Electronic Journal of Statistics
Vol. 12 (2018) 1299–1329
ISSN: 1935-7524
https://doi.org/10.1214/18-EJS1423

Conditional kernel density estimation for some incomplete data models

Ting Yan and Liangqiang Qu

Central China Normal University
Wuhan, China
e-mail: tingyanty@mail.ccnu.edu.cn; qulianq@mail.ccnu.edu.cn

Zhaohai Li

The George Washington University
Washington DC, USA
e-mail: zli@gwu.edu

Ao Yuan

Georgetown University
Washington DC, USA
e-mail: ay312@georgetown.edu

Abstract: A class of density estimators based on observed incomplete data are proposed. The method is to use a conditional kernel, defined as the expectation of a given kernel for the complete data conditioning on the observed data, to construct the density estimator. We study such kernel density estimators for several commonly used incomplete data models and establish their basic asymptotic properties. Some characteristics different from the classical kernel estimators are discovered. For instance, the asymptotic results of the proposed estimator do not depend on the choice of the kernel $k(\cdot)$. Simulation study is conducted to evaluate the performance of the estimator and compared with some exising methods.

Keywords and phrases: Conditional kernel, density estimate, incomplete data model, NPMLE.

Received November 2016.

Contents

1. Introduction

Estimating the density function is one of the fundamental problems in nonparametric statistics. Many approaches are proposed to address this issue, such as the kernel method, histogram, orthogonal series method and wavelet method. Among them, the kernel estimator proposed by Rosenblatt[1] and Parzen[2], is perhaps most popular. Devroye[3] derived basic asymptotic properties of the kernel estimator. Chai[4] proposed random window-width kernel and proved its consistency. For a comprehensive description of theoretical aspects on the kernel estimation, see the book by Rao.[5]. The monograph of Silverman[6] provided detailed accounts of various density estimation methods and their applications.

In many situations such as medical follow-up studies, clinical trials, economics and reliability studies, incomplete/missing data are frequently encountered. In these cases, the original data of interest are partially or completely missing and we only observe a functional part of them or their status and with accompanying data. Common models of incomplete/missing data include the left truncated, right censoring, doubly censoring, interval censoring of types I (or the current status data) and II, multiplicative censoring and convolution model.

For these types of data, estimates of the survival function, distribution function and density function has been extensively explored. Various estimators are proposed, including the Kaplan-Meier estimator of survival function for censored data [7], survival function estimator for doubly censored data [8], kernel density estimator with right censored data [9], nonparametric maximum likelihood estimators with truncated data [10], density and hazard rate estimation for censored data [11], density estimation with interval censoring data [12], survival function estimator for truncated and censored data [13], and ROC curve estimation for survival data [14].

The kernel smoothing type estimator is one of the commonly method for missing data. For example, for survival data, [15] introduced the kernel estimator of density f by

$$\hat{f}_n(x) = \frac{1}{h_n} \int K(\frac{x-y}{h_n}) dF_n(y),$$

where $F_n(\cdot)$ is an estimator of the corresponding distribution function based on the observed survival data. [16] derived asymptotic properties of a similar type. Dubnicka [17, 18] studied kernel density estimator with missing data by weighting the kernel over the estimated propensity scores. Ren [19] studied such kernel density estimators for doubly censored data based on the self-consistent estimators proposed by Turnbull [8]. Gine [20] derived the convergence rate of the difference between kernel smooth estimators with and without right-censoring using the Kaplan-Meier estimator. The multiplicative censoring model introduced by Vardi [21] is another missing type and has been studied by Bickel [22], van der Vaart [23] and references wherein. Vardi and Zhang[37] derived the asymptotic behavior of solutions of its nonparametric score equation. Asgharian et.al [24] established the large-sample properties of kernel density estimators.

Other common methods for density estimation in incomplete data models include the nonparametric maximum likelihood estimator (NPMLE), or taking the left/right derivative of an estimated distribution function. Huang and Wellner [25, 26] studied such estimators (with shape constraint such as monotone for the NPMLE). Such estimators are intuitively appealing and have nice asymptotic properties. However, they are piecewise linear and not continuous.

One method of density estimation with missing data is to estimate the hazard rate function $\lambda(x) = f(x)/(1 - F(x))$, where f and F are the density and distribution functions. When estimates $\hat{\lambda}_n(x)$ for $\lambda(x)$ and $\hat{F}_n(x)$ for $F(x)$ are available, estimate $f(x)$ by $\hat{f}_n(x) = \hat{\lambda}_n(x)(1 - \hat{F}_n(x))$. Another way is to estimate the cumulative hazard function $\Lambda(x)$ by $\hat{\Lambda}(x)$, such as the Nelson-Aalen estimator [27, 28], then estimate $\lambda(x)$ by smoothing $\hat{\Lambda}$, such as kernel method. These methods and the kernel smoothing methods mentioned above are the same in principle.

Here we propose a class of density estimator using the observed incomplete data. Instead of kernel smoothing an estimated distribution or survival function, we use the conditional kernel, which is the expectation of the kernel for the complete data conditioning on the observed data, and then construct the estimator based on the conditional kernel.

Formally, let the interested complete data are $X_1, \ldots, X_n$ from the underlying distribution $F(\cdot)$, but the complete data are unobserved. We only observe data $Y_1, \ldots, Y_n$ that implicitly contain information about the complete data. The aim is to estimate the density $f(\cdot)$ of the original data using the observed data $Y_1, \ldots, Y_n$. The common kernel density estimator using the complete data is $f_n(x) = (nh)^{-1} \sum_{i=1}^{n} k((X_i - x)/h)$, where the kernel $k(\cdot)$ is a given function (often a known density function) and h $(= h_n \to 0$ as $n \to \infty)$ is the bandwidth. This estimator can not be used since we do not observe the original complete data X_i's. Instead, we use a conditional kernel $K(x|F, h, Y) = h^{-1}\mathbb{E}_F[k(X - x)/h)|Y]$ based on the observed data, and construct the estimator of $f(x)$ as $f_n(x) = n^{-1} \sum_{i=1}^{n} K(x|F, h, Y_i)$. Since the conditional kernel $K(\cdot|F, h, Y)$ involves the underlying unknown distribution F, we plug in an estimator $\hat{F}_n$ based on the observed data such as NPMLE. This kernel estimator has some features different from the other methods, such as its asymptotic distribution does not depend on the sujectively chosen kernel, this is in contrast to most existing methods using kernel smoothing. As far as we know, the proposed method hasn't been seen elsewhere, except that Yuan et.al [29] used conditional kernel to construct U-statistics with missing data on a different topic.

In Section 2 we introduce the framework of the proposed estimators for five commonly used incomplete data models, type I and type II interval censoring, convolution model, double censoring model, and multiplicative censoring model, and investigate their basic asymptotic properties. Numerical simulation and comparison with other existing methods are provided in Section 3. Further discussion is given in Section 4. All proofs are given in the Appendix.

2. The proposed method

We first give a brief review of the kernel density estimator for complete data. Let $X_1, \dots, X_n$ be i.i.d. observations from density function $f(\cdot)$. The kernel estimator for original complete data is [1, 2, 30]

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^{n} k(\frac{X_i - x}{h}),$$

where the kernel $k(\cdot)$ is any given density function, and h $(= h_n \to 0$ as $n \to \infty)$ depending on n, is the bandwidth. The large-sample theories of $f_n(x)$ have been established by [2], [31], [32], [3] and among others. Chai [4] studied the case with data dependent bandwidth.

Density estimation for incomplete data has been widely studied, mostly by kernel smoothing some existing estimator of distribution function, survival function, or cumulative hazard function. These methods do not use the observed data directly. Although the NPMLE density estimator uses the observed data directly, it often needs some shape constraint and is non-smooth.

Below we introduce conditional kernel density estimators based on the observed data for five commonly used incomplete data models, interval censoring types I and II, convolution model, and investigate their large-sample properties. The results for double censoring and multiplicative censoring models are basically parallel, so we only given a brief presentation for these two models, with details given in a separate supplementary document. At the end of this section, we summarize the results for general incomplete data models.

2.1. Interval censoring type I

The interval censoring model type I is also called the current status model, see [12] for the background of this model. In this model, the original random variables are $(X, T) \in (R^+)^2$, where X and T are independent, with distribution functions F and G, and densities f and g with respect to the Lebesque measure on R^+. The observed incomplete data are $\{(T_i, 1_{[X_i<T_i]}) : i = 1, \dots, n\}$ i.i.d. with $(T, 1_{[X\le T]}) := (T, \delta)$, and $0 < P(\delta = 1) < 1$. The density-mass function of (T, δ) is

$$p_F(t, \delta) = F(t)^\delta (1 - F(t))^{1-\delta} g(t).$$

Remark 1. *The observed data is a sufficient statistic for G (or g), thus conditioning on the observed data, the resulting kernel is free of G (or g), so is the conditioning kernel density estimator. Thus, the estimator is adaptive, i.e., G (or g) is known or not does not affect the behavior of the estimator, even though G (or g) will appear in the asymptotic results. Moreover, since the data T_i $(i = 1, \dots, n)$ are fully observed, the underlying distribution (or density) G (or g) can be estimated by standard methods if needed.*

The same remark applies to the interval censoring model type II model, the double censoring model, and the multiplicative censoring model to be considered latter.

Let $Y_i = (T_i, \delta_i)$ $(i = 1, \dots, n)$ be the observed data. The aim is to estimate the density function $f(\cdot)$ of the unobserved X_i's. Since the X_i's are not directly observable, the existing kernel density estimator cannot be directly obtained. Instead, we define the conditional kernel based on the data Y, evaluated at x, as

$$K(x|F, h, Y) = \frac{1}{h}\mathbb{E}_F[k(\frac{X-x}{h})|Y].$$

For fixed F, h and Y, $K(\cdot|F, h, Y)$ is a density function, and hence a valid kernel. For this model, the conditional kernel is

$$K(x|F, h, Y) = \frac{1}{h}\int k(\frac{s-x}{h})\Big(\delta\frac{1_{[0,T]}(s)}{F(T)} + (1-\delta)\frac{1_{(T,\infty)}(s)}{1-F(T)}\Big)F(ds).$$

However, F is unknown. We use the NPMLE $\hat{F}_n$ of F based on the observed data $Y_1, \dots, Y_n$ as in [12] to replace F, and get

$$K(x|\hat{F}_n, h, Y) = \frac{1}{h}\int k(\frac{s-x}{h})\Big(\delta\frac{1_{[0,T]}(s)}{\hat{F}_n(T)} + (1-\delta)\frac{1_{(T,\infty)}(s)}{1-\hat{F}_n(T)}\Big)\hat{F}_n(ds). \quad (2.1)$$

We define the kernel estimator $f_n(x|\hat{F}_n)$ of $f(x)$ as

$$f_n(x|\hat{F}_n) = \frac{1}{n}\sum_{i=1}^{n} K(x|\hat{F}_n, h, Y_i). \quad (2.2)$$

For this model, it is known that $\hat{F}_n = F + O_p(n^{-1/3})$, and we hope that $f_n(x|\hat{F}_n)$ will have desirable asymptotic behavior.

Let $\mathbb{B}(h)$ be the two-sided Brownian motion, originating from zero, i.e., it is a zero-mean Gaussian process on R and the increment $\mathbb{B}(r) - \mathbb{B}(h)$ has variance $|r-h|$, and denote

$$A(t) = \left(\frac{4F(t)(1-F(t))f(t)}{g(t)}\right)^{1/3} \quad \text{and} \quad Z = \arg\min_h\{\mathbb{B}(h) + h^2\}.$$

Then under suitable conditions, it is known (as in [12])

$$\sup_t |\hat{F}_n(t) - F(t)| \to 0, \quad a.s., \qquad \text{and} \quad n^{1/3}(\hat{F}_n(t) - F(t)) \xrightarrow{D} A(t)Z.$$

To study the asymptotic behavior of the estimator, we need the notion of Hadamard differentiability. There are several different equivalent definitions of this notion, we adopt a simpler one as below. For a map $\phi : \mathbb{D} \mapsto \mathbb{E}$ between Banach spaces $\mathbb{D}$ and $\mathbb{E}$, ϕ is *Hadamard differentiable* at $g \in \mathbb{D}$ in the direction $h \in \mathbb{D}$, if there exists a map $\phi^{(1)} : \mathbb{D} \mapsto \mathbb{E}$ such that, for all sequences $h_n \to h$ and real numbers $t_n \searrow 0$,

$$\frac{\phi(g + t_n h_n) - \phi(g)}{t_n} \to \phi^{(1)}(h).$$

$\phi^{(1)}$ is called the first order Hadamard differential of ϕ in the direction h. Higher order Hadamard differential is defined similarly.

For fixed (h, Y), let $K^{(r)}(x|F, h, Y; A, \dots, A)$ be the r-th Hadamard differential of $K(x|F, h, Y)$ with respect to F in the direction A. For this model,

$$K^{(1)}(x|F,h,Y;A) = \frac{1}{h}\int k(\frac{s-x}{h})\Big(\delta\frac{1_{[0,T]}(s)[F(T)A(ds)-A(T)F(ds)]}{F^2(T)} + (1-\delta)\frac{1_{(T,\infty)}(s)[(1-F(T))A(ds)+A(T)F(ds)]}{(1-F(T))^2}\Big). \tag{2.3}$$

Recall that for a measure K (with density $k(\cdot)$), its total variation is defined as $|K| = \sup\sum_i K(E_i)$, where the supreme is taken over all partitions $\cup E_i$ of the support of K. We say that $K(\cdot|F, h_n, Y)$ is of *order* r at F and x in the direction $A(\cdot)$, if it is r-th Hadamard differentiable at F and x in the direction A, and for $\tilde{F}(\cdot)$ in a small neighborhood of $F(\cdot)$, i.e. $||\tilde{F} - F|| \le \epsilon$ for some small ϵ,

$$\lim_{h\to 0}\mathbb{E}K^{(m)}(X|\tilde{F},h,Y;A,\dots,A) = 0, \quad (m = 1,\dots,r-1) \quad \text{and}$$

$$0 \neq \lim_{h\to 0}\mathbb{E}K^{(r)}(X|F,h,Y;A,...,A) < \infty.$$

Define,

$$L_r(x|F,Y;A) = \lim_{h\to 0}K^{(r)}(x|F,h,Y;A,\dots,A), \quad \text{and}$$

$$L_r(x|F;A) = \mathbb{E}[L_r(x|F,Y;A)].$$

Assume the above limits exist. Denote by $\stackrel{D}{\to}$ for convergence in distribution. We list the following conditions:

(C1). $h = h_n \to 0$ as $n \to \infty$.
(C2). $\sum_{n=1}^{\infty}\exp(-\gamma n h^2) < \infty$ for every $\gamma > 0$.
(C3). $k(\cdot)$ is of bounded variation, $\int k(s)ds = 1$ and $|x|k(x) \to 0$ as $|x| \to \infty$.
(C4). $f(\cdot)$ is uniformly continuous on R^+.
(C5). $f(\cdot)$ and $g(\cdot)$ are bounded.
(C6). $A(\cdot)$ is bounded and has a continuous derivative $a(\cdot)$.
(C7). The sequence $\{n^{1/3}(\hat{F}_n(\cdot) - F(\cdot))\}$ is asymptotically tight.

Theorem 1. *(i). Assume (C1)-(C4), then we have*

$$\sup_{x\in[a,b]}|f_n(x|\hat{F}_n) - f(x)| \to 0. \quad a.s., \quad \textit{for all} \;\; 0 < a < b < \infty.$$

(ii). Assume (C1), (C5)-(C7), and that $K(\cdot|F,\cdot)$ is of order r at F and x in the direction $A(\cdot)$. Then as $nh \to \infty$,

(a). If $r = 1$ & $nh_n^5 \to 0$ or $r > 1$,

$$(h_n n)^{1/2}\big(f_n(x|\hat{F}_n) - f(x)\big) \stackrel{D}{\to} N(0,\sigma^2),$$

where, $\sigma^2 = f^2(x)\Big(\int_0^x \frac{1}{1-F(t)} + \int_x^\infty \frac{1}{F(t)}\Big)g(t)dt.$

(b). If $r = 1$ & $hn^{1/3} \to \infty$,

$$n^{1/3}\big(f_n(x|\hat{F}_n) - f(x) + b_n\big) \xrightarrow{D} L_1(x|F;A)Z,$$

where, $L_1(x|F;A) = a(x) + f(x)\Big(\int_0^x \frac{1}{1-F(t)} - \int_x^\infty \frac{1}{F(t)}\Big)A(t)g(t)dt,$
$a(x) = dA(x)/dx$ *and the bias* $b_n = h^2[f^{(2)}(x)/2]\int t^2 k(t)dt + o(h^2)$.

Remark 2. *(1) In Theorem 1 (ii) we omitted the case* $h = O(n^{-1/3})$. *In this case the weak limit is determined by the two limits displayed and appears more complex. Since we can choose h to avoid this case, we don't explore this problem here.*

(2) The weak limit of the proposed estimator does not depend on the kernel $k(\cdot)$. *In contrast, for the commonly used kernel method for such data, the asymptotic variance depends on* $k(\cdot)$, *often a term like* $\int k^2(u)du$ *appearing.*

(3) Conditions (C1)-(4) are commonly used for uniform consistency of kernel density estimator, for example, as given in Rao, B.L.S. (1983, Theorems 2.1.1-2.1.3, p.35-37). (C5) is satisfied by most of the commonly used density functions. (C6) is satisfied for most commonly used density functions with bounded derivative. Note that $\big(\int_0^x \frac{1}{1-F(t)} + \int_x^\infty \frac{1}{F(t)}\big)g(t)dt \le (1-F(x))^{-1}\int_0^x g(t)dt + (F(x))^{-1}\int_x^\infty g(t)dt = (1-F(x))^{-1}G(x) + (F(x))^{-1}(1-G(x)) < \infty$, *so* σ^2 *in Theorem 1 (ii) (a) is finite. Similarly,* $L_1(x|F;A)$ *in Theorem 1 (ii) (b) is finite. Condition (C7) is a technical assumption to ensure the functional delta method can be used. It is pointed out by Huang and Wellner [33] (p.9) that* $\{n^{1/3}(\hat{F}_n(\cdot) - F(\cdot))\}$ *is generally not tight. We make (C7) as an assumption and intuitively it can be true under sufficient smoothness conditions for* $F(\cdot)$, *although we are not clear about the exact conditions.*

It is known ([33, 34, 35]) that under some conditions, for some smooth functional $G(\cdot)$, $\sqrt{n}(G(\hat{F}_n) - G(F_0)) \xrightarrow{D} N(0,\tau^2)$ *for some* $\tau^2 < \infty$, *despite the convergence rate of* $\hat{F}_n$ *is only* $n^{1/3}$. *So, if (C7) is not satisfied, alternatively one may investigate conditions such that* $\sqrt{nh}\big(f_n(x|\hat{F}_n) - f(x)\big) \xrightarrow{D}$ *a Gaussian weak limit, and the result of (ii) in Theorems 1-6 will be modified accordingly. This can be a future research topic.*

(4) By Theorem 1 (ii), an approximate $(1-\alpha)$ *confidence interval for the density* $f(x)$ *is* $f_n(x|\hat{F}_n) \pm (nh_n)^{-1/2} z_{\alpha/2}\hat{\sigma}$, *where* z_α *denote the upper* α*-quantile of the standard normal distribution and*

$$\hat{\sigma}^2 = [f_n(x|\hat{F}_n)]^2\Big(\int_0^x \frac{1}{1-\hat{F}_n(t)} + \int_x^\infty \frac{1}{\hat{F}_n(t)}\Big)g(t)dt.$$

For the other models considered in this paper, there are also similar results for the confidence interval of the density and we will not emphasize them.

2.2. *Interval censoring type II*

For this model, $(X, U, V) \in R^+ \times (R^+)^2$, where $X \sim F$ and $(U, V) \sim G$ are independent, and $U < V$ a.s.. G has a density g with respect to the Lebesgue measure on $(R^+)^2$, G and g are assumed known. See [12] for background for this model. We observe $\{y_i = (u_i, v_i, 1_{[x_i \le u_i]}, 1_{[u_i < x_i \le v_i]}) : i = 1, \dots, n\}$ i.i.d. with $(U, V, 1_{[X \le U]}, 1_{[U < X \le V]}) := (U, V, \delta, \gamma)$, and $0 < P(\delta = 1), P(\gamma = 1) < 1$. The density-mass function of (U, V, δ, γ) is

$$p_F(u, v, \delta, \gamma) = F^\delta(u)(F(v) - F(u))^\gamma (1 - F(v))^{1-\delta-\gamma} g(u, v).$$

The algorithm for computing the NPMLE $\hat{F}_n$ for type I and II interval censoring model is given in [12] To study the asymptotic distribution of $\hat{F}_n(\cdot)$, they used the following working hypothesis (W1) and condition (W2).

(W1). Starting from the real underlying distribution function F, the iterative convex minorant algorithm will give at the first iteration step the estimator $\hat{F}_n^{(1)}(\cdot)$, which is asymptotically equivalent to the NPMLE $\hat{F}_n(\cdot)$.
(W2). $f(t) > 0$, $g(t, t) > 0$ and $g(t, \cdot)$ is left continuous at t.

Under (W1) and (W2), [12] obtained (Theorem 5.3, p.100)

$$(n \log n)^{1/3} (\hat{F}_n^{(1)}(t) - F(t)) \xrightarrow{D} \left(\frac{6 f^2(t)}{g(t, t)} \right)^{1/3} \arg\max_h \{\mathbb{B}(h) - h^2\} := A(t) Z,$$

where $\mathbb{B}(h)$ and Z are defined in section 2.1 for the interval censoring model type I.

When $(\delta, \gamma) = (0, 0)$ or equivalently $\{X > V\}$, $F(x|u, v, 0, 0) = 1_{[x > v]} F(x)/(1 - F(v))$; $(\delta, \gamma) = (0, 1)$ or $\{U < X \le V\}$, $F(x|u, v, 0, 1) = 1_{[u < x \le v]} F(x)/(F(v) - F(u))$ and $(\delta, \gamma) = (1, 0)$ or $\{X \le U\}$, $F(x|u, v, 1, 0) = 1_{[x \le u]} F(x)/F(u)$, the conditional density/mass of $X|Y$ is

$$f(x|u, v, \delta, \gamma) =$$

$$\left((1 - \delta)(1 - \gamma) \frac{1_{(v, \infty)}(x)}{1 - F(v)} + \gamma(1 - \delta) \frac{1_{(u, v]}(x)}{F(v) - F(u)} + \delta(1 - \gamma) \frac{1_{[0, u]}(x)}{F(u)} \right) f(x).$$

So the conditional kernel is

$$\begin{aligned} K(x|F, h, Y) &= \frac{1}{h} \mathbb{E}_F [k(\frac{X - x}{h}) | Y] \\ &= \frac{1}{h} \int k(\frac{s - x}{h}) \bigg((1 - \delta)(1 - \gamma) \frac{1_{(V, \infty)}(s)}{1 - F(V)} \\ &\quad + \gamma(1 - \delta) \frac{1_{(U, V]}(s)}{F(V) - F(U)} + \delta(1 - \gamma) \frac{1_{[0, U]}(s)}{F(U)} \bigg) F(ds). \end{aligned}$$

In the above equation, F is unknown. We use the NPMLE $\hat{F}_n$ as in [12] to replace F in $K(x|F,h,Y)$. Let $K(\cdot|\hat{F}_n,h,Y)$ be $K(\cdot|F,h,Y)$ with F replaced by $\hat{F}_n$, i.e.

$$K(x|\hat{F}_n,h,Y) = \frac{1}{h}\int k(\frac{s-x}{h})\Big((1-\delta)(1-\gamma)\frac{1_{(V,\infty)}(s)}{1-\hat{F}_n(V)} + \gamma(1-\delta)\frac{1_{(U,V]}(s)}{\hat{F}_n(V)-\hat{F}_n(U)} + \delta(1-\gamma)\frac{1_{[0,U]}(s)}{\hat{F}_n(U)}\Big)\hat{F}_n(ds). \tag{2.4}$$

and define the kernel estimator $f_n(x|\hat{F}_n)$ of $f(x)$ as

$$f_n(x|\hat{F}_n) = \frac{1}{n}\sum_{i=1}^{n} K(x|\hat{F}_n,h,Y_i). \tag{2.5}$$

Let $K^{(1)}(x|F,h,Y;A)$ be the 1-st Hadamard differential of $K(x|F,h,Y)$ at x and F with respect to F in the direction A,

$$\begin{aligned}
&K^{(1)}(x|Y,h,F;A)\\
&= \frac{1}{h}\int k(\frac{s-x}{h})\Big((1-\delta)(1-\gamma)\frac{1_{(V,\infty)}(s)[(1-F(s))A(ds)+A(s)F(ds)]}{(1-F(V))^2}\\
&\quad +\gamma(1-\delta)\frac{1_{(U,V]}(s)[(F(V)-F(U))A(ds)-(A(V)-A(U))F(ds)]}{(F(V)-F(U))^2}\\
&\quad +\delta(1-\gamma)\frac{1_{[0,U]}(s)[F(U)\alpha_1(ds)-\alpha_1(U)F(ds)]}{F^2(U)}\Big)\\
&= (1+o(1))\Big\{(1-\delta)(1-\gamma)\Big(\frac{1_{(V,\infty)}(x)[(1-F(x))a(x)dx+A(x)f(x)]}{(1-F(V))^2}\Big)\\
&\quad +\gamma(1-\delta)\frac{1_{(U,V]}(x)[(F(V)-F(U))a(x)-(A(V)-A(U))f(x)]}{(F(V)-F(U))^2}\\
&\quad +\delta(1-\gamma)\frac{1_{[0,U]}(x)[F(U)a(x)-A(U)f(x)]}{F^2(U)}\Big\} := (1+o(1))L_1(x|Y,F),
\end{aligned} \tag{2.6}$$

where

$$L_1(x|F,Y;A) = \lim_{h\to 0} K^{(1)}(\cdot|F,h,Y;A)$$

and

$$L_1(x|F;A) = E[L_1(x|F,Y;A)].$$

Consider the following conditions

(C8). $\Big(\int_0^x\int_0^v \frac{1}{1-F(v)} + \int_x^\infty\int_0^x \frac{1}{F(v)-F(u)} + \int_x^\infty\int_x^v \frac{1}{F(u)}\Big)g(u,v)dudv < \infty.$

Note that $\int_0^x\int_0^v[g(u,v)/(1-F(v))]dudv \le (1-F(x))^{-1}\int_0^x\int_0^v g(u,v)dudv < \infty$, and $\int_x^\infty\int_x^v[g(u,v)/F(u)]dudv \le (F(x))^{-1}\int_x^\infty\int_x^v g(u,v)dudv < \infty$, (C8) is satisfied if $\int_x^\infty\int_0^x[g(u,v)/(F(v)-F(u))]dudv < \infty$, which is automatic if $U \le V + c$ for some constant $c > 0$ and $F(\cdot)$ is strictly increasing. These conditions are very reasonable for this model.

Theorem 2. *(i). Assume (C1)-(C4), then we have*

$$\sup_{x\in[a,b]} |f_n(x|\hat{F}_n) - f(x)| \to 0 \quad a.s. \ \ for\ all\ \ 0 < a < b < \infty.$$

(ii). Assume (C1), (C6), (C7)-(C8) ((C7) with $\hat{F}_n$ and F for the model here), (W1)-(W2), and that $K(\cdot|F,Y)$ is of order r at F and x in the direction $A(\cdot)$. Then as $nh \to \infty$,

(a). If $r = 1$ & $h_n(n\log n)^{1/3} \to 0$ or $r > 1$,

$$(h_n n)^{1/2}\big(f_n(x|\hat{F}_n) - f(x)\big) \xrightarrow{D} N(0,\sigma^2),$$

where,

$$\sigma^2 = f^2(x)\Big(\int_0^x\int_0^v \frac{g(u,v)}{1-F(v)}dudv + \int_x^\infty\int_0^x \frac{g(u,v)}{F(v)-F(u)}dudv + \int_x^\infty\int_x^v \frac{g(u,v)}{F(u)}dudv\Big).$$

(b). If $r = 1$ & $h_n(n\log n)^{1/3} \to \infty$,

$$(n\log n)^{1/3}\big(f_n(x|\hat{F}_n) - f(x) + b_n\big) \xrightarrow{D} L_1(x|F;A)Z,$$

where,

$$L_1(x|F;A) = a(x) + f(x)\Big(\int_0^x\int_0^v \frac{A(v)}{1-F(v)} - \int_x^\infty\int_0^x \frac{A(v)-A(u)}{F(v)-F(u)} - \int_x^\infty\int_x^v \frac{A(u)}{F(u)}\Big)g(u,v)dudv,$$

$a(x) = dA(x)/dx$ and b_n is given in Theorem 1 (b).

2.3. Convolution model

For this model, $(X,W) \sim F\times G$. F is unknown and G, with density g, is known (otherwise the model is not identifiable). It is also a type of measurement error model. We observe $Y_1,...,Y_n$ i.i.d. $Y = X + W$. The density of Y is

$$q(y) = \int g(y-w)F(dw).$$

Since $P(X < x, X+W < y) = \int_{-\infty}^x\int_{-\infty}^{y-s} f(s)g(w)dsdw = \int_{-\infty}^x f(s)G(y-s)ds$, the joint density of (X,Y) is $p(x,y) = f(x)g(y-x)$, the conditional density of $X|Y$ is $p(x|y) = p(x,y)/p(y) = f(x)g(y-x)/\int g(y-w)F(dw)$, and the conditional kernel is

$$K(x|F,h,Y) = \frac{\int k(\frac{s-x}{h})g(Y-s)F(ds)}{h\int g(Y-w)F(dw)}.$$

In the above equation, F is unknown. We can use an existing estimate, for example the one-step NPMLE $\hat{F}_n$ of F based on the observed data $Y_1,...,Y_n$ as in [12] in which the following condition will be used.

(W3). g be a right-continuous decreasing density on $[0,\infty)$, having only a finite number of discontinuity points at $a_0 = 0 < a_1 < \cdots < a_k$; and g has a derivative $g'(x)$ at $x \neq a_i, i = 0,...,k$, satisfying $\int_0^\infty g'(x)^2/g(x)dx < \infty$, where the integrand is defined to be 0 at a_i and at points where $g = 0$; g' is bounded and continuous on (a_{i-1}, a_i) for $i = 1,...,m+1$, with $a_{m+1} := \infty$.

Let Z as in Theorem 1, $q(\cdot)$ be the density of Y,

$$A(t) = \left(4f(t)\sum_{j=0}^{k} q(t+a_j)[g(a_j)-g(a_j-)]^{-1}\right)^{1/3}.$$

By Theorem 5.4 of [12], under (W3) we have

$$n^{1/3}(\hat{F}_n(t) - F(t)) \stackrel{D}{\to} A(t)Z.$$

Define the estimator of $f(x)$ based on the observed data Y_i's by

$$f_n(x|\hat{F}_n) = \frac{1}{n}\sum_{i=1}^{n} K(x|\hat{F}_n, h, Y_i). \tag{2.7}$$

The first order of Hadamard differential of $K(x|F,h,Y,\alpha)$ at F and x in the direction α is

$$\begin{aligned} K^{(1)}(x|F,h,Y;\alpha) = \frac{1}{h}[\int g(Y-w)F(dw)]^{-2} \\ \times\Big[\int k(\frac{s-x}{h})g(Y-s)\alpha(ds)\int g(Y-w)F(dw) \\ -\int k(\frac{s-x}{h})g(Y-s)F(ds)\int g(Y-w)\alpha(dw)\Big]. \end{aligned} \tag{2.8}$$

The following conditions will be used.

(C9). $\sup_x \int \frac{|g'(y-x)\int g(y-w)f(w)dw - g(y-x)\int g'(y-x)f(w)dw|}{[\int g(y-w)f(w)dw]^2} dy < \infty$.

(C10). $\int \left(\frac{g^2(y-x)-[\int g(y-w)f(w)dw]^2}{\int g(y-w)f(w)dw}\right) dy < \infty$.

Note that $\int g(y-w)f(w)dw = q(y)$, if $q(\cdot)$ is bounded below, (C10) will be true; if in addition $g'(\cdot)$ is bounded (C9) will be true.

Theorem 3. *Assume (C1)-(C4) and (C9), then we have*

$$\sup_{x\in[a,b]} |f_n(x|\hat{F}_n) - f(x)| \to 0, \quad a.s. \quad \text{for all} \ \ 0 < a < b < \infty.$$

ii) Assume (C1), (C5)-(C7) ((C7) with $\hat{F}_n$ and F for the model here), (C10), and that $K(x|F,h,Y,A)$ is of order r at F and x in the direction $A(\cdot)$. Then as $nh \to \infty$, (a). If $r = 1$ & $h_n n^{1/3} \to 0$ or $r > 1$,

$$(h_n n)^{1/2}\big(f_n(x|\hat{F}_n) - f(x)\big) \xrightarrow{D} N(0,\sigma^2),$$

where, $\sigma^2 = f^2(x)\int \Big(\frac{g^2(y-x)}{\int g(y-w)F(dw)}\Big)dy$ and $a(x) = dA(x)/dx$.
(b). If $r = 1$ & $h_n n^{1/3} \to \infty$,

$$n^{1/3}\big(f_n(x|\hat{F}_n) - f(x) + b_n\big) \xrightarrow{D} L_1(x|F;A),$$

where, $L_1(x|F;A) = a(x) - f(x)\int \frac{g(y-x)\int g(y-w)A(dw)}{\int g(y-w)F(dw)}dy$ and b_n is given in Theorem 1 (b).

2.4. Double censoring

For this model, we only briefly present the results. In this model, the original data is $(X,U,V) \in (R^+)^3$, $U < V$ a.s., X and (U,V) are independent, $X \sim F(\cdot)$, $(U,V) \sim G$, assumed known. We observe $Y = ((X \vee U) \wedge V, 1_{[X \le U]}, 1_{[U<X\le V]}) = (Z,\delta,\gamma)$. Here (δ,γ) can only take values $(0,0)$, $(0,1)$ and $(1,0)$. The density-mass function $p_{F,G}$ for y is

$$p_F(y) = p_F(z,\delta,\gamma) = [M(z)f(z)]^\gamma[F(z)g_U(z)]^\delta[(1-F(z))g_V(z)]^{1-\gamma-\delta},$$

where $M(z) = P(U < z \le V) = G_U(z) - G(z,z)$, G_U is the marginal distribution of U, and g_U and g_V are the marginal densities of U and V, respectively. Let $G_{U|V}(\cdot|v)$ be the conditional distribution of U given $V = v$, and let $G_{V|U}(\cdot|u)$ be similarly defined. This model was studied by Turnbull [8], Tsai and Crowley [32], among others. When $(\delta,\gamma) = (0,0),(0,1)$ and $(1,0)$, we observe V, X and U, respectively.

For $(\delta_i,\gamma_i) = (0,1)$, we observe the original data X_i, and one possibility is to define the density estimator using only the observed original data as

$$f_{1n}(x) = (n_1h_1)^{-1}\sum_{i\in D_1} k(\frac{X_i - x}{h_1}),$$

where $n_1 = \sum_{i=1}^n (1-\delta_i)\gamma_i$ and D_1 is the set of data corresponding to the subset for which $(\delta_i,\gamma_i) = (0,1)$. Thus by standard kernel density estimator theory, under suitable conditions we have $\sup_x |f_{1,n}(x) - f(x)| \to 0$ (a.s.) and

$$\sqrt{n_1h_1}(f_{1n}(x) - \mathbb{E}f_n(x)) \xrightarrow{D} N(0,\sigma_1^2),$$

with $\sigma_1^2 = f(x)\int_{-\infty}^{\infty} k^2(w)dw$.

However, since $f_{1,n}(x)$ uses only the original data, the data that is not directly observed (with $(\delta_i, \gamma_i) \neq (0,1)$) is ignored. We want to use all the data to construct the estimator. The conditional kernel is

$$K(x|F,h,Y) = \frac{1}{h}\mathbb{E}(k(\frac{X-x}{h})|Y)$$

$$= \frac{1}{h}\bigg((1-\delta-\gamma)\int_0^\infty \frac{1_{(s>Z)}}{1-F(Z)}k(\frac{s-x}{h})F(ds) + \delta\int_0^\infty \frac{1_{(s\le Z)}}{F(Z)}k(\frac{s-x}{h})F(ds)\bigg).$$

In the above expression, $F(\cdot)$ is unknown and it needs to be estimated. Let $Y_i, i=1,\dots,n_1$ be the data that X is observed and $Y_i, i=n_1+1,\dots,n$ be the data that X is unobserved. Denote $n_2 = n - n_1$.

Tsai [36] studied the NPMLE $\hat{S}_n(x)$ of the survival function $S(x)$ for this model, based on the observed data $Y_1,\dots,Y_n$. Since $F(\cdot) = 1 - S(\cdot)$, $\hat{F}_n(\cdot) = 1 - \hat{S}_n(\cdot)$ is the NPMLE of $F(\cdot)$. Define the kernel estimator $f_n(x|\hat{F}_n)$ of $f(x)$ as

$$\begin{aligned} f_n(x|\hat{F}_n) &= \frac{n_1}{n}\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{h_1}k(\frac{X_i-x}{h_1}) + \frac{n_2}{n}\frac{1}{n_2}\sum_{i=n_1+1}^{n}K_2(x|\hat{F}_n,h_2,Y_i) \\ &:= \frac{n_1}{n}f_{1n}(x) + \frac{n_2}{n}f_{2n}(x|\hat{F}_n). \end{aligned} \tag{2.9}$$

Dubnicka [17, 18] studied kernel density estimator with missing data, which is a weighted kernel method, with the inverse estimated propensity scores as the weights. Our estimator (2.9) can also be viewed as a weighted kernel estimator, with weights in the second term being the conditional kernel evaluated at the observations.

The kernel density estimator is biased. If we use $f_{1n}(x)$ to estimate $f(x)$, the bias is $O(h_{n_1})$. Similarly, if only $f_{2n}(x|\hat{F}_n)$ is used, the bias is $O(h_{n_2})$. If both are used, the bias is $O(h_n)$, which is smaller than using f_{1n} or $f_{2n}(x|\hat{F}_n)$ alone. We will see that under suitable conditions $f_n(x|\hat{F}_n)$ is $\sqrt{n}$-consistent, and a $(1-\alpha)\%$ confidence interval for $f(x)$ can be obtained as below. Since

$$\begin{aligned} & n^{1/2}\big(f_n(x|\hat{F}_n) - \mathbb{E}f_n(x|F)\big) \\ = \;& h_{n_1}^{-1/2}\sqrt{n_1}n(n_1h_{n_1})^{1/2}(f_{1n}(x) - \mathbb{E}f_n(x)) \\ + \;& \sqrt{n_2/n}\sqrt{n_2}(f_{2n}(x|\hat{F}_{n_2}) - \mathbb{E}f_n(x|F)) \\ \approx \;& h_{n_1}^{-1/2}\alpha_1^{3/2}N(0,\sigma_1^2) + \alpha_2^{3/2}N(0,\sigma_2^2) \approx N(0, h_{n_1}^{-1}\alpha_1^3\sigma_1^2 + \alpha_2^3\sigma_2^2). \end{aligned}$$

So a $(1-\alpha)\%$ confidence interval for $f(x)$ is

$$[f_n(x|\hat{F}_n) + O(h) \pm z_{1-\alpha/2}\sqrt{\frac{h_{n_1}^{-1}\alpha_1^3\sigma_1^2 + \alpha_2^3\sigma_2^2}{n}}].$$

Let $\Lambda(r,t) = \mathbb{E}[S(r)S(t)]$ be the covariance function of $S(\cdot)$ and $\lambda(r,t) = \partial^2\Lambda(r,t)/(\partial r\partial t)$. Assume it exists. Denote $g_u(\cdot)$ and $g_v(\cdot)$ the U and V margins of $g(\cdot,\cdot)$ respectively. In the following Theorems 4 and 6, conditions (C11)-(C18) are given in the Appendix.

Theorem 4. *(i) Assume (C1) - (C4), then*

$$\sup_{x\in[a,b]} |f_n(x|\hat{F}_n) - f(x)| \to 0, \quad a.s. \quad \text{for all} \quad 0 < a < b < \infty.$$

(ii), Assume (C1), (C12)-(C14), (W4) and that the operator $H_4^{*'}$ *from Tsai and Crowley (1985) is invertible, and that* $K(\cdot|F,\cdot)$ *is Hadamard differentiable at* F *and* x *in the direction* $S(\cdot)$*. Then, with* $\alpha_1 = 1-\alpha_2 = P(U < X \le V) = \int_0^\infty (G_U(x) - G(x,x))F(dx)$*, as* $n_k h_k \to \infty$ $(k = 1,2)$*,*

$$\frac{n_1}{n}(n_1h_1)^{1/2}\big(f_{1n}(x) - f(x) + b_{n_1}\big) + \frac{n_2}{n}(h_2n_2)^{1/2}\big(f_{2n}(x|\hat{F}_n) - f(x) + b_{n_2}\big)$$

$$\xrightarrow{D} \alpha_1 N(0,\sigma_1^2) + \alpha_2 N(0,\sigma_2^2) + \sqrt{\alpha_2}N(0,\sigma_3^2),$$

where, b_n *is given in Theorem 1,* $\sigma_1^2 = f(x)\int k^2(s)ds$*,* $\sigma_2^2 = f^2(x)\big(\int_0^x \frac{g_v(z)}{1-F(z)}dz + \int_x^\infty \frac{g_u(z)}{F(z)}dz - 1\big)$*, and with* $A(x) = G_U(x) + G_V(x) - 1$*,*

$$\sigma_3^2 = f^2(x)\Big(\int_0^x \frac{g_v(s)}{1-F(s)} - \int_x^\infty \frac{g_u(s)}{F(s)}\Big)\Big(\int_0^x \frac{g_v(t)}{1-F(t)} - \int_x^\infty \frac{g_u(t)}{F(t)}\Big)\Lambda(s,t)dsdt$$

$$+2f(x)A(x)\Big(\int_0^x \frac{g_v(s)}{1-F(s)} - \int_x^\infty \frac{g_u(s)}{F(s)}\Big)\Lambda(ds,x) + A^2(x)\lambda(x,x).$$

2.5. Multiplicative censoring

For this model, we only briefly present the results. Let $0 < p < 1$ be known, and $\delta_{\{1\}}$ be the measure with unit mass at 1. Suppose $(X,W) \sim F\times G$, F on R^+ is unknown with density f, and $G = p\delta_{\{1\}} + (1-p)U(0,1)$. Let $T = XW$, $\delta = 1_{(W=1)}$. Thus $T = X$ with probability p and $T = XW1_{(W<1)}$ with probability $1-p$. For this model we observe $Y_i = (T_i,\delta_i)$ $(i = 1,\dots,n)$ iid $Y = (T,\delta)$. The mass-density function of Y is

$$p_F(t,\delta) = (pf(t))^\delta\left((1-p)\int_t^\infty \frac{1}{s}dF(s)\right)^{1-\delta}.$$

With $\delta = 0$, the density for T is $g(t) = \int_0^1 (1/w)f(t/w)dw = \int_t^\infty (1/s)F(ds)$, the joint distribution function for (X,T) is $P(x,t) = \int_0^x (t/s)f(s)ds$, the joint density is $p(x,t) = (f(x)/x)1_{[0,x]}(t)$, and the conditional density for $X|(T,\delta=0)$ is $(f(x)/x)1_{[0,x]}(t)/\int_t^\infty (1/s)F(ds)$. For this model, the conditional kernel is

$$K(x|F,h,Y) = \frac{1}{h}\left(\delta k(\frac{T-x}{h}) + (1-\delta)\frac{\int_T^\infty k(\frac{s-x}{h})\frac{1}{s}F(ds)}{\int_T^\infty \frac{1}{s}F(ds)}\right)$$

$$:= \frac{\delta}{h}k(\frac{T-x}{h}) + K_2(x|F,h,Y).$$

Let $n_1 = \sum_{i=1}^n \delta_i$, $n_2 = n - n_1$. Define the kernel estimator $f_n(x|\hat{F}_n)$ of $f(x)$ as

$$f_n(x|\hat{F}_n) = \frac{n_1}{n}\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{1}{h}k(\frac{X_i - x}{h}) + \frac{n_2}{n}\frac{1}{n_2}\sum_{i=n_1+1}^{n} K_2(x|\hat{F}_{n_2}, h, Y_i)$$
$$:= \frac{n_1}{n} f_{1n}(x) + \frac{n_2}{n} f_{2n}(x|\hat{F}_{n_2}). \tag{2.10}$$

Let $\Lambda(r,t) = \mathbb{E}[S(r)S(t)]$ be the covariance function of $S(\cdot)$ and $\lambda(r,t) = \partial^2\Lambda(r,t)/(\partial r\partial t)$.

The iterative convex minorant algorithm in [37] can be used to compute the NPMLE $\hat{F}_n$ of F based on the observed data $Y_1, \ldots, Y_n$ for this model.

Theorem 5. *i). Assume (C1)-(C4) and (C14), then we have*

$$\sup_{x\in[a,b]} |f_n(x|\hat{F}_n) - f(x)| \to 0, \quad a.s. \quad \text{for all} \quad 0 < a < b < \infty.$$

ii) Assume (C1), (C13) and (C15), then as $n_k h_k \to \infty$ $(k = 1, 2)$,

$$\frac{n_1}{n}(n_1h_1)^{1/2}\big(f_{n_1}(x|\hat{F}_{n_1}) - f(x) + b_{n_1}\big) + \frac{n_2}{n}(h_2n_2)^{1/2}\big(f_{n_2}(x|\hat{F}_{n_2}) - f(x) + b_{n_2}\big)$$
$$\xrightarrow{D} pN(0,\sigma_1^2) + (1-p)N(0,\sigma_2^2) + (1-p)^{1/2}N(0,\sigma_3^2),$$

where b_n *is given in Theorem 1,* $\sigma_1^2 = f(x)\int k^2(s)ds$, $\sigma_2^2 = f^2(x)\big(x^{-2} \times \int_0^x g^{-1}(t)dt - 1\big)$, *and*

$$\sigma_3^2 = \frac{\lambda(x,x)}{x^2}\int_0^x g^{-1}(t)dt - 2\frac{f(x)}{x^2}\int_0^x g^{-2}(t)\int_t^\infty \frac{1}{v}\lambda(x,v)dvdt$$
$$+\frac{f^2(x)}{x^2}\int_0^x g^{-3}(t)\int_t^\infty\int_t^\infty \frac{1}{uv}\lambda(u,v)dudvdt.$$

For the multiplicative censoring model, Asgharian et.al [24] investigated kernel smoothed estimator of the density function. Put $m = n = k/2$ in their case, as they mentioned (p.170), using their Theorem 4, they may show that their estimator is asymptotically Gaussian with rate $(h_n n)^{1/2}$, and with mean zero and covariance function σ_g (given in line -8, p.170), which again depends on the subjectively chosen kernel $K(\cdot)$. If take $n_1 = n_2 = n/2$, our estimator, in the mixture format, is asymptotically a mixture normal, with rate $(h_n n)^{1/2}$ which is the same as in [24]. Our asymptotic variance partially depend on the kernel $k(\cdot)$ only in the first component in the equation (2.10).

2.6. Summary

After studying the conditional kernel density estimation for the above five incomplete data models, we summarize the results for general incomplete data models as in Theorem 6, without proof.

Suppose the conditional kernel $K(x|H,h,Y)$ depends on an unknown quantity $H(\cdot)$, defined on T, which can be consistently estimated by $\hat{H}_n$ based on the observed data $Y_1,\dots,Y_n$, define the estimator $f_n(x|\hat{H}_n)$ for the density $f(x)$ of the original data X as

$$f_n(x|\hat{H}_n) = \frac{1}{n}\sum_{i=1}^{n} K(x|\hat{H}_n,h,Y_i). \tag{2.11}$$

Recall $l^\infty(T)$ is the space of all bounded functions on T equipped with the supremum norm. Consider the following conditions.

(C16). $\sup_{t\in T}|\hat{H}_n(t)-H(t)|\to 0$, a.s..
(C17). $\sup_{x,y}|K(x|\hat{H}_n,h,y)-K(x|H,h,y)|\to 0$, a.s..
(C18). $\sup_x|f_n(x)-f(x)|\to 0$, a.s..
(C19). For some tight stochastic process $S(\cdot)$ on T, and $b_n\to\infty$, $b_n(\hat{H}_n(\cdot)-H(\cdot))\overset{D}{\to}$
$S(\cdot)$ on $l^\infty(T)$.
(C20). $(nh)^{1/2}(f_n(x)-\mathbb{E}f_n(x))$ is asymptotic normal.
(C21). $K(x|H,h,Y)$ is of order r ($r\geq 1$) at x and H in the direction S; $L_r(x|F,Y;S,...,S)$
$=\lim_{h\to 0}K^{(r)}(H,h,Y;S,...,S)$ exists, and $L_r(x|F;S,...,S)=$
$E_Y[L_r(x|F,Y;S,...,S)]$ exists.

Theorem 6. *i) Assume (C16) - (C18), then*

$$\sup_x|f_n(x|\hat{H}_n)-f(x)|\to 0,\ \ a.s..$$

ii) Assume (C19)-(C21). Then, with $0<\sigma^2=\lim_{h\to 0}Var[K(x|H,h,Y)]<\infty$,

$$a_n\big(f_n(x|\hat{H}_n)-f(x)+b_n\big)\overset{D}{\to}\begin{cases}\frac{1}{r!}L_r(x|H;S,...,S)],\\ \frac{1}{r!}L_r(x|H;S,...,S)]+N(0,\sigma^2),\\ N(0,\sigma^2),\\ N(0,\sigma_1^2),\end{cases}$$

$$where,\qquad \begin{cases}a_n=b_n^r, & if\ b_n^r/\sqrt{hn}\to 0;\\ a_n=\sqrt{hn}, & if\ b_n^r/\sqrt{hn}\to C,\ 0<C<\infty;\\ a_n=\sqrt{hn}, & if\ b_n^r/\sqrt{hn}\to\infty;\\ a_n=\sqrt{hn}, & if\ 0<\sigma_1^2=\lim_{h\to 0}hVar[K(x|H,h,Y)]<\infty.\end{cases}$$

3. Numerical studies

In this section, we conduct the simulation studies to evaluate the performance of the proposed conditional kernel density estimators, and compare them with some existing estimators for the corresponding models. The limiting distributions of conditional kernel density estimators are functional of Chernoff distribution for type I and II interval censoring models, and functional of Gaussian process for other models. Instead of studying these weak limits, we investigate the finite sample performances of the estimators. The purpose of the simulations is to provide a straight perception on the conditional estimator. We draw the plots of the conditional kernel density estimators against the true density. We compare our method with some existing methods in terms of the integrated square errors (ISE).

For Type I, II interval censoring models, we assume that the variable of interest X and the observed variables U and V all from Gamma distributions $Gamma(k, \theta)$ with density

$$g(x; k, \theta) = [\theta^k \Gamma(k)]^{-1} x^{k-1} e^{-x/\theta} 1_{(x>0)}.$$

In our simulations, we chose $k = 5$ as the shape parameter and $\theta = 1$ as the scale parameter for the random variables X and U and let $V = U + Gamma(5, 1)$. For type I interval censoring model, we compare our esitmator with that of [39] (GJW); for type II interval censoring, we compare with the kernel smoothing estimator.

For the convolution model, we assume X from the normal distribution

$$g(x; \mu, \sigma) = e^{-(x-\mu)^2/(2\sigma^2)}/\sqrt{2\pi}\sigma,$$

We chose $\mu = 0$ and $\sigma = 1$ (i.e., the standard normal distribution) for X and W in this model. For this model, we compare our estimator with the kernel smoothing estimator.

For the double censoring model, we chose the Gamma distributions for U, X with the respective parameters $k = 2, \theta = 1$ and $k = 5, \theta = 1$ and let $V = U + Gamma(1, 1)$. For the multiplicative censoring model, we chose the Gamma distribution $Gamma(5, 1)$ for the latent distribution for X and X was censored with probability $1 - p = 1/2$. For this model, we compare our estimator with that of [19].

For the multiplicative censoring model, our method is compared with that of [24] (ACF).

For all the models we used the standard normal density as the kernel $k(\cdot)$. Various sample sizes and bandwidths were considered as given in Table 1. We presented 50 sample paths for the conditional kernel density estimators. The plots for the Type I, II interval censoring models and convolution model are drawn in Figures 1, 2 and 3, respectively, and those for the the doubly censoring model and multiplicative censoring model are putted in the supplementary material.

The first, second and third rows were generated from datasets of 200, 500, and 1000 total observations, respectively. Plots of the first, second and third columns were based on the different bandwidths $h = n^{-1/5}$, $n^{-1/10}$, and $n^{-1/20}$, respectively. In Figures 1, 2 and 3, we also plot the average 95% confidence interval given in Remark 2 (4). The integrated square errors (ISEs) are reported in Table 1.

Figures 1 and 2 have similar behavior. When $h = n^{-1/5}$, the points of the $f_n(x|\widehat{F}_n)$ deviate the true values very much with a larger range in the case $n =$

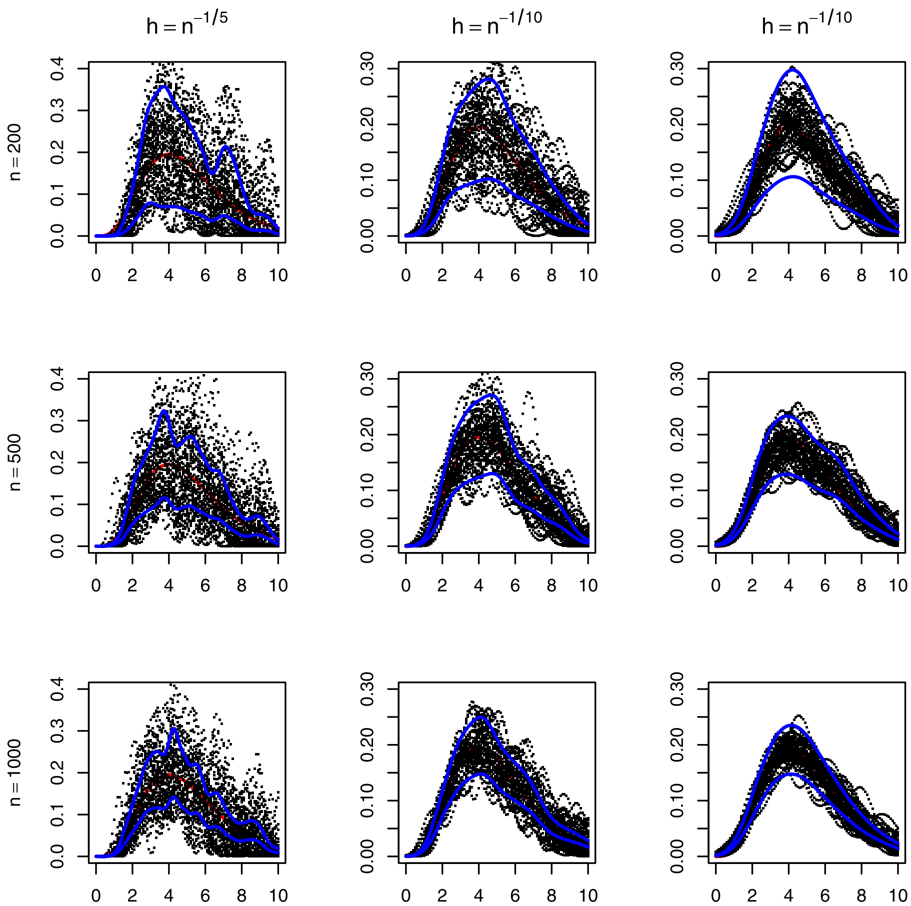

FIG 1. *Sample paths for Type I interval censoring model. The red line is the true density and the blue line is the average 95% confidence interval.*

200. This is different from the traditional kernel density estimator for uncensored samples, in which the optimal bandwidth is in the magnitude of $n^{-1/5}$. But when $h = n^{-1/10}$ and $n^{-1/20}$, $f_n(x|\widehat{F}_n)$ fit $f(x)$ well, with the pointwise average of the sample plots very closing to the true plots (in red color). On the other hand, as the sample size increases, the deviation range of the $f_n(x|\widehat{F}_n)$ becomes smaller as expected. These phenomena can also be observed from the integrated square errors in Table 1. The average length of confidence interval decreases as n increases. We also observe that for most models, except the convolution model, the ISEs under $h = h^{-1/20}$ are smaller than those under $h = n^{-1/10}$ for all the methods.

In Figure 3 for the convolution model, X and W both are assumed the normal distribution $N(0, 1)$. We used the R-package "Decon" developed by

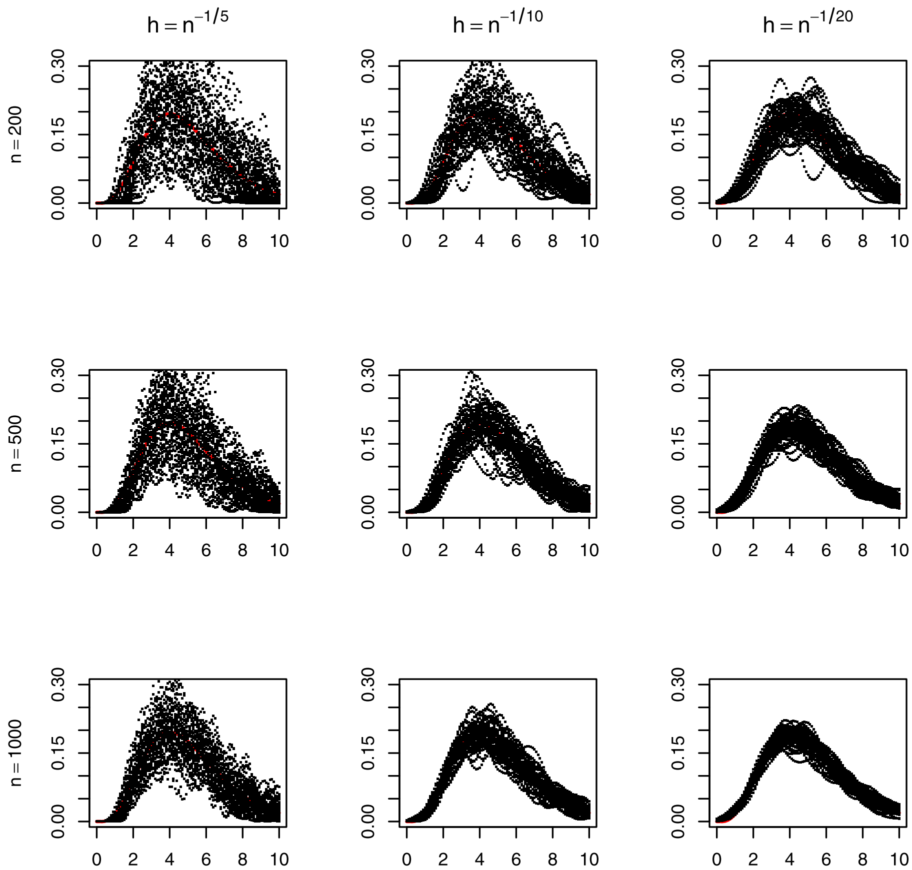

FIG 2. *Sample paths for Type II interval censoring model. The red line is the true density and the blue line is the average 95% confidence interval.*

[38] to compute the NPMLE $\hat{F}_n$ of F. When $h = n^{-1/10}, n^{-1/20}$, the plots of the $f_n(x|\widehat{F}_n)$ are evidently lower than those of the true density at around the unimodal point, although the plots depict the shape of the normal distribution. In the case $h = n^{-1/5}$, the plots of the $f_n(x|\widehat{F}_n)$ look much better than those of $h = n^{-1/10}, n^{-1/20}$, although there is some deviation at the mode. For all the methods, the ISEs under $h = n^{-1/5}$ are smaller than those under $h = n^{-1/10}, n^{-1/20}$.

For the doubly censoring model and the multiplicative censoring model, we can see that for all the methods, the bandwidths $h = n^{-1/10}$ and $h = n^{-1/20}$ are better than $h = n^{-1/5}$ from the figures of the supplementary material and Table 1.

From Table 1, we can see that the ISEs of our method are smaller than or comparable to those of the other methods under the Type I/II interval censoring model, convolution model and double censoring model. For example, when $n =$

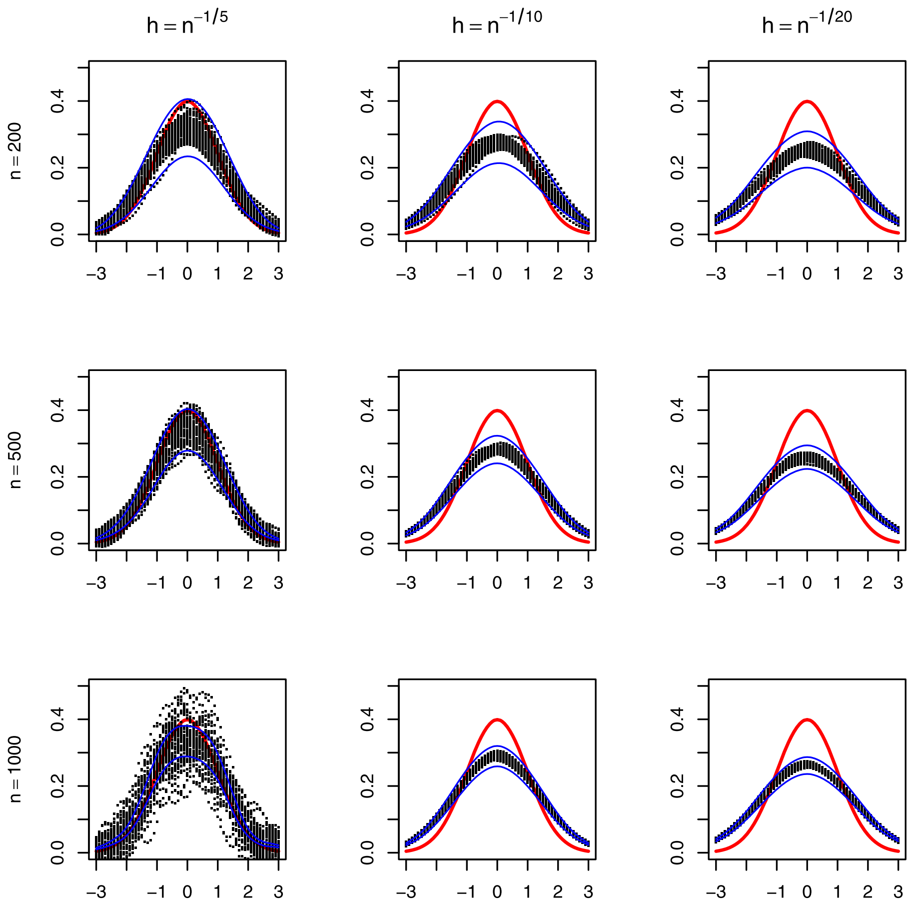

Fig 3. *Sample paths for the convolution model. The red line is the true density and the blue line is the average 95% confidence interval.*

200 and $h = n^{-1/5}$ under the convolution model, the ISE of our method is 1.21, which is smaller than the ISE of the kernel smoothing estimator (= 2.10). However, for the multiplicative censoring model, the ACF's method gives better performance than ours. Part of the reason may due to the fact that the ISE of the ACF estimator depends on the variance square of the chosen kernel (Theorem 5 in ACF). Our estimator in this case is a mixture of kernel smoothing estimator and a conditional kernel estimator, the ISE of the latter part does not depend on the kernel. The ISE of our estimator in this case partially depends on the variance square of the chosen kernel, so the ISE's of the two estimator varies with the choosen kernel. With samll variance of the kernel, the ACF method is expected to have smaller ISE than ours, and vice versa.

TABLE 1
Integrated squared errors (multiplied by 10^{-2}).

Type I interval censoring model			
n	$h = n^{-1/5}$	$h = n^{-1/10}$	$h = n^{-1/20}$
200	5.38	1.63	0.87
500	3.83	0.01088422	0.006131349
1000	0.03489001	0.006388937	0.003422104

Type II interval censoring model			
n	$h = n^{-1/5}$	$h = n^{-1/10}$	$h = n^{-1/20}$
200	0.039184744	0.017096865	0.006979607
500	0.023368021	0.006844263	0.003465765
1000	0.012759035	0.003581359	0.001820693

Convolution model			
n	$h = n^{-1/5}$	$h = n^{-1/10}$	$h = n^{-1/20}$
200	0.01209222	0.02438077	0.0309852
500	0.03192705	0.03992805	0.06300134
1000	0.007925386	0.0233965	0.03142238

Double censoring			
n	$h = n^{-1/5}$	$h = n^{-1/10}$	$h = n^{-1/20}$
200	0.046281770	0.021137848	0.028588153
500	0.02858815	0.01140205	0.00846038
1000	0.01515069	0.008508595	0.005758564

multiplicative censoring			
n	$h = n^{-1/5}$	$h = n^{-1/10}$	$h = n^{-1/20}$
200	0.007722237	0.005886002	0.006067418
500	0.004929110	0.004117460	0.004323686
1000	0.003320455	0.003735651	0.003769296

4. Discussion

In this paper, we have proposed the conditional kernel density estimation for a class of incomplete data models and derived the uniform consistency and limiting distributions for the proposed estimators. Simulation studies show that the finite sample performances of the conditional kernel estimators crucially depend on the choices of bandwidth. In our simulations, when the bandwidths are chosen between $n^{-1/10}$ and $n^{-1/20}$, the plots of the conditional kernel estimator fit the true density well for Type I, II censoring models. But for the convolution model, the appropriate bandwidth is approximate $n^{-1/5}$.

The basic properties of the proposed estimators are established. Condition (C7) is not easy to check. An alternative is to consider different conditions as discussed in last paragraph of Remark 2 (3), and the results will be modified and will be our future work.

For interval censoring data type I, [39] studied two versions of kernel smoothed estimator of the density function $f(\cdot)$. Their density estimator is the derivative of $\hat{F}_n^{MS}(\cdot)$, which is the maximizer of their (3.1), and not given in closed form. As in their Theorem 3.6 and Theorem 4.3, the convergence rate of their density estimators is $n^{2/7}$, and the asymptotic variances depend on $\int k'(u)^2 du$, where $k(\cdot)$ is the subjectively chosen kernel, and so the asymptotic variances of their estimators can have a range of values depending on the choices of $k(\cdot)$, a phenomenon shared by all the kernel smooth methods. For our estimator, it is given in closed form, the convergence rate is either $n^{1/3}$ or $(h_n n)^{1/2} = n^{2/5}$, if we choose the commonly used rate for bandwidth $h_n = O(n^{-1/5})$. In either case, our convergence rate is faster than $n^{2/7}$, and the asymptotic variance does not depend on the kernel $k(\cdot)$, a subjective input.

Ren [19] studied kernel density estimator with doubly censored data, by smoothing an estimated distribution function. It is asymptotic normal (Theorem 2) with rate $(nh)^{1/2}$ and asymptotic variance $(f(x)/[S_Y(x) - S_Z(x)]) \int k^2(u)du$, which depends on the subjective choice $k(\cdot)$. In contrast, the asymptotic behavior of our estimator for this type of data is given in Theorem 4 here, it is asymptotic normal with asymptotic variance not depending on the subjective input $k(\cdot)$, though with much more complicated form than that in [19].

Asgharian et.al [24] studied kernel density estimator for multiplicative censoring data. Their observed data is $(X_1, \ldots, X_m; Y_1, \ldots, Y_n)$, $Y_i = Z_i U_i \sim F$, $X_j, Z_i \sim G$ and $U_i \sim N(0, 1)$. Their data is different from what we consider here. They first compute the empirical distribution functions G_m and F_n, then construct an $\hat{G}(\cdot)$ and based on $\hat{G}$ construct estimator $\hat{g}_{m,n}$ of $g(\cdot)$ by kernel smoothing $\hat{G}$. They studied strong consistency, convergence rate and integrated squared error of their estimator, without providing asymptotic distribution. We provided asymptotic distribution for our estimator.

In recent years, the multivariate current status and other multivariate interval censoring data, generalizations of one dimensional interval censoring, has received much attention. The multivariate current status data can be briefly described as follows. Let $X = (X_1, \ldots, X_d) \sim F$ on $R^{+d} = [0, \infty)^d$ and $T = (T_1, \ldots, T_d) \sim G$ on R^{+d} be independent of X. The interest is to estimate the unknown multivariate distribution F and G is assumed to be known. The observation is $Y = (\Delta, T)$ where $\Delta = (\Delta_1, \ldots, \Delta_d)$ is given by $\Delta_j = 1_{[X_j \le T_j]}$. The characterizations and computations of the NPMLE for the case $d = 2$ have been studies by [40], [41], [42]. [43] established the consistency of the generalized MLE for a class of multivariate interval censoring models. [44] have obtained global rates of convergence of the MLE for multivariate current status data. The conditional density estimation in the present paper could be directly extended into the multivariate case without much difficulty. But the investigation of the large-sample theories may be challenging.

Appendix

The following more conditions are needed for Theorems 4 and 5.

(C11). $\int_0^x \frac{f(z)}{(1-F(z))^2} dz + \int_x^\infty \frac{f(z)}{F^2(z)} dz < \infty$.

(C12). $\int_0^x \frac{g_v(z)}{1-F(z)}dz + \int_x^\infty \frac{g_u(z)}{F(z)}dz < \infty$.
(C13). $\Lambda(r,t)$ is second order continuous differentiable.

Recall $g(t) = \int_t^\infty (1/s)F(ds)$. Some conditions are listed below.
(C14). $\sup_{x\in R^+} \frac{f(x)}{x} \int_0^x \frac{1}{tg^2(t)}F(dt) < \infty$.
(C15). $\int_0^x g^{-1}(t)dt + \int_0^x g^{-2}(t) \int_t^\infty \frac{1}{v}\lambda(x,v)dvdt$
$+ \int_0^x g^{-3}(t) \int_t^\infty \int_t^\infty \frac{1}{uv}\lambda(u,v)dudvdt < \infty$.

We first describe the basic strategy for the proofs of Theorems 1-3, parts (i) and (ii). For part (i), we decompose $f_n(x|\hat{F}_n) - f(x)$ into three parts, i.e.,

$$\begin{aligned} f_n(x|\hat{F}_n) - f(x) &= \Big(f_n(x|\hat{F}_n) - f_n(x|F)\Big) + \Big(f_n(x|F) - \mathbb{E}[f_n(x|F)]\Big) \\ &\quad + \Big(\mathbb{E}[f_n(x|F)] - f(x)\Big) := V_{1,n}(x) + V_{2,n}(x) + V_{3,n}(x). \end{aligned}$$

It is sufficient to show that under the given conditions, the followings hold:

$$\sup_{x\in[a,b]} |V_{1,n}(x)| \to 0, \quad a.s, \tag{A.1}$$

$$\sup_{x\in[a,b]} |V_{2,n}(x)| \to 0, \quad a.s, \tag{A.2}$$

$$\sup_{x\in[a,b]} |V_{3,n}(x)| \to 0, \quad a.s. \tag{A.3}$$

Since $\mathbb{E}[f_n(x|F)] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(h^{-1}k[(X_i - x)/h])$ is the expectation of the classical kernel estimator, by results of standard density estimation (e.g., Theorem 2.1.1 in [5]), under (C1) and (C3), we have (A.3). Therefore, we only prove (A.1) and (A.2).

For part (ii), the proofs of the central limit theorem of $f_n(x|\hat{F}_n)$ take two steps. First, we show

$$(nh)^{1/2}(f_n(x|F) - \mathbb{E}f_n(x|F)) \xrightarrow{D} N(0,\sigma^2), \tag{A.4}$$

by constructing standard double arrays. Second, when $K(x|F,h,Y)$ has order $r \geq 1$ at F and x in the direction A, then by (C7) and the functional delta method,

$$n^{r/3}(f_n(x|\hat{F}_n) - f_n(x|F)) \xrightarrow{D} \lim_{n\to\infty} \frac{1}{r!}\mathbb{E}[K^{(r)}(x|F,h_n,Y;AZ,...,AZ)], \tag{A.5}$$

Combining (A.4) and (A.5) will yield Theorems 1-3 (ii). Note that for $r=1$, $\mathbb{E}[K^{(1)}(x|F,h_n,Y;AZ)] = \mathbb{E}[K^{(1)}(x|F,h_n,Y;A)]Z = L_1(x|F;A)Z$.

The kernel density estimators under doubly censoring and multiplicative models are the sum of the classical estimators and conditional estimators. The large-sample results for the part of classical estimators could be derived from the standard density estimation theory. For the asymptotic results of conditional parts, the proofs are similar to those of Theorems 1-3. These ingredients combine to prove Theorems 4-5.

Proof of Theorem 1. (i) Let $Q_n(\delta,t)$ be the empirical distribution of the observations $Y_1,\ldots,Y_n$, $Q(\delta,t)$ be the distribution of (δ,T). Then as in [5]

$$\begin{aligned}
&\sup_{x\in[a,b]}|V_{2,n}(x)| \\
&=\sup_{x\in[a,b]}\Big|\sum_{\delta=0,1}\int_{-\infty}^{\infty}K(x|F,h,\delta,t)Q_n(\delta,dt)-\sum_{\delta=0,1}\int_{-\infty}^{\infty}dK(x|F,h,\delta,t)Q(\delta,t)\Big| \\
&\le\sup_{x\in[a,b]}\sum_{\delta=0,1}\int_{-\infty}^{\infty}|Q_n(\delta,t)-Q(\delta,t)|\Big|K(x|F,h,\delta,dt)\Big| \\
&\le\sup_{x}\sum_{\delta=0,1}\sup_{t}|Q_n(\delta,t)-Q(\delta,t)|\frac{1}{h}\mu(x), \qquad\qquad (A.6)
\end{aligned}$$

where $\mu(x)=\max\{\mu_0(x),\mu_1(x)\}$ and $\mu_\delta(x)$ is the total variation of $K(x|F,h,\delta,t)$ as a function of t (for a function $g(x)$, its total variation is defined as $\sup_{\mathcal{P}}\sum_i|g(x_{+1})-g(x_i)|$, where supreme is over all partitions over the support of $g(\cdot)$). Note

$$\begin{aligned}
K(x|F,h,0,t)&=\frac{1}{h}\int k(\frac{s-x}{h})\frac{1_{(t,\infty)}(s)F(ds)}{1-F(t)} \\
&=\int k(v)\frac{1_{(t,\infty)}(x+hv)f(x+hv)dv}{1-F(t)} \\
&=\frac{1}{1-F(t)}\int_{(t-x)/h}^{\infty}k(v)f(x+hv)dv.
\end{aligned}$$

Note that $x>0$. If $0<x<t$, then $\int_{(t-x)/h}^{\infty}k(v)f(x+hv)dv=o(1)$; if $x>t$, then $\int_{(t-x)/h}^{\infty}k(v)f(x+hv)dv=(1+o(1))f(x)$. Therefore,

$$K(x|F,h,0,t)=(1+o(1))\frac{f(x)1_{(t,\infty)}(x)}{1-F(t)}:=(1+o(1))K_0(t). \qquad (A.7)$$

The derivative of $K_0(t)$ on t is

$$K_0'(t)=-1_{(t,\infty)}(x)f(x)\times\frac{f(t)}{(1-F(t))^2}.$$

By Condition (C4) and $\int f(x)=1$, it is easy to show that $f(\cdot)$ is bounded. Further, $\int_0^\infty|K_0'(t)|dt=f(x)\int_0^x[1/(1-F(t)^2]dF(t)=f(x)F(x)/(1-F(x))<C$ uniformly over $[a,b]$ for some constant $0<C<\infty$. Thus $K_0(\cdot)$ has uniformly (over x) bounded variation on $[a,b]$, which in turn implies $K(x|F,h,0,t)$ has a uniformly (over x) bounded variation on $[a,b]$, i.e, $\mu_0:=\sup_{x\in[a,b]}\mu_0(x)<\infty$.

By the similar argument, $\mu_1 := \sup_{x\in[a,b]} \mu_1(x) < \infty$, and so $\mu = \sup_{x\in[a,b]} \mu(x) < \infty$. Now we have

$$\sup_{x\in[a,b]} |V_{2,n}(x)| \le \mu \sum_{\delta=0,1} \sup_t |Q_n(\delta,t) - Q(\delta,t)|.$$

Let

$$D_n(j) = \sup_{-\infty<t<\infty} |Q_n(j,t) - Q(j,t)|, \quad (j = 0, 1),$$

by the results in [45], there exist positive constants C and α, $0 < \alpha \le 2$ such that

$$P(D_n(j) > \lambda n^{-1/2}) \le C \exp(-\alpha\lambda^2), \quad (j = 0, 1)$$

for every $\lambda > 0$ and any continuous distribution function F. Now we get

$$\begin{aligned} P\Big(\sup_{x\in[a,b]} |V_{2,n}(x)| > \varepsilon\Big) &\le P\Big(\sup_t \sum_{\delta=0,1} |Q_n(\delta,t) - Q(\delta,t)| > \varepsilon h\mu^{-1}\Big) \\ &\le 2C \exp(-\beta n), \end{aligned}$$

where $\beta = \alpha\varepsilon^2\mu^{-2}/4$. Since $\sum_{n=1}^{\infty} \exp(-\beta n h^2) < \infty$ for any given $\varepsilon > 0$, by the Borel-Cantelli lemma, we have (A.1).

Since $\sup_t |\hat{F}_n(t) - F(t)| \to 0$ (a.s.), we have $K(x|\hat{F}_n, h, y_i) = K(x|F, h, y_i) + o(1)$ (a.s.) uniformly over (x, y_i). This gives

$$V_{1,n}(x) = \frac{1}{n}\sum_{i=1}^{n}[K(x|\hat{F}_n, h, Y_i) - K(x|F, h, Y_i)] = o(1), \quad a.s.$$

This proves part (i).

(ii). Let $Z_{i,h} = K(x|F, h, Y_i) - \mathbb{E}[K(x|F, h, Y_i)]$, then $Z_{1,h}, \dots, Z_{n,h}$ are i.i.d, with $\mathbb{E}(Z_{i,h}) = 0$. Since $\mathbb{E}[K(x|F, h, Y_i)] = \frac{1}{h}\mathbb{E}[k(\frac{X_i - x}{h})]$, we have

$$f_n(x|F) - \mathbb{E}f_n(x|F) = \frac{1}{n}\sum_{i=1}^{n} Z_{i,h}$$

is the average of the i.i.d. double array $\{Z_{i,h}\}$. By (C3), we have

$$\mathbb{E}[K(x|F, h, Y_i)] = \frac{1}{h}\mathbb{E}[k(\frac{X_i - x}{h})] = \int k(v) f(x + hv) dv \to f(x), \quad \text{as} \quad h \to 0.$$

Note the joint distribution-mass function of $(T, 1_{(X<T)})$ is

$$P(T \le t, \delta = 1) = \int_0^t \int_0^y f(x) dx g(y) dy = \int_0^t F(y) g(y) dy,$$

and so the density-mass function of $(T, 1_{(X<T)})$ is

$$\frac{d}{dt}\int_0^t F(y) g(y) dy = F(t) g(t).$$

Similarly,the joint distribution-mass function of $(T, 1_{(X \geq T)})$ is

$$P(T \leq t, \delta = 0) = \int_0^t \int_y^\infty f(x)dxg(y)dy = \int_0^t (1 - F(y))g(y)dy,$$

and so the density-mass function of $(T, 1_{(X \geq T)})$ is

$$\frac{d}{dt} \int_0^t (1 - F(y))g(y)dy = (1 - F(t))g(t).$$

Thus,

$$\begin{aligned}
\mathbb{E}[K^2(x|F, h, Y_i)] &= \frac{1}{h} \int \left(\int k(v) \frac{1_{[0,t]}(x + hv)f(x + hv)dv}{F(t)} \right)^2 F(t)g(t)dt \\
&\quad + \frac{1}{h} \int \left(\int k(v) \frac{1_{(t,\infty)}(x + hv)f(x + hv)dv}{1 - F(t)} \right)^2 (1 - F(t))g(t)dt \\
&= \frac{1}{h} \int \left(\int_{-x/h}^{(t-x)h} k(v)f(x + hv)dv \right)^2 F^{-1}(t)g(t)dt \\
&\quad + \frac{1}{h} \int \left(\int_{(t-x)/h}^{\infty} k(v)f(x + hv)dv \right)^2 (1 - F(t))^{-1}g(t)dt.
\end{aligned}$$

Note $x > 0$, so if $t < x$, then $\int_{-x/h}^{(t-x)/h} k(v)f(x + hv)dv = o(1)$; if $t > x$, then $\int_{-x/h}^{(t-x)/h} k(v)f(x + hv)dv = f(x) + o(1)$, so

$$\int \left(\int_{-x/h}^{(t-x)h} k(v)f(x + hv)dv \right)^2 F^{-1}(t)g(t)dt = [f^2(x) + o(1)] \int_x^\infty \frac{g(t)}{F(t)}dt.$$

Similarly,

$$\begin{aligned}
&\int \left(\int_{(t-x)/h}^{\infty} k(v)f(x + hv)dv \right)^2 (1 - F(t))^{-1}g(t)dt \\
&= [f^2(x) + o(1)] \int_0^x \frac{g(t)}{(1 - F(t))}dt.
\end{aligned}$$

Thus, with "$\sim$" standing for asymptotically equivalent,

$$\begin{aligned}
Var(Z_{i,h}) &= \mathbb{E}[K^2(x|F, h, Y_i)] - \{\mathbb{E}[K(x|F, h, Y_i)]\}^2 \sim \mathbb{E}[K^2(x|F, h, Y_i)] \\
&= h^{-1}f^2(x)\left(\int_0^x \frac{g(t)}{1 - F(t)}dt + \int_x^\infty \frac{g(t)}{F(t)}dt \right) = h^{-1}\sigma^2. \quad \text{(A.8)}
\end{aligned}$$

Consequently, (A.4) comes immediately from the central limit theorem for i.i.d. double array if (C3) hold.

Now we prove (A.5) for $r = 1$. First we have

$$
\begin{aligned}
& L_1(x|F,Y;A) = \lim_{h\to 0} K^{(1)}(x|F,h,Y;A) \\
= \; & \lim_{h\to 0}\int k(v)\Bigg(\delta\frac{1_{[0,T]}(x+hv)[F(T)a(x+hv)-A(T)f(x+hv)]}{F^2(T)} \\
& +(1-\delta)\frac{1_{(T,\infty)}(x+hv)[(1-F(T))a(x+hv)+A(T)f(x+hv)]}{(1-F(T))^2}\Bigg)dv \\
= \; & \lim_{h\to 0}(1+o(1))\int k(v)\Bigg(\delta\frac{1_{[0,T]}(x)[F(T)a(x)-A(T)f(x)]}{F^2(T)} \\
& +(1-\delta)\frac{1_{(T,\infty)}(x)[(1-F(T))a(x)+A(T)f(x)]}{(1-F(T))^2}\Bigg)dv \\
= \; & \delta\frac{1_{[0,T]}(x)[F(T)a(x)-A(T)f(x)]}{F^2(T)} + \\
& (1-\delta)\frac{1_{(T,\infty)}(x)[(1-F(T))a(x)+A(T)f(x)]}{(1-F(T))^2}.
\end{aligned}
$$

Let $l^\infty(R^+)$ be the space of all bounded functions on R^+ equipped with the supremum metric. For fixed t with $f(t) > 0$ and $g(t) > 0$ and conditions (C5) and (C6), it is known (see, for example, [12], that

$$
n^{1/3}[\hat{F}_n(t) - F(t)] \xrightarrow{D} A(t)Z. \quad \text{in} \quad l^\infty(R^+),
$$

so by the functional delta method (e.g. [46]), for fixed Y, we have

$$
\begin{aligned}
n^{1/3}[K(x|\hat{F}_n,h,Y) - K(x|F,h,Y)] &= K^{(1)}(x|F,h,Y;AZ) + o_p(1) \\
&= L_1(x|F,Y;A)Z + o_p(1).
\end{aligned}
$$

Since the weak convergence of $n^{1/3}[\hat{F}_n(t) - F(t)]$ is in $l^\infty(R^+)$, in the above the $o(1)$ is uniform over Y, by the strong law of large numbers, for small b and large d we obtain

$$
\begin{aligned}
& n^{1/3}(f_n(x|\hat{F}_n) - f_n(x|F)) \\
= \; & \frac{1}{n}\sum_{i=1}^n n^{1/3}[K(x|\hat{F}_n,h,Y_i) - K(x|F,h,Y_i)](1_{[b,d]}(Y_i) + 1_{[b,d]^c}(Y_i)) \\
= \; & \frac{1}{n}\sum_{i=1}^n [L_1(x|F,Y_i;A)Z1_{[b,d]}(Y) + o_p(1)] \\
\xrightarrow{D} \; & \mathbb{E}[L_1(x|F,Y;A)]Z + \epsilon,
\end{aligned}
$$

where ϵ is such that $\frac{1}{n}\sum_{i=1}^n n^{1/3}[K(x|\hat{F}_n,h,Y_i) - K(x|F,h,Y_i)]1_{[b,d]^c}(Y_i) \sim \epsilon$, which can be arbitrarily small, and so $\mathbb{E}[L_1(x|F,Y;A)1_{[b,d]}(Y)]Z$ is arbitrarily close to $\mathbb{E}[L_1(x|F,Y;A)]Z$. Thus we have

$$
n^{1/3}(f_n(x|\hat{F}_n) - f_n(x|F)) \xrightarrow{D} \mathbb{E}[L_1(x|F,Y;A)]Z.
$$

Note $L_1(x|F;A)$ is finite and is given by

$$\begin{aligned} L_1(x|F;A) & \\ =\quad & \mathbb{E}[L_1(x|F,Y;A)] \\ =\quad & a(x)+f(x)\bigg(\int_0^x \frac{A(t)}{1-F(t)}g(t)dt-\int_x^\infty \frac{A(t)}{F(t)}g(t)dt\bigg). \end{aligned}$$

Similarly, when $K^{(r)}(\cdot|F,h,Y)$ has order $r>1$, we have

$$n^{r/3}(f_n(x|\hat{F}_n)-f_n(x|F))\xrightarrow{D}\frac{1}{r!}\mathbb{E}[L_r(x|F,Y;A,...,A)]Z^r.$$

i.e., we proved (A.5).

If $r=1$ and $hn^{1/3}\to\infty$, with (A.4) and (A.5), then $[n^{1/3}/(nh)^{1/2}] = 1/\sqrt{hn^{1/3}}\to 0$, and by standard result for kernel density estimator, $(nh)^{1/2}(f_n(x|F)-\mathbb{E}f(x|F))$ is asymptotic normal, so we have

$$\begin{aligned} n^{1/3}(f_n(x|\hat{F}_n)-\mathbb{E}f_n(x|F)) &= n^{1/3}(f_n(x|\hat{F}_n)-f_n(x|F)) \\ &+[n^{1/3}/(nh)^{1/2}](nh)^{1/2}(f_n(x|F)-\mathbb{E}f_n(x|F)) \\ &= n^{1/3}(f_n(x|\hat{F}_n)-f_n(x|F))+o_p(1)\xrightarrow{D}\mathbb{E}[L_1(x|F,Y;A)]Z; \end{aligned}$$

If $r=1$ and $hn^{1/3}\to 0$, then $(nh)^{1/2}/n^{1/3}=(hn^{1/3})^{1/2}\to 0$, and

$$\begin{aligned} (nh)^{1/2}(f_n(x|\hat{F}_n)-\mathbb{E}f_n(x|F)) &= [(nh)^{1/2}/n^{1/3}]n^{1/3}(f_n(x|\hat{F}_n)-f_n(x|F)) \\ &+(nh)^{1/2}(f_n(x|F)-\mathbb{E}f_n(x|F)) \\ &= (nh)^{1/2}(f_n(x|F)-\mathbb{E}f_n(x|F))+o_p(1)\xrightarrow{D}N(0,\sigma^2). \end{aligned}$$

Similarly, if $r>1$, then $(nh)^{1/2}/n^{r/3}=(hn^{1-2r/3})^{1/2}\to 0$ and

$$\begin{aligned} (nh)^{1/2}[f_n(x|\hat{F}_n)-\mathbb{E}f(x|F)] &= [(nh)^{1/2}/n^{r/3}]n^{r/3}(f_n(x|\hat{F}_n)-f_n(x|F)) \\ &+(nh)^{1/2}(f_n(x|F)-\mathbb{E}f_n(x|F)) \\ &= (nh)^{1/2}(f_n(x|F)-\mathbb{E}f_n(x|F))+o_p(1)\xrightarrow{D}N(0,\sigma^2). \end{aligned}$$

Also, recall $f_n(x)$ is the kernel density estimate of $f(x)$ based on $X_1,\dots,X_n$, assume $\int x^r k(x)dx=0$ $(1\le r\le m-1)$, $\int x^m k(x)dx\ne 0$ and f being m-th differentiable, then $Ef_n(x|F)=Ef_n(x)=f(x)-h_n^m A_m f^{(m)}(x)/m!+o(h_n^m)$, with $A_m=\int x^m k(x)dx$ [5], In the case $m=2$ and $nh_n^5\to 0$, we have $(hn)^{1/2}h=(nh_n^5)^{1/2}\to 0$. This completes the proof of Theorem 1 (ii). □

The proofs of Theorems 2-5 are similar and omitted, and can be provided upon request.

Acknowledgments

We are very grateful to two anonymous referees, an associate editor, and the Editor for their valuable comments that have greatly improved the manuscript.

References

[1] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, 2nd ed. **27** 832–837. MR0079873

[2] Parzen E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.*, 2nd ed. **33** 1065–1076. MR0143282

[3] Devroye, L. (1983). The equivalence of weak, strong and complete convergence in L1 for kernel density estimates. *Annals of Statistics.*, **11** 896–904. MR0707939

[4] Chai, G. (1984). Consistent estimation of random window-width kernel of density function. *Science in China.*, Ser. A **27** 1155–1163. MR0794287

[5] Rao, B. L. S. (1983). *Nonparametric functional estimation.*.Academic Press: Orlando, Florida.

[6] Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. *Chapman and Hall, London.*, x+175 pp. ISBN: 0-412-24620-1. MR0848134

[7] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481. MR0093867

[8] Turnbull, B.W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of American Statistical Association.* **69** 169–173. MR0381120

[9] Marron J. S. and Padgett W. J. (1987). Asymptotically Optimal Bandwidth Selection for Kernel Density Estimators from Randomly Right-Censored Samples. *Annals of Statistics.*, **15** 1520–1535. MR0913571

[10] Chao M. T. and Lo S. H. (1988). Some representations of the nonparametric maximum likelihood estimators with truncated data. *Annal of Statistics.*, Ser. A **16** 661–668.

[11] Lo S. H., Mack Y. P. and Wang J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Probab. Th. Rel. Fields.*, **80** 461–473. MR0794287

[12] Groeneboom, P. and Wellner, J. (1992). Information Bounds and Nonparametric Maximum Likelihood Estimation. *Birkhauser Verlag, Basel.*, viii+126 pp. ISBN: 3-7643-2794-4 MR1180321

[13] Gijbels I. and Wang J.-L. (1993). Strong representations of the survival function estimator for truncated and censored data with applications. *Journal of Multivariate analysis.*, **47** 210–229. MR1247375

[14] Heagerty, P. and Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics.*, **61** 92–105. MR2135849

[15] Blum, J. R. and Susarla, V. (1980). Maximal deviation theory of density and failure estimates based on censored data. *In Multivariate Analysis.V (Proc. Fifth Internat. Sympos., Univ. Pittsburgh, Pittsburgh, Pa., 1978)*, 213–222. MR0566340

[16] Mielniczuk, J. (1986). Some asymptotic properties of kernel estimators of a density function in case of censored data. *Annals of Statistics.*,**14** 766–773. MR0840530

[17] Dubnicka, S. R. (2009). Kernel density estimation with missing data: misspecifying the missing data mechanism. *In Nonparametric Statistics and Mixture Models: A Festschrift in Honor of Thomas P. Hettmansperger, ed. D.R. Hunter, D.St.P. Richards, and J.L. Rosenberger*, 114–135. MR2838723

[18] Dubnicka, S. R. (2009). Kernel density estimation with missing data and auxiliary variables.Australian and New Zealand Journal of Statistics. *Australian and New Zealand Journal of Statistics.*, **51** 247–270. MR2569798

[19] Ren J. (1997). On self-consistent estimators and kernel density estimators with doubly censored data. *Journal of Statistical Planning and Inference.* **64** 27–43. MR1492359

[20] Gine E. and Guillou A. (2001). On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals. *Annals de l'Institut Henri Poincare (B) Probability and Statistics.*, **37** 503–522. MR1876841

[21] Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika.*, **76** 751–761. MR1041420

[22] Bickel , P. J., Klaassen , A. J., Ritov , Y. and Wellner , J. A. (1993). Consistent estimation of random window-width kernel of density function. *Science in China.*,Johns Hopkins Univ. Press, Baltimore.

[23] van der Vaart, A. (1994). Maximum likelihood estimation with partially censored data. *Annals of Statistics.*, **22** 1896–1916. MR1329174

[24] Asgharian M, Carone M. And Fakoor V. (2012). Consistent estimation of random window-width kernel of density function. *Ann. Statist.* **40** 159–187. MR3013183

[25] Huang, J. and Wellner, J. A. (1995a). Asymptotic normality of the npmle of linear functionals for interval censored data, case 1. *Statistica Neerlandica.*, **49** 153–163. MR1345376

[26] Huang, J. and Wellner, J. A. (1995b). Estimation of a monotone density or monotone hazard under random censoring.*Scandinavian Journal of Statistics.* **22** 3–33. MR1334065

[27] Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics.* **14** 945–965.

[28] Aalen, O. O. (1984). Nonparametric inference for a family of counting processes. *Annals of Statistics.*, **6** 701–726. MR0491547

[29] Yuan, A., Giurcanu, M., Luta, G., Tan, M. (2017). U-statistic with conditional kernels for incomplete data models. *Annals of the Institute of Statistical Mathematics.*, **69** 271–302. MR3611521

[30] Watson G. S. (1964). Smooth regression analysis *Sankhya.*, Series A **26** 359–372. MR0185765

[31] Nadaraya, E. A. (1965). On nonparametric estimation of density functions and regression curves. *Theory Prob.*, Appl. **10** 186–190. MR0172400

[32] Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Annals of Statistics.* **6** 177–184. MR0471166

[33] Huang, J. and Wellner, J. A. (1997). Interval Censored Survival Data: A Review of Recent Progress.*Proceedings of the First Seattle Symposium in Biostatistics Lecture Notes in Statistics.* **123** 123–169.

[34] van der Laan, M. J. (1993). General identity for linear parameters in convex models with application to eciency of the (NP)MLE. *Preprint No. 765*, Department of Mathematics, University of Utrecht, The Netherlands.

[35] Geskus, R. and Groeneboom, P. (1999). Asymptotically optimal estimation of smooth functionals for interval censoring, case 2. *Annals of Statistics.* **27** 627–674. MR1714713

[36] Tsai, W.-Y. and Crowley, J. (1985). A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency. *Annals of Statistics.*, **13** 1317–1334. MR0811495

[37] Vardi, Y. and Zhang, C-H. (1992). Large sample study of empirical distributions in a random-multiplicative censoring model. *Annals of Statistic.* **20** 1022–1039. MR1165604

[38] Wang, X. F. and Wang, B. (2011). Deconvolution estimation in measurement error models: The R package decon. *Science in China.* **39(10)** 1–24.

[39] Groeneboom, P., Jongbloed, G., Witte, B.I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model.*Annals of Statistic.*,**38(1)** 352–387. MR2589325

[40] Song, S. (2001). Estimation with bivariate interval-censored data. *PhD thesis.*, University of Washington, Department of Statistics. MR2717005

[41] Gentleman, R. and Vandal, A. C. (2002). Nonparametric estimation of the bivariate CDF for arbitrarily censored data. *Canad. J. Statist* **30** 557–571. MR1964427

[42] Maathuis M. H. (2005). eduction algorithm for the NPMLE for the distribution function of bivariate interval-censored data. *J. Comput. Graph. Statist.* **14** 352–362. MR2160818

[43] Yu S., Yu Q. and Wong G. Y. C. (2006). Consistency of the generalized MLE of a joint distribution function with multivariate interval-censored data. *J. Multivariate Anal.*, Ser. A **27** 1155–1163. MR2236498

[44] Gao F. and Wellner J. A. (2013). Global rates of convergence of the MLE for multivariate interval censoring. *Electronic Journal of Statistics.* **7** 364–380. MR3020425

[45] Dvoretzky A., Kiefer J., and Wolfowitz J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics.* **27** 642–669. MR0083864

[46] van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.*, Springer-Verlag, New York. MR1385671