

Approximate Nonparametric Maximum Likelihood Estimation for Interval Censoring Model Case II (Running Head: NPMLLE for Interval Censoring Case II)

Ao Yuan¹, Yizheng Wei², Kephher Makambi¹ & Hongkun Wang¹

¹ Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington DC, USA

² Department of Statistics, University of South Carolina, Columbia SC, USA

Correspondence: Ao Yuan, Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University, Washington DC 20057, USA. E-mail: yuanao@hotmail.com

Received: April 16, 2014 Accepted: May 29, 2014 Online Published: June 11, 2014

doi:10.5539/ijsp.v3n3p44

URL: <http://dx.doi.org/10.5539/ijsp.v3n3p44>

Abstract

We study the nonparametric maximum likelihood estimate of the distribution function in a type II interval censoring model. We propose an approximate solution of the problem under a technical assumption. Some basic asymptotic properties of the estimator are investigated.

Keywords: interval censoring case II, isotonic regression, nonparametric maximum likelihood estimate

1. Introduction

Interval censoring models are commonly used in practice, especially in the biomedical sciences, such as in many clinical trials and longitudinal studies, in acquired immune deficiency syndrome studies, or the studies of HIV infection times, etc. There are two basic types of interval censoring models. In the interval censoring model case I, also termed current status data model, the left (or right) censoring and the status variables of the original data are observed; while in the interval censoring model case II, two censoring variables and the statuses are observed. The type I interval censoring model is much more commonly used in practice, and there are rich statistical research and application literature about this model, for example, Huang (1996), Bechuk and Betensky (2000), Chen and Jewell (2001), Yu et al. (2000), Pan and Chappell (2002). Fang and Sun (2001) studied the nonparametric maximum likelihood estimation (NPMLLE) of the doubly interval censored model. Type II interval censoring model is less commonly used in practice, and its mechanism is more complicated, and there are some research literature on it, for example, in the book of Groeneboom and Wellner (1992), Wellner (1995), Groeneboom (1996), Jongbloed (1998). To our knowledge, investigation of this model is still incomplete, NPMLLE of the distribution function exists, can be given by a nonparametric EM algorithm, and its strong consistency is obtained, but its asymptotic distribution seems not available yet. A good review of the topic of the two types of censoring models can be found in Sun (2012). Also, these models, along with the other survival models, provide practical background for the theory of the estimation of Banach space valued parameters.

Unlike Euclidean parameters, Banach parameters are often not \sqrt{n} estimable, in that there exists no estimator which is consistent at rate \sqrt{n} , and when the weak limits of estimators exist, they are often non-Gaussian. Also, unlike maximum likelihood estimators for Euclidean parameters, Banach parameters have no standard method of estimation. For some models, setting the normal equation of the Hadamard differential (along some particular selected direction) to zero will yield a version of the nonparametric maximum likelihood estimation (NPMLLE), although it may not be unique. The isotonic regression is another commonly used method for this problem. Originally this method is used in the optimization of a class of regression problem, such as weighted least squares problem in the regression context, in which the response and covariate variables have some particular relationship (isotonic relationship, as characterized by Theorem 1.4.1 in Robertson, Wright and Dykstra, 1988), such as monotone relationship. Then it was found that the method can be used to solve the NPMLLE for some models, as a optimization procedure. See Robertson, Wright and Dykstra (1988) for a systematic account of this method. Early usage of this method to compute NPMLLE can be found in Grenander (1956), Chernoff (1964), Prakasa Rao (1969).

This method works only for certain models with log-likelihood satisfying some convexity condition. For some models, one of the methods works; for some models, both methods work; for many other models, none of the methods works, and often approximate NPMLE can be found which maximizes the log-likelihood up to a small constant ϵ_n , with $\epsilon_n \rightarrow 0$. Such estimates are also called ϵ_n -MLE, studied by a number of authors, such as Wald (1949), Bahadur (1967), and Wong and Severini (1991), and are shown to have some nice properties. For interval censoring model case II, Groeneboom and Wellner (1992), hereafter GW, studied the NPMLE of the distribution function, which uses iterative procedures for its computation. They showed the convergence rate of $n^{1/3} \log n$ for the NPMLE when the data have some mass distributed around the diagonal, otherwise the convergence rate is $n^{1/3}$, like that for censoring model case I, but the formal identification of the asymptotic distribution is not easy, although there is a conjecture about it. Another NPMLE is the one step iteration in the iterative convex minorant algorithm, with the true distribution as starting value. Groeneboom and Wellner justified the asymptotic distribution of the one-step iteration “estimator”, obtained its asymptotic distribution. Since this one-step iteration is not an actual estimator, the result can only be used as a technical tool in the further study of the NPMLE. The iterative convex minorant is one of the computational algorithms of the isotonic regression, it makes the optimization problem into a sequence of weighted isotonic regression problems, with the weights obtained by the values at the preceding step. It is known that the parameter (a distribution function) in this model is not rate- \sqrt{n} estimable, not even for many smooth functionals of it (for example, Yuan, Xu, & Zheng, 2012). Geskus and Groeneboom (1999) showed rate \sqrt{n} efficient estimability for certain smooth functionals (mostly linear functionals) of this model. In this article we study a slightly different version of this model in which we observe the statuses of two right censoring variables. It is known that the (NP)MLE is neither always efficient, nor consistent, nor optimal (Wellner, 2005). Using the isotonic regression method, we find an approximate NPMLE of our model in that it maximizes the averaged log-likelihood up to a $o(n^{-1/\alpha})$ (a.s.) term ($0 < \alpha < 1/2$). It can be computed in closed form, we show rate $n^{1/3}$ weak consistency of it and evaluate its asymptotic distribution, study its strong consistency, convergence rate in probability, rate \sqrt{n} estimability for certain smooth functionals of it, and for some submodels of it.

In Section 2 we describe the model, the approximate NPMLE of the underlying distribution function, and some basic results. Section 3 gives some illustration for the uses of this model and its estimation. Relevant derivations are given in the Appendix.

2. Approximate NPMLE of Interval Censoring Model Case II

In this model, let $(Z, U, V) \in R^+ \times (R^+)^2$, where $Z \sim F$ and $(U, V) \sim G$ are independent, and $U < V$ a.s.. G has a density g with respect to the Lebesgue measure on $(R^+)^2$. We observe $[(u_i, v_i, 1_{[z_i \leq u_i]}, 1_{[u_i < z_i \leq v_i]}) : i = 1, \dots, n]$ i.i.d. with $(U, V, 1_{[Z \leq U]}, 1_{[U < Z \leq V]}) := (U, V, \delta, \gamma)$, with $0 < P(\delta = 1) < P(\gamma = 1) < 1$. The density-mass function of $(U, V, 1_{[Z \leq U]}, 1_{[U < Z \leq V]})$ is (Groeneboom & Weller, 1992, Example 1.6, p. 5)

$$p_F(u, v, \delta, \gamma) = F^\delta(u)(F(v) - F(u))^\gamma(1 - F(v))^{1-\delta-\gamma}g(u, v). \quad (0)$$

They used an iterative convex minorant algorithm to compute the NPMLE $\hat{F}_n(\cdot)$ of $F(\cdot)$ based on the above model. To study the asymptotic distribution of $\hat{F}_n(\cdot)$, to study the asymptotic distribution of $\hat{F}_n(\cdot)$, they used the following *working hypothesis*: starting from the real underlying distribution function F , the iterative convex minorant algorithm will give at the first iteration step estimator $\hat{F}_n^{(1)}(\cdot)$, which is asymptotically equivalent to the maximum likelihood estimator $\hat{F}_n(\cdot)$.

Let \xrightarrow{D} denote convergence in distribution. GW justified the working hypothesis as below. Assuming that $f(t) > 0$, $g(t, t) > 0$ and that $g(t, \cdot)$ is left continuous at t , they obtained (Theorem 5.3, GW, p. 100)

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

where $\mathbb{B}(h)$ is the two-sided Browning 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 $|r - h|$.

The proposed model and method. Our model below is slightly different from (0), in that we define $\gamma = 1_{[Z \leq V]}$ (instead of $1_{[U < Z \leq V]}$). In Example 1.6 of GW, the density-mass function of $(U, V, 1_{[Z \leq U]}, 1_{[Z \leq V]})$ is given as in (0), but we find that (0) is actually the density-mass of $(U, V, 1_{[Z \leq U]}, 1_{[U < Z \leq V]})$. The distribution P_F of (U, V, δ, γ) has the following density/mass function

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

on $(R^+)^2 \times \{0, 1\}^2$. Here our main interest is to estimate F in model (1), as $g(u, v)$ is factored out in the model, its estimate is straight-forward using the data $(u_1, v_1), \dots, (u_n, v_n)$ by many existing methods, such as the kernel estimator. So we assume $g(\cdot, \cdot)$ is known. The averaged log-likelihood for F is

$$l_n(F) = \frac{1}{n} \sum_{i=1}^n \left(\delta_i \log F(u_i) + (\gamma_i - \delta_i) \log(F(v_i) - F(u_i)) + (1 - \gamma_i) \log(1 - F(v_i)) \right).$$

The approximate NPMLE $\hat{F}_n(\cdot)$ of $F(\cdot)$ is the maximizer of $l_n(F)$ with respect to F , up to a $o(n^{-1/\alpha})$ (a.s.) term ($0 < \alpha < 1/2$). Unlike MLE for Euclidean parameters, NPMLE for infinite-dimensional parameter in likelihood model (like F in the above model) is not straight forward. Setting the corresponding Hadamard differential of the log-likelihood to zero often leads to nowhere. Instead, we use the method of isotonic regression. For this, rearrange $(u_1, \dots, u_n; v_1, \dots, v_n)$ in increasing order, denoted as (x_1, \dots, x_{2n}) , and $\{\Delta_i: i = 1, \dots, 2n\}$ as the concomitants of the $\{u_i, v_i: i = 1, \dots, n\}$, i.e., $\Delta_i = \delta_j$ if $x_i = u_j$ for some j , and $\Delta_i = \gamma_k$ if $x_i = v_k$ for some k . In the proof of Theorem 1 we will see that

$$l_n(F) = \frac{1}{n} \sum_{i=1}^{2n} \left(\Delta_i \log F(x_i) + (1 - \Delta_i) \log(1 - F(x_i)) \right) + R_n := \tilde{l}_n(F) + R_n$$

with $R_n = o(n^{-\alpha})$ (a.s.) for some $0 < \alpha < 1/2$. Let $\hat{F}_n(\cdot)$ be the greatest convex minorant (Robertson, Wright, & Dykstra, 1988) based on $(\Delta_1, \dots, \Delta_{2n})$, (x_1, \dots, x_{2n}) and $\tilde{l}_n(F)$, as

$$\hat{F}_n = \arg \max_F \tilde{l}_n(F).$$

Let $g_1(\cdot)$ and $g_2(\cdot)$ be the margins of $g(\cdot, \cdot)$. Below, we study the asymptotic distribution of \hat{F}_n . We need the following conditions, in which (C2) is for the averaged log-likelihood $l_n(F)$ to satisfy a convexity condition, up to a $o(n^{-1/\alpha})$ (a.s.) term ($0 < \alpha < 1/2$), so that the isotonic regression method can be applied.

(C1) $f(t) > 0, g_1(t) + g_2(t) > 0$.

(C2) $E\left((F(V) - F(U)) \log(F(V) - F(U)) - (1 - F(U)) \log(1 - F(U)) - F(V) \log F(V)\right) = 0$, where the expectation is for (U, V) .

Theorem 1 Assuming (C1) and (C2), we have

$$n^{1/3}(\hat{F}_n(t) - F(t)) \xrightarrow{D} \left(\frac{4F(t)f(t)}{g_1(t) + g_2(t)} \right)^{1/3} \arg \min_h \{\mathbb{B}(h) + h^2\}.$$

Let $\stackrel{D}{=}$ denote equality in distribution. Note $-\mathbb{B}(h) \stackrel{D}{=} \mathbb{B}(h)$, we have $\arg \min_h \{\mathbb{B}(h) + h^2\} = \arg \max_h \{-\mathbb{B}(h) - h^2\} \stackrel{D}{=} \arg \max_h \{\mathbb{B}(h) - h^2\}$, which has a density symmetric about zero, as given in Corollary 3.3 of Groeneboom (1989).

Next we study the almost sure behavior of the estimator \hat{F}_n , and of any general NPMLE \widehat{F}_n of F of model (1) with or without assumption (2), we are not confined to the isotonic regression method, only assuming its existence. We need the following two conditions.

(C3) $\int_0^\infty \int_0^v \left(\sqrt{F(u)} + \sqrt{F(v) - F(u)} + \sqrt{1 - F(v)} \right) \sqrt{g(u, v)} du dv < \infty$.

(C4) $\int_0^\infty \sqrt{F(t)} (\sqrt{g_1(t)} + \sqrt{g_2(t)}) dt < \infty$.

Theorem 2 (i) Let $\widehat{F}_n(\cdot)$ be any NPMLE of $F(\cdot)$ in model (1) (with or without condition (C2)). Assume (C3), then

$$\lim_{n \rightarrow \infty} \int_0^\infty |\widehat{F}_n(t) - F(t)|(g_1(t) + g_2(t)) dt = 0 \quad a.s.$$

Further, when $F(\cdot)$ is continuous, then

$$\limsup_{n \rightarrow \infty} \sup_t |\widehat{F}_n(t) - F(t)| \rightarrow 0, \quad a.e. (G_1 + G_2), \quad a.s.$$

(ii) Let \hat{F}_n be as in Theorem 1. Assume (C2) and (C4). Then

$$\lim_{n \rightarrow \infty} \int_0^\infty |\hat{F}_n(t) - F(t)|(g_1(t) + g_2(t)) dt = 0 \quad a.s.$$

and when $F(\cdot)$ is continuous, then

$$\limsup_{n \rightarrow \infty} \sup_t |\hat{F}_n(t) - F(t)| \rightarrow 0, \quad a.e. (G_1 + G_2), \quad a.s.$$

Below we investigate the convergence rate of the approximate NPMLE \hat{F}_n and any NPMLE \hat{F}_n of F based on model (1) in probability. For a parameter $b \in \mathcal{B}$, a Banach space, and a probability model p_b , let b_n be an estimator of the true parameter b_0 based on n i.i.d. observations from p_{b_0} , $h(p_{b_n}, p_{b_0})$ be the Hellinger distance between p_{b_n} and p_{b_0} : $h^2(p_{b_n}, p_{b_0}) = (1/2) \int (p_{b_n}^{1/2}(x) - p_{b_0}^{1/2}(x))^2 dx$, $N_{[\cdot]}(\epsilon, \mathcal{B}, \rho)$ be the bracketing covering number of the family \mathcal{B} , of size ϵ with respect some semi-metric ρ on \mathcal{B} , and $H_{[\cdot]}(\epsilon, \mathcal{B}, \rho)$ be the bracketing entropy. It is known (LeCam, 1973; Birgé, 1983; as stated in Wellner, 2005) that the optimal rate r_n of convergence in probability, of estimating b_0 , in the sense $r_n h(p_{b_n}, p_{b_0}) = O_P(1)$, is determined by

$$nr_n^{-2} = H_{[\cdot]}(r_n^{-1}, \mathcal{B}, \rho).$$

Typically, $H_{[\cdot]}(\epsilon, \mathcal{B}, \rho) = O(\epsilon^{-d/\alpha})$, where d is the dimension of the argument of b and α is a smoothness measure of p_b . Thus the optimal rate in this case is $r_n = n^{\alpha/(2\alpha+d)}$.

For NPMLE and more generally the minimum contrast estimator, the best achievable rate r'_n is determined by Birgé and Massart (1993), for some $0 < c < \infty$, as

$$\sqrt{n}(r'_n)^{-2} = \int_{c(r'_n)^{-2}}^{(r'_n)^{-1}} H_{[\cdot]}^{1/2}(\epsilon, \mathcal{B}, \rho) d\epsilon.$$

When $H_{[\cdot]}(\epsilon, \mathcal{B}, \rho) = O(\epsilon^{-d/\alpha})$, the above gives $r'_n = n^{\alpha/(2\alpha+d)}$ if $\alpha/d > 1/2$; $r'_n = n^{\alpha/(2d)}$ if $\alpha/d < 1/2$; and $\log r'_n = \sqrt{n}(r'_n)^{-2}$ if $\alpha/d = 1/2$. So, in this case the NPMLE can achieve the optimal rate of convergence only if $\alpha/d > 1/2$.

Let \mathcal{F} be the collection of distribution functions F in model (1). For $F, F' \in \mathcal{F}$, let $D(p_F \| p_{F'}) = E_F(\log(p_F/p_{F'}))$ be the Kullback-Leibler divergence between p_F and $p_{F'}$ under p_F . Let $p_{1,F}(\delta, u) = F^\delta(u)(1 - F(u))^{1-\delta} g_1(u)$ and $p_{2,F}(\gamma, v) = F^\gamma(v)(1 - F(v))^{1-\gamma} g_2(v)$, then $p_{1,F}$ and $p_{2,F}$ are density-mass functions. We need the following assumptions:

(C5) $E_{p_F}(\log(p_F/p_{F'}))^\alpha < \infty$ for some $\alpha > 2$, for all F' with $D(p_F \| p_{F'}) < \delta$ for all sufficiently small δ .

(C6) $E_{p_{k,F}}(\log(p_{k,F}/p_{k,F'}))^\alpha < \infty$ for some $\alpha > 2$, for all F' with $D(p_{k,F} \| p_{k,F'}) < \delta$ for all sufficiently small δ , for $k = 1, 2$.

Theorem 3 (i) Let \hat{F}_n be as in Theorem 2. Assume (C1) and (C5). Then,

$$\int_0^\infty |\hat{F}_n(t) - F(t)| [g_1(t) + g_2(t)] dt = O_P(n^{-1/3}), \quad \text{and}$$

$$\sup_t |\hat{F}_n(t) - F(t)| = O_P(n^{-1/3}), \quad a.e. (G_1 + G_2), \quad \text{when } F(\cdot) \text{ is continuous.}$$

(ii) Let \hat{F}_n as in Theorem 1. Assume (C1)-(C2), and (C6). Then

$$\int_0^\infty |\hat{F}_n(t) - F(t)| [g_1(t) + g_2(t)] dt = O_P(n^{-1/3}), \quad \text{and}$$

$$\sup_t |\hat{F}_n(t) - F(t)| = O_P(n^{-1/3}), \quad a.e. (G_1 + G_2), \quad \text{when } F(\cdot) \text{ is continuous.}$$

In both (i) and (ii) the rate of $n^{1/3}$ is optimal.

Note that F in model (1) is not rate \sqrt{n} estimable, like many smooth functionals of it. There may still be certain smooth functionals of it which are rate \sqrt{n} estimable. Geskus and Groeneboom (1999) showed this kind of result for model (0). Groeneboom and Wellner (1992, Chapter 5.4) showed such a result for the mean of F in case I interval censoring model. Huang and Wellner (1995) showed such a result for smooth linear functionals of the form $\nu(F) = \int H dF$, with fixed H , for type I interval censoring model, and that the plug-in estimator is efficient. In fact, one of their purposes is to help treating estimation of smooth functionals in the case II interval censoring model. Yuan, Xu, and Zheng (2012) showed the same functional is rate \sqrt{n} estimable for case II interval censoring model,

without knowing the result of Huang and Wellner. Note that not all linear functionals of F are rate \sqrt{n} estimable. For example, for fixed t_0 , $\nu(F) = F(t_0)$ is not rate \sqrt{n} estimable, but the moments of F are (see for example, Theorem 2 in Yuan, Xu, & Zheng, 2012). Under some conditions, van der Laan (1993) obtained the following relationship between smooth functional, canonical gradient and empirical process: $\nu(\hat{\theta}_n) - \nu(\theta) = (P_n - P_\theta)(\tilde{l}_\nu(\cdot; \hat{\theta}_n))$, where θ is a general parameter (Euclidean or Banach), $\hat{\theta}_n$ an estimator of it, and \tilde{l}_ν the canonical gradient (also called efficient influence function) of θ . This relationship allows one to investigate a class of rate \sqrt{n} estimable smooth functionals. Below, we give a similar result for model (1) as in Huang and Wellner (1995) for model (0). For fixed $h(\cdot)$ and $M > 0$, let the smooth (and linear) functional of F be

$$\nu(F) = \int_0^M (1 - F(t))h(t)dt = \int_0^M H(t)dF(t), \quad H(t) = \int_0^t h(s)ds.$$

Below we give the asymptotic distribution of the plug-in estimator $\hat{\nu}_n = \nu(\hat{F}_n)$, which has nothing to do with the asymptotic distribution of $n^{1/3}(\hat{F}_n - F)$, it only needs $\|\hat{F}_n - F\|_{L_2(G_1+G_2)}^2 = o_P(n^{-1/2})$. The following conditions are from Huang and Wellner (1995).

(C7) The support of F is a bounded interval $[0, M]$, both G and F are dominated by the Lebesgue measure.

(C8) F , g_1 , g_2 and h satisfy

$$\sigma^2 := 2 \int_0^M \frac{F(t)(1 - F(t))}{g_1(t) + g_2(t)} h^2(t)dt < \infty.$$

(C9) $2(h/(g_1 + g_2)) \circ F^{-1}$ is bounded and is Lipschitz on $[0, 1]$.

Theorem 4 Let \hat{F}_n as in Theorem 1, assume the conditions there and (C7)-(C9). Then

$$\sqrt{n}(\nu(\hat{F}_n) - \nu(F)) \xrightarrow{D} N(0, \sigma^2),$$

and generally $\nu(\hat{F}_n)$ is not efficient for $\nu(F)$.

Since models with Banach valued parameters are often not rate \sqrt{n} estimable, and their weak limits, if exist, are often non-Gaussian, we cannot talk about efficiency of their estimators. However, this may be possible for some submodels. Below we explore such a scenario. For this we first review some basic facts of efficient estimation for Euclidean parameters. For the estimation of Euclidean parameter θ , let $\theta_n = \theta_0 + n^{-1/2}b$ for some $b \in C$, the complex plane. A rate $n^{1/2}$ consistent estimator $T_n = T_n(X_1, \dots, X_n)$ is said to be *regular*, if under $f(\cdot|\theta_n)$, $W_n := \sqrt{n}(T_n - \theta_n) \xrightarrow{D} W$ for some random variable W , and the result does not depend on the sequence $\{\theta_n\}$. Let $Z \oplus V$ denote the summation of two independent random variables Z and V , $I(\theta)$ be the Fisher information for $f(\cdot|\theta)$ at θ . The convolution Theorem (Hájek, 1970) states that for any regular estimator T_n with weak limit W , there is a random variable V such that

$$W = Z \oplus V, \quad Z \sim N(0, I^{-1}(\theta_0)).$$

The Cramer-Rao theorem gives the lower bound of the asymptotic variance of any asymptotically unbiased estimators. The convolution theorem further characterizes the weak limit of an asymptotically optimal estimator: it is a normal random variable with mean zero and variance $I^{-1}(\theta_0)$. An estimator is efficient iff $V = 0$. In many cases, the convergence rate of Euclidean or infinite-dimensional parameters can be different from \sqrt{n} . For example, for distributions with singularity of order α , the optimal convergence rate an estimator of its Euclidean parameter in the model is $r_n = n^{1/(1+\alpha)}$, $-1 < \alpha < 1$ ($\alpha \neq 0$). In this case, the local parameter is defined as $\theta_n = \theta_0 + r_n^{-1}b$, and the local likelihood ratio is often asymptotically non-normal, see Ibragimov and Has'minskii (1981). For Euclidean parameter taking only finite number of possible values, the optimal convergence rate r_n of an estimator is exponential (for example, Hammersley, 1950; Robson, 1958). For infinite-dimensional parameters, the convergence rates of their estimators are often slower than \sqrt{n} and the weak limits are often non-Gaussian, although in some few cases rate \sqrt{n} exist with Gaussian weak limits. When the convergence rate is not \sqrt{n} , the problem is much harder, as the locally asymptotical normal (LAN) property no longer holds with such rates for the full model. However a number of papers have tackled this question, such as in Millar (1985) and LeCam (1994). These authors considered very general parameter spaces, established convolution results for estimators regardless of their convergence rates or forms of their weak limits. But these results are mostly of the existence type, not the specific type. Also, it is unclear whether one of the two components in their convolution representation is optimally achievable. For example, given an infinite-dimensional parameter and/or the corresponding likelihood model, although the optimal convergence rate for estimators of this parameter can be determined in principle (LeCam, 1973; Birgé, 1983),

but it is still unknown if there is an optimal weak limit of its estimators, and what is its specific form if it exists. Pötzelberger, Schachermayer and Strasser (2000) gave examples in which the infinite-dimensional version of the convolution theorem does not hold in general abstract space, but does hold under some regularity conditions. The results are of existence type. Janssen and Ostrovski (2005) gave more detailed account of the optimal weak limit. Their Theorem 2.3 gives a convolution result for arbitrary convergence rate r_n , for linear functions of infinite-dimensional parameter and their estimate, with the assumption that the two involved estimators are asymptotically joint Gaussian. They gave the optimal weak limit as the minimal variance random element defined in their condition (a), but how to find this random element is still not clear. Also, the joint asymptotic Gaussian assumption can only be satisfied for a few parameters in the infinite-dimensional spaces, and in these cases often $r_n = \sqrt{n}$. Their Theorem 3.1 and 4.1 established convolution results for infinite-dimensional parameters in abstract spaces, but again the results are of existence type.

A non-negative function $l(\cdot)$ is said bowl shaped if (i) $l(x) = l(-x)$. (ii) $\{x: l(x) \leq c\}$ is convex for every $c \geq 0$. For rate \sqrt{n} estimable parameter θ , the asymptotic minimax theorem (Hájek, 1972) states that for arbitrary (not necessarily regular) estimator T_n of θ , and any bowl-shaped function $l(\cdot)$,

$$\lim_{b \rightarrow \infty} \liminf_{n \rightarrow \infty} \sup_{\theta: \sqrt{n}|\theta - \theta_0| \leq b} E_{\theta} l(\sqrt{n}(T_n - \theta)) \geq El(Z),$$

with Z as given before. Also, for non- \sqrt{n} consistent estimators, such results are unclear.

Let $\theta_n = \theta_0 + r_n^{-1}b$ for some $b \in C$, the complex plane, $r_n \rightarrow \infty$. A rate r_n consistent estimator $T_n = T_n(X_1, \dots, X_n)$ is said to be *regular*, if under $f(\cdot|\theta_n)$, $W_n := r_n(T_n - \theta_n) \xrightarrow{D} W$ for some random variable W , and the result does not depend on the sequence $\{\theta_n\}$. It is possible that on some submodels of the original one, both the convolution and asymptotic minimax result can hold with rate $r_n \neq \sqrt{n}$, although they are not on the original model. Motivated from the exercises of Chapter 2 in Groeneboom and Wellner (1992), below we give such results for the type II censoring model (1), although it is not clear whether such optimal weak limit Z is achievable by some estimator(s). Consider the following parametric submodels of (1): fix $t > 0$ with $f(t) > 0$, and a_n with $a_n \rightarrow \infty$, let $h_n(s) = f(t)[1(t - ca_n^{-1} \leq s \leq t) - 1(t < s \leq t + ca_n^{-1})]$, with $0 < c < \infty$, $H_n(t) = \int_0^t h_n(s)ds$, and for $|\theta| < 1$, let

$$F_n(t|\theta) = F(t) + \theta H_n(t).$$

$F_n(t|\theta)$ has a density (derivative) $f(t|\theta) = f(t) + \theta h_n(t) = (1 + \theta)f(t) > 0$ for all n , so $F_n(\cdot|\theta)$ is increasing, $F_n(0|\theta) = 0$. Since $H_n(\infty) = f(t)(\int_{t-a_n^{-1}}^t ds - \int_t^{t+a_n^{-1}} ds) = 0$, $F_n(\infty|\theta) = 1$, i.e., $F_n(\cdot|\theta)$ is a proper distribution function, and we parametrized $F(\cdot)$ by a Euclidean parameter θ . Let $Q(t) = \frac{1}{3} \left(\frac{2f^2(t)g_1(t)}{F(t)} + \frac{2f^2(t)g_2(t)}{1-F(t)} + f(t)g(t,t) \right)$.

Theorem 5 Assume $a_n = o(n^{1/3})$, $r_n = n^{1/2}c^{3/2}a_n^{-3/2}Q^{1/2}(t)$. Then the local log-likelihood of $F_n(t, \theta)$ satisfies the LAN condition with rate r_n , and

(i) For any rate- r_n regular estimator T_n of $F(t)$, with $r_n(T_n - F(t)) \xrightarrow{D} W$, we have, for some random variable V ,

$$W = Z \oplus V, \quad Z \sim N(0, 1).$$

(ii) For any bowl-shaped function $l(\cdot)$, for any estimator T_n of $F(t)$, and $b > 0$,

$$\liminf_n \sup_{\theta: r_n|\theta| \leq b} E_{\theta} l(r_n(T_n - F_n(t, \theta))) \geq \frac{1}{2} E[l(Z)1(|Z| \leq b/2)].$$

(iii) Especially, take $c^{3/2} = Q^{-7/6}(t)\delta$ for $\delta > 0$, $b = Q^{1/2-\alpha}(t)\delta^2$, $r_n = n^{1/3}Q^{1/2-\alpha}(t)\delta$, and $l(\cdot) = |\cdot|$, we have

$$\sup_{\delta > 0} \liminf_n \sup_{\theta: n^{1/3}|\theta| \leq \delta} E_{\theta} \left| n^{1/3}(T_n - F_n(t|\theta)) \right| \geq \frac{0.4344491}{\sqrt{2\pi}} Q^{1/3}(t).$$

3. Illustration for Applications

Interval censoring model case II is a generalization of that of case I. A common example of interval-censored survival data occurs in medical or health studies. In clinical trials, an individual due for the scheduled observations may miss some observations and may return with a changed status, thus contributing an interval-censored time of the occurrence of the change. As another example, in the acquired immune deficiency syndrome (AIDS) studies,

if a subject is HIV positive at the beginning of the study, the subject's HIV infection time is usually determined by a retrospective study of the subject's history. This an interval censoring case II, with censoring variables as the first HIV positive test and the last HIV negative test. More practical background can be found in Sun (2012).

Although the asymptotics of $\hat{F}_n(\cdot)$ is non standard, with a rate of $n^{1/3}$ instead of the common $n^{1/2}$, and with a non-Gaussian weak limit, these making \hat{F}_n harder to use, but the application of $\hat{F}_n(\cdot)$ is not hindered by this phenomenon. Like most standard estimators, it can be used to predict the probability of events related in this model, or test hypothesis about a given null distribution F . When the corresponding densities f and g_1 and g_2 are given, then Theorem 1 can be used to test F , construct confidence interval for F and error bounds for \hat{F}_n . The key is to evaluate the distribution of $H := \arg \min_h \{\mathbb{B}(h) + h^2\}$, known as the Chernoff distribution, its density function is given in Corollary 3.3 of Groeneboom (1989). For the NPMLE $\hat{F}_n(\cdot)$, there are existing softwares to compute it, for example the R-package "Decon" developed by Wang and Wang (2011). For given α , let $H^{-1}(1 - \alpha)$ be the $(1 - \alpha)$ -th upper quantile of the Chernoff distribution, by Theorem 1, the $(1 - \alpha)$ -th confidence interval for $F(t)$ is

$$[\hat{F}_n(t) \pm n^{-1/3} \hat{A}(t) H^{-1}(1 - \alpha)],$$

where $\hat{A}(t) = \left(\frac{4\hat{F}_n(t)f_n(t)}{g_1(t)+g_2(t)} \right)^{1/3}$ is an estimate of $A(t) := \left(\frac{4F(t)f(t)}{g_1(t)+g_2(t)} \right)^{1/3}$, here an estimate $f_n(t)$ of $f(t)$ is also needed and may be obtained by differencing of $\hat{F}_n(t)$; $g_1(t)$ and $g_2(t)$ are either known, or can be easily estimated as the data (U_i, V_i) ($i = 1, \dots, n$) are directly observed. Similarly for testing $H_0 : F(t) = F_0(t)$, a test statistic is given by

$$T_n = n^{1/3}(\hat{F}_n(t) - F_0(t)).$$

Under H_0 , T_n is asymptotically distributed with weak limit given in Theorem 1. If $|T_n| > \hat{A}(t)H^{-1}(1 - \alpha/2)$, H_0 is rejected with significance level α ; otherwise H_0 is accepted.

In fact, since the random variable H does not depend on t , and $A(t)$ is deterministic, under some more conditions, we can have that $n^{1/3}(\hat{F}_n(\cdot) - F(\cdot))$ being a tight stochastic sequence on the support space, and the resulting weak convergence can be strengthened to that on the corresponding metric space equipped with the supreme norm. Thus any linear functional of $n^{1/3}(\hat{F}_n(\cdot) - F(\cdot))$ converges weakly, and the confidence interval described above can be strengthened to confidence band, the point-wise test of H_0 can be generalized to test of $H'_0 : F(\cdot) = F_0(\cdot)$ using these linear functionals.

The interested variable X and the observed variables U and V can all be generated from the Gamma densities, with different parameters for them to satisfy the underlying conditions,

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

The condition $V > U$ (a.s.) can be satisfied by specifying $V = U + Z$ for some non-negative random variable Z , for example from another independent Gamma distribution. In some literatures, the censoring variable are just generated from uniform distributions. For the Gamma distribution, condition (C1) is automatic, (C3)-(C4) are commonly used conditions for the proof of consistency of NPMLE, as in van de Geer (1993). (C5)-(C6) are easily satisfied for the Gamma families, condition (C2) can be satisfied with suitable chosen parameters in two different Gamma distributions for U and V . Conditions (C7)-(C9) are satisfied if we constraint the Gamma distributions on a finite interval.

References

- Bahadur, R. R. (1967). Rates of convergence of estimates and test statistics. *Annals of Mathematical Statistics*, 38, 303-324. <http://dx.doi.org/10.1214/aoms/1177698949>
- Bebchuk, J. D., & Betensky, R. A. (2000). Multiple imputation for simple estimation of the hazard function based on interval censored data. *Statistics in Medicine*, 19, 405-419. [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(20000215\)19:3<405::AID-SIM325>3.0.CO;2-2](http://dx.doi.org/10.1002/(SICI)1097-0258(20000215)19:3<405::AID-SIM325>3.0.CO;2-2)
- Bickel, P. J., Klaassen, C. A., Ritov, Y., & Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, Maryland: Johns Hopkins University Press.
- Birgé, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 65, 181-237. <http://dx.doi.org/10.1007/BF00532480>
- Birgé, L., & Massart, P. (1993). Rates of convergence for minimum contrast estimators. *Probability Theory and Related Fields*, 97, 113-150. <http://dx.doi.org/10.1007/BF01199316>

- Chen, Y. Q., & Jewell, N. P. (2001). On a general class of semiparametric hazards regression models. *Biometrika*, 88, 687-702. <http://dx.doi.org/10.1093/biomet/88.3.687>
- Chernoff, H. (1964). Estimation of the mode. *Annals of the Institute of Statistical Mathematics*, 16, 31-41. <http://dx.doi.org/10.1007/BF02868560>
- Csörgő, M., & Révész, P. (1981). *Strong Approximations in Probability and Statistics*. New York, San Francisco, London: Academic Press.
- Droste, W., & Wefelmeyer, W. (1984). On Hájek's convolution theorem. *Statistical Decisions*, 2, 131-144.
- Fang, H., & Sun, J. (2001). Consistency of nonparametric maximum likelihood estimation of a distribution function based on doubly interval-censored failure time data. *Statistics and Probability Letters*, 55, 311-318. [http://dx.doi.org/10.1016/S0167-7152\(01\)00160-2](http://dx.doi.org/10.1016/S0167-7152(01)00160-2)
- Geskus, R., & Groeneboom, P. (1999). Asymptotically optimal estimation of smooth functionals for interval censoring, case 2. *Annals of Statistics*, 27, 627-674.
- Grenander, U. (1956). On the theory of mortality measurement. Part II. *Skand. Akt.*, 39, 125-153.
- Groeneboom, P. (1989). Brownian motion with a parabolic drift and airy functions. *Probability Theory and Related Fields*, 81, 79-109. <http://dx.doi.org/10.1007/BF00343738>
- Groeneboom, P. (1996). Lectures on inverse problems. *Lectures on Probability Theory. Ecole d'Etéde Probabilités de Saint-Flour XXIV*. Berlin: Springer. <http://dx.doi.org/10.1007/BFb0095675>
- Groeneboom, P., & Wellner, J. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Basel: Birkhäuser Verlag. <http://dx.doi.org/10.1007/978-3-0348-8621-5>
- Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. und Verw. Gebiete*, 14, 323-330. <http://dx.doi.org/10.1007/BF00533669>
- Hájek, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, 1, 175-194. Berkeley: University of California Press.
- Hammersley, J. M. (1950) On estimating restricted parameters. *Journal of the Royal Statistical Society, Series B*, 12, 192-240.
- Huang, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Annals of Statistics*, 24, 540-568. <http://dx.doi.org/10.1214/aos/1032894452>
- Huang, J., & Wellner, J. (1995). Asymptotic normality of the NPMLE of linear functionals for interval censored data, case I. *Statistica Neerlandica*, 49, 153-163. <http://dx.doi.org/10.1111/j.1467-9574.1995.tb01462.x>
- Ibragimov, I. A., & Has'minskii, R. Z. (1981). *Statistical Estimation Asymptotic Theory*. New York: Springer-Verlag. <http://dx.doi.org/10.1007/978-1-4899-0027-2>
- Janssen, A., & Ostrovski, V. (2005). The convolution theorem of Hájek and LeCam-revisited. *Statistics and Decision*, 1, submitted.
- Jeganathan, P. (1982). On the asymptotic theory of estimation when the limit of the log-likelihood ratio is mixed normal. *Sankhā*, 44, 173-212.
- Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *Journal of Computational and Graphical Statistics*, 7, 310-321.
- LeCam, L. (1973). Convergence of estimates under dimensionality restrictions. *Annals of Statistics*, 1, 38-53. <http://dx.doi.org/10.1214/aos/1193342380>
- LeCam, L. (1994). An infinite dimensional convolution theorem. Proc. 5. Purdue Int. Symp. Stat. Decis. Theory and Related. Topics, Purdue Univ./USA, Springer, Vol. 5, 401-411.
- Millar, P. W. (1985). Non-parametric applications of an infinite-dimensional convolution theorem. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 68, 545-556.
- Pan, W., & Chappell, R. (2002). Estimation in the Cox proportional hazards model with left truncated and interval censored data. *Biometrics*, 58, 64-70.
- Pötzelberger, K., Schachermayer, W., & Strasser, H. (2000). On the convolution theorem for infinite-dimensional

parameter spaces, manuscript.

- Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer.
- Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhya, Ser. A.*, 31, 23-36.
- Robertson, T., Wright, F. T., & Dykstra. (1988). *Order Restricted Statistical Inference*. Chichester, New York, Brisbane, Toronto, Singapore: John Wiley & Sons.
- Robson, D. S. (1958). Admissible and minimax integer-valued estimators of an integer-valued parameters. *Annals of Mathematical Statistics*, 29, 801-812. <http://dx.doi.org/10.1214/aoms/1177706537>
- Sun, J. (2012). Interval censoring. Manuscript.
- Van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimates. *Annals of Statistics*, 21, 14-44. <http://dx.doi.org/10.1214/aos/1176349013>
- Van der Laan, M. J. (1993). General identity for linear parameters in convex models with application to efficiency of the (NP)MLE, Preprint No. 765, Department of Mathematics, University of Utrecht, The Netherlands.
- Van der Vaart, A. W., & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer-Verlag. <http://dx.doi.org/10.1007/978-1-4757-2545-2>
- Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics*, 18, 309-348.
- Wang, X. F., & Wang, B. (2011). Deconvolution estimation in measurement error models: The R package decon. *Journal of Statistical Software*, 39(10), 1-24.
- Wellner, J. (1995). Interval censoring case 2: Alternative hypotheses. In H. L. Koul, & J. V. Deshpande (Eds.), *Analysis of Censored Data* (pp. 271-291). IMS, Hayward, CA. Retrieved from <http://www.stat.washington.edu:80/jaw/jaw.research.available.html>
- Wellner, J. (2005). Maximum likelihood: Counterexamples, examples, and open problems. Talk at BeNeLuxFra Mathematics Meeting, 21 May, 2005.
- Wong, W. H., & Severini, T. A. (1991). On maximum likelihood estimation in infinite dimensional parameter spaces *Annals of Statistics*, 19, 603-632. <http://dx.doi.org/10.1214/aos/1176348113>
- Yu, Q. Q., Li, L. X., & Wong, G. Y. C. (2000). On consistency of the self-consistent estimator of survival functions with interval censored data. *Scandinavian Journal of Statistics*, 27, 35-44. <http://dx.doi.org/10.1111/1467-9469.00176>
- Yuan, A., Xu, J., & Zheng, G. (2012). Root- n estimability of some missing data models. *Journal of Multivariate Analysis*, 106, 147-166. <http://dx.doi.org/10.1016/j.jmva.2011.11.007>

Appendix

Proof of Theorem 1. Let

$$\tilde{l}_n(F) = \frac{1}{n} \sum_{i=1}^n \left(\delta_i \log F(u_i) + (1 - \delta_i) \log(1 - F(u_i)) + \gamma_i \log F(v_i) + (1 - \gamma_i) \log(1 - F(v_i)) \right)$$

and

$$R_n = \frac{1}{n} \sum_{i=1}^n \left((\gamma_i - \delta_i) \log(F(v_i) - F(u_i)) - (1 - \delta_i) \log(1 - F(u_i)) - \gamma_i \log F(v_i) \right).$$

Then

$$l_n(F) = \tilde{l}_n(F) + R_n,$$

and by (C2), $R_n \stackrel{a.s.}{=} o(n^{-\alpha})$, $\forall \alpha < 1/2$. So we only need to maximize $\tilde{l}_n(F)$ to get an approximate NPMLE \hat{F}_n of F . It is known that approximate NPMLE enjoys many properties of the exact NPMLE (Wong & Severini, 1991), often the former is relatively easy to compute, while the latter is often hard. To use the method of isotonic regression, we re-written $\tilde{l}_n(F)$ in terms of the Δ_i 's and x_i 's as

$$\tilde{l}_n(F) = \frac{1}{n} \sum_{i=1}^{2n} \left(\Delta_i \log F(x_i) + (1 - \Delta_i) \log(1 - F(x_i)) \right).$$

Thus $\{y_i := F(x_i)\}$ is an increasing sequence, and we are to maximizing, over y_i 's

$$\tilde{l}_n(F) = \frac{1}{n} \sum_{i=1}^{2n} \left(\Delta_i \log y_i + (1 - \Delta_i) \log(1 - y_i) \right)$$

subject to $0 \leq y_1 \leq \dots \leq y_{2n} \leq 1$. Let $\Phi(y) = y \log y + (1 - y) \log(1 - y)$, it is a strictly convex function with derivative $\phi(y)$. By Theorem 1.5.1 of Robertson, Wright, and Dykstra (1988), especially expression (1.5.4) of p. 31, the maximization of

$$\sum_{i=1}^{2n} \left(\Delta_i \log y_i + (1 - \Delta_i) \log(1 - y_i) \right) = \sum_{i=1}^{2n} \left(\Phi(y_i) + (\Delta_i - y_i) \phi(y_i) \right)$$

is the same as minimization of $\sum_{i=1}^{2n} (\Delta_i - y_i)^2$ subject to the same constraint, and the unique solution $\hat{y}_i = \hat{F}_n(x_i)$'s is the slope (or left derivative) of the greatest convex minorant of the cumulative-sum diagram of the points $(2n)^{-1}(i, \sum_{j \leq i} \Delta_j)$. This cumulative-sum diagram is given by the function $c(0) = 0$ and $c(s) = (2n)^{-1} \sum_{j \leq i} \Delta_j$ if $s \in (2n)^{-1}(i - 1, i]$. It follows that $\hat{y}_i \leq a$ if and only if $i/(2n) \leq \arg \min_s \{c(s) - as\}$ (van der Vaart & Wellner, 1996, p. 298). Let

$$V_n(t) = \frac{1}{2n} \sum_{i=1}^{2n} \Delta_i 1(x_i \leq t), \quad G_n(t) = \frac{1}{2n} \sum_{i=1}^{2n} 1(x_i \leq t),$$

$$V_{1,n}(t) = \frac{1}{n} \sum_{i=1}^n \delta_i 1(u_i \leq t), \quad G_{1,n}(t) = \frac{1}{n} \sum_{i=1}^n 1(u_i \leq t),$$

$$V_{2,n}(t) = \frac{1}{n} \sum_{i=1}^n \gamma_i 1(v_i \leq t), \quad G_{2,n}(t) = \frac{1}{n} \sum_{i=1}^n 1(v_i \leq t).$$

Then

$$V_n(t) = \frac{1}{2} [V_{1,n}(t) + V_{2,n}(t)], \quad G_n(t) = \frac{1}{2} [G_{1,n}(t) + G_{2,n}(t)]$$

and

$$\hat{F}_n(x_i) \leq a \quad \text{iff} \quad \arg \min_s \{V_n(s) - aG_n(s)\} \geq x_i \quad (i = 1, \dots, 2n).$$

For fixed t , $\hat{F}_n(t) = \hat{F}_n(x_i)$ for some x_i that is just at left of t . By (C1), $t - x_i$ is asymptotically negligible. Thus,

$$\hat{F}_n(t) \leq a \quad \text{iff} \quad \arg \min_s \{V_n(s) - aG_n(s)\} - t \geq 0. \quad (\text{A.1})$$

We are to find the limit distribution of $n^{-1/3}(\hat{F}_n(t) - F(t))$. (Why the convergence rate of $n^{1/3}$? We will see this latter after we computed the asymptotic variance of $m_{n,h}(Z, U, V)$, to be defined latter). Now, let's evaluate the limit of probability of the event $\{n^{-1/3}(\hat{F}_n(t) - F(t)) \leq x\} = \{\hat{F}_n(t) \leq F(t) + xn^{-1/3}\}$ for each fixed x . Thus we need to evaluate the limit of the arg \max_s event in (A.1) for $a = F(t) + xn^{-1/3}$.

Make the change of variable $s = t + n^{-1/3}h$, and let $\hat{s}_n = \arg \min_s \{V_n(s) - aG_n(s)\}$ and $\hat{h}_n = \arg \min_h \{V_n(t + n^{-1/3}h) - aG_n(t + n^{-1/3}h)\}$. Then $\hat{s}_n = t + n^{-1/3}\hat{h}_n$, and so

$$\left\{ \arg \min_s [V_n(s) - aG_n(s)] - t \geq 0 \right\} = \left\{ \arg \min_h [V_n(t + n^{-1/3}h) - aG_n(t + n^{-1/3}h)] \geq 0 \right\}.$$

Thus we are to evaluate the limit of $P(\hat{h}_n \geq 0)$. Let $\mathbb{P}_{1,n}$ be the empirical measure of $(z_1, u_1), \dots, (z_n, u_n)$, $\mathbb{P}_{2,n}$ be that of $(z_1, v_1), \dots, (z_n, v_n)$, $\mathbb{P}_n = (\mathbb{P}_{1,n} + \mathbb{P}_{2,n})/2$, P the theoretical measure corresponding to \mathbb{P}_n , $A = \{(z, u) : z \leq u\}$, and Ph means $E_P(h)$ for any measure P and function h . We have

$$\begin{aligned} & \arg \min_h [V_n(t + n^{-1/3}h) - (F(t) + xn^{-1/3})G_n(t + n^{-1/3}h)] \\ &= \arg \min_h \left[\mathbb{P}_n 1_A 1_{R \times [0, t + hn^{-1/3}]} - (F(t) + xn^{-1/3}) \mathbb{P}_n 1_{R \times [0, t + hn^{-1/3}]} \right] \\ &= \arg \min_h \left[\mathbb{P}_n 1_A (1_{R \times [0, t + hn^{-1/3}]} - 1_{R \times [0, t]}) - (F(t) + xn^{-1/3}) \mathbb{P}_n (1_{R \times [0, t + hn^{-1/3}]} - 1_{R \times [0, t]}) \right. \\ & \quad \left. + \mathbb{P}_n 1_A 1_{R \times [0, t]} - (F(t) + xn^{-1/3}) \mathbb{P}_n 1_{R \times [0, t]} \right] \\ &= \arg \min_h \left[\mathbb{P}_n 1_A (1_{R \times [0, t + hn^{-1/3}]} - 1_{R \times [0, t]}) - (F(t) + xn^{-1/3}) \mathbb{P}_n (1_{R \times [0, t + hn^{-1/3}]} - 1_{R \times [0, t]}) \right] \\ &= \arg \min_h \left[(\mathbb{P}_n - P) 1_A (1_{R \times [0, t + hn^{-1/3}]} - 1_{R \times [0, t]}) + [P 1_A (1_{R \times [0, t + hn^{-1/3}]} - 1_{R \times [0, t]}) \right. \\ & \quad \left. - F(t) \mathbb{P}_n (1_{R \times [0, t + hn^{-1/3}]} - 1_{R \times [0, t]}) - xn^{-1/3} \mathbb{P}_n (1_{R \times [0, t + hn^{-1/3}]} - 1_{R \times [0, t]}) \right] \\ &= \arg \min_h \left[n^{2/3} (\mathbb{P}_n - P) 1_A (1_{R \times [0, t + hn^{-1/3}]} - 1_{R \times [0, t]}) + n^{2/3} [P 1_A (1_{R \times [0, t + hn^{-1/3}]} - 1_{R \times [0, t]}) \right. \\ & \quad \left. - F(t) \mathbb{P}_n (1_{R \times [0, t + hn^{-1/3}]} - 1_{R \times [0, t]}) - xn^{1/3} \mathbb{P}_n (1_{R \times [0, t + hn^{-1/3}]} - 1_{R \times [0, t]}) \right] \\ &:= \arg \min_h [B_{1,n}(h) + B_{2,n}(h) - B_{3,n}(h)] := \arg \min_h B_n(h). \end{aligned}$$

Below, without confusion, for $h < 0$ by the notation $[t, t + hn^{-1/3}]$ we actually mean $[t + hn^{-1/3}, t]$. We have

$$\begin{aligned} B_{3,n}(h) &\sim xn^{1/3} \text{sign}(h) \frac{1}{2} \left(P(U \in [t, t + hn^{-1/3}]) + P(V \in [t, t + hn^{-1/3}]) \right) \\ &= xn^{1/3} \frac{1}{2} \left(\int_t^{t+hn^{-1/3}} g_1(u) du + \int_t^{t+hn^{-1/3}} g_2(v) dv \right) \sim \frac{x}{2} [g_1(t) + g_2(t)]h; \\ B_{2,n}(h) &\sim n^{2/3} \frac{1}{2} \text{sign}(h) \left(E[1(Z \leq U)1(U \in [t, t + hn^{-1/3}])] - F(t)E[1(U \in [t, t + hn^{-1/3}])] \right. \\ & \quad \left. + E[1(Z \leq V)1(V \in [t, t + hn^{-1/3}])] - F(t)E[1(V \in [t, t + hn^{-1/3}])] \right) \\ &= n^{2/3} \frac{1}{2} \text{sign}(h) \left(E[(1(Z \leq U) - F(t))1(U \in [t, t + hn^{-1/3}])] \right. \\ & \quad \left. + E[(1(Z \leq V) - F(t))1(V \in [t, t + hn^{-1/3}])] \right) \\ &= n^{2/3} \frac{1}{2} \int_t^{t+hn^{-1/3}} (F(u) - F(t)) [g_1(u) + g_2(u)] du \\ &\sim n^{2/3} \frac{1}{2} f(t) [g_1(t) + g_2(t)] \int_t^{t+hn^{-1/3}} (u - t) du \\ &= \frac{1}{4} f(t) [g_1(t) + g_2(t)] h^2. \end{aligned}$$

For fixed t , let $m_{n,h}(z, u, v) = n^{1/6}(1/2)\text{sign}(h)\left(1(z \leq u)1(u \in [t, t + hn^{-1/3}]) + 1(z \leq v)1(v \in [t, t + hn^{-1/3}])\right)$, we have

$$B_{1,n}(h) = n^{-1/2} \sum_{i=1}^n (m_{n,h}(z_i, u_i, v_i) - Em_{n,h}(Z, U, V)).$$

Let $\mathcal{M}_n = \{m_{n,h}: h \in [-K, K]\}$. Then $M_n(z, u, v) = n^{1/6}(1/2)\left(1(u \in [t \pm Kn^{-1/3}]) + 1(v \in [t \pm Kn^{-1/3}])\right)$ is an envelope function of \mathcal{M}_n . We have

$$EM_n^2(Z, U, V) = O(1), \\ E[M_n^2 1(M_n > \eta \sqrt{n})] \rightarrow 0, \quad \forall \eta > 0,$$

and

$$\begin{aligned} E(m_{n,s}(Z, U, V) - m_{n,h}(Z, U, V))^2 &= \frac{1}{4}n^{1/3} E\left(1(Z \leq U)1(t + (s \wedge h)n^{-1/3} < U \leq t + (s \vee h)n^{-1/3}) \right. \\ &\quad \left. + 1(Z \leq V)1(t + (s \wedge h)n^{-1/3} < V \leq t + (s \vee h)n^{-1/3})\right)^2 \\ &\leq \frac{1}{2}n^{1/3} E\left(1(Z \leq U)1(t + (s \wedge h)n^{-1/3} < U \leq t + (s \vee h)n^{-1/3}) \right. \\ &\quad \left. + 1(Z \leq V)1(t + (s \wedge h)n^{-1/3} < V \leq t + (s \vee h)n^{-1/3})\right) \\ &= \frac{1}{2}n^{1/3} \int_{t+(s \wedge h)n^{-1/3}}^{t+(s \vee h)n^{-1/3}} F(y)[g_1(y) + g_2(y)]dy \sim 2^{-1}F(t)[g_1(t) + g_2(t)]|s - h|, \end{aligned}$$

thus

$$\sup_{|s-h| < \delta_n} E(m_{n,s}(Z, U) - m_{n,h}(Z, U))^2 \rightarrow 0, \quad \forall \delta_n \rightarrow 0.$$

Let $l^\infty(-K, K) = \{s(\cdot): \|s\| := \sup_{t \in [-K, K]} |s(t)| < \infty\}$ be the space of all bounded real functions on $[-K, K]$, it is indexed by the totally bounded space (for definition c.f., VW, p. 17) $([-K, K], \rho)$ with $\rho(\cdot, \cdot) = |\cdot - \cdot|$. Let $N_{[]}(\epsilon \|M_n\|_{P,2}, \mathcal{M}_n, L_2(P))$ be the bracketing number (c.f., VW, p. 83) of \mathcal{M}_n , of size $\epsilon \|M_n\|_{P,2}$, with respect to the norm $L_2(P)$. Note $\|M_n\|_{P,2} = [EM_n^2(Z, U)]^{1/2} = O(1)$, so $\epsilon \|M_n\|_{P,2} = O(\epsilon)$. It is obvious that the minimum number of balls with radius $\epsilon \|M_n\|_{P,2}$ needed to cover \mathcal{M}_n under the norm $L_2(P)$ is $O(K/\epsilon)$. Thus, $N_{[]}(\epsilon \|M_n\|_{P,2}, \mathcal{M}_n, L_2(P)) \leq C/\epsilon$ for some $0 < C < \infty$ and so, for all $\delta_n \searrow 0$,

$$\int_0^{\delta_n} \sqrt{\log N_{[]}(\epsilon \|M_n\|_{P,2}, \mathcal{M}_n, L_2(P))} d\epsilon \leq \int_0^{\delta_n} \sqrt{\log(C/\epsilon)} d\epsilon = C \int_{\log(C/\delta_n)}^{\infty} \sqrt{y} e^{-y} dy \rightarrow 0.$$

Thus, by Theorem 2.11.23 (VW, p. 221), $\{B_{1,n}(h): h \in [-K, K]\}$ is asymptotically tight, and converges in distribution to the Gaussian process G on $l^\infty([-K, K])$, with mean zero and covariance function $R(s, h)$ on $[-K, K]^2$ given by

$$R(s, h) = \lim_n \left(E[m_{n,s}(Z, U, V)m_{n,h}(Z, U, V)] - E[m_{n,s}(Z, U, V)]E[m_{n,h}(Z, U, V)] \right).$$

For fixed h ,

$$\begin{aligned} E[m_{n,h}(Z, U, V)] &= n^{1/6} \frac{1}{2} \text{sign}(h) \left(E[1(Z \leq U)1(U \in [t, t + hn^{-1/3}])] + E[1(Z \leq V)1(V \in [t, t + hn^{-1/3}])] \right) \\ &= n^{1/6} \frac{1}{2} \text{sign}(h) \int_t^{t+hn^{-1/3}} F(u)[g_1(u) + g_2(u)] du \sim n^{-1/6} \frac{1}{2} F(t)[g_1(t) + g_2(t)]h \rightarrow 0, \end{aligned}$$

and for $s, h > 0$, note $U \leq V$ a.s.,

$$\begin{aligned} E[m_{n,s}(Z, U, V)m_{n,h}(Z, U, V)] &= n^{1/3} \frac{1}{4} E\left(1(Z \leq U)1(t \leq U \leq t + (h \wedge s)n^{-1/3}) \right. \\ &\quad \left. + 1(Z \leq U)1(t \leq U \leq t + hn^{-1/3})1(t \leq V \leq t + sn^{-1/3}) \right. \\ &\quad \left. + 1(Z \leq U)1(t \leq U \leq t + sn^{-1/3})1(t \leq V \leq t + hn^{-1/3}) \right. \\ &\quad \left. + 1(Z \leq V)1(t \leq V \leq t + (h \wedge s)n^{-1/3})\right). \end{aligned}$$

Note

$$\begin{aligned} & n^{1/3} E\left(1(Z \leq U)1(t \leq U \leq t + (h \wedge s)n^{-1/3}) + 1(Z \leq V)1(t \leq V \leq t + (h \wedge s)n^{-1/3})\right) \\ &= n^{1/3} \int_t^{t+(h \wedge s)n^{-1/3}} F(u)[g_1(u) + g_2(u)]du \sim F(t)[g_1(t) + g_2(t)](h \wedge s), \end{aligned}$$

and

$$\begin{aligned} & n^{1/3} E\left(1(Z \leq U)1(t \leq U \leq t + hn^{-1/3})1(t \leq V \leq t + sn^{-1/3})\right) \\ &= n^{1/3} \int_t^{t+sn^{-1/3}} \int_t^{v \wedge (t+hn^{-1/3})} F(u)g(u, v)dudv \\ &\sim n^{1/3} F(t) \int_t^{t+sn^{-1/3}} g(t, v)[(v \wedge (t + hn^{-1/3}) - t)]dv \\ &= \left(\frac{1}{2}F(t)g(t, t)(s \wedge h)^2 + F(t)g(t, t)h(s - h)1(s > h)\right)n^{-1/3} \rightarrow 0. \end{aligned}$$

Similarly,

$$n^{1/3} E\left(1(Z \leq U)1(t \leq U \leq t + sn^{-1/3})1(t \leq V \leq t + hn^{-1/3})\right) = O(n^{-1/3}) \rightarrow 0.$$

By the same way, for $s, h < 0$,

$$E[m_{n,s}(Z, U, V)m_{n,h}(Z, U, V)] \sim F(t)[g_1(t) + g_2(t)](-h \wedge -s).$$

For s, h with $sh < 0$ (suppose $s < 0 < h$. The case $h < 0 < s$ is similar),

$$\begin{aligned} E[m_{n,s}(Z, U, V)m_{n,h}(Z, U, V)] &= -\frac{1}{4}n^{1/3} E[1(Z \leq U)1(t + sn^{-1/3} \leq U \leq t)1(t \leq V \leq t + hn^{-1/3})] \\ &= -\frac{1}{4}n^{1/3} \int_t^{t+hn^{-1/3}} \int_{t+sn^{-1/3}}^t F(u)g(u, v)dudv \sim \frac{1}{4}F(t)g(t, t)shn^{-1/3} \rightarrow 0. \end{aligned}$$

Thus we have

$$R(s, h) = \frac{1}{4}F(t)[g_1(t) + g_2(t)]r(s, h), \quad r(s, h) = (|h| \wedge |s|)1(sh > 0).$$

It is easy to check that $r(s, h)$ is the covariance function of $\{\mathbb{B}(h); h \in R\}$. Here we can see why the convergence rate of $\hat{F}_n(t)$ is $n^{1/3}$. If a rate $a(n)$ is used with $a(n) \rightarrow \infty$, then in the definition of $m_{n,h}(u, v, z)$ we should replace $n^{-1/3}$ and $n^{2/3}/n^{1/2} = n^{1/6}$ by $a^{-1}(n)$ and $a^2(n)/n^{1/2}$, and the asymptotic variance of $m_{n,h}(U, V, Z)$ would then be $O(a^{-1}(n)[a^2(n)/n^{1/2}]^2) = O(a^3(n)/n)$, which is a non-zero constant only if $a(n)^3 = n$.

Now we have, for all finite $K > 0$, on $I^\infty([-K, K])$,

$$B_n(h) \xrightarrow{D} \frac{1}{2} \sqrt{F(t)[g_1(t) + g_2(t)]} \mathbb{B}(h) + \frac{1}{4}f(t)[g_1(t) + g_2(t)]h^2 - \frac{x}{2}[g_1(t) + g_2(t)]h. \quad (\text{A.2})$$

Let \hat{h} be the minimizer of the right hand side of (A.2) above. To simplify the expression for \hat{h} , the following Lemma will be used.

Lemma Let $\mathbb{B}(h)$ be the two-sided Brownian motion, originating from zero, $a > 0, b > 0$ and c are constants, then

$$\arg \min_h \{a\mathbb{B}(h) + bh^2 + ch\} \stackrel{d}{=} (a/b)^{2/3} \arg \min_r \{\mathbb{B}(r) + r^2\} - \frac{1}{2} \frac{c}{b}.$$

Proof. Let $h = (a/b)^{2/3}r - \frac{1}{2} \frac{c}{b}$, using the fact that $\mathbb{B}(\sigma h - \mu) = \sqrt{\sigma} \mathbb{B}(h) - \mathbb{B}(\mu)$, we have

$$a\mathbb{B}(h) + bh^2 + ch = a(a/b)^{1/3} \mathbb{B}(r) - a\mathbb{B}(c/(2b)) + a(a/b)^{1/3} r^2 - \frac{1}{4} \frac{c^2}{b},$$

and

$$\min_h \{a\mathbb{B}(h) + bh^2 + ch\} = \min_r \{a(a/b)^{1/3} \mathbb{B}(r) - a\mathbb{B}(c/(2b)) + a(a/b)^{1/3} r^2 - \frac{1}{4} \frac{c^2}{b}\}.$$

Let h^* and g^* be the minimizers of the left hand side and right hand side above, then $h^* = (a/b)^{2/3}r^* - \frac{1}{2}\frac{c}{b}$ simply by the given relationship between h and r . Since, as $a(a/b)^{1/3} > 0$,

$$\arg \min_r \{a(a/b)^{1/3}\mathbb{B}(r) - a\mathbb{B}(c/(2b)) + a(a/b)^{1/3}r^2 - \frac{1}{4}\frac{c^2}{b}\} = \arg \min_r \{\mathbb{B}(r) + r^2\} = r^*,$$

the desired conclusion follows. \square

Using this Lemma, we have

$$\begin{aligned} \{\hat{h} \geq 0\} &= \left\{ \left(\frac{4F(t)}{f^2(t)[g_1(t) + g_2(t)]} \right)^{1/3} \arg \min_h \{\mathbb{B}(h) + h^2\} + \frac{x}{f(t)} \geq 0 \right\} \\ &= \left\{ \left(\frac{4F(t)f(t)}{g_1(t) + g_2(t)} \right)^{1/3} [-\arg \min_h \{\mathbb{B}(h) + h^2\}] \leq x \right\} \\ &= \left\{ \left(\frac{4F(t)f(t)}{g_1(t) + g_2(t)} \right)^{1/3} \arg \min_h \{\mathbb{B}(h) + h^2\} \leq x \right\}. \end{aligned}$$

In the last step above, with the notation $\stackrel{D}{=}$ denotes equal in distribution, we used the fact that $-\mathbb{B}(h) \stackrel{D}{=} \mathbb{B}(-h) \stackrel{D}{=} \mathbb{B}(h)$, and hence $\mathbb{B}(h) + h^2$ is symmetric about 0, in distribution, and so $-\arg \min_h \{\mathbb{B}(h) + h^2\} \stackrel{D}{=} \arg \min_h \{\mathbb{B}(h) + h^2\}$. \square

Below we will show that $\hat{h}_n = O_p(1)$, so by (A.2) and the continuity of $\arg \min\{\cdot\}$, we have, for all x ,

$$\begin{aligned} P\left(n^{1/3}(\hat{F}_n(t) - F(t)) \leq x\right) &= P(\hat{h}_n \geq 0) \rightarrow P(\hat{h} \geq 0) \\ &= P\left(\left(\frac{4F(t)f(t)}{g_1(t) + g_2(t)}\right)^{1/3} \arg \min_h \{\mathbb{B}(h) + h^2\} \leq x\right), \end{aligned}$$

this gives the desired result.

Now, we show that $\hat{h}_n = O_p(1)$. Recall $\hat{s}_n = t + n^{-1/3}\hat{h}_n$ or $\hat{h}_n = n^{1/3}(\hat{s}_n - t)$, where $\hat{s}_n = \arg \min_s \{V_n(s) - (F(t) + xn^{-1/3})G_n(s)\}$. So we only need to show $n^{1/3}(\hat{s}_n - t) = O_p(1)$. For this, let $G_1(\cdot)$ and $G_2(\cdot)$ be the margins of $G(\cdot, \cdot)$, and

$$\begin{aligned} M_n(s) &= (F(t) + xn^{-1/3})G_n(s) - V_n(s), \\ M(s) &= \frac{1}{2}F(t)[G_1(s) + G_2(s)] - \frac{1}{2} \int_{-\infty}^s F(y)[g_1(y) + g_2(y)]dy. \end{aligned}$$

Then $\hat{s}_n = \arg \max_s M_n(s)$, and it is easy to check that $\arg \max_s M(s) = t$. Since $\|G_n(s) - \frac{1}{2}(G_1(s) + G_2(s))\|_R \leq \frac{1}{2}[\sup_s |G_{1,n}(s) - G_1(s)| + \sup_s |G_{2,n}(s) - G_2(s)|] \rightarrow 0$ (a.s.), $V(s) := \frac{1}{2} \int_{-\infty}^s F(y)[g_1(y) + g_2(y)]dy = \frac{1}{2}[\int_{-\infty}^s F(y)g_1(y)dy + \int_{-\infty}^s F(y)g_2(y)dy] := \frac{1}{2}[V_1(s) + V_2(s)]$, $V_1(s)$ and $V_2(s)$ are the conditional distribution function $P(U \leq s|Z \leq U)$ and $P(V \leq s|Z \leq V)$, and $V_n(s)$ is the empirical version of $V(s)$, thus $\|V_n - V\|_R \rightarrow 0$ a.s.. Thus for fixed x and t , $\|M_n - M\|_R \rightarrow 0$ a.s., and so by Corollary 3.2.3 in VW (p. 287), we have $\hat{s}_n \rightarrow t$ in probability.

Also, for s in a small neighborhood of t , for some $0 < C < \infty$,

$$\begin{aligned} M(s) - M(t) &= \frac{1}{2} \text{sign}(s - t) \int_t^s (F(t) - F(y))(g_1(y) + g_2(y))dy \\ &\sim \frac{1}{2} \text{sign}(t - s) \int_t^s f(t)(y - t)(g_1(y) + g_2(y))dy \leq -C(t - s)^2. \\ M_n(s) - M_n(t) &= \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \text{sign}(s - t) [(F(t) - \delta_i)1(t \leq u_i \leq s) + (F(t) - \gamma_i)1(t \leq v_i \leq s)] \\ &\quad + xn^{-1/3} \sum_{i=1}^n \frac{1}{2} \text{sign}(s - t) [1(t \leq u_i \leq s) + 1(t \leq v_i \leq s)] \\ &:= \frac{1}{n} \sum_{i=1}^n \xi_i + xn^{-1/3} \sum_{i=1}^n \zeta_i. \end{aligned}$$

Then $E(\xi_i) = M(s) - M(t)$ and

$$(M_n - M)(s) - (M_n - M)(t) = n^{-1/2} \left(n^{-1/2} \sum_{i=1}^n (\xi_i - E(\xi_i)) + xn^{1/6} \frac{1}{n} \sum_{i=1}^n \zeta_i \right).$$

Note

$$E|\zeta_i| = \frac{1}{2} (|G_1(t) - G_1(s)| + |G_2(t) - G_2(s)|) = O(|t - s|)$$

and for $|s - t| < \delta$ with δ small,

$$\begin{aligned} \sigma^2(s, t) := E\xi_i^2 &= \frac{1}{4} \left(\text{sign}(s - t) \int_t^s [F^2(t) - 2F(t)F(y) + F(y)][g_1(y) + g_2(y)] dy \right. \\ &\quad \left. + 2 \int_t^s \int_t^s [F^2(t) - F(t)F(u) - F(t)F(v) + F(u)]g(u, v) dudv \right) \\ &= O(|s - t|). \end{aligned}$$

Now, let $\eta_i = \xi_i/\sigma(s, t)$, then the η_i 's are i.i.d. with $E(\eta_i) = 0$, $E(\eta_i^2) = 1$ and $E|\eta_i|^\alpha < \infty$ ($2 < \alpha \leq 4$). Let $S_n = \sum_{i=1}^n \eta_i$, by Theorem 2.2.4 in Csörgö and Révész (1981, p.94), there is a Wiener process $W(\cdot)$ such that

$$\frac{|S_n - W(n)|}{n^{1/\alpha}(\log n)^{1/2}} \xrightarrow{a.s.} 0.$$

So

$$\begin{aligned} n^{-1/2}S_n &= n^{-1/2}W(n) + n^{-(1/2-1/\alpha)}(\log n)^{1/2} \frac{S_n - W(n)}{n^{1/\alpha}(\log n)^{1/2}} \\ &= n^{-1/2}W(n) + o(1), \quad a.s. \end{aligned}$$

Note $n^{-1/2}W(n) \stackrel{D}{=} W(1) \sim N(0, 1)$. Thus there is $0 < C < \infty$ such that for large n ,

$$\begin{aligned} \sup_{|t-s|<\delta} |(M_n - M)(s) - (M_n - M)(t)| &= n^{-1/2} \sup_{|t-s|<\delta} \left| n^{-1/2}\sigma(s, t)S_n + xn^{1/6} \frac{1}{n} \sum_{i=1}^n \zeta_i \right| \\ &\leq n^{-1/2} \sup_{|t-s|<\delta} \left(\sigma(s, t)|n^{-1/2}S_n| + |x|n^{1/6} \frac{1}{n} \sum_{i=1}^n |\zeta_i| \right) \\ &\leq n^{-1/2} \sup_{|t-s|<\delta} \left(\sigma(s, t)[C + C|W(1)|] + C|x|n^{1/6}|t - s| \right), \quad (a.s.). \end{aligned}$$

Since $E|W(1)| \leq (EW^2(1))^{1/2} = 1$. Let $\phi_n(\delta) = \sqrt{\delta} + |x|n^{1/6}\delta$, then $\phi_n(\delta)/\delta^\alpha$ is decreasing on $(0, \infty)$ for $1 < \alpha < 2$. We have

$$E \left(\sup_{|t-s|<\delta} |(M_n - M)(s) - (M_n - M)(t)| \right) \leq \frac{C\phi_n(\delta)}{\sqrt{n}}.$$

Let r_n be the largest sequence satisfies

$$r_n^2 \phi_n \left(\frac{1}{r_n} \right) \leq \sqrt{n},$$

which gives $r_n = n^{1/3}$. Since $\hat{s}_n = \arg \max_s M_n(s)$, $M_n(\hat{s}_n) \geq M_n(t) \geq M_n(t) - O_p(r_n^{-2})$, so by Theorem 3.2.5 in VW (p. 289-290), we have $r_n|\hat{s}_n - t| = O_p(1)$, as we wanted.

Proof of Theorem 2. (i). Let $\mu = \mu_1 \times \mu_2$, where μ_1 is the Lebesgue measure on $(R^+)^2$ and μ_2 the counting measure on $D := \{(0, 0), (0, 1), (1, 1)\}$, P_F be the distribution of p_F , \mathcal{B} be the Borel filed on $(R^+)^2 \times D$, and $H(p_{\hat{F}_n}, p_F)$ be the Hellinger distance as given before Theorem 3, and $\|p_{\hat{F}_n} - p_F\|$ be the variational distance, between $p_{\hat{F}_n}(\cdot)$ and p_F ,

$$\|p_{\hat{F}_n} - p_F\| = 2 \sup\{|P_{\hat{F}_n}(B) - P_F(B)| : B \in \mathcal{B}\} = \int |p_{\hat{F}_n}(u, v, \delta, \gamma) - p_F(u, v, \delta, \gamma)| d\mu.$$

It is known that $\|p_{\hat{F}_n} - p_F\| \leq 2h(p_{\hat{F}_n}, p_F)$ (cf. Bickel, Klaassen, Ritov, & Wellner, 1993, p. 464). We will show that $H(p_{\hat{F}_n}, p_F) \rightarrow 0$ a.s., so that $\|p_{\hat{F}_n} - p_F\| \rightarrow 0$ (a.s.). Note p_F can be re-written as

$$p_F(u, v, \delta, \gamma) = \left(\delta F(u) + (\gamma - \delta)(F(v) - F(u)) + (1 - \gamma)(1 - F(v)) \right) g(u, v).$$

So we will have

$$\begin{aligned} & \int_0^\infty |\widehat{F}_n(t) - F(t)|(g_1(t) + g_2(t))dt \\ & \leq \int_0^\infty \int_0^v \left(|\widehat{F}_n(u) - F(u)| + |\widehat{F}_n(v) - F(v) - \widehat{F}_n(u) + F(u)| + |\widehat{F}_n(v) - F(v)| \right) g(u, v) dudv \\ & = \sum_{(\delta, \gamma) \in D} \int_0^\infty \int_0^v |p_{\widehat{F}_n}(u, v, \delta, \gamma) - p_F(u, v, \delta, \gamma)| dudv \\ & = \|p_{\widehat{F}_n} - p_F\| \rightarrow 0, \quad (a.s.). \end{aligned}$$

The above implies $\widehat{F}_n(t) - F(t) \rightarrow 0$ a.e. $(G_1 + G_2)$, a.s. When $F(\cdot)$ is a continuous distribution function, this in turn implies

$$\limsup_n \sup_t |\widehat{F}_n(t) - F(t)| \rightarrow 0, \quad a.e. (G_1 + G_2), \quad a.s.$$

Below we show $H(p_{\widehat{F}_n}, p_F) \rightarrow 0$ a.s. For fixed F , let $h_{F'} = (\sqrt{p_{F'}/p_F} - 1)1(p_F > 0)$, $\mathcal{G} = \{h_{F'} : F' \in \mathcal{F}\}$, and P_n and P are as given in the proof of Theorem 1. By Lemma 1.1 of van de Geer (1993), since \widehat{F}_n is the NPMLE of F in model (1),

$$H^2(p_{\widehat{F}_n}, p_F) \leq 2(P_n - P) \left(1(p_F > 0) [\sqrt{p_{\widehat{F}_n}/p_F} - 1] \right).$$

So we only need to show

$$\sup_{h \in \mathcal{G}} |(P_n - P)h| \rightarrow 0, \quad a.s.$$

i.e., \mathcal{G} is a Glivenko-Cantelli class with respect to P .

For this, given a probability measure Q and $r > 0$, let $\|g\|_{Q,r} = [Q|g|^r]^{1/r} := [E_Q|g(T)|^r]^{1/r}$, $N(\epsilon, \mathcal{G}, \|\cdot\|_{Q,r})$ be the covering number of the set \mathcal{G} with respect to the norm $\|\cdot\|_{Q,r}$, and $H(\epsilon, \mathcal{G}, \|\cdot\|_{Q,r}) = \log N(\epsilon, \mathcal{G}, \|\cdot\|_{Q,r})$ be the entropy of \mathcal{G} with respect to $\|\cdot\|_{Q,r}$. Similarly, let $N_{[\cdot]}(\epsilon, \mathcal{G}, \|\cdot\|_{Q,r})$ be the minimum number of brackets to cover \mathcal{G} under norm $\|\cdot\|_{Q,r}$, i.e. the minimum number k of pairs (l_k, u_k) , $l_k, u_k \in \mathcal{G}$ such that $\forall g \in \mathcal{G}$, there is (l_k, u_k) with $l_k \leq g \leq u_k$ and $\|u_k - l_k\|_{Q,r} \leq \epsilon$, and $H_{[\cdot]}(\epsilon, \mathcal{G}, \|\cdot\|_{Q,r}) = \log N_{[\cdot]}(\epsilon, \mathcal{G}, \|\cdot\|_{Q,r})$.

Below we need to evaluate $H(\epsilon, \mathcal{G}, \|\cdot\|_{P_{n,1}})$. Since for all $F_1, F_2 \in \mathcal{F}$,

$$\begin{aligned} & \left\| \left(\sqrt{\frac{p_{F_1}}{p_F}} - 1 \right) 1(p_F > 0) - \left(\sqrt{\frac{p_{F_2}}{p_F}} - 1 \right) 1(p_F > 0) \right\|_{P_1} \\ & = \sum_{(\delta, \gamma) \in D} \int_0^\infty \int_0^v \frac{|p_{F_1}^{1/2}(u, v, \delta, \gamma) - p_{F_2}^{1/2}(u, v, \delta, \gamma)|}{p_F^{1/2}(u, v, \delta, \gamma)} p_F(u, v, \delta, \gamma) dudv \\ & = \sum_{(\delta, \gamma) \in D} \int_0^\infty \int_0^v |p_{F_1}^{1/2}(u, v, \delta, \gamma) - p_{F_2}^{1/2}(u, v, \delta, \gamma)| p_F^{1/2}(u, v, \delta, \gamma) dudv \\ & = \|\sqrt{p_{F_1}} - \sqrt{p_{F_2}}\|_{Q_1}, \end{aligned}$$

where Q is the probability measure corresponding to $\sqrt{p_F}$ (after normalization) and μ_2 , as by condition (C3) this measure is well defined. Now, let $\mathcal{G}_0 = \{\sqrt{p_{F'}} : F' \in \mathcal{F}\}$. Since F is fixed, we have $H(\epsilon, \mathcal{G}, \|\cdot\|_{P_1}) = H(\epsilon, \mathcal{G}_0, \|\cdot\|_{Q_1}) = O(H(\epsilon, \mathcal{G}_0, \|\cdot\|_{P_1}))$. Thus, there is $0 < C < \infty$, for large n ,

$$H(\epsilon, \mathcal{G}, \|\cdot\|_{P_{n,1}}) \leq CH(\epsilon, \mathcal{G}_0, \|\cdot\|_{P_{n,1}}) \quad a.s.$$

Now we only need to evaluate $H(\epsilon, \mathcal{G}_0, \|\cdot\|_{P_{n,1}})$. Since μ_2 is counting measure for (δ, γ) , in computing the entropy we can treat $p_{F'}$ as a function of (u, v) only, and \mathcal{G}_0 is a collection of bounded two-dimensional functions on $(R^+)^2$, so by Corollary 2.7.4 in VW (p. 158), we have

$$H_{[\cdot]}(\epsilon, \mathcal{G}_0, \|\cdot\|_{P_{n,1}}) = O\left(\frac{1}{\epsilon^2}\right).$$

Since $N(\epsilon, \mathcal{G}_0, \|\cdot\|) \leq N_{[\cdot]}(2\epsilon, \mathcal{G}_0, \|\cdot\|)$ (VW, p. 84) for any norm $\|\cdot\|$, we have $H(\epsilon, \mathcal{G}_0, \|\cdot\|_{P_{n,1}}) = O(1/(4\epsilon^2)) < \infty$ (a.s.), and so for all fixed $\epsilon > 0$,

$$\frac{1}{n} H(\epsilon, \mathcal{G}, \|\cdot\|_{P_{n,1}}) \xrightarrow{P} 0$$

and by Theorem 24 in Pollard (1984, p. 23), \mathcal{G} is a Glivenko-Cantelli class with respect to P .

(ii). Recall the notation $\tilde{l}_n(F)$ in the proof of Theorem 1. It is not a proper log-likelihood. Let

$$\tilde{l}_{1,n}(F) = \sum_{i=1}^n [\delta_i F(u_i) + (1 - \delta_i)(1 - F(u_i))] g_1(u_i),$$

$$\tilde{l}_{2,n}(F) = \sum_{i=1}^n [\gamma_i F(v_i) + (1 - \gamma_i)(1 - F(v_i))] g_2(v_i).$$

Then $\tilde{l}_n(F) = \tilde{l}_{1,n}(F) + \tilde{l}_{2,n}(F)$, and $\tilde{l}_{1,n}(F)$ and $\tilde{l}_{2,n}(F)$ are log-likelihoods corresponding to densities $p_{1,F}(u, \delta) = F^\delta(u)(1 - F(u))^{1-\delta} g_1(u)$ and $p_{2,F}(v, \gamma) = F^\gamma(v)(1 - F(v))^{1-\gamma} g_2(v)$. As in the proof of Theorem 1, the same \hat{F}_n maximizes both $\tilde{l}_{1,n}(F)$ and $\tilde{l}_{2,n}(F)$. Let $\mu = \mu_1 \times \mu_2$, with μ_1 the Lebesgue measure on R^+ and μ_2 the counting measure on $\{0, 1\}$, and let

$$H^2(p_{1,\hat{F}_n}, p_{1,F}) = \int \left(p_{1,\hat{F}_n}^{1/2}(u, \delta) - p_{1,F}^{1/2}(u, \delta) \right)^2 d\mu,$$

$$H^2(p_{2,\hat{F}_n}, p_{2,F}) = \int \left(p_{2,\hat{F}_n}^{1/2}(v, \gamma) - p_{2,F}^{1/2}(v, \gamma) \right)^2 d\mu.$$

Let $h_{k,F'} = (\sqrt{p_{k,F'}/p_{k,F}} - 1)$, P_k be the probability corresponding to $p_{k,F}$ ($k = 1, 2$), $P_{1,n}$ be the empirical measure for (u_i, δ_i) ($i = 1, \dots, n$), and $P_{2,n}$ for that of (v_i, γ_i) ($i = 1, \dots, n$). Then as before

$$H^2(p_{k,\hat{F}_n}, p_{k,F}) \leq 2(P_{k,n} - P_k)h_{k,\hat{F}_n}, \quad (k = 1, 2).$$

Similarly, we have $H^2(p_{k,\hat{F}_n}, p_{k,F}) \rightarrow 0$ (a.s.), and consequently $\|p_{k,\hat{F}_n} - p_{k,F}\| \rightarrow 0$ (a.s.), and note

$$\|p_{1,\hat{F}_n} - p_{1,F}\| = 2 \int_0^\infty |\hat{F}_n(u) - F(u)| g_1(u) du, \quad \|p_{2,\hat{F}_n} - p_{2,F}\| = 2 \int_0^\infty |\hat{F}_n(v) - F(v)| g_2(v) dv,$$

the result follows. □

Proof of Theorem 3. (i) Let $d(F, F') = D^\beta(p_F \| p_{F'})$ with $\beta = 5/8$. Note $d(F, F') \geq 0$ with “=” iff $F \equiv F'$, though $d(\cdot, \cdot)$ is not a distance. For $F' \in \mathcal{F}$, let $l(F') = E_{F'}(\log p_{F'}) = E_{F'}(l_n(F'))$. It is seen that $F = \arg \max_{F'} l(F')$. Also, for $F' \in \mathcal{F}$ within a small neighborhood of F , i.e., with $D(p_F \| p_{F'}) \leq 1$, we have

$$l(F') - l(F) = -D(p_F \| p_{F'}) \leq -d^2(F, F').$$

Denote $x_i = (u_i, v_i, \delta_i, \gamma_i)$, then

$$\begin{aligned} (l_n(F) - l(F)) - (l_n(F') - l(F')) &= n^{-1/2} \left(n^{-1/2} \sum_{i=1}^n [\log \frac{p_{F'}(x_i)}{p_F(x_i)} - (l(F) - l(F'))] \right) \\ &:= n^{-1/2} \left(n^{-1/2} \sum_{i=1}^n [\xi_i - E_F(\xi_i)] \right). \end{aligned}$$

The ξ_i 's are i.i.d.. Let $\sigma^2(F, F') = \text{Var}_F(\xi_i)$, which exists by (C5) for all F' in a small neighborhood $B(\delta) = \{F' : d(F, F') < \delta\} = \{F' : D(p_F \| p_{F'}) < \delta^{1/\beta}\}$ of F , we have, for some $0 < C < \infty$,

$$\begin{aligned} \sigma^2(F, F') \leq E_F(\xi_i^2) &= E_F \left(\log \frac{p_{F'}(x_i)}{p_F(x_i)} \right)^2 \leq E_F \left(\left[\frac{p_{F'}(x_i)}{p_F(x_i)} - 1 \right] \log \frac{p_{F'}(x_i)}{p_F(x_i)} \right) \\ &= D(p_{F'} \| p_F) + D(p_F \| p_{F'}) \leq CD(p_F \| p_{F'}) \\ &= Cd^{1/\beta}(F, F') \leq C\delta^{1/\beta}. \end{aligned}$$

Thus, $\sigma(F, F') \leq C\delta^{1/(2\beta)}$ for $F' \in B(\delta)$. Now, let $\eta_i = (\xi_i - E_F(\xi_i))/\sigma(F, F')$, then the η_i 's are i.i.d. with $E_F(\eta_i) = 0$, $\text{Var}_F(\eta_i) = E_{p_F}(\eta_i^2) = 1$ and $E_F|\eta_i|^\alpha < \infty$ (by (C5)). Let $S_n = \sum_{i=1}^n \eta_i$, then, as in the proof of Theorem 1, for some $0 < C < \infty$,

$$E_F \sup_{d(F,F') < \delta} |l_n(F) - l(F)) - (l_n(F') - l(F'))| \leq Cn^{-1/2} \sup_{d(F,F') < \delta} \sigma(F, F') E_F |n^{-1/2} S_n| \leq Cn^{-1/2} \phi_n(\delta),$$

with $\phi_n(\delta) = \delta^{1/(2\beta)}$. Note $\phi_n(\delta)/\delta^\alpha$ is decreasing on $(0, \infty)$ for some $\alpha < 2$. The largest sequence r_n satisfies

$$r_n^2 \phi_n\left(\frac{1}{r_n}\right) \leq \sqrt{n}$$

is $r_n = n^{\beta/(4\beta-1)}$. Also, $\hat{F}_n = \arg \max_{F'} l_n(F')$, and for all $F' \in \mathcal{F}$,

$$l_n(\hat{F}_n) \geq l_n(F') \geq l_n(F') - O_P(r_n^{-2}),$$

so by Theorem 3.2.5 in VW (p.289-290), $r_n d(F, \hat{F}_n) = O_P(1)$, or $n^{2/3} D(p_F \| p_{\hat{F}_n}) = O_P(1)$.

Now, let $\|p_{\hat{F}_n} - p_F\| = \int |p_{\hat{F}_n}(x) - p_F(x)| dx$ be the variational distance between $p_{\hat{F}_n}$ and p_F . By Pinsker's inequality $\|p_{\hat{F}_n} - p_F\| \leq \sqrt{2D(p_F \| p_{\hat{F}_n})}$, this gives

$$\begin{aligned} O_P(1) &= n^{1/3} \|p_{\hat{F}_n} - p_F\| \\ &= n^{1/3} \int_0^\infty \int_0^v \left(|\hat{F}_n(v) - F(v)| |\hat{F}_n(v) - F(v) - \hat{F}_n(u) + F(u)| + |\hat{F}_n(u) - F(u)| \right) g(u, v) du dv \\ &\geq n^{1/3} \int_0^\infty |\hat{F}_n(t) - F(t)| (g_1(t) + g_2(t)) dt. \end{aligned}$$

The above implies $n^{1/3} |\hat{F}_n(t) - F(t)| = O_P(1)$, a.e. $G_1 + G_2$. When $F(\cdot)$ is continuous, $\hat{F}_n(t) \xrightarrow{P} F(t)$ is uniform, thus $\sup_t |\hat{F}_n(t) - F(t)| = O_P(n^{-1/3})$, a.e. $G_1 + G_2$.

(ii). Recall the notation $\tilde{l}_n(F)$ in the proof of Theorem 1. For $F' \in \mathcal{F}$, let $\tilde{l}(F') = E_{p_{1,F'}}(\log p_{1,F'}) + E_{p_{2,F'}}(\log p_{2,F'})$. Then $F = \arg \max_{F'} \tilde{l}(F')$. Let $d(F, F') = [D(p_{1,F} \| p_{1,F'}) + D(p_{2,F} \| p_{2,F'})]^\beta$, with $\beta = 5/8$ as in the proof in (i). We have

$$\tilde{l}(F') - \tilde{l}(F) = -[D(p_{1,F} \| p_{1,F'}) + D(p_{2,F} \| p_{2,F'})] \leq -d^2(F, F').$$

Denote $x_i = (u_i, \delta_i)$, $y_i = (v_i, \gamma_i)$, then

$$\begin{aligned} (\tilde{l}_n(F) - \tilde{l}(F)) - (\tilde{l}_n(F') - \tilde{l}(F')) &= n^{-1/2} \left(n^{-1/2} \sum_{i=1}^n \left[\log \frac{p_{1,F}(x_i)}{p_{1,F'}(x_i)} + \log \frac{p_{2,F}(y_i)}{p_{2,F'}(y_i)} - (l(F) - l(F')) \right] \right) \\ &:= n^{-1/2} \left(n^{-1/2} \sum_{i=1}^n [\xi_i - E_F(\xi_i)] \right). \end{aligned}$$

The ξ_i 's are i.i.d.. Let $\sigma^2(F, F') = \text{Var}_F(\xi_i)$, which exists by (C6) for all F' in a small neighborhood $B(\delta) = \{F': d(F, F') < \delta\} = \{F': D(p_{1,F} \| p_{1,F'}) + D(p_{2,F} \| p_{2,F'}) < \delta^{1/\beta}\}$ of F . Similarly as in the proof of (i), we have, for some $0 < C < \infty$,

$$\sigma(F, F') \leq 2 \sum_{k=1}^2 [D(p_{k,F} \| p_{k,F'}) + D(p_{k,F'} \| p_{k,F})] \leq C \delta^{1/\beta}.$$

Also, with $r_n = n^{\beta/(4\beta-1)}$, $\forall F' \in \mathcal{F}$, $\tilde{l}_n(\hat{F}_n) \geq \tilde{l}_n(F') - O_P(r_n^{-2})$. So, as before, we get $r_n d(F, \hat{F}_n) = O_P(1)$, or $n^{1/3} [D(p_{1,F} \| p_{1,\hat{F}_n}) + D(p_{2,F} \| p_{2,\hat{F}_n})] = O_P(1)$, and consequently, by Pinsker's inequality,

$$\begin{aligned} O_P(1) &= n^{1/3} (\|p_{1,\hat{F}_n} - p_{1,F}\| + \|p_{2,\hat{F}_n} - p_{2,F}\|) \\ &= 2n^{1/3} \int_0^\infty |\hat{F}_n(t) - F(t)| (g_1(t) + g_2(t)) dt. \end{aligned}$$

Lastly we check that the rate $r_n = n^{1/3}$ is optimal. We only check this for (ii) and that for (i) is the same. By Theorem 2.7.5 in van der Vaart and Wellner (1996, p. 159), the bracketing entropy $H_{[\cdot]}(\epsilon, \mathcal{F}, L_r(Q)) = O(\epsilon^{-1})$, i.e., we have $d = \alpha = 1$ (recall the paragraph for optimal rate before Theorem 3), and the best rate r_n of convergence is $r_n = n^{\alpha/(2\alpha+1)} = n^{1/3}$, or $n^{1/3} h(p_{F_n}, p_{F_0}) = O_P(1)$ for any estimator F_n of F_0 . Also, since $\alpha/d = 1 > 1/2$, $n^{1/3}$ is also the best achievable rate for NPML of F_0 . Similarly as in the proof of Theorem 2 (i), we have

$$\begin{aligned} h(p_{F_n}, p_{F_0}) &\geq \left[\frac{1}{2} \int (F_n^{1/2}(u) - F_0^{1/2}(u))^2 d(G_1(u) + G_2(u)) \right]^{1/2} \\ &\geq \frac{1}{\sqrt{2}} \int |F_n^{1/2}(u) - F_0^{1/2}(u)| d(G_1(u) + G_2(u)) \\ &\geq \frac{1}{2\sqrt{2}} \int |F_n^{1/2}(u) - F_0^{1/2}(u)| |F_n^{1/2}(u) + F_0^{1/2}(u)| d(G_1(u) + G_2(u)) \\ &= \int |F_n(u) - F_0(u)| d(G_1(u) + G_2(u)), \end{aligned}$$

thus the best achievable rate of convergence of estimating F_0 in terms of $L_1(G_1 + G_2)$ distance is $n^{1/3}$.

Proof of Theorem 4. Let \mathbb{P}_n, P and Δ_i 's as in the proof of Theorem 1. Note P_F is the distribution in model (1), not to confuse it with P . Then as in Huang and Wellner (1995), hereafter HW, for any function $r(\cdot)$, we have

$$\int (\Delta - \hat{F}_n(t))r(\hat{F}_n(t))d\mathbb{P}_n(t, \Delta) = 0.$$

Note $dP(t, \delta) = (1/2)F^\delta(t)(1 - F(t))^{1-\delta}g_1(t)dt$ and $dP(t, \gamma) = (1/2)F^\gamma(t)(1 - F(t))^{1-\gamma}g_2(t)dt$. Let $g(t) = (g_1(t) + g_2(t))/2$, and $G(t) = (G_1(t) + G_2(t))/2$. Then

$$\begin{aligned} \int \frac{(1-\Delta)h(t)}{g(t)}dP(t, \Delta) &= \frac{1}{2} \left(\int \frac{(1-0)h(t)}{g(t)}(1 - F(t))g_1(t)dt + \int \frac{(1-0)h(t)}{g(t)}(1 - F(t))g_2(t)dt \right) \\ &= \int h(t)(1 - F(t))dt. \end{aligned}$$

Similarly,

$$\int \frac{(1 - \hat{F}_n(t))h(t)}{g(t)}dP(t, \Delta) = \int (1 - \hat{F}_n(t))h(t)dt.$$

So we have, as in HW,

$$\begin{aligned} \sqrt{n}(\nu(\hat{F}_n) - \nu(F)) &= \sqrt{n} \int [(1 - \hat{F}_n(t)) - (1 - F(t))]h(t)dt \\ &= \sqrt{n} \int \frac{1 - \hat{F}_n(t) - (1 - \Delta)}{g(t)}h(t)dP(t, \Delta) \\ &= \sqrt{n} \int \frac{\Delta - F(t)}{g(t)}h(t)d(\mathbb{P}_n - P)(t, \Delta) + o_P(1). \end{aligned}$$

Now the rest proofs are the same as in HW. We only point out that, in the proof of the $o_P(1)$ part, to show $\int [\hat{F}_n(t) - F(t)]^2 dG(t) \xrightarrow{P} 0$, we can use our Theorem 2(ii) directly when the conditions there are met, in stead of the more complicated arguments there; to show $K\sqrt{n}\|\hat{F}_n - F\|_{L_2(G)}^2 = O_P(n^{-1/6})$, we only need to apply our Theorem 3(ii) directly when the conditions there are met, in stead of the arguments there. It is easy to check that the asymptotic variance of $\sqrt{n} \int \frac{\Delta - F(t)}{g(t)}h(t)d(\mathbb{P}_n - P)(t, \Delta)$ is σ^2 .

Now we compute the efficient influence function $\tilde{I}_\nu(\cdot)$ and the information bound for estimating $\nu(F)$ via model (1) and constraint (C2). We first compute the efficient influence function of $\nu(F)$ without constraint (C2), the extended version $\tilde{I}_{\nu,e}(\cdot)$. By Theorem 2 (iii) in Yuan, Xu and Zheng (2011), with $x = (u, v, \delta, \gamma)$, we have

$$\tilde{I}_{\nu,e}(x) = \frac{g_1(u)h(v)}{g(u, v)}[(1 - \delta)(1 - \gamma) - (1 - F(v))].$$

Let $r(u, v) = (F(v) - F(u)) \log(F(v) - F(u)) - (1 - F(u)) \log(1 - F(u)) - F(v) \log F(v)$. Then (C2) is $\eta(F) := E_{P_F}[r(U, V)] = 0$. Let $\Pi(s|s_1)$ be the projection of s onto $[s_1]$, the linear span of s_1 , s_1^\perp the orthogonal complement of $[s_1]$ with respect to P_P , $\langle s_1, s_2 \rangle_{P_F} = E_{P_F}(s_1, s_2)$ and $\|s\|_{P_F}^2 = \langle s, s \rangle_{P_F}$. By Proposition A.5.2 in Bickel, Klaassen, Ritov and Wellner (1993), the canonical gradient of η (the adjoint of the pathwise differential $\dot{\eta}[\cdot](F)$ of $\eta(F)$, evaluated at 1) is $\dot{\eta}^*(x) = \dot{\eta}^*[1](x) = r(u, v)$, and

$$\tilde{I}_\nu(x) = \Pi(\tilde{I}_{\nu,e}|(\dot{\eta}^*)^\perp) = \tilde{I}_{\nu,e}(x) - \langle \tilde{I}_{\nu,e}, \dot{\eta}^* \rangle_{P_F} \|\dot{\eta}^*\|_{P_F}^{-2} \dot{\eta}^*(x),$$

and the information bound for estimating $\nu(F)$ in model (1) with constraint (C2) is $E_{P_F}(\tilde{I}_\nu^2(x))$. Since generally $E_{P_F}(\tilde{I}_\nu^2(x)) \neq \sigma^2$, $\nu(\hat{F}_n)$ is not efficient for $\nu(F)$, unlike the corresponding result for the case I interval censoring model. \square

Proof of Theorem 5. The log-likelihood for θ under this submodel is

$$L_n(\theta) = \sum_{i=1}^n \left(\delta_i \log F_n(u_i|\theta) + (\gamma_i - \delta_i) \log(F_n(v_i|\theta) - F_n(u_i|\theta)) + (1 - \gamma_i) \log(1 - F_n(v_i|\theta)) \right),$$

with local log-likelihood ratio

$$L_n(\theta + hr_n^{-1}) - L_n(\theta) = r_n^{-1}L'_n(\theta)h + \frac{1}{2}r_n^{-2}L''_n(\theta)h^2 + o(1)r_n^{-2}L''_n(\theta)h^2.$$

where $L'_n(\theta)$ and $L''_n(\theta)$ are the first two derivatives,

$$L'_n(\theta) = \sum_{i=1}^n \left(\frac{\delta_i H_n(u_i)}{F_n(u_i|\theta)} + \frac{(\gamma_i - \delta_i)(H_n(v_i) - H_n(u_i))}{F_n(v_i|\theta) - F_n(u_i|\theta)} - \frac{(1 - \gamma_i)H_n(v_i)}{1 - F_n(v_i|\theta)} \right) := \sum_{i=1}^n \xi_{ni},$$

$$L''_n(\theta) = - \sum_{i=1}^n \left(\frac{\delta_i H_n^2(u_i)}{F_n^2(u_i|\theta)} + \frac{(\gamma_i - \delta_i)(H_n(v_i) - H_n(u_i))^2}{(F_n(v_i|\theta) - F_n(u_i|\theta))^2} + \frac{(1 - \gamma_i)H_n^2(v_i)}{(1 - F_n(v_i|\theta))^2} \right) := - \sum_{i=1}^n \eta_{ni}.$$

Note the ξ_{ni} 's are i.i.d., with $E_\theta(\xi_{ni}) = 0$; $H_n(t) = f(t)ca_n^{-1}$, $F_n(t|\theta) = F(t) + \theta f(t)ca_n^{-1} \sim F(t)$ and

$$H_n(s) = \begin{cases} 0, & s < t - ca_n^{-1} \text{ or } s \geq t + ca_n^{-1}; \\ f(t)(s - t + ca_n^{-1}), & t - ca_n^{-1} < s \leq t; \\ f(t)(ca_n^{-1} - s + t), & t < s < t + ca_n^{-1}. \end{cases}$$

We have

$$\begin{aligned} Var_\theta(\xi_{ni}) = E_\theta(\eta_{ni}^2) &= E_\theta \left(\frac{\delta_i H_n^2(u_i)}{F_n^2(u_i|\theta)} \right) + E_\theta \left(\frac{(\gamma_i - \delta_i)(H_n(v_i) - H_n(u_i))^2}{(F_n(v_i|\theta) - F_n(u_i|\theta))^2} \right) + E_\theta \left(\frac{(1 - \gamma_i)H_n^2(v_i)}{(1 - F_n(v_i|\theta))^2} \right) \\ &:= J_{n1} + J_{n2} + J_{n3}. \end{aligned}$$

$$J_{n1} \sim 2 \int_{t-ca_n^{-1}}^t \frac{f^2(t)g_1(u)(u - t + ca_n^{-1})^2}{F(u)} du \sim \frac{2}{3} \frac{f^2(t)g_1(t)}{F(t)} c^3 a_n^{-3};$$

$$J_{n3} \sim 2 \int_{t-ca_n^{-1}}^t \frac{f^2(t)g_2(v)(v - t + ca_n^{-1})^2}{1 - F(v)} dv \sim \frac{2}{3} \frac{f^2(t)g_2(t)}{1 - F(t)} c^3 a_n^{-3}.$$

$$\begin{aligned} J_{n2} &\sim 2 \int_{t-ca_n^{-1}}^t \int_{t-ca_n^{-1}}^v \frac{f^2(t)g(u,v)(v - u)^2}{F(v) - F(u)} dudv + \int_t^{t+ca_n^{-1}} \int_{t-ca_n^{-1}}^t \frac{f^2(t)g(u,v)[(v - t) - (t - u)]^2}{F(v) - F(u)} dudv \\ &= 2 \int_{t-ca_n^{-1}}^t \int_{t-ca_n^{-1}}^v \frac{f^2(t)g(u,v)(v - u)^2}{f(u)(v - u) + O((v - u)^2)} dudv + \int_t^{t+ca_n^{-1}} \int_{t-ca_n^{-1}}^t \frac{f^2(t)g(u,v)[(v - t) - (t - u)]^2}{f(u)(v - u) + O((v - u)^2)} dudv \\ &\sim 2 \int_{t-ca_n^{-1}}^t \int_{t-ca_n^{-1}}^v f(t)g(t,t)(v - u)dudv + \int_t^{t+ca_n^{-1}} \int_{t-ca_n^{-1}}^t \frac{f(t)g(t,t)[(v - t) - (t - u)]^2}{(v - u)} dudv \\ &= \frac{1}{3} f(t)g(t,t)c^3 a_n^{-3} + c^4 a_n^{-4} \int_0^1 \int_0^1 \frac{f(t)g(t,t)(v - u)^2}{2t + ca_n^{-1}(v - u)} dudv \\ &\sim \frac{1}{3} f(t)g(t,t)c^3 a_n^{-3}. \end{aligned}$$

Thus we get

$$Var_\theta(\xi_{ni}) \sim \frac{1}{3} \left(\frac{2f^2(t)g_1(t)}{F(t)} + \frac{2f^2(t)g_2(t)}{1 - F(t)} + f(t)g(t,t) \right) c^3 a_n^{-3} := Q(t)c^3 a_n^{-3}.$$

Take $r_n^{-1} = n^{-1/2} c^{3/2} a_n^{3/2} Q^{-1/2}(t)$, by central limit theory and strong law of large numbers, we have

$$r_n^{-1} L'_n(\theta) \xrightarrow{D} N(0, 1) \text{ and } r_n^{-2} L''_n(\theta) \xrightarrow{a.s.} -1,$$

i.e., the local log-likelihood ratio satisfies the LAN condition with rate r_n and variance 1. (i). Since $a_n = o(n^{1/3})$, $r_n \rightarrow \infty$. The conclusion follows from standard convolution result with LAN condition.

(ii) With LAN, this follows by Remark 12.3 in Ibragimov and Has'minskii (1981, p. 169), or Theorem 2.3 in Groeneboom and Wellner (1992, p. 15).

(iii) Take $c^{3/2} = Q^{-\alpha}(t)\delta$ for some $\alpha > 0$ to be determined, and $l(\cdot) = |\cdot|$, we have $r_n = n^{1/3}Q^{1/2-\alpha}(t)\delta$. Take $b = Q^{1/2-\alpha}(t)\delta^2$, by Theorem 4 (ii), we get

$$\begin{aligned} & \liminf_n \sup_{\theta: n^{1/3}|\theta| \leq \delta} E_\theta \left| n^{1/3}(T_n - F_n(t|\theta)) \right| \geq \frac{1}{2} Q^{\alpha-1/2}(t)\delta^{-1} E[|Z|1(|Z| \leq Q^{1/2-\alpha}\delta^2/2)] \\ &= Q^{\alpha-1/2}\delta^{-1} \sqrt{2\pi} \int_0^{Q^{1/2-\alpha}\delta^2/2} z e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} Q^{\alpha-1/2}\delta^{-1} (1 - e^{-Q^{1-2\alpha}\delta^4/8}) \\ &= \frac{1}{\sqrt{2\pi}} Q^{\alpha-1/2+(1-2\alpha)/4} a^{-1} (1 - e^{-a^4/8}), \end{aligned}$$

where $a = Q^{(1-2\alpha)/4}\delta$. Take $\alpha = 7/6$, then $Q^{\alpha-1/2+(1-2\alpha)/4} = Q^{1/3}$. Maximizing $R(a) := a^{-1}(1 - e^{-a^4/8})$, or solve $2e^{a^4/8} - a^4 - 2 = 0$ for $a > 0$, with $a_0 = \arg \max_{a>0} R(a)$, we get $a_0^4 = 18.6933$ and $R(a_0) = 0.434449$. Thus

$$\sup_{\delta>0} \liminf_n \sup_{\theta: n^{1/3}|\theta| \leq \delta} E_\theta \left| n^{1/3}(T_n - F_n(t|\theta)) \right| \geq \frac{0.434449}{\sqrt{2\pi}} Q^{1/3}(t).$$

□

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).