

On asymptotic equivalence of the NPMLE of a monotone density and a Grenander-type estimator in multi-sample biased sampling models

Kwun Chuen Gary Chan

*Department of Biostatistics, University of Washington,
Seattle, United States
e-mail: kcgchan@uw.edu*

Hok Kan Ling

*Department of Mathematics and Statistics, Queen's University,
Kingston, Canada
e-mail: brian.ling@queensu.ca*

Tony Sit and Sheung Chi Phillip Yam

*Department of Statistics, The Chinese University of Hong Kong,
Hong Kong
e-mail: tonysit@sta.cuhk.edu.hk; scpyam@sta.cuhk.edu.hk*

Abstract: In this article, we show that the nonparametric maximum likelihood estimator (NPMLE) of the decreasing density function in the s -sample biased sampling models, is asymptotically equivalent to a Grenander-type estimator, namely the left-continuous slope of the least concave majorant of the NPMLE of the distribution function in the larger model without imposing the monotonicity assumption. Since the two estimators favor different proof directions in establishing weak convergence, we require additional results for both estimators so that the two estimators can be considered jointly in a unified approach. For instance, we employ an analytic argument for showing the tightness of an inverse processes associated with the NPMLE, since a conventional geometric approach used in the literature cannot be employed due to multiple biased samples. We demonstrate other results using numerical simulation and a real data illustration.

MSC2020 subject classifications: Primary 62G07, 62G20.

Keywords and phrases: Argmax continuous mapping theorem, asymptotic equivalence, density estimation, Karush-Kuhn-Tucker conditions, shape-constrained estimation, s -sample biased sampling.

Received June 2020.

Contents

1	Introduction	2877
2	Setting and notation	2879

3	Consistency of Grenander-type estimator \tilde{g}_n	2880
4	Asymptotic equivalence of \tilde{g}_n and \hat{g}_n	2881
5	Tightness of the inverse processes \hat{S}_n and \tilde{S}_n	2890
6	Numerical studies	2892
6.1	Simulations	2892
6.2	Data analysis	2894
7	Discussion	2895
A	Proofs of auxiliary lemmas	2896
	Acknowledgments	2902
	References	2903

1. Introduction

Biased sampling problems have long been an important issue in a wide array of scientific studies. One of the most popular types of biased sampling is length-biased sampling which has been recognized in statistics for almost half a century in the studies of ecology [20, 21], fiber length [5] and economic duration data [15, 13]. This kind of biased sampling arises when a positive-valued outcome variable is sampled with selection probability proportional to its size/survival time, which occurs usually in cross-sectional studies. Another example of biased sampling is discussed in [22] who considered a generalized version of length-bias in a melanoma study. [23] and [4] considered observed blood alcohol concentration of drivers in traffic accidents as bias samples of the blood alcohol concentration of all drivers, with different biasing functions for various age groups. [7] considered the distribution of amino acid strain distance in vaccine studies as a multisample biased sampling problem.

A general formulation of s -sample ($s \geq 2$) biased sampling problem is given as follows. Let G_0 be an unknown distribution function on \mathbb{R} and there are s positive known weight functions w_i ($i = 1, \dots, s$). Suppose s independent samples X_{i1}, \dots, X_{in_i} ($i = 1, \dots, s$) are observed, where each X_{ij} independently follows the biased distribution F_i ($i = 1, \dots, s, j = 1, \dots, n_i$) given by:

$$F_i(x) \triangleq \frac{\int_{-\infty}^x w_i(y) dG_0(y)}{\int_{-\infty}^{\infty} w_i(y) dG_0(y)},$$

where $0 < W_i \triangleq \int_{-\infty}^{\infty} w_i(y) dG_0(y) < \infty$ for $i = 1, \dots, s$. In the absence of the assumption on the shape of the distribution function or its underlying density, [28] established the unique existence of the nonparametric maximum likelihood estimator (NPML) \mathbb{G}_n of the unbiased distribution function G_0 in s -sample biased sampling models. Large sample theory of this NPML was investigated in [8]. Recently, [3] established the unique existence of the decreasing NPML \hat{g}_n in s -sample biased sampling models and also gave its asymptotic distribution at a fixed interior point where the underlying density has a strictly negative derivative; such a problem has been open in the literature due to certain non-standard structures of the likelihood function, such as non-separability and

a lack of strictly positive second order derivatives of the negative of the log-likelihood function. Formally, denote \mathcal{G} to be the set of all decreasing densities. For any $g \in \mathcal{G}$, the likelihood evaluated at this g of the s -sample is proportional to

$$L_n(g) \triangleq \prod_{i=1}^s \frac{\prod_{j=1}^{n_i} g(X_{ij})}{[\int_a^b w_i(x)g(x)dx]^{n_i}}.$$

The decreasing NPMLE $\hat{g}_n \in \mathcal{G}$ is defined such that $L_n(\hat{g}_n) \geq L_n(g)$ for all $g \in \mathcal{G}$.

From an alternative perspective, another natural estimator for such a decreasing density is a Grenander-type estimator, which is the left-continuous slope, denoted by \tilde{g}_n , of the least concave majorant of \mathbb{G}_n , i.e., the NPMLE of G_0 without the monotonicity assumption. In general, suppose that r is the underlying function of interest, for example, a density, hazard rate function or regressor, and $\hat{R}(t)$ is an estimator of $\int_{-\infty}^t r(s)ds$. A Grenander-type estimator is an estimator of r being monotone decreasing (resp. increasing), which is the left-continuous (resp. right-continuous) slope of the least concave majorant (resp. greatest convex minorant) of \hat{R} when r is decreasing (resp. increasing). There have been ongoing works related to Grenander-type estimators; see also [11] and the references therein. In the case when there is only one unbiased sample under the decreasing density assumption, [9] already showed that the NPMLE is exactly the left-continuous slope of the least concave majorant of the empirical distribution function, i.e., the usual NPMLE of the distribution function without constraint. Generally speaking, one may not expect this kind of correspondence to hold exactly, but perhaps just asymptotically; for instance, in the random right censorship model, [14] showed that the NPMLE of a decreasing density function is asymptotically equivalent to the left-continuous slope of the Kaplan-Meier estimator, the NPMLE of the distribution function in the absence of the monotonicity assumption. A similar result was also shown in [14] to hold for the NPMLE of a decreasing hazard rate, where the NPMLE is asymptotically equivalent to the estimator obtained as the left-continuous slope of the least concave majorant of the Nelson-Aalen estimator, i.e., the NPMLE of the cumulative hazard function without the monotonicity assumption. More recently, [18] showed that similar results are true in the Cox model; the NPMLE of an increasing baseline hazard and the left-hand slope of the greatest convex minorant of the Breslow estimator are asymptotically equivalent.

It would be of both theoretical and practical interests to see if this asymptotic equivalence also holds in other models such as that in [3]. Practically, for the s -sample biased modeling models, the computation of the NPMLE is done iteratively based on a self-characterization given in [3], where an initial consistent estimator is required since the corresponding optimization problem is non-convex. In [3], we suggested to use the Grenander-type estimator as an initial estimator, where a numerically efficient method of finding \mathbb{G}_n has already been discussed in [28]. The asymptotic equivalence of the two estimators implies that that this initial estimator is already as good as the NPMLE asymptotically. The main goal in the present article is to show that \tilde{g}_n and \hat{g}_n in s -sample biased sam-

pling models are asymptotically equivalent in the sense that $n^{1/3}[\tilde{g}_n(t_0) - \hat{g}_n(t_0)]$ converges to 0 in probability as $n \rightarrow \infty$. The rate $n^{1/3}$ is sought here because $n^{1/3}[\hat{g}_n(t_0) - g_0(t_0)]$ converges in distribution to a nondegenerate limiting distribution (Theorem 1.1 in [3]), where g_0 is the true underlying density. To study the asymptotic equivalence of the estimators, one needs to consider the two estimators jointly. The Grenander-type estimator favors a geometric method by considering a switching relation and an inverse process, which is similar to [14]. On the other hand, the approach we took in [3] for the NPMLE made use of the continuous mapping argument for slopes of least concave majorant as illustrated by [1]. To establish the asymptotic equivalence of the estimators using a unified approach of making use of the switching relation and inverse process, we develop additional results for both estimators. In particular, we show the tightness of the corresponding inverse process of the NPMLE, by a pure analytic argument instead of a commonly-employed geometric argument.

The organization of our work is as follows. In Section 2, we describe the setting and notation. Consistency of \tilde{g}_n will be established in Section 3. The main results on asymptotic equivalence of \tilde{g}_n and \hat{g}_n will be proven in Section 4. The main results depend on the tightness of two inverse processes which will be shown in Section 5. In Section 6, numerical studies including simulation and an analysis of a real data set are performed to compare the effectiveness of the two estimators. Some concluding remarks are given in Section 7.

2. Setting and notation

Assume that the distribution function G_0 has a density function g_0 with respect to the Lebesgue measure which is known to be decreasing. Denote f_i to be the density of the biased distribution F_i with respect to the Lebesgue measure. The total sample size is $n \triangleq n_1 + \dots + n_s$. To study the asymptotic behavior of the estimators \tilde{g}_n and \hat{g}_n for a nondegenerate model, we shall also assume that $\lambda_{ni} \triangleq n_i/n \rightarrow \lambda_i > 0$ as $n \rightarrow \infty$, corresponding to Assumptions 2.1 (A) in [3]. Let $\mathbb{F}_n \triangleq \sum_{i=1}^s \lambda_{ni} \mathbb{F}_{i,n_i}$, where \mathbb{F}_{i,n_i} is the empirical distribution function from the i -th sample, that is, $\mathbb{F}_{i,n_i}(x) \triangleq \frac{1}{n_i} \sum_{j=1}^{n_i} 1(X_{ij} \leq x)$. Equality in distribution and convergence in distribution will be denoted by $\stackrel{d}{=}$ and \xrightarrow{d} respectively; convergence in probability will be denoted by $\xrightarrow{\mathbb{P}}$; almost sure convergence will be denoted by $\xrightarrow{a.s.}$. For any function K on $[a, b]$ with $K(a) = 0$, the least concave majorant \tilde{K} of K is defined to be the smallest concave function that dominates K , i.e., $\tilde{K} \geq K$, over $[a, b]$ with $\tilde{K}(a) = 0$.

Let $V_0 \triangleq (V_{01}, \dots, V_{0,s-1}, 1)^T$ and $V_{0i} \triangleq W_i/W_s$. For $i = 1, \dots, s-1$, define

$$\mathbb{H}_{ni}(u_1, \dots, u_s) \triangleq \frac{1}{u_i} \int_{-\infty}^{\infty} \frac{w_i(y)}{\sum_{k=1}^s \lambda_{nk} w_k(y) u_k^{-1}} d\mathbb{F}_n(y)$$

over the domain so that $u_1, \dots, u_s > 0$. Without imposing the decreasing assumption on the unbiased density, [28] showed that the NPMLE \mathbb{G}_n can be

written as

$$\mathbb{G}_n(x) = \frac{\int_{-\infty}^x [\sum_{k=1}^s \lambda_{nk} w_k(y) \mathbb{V}_{nk}^{-1}]^{-1} d\mathbb{F}_n(y)}{\int_{-\infty}^{\infty} [\sum_{k=1}^s \lambda_{nk} w_k(y) \mathbb{V}_{nk}^{-1}]^{-1} d\mathbb{F}_n(y)}, \tag{2.1}$$

where $\mathbb{V}_n \triangleq (\mathbb{V}_{n1}, \dots, \mathbb{V}_{n,s-1}, 1)^T$ solves

$$1 = \mathbb{H}_{ni}(\mathbb{V}_{n1}, \dots, \mathbb{V}_{n,s-1}, 1), \quad \text{for } i = 1, \dots, s - 1.$$

Let $\tilde{\mathbb{G}}_n$ be the least concave majorant of \mathbb{G}_n . Then the Grenander-type estimator \tilde{g}_n is the left-continuous slope of $\tilde{\mathbb{G}}_n$. An important condition for the asymptotic properties of \mathbb{G}_n is the connectedness of the graph \mathbf{G} on the s vertices $i = 1, \dots, s$ formed by defining an edge between i and j if and only if $\int 1(w_i > 0)1(w_j > 0)dG_0 > 0$.

3. Consistency of Grenander-type estimator \tilde{g}_n

The consistency of the NPMLE \hat{g}_n has been shown in [3]. Before we proceed to the main results on asymptotic distribution in the next section, we show the consistency of the Grenander-type estimator \tilde{g}_n . Define $\|\cdot\|_\infty$ to be the supremum norm. Theorem 2.1 of [8] implies that $\|\mathbb{G}_n - G_0\|_\infty \xrightarrow{a.s.} 0$ when \mathbf{G} is connected. The same result is true for $\tilde{\mathbb{G}}_n$ according to Marshall's Lemma ([19]) as stated in Lemma 3.1.

Lemma 3.1. *Suppose that \mathbf{G} is connected. There is a contraction property such that*

$$\|\tilde{\mathbb{G}}_n - G_0\|_\infty \leq \|\mathbb{G}_n - G_0\|_\infty \xrightarrow{a.s.} 0, \quad \text{as } n \rightarrow \infty.$$

To connect the consistency of $\tilde{\mathbb{G}}_n$ with that of \tilde{g}_n , recall the following elementary convex analysis result (see e.g., p.330 in [25]):

Lemma 3.2. *Let H and H_n , $n = 1, 2, \dots$, be concave functions defined on an open interval I . If*

$$\lim_{n \rightarrow \infty} H_n(x) = H(x)$$

uniformly in $x \in I$, then

$$H^-(x) \geq \limsup_{n \rightarrow \infty} H_n^-(x) \geq \liminf_{n \rightarrow \infty} H_n^+(x) \geq H^+(x),$$

for all $x \in I$, where H^- and H^+ denote the left and right derivatives of H respectively.

The following proposition is an immediate consequence of Lemma 3.1 and Lemma 3.2.

Proposition 3.3. *Suppose that \mathbf{G} is connect. For any closed interval $[\sigma, \tau]$ contained in the interior of the support of G_0 ,*

$$\sup_{x \in [\sigma, \tau]} |\tilde{g}_n(x) - g_0(x)| \xrightarrow{a.s.} 0 \text{ as } n \rightarrow \infty.$$

Proof. By considering an open interval, a subset of the interior of the support of G_0 , slightly larger than $[\sigma, \tau]$, the pointwise result follows from Lemma 3.1 and Lemma 3.2. The uniform result holds as $\{\tilde{g}_n\}$ is a sequence of decreasing functions converging to the continuous decreasing density function g_0 , for example, see the auxiliary results as found in [2] or [24]. \square

4. Asymptotic equivalence of \tilde{g}_n and \hat{g}_n

In the rest of this article, we assume that the support of g_0 is $[a, b]$, where $-\infty < a < b < \infty$, and t_0 will be a fixed interior point of $[a, b]$. The following regularity conditions are assumed for establishing the asymptotic equivalence of \tilde{g}_n and \hat{g}_n .

Assumption 1.

- (i) The unbiased decreasing density g_0 is differentiable on the interior of its support (a, b) , with $0 < \inf_{t \in (a, b)} |g'_0(t)| \leq \sup_{t \in (a, b)} |g'_0(t)| < \infty$;
- (ii) There exist constant m, M with $0 < m < M < \infty$ such that for all $i = 1, \dots, s$, $m \leq g_0, w_i \leq M$;
- (iii) For each $i = 1, \dots, s$, $w_i, w_i \circ g_0^{-1}$ are Lipschitz continuous.

Assumptions (i)–(iii) correspond to Assumption 2.1 (B)–(D) in [3] for establishing the asymptotic distribution of \hat{g}_n . A direct consequence of Assumption (ii) warrants that \mathbf{G} is a complete graph and hence connected. As a result, the s -sample biased sampling model without the monotonicity assumption is identifiable (Proposition 1.1 in [8]) and $\sqrt{n}(\mathbb{V}_n - V_0)$ follows asymptotically a multivariate normal distribution (Proposition 2.3 in [8]). In this section we will show the following main result of this paper.

Theorem 4.1. *Under Assumption 1, we have*

$$n^{1/3}[\tilde{g}_n(t_0) - \hat{g}_n(t_0)] \xrightarrow{\mathbb{P}} 0.$$

Furthermore,

$$n^{1/3} \left[\sum_{i=1}^s \lambda_i \frac{w_i(t_0)}{W_i} \right]^{1/3} \left| \frac{1}{2} g'_0(t_0) g_0(t_0) \right|^{-1/3} [\tilde{g}_n(t_0) - g_0(t_0)] \xrightarrow{d} 2\mathbb{Y}, \quad (4.1)$$

where $\mathbb{Y} \triangleq \arg \max_t \{W(t) - t^2\}$, $\{W(t), t \in \mathbb{R}\}$ is a standard two-sided Brownian motion with $W(0) = 0$.

We first provide a roadmap on how Theorem 4.1 will be proven. One of the main difficulties is that both \hat{g}_n and \tilde{g}_n are implicitly defined without explicit simple analytic forms to analyse the difference between them directly. For example, it is not clear if some terms will cancel when we consider $n^{1/3}(\hat{g}_n(t_0) - \tilde{g}_n(t_0))$ directly. Our strategy, on the other hand, is to first show that the joint distribution $(n^{1/3}\{\hat{g}_n(t_0) - g_0(t_0)\}, n^{1/3}\{\tilde{g}_n(t_0) - g_0(t_0)\})$ converges to the same limit, serving as a proxy, so that the difference of them converges to 0 in probability.

Since \tilde{g}_n is defined as the slope of the least concave majorant of \mathbb{G}_n , to study its asymptotic distribution, we define, for $t \in \mathbb{R}$, the local processes

$$\begin{aligned}\tilde{G}_n(t) &\triangleq t; \\ \tilde{U}_n(t) &\triangleq n^{2/3}[\mathbb{G}_n(t_0 + tn^{-1/3}) - \mathbb{G}_n(t_0) - g_0(t_0)tn^{-1/3}].\end{aligned}\quad (4.2)$$

Define $\lambda \triangleq \lambda(t_0) = \left(\frac{g_0(t_0)}{\sum_{i=1}^s \lambda_i \frac{w_i(t_0)}{W_i}}\right)^{-1}$. The corresponding local processes for \hat{g}_n were defined in [3] and are given by:

$$\begin{aligned}\hat{G}_{n,\hat{g}_n}(t) &\triangleq \frac{n^{1/3}}{\lambda} \frac{1}{n} \sum_{j=1}^n \frac{1}{\hat{g}_n^2(T_j)} [1(T_j \leq t_0 + tn^{-1/3}) - 1(T_j \leq t_0)]; \\ \hat{U}_{n,\hat{g}_n}(t) &\triangleq \frac{n^{2/3}}{\lambda} \frac{1}{n} \sum_{j=1}^n \left(\frac{\hat{g}_n(T_j) - g_0(t_0)}{\hat{g}_n^2(T_j)} + \frac{1}{\hat{g}_n(T_j)} - \sum_{i=1}^s \frac{n_i c_{ij}}{\sum_{k=1}^n \hat{g}_n(T_k) c_{ik}} \right) \\ &\quad \times [1(T_j \leq t_0 + tn^{-1/3}) - 1(T_j \leq t_0)],\end{aligned}\quad (4.3)$$

where T_1, \dots, T_n are the order statistics of all the samples X_{ij} 's and $c_{ik} \triangleq \int_{T_{k-1}}^{T_k} w_i(x) dx$. Furthermore, for $(t, a) \in \mathbb{R}^2$, define

$$\begin{aligned}\tilde{Z}_n(t, a) &\triangleq \tilde{U}_n(t) - a\tilde{G}_n(t) = \tilde{U}_n(t) - at, \\ \hat{Z}_n(t, a) &\triangleq \hat{U}_{n,\hat{g}_n}(t) - a\hat{G}_{n,\hat{g}_n}(t),\end{aligned}$$

where $\tilde{G}_n, \tilde{U}_n, \hat{U}_{n,\hat{g}_n}$ and \hat{G}_{n,\hat{g}_n} are defined in (4.2) and (4.3). The proof of Theorem 4.1 will further depend on the following facts which will be proven:

$$\left\{ \arg \max_t \tilde{Z}_n(t, a) \leq 0 \right\} = \{n^{1/3}(\tilde{S}_n(n^{-1/3}a + g_0(t_0)) - t_0) \leq 0\} \quad (4.4)$$

and

$$\left\{ \arg \max_t \hat{Z}_n(t, a) \leq 0 \right\} = \{n^{1/3}(\hat{S}_n(n^{-1/3}a + g_0(t_0)) - t_0) \leq 0\}, \quad (4.5)$$

where for $a > 0$,

$$\begin{aligned}\tilde{S}_n(a) &\triangleq \arg \max_{t \geq 0} \{\mathbb{G}_n(t) - at\}, \\ \hat{S}_n(a) &\triangleq \arg \max_{t \geq 0} \{U_{n,\hat{g}_n}(t) - aG_{n,\hat{g}_n}(t)\}, \\ G_{n,\hat{g}_n}(t) &\triangleq \frac{1}{n} \sum_{j=1}^n \frac{1}{\hat{g}_n^2(T_j)} 1(T_j \leq t), \\ U_{n,\hat{g}_n}(t) &\triangleq \frac{1}{n} \sum_{j=1}^n \left(\frac{2}{\hat{g}_n(T_j)} - \sum_{i=1}^s \frac{n_i c_{ij}}{\sum_{k=1}^n \hat{g}_n(T_k) c_{ik}} \right) 1(T_j \leq t).\end{aligned}$$

Then by the definition of \tilde{g}_n and Proposition 3.2 in [3], with probability one, we have the following switch relations [see [10]]:

$$\tilde{g}_n(t_0) \leq a \Leftrightarrow \tilde{S}_n(a) \leq t_0; \quad (4.6)$$

$$\hat{g}_n(t_0) \leq a \Leftrightarrow \hat{S}_n(a) \leq t_0. \quad (4.7)$$

By (4.6) and (4.7), the convergence of (\tilde{g}_n, \hat{g}_n) and (\tilde{S}_n, \hat{S}_n) can be related by the following argmax continuous mapping theorem. Let \mathbb{B}_{loc} be the space of all locally bounded real functions on \mathbb{R} endowed with the topology of uniform convergence on compacta. That is, for $h_n, h \in \mathbb{B}_{loc}(\mathbb{R})$, h_n converges to h if for every $M > 0$, $\sup_{t \in [-M, M]} |h_n(t) - h(t)| \rightarrow 0$ as $n \rightarrow \infty$. Let $\mathbb{C}_{max}(\mathbb{R})$ denote the (separable) subset of continuous function x in $\mathbb{B}_{loc}(\mathbb{R})$ which satisfies $x(t) \rightarrow -\infty$ as $|t| \rightarrow \infty$, and x achieves its maximum at a unique point in \mathbb{R} .

Proposition 4.2 (Theorem 6.1 in [14]). *Let (J_{1n}, J_{2n}) be a sequence of a pair of random mappings valued in $\mathbb{B}_{loc}(\mathbb{R}) \times \mathbb{B}_{loc}(\mathbb{R})$ and (T_{1n}, T_{2n}) be another sequence of random mappings into $\mathbb{R} \times \mathbb{R}$ such that:*

- (i) $(J_{1n}, J_{2n}) \xrightarrow{d} (J_1, J_2)$, $\mathbb{P}((J_1, J_2) \in \mathbb{C}_{max}(\mathbb{R}) \times \mathbb{C}_{max}(\mathbb{R})) = 1$;
- (ii) $T_{1n}, T_{2n} = O_p(1)$;
- (iii) $J_{1n}(T_{1n}) \geq \sup_t J_{1n}(t) - \alpha_n$, and $J_{2n}(T_{2n}) \geq \sup_t J_{2n}(t) - \beta_n$ where $\alpha_n, \beta_n = o_p(1)$.

Then $(T_{1n}, T_{2n}) \xrightarrow{d} (T_1, T_2) \triangleq (\arg \max(J_1), \arg \max(J_2))$.

The remaining gap is to show the weak convergence of $(\tilde{Z}_n(t, a), \hat{Z}_n(t, b))$, which is closely related to that weak convergence of \tilde{U}_n and \hat{U}_{n, \hat{g}_n} which is given in the following proposition:

Proposition 4.3. *Under Assumption 1, $\sup_{t \in [-K, K]} |\tilde{U}_n(t) - \hat{U}_{n, \hat{g}_n}(t)| \xrightarrow{\mathbb{P}} 0$ and both \tilde{U}_n and \hat{U}_{n, \hat{g}_n} converge weakly in $\mathbb{B}_{loc}(\mathbb{R})$ to the process U defined by*

$$U(t) \triangleq \frac{1}{\lambda^{1/2}} W(t) + \frac{g'(t_0)}{2} t^2.$$

The proof of Proposition 4.3 will require an additional lemma:

Lemma 4.4. *Under Assumption 1 (ii), for any measurable set A , we have*

$$\mathbb{G}_n(A) = \int_A \left[\sum_{k=1}^s \lambda_{nk} \frac{w_k(y)}{W_k} \right]^{-1} d\mathbb{F}_n(y) + O_p(n^{-1/2}) \int_A d\mathbb{F}_n(y),$$

where $O_p(n^{-1/2})$ is independent of the choice of A .

and the proof of Lemma 4.4 will require an additional result:

Lemma 4.5. *Under Assumption 1 (ii), we have*

$$\int_a^b \left[\sum_{k=1}^s \lambda_{nk} w_k(y) \mathbb{V}_{nk}^{-1} \right]^{-1} d\mathbb{F}_n(y) = \frac{1}{W_s} + O_p(n^{-1/2}).$$

An interpretation of these lemmas are as follows. From Proposition 2.3 in [8], we know that $\sqrt{n}(\mathbb{V}_n - V_0) = O_p(1)$. Using this result, in Lemma 4.5, we shall show that the denominator of \mathbb{G}_n is essentially W_s^{-1} . As a result, we can write \mathbb{G}_n as a sum of two terms, where the first involves only the weighted empirical distribution \mathbb{F}_n and the true quantity W_k instead of the estimator \mathbb{V}_n , and the second term is an error term (see Lemma 4.4). This more convenient expression of \mathbb{G}_n will facilitate the proof of Proposition 4.3. In addition to Proposition 4.3, the proof of Theorem 4.1 will require the tightness of the inverse processes \tilde{S}_n and \hat{S}_n which will be given in Section 5 and further results on the process $\mathcal{S}_{\alpha,\beta}(x) \triangleq \arg \max_t \{\alpha W(t) + \beta t^2 - xt\}$ for any $\alpha \in \mathbb{R}$ and $\beta < 0$:

Lemma 4.6. $\mathcal{S}_{\alpha,\beta}(x) \stackrel{d}{=} \mathcal{S}_{\alpha,\beta}(0) + \frac{x}{2\beta}$.

Lemma 4.7. $\mathcal{S}_{\alpha,\beta}(0) = \alpha^{2/3}|\beta|^{-2/3}\mathbb{Y}$.

The results in Lemmas 4.6 and 4.7 are known but proofs of them are given in the Appendix for completeness.

Since the results are nested, we will show the proofs in the following order: Lemma 4.5, Lemma 4.4, Proposition 4.3, then finally Theorem 4.1. Denote $\bar{F}_n \triangleq \sum_{i=1}^s \lambda_{ni} F_i$.

Proof of Lemma 4.5. By telescoping the terms,

$$\int_a^b \left[\sum_{k=1}^s \lambda_{nk} w_k(y) \mathbb{V}_{nk}^{-1} \right]^{-1} d\mathbb{F}_n(y) = C_1 + C_2 + C_3,$$

where

$$\begin{aligned} C_1 &\triangleq \int_a^b \left[\sum_{k=1}^s \lambda_{nk} w_k(y) V_{0k}^{-1} \right]^{-1} d\bar{F}_n(y), \\ C_2 &\triangleq \int_a^b \left[\sum_{k=1}^s \lambda_{nk} w_k(y) V_{0k}^{-1} \right]^{-1} d(\mathbb{F}_n(y) - \bar{F}_n(y)), \\ C_3 &\triangleq \int_a^b \left\{ \left[\sum_{k=1}^s \lambda_{nk} w_k(y) \mathbb{V}_{nk}^{-1} \right]^{-1} - \left[\sum_{k=1}^s \lambda_{nk} w_k(y) V_{0k}^{-1} \right]^{-1} \right\} d\mathbb{F}_n(y). \end{aligned}$$

Simple algebra gives that

$$C_1 = \sum_{i=1}^s \frac{n_i}{n} \int_a^b \frac{f_i(y) dy}{\sum_{k=1}^s \frac{n_k w_k(y)}{n} W_k} = \int_a^b \frac{g_0(y) \sum_{i=1}^s \frac{n_i w_i(y)}{n} dy}{\sum_{k=1}^s \frac{n_k w_k(y)}{n} W_k} = \frac{1}{W_s}.$$

Note the class of functions involved in C_2 is, for large n , a subset of

$$\mathcal{H} \triangleq \left\{ \left[\sum_{k=1}^s h_k(y) w_k(y) V_{0k}^{-1} \right]^{-1} : h_k(y) \equiv r_k, r_k \in \left[\frac{\lambda_k}{2}, 2\lambda_k \right] \right\}.$$

Define also

$$\mathcal{H}^{-1} \triangleq \left\{ \sum_{k=1}^s h_k(y) w_k(y) V_{0k}^{-1} : h_k(y) \equiv r_k, r_k \in \left[\frac{\lambda_k}{2}, 2\lambda_k \right] \right\}.$$

Note that \mathcal{H}^{-1} is a Donsker class as each function in \mathcal{H}^{-1} is a finite sum of products of bounded functions from Donsker classes. That \mathcal{H} is also a Donsker class follows as the functions in \mathcal{H}^{-1} are bounded away from zero. Therefore, we have $C_2 = O_p(n^{-1/2})$. For C_3 , using the bounds on w_i and the fact that $\sqrt{n}(\mathbb{V}_n - V_0) = O_p(1)$,

$$\begin{aligned} |C_3| &\leq \int_a^b \frac{\sum_{k=1}^s \lambda_{nk} w_k(y) |V_{0k}^{-1} - \mathbb{V}_{nk}^{-1}|}{\left[\sum_{k=1}^s \lambda_{nk} w_k(y) \mathbb{V}_{nk}^{-1} \right] \left[\sum_{k=1}^s \lambda_{nk} w_k(y) V_{0k}^{-1} \right]} d\mathbb{F}_n(y) \\ &\leq \frac{\sum_{k=1}^s \lambda_{nk} M |V_{0k}^{-1} - \mathbb{V}_{nk}^{-1}|}{\left[\sum_{k=1}^s \lambda_{nk} m \mathbb{V}_{nk}^{-1} \right] \left[\sum_{k=1}^s \lambda_{nk} m V_{0k}^{-1} \right]} \int_a^b d\mathbb{F}_n(y) = O_p(n^{-1/2}). \quad \square \end{aligned}$$

Proof of Lemma 4.4. By telescoping the terms, we can write

$$\mathbb{G}_n(A) = W_s \int_A \left[\sum_{k=1}^s \lambda_{nk} w_k(y) V_{0k}^{-1} \right]^{-1} d\mathbb{F}_n(y) + A_1 + A_2,$$

where

$$\begin{aligned} A_1 &\triangleq \frac{\int_A \left[\sum_{k=1}^s \lambda_{nk} w_k(y) V_{0k}^{-1} \right]^{-1} d\mathbb{F}_n(y)}{\int_a^b \left[\sum_{k=1}^s \lambda_{nk} w_k(y) \mathbb{V}_{nk}^{-1} \right]^{-1} d\mathbb{F}_n(y)} - W_s \int_A \left[\sum_{k=1}^s \lambda_{nk} w_k(y) V_{0k}^{-1} \right]^{-1} d\mathbb{F}_n(y), \\ A_2 &\triangleq \frac{\int_A \left[\sum_{k=1}^s \lambda_{nk} w_k(y) \mathbb{V}_{nk}^{-1} \right]^{-1} d\mathbb{F}_n(y)}{\int_a^b \left[\sum_{k=1}^s \lambda_{nk} w_k(y) \mathbb{V}_{nk}^{-1} \right]^{-1} d\mathbb{F}_n(y)} - \frac{\int_A \left[\sum_{k=1}^s \lambda_{nk} w_k(y) V_{0k}^{-1} \right]^{-1} d\mathbb{F}_n(y)}{\int_a^b \left[\sum_{k=1}^s \lambda_{nk} w_k(y) \mathbb{V}_{nk}^{-1} \right]^{-1} d\mathbb{F}_n(y)}. \end{aligned}$$

Clearly,

$$W_s \int_A \left[\sum_{k=1}^s \lambda_{nk} w_k(y) V_{0k}^{-1} \right]^{-1} d\mathbb{F}_n(y) = \int_A \left[\sum_{k=1}^s \lambda_{nk} \frac{w_k(y)}{W_k} \right]^{-1} d\mathbb{F}_n(y).$$

For A_1 , by Lemma 4.5,

$$\begin{aligned} |A_1| &= \int_A \left[\sum_{k=1}^s \lambda_{nk} w_k(y) V_{0k}^{-1} \right]^{-1} d\mathbb{F}_n(y) \cdot \left| W_s - \frac{1}{\frac{1}{W_s} + O_p(n^{-1/2})} \right| \\ &\leq O_p(n^{-1/2}) \frac{\int_A d\mathbb{F}_n(y)}{m \sum_{k=1}^s \lambda_{nk} V_{0k}^{-1}} = O_p(n^{-1/2}) \int_A d\mathbb{F}_n(y). \end{aligned}$$

For A_2 , by using the same argument and similar calculation leading to bounding C_3 in Lemma 4.5, we also have $A_2 = O_p(n^{-1/2}) \int_A d\mathbb{F}_n(y)$. \square

Proof of Proposition 4.3. Fix $K > 0$. It suffices to show that $\sup_{t \in [-K, K]} |\tilde{U}_n(t) - \hat{U}_{n, \hat{g}_n}(t)| \xrightarrow{\mathbb{P}} 0$. The second statement follows from the first statement and the fact that \hat{U}_{n, \hat{g}_n} converges weakly in $\mathbb{B}_{\text{loc}}(\mathbb{R})$ to U as shown in Lemma 6.5 in [3]. Consider $t \in [-K, K]$. Using Lemma 4.4, we have

$$\begin{aligned} & \mathbb{G}_n(t_0 + tn^{-1/3}) - \mathbb{G}_n(t_0) \\ &= \int_{t_0}^{t_0+tn^{-1/3}} \left[\sum_{k=1}^s \lambda_{nk} \frac{w_k(y)}{W_k} \right]^{-1} d\mathbb{F}_n(y) + O_p(n^{-1/2}) \int_{t_0}^{t_0+tn^{-1/3}} d\mathbb{F}_n(y) \\ &= \int_{t_0}^{t_0+tn^{-1/3}} \left[\sum_{k=1}^s \lambda_{nk} \frac{w_k(y)}{W_k} \right]^{-1} d\mathbb{F}_n(y) + O_p(n^{-5/6}), \end{aligned} \tag{4.8}$$

where the O_p term is independent of t since $\int_{t_0}^{t_0+tn^{-1/3}} d\mathbb{F}_n(y) = O_p(n^{-1/3})$ is independent of t in $[-K, K]$; indeed,

$$\begin{aligned} \int_{t_0}^{t_0+tn^{-1/3}} d\mathbb{F}_n(y) &= \int_{t_0}^{t_0+tn^{-1/3}} d(\mathbb{F}_n - \bar{F}_n)(y) + \int_{t_0}^{t_0+tn^{-1/3}} d\bar{F}_n(y) \\ &= \sum_{i=1}^s \frac{n_i}{n} \int_{t_0}^{t_0+tn^{-1/3}} d(\mathbb{F}_{i, n_i} - F_i)(y) \\ &\quad + \sum_{i=1}^s \frac{n_i}{n} [F_i(t_0 + tn^{-1/3}) - F_i(t_0)]. \end{aligned}$$

The first term is $O_p(n^{-1/2})$ independent of t as $\{1(x \leq t_0 + tn^{-1/3}) - 1(x \leq t_0) : n \in \mathbb{N}, t \in [-K, K]\}$ is a subset of a Donsker class. The second term is $O(n^{-1/3})$, independent of t , as

$$|F_i(t_0 + tn^{-1/3}) - F_i(t_0)| = |f_i(t_n)n^{-1/3}t| \leq \left| \frac{M^2}{m} n^{-1/3} K \right| = O(n^{-1/3}),$$

where t_n is between $t_0 + tn^{-1/3}$ and t_0 after applying mean value theorem. Hence, by (4.8),

$$\begin{aligned} \tilde{U}_n(t) &= n^{2/3} \left\{ \int_{t_0}^{t_0+tn^{-1/3}} \left[\sum_{k=1}^s \lambda_{nk} \frac{w_k(y)}{W_k} \right]^{-1} d\mathbb{F}_n(y) - g_0(t_0)tn^{-1/3} \right\} \\ &\quad + O_p(n^{-1/6}) \\ &= B_1(t) + B_2(t) + O_p(n^{-1/6}), \end{aligned}$$

where

$$B_1(t) \triangleq n^{2/3} \left\{ \int_{t_0}^{t_0+tn^{-1/3}} \left[\sum_{k=1}^s \lambda_{nk} \frac{w_k(y)}{W_k} \right]^{-1} d(\mathbb{F}_n - \bar{F}_n)(y) \right\}; \tag{4.9}$$

$$B_2(t) \triangleq n^{2/3} \left\{ \int_{t_0}^{t_0+tn^{-1/3}} \left[\sum_{k=1}^s \lambda_{nk} \frac{w_k(y)}{W_k} \right]^{-1} d\bar{F}_n(y) - g_0(t_0)tn^{-1/3} \right\}. \quad (4.10)$$

We first rewrite $B_1(t)$ as

$$B_1(t) = \sum_{i=1}^s \frac{n_i}{n} n^{1/2} \int p_{n,t}(y) d(\mathbb{F}_{i,n_i} - F_i)(y),$$

where

$$p_{n,t}(y) \triangleq n^{1/6} \left[\sum_{k=1}^s \lambda_{nk} \frac{w_k(y)}{W_k} \right]^{-1} [1(y \leq t_0 + tn^{-1/3}) - 1(y \leq t_0)].$$

On the other hand, from the proof of Lemma 6.5 in [3], we have

$$\hat{U}_{n,\hat{g}_n}(t) = A_1(t) + A_2(t) + A_3(t),$$

where

$$\begin{aligned} A_1(t) &\triangleq \sum_{i=1}^s \frac{n_i}{n} n^{1/2} \int q_{n,t}(y) d(\mathbb{F}_{i,n_i} - F_i)(y), \\ A_2(t) &\triangleq \frac{1}{\lambda} \sum_{i=1}^s \frac{n_i}{n} n^{2/3} \int_{t_0}^{t_0+tn^{-1/3}} \frac{g_0(y) - g_0(t_0)}{g_0^2(y)} dF_i(y), \\ A_3(t) &\triangleq \sum_{i=1}^s \frac{n_i}{n} n^{2/3} \int \frac{1}{g_0(y)} [1(y \leq t_0 + tn^{-1/3}) - 1(y \leq t_0)] dF_i(y) \\ &\quad - n^{2/3} \sum_{i=1}^s \frac{n_i}{n} \frac{\sum_{j=1}^n c_{ij} [1(y \leq t_0 + tn^{-1/3}) - 1(y \leq t_0)]}{\sum_{k=1}^n c_{ik} g_0(T_k)}, \\ q_{n,t}(y) &\triangleq n^{1/6} \frac{1}{\lambda} \left[\frac{2g_0(y) - g_0(t_0)}{g_0^2(y)} \right] [1(y \leq t_0 + tn^{-1/3}) - 1(y \leq t_0)], \end{aligned}$$

and $A_3(t) = o_p(1)$ which is independent of t .

- (i) Since $B_1(t) - A_1(t) = \sum_{i=1}^s \frac{n_i}{n} n^{1/2} \int (p_{n,t}(y) - q_{n,t}(y)) d(\mathbb{F}_{i,n_i} - F_i)(y)$, we shall show that, for each $i = 1, \dots, s$, $n^{1/2} \int (p_{n,t} - q_{n,t})(y) d(\mathbb{F}_{i,n_i} - F_i)(y)$ converges to a Gaussian process with covariance function $K_i(u, t) = 0$ on $l^\infty([-K, K])$, where $l^\infty(\tilde{T})$ denotes the space of all real-valued bounded functions on \tilde{T} equipped with the uniform norm. To this end, it suffices to show the validity of the three items in Condition (2.11.21) and the entropy integral condition in Theorem 2.11.22 in [26]. These can be checked in the same way as in the proof of Lemma 6.4 in [3] and are therefore omitted. We then see that, $n^{1/2} \int \{p_{n,t}(y) - q_{n,t}(y)\} d(\mathbb{F}_{i,n_i} - F_i)(y)$ for $i = 1, \dots, s$ is asymptotically tight in $l^\infty([-K, K])$ and converges in distribution to a Gaussian process with covariance function

$$K_i(u, t) = \lim_{n \rightarrow \infty} \{ \mathbb{E}_{F_i} [(p_{n,u} - q_{n,u})(p_{n,t} - q_{n,t})] \}$$

$$- \mathbb{E}_{F_i}(p_{n,u} - q_{n,u}) \mathbb{E}_{F_i}(p_{n,t} - q_{n,t})\}.$$

For $u > 0, t > 0$,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{E}_{F_i}[(p_{n,u} - q_{n,u})(p_{n,t} - q_{n,t})] \\ &= \lim_{n \rightarrow \infty} n^{1/3} \int_{t_0}^{t_0 + (u \wedge t)n^{-1/3}} (p_{n,u} - q_{n,u})(p_{n,t} - q_{n,t}) f_i(y) dy \\ &= (u \wedge t) \times \left\{ \left(\sum_{k=1}^s \lambda_k \frac{w_k(t_0)}{W_k} \right)^{-2} f_i(t_0) + \frac{f_i(t_0)}{\lambda^2 g_0^2(t_0)} \right. \\ & \quad \left. - 2 \left(\sum_{k=1}^s \lambda_k \frac{w_k(t_0)}{W_k} \right)^{-1} \frac{f_i(t_0)}{\lambda g_0(t_0)} \right\} \\ &= 0, \end{aligned}$$

as $\lambda g_0(t_0) = \sum_{k=1}^s \lambda_k \frac{w_k(t_0)}{W_k}$. Similarly, it is also straightforward to see that $\lim_{n \rightarrow \infty} \mathbb{E}_{F_i}(p_{u,n} - q_{u,n}) \mathbb{E}_{F_i}(p_{n,t} - q_{n,t}) = 0$. Similarly, when $u < 0$ and $t < 0$, $K_i(u, t) = 0$. When u and t are of opposite signs, $K_i(u, t)$ is also 0. In summary, $n^{1/2} \int (p_{n,t} - q_{n,t})(y) d(\mathbb{F}_{i,n_i} - F_i)(y)$ converges in $l^\infty([-K, K])$ to 0. By the independence of different samples, as an independent sum, $B_1 - A_1$ also converges in $l^\infty([-K, K])$ to 0.

(ii) Note that B_2 converges uniformly to $\frac{1}{2}g_0'(t_0)t^2$ on $[-K, K]$ as

$$\begin{aligned} B_2(t) &= n^{2/3} \left[\int_{t_0}^{t_0 + tn^{-1/3}} \frac{\sum_{i=1}^s \lambda_{ni} f_i(y)}{\sum_{k=1}^s \lambda_{nk} \frac{w_k(y)}{W_k}} dy - g_0(t_0)tn^{-1/3} \right] \\ &= n^{2/3} \left[\int_{t_0}^{t_0 + tn^{-1/3}} g_0(y) dy - g_0(t_0)tn^{-1/3} \right]. \end{aligned}$$

From the proof of Lemma 6.5 in [3], we know A_2 also converges uniformly to $\frac{1}{2}g_0'(t_0)t^2$. Therefore, $B_2 - A_2$ converges uniformly to 0 on $[-K, K]$.

Combining (i) and (ii), we have shown that $\sup_{t \in [-K, K]} |\tilde{U}_n(t) - \hat{U}_{n, \hat{g}_n}(t)| \xrightarrow{\mathbb{P}} 0$. \square

Proof of Theorem 4.1. We first show (4.4) and (4.5). Using the definition of \tilde{S}_n , the fact that adding or multiplying a constant will not affect the maximum of a process and the definition of \tilde{Z}_n , we obtain

$$\begin{aligned} & \{n^{1/3}(\tilde{S}_n(n^{-1/3}a + g(t_0)) - t_0) \leq 0\} \\ &= \left\{ \arg \max_t \{ \mathbb{G}_n(t_0 + tn^{-1/3}) - n^{-2/3}at - g_0(t_0)tn^{-1/3} \} \leq 0 \right\} \\ &= \left\{ \arg \max_t \{ n^{2/3}[\mathbb{G}_n(t_0 + tn^{-1/3}) - \mathbb{G}_n(t_0) - g_0(t_0)tn^{-1/3}] - at \} \leq 0 \right\} \\ &= \left\{ \arg \max_t \{ \tilde{Z}_n(t, a) \} \leq 0 \right\}. \end{aligned}$$

Similarly,

$$\begin{aligned}
& \{n^{1/3}(\hat{S}_n(n^{-1/3}a + g_0(t_0)) - t_0) \leq 0\} \\
= & \left\{ \arg \max_t \{U_{n,\hat{g}_n}(t_0 + tn^{-1/3}) - g_0(t_0)G_{n,\hat{g}_n}(t_0 + tn^{-1/3}) \right. \\
& \quad \left. - n^{-1/3}aG_{n,\hat{g}_n}(t_0 + tn^{-1/3})\} \leq 0 \right\} \\
= & \left\{ \arg \max_t \{U_{n,\hat{g}_n}(t_0 + tn^{-1/3}) - U_{n,\hat{g}_n}(t_0) - \right. \\
& \quad g_0(t_0)(G_{n,\hat{g}_n}(t_0 + tn^{-1/3}) - G_{n,\hat{g}_n}(t_0)) \\
& \quad \left. - n^{-1/3}a(G_{n,\hat{g}_n}(t_0 + tn^{-1/3}) - G_{n,\hat{g}_n}(t_0))\} \leq 0 \right\} \\
= & \left\{ \arg \max_t \{\hat{U}_{n,\hat{g}_n}(t) - a\hat{G}_{n,\hat{g}_n}(t)\} \leq 0 \right\} \\
= & \left\{ \arg \max_t \{\hat{Z}_n(t, a)\} \leq 0 \right\}.
\end{aligned}$$

Proposition 4.3 gives that $(\tilde{U}_n(t), \hat{U}_{n,\hat{g}_n}(t))$ converges to $(U(t), U(t))$ in $\mathbb{B}_{\text{loc}}(\mathbb{R}) \times \mathbb{B}_{\text{loc}}(\mathbb{R})$. Lemma 6.1 and Lemma 6.3 in [3] imply that $\sup_{t \in [-K, K]} |\hat{G}_{n,\hat{g}_n}(t) - t| \xrightarrow{\mathbb{P}} 0$. These together imply that $(\tilde{Z}_n(t, a), \hat{Z}_n(t, b))$ converges to $(U(t) - at, U(t) - bt)$ in $\mathbb{B}_{\text{loc}}(\mathbb{R}) \times \mathbb{B}_{\text{loc}}(\mathbb{R})$. We can then apply Proposition 4.2 with $(J_{1n}(t), J_{2n}(t)) = (\tilde{Z}_n(t, a), \hat{Z}_n(t, b))$, $J_1(t) = U(t) - at$, $J_2(t) = U(t) - bt$ and

$$\begin{aligned}
T_{1n}(a) &= n^{1/3}(\tilde{S}_n(n^{-1/3}a + g_0(t_0)) - t_0), \\
T_{2n}(b) &= n^{1/3}(\hat{S}_n(n^{-1/3}b + g_0(t_0)) - t_0),
\end{aligned}$$

where $T_{1n}(a) = O_p(1)$ and $T_{2n}(b) = O_p(1)$ to be shown in Lemmas 5.1 and 5.2, respectively. This gives

$$(T_{1n}(a), T_{2n}(b)) \xrightarrow{d} (\mathcal{S}_{\alpha,\beta}(a), \mathcal{S}_{\alpha,\beta}(b)), \quad (4.11)$$

with $\alpha = \frac{1}{\lambda^{1/2}}$ and $\beta = \frac{g'_0(t_0)}{2}$ in the definition of $\mathcal{S}_{\alpha,\beta}$. Using (4.6), (4.7), (4.11) and Lemma 4.6,

$$\begin{aligned}
& \mathbb{P}(n^{1/3}(\tilde{g}_n(t_0) - g_0(t_0)) \leq a, n^{1/3}(\hat{g}_n(t_0) - g_0(t_0)) \leq b) \\
= & \mathbb{P}(n^{1/3}(\tilde{S}_n(n^{-1/3}a + g_0(t_0)) - t_0) \leq 0, \\
& \quad n^{1/3}(\hat{S}_n(n^{-1/3}b + g_0(t_0)) - t_0) \leq 0)
\end{aligned} \quad (4.12)$$

$$\begin{aligned}
& \rightarrow \mathbb{P}(\mathcal{S}_{\alpha,\beta}(a) \leq 0, \mathcal{S}_{\alpha,\beta}(b) \leq 0) \\
= & \mathbb{P}(-2\beta\mathcal{S}_{\alpha,\beta}(0) \leq a, -2\beta\mathcal{S}_{\alpha,\beta}(0) \leq b), \quad \text{for any } a, b.
\end{aligned} \quad (4.13)$$

Hence,

$$(n^{1/3}(\tilde{g}_n(t_0) - g_0(t_0)), n^{1/3}(\hat{g}_n(t_0) - g_0(t_0))) \xrightarrow{d} 2|\beta|(\mathcal{S}_{\alpha,\beta}(0), \mathcal{S}_{\alpha,\beta}(0)).$$

Therefore,

$$n^{1/3}(\tilde{g}_n(t_0) - \hat{g}_n(t_0)) \xrightarrow{d} 0,$$

which implies that $n^{1/3}(\tilde{g}_n(t_0) - \hat{g}_n(t_0)) \xrightarrow{\mathbb{P}} 0$. Finally, using Lemma 4.7, we have

$$n^{1/3}(\tilde{g}_n(t_0) - g_0(t_0)) \xrightarrow{d} 2|\beta|\mathcal{S}_{\alpha,\beta}(0) = 2\alpha^{2/3}|\beta|^{1/3}\mathbb{Y} = \lambda^{-1/3}\left|\frac{g'_0(t_0)}{2}\right|^{1/3}2\mathbb{Y},$$

which is equivalent to (4.1). □

5. Tightness of the inverse processes \tilde{S}_n and \hat{S}_n

To close the final gap in the proof of Theorem 4.1, we show in this section the tightness of the inverse processes \tilde{S}_n and \hat{S}_n in Lemmas 5.1 and 5.2.

Lemma 5.1. *Under Assumption 1 (i)–(ii), for all $\varepsilon > 0$ and $M_1 > 0$, there is an $M_2 > 0$ such that*

$$\mathbb{P}\left(\max_{|a|\leq M_1} n^{1/3}\left|\tilde{S}_n(n^{-1/3}a + g_0(t_0)) - t_0\right| > M_2\right) < \varepsilon.$$

Lemma 5.2. *Under Assumption 1, for all $\varepsilon > 0$ and $M_1 > 0$, there is an $M_2 > 0$ such that*

$$\mathbb{P}\left(\max_{|a|\leq M_1} n^{1/3}\left|\hat{S}_n(n^{-1/3}a + g_0(t_0)) - t_0\right| \geq M_2\right) < \varepsilon.$$

The proof of Lemma 5.1 is similar to the proof of Lemma 5.3 in [12] (see also Lemma 7.1 in [14]), we therefore put it in the Appendix. The proof of Lemma 5.2, however, employs a different argument than those in the literature, by making the direct uses of the switch relation and the analytic properties of \hat{g}_n through its accompanied Karush-Kuhn-Tucker condition, without the need to study the geometric relationship between the points $\{G_{n,\hat{g}_n}(T_i), U_{n,\hat{g}_n}(t)\}$ and its least concave majorant. In particular, we will first show the following result:

Lemma 5.3. *Under Assumption 1, for any $\varepsilon > 0$, there exists $C_0 > 0$ such that for all $C \geq C_0$ and all large n ,*

$$\mathbb{P}(\hat{g}_n(t_0 + 2Cn^{-1/3}) \leq g_0(t_0 + Cn^{-1/3})) \geq 1 - \varepsilon.$$

To prove Lemma 5.3, we make use of Lemma 5.4 which assures that a certain event related to the Karush-Kuhn-Tucker condition happens with a small probability.

Lemma 5.4. *Under Assumption 1, for any $\varepsilon > 0$, there exists $C_0 > 0$ and $R_0 > 0$ such that for any $C \geq C_0$ and $0 < R \leq R_0$, we have for all sufficiently large n ,*

$$\mathbb{P}\left(\inf_{t \in I_n} \sum_{j:t_0 + Cn^{-1/3} \leq T_j < t} \left(-\frac{1}{ng_0(t_0 + Cn^{-1/3}n)} + \sum_{i=1}^s \frac{n_i}{n} \frac{c_{ij}}{f w_i \hat{g}_n}\right) \leq 0\right) \leq \varepsilon,$$

where $I_n \triangleq [t_0 + 2Cn^{-1/3}, t_0 + R)$.

Lemma 5.4 is similar to Lemma 5.11 and Lemma 5.12 in [3], but now on a different shifted interval. The proofs of Lemmas 5.3 and 5.4 will be given in the appendix.

Proof of Lemma 5.2. Fix $\varepsilon > 0$ and $M_1 > 0$. We shall show only that there exists $M_2 > 0$ such that

$$\mathbb{P}\left(\max_{|a| \leq M_1} n^{1/3} \left(\hat{S}_n(n^{-1/3}a + g_0(t_0)) - t_0 \right) \geq M_2\right) < \varepsilon,$$

as another case can be established similarly. By observing that $\hat{S}_n(a)$ is decreasing in a , we have

$$\begin{aligned} & \mathbb{P}\left(\max_{|a| \leq M_1} n^{1/3} \left(\hat{S}_n(n^{-1/3}a + g_0(t_0)) - t_0 \right) \geq M_2\right) \\ &= \mathbb{P}(n^{1/3}(\hat{S}_n(-n^{-1/3}M_1 + g_0(t_0)) - t_0) \geq M_2). \end{aligned}$$

Simple algebra and the switch relation (4.7) give that

$$\begin{aligned} & \mathbb{P}(n^{1/3}(\hat{S}_n(-n^{-1/3}M_1 + g_0(t_0)) - t_0) \geq M_2) \\ &= \mathbb{P}(\hat{S}_n(-n^{-1/3}M_1 + g_0(t_0)) \geq n^{-1/3}M_2 + t_0) \\ &= \mathbb{P}(\hat{g}_n(n^{-1/3}M_2 + t_0) \geq -n^{-1/3}M_1 + g_0(t_0)) \\ &= \mathbb{P}(n^{1/3}(g_0(t_0) - \hat{g}_n(n^{-1/3}M_2 + t_0)) \leq M_1). \end{aligned}$$

We claim in Lemma 5.3 that exists $C_0 > 0$ such that for all $C \geq C_0$ and all large n ,

$$\mathbb{P}(\hat{g}_n(t_0 + 2Cn^{-1/3}) \leq g_0(t_0 + Cn^{-1/3})) \geq 1 - \varepsilon.$$

By definition of $g'_0(t_0)$, there exists $\delta > 0$ such that for all $0 < u < \delta$,

$$\left| \frac{g_0(t_0 + u) - g_0(t_0)}{u} - g'_0(t_0) \right| < \frac{1}{2}g'_0(t_0).$$

This implies that

$$\frac{g_0(t_0 + u) - g_0(t_0)}{u} < \frac{1}{2}g'_0(t_0).$$

Now, we choose $M_2 = 2C$, where $C \geq \max\{C_0, -\frac{2M_1}{g'_0(t_0)}\}$.

$$\begin{aligned} & \mathbb{P}(n^{1/3}(g_0(t_0) - \hat{g}_n(2Cn^{-1/3} + t_0)) \leq M_1) \\ &\leq \mathbb{P}(n^{1/3}(g_0(t_0) - \hat{g}_n(2Cn^{-1/3} + t_0)) \leq M_1, \\ &\quad \hat{g}_n(t_0 + 2Cn^{-1/3}) \leq g_0(t_0 + Cn^{-1/3})) \\ &\quad + \mathbb{P}(\hat{g}_n(t_0 + 2Cn^{-1/3}) > g_0(t_0 + Cn^{-1/3})) \\ &\leq \mathbb{P}(n^{1/3}(g_0(t_0) - g_0(t_0 + Cn^{-1/3})) \leq M_1) + \varepsilon. \end{aligned}$$

Note that for all large enough n , we have $Cn^{-1/3} < \delta$ and so

$$n^{1/3}(g_0(t_0) - g_0(t_0 + Cn^{-1/3})) = -C \frac{g_0(t_0 + Cn^{-1/3}) - g_0(t_0)}{Cn^{-1/3}}$$

$$> -\frac{C}{2}g_0'(t_0) \geq M_1,$$

by the choice of C . Therefore, $\mathbb{P}(n^{1/3}(g_0(t_0) - g_0(t_0 + Cn^{-1/3})) \leq M_1) = 0$ for all large enough n and so the result of the lemma follows. \square

6. Numerical studies

6.1. Simulations

In this section, we illustrate the finite sample performance of the proposed monotone MLE and compare it with the Grenander-type estimator. For the Grenander-type estimator, we first obtained the NPML estimator \mathbb{G}_n for the unbiased distribution function G_0 [27, 28]. The Grenander-type estimator \tilde{g}_n is then obtained as the slope of the least concave majorant of \mathbb{G}_n . To obtain the monotone MLE via the self-characterization introduced in [3], a set of initial guesses for pointwise density values is required. Given the initial guesses, say $\hat{z}^{(0)}$, an updated set of estimates, denoted as $\hat{z}^{(1)}$, is defined as the solution of the right hand side of (3.5) of [3]. These updated values $\hat{z}^{(1)}$ will then serve as the initial values for the next iteration and the procedure will continue iteratively until convergence.

Our simulation studies analyzed the performance of (i) \tilde{g}_n , (ii) $\hat{g}_{n,V}$, the monotone MLE based on \tilde{g}_n as the initial guess, and (iii) $\hat{g}_{n,R}$, the monotone MLE based on the density estimated from randomly drawn samples as the initial value, denoted by $\hat{g}_{n,R}^{(0)}$. For the two-sample setting considered, we simulated n_1 exponentially-distributed samples with rate 0.5, which represented the true unbiased distribution, i.e. $w_1(t) = 1$. In addition, we also generated $n_2 = n_1$ samples from the length-biased variation of the unbiased distribution, i.e. the weight function is set to be $w_2(t) = t$. Define $n \triangleq n_1 + n_2$, we generated samples with $n = 100, 200, 500$ and 1000 with a balanced design. 500 iterations were carried out for each of the simulation exercises.

As we can see from the Table 1, the MLE procedure, regardless of which initial values chosen, produces virtually unbiased estimates whose standard errors are smaller than its counterpart estimated via the Grenander-type approach. This is sensible as the maximum likelihood procedure should give the most efficient estimates. Our numerical experience also suggests that the convergence of the self-characterization can be achieved after three to five iterations. Figure 1 demonstrates the initial values, the maximum likelihood estimator obtained after four iterations and the true density in one simulation scenario.

We also conducted simulation results to examine the validity of Theorem 4.1. We simulated samples from each of the following two distributions, namely (a) an unbiased exponential distribution ($f_1 = g$) with mean $\eta = 2$ and (b) a length biased exponential distribution f_2 whose density can be written as $f_2(x) = \{xg(x; \eta)\} \left\{ \int_0^\infty \xi g(\xi; \eta) d\xi \right\}^{-1}$. In such a setting, $n_1 = n_2$ (i.e. $\lambda_1 = \lambda_2 = 0.5$). Different sample sizes were generated to examine the corresponding

TABLE 1
Biases and standard errors of the unbiased density estimates \tilde{g}_n , $\hat{g}_{n,R}$ and $\hat{g}_{n,V}$

$n = 100$	$x = 0.25$	$x = 0.50$	$x = 0.75$	$x = 1.00$	$x = 1.25$	$x = 1.50$
$\tilde{g}_n(x)$	-0.046 (0.222)	-0.014 (0.149)	-0.010 (0.105)	-0.002 (0.067)	0.000 (0.046)	0.005 (0.034)
$\hat{g}_{n,R}(x)$	-0.015 (0.198)	0.003 (0.134)	-0.001 (0.095)	0.005 (0.062)	0.004 (0.043)	0.007 (0.032)
$\hat{g}_{n,V}(x)$	0.010 (0.199)	0.007 (0.135)	-0.001 (0.096)	0.004 (0.062)	0.003 (0.043)	0.006 (0.032)
$n = 250$	$x = 0.25$	$x = 0.50$	$x = 0.75$	$x = 1.00$	$x = 1.25$	$x = 1.50$
$\tilde{g}_n(x)$	-0.035 (0.159)	-0.014 (0.109)	-0.005 (0.076)	0.003 (0.050)	0.001 (0.034)	0.002 (0.023)
$\hat{g}_{n,R}(x)$	-0.020 (0.144)	-0.006 (0.100)	-0.001 (0.071)	0.006 (0.048)	0.003 (0.032)	0.004 (0.022)
$\hat{g}_{n,V}(x)$	-0.014 (0.144)	-0.005 (0.101)	-0.002 (0.071)	0.005 (0.048)	0.002 (0.032)	0.003 (0.022)
$n = 500$	$x = 0.25$	$x = 0.50$	$x = 0.75$	$x = 1.00$	$x = 1.25$	$x = 1.50$
$\tilde{g}_n(x)$	-0.026 (0.128)	-0.004 (0.082)	-0.005 (0.057)	0.001 (0.037)	-0.002 (0.025)	0.001 (0.018)
$\hat{g}_{n,R}(x)$	-0.019 (0.120)	0.000 (0.078)	-0.003 (0.054)	0.002 (0.036)	-0.001 (0.025)	0.001 (0.017)
$\hat{g}_{n,V}(x)$	-0.014 (0.120)	0.000 (0.078)	-0.004 (0.054)	0.001 (0.036)	-0.002 (0.025)	0.001 (0.017)
$n = 1000$	$x = 0.25$	$x = 0.50$	$x = 0.75$	$x = 1.00$	$x = 1.25$	$x = 1.50$
$\tilde{g}_n(x)$	-0.001 (0.097)	-0.006 (0.067)	-0.005 (0.046)	-0.001 (0.032)	-0.001 (0.020)	0.000 (0.014)
$\hat{g}_{n,R}(x)$	0.005 (0.093)	-0.003 (0.064)	-0.003 (0.045)	0.000 (0.032)	-0.001 (0.020)	0.000 (0.013)
$\hat{g}_{n,V}(x)$	0.007 (0.093)	-0.003 (0.064)	-0.004 (0.045)	-0.001 (0.031)	-0.001 (0.020)	0.000 (0.013)

finite sample performances. To verify the convergence properties of Theorem 4.1, namely (4.1), we also calculated the Kolmogorov-Smirnov (KS) test statistics for two-sided tests between the distribution of Δ_n evaluated at $t_0 = 0.5$ and $2\mathbb{Y}$ based on 500 repetitions with different sample sizes, where Δ_n denotes the LHS of (4.1); see Table 2. All the KS tests are not rejected at $\alpha = 0.05$.

TABLE 2
 p -value of two-sided KS test between Δ_n and $2\mathbb{Y}$ for $t_0 = 0.5$, $n = \{100, 250, 500, 1000\}$

n	p -value
100	0.062
250	0.536
500	0.181
1000	0.263

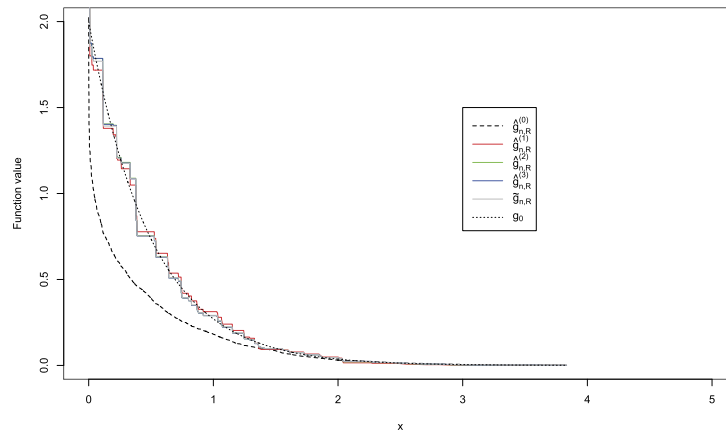


FIG 1. Estimation of the proposed monotone MLE after three iterations $\hat{g}_{n,R}^{(3)}(\cdot)$.

6.2. Data analysis

We apply our methodology to a real-life application. As discussed in [23] and [4], the blood alcohol concentration (BAC) of drivers involved in fatal car accidents in the US demonstrate size bias between younger (< 30 years old) and older (≥ 30 years old) population. Drivers with higher BACs usually incur higher chances of being involved in fatal traffic accidents.

We illustrate using a dataset obtained from the National Highway Traffic Safety Administration Department of Transportation (NHTSA) in the United States. In particular, the Fatality Analysis Reporting System (FARS) provides a collection of raw statistics recording all the qualifying fatal car crashes that occurred with the 50 states and the District of Columbia. The FARS database (<https://www.nhtsa.gov/crash-data-systems/fatality-analysis-reporting-system>) stores three sections of incidents including the accident, the vehicle and person files in which our variable of interest, namely the BAC, measured in grams /decilitre (g/dL) and the ages of the associated drivers can be found. We focus our analysis to whole blood test results valued at or above 0.08 g/dL, the legal limit to define driving under influence after 2004, using drivers involved in accidents in all 50 states during 2009. The total number of samples considered in this analysis is 5,385 in which 3,086 (57.3%) of them were aged 30 or above.

Since the drivers with a higher BAC are more likely to be involved in an accident, the observed BAC in FARS data are biased towards larger values. To study the distribution of BAC among all drivers, [4] and [23] showed that it is reasonable to use the biasing functions $w_1(x) = \sqrt{x}$ for younger drivers (age < 30) and $w_2(x) = x$ for older drivers (age ≥ 30). It is also reasonable to assume that BAC density is decreasing past the legal limit (BAC ≥ 0.08) since a high BAC is potentially lethal. The estimates for the population density

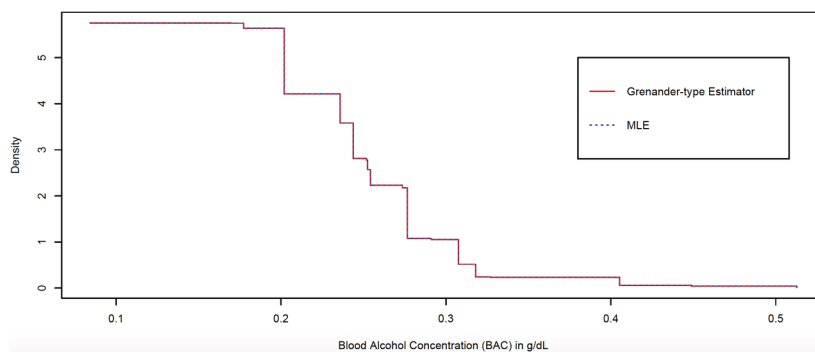


FIG 2. Estimated density of blood alcohol concentration via the Grenander-type estimator and the MLE.

of BAC conditional on $\text{BAC} \geq 0.08$ is given in Figure 2. The Grenander-type estimator and the MLE gave almost identical results.

7. Discussion

This paper establishes the asymptotic equivalence between the nonparametric maximum likelihood estimate (NPMLE) of the decreasing density function and the Grenander-type estimator without any shape constraint under the s -sample biased sampling models considered in [3]. In particular, we show that \tilde{g}_n and \hat{g}_n in s -sample biased sampling models are asymptotically equivalent in the sense that $n^{1/3}[\tilde{g}_n(t_0) - \hat{g}_n(t_0)] \xrightarrow{\mathbb{P}} 0$ as $n \rightarrow \infty$. Instead of adopting the traditional approach of showing the tightness of the corresponding inverse processes, we develop a new technique of direct use of the switch relation for the NPMLE case. One possible extension is to investigate the corresponding inference procedure when the biasing functions $w_i(\cdot)$, $i = 1, \dots, s$ contain unknown parameters. We shall explore this matter in the future.

A reviewer suggested to consider the absolute difference between \hat{g}_n and \tilde{g}_n . While the absolute difference between the Grenander estimator and true density is studied in [6], such results are not readily extendable to study the absolute difference between \hat{g}_n and \tilde{g}_n . Moreover, as the Grenander estimator is known to be inconsistent at the boundaries of the support, it is impossible for $\sup_{t \in [a, b]} |\hat{g}_n(t) - g_0(t)| = o_p(1)$. However, it is possible that $\sup_{t \in [a, b]} |\hat{g}_n(t) - \tilde{g}_n(t)| = o_p(1)$. To obtain a rate of convergence of 0 of $|\hat{g}_n(t) - \tilde{g}_n(t)|$ for t in a neighborhood of 0 and 1, additional analysis on \hat{g}_n and \tilde{g}_n near the boundaries are required. This is possibly related to [17], where they have established some results for the Grenander estimator near the boundaries of the support. We expect that the technique of proofs developed in [17] and [6] together with the results in this paper and [3] are required to tackle this problem.

Appendix A: Proofs of auxiliary lemmas

Proof of Lemma 4.6. Using the fact that adding a constant will not affect the location of the maximum of a process and the stationarity of Brownian motion,

$$\begin{aligned} \mathcal{S}_{\alpha,\beta}(x) - \frac{x}{2\beta} &= \arg \max_t \left\{ \alpha W\left(t + \frac{x}{2\beta}\right) + \beta\left(t + \frac{x}{2\beta}\right)^2 - x\left(t + \frac{x}{2\beta}\right) \right\} \\ &= \arg \max_t \left\{ \alpha W\left(t + \frac{x}{2\beta}\right) - \alpha W\left(\frac{x}{2\beta}\right) + \beta t^2 \right\} \\ &\stackrel{d}{=} \arg \max_t \{ \alpha W(t) + \beta t^2 \} = \mathcal{S}_{\alpha,\beta}(0). \end{aligned}$$

□

Proof of Lemma 4.7. Let δ and γ be real constants. Using the scaling property of Brownian motion, we have

$$\begin{aligned} \mathcal{S}_{\alpha,\beta}(0) &= \alpha^{-\delta} |\beta|^{-\gamma} \arg \max_t \{ W(\alpha^{2-\delta} |\beta|^{-\gamma} t) + \beta(\alpha^{-\delta} |\beta|^{-\gamma} t)^2 \} \\ &\stackrel{d}{=} \alpha^{-\delta} |\beta|^{-\gamma} \arg \max_t \{ \alpha^{1-\delta/2} |\beta|^{-\gamma/2} W(t) - \alpha^{-2\delta} |\beta|^{1-2\gamma} t^2 \}. \end{aligned}$$

Therefore, if $1 - \delta/2 = -2\delta$ and $-\gamma/2 = 1 - 2\gamma$, we have $\mathcal{S}_{\alpha,\beta}(0) \stackrel{d}{=} \alpha^{-\delta} |\beta|^{-\gamma} \Upsilon$ as multiplying a constant will not affect the location of the maximum of a process. Finally, the two equations imply that $\delta = -2/3$ and $\gamma = 2/3$. □

Proof of Lemma 5.1. Fix $\varepsilon > 0$. We shall show only that

$$\mathbb{P}\left(\max_{|a| \leq M_1} n^{1/3} \left(\tilde{S}_n(n^{-1/3}a + g_0(t_0)) - t_0 \right) > M_2\right) < \varepsilon,$$

as the other case can be proved similarly. By observing that $\tilde{S}_n(a)$ is decreasing in a , the fact that $n^{1/3}(\tilde{S}_n(-n^{-1/3}M_1 + g_0(t_0)) - t_0) = \arg \max_t \{ \tilde{Z}_n(t, -M_1) \}$ and $\tilde{Z}_n(0, -M_1) = 0$, we have

$$\begin{aligned} &\mathbb{P}\left(\max_{|a| \leq M_1} n^{1/3} \left(\tilde{S}_n(n^{-1/3}a + g_0(t_0)) - t_0 \right) > M_2\right) \\ &= \mathbb{P}(n^{1/3}(\tilde{S}_n(-n^{-1/3}M_1 + g_0(t_0)) - t_0) > M_2) \\ &\leq \mathbb{P}(\tilde{Z}_n(t, -M_1) \geq 0 \text{ for some } t \geq M_2). \end{aligned}$$

Note that there exists $u_0 > 0$ such that for any $|y - t_0| < u_0$, $|\frac{g_0(y) - g_0(t_0)}{y - t_0} - g'_0(t_0)| < -g'_0(t_0)/4$ and so $\frac{g_0(y) - g_0(t_0)}{y - t_0} < \frac{3}{4}g'_0(t_0)$. Using Lemma 4.4, we have

$$\begin{aligned} &\tilde{Z}_n(t, -M_1) \\ &= n^{2/3}(\mathbb{G}_n(t_0 + tn^{-1/3}) - \mathbb{G}_n(t_0) - g_0(t_0)tn^{-1/3}) + M_1t \\ &= n^{2/3} \int_{t_0}^{t_0 + tn^{-1/3}} \left[\sum_{k=1}^s \lambda_{nk} \frac{w_k(y)}{W_k} \right]^{-1} d\mathbb{F}_n(y) \end{aligned}$$

$$\begin{aligned}
& +O_p(n^{1/6}) \int_{t_0}^{t_0+tn^{-1/3}} d\mathbb{F}_n(y) - g_0(t_0)tn^{1/3} + M_1t \\
= & B_1(t) + B_2(t) + O_p(n^{1/6}) \int_{t_0}^{t_0+tn^{-1/3}} d\bar{F}_n(y) \\
& +O_p(n^{1/6}) \int_{t_0}^{t_0+tn^{-1/3}} d(\mathbb{F}_n - \bar{F}_n)(y) + M_1t,
\end{aligned}$$

where $B_1(t)$ and $B_2(t)$ were defined in (4.9) and (4.10). Consider $n^{-1/3}t \leq u_0$. Then $B_2(t) < \frac{3}{4}g'_0(t_0)n^{2/3} \int_{t_0}^{t_0+tn^{-1/3}} (y - t_0)dy = \frac{3}{8}g'_0(t_0)t^2$. There exists $K > 0$ such that for any $t > K$,

$$\frac{3}{8}g'_0(t_0)t^2 + M_1t < \frac{1}{4}g'_0(t_0)t^2.$$

Using Lemma 4.1 in [16] for $\alpha = -\frac{1}{8s}g'_0(t_0) > 0$ and $n^{-1/3}t \leq u_0$, there exist s tight sequences of random variables $\{A_{i,n}\}_{n=1}^\infty$, $i = 1, \dots, s$ such that for each $i = 1, \dots, s$,

$$\left| \int_{t_0}^{t_0+tn^{-1/3}} \frac{n_i}{n} \left[\sum_{k=1}^s \lambda_{nk} \frac{w_k(y)}{W_k} \right]^{-1} d(\mathbb{F}_{i,n_i} - F_i)(y) \right| \leq \alpha n^{-2/3}t^2 + n^{-2/3}A_{i,n}^2.$$

Hence,

$$|B_1(t)| \leq -\frac{1}{8}g'_0(t_0)t^2 + \sum_{i=1}^s A_{i,n}^2.$$

Finally, it is easy to see that $\int_{t_0}^{t_0+tn^{-1/3}} d\bar{F}_n(y) = O(n^{1/3})t$ and $\int_{t_0}^{t_0+tn^{-1/3}} d(\mathbb{F}_n - \bar{F}_n)(y) = O_p(n^{-1/2})$ (see also the proof of Lemma 4.3). Therefore, for $t > K$, we have

$$\begin{aligned}
\tilde{Z}_n(t, -M_1) & < \frac{1}{4}g'_0(t_0)t^2 + O_p(n^{-1/6})t + O_p(n^{-1/3}) - \frac{1}{8}g'_0(t_0)t^2 + \sum_{i=1}^s A_{i,n}^2 \\
& = \frac{1}{8}g'_0(t_0)t^2 + O_p(n^{-1/6})t + O_p(n^{-1/3}) + \sum_{i=1}^s A_{i,n}^2,
\end{aligned}$$

where in the last line the O_p terms does not depend on t . Since $\sum_{j=1}^s A_{j,n}^2 = O_p(1)$, there exists $M_2 > K$ such that for $t > M_2$ and all sufficiently large n ,

$$\mathbb{P}\left(\frac{1}{8}g'_0(t_0)t^2 + O_p(n^{-1/6})t + O_p(n^{-1/3}) + \sum_{j=1}^s A_{j,n}^2 \geq 0\right) \leq \varepsilon/3.$$

Now, for this choice of M_2 and all sufficiently large n ,

$$\mathbb{P}(\tilde{Z}_n(t, -M_1) \geq 0 \text{ for some } M_2 \leq t \leq n^{1/3}u_0)$$

$$\leq \varepsilon/3. \tag{A.1}$$

For $n^{-1/3}t > u_0$, we claim that for all sufficiently large n , we have

$$\mathbb{P}\left(\tilde{Z}_n(t, -M_1) \leq \tilde{Z}_n\left(n^{1/3}\frac{u_0}{2}, -M_1\right)\right) = 1 - \varepsilon/3, \tag{A.2}$$

see the proof below. Then, for all sufficiently large n ,

$$\begin{aligned} & \mathbb{P}(\tilde{Z}_n(t, -M_1) \geq 0 \text{ for some } t > n^{1/3}u_0) \\ & \leq \mathbb{P}(\tilde{Z}_n(n^{1/3}\frac{u_0}{2}, -M_1) \geq 0) + \varepsilon/3 \\ & \leq \mathbb{P}(\tilde{Z}_n(t, -M_1) \geq 0 \text{ for some } M_2 \leq t \leq n^{1/3}u_0) + \varepsilon/3 \\ & \leq 2\varepsilon/3. \end{aligned} \tag{A.3}$$

Combining (A.1) and (A.3), we have

$$\mathbb{P}(\tilde{Z}_n(t, -M_1) \geq 0 \text{ for some } t \geq M_2) \leq \varepsilon.$$

To show (A.2), recall that Lemma 3.1 implies that $|\tilde{\mathbb{G}}_n(t_0 + \frac{u_0}{2}) - \mathbb{G}_n(t_0 + \frac{u_0}{2})| \xrightarrow{a.s.} 0$ and Lemma 3.3 gives $|\tilde{g}_n(t_0 + \frac{u_0}{2}) - g_0(t_0 + \frac{u_0}{2})| \xrightarrow{a.s.} 0$. Hence, for each sample point ω ,

$$\mathbb{G}_n\left(t_0 + \frac{u_0}{2}\right) = \tilde{\mathbb{G}}_n\left(t_0 + \frac{u_0}{2}\right) + o_\omega(1)$$

and

$$\tilde{g}_n\left(t_0 + \frac{u_0}{2}\right) = g_0\left(t_0 + \frac{u_0}{2}\right) + o_\omega(1).$$

Thus, by the definition of the least concave majorant and the concavity of $\tilde{\mathbb{G}}_n$,

$$\begin{aligned} & \mathbb{G}_n\left(t_0 + n^{-1/3}t\right) - \mathbb{G}_n\left(t_0 + \frac{u_0}{2}\right) \\ & \leq \tilde{\mathbb{G}}_n\left(t_0 + n^{-1/3}t\right) - \tilde{\mathbb{G}}_n\left(t_0 + \frac{u_0}{2}\right) + o_\omega(1) \\ & \leq \tilde{g}_n\left(t_0 + \frac{u_0}{2}\right)\left(n^{-1/3}t - \frac{u_0}{2}\right) + o_\omega(1) \\ & = \left[g_0\left(t_0 + \frac{u_0}{2}\right) + o_\omega(1)\right]\left(n^{-1/3}t - \frac{u_0}{2}\right) + o_\omega(1). \end{aligned} \tag{A.4}$$

For all sufficiently large n , using the definition of $\tilde{Z}_n(t, a)$,

$$\begin{aligned} & \tilde{Z}_n(t, -M_1) - \tilde{Z}_n(n^{1/3}u_0, -M_1) \\ & = n^{2/3}\left[\mathbb{G}_n(t_0 + n^{-1/3}t) - \mathbb{G}_n\left(t_0 + \frac{u_0}{2}\right) - g_0(t_0)tn^{-1/3} + g_0(t_0)\frac{u_0}{2}\right] \\ & \quad + M_1t - M_1n^{1/3}\frac{u_0}{2} \end{aligned}$$

$$\begin{aligned}
&= n^{2/3} \left[\mathbb{G}_n(t_0 + n^{-1/3}t) - \mathbb{G}_n\left(t_0 + \frac{u_0}{2}\right) - g_0(t_0) \left(tn^{-1/3} - \frac{u_0}{2} \right) \right. \\
&\quad \left. + M_1 n^{-1/3} \left(tn^{-1/3} - \frac{u_0}{2} \right) \right] \\
&= n^{2/3} \left[\mathbb{G}_n(t_0 + n^{-1/3}t) - \mathbb{G}_n\left(t_0 + \frac{u_0}{2}\right) \right. \\
&\quad \left. - (g_0(t_0) - M_1 n^{-1/3}) \left(tn^{-1/3} - \frac{u_0}{2} \right) \right] \\
&\leq n^{2/3} \left[\left(g_0(t_0 + u_0) - g_0(t_0) + o_\omega(1) + M_1 n^{-1/3} \right) \right. \\
&\quad \left. \times \left(n^{-1/3}t - \frac{u_0}{2} \right) + o_\omega(1) \right] \\
&\leq 0,
\end{aligned}$$

where the second last inequality follows from (A.4) and the last inequality follows from the fact that g_0 is strictly decreasing at t_0 . This implies that (A.2) holds. \square

Proof of Lemma 5.4. Let

$$J_i(t) \triangleq \sum_{j:t_0+Cn^{-1/3} \leq X_{ij} < t} \left(-\frac{1}{n_i g_0(t_0 + Cn^{-1/3}n)} \right) + \frac{\sum_{j:t_0+Cn^{-1/3} \leq T_j < t} c_{ij}}{\int w_i \hat{g}_n}.$$

Then

$$\sum_{j:t_0+Cn^{-1/3} \leq T_j < t} \left(-\frac{1}{n g_0(t_0 + Cn^{-1/3}n)} + \sum_{i=1}^s \frac{n_i}{n} \frac{c_{ij}}{\int w_i \hat{g}_n} \right) = \sum_{i=1}^s \frac{n_i}{n} J_i(t).$$

We first claim that $\mathbb{P}(\inf_{t \in I_n} J_i(t) \leq 0) \leq \varepsilon$ for all $i = 1, \dots, s$. The claim of our present lemma will then follow because

$$\begin{aligned}
\mathbb{P} \left(\inf_{t \in I_n} \sum_{i=1}^s \frac{n_i}{n} J_i(t) \leq 0 \right) &\leq \mathbb{P} \left(\sum_{i=1}^s \frac{n_i}{n} \inf_{t \in I_n} J_i(t) \leq 0 \right) \\
&\leq \sum_{i=1}^s \mathbb{P} \left(\inf_{t \in I_n} J_i(t) \leq 0 \right) \leq s\varepsilon.
\end{aligned}$$

Now, we verify the first claim. Define

$$\begin{aligned}
\mathcal{A}(t) &\triangleq - \int \frac{1}{g_0(t_0 + Cn^{-1/3})} I(t_0 + Cn^{-1/3} \leq x < t) dF_i(x) \\
&\quad + \frac{\int_{t_0+Cn^{-1/3}}^t w_i(x) dx}{W_i}, \\
\mathcal{B}(t) &\triangleq - \int \frac{1}{g_0(t_0 + Cn^{-1/3})} I(t_0 + Cn^{-1/3} \leq x < t) d(\mathbb{F}_{i,n_i} - F_i)(x),
\end{aligned}$$

$$\mathcal{C}(t) \triangleq \frac{\int_{T_L}^t w_i(x) dx}{\int w_i \hat{g}_n} - \frac{\int_{T_U}^{t_0 + Cn^{-1/3}} w_i(x) dx}{\int w_i \hat{g}_n},$$

$$\mathcal{D}(t) \triangleq \frac{\int_{t_0 + Cn^{-1/3}}^t w_i(x) dx}{\int w_i \hat{g}_n} - \frac{\int_{t_0 + Cn^{-1/3}}^t w_i(x) dx}{W_i},$$

where T_L and T_U denote the maxima of X_{ij} less than t and $t_0 + Cn^{-1/3}$ respectively. Clearly, $J_i(t) = \mathcal{A}(t) + \mathcal{B}(t) + \mathcal{C}(t) + \mathcal{D}(t)$.

(i) Note that there exists $R_1 > 0$ such that for all $0 < t - t_0 < R_1$,

$$\left| \frac{g_0(t) - g_0(t_0)}{t - t_0} - g'_0(t_0) \right| < -\frac{1}{2}g'_0(t_0).$$

Now, for $R < R_1$ and all sufficiently large n such that $t_0 + Cn^{-1/3} < R_1$,

$$\begin{aligned} & \mathcal{A}(t) \\ & \geq -\frac{m}{M^3} \int_{t_0 + Cn^{-1/3}}^t \frac{g_0(x) - g_0(t_0 + Cn^{-1/3})}{x - (t_0 + Cn^{-1/3})} [x - (t_0 + Cn^{-1/3})] dx \\ & > -\frac{m}{M^3} \frac{g'_0(t_0)}{2} \int_{t_0 + Cn^{-1/3}}^t [x - (t_0 + Cn^{-1/3})] dx \\ & = \alpha [t - (t_0 + Cn^{-1/3})]^2, \end{aligned}$$

where $\alpha \triangleq \frac{m}{M^3} \frac{|g'_0(t_0)|}{4}$, the first inequality follows from the bounds on w_i and g_0 and the second inequality follows as for $t_0 + Cn^{-1/3} \leq x < t_0 + R$,

$$\begin{aligned} \frac{g_0(x) - g_0(t_0 + Cn^{-1/3})}{x - (t_0 + Cn^{-1/3})} &= \frac{g_0(x) - g_0(t_0) + g_0(t_0) - g_0(t_0 + Cn^{-1/3})}{x - (t_0 + Cn^{-1/3})} \\ &> \frac{\frac{1}{2}g'_0(t_0)(x - t_0) - \frac{1}{2}g'_0(t_0)Cn^{-1/3}}{x - (t_0 + Cn^{-1/3})} \\ &= \frac{1}{2}g'_0(t_0). \end{aligned}$$

(ii) For $\mathcal{B}(t)$, by the same argument as used in the proof of Lemma 4.1 of [16], for any $\delta > 0$, there exists a tight sequence of random variables $\{M_n\}$ such that for any $t \in I_n$,

$$|\mathcal{B}(t)| \leq \delta [t - (t_0 + Cn^{-1/3})]^2 + n^{-2/3} M_n^2.$$

(iii) For $\mathcal{C}(t)$, it is clear that $\frac{\int_{T_L}^t w_i(x) dx}{\int w_i \hat{g}_n}$ and $\frac{\int_{T_U}^{t_0 + Cn^{-1/3}} w_i(x) dx}{\int w_i \hat{g}_n}$ are of order $O_p(n^{-1})$, which are also independent of choice of $t \in I_n$; see also the proof of Lemma 5.11 in [3].

(iv) For $\mathcal{D}(t)$, by Proposition 5.8 in [3],

$$\frac{\int_{t_0 + Cn^{-1/3}}^t w_i(x) dx}{\int w_i \hat{g}_n} - \frac{\int_{t_0 - Cn^{-1/3}}^t w_i(x) dx}{W_i} = O_p(n^{-1/2}) [t - (t_0 + Cn^{-1/3})].$$

Combining (i)–(iv) and choosing $\delta = \alpha/2$ in (ii), we have for $t \in I_n = [t_0 + 2Cn^{-1/3}, t_0 + R)$,

$$\begin{aligned} J_i(t) &\geq \frac{\alpha}{2}[t - (t_0 + Cn^{-1/3})]^2 + n^{-2/3}M_n^2 + O_p(n^{-1}) \\ &\quad + O_p(n^{-1/2})[t - (t_0 + Cn^{-1/3})] \\ &= \left(\frac{\alpha}{2} + O_p(n^{-1/2})[t - (t_0 + Cn^{-1/3})]^{-1}\right)[t - (t_0 + Cn^{-1/3})]^2 \\ &\quad + O_p(n^{-2/3}). \end{aligned}$$

With probability more than $1 - \varepsilon/2$, there exists some constant $D > 0$ such that for all large n ,

$$\begin{aligned} J_i(t) &\geq \left(\frac{\alpha}{2} - Dn^{-1/2}[t - (t_0 + Cn^{-1/3})]^{-1}\right)[t - (t_0 + Cn^{-1/3})]^2 \\ &\quad + O_p(n^{-2/3}). \end{aligned}$$

Hence, for any $t \in I_n$,

$$J_i(t) \geq \left(\frac{\alpha}{2} - DCn^{-1/6}\right)C^2n^{-2/3} + O_p(n^{-2/3}).$$

Hence, there exists $C_0 > 0$ such that for all $C \geq C_0$, we have for all large n ,

$$\mathbb{P}\left(\inf_{t \in I_n} J_i(t) > 0\right) \geq 1 - \varepsilon. \quad \square$$

Proof of Lemma 5.3. Take $C \geq C_0$ and $R \leq R_0$ as specified in Lemma 5.4. Define $E_n \triangleq \{\hat{g}_n(t_0 + 2Cn^{-1/3}) > g_0(t_0 + Cn^{-1/3})\}$. Note that for large enough n , $g_0(t_0 + 2Cn^{-1/3}) > g_0(t_0 + R)$ as $g'_0(t_0)$ is strictly negative. By Proposition 4.1 (iv) in [3], we see that with a probability more than $1 - \varepsilon$, for large enough n , $\hat{g}_n(t_0 + 2Cn^{-1/3}) > \hat{g}_n(t_0 + R)$, implying that \hat{g}_n has a jump in $I_n = [t_0 + 2Cn^{-1/3}, t_0 + R)$. Define \mathcal{U}_n to be the event that \hat{g}_n has a jump in I_n . Thus, $\mathbb{P}(\mathcal{U}_n) \geq 1 - \varepsilon$ for large enough n . Now, fix a sample point $\omega \in E_n \cap \mathcal{U}_n$. Denote $\tau_n = \tau_n(\omega)$ the first jump point of \hat{g}_n in I_n . Denote the negative log-likelihood function to be ψ_n . This can be written as

$$\psi_n(z_1, \dots, z_n) = -\sum_{j=1}^n \log z_j + \sum_{i=1}^s n_i \log \left(\sum_{k=1}^n z_k c_{ik} \right).$$

Define $\hat{z} = (\hat{g}_n(T_1), \dots, \hat{g}_n(T_n))$; see Section 3 in [3] for the conversion from \hat{g}_n to a finite number of auxiliary variables \hat{z} in the likelihood function. From the Karush-Kuhn-Tucker condition,

$$0 = \sum_{j: a < T_j < \tau_n} \frac{\partial \psi_n}{\partial z_j}(\hat{z}) = \sum_{j: a < T_j: t_0 + 2Cn^{-1/3}} \frac{\partial \psi_n}{\partial z_j}(\hat{z}) + \sum_{j: t_0 + 2Cn^{-1/3} \leq T_j < \tau_n} \frac{\partial \psi_n}{\partial z_j}(\hat{z})$$

and

$$\sum_{j:a < T_j < t_0 + 2Cn^{-1/3}} \frac{\partial \psi_n}{\partial z_j}(\hat{z}) \geq 0.$$

Therefore,

$$\sum_{j:t_0 + 2Cn^{-1/3} \leq T_j < \tau_n} \frac{\partial \psi_n}{\partial z_j}(\hat{z}) \leq 0.$$

For $t \in [t_0 + 2Cn^{-1/3}, \tau_n)$, by the monotonicity of \hat{g}_n , the fact that τ_n is the first jump point and $\omega \in E_n \cap \mathcal{U}_n$, we have

$$\hat{g}_n(t) = \hat{g}_n(\tau_n-) = \hat{g}_n(t_0 + 2Cn^{-1/3}) > g_0(t_0 + Cn^{-1/3}).$$

Therefore,

$$\begin{aligned} & \sum_{j:t_0 + Cn^{-1/3} \leq T_j < \tau_n} \left(-\frac{1}{g_0(t_0 + Cn^{-1/3})} + \sum_{i=1}^s \frac{n_i c_{ij}}{\int w_i \hat{g}_n} \right) \\ \leq & \sum_{j:t_0 + Cn^{-1/3} \leq T_j < \tau_n} \left(-\frac{1}{\hat{g}_n(T_j)} + \sum_{i=1}^s \frac{n_i c_{ij}}{\int w_i \hat{g}_n} \right) \\ = & \sum_{j:t_0 + 2Cn^{-1/3} \leq T_j < \tau_n} \frac{\partial \psi_n}{\partial z_j}(\hat{z}) \leq 0. \end{aligned}$$

Since $\tau_n \in I_n$, we have

$$\inf_{t \in I_n} \sum_{j:t_0 + Cn^{-1/3} \leq T_j < t} \left(-\frac{1}{g_0(t_0 + Cn^{-1/3})} + \sum_{i=1}^s \frac{n_i c_{ij}}{\int w_i \hat{g}_n} \right) \leq 0. \quad (\text{A.5})$$

However, from Lemma 5.4, for large enough n , we know that (A.5) happens with a probability at most ε , implying that $\mathbb{P}(E_n) = \mathbb{P}(E_n \cap \mathcal{U}_n) + \mathbb{P}(E_n \cap \mathcal{U}_n^c) \leq \varepsilon + \mathbb{P}(\mathcal{U}_n^c) \leq 2\varepsilon$. \square

Acknowledgments

The first author acknowledges the support by US National Institutes of Health Grant R01HL122212 and US National Science Foundation Grant DMS1711952. The second author thanks for the hospitality at the Chinese University of Hong Kong during the final draft of this work. The third author acknowledges the financial support of HKSAR GRF Grants 14317716 and 14301618. The fourth author acknowledges the financial support from HKGRF- Project Number 14300319 with the project title: Shape-constrained Inference: Testing for Monotonicity, 14300717 with the project title: New Kinds of Forward-Backward Stochastic Systems with Applications. He also thanks Columbia University for the kind invitation to be a visiting faculty member in the Department of Statistics during his sabbatical leave. Phillip Yam also recalls the unforgettable moments and the happiness shared with his beloved father during the drafting of

the present article at their home. Although he lost his father with the deepest sadness at the final stage of the review of this work, his father will never leave the heart of Phillip Yam; and he used this work in memory of his father's brave battle against liver cancer.

References

- [1] BANERJEE, M. (2007). Likelihood based inference for monotone response models. *The Annals of Statistics* **35** 931–956.
- [2] BINGHAM, N. H., GOLDIE, C. M. and TEUGELS, J. L. (1989). *Regular Variation* **27**. Cambridge University Press.
- [3] CHAN, K. C. G., LING, H. K., SIT, T. and YAM, S. C. P. (2018). Estimation of a monotone density in s -sample biased sampling models. *The Annals of Statistics* **46** 2125–2152.
- [4] CHANG, H. W., EL BARMÍ, H. and MCKEAGUE, I. W. (2016). Tests for stochastic ordering under biased sampling. *Journal of Nonparametric Statistics* **28** 659–682.
- [5] COX, D. R. (1969). *Some Sampling Problems in Technology*, eds. Johnson and Smith. New York: Wiley.
- [6] DUROT, C., KULIKOV, V. N., LOPUHAÄ, H. P. ET AL. (2012) The limit distribution of the L_1 -error of Grenander-type estimators. *The Annals of Statistics* **40** 1578–1608.
- [7] GILBERT, P. B., LELE, S. R. and VARDI, Y. (1999). Maximum likelihood estimation in semiparametric selection bias models with application to AIDS vaccine trials. *Biometrika* **86** 27–43.
- [8] GILL, R. D., VARDI, Y. and WELLNER, J. A. (1988). Large sample theory of empirical distributions in biased sampling models. *The Annals of Statistics* **16** 1069–1112.
- [9] GRENANDER, U. (1956). On the theory of mortality measurement, Part II. *Skand. Aktuarietidskr* **39** 125–153.
- [10] GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II* 539–555.
- [11] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints* **38**. Cambridge University Press.
- [12] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation* **19**. Springer Science & Business Media.
- [13] HELSEN, K. and SCHMITTLEIN, D. C. (1993). Analyzing duration times in marketing: Evidence for the effectiveness of hazard rate models. *Marketing Science* **11** 395–414.
- [14] HUANG, J. and WELLNER, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scandinavia Journal of Statistics* **22** 3–33.
- [15] KIEFER, N. M. (1988). Economic duration data and hazard functions. *Journal of Economic Literature* **26** 646–679.

- [16] KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *The Annals of Statistics* **18** 191–219.
- [17] KULIKOV, V. N. and LOPUHAÄ, H. P. (2006). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *The Annals of Statistics* **34** 742–768.
- [18] LOPUHAÄ, H. P. and NANE, G. F. (2013). Shape constrained non-parametric estimators of the baseline distribution in Cox proportional hazards model. *Scandinavian Journal of Statistics* **40** 619–646.
- [19] MARSHALL, A. W. (1970). Discussion of Barlow and van Zwet's paper. In *Nonparametric Techniques in Statistical Inference. Proceedings of the First International Symposium on Nonparametric Techniques held at Indiana University, June, 1969*, 174–176. Cambridge University Press.
- [20] MCFADDEN, J. A. (1962). On the lengths of intervals in a stationary point process. *Journal of the Royal Statistical Society, Series B* **24** 364–382.
- [21] MUTTLAK, H. A. and McDONALD, L. L. (1990). Ranked set sampling with size-biased probability of selection. *Biometrics* **46** 435–446.
- [22] QIN, J., BERWICK, M., ASHBOLT, R. and DWYER, T. (2002). Quantifying the change of Melanoma incidence by Breslow thickness. *Biometrics* **58** 665–670.
- [23] RAMÍREZ, P. and VIDAKOVIC, B. (2010). Wavelet density estimation for stratified size-biased sample. *Journal of Statistical Planning and Inference* **140** 419–432.
- [24] RESNICK, S. I. (2013). *Extreme Values, Regular Variation and Point Processes*. Springer.
- [25] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley.
- [26] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer-Verlag New York Inc.
- [27] VARDI, Y. (1982). Nonparametric estimation in the presence of length bias. *The Annals of Statistics* **10** 612–620.
- [28] VARDI, Y. (1985). Empirical distributions in selection bias models. *The Annals of Statistics* **13** 178–203.