NONPARAMETRIC BIVARIATE DENSITY ESTIMATION FOR CENSORED LIFETIMES

BY SAM EFROMOVICH[??],

[??]*Department of Mathematical Sciences, University of Texas at Dallas,* ??

It is well known that estimation of a bivariate cumulative distribution function of a pair of right censored lifetimes presents challenges unparalleled to the univariate case where a product-limit Kaplan-Meyer's methodology typically yields optimal estimation, and the literature on optimal estimation of the joint probability density is next to none. The paper, for the first time in the survival analysis literature, develops the theory and methodology of sharp minimax and adaptive nonparametric estimation of the joint density under the mean integrated squared error (MISE) criterion. The theory shows how an underlying joint density, together with the bivariate distribution of censoring variables, affect the estimation, and what and how may or may not be estimated in the presence of censoring. Practical example illustrates the problem.

1. Introduction. The problem of estimation of the distribution of a random variable based on direct data is a classical one. For a sample $X_1, \ldots, X_n$ from a univariate random variable X, the empirical cumulative distribution function $\check{F}^X(x) := n^{-1} \sum_{l=1}^n I(X_l \le x)$ is the classical estimator of the cumulative distribution function $F^X(x) := \mathbb{P}(X \le x) = \mathbb{E}\{I(X \le x)\}$. Here $\mathbb{P}(\cdot)$, $\mathbb{E}\{\cdot\}$ and $I(\cdot)$ are the probability, the expectation and the indicator function, respectively. Note that the estimator is based on the idea of a sample mean (method of moments) estimation which yields a bouquet of excellent statistical properties. The beauty of the sample mean approach is that it is effortlessly extended to the case of a bivariate distribution. Indeed, if $(X_1, Y_1), \ldots, (X_n Y_n)$ is a sample from a pair (X, Y), then the empirical joint cumulative distribution function $\check{F}^{X,Y}(x, y) := n^{-1} \sum_{l=1}^n I(X_l \le x, Y_l \le y)$ is again a good estimator of the joint cumulative distribution function $F^{X,Y}(x, y) := \mathbb{P}(X \le x, Y \le y)$. Further, either smoothing of empirical distributions or again using a sample mean methodology yields efficient estimation of the corresponding univariate f^X and bivariate $f^{X,Y}$ densities. In particular, let us recall that if joint density $f^{X,Y}$ of the pair (X, Y) is m_X-fold differentiable in x and m_Y-fold differentiable in y, then based on a sample of size n it is possible to estimate the bivariate density with the MISE (mean integrated squared error) decreasing with the optimal rate $n^{-2\beta/(2\beta+1)}$ where

$$\beta := m_X m_Y / (m_X + m_Y) \tag{1.1}$$

is the effective smoothness. More results for direct data may be found in books Efromovich (1999,2018) and Wasserman (2006).

The situation changes rather dramatically if available observations are right censored. Because we are interested only in right censoring, in what follows we may simply say censoring in place of right censoring. For a univariate censoring setting we observe a sample of size n from a pair $(V, \Delta) := (\min(X, C), I(X \le C))$, and in a bivariate censoring we observe a sample of size n from a quartet

$$(V, W, \Delta, \Gamma) := (\min(X, C), \min(Y, D), I(X \le C), I(Y \le D)). \tag{1.2}$$

MSC2020 subject classifications: Primary 62G05, 62J05; secondary 62F35.
Keywords and phrases: Adaptation, MISE, sharp minimax, survival analysis.

Here (C, D) is a pair of continuous random censoring lifetimes, and it is assumed that the pair is independent of the lifetimes of interest (X, Y).

There are many examples in survival analysis where bivariate estimation is explored under censoring. In biomedical research, study subjects from the same cluster (e.g. family or twins) share common genetic and/or environmental factors. Another example is the time to a deterioration level or the time to reaction of a treatment in pairs of lungs, kidneys, eyes or ears of humans, or time from initiation of treatment until first response in two successive courses of a treatment in the same patient. Bivariate data are also recorded when two related diseases happened in one patient or two recurrence times of a certain disease are encountered. In all these studies the censoring may arise for a number of reasons including withdraw from the study, a change of health status or contamination, or by death from a cause unrelated to the study. In actuarial science an important example is the joint life annuities issued to married couples who tend to be exposed to similar risks and likely to have the same living habits. The insurance data is censored and the interest is in estimation of the joint distribution of future lifetimes. A reliability example for anaerobic treatment of municipal wastewater will be considered shortly in Section 5.

If we are interested in recovery of the cumulative distribution function F^X of a censored X, the underlying idea of the empirical cumulative distribution function is no longer applicable, and instead a product-limit methodology is used. See the original paper Kaplan and Meier (1958) and a discussion in books Moore (2016) and Efromovich (2018). Further, there is no straightforward extension of the univariate product-limit methodology to the bivariate case. Instead, a number of sophisticated and mathematically involved procedures, ranging from EM, hazard gradient, partial differential equations and copula to nonparametric MLE and solving an inhomogeneous Volterra equation via Peano series, have been proposed. The interested reader can find an insightful discussion of this topic in Campbell (1981), Dabrowska (1988), Oakes (1989), Prentice and Cai (1992), Pruitt (1993), Frees, Carriere and Valdez (1995), Hougaard (2000), Akritas and Van Kellogom (2003), Collett (2003), Crowder (2012), Lopez (2012), Li and Ma (2013), Prentice (2016), Prentice and Zhao (2018).

Nonparametric estimation of the joint probability density $f^{X,Y}$ in the presence of censoring is less explored. The nonparametric literature is primarily devoted to differentiation of known estimators of the cumulative distribution function, and there is no theory which sheds light on optimal density estimation. See a discussion in Wells and Yeo (1996), Dabrowska, Duffy and Zhang (1998), Kooperberg (1998), Crowder (2012), Seok, Tian and Wong (2014), Ghosal and van der Vaart (2017).

The paper extends to censored data the classical sharp minimax theory and methodology of estimation of a bivariate density under MISE criterion. Let us briefly explain an underlying idea of the proposed joint density estimator that also sheds light on context of the paper. The estimator is motivated by a sharp minimax lower bound for the MISE of an oracle-estimator that knows the survival function S of censoring variables (C, D) and smoothness of an estimated joint density f. The key oracle's conclusion is that using only uncensored pairs, when $(V, W) = (X, Y)$, is sufficient for attaining the lower bound. Then the oracle proposes a relatively simple series estimator based on the survival function S and unbiased sample mean Fourier estimates. To mimic the oracle, the proposed data-driven estimator uses the exponent of a negative sample mean estimate of the cumulative hazard function of (C, D) in place of S.

The context of the paper reflects the above-explained motivation. Section 2 presents a sharp lower bound for the oracle. Sharp-minimax oracle estimation is discussed in Section 3 which helps us to understand a proposed data-driven estimator of Section 4. Analysis of a real censored data and simulated examples can be found in Section 5. Section 6 contains proofs.

Now let us introduce main notations and shortcuts used in the paper. In what follows we distinguish between a function g and its value $g(\cdot)$. If no confusion can occur, we may write $f := f^{X,Y}$ for the joint density of interest and $S(x,y) := S^{C,D}(x,y) := \mathbb{P}(C > x, D > y)$ for the value of joint survival function of the pair (C,D) of censoring lifetimes. Denote by $f^{V,W}$ the joint density of (V,W), and denote the mixed joint density of the quartet (V,W,Δ,Γ) as

$$p(v,w,\delta,\gamma) := f^{V,W}(v,w)\mathbb{P}(\Delta = \delta, \Gamma = \gamma | V = v, W = w). \tag{1.3}$$

We may write $\mathbb{E}_{f,S}\{\cdot\}$ or $\mathbb{E}_f\{\cdot\}$ to stress that the expectation is taken given joint density f and joint survival function S or given f, respectively.

Set $q := q_n := \lceil \ln(n+20) \rceil$ for the minimal integer larger than or equal to $\ln(n+20)$, similarly $s := s_n := \lceil \ln(q_n) \rceil$. In what follows Q, a and b are positive constants, m_X and m_Y are positive integers, $R := [0,a] \times [0,b]$ is a rectangle, $R_n := \big[[3s_n]^{-2}, a(1-[3s_n]^{-2})\big] \times \big[[3s_n]^{-2}, b(1-[3s_n]^{-2})\big]$ is a smaller rectangle not including boundary strips, and we may write $\int_R g(x,y)dxdy := \int_0^a \int_0^b g(x,y)dxdy$. Set

$$a_{ij} := a_{ij}(m_X, m_Y, a, b) := 1 + (\pi i/a)^{2m_X} + (\pi j/b)^{2m_Y}. \tag{1.4}$$

The following functional will be referred to as the coefficient of difficulty,

$$d := d(f,S,a,b) := \frac{1}{ab}\int_R \frac{f(x,y)}{S(x,y)}dxdy. \tag{1.5}$$

Also introduce

$$P := P(m_X, m_Y, Q, a, b) := \Big[\frac{ab}{\pi^2}\Big]^{2\beta/(2\beta+1)}\Big[\frac{Q}{C_1(m_X,m_Y)}\Big]^{1/(2\beta+1)} C_2(m_X,m_Y). \tag{1.6}$$

Here for positive k and r,

$$C_1(k,r) := \int_{\{(u,v):\, u^{2k}+v^{2r} \le 1;\, u,v \ge 0\}} [(u^{2k}+v^{2r})^{1/2} - (u^{2k}+v^{2r})]dudv, \tag{1.7}$$

$$C_2(k,r) := \int_{\{(u,v):\, u^{2k}+v^{2r} \le 1;\, u,v \ge 0\}} [1 - (u^{2k}+v^{2r})^{1/2}]dudv. \tag{1.8}$$

To honor the pioneering paper Pinsker (1980) on sharp minimax estimation, the P in (1.6) will be referred to as the Pinsker constant.

For square-integrable functions on a rectangle R we will use a tensor-product cosine basis

$$\varphi_{i,j}(x,y) := \varphi_{ij}(x,y) := \varphi_i(x|a)\varphi_j(y|b), \;\; (x,y) \in R, \; i,j = 0,1,\dots \tag{1.9}$$

where

$$\varphi_k(z|c) := \frac{1}{\sqrt{c}}\big[I(k=0) + \sqrt{2}\cos\big(\frac{\pi k z}{c}\big)I(k \ge 1)\big].$$

Finally, let us present an assumption that is used throughout the paper.

Assumption 1. *Pair of continuous lifetimes of interest (X,Y) is independent of pair of continuous censoring lifetimes (C,D). The joint density f of the lifetimes of interest is estimated on a rectangle $R = [0,a] \times [0,b]$. The joint survival function S of the censoring lifetimes is positive on R, that is $S(a,b) > 0$. If S is unknown, then the joint survival function of the lifetimes of interest is positive on R.*

Note that the made assumption about a positive joint survival function is traditional. Indeed, the rectangle R must be a subset of the support of (C,D) for consistent estimation of the distribution of (X,Y) on R. A similar conclusion holds for estimation of the distribution of (C,D).

2. Sharp-mimimax oracle's lower bound. The aim is to estimate the joint density f of a pair of continuos random lifetimes (X, Y) over a rectangle $R = [0, a] \times [0, b]$ with a minimal MISE. We observe the pair in the presence of right censoring, namely we observe a sample of size n from the quartet $(V, W\Delta, \Gamma)$ defined in (1.2). As we will see shortly, the density of interest affects the MISE, and accordingly we consider a minimax approach over a class of possible underlying densities shrinking, as n increases, toward a pivotal joint density $f_0(x, y)$, $(x, y) \in [0, \infty)^2$. In what follows it is always assumed that f_0 is continuous and positive on R.

To define a shrinking minimax approach, we introduce a class of additive perturbations of the pivot f_0 on the rectangle R. Recall notation (1.4) for a_{ij}, and following Nikolskii (1975) and Hoffmann and Lepski (2002), we begin with a classical anisotropic global Sobolev class of m_X-fold differentiable in x and m_Y-fold differentiable in y bivariate functions on $[0, \infty)^2$,

$$\mathcal{G}(m_X, m_Y, Q, a, b) := \{g: \ g(x, y) = \sum_{i,j=0}^{\infty} \theta_{ij}\varphi_{ij}(x, y), \ \sum_{i,j=0}^{\infty} a_{ij}\theta_{ij}^2 \le Q\}. \tag{2.1}$$

Recall that sequence q_n and rectangles R and R_n are defined at the end of the Introduction, and set $I_R := I((x, y) \in R)$. Introduce a shrinking local Sobolev class of joint densities for two lifetimes,

$$\mathcal{F}_n := \mathcal{F}_n(f_0, m_X, m_Y, Q, a, b) \tag{2.2}$$

$$:= \Big\{f: \ f(x, y) = f_0(x, y) + g(x, y)I_R, \ (x, y) \in [0, \infty)^2,$$

$$g \in \mathcal{G}(m_X, m_Y, Q, a, b), \ \max_{(x,y)\in R} |g(x, y)| \le \min_{(x,y)\in R} f_0(x, y)/q_n,$$

$$g(x, y) = 0 \ \text{ whenever} (x, y) \notin R_n, \ \int_0^a g(x, y)dx = \int_0^b g(x, y)dy = 0\Big\}.$$

Let us comment on the shrinking local Sobolev class (2.2). The second from the top line defines an underlying joint density f as an additive perturbation of f_0 on the rectangle R. The third line states that a perturbation g must belong to the Sobolev class, and that considered densities f shrink in L_∞-norm toward the pivotal density f_0 as n increases. The first requirement on g in the bottom line allows us to preserve smoothness of f near boundaries of R. The second requirement in the bottom line, together with the third line, imply that all functions f are bona fide densities on $[0, \infty)^2$. In short, the pivotal density is preserved beyond the smaller rectangle R_n, within R_n it is additively perturbated by shrinking Sobolev functions, and all perturbations are bona fide densities on $[0, \infty)^2$.

Theorem 1 (Oracle's Lower Bound). *The problem is to estimate a joint density f of lifetimes of interest (X, Y) by an oracle-estimator $\tilde{f}_*$ on a rectangle $R = [0, a] \times [0, b]$. Suppose that the survival function S of censoring lifetimes (C, D) is known, Assumption 1 holds, and f belongs to a shrinking local Sobolev class $\mathcal{F}_n$ defined in (2.2) where the pivotal density f_0 is continuous and positive on R. The oracle knows a censored sample $(V_l, W_l, \Delta_l, \Gamma_l)$, $l = 1, 2, \dots, n$ from the quartet (1.2) and everything about the class $\mathcal{F}_n$. Then the following lower bound for minimax MISE holds,*

$$\inf_{\tilde{f}_*} \sup_{f\in\mathcal{F}_n} [n/d]^{2\beta/(2\beta+1)}\mathbb{E}_{f,S}\{\int_R (\tilde{f}_*(x, y) - f(x, y))^2dxdy\} \le P(1 + o_n(1)). \tag{2.3}$$

Here the effective smoothness β, the coefficient of difficulty d and the Pinsker constant P are defined in (1.1), (1.5) and (1.6), respectively. Further, the Pinsker constant does not increase if only a subsample of uncensored observations with $\Delta_l\Gamma_l = 1$ is used.

It will be proved shortly that the oracle's lower bound (2.3) is sharp, meaning that an oracle-estimator and a data-driven estimator attain it. Note that the rate $n^{-2\beta/(2\beta+1)}$ is the same as for direct data, and the coefficient of difficulty indicates that the MISE is affected by the ratio f/S. Further, if the survival function S is known, then only uncensored observations may be used for sharp estimation. As we will see shortly in Section 3, the latter yields a relatively simple series oracle-estimator based on an unbiased sample mean estimate of Fourier coefficients.

3. Sharp minimax oracle-estimator. This section explains, with the help of a friendly oracle, how a sample mean methodology, so efficiently used for direct data, may be also utilized for estimating a bivariate density in the presence of censoring. Further, to help us with finding a feasible data-driven estimator, the oracle will step-by-step remove facts known only to the oracle. At the end of this section the oracle will use only a triplet (m_X, m_Y, Q), and then the case of an unknown triplet will be addressed in the next section.

We are interested in estimation of a joint density f of (X, Y) on a rectangle $R = [0, a] \times [0, b]$. It is assumed that f is square integrable on R, and recall notation (1.9) for the tensor-product cosine basis $\varphi_{ij}(x, y)$ on R. Then for $(x, y) \in R$ a density f can be written as a Fourier series via its Fourier coefficients θ_{ij},

$$f(x,y) = \sum_{i,j=0}^{\infty} \theta_{ij}\varphi_{ij}(x,y), \quad \theta_{ij} := \int_R f(x,y)\varphi_{ij}(x,y)dxdy. \tag{3.1}$$

Our aim is to suggest a series sharp-minimax oracle-estimator of f whose MISE attains the lower bound of Theorem 1. We also are going to use the hint of Theorem 1 that an oracle may use only uncensored observations. Using notation (1.3) for the mixed joint density p of the quartet (V, W, Δ, Γ) defined in (1.2), we can write

$$p(x, y, 1, 1) = f(x, y)S(x, y). \tag{3.2}$$

This relation and (3.1) yield

$$\theta_{ij} = \int_R \frac{p(x,y,1,1)\varphi_{ij}(x,y)}{S(x,y)}dxdy = \mathbb{E}\Big\{\frac{\Delta\Gamma\varphi_{ij}(V,W)I((V,W)\in R)}{S(V,W)}\Big\}. \tag{3.3}$$

The oracle knows the joint survival function S, and we get the following sample mean Fourier estimate based solely on uncensored observations,

$$\hat{\theta}^*_{ij} := n^{-1}\sum_{l=1}^{n} \frac{\Delta_l\Gamma_l\varphi_{ij}(V_l,W_l)I((V_l,W_l)\in R)}{S(V_l,W_l)}. \tag{3.4}$$

If the oracle does not want to use S, then the following estimate may be plugged in (3.4),

$$\hat{S}(V_l,W_l) := n^{-1} + \exp\Big\{-\sum_{k=1}^{n}\Big[\frac{(1-\Delta_k)I(V_k\le V_l)}{\sum_{r=1}^{n} I(V_r\ge V_k)}\Big] - \sum_{k=1}^{n}\Big[\frac{(1-\Gamma_k)I(V_k>V_l)I(W_k\le W_l)}{1+\sum_{r=1}^{n} I(V_r\ge V_l)I(W_r\ge W_k)}\Big]\Big\}, \quad (V_l,W_l)\in R. \tag{3.5}$$

We postpone discussion of this estimator until the end of this section.

Now we plug $\hat{S}$ in (3.4) and get a Fourier estimator

$$\hat{\theta}_{ij} := n^{-1} \sum_{l=1}^{n} \frac{\Delta_l \Gamma_l \, \varphi_{ij}(V_l, W_l) I((V_l, W_l) \in R)}{\hat{S}(V_l, W_l)}. \tag{3.6}$$

Note that this Fourier estimator uses all available observations.

To define a series oracle-estimator of f we need several new notations. Set

$$a_n := a_n(Q) := \Big[\frac{\pi^2 Q}{abC_1(m_X, m_Y)} \frac{n}{d}\Big]^{2\beta/(2\beta+1)}, \tag{3.7}$$

where $C_1(m_X, m_Y)$ is defined in (1.7) and the coefficient of difficulty d in (1.5). Recall that the coefficient of difficulty depends on both f and S. Because both these functions may be unknown and d is used in the denominator of (3.7), the oracle proposes to evaluate the coefficient of difficulty by a truncated from below sample mean estimate

$$\hat{d} := \hat{d}(a,b) := \max\big(q_n^{-1}, \tilde{d}(a,b)\big), \quad \tilde{d}(a,b) := \frac{1}{n} \sum_{l=1}^{n} \frac{\Delta_l \Gamma_l I((V_l, W_l) \in R)}{ab[\hat{S}(V_l, W_l)]^2}. \tag{3.8}$$

Recall that sequence q_n is defined in the Introduction. Then we plug $\hat{d}$ in (3.7) and get

$$\hat{a}_n := \Big[\frac{\pi^2 Q}{abC_1(m_X, m_Y)} \frac{n}{\hat{d}}\Big]^{2\beta/(2\beta+1)}. \tag{3.9}$$

Now we are ready to present our first oracle-estimator of density f based on censored data and parameters (m_X, m_Y, Q) of an underlying functional class. The oracle-estimator is

$$\tilde{f}_*(x,y) := \sum_{i,j \geq 0} \lambda_{ij} \hat{\theta}_{ij} \varphi_{ij}(x,y), \tag{3.10}$$

where the shrinkage coefficients are

$$\lambda_{ij} := I(i \leq q_n) I(j \leq q_n) + [1 - (a_{ij}/\hat{a}_n)^{1/2}] I(\max(i,j) > q_n) I(a_{ij} \leq \hat{a}_n). \tag{3.11}$$

Note that, to help the statistician, the oracle does not use the pivotal density f_0 which was used in establishing the lower bound. To offset this knowledge in establishing an upper bound, it will be assumed that on the rectanle R the pivot is smoother than a perturbation. The latter is a traditional assumption in the local minimax literature that goes back to Golubev (1991). Recall that $\mathcal{F}_n$ and P are defined in (2.2) and (1.6), respectively.

Theorem 2 (Oracle's upper bound). *Let Assumption 1 hold and $f_0 \in \mathcal{G}(m_X + 1, m_Y + 1, Q', a, b)$. Then the MISE of oracle-estimator (3.10) attains the lower bound (2.3) and*

$$\sup_{f \in \mathcal{F}_n} [n/d]^{2\beta/(2\beta+1)} \, \mathbb{E}_{f,S}\Big\{ \int_R (\tilde{f}_*(x,y) - f(x,y))^2 dx dy \Big\} = P(1 + o_n(1)). \tag{3.12}$$

This result proves that the oracle's lower bound (2.3) is sharp and attainable by an oracle-estimator that knows data and smoothness of density f. The next section explains how to develop a data-driven estimator that adapts to the smoothness.

We are finishing this section by commenting upon estimate (3.5) of the joint survival function S of censoring lifetimes (C, D). Let c^*, c_k^*, $k = 0, 1, \ldots$ denote generic positive constants that may depend on $S^{V,W}(a,b)$. The following lemma sheds light on basic properties of $\hat{S}(V_l, W_l)$ used in the denominator of (3.6).

Lemma 1. *Let S be unknown and Assumption 1 hold. Consider $(x,y) \in R$, $Z_1 := (V_1, W_1, \Delta_1, \Gamma_1)$, $z := (x, y, \delta, \gamma)$. Then*

$$|\mathbb{E}_{f,S}\{[\hat{S}(V_1, W_1) - S(V_1, W_1)] | Z_1 = z\}| \leq c_0^* n^{-1}, \tag{3.13}$$

for a positive integer k

$$\mathbb{E}_{f,S}\{[\hat{S}(V_1,W_1)\} - S(V_1,W_1)]^{2k}|Z_1 = z\} \le c_k^*\, n^{-k}, \tag{3.14}$$

and for $\epsilon > 0$

$$\mathbb{P}_{f,S}(|\hat{S}(V_1,W_1) - S(V_1,W_1)| \ge \epsilon | Z_1 = z) \le c^* n^2 e^{-n\epsilon^2/c^*}. \tag{3.15}$$

Several more comments about $\hat{S}$ are due. First, we add n^{-1} in (3.5) to make the estimate bounded below from zero, and this choice is explained by the fact that bias of the exponential part is of order n^{-1}, see (3.13). Second, we may write (3.5) as $\hat{S}(x,y) =: n^{-1} + e^{-\hat{H}(x,y)}$. Here $\hat{H}$ is a sample mean estimate of the cumulative hazard $H := -\ln(S)$. Third, the denominators in (3.5) are at least 1. Finally, (3.15) implies that if $(V_l, W_l) \in R$ then the probability that $\hat{S}(V_l, W_l) < S(a,b)/2$ is exponentially small in n.

4. Data-driven estimation. To understand how to construct a data-driven sharp minimax estimator, we again begin with an oracle-estimator that instead of an unknown triplet (m_X, m_Y, Q) uses a functional of an underlying bivariate density of interest f. As we will see shortly, this approach will lead us to a relatively simple data-driven estimator that mimics the oracle and does not require solving numerical optimization problems. Recall our notations $q = q_n$ and $s = s_n$ defined at the end of the Introduction and introduce an increasing sequence of integers $b_1 = 0, b_2 = b_1 + 1, \ldots, b_q = b_{q-1} + 1$, and $b_{q+k} = b_{q+k-1} + \lceil (1+1/s)^k \rceil$ for $k = 1, 2, \ldots$ Set $L_k := b_{k+1} - b_k$ and define $K := K_n$ as the smallest integer such that $\sum_{k=1}^K L_k > n^{1/4}s$. Next, for positive integers k and τ introduce blocks of nonnegative integers $B_{k\tau} := \{(i,j) : b_k \le i < b_{k+1},\ b_\tau \le j < b_{\tau+1}\}$, denote the cardinality (number of elements) of $B_{k\tau}$ as $L_{k\tau} := L_k L_\tau$, and set $t_{k\tau} := 1/\ln(\ln((k+20)(\tau+20)))$.

An oracle-estimator, that does not use the triplet (m_X, m_Y, Q), is defined as

$$\hat{f}_*(x,y) := \sum_{k,\tau=1}^{K} \Lambda_{k\tau} \sum_{(i,j)\in B_{k\tau}} \hat{\theta}_{ij}\varphi_{ij}(x,y). \tag{4.1}$$

Here

$$\Lambda_{k\tau} := I(k \le q, \tau \le q) + \frac{\Theta_{k\tau}}{\Theta_{k\tau} + \hat{d}n^{-1}} I(\Theta_{k\tau} > t_{k\tau}\hat{d}n^{-1}) I(\max(k,\tau) > q) \tag{4.2}$$

are smoothing weights and

$$\Theta_{k\tau} := L_{k\tau}^{-1} \sum_{i,j\in B_{k\tau}} \theta_{ij}^2 \tag{4.3}$$

are classical Sobolev functionals. The estimates $\hat{\theta}_{ij}$ are defined in (3.6) and $\hat{d}$ in (3.8). Also recall that $\mathcal{F}_n$ and P are defined in (2.2) and (1.6), respectively.

Theorem 3 (Oracle's upper bound). *Let assumptions of Theorem 2 hold. Then the oracle-estimator (4.1) is sharp-minimax and*

$$\sup_{f\in\mathcal{F}_n} [n/d]^{2\beta/(2\beta+1)}\, \mathbb{E}_{f,S}\{\int_R (\hat{f}_*(x,y) - f(x,y))^2 dxdy\} \le P(1 + o_n(1)). \tag{4.4}$$

To mimic the oracle-estimator, one only needs to estimate the Sobolev functionals (4.3). This is done by an asymptotically unbiased estimator

$$\hat{\Theta}_{k\tau} := L_{k\tau}^{-1} \sum_{(i,j)\in B_{k\tau}} \hat{\theta}_{ij}^2 - \hat{d}n^{-1}. \tag{4.5}$$

Then the recommended data-driven estimator is

$$\hat{f}(x,y) := \sum_{k,\tau=1}^{K} \hat{\Lambda}_{k\tau} \sum_{(i,j)\in B_{k\tau}} \hat{\theta}_{ij}\varphi_{ij}(x,y), \tag{4.6}$$

where

$$\hat{\Lambda}_{k\tau} := I(k \le q, \tau \le q) + \frac{\hat{\Theta}_{k\tau}}{\hat{\Theta}_{k\tau} + \hat{d}n^{-1}} I(\hat{\Theta}_{k\tau} > t_{k\tau}\hat{d}n^{-1}) I(\max(k,\tau) > q). \tag{4.7}$$

Theorem 4 (Sharp minimax data-driven estimator). *Under assumptions of Theorem 2, the MISE of data-driven estimator (4.6) satisfies*

$$\sup_{f\in\mathcal{F}_n} [n/d]^{2\beta/(2\beta+1)}\, \mathbb{E}_{f,S}\{\int_R (\hat{f}(x,y) - f(x,y))^2 dxdy\} \le P(1 + o_n(1)). \tag{4.8}$$

This implies that the estimator is sharp-minimax and adapts to unknown smoothness of an underlying joint density of interest f and an unknown nuisance joint survival function S.

Note how simple the adaptive density estimator (4.6) is and that it does not require solving optimization problems traditionally used for adaptation, see Wassermann (2006).

Remark 1. In (4.6) the unit weights $\hat{\Lambda}_{k\tau} = 1$, used for low frequencies $i, j \le q$, may be replaced by classical hard thresholds $I(\hat{\theta}_{ij}^2 > 2q\hat{d}n^{-1})$. It will be shown in Section 6 that the replacement does not change (4.8), and it is recommended in Efromovich (1999) for small samples. Also, if a density estimate takes on negative values, then its bona fide projection is used, see Efromovich (1999). This modification is used in the next section.

5. Practical example and simulations. Aeration is an essential process in the majority of wastewater treatment plants, see a discussion in Rosso et al. (2008) and Albu et al. (2021). Aeration introduces air into a wastewater, providing an aerobic environment for microbial degradation of organic matter. The purpose of aeration is to supply the required oxygen to the metabolizing microorganisms and to provide mixing so that microorganisms come into contact with the dissolved and suspended organic matter. For now the most common aeration system introduces air by fine pore diffusers submerged in the wastewater. The diffusers produce very small bubbles, and smaller bubbles result in more bubble surface area per unit volume and greater oxygen transfer efficiency. On the other hand, the diffusers are susceptible to chemical and biological fouling, and as a result require routine cleaning/replacement.

The environmental company BIFAR was interested in comparing lifetimes of two types of diffusers. Let us refer to these two types as X-diffusers and Y-diffusers. In an experiment, BIFAR studied the diffusers in pairs under the same wastewater characteristics and quality of maintenance. Accordingly, in the experiment lifetimes X and Y of the diffusers may be dependent. A serious complication of the BIFAR's experiment is that lifetime of a diffuser is comparable with lifetimes of other parts of an aeration system. Accordingly, in the BIFAR experiment lifetimes (X, Y) may be right censored by censoring lifetimes (C, D) of the aeration system.

The top diagram in Figure 1 shows the BIFAR data, and the caption explains all four different types of observations available for a pair of censored variables. Let us look at the diagram. Probably the first what catches the eye is the straight line of crosses and that other crosses are above the line. A cross corresponds to a pair of diffusers whose lifetimes are censored, that is $\Delta = \Gamma = 0$. This particular pattern of crosses points upon a possibility that $C = \min(D, \xi)$, and indeed the latter is the case due to using an additional pressure intensifier for X-diffusers. Another interesting observation is that overall an aeration equipment may fail much earlier

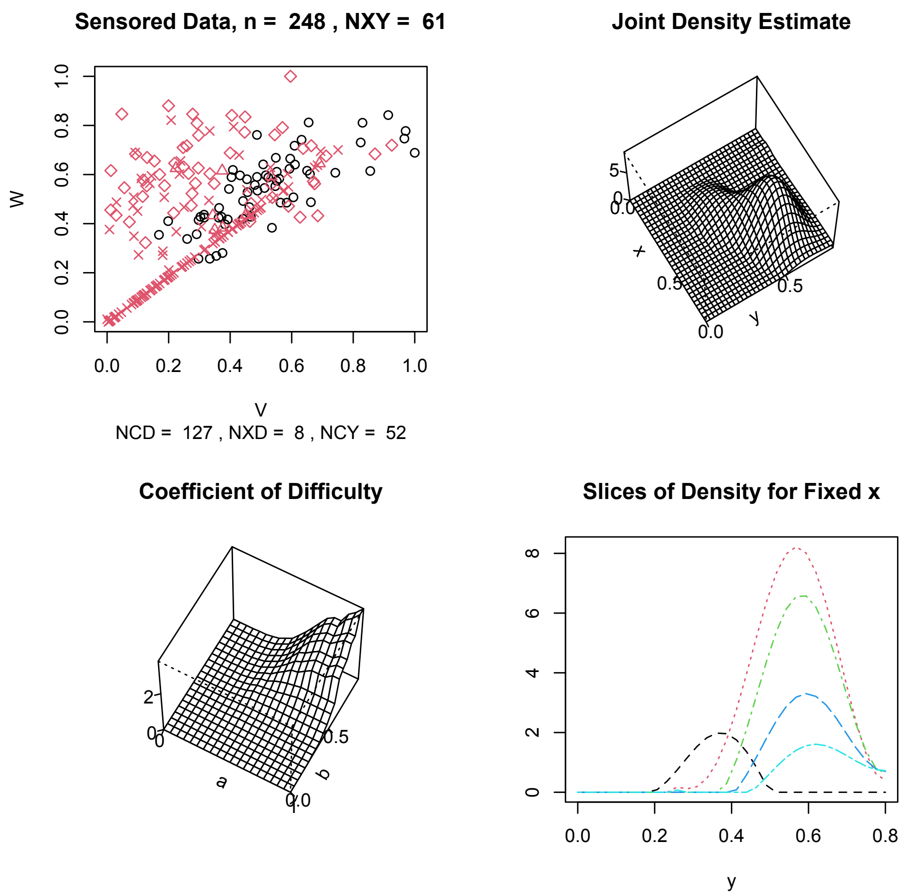

FIG 1. *Bivariate density estimation based on BIFAR aeration diffusers data. Data are rescaled onto the unit square. The left-top diagram exhibits censored data. The circles show uncensored observations* (X,Y) *corresponding to* $\Delta\Gamma = 1$. *The crosses, triangles and rhombuses show pairs* (V,W) *with* (Δ,Γ) *equal to* $(0,0)$, $(1,0)$ *and* $(0,1)$, *respectively. Further, the diagram indicates the total sample size* $n = 248$ *and sizes of the above-mentioned four subsamples as NXY, NCD, NXD and NCY, respectively. In particular, NXY* $:= \sum_{l=1}^{n} \Delta_l \Gamma_l$ *is the number of uncensored pairs while NCD* $:= \sum_{l=1}^{n}(1-\Delta_l)(1-\Gamma_l)$ *is the number of pairs where the both lifetimes of interest are censored. The left-bottom diagram shows an estimated coefficient of difficulty. The proposed bivariate density estimate is shown in the right-top diagram. Slices of that bivariate estimate are shown by the short-dashed, dotted, short-dashed-dotted, long-dashed and long-dashed-dotted lines for* x *equal to 0.25, 0.5, 0.625, 0.7 and 0.75, respectively.*

than a diffuser. The left-bottom diagram shows us an estimated coefficient of difficulty $\hat{d}(a,b)$ defined in (3.8). We see a sharply increasing function in a and b. Accordingly, the bivariate density is estimated on the square $[0, 0.8]^2$, and the estimate is shown in the right-top diagram. The right-bottom diagram sheds light on the bivariate density via its slices for fixed values of x, and explanation of the slices can be found in the figure's caption. As it could be expected, an increase in the lifetime X implies an increase in the lifetime Y but the growth slows down as we may notice from the modes of slices. After this remark we may return to the right-top diagram of the joint density and recognize this interesting feature of the lifetime of diffusers. BIFAR has found these results and the density shape reasonable and insightful.

Now let us present a numerical study that sheds additional light on the problem. Our aim is to understand how the plug-in estimate $\hat{S}$ of the nuisance survival function S performs and how censoring affects estimation of the joint density f. Accordingly, we study the proposed data-driven estimator, the same estimator only based on a known survival function S, and estimator of Efromovich (1999) based on underlying sample from (X,Y). We denote considered models as (k,r), where $k=1,2$ and $r=1,2,3,4$ define two estimated joint densities f and four nuisance joint survival function S, respectively. k denotes density number k shown and defined in Figure 6.3 of Efromovich (1999), the both densities are supported on $[0,1]$ and will be estimated on $[0,.8]^2$. Distributions of censoring (C,D) are denoted as: $r=1$ for uniform distribution on $[0,1]^2$; $r=2$ for independent exponential with rate 1; $r=3$ for independent exponential with rate 0.5; $r=4$ for $S(x,y)=(1+y)^{-1}e^{-(1+y)x}I(x>0,y>0)$. For each model and a sample size $n=100,200,300,400$ we repeat a simulation 1000 times, and for tth simulation calculate integrated squared errors ISE_{1t}, ISE_{2t} and ISE_{3t} of the proposed estimator, the proposed estimator with known survival function, and the estimator based on a sample from (X,Y). Then an entry in Table 1 is written as A/B where $A=(1000)^{-1}\sum_{t=1}^{1000} ISE_{1t}/ISE_{2t}$ and $B=(1000)^{-1}\sum_{t=1}^{1000} ISE_{1t}/ISE_{3t}$. The coefficient of difficulty is also shown.

TABLE 1
Numerical study

Model	n				d
	100	200	300	400	
(1,1)	1.15/5.7	1.11/5.4	1.06/5.3	1.04/5.3	10.3
(1,2)	1.11/4.9	1.05/4.5	1.03/3.8	1.02/3.8	6.0
(1,3)	1.12/3.4	1.08/2.9	1.04/2.3	1.02/2.2	3.6
(1,4)	1.14/5.4	1.09/4.9	1.06/4.4	1.03/4.2	7.1
(2,1)	1.12/2.9	1.09/3.5	1.05/4.1	1.02/4.4	13.1
(2,2)	1.10/1.7	1.07/2.2	1.04/2.7	1.02/2.9	6.4
(2,3)	1.11/1.3	1.08/1.7	1.04/2.1	1.02/2.4	3.7
(2,4)	1.12/1.9	1.07/2.5	1.05/3.0	1.03/3.5	7.8

Let us look at the results. First of all, we note that the use of estimated joint survival function $\hat{S}$ in place of an unknown joint survival function S is a feasible approach in the joint density estimation. The data also highlight the dramatic effect of censoring on density estimation, especially for the first density which, according to Efromovich (1999), is estimated very well for uncensored (X,Y) and small samples. Finally, note that despite the asymptotic-theoretical nature of the coefficient of difficulty d, it sheds light on complexity of estimating a particular density.

6. Proofs. In this section we continue to use notations introduced in the Introduction. In particular, recall notation $q:=q_n$, $s:=s_n$, $\varphi_{ij}(x,y)$ and $\varphi_i(z|c)$. Sequence a_{ij} is defined in (1.4), $a_n:=a_n(Q)$ in (3.7), and b_k, L_k, B_{kr} and t_{kr} are introduced in Section 4. In what follows c and c_k are generic positive constants whose value is not of interest to us, and $o_k(1)\to 0$ as $k\to\infty$.

Proof of Theorem 1. We are considering a sequence in n of classes of additive perturbations of the pivot. Namely, we are considering perturbations on an increasing (as $n\to\infty$) number of subrectangles of the rectangle $R:=[0,a]\times[0,b]$. These subrectangles are created by dividing the rectangle R into s^2 subrectangles of sizes $(a/s)\times(b/s)$. To satisfy the restriction $g(x,y)=0$ for $(x,y)\notin R_n$ in the bottom line of definition (2.2) of $\mathcal{F}_n$, no perturbation is done for boundary subrectangles. Now we introduce several new notations. We begin with a sequence of function classes on $[0,\infty)^2$,

$$(6.1)\qquad \mathcal{H}_s=\Big\{f:\ f=f_0+I_R\sum_{k,r=1}^{s-2}[f_{(kr)}-Af_{(kr)}],\ f_{(kr)}\in\mathcal{H}_{skr},\ f\geq 0\Big\},$$

where $Ah(x,y) := a^{-1}\int_0^a h(u,y)du + b^{-1}\int_0^b h(x,v)dv - (ab)^{-1}\int_R h(u,v)dudv$. The function classes $\mathcal{H}_{skr}$ in (6.1) are defined as follows. Let $\tilde{\phi}(x) := \phi(n,x)$ be a sequence of flattop nonnegative kernels defined on the real line such that for a given n: the kernel is zero beyond $(0,1)$, it is m_X-fold continuously differentiable on $(-\infty,\infty)$, $0 \le \tilde{\phi}(x) \le 1$, $\tilde{\phi}(x) = 1$ for $2/q^2 \le x \le 1 - 2/q^2$, and its lth derivative satisfies $\max_x |\tilde{\phi}^{(l)}(x)| \le Cq^{2l}$, $l = 1,\ldots,m_X$. For instance, such a kernel may be constructed using so-called mollifiers discussed in Efromovich (1999). Then set $\tilde{\phi}_{sk}(x) := \tilde{\phi}(sa^{-1}x - k)$. Absolutely similarly define $\hat{\phi}_{sr}(y)$ only with m_X being replaced by m_Y and a by b. Set $\varphi^*_{ski}(x) := \varphi_i(x - ka/s|a/s)$ and $\varphi'_{srj}(y) := \varphi_j(y - rb/s|b/s)$. For a (k,r)th subrectangle $R_{skr} := [ak/s, a(k+1)/s] \times [br/s, b(r+1)/s]$ (as we will see shortly, common boundaries are irrelevant for the proof and simplify formulae), $1 \le k, r \le s-2$, set $\phi_{skr}(x,y) := \tilde{\phi}_{sk}(x)\hat{\phi}_{sr}(y)$, $\varphi_{skrij}(x,y) := \varphi^*_{ski}(x)\varphi'_{srj}(y)$, and

$$f_{[kr]}(x,y) := \sum_{(i,j)\in T(s,k,r)} \nu_{skrij}\varphi_{skrij}(x,y), \quad f_{(kr)}(x,y) := f_{[kr]}(x,y)\phi_{skr}(x,y). \tag{6.2}$$

Here the set $T(s,k,r) := \{(i,j):\ \min(i,j) > q^s, n^{\beta/(2\beta+1)}s^{-4} \le a_{ij} \le a_n(Q_{skr})\}$, $Q_{skr} := Q(1-1/s)(\overline{I_s^{-1}}I_{skr})^{-1}$, $I_{skr} := S(ak/s, br/s)/f_0(ak/s, br/s)$, $\overline{I_s^{-1}} = \sum_{k,r=1}^{s-2}(1/I_{skr})$. Note via analysis of (6.1) and (6.2) how flattop kernels ϕ_{skr} "sew" together and smooth additive perturbations on R, and that $f = f_0$ on the boundary subrectangles.

Using the above-introduced notations we can define function classes used in (6.1) as

$$\mathcal{H}_{skr} := \Big\{ f_{(kr)} : \sum_{(i,j)\in T(s,k,r)} a_{ij}\nu^2_{skrij} \le Q_{skr}, \tag{6.3}$$

$$s^{-1} < n\nu^2_{skrij} \le s, \ \max_{(x,y)\in R_{skr}} |f_{[kr]}(x,y)|^2 \le s^4 q n^{-2\beta/(2\beta+1)} \Big\}.$$

Here a_{ij} are defined in (1.4), and deterministic ν^2_{skrij} satisfying (6.3) will be defined shortly.

Let us verify that for sufficiently large n we have $\mathcal{H}_s \subset \mathcal{F}_n$. Definition of the flattop kernel implies that for $(x,y) \in R$ the difference $f(x,y) - f_0(x,y)$ is m_X-fold differentiable with respect to x and m_Y-fold differentiable with respect to y. Second, let us verify that for $f \in \mathcal{H}_s$ this difference belongs to $\mathcal{G}(m_X, m_Y, Q, a, b)$. Set $m = m_X$, begin with the differentiation with respect to x, and we will use notation $\psi^{(l)}(x,y) := \partial^l\psi(x,y)/\partial x^l$ for the lth derivative in several following lines. By the Leibniz rule $(f_{[kr]}(x,y)\phi_{skr}(x,y))^{(m)} = \sum_{l=0}^m \mathbf{C}_l^m f_{[kr]}^{(m-l)}(x,y)\phi_{skr}^{(l)}(x,y)$ where $\mathbf{C}_l^m := m!/((m-l)!l!)$. For $0 < l \le m$ we have $(\phi_{skr}^{(l)}(y,x))^2 \le C(s(\ln(n))^2)^{2l}$, and for $f_{(kr)} \in \mathcal{H}_{skr}$

$$\int_R [f_{[kr]}^{(m-l)}(x,y)\phi_{skr}^{(l)}(x,y)]^2 dxdy \le cs^{2l}q^{4l}\int_{R_{skr}} [f_{[kr]}^{(m-l)}(x,y)]^2 dxdy \tag{6.4}$$

$$\le cs^{2l}q^{4l}\sum_{(i,j)\in T(s,k,r)} i^{2(m-l)}\nu^2_{skrij} \le cq^{4m+1}\max_{(i,j)\in T(s,k,r)} \frac{i^{2(m-l)}}{a_{ij}} Q_{srk} = o_n(1)q^{-2}Q_{skr}.$$

In the last equality we used definition of a_{ij} and inequality $i > q^s$ which holds for $(i,j) \in T(s,k,r)$. Differentials with respect to y yield the same result, and using the Parseval identity we get for $f_{(kr)} \in \mathcal{H}_{skr}$

$$\int_R \Big[f^2_{[kr]}(x,y) + (\partial^{m_X} f_{[kr]}(x,y)/\partial x^{m_X})^2 \tag{6.5}$$

$$+(\partial^{m_Y} f_{[kr]}(x,y)/\partial y^{m_Y})^2\Big]\phi^2_{skr}(x,y)dxdy \le \sum_{(i,j)\in T(s,k,r)} a_{ij}\nu^2_{skrij} \le Q_{skr}.$$

This inequality, the fact that the function $\sum_{k,r=1}^{s-1} f_{(kr)}$ and its corresponding derivatives are zero at the boundary of R, Proposition 1 of Efromovich (2001), and $\sum_{k,r=1}^{s-2} Q_{skr} = Q(1-s^{-1})$ yield $\sum_{k,r=1}^{s-2} f_{(kr)} \in \mathcal{G}(m_X, m_Y, Q(1-s^{-1}), a, b)$. The last step in checking the new function class is to verify (recall (6.1)) that

$$g_s := I_R \sum_{k,r=1}^{s-2} Af_{(sk)} \tag{6.6}$$

belongs to $\mathcal{G}(m_X, m_Y, o_n(1)s^{-1}, a, b)$. This result follows from definition (6.2) of $f_{(kr)}(x,y)$, relation $\int_{ak/s}^{a(k+1)/s} f_{[kr]}(x,y)dx = \int_{br/s}^{b(r+1)/s} f_{[kr]}(x,y)dy = 0$, and definition of the flattop kernel $\phi_{skr}(x,y)$. The relation $\mathcal{H}_s \subset \mathcal{F}_n$ is verified for all sufficiently large n. In what follows we use $\mathcal{H}_s$ to establish the lower bound of Theorem 1.

Set $\hat{f} =: f_0 + \tilde{f}$, recall notation (6.6), and write for $f \in \mathcal{H}_s$,

$$\int_{R_{skr}} (\hat{f}(x,y) - f(x,y))^2 dxdy = \int_{R_{skr}} (\tilde{f}(x,y) - f_{(kr)}(x,y) + g_s(x,y))^2 dxdy$$

$$\ge (1-s^{-1}) \int_{R_{skr}} (\tilde{f}(x,y) - f_{[kr]}(x,y))^2 dxdy$$

$$-s \int_{R_{skr}} [f_{[kr]}(x,y)(1-\phi_{skr}(x,y)) + g_s(x,y)]^2 dxdy$$

$$= (1-s^{-1}) \int_{R_{skr}} (\tilde{f}(x,y) - f_{[kr]}(x,y))^2 dxdy + o_n(1) s q^{-1/2} n^{-2\beta/(2\beta+1)}.$$

This and notation $\tilde{\nu}_{skrij} := \int_{R_{skr}} \tilde{f}(x,y)\varphi_{skrij}(x,y)dxdy$ allow us to write,

$$\sup_{f\in\mathcal{F}_n} \mathbb{E}_{f,S}\left\{\int_R (\hat{f}(x,y) - f(x,y))^2 dxdy\right\} \tag{6.7}$$

$$\ge \sup_{f\in\mathcal{H}_s} \mathbb{E}_{f,S}\left\{\int_R (\hat{f}(x,y) - f(x,y))^2 dxdy\right\}$$

$$= \sup_{f\in\mathcal{H}_s} \sum_{k,r=1}^{s-2} \mathbb{E}_{f,S}\left\{\int_{R_{skr}} (\hat{f}(x,y) - f(x,y))^2 dxdy\right\}$$

$$\ge (1-s^{-1}) \sum_{k,r=1}^{s-2} \sup_{f\in\mathcal{H}_{skr}} \sum_{(i,j)\in T(s,k,r)} \mathbb{E}_{f,S}\left\{(\tilde{\nu}_{skrij} - \nu_{skrij})^2\right\} + o_n(1) n^{-2\beta/(2\beta+1)}$$

$$=: (1-s^{-1}) \sum_{k,r=1}^{s-2} A_{kr} + o_n(1) n^{-2\beta/(2\beta+1)}.$$

Now we need to establish a lower bound for a term A_{kr} corresponding to the (k,r)th subrectangle. The rational of converting the original nonparametric MISE into the local mean squared error (MSE) is that: (i) Due to Assumption 1 the pivot f_0 and the joint survival function S are almost "constant" on a subrectangle; (ii) The nonparametric setting is converted into a multivariate parametric setting.

The above-made remark explains our next steps in obtaining a lower minimax bound for the parametric model. Introduce an array of independent normal random variables ζ_{skrij} with zero mean and variance $(1-\gamma_n)\nu^2_{skrij}$ where the positive sequence γ_n tends to zero as slowly as desired. Using these variables we introduce a stochastic process $\bar{f}^*(x,y)$, $(x,y) \in [0,\infty)^2$, defined as the studied $f \in \mathcal{H}_s$ but with random ζ_{skrij} used in place of deterministic ν_{skrij}. The idea of considering such a stochastic process goes back to Pinsker (1980) and specifically for density estimation to Efromovich and Pinsker (1982), and the rest of the proof is based on using steps of those papers. First, we choose deterministic ν^2_{skrij} as explained in Efromovich and Pinsker (1982). Second, following along lines of establishing (A.18) in Pinsker (1980) we get

$$\mathbb{P}\Big((\bar{f}^* - f_0) \in \mathcal{G}(m_Y, m_X, Q, a, b)\Big) \geq 1 - |o_n(1)|. \tag{6.8}$$

Using $\nu^2_{skrij} \leq sn^{-1}$ and that $T(s,k,r)$ has the cardinality of order $n^{1/(2\beta+1)}$ we get

$$\sum_{(i,j)\in T(s,k,r)} \sup_{(x,y)\in R_{skr}} [\nu_{skrij}\varphi_{skrij}(x,y)]^2 \leq cs^3 n^{-2\beta/(2\beta+1)}.$$

Further, introducing a similarly defined stochastic process $\bar{f}^*_{[kr]}$, and using the above-presented calculations together with Theorem 6.2.3 in Kahane (1985) we get

$$\mathbb{P}\Big(\sup_{(x,y)\in R} |\bar{f}^*_{[kr]}(x,y)|^2 \leq s^4 q n^{-2\beta/(2\beta+1)}\Big) \geq 1 - |o_n(1)|s^{-2}.$$

Our next step is to compute local parametric Fisher informations for $f \in \mathcal{H}_s$. We calculate them for the quartet (V, W, Δ, Γ) via considering its additive components corresponding to different values of pair (Δ, Γ). We begin with the case $\Delta\Gamma = 1$ when both X and Y are observed directly. As we will see shortly, this is the main case. Using (1.3) we get

$$p(x,y,1,1) = f(x,y)S(x,y). \tag{6.9}$$

Note that an observation of (X,Y) is biased by $S(X,Y)$. The corresponding to (6.9) component of a local Fisher information that the observation carries about parameter ν_{skrij} is

$$I_{skrij11} := \mathbb{E}_{f_*S}\{\Delta\Gamma[\partial \ln(p(X,Y,1,1))/\partial\nu_{skrij}]^2\} \tag{6.10}$$

$$= \mathbb{E}_{f_*S}\{\Delta\Gamma[\partial \ln(f(X,Y))/\partial\nu_{skrij}]^2\}.$$

Here f_* is an underlying f where we set $\nu_{skrij} = 0$. Let us consider the derivative on the right side of (6.10),

$$\frac{\partial \ln(f(x,y))}{\partial \nu_{skrij}} = \frac{\partial \ln\Big(f_0(x,y) + I_R \sum_{k',r'=1}^{s-2}[f_{(k'r')}(x,y) - Af_{(k'r')}]\Big)}{\partial\nu_{skrij}}. \tag{6.11}$$

Now recall that

$$f_{(kr)}(x,y) := f_{[kr]}(x,y)\phi_{skr}(x,y)I((x,y)\in R_{skr}), \tag{6.12}$$

$\partial f_{[kr]}(x,y)/\partial\nu_{skrij} = \varphi_{skrij}(x,y)$, $\phi_{skr}(x,y)$ is a flattop kernel described in the beginning of the proof, $\varphi_{skrij}(x,y)$ are elements of the tensor-product cosine basis on the (k,r)th subrectangle and accordingly $\int_{ak/s}^{a(k+1)/s} \varphi_{skrij}(x,y)dx = 0$ and $\int_{br/s}^{b(r+1)/s} \varphi_{skrij}(x,y)dy = 0$.

Also recall that on the subrectangle functions f_0 and S are continuous and $|f - f_0| \leq 1/q$. Using these remarks and a straightforward calculation we continue (6.10) and get,

$$I_{skrij11} := \frac{S(ak/s, br/s)}{f_0(ak/s, br/s)}(1 + o_n(1)) = I_{skr}(1 + o_n(1)). \tag{6.13}$$

Here $o_n(1) \to 0$ as $n \to \infty$ uniformly over considered (s, k, r, i, j).

Now consider the case when $\Delta = 1$ and $\Gamma = 0$. Using(1.3) we can write,

$$p(x, y, 1, 0) = f^{X,Y>y}(x) f^{C>x,D}(y). \tag{6.14}$$

Here we use new notations $f^{X,Y>y}(x) := \int_y^\infty f(x, z)dz$, $f^{C>x,D}(y) := \int_x^\infty f^{C,D}(z, y)dz$ and $f^{C,D}(x, y)$ is the joint density of (C, D). Also recall that f_* is an underlying f with $\nu_{skrij} = 0$. Then the corresponding component of Fisher information is

$$I_{skrij10} := \mathbb{E}_{f_*S}\{\Delta(1-\Gamma)[\partial \ln(p(X, Y, 1, 0))/\partial\nu_{skrij}]^2\} \tag{6.15}$$

$$= \mathbb{E}_{f_*S}\Big\{\Delta(1-\Gamma)\frac{[\partial f^{X,Y>Y}(X, Y)/\partial\nu_{skrij}]^2}{[f^{X,Y>Y}(X, Y)]^2}\Big\}.$$

To evaluate the derivative on the right side of (6.15), we begin with analysis of $f^{X,Y>y}(x, y)$. Using notation (6.6) and that $f_{(kr)}$ is supported on R_{skr} we may write,

$$f^{X,Y>y}(x) = f_0^{X,Y>y}(x) + [\sum_{k,r=1}^{s-2} \int_y^b f_{(kr)}(x, z)dz - \int_y^b g_s(x, z)dz] I(0 \leq y \leq b). \tag{6.16}$$

This expression, together with $f_{(kr)}(x, y) = f_{[kr]}(x, y)\phi_{skr}(x, y)$, $\partial f_{[k,r]}(x, y)/\partial\nu_{skrij} = \varphi_{skrij}(x, y)$ and recalling that $\min(i, j) > q^s$, allow us to write via integration by parts for $y \in [0, b]$,

$$\frac{\partial \sum_{k,r=1}^{s-2} \int_y^b f_{(kr)}(x, z)dz}{\partial\nu_{skrij}} = \int_y^b \varphi_{skrij}(x, z)\phi_{skr}(x, z)dz \tag{6.17}$$

$$= \sqrt{\frac{2s}{b}}\,\frac{b}{\pi js}\,\varphi^*_{ski}(x)\sin(\pi jb^{-1}(sz - br))\phi_{skr}(x, z)\Big|_{z=y}^{z=b} - \sqrt{\frac{2s}{b}}\,\frac{b}{\pi js}$$

$$\times \int_y^b \varphi^*_{ski}(x)\sin(\pi jb^{-1}(sz - br))[\partial\phi_{skr}(x, z)/dz]dz = O_n(1)q^{-s} I(y \leq b(1 - s^{-1})).$$

In a similar manner we establish that $|\frac{\partial \sum_{k,r=1}^{s-2} \int_y^b g_s(x,z)dz}{\partial\nu_{skrij}}| = O_n(1)q^{-s} I(y \leq b(1 - s^{-1}))$.

Now we need to bound from below the denominator inside the expectation on the right side of (6.15). The following remarks are due. First, we consider $r < s - 1$ and thus exclude all boundary subrectangles $R_{sk(s-1)}$ where we have $f = f_* = f_0$. As a result, we need to consider only $y \in [0, b(1 - s^{-1})]$. Second, recall that $\min_{(x,y)\in R} f_0(x, y) =: c_* > 0$ due to Assumption 1. Using these remarks and (2.2) we can write for $(x, y) \in [0, a] \times [0, b(1 - s^{-1})]$,

$$\int_y^\infty f(x, z)dz \geq \int_{b(1-s^{-1})}^b f_0(x, z)dz \geq c_* bs^{-1}.$$

Using the obtained relations in (6.15) we conclude that uniformly over considered (s, k, r, i, j)

$$I_{skrij10} := o_n(1). \tag{6.18}$$

Absolutely similarly we establish that $I_{skrij01}$ and $I_{skrij00}$ vanish as $n \to \infty$. Combining the obtained results we conclude that a local Fisher information contained in the quartet (V, W, Δ, Γ) about parameter ν_{skrij} is

$$I_{skrij} := \mathbb{E}_{f_* S}\{[\partial \ln(p(V, W, \Delta, \Gamma))/\partial \nu_{skrij}]^2\} \tag{6.19}$$

$$= I_{skrij11} + I_{skrij10} + I_{skrij01} + I_{skrij00} = I_{skr}(1 + o_n(1)).$$

Here $o_n(1) \to 0$ as $n \to \infty$ uniformly over considered (s, k, r, i, j).

Remark 2. The above-presented calculation of Fisher information indicates that estimation of parameters ν_{skrij} based on observations from (V, W, Δ, Γ) with $\Delta\Gamma = 0$, that is when at least one of X or Y is censored, yields an ill-posed problem with a slower rate of convergence. As we will see shortly, this conclusion also verifies the assertion of Theorem 1 about sufficiency of using uncensored observations.

Now we need to establish several technical results. Recall notations (1.1), (1.4), (1.7), (1.8) and (3.7). Consider an equation $\sum_{\{i,j:\, 0<a_{ij}\le a_n^*\}}[(a_{ij}a_n^*)^{1/2} - a_{ij}] = Qd^{-1}n$. The sum can be approximated for large n by the integral

$$M_n := \int_{\{(x,y):\, 1+(\pi x/a)^{2m_X}+(\pi y/b)^{2m_Y}\le a_n^*;\, x,y>0\}} \Big([(\pi x/a)^{2m_X} + (\pi y/b)^{2m_Y}]^{1/2} \tag{6.20}$$

$$\times [a_n^*]^{1/2} - [(\pi x/a)^{2m_X} + (\pi y/b)^{2m_Y}]\Big) dxdy.$$

Using the change of variables $v = \pi x a^{-1}[a_n^*]^{-1/(2m_X)}$ and $u = \pi y b^{-1}[a_n^*]^{-1/(2m_Y)}$ we continue,

$$M_n = [ab/\pi^2][a_n^*]^{\frac{1}{2m_X}+\frac{1}{2m_Y}+1}$$

$$\times \int_{\{(v,u):\, v^{2m_X}+u^{2m_Y}\le 1-1/a_n^*;\, v,u>0\}} ([v^{2m_X} + u^{2m_Y}]^{1/2} - [v^{2m_X} + u^{2m_Y}]) dvdu.$$

These calculations yield $a_n^* = a_n(1 + o_n(1))$ where a_n is defined in (3.7).

A similar calculation, when we approximate a sum by an appropriate integral, yields

$$\sum_{\{i,j:\, 0<a_{ij}\le a_n^*\}} [1 - (a_{ij}/a_n^*)^{1/2}] \tag{6.21}$$

$$= P(m_X, m_Y, 1, a, b) Q^{1/(2\beta+1)}(n/d)^{1/(2\beta+1)}(1 + o_n(1)).$$

These are technical results that allow us to proceed along lines of the proof of Theorem 1 in Efromovich (1989) and evaluate from below terms A_{kr} introduced in (6.7). Namely, we get that uniformly over $k, r \in \{1, \ldots, s-2\}$

$$\inf A_{kr} \ge (s^{-4\beta} Q_{skr})^{1/(2\beta+1)}(nI_{skr})^{-2\beta/(2\beta+1)} P(m_X, m_Y, 1, a, b)(1 + o_n(1)). \tag{6.22}$$

Here the infimum is over all possible nonparametric oracle-estimates of f considered in Theorem 1. Now we plug in values of Q_{skr} introduced at the beginning of the proof and get

$$\inf \sum_{k,r=1}^{s-2} A_{kr} \ge P(m_X, m_Y, 1, a, b) Q^{1/(2\beta+1)} n^{-2\beta/(2\beta+1)} s^{-4\beta/(2\beta+1)} \tag{6.23}$$

$$\times\Big[\sum_{k,r=1}^{s-2}(\overline{I_s^{-1}}I_{skr})^{-1/(2\beta+1)}I_{skr}^{-2\beta/(2\beta+1)}\Big](1+o_n(1)).$$

For the sum on the right side of (6.23) we may write,

$$\sum_{k,r=1}^{s-2}(\overline{I_s^{-1}}I_{skr})^{-1/(2\beta+1)}I_{skr}^{-2\beta/(2\beta+1)}=(\overline{I_s^{-1}})^{-1/(2\beta+1)}\sum_{k,r=0}^{s-2}I_{skr}^{-1} \tag{6.24}$$

$$=(\overline{I_s^{-1}})^{2\beta/(2\beta+1)}=\Big[\sum_{k,r=1}^{s-2}\frac{f_0(ka/s,br/s)}{S(ak/s,br/s)}\Big]^{2\beta/(2\beta+1)}.$$

Recall that the pivotal joint density and the joint survival function of the censoring variables are continuous on R, and write

$$s^{-4\beta/(2\beta+1)}\Big[\sum_{k,r=1}^{s-2}\frac{f_0(ka/s,rb/s)}{S(ak/s,br/s)}\Big]^{2\beta/(2\beta+1)}$$

$$=\Big[s^{-2}\sum_{k,r=1}^{s-2}\frac{f_0(ak/s,rb/s)}{S(ak/s,br/sr)}\Big]^{2\beta/(2\beta+1)}=\Big[\frac{1}{ab}\int_R\frac{f_0(x,y)}{S(x,y)}dxdy\Big]^{2\beta/(2\beta+1)}(1+o_n(1)).$$

Using these calculations, together with (1.6), in (6.23) yield

$$\inf\sum_{k,r=1}^{s-2}A_{kr}\ge P\Big[n^{-1}\,\frac{1}{ab}\int_R\frac{f_0(x,y)}{S(x,y)}dxdy\Big]^{2\beta/(2\beta+1)}(1+o_n(1)).$$

This lower bound, together with (1.6), (2.2) and (6.7), verify Theorem 1.

In what follows we may write $\mathbb{E}\{\cdot\}:=\mathbb{E}_{fS}\{\cdot\}$ whenever no confusion occurs. Also recall that i and j are nonnegative integers.

Proof of Theorem 2. We begin with the following proposition which is of interest on its own whenever the bivariate survival function S of censoring lifetimes is known.

Lemma 2. *Let S be known and the assumption of Theorem 2 holds. Consider an oracle-estimator*

$$\bar f_*(x,y):=\sum_{i,j\ge 0}\lambda_{ij}^*\hat\theta_{ij}^*\varphi_{ij}(x,y), \tag{6.25}$$

where λ_{ij}^* is defined as in (3.11) only with $\hat a_n$ being replaced by a_n defined in (3.7), and $\hat\theta_{ij}^*$ is defined in (3.4). Then the MISE of $\bar f_*$ satisfies (3.12) and the oracle-estimator is sharp minimax.

Proof of Lemma 2. Using the Parseval identity we can write

$$\mathbb{E}\Big\{\int_R(\bar f_*(x,y)-f(x,y))^2dxdy\Big\}=\sum_{i,j\le q}\mathbb{E}\{(\hat\theta_{ij}^*-\theta_{ij})^2\} \tag{6.26}$$

$$+\sum_{\{i,j:\max(i,j)>q,a_{ij}\le a_n\}}\mathbb{E}\{[(1-[\frac{a_{ij}}{a_n}]^{1/2})\hat\theta_{ij}^*-\theta_{ij}]^2\}+\sum_{a_{ij}>a_n}\theta_{ij}^2=\sum_{i,j\le q}\mathbb{E}\{(\hat\theta_{ij}^*-\theta_{ij})^2\}$$

$$+\sum_{\{i,j:\,\max(i,j)>q,a_{ij}\le a_n\}}\mathbb{E}\{[(1-[\frac{a_{ij}}{a_n}]^{1/2})(\hat\theta_{ij}^*-\theta_{ij})-[\frac{a_{ij}}{a_n}]^{1/2}\theta_{ij}]^2\}+\sum_{a_{ij}>a_n}\theta_{ij}^2.$$

Now we need to study moments of Fourier estimator $\hat{\theta}^*_{ij}$. Write

$$\mathbb{E}\{\hat{\theta}^*_{ij}\} = \mathbb{E}\Big\{\frac{\Delta\Gamma\varphi_{ij}(V,W)I((V,W)\in R)}{S(V,W)}\Big\} \tag{6.27}$$

$$= \int_R f(x,y)S(x,y)\varphi_{ij}(x,y)[S(x,y)]^{-1}dxdy = \int_R f(x,y)\varphi_{ij}(x,y)dxdy = \theta_{ij}.$$

We conclude that the Fourier estimator is unbiased. For the MSE we write,

$$n\mathbb{E}\{(\hat{\theta}^*_{ij}-\theta_{ij})^2\} = \mathbb{E}\Big\{\Big[\frac{\Delta\Gamma\varphi_{ij}(V,W)I((V,W)\in R)}{S(V,W)}\Big]^2\Big\} - \theta^2_{ij} \tag{6.28}$$

$$= \int_R \frac{f(x,y)S(x,y)[\varphi_{ij}(x,y)]^2}{[S(x,y)]^2}dxdy - \theta^2_{ij} = \int_R \frac{f(x,y)[\varphi_{ij}(x,y)]^2}{S(x,y)}dxdy - \theta^2_{ij}.$$

Introduce notation $i\wedge j := \min(i,j)$. Our next step is to show that (6.28) implies

$$\sup_{f\in\mathcal{F}} |n\mathbb{E}\{(\hat{\theta}^*_{ij}-\theta_{ij})^2\} - d| = o_{i\wedge j}(1). \tag{6.29}$$

The verification is based on using several classical trigonometric formulas. First, recall that

$$\varphi_{ij}(x,y) = a^{-1/2}[I(i=0) + 2^{1/2}\cos(\pi ix/a)I(i>0)]$$

$$\times b^{-1/2}[I(j=0) + 2^{1/2}\cos(\pi jy/b)I(j>0)],$$

and then using $\cos^2(x) = [1+\cos(2x)]/2$ we get

$$\varphi^2_{ij}(x,y) = (ab)^{-1} \tag{6.30}$$

$$+(2ab)^{-1/2}[\varphi_{2i,0}(x,y)I(i>0) + \varphi_{0,2j}(x,y)I(j>0) + 2^{-1/2}\varphi_{2i,2j}(x,y)I(ij>0)].$$

Second, set $\kappa_{i,j} := \int_R [f(x,y)/S(x,y)]\varphi_{ij}(x,y)dxdy$, and then (6.30) implies

$$\int_R \frac{f(x,y)[\varphi_{ij}(x,y)]^2}{S(x,y)}dxdy \tag{6.31}$$

$$= d + (2ab)^{-1/2}[\kappa_{2i,0}I(i>0) + \kappa_{0,2j}I(j>0) + 2^{-1/2}\kappa_{2i,2j}I(ij>0)].$$

Let us show that $\sup_{f\in\mathcal{F}_n}|\kappa_{ij}| = o_{i\wedge j}(1)$, and then this, (6.31) and $\sup_{f\in\mathcal{F}_n}\theta^2_{ij} = o_{i\wedge j}(1)$ will imply the verified (6.29). Recall that $f(x,y) = \sum_{k,r=0}^{\infty}\theta_{kr}\varphi_{kr}(x,y)$, $(x,y)\in R$, c denotes generic positive constants, and set $\nu_{i,j} := \int_R [S(x,y)]^{-1}\varphi_{ij}(x,y)dxdy$. Using Cauchy-Schwarz inequality and $\cos(\alpha)\cos(\beta) = [\cos(\alpha+\beta)+\cos(\alpha-\beta)]/2$ we may write,

$$|\kappa_{ij}| = |\sum_{k,r=0}^{\infty}\theta_{kr}\int_R [S(x,y)]^{-1}\varphi_{kr}(x,y)\varphi_{ij}(x,y)dxdy| \tag{6.32}$$

$$\leq c\sum_{k,r=0}^{\infty}|\theta_{kr}|[|\nu_{i-k,j-r}| + |\nu_{i-k,j+r}| + |\nu_{i+k,j-r}| + |\nu_{i+k,j+r}|]$$

$$\leq c[\sum_{\{k>i/2\}\cup\{r>j/2\}}\theta^2_{kr}]^{1/2} + c\Big[\sum_{0\leq k\leq i/2, 0\leq r\leq j/2}[\nu^2_{i-k,j-r} + \nu^2_{i-k,j+r}$$

$$+\nu^2_{i+k,j-r}+\nu^2_{i+k,j+r}]\Big]^{1/2}\le c[\sum_{\{k>i/2\}\cup\{r>j/2\}}\theta^2_{kr}]^{1/2}+c[\sum_{k\ge i/2,r\ge j/2}\nu^2_{k,r}]^{1/2}.$$

The first term on the right side of (6.32) is $o_{i\wedge j}(1)$ uniformly over $f\in\mathcal{F}_n$, and the second term is also $o_{i\wedge j}(1)$ because S^{-1} is square-integrable on R. Relation (6.29) is verified.

Now recall that $q:=q_n$ is of order $\ln(n)$, and using (6.27) and (6.29) we continue (6.26),

$$\mathbb{E}\Big\{\int_R(\bar f_*(x,y)-f(x,y))^2dxdy\Big\}\le\sum_{\{i,j:\min(i,j)\le q,a_{ij}\le a_n\}}\mathbb{E}\{(\hat\theta^*_{ij}-\theta_{ij})^2\} \tag{6.33}$$

$$+\sum_{\{i,j:\,\min(i,j)>q,\,a_{ij}\le a_n\}}(1-(a_{ij}/a_n)^{1/2})^2n^{-1}[d+o_{i,j}(1)]+a_n^{-1}\sum_{\max(i,j)>q}a_{ij}\theta^2_{ij}$$

$$\le o_n(1)n^{-2\beta/(2\beta+1)}+\sum_{a_{ij}\le a_n}(1-(a_{ij}/a_n)^{1/2})^2n^{-1}[d+o_n(1)]+a_n^{-1}\sum_{\max(i,j)>q}^{\infty}a_{ij}\theta^2_{ij}.$$

Let us evaluate the second sum on the right side of (6.33). According to (2.2) and the assumption about the pivot f_0, an underlying function f has Fourier coefficients $\theta_{ij}=\theta_{0ij}+\nu_{ij}$ where

$$\sum_{i,j\ge0}[1+(\pi i/a)^{2m_X+1}+(\pi j/b)^{2m_Y+1}]\theta^2_{0ij}\le Q',\quad\sum_{i,j\ge0}a_{ij}\nu^2_{ij}\le Q. \tag{6.34}$$

Here, recalling notation $f=f_0+gI_R$ introduced in (2.2), θ_{0ij} are Fourier coefficients of the pivot f_0 and ν_{ij} are Fourier coefficients of a perturbation g. For a pair (i,j) such that $\max(i,j)>q$ we may write for some absolute positive constant $c_*:=c_*(a,b,m_X,m_Y)$,

$$[1+(\pi i/a)^{2m_X+1}+(\pi j/b)^{2m_Y+1}]\ge c_*q[1+(\pi i/a)^{2m_X}+(\pi j/b)^{2m_Y}]=c_*qa_{ij}.$$

Using this relation and the Cauchy inequality we get

$$\sum_{\max(i,j)>q}a_{ij}\theta^2_{ij}\le\sum_{\max(i,j)>q}a_{ij}[(1+q^{1/2})\theta^2_{0ij}+(1+q^{-1/2})\nu^2_{ij}]$$

$$\le(c_*q)^{-1}(1+q^{1/2})\sum_{\max(i,j)>q}[1+(\pi i/a)^{2m_X+1}+(\pi j/b)^{2m_Y+1}]\theta^2_{0ij}$$

$$+(1+q^{-1/2})\sum_{\max(i,j)>q}a^2_{ij}\nu^2_{ij}.$$

The last inequality and (6.34) yield

$$\sum_{\max(i,j)>q}a_{ij}\theta^2_{ij}\le Q(1+o_n(1)). \tag{6.35}$$

Using (6.35) in (6.33) we conclude that

$$\mathbb{E}\{\int_R(\bar f_*(x,y)-f(x,y))^2dxdy\} \tag{6.36}$$

$$\le[\sum_{a_{ij}\le a_n}(1-(a_{ij}/a_n)^{1/2})^2n^{-1}d+a_n^{-1}Q](1+o_n(1)).$$

The final step is a straightforward calculation, based on approximation of the sum by a double integral, and it is performed similarly to calculations (6.20)–(6.21). Lemma 2 is proved.

Now we are considering cases of unknown S and d in turn. Suppose that Lemma 1 is valid and also

$$\Big|\mathbb{E}\Big\{\prod_{t=1}^{2}[\hat{S}(x_t,y_t)-S(x_t,y_t)]|Z_1=z_1,Z_2=z_2\Big\}\Big|\le c^*n^{-1}. \tag{6.37}$$

In (6.37) the notation $Z_t := (V_t, W_t, \Delta_t, \Gamma_t)$ and $z_t := (x_t, y_t, \delta_t, \gamma_t)$ of Lemma 1 is used. Formula

$$\frac{1}{v}=\frac{1}{w}+\frac{w-v}{w^2}+\frac{(w-v)^2}{w^2v} \quad \text{where} \quad wv\neq 0, \tag{6.38}$$

and notation $I_l := I((V_l, W_l) \in R)$ allow us to write for Fourier estimate (3.6),

$$\hat{\theta}_{ij}=n^{-1}\sum_{l=1}^{n}\frac{\Delta_l\Gamma_l\varphi_{ij}(V_l,W_l)I_l}{\hat{S}(V_l,W_l)}=n^{-1}\sum_{l=1}^{n}\frac{\Delta_l\Gamma_l\varphi_{ij}(V_l,W_l)I_l}{S(V_l,W_l)} \tag{6.39}$$

$$+n^{-1}\sum_{l=1}^{n}\frac{\Delta_l\Gamma_l\varphi_{ij}(V_l,W_l)I_l[S(V_l,W_l)-\hat{S}(V_l,W_l)]}{[S(V_l,W_l)]^2}$$

$$+n^{-1}\sum_{l=1}^{n}\frac{\Delta_l\Gamma_l\varphi_{ij}(V_l,W_l)I_l[S(V_l,W_l)-\hat{S}(V_l,W_l)]^2}{[S(V_l,W_l)]^2\hat{S}(V_l,W_l)}=:\hat{\theta}^*_{ij}+A_1+A_2.$$

Here $\hat{\theta}^*_{ij}$ is the oracle's Fourier estimator (3.4) that has been already studied in (6.27) and (6.29). Let us evaluate second moments of A_1 and A_2 in turn. Write,

$$n^2\mathbb{E}\{A_1^2\}=\mathbb{E}\Big\{\Big[\sum_{l=1}^{n}\frac{\Delta_l\Gamma_l\varphi_{ij}(V_l,W_l)I_l[S(V_l,W_l)-\hat{S}(V_l,W_l)]}{[S(V_l,W_l)]^2}\Big]^2\Big\} \tag{6.40}$$

$$=n\mathbb{E}\Big\{\Big[\frac{\Delta_1\Gamma_1\varphi_{ij}(V_1,W_1)I_1[S(V_1,W_1)-\hat{S}(V_1,W_1)]}{[S(V_1,W_1)]^2}\Big]^2\Big\}$$

$$+n(n-1)\mathbb{E}\Big\{\prod_{l=1}^{2}\frac{\Delta_l\Gamma_l\varphi_{ij}(V_l,W_l)I_l}{[S(V_l,W_l)]^2}[S(V_l,W_l)-\hat{S}(V_l,W_l)]\Big\}.$$

Set $S_* := \mathbb{P}(V > a, W > b)$. Recall that $S_* > 0$ due to Assumption 1, and c^* are generic positive constants that may depend on S_*. The first expectation on the right side of (6.40) is bounded by c^*n^{-1} due to (3.14). To evaluate the second expectation, note that $\varphi_{ij}(x_1, y_1)\varphi_{ij}(x_2, y_2)$ are elements of a cosine tensor-product basis on R^2. These remarks and (6.37) imply that $\mathbb{E}\{A_1^2\} \le c^*n^{-1}[o_n(1) + o_{i\wedge j}(1)]$. Now we evaluate A_2. Let us make a remark that simplifies the proof and formulas. Exponential inequality (3.15) yields

$$\mathbb{P}\Big(\bigcup_{l=1}^{n}\{\hat{S}(V_l,W_l)<S(a,b)/2,\ (V_l,W_l)\in R\}\Big)\le c^*n^3e^{-n/c^*}. \tag{6.41}$$

The exponential inequality (6.41), together with $\hat{S}(x, y) \ge n^{-1}$, imply that in the proofs we can restrict our attention to the case

$$\hat{S}(V_l,W_l)>S(a,b)/2 \ \text{ whenever } \ (V_l,W_l)\in R. \tag{6.42}$$

This and (3.14) yield $\mathbb{E}\{A_2^2\} \le c^* n^{-2}$. Combining the obtained results in (6.39) we get

$$\sup_{f\in\mathcal{F}_n} |n\mathbb{E}\{(\hat\theta_{ij}-\theta_{ij})^2\} - d| = o_n(1) + o_{i\wedge j}(1). \quad (6.43)$$

Further, Lemma 1, (6.41) and (6.42) allow us to conclude that $\mathbb{E}\{(\hat d - d)^2\} \le c^* n^{-1}$. This and a straightforward algebra, identical to Efromovich and Pinsker (1982), verify Theorem 2 if Lemma 1 and (6.37) are valid.

Proof of Lemma 1 and (6.37). Recall notation $S(x,y) := \mathbb{P}(C > x, D > y)$ and the assumed $S^{V,W}(a,b) = S^{X,Y}(a,b)S(a,b) =: S_* > 0$. In what follows we are considering $(x,y) \in R$ because we need to estimate $S(V_l,W_l)$ only for $(V_l,W_l) \in R$. Further, note that the probability in (3.15) is zero whenever $\epsilon > 1 + n^{-1}$, and accordingly we will consider only $0 < \epsilon \le 1 + n^{-1}$.

We begin with explanation of the underlying idea of the estimate $\hat S$. Introduce a bivariate cumulative hazard

$$H(x,y) := -\ln(S(x,y)). \quad (6.44)$$

Using notation $f^{C>x,D}(t_2) := d\mathbb{P}(C > x, D \le t_2)/dt_2$ and a line integration we write,

$$H(x,y) = \int_0^x \Big[\frac{\partial}{\partial t_1} H(t_1,0)\Big] dt_1 + \int_0^y \Big[\frac{\partial}{\partial t_2} H(x,t_2)\Big] dt_2 \quad (6.45)$$

$$= \int_0^x \frac{f^C(t_1)}{S^C(t_1)} dt_1 + \int_0^y \frac{f^{C>x,D}(t_2)}{S(x,t_2)} dt_2 =: H_1(x) + H_2(x,y).$$

The two integrals can be estimated by method of moments. We begin with $H_1(x)$. Using $S^V(x) = S^C(x)S^X(x)$ we write,

$$f^{V,\Delta}(x,0) = f^C(x)S^X(x) = \frac{f^C(x)S^V(x)}{S^C(x)}. \quad (6.46)$$

This formula implies that the integrand in $H_1(x)$ may be written as

$$\frac{f^C(x)}{S^C(x)} = \frac{f^{V,\Delta}(x,0)}{S^V(x)}. \quad (6.47)$$

In its turn, the last relation yields for $H_1(x)$,

$$H_1(x) = \int_0^x \frac{f^C(t)}{S^C(t)} dt = \int_0^x \frac{f^{V,\Delta}(t,0)}{S^V(t)} dt = \mathbb{E}\Big\{\frac{(1-\Delta)I(V \le x)}{S^V(V)}\Big\}. \quad (6.48)$$

We do not know the univariate survival function $S^V(x)$, but because V is directly observed we estimate it by an empirical survival function

$$\hat S^V(x) := n^{-1}\sum_{l=1}^n I(V_l \ge x). \quad (6.49)$$

To analyze $\hat S^V$, let us mention several classical results that may be found in Efromovich (2018). First, the Hoeffding inequality states that if $\eta_1, \eta_2, \dots, \eta_m$ are independent mean zero random variables with bounded ranges, that is, $\mathbb{P}(\eta_i \in [a_i,b_i]) = 1$, $-\infty < a_i < b_i < \infty$, $i = 1,2,\dots,m$, then for any $\epsilon > 0$

$$\mathbb{P}\Big(\sum_{i=1}^m \eta_i \ge \epsilon\Big) \le e^{-2\epsilon^2/\sum_{i=1}^m (b_i-a_i)^2}. \quad (6.50)$$

Second, for independent and mean zero random variables $\zeta_1, \ldots, \zeta_m$

$$\mathbb{E}\{|m^{-1}\sum_{i=1}^{m}\zeta_i|^p\} \le c'_p m^{-p/2-1}\sum_{i=1}^{m}\mathbb{E}\{|\zeta_i|^p\},\ \ p \ge 2, \tag{6.51}$$

where c'_p is a finite absolute constant depending only on p. Third,

$$|\sum_{i=1}^{n}\xi_i|^p \le n^{p-1}\sum_{i=1}^{n}|\xi_i|^p,\ p \ge 1. \tag{6.52}$$

Now we are ready to formulate properties of $\hat{S}^V(V_1)$. First of all, $\hat{S}^V(V_1)$ is not smaller than n^{-1}, and accordingly may be used in a denominator. Second, using (6.51) implies

$$\mathbb{E}\{\hat{S}^V(V_1)|V_1\} = S^V(V_1)(1-n^{-1})+n^{-1},\ \ \mathbb{E}\{(\hat{S}^V(V_1)-S^V(V_1))^{2k}|V_1\} \le c_k n^{-k}, \tag{6.53}$$

Using (6.50) we get for any $\epsilon > 2n^{-1}$

$$\mathbb{P}(|\hat{S}^V(V_1) - S^V(V_1)| > \epsilon|V_1) \le 2e^{-n\epsilon^2/2}. \tag{6.54}$$

These results and (6.48) yield feasibility of the following sample mean estimate of H_1,

$$\hat{H}_1(x) := n^{-1}\sum_{l=1}^{n}\frac{(1-\Delta_l)I(V_l \le x)}{\hat{S}^V(V_l)} = \sum_{l=1}^{n}\frac{(1-\Delta_l)I(V_l \le x)}{\sum_{r=1}^{n} I(V_r \ge V_l)}. \tag{6.55}$$

Now we are considering estimation of $H_2(x,y)$ defined in (6.45). Write,

$$\mathbb{P}(V > x, W \le y, \Gamma = 0) = \int_0^y f^{V>x,W,\Gamma}(t,0)dt \tag{6.56}$$

$$= \int_0^y S^{X,Y}(x,t)f^{C>x,D}(t)dt = \int_0^y \frac{S^{V,W}(x,t)f^{C>x,D}(t)}{S(x,t)}dt.$$

Relation (6.56) implies that

$$f^{V>x,W,\Gamma}(y,0) = \frac{S^{V,W}(x,y)f^{C>x,D}(y)}{S(x,y)}. \tag{6.57}$$

Using (6.57) and definition (6.45) of the integral H_2 yield

$$H_2(x,y) = \int_0^y \frac{f^{V>x,W,\Gamma}(t,0)}{S^{V,W}(x,t)}dt = \mathbb{E}\Big\{\frac{(1-\Gamma)I(V > x)I(W \le y)}{S^{V,W}(x,W)}\Big\}. \tag{6.58}$$

To use (6.58) for sample mean estimation we need to propose an estimate of $S^{V,W}$. Set

$$\hat{S}^{V,W}(x,y) := n^{-1}\Big[1 + \sum_{r=1}^{n} I(V_r \ge x)I(W_r \ge y)\Big]. \tag{6.59}$$

Note that $\hat{S}^{V,W}(V_1,W_2) \ge n^{-1}$, and hence it may be used in a denominator. Further, similarly to (6.53) and (6.54) we have for $(t,r) \in \{1,2\}^2$ that almost sure,

$$|\mathbb{E}\{(\hat{S}^{V,W}(V_t,W_r) - S^{V,W}(V_t,W_r))|V_t,W_t,V_r,W_r\}| \le 5n^{-1}, \tag{6.60}$$

$$\mathbb{E}\{(\hat{S}^{V,W}(V_t,W_r) - S^{V,W}(V_t,W_r))^{2k}|V_t,W_t,V_r,W_r\} \le c_k n^{-k}, \tag{6.61}$$

and for $\epsilon > 10n^{-1}$

$$\mathbb{P}(|\hat{S}^{V,W}(V_t,W_r) - S^{V,W}(V_t,W_r)| > \epsilon|V_t,W_t,V_r,W_r) \le 2e^{-n\epsilon^2/2}. \tag{6.62}$$

Recall our notation $S_* := S^{V,W}(a,b)$ and that $S_* > 0$ according to Assumption 1. There is a useful corollary from (6.54) and (6.62) which states that for $n > 20/S_*$ we have

$$\sum_{t,r=1}^{n} \mathbb{P}(\{\hat{S}^{V,W}(V_t,W_r) < S_*/2\} \cap \{(V_t,W_r) \in R\}) \tag{6.63}$$

$$+\sum_{t=1}^{n} \mathbb{P}(\{\hat{S}^{V}(V_t) < S_*/2\} \cap \{V_t \le a\}) \le 2n(n+1)e^{-2nS_*^2}.$$

Due to (6.63) and because the survival density estimates are at least n^{-1}, in the following proof we may assume that

$$\min\left(\hat{S}^V(V_t), \hat{S}^{V,W}(V_t,W_r)\right) > S_*/2 \text{ for } (V_t,W_r) \in R,\ (t,r) \in \{1,\dots,n\}^2. \tag{6.64}$$

Recall (6.58) and (6.59), and we are in a position to introduce a sample mean estimate of H_2,

$$\hat{H}_2(x,y) := n^{-1} \sum_{l=1}^{n} \frac{(1-\Gamma_l)I(V_l > x)I(W_l \le y)}{\hat{S}^{V,W}(x,W_l)} \tag{6.65}$$

$$= \sum_{l=1}^{n} \frac{(1-\Gamma_l)I(V_l > x)I(W_l \le y)}{1+\sum_{r=1}^{n} I(V_r \ge x)I(W_r \ge W_l)},\ \ (x,y) \in R.$$

This estimate, together with (6.55) and the above-presented properties of $\hat{S}^V$ and $\hat{S}^{V,W}$, shed light on the estimate $\hat{S}$ defined in (3.5) and whose properties we are verifying. Using a Taylor formula we may write for $(x,y) \in R$,

$$\hat{S}(x,y) - S(x,y) = e^{-\hat{H}_1(x)-\hat{H}_2(x,y)} - e^{-H_1(x)-H_2(x,y)} \tag{6.66}$$

$$= M(x,y)S(x,y) + (1/2)M^2(x,y)S(x,y) + \rho(x,y).$$

Here $|\rho(x,y)| \le |M(x,y)|^3$ and

$$M(x,y) := M_1(x) + M_2(x,y) := [H_1(x) - \hat{H}_1(x)] + [H_2(x,y) - \hat{H}_2(x,y)]. \tag{6.67}$$

We are evaluating M_1 and M_2 in turn. Using (6.48), (6.55) and (6.38) we can write,

$$M_1(x) = \mathbb{E}\{\frac{(1-\Delta)I(V \le x)}{S^V(V)}\} - n^{-1}\sum_{l=1}^{n} \frac{(1-\Delta_l)I(V_l \le x)}{\hat{S}^V(V_l)} \tag{6.68}$$

$$= \Big[\mathbb{E}\{\frac{(1-\Delta)I(V \le x)}{S^V(V)}\} - n^{-1}\sum_{l=1}^{n} \frac{(1-\Delta_l)I(V_l \le x)}{S^V(V_l)}\Big]$$

$$-n^{-1}\sum_{l=1}^{n} \frac{(1-\Delta_l)I(V_l \le x)[S^V(V_l) - n^{-1}\sum_{r=1}^{n} I(V_r \ge V_l)]}{[S^V(V_l]^2}$$

$$-n^{-1}\sum_{l=1}^{n} \frac{(1-\Delta_l)I(V_l \le x)[S^V(V_l) - n^{-1}\sum_{r=1}^{n} I(V_r \ge V_l)]^2}{[S^V(V_l]^2\hat{S}^V(V_l)}$$

$$=: M_{11}(x) - M_{12}(x) - M_{13}(x).$$

Evaluating of M_{11} is straightforward,

$$|\mathbb{E}\{M_{11}(x_1)|Z_1=z_1\}| = |\mathbb{E}\{\frac{1-\Delta)I(V\le x_1)}{S^V(V)}\} \tag{6.69}$$

$$-n^{-1}(n-1)\mathbb{E}\{\frac{1-\Delta)I(V\le x_1)}{S^V(V)}\} - n^{-1}\frac{(1-\delta_1)I(V_1\le x_1)}{S^V(x_1)}\}| \le n^{-1}/S_*.$$

Similarly we conclude that $|\mathbb{E}\{M_{12}(x_1)|Z_1=z_1\}| \le cn^{-1}/S_*^2$, and using (6.64) we get $|\mathbb{E}\{M_{13}(x_1)|Z_1=z_1\}| \le cn^{-1}/S_*^3$. Combining the results yields (compare with the verified (3.13))

$$|\mathbb{E}\{M_1(x_1)|Z_1=z_1\}| \le c^* n^{-1}. \tag{6.70}$$

Further, using (6.53) and (6.64) we conclude (compare with the verifies (3.14)) that

$$\mathbb{E}\{[M_1(x_1)]^{2k}|Z_1=z_1\} \le c_k^* n^{-1}. \tag{6.71}$$

Using (6.50) to evaluate M_{11} and (6.54) to evaluate M_{12} and M_{13} we get

$$\mathbb{P}(|M_1(x_1)|>\epsilon|Z_1=z) \le c^* n e^{-n\epsilon^2/c^*}. \tag{6.72}$$

Now consider the term $M_2(x,y)$ defined in (6.67). Using (6.38), (6.45) and (6.65) we may write,

$$M_2(x,y) = \mathbb{E}\{\frac{(1-\Gamma)I(V>x)I(W\le y)}{S^{V,W}(x,W)}\} - n^{-1}\sum_{l=1}^n \frac{(1-\Gamma_l)I(V_l>x)I(W_l\le y)}{\hat{S}^{V,W}(x,W_l)}$$

$$= [\mathbb{E}\{\frac{(1-\Gamma)I(V>x)I(W\le y)}{S^{V,W}(x,W)}\} - n^{-1}\sum_{l=1}^n \frac{(1-\Gamma_l)I(V_l>x)I(W_l\le y)}{S^{V,W}(x,W_l)}]$$

$$-n^{-1}\sum_{l=1}^n \frac{(1-\Gamma_l)I(V_l>x)I(W_l\le y)[S^{V,W}(x,W_l)-\hat{S}^{V,W}(x,W_l)]}{[S^{V,W}(x,W_l)]^2}$$

$$-n^{-1}\sum_{l=1}^n \frac{(1-\Gamma_l)I(V_l>x)I(W_l\le y)[S^{V,W}(x,W_l)-\hat{S}^{V,W}(x,W_l)]^2}{[S^{V,W}(x,W_l)]^2\hat{S}^{V,W}(x,W_l)}$$

$$=: M_{21}(x,y) - M_{22}(x,y) - M_{23}(x,y). \tag{6.73}$$

We begin our analysis with conditional expectation of M_2. For M_{21} we have

$$|\mathbb{E}\{M_{21}(x_1,y_1)|Z_1=z_1\}| \tag{6.74}$$

$$= |n^{-1}\mathbb{E}\{\frac{(1-\Gamma)I(V>x_1)I(W\le y_1)}{S^{V,W}(x_1,W)}\} - \frac{(1-\gamma_1)I(V_1>x_1)I(W_1\le y_1)}{S^{V,W}(x_1,W_1)}| \le \frac{1}{nS_*}.$$

Now compare this inequality with (6.69) and realize the similarity between analysis of M_1 and M_2. Then following the analysis of M_1 we conclude that

$$\mathbb{E}\{|M_2(x_1,y_1)|^{2k}|Z_1=z_1\} \le c^* n^{-1}, \quad \mathbb{E}\{|M_2(x_1,y_1)|^{2k}|Z_1=z_1\} \le c_k^* n^{-k}, \tag{6.75}$$

and due to (6.62) we get a rough but sufficient for our purpose inequality

$$\mathbb{P}(|M_2(x_1,y_1)|>\epsilon|Z_1=z_1) \le c^* n^2 e^{-n\epsilon^2/c^*}. \tag{6.76}$$

Combining the above-obtained results in (6.66), together with a simple calculation, verifies Lemma 1. To verify (6.37) we are using the Cauchy-Schwarz inequality and write,

$$\mathbb{E}\Big\{[\hat{S}(x_1,y_1)-S(x_1,y_1)][\hat{S}(x_2,y_2)-S(x_2,y_2)]|Z_1=z_1,Z_2=z_2\Big\}$$

$$\leq \prod_{t=1}^{2}\Big[\mathbb{E}\Big\{[\hat{S}(x_t,y_t)-S(x_t,y_t)]^2|Z_1=z_1,Z_2=z_2\Big\}\Big]^{1/2}\leq c^* n^{-1}.$$

The last inequality is established identically to (3.14) with $k=1$. What was wished to prove.

The assertion of Theorem 3 follows from Efromovich and Pinsker (1982). Proof of Theorem 4 follows along lines of Efromovich (1985) with the use of Lemma 6.2.

Proof of Remark 1 in Section 4. Write for MSE of a low-frequency thresholded Fourier estimate,

$$\mathbb{E}\{(I(\hat{\theta}_{ij}^2>2q\hat{d}n^{-1})\hat{\theta}_{ij}-\theta_{ij})^2\}\leq \mathbb{E}\{(\hat{\theta}_{ij}-\theta_{ij})^2\}+\theta_{ij}^2\mathbb{E}\{I(\hat{\theta}_{ij}^2\leq 2q\hat{d}n^{-1})\} \tag{6.77}$$

$$\leq cn^{-1}+\theta_{ij}^2\mathbb{E}\{I(\theta_{ij}^2<8q\hat{d}n^{-1})\}+\theta_{ij}^2\mathbb{E}\{I(|\theta_{ij}-\hat{\theta}_{ij}|>|\theta_{ij}|/2\}$$

$$\leq Cn^{-1}+8qn^{-1}\mathbb{E}\{\hat{d}\}+\theta_{ij}^2 4\mathbb{E}\{(\hat{\theta}_{ij}-\theta_{ij})^2\}/\theta_{ij}^2\leq cqn^{-1}.$$

Here we used the already proved inequalities $\mathbb{E}\{(\hat{\theta}_{ij}-\theta_{ij})^2\}\leq cn^{-1}$ and $\mathbb{E}\{\hat{d}\}<c$. Further, recall that only $(1+q)^2$ low-frequency Fourier coefficient estimates $\hat{\theta}_{ij}$ are hard-thresholded. This and (6.77) prove the remark. Note that (6.77) points upon a large choice of feasible low-frequency estimates.

Acknowledgements. The research is supported in part by NSF Grant DMS-1915845 and Grants from CAS and BIFAR. Valuable comments of reviewers, the associate editor and the editor are greatly appreciated.

REFERENCES

[1] Akritas, M. and Van Kellogom, I. (2003). Estimation of Bivariate and Marginal Distribution with Censored Data. *Journal of Royal Statistical Society Ser. B* **65** 457-471.

[2] Albu, N., Barani, N. and Constantin, M. (2021). Choosing an Economical Solution for Water Aeration. *Hydraulica* **145** 32-37.

[3] Campbell, G. (1981). Nonparametric Bivariate Estimation with Randomly Censored Data. *Biometrika* 68 417-422.

[4] Collett, D. (2003). *Modelling Survival Data in Medical Research.* Chapman and Hall, Boca Raton.

[5] Crowder, M. (2012). *Multivariate Survival Analysis and Competing Risks*. CRC Press, Boca Raton.

[6] Dabrowska, D. (1988). Kaplan-Meier estimate on the plane *The Annals of Statistics* **16** 1475-1489.

[7] Dabrowska, D., Duffy, D. and Zhang, Z. (1998). Hazard and Density Estimation from Bivariate Censored Data, *Journal of Nonparametric Statistics* **10** 67-93.

[8] Efromovich, S. and Pinsker, M. (1982). Estimation of a square integrable probability density of a random variable. *Problems of Information Transmission* **18** 19-38.

[9] Efromovich, S. (1985). Adaptive estimation of a density with unknown smoothness. *Theory of Probability and its Applications* **30** 557-568.

[10] Efromovich, S. (1989). On sequential nonparametric estimation of a density. *Theory Probab. Applications* **34** 228–239.

[11] Efromovich, S. (1999). *Nonparametric Curve Estimation: Methods, Theory and Applications.* Springer, New York.

[12] Efromovich, S. (2001). Density estimation under random censorship and order restrictions: from asymptotic to small samples. *JASA* **96** 667-684.

[13] Efromovich, S. (2018). *Missing and Modified Data in Nonparametric Estimation.* Chapman & Hall, Boca Raton.

[14] Frees, E., Carriere, J. and Valdez, E. (1995). Annuity Valuation with Dependent Mortality. *Actuarial Research Clearing House* **2** 31-80.

[15] Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference.* Cambridge Univ. Press, Cambridge.

[16] Golubev, G.K. (1991). LAN in problems of non-parametric estimation of functions and lower bounds for quadratic risks. *Problems of Information Transmission* **36** 152-157.

[17] Hoffmann, M. and Lepski, O. (2002). Random rates in anisotropic regression. *Annals of Statistics* **30** 325-396.

[18] Hougaard, P. (2000). Analysis of Multivariate Survival Data, Springer: New York.

[19] Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *JASA* **53** 457-481.

[20] Kooperberg, C. (1998). Bivariate Density Estimation With an Application to Survival Analysis. *Journal of Computational and Graphical Statistics*, 7, 322-341.

[21] Li, J. and Ma, S. (2013). *Survival Analysis in Medicine and Genetics*. CRC Press, Boca Raton.

[22] Lopez, O. (2012). A generalization of Kaplan-Meier estimator for analyzing bivariate mortality under right-censoring and left-truncation with applications to model-checking for survival copula models. *Insurance: Mathematics and Economics* **51** 505-516.

[23] Moore, D. (2016). *Applied Survival Analysis Using R*. Springer, New York.

[24] Nikolskii, S. M. (1975). *Approximation of Functions of Several Variables and Imbedding Theorems*. Springer, Berlin.

[25] Pinsker, M. (1980). Optimal filtration of functions from L_2 in Gaussian noise. *Problems of Information Transmission* **16** 970-983.

[26] Prentice, R. and Cai, J. (1992). Covariance and survival function estimation using censored multivariate failure time data. *Biometrika* 79 495-512.

[27] Prentice R. (2016). Higher dimensional Clayton–Oakes models for multivariate failure time data. *Biometrika* **103** 231-236

[28] Prentice, R. and Zhao, S. (2018). Nonparametric estimation of the multivariate survivor function: the multivariate Kaplan–Meier estimator, *Lifetime Data Analysis* **24** 3-27.

[29] Pruitt, R. (1993). Identifiability of bivariate survival curves from censored data. *JASA* **88** 573-579.

[30] Rosso, D., Larson, L. and Stenstrom, M. (2008). Aeration of large-scale municipal wastewater treatment plants: state of the art. *Water Science and Technology* **57** 973-978.

[31] Tsai, W., Leurgans, S. and Crowley, J. (1986). Nonparametric Estimation of a Bivariate Survival Function in the Presence of Censoring. *Ann. Statist.* **14** 1351-1365.

[32] Saraiva, E., Suzuki, A., and Milan, L. (2018). Bayesian computational methods for sampling from the posterior distribution of a bivariate survival model, based on AMH copula in the presence of right-censored data. *Entropy*, **20** 642-654.

[33] Seok, J., Tian, L. and Wong, W. (2014). Density estimation on multivariate censored data with optional Polya tree. *Biostatistics*, 15, 182-195.

[34] Oakes, D. (1989). Bivariate survival models induced by frailties. *JASA* **84** 487-493.

[35] Wang, W. and Wells, M. (1997). Nonparametric estimators of the bivariate survival function under simplified censoring conditions *Biometrika* **84** 863-880.

[36] Wasserman, L. (2006). *All About Nonparametric Statistics*. Springer, New York.

[37] Wells, M. and Yeo, K. (1996). Density estimation with bivariate censored data. *JASA* **91** 1566-1574.