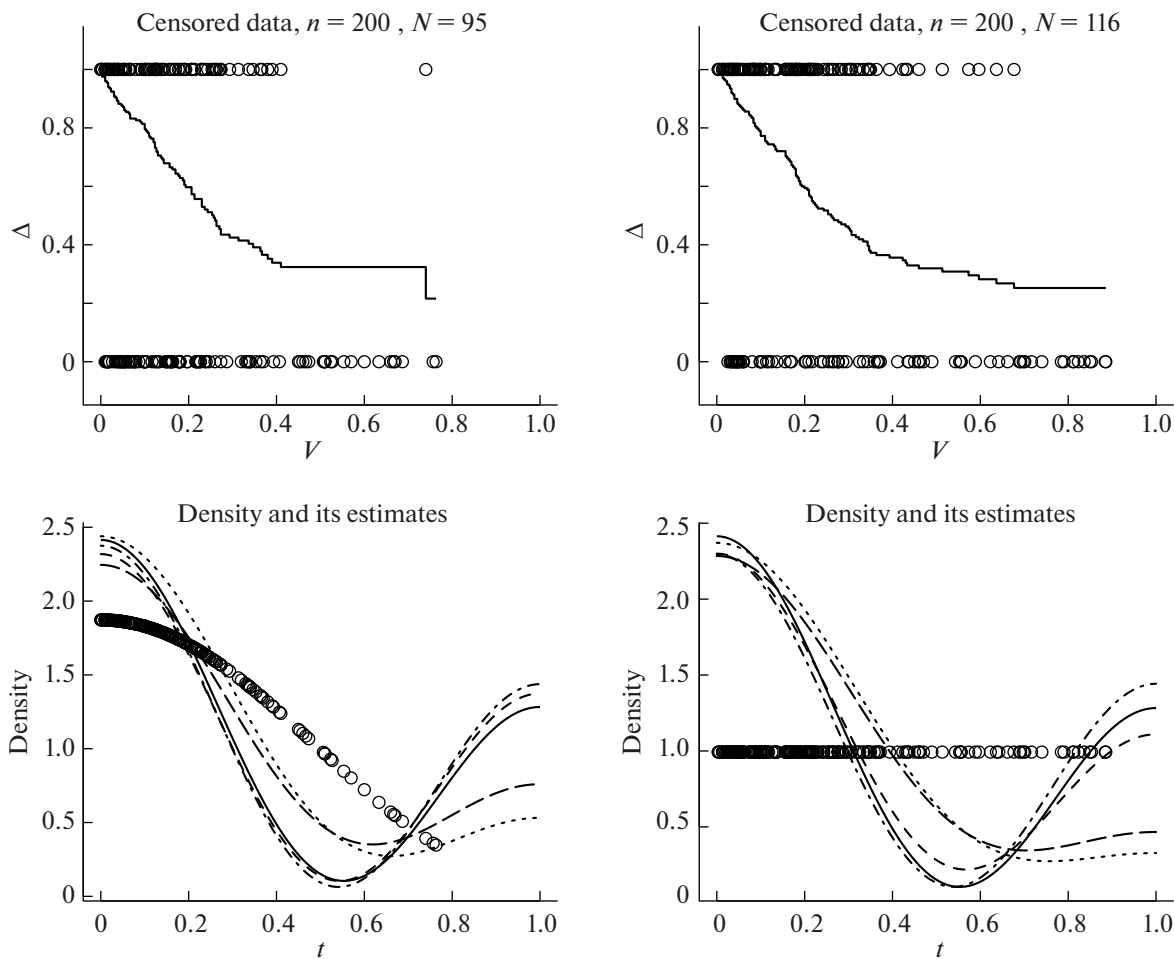
Fig. 2. Density estimation for two simulated examples with high censoring of right tail. Curves are the same as in Fig. 1, and the data diagrams additionally show the Kaplan−Meier cdf estimate.

classical Kaplan−Meier estimator of the cdf shown in the top diagram by the stepwise solid line. The uncensored-data estimate (the dotted line) cannot correctly evaluate the right tail but the censored-data estimate (the dot-dashed line) may and does. Further, note that the censored-data estimate also helps the aggregated estimate (the long-dashed line) to better exhibit the right tail. At the same time, all estimates nicely show the left tail of the underlying Bathtub density (the solid line). It is also insightful to use this example and shed extra light on the theory presented in Theorem 2. A direct calculation shows that for the problem at hand the coefficients of difficulty (8) for the uncensored and censored observations are $d_u = 7.9$ and $d_c = 0.4$, respectively. At first glance this creates a huge advantage for the uncensored-data estimator, but the core issue is that this estimator is ill-posed and lines (15) and (16) allow us to appreciate that. Indeed, while the constant in variance of the uncensored-data Fourier estimate $\tilde{\theta}^*_{uj}$ is $d_u = 7.9$, for censored observations and $\tilde{\theta}^*_{cj}$ it is $(\pi j)^2 d_c$ which, as an example, yields values 3.9, 15.8, and 35.5 for $j =$1, 2, and 3, respectively. This is why censored observations are called ill-posed, and at the same time the dramatically smaller d_c gives a chance to small samples.

Now let us look at the second experiment in Fig. 2. The right column of diagrams in Fig. 2 exhibits a simulation similar to the top one in Fig. 1 only here the distribution of censoring C is Uniform(0, 0.9). Note that no consistent estimation of f^T over its support is possible in this case, but still the estimators can be used because all their denominators are bounded from zero. In the top-right diagram we observe the extremely challenging data for right tail estimation, and the Kaplan−Meier estimator (the step-wise solid line) sheds an extra light on the complexity. Nonetheless, the censored-data estimate (the dot-dashed line) nicely exhibits the underlying density over its support, and it is dramatically better than the uncensored-data estimate (the dotted line). Further, the aggregated estimate is again better than the

Table 1. Numerical analysis of the three experiments in Fig. 1 (experiments 1−3) and two experiments in Fig. 2 (experiments 4, 5). Each entry in columns 2-5 is written as a ratio where: The numerator is the integrated squared error (ISE) of the estimate shown in Figs. 1, 2. The denominator is the median, over 5000 repeated simulations, of ratios between ISEs of the hidden-data estimate and the estimate indicated in the column. The last column shows theoretical rates of censoring

Experiment	Hidden-data estimate	Uncensored-data estimate	Censored-data estimate	Aggregated estimate	$\mathbb{E}\{1 - \Delta\}$
1	0.0002/1	0.003/0.80	0.026/0.03	0.0024/0.81	0.25
2	0.0038/1	0.012/0.25	0.36/0.05	0.0027/0.34	0.38
3	0.051/1	0.092/0.52	0.33/0.13	0.091/0.54	0.50
4	0.0045/1	0.141/0.08	0.0064/0.10	0.074/0.13	0.49
5	0.0095/1	0.229/0.06	0.010/0.07	0.17/0.08	0.42

uncensored-data estimate. Of course, here for consistent estimation the interval of estimation should be decreased, see [18, 21, 32, 34].

To finish discussion of Figs. 1 and 2, let us present integrated squared errors of the estimates for the five simulations, and then compare them with results of an intensive numerical study based on 5000 repeated simulations for each experiment. Results for empirical integrated squared errors (ISE) are shown in Table 1 whose caption explains the entries. We begin with experiment in the top row of Fig. 1, that is with Experiment 1. For the particular data shown in Fig. 1, the hidden-data estimate clearly dominates the others, and the uncensored-data estimate is dramatically better than the censored-data estimate. Nonetheless, the aggregated estimate is much better (in terms of its ISE) than the uncensored-data estimate and also better than the censored-data estimate. Overall, as the repeated simulations show (see the denominators), the particular experiment in Fig. 1 shows us a better side of the aggregation, because in a long run the aggregated estimator is only a bit better than the uncensored-data estimator. Further, on average the censored-data estimate is dramatically worse than the one shown in the right-top diagram in Fig. 1. This is the essence of ill-posedness. We will return to discussion of the outcome shortly, and now let us look at Experiment 2 which is similar to the previous only now the rate of censoring jumps from 25% to 38%. Due to the smaller number of uncensored observations, the uncensored-data estimator performs dramatically worse than in the Experiment 1 but still dominates the censored-data estimate. The good news is that the aggregated estimator performs significantly better than the uncensored-data estimator due to the high rate of censoring. For the experiment 3, as it could be expected, the censored-data estimator performs worse than the uncensored-data one, but again the aggregation improves the censored-data and uncensored-data estimators. Finally, for the challenging, due to sparse right-tail observations, experiments 4 and 5 we see the clear dominance of the censored-data estimator over the uncensored-data estimator due to the better right tails. Further, similarly to the other experiments, the aggregated estimator dominates the two others.

We can make the following conclusions for density estimation: (i) under a mild assumption, asymptotically uncensored observations dominate censored ones due to ill-posedness of censored observations; (ii) for small samples and high rates of censoring it may be beneficial to aggregate these observations; (iii) aggregation may not benefit estimation of high-frequency densities but it definitely does not hurt the estimation; and (iv) the censored-data estimate may help in analysis of censored data with sparse right-tail observations.

2.4. Lung Cancer Data

Let us complement the above-presented numerical study by analysis of the Arm A small cell lung cancer (SCLC) clinical study data presented in the *JASA* [49]. The data contains right-censored survival lifetimes, in days, and age, in years. The censoring is caused by administrative end of the study, and according to the paper it is independent of the survival lifetime and the age. Here we are interested in the density of survival lifetimes, and in the next section consider the regression problem. Let us note that
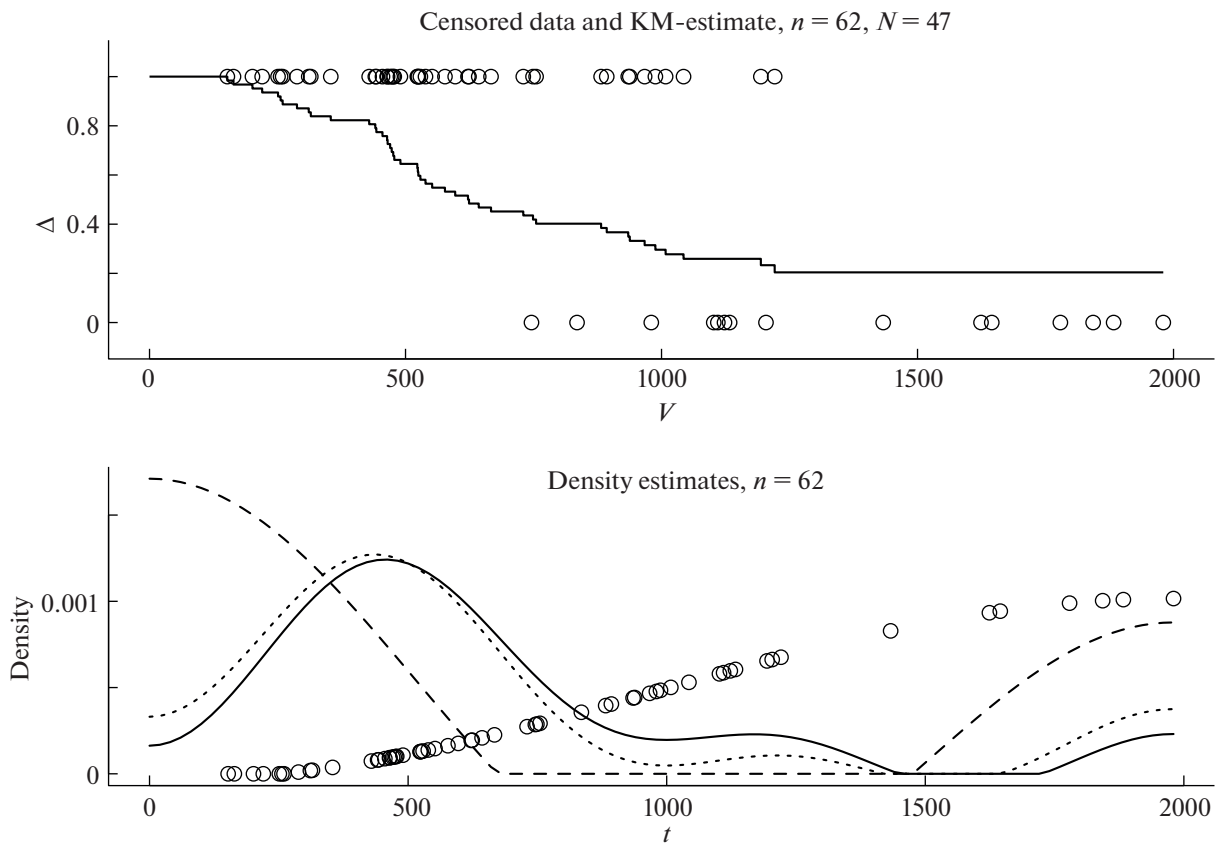
Fig. 3. Data and density estimates for lung cancer data. In the bottom diagram the solid, dashed, and dotted lines are the uncensored-data, censored-data, and aggregated estimates of f^T, respectively, and the circles show estimates of $f^C(V_l)$.

about 15% of all lung cancer cases are the SCLC and this is the most aggressive type of lung cancer with extremely short survival times after the cancer diagnosis. Figure 3 shows the right-censored lifetimes and estimates. The data resembles the sparse right-tail simulations of Fig. 2, only now we do not know the underlying density.

The top diagram exhibits the data. We are dealing with a small sample, rate of censoring is 24%, and right-tail observations are sparse. Note that the seven largest observations of V are censoring times, and they point upon a subset of relatively large underlying lifetimes T. The largest $V_l \Delta_l = 1221$ days, and note that the Kaplan–Meier (KM) estimator of $1 - S^T$ provides no information about the distribution of T beyond that time. In the bottom diagram the circles show the estimates of $f^C(V_l), l = 1, \dots, n$, and the lines are estimates of $f^T(t)$ for $t \in [0, V_{(n)}]$ where $V_{(n)}$ denotes the largest order statistic. The solid line is the uncensored-data estimate. It indicates two strata of underlying lifetimes. The stratum of smaller lifetimes has the mode near 500, and the stratum of larger lifetimes begins with values exceeding 1700. The dashed line shows the censored-data estimate, and it even more articulately points upon two strata of lifetimes. At the same time, keeping in mind its ill-posedness and the extremely small size of the censored-data, its overall shape should be considered with the grain of salt. The aggregated estimate is shown by the dotted line which makes the right strata a bit more pronounced with respect to the uncensored-data estimate. A plausible explanation of the two strata can be found in the publication [37] in the *Journal of Clinical Oncology* devoted to the lung cancer study. According to the publication, the survival of participants was primarily defined by the binary stage of cancer, limited or extensive. This is what the density estimates tell us about. Interestingly, the conclusion about two strata is different from the unimodal on $[0, \infty)$ Bayesian density estimate in [42]. We will continue discussion of the lung cancer data in Subsection 3.4.2 where we look at regression of the survival lifetime on the patient's age.

3. NONPARAMETRIC REGRESSION WITH CENSORED RESPONSES

The structure of this section is identical to the previous one, that is we begin with the literature review which is followed by asymptotic theory, methodology, numerical study, and analysis of real data.

3.1. Literature Review

It is well documented in the literature that the Kaplan–Meier's understanding of the dominance of uncensored observations was also pathbreaking in regression estimation. The dominance is at the core of the seminal Cox's papers [11, 12] where the methodology of partial likelihood was proposed. Later, using information calculations, it was established in [22, 41] that the Cox's estimator is nearly fully efficient. Using the dominance principle, Buckley and James [8] suggested to replace censored responses by their conditional expectations calculated using uncensored observation. This novel imputation approach and the estimator got their name, and the estimator was rigorously studied in [29, 44], and also see an interesting discussion of the imputation in [45]. Discussion of several other related ideas, all of whom are based on the dominance principle, can be found in [29, 33, 39, 48]. There is also a large literature specifically devoted to nonparametric regression with censored responses. Fan and Gijbels [23] use uncensored observations to construct nonparametric imputation of censored responses, and then apply a nonparametric estimator to the transformed responses. This is an interesting and technically challenging nonparametric development of the Buckley–James imputation method. Interesting and sophisticated regression estimators, based on the dominance of uncensored observations, Kaplan–Meier methodology and Buckley–James imputation, can be found in [5, 16, 20, 25, 26, 31, 36, 48].

Surprisingly, it will be shown in the next Subsection 3.2 that, in contrary to the above-discussed problem of density estimation, censored observations are no longer ill-posed for nonparametric regression estimation. Namely, censored observations allow the oracle to estimate nonparametric regression with the same rate as uncensored observations. Accordingly, there is no superiority of uncensored observations over censored ones, and the oracle recommends to use an aggregated, in frequency domain, regression estimator. Then in Subsections 3.3 and 3.4 we will have a chance to evaluate this recommendation using a numerical study when we know the underlying regression as well as analysis of real-life examples.

3.2. Asymptotic Theory and Methodology

There is an underlying pair of interest (X, T) where X is the predictor and lifetime T is the response, and the problem is to estimate nonparametric regression $m(x) = \mathbb{E}\{T | X = x\}$. The response T is not observed directly. Instead, we observe a sample of size n of independent and identically distributed observations from the triplet (X, V, Δ) where $V := \min(T, C)$, $\Delta := I(T \leq C)$ and C is the censoring variable. The censoring partitions data into uncensored and censored observations when we observe realizations of $(X, T, \Delta = 1)$ and $(X, C, \Delta = 0)$, respectively.

As we know from the literature review presented in Subsection 3.1, the principle of dominance of uncensored observations is believed to be valid for nonparametric regression. Let us explore this issue using the oracle approach. We begin with several assumptions that resemble assumptions of Section 2. More general settings are considered in Section 5. In what follows we use sequences q_n and s_n introduced in Section 2.

Assumption 2. *The conditional density $f^{T|X}(t|x)$ is supported on $[0, t^*) \times [0, 1]$ where t^* is either a finite number or infinity. Censoring variable C is a continuous lifetime, its density f^C is positive and continuous on $[0, t^*)$, and C is independent of (X, T). Predictor X is a continuous variable with density f^X which is continuous, positive, and supported on $[0, 1]$.*

Our next assumption is about the oracle.

Assumption 3. *The oracle knows an anchor conditional survival function $S_0(t|x)$, $(t, x) \in [0, t^*) \times [0, 1]$. Anchor $S_0(t|x)$ is continuous in (t, x), differentiable in t, and for any positive constant $a < t^*$*

$$\min_{x \in [0,1]} \min_{t \in [0,a]} S_0(t|x) \geq u_1(a) > 0, \quad \max_{x \in [0,1]} \max_{t \in [0,a]} \partial S_0(t|x)/\partial t \leq -u_2(a) < 0. \tag{26}$$

Set $m_0(x) := \int_0^\infty S_0(t|x)dt$. *The oracle knows that an underlying $S^{T|X}$ belongs to the following local shrinking Sobolev class*

$$\mathcal{F}(\alpha, Q, m_0, n) := \left\{ S^{T|X}(t|x) : \int_0^\infty S^{T|X}(t|x)dt \in \mathcal{M}(\alpha, Q, m_0, n) \right\}, \qquad (27)$$

where

$$\mathcal{M}(\alpha, Q, m_0, n)$$
$$:= \{ m(x) : m(x) = m_0(x) + g(x), \ g(x) \in \mathcal{M}(\alpha, Q), \ |g(x)| \leq 1/s_n, \ x \in [0,1] \}, \qquad (28)$$

$$\mathcal{M}(\alpha, Q) := \left\{ g(x) : g(x) = \sum_{j=0}^\infty \theta_j \varphi_j(x), \ \sum_{j=0}^\infty [1 + (\pi j)^{2\alpha}]\theta_j^2 \leq Q < \infty, \ x \in [0,1] \right\}. \qquad (29)$$

Let us explain these assumptions. The conditional survival function $S_0(t|x)$ anchors all possible underlying survival functions whose regression functions satisfy the additive perturbation (28). Because a conditional survival function must be bona fide (nonnegative and nonincreasing in t), restriction (26) on the anchor is introduced. Also note that the second inequality in (26) implies that the anchor conditional density $f_0^{T|X}(t|x)$ is positive on $[0,1] \times [0,a]$. Let us note that a may depend on n. Line (29) defines a global Sobolev class of α-fold differentiable functions traditionally studied in the classical nonparametric regression theory devoted to the model $T = m(X) + \sigma(x)\xi$ where ξ is a standard normal variable (error) independent of X, see [27]. Then it is known that, based on a direct sample of size n from (X, T), the regression function can be estimated with the classical rate $n^{-2\alpha/(2\alpha+1)}$ of the MISE convergence.

Let us present a lower bound for MISE of the oracle who uses only censored observations.

Theorem 3. *Consider a nonparametric regression problem of estimating $m(x) = \mathbb{E}\{T|X = x\}$ by the oracle who knows the nuisance functions f^X, f^C and the function class $\mathcal{F}(\alpha, Q, m_0, n)$. The oracle uses only censored observations from a sample of size n from (X, V, Δ). Suppose that Assumptions $2, 3$ hold and*

$$D_c := \int_0^1 \frac{\int_0^\infty \frac{S^{T|X}(t|x)}{f^C(t)}dt}{f^X(x)}dx < \infty. \qquad (30)$$

Then

$$\inf_{\tilde{m}^*} \sup_{S^{T|X} \in \mathcal{F}(\alpha, Q, m_0, n)} [n/D_c]^{2\alpha/(2\alpha+1)} \mathbb{E}_{S^{T|X}} \left\{ \int_0^1 (\tilde{m}^*(x) - m(x))^2 dx \right\} \geq P(1 + o_n(1)). \qquad (31)$$

Here the infimum is taken over all possible oracle-estimators and P is equal to the right side of (6). Further, if the anchor $m_0 \in \mathcal{M}(\alpha + \beta, Q')$, $\beta > 0$, $Q' < \infty$, then the lower bound is attainable by an oracle-estimator that does not use the anchor.

Several comments are due. First, $n^{-2\alpha/(2\alpha+1)}$ is the optimal rate of regression estimation for the case of a directly observed sample from (X, T), and Theorem 3 asserts that using censored observations yields the same rate. Accordingly, for the regression problem censored observations are no longer ill-posed and the idea of aggregation is fertile. Second, for direct observations and a classical regression $Y = m(x) + \sigma(x)\xi$ with standard Normal ξ we would see in the lower bound (31) the functional $D := \int_0^1 \sigma^2(x)[f^X(x)]^{-1}dx$ in place of D_c. This allows us to conclude that using only censored observations is similar to the classical regression with normal regression errors and the scale function

$$\sigma(x) = \left[\int_0^\infty \frac{S^{T|X}(t|x)}{f^C(t)}dt \right]^{1/2}. \qquad (32)$$

This is an interesting outcome of the theory which sheds a new light on regression with right-censored responses. Third, the integral (30) can be finite even if T is supported on $[0, \infty)$, in this case the conditional survival $S^{T|X}(t|x)$ should decrease in t a "bit" faster than $f^C(t)$, for instance if for large t we have $S^{T|X}(t|x)/f^C(t) \leq Bt^{-1-\nu}$, $\nu > 0$. The corresponding example will be considered shortly in Subsection 3.3.

Let us present a series oracle-estimator (compare with the estimators in Section 2) which is based on censored observations and attains the lower bound (31). The proof of this assertion can be found in the next section. Set

$$\hat{m}^*(x) := \sum_{j=0}^{J_n} \hat{\theta}_{cj}^* I((\hat{\theta}_{cj}^*)^2 > c_{TH} D_c n^{-1}) \varphi_j(x) + \sum_{j=J_n+1}^{J'_{cn}} (1 - (j/J'_{cn})^\alpha) \hat{\theta}_{cj}^* \varphi_j(x), \qquad (33)$$

where J_n and c_{TH} are the same as in Section 2, J'_{cn} is equal to the right side of (11) with d_u being replaced by D_c, and

$$\hat{\theta}_{cj}^* := n^{-1} \sum_{l=1}^n \frac{(1 - \Delta_l)\varphi_j(X_l)}{f^X(X_l)f^C(V_l)}. \qquad (34)$$

It will be also shown in the next section that $\hat{\theta}_{cj}$ is unbiased estimate of Fourier coefficient $\theta_j := \int_0^1 m(x)\varphi_j(x)dx$ and

$$\mathbb{E}\{(\hat{\theta}_{cj}^* - \theta_j)^2\} = n^{-1}D_c(1 + o_j(1)). \qquad (35)$$

Then, following Section 2, we can use the data-driven E-estimator for small samples and the blockwise-shrinkage estimator for sharp-minimax estimation. Further, in place of unknown densities f^X and f^C we can use the E-estimates presented in Section 2.

Nonparametric regression based on uncensored observations is discussed in [20]. The aggregation of Fourier coefficient estimates in frequency domain is the same as in Section 2 and it is based on the variances of aggregated Fourier coefficient estimates.

Let us check performance of the proposed regression E-estimators for small samples.

3.3. Numerical Study

Figure 4 presents data and regression estimates for a simulated example. Let us describe the experiment and the diagrams. The response T has exponential distribution with mean $m(x) = \mathbb{E}\{T|X = x\}$ being the Bimodal density (recall Fig. 1) plus 0.3. The censoring variable is exponential with mean 2, and the predictor X is uniform on $[0, 1]$. Let us explain the diagrams. The top diagram shows censored observations by the crosses and uncensored by the circles, N is the number of uncensored observations. The middle diagram shows the underlying (hidden) scattergram from (X, T), and the solid and short-dashed lines are the underlying regression and the estimate of the R-package [20]. In the bottom diagram we see the same solid and short-dashed lines as in the middle diagram. Further, the dotted line is the uncensored-data estimate, the dot-dashed line is the censored-data estimate, the long-dashed line is the aggregated estimate; all these estimates are data-driven and based on the scattergram shown in the top diagram.

Now we are ready to analyze the data and the estimates. In the top diagram we see a sample of size n from $(X, V, \Delta) := (X, \min(T, C), I(T \leq C))$, and the problem is to estimate the nonparametric regression $m(x) := \mathbb{E}\{T|X = x\}$. Note that about a third of observations are censored. It is of interest to compare the scattergram of right-censored data with the underlying (hidden) scattergram from (X, T) shown in the middle diagram. The underlying scattergram exhibits a complicated heteroscedastic regression. The regression function, shown by the solid line, has the shape of Bimodal density studied in Fig. 1. Accordingly, we know that the regression is a high-frequency function, and recall that censored observations could not help to estimate the Bimodal density. In the bottom diagram of Fig. 4 the solid and short-dashed lines are the same as in the middle diagram, that is we see the underlying regression and its estimate based on the hidden data. The hidden-data estimate serves as a benchmark. The
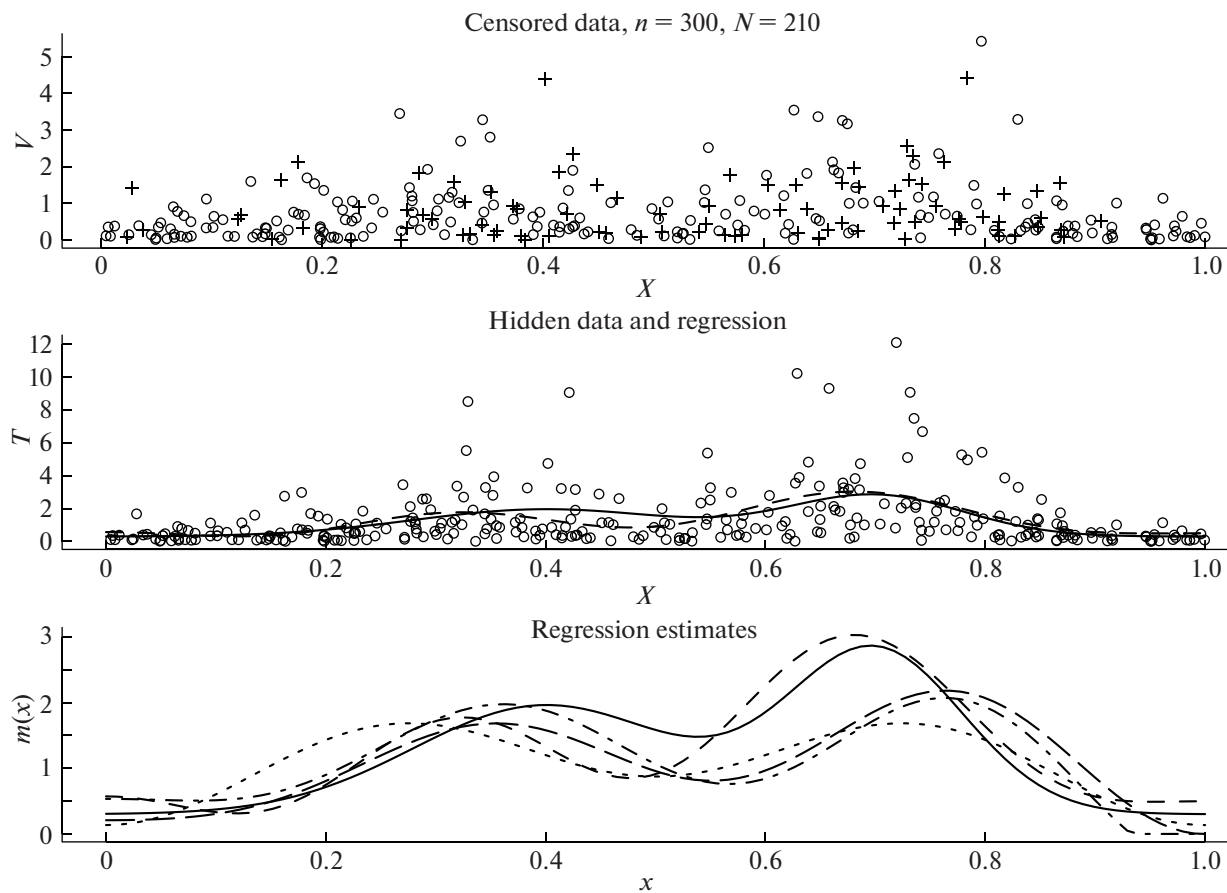
Fig. 4. Simulated nonparametric regression with predictor X and randomly right-censored response T.

estimate is relatively good and indicates a "fair" sample from $(X,\ T)$. The dotted line is the uncensored-data estimate, the dot-dashed line is the censored-data estimate, and the long-dashed line shows the aggregated estimate. These three estimates are data-driven, and let us look at them more closely. They correctly show the bimodal shape but the magnitude of the right mode is small. This could be predicted from analysis of the two scattergrams because all realizations of T larger than 6 are censored. As about another mode, it is shifted to the left and the reason for this is clear from the scattergram. Surprisingly, the estimate based on censored observations, despite their relatively small number, is better than the estimate based on uncensored observations. Also, note that while for density estimation the high-frequency nature of the Bimodal prevented its fair estimation based on censored observations, there is no such issue for the regression. Aggregated estimate (the long-dashed line) is the best and it is dramatically better than the uncensored-data estimate. The visual analysis is supported by the empirical ISEs. Namely, the integrated squared errors of the hidden-data, uncensored-data, censored-data and aggregated estimates are 0.12, 0.36, 0.30, and 0.29, respectively.

Now let us present results of a numerical study based on 5000 repeated simulations of Fig. 4. The mean rate of censoring is 36%, that is in Fig. 3 we see a bit less than the mean number of censored observations. The average ISEs are 0.14, 0.36, 0.27, and 0.25 for the hidden-data, uncensored-data, censored-data, and aggregated estimates, respectively. We may conclude that the shown in Fig. 4 estimates are typical in terms of their ISEs, and that the idea of aggregating uncensored and censored observations in nonparametric regression is feasible even for high-frequency regression functions.

3.4. Analysis of Real Data

We are considering in turn two practical examples with high rates of censoring. The first example is the environmental study by BIFAR. It is of a special interest to us because, due to the small number of

observations ($n = 34$, $N = 16$), the BIFAR provided an extra sample of censoring variable. Accordingly, we can use the oracle's methodology of estimation presented in Subsection 3.2. The second example is continuation of the lung cancer example of Subsection 2.4.

3.4.1. Environmental example. Wastewater treatment facilities are designed to speed up the natural process of purifying water. With billions of people and even more wastewater, the natural process is overloaded. Without wastewater treatment, the amount of wastewater would cause environmental devastation by discharging into the environment. Moreover, wastewater treatment plays critical role in climate change mitigation by reducing greenhouse gas emission, see [15].

Wastewater centrifuge is a part of industrial wastewater treatment plant, see [15, 24, 35]. Centrifugal thickening and dewatering of sludge is a high speed process that uses the force from rapid rotation of a cylindrical bowl to separate wastewater solids from liquid. The sludge accumulates on the bowl periphery, and the internal conveyer scrapes towards the sludge discharge ports to produce a non-liquid material referred to as the cake. Because of the abrasive nature of many sludges, especially some mining, industrial and sewage sludges, hard-facing materials are applied to the leading edges of the conveyer blades. The wearing surfaces are replaceable but this can be done only by the manufacturer due to necessity to balance the conveyer. Accordingly, it is important to know lifetime of the conveyer blade.

Another important issue, related to the lifetime of the conveyer blade, is the level of grit in the treated waste. Grit is the heavy inorganic solids that could cause excessive mechanical wear. Grit is heavier than organic solids and includes sand, gravel, clay, metal filings, seeds, and other similar materials. Several processes are used for grit removal. All of the processes are based on the fact that grit is heavier than organic solids, which should be kept in suspension for treatment in subsequent processes. Grit removal is done by grit separators that are relatively expensive and may slow waste water flow.

The environmental company BIFAR conducted a controlled study devoted to exploring how lifetime T of a conveyer blade depends on concentration of grit. There are two facts about the study that should be mentioned. First, the lifetime of modern blades is relatively long and may last several years. This explains why only $n = 34$ observations are available. Second, while industrial centrifuge is a device with many parts that may break down, only lifetime C of bearings right censors T. In the BIFAR experiment only 16 observations of T are uncensored, that is $N = \sum_{l=1}^{34} \Delta_l = 16$. These n and N are very small for nonparametric estimation. To help with the estimation, BIFAR provided data about directly observed $n_E = 54$ lifetimes of bearings. These observations were used to construct E-estimate of f^C. The final remark is about used predictor. It is difficult to control the level of grit in the waste supplied to the centrifuge, but plain to define cost of a preliminary grit separation. Accordingly, BIFAR provided cost X of grit separation and recommended it as the predictor.

The available data and corresponding estimates are shown in Fig. 5 and its caption explains the diagrams. Uncensored observations are shown by circles in the left-top diagram, censored observations are shown by crosses in the right-top diagram, and these observations together are shown in the left-bottom diagram. The above-explained extra observations of C are shown by triangles in the right-bottom diagram, and the estimated density f^C is shown by the solid line. Note that the support of C in the extra sample is clearly larger than $[0, V_{(n)}]$, and hence $S^C(V_{(n)}) > 0$. This yields validity of using the developed estimators.

The uncensored-data regression estimate (the solid line in the left-top diagram) looks reasonable for the shown scattegram, but please keep in mind that the observations are biased because $f^{V|\Delta}(t|1) = \frac{f^T(t)S^C(t)}{\mathbb{P}(\Delta=1)}$. Accordingly, visualization should be used with vigilance. The right top diagram shows us 18 censored observations, that is the lifetimes of bearings. Note that the censored-data regression estimate (the solid line) is not "supported" by the data visualization, and we already know that for censored data this is not a defining factor in judging the estimate.

Figure 5d shows us the total right-censored BIFAR data and the aggregated regression estimate (the solid line). Let us stress that here all available observations are used to construct the estimate.

Table 2 sheds extra light on the BIFAR data and the three regression estimates. Estimates of Fourier coefficient $\theta_0 = \int_0^1 m(x)dx$ are presented in the second column. The estimates are very close, and note how aggregation of the uncensored and censored observations decreases the standard deviation. The latter is not a surprise due to the high rate of censoring. Estimates of Fourier coefficient $\theta_1 = \int_0^1 m(x)2^{1/2}\cos(\pi jx)dx$ are presented in the third column. This parameter defines "slope" of the
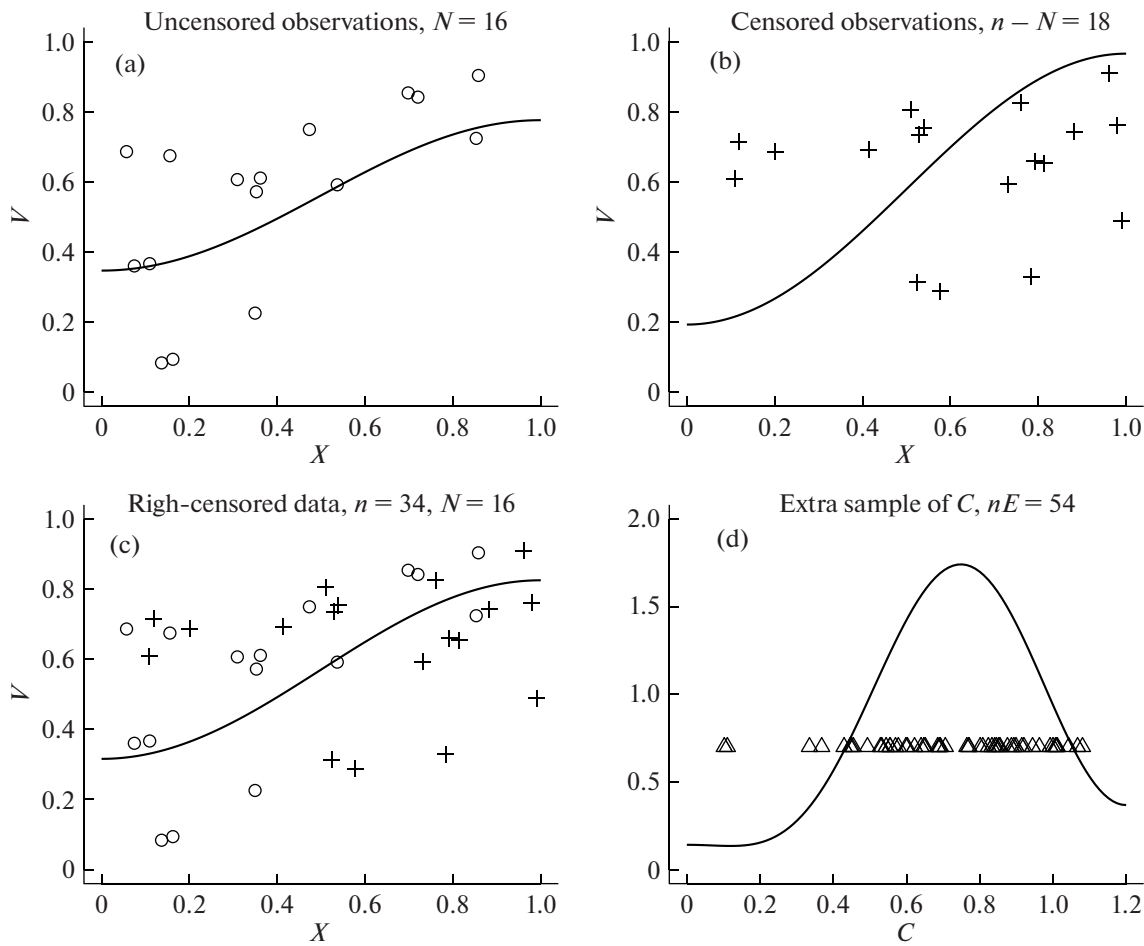
Fig. 5. BIFAR data. (a) Shows uncensored observations of T overlaid by the uncensored-data regression estimate. (b) Shows censored observations, that is observations of C, overlaid by the censored-data regression estimate. (c) Shows the total right-censored BIFAR data overlaid by the aggregated regression estimate. (d) Shows the extra sample of the censoring variable and the density estimate. All observations are linearly transformed by the BIFAR.

regression. Note the difference in conclusions of the uncensored-data and censored-data regression estimators about θ_1, and we can also see this in the slopes of the corresponding regressions shown in Fig. 5. The aggregated Fourier estimate is more close to the uncensored-data estimate of θ_1 than to the censored-data one, and this is because the standard deviation of the uncensored-data Fourier estimate is almost twice smaller. All other Fourier estimates are insignificant and hard-thresholded by the regression E-estimators. Accordingly, all three estimates are of the form $\hat{m}(x) = \hat{\theta}_0 + \hat{\theta}_1 \varphi_1(x)$.

The BIFAR example is of a special interest because it shows how the extra sample from C may help

Table 2. Estimated Fourier coefficients and corresponding standard deviations for the three regression estimates shown in Fig. 5. Each entry is written as A/B where A is the estimate and B is its standard deviation

Estimand	θ_0	θ_1	θ_2	θ_3	θ_4
Uncensored-data	0.56/0.16	−0.15/0.07	−0.01/0.05	0.04/0.05	−0.05/0.05
Censored-data	0.58/0.16	−0.27/0.13	−0.17/0.21	0.16/0.14	0.18/0.21
Aggregated	0.57/0.11	−0.18/0.06	−0.02/0.05	0.06/0.05	0.05/0.05

Censored data, $n = 62$, $N = 47$
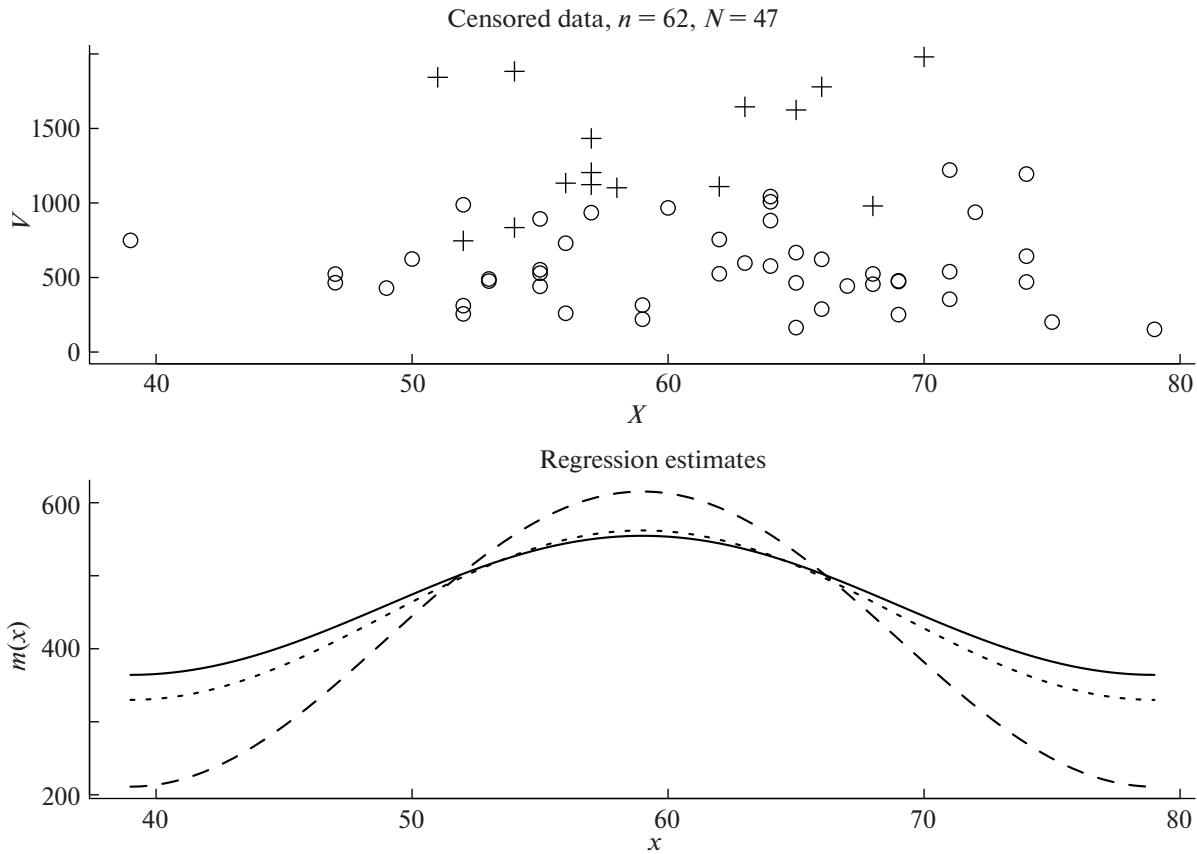

Regression estimates

Fig. 6. Regression estimates for the lung cancer clinical study. The circles and the crosses show the uncensored and censored lifetimes. The solid, dashed and dotted lines are the uncensored-data, censored-data and aggregated estimates.

to deal with extremely small, for nonparametric estimation, samples of right censored data. This is also the real example which shows feasibility of the oracle's methodology.

3.4.2. Lung cancer example. Here we look at the regression for the lung cancer study for which we already estimated densities of survival and censoring lifetimes in Section 2. Recall that in the *JASA* paper [49] the regression data is provided with predictor X being the age of a participant, and it is explained that the censoring lifetime C does not depend on the predictor. The data and the estimates are shown in Fig. 6. Let us look at them.

The top diagram in Fig. 6 shows the scattergram of the right-censored data. Several interesting observations can be made about the data. First, note that the lifetimes are relatively small for the five youngest and four oldest participants. Second, the largest lifetimes are the censored lifetimes (the crosses) for the middle age participants. Accordingly, we can expect that the regression should have a pronounced maximum for the middle age and decreasing tails. And indeed, these observations are reflected by the all regression estimates. Further, as it could be expected from the scattergram, the censored-data estimate (the dashed line) is the most pronounced, but due to the smaller number of censored lifetimes (the crosses) its effect on the aggregated estimate is minor. According to [37], the small cell lung cancer is extremely aggressive and smoking is one of the main factors. This may explain the data and the underlying message of the regression estimates. In the *JASA* paper [49] a log-linear regression was studied, and correspondingly this interesting effect of the age could not be revealed.

We may conclude that the proposed methodology of regression for right-censored responses is feasible and can be recommended for analysis of real data.

4. PROOFS

Proof of Theorem 1. Lower bound (4) and its sharpness for uncensored observations (a subsample with $\Delta = 1$) is established in [18]. Let us prove lower bound (5) for censored observations which establishes ill-posedness of the data. The proof is involved and it is worthwhile begin with the heuristic. The main step is to replace the nonparametric minimax by a parametric minimax for increasing number of parameters, and then bound from below the parametric minimax by a corresponding Bayes risk. The choice of parameters should be such that the corresponding classical parametric Fisher informations are constants as functions in t. As we will see shortly, Fisher informations are functionals of $f^C(t)/S^T(t)$. This triggers the necessity to divide the studied interval $[0, 1]$ into a sequence of subintervals with decreasing length as $n \to \infty$. This step allows us to deal with almost constant Fisher information for each subinterval. Correspondingly, for each subinterval its own Sobolev class of parametric densities is proposed. Then the main issue is how to spread the power Q of the global Sobolev class over the subintervals, and this is done inversely proportional to the local Fisher informations. Several other comments are also due. The density must be from the class $\mathcal{F}_n$, and to achieve that we sew local functions at boundaries of the subintervals using so-called flattop kernels, the latter is a standard technique in harmonic analysis. Another issue is the right boundary where $S^T(t)$ may be too close to zero. To deal with this issue we bound from below the MISE considered over interval $[0, 1]$ by the MISE considered over a subinterval $[0, a]$ with some fixed $a \in (0, 1)$. To implement this idea, we divide the unit interval into s_n subintervals and then consider only subintervals within $[0, a]$. Finally, to highlight steps that shed light on ill-posedness and to make the proof shorter, whenever possible we are using technical results of [18] obtained in the proof of lower bound (4).

Now we begin the outlined steps of the proof of lower bound (5). Set $s := s_n := 3 + \lceil \ln(\ln(n + 3)) \rceil$, where $\lceil x \rceil$ is the smallest integer larger or equal to x. Recall that f_0 is the anchor density of the considered local Sobolev class of underlying densities, $S_0(t) := \int_t^1 f_0(u)du$ is the anchor survival function, B denotes a generic positive constant, $a \in (0, 1)$ is a constant. Let $\phi(x) = \phi(n, x)$ denote a sequence of flattop kernels such that for a given n: the kernel is zero beyond $(0, 1)$, α-fold continuously differentiable on $(-\infty, \infty)$, $0 \le \phi(x) \le 1$, $\phi(x) = 1$ for $2(\ln(n))^{-2} \le x \le 1 - 2(\ln(n))^{-2}$, and $|\phi^{(m)}| \le B(\ln(n))^{2m}$, see examples of the kernel in [19]. We divide the unit interval $[0, 1]$ into s equal subintervals and numerate them using index $k = 0, 1, \ldots, s - 1$. On each subinterval we introduce the sine basis $\psi_{skj}(t) := s^{1/2}\psi_j(st - k)$, $j = 1, 2, \ldots$, $\psi_j(t) := 2^{1/2}\sin(\pi j t)$ and the corresponding flattop kernel $\phi_{sk}(t) := \phi(st - k)$. Also set $Q_{sk} = (Q - 1/s)(\overline{I_s^{-1}}I_{sk})^{-1}$, $I_{sk} = f^C(k/s)/S_0(k/s)$, $\overline{I_s^{-1}} = \sum_{k=0}^{\lfloor sa \rfloor}(1/I_{sk})$ where $\lfloor x \rfloor$ is the largest integer not exceeding x. Note that, as was explained in the heuristic, only subintervals of $[0, a]$ are considered. Also set

$$J_{sk} := \lceil [(\alpha + 3)(2\alpha + 3)\alpha^{-1}(s\pi)^{-2\alpha-2}Q_{sk}I_{sk}n]^{1/(2\alpha+3)} \rceil.$$

We begin with replacing the studied local Sobolev class by a sequence in n of parametric classes of densities that are subclasses of the local Sobolev class. Set $S_0(t) := 1 - \int_0^t f_0(v)dv$,

$$\mathcal{H}_s = \left\{ f : S(t) = \int_t^1 f(v)dv, \ S(t) = S_0(t) + \sum_{k=0}^{\lfloor as \rfloor} g_{sk}(t)\phi_{sk}(t), \right.$$

$$g_{sk}(t) = \sum_{j=\lfloor J_{sk}/\ln(n) \rfloor}^{J_{sk}} (\pi j s)^{-1}\nu_{skj}\psi_{skj}(t),$$

$$|dg_{sk}(t)/dt|^2 \le s^3 \ln(n) J_{sk} n^{-1},$$

$$\left. \sum_{j=\lfloor J_{sk}/\ln(n) \rfloor}^{J_{sk}} (\pi j)^{2\alpha}\nu_{skj}^2 \le s^{-2\alpha}Q_{sk}, \ f \ge 0, \ t \in [0, 1] \right\}.$$

Let us comment on the class $\mathcal{H}_s$. The flattop kernels are used to smoothly "sew" the additive permutations g_{sk} at the boundaries, and also note that the permutations are zero (vanish) at the

boundary points. If we "ignore" the flattop kernels and differentiate a survival function from $\mathcal{H}_s$, then we get an additive permutation studied in [18] and matching the underlying local Sobolev class of densities. The reason why we are dealing with the specific class of survival functions is because the likelihood of censored observations is defined by the survival function and not the density, recall (1), and accordingly it is convenient to define the class $\mathcal{H}_s$ via survival functions. Then following [18] we get $\mathcal{H}_s \subset \mathcal{F}_n$ for all sufficiently large n, and accordingly in (5) we can replace the supremum over $\mathcal{F}_n$ by supremum over $\mathcal{H}_s$. Our final remark is that $\mathcal{H}_s$ is a class of densities created by additive perturbations of the anchor density f_0 on each of the first $\lfloor as \rfloor + 1$ subintervals of $[0, 1]$. Note that the additive perturbations are independent of each other. Thus we get the inequality,

$$\sup_{f \in \mathcal{F}_n} \mathbb{E}_f \left\{ \int_0^1 (\tilde{f}^*(t) - f(t))^2 dt \right\} \geq \sup_{f \in \mathcal{H}_s} \sum_{k=0}^{\lfloor as \rfloor} \mathbb{E}_f \left\{ \int_{k/s}^{(k+1)/s} (\tilde{f}^*(t) - f(t))^2 dt \right\}. \tag{36}$$

Now we can use the classical approach of bounding from below the studied minimax risk by a Bayesian risk using independent zero mean Normal priors for parameters ν_{skj}. Namely, for ν_{skj} the variance of normal prior is set to

$$\tau_{skj}^2 = (\pi j s)^2 n^{-1} (1 - 3q_*^{-1}) I_{sk}^{-1} \max(q_*^{-1}, \min(q_*, [(J_{sk}/j)^\alpha - 1])),$$

where $q_* > 3$ is a constant that may be as large as desired. Next we need to make several straightforward calculations. We begin with calculating Fisher information for parameter ν_{skj} and censored pair $((1 - \Delta)C, \Delta)$, recall that we are verifying the lower bound (5) for the oracle who uses only censored observations. This is the step that will shed light on ill-posedness of censored observations. The corresponding mixed density is

$$f^{(1-\Delta)C, \Delta}(t, \delta) = [f^C(t)S^T(t)]^{1-\delta} [\mathbb{P}(\Delta = 1)]^\delta$$

$$= [f^C(t)S^T(t)]^{1-\delta} \left[1 - \int_0^1 f^C(v)S^T(v)dv \right]^\delta. \tag{37}$$

Here the first factor corresponds to density of censored pair (V, Δ) with $\Delta = 0$ while the second factor is the corresponding value of the probability mass function of the Bernoulli random variable Δ, as we will see shortly the second factor yields a negligibly small component of the Fisher information. The parametric Fisher information is

$$I_{skj} := \mathbb{E} \left\{ \left[\partial \ln([f^C(C)S(C)]^{1-\Delta} \left[1 - \int_0^1 f^C(u)S(u)du \right]^\Delta) / \partial \nu_{skj} \right]^2 \right\}$$

$$= \mathbb{E} \left\{ (1 - \Delta) \frac{[\partial S(C)/\partial \nu_{skj}]^2}{[S(C)]^2} \right\} + \mathbb{E} \left\{ \Delta \frac{[\int_0^1 f^C(u)(\partial S(u)/\partial \nu_{skj})du]^2}{[\mathbb{P}(\Delta = 1)]^2} \right\}$$

$$= (\pi j s)^{-2} \left[\mathbb{E}_{f_0} \left\{ (1 - \Delta) \frac{[\psi_{skj}(C)\phi_{sk}(C)]^2}{[S(C)]^2} \right\} \right.$$

$$\left. + \mathbb{E} \left\{ \Delta \frac{[\int_0^1 f^C(u)\psi_{skj}(u)\phi_{sk}(u)du]^2}{[\mathbb{P}(\Delta = 1)]^2} \right\} \right]$$

$$= (\pi j s)^{-2} \left[\int_0^1 [f^C(u)/S_0(u)][\psi_{skj}(u)\phi_{sk}(u)]^2 du \right.$$

$$\left. + \int_0^1 f^C(u)\psi_{skj}(u)\phi_{sk}(u)du/\mathbb{P}(\Delta = 1) \right]$$

$$= (\pi j s)^{-2}[f^C(k/s)/S_0(k/s)](1 + o_n(1))$$

$$= (\pi j s)^{-2} I_{sk}(1 + o_n(1)). \tag{38}$$

In the next to last line we used definition of ψ_{skj}, which is element of the sine basis on the kth subinterval, and that ϕ_{sk} is the flattop kernel on that subinterval, as well as the assumption about continuity and smoothness of f^C and S_0. Formula (38) sheds light on the ill-posedness because we see that the Fisher information decreases as the frequency j increases. This is what creates the ill-posedness in frequency domain for censored observations.

Now we make several more calculations. First we get via approximation of a sum by a corresponding integral,

$$R(J, n, d) := n^{-1}d\sum_{j=0}^{J}(\pi j s)^2[1 - (j/J)^\alpha] = n^{-1}J^3 d(\pi s)^2 \frac{\alpha}{3(\alpha + 3)}(1 + o_J(1)). \tag{39}$$

Second, using the same calculation technique, we find that solution with respect to J of the equation $dn^{-1}\sum_{j=1}^{J}(\pi j s)^{2\alpha+2}[(J/j)^\alpha - 1] = Q_*$ is

$$J(n, s, d, Q_*) = \left[\frac{nQ_*(\alpha + 3)(2\alpha + 3)}{d(\pi s)^{2\alpha+2}\alpha}\right]^{1/(2\alpha+3)}(1 + o_n(1)). \tag{40}$$

Using (36), the above-defined Bayesian approach, (38)–(40), and following steps of the proof in [18] we get

$$\sup_{f\in\mathcal{F}_n} \mathbb{E}_f\left\{\int_0^1 (\tilde{f}^*(t) - f(t))^2 dt\right\} \geq \sum_{k=1}^{\lfloor as\rfloor} A_k + o_n(1)n^{-2\alpha/(2\alpha+3)}, \tag{41}$$

where

$$A_k \geq R(J(n, s, I_{sk}^{-1}, Q_{sk}), n, I_{sk}^{-1})(1 + o_n(1))$$

$$= n^{-1}[J(n, s, I_{sk}^{-1}, Q_{sk})]^3 I_{sk}^{-1}(\pi s)^2\alpha[3(\alpha + 3)]^{-1}(1 + o_n(1))$$

$$= P_c n^{-2\alpha/(2\alpha+3)}[s^{2\alpha+2}\overline{I_s^{-1}}]^{-3/(2\alpha+3)}s^2 I_{sk}^{-1}(1 + o_n(1))$$

$$= P_c n^{-2\alpha/(2\alpha+3)}[s^{-1}\overline{I_s^{-1}}]^{-3/(2\alpha+3)}s^{-1} I_{sk}^{-1}(1 + o_n(1)).$$

Now note that $s^{-1}\overline{I_s^{-1}} = s^{-1}\sum_{k=0}^{\lfloor as\rfloor} I_{sk}^{-1} = \int_0^a [S_0(v)/f^C(v)]dv(1 + o_n(1))$. We conclude that

$$\sum_{k=0}^{\lfloor as\rfloor} A_k \geq P_c\left[n/\int_0^a [S_0(v)/f^C(v)]dv\right]^{-2\alpha/(2\alpha+3)}(1 + o_n(1)).$$

Now recall that $d_c = \int_0^1 [S_0(v)/f^C(v)]dv$. Because a may be chosen as close to 1 as desired, this finishes the proof of lower bound (5). Sharpness of the lower bounds will follow from the verified below Theorem 2. Theorem 1 is proved.

Proof of Theorem 2. We begin with the following assertion that evaluates MISE of the low-frequency component of the density estimate. Suppose that $\mathbb{E}\{(\tilde{\kappa}_j - \kappa_j)^2\} \leq B_*n^{-1}$. Then we can write that

$$\mathbb{E}\{(\tilde{\kappa}_j I(\tilde{\kappa}_j^2 > c_{TH}dn^{-1}) - \kappa_j)^2\} \leq 2[\mathbb{E}\{(\tilde{\kappa}_j - \kappa_j)^2\} + \mathbb{E}\{\kappa_j^2 I(\tilde{\kappa}_j^2 \leq c_{TH}dn^{-1})\}]$$

$$\leq 2[B_*n^{-1} + 4\mathbb{E}\{[\tilde{\kappa}_j^2 + (\tilde{\kappa}_j - \kappa_j)^2]I(\tilde{\kappa}_j^2 \leq c_{TH}dn^{-1})\} = o_n(1)s_n n^{-1}.$$

Accordingly, the MISE of the low-frequency component of the estimate is $o_n(1)n^{-2\alpha/(2\alpha+1)}$, and we need only to study MISE of the high-frequency component. Note that this is an interesting result because the low-frequency component inspires the E-estimator for small samples while the high-frequency component yields asymptotic efficiency (sharp constant and optimal rate). Further, note that c_{TH} may depend on n and this does not change the result. For instance, $c_{TH} = 2\ln(n)$ implies the classical hard thresholding [17]. Now let us return to the proof.

The case of using uncensored observations is considered in [18]. The new here is the case of censored observations and the aggregation. We begin with analysis of the Fourier estimate $\tilde{\theta}_{cj}^*$ defined in (12). Recall that $\{\psi_j(x)\}$ and $\{\varphi_j(x)\}$ are the sine and cosine bases on $[0,1]$. Using iid of censored observations and (1) we can write for $j \geq 1$,

$$\mathbb{E}\{\tilde{\theta}_{cj}^*\} = 2^{1/2} - \mathbb{E}\left\{n^{-1}(\pi j)\sum_{l=1}^{n}(1-\Delta_l)\frac{\psi_j(V_l)}{f^C(V_l)}\right\}$$

$$= 2^{1/2} - (\pi j)\mathbb{E}\left\{(1-\Delta)\frac{\psi_j(V)}{f^C(V)}\right\}$$

$$= 2^{1/2} - (\pi j)\int_0^\infty S^T(t)\psi_j(t)dt.$$

We continue using $\varphi_j(0) = 2^{1/2}$, integration by parts, and the assumed support $[0,1]$ of T,

$$\mathbb{E}\{\tilde{\theta}_{cj}^*\} = 2^{1/2} - (\pi j)[-S^T(t)(\pi j)^{-1}\varphi_j(t)|_{t=0}^1 - (\pi j)^{-1}\int_0^1 f^T(t)\varphi_j(t)dt]$$

$$= 2^{1/2} - [2^{1/2} - \int_0^1 f^T(t)\varphi_j(t)dt] = \theta_j.$$

This proves that the estimate is unbiased. Next, again using (1) we write,

$$n(\pi j)^{-2}\mathbb{E}\{(\tilde{\theta}_{cj}^* - \theta_j)^2\} = \mathbb{E}\left\{\frac{(1-\Delta)\psi_j^2(V)}{[f^C(V)]^2}\right\} = \int_0^1 \frac{S^T(t)\psi_j^2(t)}{f^C(t)}dt = d_c(1+o_j(1)).$$

This verifies (16). To verify (17), (18) we note that if random variables Z and Y have the same mean and variances σ_Z^2 and σ_Y^2, then the aggregation $\lambda Z + (1-\lambda)Y$ has the minimal variance when $\lambda = \sigma_Y^2/[\sigma_Z^2 + \sigma_Y^2]$. Now we verify that the density estimate (13) attains the lower bound (5). Let B denotes generic positive constants. Using the Parseval identity and the above-verified properties of $\tilde{\theta}_{cj}$ we get,

$$\mathbb{E}\left\{\int_0^1 (\tilde{f}_c^*(t) - f(t))^2 dt\right\}$$

$$= \sum_{j=1}^{J_n}\mathbb{E}\{\tilde{\theta}_{cj}^* - \theta_j)^2\} + \sum_{j=J_n+1}^{J_{cn}^*}\mathbb{E}\{[(1-(j/J_{cn}^*)^\alpha)\tilde{\theta}_j - \theta_j]^2\} + \sum_{j>J_{cn}^*}\theta_j^2$$

$$= \sum_{j=1}^{J_{cn}^*}\mathbb{E}\{(\tilde{\theta}_{cj}^* - \theta_j)^2\}$$

$$+ \mathbb{E}\left\{\sum_{j=J_n+1}^{J_{cn}^*}[(1-(j/J_{cn}^*)^\alpha)(\tilde{\theta}_{cj}^* - \theta_j) + (j/J_{cn}^*)^\alpha\theta_j]^2\right\} + \sum_{j>J_{cn}^*}\theta_j^2$$

$$\leq Bn^{-1}J_n^2 + n^{-1}\sum_{j=1}^{J_{cn}^*}(\pi j)^2(1-(j/J_{cn}^*)^\alpha)^2 d_c(1+o_j(1)) + \sum_{j>J_n}(j/J_{cn}^*)^{2\alpha}\theta_j^2. \qquad (42)$$

For the considered functional class of densities we have

$$\sum_{j>J_n}(j/J_{cn}^*)^{2\alpha}\theta_j^2 \leq (\pi J_{cn}^*)^{-2\alpha}Q(1+o_n(1))$$

due to the assumed smoothness of the anchor. Using this relation as well as (40) with $s = 1$ and $Q_* = Q$, we get

$$d_c n^{-1} \sum_{j=J_n+1}^{J_{cn}^*} (\pi j)^{2\alpha+2}[(J_{cn}^*/j)^\alpha - 1] = Q(1 + o_n(1)),$$

and then can write,

$$\sum_{j>J_n} (j/J_{cn}^*)^{2\alpha}\theta_j^2$$

$$\leq (\pi J_{cn}^*)^{-2\alpha} d_c n^{-1} \sum_{j=J_n+1}^{J_{cn}^*} (\pi j)^{2\alpha+2}[(J_{cn}^*/j)^\alpha - 1](1 + o_n(1))$$

$$= n^{-1}d_c \sum_{j=J_n+1}^{J_{cn}^*} (\pi j)^2[(j/J_{cn}^*))^\alpha - (j/J_{cn}^*)^{-2\alpha}](1 + o_n(1)).$$

Using this relation in (42), together with (7), (14), and (39) we conclude that

$$\mathbb{E}\left\{\int_0^1 (\tilde{f}_c^*(t) - f(t))^2 dt\right\}$$

$$\leq Bn^{-1}J_n^3 + n^{-1}d_c \sum_{j=1}^{J_{cn}^*}(\pi j)^2[1 - (j/J_{cn}^*)^\alpha]$$

$$\leq n^{-1}d_c[J_{cn}^*]^3\pi^2 \frac{\alpha}{3(\alpha + 3)}(1 + o_n(1))$$

$$= (n/d)^{-2\alpha/(2\alpha+3)}P_c(1 + o_n(1)).$$

Sharp minimax property of $\tilde{f}_c^*$ is established. Theorem 2 is proved.

Proof of Lemma 1. Let us look at the estimate (22) of the survival function $S^V(t) := \mathbb{E}\{I(V \geq t)\}$ of the continuos and always observed random variable V,

$$\hat{S}^V(t) := n^{-1}\sum_{l=1}^n I(V_l \geq t).$$

This is the classical sample mean estimate based on the sample of n Bernoulli random variables $I(V_l \geq t)$, $l = 1, 2, \ldots, n$, and it is assumed that we are considering this estimate for $t \in [0, a]$ and $S^V(a) > 0$. Then the relations (23)–(25) of the verified Lemma 1, where $(\hat{S}^V, S^V)$ are used in place of $(\hat{S}^C, S^C)$, hold according to classical properties of a sum of Bernoulli random variables, see [20], Sect. 1.3]. Next we note that

$$S^C(t) = e^{-\int_0^t [f^C(v)/S^C(v)]dv} = e^{-\int_0^t f^{V,\Delta}(v,0)/S^V(v)]dv} = e^{-\mathbb{E}\{(1-\Delta)I(V\in[0,t])/S^V(V)\}}.$$

Now we can note that (21) is the plug-in sample mean estimate, and the assertion of Lemma 1 follows from the Taylor formula and a straightforward calculation. Lemma 1 is verified.

Proof of Theorem 3. The proof follows along lines of the proof of Theorem 1 and, whenever possible, it uses the same notations. The main difference is in the necessity to use more complicated parametric classes because those used in the proof of Theorem 1 are too "simple" and yield a lower bound smaller than the verified (31). Nonetheless, to simplify the presentation, we first introduce a parametric class that is similar to $\mathcal{H}_s$, used in the proof of Theorem 1, and then it will be explained how to modify it for more "complicated", or we can say less favorable, estimation. Introduce a class of additive perturbations of the anchor regression,

$$\mathcal{M}_s := \left\{m(x): \ m(x) := m_0(x) + \sum_{k=1}^{s-2} g_k(x)I(1/s \leq x \leq 1 - 1/s), \ g_k(x) \in \mathcal{M}_{sk}\right\},$$

where classes $\mathcal{M}_{sk}$ will be defined shortly. In what follows, similarly to the proof of Theorem 1, we are dividing the interval $[0,1]$ into s subintervals, and then use the additive perturbation only at the inner intervals. Then we are following steps of the proof of Theorem 1 and use the same flattop kernels to smoothly sew additive perturbations on the subintervals of $[0,1]$. Namely, for $1 \le k \le s-2$ we introduce $\varphi_{skj}(x) := \sqrt{s}\varphi_j(sx - k)$,

$$g_{[k]}(x) := \sum_{j=J'(k)}^{J(k)} \nu_{skj}\varphi_{skj}(x),$$

$$g_{(k)}(x) := g_{[k]}(x)\phi_{sk}(x),$$

$$J(k) := \lceil [n(2\alpha + 1)(\alpha + 1)s^{-2\alpha}Q_{sk}(\alpha\pi^{2\alpha})^{-1}]^{1/(2\alpha+1)} \rceil,$$

$$J'(k) := \lceil J(k)/\ln(n) \rceil, \quad Q_{sk} := (Q - 1/s)(\overline{I_s^{-1}}I_{sk})^{-1},$$

$$I_{sk}^{-1} := \int_0^{t^*} \frac{S_0(t|k/s)}{f^X(k/s)f^C(t)}dt,$$

and $\overline{I_s^{-1}} := \sum_{k=1}^{s-2}(1/I_{sk})$.

Using these sequences we define classes $\mathcal{M}_{sk}$ used in the definition of $\mathcal{M}_s$,

$$\mathcal{M}_{sk} := \left\{ g : g(x) = g_{(k)}(x)I(k/s \le x \le (k+1)/s), \right.$$

$$\left. \sum_{j=J'(k)}^{J(k)} (\pi j)^{2\alpha}\nu_{skj}^2 \le s^{-2\alpha}Q_{sk}, \ |g_{[k]}(x)|^2 \le s^3\ln(n)J(k)n^{-1} \right\}.$$

Note that a regression function from the class $\mathcal{M}_{sk}$ can be written as

$$m(x) = m_0(x) + \sum_{k=1}^{s-2}\sum_{j=J'(k)}^{J(k)} \nu_{skj}\varphi_{skj}(x)\phi_{sk}(x).$$

Next five steps follow along lines of the proof of Theorem 1. First, direct calculations show that the class $\mathcal{F}$ includes $\mathcal{M}_s$. Second, we introduce

$$\tau_{skj} := [n^{-1}(1 - 3q_*^{-1})I_{sk}^{-1}\max(q_*^{-1}, \min(q_*, (J(k)/j)^\alpha - 1))]^{1/2},$$

where $q_* > 3$ is a constant that may be as large as desired. Note that $\nu_{skj} = \tau_{skj}$ satisfy the definition of classes $\mathcal{M}_{sk}$, and for $k = 1, \ldots, s-2$ we can introduce the following sets of these parameters,

$$\Theta_{sk} := \left\{ \vec{\nu}_{sk} : \sum_{j=J'(k)}^{J(k)} (\pi j)^{2\alpha}\nu_{skj}^2 \le s^{-2\alpha}Q_{sk}, \ |g_{[k]}(x)|^2 \le s^3\ln(n)J(k)n^{-1} \right\}.$$

Here $\vec{\nu}_{sk} := \{\nu_{skJ'(k)}, \ldots, \nu_{skJ(k)}\}$, and note that $\vec{\tau}_{sk} := \{\tau_{skJ'(k)}, \ldots, \tau_{skJ(k)}\} \in \Theta_{sk}$. The third step is to use the Parseval identity, notation $\tilde{\nu}_{skj}^* := \int_{k/s}^{(k+1)/s} \tilde{g}^*(x)\varphi_{skj}(x)dx$, the fact that the oracle knows the anchor m_0, and conclude that

$$\sup_{S^{T|X} \in \mathcal{F}(\alpha,Q,m_0,n)} \mathbb{E}\left\{ \int_0^1 (\tilde{m}^*(x) - m(x))^2 dx \right\}$$

$$\ge (1 - s^{-1}) \sum_{k=1}^{s-2} \sup_{\vec{\nu}_{sk} \in \Theta_{sk}} \sum_{j=J'(k)}^{J(k)} \mathbb{E}\left\{ (\tilde{\nu}_{skj}^* - \nu_{skj})^2 \right\} + o(1)n^{-2\alpha/(2\alpha+1)}. \tag{43}$$

The fourth step, motivated by the proof of Theorem 1, is to bound from below the supremum of expectations on the right side of (43) by Bayesian risks via introducing independent and zero mean normal random variables ζ_{skj} with the above-defined corresponding variances τ_{skj}^2. Then a direct calculation shows that for the considered regression this step yields a smaller lower bound than the verified one. This is the earlier-mentioned place were a more complicated parametric class and less favorable prior is needed. We do that by creating another layer of parameters, and then by using new normal variables for defining the desired least favorable prior. Set

$$S(t|x) := S_0(t|x) + \sum_{k=1}^{s-2} \sum_{j=J'(k)}^{J(k)} \sum_{r=1}^{s-2} \sum_{i=1}^{\lceil \ln(s) \rceil} \kappa_{skjri} \varphi_{skj}(x) \phi_{sk}(x) \psi_{sri}(t).$$

Here $\psi_{sri}(t) := (s/t^*)^{1/2} \psi_i(st/t^* - r)$, $\psi_i(t) = 2^{1/2} \sin(\pi i t) I(t \in [0,1])$. Note that if $|\kappa_{skjri}| \leq n^{-1/3}/s^5$ (and compare this bound with τ_{skj} being of order $n^{-1/2}$ in the proof of Theorem 1), then $|\partial[S(t|x) - S_0(t|x)]/\partial t| = o_n(1)$. This and Assumption 3 allow us to conclude that $S(t|x)$ is a bona fide survival function for all large n. Further, the corresponding regression is

$$m(x) = m_0(x) + \sum_{k=1}^{s-2} \sum_{j=J'(k)}^{J(k)} \left[\sum_{r=1}^{s-2} \sum_{i=1}^{\lceil \ln(s) \rceil} b_{sri} \kappa_{skjri} \right] \varphi_{skj}(x) \phi_{sk}(x),$$

where $b_{sri} := \int_0^{t^*} \psi_{sri}(t) dt$. Using the Parseval identity we get $\sum_{i=1}^{\infty} b_{sri}^2 = t^*/s$. The last equality allows us to introduce independent Normal random variables ζ_{skjri} with zero mean and variance

$$\frac{S_0(r/s|k/s)}{f^X(k/s) f^C(r/s)} \left[n^{-1}(1 - 3q_*^{-1}) \max(q_*^{-1}, \min(q_*, (J(k)/j)^\alpha - 1)) \right],$$

compare with τ_{skj}^2. Further, a direct calculation shows that

$$\mathbb{E} \left\{ \left(\sum_{r=1}^{s-2} \sum_{i=1}^{\lceil \ln(s) \rceil} b_{sri} \zeta_{skjri} \right)^2 \right\} = \tau_{skj}^2 (1 + o_n(1)).$$

The fifth step is to calculate parametric Fisher informations. Recall that the oracle uses only uncensored observations, and this is equivalent to having a sample of size n from the triplet $(X, (1 - \Delta)C, \Delta)$. Consider corresponding elements of the Fisher matrix,

$$\mathcal{I}_{skj}(r_1, i_1, r_2, i_2) := \mathbb{E} \left\{ \prod_{l=1}^{2} [\partial \ln(f^X(X) f^{X,(1-\Delta)C,\Delta}(X, (1-\Delta)C, \Delta)/\partial \kappa_{skjr_l i_l}] \right\}.$$

Direct calculations, similar to those in the proof of Theorem 1, yield that $\mathcal{I}_{skj}$ is a block-diagonal matrix and $\mathcal{I}_{skj} = \mathrm{diag}(B_1, \ldots, B_{s-2})$ where each B_r is a $\lceil \ln(s) \rceil \times \lceil \ln(s) \rceil$ matrix with diagonal elements $B_r(i_1, i_1) = [S_0(r/s|k/s)/f^X(k/s) f^C(r/s)]^{-1}(1 + o_n^*(1))$, where for some finite constant B_* we have $|o_n(1)| < B_*/s$ uniformly over all considered parameters, and absolute values of all other elements are bounded by B_*/s. Accordingly, the inverse Fisher matrix $\mathcal{I}_{skj}^{-1}$ satisfies for the vector-row $\vec{b}_{sk} := (b_{sk11}, \ldots, b_{sk(s-2)\lceil \ln(s) \rceil})$ the relation

$$\vec{b}_{sk} \mathcal{I}_{skj}^{-1} \vec{b}_{sk}^T = \int_0^{t^*} \frac{S_0(t|k/s)}{f^X(k/s) f^C(t)} dt (1 + o_n(1)) = I_{sk}^{-1}(1 + o_n(1)).$$

This, (41), (42) and direct calculations yield

$$\inf_{\vec{\nu}_{sk} \in \Theta_{sk}} \sup \sum_{j=J'(k)}^{J(k)} \mathbb{E}\{(\tilde{\nu}_{skj}^* - \nu_{skj})^2\} \geq (n I_{sk})^{-2\alpha/(2\alpha+1)} P(1 + o_n(1)), \tag{44}$$

where the infimum is over all possible oracle-estimators of $\vec{v}_{sk}$ considered in Theorem 3. The rest of the proof of the lower bound (31) follows along lines of the proof of Theorem 1.

Now let us show that the lower bound is sharp and is attainable by the series estimate (33). For the proposed Fourier coefficient estimator (34) we can write,

$$\mathbb{E}\{\hat{\theta}_{cj}^*\} = \mathbb{E}\left\{\frac{(1-\Delta)\varphi_j(X)}{f^X(X)f^C(V)}\right\}$$

$$= \int_0^1 \int_0^\infty \frac{f^X(x)S^{T|X}(t|x)f^C(t)\varphi_j(x)}{f^X(x)f^C(t)}dtdx$$

$$= \int_0^1 \left[\int_0^\infty S^{T|X}(t|x)dt\right]\varphi_j(x)dx = \int_0^1 m(x)\varphi_j(x)dx = \theta_j.$$

Thus the estimate is unbiased. Further,

$$\mathbb{E}\{(\hat{\theta}_{cj}^* - \theta_j)^2\} = n^{-1}\left[\mathbb{E}\left\{\frac{(1-\Delta)\varphi_j^2(X)^2}{f^X(X)f^C(V)}\right\} - \theta_j^2\right] = n^{-1}D_c(1 + o_j(1)).$$

This verifies (35). Using these two results, we can use the proof of Theorem 2 to verify efficiency of (33) and sharpness of the lower bound. Theorem 3 is proved.

5. CONCLUSIONS

Right-censoring partitions data into uncensored and censored observations when either the lifetime of interest T or the censoring variable C are observed. The dominance of uncensored observations over censored ones is a familiar principle in the survival analysis literature. The paper addresses the dominance principle both theoretically, using the oracle's approach, and numerically. The obtained answer is two-fold. First, for nonparametric density estimation the dominance principle is correct, and the problem of density estimation based on censored-data is ill-posed. This is the bad news, the good news is that the ill-posedness is defined in frequency domain with its onset on low frequencies. Accordingly, it may be beneficial to aggregate uncensored and censored observations for estimating low-frequency components of the density. Second, for nonparametric regression censored-data are not ill-posed, and then their special aggregation in frequency domain is beneficial.

It is important to stress that the proposed aggregation is different from the classical one performed in the time domain. The time-domain aggregation uses a set of already calculated nonparametric estimates and then tries to create a better one with the main application being adaptation to unknown smoothness and/or dimensionality. The proposed methodology suggests to aggregate individual Fourier coefficient estimates based on complementary subsamples of uncensored and censored observations, that is the aggregation is done in the frequency domain.

The developed aggregation methodology is beneficial for data with high rate censoring and/or small sample sizes. Further, the improvement may be dramatic for data with sparse right-tail observations.

Let us also comment on one interesting byproduct of the developed oracle's theory. It shows that if distribution of the censoring variable is known, then the oracle recommends to use relatively simple efficient estimators. Accordingly, whenever possible it is desirable to get an extra information about the censoring variable and then mimic the oracle. The environmental example of Subsection 3.4.1 illustrates this possibility.

Now let us briefly comment on topics for future research.

(1) Considered model of nonparametric regression assumes that the predictor and response (X,T) are independent of the censoring variable C. This is a classical model and occurs in many applications, see [32, 34, 39]. A more general model is when T and C are conditionally independent given predictor X. In this case the developed methodology is still applicable with the estimates $\hat{S}^C$ and $\hat{f}^C$ being replaced by $\hat{S}^{C|X}$ and $\hat{f}^{C|X}$, respectively.

(2) Estimation of conditional density $f^{T|X}$ is theoretically challenging and practically important problem. Note that both the classical and quantile nonparametric regressions are functionals of the conditional density.

(3) Nonparametric estimation of the hazard rate and conditional hazard rate. This topic is of a particular interest in actuarial science, biostatistics and reliability theory, see [20].

(4) Missing data is a traditional complication in survival analysis. It is known that in the case of a classical (no censoring) regression, cases of missing responses and predictors require different methods of estimation. It will be of interest to understand how missing and right-censoring interact, and then develop optimal estimators.

ACKNOWLEDGEMENTS

FUNDING

CONFLICT OF INTEREST

The author of this work declares that he has no conflicts of interest.

REFERENCES

1. O. Aalen, "Nonparametric inference for a family of counting processes," Ann. Statist. **6**, 701−726 (1978).

2. O. Aalen, O. Borgan, and H. Gjessing, *Survival and Event History Analysis, A Process Point of View* (Springer, New York, 2008).

3. B. Bayowa, O. Ebenezer, E. Osafu, and A. Ayantunji, "A simulated based comparative study of some tests for checking homogeneity of non crossing survival curves under high censoring rates," J. Appl. Probab. Statist. **17**, 87−99 (2022).

4. N. Breslow and J. Crowley, "A large sample study of the life-table and product-limit estimates under random censorship," Ann. Statist. **3**, 437−453 (1074).

5. E. Brunel and F. Comte, "Nonparametric adaptive regression estimation in presence of censoring," Math. Meth. Statist. **15**, 233−255 (2006).

6. E. Brunel, F. Comte, and A. Guilloux, "Nonparametric density estimation in presence of bias and censoring," Test **18**, 166−194 (2009).

7. W. Buon, C. Fletcher, and R. Brundage, "Impact of censoring data below an arbitrary quantification limit on structural model misspecification," J. Pharmacokinet Pharmacodyn **35**, 101−116 (2008).

8. J. Buckley and I. James, "Linear regression with censored data," Biometrika **66**, 429−436 (1979).

9. X. Cheng, S. Wang, H. Wang, and S. Ng, "Deep survival forests for extremely high censored data," Applied Intelligence **53**, 7041−7055 (2023).

10. P. Chu, J. Wang, J. Hwang, and Y. Chang, "Estimation of life expectancy and the expected years of life lost in patients with major cancers: extrapolation of survival curves under high-censored rates," Value in Health **11**, 1102−1109 (2008).

11. D. Cox, "Regression models and life-tables, with discussion," J. R. Statist. Soc. B **34**, 187−220 (1972).

12. D. Cox, "Partial likelihood," Biometrika **62**, 269−276 (1975).

13. Y. Cui and J. Hannig, (2019). "Nonparametric generalized fiducial inference for survival functions under censoring," Biometrika **106**, 501−518 (2019).

14. S. Cummings, *The Use of Survival Analysis Techniques among Highly Censored Data Sets,* 1−38 (2013), www.witenberg.edu.

15. M. Davis, *Water and Wastewater Engineering*, 2nd ed. (McGraw Hill, New York, 2020).

16. M. Delecroix, O. Lopez, and V. Patilea, "Nonlinear censored regression using synthetic data," Scand. J. Stat. **35**, 248−265 (2008).

17. S. Efromovich, *Nonparametric Curve Estimation* (Springer, New York, 1999).

18. S. Efromovich, "Density estimation under random censorship and order restrictions," JASA **96**, 667−684 (2001).

19. S. Efromovich, " Minimax theory of nonparametric hazard rate estimation: Efficiency and adaptation," Ann. Instit. Mathem. Statist. **68**, 25−75 (2016).

20. S. Efromovich, *Missing and Modified Data in Nonparametric Estimation with R Examples* (Chapman and Hall, Boca Raton, 2018).

21. B. Efron, "The two sample problem with censored data," Proc. 5th Berkeley Symp. **4**, 831−854 (1967).

22. B. Efron, "The efficiency of Cox's likelihood function for censored data," JASA **72**, 557−565 (1977).

23. J. Fan and I. Gijbels, "Censored regression: Local linear approximations and their applications," JASA **89**, 560−570 (1994).

24. C. Foster, *Wastewater Treatment and Technology* (Thomas Telford, London, 2003).

25. T. Gross and T. Lai, "Nonparametric estimation and regression analysis with left-truncated and right-censored data," JASA **91**, 1166−1180 (1996).

26. Z. Guessoum and E. Said, "Kernel regression uniform rate estimation for censored data under α-mixing condition," Electr. J. Statist. **4**, 117−132 (2010).

27. M. Hoffmann and O. Lepski, "Random rates in anisotropic regression (with discussion)," Ann. Statist. **30**, 325−396 (2022).

28. H. Jankowski and J. Wellner, "Nonparametric estimation of a convex bathtab-shaped hazard function," Bernoulli **15**, 1010−1035 (2009).

29. Z. Jin, D. Lin, and Z. Ying, "On least-squares regression with censored data," Biometrika **93**, 147−161 (2006).

30. E. Kaplan, and P. Meier, "Nonparametric estimation with incomplete observations," JASA **53**, 457−481 (1958).

31. H. Kim and Y. Truong, "Nonparametric regression estimates with censored data: Local linear smoothers and their applications," Biometrics **54**, 1434−1444 (1998).

32. J. Klein, H. van Houwelingen, J. Ibrahim, and T. Sheike, *Handbook of Survival Analysis* (Chapman and Hall, Boca Raton, 2014).

33. M. Kohler, K. Mathe, and M. Pinter, "Prediction from randomly right censored data," J. Mult. Anal. **80**, 73−100 (2002).

34. C. Legrand, *Advanced Survival Models* (Chapman and Hall, Boca Raton, 2021)

35. A. Letki, "Know when to turn to centrifugal separation," Chem. Engng. Progr. **94**, 29−44 (1998).

36. L. Li, B. MacGibbon, and C. Valenta, "On the optimality of wavelet-based nonparametric regression with censored data," J. Appl. Probab. Statist. **3**, 243−261 (2008).

37. A. Maksymiuk et al., "Sequencing and schedule effects of cisplatin plus etoposide in small cell lung cancer results of a north central cancer treatment group randomized clinical trial," J. Clinical Oncology **12**, 70−76 (1994).

38. P. Meier, "Estimation of a distribution function from incomplete observations," J. of Appl. Probab. **12**, 67−87 (1975).

39. R. Miller and J. Halperin, "Regression with censored data," Biometrika **69**, 521−531 (1982).

40. A. Nemirovski, *Topics in Non-parametric Statistics,* Lecture Notes in Mathematics, (Springer, New York, 2000), Vol. 1738.

41. D. Oakes, "The asymptotic information in censored survival data," Biometrika **64**, 441−448 (1977).

42. V. Poynor and A. Kottas, "Nonparametric bayesian inference for mean residual life functions in survival analysis," Biostatistics **20**, 240−255 (2019).

43. P. Rigollet and A. Tsybakov, "Linear and convex aggregation of density estimators," Mathem. Meth. Statist. **16**, 260−280 (2007).

44. Y. Ritov, "Estimation in a linear regression model with censored data," Ann. Statist. **18**, 303−328 (1990).

45. D. Rubin, *Multiple Imputation for Nonresponse in Surveys* (Wiley, New York, 1987).

46. K. Salah and S.Yousri, "Nonparametric relative regression under random censorship model," Statist. Probab. Letters **151**, 116−122 (2019).

47. A. Samarov and A.Tsybakov, "Aggregation of density estimators and dimension reduction," in *Advances in Statistical Modeling and Inference,* Essays in Honor of K. Doksum, V. Nair, ed., 233−251 (2007).

48. E. Shirazi, H. Doosti , H. Niroumand, and N. Hosseinioun, (2013). "Nonparametric regression estimates with censored data based on block thresholding method," J. Statist. Plan. Infer. **143**, 1150−1165 (2013).

49. Z. Ying, S. Jung, and L. Wei, "Survival analysis with median regression models," JASA **90**, 178−184 (1995).

50. M. Yu, W. Zhao, Y. Zhou and C. Wu, "Robust online detection on highly censored data using a semi-parametric EWMA chart," J. Statist. Comp. Simul. **93**, 1403−1419 (2022).

51. Y. Yildiray, "Estimating default probabilities of CMBS loans with clustering and heavy censoring," J. Real Estate Fin. Econ. **37**, 93−111 (2008).

52. Q. Zhong, J. Mueller, and J. Wang, "Deep learning for the partially linear Cox model," Ann. Statist. **50**, 1348−1375 (2022).

53. M. Zhou, *Empirical Likelihood Method in Survival Analysis* (Chapman and Hall, Boca Raton, 2019).