Efficient nonparametric estimation of distribution for current status censoring

Sam Efromovich[1]

[1]*Department of Mathematical Sciences, University of UTDallas, Richardson, USA. e-mail:*
efrom@utdallas.edu

Abstract: Current status censoring (CSC) implies that there is no direct access to the lifetime of an event of interest. Instead it is known if the event already occurred or not at a random monitoring time. CSC is a simple sampling procedure and in many cases the only possibility to assess the lifetime of interest. At the same time, the absence of a direct measurement of a lifetime of interest makes the problem of nonparametric distribution estimation ill-posed. A simple, adaptive and sharp minimax estimator of the density and cumulative distribution function is proposed. The simplicity of estimator also allows us to relax assumptions. Practical examples illustrate CSC problem and the proposed estimator.

Keywords and phrases: Adaptation, cumulative distribution function, density, robustness, sharp minimax, survival anaysis.

1. Introduction

Considered problem is a nonparametric data-driven and sharp minimax estimation of the probability density and the cumulative distribution function (cdf) of the lifetime of interest X (nonnegative random variable) which is not observed directly. Instead, there is a possibility to check status of the event at some random moment of time Z, called the *monitoring* time. Then the available current status censoring (CSC) observation is a pair of random variables (Z, Δ) where Z is the monitoring time and $\Delta := I(X \leqslant Z)$ is the status of the event of interest, namely the status (indicator) is equal to 1 if the event of interest already occurred at moment Z and the status is 0 otherwise. Available data is a sample of size n from (Z, Δ).

Current status censoring (CSC), also known as "case I" interval censoring, is a classical problem in survival analysis, see a discussion in books [9,15,21,23,25, 32,36] and thorough reviews of CSC in papers [7,10,16,24,27,28,34] where further references may be found. It is well known that the stated nonparametric problem of density and cumulative distribution function estimation is ill-posed with slower rates of risk convergence than for the case of direct observations. Due to the slower rates of convergence, it is always important to study not only rates but sharp (minimal) constants. Recently [16] established sharp minimax lower bounds for the Mean Integrated Squared Error (MISE) convergence for an oracle that knows CSC data, information about distribution of X, and density of monitoring time Z. The aim of this paper is to propose density and cdf

estimators that match performance of the oracle and attain the sharp constant and the rate of the MISE convergence.

The oracle lower bounds and proposed data-driven estimators matching the oracle will be presented shortly, and now let us review the related nonparametric literature. The literature is vast and primarily devoted to estimating the cdf. Let us begin with [33] where it is established that, under a mild assumption on differentiability of the cdf, it may be estimated pointwise with rate $n^{-1/3}$ by a nonparametric maximum likelihood estimate (NPMLE). Minimal assumptions under which the rate is optimal can be found in [18]. Note that the rate is dramatically slower than $n^{-1/2}$ for the case of direct observations, and more discussion and a thorough review of previous results can be found in these papers. Another seminal paper, devoted to estimation under minimal assumptions, is [5] where a piecewise constant (histogram-type) estimator of cdf is proposed. [37] studied locally linear smoothers. Spline methods were introduced in [26]. A novel kernel method was proposed in [20]. In [8] a warped adaptation for a kernel estimator was motivated by Goldenshluger- Lepskii procedure which yielded the squared-bias and variance trade-off. Log-concave constraint was proposed in [2]. Bootstrapped confidence bands are developed in [22]. A number of interesting and thought-provoking papers are devoted to orthogonal series estimation. In [7] a rigorous analysis of a so-called quotient estimator is performed. The underlying idea is to write the cumulative distribution function of interest as a ratio of two densities of directly observed random variables. Then each density is estimated via a series projection estimator with cutoffs chosen via minimization of a penalized contrast function. It is shown that the adaptive estimator attains optimal nonparametric rates whenever smoothnesses of the two densities are the same. A regression-type estimator was also explored. Further development and literature review on series estimation can be found in [6] where both compact and non compactly supported bases are considered. [29] used penalization for a projection series estimator of a conditional cumulative distribution function.

Density estimation problem for CSC data is dramatically less explored. Let us mention [3] where the authors consider a kernel density estimation via solving iterative equations, nonparametric maximum likelihood estimation, and also via a local EM approach. The latter requires an explicit solution of the local likelihood equations which is done via the symbolic Newton-Raphson algorithm. Convergence of proposed algorithms is studied. In [4] data sharpening is proposed to increase robustness of a kernel density estimator to bandwidth misspecification and measurement errors. [19] used a maximum smoothed likelihood approach and a smoothing the (discrete) MLE of the distribution function approach. In particular, under assumption that the cumulative distribution function is three-fold differentiable, the density is estimated with the rate $n^{-4/7}$. Asymptotic distribution of the estimate was further explored in [20]. In [34] a kernel estimator is proposed under an assumption that the density of interest is twice differentiable and the density of monitoring time is three-fold differentiable. A smart procedure of data transformation is used to convert the problem into deconvolution. Then the optimal rate $n^{-4/7}$ is achieved, and expansions of the expectation and variance as well as asymptotic normality are derived. This pa-

per also contains a nice literature review.

It is important to stress that current status sampling may be dramatically simpler than direct sampling of a lifetime of interest, and in many cases it is the only available option, see a discussion in [9,25,32]. At the same time, CSC makes cdf and density estimation problems ill-posed. As a result, it is prudent to develop a data-driven estimator that attains both the optimal rate and the constant of a risk convergence, and does that without a requirement for the density of monitoring time Z to have a smoothness matching smoothness of an underlying cdf of the lifetime of interest X.

The paper proposes a simple adaptive and sharp-minimax estimator of the density and cdf of a lifetime of interest. The simplicity allows us to prove robustness of the estimator with respect to smoothness of the density of monitoring time. In [16], where a sharp minimax lower bound for CSC is obtained, it is assumed that the density of Z is as smooth as an estimated cdf of the lifetime of interest, and this is a serious restriction because we never know how smooth an underlying cdf of interest is. In this paper only a bounded derivative of the density is assumed, and this makes the proposed estimator robust.

The context of the paper is as follows. Section 2 presents assumptions and a known sharp lower bound. A proposed data-driven estimator for the cdf may be found in Section 3. Density estimation is considered in Section 4. Practical examples and a numerical study are presented in Section 5. Proofs are in Section 6. Conclusion and open problems are in Section 7.

Finally let us introduce several notations used in the paper. Sample size is denoted as n, $o_n(1)$ is a traditional notation for vanishing sequences in n,

$$s := s_n := 3 + \lceil \ln(n+3) \rceil \tag{1.1}$$

and $\lceil a \rceil$ denotes the smallest integer larger or equal to a, $I(A)$ is the indicator of event A. The cosine basis on $[0,1]$ is

$$\varphi_0(x) := 1, \quad \varphi_j(x) := 2^{1/2} \cos(\pi j x), \quad j = 1, 2, \ldots \tag{1.2}$$

We also use notation $g^{(\alpha)}(x)$ or $(g(x))^{(\alpha)}$ for the αth derivative of $g(x)$, while $g^k(x)$ is a notation for $[g(x)]^k$. In the paper specific sequences in n and constants are primarily chosen to make an orthogonal series estimator feasible for small samples. The interested reader is referred to book [13] where series estimation for small samples is discussed in depth.

2. Lower Bound

This section is based on [16] and serves as a reference. Considered nonparametric CSC model is as follows. There is an underlying and hidden sample $X_1, \ldots, X_n$ from a lifetime of interest X. There is also a sample $Z_1, \ldots, Z_n$ from a monitoring variable Z. The available CSC sample is $(Z_1, \Delta_1), \ldots, (Z_n, \Delta_n)$ where $\Delta_l := I(X_l \leqslant Z_l)$ is the status. The aim is to estimate the cumulative distribution function $F^X(x) := \mathbb{P}(X \leqslant x)$ and the density $f^X(x) := dF^X(x)/dx$ of the

lifetime of interest X under the mean integrated squared error (MISE) criterion. We restrict our attention to bounded variables, and without loss of generality may assume that they are supported on $[0, 1]$.

Let us formulate the main assumption.

Assumption 1. *The lifetime of interest X and the monitoring time Z are independent continuous random variables, $\mathbb{P}(X \in [0, 1]) = 1$, a known density $f^Z(z)$ of the monitoring time has a bounded derivative on $[0, 1]$, $\int_0^1 f^Z(z)dz = 1$, and $\min_{z \in [0,1]} f^Z(z) \geqslant c_* > 0$.*

Let us comment on the assumption. Given $\mathbb{P}(Z \leqslant 1) = 1$, no consistent estimation of the distribution of X is possible if $\mathbb{P}(X > 1) > 0$. In other words, for consistent estimation the support of X must be the subset of the support of Z. Similarly, the assumed independence between X and Z is necessary for consistent estimation, see a discussion in [16]. In the presented below lower bound continuity of $f^Z(z)$ is sufficient, but in upper bounds we will use the differentiability. More comments can be found in Section 7.

Assumption 1 implies that the joint (mixed) density of (Z, Δ) is

$$f^{Z,\Delta}(z, \delta) = f^Z(z)[F^X(z)]^\delta[1 - F^X(z))]^{1-\delta}, \quad \delta \in \{0, 1\}. \tag{2.1}$$

The formula sheds additional light on why for consistent estimation the support of X must be a subset of the support of Z.

Now we introduce two function classes. The former is a classical global Sobolev class (ellipsoid) of α-fold differentiable functions on $[0, 1]$

$$\mathcal{F}(\alpha, Q) := \Big\{ g : \; g^{(\alpha)}(x) \text{ exists and finite on [0,1] and} \tag{2.2}$$

$$\sum_{j=1}^n (\pi j)^{2\alpha} \kappa_j^2 \leqslant Q < \infty, \quad \kappa_j = \int_0^1 g(x)\varphi_j(x)dx \Big\}. \tag{2.3}$$

Note that in (2.3) κ_j are Fourier coefficients of the function g. Sobolev function classes are traditionally considered in upper bounds, and this is what will be done shortly.

The latter function class is a local one where considered functions are close to a pivot in L_∞-norm. Let $F_0(x)$ be the cdf of a random variable (lifetime) supported on $[0, 1]$. Introduce a class of cumulative distribution functions supported on $[0, 1]$ and created by additive perturbations of F_0,

$$\mathcal{F}(F_0, \alpha, Q, c_0, c_1, \rho) := \Big\{ F : \; F(x) = F_0(x) + g(x)I(0 \leqslant x \leqslant 1), \; F^{(1)}(x) \geqslant 0,$$

$$g(0) = g(1) = 0, \; g \in \mathcal{F}(\alpha, Q), \; \max_{x \in [0,1]} \max(|g(x)|, |g^{(1)}(x)|) \leqslant \rho,$$

$$\min_{x \in [0,1]} F_0^{(1)}(x) \geqslant c_0, \quad \max_{x \in [0,1]} F_0^{(1)}(x) \leqslant c_1 \Big\}. \tag{2.4}$$

The local functional class (2.4) was introduced in [16], function $F_0(x)$ is called the pivot, and parameter ρ defines the local nature of the class. Local

function classes and a corresponding local minimax approach for establishing lower bounds were pioneered in [17] for the case of direct observations, and for CSC data in [18] for finding an optimal rate for cdf estimation.

Now we are in a position to present the lower bound of [16]. We are using notation $(F^X(x))^{(\beta)}$ with $\beta = 0$ and $\beta = 1$ corresponding to the cdf and the density, and in what follows the parameter β is used solely for this purpose.

Introduce a nonparametric Fisher information

$$\mathcal{J}(\alpha, Q, d, \beta)$$

$$:= \left[(Q(2\alpha + 1))^{-\frac{2\beta+1}{2(\alpha-\beta)}} (2\beta + 1)^{\frac{2\alpha+1}{2(\alpha-\beta)}} \pi(\alpha + 1 + \beta)(\alpha - \beta)^{-1} \right]^{\alpha/(\alpha-\beta)} / d, \quad (2.5)$$

where

$$d := \int_0^1 \frac{F^X(x)(1 - F^X(x))}{f^Z(x)} dx \qquad (2.6)$$

is the so-called coefficient of difficulty of CSC. Note that the nonparametric Fisher information (2.5) is a ratio. The numerator is defined by an underlying class of estimated functions (by parameters α and Q) and by the estimand specified by parameter β, and the denominator (coefficient of difficulty) captures the effect of an underlying distribution of interest (here cdf F^X) and a nuisance function (here density f^Z of the monitoring time). One may think that a nonparametric Fisher information is an analog of a classical parametric Fisher information $1/\sigma^2$ in a problem of estimating the mean of a normal variable with variance σ^2, see [1]. Further, the coefficient of difficulty d is a nonparametric analog of σ^2 because, as we will see shortly, d is a factor in the asymptotic variance of an efficient Fourier estimator implying sharp minimax nonparametric estimation. The larger the coefficient of difficulty, the more complex an estimated nonparametric problem is, and this explains the "coefficient of difficulty" terminology introduced in [13].

Theorem 2.1 (Lower Bound). *Let Assumption 1 hold and $\alpha \geqslant 2$. Then the following lower bound holds for estimation of $(F^X(x))^{(\beta)}$ based on an CSC sample of size n,*

$$\inf_{\widetilde{\Psi}_\beta} \sup_{F^X} \mathbb{E}_{F^X} \left\{ [n\mathcal{J}(\alpha, Q, d, \beta)]^{2(\alpha-\beta)/(2\alpha+1)} \int_0^1 [\widetilde{\Psi}_\beta(x) - (F^X(x))^{(\beta)}]^2 dx \right\}$$

$$\geqslant (1 + o_n(1)). \qquad (2.7)$$

Here parameter β is either 0 or 1 for the estimand being either the cdf or the density, respectively, the supremum is over $F^X \in \mathcal{F}(F_0, \alpha, Q, 1/\ln(\ln(n)),$ $[\ln(\ln(n))]^{1/2}, 1/\ln(\ln(n)))$ defined in (2.4), the infimum is taken over all possible oracle-estimators $\widetilde{\Psi}_\beta$ knowing the sample, density $f^Z(x)$ of the monitoring time Z and everything about the class (2.4), namely F_0, α and Q.

In Theorem 2.1 parameter $\rho = 1/\ln(\ln(n)) = o_n(1)$, and accordingly (2.7) is the lower bound for a shrinking local minimax which makes the bound more challenging to match for an adaptive estimator and a global minimax.

It follows from Theorem 2.1 that the classical rate n^{-1} for estimating cdf, known for direct observations, slows down to $n^{-2\alpha/(2\alpha+1)}$. This is why a CSC sampling is called ill-posed with respect to direct sampling a lifetime of interest. Further, nonparametric Fisher information (2.5) and coefficient of difficulty d are important outcomes of the oracle's lower bound. In particular, the coefficient of difficulty captures the effect of the underlying distribution of interest and the distribution of the monitoring time on the Fisher information and the MISE convergence.

Let us comment on what is known about sharpness of the lower bound (2.7). It is a lower bound for an oracle-estimator that knows data, smoothness of an underlying cdf, and the nuisance density f^Z of the monitoring time. The lower bound is sharp and it is attained by an oracle-estimator for a local Sobolev class. The estimator is too complicated to present here, it is motivated by a proof of the lower bound and uses an aggregation of bases for subsets of the support of X. In next two sections a simple data-driven and robust estimator is proposed that attains the sharp lower bound for a global Sobolev class (2.2) with unknown parameters α and Q. The simplicity will also allow us to relax assumptions of [16] about smoothness of the nuisance density f^Z of the monitoring time. In other words, we will see shortly that it is possible to match performance of the minimax oracle.

3. CDF Estimator

The aim of this section is to present a simple sharp-minimax cdf estimator that adapts to smoothness of an underlying cdf and may be used for small samples. Further, the estimator preserves its properties under a mild assumption of differentiability of the density f^Z of the monitoring time Z.

In what follows we first present the estimator and its properties, and then explain the estimator and discuss the results.

Recall that we observe a sample $(Z_1, \Delta_1), \ldots, (Z_n, \Delta_n)$ from a CSC pair (Z, Δ) where Z is the monitoring time, $\Delta := I(X \leqslant Z)$ is the status, and X is the unobserved lifetime of interest. The aim is to estimate an underlying cdf $F^X(x) := \mathbb{P}(X \leqslant x)$ of X.

We begin with the case of a controlled CSC when density $f^Z(z)$ of the monitoring time is known. Recall that sequence $s := s_n$ and elements of cosine basis $\varphi_j(x)$, $j = 0, 1, \ldots$ are defined in (1.1) and (1.2), respectively. Set

$$\theta_j := \int_0^1 F^X(x)\varphi_j(x), \quad j = 0, 1, \ldots \tag{3.1}$$

for Fourier coefficient of the cdf of interest $F^X(x)$. Then (2.1) implies that a pilot Fourier estimator

$$\check{\theta}_j := n^{-1} \sum_{l=1}^n \frac{\Delta_l \varphi_j(Z_l)}{f^Z(Z_l)} \tag{3.2}$$

is unbiased estimator of θ_j, and accordingly set

$$\check{F}_{-j}(x) := \sum_{i\in\{\{0,1,\dots,s\}\backslash\{j\}\}} \check{\theta}_i\varphi_i(x)I(j\leqslant s) + \sum_{i=0}^{s}\check{\theta}_i\varphi_i(x)I(j>s) \qquad (3.3)$$

for unbiased estimate of function

$$F_{-j}(x) := \sum_{i\in\{\{0,1,\dots,s\}\backslash\{j\}\}} \theta_i\varphi_i(x)I(j\leqslant s) + \sum_{i=1}^{s}\theta_i\varphi_i(x)I(j>s). \qquad (3.4)$$

This function approximates $F^X(x)$ as both s and j increase to infinity, and note that $\int_0^1 F_{-j}(x)\varphi_j(x)dx = \int_0^1 \check{F}_{-j}(x)\varphi_j(x)dx = 0$. In what follows $\check{F}_{-j}(x)$ is referred to as the pilot cdf estimator, and note that this estimator and $F_{-j}(x)$ do not depend on j whenever $j > s$.

Now we can define a Fourier estimator

$$\hat{\theta}_j := n^{-1}\sum_{l=1}^{n}\frac{(\Delta_l - \check{F}_{-j}(Z_l))\varphi_j(Z_l)}{f^Z(Z_l)}. \qquad (3.5)$$

Note how simple the estimator is, hardly more complicated than the pilot (3.2).

The proposed blockwise-shrinking cdf estimator is based in Fourier estimates (3.5) and the following three statistics. Introduce blocks of positive integers/frequencies $B_0 := \{0, 1, \dots, s\}$, $B_k := \{0, 1, \dots, s\lfloor 1 + 1/\ln(s)\rfloor^k\}\backslash \cup_{r=0}^{k-1}B_r$, $k = 1, 2, \dots$, and denote the length (cardinality) of B_k by L_k. Set

$$\widetilde{\Theta}_k := \frac{2}{L_k n(n-1)}$$

$$\times \sum_{1\leqslant l<l'\leqslant n}\sum_{j\in B_k}\frac{(\Delta_l - \check{F}_{-j}(Z_l))\varphi_j(Z_l)(\Delta_{l'} - \check{F}_{-j}(Z_{l'}))\varphi_j(Z_{l'})}{f^Z(Z_l)f^Z(Z_{l'})}, \qquad (3.6)$$

$$\widehat{\Theta}_k := L_k^{-1}\sum_{j\in B_k}\hat{\theta}_j^2, \qquad (3.7)$$

and

$$\tilde{d} := \min\left(3s, n^{-1}\sum_{l=1}^{n}\frac{\check{F}_{-s}(Z_l)(1 - \check{F}_{-s}(Z_l))}{[f^Z(Z_l)]^2}\right). \qquad (3.8)$$

Note that $\tilde{d}$ is a bounded plug-in sample mean estimate of the coefficient of difficulty (2.6).

The cdf estimator, based on an underlying density $f^Z(z)$, is defined as

$$\widetilde{F}^X(x, f^Z) := \sum_{j=0}^{s}\hat{\theta}_j I(\hat{\theta}_j^2 > 2\tilde{d}sn^{-1})\varphi_j(x)$$

$$+ \sum_{k=1}^{r_n}\min(1, \frac{\widetilde{\Theta}_k}{\widehat{\Theta}_k})I(\widetilde{\Theta}_k > 10s^{-1}n^{-1})\sum_{j\in B_k}\hat{\theta}_j\varphi_j(x). \qquad (3.9)$$

Here r_n is the largest integer r such that $s\lfloor 1 + 1/\ln(s)\rfloor^r \leqslant \ln(s)n^{1/3}/4$. Note that the number of blocks r_n is of order $(\ln(n))^2$. Here and in what follows, all specific constants allow us to use an estimator for small samples. The estimator will be commented on shortly.

In the following proposition we use notations of Theorem 2.1 introduced in Section 2.

Theorem 3.1 (Upper Bound for the MISE of $\widetilde{F}^X(x, f^Z)$). *Let Assumption 1 hold. Then for a CSC sample of size n,*

$$\sup_{F^X \in \mathcal{F}(\alpha, Q)} \mathbb{E}_{F^X} \left\{ [n\mathcal{J}(\alpha, Q, d, 1)]^{2\alpha/(2\alpha+1)} \int_0^1 [\widetilde{F}^X(x, f^Z) - F^X(x)]^2 dx \right\}$$

$$\leqslant (1 + o_n(1)). \tag{3.10}$$

If f^Z is unknown, then its estimation is based on available sample $Z_1, \ldots, Z_n$ from Z. Accordingly, define a bounded from below projection estimate

$$\bar{f}^Z(z) := \max(\gamma_n, \check{f}^Z), \quad \text{where} \quad \check{f}^Z(z) := n^{-1} \sum_{j=0}^{J_n} \sum_{l=1}^n \varphi_j(Z_l)\varphi_j(z), \tag{3.11}$$

and

$$J_n := \lfloor \ln(s)n^{1/3}/2 \rfloor, \quad \gamma_n := 1/(3\ln(s)). \tag{3.12}$$

Theorem 3.2 (Upper Bound for Plug-In Estimator $\widetilde{F}^X(x, \bar{f}^Z)$). *Let Assumption 1 hold. Then the upper bound (3.10) holds for $\widetilde{F}^X(x, \bar{f}^Z)$.*

Note that the plug-in estimator is completely data-driven and it adapts to unknown smoothness of an underlying cdf of X. Important theoretical achievement of Theorem 3.2 is that only differentiability of $f^Z(z)$ is assumed (versus α-fold differentiability assumed in [16]).The improvement is due to the simpler Fourier estimator (3.5) and a more advanced proof.

Now let us comment on the introduced estimator and statistics. It is worthwhile to begin with a brief introduction to series estimation.

Estimator (3.9) belongs to a class of orthogonal series estimators. A series estimator employs a familiar from the function analysis result that a square integrable on $[0, 1]$ function $q(x)$ can be written as $q(x) = \sum_{j=0}^\infty \kappa_j \varphi_j(x)$ whenever $\{\varphi_j(x), j = 0, 1, \ldots\}$ is a basis on $[0, 1]$ and $\kappa_j = \int_0^1 q(x)\varphi_j(x)dx$ are corresponding Fourier coefficients. Then a traditional nonparametric rate-optimal estimation paradigm, based on a sample of size n, is as follows. Find a Fourier estimator $\bar{\kappa}_j$ whose mean squared error is of order n^{-1}, that is $\mathbb{E}\{(\bar{\kappa}_j - \kappa_j)^2\} \leqslant Cn^{-1}$. Then use a projection estimator $\bar{q}(x) = \sum_{j=1}^{M_n} \bar{\kappa}_j \varphi_j(x)$ where M_n is called a cutoff. A popular add-hoc choice of the cutoff is $M_n = n^{1/5}$ which yields rate-optimal estimation for $q(x) \in \mathcal{F}(2, Q)$. Indeed, the Parseval identity implies that

the mean integrated squared error (MISE) of the estimator is

$$\mathbb{E}\{\int_0^1 (\bar{q}(x) - q(x))^2\} = \sum_{j=0}^{M_n} \mathbb{E}\{(\bar{\kappa}_j - \kappa_j)^2\} + \sum_{j>M_n} \kappa_j^2 =: V_n + B_n. \qquad (3.13)$$

Here V_n is called the variance component of the MISE and it is proportional to $M_n n^{-1}$, and B_n is called the integrated squared bias (or simply bias) component and it is proportional to M_n^{-4}. This yields the optimal MISE convergence of order $n^{-4/5}$. Note that for $q \in \mathcal{F}(\alpha, Q)$ the optimal cutoff is of order $n^{1/(2\alpha+1)}$. An interesting and practically important property of a rate-optimal projection estimator using a trigonometric basis is that its derivative is elementary calculated and it is a rate-optimal estimate of the derivative of an estimated function. This is due to the fact that the optimal cutoff is the same for a function and its derivative, see [15]. Knowing this fact will be handy in understanding the next Section 4.

In applications smoothness of $q(x)$, defined by parameter α, is unknown and then numerical procedures for data-driven (adaptive) choosing M_n are developed based on a variance-bias tradeoff. The main technical element here is to note that due to the Parseval identity a bias in (3.13) can be written as $B_n = \int_0^1 q^2(x)dx - \sum_{j=0}^{M_n} \kappa_j^2$. Now note that $\int_0^1 q^2(x)dx$ does not depend on M_n, and accordingly the problem of finding a data-driven cutoff is converted into minimizing $\sum_{j=0}^{M} [\mathbb{E}_q\{(\bar{\kappa}_j - \kappa_j)^2\} - \kappa_j^2]$ with respect to M. Another interesting approach, popular for wavelet bases, is to use a thresholding like the one in the first sum on the right side of (3.9). A thresholding may yield almost rate optimal (within a logarithmic factor) estimation. The interested reader can find more about rate-optimal estimation in books [13,35].

The problem becomes more complex if sharp-minimax is of interest when the aim is to achieve both optimal rate and constant of the MISE convergence. For a global Sobolev class $\mathcal{F}(\alpha, Q)$ defined in (2.2), a linear Pinsker's oracle

$$\tilde{q}^*(x) := \sum_{j=0}^{J_n(\alpha,Q)} [1 - (j/J_n(\alpha, Q))^\alpha]\hat{\kappa}_j \varphi_j(x) \qquad (3.14)$$

is sharp-minimax whenever the Fourier estimator $\hat{\kappa}_j$ is efficient, namely

$$\mathbb{E}_q\{(\hat{\kappa}_j - \kappa_j)^2\} = d'n^{-1}[1 + o_j(1) + o_n(1)]. \qquad (3.15)$$

Here $J_n(\alpha, Q)$ is a special cutoff specific for an underlying Sobolev class, and d' is a coefficient of difficulty specific for an underlying statistical problem. For instance, for estimation of the density $f^X(x)$ based on direct observations of X this coefficient is 1. Note that the linear oracle (3.14) "smooths" Fourier estimates. While being very simple, the linear oracle has two drawbacks. The former is that it is possible but extremely difficult to estimate the cutoff, the latter is that derivative of the linear oracle is not a sharp-minimax estimate of the derivative of $q(x)$.

Nonetheless, the Pinsker's linear oracle has been a motivation for a blockwise-shrinkage adaptation used by the proposed estimator (3.9). Let us explain the underlying idea following [11] where the problem of density estimation based on direct observations is considered. First, the Pinsker smoothing coefficient $(1 - (j/J_n(\alpha, Q))^\alpha)$ is dominated by an oracle-coefficient $\kappa_j^2/[\kappa_j^2 + d'n^{-1}]$ where $d' = 1$ is the coefficient of difficulty for the density estimation problem. Second, we may estimate the oracle-coefficient but the accuracy is not sufficient for its mimicking. Instead, using the fact that the Pinsker's smoothing coefficients with neighboring frequencies j are close to each other, we can propose a single smoothing coefficient for a block of Fourier coefficients. Consider a block of frequencies B which includes L frequencies, then the corresponding smoothing blockwise oracle-coefficient is

$$w^* := \frac{L^{-1} \sum_{j \in B} \kappa_j^2}{L^{-1} \sum_{j \in B} \kappa_j^2 + d'n^{-1}}. \tag{3.16}$$

Due to the averaging over the block, the Sobolev functional $L^{-1} \sum_{j \in B} \kappa_j^2$ can be estimated with a sufficient accuracy for sharp-minimax estimation. The oracle-coefficient w^* is convenient when the coefficient of difficulty is known as in the density estimation for direct observations. Otherwise, it may be more convenient to mimic a closely related, due to (3.15), an oracle-coefficient

$$w_* := \frac{L^{-1} \sum_{j \in B} \kappa_j^2}{L^{-1} \sum_{j \in B} \mathbb{E}\{\hat{\kappa}_j^2\}}. \tag{3.17}$$

A nice feature of this oracle-coefficient is that no estimation of d' is required, and more discussion and specific results can be found in Section 6, see Lemmas 6.3-6.6. Further, it is established in [12] that derivative of a sharp-minimax blockwise estimator is a sharp-minimax estimate of the derivative. This is another attractive feature of a blockwise adaptive estimation that will be used shortly in Section 4.

This ends our brief overview of orthogonal series estimation which sheds light on the proposed methodology, and the interested reader can be find more information in books [13,15,35].

Now let us comment on specifics of the proposed estimator (3.9). We are explaining statistics in the order as they were introduced. The pilot Fourier estimate (3.2) is a simple sample mean estimate based on the formula (2.1), its properties are highlighted in Lemma 6.1 of Section 6. Using the terminology of [24] we may say that the pilot estimate is based on "cases". The pilot cdf estimator (3.3) is the above-discussed projection series estimator with cutoff s and "removed" frequency $j \leq s$. As a result, we have important relation $\int_0^1 \breve{F}_{-j}(x)\varphi_j(x)dx = 0$ for all $j = 0, 1, \ldots$ Another property of the pilot estimator is that it is unbiased estimate of $F_{-j}(x)$ defined in (3.4), and also see Corollary 6.1 in Section 6. Note that the mean squared error of the pilot Fourier estimate is $\int_0^1 [F^X(x)/f^Z(x)]dx n^{-1}(1 + o_j(1) + o_n(1))$, while the mean squared error of the proposed Fourier estimate (3.5) is $dn^{-1}(1 + o_j(1) + o_n(1))$, see also Lemmas

6.1 and 6.2. As a result, recalling (3.15) we may refer to the Fourier estimate $\widehat{\theta}_j$ as efficient. Note how subtracting the pilot cdf estimate from the status Δ in (3.5) decreased the mean squared error. Now we are ready to look at the used blockwise smoothing coefficients. If we replace in statistic (3.6) the pilot estimate $\breve{F}_{-j}$ by F_{-j}, then this statistic becomes U-statistic and unbiased estimate of the Sobolev functional $\Theta_k := L_k^{-1} \sum_{j \in B_k} \theta_j^2$, see a discussion in the proof of Lemma 6.6 in Section 6. This fact, together with the oracle's smoothing coefficient (3.17), explain the blockwise smoothing used in the second sum on the right side of (3.9). Lemma 6.4 in Section 6 sheds light on relationship between statistics $\widehat{\Theta}_k$ and $\widetilde{\Theta}_k$, while Lemma 6.5 explains how well the blockwise shrinkage mimics a blockwise oracle in L_2-norm. The used thresholding is necessary to attain sharp-minimaxity as shown in [11]. The choice of r_n in (3.9) is such that the estimator includes all frequencies that a sharp-minimax oracle does.

Now let us comment on the low-frequency component in (3.9). It is created solely for small samples using recommendation of [13], and its effect on the asymptotic MISE is negligible, see details in (6.27), (6.30) and Corollary 6.2 in Section 6. Further, thresholding $I(\widehat{\theta}_j^2 > 2\widetilde{d}sn^{-1})$ may be skipped or replaced by $I(\widehat{\theta}_j^2 > Csn^{-1})$ with no effect on the sharp-minimax. The used estimate (3.8) may be replaced by others. For instance, note that $\mathbb{E}\{\widehat{\theta}_j^2\} = dn^{-1}[1 + o_j(1) + o_n(1)]$, and hence using a sample mean or sample median of $\{\widehat{\theta}_j^2, j = s+1, \ldots, 2s\}$ is applicable. The reader familiar with wavelet estimators may recall that the median approach is popular in wavelet statistical packages. Overall, there is a large flexibility in choosing a lower frequency component of a blockwise shrinkage estimator, and in (3.9) this component is chosen based on recommendations of [13] and analysis of examples presented in Section 5. We will continue the discussion in Section 7.

4. Density Estimation

There are two classical approaches to the problem of density estimation, and we are considering them in turn. The former is to propose a smooth estimate of cdf and then take derivative. This approach looks natural for CSC data due to the underlying likelihood (2.1). An interesting example of using this approach for CSC data is [19] where derivative of a maximum smoothed likelihood cdf estimate is used to estimate the density. Under the assumption that the third derivative of $F^X(x)$ is continuous, it is shown that the density estimate attains the rate $n^{-4/7}$. The latter approach is to bypass estimation of the cdf and consider density estimation as a self-defined nonparametric problem, recall literature review in the Introduciton. We consider these approaches in turn.

Using estimates (3.9) and (3.11) we can define a density estimate

$$\widetilde{f}^X(x) := \begin{cases} d\widetilde{F}^X(x, f^Z)/dx & \text{if } f^Z \text{ is known,} \\ d\widetilde{F}^X(x, \bar{f}^Z)/dx & \text{otherwise.} \end{cases} \tag{4.1}$$

Next proposition follows from [12] and results of Section 3.

Corollary 4.1 (Upper Bound for the MISE of Density Estimator (4.1)).
Let Assumption 1 hold and $\alpha \geqslant 2$. Then

$$\sup_{F^X \in \mathcal{F}(\alpha, Q)} \mathbb{E}_{F^X} \left\{ [n\mathcal{J}(\alpha, Q, d, 2)]^{2(\alpha-1)/(2\alpha+1)} \int_0^1 [\widetilde{f}^X(x) - f^X(x)]^2 dx \right\}$$

$$\leqslant (1 + o_n(1)). \tag{4.2}$$

This result presents good and bad news about density estimation for CSC data. The good one is that adaptive sharp-minimax estimation is possible. The bad one is that rate of the MISE convergence is dramatically slower than for the case of a direct sampling from X. Indeed, recall that if density $f^X(x)$ has γ derivatives then based on a direct sample of size n it may be estimated with the rate $n^{-2\gamma/(2\gamma+1)}$. In (4.2) α is the number of derivatives of $F^X(x)$, and accordingly density $f^X(x)$ has $\gamma = \alpha - 1$ derivatives. Then Corollary 4.1 asserts that the best rate of the MISE convergence for a CSC sample is $n^{-2\gamma/(2\gamma+3)}$. According to [13], for a direct sample of size n from X this rate is the same as for estimation of a trivariate density having γ derivatives in each variable. Now recall a familiar curse of multidimensionality in nonparametric estimation, see a discussion in [35], and then Corollary 4.1 sheds a new light on complexity of CSC data analysis. We will continue this discussion in Section 5.

Under the second approach, when estimation of cdf is bypassed, we estimate density $f^X(x)$ directly. The proposed density estimator is again a blockwise-shrinkage cosine series estimator discussed in Section 3, and we continue to use notation of that section. Further, for the reader's convenience we will compare steps in construction of a blockwise density estimator with those for the cdf estimator (3.9).

The main principal step is to understand how Fourier coefficients

$$\zeta_j := \int_0^1 f^X(x)\varphi_j(x)dx \tag{4.3}$$

of the density f^X can be estimated. We know that $\zeta_0 = \int_0^1 f^X(x)dx = 1$ because X is supported on $[0, 1]$. Accordingly, we need to propose a Fourier estimator for $j \geqslant 1$, and from now on we are considering only $j \geqslant 1$. Using integration by parts we can write,

$$\zeta_j = \int_0^1 f^X(x)\varphi_j(x)dx = [\varphi_j(1)F^X(1) - \varphi_j(0)F^X(0)] - \int_0^1 \varphi_j^{(1)}(x)F^X(x)dx$$

$$= \varphi_j(1) - \mathbb{E}_{F^X}\left\{\Delta \frac{\varphi_j^{(1)}(Z)}{f^Z(Z)}\right\} = \varphi_j(1) + (\pi j)\mathbb{E}_{F^X}\left\{\Delta \frac{\psi_j(Z)}{f^Z(Z)}\right\}. \tag{4.4}$$

In the last equality we used $F^X(0) = 0$, $F^X(1) = 1$, and $\varphi_j^{(1)}(x) = (-\pi j)\psi_j(x)$. Here

$$\psi_j(x) := 2^{1/2} \sin(\pi j x), \quad j = 1, 2, \dots \tag{4.5}$$

are elements of the classical sine basis on $[0, 1]$. Now note that the expectation on the right side of (4.4) can be estimated by a sample mean estimate. This yields a pilot Fourier estimator (compare with (3.2))

$$\check{\zeta}_j := \varphi_j(1) + \frac{\pi j}{n} \sum_{l=1}^{n} \frac{\Delta_l \psi_j(Z_l)}{f^Z(Z_l)}. \tag{4.6}$$

Note how similar pilot Fourier estimators (4.6) and (3.2) are despite different estimands ζ_j and θ_j, respectively. They are both unbiased estimates of their estimands, but a critical difference is in the factor πj in (4.6) which yields an increasing by factor $(\pi j)^2$ mean squared error (MSE) of $\check{\zeta}_j$ while the MSE of $\check{\theta}_j$ is bounded.

Following (3.3), we use Fourier estimates (4.6) to construct a pilot cdf estimator

$$\widehat{F}_{-j}(x) := \sum_{i \in \{\{1,\ldots,s\}\setminus\{j\}\}} \check{\zeta}_i \psi_i(x) I(j \leqslant s) + \sum_{i=1}^{s} \check{\zeta}_i \psi_i(x) I(j > s). \tag{4.7}$$

Note that $\widehat{F}_{-j}(x)$ is unbiased estimate of a function

$$F^*_{-j}(x) = \sum_{i \in \{\{1,\ldots,s\}\setminus\{j\}\}} \zeta_i \psi_i(x) I(j \leqslant s) + \sum_{i=1}^{s} \zeta_i \psi_i(x) I(j > s). \tag{4.8}$$

This function mimics $F_{-j}(x)$ defined in (3.4) only here the sine basis is used in place of the cosine basis. The sine basis is used because, due to using sines in (4.6), the new pilot estimate and the approximation F^*_{-j} of $F^X(x)$ have the desired property

$$\int_0^1 \widehat{F}_{-j}(x)\psi_j(x)dx = \int_0^1 F^*_{-j}(x)\psi_j(x)dx = 0. \tag{4.9}$$

Note that (4.9) matches the same property of $\check{F}_{-j}(x)$ and $F_{-j}(x)$ with respect to $\varphi_j(x)$, see the paragraph below line (3.4).

Following (3.5) we define a Fourier estimator

$$\widehat{\zeta}_j := \varphi_j(1) + \frac{\pi j}{n} \sum_{l=1}^{n} \frac{(\Delta_l - \widehat{F}_{-j}(Z_l))\psi_j(Z_l)}{f^Z(Z_l)}, \ j \geqslant 1. \tag{4.10}$$

It is easy to check that if in (4.10) the estimate $\widehat{F}_{-j}(Z_l)$ is replaced by $F^*_{-j}(Z_l)$, then due to (4.9) estimate $\widehat{\zeta}_j$ becomes unbiased estimate of ζ_j. Another important property of the Fourier estimate will be presented shortly in Lemma 4.1, and also see Lemmas 6.7 and 6.8 in Section 6.

Now we introduce three new statistics. The first one is analog of $\widetilde{\Theta}_k$,

$$\widetilde{\mathcal{Z}} := \frac{2}{L_k n(n-1)} \sum_{1 \leqslant l < l' \leqslant n} \sum_{j \in B_k} \left[\varphi_j(1) + \frac{(\pi j)(\Delta_l - \widehat{F}_{-j}(Z_l))\psi_j(Z_l)}{f^Z(Z_l)} \right]$$

$$\times \left[\varphi_j(1) + \frac{(\pi j)(\Delta_{l'} - \widehat{F}_{-j}(Z_{l'}))\psi_j(Z_{l'})}{f^Z(Z_{l'})} \right]. \tag{4.11}$$

The second statistic is analog of $\widehat{\Theta}_k$,

$$\widehat{\mathcal{Z}}_k := L_k^{-1} \sum_{j \in B_k} \widehat{\zeta}_j^2. \tag{4.12}$$

The third statistic is analog of $\widetilde{d}$,

$$\widehat{d} := \min \left(3s, n^{-1} \sum_{l=1}^{n} \frac{\widehat{F}_{-s}(Z_l)(1 - \widehat{F}_{-s}(Z_l))}{[f^Z(Z_l)]^2} \right). \tag{4.13}$$

Now we are in a position to introduce a blockwise density estimator that mimics the cdf estimator (3.9),

$$\widehat{f}^X(x, f^Z) := 1 + \sum_{j=1}^{s} \widehat{\zeta}_j I(\widehat{\zeta}_j^2 > (\pi j)^2 2 \widehat{d} s n^{-1}) \varphi_j(x)$$

$$+ \sum_{k=1}^{r_n} \min(1, \frac{\widetilde{\mathcal{Z}}_k}{\widehat{\mathcal{Z}}_k}) I(\widetilde{\mathcal{Z}}_k > 10[\pi \sum_{r=0}^{k} L_r]^2 s^{-1} n^{-1}) \sum_{j \in B_k} \widehat{\zeta}_j \varphi_j(x). \tag{4.14}$$

Finally, if $f^Z(z)$ is unknown, then we use its estimate $\bar{f}^Z(z)$ defined in (3.11).

Theorem 4.1 (Upper Bound for the MISE of Density Estimator (4.14)).
Let Assumption 1 hold and $\alpha \geqslant 2$. Set $\widehat{f}^X(x) = \widehat{f}^X(x, f^Z)$ if f^Z is known and $\widehat{f}^X(x) = \widehat{f}^X(x, \bar{f}^Z)$ otherwise. Then

$$\sup_{F^X \in \mathcal{F}(\alpha, Q)} \mathbb{E}_{F^X} \left\{ [n \mathcal{J}(\alpha, Q, d, 2)]^{2(\alpha-1)/(2\alpha+1)} \int_0^1 [\widehat{f}^X(x) - f^X(x)]^2 dx \right\}$$

$$\leqslant (1 + o_n(1)). \tag{4.15}$$

We may conclude that the both methods (taking derivative of the blockwise cdf estimate and a direct blockwise density estimation) lead to a sharp-minimax estimation that matches performance of oracles. Overall the second method (direct density estimation) is preferable because this technique is widely used and many innovations are developed, see [13,15,35]. Another comment is that integration of the proposed density estimator yields a sharp-minimax cdf estimator.

Now let us present a technical result which explains our choice of Fourier estimator $\widehat{\zeta}_j$ and sheds light on Theorem 4.1.

Lemma 4.1. *Under assumption of Theorem 4.2 the mean squared error of Fourier estimator $\widehat{\zeta}_j$ satisfies*

$$\mathbb{E}_{F^X}\{(\widehat{\zeta}_j - \zeta_j)^2\} = \frac{(\pi j)^2}{n} \left[\int_0^1 \frac{F^X(x)(1 - F^X(x))}{f^Z(x)} dx + o_j(1) + o_n(1) \right]. \tag{4.16}$$

The interested reader may also compare statistical properties of the Fourier estimates $\check{\zeta}_j$ and $\hat{\zeta}_l$ using Lemmas 6.7 and 6.8 in Section 6.

Remark 4.1. Recall that $d = \int_0^1 F^X(x)(1 - F^X(x))[f^Z(x)]^{-1}dx$ is the coefficient of difficulty for cdf estimation defined in (2.6). On the right side of (4.16) we see this coefficient of difficulty multiplied by factor $(\pi j)^2$ which is due to $\int_0^1 [d\varphi_j(x)/dx]^2 dx = (\pi j)^2$. This observation explains similarity between the two approaches for density estimation. Lemma 4.1 also helps us to understand why for $F^X \in \mathcal{F}(\alpha, Q)$ rate optimal cdf and density projection estimators can use the same cutoff $J_n(\alpha)$. Indeed, let us make several simple calculations. The MISE of a projection cdf estimator $\bar{F}(x) := \sum_{j=0}^J \hat{\theta}_j \varphi_j(x)$ is of order $Jn^{-1} + J^{-2\alpha}$. The latter yields rate-optimal cutoff $J_n(\alpha) = J^{1/(2\alpha+1)}$ and the rate $n^{-2\alpha/(2\alpha+1)}$ for the MISE convergence. The MISE of a projection density estimator $\bar{f}(x) := 1 + \sum_{j=1}^J \hat{\zeta}_j \varphi_j(x)$ is of order $\sum_{j=1}^J (\pi j)^2 n^{-1} + J^{-2(\alpha-1)}$. This yields the same rate-optimal cutoff $J_n(\alpha) = n^{1/(2\alpha+1)}$ and the corresponding rate $n^{-2(\alpha-1)/(2\alpha+1)}$ for the MISE convergence. Note that the above-presented rates of MISE convergence are optimal according to Theorem 2.1.

Let us finish our theoretical sections with a practical comment. In general a series estimate may be non bona fide. If this is an issue, then an L_2-projection on a corresponding bona fide class may be performed, see a discussion in book [13] and use its R software.

5. Examples

The section sheds additional light on estimation for CSC data via analysis of real and simulated data. We begin with real-data examples.

Environmental company BIFAR has been interested in exploring aerobic treatment of municipal wastewater, see a discussion of the treatment and BIFAR's CSC experiments in [16]. In one of the experiments, a random variable of interest was time X when a chemical pollutant appears at a sludge tank. Because it was impossible to observe the time directly, a CSC study was conducted. BIFAR's CSC observations $(Z_1, \Delta_1), \ldots, (Z_n, \Delta_n)$ are shown in the top diagram in Figure 1. Before proceeding to estimates, let us look at the data and try to guess an underlying density. Recall that guessing an underlying density is often possible for direct observations, and this is a recommended first step in nonparametric estimation, see chapter 3 in [13]. In the diagram we can directly observe monitoring times, and it is reasonable to assume that an underlying density is uniform. The estimate $\bar{f}^Z(z)$ (the solid line) supports the conclusion.

Now let us try to analyze the CSC data. Because $\Delta = I(X \leqslant Z)$ and all Δs with $Z > 0.65$ are equal to 1, the CSC data, together with the uniform $\bar{f}^Z$, tells us that an underlying density has a vanishing right tail. Next we note that Δs with $Z < 0.3$ are zero and hence all corresponding underlying times of interest satisfy $X > Z$. This points upon a vanishing left tail. Unfortunately, there is nothing else that a visualization can produce. In particular, it is difficult to answer questions about symmetry or multimodality. This complexity is explained
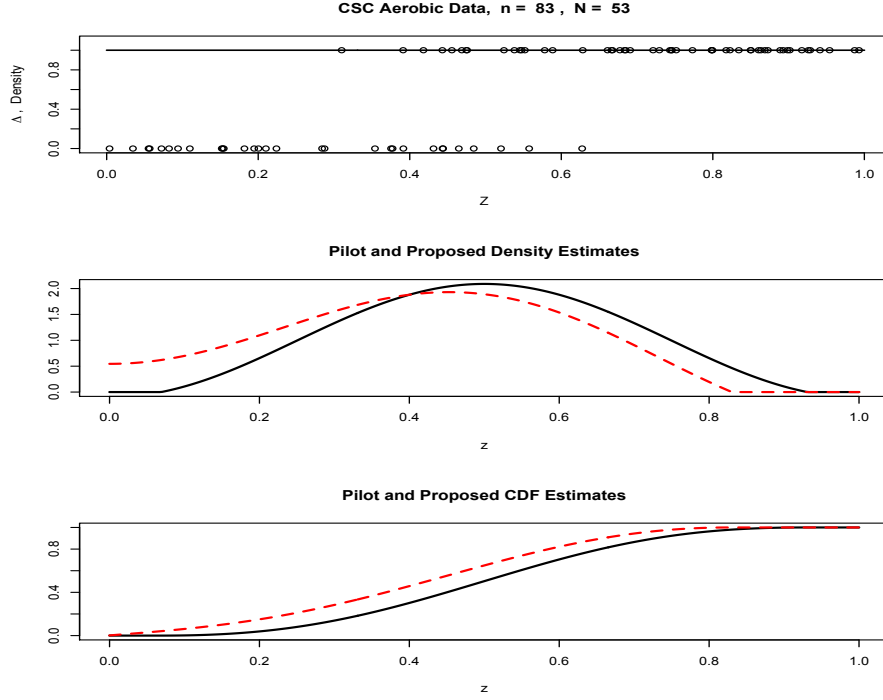
CSC Aerobic Data, n = 83 , N = 53

Pilot and Proposed Density Estimates

Pilot and Proposed CDF Estimates

Fɪɢ 1. *Analysis of BIFAR CSC aerobic data. Monitoring times are rescaled onto [0,1]. The top diagram shows CSC observations by the circles, n is the sample size, $N := \sum_{l=1}^{n} \Delta_l$, the solid line shows the estimated density of the monitoring time Z. In the middle and bottom diagrams a solid line is the proposed estimate and a dashed line is the pilot estimate.*

by formula (2.1) implying that to visualize an underlying density, one first need to visualize an underlying cdf and then visualize its derivative. The derivative step is too complicated for visual analysis. We may conclude that only statistical estimators may help us to gain understanding of CSC data.

The middle and bottom diagrams exhibit pilot (the dashed line) and proposed (the solid line) estimators of the density and cdf. Recall that all estimates are series estimates and they are consistent, the proposed estimator is more accurate while the pilot estimator is very simple. For the data at hand, based on the estimated F^X and f^Z, the ratio of the standard deviation of the pilot Fourier estimator (3.2) to the standard deviation of the proposed Fourier estimator (3.5) is 2.3. This sheds light on improvement in estimating Fourier coefficients. Further, let us look at the two density estimates. The pilot estimate is skewed to the left and asymmetric. At the same time, the proposed estimate is unimodal and symmetric. BIFAR concurred the symmetric and unimodal shape of an underlying density of interest.
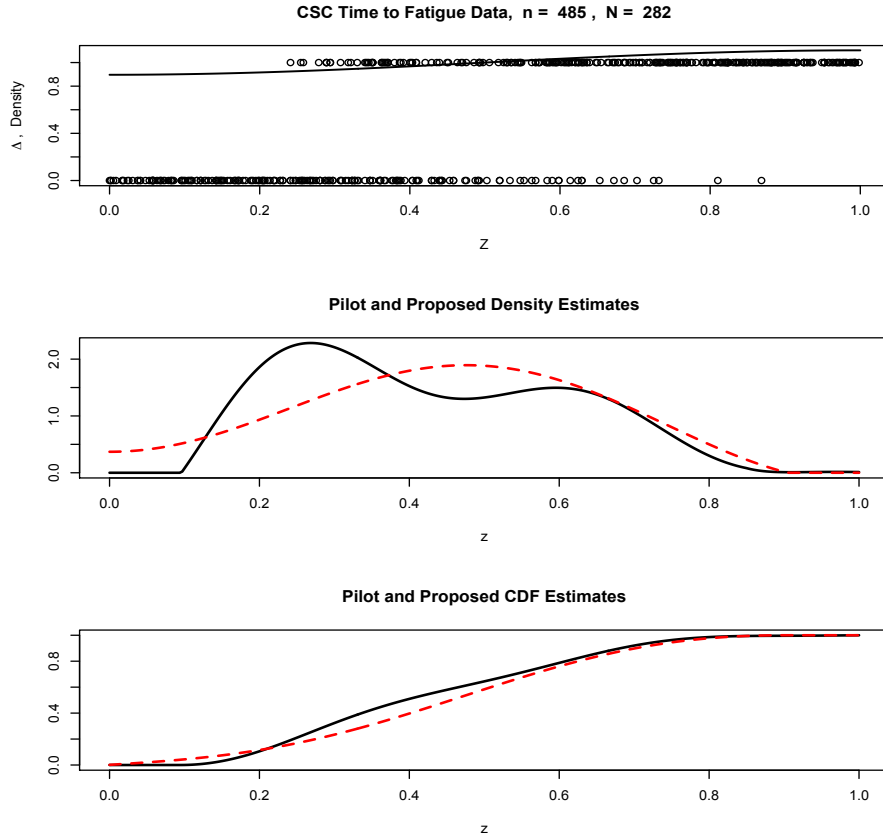
Fɪɢ 2. *Analysis of time until "drowsy driving." Monitoring times are rescaled onto [0,1]. The top diagram shows the CSC data by the circles, n is the sample size, $N := \sum_{l=1}^{n} \Delta_l$, the solid line shows the estimated density of Z. In the middle and bottom diagrams a solid line is the proposed estimate and a dashed line is the pilot estimate.*

Our second example is more interesting in terms of the shape of an underlying distribution, and it involves a larger sample size. Operating a motor vehicle while fatigued or sleepy is commonly referred to as "drowsy driving." There is an extensive literature devoted to the topic with a primary interest in understanding the drowsy driving, driver assistance technologies in vehicles, and the driver's fatigue detection by monitoring systems, see reviews and interesting data in [31,38]. Here our aim is less ambitious, and we would like to evaluate distribution of a driving time X until first yawning. This is a simple and inexpensive CSC experiment which uses a call to a driver and a question about yawning prior to the call. CSC data for $n = 485$ commercial drivers are shown in the top diagram in Figure 2. The structure of Figure 2 is identical to Figure 1, and this allows us to compare the two CSC datasets. We again see vanishing tails, but the difference is in asymmetry of the fatigue data. There are too many observations to visualize an underlying density f^Z, and the estimate $\bar{f}^Z$ (the solid line) indicates an increasing density. Based on the top diagram, it is difficult to add something else to these remarks.

Now let us look at the middle diagram. The left tail of the pilot estimate (the dashed line) looks strange, there is no reason to believe that it is correct. The more accurate proposed density estimate (the solid line) tells us an interesting story about the data. It reveals two modes in the lifetime of interest X. A reasonable explanation is that the left and right modes are created by driving at night and daytime, correspondingly. The bimodal density is also supported by a known fact (see the above-mentioned literature) that a driver three times more likely to have a fatal accident at night than during the day. Let us also note that the pilot estimate is a smoothed version of the proposed estimate. The bottom diagram shows us estimates of cdf. Here the ratio of the standard deviation of the pilot Fourier estimator to the standard deviation of the proposed Fourier estimator is 2.2.

Now let us present results of a numerical study when we know an underlying model. We would like to understand how CSC effects density estimation and relative performance of the proposed CSC density estimator. Figure 3 exhibits two simulations with underlying densities shown by the solid lines in the bottom diagrams. In the left diagram the density is unimodal and in the right it is bimodal. In both experiments the monitoring time is uniform. Considered estimates are: (i) Estimate of [13] based on an underlying sample from X. Recall that we are dealing with simulated CSC and hence know underlying lifetimes of interest. Let us refer to this estimate as an oracle. The oracle is shown by the dashed line; (ii) Proposed estimate. The estimate is shown by the dotted line; (iii) Aggregated estimate of [15]. This estimate aggregates two series estimates based on observations with $\Delta = 1$ and $\Delta = 0$, respectively, and this is a reliable rate optimal CSC density estimate supported by the software [15].

Now we repeat a simulation 5000 times, for each simulation and each estimate calculate an integrated squared error (ISE), calculate ratio A of ISE of the proposed estimate to ISE of the oracle, calculate ratio B of ISE of the aggregated estimate to ISE of the oracle, and then calculate sample means $\bar{A}$ and $\bar{B}$ of the ratios over 5000 simulations. Results are shown in Table 1, and the underlying
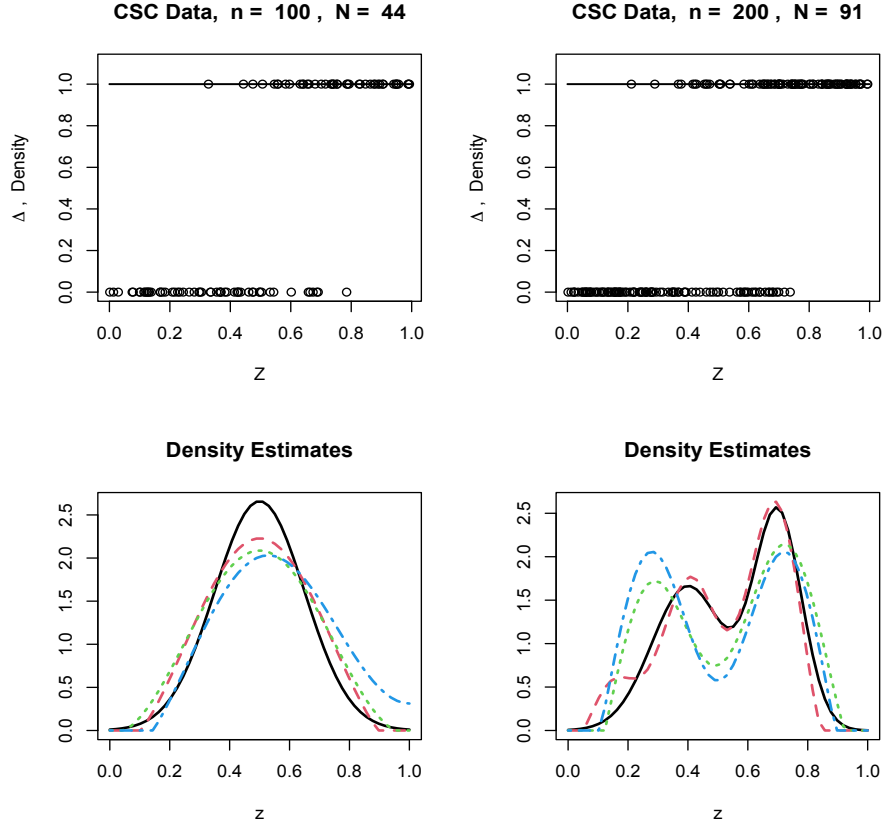
FIG 3. *Two simulated CSC examples. A top diagram shows CSC data by the circles, n is the sample size, $N := \sum_{l=1}^{n} \Delta_l$, the solid line shows the estimated density of Z. A bottom diagram shows by the solid, dashed, dotted and dot-dashed lines an underlying density of interest f^X, the oracle (density estimate of [13]) based on an underlying sample from X, the proposed estimate, and aggregated estimate of [15], respectively.*

models are as follows. Model 1 is for the unimodal underlying density, and the left column in Figure 3 exhibits a particular simulation. Model 2 is for the bimodal underlying density, and the right column in Figure 3 exhibits a particular simulation. Models 3 and 4 use a new density for the monitoring time Z which is a mixture with weight 0.3 of the uniform density and with weight 0.7 of the unimodal density shown by the solid line in the left bottom diagram in Figure 3. Then model 3 in Table 1 corresponds to the unimodal underlying density of interest, and model 4 to the bimodal underlying density of interest shown by solid lines in the bottom diagrams of Figure 3, respectively. Let us look at the presented results. As we already know from the theory, small samples may present only onset of ill-posedness, and the results support this possibility. For larger samples CSC creates more dramatic complications with respect to direct observations. Relative performance of the two CSC density estimators becomes worse with increased sample size, and this outcome coincides with the theory. Recall that asymptotically CSC density estimation is equivalent to estimating a trivariate density, and the results reflect this. On a positive note, the study shows that the used series density estimator is robust toward an underlying density of the monitoring time, and there is only a minor bump in performance of the estimator when the density of monitoring time changes from the uniform to the unimodal. To see the latter, compare outcomes for odd and even models. Another interesting observation is that the relative performance of the CSC estimates is better for the bimodal underlying density than for the uniform underlying density. The latter is due to the fact that even for the oracle estimation of the bimodal density is a notoriously complicated task, see a discussion in [13].

TABLE 1

Results of a numerical study of two density estimates with respect to an oracle that knows underlying direct observations of the lifetime of interest. Each entry shows $\bar{A}/\bar{B}$ explained in the text.

Model	n				
	100	200	300	400	500
1	7.4/8.2	12.3/14.7	26.5/28.6	43.1/50.5	56.9/67.3
2	4.3/5.3	7.3/10.2.	11.8/15.2.	14.4/17.8	18.4/23.0.
3	8.7/9.4	12.8/15.2	27.3/29.2	44.7/51.6	57.3/67.7
4	5.1/6.2	7.6/10.6	12.7/15.6	15.9/18.1	19.3/23.8

In conclusion, the examples show that estimation based on CSC data is a feasible but ill-posed task. Fortunately, the ill-posedness is relatively mild for small samples. Accordingly practitioners should not shine from employing CSC experiments and nonparametric analysis of CSC data because visualization of nonparametric density estimates may shed a useful light on the lifetime of interest.

6. Proofs

In what follows Cs are generic finite constants, d is the coefficient of difficulty CSC defined in (2.6), $\mathbb{E}_{F^X}\{\cdot\}$ and $\mathbb{V}_{F^X}(\cdot)$ denote the expectation and the variance given the underlying cdf F^X whenever we would like to stress an underlying cdf, and we continue using notations introduced at the end of the Introduction.

Proof of Theorem 3.1. The proof is structured via a sequence of technical lemmas. This simplifies understanding main steps in the proof. We begin with evaluation of the mean and variance of the pilot estimator $\breve{\theta}_j$ of Fourier coefficient θ_j defined in (3.2) and (3.1), respectively.

Lemma 6.1. *Let Assumption 1 hold. Then the pilot estimator $\breve{\theta}_j$ of θ_j is unbiased and rate optimal, namely*

$$\mathbb{E}_{F^X}\{\breve{\theta}_j\} = \theta_j, \tag{6.1}$$

and

$$\mathbb{V}_{F^X}(\breve{\theta}_j) = n^{-1}\int_0^1 F^X(x)[f^Z(x)]^{-1}[\varphi_j(x)]^2 dx \leqslant Cn^{-1}. \tag{6.2}$$

Proof of Lemma 6.1. For the expectation we write using (2.1),

$$\mathbb{E}_{F^X}\{\breve{\theta}_j\} = \mathbb{E}_{F^X}\{\Delta\varphi_j(Z)[f^Z(Z)]^{-1}\}$$

$$= \int_0^1 [f^Z(z)F^X(z)/f^Z(z)]\varphi_j(z)dz = \int_0^1 F^X(z)\varphi_j(z)dz = \theta_j.$$

This verifies unbiasedness of the pilot Fourier estimator $\breve{\theta}_j$. Variance of the pilot Fourier estimator is

$$\mathbb{V}_{F^X}(\breve{\theta}_j) = n^{-1}\mathbb{V}_{F^X}\left(\Delta\varphi_j(Z)[f^Z(Z)]^{-1}\right)$$

$$= n^{-1}[\mathbb{E}_{F^X}\{\Delta^2\varphi_j^2(Z)[f^Z(Z)]^{-2}\} - \theta_j^2]$$

$$= n^{-1}\int_0^1 F^X(z)[\varphi_j(z)]^2[f^Z(z)]^{-1}dz \leqslant Cn^{-1}.$$

We conclude that the pilot Fourier estimator is unbiased and rate optimal. Lemma 6.1 is verified.

Now let us consider Fourier estimator $\widehat{\theta}_j$ defined in (3.5). Next proposition sheds light on why this estimator may be referred to as efficient for the considered nonparametric problem of the cdf estimation.

Lemma 6.2. *Let Assumption 1 hold and $F^X \in \mathcal{F}(\alpha, Q)$. Then*

$$|\mathbb{E}_{F^X}\{\widehat{\theta}_j\} - \theta_j| \leqslant Csn^{-1}, \tag{6.3}$$

and

$$\mathbb{E}_{F^X}\{(\widehat{\theta}_j - \theta_j)^2\} \leqslant dn^{-1}[1 + C((j+1)^{-2} + s^{-1})]. \qquad (6.4)$$

Here $d = \int_0^1 F^X(x)(1 - F^X(x))[f^Z(x)]^{-1}dx$ is the coefficient of difficulty CSC introduced in (2.6).

Remark 6.1. In what follows it is sufficient to have $dn^{-1}[1 + o_j(1) + o_n(1)]$ on the right side of (6.4)

Proof of Lemma 6.2. Recall statistic $\check{F}_{-j}(x)$ is defined in line (3.3) and used to estimate function $F_{-j}(x)$ which approximates $F^X(x)$ and is defined in (3.4). Write,

$$\widehat{\theta}_j = n^{-1}\sum_{l=1}^n \frac{(\Delta_l - F_{-j}(Z_l))\varphi_j(Z_l)}{f^Z(Z_l)}$$

$$+ n^{-1}\sum_{l=1}^n \frac{(F_{-j}(Z_l) - \check{F}_{-j}(Z_l))\varphi_j(Z_l)}{f^Z(Z_l)} =: \widehat{\theta}_{1j} + \widehat{\theta}_{2j}. \qquad (6.5)$$

Let us evaluate mean and mean squared error (MSE) of the two terms in turn. Using (2.1) and a straightforward calculation we get for the mean,

$$E_{F^X}\{\widehat{\theta}_{1j}\} = \int_0^1 [F^X(x) - F_{-j}(x)]\varphi_j(x)dx = \int_0^1 F^X(x)\varphi_j(x)dx = \theta_j. \qquad (6.6)$$

We conclude that $\widehat{\theta}_{1j}$ is unbiased estimate of θ_j. Using (2.1) and $\Delta_l^2 = \Delta_l$ we write for MSE,

$$\mathbb{E}_{F^X}\{(\widehat{\theta}_{1j} - \theta_j)^2\} = n^{-1}\Big[\mathbb{E}_{F^X}\Big\{\Big((\Delta_l - F_{-j}(Z_l))\varphi_j(Z_l)[f^Z(Z_l)]^{-1}\Big)^2\Big\} - \theta_j^2\Big]$$

$$= n^{-1}\Big[\mathbb{E}_{F^X}\Big\{(\Delta_l^2 - 2\Delta_j F_{-j}(Z_l) + F_{-j}^2(Z_l))\varphi_j^2(Z_l)[f^Z(Z_l)]^{-2}\Big\} - \theta_j^2\Big]$$

$$= n^{-1}\Big[\int_0^1 [F^X(x) - 2F^X(x)F_{-j}(x) + F_{-j}^2(x)]\varphi_j^2(x)[f^Z(x)]^{-1}dx - \theta_j^2\Big]$$

Now we write $F_{-j}(x)$ as $F^X(x) + [F_{-j}(x) - F^X(x)]$ and continue

$$\mathbb{E}_{F^X}\{(\widehat{\theta}_{1j} - \theta_j)^2\} \leqslant n^{-1}\int_0^1 [F^X(x)(1 - F^X(x))]\varphi_j^2(x)[f^Z(x)]^{-1}dx$$

$$+ n^{-1}\Big|\int_0^1 [-2F^X(x)(F_{-j}(x) - F^X(x))]\varphi_j^2(x)[f^Z(x)]^{-1}dx\Big|$$

$$+ n^{-1}\int_0^1 [F_{-j}^2(x) - (F^X(x))^2]\varphi_j^2(x)[f^Z(x)]^{-1}dx =: v_1 + v_2 + v_3. \qquad (6.7)$$

To evaluate v_1 we use trigonometric formula $2\cos^2(\gamma) = 1 + \cos(2\gamma)$ which implies

$$\varphi_j^2(x) = 1 + 2^{-1/2}\varphi_{2j}(x). \qquad (6.8)$$

We also use differentiability of both $F^X(x)$ and $f^Z(x)$, and relation (2.2.5) in [13]. We get

$$v_1 \leqslant n^{-1}[d + c_j(j+1)^{-1}], \text{ where } \sum_{j=0}^{\infty} c_j^2 < C. \tag{6.9}$$

Note that we need only $c_j^2 < C$ to verify (6.4).

To evaluate v_2 we first note that

$$F^X(x) - F_{-j}(x) = \theta_j \varphi_j(x) I(j \leqslant s) + \sum_{r>s} \theta_r \varphi_r(x). \tag{6.10}$$

Second we make several preliminary calculations. We begin with

$$\theta_j \int_0^1 F^X(x) \varphi_j(x) \varphi_j^2(x) [f^Z(x)]^{-1} dx I(j \leqslant s)$$

$$= \theta_j \int_0^1 F^X(x) [f^Z(x)]^{-1} [\varphi_j(x) + \varphi_j(x) 2^{-1/2} \varphi_{2j}(x)] dx I(j \leqslant s)$$

$$=: \theta_j \kappa_j I(j \leqslant s). \tag{6.11}$$

To evaluate the right side of (6.11) we use $\varphi_i(x)\varphi_{2j}(x) = 2^{-1/2}[\varphi_j(x) + \varphi_{3j}(x)]$ and inequality (2.2.8) in [13]. We get

$$|\theta_j \kappa_j| I(j \leqslant s) \leqslant C(j+1)^{-2} I(j \leqslant s).$$

Our next preliminary calculation is for the second term on the right side of (6.10). Write,

$$\sum_{r>s} \theta_r \int_0^1 F^X(x) \varphi_r(x) \varphi_j^2(x) [f^Z(x)]^{-1} dx =: \sum_{r>s} \theta_r \nu_r. \tag{6.12}$$

Using Cauchy-Schwarz inequality and $F^X \in \mathcal{F}(\alpha, Q)$ we evaluate the right side of (6.12) by

$$|\sum_{r>s} \theta_r \nu_r| \leqslant [\sum_{r>s} \theta_r^2]^{1/2} [\sum_{r>s} \nu_r^2]^{1/2} \leqslant Cs^{-1}. \tag{6.13}$$

Now we have all technical relations needed for the analysis v_2. Using (6.10)-(6.13) yields

$$v_2 \leqslant Cn^{-1}[(j+1)^{-2} + s^{-1}]. \tag{6.14}$$

The last term to evaluate on the right side of (6.7) is v_3. Note that

$$F_{-j}^2(x) = [F^X(x)]^2 - 2F^X(x)[\theta_j \varphi_j(x) I(j \leqslant s) + \sum_{r>s} \theta_r \varphi_r(x)]$$

$$+ [\theta_j \varphi_j(x) I(j \leqslant s) + \sum_{r>s} \theta_r \varphi_r(x)]^2.$$

This yields that

$$v_3 \leqslant C\Big[v_2 + \int_0^1 [\theta_j \varphi_j(x) I(j \leqslant s) + \sum_{r>s} \theta_r \varphi_r(x)]^2 \varphi_j^2(x) [f^Z(x)]^{-1} dx\Big]$$

$$\leqslant C\Big[v_2 + \theta_j^2 I(j \leqslant s) + [\sum_{r>s} |\theta_r|]^2\Big] \leqslant C[(j+1)^{-2} + Cs^{-1}]. \qquad (6.15)$$

Combining (6.9), (6.14) and (6.15) on the right side of (6.7) we get

$$\mathbb{E}_{F^X}\{(\widehat{\theta}_{1j} - \theta_j)^2\} \leqslant n^{-1}d$$

$$+ Cn^{-1}[c_j(j+1)^{-1} + (j+1)^{-2} + s^{-1}], \quad \text{where } \sum_{j=0}^{\infty} c_j^2 < C. \qquad (6.16)$$

This ends our analysis of $\widehat{\theta}_{1j}$. Now we are evaluating the mean and MSE of $\widehat{\theta}_{2j}$ defined in (6.5). First of all a comment is due. This is not a simple task to evaluate this statistic with the required accuracy. On one hand, we have a plain and very strong property

$$\mathbb{E}_{F^X}\Big\{\int_0^1 (\breve{F}_{-j}(z) - F_{-j}(z))^2 dz\Big\} \leqslant \sum_{j=0}^{s} \mathbb{E}\{(\breve{\theta}_j - \theta_j)^2\} \leqslant Csn^{-1}, \qquad (6.17)$$

which holds due to the Parceval identity and Lemma 6.1. On the other hand, this result alone is not sufficient for verification of Lemma 6.2 and making a conclusion that term $\widehat{\theta}_{2j}$ is negligible with respect to the main term $\widehat{\theta}_{1j}$ in (6.5).

To simplify formulas for analysis of $\widehat{\theta}_{2j}$, we need several new notations. Set $S_{-j} := \{0, 1, \ldots, s\}\backslash\{j\}$ for $j \leqslant s$ and $S_{-j} := \{0, 1, \ldots, s\}$ for $j > s$, and $\mathcal{N}_{-l} := \{1, \ldots, n\}\backslash\{l\}$. Using these notations we may rewrite $\breve{F}_{-j}(Z_l)$,

$$\breve{F}_{-j}(Z_l) = n^{-1} \sum_{k \in \mathcal{N}_{-l}} \Delta_k [f^Z(Z_k)]^{-1} \sum_{i \in S_{-j}} \varphi_i(Z_k) \varphi_i(Z_l)$$

$$+ n^{-1} \Delta_l [f^Z(Z_l)]^{-1} \sum_{i \in S_{-j}} \varphi_i^2(Z_l) =: \breve{F}_1(Z_l) + \breve{F}_2(Z_l). \qquad (6.18)$$

Using orthogonality of elements $\{\varphi_j(x), j = 0, 1, \ldots\}$ of the cosine basis we get

$$\mathbb{E}_{F^X}\{[F_{-j}(Z_l) - \breve{F}_1(Z_l)][f^Z(Z_l)]^{-1}\varphi_j(Z_l)|\{(\Delta_k, Z_k), \, k \in \mathcal{N}_{-l}\}\} = 0. \qquad (6.19)$$

Further, using Assumption 1 we get

$$|\mathbb{E}_{F^X}\{\breve{F}_2(Z_l)\varphi_j(Z_l)[f^Z(Z_l))]^{-1}\}|$$

$$= n^{-1}|\mathbb{E}_{F^X}\{\Delta_l [f^Z(Z_l)]^{-2} \sum_{i \in S_{-j}} \varphi_i^2(Z_l)\varphi_j(Z_l)\}| \leqslant Csn^{-1}. \qquad (6.20)$$

Combining the obtained results yields

$$|\mathbb{E}_{F^X}\{\widehat{\theta}_{2j}\}| \leqslant Csn^{-1}.$$

This and (6.6) yields (6.3).

Now we are evaluating the second moment of $\widehat{\theta}_{2j}$. First we note that

$$\mathbb{E}_{F^X}(F_{-j}(Z_l) - \check{F}_1(Z_l))^2 \leqslant Csn^{-1}. \tag{6.21}$$

Second we have

$$\mathbb{E}_{F^X}\left\{\left(n^{-1}\sum_{l=1}^{n}\check{F}_2(Z_l)\varphi_j(Z_l)[f^Z(Z_l)]^{-1}\right)^2\right\} \leqslant Cn^{-2}s^2. \tag{6.22}$$

Third, using notation (6.8) for Z_l and Z_m where $l \neq m$ and $l, m \in \{1, 2, \ldots, n\}$, we write

$$[F_{-j}(Z_l) - \check{F}_{-j}(Z_l)][F_{-j}(Z_m) - \check{F}_{-j}(Z_m)]$$

$$= [F_{-j}(Z_l) - \check{F}_1(Z_l) - \check{F}_2(Z_l)][F_{-j}(Z_m) - \check{F}_1(Z_m) - \check{F}_2(Z_m)]$$

$$= [F_{-j}(Z_l) - \check{F}_1(Z_l)][F_{-j}(Z_m) - \check{F}_1(Z_m)] - [F_{-j}(Z_l) - \check{F}_1(Z_l)]\check{F}_2(Z_m)$$

$$-\check{F}_2(Z_l)[F_{-j}(Z_m) - \check{F}_1(Z_m)] + \check{F}_2(Z_l)\check{F}_2(Z_m)$$

$$=: \check{A}_1 + \check{A}_2 + \check{A}_3 + \check{A}_4. \tag{6.23}$$

This expansion and (6.17) allows us to write for any $1 \leqslant l < m \leqslant n$,

$$\mathbb{E}_{F^X}\{\widehat{\theta}_{2j}^2\} \leqslant n^{-1}C\mathbb{E}_{F^X}\{(F_{-j}(Z_l) - \check{F}_{-j}(Z_l))^2\}$$

$$+C\mathbb{E}_{F_X}\{[\check{A}_1 + \check{A}_2 + \check{A}_3 + \check{A}_4]\varphi_j(Z_l)\} \leqslant Csn^{-2}$$

$$+ C\mathbb{E}_{F_X}\{[\check{A}_1 + \check{A}_2 + \check{A}_3 + \check{A}_4]\varphi_j(Z_l)\varphi_j(Z_m)[f^Z(Z_l)f^Z(Z_m)]^{-1}\}. \tag{6.24}$$

We need an extra notation to analyze the last expectation in (6.24). Following (6.18) we set

$$\check{F}_1(Z_l, -m) := \check{F}_1(Z_l) - n^{-1}\Delta_m[f^Z(Z_m)]^{-1}\sum_{i\in S_{-j}}\varphi_i^2(Z_m)$$

$$=: \check{F}_1(Z_l) - \check{F}_1(Z_l, m). \tag{6.25}$$

The reason for this new decomposition of $\check{F}_1(Z_l)$ is that $\check{F}_1(Z_l, -m)$ does not depend on $(\Delta_l, Z_l, \Delta_m, Z_m)$ and $|\check{F}_1(Z_l, m)| \leqslant Csn^{-1}$ almost sure. Using this notation we write,

$$A_1' := \mathbb{E}_{F^X}\{\check{A}_1\varphi_j(Z_l)\varphi_j(Z_m)[f^Z(Z_l)f^Z(Z_m)]^{-1}\}$$

$$= \mathbb{E}_{F_X}\left\{[F_{-j}(Z_l) - \check{F}_1(Z_l, -m) - \check{F}_1(Z_l, m)][F_{-j}(Z_m) - \check{F}_1(Z_m, -l) - \check{F}_1(Z_m, l)]\right.$$

$$\times \varphi_j(Z_l)\varphi_j(Z_m)[f^Z(Z_l)f^Z(Z_m)]^{-1}\Big\}.$$

To continue evaluation of A_1' we need several technical relations. Using (6.19) we can write,

$$\mathbb{E}_{F_X}\Big\{[F_{-j}(Z_l) - \check{F}_1(Z_l, -m)][F_{-j}(Z_m) - \check{F}_1(Z_m, -l)]$$

$$\times \varphi_j(Z_l)\varphi_j(Z_m)[f^Z(Z_l)f^Z(Z_m)]^{-1}\Big\} = 0.$$

Recall the rough inequality $|[\check{F}_1(Z_l, m)| \leqslant Csn^{-1}$, and using the Cauchy-Schwarz inequality together with (6.21) we get

$$|\mathbb{E}_{F_X}\{[F_{-j}(Z_l) - \check{F}_1(Z_l, -m)]\check{F}_1(Z_m, l)\}|$$

$$\leqslant [\mathbb{E}_{F_X}\{[F_{-j}(Z_l) - \check{F}_1(Z_l, -m)]^2\}E_{F^X}\{[\check{F}_1(Z_m, l)]^2\}]^{1/2}$$

$$\leqslant C[(sn^{-1}(s^2n^{-2}]^{1/2} \leqslant Cs^{3/2}n^{-3/2}.$$

We conclude that $A_1' \leqslant Cs^{3/2}n^{-3/2}$. Analysis of $\check{A}_2$ is similar. Using the Cauchy-Schwarz inequality we get

$$\mathbb{E}_{F^X}\{|\check{A}_2|\} \leqslant [\mathbb{E}_{F_X}\{(F_{-j}(Z_l) - \check{F}_1(Z_l))^2\}\mathbb{E}_{F^X}\{[\check{F}_2(Z_m)]^2\}]^{1/2} \leqslant Cs^2n^{-3/2}.$$

Due to symmetry the same inequality holds for $\check{A}_3$. Finally, we have $\mathbb{E}_{F^X}\{|\check{A}_4|\} \leqslant Cs^2n^{-2}$. Combining the above-presented relations we get

$$\mathbb{E}_{F^X}\{\widehat{\theta}_{2j}^2\} \leqslant Cs^2n^{-3/2}.$$

As we see, the component $\widehat{\theta}_{2j}$ is indeed negligible with respect to $\widehat{\theta}_{1j}$.

Lemma 6.2 is proved.

Corollary 6.1. There are three technical conclusions from the above-presented proof. The first one is that according to (6.5) we have a representation $\widehat{\theta}_j = \widehat{\theta}_{j1} + \widehat{\theta}_{j2}$ where the second term is negligible and may be skipped. Using (6.19), a similar conclusion can be made for the pilot cdf estimate $\check{F}_{-j}(Z_l) = \check{F}_1(Z_l) + \check{F}_2(Z_l)$ where the second term $\check{F}_2(Z_l)$ is negligible. Further, note that the pivotal estimate is based on at most s estimated Fourier coefficients, and then for any integer k we have the following directly verified rough inequality

$$\sup_{F^X \in \mathcal{F}(\alpha, Q)} \mathbb{E}_{F^X}\{[\check{F}_{-j}(Z_l) - F_{-j}(Z_l)]^{2k}\} \leqslant C_k s^{2k} n^{-k}, \ \ C_k < \infty. \tag{6.26}$$

The reader, who wants to understand the following proof without going into detailed analysis of negligible terms, may replace $\widehat{\theta}_j$ by $\widehat{\theta}_{1j}$ and $\check{F}_{-j}^X$ by F_{-j}^X. Finally, in Lemma 6.2 the upper bound (6.4) may be replaced by the right side of (6.16). While we do not need that accuracy, it is an interesting upper bound that sheds light on the Fourier estimator which yields sharp minimax cdf estimation. Recall that in nonparametric curve estimation literature such a Fourier estimator is called efficient.

Now we are ready to begin analysis of the MISE of $\widetilde{F}^X(x, f^Z)$. Using the Parseval identity we can write,

$$E_{F^X}\left\{ \int_0^1 (\widetilde{F}(x, f^Z) - F^X(x))^2 dx \right\}$$

$$= \sum_{j=0}^s \mathbb{E}_{F^X}\left\{ (\widehat{\theta}_j I(\widehat{\theta}_j^2 > 2\widetilde{d}sn^{-1}) - \theta_j)^2 \right\}$$

$$+ \sum_{k=1}^{r_n} \sum_{j \in B_k} \mathbb{E}_{F^X}\left\{ \left(\min(1, \widetilde{\Theta}_k \widehat{\Theta}_k^{-1}) I(\widetilde{\Theta}_k > 10 s^{-1} n^{-1}) \widehat{\theta}_j - \theta_j \right)^2 \right\}$$

$$+ \sum_{k > r_n} L_k \Theta_k =: A_1(F^X) + A_2(F^X) + A_3(F^X). \tag{6.27}$$

The three terms on the right side of (6.27) correspond to the MISE components on low, middle and high frequencies, and they are explored in turn. Using $\widetilde{d} \leqslant 3s$ and Lemma 6.2 we get for a jth term in $A_1(F^X)$,

$$\mathbb{E}_{F^X}\left\{ [\widehat{\theta}_j I(\widehat{\theta}_j^2 > 2\widetilde{d}sn^{-1}) - \theta_j]^2 \right\}$$

$$\leqslant \mathbb{E}_{F^X}\left\{ (\widehat{\theta}_j - \theta_j)^2 \right\} + \theta_j^2 \mathbb{P}_{F^X}\left(\widehat{\theta}_j^2 \leqslant 2\widetilde{d}sn^{-1} \right)$$

$$\leqslant Cn^{-1} + \theta_j^2 \mathbb{P}_{F^X}\left(\widehat{\theta}_j^2 \leqslant 6s^2 n^{-1} \right) \leqslant Cn^{-1} + 12s^2 n^{-1}$$

$$+ \theta_j^2 \mathbb{P}_{F^X}\left(\widehat{\theta}_j^2 \leqslant 6s^2 n^{-1}, \theta_j^2 > 12s^2 n^{-1} \right). \tag{6.28}$$

Let us consider the last term on the right side of (6.18). Using the Chebyshev inequality and Lemma 6.2 we get,

$$\theta_j^2 \mathbb{P}_{F^X}\left(\widehat{\theta}_j^2 \leqslant 6s^2 n^{-1}, \theta_j^2 > 12s^2 n^{-1} \right)$$

$$\leqslant \theta_j^2 \mathbb{P}_{F^X}\left(\theta_j^2 - \widehat{\theta}_j^2 > \theta_j^2/2, \theta_j^2 > 12s^2 n^{-1} \right)$$

$$\leqslant \theta_j^2 \mathbb{P}_{F^X}\left(|\theta_j - \widehat{\theta}_j|(2|\theta_j|) > \theta_j^2/2, \theta_j^2 > 12s^2 n^{-1} \right)$$

$$\leqslant C\theta_j^2 \frac{\mathbb{E}_{F^X}\{(\theta_j - \widehat{\theta}_j)^2\}}{\theta_j^2} \leqslant Cn^{-1}. \tag{6.29}$$

Using (6.29) in (6.28) we conclude that

$$A_1(F^X) < Cs^3 n^{-1} = o_n(1) n^{-2\alpha/(2\alpha+1)}. \tag{6.30}$$

Corollary 6.2. As we see, the MISE of the low-frequency part of the proposed cdf estimator is negligible with respect to the verified rate $n^{-2\alpha/(2\alpha+1)}$. Another important conclusion is that there is a wide choice of thresholds and a boundary frequency (which is currently s) between the low and middle frequency

components of a blockwise estimator that yield sharp-minimax estimation. The currently used boundary frequency s is chosen based on recommendation of [13], because of its simplicity, and using a numerical study of experiments similar to those in Section 5. Discussion of more complicated procedures can be found in [13].

Now we turn our attention to the second term $A_2(F^X)$ in (6.27). This term is the MISE of the block-shrinkage part of the proposed cdf estimator.

Lemma 6.3. *Let Assumption 1 hold. Then*

$$\sup_{F^X \in \mathcal{F}(\alpha,Q)} \left\{ [n\mathcal{J}(\alpha,Q,d,1)]^{2\alpha/(2\alpha+1)} A_2(F^X) \right\} \leqslant (1 + o_n(1)). \qquad (6.31)$$

Proof of Lemma 6.3. We begin with a remark that the studied risk $A_2(F^X)$ is the MISE of a blockwise estimator

$$\widetilde{\psi}(x) := \sum_{k=1}^{r_n} \sum_{j \in B_k} \min(1, \widetilde{\Theta}_k \widehat{\Theta}_k^{-1}) I(\widetilde{\Theta}_k > 10s^{-1}n^{-1}) \widehat{\theta}_j \varphi_j(x) \qquad (6.32)$$

which estimates a function

$$\psi(x) := \sum_{k=1}^{r_n} \sum_{j \in B_k} \theta_j \varphi_j(x). \qquad (6.33)$$

Note that (6.33) is the intermediate frequency component of an underlying cdf $F^X(x)$.

Analysis of the MISE of $\widetilde{\psi}(x)$ includes several steps. The first step is to evaluate MISE of an oracle-estimator

$$\widehat{\psi}^*(x) := \sum_{k=1}^{r_n} \sum_{j \in B_k} \mu_k \widehat{\theta}_j \varphi_j(x), \qquad (6.34)$$

where

$$\mu_k = \frac{\Theta_k}{\Theta_k + dn^{-1}} \qquad (6.35)$$

is oracle's smoothing coefficient,

$$\Theta_k := L_k^{-1} \sum_{j \in B_k} \theta_j^2 \qquad (6.36)$$

is the Sobolev statistic introduced in Section 3, and d is the coefficient of difficulty (2.6). The Parseval identity allows us to express the oracle's MISE as

$$E\left\{ \int_0^1 (\widehat{\psi}^*(x) - \psi(x))^2 dx \right\} = \sum_{k=1}^{r_n} \sum_{j \in B_k} E\{(\mu_k \widehat{\theta}_j - \theta_j)^2\}$$

$$= \sum_{k=1}^{K} \sum_{j \in B_k} [\mu_k^2 E\{(\widehat{\theta}_j - \theta_j)^2\} + (\mu_k - 1)^2 \theta_j^2 - 2\mu_k(1 - \mu_k)\theta_j E\{\widehat{\theta}_j - \theta_j\}]. \quad (6.37)$$

Using Lemma 6.2 we evaluate terms in (6.37). Write,

$$E\{\int_0^1 (\widehat{\psi}^*(x) - \psi(x))^2 dx\} \leqslant \Big[\sum_{k=1}^{r_n} L_k[\mu_k^2 dn^{-1} + \frac{d^2 n^{-2}\Theta_k}{(\Theta_k + dn^{-1})^2}]\Big]$$

$$+ \Big[\sum_{k=1}^{r_n} \mu_k^2 dn^{-1} \sum_{j \in B_k} C[(j+1)^{-1} + s^{-1}]\Big]$$

$$+ \Big[2| \sum_{k=1}^{r_n} \mu_k dn^{-1}(\Theta_k + dn^{-1})^{-1} \sum_{j \in B_k} \theta_j E\{\widehat{\theta}_j - \theta_j\}|\Big] =: T_1 + T_2 + T_3. \quad (6.38)$$

We are considering the three terms on the right side of (6.38) in turn. The first term is simplified into

$$T_1 = \sum_{k=1}^{r_n} L_k[\mu_k^2 dn^{-1} + \frac{d^2 n^{-2}\Theta_k}{(\Theta_k + dn^{-1})^2}] = \sum_{k=1}^{r_n} L_k\mu_k dn^{-1}.$$

What we see is the MISE of a classical blockwise oracle studied in [11]. Line (3.1) of that paper yields

$$\sup_{F^X \in \mathcal{F}(\alpha,Q)} \Big([n\mathcal{J}(\alpha, Q, d, 1)]^{2\alpha/(2\alpha+1)} T_1\Big)$$

$$= \sup_{F^X \in \mathcal{F}(\alpha,Q)} \Big([n\mathcal{J}(\alpha, Q, d, 1)]^{2\alpha/(2\alpha+1)} \sum_{k=1}^{r_n} L_k\mu_k dn^{-1}\Big) \leqslant 1 + o_n(1)). \quad (6.39)$$

To consider T_2 we note that $\mu_k \leqslant 1$. This, together with $(j+1)^{-1} \leqslant s^{-1}$ for $j \in B_k$ and (6.39), yield that $T_2 \leqslant Cs^{-1}n^{-2\alpha/(2\alpha+1)} = o_n(1)n^{-2\alpha/(2\alpha+1)}$.

The third term T_3 is far from being simple for evaluation because of the factor $(\Theta_k + dn^{-1})^{-1}$ which may be of order n. The idea of evaluating T_3 is to correctly bound the sum in j. To do that we use Lemma 6.2 and Cauchy inequality. Write,

$$| \sum_{j \in B_k} \theta_j \mathbb{E}\{\widehat{\theta}_j - \theta_j\}| \leqslant s^{-1} \sum_{j \in B_k} \theta_j^2 + s \sum_{j \in B_k} [\mathbb{E}\{\widehat{\theta}_j - \theta_j\}]^2$$

$$\leqslant L_k[s^{-1}\Theta_k + Cs^3 n^{-2}].$$

Now note that

$$\max\Big(\frac{dn^{-1}}{\Theta_k + dn^{-1}}, \frac{\Theta_k}{\Theta_k + dn^{-1}}\Big) \leqslant 1.$$

Using this inequality and (6.39) we write,

$$\sup_{F^X \in \mathcal{F}(\alpha,Q)} T_3 = \sup_{F^X \in \mathcal{F}(\alpha,Q)} 2| \sum_{k=1}^{r_n} \mu_k dn^{-1}(\Theta_k + dn^{-1})^{-1} \sum_{j \in B_k} \theta_j E\{\widehat{\theta}_j - \theta_j\}|$$

$$\leqslant 2 \sup_{F^X \in \mathcal{F}(\alpha, Q)} \Big[s^{-1} \sum_{k=1}^{r_n} L_k \mu_k dn^{-1} \frac{\Theta_k}{\Theta_k + dn^{-1}}$$

$$+ C s^3 n^{-1} \sum_{k=1}^{r_n} L_k \mu_k n^{-1} \frac{dn^{-1}}{\Theta_k + dn^{-1}} \Big]$$

$$= o_n(1) n^{-2\alpha/(2\alpha+1)}. \tag{6.40}$$

Combining the bounds for T_1, T_2 and T_3 in (6.38) we conclude that the oracle $\widehat{\psi}^*(x)$ is sharp-minimax. This ends step 1 of the analysis of $A_2(F^X)$.

Step 2 is to introduce a new oracle-estimator $\widetilde{\psi}^*(x)$ which is more convenient for studying the blockwise-shrinkage estimate $\widetilde{\psi}(x)$ defined in (6.32) and which sheds light on the chosen blockwise smoothing. Set

$$\widetilde{\psi}^*(x) := \sum_{k=1}^{r} \sum_{j \in B_k} \lambda_k \widehat{\theta}_j \varphi_j(x), \quad \lambda_k := \frac{\Theta_k}{\mathbb{E}\{\widehat{\Theta}_k\}}. \tag{6.41}$$

Here $\widehat{\Theta}_k = L_k^{-1} \sum_{j \in B_k} \widehat{\theta}_j^2$ was introduced in (3.7), and recall that it is used in the denominator of the smoothing ratio $\widetilde{\Theta}_k / \widehat{\Theta}_k$ of the proposed estimate (3.9). To analyze MISE of the new oracle, the Parseval identity allows us to consider a particular block B_k. Write for any constant $\lambda \in [0, 1]$,

$$\sum_{j \in B_k} \mathbb{E}\{(\lambda \widehat{\theta}_j - \theta_j)^2\} = \sum_{j \in B_k} \mathbb{E}\{\lambda^2 \widehat{\theta}_j^2 - 2\lambda \widehat{\theta}_j \theta_j + \theta_j^2\}.$$

Using Lemma 6.2 we continue,

$$\sum_{j \in B_k} \mathbb{E}\{(\lambda \widehat{\theta}_j - \theta_j)^2\}$$

$$= \lambda^2 L_k \mathbb{E}\{\widehat{\Theta}_k\} - (2\lambda - 1) L_k \Theta_k - 2\lambda \sum_{j \in B_k} \theta_j c_j', \quad |c_j'| \leqslant C s n^{-1}. \tag{6.42}$$

We begin with evaluating the last sum on the right side of (6.32). For $F^X \in \mathcal{F}(\alpha, Q)$ we can write using the Cauchy-Schwarz inequality,

$$\sum_{k=1}^{r_n} |2\lambda \sum_{j \in B_k} \theta_j c_j'| \leqslant \sum_{k=1}^{r_n} 2\lambda \sum_{j \in B_k} |\theta_j c_j'| \leqslant C s n^{-1} \sum_{k=1}^{r_n} \sum_{j \in B_k} |\theta_j|$$

$$\leqslant C s n^{-1} \Big[\sum_{j > s} j^{2\alpha} \theta_j^2 \sum_{j > s} j^{-2\alpha} \Big]^{1/2} \leqslant C n^{-1}.$$

For the first two terms on the right side of (6.42) we note that λ^* that minimizes $\lambda^2 L_k \mathbb{E}\{\widehat{\Theta}_k\} - (2\lambda - 1) L_k \Theta_k$ is $\lambda^* = \Theta_k / \mathbb{E}\{\widehat{\Theta}_k\}$. Thus $\lambda^* = \lambda_k$ introduced in (6.41). This and sharp minimaxity of the oracle $\widehat{\psi}^*(x)$ yields that

$$\sup_{F^X \in \mathcal{F}(\alpha, Q)} \mathcal{J}(\alpha, Q, d, 1) \mathbb{E}_{F_X} \Big\{ \int_0^1 (\widetilde{\psi}^*(x) - \psi(x))^2 dx \Big\} \leqslant (1 + o_n(1)). \tag{6.43}$$

We conclude that the oracle $\widetilde{\psi}^*(x)$ is sharp-minimax.

To continue evaluation of the term $A_2(F^X)$ on the right side of (6.27) we need several more technical lemmas. These lemmas are also of interest on their own because they shed light on the Sobolev statistics introduced in Section 3 and also on blockwise oracles.

Lemma 6.4. *Statistics $\widetilde{\Theta}_k$ and $\widehat{\Theta}_k$, defined in (3.6) and (3.7), satisfy the following relation*

$$\widehat{\Theta}_k = \widetilde{\Theta}_k(1 - n^{-1}) + n^{-1}\Big\{L_k^{-1}\sum_{j \in B_k}\Big[n^{-1}\sum_{l=1}^{n}\frac{(\Delta_l - \check{F}_{-j}(Z_l))^2\varphi_j^2(Z_l)}{(f^Z(Z_l))^2}\Big]\Big\}$$

$$=: \widetilde{\Theta}_k(1 - n^{-1}) + n^{-1}\widetilde{d}_k. \tag{6.44}$$

If Assumption 1 holds and $F^X \in \mathcal{F}(\alpha, Q)$ then

$$|\mathbb{E}_{F^X}\{\widetilde{d}_k\} - d| \leqslant Cs^{-1} \tag{6.45}$$

and

$$\mathbb{E}_{F^X}\{(\widetilde{d}_k - d)^2\} = o_n(1)s^{-1}, \tag{6.46}$$

where d is the CSC coefficient of difficulty (2.6).

Proof of Lemma 6.4. We begin with verification of (6.44). Using (3.6) and (3.7) we write,

$$\widehat{\Theta}_k = L_k^{-1}\sum_{j \in B_k}n^{-2}\sum_{l,l'=1}^{n}\frac{(\Delta_l - \check{F}_{-j}(Z_l))\varphi_j(Z_l)(\Delta_{l'} - \check{F}_{-j}(Z_{l'}))\varphi_j(Z_{l'})}{f^Z(X_l)f^Z(X_{l'})}$$

$$= \widetilde{\Theta}_k(1 - n^{-1}) + n^{-1}L_k^{-1}\sum_{j \in B_k}\Big[n^{-1}\sum_{l=1}^{n}\frac{(\Delta_l - \check{F}_{-j}(Z_l))^2\varphi_j^2(Z_l)}{(f^Z(Z_l))^2}\Big].$$

This proves (6.34).

The verified inequalities are rough but sufficient for our purposes. We begin with (6.45). Write,

$$\mathbb{E}\{\widetilde{d}_k\} = L_k^{-1}\sum_{j \in B_k}\mathbb{E}\Big\{\frac{(\Delta_l - \check{F}_{-j}(Z_l))^2\varphi_j^2(Z_l)}{(f^Z(Z_l))^2}\Big\}$$

$$= L_k^{-1}\sum_{j \in B_k}\mathbb{E}\Big\{\frac{(\Delta_l - F_{-j}(Z_l) + F_{-j}(Z_l) - \check{F}_{-j}(Z_l))^2\varphi_j^2(Z_l)}{(f^Z(Z_l))^2}\Big\}$$

$$= L_k^{-1}\sum_{j \in B_k}\Big[\mathbb{E}\Big\{\frac{(\Delta_l - F_{-j}(Z_l))^2\varphi_j^2(Z_l)}{(f^Z(Z_l))^2}\Big\} + \mathbb{E}\Big\{\frac{(F_{-j}(Z_l) - \check{F}_{-j}(Z_l))^2\varphi_j^2(Z_l)}{(f^Z(Z_l))^2}\Big\}$$

$$+2\mathbb{E}\Big\{\frac{(\Delta_l - F_{-j}(Z_l))(F_{-j}(Z_l) - \check{F}_{-j}(Z_l))\varphi_j^2(Z_l)}{(f^Z(Z_l))^2}\Big\}\Big]$$

$$= L_k^{-1} \sum_{j \in B_k} [a_{1j} + a_{2j} + a_{3j}].$$

The term a_{1j} was already studied as a component of $\mathbb{E}_{F^X}(\hat\theta_{1j} - \theta_j)^2\}$, see the line below (6.6) and then (6.16). Using notation of (6.16) we can write that

$$L_k^{-1} \sum_{j \in B_k} a_{1j} \leqslant C[L_k^{-1} \sum_{j \in B_k} [c_j j^{-1} + j^{-2} + s^{-1}] < Cs^{-1}.$$

The last inequality is valid because $c_j < C$ and $j > s$ whenever $j \in B_k$. Inequality (6.26) implies that $L_k^{-1} \sum_{j \in B_k} a_{2j} \leqslant Cs^2 n^{-1}$, and similarly $L_k^{-1} \sum_{j \in B_k} |a_3| \leqslant Csn^{-1/2}$. Inequality (6.45) is verified.

Now let us check inequality (6.46). Write using (6.45),

$$\mathbb{E}_{F^X}\{(\tilde d_k - d)^2\} = E_{F^X}\{[\tilde d_k - \mathbb{E}_{F^X}\{\tilde d_k\} + \mathbb{E}_{F^X}\{\tilde d_k\} - d]^2\}$$

$$\leqslant 2\mathbb{E}_{F^X}\{(\tilde d_k - \mathbb{E}_{F^X}\{\tilde d_k\})^2\} + o_n(1)s^{-1}.$$

To evaluate the expectation we use the expansion

$$\check F_{-j}(Z_l) = F_{-j}(Z_l) + (\check F_{-j}(Z_l) - F_{-j}(Z_l)),$$

which allows us to write for a factor in $\tilde d_k$,

$$[\Delta_l - \check F_{-j}(Z_l)]^2 = [\Delta_l - F_{-j}(Z_l)]^2$$

$$- \{(\check F_{-j}(Z_l) - F_{-j}(Z_l))[2(\Delta_l - \check F_{-j}(Z_l)) - (\check F_{-j}(Z_l) - F_{-j}(Z_l))]\}. \quad (6.47)$$

Let us look at terms on the right side of (6.47). The term $[\Delta_l - F_{-j}(Z_l)]^2$ depends only on pair (Δ_l, Z_l), and hence the central second moment of a sample mean

$$n^{-1} \sum_{l=1}^n \left[L_k^{-1} \sum_{j \in B_k} \frac{(\Delta_l - F_{-j}(Z_l))^2 \varphi_j^2(Z_l)}{(f^Z(Z_l))^2} \right]$$

is proportional to n^{-1}. Further, the term in curly brackets in (6.47) is negligibly small with respect to s^{-1} due to inequality $\mathbb{E}_{F^X}\{(\check F_{-j}(Z_l) - F_{-j}(Z_l))^{2k}\} \leqslant Cs^{2k}n^{-k}$, see (6.16). Inequality (6.36) is verified. Lemma 6.4 is proved.

Lemma 6.5. *Let Assumption 1 hold. Recall notations*

$$\tilde\psi(x) := \sum_{k=1}^{r_n} \min(1, \frac{\tilde\Theta_k}{\widehat\Theta_k}) I(\tilde\Theta_k > 10s^{-1}n^{-1}) \sum_{j \in B_k} \hat\theta_j \varphi_j(x), \quad (6.48)$$

and

$$\widehat\psi^*(x) := \sum_{k=1}^{r_n} \frac{\Theta_k}{\Theta_k + dn^{-1}} \sum_{j \in B_k} \hat\theta_j \varphi_j(x), \quad \Theta_k := L_k^{-1} \sum_{j \in B_k} \theta_j^2. \quad (6.49)$$

Then

$$\sup_{F^X \in \mathcal{F}(\alpha,Q)} \mathbb{E}\{\int_0^1 (\widetilde{\psi}(x) - \widehat{\psi}^*(x))^2 dx\} = o_n(1)n^{-2\alpha/(2\alpha+1)}. \qquad (6.50)$$

Proof of Lemma 6.5. The Parseval identity yields,

$$\int_0^1 (\widetilde{\psi}(x) - \widehat{\psi}^*(x))^2 dx$$

$$= \sum_{k=1}^{r_n} \left(\min(1, \frac{\widetilde{\Theta}_k}{\widehat{\Theta}_k})I(\widetilde{\Theta}_k > 10s^{-1}n^{-1}) - \frac{\Theta_k}{\Theta_k + dn^{-1}} \right)^2 \sum_{j \in B_k} \widehat{\theta}_j^2$$

$$= \sum_{k=1}^{r_n} \left(\min(1, \frac{\widetilde{\Theta}_k}{\widehat{\Theta}_k}) - \frac{\Theta_k}{\Theta_k + dn^{-1}} \right)^2 \sum_{j \in B_k} \widehat{\theta}_j^2 I(\widetilde{\Theta}_k > 10s^{-1}n^{-1})$$

$$+ \sum_{k=1}^{r_n} \left(\frac{\Theta_k}{\Theta_k + dn^{-1}} \right)^2 \sum_{j \in B_k} \widehat{\theta}_j^2 I(\widetilde{\Theta}_k \leqslant 10s^{-1}n^{-1}).$$

Now note that $\Theta_k/(\Theta_k + dn^{-1}) < 1$ and $\sum_{j \in B_k} \widehat{\theta}_j^2 = L_k \widehat{\Theta}_k$. Using these facts we continue,

$$\int_0^1 (\widetilde{\psi}(x) - \widehat{\psi}^*(x))^2 dx$$

$$\leqslant \sum_{k=1}^{r_n} \left(\frac{\widetilde{\Theta}_k}{\widehat{\Theta}_k} - \frac{\Theta_k}{\Theta_k + dn^{-1}} \right)^2 L_k \widehat{\Theta}_k I(\widetilde{\Theta}_k > 10s^{-1}n^{-1})$$

$$+ \sum_{k=1}^{r_n} \left(\frac{\Theta_k}{\Theta_k + dn^{-1}} \right)^2 L_k \widehat{\Theta}_k I(\widetilde{\Theta}_k \leqslant 10s^{-1}n^{-1})$$

$$=: D_1 + D_2. \qquad (6.51)$$

We begin with analysis of D_1. Using Lemma 6.4 we may write,

$$D_1 = \sum_{k=1}^{r_n} \frac{L_k(\widetilde{\Theta}_k dn^{-1} - \Theta \widetilde{d}_k n^{-1})^2}{\widehat{\Theta}_k(\Theta_k + dn^{-1})^2} I(\widetilde{\Theta}_k > 10s^{-1}n^{-1})$$

$$= \sum_{k=1}^{r_n} \frac{L_k[(\widetilde{\Theta}_k - \Theta_k)dn^{-1} + \Theta_k(d - \widetilde{d}_k)n^{-1}]^2}{(\widetilde{\Theta}_k(1 - n^{-1}) + \widetilde{d}_r n^{-1})(\Theta_k + dn^{-1})^2} I(\widetilde{\Theta}_k > 10s^{-1}n^{-1})$$

$$\leqslant C \sum_{k=1}^{r_n} \frac{sL_k n^{-2}(\widetilde{\Theta}_k - \Theta_k)^2}{n^{-1}(\Theta_k + dn^{-1})^2} I(\widetilde{\Theta}_k > s^{-1}n^{-1})$$

$$+ C \sum_{k=1}^{r_n} \frac{sL_k \Theta_k^2 n^{-2}(d - \widetilde{d}_k)^2}{n^{-1}(\Theta_k + dn^{-1})^2} I(\widetilde{\Theta}_k > s^{-1}n^{-1}). \qquad (6.52)$$

Now we evaluate expectations of terms on the right side of (6.52). To do that we need the following technical result that will be proved later.

Lemma 6.6. *Let Assumption 1 hold and $F^X \in \mathcal{F}(\alpha, Q)$. Then*

$$\mathbb{E}_{F^X}\{(\widetilde{\Theta}_k - \Theta_k)^2\} \leqslant C L_k^{-1} n^{-2}(\Theta_k + n^{-1}). \tag{6.53}$$

Using (6.46) and (6.53) to evaluate expectation of the right side of (6.52) we get,

$$\mathbb{E}_{F^X}\{D_1\} \leqslant C \sum_{k=1}^{r_n} \frac{sL_k n^{-2} E_{F^X}\{(\widetilde{\Theta}_k - \Theta_k)^2\}}{n^{-1}(\Theta_k + dn^{-1})^2}$$

$$+ C \sum_{k=1}^{r_n} \frac{sL_k \Theta_k^2 n^{-2} \mathbb{E}_{F^X}\{(d - \widetilde{d}_k)^2\}}{n^{-1}(\Theta_k + dn^{-1})^2}$$

$$\leqslant C \sum_{k=1}^{r_n} \frac{sL_k n^{-2} L_k^{-1} n^{-1}(\Theta_k + n^{-1})}{n^{-1}(\Theta_k + dn^{-1})^2}$$

$$+ o_n(1) \sum_{k=1}^{r_n} \frac{ss^{-1}L_k \Theta_k^2 n^{-2}}{n^{-1}(\Theta_k + dn^{-1})^2}$$

$$\leqslant C s r_n n^{-1} + o_n(1)\Big[n^{-1}\sum_{k=1}^{n} L_k \frac{\Theta_k}{\Theta_k + dn^{-1}}\Big]$$

$$= o_n(1)n^{-2\alpha/(2\alpha+1)}. \tag{6.54}$$

In the last line we used $sr_n \leqslant Cs^4$ and (6.39).

Now we are estimating the expectation of D_2. Write using (6.44),

$$\mathbb{E}_{F^X}\{D_2\} = \sum_{k=1}^{r_n} \Big(\frac{\Theta_k}{\Theta_k + dn^{-1}}\Big)^2 L_k \mathbb{E}_{F^X}\{(\widetilde{\Theta}_k + \widetilde{d}_k n^{-1})I(\widetilde{\Theta}_k \leqslant 10s^{-1}n^{-1})\}$$

$$\leqslant 10s^{-1}n^{-1}\sum_{k=1}^{r_n} L_k \frac{\Theta_k}{\Theta_k + dn^{-1}}$$

$$+ dn^{-1}\sum_{k=1}^{r_n} L_k \frac{\Theta_k^2}{(\Theta_k + dn^{-1})^2} \mathbb{E}_{F_X}\{I(\widetilde{\Theta}_k \leqslant 10s^{-1}n^{-1})\}$$

$$+ n^{-1}\sum_{k=1}^{r_n} L_k \frac{\Theta_k}{\Theta_k + dn^{-1}} \mathbb{E}_{F^X}\{|\widetilde{d}_k - d|\}.$$

Using Lemma 6.4 and (6.39) we get,

$$\sup_{F^X \in \mathcal{F}(\alpha, Q)} \mathbb{E}_{F^X}\{D_2\} = o_n(1)n^{-2\alpha/(2\alpha+1)}$$

$$+ dn^{-1} \sum_{k=1}^{r_n} L_k \frac{\Theta_k^2}{(\Theta_k + dn^{-1})^2} \mathbb{E}_{F^X} \{I(\widetilde{\Theta}_k \leqslant 10s^{-1}n^{-1})\}. \qquad (6.55)$$

To evaluate the expectation on the right side of (6.55) we use the Chebyshev inequality and Lemma 6.6. Write,

$$\mathbb{E}_{F^X} \{I(\widetilde{\Theta}_k \leqslant 10s^{-1}n^{-1})\}$$

$$\leqslant \mathbb{E}_{F^X} \{I(\Theta_k > 20s^{-1}n^{-1})I(\widetilde{\Theta}_k \leqslant 10s^{-1}n^{-1})\} + I(\Theta_k \leqslant 20s^{-1}n^{-1})$$

$$\leqslant C \frac{L_k^{-1}n^{-1}(\Theta_k + n^{-1})}{[\Theta_k/2]^2} I(\Theta_k > 20s^{-1}n^{-1}) + I(\Theta_k \leqslant 20s^{-1}n^{-1}). \qquad (6.56)$$

Now note that $x/(x + dn^{-1})$ is increasing function in $x > 0$. Using this observation and (6.56) in (6.55) we conclude that

$$\mathbb{E}_{F^X} \{D_2\} \leqslant o_n(1)n^{-2\alpha/(2\alpha+1)}$$

$$+ Cn^{-1} \sum_{k=1}^{r_n} \frac{n^{-1}}{\Theta_k + dn^{-1}} + Cn^{-1} \sum_{k=1}^{r_n} L_k \frac{\Theta_k}{\Theta_k + dn^{-1}} \frac{s^{-1}n^{-1}}{(s^{-1} + d)n^{-1}}$$

$$= o_n(1)n^{-2\alpha/(2\alpha+1)}. \qquad (6.57)$$

In the last relation we used the earlier mentioned inequality $r_n < Cs^2$.

Using (6.54) and (6.57) in (6.51) verifies Lemma 6.5 given validity of Lemma 6.6.

Proof of Lemma 6.6. Consider a fixed $k \in \{1, 2, \ldots, r_n\}$ and introduce an oracle

$$\widetilde{U}_k := \frac{2}{L_k n(n-1)}$$

$$\times \sum_{1 \leqslant l < l' \leqslant n} \sum_{j \in B_k} \frac{(\Delta_l - F_{-j}(Z_l))\varphi_j(Z_l)(\Delta_{l'} - F_{-j}(Z_{l'}))\varphi_j(Z_{l'})}{f^Z(Z_l)f^Z(Z_{l'})}. \qquad (6.58)$$

This oracle is a U-statistic which is unbiased estimate of Θ_k and a benchmark for the studied statistic

$$\widetilde{\Theta}_k := \frac{2}{L_k n(n-1)}$$

$$\times \sum_{1 \leqslant l < l' \leqslant n} \sum_{j \in B_k} \frac{(\Delta_l - \check{F}_{-j}(Z_l))\varphi_j(Z_l)(\Delta_{l'} - \check{F}_{-j}(Z_{l'}))\varphi_j(Z_{l'})}{f^Z(Z_l)f^Z(Z_{l'})}. \qquad (6.59)$$

While it is possible to analyze variance of $\widetilde{U}_k$ directly following [11], it is faster and simpler to follow a standard methodology of calculating moments of a U-statistic. To do that, for the reader's convenience we will use terminology, notation and results of the book [30]. First of all, note that the U-statistic $\widetilde{U}_k$ is based on a symmetric kernel

$$h_k((\Delta_l, Z_l), (\Delta_{l'}, Z_{l'}))$$

$$:= L_k^{-1} \sum_{j \in B_k} \frac{(\Delta_l - F_{-j}(Z_l))\varphi_j(Z_l)(\Delta_{l'} - F_{-j}(Z_{l'}))\varphi_j(Z_{l'})}{f^Z(Z_l)f^Z(Z_{l'})}. \tag{6.60}$$

A relation between the studied statistic and the oracle is straightforward,

$$\widetilde{\Theta}_k = \widetilde{U}_k$$

$$+ \frac{4}{n(n-1)L_k} \sum_{1 \leqslant l < l' \leqslant n} \sum_{j \in B_k} \frac{(\Delta_l - F_{-j}(Z_l))\varphi_j(Z_l)(F_{-j}(Z_{l'}) - \check{F}_{-j}(Z_{l'}))\varphi_j(Z_{l'})}{f^Z(Z_l)f^Z(Z_{l'})}$$

$$+ \frac{2}{n(n-1)L_k}$$

$$\times \sum_{1 \leqslant l < l' \leqslant n} \sum_{j \in B_k} \frac{(F_{-j}(Z_l) - \check{F}_{-j}(Z_l))\varphi_j(Z_l)(F_{-j}(Z_{l'}) - \check{F}_{-j}(Z_{l'}))\varphi_j(Z_{l'})}{f^Z(Z_l)f^Z(Z_{l'})}$$

$$=: \widetilde{U}_k + T_1 + T_2. \tag{6.61}$$

Recall that for considered $j \in B_k$ neither $F_{-j}(Z)$ nor $\check{F}_{-j}(Z)$ depend on j, but the notation is still useful because it reminds us that

$$\int_0^1 F_{-j}(z)\varphi_j(z)dz = \int_0^1 \check{F}_{-j}(z)\varphi_j(z)dz = 0. \tag{6.62}$$

Note that $\mathbb{E}_{F_X}\{\widetilde{U}_k\} = \Theta_k$, and accordingly our first step is to estimate variance of $\widetilde{U}_k$. In [30] there is an explicit formula for the variance, and to introduce it we need several more notations of [30]. Set $X := (\Delta, Z)$, $X_l := (\Delta_l, Z_l)$,

$$V_j(X) := \frac{(\Delta - F_{-j}(Z))\varphi_j(Z)}{f^Z(Z)}, \tag{6.63}$$

$$g_1(X) := L_k^{-1} \sum_{j \in B_k} V_j(X)\theta_j - \Theta_k, \tag{6.64}$$

$$g_2(X_1, X_2) := h_k(X_1, X_2) - g_1(X_1) - g_2(X_2) - \Theta_k. \tag{6.65}$$

Using new notations we can write,

$$h_k(X_1, X_2) = L_k^{-1} \sum_{j \in B_k} V_j(X_1)V_j(X_2), \tag{6.66}$$

$$\mathbb{E}_{F^X}\{V_j(X)\} = \theta_j, \tag{6.67}$$

$$\mathbb{E}_{F^X}\{g_1(X)\} = 0, \tag{6.68}$$

$$\mathbb{E}_{F^X}\{g_2(X_1, X_2)|X_2 = x_2\} = 0. \tag{6.69}$$

The Hoeffding lemma (section 5.1.4 in [30]) gives us a representation

$$\widetilde{U}_n - \Theta_k = \frac{2}{n} \sum_{l=1}^n g_1(X_l) + \frac{2}{n(n-1)} \sum_{1 \leqslant l_1 < l_2 \leqslant n} g_2(X_{l_1}, X_{l_2}). \tag{6.70}$$

We need to introduce several more notations. Set

$$\zeta_1 := \mathbb{E}_{F^X}\{g_1^2(X)\} = L_k^{-2} \sum_{j,i \in B_k} \mathbb{E}_{F_X}\{V_j(X_1)V_i(X_1)\}\theta_j\theta_i - \Theta_k^2. \tag{6.71}$$

Using trigonometric formula

$$\varphi_j(z)\varphi_i(z) = 2^{-1/2}[\varphi_{j-i}(z) + \varphi_{j+i}(z)], \tag{6.72}$$

we continue simplification of the expression for ζ_1,

$$\zeta_1 = 2^{-1/2}L_k^{-2} \sum_{j,i \in B_k} \mathbb{E}_{F^X}\left\{\frac{(\Delta_1 - F_{-j}(Z_1))^2[\varphi_{j+i}(Z_2) + \varphi_{j-i}(Z_2)]}{[f^Z(Z_1)]^2}\right\}\theta_j\theta_i$$

$$- \Theta_k^2. \tag{6.73}$$

Here we used the fact that $F_{-j}(x) = F_{-i}(x)$ for the considered i and j from B_k.

Similarly set

$$\zeta_2 := \mathbb{E}_{F^X}\{[h(X_1, X_2) - \Theta_k]^2\}$$

$$= L_k^{-2}\mathbb{E}_{F^X}\left\{\sum_{j,i \in B_k} V_j(X_1)V_i(X_1)V_j(X_2)V_i(X_2)\right\} - \Theta_k^2$$

$$= 2^{-1}L_k^{-2} \sum_{j,i \in B_k} \mathbb{E}_{F^X}\left\{\frac{(\Delta_1 - F_{-j}(Z_1))(\Delta_1 - F_{-i}(Z_1))[\varphi_{j-i}(Z_1) + \varphi_{j+i}(Z_1)]}{[f^Z(Z_1)]^2}\right\}$$

$$\times \mathbb{E}_{F^X}\left\{\frac{(\Delta_2 - F_{-j}(Z_2))(\Delta_2 - F_{-i}(Z_2))[\varphi_{j-i}(Z_2) + \varphi_{j+i}(Z_2)]}{[f^Z(Z_2)]^2}\right\}$$

$$- \Theta_k^2. \tag{6.74}$$

The Hoeffding formula (Lemma A in section 5.2.1 of [30]) yields that

$$\mathbb{V}_{F_X}(\widetilde{U}_k) = \frac{2}{n(n-1)}[2(n-2)\zeta_1 + \zeta_2]. \tag{6.75}$$

This is an exact formula, and while we can use it for calculating the variance, here we need only a rough upper bound. Note that a function

$$\mathbb{E}_{F^X}\left\{\frac{(\Delta - F_{-j}(Z))^2}{[f^Z(Z)]^2}\Big| Z = z\right\} = \frac{F^X(z) - 2F^X(z)F_{-j}(z) + F_{-j}^2(z)}{[f^Z(z)]^2} \tag{6.76}$$

has a bounded derivative whenever $z \in [0, 1]$. This yields that Fourier coefficients of the function are absolutely summable, and the reader may recall that this property is the Bernstein inequality, see [13]. This fact, together with inequality $2|\theta_j\theta_i| \leqslant \theta_j^2 + \theta_i^2$, yield a rough inequality

$$\zeta_1 \leqslant CL_k^{-1}\Theta_k. \tag{6.77}$$

Similarly we conclude that

$$\zeta_2 \leqslant CL_k^{-1}. \tag{6.78}$$

Combining the results in (6.75) we get

$$\mathbb{V}_{F^X}(U_n) \leqslant C[L_k^{-1}\Theta_k n^{-1} + L_k^{-1}n^{-2}] = CL_k^{-1}n^{-1}[\Theta_k + n^{-1}]. \qquad (6.79)$$

We proved that the U-statistic $\widetilde{U}_k$ satisfies the verified inequality for the variance of $\widetilde{\Theta}_k$. Further, this is the main term in the expansion (6.61) as we will see shortly. The interested reader may note that [30] gives a simple upper bound $C_* n^{-1}$ for the variance, but the issue is that we need to get a specific constant C_* for the considered U-statistic which is $CL_k^{-1}[\Theta_k + n^{-1}]$. This is why all these lengthy calculations are presented.

Now we are evaluating the second moment of the term T_1 on the right side of (6.61). This is not a U-statistic and we need to evaluate it directly. Nonetheless, similarly to analysis of a U-statistic the main tools are combinatoric and combining similar terms in a few groups. Several new notations are needed. Instead of a summation over $l < l'$, we consider the summation over $l_1 < l_2$, and when T_1^2 is considered we use a double sum over $l_1 < l_2$ and $l_3 < l_4$. Second, recall notation (6.18) and write,

$$\check{F}_{-j}(Z_l) = \sum_{t=1}^{n}[n^{-1}\sum_{r=0}^{s}\Delta_t\varphi_r(Z_t)(f^Z(Z_t))^{-1}\varphi_r(Z_l)]$$

$$= \sum_{t\in\{1,2,\ldots,n\}\backslash\{l\}}[n^{-1}\sum_{r=0}^{s}\Delta_t\varphi_r(Z_t)(f^Z(Z_t))^{-1}\varphi_r(Z_l)]$$

$$+n^{-1}\Delta_l(f^Z(Z_l))^{-1}\sum_{r=0}^{s}\varphi_r^2(Z_l)$$

$$=: \check{F}_1(Z_l) + \check{F}_2(Z_l). \qquad (6.80)$$

Let us comment on (6.80). First of all, for the considered $j \in B_k$ neither $\check{F}_1(Z_l)$ nor $\check{F}_2(Z_l)$ depend on j. Second, $\mathbb{E}_{F^X}\{\check{F}_1(Z_l)\} = 0$. Finally, $\check{F}_2(Z_l)$ depends only on (Δ_l, Z_l) and not on any other observation (Δ_t, Z_t) with $t \neq l$, and also $|\check{F}_2(Z_l)| \leqslant Csn^{-1}$ almost sure. These facts will help us to analyze the second moment of T_1.

When we square T_1, we get

$$T_1^2 = \frac{16}{n^2(n-1)^2}L_k^{-2}\sum_{1\leqslant l_1<l_2\leqslant n}\sum_{1\leqslant l_3<l_4\leqslant n}T_1(l_1,l_2,l_3,l_4), \qquad (6.81)$$

where

$$T_1(l_1,l_2,l_3,l_4) := \sum_{j,i\in B_k}\frac{[\Delta_{l_1} - F_{-j}(Z_{l_1})]\varphi_j(Z_{l_1})}{f^Z(Z_{l_1})}$$

$$\times \frac{[F_{-j}(Z_{l_2}) - \sum_{t=1}^{n}n^{-1}\sum_{r=0}^{s}\Delta_t\varphi_r(Z_t)(f^Z(Z_t))^{-1}\varphi_r(Z_{l_2})]\varphi_j(Z_{l_2})}{f^Z(Z_{l_2})}$$

$$\times \frac{[\Delta_{l_3} - F_{-i}(Z_{l_3})]\varphi_i(Z_{l_3})}{f^Z(Z_{l_3})}$$

$$\times \frac{[F_{-i}(Z_{l_4}) - \sum_{t=1}^{n} n^{-1} \sum_{r=0}^{s} \Delta_t \varphi_r(Z_t)(f^Z(Z_t))^{-1} \varphi_r(Z_{l_4})] \varphi_i(Z_{l_4})}{f^Z(Z_{l_4})}. \quad (6.82)$$

We can also rewrite (6.82) in a more compact form,

$$T_1(l_1, l_2, l_3, l_4)$$

$$:= \sum_{j,i \in B_k} \frac{[\Delta_{l_1} - F_{-j}(Z_{l_1})] \varphi_j(Z_{l_1})}{f^Z(Z_{l_1})} \frac{[F_{-j}(Z_{l_2}) - \check{F}_1(Z_{l_2}) - \check{F}_2(Z_{l_2})] \varphi_j(Z_{l_2})}{f^Z(Z_{l_2})}$$

$$\times \frac{[\Delta_{l_3} - F_{-i}(Z_{l_3})] \varphi_i(Z_{l_3})}{f^Z(Z_{l_3})} \frac{[F_{-j}(Z_{l_4}) - \check{F}_1(Z_{l_4}) - \check{F}_2(Z_{l_4})] \varphi_i(Z_{l_4})}{f^Z(Z_{l_4})}. \quad (6.83)$$

What we want to show is that

$$\mathbb{E}_{F^X}\{T_1^2\} = \frac{16}{n^2(n-1)^2} L_k^{-2} \sum_{1 \leqslant l_1 < l_2 \leqslant n} \sum_{1 \leqslant l_3 < l_4 \leqslant n} \mathbb{E}_{F^X}\{T_1(l_1, l_2, l_3, l_4)\}$$

$$\leqslant C L_k^{-1} n^{-1} [\Theta_k + n^{-1}]. \quad (6.84)$$

Similarly to analysis of a U-statistic, while there are of order n^4 terms $\mathbb{E}\{T_1(l_1, l_2, l_3, l_4)\}$ in the sum (6.84), they may be grouped in 4 categories that we are considering in turn. The first one is when $l_1 \neq l_2 \neq l_3 \neq l_4$, and recall that always $l_1 < l_2$ and $l_3 < l_4$. This is the largest category that contains of order n^4 terms. We want to show that each term from this category satisfies

$$\mathbb{E}\{T_1(l_1, l_2, l_3, l_4)\} \leqslant C L_k n^{-1}(\Theta_k + n^{-1}), \ l_1 \neq l_2 \neq l_3 \neq l_4. \quad (6.85)$$

Let us explain why such a term satisfies (6.85). The analysis is primarily based on two facts. The former is that

$$\mathbb{E}_{F^X}\left\{ \frac{[F_{-j}(Z_l) - \check{F}_{-j}(Z_l)] \varphi_j(Z_l)}{f^Z(Z_l)} \right\} = 0. \quad (6.86)$$

The latter is that

$$\mathbb{E}\left\{ \frac{[\Delta_l - F_{-j}(Z_l)] \varphi_j(Z_l)}{f^Z(Z_l)} \right\} = \theta_j. \quad (6.87)$$

Now we can analyze a term. In (6.83) we have L_k^2 terms that include a factor $[F_{-j}(Z_{l_2}) - \check{F}_1(Z_{l_2})]$ and L_k^2 terms that include a factor $\check{F}_2(Z_{l_2})$. Consider the first type of terms. Such a term is not zero due to (6.86) if there is no extra factor containing Z_{l_2}. This implies that the only possible nonzero term is

$$T_{12} := \mathbb{E}\left\{ \sum_{j,i \in B_k} \frac{[\Delta_{l_1} - F_{-j}(Z_{l_1})] \varphi_j(Z_{l_1})}{f^Z(Z_{l_1})} \frac{[F_{-j}(Z_{l_2}) - \check{F}_1(Z_{l_2})] \varphi_j(Z_{l_2})}{f^Z(Z_{l_2})} \right.$$

$$\left. \times \frac{[\Delta_{l_3} - F_{-i}(Z_{l_3})] \varphi_i(Z_{l_3})}{f^Z(Z_{l_3})} \right.$$

$$\times \frac{[-n^{-1}\sum_{r=0}^{s}\Delta_{l_2}\varphi_r(Z_{l_2})\varphi_r(Z_{l_4})(f^Z(Z_{l_2}))^{-1}]\varphi_i(Z_{l_4})}{f^Z(Z_{l_4})}\Big\}. \tag{6.88}$$

Repeating the same argument of using (6.86), only now with Z_{l_4}, the term T_{12} is simplified into

$$T_{12} = \mathbb{E}\Big\{\sum_{j,i\in B_k}\frac{[\Delta_{l_1}-F_{-j}(Z_{l_1})]\varphi_j(Z_{l_1})}{f^Z(Z_{l_1})}\frac{[\Delta_{l_3}-F_{-i}(Z_{l_3})]\varphi_i(Z_{l_3})}{f^Z(Z_{l_3})}$$

$$\times\frac{[-n^{-1}\sum_{r=0}^{s}\Delta_{l_4}\varphi_r(Z_{l_4})\varphi_r(Z_{l_2})(f^Z(Z_{l_4}))^{-1}]\varphi_j(Z_{l_2})}{f^Z(Z_{l_2})}$$

$$\times\frac{[-n^{-1}\sum_{r=0}^{s}\Delta_{l_2}\varphi_r(Z_{l_2})\varphi_r(Z_{l_4})(f^Z(Z_{l_2}))^{-1}]\varphi_i(Z_{l_4})}{f^Z(Z_{l_4})}\Big\}. \tag{6.89}$$

Using (6.87) we simplify (6.89) into

$$T_{12} = \sum_{j,i\in B_k}\theta_j\theta_i$$

$$\times\mathbb{E}_{F^X}\Big\{\frac{[-n^{-1}\sum_{r=0}^{s}F^X(Z_{l_4})\varphi_r(Z_{l_4})\varphi_r(Z_{l_2})(f^Z(Z_{l_4}))^{-1}]\varphi_j(Z_{l_2})}{f^Z(Z_{l_2})}$$

$$\times\frac{[n^{-1}\sum_{r=0}^{s}F^X(Z_{l_2})\varphi_r(Z_{l_2})\varphi_r(Z_{l_4})(f^Z(Z_{l_2}))^{-1}]\varphi_i(Z_{l_4})}{f^Z(Z_{l_4})}\Big\}. \tag{6.90}$$

Now we use a rough inequality

$$\sup_{F^X\in\mathcal{F}(\alpha,Q)}T_{12} \leqslant \sup_{F^X\in\mathcal{F}(\alpha,Q)}\sum_{j,i\in B_k}[\theta_j^2+\theta_i^2]Cn^{-2}s^2 \leqslant Cn^{-2}L_ks^{-2\alpha}s^2. \tag{6.91}$$

We conclude that the first type of terms in T_1^2 satisfies (6.85). Let us also note that using a more thorough analysis of the expectation in (6.90) as a Fourier coefficient, inequality (6.91) can be improved in order.

Now consider expectation of the second type with the factor $-\check{F}_2(Z_{l_2})$,

$$\mathbb{E}_{F^X}\{T'_{12}\} := \sum_{j,i\in B_k}\mathbb{E}_{F^X}\Big\{\frac{[\Delta_{l_1}-F_{-j}(Z_{l_1})]\varphi_j(Z_{l_1})}{f^Z(Z_{l_1})}\frac{[-\check{F}_2(Z_{l_2})]\varphi_j(Z_{l_2})}{f^Z(Z_{l_2})}$$

$$\times\frac{[\Delta_{l_3}-F_{-i}(Z_{l_3})]\varphi_i(Z_{l_3})}{f^Z(Z_{l_3})}\frac{[F_{-j}(Z_{l_4})-\check{F}_1(Z_{l_4})-\check{F}_2(Z_{l_4})]\varphi_i(Z_{l_4})}{f^Z(Z_{l_4})}\Big\}. \tag{6.92}$$

Recall that $\check{F}_2(Z_{l_2})$ is a function only in (Δ_{l_2},Z_{l_2}), and then (6.86) used for $l=l_4$ yields that

$$\mathbb{E}_{F^X}\{T'_{12}\} = \sum_{j,i\in B_k}\mathbb{E}_{F^X}\Big\{\frac{[\Delta_{l_1}-F_{-j}(Z_{l_1})]\varphi_j(Z_{l_1})}{f^Z(Z_{l_1})}\frac{[-\check{F}_2(Z_{l_2})]\varphi_j(Z_{l_2})}{f^Z(Z_{l_2})}$$

$$\times\frac{[\Delta_{l_3}-F_{-i}(Z_{l_3})]\varphi_i(Z_{l_3})}{f^Z(Z_{l_3})}\frac{[-\check{F}_2(Z_{l_4})]\varphi_i(Z_{l_4})}{f^Z(Z_{l_4})}\Big\}$$

$$= \sum_{j,i \in B_k} \theta_j \theta_i \mathbb{E}_{F^X} \left\{ \frac{\check{F}_2(Z_{l_2}) \varphi_j(Z_{l_2})}{f^Z(Z_{l_2})} \frac{\check{F}_2(Z_{l_4}) \varphi_i(Z_{l_4})}{f^Z(Z_{l_4})} \right\} \qquad (6.93)$$

Then identically to (6.91) we get the wished upper bound for T'_{12}. This and (6.91) yield the wished (6.85) for the first category of indexes where $l_1 \neq l_2 \neq l_3 \neq l_4$. Note that we can improve in order the upper bound by considering the expectations on the right side of (6.90) and (6.93) as Fourier coefficients, but the obtained rough inequalities are sufficient for our purposes.

Now we are considering terms in (6.81) from the second group where $l_1 < l_2 = l_3 < l_4$ or $l_1 = l_4$ and $l_2 \neq l_3$. We present analysis of the former case because the latter is analyzed similarly. There are of order n^3 such terms, accordingly it is sufficient to show that the expectation of a particular term $T_1(l_1, l_2, l_2, l_4)$ is bounded by $CL_k(\Theta_k + n^{-1})$. Using (6.83) we can write that

$$E_{F^X}\{T_1(l_1, l_2, l_2, l_4)\}$$

$$= \sum_{j,i \in B_k} \mathbb{E}_{F^X} \left\{ \frac{[\Delta_{l_1} - F_{-j}(Z_{l_1})] \varphi_j(Z_{l_1})}{f^Z(Z_{l_1})} \frac{[F_{-j}(Z_{l_2}) - \check{F}_1(Z_{l_2}) - \check{F}_2(Z_{l_2})] \varphi_j(Z_{l_2})}{f^Z(Z_{l_2})} \right.$$

$$\left. \times \frac{[\Delta_{l_2} - F_{-i}(Z_{l_2})] \varphi_i(Z_{l_2})}{f^Z(Z_{l_2})} \frac{[F_{-j}(Z_{l_4}) - \check{F}_1(Z_{l_4}) - \check{F}_2(Z_{l_4})] \varphi_i(Z_{l_4})}{f^Z(Z_{l_4})} \right\}. \qquad (6.94)$$

Using (6.87) and (6.86) with $l = l_4$ we simplify the last expectation into

$$E_{F^X}\{T_1(l_1, l_2, l_2, l_4)\}$$

$$= \sum_{j,i \in B_k} \theta_j \mathbb{E}_{F^X} \left\{ \frac{[-n^{-1} \sum_{r=0}^{s} F^X(Z_{l_4}) \varphi_r(Z_{l_4}) \varphi_r(Z_{l_2})(f^Z(Z_{l_4}))^{-1}] \varphi_j(Z_{l_2})}{f^Z(Z_{l_2})} \right.$$

$$\left. \times \frac{[\Delta_{l_2} - F_{-i}(Z_{l_2})] \varphi_i(Z_{l_2})}{f^Z(Z_{l_2})} \frac{[F_{-j}(Z_{l_4}) - \check{F}_1(Z_{l_4})] \varphi_i(Z_{l_4})}{f^Z(Z_{l_4})} \right\}$$

$$+ \sum_{j,i \in B_k} \mathbb{E}_{F^X} \left\{ \frac{[\Delta_{l_1} - F_{-j}(Z_{l_1})] \varphi_j(Z_{l_1})}{f^Z(Z_{l_1})} \frac{[F_{-j}(Z_{l_2}) - \check{F}_1(Z_{l_2}) - \check{F}_2(Z_{l_2})] \varphi_j(Z_{l_2})}{f^Z(Z_{l_2})} \right.$$

$$\left. \times \frac{[\Delta_{l_2} - F_{-i}(Z_{l_2})] \varphi_i(Z_{l_2})}{f^Z(Z_{l_2})} \frac{[-\check{F}_2(Z_{l_4})] \varphi_i(Z_{l_4})}{f^Z(Z_{l_4})} \right\}. \qquad (6.95)$$

To evaluate the first sum on the right side of (6.95) we use a relation following from the Cauchy-Schwarz inequality,

$$\sup_{F^X \in \mathcal{F}(\alpha, Q)} \sum_{j \in B_k} |\theta_j|$$

$$\leqslant \sup_{F^X \in \mathcal{F}(\alpha, Q)} \left[\sum_{j \in B_k} j^{2\alpha} \theta_j^2 \sum_{j \in B_k} j^{-2\alpha} \right]^{1/2} \leqslant C s^{(-2\alpha+1)/2}, \qquad (6.96)$$

and following from (6.26)

$$\sup_{F^X \in \mathcal{F}(\alpha,Q)} \mathbb{E}\{\int_0^1 |F_j(z) - \breve{F}_1(z)|dz\} \leqslant Csn^{-1/2}. \tag{6.97}$$

The inequalities imply the wished upper bound. Similarly, we use the previous inequality and $|\breve{F}_2(z)| \leqslant Csn^{-1}$ to get the wished upper bound for the second sum in (6.95). This ends evaluation of a term from the second category. Note that we have established a stronger in order inequality than needed.

The third category of terms is when $l_1 = l_3$, $l_1 < l_2$, $l_3 < l_4$ and $l_2 \neq l_4$. Similarly to the second category, there are of order n^3 terms in the category, and we need to verify the same upper bound for a term. Write for such a term,

$$E_{F^X}\{T_1(l_1,l_2,l_1,l_4)\}$$

$$= \sum_{j,i \in B_k} \mathbb{E}_{F^X}\left\{ \frac{[\Delta_{l_1} - F_{-j}(Z_{l_1})]\varphi_j(Z_{l_1})}{f^Z(Z_{l_1})} \frac{[F_{-j}(Z_{l_2}) - \breve{F}_1(Z_{l_2}) - \breve{F}_2(Z_{l_2})]\varphi_j(Z_{l_2})}{f^Z(Z_{l_2})} \right.$$

$$\left. \times \frac{[\Delta_{l_1} - F_{-i}(Z_{l_1})]\varphi_i(Z_{l_2})}{f^Z(Z_{l_2})} \frac{[F_{-j}(Z_{l_4}) - \breve{F}_1(Z_{l_4}) - \breve{F}_2(Z_{l_4})]\varphi_i(Z_{l_4})}{f^Z(Z_{l_4})} \right\}. \tag{6.98}$$

Here we simply repeat analysis of a term from the first category, namely we get (6.89) and (6.92) only now with $l_1 = l_3$, and this yields the wished upper bound.

The final fourth category is when $l_1 = l_3$ and $l_2 = l_4$. There are of order n^{-2} terms in this category. For a particular term we have

$$E_{F^X}\{T_1(l_1,l_2,l_1,l_2)\}$$

$$= \sum_{j,i \in B_k} \mathbb{E}_{F^X}\left\{ \frac{[\Delta_{l_1} - F_{-j}(Z_{l_1})]\varphi_j(Z_{l_1})}{f^Z(Z_{l_1})} \frac{[F_{-j}(Z_{l_2}) - \breve{F}_1(Z_{l_2}) - \breve{F}_2(Z_{l_2})]\varphi_j(Z_{l_2})}{f^Z(Z_{l_2})} \right.$$

$$\left. \times \frac{[\Delta_{l_1} - F_{-i}(Z_{l_1})]\varphi_i(Z_{l_2})}{f^Z(Z_{l_2})} \frac{[F_{-j}(Z_{l_2}) - \breve{F}_1(Z_{l_2}) - \breve{F}_2(Z_{l_2})]\varphi_i(Z_{l_2})}{f^Z(Z_{l_2})} \right\}. \tag{6.99}$$

The wished upper bound follows immediately from (6.26).

Combing the upper bounds for the four categories of terms we verify that

$$\sup_{F^X \in \mathcal{F}(\alpha,Q)} \mathbb{E}_{F^X}\{T_1^2\} \leqslant CL_k^{-1}n^{-1}(\Theta_k + n^{-1}). \tag{6.100}$$

We are left with evaluating the second moment of T_2 defined in (6.61). We need to show that

$$\sup_{F^X \in \mathcal{F}(\alpha,Q)} \mathbb{E}_{F^X}\{T_2^2\} = \frac{4}{n^2(n-1)^2}L_k^{-2} \sum_{1 \leqslant l_1 < l_2 \leqslant n} \sum_{1 \leqslant l_3 < l_4 \leqslant n} \sum_{j,i \in B_k}$$

$$\times \mathbb{E}_{F^X}\left\{ \frac{(F_{-j}(Z_{l_1}) - \breve{F}_{-j}(Z_{l_1}))\varphi_j(Z_{l_1})}{f^Z(Z_{l_1})} \frac{(F_{-j}(Z_{l_2}) - \breve{F}_{-j}(Z_{l_2}))\varphi_j(Z_{l_2})}{f^Z(Z_{l_2})} \right.$$

$$\times \frac{(F_{-i}(Z_{l_3}) - \check{F}_{-i}(Z_{l_3}))\varphi_i(Z_{l_3})}{f^Z(Z_{l_3})} \frac{(F_{-i}(Z_{l_4}) - \check{F}_{-i}(Z_{l_4}))\varphi_i(Z_{l_4})}{f^Z(Z_{l_4})}\Big\}$$

$$\leqslant CL_k^{-1}n^{-1}(\Theta_k + n^{-1}). \tag{6.101}$$

First of all, let us note that using (6.26) we get $\mathbb{E}_{F_X}\{T_2^2\} < Cs^2 n^{-2}$ which is close to what we want. To get (6.101), we use (6.86) and follow our analysis of T_{12} which gives us a very rough but sufficient upper bound $Cn^{-5/2}s^4$. This and $s^4 L_{r_n} 0 = o_n(1)n^{1/2}$ prove (6.101).

Using (6.79), (6.100) and (6.101) in (6.61) verifies Lemma 6.6.

Theorem 3.1 is proved.

Proof of Theorem 3.2. The assertion of Theorem 3.2 is not obvious and intriguing because according to (2.1) the joint density $f^{Z,\Delta}(z,\delta)$ of observed pair (Z,Δ) is equal to the product of an unknown density $f^Z(z)$ and $[F^X(z)]^\delta[1 - F^X(z)]^{1-\delta}$. Accordingly, it is natural to conjecture that, for efficient estimation of $F^X(z)$, an unknown density $f^Z(z)$ of the monitoring time should be as smooth as the cdf of interest $F^X(z)$ because the smaller smoothness defines the smoothness of $f^{Z,\Delta}(z,\delta)$ in z. Fortunately, as we will establish shortly, the latter is not the case and it suffices for the density $f^Z(z)$ to be differentiable regardless of how smooth the cdf is. This is a remarkable outcome in comparison with [16], and it will be explained why this is the case.

We begin with recalling Lemma 1 in [14] which allows us to evaluate quality of estimators $\bar{f}^Z(z)$ and $\check{f}^Z(z)$ for any positive integer t,

$$\max\Big(\max_{z\in[0,1]}\mathbb{E}_{f^Z}\{(\bar{f}^Z(z) - f^Z(z))^{2t}\}, \max_{z\in[0,1]}\mathbb{E}_{f^Z}\{(\check{f}^Z(z) - f^Z(z))^{2t}\}\Big)$$

$$= o_n(1)s^{4t+2}n^{-2t/3}. \tag{6.102}$$

Because the density estimator is plugged in denominator, we use an expansion that will allow us to utilize (6.102),

$$\frac{1}{\bar{f}^Z(z)} = \frac{1}{f^Z(z)} + \frac{f^Z(z) - \bar{f}^Z(z)}{[f^Z(z)]^2} + \frac{(f^Z(z) - \bar{f}^Z(z))^2}{\bar{f}^Z(z)[f^Z(z)]^2}. \tag{6.103}$$

We begin our analysis with the plugged-in pilot Fourier estimator. Using (6.103) we write,

$$\check{\theta}_j^* := n^{-1}\sum_{l=1}^{n}\frac{\Delta_l\varphi_j(Z_l)}{\bar{f}^Z(Z_l)}$$

$$= n^{-1}\sum_{l=1}^{n}\frac{\Delta_l\varphi_j(Z_l)}{f^Z(Z_l)} + n^{-1}\sum_{l=1}^{n}\frac{\Delta_l\varphi_j(Z_l)(f^Z(Z_l) - \bar{f}^Z(Z_l))}{[f^Z(Z_l)]^2}$$

$$+ n^{-1}\sum_{l=1}^{n}\Delta_l\varphi_j(Z_l)\frac{(f^Z(Z_l) - \bar{f}^Z(Z_l))^2}{\bar{f}^Z(Z_l)[f^Z(Z_l)]^2} =: \check{\theta}_j + A_{j1} + A_{j2}. \tag{6.104}$$

Here $\check{\theta}_j$ is the estimate based on the underlying f^Z and studied in the proof of Theorem 3.1, see Lemma 6.1.

It is simple to evaluate A_{j2} using (6.102),

$$\mathbb{E}_{fz}\{A_{j2}^2\} \leqslant Cn^{-1}, \quad \sum_{j=0}^{J_n} \mathbb{E}_{fz}\{A_{j2}^2\} = o_n(1)s^{11}n^{-1}. \tag{6.105}$$

Our next step is to evaluate the second moment of the term A_{j1} in (6.104). Write,

$$\mathbb{E}\{A_{j1}^2\} \leqslant 2n^{-2}\mathbb{E}_{fz}\left\{\left[\sum_{l=1}^{n} \frac{\Delta_l\varphi_j(Z_l)(f^Z(Z_l) - \check{f}^Z(Z_l))}{[f^Z(Z_l)]^2}\right]^2\right\}$$

$$+ 2n^{-2}\mathbb{E}_{fz}\left\{\left[\sum_{l=1}^{n} \frac{\Delta_l\varphi_j(Z_l)(\check{f}^Z(Z_l) - \bar{f}^Z(Z_l))}{[f^Z(Z_l)]^2}\right]^2\right\} =: 2B_{j1} + 2B_{j2}. \tag{6.106}$$

According to Assumption 1, density $f^Z(z)$ is bounded below from zero, and $\bar{f}^Z(z) \neq \check{f}^Z(z)$ only if $\check{f}^Z(z) < \gamma_n$. These facts and (6.102) yield

$$\sum_{j=0}^{J_n} B_{j2} = o_n(1)n^{-1}. \tag{6.107}$$

Evaluation of B_{j1} is more involved because we only assume that the density $f^Z(z)$ is differentiable. Write,

$$B_{j1} = n^{-1}\mathbb{E}_{fz}\left\{\frac{[\Delta_1\varphi_j(Z_1)(f^Z(Z_1) - \check{f}^Z(Z_1))]^2}{[f^Z(Z_l)]^4}\right\}$$

$$+ n^{-2}n(n-1)\mathbb{E}_{fz}\left\{\prod_{l=n-1}^{n} \frac{\Delta_l\varphi_j(Z_l)(f^Z(Z_l) - \check{f}^Z(Z_l))}{[f^Z(Z_l)]^2}\right\} =: B_{j11} + B_{j12}. \tag{6.108}$$

Using (6.102) we get that B_{j11} is bounded by $o_n(1)n^{-3/2}$. To estimate B_{j12} we begin with definition of $\check{f}^Z(z)$ and write

$$\check{f}^Z(z) = n^{-1}\sum_{j=0}^{J_n}\sum_{l=1}^{n-2} \varphi_j(Z_l)\varphi_j(z)$$

$$+ n^{-1}\sum_{j=0}^{J_n}\sum_{l=n-1}^{n} \varphi_j(Z_l)\varphi_j(z) =: \tilde{h}'(z) + \tilde{h}''(z). \tag{6.109}$$

This implies

$$f^Z(z) - \check{f}^Z(z) = [(n-2)n^{-1}f^Z(z) - \tilde{h}'(z)] + [2n^{-1}f^Z(z) - \tilde{h}''(z)].$$

Using this representation we can rewrite B_{j12} as

$$B_{j12} = n^{-1}(n-1)\mathbb{E}_{fz}\left\{\prod_{l=n-1}^{n} \frac{\Delta_l\varphi_j(Z_l)}{[f^Z(Z_l)]^2}[(n-2)n^{-1}f^Z(Z_l) - \tilde{h}'(Z_l))\right.$$

$$+ (2n^{-1}f^Z(Z_l) - \widetilde{h}''(Z_l))]\Big\}$$

$$= \frac{(n-1)}{n}\mathbb{E}_{f^Z}\Big\{\prod_{l=n-1}^{n}\frac{\Delta_l\varphi_j(Z_l)}{[f^Z(Z_l)]^2}[(n-2)n^{-1}f^Z(Z_l) - \widetilde{h}'(Z_l)]\Big\}$$

$$+ \frac{2(n-1)}{n}\mathbb{E}_{f^Z}\Big\{\Big[\prod_{l=n-1}^{n}\frac{\Delta_l\varphi_j(Z_l)}{[f^Z(Z_l)]^2}\Big]$$

$$\times [(n-2)n^{-1}f^Z(Z_{n-1}) - \widetilde{h}'(Z_{n-1})][(2n^{-1}f^Z(Z_n) - \widetilde{h}''(Z_n))]\Big\}$$

$$+ \frac{(n-1)}{n}\mathbb{E}_{f^Z}\Big\{\prod_{l=n-1}^{n}\frac{\Delta_l\varphi_j(Z_l)}{[f^Z(Z_l)]^2}[2n^{-1}f^Z(Z_l) - \widetilde{h}''(Z_l)]\Big\}$$

$$=: \frac{(n-1)}{n}D_{j1} + \frac{2(n-1)}{n}D_{j2} + D_{j3}. \qquad (6.110)$$

Inequality $|2n^{-1}f^Z(z) - \widetilde{h}''(z)| \leqslant C\ln(s)n^{-2/3}$ yields

$$\sum_{j=0}^{J_n} D_{j3} = o_n(1)[\ln(s)]^3 n^{-1}.$$

Next we consider D_{j1},

$$D_{j1} = \mathbb{E}_{f^Z}\Big\{\prod_{l=n-1}^{n}\frac{\Delta_l\varphi_j(Z_l)}{[f^Z(Z_l)]^2}[(n-2)n^{-1}f^Z(Z_l) - \widetilde{h}'(Z_l)]\Big\}$$

$$= \mathbb{E}_{f^Z}\Big\{\Big[\int_0^1 F^X(z)\varphi_j(z)[f^Z(z)]^{-1}[(n-2)n^{-1}f^Z(z) - \widetilde{h}'(z)]dz\Big]^2\Big\}. \quad (6.111)$$

Set

$$\widetilde{\kappa}_i := (n-2)^{-1}\sum_{l=1}^{n-2}\varphi_i(Z_l), \quad \kappa_i := \int_0^1 f^Z(z)\varphi_i(z)dz = \mathbb{E}_{f^Z}\{\varphi_i(Z)\}. \quad (6.112)$$

Then $\widetilde{h}'(z) = (n-2)n^{-1}\sum_{i=1}^{J_n}\widetilde{\kappa}_i\varphi_i(z)$, and for the difference in (6.111) we get

$$\frac{n-2}{n}f^Z(z) - \widetilde{h}'(z) = \frac{n-2}{n}\sum_{i=1}^{J_n}(\kappa_i - \widetilde{\kappa}_i)\varphi_i(z) + \frac{n-2}{n}\sum_{i>J_n}\kappa_i\varphi_i(z). \quad (6.113)$$

Substituting (6.113) in (6.111) and then using the Cauchy inequality we continue (6.111),

$$D_{j1} = \frac{(n-2)^2}{n^2}$$

$$\times\mathbb{E}_{f^Z}\Big\{\Big[\int_0^1 F^X(z)[f^Z(z)]^{-1}\varphi_j(z)[\sum_{i=1}^{J_n}(\kappa_i - \widetilde{\kappa}_i)\varphi_i(z) + \sum_{i>J_n}\kappa_i\varphi_i(z)]dz\Big]^2\Big\}$$

$$\leqslant 2\mathbb{E}_{f^Z}\Big\{\Big[\sum_{i=1}^{J_n}(\kappa_i - \widetilde{\kappa}_i)\int_0^1 \frac{F^X(z)\varphi_j(z)\varphi_i(z)}{f^Z(z)}dz\Big]^2\Big\}$$

$$+2\Big[\sum_{i>J_n}\kappa_i\int_0^1 \frac{F^X(z)\varphi_j(z)\varphi_i(z)}{f^Z(z)}dz\Big]^2$$

$$=: 2D_{j11} + 2D_{j12}. \tag{6.114}$$

To evaluate the first term, recall a familiar trigonometric formula $\varphi_j(z)\varphi_i(z) = 2^{-1/2}[\varphi_{j-i}(z) + \varphi_{j+i}(z)]$, definitions (6.112), and write for the expectation,

$$E\{(\kappa_i - \widetilde{\kappa}_i)(\kappa_t - \widetilde{\kappa}_t)\}$$

$$= (n-2)^{-2}E\{\sum_{l,m=1}^{n-2}(\kappa_i - \varphi_i(Z_l))(\kappa_t - \varphi_t(Z_m))\}$$

$$= (n-2)^{-1}[2^{-1/2}(\kappa_{i-t} + \kappa_{i+t}) - \kappa_i\kappa_t]. \tag{6.115}$$

Now we can finish evaluation of D_{j11}. Set, with the use of the above-presented trigonometric formula,

$$\int_0^1 \frac{F^X(z)\varphi_j(z)\varphi_i(z)}{f^Z(z)}dz = 2^{-1/2}[b_{j-i} + b_{j+i}], \tag{6.116}$$

where

$$b_i := \int_0^1 \frac{F^X(z)\varphi_i(z)}{f^Z(z)}dz.$$

Using (6.115) and (6.116) we get

$$D_{j11} = (n-2)^{-1}\sum_{i,t=1}^{J_n}[2^{-1/2}(\kappa_{i-t} + \kappa_{i+t}) - \kappa_i\kappa_t]$$

$$\times [2^{-1}(b_{j-i} + b_{j+i})(b_{j-t} + b_{j+t})] \leqslant Cn^{-1}. \tag{6.117}$$

The last inequality holds due to the famous Bernstein inequality which states that Fourier coefficients of differentiable functions are absolutely summable. Furthermore, Lipschitz functions of order larger than $1/2$ can be considered as well, see [13].

Now consider D_{j12}. Using (6.116) and the Cauchy-Schwarz inequality yields

$$D_{j12} = \Big[\sum_{i>J_n}\kappa_i\int_0^1 \frac{F^X(z)\varphi_j(z)\varphi_i(z)}{f^Z(z)}dz\Big]^2 \leqslant \sum_{i>J_n}\kappa_i^2\sum_{i>J_n}[b_{j-i}^2 + b_{j+i}^2]$$

$$\leqslant CJ_n^{-2}(J_n - j)^{-2} \leqslant C[\ln(s)]^{-4}n^{-4/3}. \tag{6.118}$$

Using the obtained results in (6.114) we conclude that

$$D_{j1} \leqslant Cn^{-1}. \tag{6.119}$$

As we will see shortly, this is the dominant term in A_{j1}.

We are left with evaluation of the cross-product term D_{j2} defined in (6.110). Write,

$$D_{j2} = \mathbb{E}_{f^Z}\Big\{\frac{\Delta_{n-1}\varphi_j(Z_{n-1})}{[f^Z(Z_{n-1})]^2}[\frac{n-2}{n}f^Z(Z_{n-1}) - \widetilde{h}'(Z_{n-1})]$$

$$\times\frac{\Delta_n\varphi_j(Z_n)}{[f^Z(Z_n)]^2}[2n^{-1}f^Z(Z_n) - n^{-1}\sum_{i=0}^{J_n}\varphi_i(Z_{n-1})\varphi_i(Z_n) - n^{-1}\sum_{i=0}^{J_n}\varphi_i^2(Z_n)]\Big\}$$

$$= \int_0^1 \frac{F^X(z)\varphi_j(z)}{f^Z(z)}[\frac{2n-2}{n}f^Z(z) - \mathbb{E}\{\widetilde{h}'(z)\}]dz$$

$$\times \int_0^1 \frac{F^X(x)\varphi_j(x)}{f^Z(x)}[2n^{-1}f^Z(x) - n^{-1}\sum_{i=0}^{J_n}\varphi_i^2(x)]dx$$

$$-n^{-1}\sum_{i=0}^{J_n}\int_0^1 \frac{F^X(z)\varphi_j(z)}{f^Z(z)}[\frac{n-2}{n}f^Z(z) - \mathbb{E}\{\widetilde{h}'(z)\}]\varphi_i(z)dz$$

$$\times \int_0^1 \frac{F^X(x)\varphi_j(x)}{f^Z(x)}\varphi_i(x)dx. \tag{6.120}$$

Several remarks about terms in (6.120) are due. Using (6.113) and the previously introduced notation we conclude that

$$\frac{n-2}{n}f^Z(z) - \mathbb{E}\{\widetilde{h}'(z)\} = \frac{n-2}{n}\sum_{i>J_n}\kappa_i\varphi_i(z). \tag{6.121}$$

Further, we have $n^{-1}|2f^Z(z) - \sum_{i=0}^{J_n}\varphi_i^2(z)| \leqslant CJ_n n^{-1}$. Using (6.116), Cauchy-Schwarz and Bessel inequalities yield

$$[\int_0^1 \frac{F^X(z)\varphi_j(z)\varphi_i(z)}{f^Z(z)}\sum_{t>J_n}\kappa_t\varphi_t(z)dz]^2$$

$$\leqslant [\sum_{t>J_n}\kappa_t^2][\sum_{t>J_n}(\int_0^1 \frac{F^X(z)\varphi_j(z)\varphi_i(z)\varphi_t(z)}{f^Z(z)}dz)^2] \leqslant CJ_n^{-1}. \tag{6.122}$$

Combining the results in (6.120) we conclude that

$$D_{j2} = o_n(1)n^{-1}. \tag{6.123}$$

As we have verified, the dominant term in A_{j1} is D_{j1}.

Using the obtained results in (6.110) yields $B_{j12} \leqslant Cn^{-1}$, and together with already evaluated $B_{j11} = o_n(1)n^{-3/2}$ and (6.108) we get $B_{j1} \leqslant Cn^{-1}$. Using this inequality and (6.107) in (6.106), and the obtained result, (6.104) and (6.105) we conclude that uniformly over all considered cumulative distribution functions of interest F^X and densities f^Z of the monitoring time we have

$$\mathbb{E}_{F^X,f^Z}\{(\breve{\theta}_j^* - \breve{\theta}_j)^2\} \leqslant Cn^{-1}. \tag{6.124}$$

This inequality yields that a projection estimate of an underlying cdf $F^X(x)$, based on either $\widetilde{\theta}_j^*$ or $\widetilde{\theta}_j$ yields the same rate of the MISE convergence. Also note that in (6.124) the subscript emphasizes that both F^X and f^Z must be known to calculate the expectation, and the constant C is the same for all considered F^X.

Now we are considering the data-driven Fourier estimate $\widehat{\theta}_j$, defined in (3.5) and where in place of $f^Z(z)$ we use $\bar{f}^Z(z)$. Namely, we are exploring the data-driven Fourier estimator

$$\widehat{\theta}_j^* := n^{-1} \sum_{l=1}^{n} \frac{(\Delta_l - \check{F}_{-j}(Z_l))\varphi_j(Z_l)}{\bar{f}^Z(Z_l)}. \tag{6.125}$$

Here $\check{F}_{-j}(z)$ is defined in (3.3), and it estimates a function $F_{-j}(z)$ defined in (3.4). The difference between the already studied Fourier estimator $\check{\theta}_j^*$ and the new Fourier estimator $\widehat{\theta}_j^*$ (compare (6.104) and (6.125)) is that in $\widehat{\theta}_j^*$ the factor Δ_l is replaced by $\Delta_l - \check{F}_{-j}(Z_l)$. The underlying idea is that with the help of this replacement the second moment inequality (6.124) for $\check{\theta}_j^* - \check{\theta}_j$ can be replaced by

$$\mathbb{E}_{F^X, f^Z}\{(\widehat{\theta}_j^* - \widehat{\theta}_j)^2\} = o_n(1)n^{-1}I(j > s) \tag{6.126}$$

and

$$\mathbb{E}_{F^X, f^Z}\{(\widehat{\theta}_j^* - \widehat{\theta}_j)^2\} \leqslant Cn^{-1}I(j \leqslant s). \tag{6.127}$$

This result is the key in proving the sharp minimax assertion of Theorem 3.2.

A comment is due. Inequality (6.127) states that $\widehat{\theta}_j^*$ mimics $\widehat{\theta}_j$ with the MSE of order n^{-1}, and this result is well expected. A challenging part is the case $j > s$ in (6.126). It follows from the above-presented proof of (6.124) that the term that prevented us from establishing (6.126) is D_{j1} which is of order n^{-1}, while all other terms are of the wished order $o_n(1)n^{-1}$. Accordingly, we are considering

$$D_{j1}^* := \mathbb{E}_{f^Z}\left\{ \prod_{l=n-1}^{n} \frac{(\Delta_l - \check{F}_{-j}(Z_l))\varphi_j(Z_l)}{[f^Z(Z_l)]^2} \left[\frac{(n-2)}{n} f^Z(Z_l) - \widetilde{h}'(Z_l) \right] \right\} \tag{6.128}$$

and would like to show that

$$D_{j1}^* \leqslant Cn^{-1}I(j \leqslant s) + o_n(1)n^{-1}I(j > s). \tag{6.129}$$

To verify (6.129), we rewrite the new factor as

$$\Delta_l - \check{F}_{-j}(Z_l) = [\Delta_l - F_{-j}(Z_l)] + [F_{-j}(Z_l) - \check{F}_{-j}(Z_l)]$$

$$=: W_1(\Delta_l, Z_l) + W_2(Z_l). \tag{6.130}$$

For $W_2(Z_l)$ we have

$$W_2(Z_l) = F_{-j}(Z_l) - \check{F}_{-j}(Z_l) = \sum_{i \in S_{-j}} (\theta_i - \check{\theta}_i^*)\varphi_i(Z_l), \tag{6.131}$$

where $S_{-j} = \{0, 1, \ldots, s\} \backslash \{j\}$ if $j \leqslant s$ and $S_{-j} = \{0, 1, \ldots, s\}$ otherwise. Then, with the help of (2.1), Lemma 6.1 and (6.124) we get a rough but sufficient for our task inequality,

$$\mathbb{E}_{F^X, f^Z}\{[W_2(Z_l)]^4\} \leqslant C n^{-2} s^4. \tag{6.132}$$

Now we are ready to evaluate term D_{j1}^* defined in (6.128). Using notation (6.130) we write,

$$D_{j1}^* := \mathbb{E}_{f^z}\Big\{ \prod_{l=n-1}^{n} \frac{(\Delta_l - \check{F}_{-j}(Z_l))\varphi_j(Z_l)}{[f^Z(Z_l)]^2} \big[\frac{(n-2)}{n} f^Z(Z_l) - \tilde{h}'(Z_l)\big]\Big\}$$

$$= \mathbb{E}_{f^z}\Big\{ \prod_{l=n-1}^{n} \frac{(W_1(\Delta_l, Z_l) + W_2(Z_l))\varphi_j(Z_l)}{[f^Z(Z_l)]^2} \big[\frac{(n-2)}{n} f^Z(Z_l) - \tilde{h}'(Z_l)\big]\Big\}. \tag{6.133}$$

We begin with considering a term with $W_1(\Delta_{n-1}, Z_{n-1}) W_1(\Delta_n, Z_n)$. Write

$$D_{j11}^* := \mathbb{E}_{f^z}\Big\{ \prod_{l=n-1}^{n} \frac{W_1(\Delta_l, Z_l)\varphi_j(Z_l)}{[f^Z(Z_l)]^2} \big[\frac{(n-2)}{n} f^Z(Z_l) - \tilde{h}'(Z_l)\big]\Big\}$$

$$= \mathbb{E}_{f^z}\Big\{ \prod_{l=n-1}^{n} \frac{(\Delta_l - F_{-j}(Z_l))\varphi_j(Z_l)}{[f^Z(Z_l)]^2} \big[\frac{(n-2)}{n} f^Z(Z_l) - \tilde{h}'(Z_l)\big]\Big\}$$

$$= \mathbb{E}_{f^z}\Big\{\Big[\int_0^1 (F^X(z) - F_{-j}(z))\varphi_j(z)[f^Z(z)]^{-1}$$

$$\times [(n-2)n^{-1}f^Z(z) - \tilde{h}'(z)]dz\Big]^2\Big\}. \tag{6.134}$$

Using (6.112), (6.113) and the Cauchy inequality we continue (6.134),

$$D_{j11}^* = \frac{(n-2)^2}{n^2} \mathbb{E}_{f^z}\Big\{\Big[\int_0^1 (F^X(z) - F_{-j}(z))[f^Z(z)]^{-1}\varphi_j(z)$$

$$\times [\sum_{i=1}^{J_n}(\kappa_i - \tilde{\kappa}_i)\varphi_i(z) + \sum_{i>J_n} \kappa_i\varphi_i(z)]dz\Big]^2\Big\}$$

$$\leqslant 2\mathbb{E}_{f^z}\Big\{\Big[\sum_{i=1}^{J_n}(\kappa_i - \tilde{\kappa}_i)\int_0^1 \frac{(F^X(z) - F_{-j}(z))\varphi_j(z)\varphi_i(z)}{f^Z(z)}dz\Big]^2\Big\}$$

$$+ 2\Big[\sum_{i>J_n} \kappa_i \int_0^1 \frac{(F^X(z) - F_{-j}(z))\varphi_j(z)\varphi_i(z)}{f^Z(z)}dz\Big]^2$$

$$=: 2B_{j11} + 2B_{j12}. \tag{6.135}$$

To evaluate B_{j11} we recall $\varphi_j(z)\varphi_i(z) = 2^{-1/2}[\varphi_{j-i}(z) + \varphi_{j+i}(z)]$ and write,

$$\int_0^1 \frac{(F^X(z) - F_{-j}(z))\varphi_j(z)\varphi_i(z)}{f^Z(z)}dz = 2^{-1/2}[b_{j-i}(j, s) + b_{j+i}(j, s)]. \tag{6.136}$$

Here

$$b_i(j,s) := \int_0^1 \frac{[F^X(z) - F_{-j}(z)]\varphi_i(z)}{f^Z(z)} dz. \qquad (6.137)$$

Let us comment on terms $b_i(j,s)$. Consider $j > s$, recall that θ_i are Fourier coefficients of an underlying F^X, and write,

$$b_i(j,s) = \int_0^1 \frac{[\sum_{t>s} \theta_t \varphi_t(z)]\varphi_i(z)}{f^Z(z)} dz = \sum_{t>s} \theta_t[\nu_{t-i} + \nu_{t+i}],$$

$$\text{where} \quad \nu_i := \int_0^1 [f^Z(z)]^{-1} \varphi_i(z) dz. \qquad (6.138)$$

Thus, using the above-mentioned Bernstein inequality we get for $j > s$

$$\sum_{i=-\infty}^{\infty} |b_i(j,s)| \leqslant \sum_{t>s} |\theta_t| \sum_{i=-\infty}^{\infty} [|\nu_{t-i}| + |\nu_{t+i}|]$$

$$\leqslant C \sum_{t>s} |\theta_t| \leqslant C[\sum_{t>s} t^{-2}]^{1/2} [\sum_{t>s} t^2 \theta_t^2]^{1/2} \leqslant C s^{-1/2}. \qquad (6.139)$$

Using the obtained relations, together with (6.112) and (6.115), we get for $j > s$

$$B_{j11} = (n-2)^{-1} \sum_{i,t=1}^{J_n} [2^{-1/2}(\kappa_{i-t} + \kappa_{i+t}) - \kappa_i \kappa_t]$$

$$\times [2^{-1}(b_{j-i}(j,s) + b_{j+i}(j,s))(b_{j-t}(j,s) + b_{j+t}(j,s))] = o_n(1)n^{-1}. \qquad (6.140)$$

Now consider B_{j12} and write,

$$B_{j12} = \Big[\sum_{i>J_n} \kappa_i \int_0^1 \frac{(F^X(z) - F_{-j}(z))\varphi_j(z)\varphi_i(z)}{f^Z(z)} dz \Big]^2$$

$$\leqslant \sum_{i>J_n} \kappa_i^2 \sum_{i>J_n} [b_{j-i}^2(j,s) + b_{j+i}^2(j,s)]$$

$$\leqslant C J_n^{-2}(J_n - j)^{-2} \leqslant C[\ln(s)]^{-4} n^{-4/3}. \qquad (6.141)$$

Using obtained results in (6.135) we conclude that

$$D_{j11}^* \leqslant C n^{-1} I(j \leqslant s) + o_n(1) n^{-1} I(j > s). \qquad (6.142)$$

Now let us consider a term with $W_2(Z_{n-1})W_2(Z_n)$ in (6.133). Using (6.132) and that $\tilde{h}'(z)$ is a projection estimate of $f^Z(z)$ with the bias (6.121), we get with the help of Cauchy and Cauchy-Schwarz inequalities,

$$\mathbb{E}_{f^Z}\Big\{ \prod_{l=n-1}^n \frac{W_2(Z_l)\varphi_j(Z_l)}{[f^Z(Z_l)]^2} [\frac{(n-2)}{n} f^Z(Z_l) - \tilde{h}'(Z_l)]\Big\}$$

$$\leqslant C\Big[\mathbb{E}_{f^Z}\{[W_2(Z_l)]^4\}\mathbb{E}\{[\frac{n-2}{n}f^Z(Z_l) - \widetilde{h}'(Z_l)]^4\}\Big]^{1/2} = o_n(1)n^{-1}. \quad (6.143)$$

Finally, we need to consider a cross-product term with $W_1(\Delta_{n-1}, Z_{n-1})W_2(Z_n)$,

$$D_{j13} := \mathbb{E}_{f^z}\Big\{\frac{(\Delta_{n-1} - F_{-j}(Z_{n-1}))\varphi_j(Z_{n-1})[\frac{n-2}{n}f^Z(Z_{n-1}) - \widetilde{h}'(Z_{n-1})]}{[f^Z(Z_{n-1}]^2}$$

$$\times \frac{W_2(Z_n))\varphi_j(Z_n)[\frac{(n-2)}{n}f^Z(Z_n) - \widetilde{h}'(Z_n)]}{[f^Z(Z_n)]^2}\Big\}. \quad (6.144)$$

Using Cauchy-Schwarz inequality, (6.102) and (6.132), we get a rough but sufficient relation

$$D_{j13}^2 \leqslant C\mathbb{E}_{f^z}\{[W_2(Z_n)]^2\}\mathbb{E}_{f^z}\{[\int_0^1 (\frac{(n-2)}{n}f^Z(z) - \widetilde{h}'(z))^2 dz]^2\}$$

$$= o_n(1)n^{-2} \quad (6.145)$$

Combining obtained results we verify (6.129), and then validity of Theorem 3.2 follows from the above-presented proof of Theorem 3.1. Let us also note that the presented proof justifies using a plug-in density estimate.

Proof of Corollary 4.1. The verified assertion follows from [12] and the already proved sharp-minimaxity of the blockwise-shrinkage estimator (3.9). The interested reader may also follow lines in the proofs of Theorems 3.1 and 3.2, and establish validity of Corollary 4.1 directly without reference on [12].

Proof of Theorem 4.1. Let us begin with a general remark. The proposed density estimator is "identical" to the cdf estimator apart of: (i) Using the sine basis for calculating statistics $\widetilde{\mathcal{Z}}_k$, $\widehat{\mathcal{Z}}_k$ and $\widehat{d}$; (ii) The new Fourier estimates $\check{\zeta}_j$ and $\widehat{\zeta}_j$. To address the first issue, let us present classical formulas for elements $\varphi_j(x)$ of the cosine basis (1.2) used in the proofs of Theorems 3.1 and 3.2, and then complement them by corresponding formulas for elements $\psi_j(x) := 2^{1/2}\sin(\pi jx)$ of the sine basis. We have

$$\varphi_j(x)\varphi_i(x) = 2^{-1/2}[\varphi_{j-i}(x) + \varphi_{j+i}(x)] \quad (6.146)$$

versus

$$\psi_j(x)\psi_i(x) = 2^{-1/2}[\varphi_{j-i}(x) - \varphi_{j+i}(x)], \quad (6.147)$$

and

$$\varphi_j^2(x) = 1 + 2^{-1/2}\varphi_{2j}(x) \quad (6.148)$$

versus

$$\psi_j^2(x) = 1 - 2^{-1/2}\varphi_{2j}(x). \quad (6.149)$$

Another useful trigonometric formula is

$$\psi_j(x)\varphi_i(x) = 2^{-1/2}[\psi_{j-i}(x) + \psi_{j+i}(x)]. \quad (6.150)$$

In what follows we also will need an analog of relation (2.2.5) in [13] which was used in the proof of Theorem 3.1. Let $g(x)$, $x \in [0, 1]$ be a function with bounded derivative, then

$$|\int_0^1 g(x)\psi_j(x)dx| \leqslant Cj^{-1}. \tag{6.151}$$

The inequality is proved via integration by parts.

The new estimators of the density Fourier coefficients ζ_j are analyzed in the presented below two lemmas. We begin with a proposition which is similar to Lemma 6.1.

Lemma 6.7. *Let Assumption 1 hold. Then*

$$\mathbb{E}_{F^X}\{\check{\zeta}_j\} = \zeta_j, \tag{6.152}$$

and

$$\mathbb{V}_{F^X}(\check{\zeta}_j) = n^{-1}(\pi j)^2 \int_0^1 F^X(x)[f^Z(x)]^{-1}\psi_j^2(x)dx \leqslant Cj^2n^{-1}. \tag{6.153}$$

Note how (6.152) and (6.153) mimic results (6.1) and (6.2) of Lemma 6.1.

Proof of Lemma 6.7. For the expectation we write using (4.4),

$$\mathbb{E}_{F^X}\{\check{\zeta}_j\} = \mathbb{E}\{\varphi_j(1) + (\pi j)\frac{\Delta\psi_j(Z)}{f^Z(Z)}\}$$

$$= \varphi_j(1) + (\pi j)\mathbb{E}_{F^X}\{\Delta\frac{\psi_j(Z)}{f^Z(Z)}\} = \zeta_j.$$

This verifies (6.152) and tells us that $\check{\zeta}_j$ is unbiased estimate of ζ_j. For the variance we get using (2.1),

$$\mathbb{V}_{F^X}(\check{\zeta}_j) = n^{-1}(\pi j)^2\mathbb{E}_{F^X}\{[\Delta\psi_j(Z)/f^Z(Z)]^2\}$$

$$= n^{-1}(\pi j)^2\int_0^1 F^X(x)[f^Z(x)]^{-1}\psi_j^2(x)dx.$$

This and Assumption 1 verify (6.153). Lemma 6.7 is proved.

Our next proposition is analog of Lemma 6.2.

Lemma 6.8. *Let assumption of Theorem 4.1 hold and $F^X \in \mathcal{F}(\alpha, Q)$. Then*

$$|\mathbb{E}_{F^X}\widehat{\zeta}_j\} - \zeta_j| \leqslant Cjsn^{-1}, \tag{6.154}$$

and

$$\mathbb{E}_{F^X}\{(\widehat{\zeta}_j - \zeta_j)^2\} \leqslant (\pi j)^2d[1 + C(j^{-1} + s^{-1})]. \tag{6.155}$$

Here $d = \int_0^1 F^X(x)(1 - F^X(x)[f^Z(z)]^{-1}dx$ is the coefficient of difficulty (2.6).

Proof of Lemma 6.8. In the proof we are following steps of the proof of Lemma 6.2 to highlight similarity between analysis of the two Fourier estimates. This also will allow us to highlight differences between the proofs.

Using the fact that statistic $\widehat{F}_{-j}(x)$ is used to estimate function $F^*_{-j}(x)$, which in its turn approximates $F^X(x)$, we write,

$$\widehat{\zeta}_j = \left[\varphi_j(1) + \frac{\pi j}{n}\sum_{l=1}^{n}\frac{(\Delta_l - F^*_{-j}(Z_l))\varphi_j(Z_l)}{f^Z(Z_l)}\right]$$

$$+ \frac{\pi j}{n}\sum_{l=1}^{n}\frac{(F^*_{-j}(Z_l) - \widehat{F}_{-j}(Z_l))\psi_j(Z_l)}{f^Z(Z_l)} =: \widehat{\zeta}_{1j} + \widehat{\zeta}_{2j}. \qquad (6.156)$$

We are considering the two terms on the right side of (6.156) in turn. Using (2.1), (4.4) and a straightforward calculation we get,

$$E_{F^X}\{\widehat{\zeta}_{1j}\} = \varphi_j(1) + (\pi j)\int_0^1 [F^X(x) - F^*_{-j}(x)]\psi_j(x)dx$$

$$= \varphi_j(1) + (\pi j)\int_0^1 F^X(x)\psi_j(x)dx = \zeta_j. \qquad (6.157)$$

Thus $\widehat{\zeta}_{1j}$ is unbiased estimate of ζ_j. Using (2.1) and $\Delta_l^2 = \Delta_l$ we can bound from above MSE of $\widehat{\zeta}_{1j}$,

$$\mathbb{E}_{F^X}\{(\widehat{\zeta}_{1j} - \zeta_j)^2\} \leqslant \frac{(\pi j)^2}{n}\mathbb{E}_{F^X}\left\{\left((\Delta_l - F^*_{-j}(Z_l))\psi_j(Z_l)[f^Z(Z_l)]^{-1}\right)^2\right\}$$

$$= \frac{(\pi j)^2}{n}\mathbb{E}_{F^X}\left\{(\Delta_l^2 - 2\Delta_j F^*_{-j}(Z_l) + [F^*_{-j}(Z_l)]^2)\psi_j^2(Z_l)[f^Z(Z_l)]^{-2}\right\}$$

$$= \frac{(\pi j)^2}{n}\int_0^1 [F^X(x) - 2F^X(x)F^*_{-j}(x) + [F^*_{-j}(x)]^2]\psi_j^2(x)[f^Z(x)]^{-1}dx. \quad (6.158)$$

Using $F^*_{-j}(x) = F^X(x) + [F^*_{-j}(x) - F^X(x)]$ we may continue,

$$\mathbb{E}_{F^X}\{(\widehat{\zeta}_{1j} - \zeta_j)^2\} \leqslant \frac{(\pi j)^2}{n}\int_0^1 [F^X(x)(1 - F^X(x))]\psi_j^2(x)[f^Z(x)]^{-1}dx$$

$$+ \frac{(\pi j)^2}{n}\left|\int_0^1 [-2F^X(x)(F^*_{-j}(x) - F^X(x))]\psi_j^2(x)[f^Z(x)]^{-1}dx\right|$$

$$+ \frac{(\pi j)^2}{n}\int_0^1 [(F^*_{-j}(x))^2 - (F^X(x))^2]\psi_j^2(x)[f^Z(x)]^{-1}dx =: u_1 + u_2 + u_3. \quad (6.159)$$

We are evaluating the three terms on the right side of (6.159) in turn. Using (6.149), differentiability of both $F^X(x)$ and $f^Z(x)$, and relation (2.2.5) in [13], we get

$$u_1 \leqslant \frac{(\pi j)^2}{n}[d + c_j j^{-1}], \text{ where } \sum_{j=1}^{\infty} c_j^2 < C. \qquad (6.160)$$

Now we evaluate u_2 using several technical relations. First,

$$F^X(x) - F^*_{-j}(x) = \zeta_j \psi_j(x) I(j \leqslant s) + \sum_{r > s} \zeta_r \psi_r(x). \qquad (6.161)$$

Second, using (6.149) we write

$$\zeta_j \int_0^1 F^X(x) \psi_j(x) \psi_j^2(x) [f^Z(x)]^{-1} dx I(j \leqslant s)$$

$$= \zeta_j \int_0^1 F^X(x) [f^Z(x)]^{-1} [\psi_j(x) - \psi_j(x) 2^{-1/2} \varphi_{2j}(x)] dx I(j \leqslant s)$$

$$=: \zeta_j \nu_j I(j \leqslant s). \qquad (6.162)$$

To continue evaluation of the right side of (6.162) we use (6.150) and note that according to that formula we get $\psi_j(x) \varphi_{2j}(x) = 2^{-1/2} [\psi_{3j}(x) - \psi_j(x)]$. This and (6.151) yield $|\zeta_j \nu_j| I(j \leqslant s) \leqslant C(j+1)^{-2} I(j \leqslant s)$.

Third, using the Cauchy-Schwarz inequality and the assumption of Lemma 6.8 we get,

$$\sum_{r > s} |\zeta_r \int_0^1 F^X(x) \psi_r(x) \psi_j^2(x) [f^Z(x)]^{-1} dx| \leqslant [\sum_{r > s} \zeta_r^2]^{1/2}$$

$$\times \Big[\sum_{r > s} [\int_0^1 F^X(x) \psi_r(x) \psi_j^2(x) [f^Z(x)]^{-1} dx]^2 \Big]^{1/2} \leqslant C s^{-1}. \qquad (6.163)$$

Using these results we establish

$$u_2 \leqslant C \frac{(\pi j)^2}{n} [(j+1)^{-2} + s^{-1}]. \qquad (6.164)$$

To evaluate u_3 we begin with based on (6.161) relation

$$[F^*_{-j}(x)]^2 = [F^X(x)]^2 - 2F^X(x) [\zeta_j \psi_j(x) I(j \leqslant s) + \sum_{r > s} \zeta_r \psi_r(x)]$$

$$+ [\zeta_j \psi_j(x) I(j \leqslant s) + \sum_{r > s} \zeta_r \psi_r(x)]^2.$$

Using it we obtain the following upper bound,

$$u_3 \leqslant C \Big[u_2 + \int_0^1 [\zeta_j \psi_j(x) I(j \leqslant s) + \sum_{r > s} \zeta_r \psi_r(x)]^2 \psi_j^2(x) [f^Z(x)]^{-1} dx \Big]$$

$$\leqslant C \Big[u_2 + \zeta_j^2 I(j \leqslant s) + [\sum_{r > s} |\zeta_r|]^2 \Big] \leqslant C[(j+1)^{-2} + Cs^{-1}]. \qquad (6.165)$$

Now we can use the obtained upper bounds for u_1, u_2 and u_3 in (6.159) and get

$$\mathbb{E}_{F^X} \{ (\widehat{\zeta}_{1j} - \zeta_j)^2 \} \leqslant \frac{(\pi j)^2}{n} d$$

$$+ C\frac{(\pi j)^2}{n}[c_j(j+1)^{-1} + (j+1)^{-2} + s^{-1}], \quad \text{where } \sum_{j=0}^{\infty} c_j^2 < C. \qquad (6.166)$$

This ends our analysis of $\widehat{\zeta}_{1j}$.

Now we analyze two moments of statistic $\widehat{\zeta}_{2j}$ introduced in (6.156). Our aim is to show that this statistic is negligible with respect to $\widehat{\zeta}_{1j}$. Again we begin with several technical results. Set $S_{-j} := \{1, \ldots, s\}\backslash\{j\}$ for $j \leqslant s$ and $S_{-j} := \{1, \ldots, s\}$ for $j > s$, and $\mathcal{N}_{-l} := \{1, \ldots, n\}\backslash\{l\}$. First, using the Parseval identity, definition of $\breve{F}_{-j}$ and F^*_{-j}, and Lemma 6.7 we can write

$$\mathbb{E}_{F^X}\Big\{ \int_0^1 (\widehat{F}_{-j}(z) - F^*_{-j}(z))^2 dz \Big\} = \sum_{i \in S_{-j}} \mathbb{E}\{\breve{\zeta}_i - \zeta_i)^2\} \leqslant Cs^3n^{-1}. \qquad (6.167)$$

Second, introduce a decomposition

$$\widehat{F}_{-j}(Z_l) = n^{-1} \sum_{k \in \mathcal{N}_{-l}} \Delta_k [f^Z(Z_k)]^{-1} \sum_{i \in S_{-j}} \psi_i(Z_k)\psi_i(Z_l)$$

$$+ n^{-1}\Delta_l[f^Z(Z_l)]^{-1} \sum_{i \in S_{-j}} \psi_i^2(Z_l) =: \widehat{F}_1(Z_l) + \widehat{F}_2(Z_l). \qquad (6.168)$$

Then using orthogonality of elements of the sine basis we can write for $\widehat{F}_1(Z_l)$,

$$\mathbb{E}_{F^X}\{[F^*_{-j}(Z_l) - \widehat{F}_1(Z_l)][f^Z(Z_l)]^{-1}\psi_j(Z_l)|\{(\Delta_k, Z_k), \ k \in \mathcal{N}_{-l}\}\} = 0. \qquad (6.169)$$

Third, Assumption 1 allows us to write

$$|\mathbb{E}_{F^X}\{\widehat{F}_2(Z_l)\psi_j(Z_l)[f^Z(Z_l))]^{-1}\}|$$

$$= n^{-1}|\mathbb{E}_{F^X}\{\Delta_l[f^Z(Z_l)]^{-2} \sum_{i \in S_{-j}} \psi_i^2(Z_l)\psi_j(Z_l)\}| \leqslant Csn^{-1}. \qquad (6.170)$$

Combining the obtained results yields

$$|\mathbb{E}_{F^X}\{\widehat{\zeta}_{2j}\}| \leqslant Cjsn^{-1}.$$

This and (6.157) verify (6.154).

We are left with evaluating the second moment of $\widehat{\zeta}_{2j}$. We again do that via establishing several technical results. First,

$$\mathbb{E}_{F^X}(F^*_{-j}(Z_l) - \widehat{F}_1(Z_l))^2 \leqslant Csn^{-1}. \qquad (6.171)$$

Second,

$$\mathbb{E}_{F^X}\Big\{ \Big(n^{-1} \sum_{l=1}^n \widehat{F}_2(Z_l)\psi_j(Z_l)[f^Z(Z_l)]^{-1}\Big)^2 \Big\} \leqslant Cn^{-2}s^2. \qquad (6.172)$$

Third, consider two integers l and m such that $l \neq m$ and $l, m \in \{1, 2, \ldots, n\}$, and write,

$$[F^*_{-j}(Z_l) - \widehat{F}_{-j}(Z_l)][F^*_{-j}(Z_m) - \widehat{F}_{-j}(Z_m)]$$

$$= [F^*_{-j}(Z_l) - \widehat{F}_1(Z_l) - \widehat{F}_2(Z_l)][F^*_{-j}(Z_m) - \widehat{F}_1(Z_m) - \widehat{F}_2(Z_m)]$$

$$= [F^*_{-j}(Z_l) - \widehat{F}_1(Z_l)][F^*_{-j}(Z_m) - \widehat{F}_1(Z_m)] - [F^*_{-j}(Z_l) - \widehat{F}_1(Z_l)]\widehat{F}_2(Z_m)$$

$$- \widehat{F}_2(Z_l)[F^*_{-j}(Z_m) - \widehat{F}_1(Z_m)] + \widehat{F}_2(Z_l)\widehat{F}_2(Z_m)$$

$$=: W_1 + W_2 + W_3 + W_4. \tag{6.173}$$

Using these notations we can write for any $1 \leqslant l < m \leqslant n$,

$$\mathbb{E}_{F^X}\{\widehat{\zeta}^2_{2j}\} \leqslant Cj^2 n^{-1} \mathbb{E}_{F^X}\{(F^*_{-j}(Z_l) - \widehat{F}_{-j}(Z_l))^2\}$$

$$+ C\mathbb{E}_{F^X}\{[W_1 + W_2 + W_3 + W_4]\psi_j(Z_l)\psi_j(Z_m)[f^Z(Z_l)f^Z(Z_m)]^{-1}\}. \tag{6.174}$$

The first term on the right side of (6.174) is at most $Cj^2 s n^{-2}$ due to (6.171). To evaluate the second term set

$$\widehat{F}_1(Z_l, -m) := \widehat{F}_1(Z_l) - n^{-1}\Delta_m[f^Z(Z_m)]^{-1}\sum_{i \in S_{-j}}\psi_i^2(Z_m)$$

$$=: \widehat{F}_1(Z_l) - \widehat{F}_1(Z_l, m). \tag{6.175}$$

Note that $\widehat{F}_1(Z_l, -m)$ does not depend on $(\Delta_l, Z_l, \Delta_m, Z_m)$ and $|\widehat{F}_1(Z_l, m)| \leqslant Csn^{-1}$ almost sure. Write,

$$V := \mathbb{E}_{F^X}\{W_1\psi_j(Z_l)\psi_j(Z_m)[f^Z(Z_l)f^Z(Z_m)]^{-1}\}$$

$$= \mathbb{E}_{F^X}\Big\{[F^*_{-j}(Z_l) - \widehat{F}_1(Z_l, -m) - \widehat{F}_1(Z_l, m)][F^*_{-j}(Z_m) - \widehat{F}_1(Z_m, -l) - \widehat{F}_1(Z_m, l)]$$

$$\times \psi_j(Z_l)\psi_j(Z_m)[f^Z(Z_l)f^Z(Z_m)]^{-1}\Big\}.$$

To continue we need several relations. First, using (6.169) allows us to write,

$$\mathbb{E}_{F^X}\Big\{[F^*_{-j}(Z_l) - \widehat{F}_1(Z_l, -m)][F^*_{-j}(Z_m) - \widehat{F}_1(Z_m, -l)]$$

$$\times \psi_j(Z_l)\psi_j(Z_m)[f^Z(Z_l)f^Z(Z_m)]^{-1}\Big\} = 0.$$

Second, note that $|[\widehat{F}_1(Z_l, m)| \leqslant Csn^{-1}$ almost sure. Third, the Cauchy-Schwarz inequality allows us to write,

$$|\mathbb{E}_{F^X}\{[F^*_{-j}(Z_l) - \widehat{F}_1(Z_l, -m)]\widehat{F}_1(Z_m, l)\}|$$

$$\leqslant [\mathbb{E}_{F^X}\{[F^*_{-j}(Z_l) - \widehat{F}_1(Z_l, -m)]^2\}E_{F^X}\{[\widehat{F}_1(Z_m, l)]^2\}]^{1/2}$$

$$\leqslant C[(sn^{-1}(s^2n^{-2}]^{1/2} \leqslant Cs^{3/2}n^{-3/2}.$$

These relations allow us to conclude that $|V| \leqslant Cs^{3/2}n^{-3/2}$.

Next, using the Cauchy-Schwarz inequality we get

$$\mathbb{E}_{F^X}\{|W_2|\} \leqslant [\mathbb{E}_{F^X}\{(F^*_{-j}(Z_l) - \widehat{F}_1(Z_l))^2\}\mathbb{E}_{F^X}\{[\widehat{F}_2(Z_m)]^2\}]^{1/2} \leqslant Cs^2 n^{-3/2}.$$

Due to symmetry the same inequality holds for W_3, and $\mathbb{E}_{F^X}\{|W_4|\} \leqslant Cs^2n^{-2}$. Combining the above-presented relations we get

$$\mathbb{E}_{F^X}\{\widehat{\zeta}_{2j}^2\} \leqslant Cj^2s^2n^{-3/2}. \tag{6.176}$$

Lemma 6.8 is proved.

We have established that the considered Fourier estimates $\check{\zeta}_j$ and $\widehat{\zeta}_j$ satisfy the desired statistical properties mimicking properties outlined in Lemmas 6.1 and 6.2 for Fourier estimates of cdf Fourier coefficients. The rest of the proof of Theorem 4.1 follows the same steps as in the proof of Theorems 3.1 and 3.2, the steps are based on using relations (6.146)–(6.151) for sines in place of cosines, and finishing the proof presents no new technical complications. Theorem 4.1 is verified.

Proof of Lemma 4.1. Using (6.156) we can write,

$$\mathbb{E}_{F_X}\{(\widehat{\zeta}_j - \zeta_j)^2\} = \mathbb{E}_{F^X}\{(\widehat{\zeta}_{1j} - \zeta_j)^2\} + 2\mathbb{E}_{F_X}\{(\widehat{\zeta}_{1j} - \zeta_j)\widehat{\zeta}_{2j}\} + \mathbb{E}_{F_X}\{\widehat{\zeta}_{2j}^2\}. \tag{6.177}$$

Now we are evaluating the expectations on the right side of (6.177) in turn. Formulas (6.158) and (6.159) yield

$$|\mathbb{E}_{F^X}\{(\widehat{\zeta}_{1j} - \zeta_j)^2\} - \frac{(\pi j)^2}{n}\int_0^1 F^X(x)(1 - F^X(x))\psi_j^2(x)[f^Z x]^{-1}dx|$$

$$\leqslant u_2 + u_3 + \zeta_j^2. \tag{6.178}$$

Using (6.149) and (6.151) we conclude that

$$\int_0^1 F^X(x)(1 - F^X(x))\psi_j^2(x)[f^Z(x)]^{-1}dx$$

$$= d - \int_0^1 F^X(x)(1 - F^X(x))f^Z(x)2^{-1/2}\varphi_{2j}(x)dx = d + o_j(1).$$

Combining this relation with upper bound (6.164) for u_2 and upper bound (6.165) for u_3 we get

$$\mathbb{E}_{F^X}\{(\widehat{\zeta}_{1j} - \zeta_j)^2\} = \frac{(\pi j)^2}{n}d[1 + o_j(1) + o_n(1)]. \tag{6.179}$$

Further, it is established in (6.176) that $\mathbb{E}_{F^X}\{\widehat{\zeta}_{2j}^2\} \leqslant Cj^2s^2n^{-3/2} = o_n(1)j^2n^{-1}$. Using this, (6.179) and the Cauchy-Schwarz inequality on the right side of (6.177) proves Lemma 4.1.

7. Conclusion and further research

Current Status Censoring (CSC) is a familiar sampling procedure when instead of observing a lifetime of interest X only its status $\Delta = I(X \leqslant Z)$ is available

at a monitoring time Z. In other words, in place of a classical sample of direct observations from X, under CSC a sample from (Δ, Z) is available. CSC sampling is a popular technique due to its simplicity, and in many applications it is the only possibility to get information about an underlying lifetime of interest. It is also well known that CSC makes estimation of the density and cdf ill-posed and dramatically slows down rate of a risk convergence.

Whenever a problem is ill-posed, it is important to investigate not only the rate but also a minimal constant of a risk convergence. The latter is due to the fact that this constant sheds light on an ill-posed problem and because for small samples we may see only the onset of ill-posedness. Recently [16] obtained sharp-minimax lower bounds for density and cdf oracle-estimators, and this paper proposed a data-driven and robust sharp-minimax estimator that attains the oracle's lower bound. Further, the estimator may be used for small samples and it is tested on real and simulated examples.

There are many interesting and practically important extensions of the considered CSC setting. Missing is a typical complications in data analysis, and then approaches of book [15] may be tested. Estimation of joint and conditional densities is another important topic to consider. CSC regression, including a functional regression, is another important applied topic to consider. Let us also mention a research devoted to more general cases of interval censoring. Here no results on sharp minimax estimation are known even for the interval censoring case II when one observes a triplet (L, U, Δ) where $L < U$ are two monitoring times and the status $\Delta = -1$ if $X \leqslant L$, $\Delta = 0$ if $L < X \leqslant U$, and $\Delta = 1$ otherwise, see [5,6,18]. It is an interesting, challenging and open problem to expand the obtained CSC sharp-minimax results to a general interval censoring data.

Now let us formulate several open problems for the considered CSC setting and make several last remarks. (i) The coefficient of difficulty (2.6) indicates that there is a serious issue when the integral $\int_0^x (1 - F^X(u))[f^Z(u)]^{-1} du$ diverges as x increases due to a light tail of $f^Z(u)$. This issue is paramount to address for extending the developed theory to unbounded support of X. While cases of unbounded lifetimes are rare in practical applications, they are of a great theoretical interest. (ii) In Assumption 1 the equality $\int_0^1 f^Z(z) dz = 1$ can be relaxed and the support of Z may be larger than $[0, 1]$. This case does not present a theoretical complication and the results still hold only now we are estimating $f^Z(z)$ over interval $[0, 1]$ using $Z_l \in [0, 1]$, see the corresponding density estimate in [13]. What is of interest here is to estimate the support of X. (iii) The case when the support of X is larger than the support of Z yields inconsistent estimation, but a sharp estimation over the support of Z is still possible. (iv) An interesting and complicated setting is when X and Z are unbounded. No sharp-minimax theory is known for this case. (v) An important applied task is to create a user-friendly R package for CSC density and cdf estimation.

8. Acknowledgements

References

[1] Abramovich, F. and Ritov, Y. (2013). *Elements of Statistical Decision Theory.* Boca Raton: Chapman and Hall.
[2] Anderson-Bergman, C. and Yu, Y. (2016). Computing the log concave npmle for interval censored data. *Statistics and Computing* 26 813–826.
[3] Braun, J., Duchesne, T. and Stafford, J. (2005). Local likelihood density estimation for interval censored data. *The Canadian Journal of Statistics* 33 39-60.
[4] Becker, D., Braun, W. and White, B. (2017). Interval-censored unimodal kernel density estimation via data sharpening. *Journal of Statistical Computation and Simulation* 87 2023-2037.
[5] Birge, L. (1999). Interval censoring: a nonasymptotic point of view. *Mathematical Methods of Statistics* 8 285-298.
[6] Bouaziz, O., Brunel, E. and Comte, F. (2019). Nonparametric survival function estimation for data subject to interval censoring, case 2. *Journal of Nonparametric Statistics* 31 952-987.
[7] Brunel E. and Comte F. (2009). Cumulative distribution function estimation under interval censoring case. *Electronic Journal of Statistics* 3 1-24.
[8] Chagny G. (2015). Adaptive warped kernel estimators. *Scandinavian Journal of Statistics* 42 336-360.
[9] Chen, D., Sun, J. and Peace, K. (2012). *Interval-Censored Time-to-Event Data: Methods and Applications.* Boca Raton: CRC Press.
[10] Cheng, G., Wang, X. (2011). Semiparametric additive transformation model under current status data. *Electronic Journal of Statistics* 5 1735–1764.
[11] Efromovich, S. (1985). Nonparametric estimation of a density with unknown smoothness. *Theory of Probability and its Applications* 30 557-568.
[12] Efromovich, S. (1998). Simultaneous sharp estimation of functions and their derivatives. *Annals of Statistics* 26 273-278.
[13] Efromovich, S. (1999). *Nonparametric Curve Estimation.* New York: Springer.
[14] Efromovich, S. (2007). Conditional density estimation. *Annals of Statistics* 35 2504-2535.
[15] Efromovich, S. (2018). *Missing and Modified Data in Nonparametric Estimation with R Examples.* Boca Raton: Chapman & Hall.
[16] Efromovich, S. (2021). Sharp minimax distribution estimation for current status censoring with or without missing. *Annals of Statistics* 49 568-589.
[17] Golubev, Y. (1991). LAN in problems of non-parametric estimation of func-

tions and lower bounds for risks. *Problems of Information Transmission* 36 152-157.

[18] Gill, R. and Levit, B. (1995). Application of the Van Trees inequality: a Bayesian Cramer-Rao bound. *Bernoulli* 1 59-79.

[19] Groeneboom, P., Jongbloed, G. and Witte, B.I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *The Annals of Statistics* 38 352-387.

[20] Groeneboom, P. and Ketelaars, T. (2011). Estimators for the interval censoring problem. *Electronic Journal of Statistics* 5 1797-1845.

[21] Groeneboom, P. and Jongbloed, G. (2014).*Nonparametric Estimation Under Shape Constraints: Estimators, Algorithms and Asymptotics.* Cambridge Series in Statistical and Probabilistic Mathematics 38. New York: Cambridge Univ.

[22] Groeneboom, P. and Hendrickx, K. (2017). The nonparametric bootstrap for the current status model. *Electronic Journal of Statistics* 11 3446-3484.

[23] Jewell, N. and van der Laan, M. (2004a). Current status data: review, recent developments and open problems. In: *Advances in Survival Analysis*, eds. N. Balakrishnan and C.R. Rao, Elsevier, Amsterdam, 625-643.

[24] Jewell, N. and van der Laan, M. (2004b). Case-control current status data. *Biometrika* 91 529-541.

[25] Klein, J., van Houwelingen, H., Ibrahim, J. and Sheike, T. (2014). *Handbook of Survival Analysis.* Boca Raton: Chapman & Hall.

[26] Kooperberg, C. and Stone, C. (1992). Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics* 1 3010328.

[27] Li, S., Hu, T., Zhao,X. and Sun, J. (2018). A class of semiparametric transformation cure models for interval-censored failure time data. *Computational Statistics and Data Analysis*, https://doi.org/10.1016/j.csda.2018.09.008.

[28] Malov, S. (2019). Nonparametric estimation for a current status right-censored data model. *Statistica Neerlandica.* https://doi.org/10.1111/stan.12180

[29] Plancade, S. (2013). Adaptive estimation of the conditional cumulative distribution function from current status data. *Journal of Statistical Planning and Inference* 143 1466-1485.

[30] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

[31] Sikander, G. and Anwar, S. (2019). Driver Fatigue Detection Systems: A Review. *IEEE Transactions on Intelligent Transportation Systems* 20 2339-2352.

[32] Sun, J. and Zhao, X. (2013). *The Statistical Analysis of Interval-Censored Failure Time Data.* New York: Springer.

[33] van de Geer, S. (1993). Hellinger-consistency of certain nonparametric likelihood estimators. *Annals of Statistics* 21, 14–44. MR1212164

[34] van Es, B. and Graafland, C. (2017). Nonparametric Kernel Density Estimation for Univariate Current Status Data. Online Manuscript, arXiv: 1707.00544v1

[35] Wasserman, L. (2006). *All of Nonparametric Statistics.* New Your: Springer.

[36] Wellner, J. A. (1995). Interval censoring, case 2: alternative hypotheses. Analysis of censored data. *IMS Lecture Notes Monogr. Ser.* 27 271-291.

[37] Yang, S. (2000). Functional estimation under interval censoring case 1. *Journal of Statistical Planning and Inference*, 89, 135-144.

[38] Xing, Y., Lv, C., Cao, D., Wang, H. and Zhao, Y. (2018). Driver workload estimation using a novel hybrid method of error reduction ratio causality and support vector machine. *Measurement*, 114, 390-397.