

ESTIMATION OF A MONOTONE DENSITY IN s -SAMPLE BIASED SAMPLING MODELS

BY KWUN CHUEN GARY CHAN^{*,1}, HOK KAN LING[†], TONY SIT^{‡,2} AND SHEUNG CHI PHILLIP YAM^{‡,3}

University of Washington^{}, Columbia University[†] and The Chinese University of Hong Kong[‡]*

We study the nonparametric estimation of a decreasing density function g_0 in a general s -sample biased sampling model with weight (or bias) functions w_i for $i = 1, \dots, s$. The determination of the monotone maximum likelihood estimator \hat{g}_n and its asymptotic distribution, except for the case when $s = 1$, has been long missing in the literature due to certain nonstandard structures of the likelihood function, such as nonseparability and a lack of strictly positive second order derivatives of the negative of the log-likelihood function. The existence, uniqueness, self-characterization, consistency of \hat{g}_n and its asymptotic distribution at a fixed point are established in this article. To overcome the barriers caused by nonstandard likelihood structures, for instance, we show the tightness of \hat{g}_n via a purely analytic argument instead of an intrinsic geometric one and propose an indirect approach to attain the \sqrt{n} -rate of convergence of the linear functional $\int w_i \hat{g}_n$.

1. Introduction.

1.1. *Background and problem formulation.* The estimation of a density function is a fundamental problem in nonparametric statistics. A monotone constraint may arise naturally and is often assumed. Grenander (1956) showed the maximum likelihood estimator (MLE) of a decreasing density function is the derivative of the least concave majorant of the empirical distribution. Its theoretical properties are studied by Prakasa Rao (1969), Barlow et al. (1972), Groeneboom (1985) and Dümbgen, Wellner and Wolff (2016), among others. Nonparametric estimation under monotone constraints in various statistical problems has been studied.

Received February 2016; revised May 2017.

¹Supported by US National Institutes of Health Grant R01HL122212 and US National Science Foundation Grant DMS1711952.

²Supported in part by CUHK Direct Grant 4053086.

³Supported by The Hong Kong RGC GRF 14301015 with the project title: Advance in Mean Field Theory, The Hong Kong RGC GRF 14300717 with the project title: New kinds of Forward-backward Stochastic Systems with Applications, and Direct Grant for Research 2015/16 with the project code 4053141 offered by CUHK.

MSC2010 subject classifications. Primary 62G20, 62E20; secondary 62G08.

Key words and phrases. Density estimation, empirical process theory, Karush–Kuhn–Tucker conditions, nonparametric estimation, order statistics from multiple samples, self-induced characterization, shape-constrained problem, s -sample biased sampling.

For instance, [Huang and Wellner \(1995b\)](#) studied monotone density and hazard estimation for right censored survival data, and [Groeneboom and Wellner \(1992\)](#) studied interval censoring data. A comprehensive survey of nonparametric estimation problems under shape constraints can also be found in [Groeneboom and Jongbloed \(2014\)](#).

Existing methods mostly require that the observed data be a simple random sample from the underlying population. Biased sampling, however, often arises in practice when a unit is preferentially sampled based on its value. This is often due to the method of selection such that units in the study population do not have equal chances of being recorded. Many examples exist in the literature. For instance, a large herd is more likely to be sampled in wildlife [[Cook and Martin \(1974\)](#)], a more long-lived survivor is more likely to appear in a prevalent cohort [[Wang \(1991\)](#)], a long textile fiber is more likely to be picked from a bulk of fibers in intersect sampling [[Cox \(1968\)](#)]. Other examples are given in [Patil and Rao \(1978\)](#) and [Drummer and McDonald \(1987\)](#) among others.

Suppose $s \geq 1$ sampling methods are adopted, each resulting in a different form of sampling bias. This is called a s -sample biased sampling model and the corresponding estimation of a distribution function was studied by [Vardi \(1985\)](#) and [Gill, Vardi and Wellner \(1988\)](#). Our primary goal is to study the estimation of a decreasing density function under a s -sample biased sampling model. It is noteworthy that if only one set of biased samples is available, that is, $s = 1$, the problem of the monotone density estimation can be carried out via the classical method [[Grenander \(1956\)](#)] based on properly transformed samples, which has been studied by [El Barmi and Nelson \(2002\)](#). This is feasible due to an invariance principle which allows a one-to-one mapping between the two densities. This approach, however, cannot be directly extended to $s > 1$ as such an invariant structure is no longer available. To the best of our knowledge, a general solution to this problem has long been missing due to several nonstandard features of the likelihood function, which will be discussed in Section 1.3.

More precisely, we consider the following estimation problem. Let G_0 be a fixed but unknown distribution function with a decreasing density g_0 supported on a bounded interval $[a, b]$, where $0 \leq a < b < \infty$. Assume that there are s known positive weight (or bias) functions w_i ($i = 1, \dots, s$). Suppose we obtain s independent samples X_{i1}, \dots, X_{in_i} ($i = 1, \dots, s$), each X_{ij} independently follows the biased distribution F_i ($i = 1, \dots, s, j = 1, \dots, n_i$), such that

$$F_i(x) \triangleq \frac{\int_a^x w_i(y)g_0(y) dy}{\int_a^b w_i(y)g_0(y) dy} \quad \text{for } x \in [a, b],$$

where $0 < \int_a^b w_i(y)g_0(y) dy < \infty$ for $i = 1, \dots, s$. Let \mathcal{G} be the class of all decreasing densities on $[a, b]$. For any $g \in \mathcal{G}$, the likelihood evaluated at this g of the s -sample is proportional to

$$(1.1) \quad L_n(g) \triangleq \prod_{i=1}^s \frac{\prod_{j=1}^{n_i} g(X_{ij})}{\left(\int_a^b w_i(x)g(x) dx\right)^{n_i}}.$$

From now on, without ambiguity, we also regard L_n as the likelihood function. The monotone MLE $\hat{g}_n \in \mathcal{G}$ is then defined such that $L_n(\hat{g}_n) \geq L_n(g)$ for all $g \in \mathcal{G}$. Our goal of this article is to characterize the properties of and to establish the limiting distribution of \hat{g}_n at a fixed point $t_0 \in (a, b)$. We shall prove the following theorem.

THEOREM 1.1. *Under Assumptions 2.1, we have*

$$n^{1/3} \left(\sum_{i=1}^s \frac{\lambda_i w_i(t_0)}{\int_a^b w_i(x) g_0(x) dx} \right)^{1/3} \left| \frac{1}{2} g_0'(t_0) g_0(t_0) \right|^{-1/3} (\hat{g}_n(t_0) - g_0(t_0)) \xrightarrow{\mathbb{D}} 2\mathbb{Y},$$

where $\mathbb{Y} \triangleq \arg \max_t \{W(t) - t^2\}$, $W(t)$ ($t \in \mathbb{R}$), is the standard two-sided Brownian motion with $W(0) = 0$, $n = \sum_{i=1}^s n_i$ and $\lambda_i = \lim_n n_i/n > 0$.

1.2. Examples of applications. Multisample selection bias often arises when data are collected from multiple sampling plans. Examples include outcome dependent enriched sampling in medical studies [Wang and Zhou (2006), Kang, Nelson and Vahl (2010)] and endogenous stratified sampling in economics [Hausman and Wise (1981), Imbens and Lancaster (1996)]. In these cases, biased samples are collected in addition to a random sample. Such sampling designs are often considered more focused and economical than random samples from a population. Multi-sample selection bias also occurs frequently in combining different data sources. For example, an unbiased sample and a duration-biased sample are formed when the Surveillance, Epidemiology and End Results (SEER) data are linked to the Medicare data [Chan and Wang (2012)]. Since Medicare data are only available from 1986 but SEER data are available from 1973, the linked data excludes the SEER subjects who died before 1986. Hence, subjects with cancer diagnosis before 1986 forms a duration-biased cohort in the combined data while subjects with cancer diagnosis on or after 1986 forms an unbiased cohort. Patil (1984) combined travel survey data collected from hotels and frontier stations in Morocco, where the samples collected from hotels are biased toward longer sojourn times. Other examples of multiple-sample selection bias include the measurement of velocities of moving vehicles by multiple moving observes traveling at different velocities [Smith and Parnes (1994)], the sampling of time from a disease incidence to a failure event from multiple prevalent cohorts with different time trends of disease incidence [Wang (1991)] and the collection of survival data from multiple study sites with different reporting delay distributions [Wang (1992)].

1.3. An overview of our approach. We first show that the maximizer of L_n must be a step function. We then show that the monotone MLE exists uniquely, and can be evaluated via a self-induced characterization through the use of an isotonic regression problem, in which we need to choose a diagonal matrix with positive entries. In particular, we choose the first term of the second-order derivative of the negative of the log-likelihood function, which is always positive and is

also the dominating term for a large enough n . There is a special feature of the presence of the linear functional $\int w_i \hat{g}_n$ in the log-likelihood function making it nonseparable; hence, we aim to establish that $|\int w_i \hat{g}_n - \int w_i g_0| = O_p(n^{-1/2})$ so that we can essentially replace $\int w_i \hat{g}_n$ with $\int w_i g_0$ in most of the calculations for deriving various estimates. To this end, we have a set of preparatory lemmas. More specifically, we first show that \hat{g}_n is bounded in probability and use this to obtain a L_2 -rate of convergence of \hat{g}_n from its rate of convergence in Hellinger distance. Unlike the one-sample unbiased case where the maximum likelihood estimator can be characterized as the left-continuous slope of the least concave majorant of the empirical distribution, such a geometric interpretation is lacking here. We cannot directly extend the tightness results of Woodroffe and Sun (1993) to the s -sample case, which will be needed for our proofs, as they use the geometric configuration of the Grenander estimator in their proof. Thus, we seek an alternative analytic argument to show the tightness of the monotone MLE. After obtaining this crucial rate of convergence of $\int w_i \hat{g}_n$, the local consistency of \hat{g}_n , and hence the asymptotic distribution of \hat{g}_n at a fixed interior point can then be obtained. More detailed explanations and relevant connections with the existing literature are provided in each section. An additional comparison with the interval censoring problem (case 2) studied in Groeneboom (1996) is given in the discussion, where his proposed problem and ours both possess the nonseparability in the log-likelihood function.

1.4. Organization. The remainder of this paper is structured as follows: Section 2 describes the notation and some relevant technical assumptions. Discussions on the existence, uniqueness and characterization of the monotone MLE are included in Section 3. Section 4 establishes the consistency of the monotone MLE. Section 5 studies the rates of convergence of the monotone MLE and its linear functionals, and contains the majority of our notable methodologies. Those results are then used for establishing the asymptotic distribution of the monotone MLE in Section 6. Concluding remarks are given in Section 7. Numerical demonstrations and part of the technical details and proofs of our theoretical results will be relegated to the supplementary material [Chan et al. (2018)].

2. Notation and assumptions. Consider the biased sampling model introduced in Section 1.1. Denote $n \triangleq \sum_{i=1}^s n_i$. For $i = 1, \dots, s$, the empirical measures from $X_{i1}, \dots, X_{in_i} \sim F_i$ will be denoted as \mathbb{F}_{i,n_i} , where the measure F_i has a density f_i with respect to Lebesgue measure. In the rest of this article (including the supplementary material [Chan et al. (2018)]), if clarity of the dependence of \hat{g}_n on an element ω of the sample space Ω is demanded, we shall explicitly write $\hat{g}_n(x, \omega)$; otherwise, we shall only write $\hat{g}_n(x)$ for simplicity. For other functions, we shall follow the same custom. The indicator for a set A will be denoted by $1(A)$.

Before we proceed further, we shall state the regularity conditions adopted in the whole article.

ASSUMPTIONS 2.1. (A) The sampling fraction $\frac{n_i}{n} \rightarrow \lambda_i > 0$, for all $i = 1, \dots, s$.

(B) The true unbiased decreasing density g_0 is differentiable on the interior of its support (a, b) , with $0 < \inf_{t \in (a,b)} |g'_0(t)| \leq \sup_{t \in (a,b)} |g'_0(t)| < \infty$.

(C) There exist constant bounds $0 < m < M < \infty$ such that for all $i = 1, \dots, s$, $m \leq g_0, w_i \leq M$.

(D) For each $i = 1, \dots, s$, $w_i, w_i \circ g_0^{-1}$ are Lipschitz continuous.

Our main theorem concerns the asymptotic distribution of \hat{g}_n at a fixed point $t_0 \in (a, b)$. The conditions on g_0 in Assumptions (B) and (C) are similar to Assumptions (A1) and (A2) imposed in Theorem 1.1 in Groeneboom, Hooghiemstra and Lopuhaä (1999), which assists in deriving the asymptotic normality of a suitably rescaled version of the L_1 error of the Grenander estimator; while the lower boundedness condition on w_i in Assumption (C) is also used in El Barmi and Nelson (2002). Assumption (D) is used to ensure the permanence of the Donsker property after Lipschitz transformation and the rate of convergence of certain integrals in connection to the L_2 -convergence $\int (\hat{g}_n - g_0)^2$ in Lemma 5.6.

For any function H on $[0, \infty)$ with $H(0) = 0$, we define the least concave majorant (LCM) \hat{H} of H to be the smallest concave function that dominates H over $[0, \infty)$ with $\hat{H}(0) = 0$. Clearly, as \hat{H} is concave, its derivative is nonincreasing. Given points $\{(x_i, y_i)\}_{i=0}^n$ with $x_0 = y_0 = 0$ and $x_0 < x_1 < \dots < x_n$, consider the right-continuous step function P such that $P(x_i) = y_i$ and it remains constant on each interval $[x_i, x_{i+1})$ for $i = 0, \dots, n - 1$. Denote the \mathbb{R}^n -vector of left-derivatives of the least concave majorant of P computed at the points (x_1, \dots, x_n) by $\text{slcmm}\{(x_i, y_i)\}_{i=0}^n$.

3. Existence, uniqueness and characterization of the monotone MLE \hat{g}_n .

Since the observations are from a continuous distribution, we assume that there are no ties. Let $a = T_0 < T_1 < \dots < T_n$ be the order statistics of all X_{ij} 's. As in many other nonparametric estimation problems with monotone shape constraints, the monotone MLE in the current problem must be a step function with jumps only at the observations as shown below.

PROPOSITION 3.1. *For any $g \in \mathcal{G}$, there exists a left continuous decreasing step function \bar{g} with jumps only at the order statistics T_i (not necessarily all) for all $i = 1, \dots, n$, and it takes value 0 whenever $x \notin [T_0, T_n]$ such that $L_n(\bar{g}) \geq L_n(g)$.*

PROOF. See the details provided in Section 8 of the supplementary material [Chan et al. (2018)]. □

3.1. *Existence and uniqueness of the monotone MLE \hat{g}_n .* For a given sequence of $T_0 < T_1 < \dots < T_n$, define \mathcal{G}^T to be the (samplewise-) set of all left continuous decreasing piecewise constant densities with jumps only at the order statistics T_i 's.

Define also $c_{jk} \triangleq \int_{T_{k-1}}^{T_k} w_j(y) dy > 0$. Proposition 3.1 suggests that we only have to look for the maximizer of L_n in \mathcal{G}^T , which is equivalent to the resolution of the following maximization problem with the objective function

$$(3.1) \quad \mathcal{L}_n(z_1, \dots, z_n) \triangleq \prod_{i=1}^n z_i \prod_{j=1}^s \left(\sum_{k=1}^n c_{jk} z_k \right)^{-n_j},$$

subject to

$$(3.2) \quad \begin{cases} \sum_{i=1}^n z_i (T_i - T_{i-1}) = 1, \\ z_1 \geq \dots \geq z_n \geq 0. \end{cases}$$

Let

$$\mathcal{K}_n \triangleq \left\{ z \in \mathbb{R}^n : z_1 \geq \dots \geq z_n \geq 0 \text{ and } \sum_{i=1}^n z_i (T_i - T_{i-1}) = 1 \right\}.$$

Then, for each $n \in \mathbb{N}$, with probability one, \mathcal{K}_n is compact. The original problem can be recovered through the transformation $z_i = g(T_i)$ and $c_{jk} = \int_{T_{k-1}}^{T_k} w_j(y) dy$. The constraint $\sum_{i=1}^n z_i (T_i - T_{i-1}) = 1$ ensures at least one of the z_i 's is positive (nonvanishing), and hence $\sum_{k=1}^n z_k c_{jk} > 0$ as $w_j > 0$. Therefore, (3.1) is well defined.

Note that the log-likelihood function of (3.1) is neither concave nor convex in z . We define $p_i \triangleq \log z_i$ for $z_i > 0$, otherwise if $z_i = 0$, $p_i \triangleq -\infty$, for $i = 1, \dots, n$ and

$$(3.3) \quad \begin{aligned} & \tilde{\mathcal{L}}_n(p_1, \dots, p_n) \\ & \triangleq \begin{cases} \sum_{j=1}^s p_j - \sum_{i=1}^n n_i \log \left(\sum_{k=1}^n c_{ik} e^{p_k} \right) & \text{if all } p_i \in (-\infty, \infty), \\ -\infty & \text{if some } p_i \text{ is equal to } -\infty. \end{cases} \end{aligned}$$

Note that at least one of the z_i 's is positive and so p_i 's cannot be all equal to $-\infty$, and hence $\tilde{\mathcal{L}}_n$ is well defined. Then it can be shown that $\tilde{\mathcal{L}}_n$ is concave in $\mathbf{p} = (p_1, \dots, p_n)$ (as shown in Proposition 8.1 in the supplementary material [Chan et al. (2018)], see also Davidov and Iliopoulos (2009)). With the concavity of $\tilde{\mathcal{L}}_n$, it can also be shown that, with probability one, the monotone MLE \hat{g}_n for the true unbiased density g_0 uniquely exists for each $n \in \mathbb{N}$; see Proposition 8.2 in the supplementary material [Chan et al. (2018)] for details.

3.2. *Characterization of the monotone MLE \hat{g}_n via theory of isotonic regression.* Self-induced characterization in the estimation of monotone function has become prevalent since the work of Groeneboom and Wellner (1992). One of the

developments is to connect isotonic regression problems with self-induced characterization; this approach relies on a suitable choice of a diagonal positive definite matrix in a quadratic programming problem that closely links to the original maximization problem.

To tackle our present problem, we here advocate the approach in Banerjee (2007) and Groeneboom and Wellner (1992). However, due to the nonseparateness of the arguments in the log-likelihood function, we propose an alternative function ψ_n which is the negative log-likelihood function of all the data instead of just one single datum. That is,

$$(3.4) \quad \psi_n(z_1, \dots, z_n) \triangleq - \sum_{j=1}^n \log z_j + \sum_{i=1}^s n_i \log \left(\sum_{k=1}^n z_k c_{ik} \right).$$

However, ψ_n is not necessarily strictly convex (also see Proposition 8.1 in the supplementary material [Chan et al. (2018)]), and $\frac{\partial^2 \psi_n}{\partial z_j^2}(\hat{z})$ may not always be strictly positive therefore $\frac{\partial^2 \psi_n}{\partial z_j^2}(\hat{z})$'s cannot immediately serve as the diagonal entries in the positive definite matrix mentioned above. A proper choice of matrix plays a critical role so that the methodologies in Banerjee (2007) and Groeneboom and Wellner (1992) can be implemented in our setting. We therefore note that the cumulative sums of these diagonal elements of the matrix are essentially the first coordinates of the corresponding isotonic regression problem (Proposition 3.2); its increasing structure results in a natural geometry of the solution of that regression problem. Therefore, to maintain this plausible approach, alternative candidates other than $\frac{\partial^2 \psi_n}{\partial z_j^2}(\hat{z})$'s have to be suggested in order to keep this monotonic structure. In particular, we here choose $\frac{1}{z_j^2}$, which is always positive and is actually the first term of $\frac{\partial^2 \psi_n}{\partial z_j^2}(\hat{z})$, to form that positive definite matrix.

PROPOSITION 3.2. *The maximizer \hat{z} of (3.1) subject to (3.2) satisfies*

$$(3.5) \quad (\hat{z}_1, \dots, \hat{z}_n) = \text{slo lcm} \left\{ \sum_{j=1}^i \frac{1}{\hat{z}_j^2}, \sum_{j=1}^i \left(\frac{1}{\hat{z}_j} - \frac{\partial \psi_n}{\partial z_j}(\hat{z}) \right) \right\}_{i=0}^n.$$

PROOF. See Section 8 of the Supplementary Material [Chan et al. (2018)] for details. □

Without ambiguity, we shall use \hat{z}_i and $\hat{g}_n(T_i)$ interchangeably depending on the actual context. Following Proposition 3.2, we consider some processes \tilde{G}_{n, \hat{g}_n} , \tilde{G}_{n, g_0} , \tilde{U}_{n, \hat{g}_n} and \tilde{U}_{n, g_0} , which will be defined next. Our main theorem concerns the asymptotic distribution of \hat{g}_n at a fixed interior point t_0 in the support of g_0 . Define

$\lambda \triangleq \sum_{i=1}^s \lambda_i f_i(t_0)/g_0^2(t_0)$ and

$$G_{n,\hat{g}_n}(t) \triangleq \frac{1}{n} \sum_{j=1}^n \frac{1}{\hat{g}_n^2(T_j)} 1(T_j \leq t),$$

$$U_{n,\hat{g}_n}(t) \triangleq \frac{1}{n} \sum_{j=1}^n \left(\frac{1}{\hat{g}_n(T_j)} - \frac{\partial \psi_n}{\partial z_j}(\hat{z}) \right) 1(T_j \leq t).$$

Then, from Proposition 3.2,

$$\{\hat{g}_n(T_1), \dots, \hat{g}_n(T_n)\} = \text{slolcm}\{G_{n,\hat{g}_n}(T_i), U_{n,\hat{g}_n}(T_i)\}_{i=0}^n.$$

Define the normalized and localized version of $G_{n,\hat{g}_n}(t)$ and $U_{n,\hat{g}_n}(t)$ as follows:

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

Finally, we define the theoretical counterpart process \tilde{U}_{n,g_0} as \tilde{U}_{n,\hat{g}_n} and \tilde{G}_{n,g_0} as \tilde{G}_{n,\hat{g}_n} , respectively, with \hat{g}_n being replaced by g_0 , that is,

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

Following established approaches, the determination of the asymptotic distribution of the monotone \hat{g}_n relies on the asymptotic distribution of \tilde{G}_{n,\hat{g}_n} and \tilde{U}_{n,\hat{g}_n} , both of which can be determined via the establishment of the asymptotic equivalence of \tilde{G}_{n,\hat{g}_n} and \tilde{G}_{n,g_0} and that of \tilde{U}_{n,\hat{g}_n} and \tilde{U}_{n,g_0} , and then through the finding of

the asymptotic distributions of \tilde{G}_{n,g_0} and \tilde{U}_{n,g_0} . However, in showing the asymptotic equivalence of \tilde{U}_{n,\hat{g}_n} and \tilde{U}_{n,g_0} , the choice of $\frac{1}{z_j^2}$ instead of $\frac{\partial^2 \psi_n}{\partial z_j^2}(\hat{z})$, and the nonseparability of the arguments in the log-likelihood function require additional analytic treatment that involves techniques beyond those illustrated in Banerjee (2007); see Section 5 for a more detailed discussion. Essentially, we develop a notable approach on first obtaining the \sqrt{n} -convergence of $\int w_i \hat{g}_n$ in Proposition 5.8 in order to establish the asymptotic equivalence of \tilde{U}_{n,\hat{g}_n} and \tilde{U}_{n,g_0} .

4. Consistency of the monotone MLE \hat{g}_n . Following the same notation as in van de Geer (2000), denote $H_{p,B}(\delta, \mathcal{I}, Q)$ as the δ -entropy with bracketing for the $L_p(Q)$ -metric of \mathcal{I} , where \mathcal{I} is a class of suitable functions and Q is a measure on a measurable space $(\mathcal{M}, \mathcal{B})$. Consistency of the monotone MLE can be established by using the notion of Hellinger distance and the theory of empirical processes. Define $\hat{f}_{i,n} \triangleq \frac{w_i \hat{g}_n}{\int w_i \hat{g}_n}$ and $f_i \triangleq \frac{w_i g_0}{\int w_i g_0}$ for $\hat{g}_n, g_0 \in \mathcal{G}$. For clarity, we shall write all the integrals, except for $\int w_i g_0$ and $\int w_i \hat{g}_n$, in the form $\int_a^b f(x) dx$, where the argument of the integrand and the domain of integration are written out explicitly. For any density functions $q_1, q_2 \in \mathcal{G}$ with support on $[a, b]$, their Hellinger distance h is defined as

$$h(q_1, q_2) \triangleq \left(\frac{1}{2} \int_a^b (q_1^{1/2}(x) - q_2^{1/2}(x))^2 dx \right)^{1/2}.$$

To obtain the consistency of \hat{g}_n with the appropriate rate of convergence, we show item (iv) in the following proposition, which is a consequence of items (i) to (iii).

PROPOSITION 4.1. *Under Assumptions 2.1, for each $i = 1, \dots, s$:*

- (i) $h(\hat{f}_{i,n}, f_i) \rightarrow 0$ a.s.
- (ii) $\int w_i \hat{g}_n \rightarrow \int w_i g_0$ a.s.
- (iii) For any $0 < \delta < b - a$, $\limsup_n \hat{g}_n(a + \delta) \leq \frac{1}{\delta}$ and there exists constants $C(\delta)$ such that $P(\limsup_n \frac{1}{\hat{g}_n(b-\delta)} \leq C(\delta)) = 1$.
- (iv) For any closed subinterval $[\sigma, \tau] \subset (a, b)$, $\sup_{x \in [\sigma, \tau]} |\hat{g}_n(x) - g_0(x)| \rightarrow 0$ a.s.

PROOF. See the details provided in Section 9 of the supplementary material [Chan et al. (2018)]. □

5. Rates of convergence of the linear functional $\int w_i \hat{g}_n$ and the monotone MLE \hat{g}_n .

5.1. *Rate of convergence of the linear functional $\int w_i \hat{g}_n$.* To establish the asymptotic equivalence of \tilde{U}_{n,\hat{g}_n} and \tilde{U}_{n,g_0} , the minimum order $o_p(n^{-1/3})$ of the

term $\sum_{k=1}^n g_0(T_k)c_{ik} - \int w_i \hat{g}_n$ is necessary. In this subsection, the main result to be obtained is the fact that $\int w_i \hat{g}_n - \int w_i g_0 = O_p(n^{-1/2})$, which is given in Proposition 5.8. The proof of this result requires several intermediate lemmas, and we here briefly describe the crux of their arguments. By invoking two noncanonical rudimentary results related to order statistics (Lemma 5.1 and Corollary 5.2), we develop a new purely analytic approach to establish the boundedness in probability of \hat{g}_n instead of using the geometric configuration of the Grenander estimator in the existing literature; see Lemma 5.3. The multisample and the self-induced characterization nature of \hat{g}_n make \hat{g}_n substantially different from the Grenander estimator, \hat{g}_n^{GE} , corresponding to the 1-sample unbiased case, in which the Grenander estimator \hat{g}_n^{GE} can be characterized as the left-continuous slope of the least concave majorant of the empirical distribution. By using this geometric interpretation of \hat{g}_n^{GE} , Woodroffe and Sun (1993) showed that if $0 < g_0(0+) < \infty$, then

$$(5.1) \quad \frac{\hat{g}_n^{\text{GE}}(0+)}{g_0(0+)} \xrightarrow{\mathbb{D}} \sup_{1 \leq k < \infty} \frac{k}{\Gamma_k},$$

where $\Gamma_1, \Gamma_2, \dots$ are partial sums of i.i.d. standard exponential random variables. In our present situation, such a geometric connection is absent and we therefore utilize some pure analytic arguments that lead to a similar bound in the spirit of (5.1), yet allow more room for further generalization; see Lemma 5.1, Corollary 5.2 and Lemma 5.3.

From (9.1) in the proof of Proposition 4.1, we can also obtain a rough rate of Hellinger distance $h(\hat{f}_{i,n}, f_i)$. Together with the boundedness in probability of \hat{g}_n , we can establish a rough, but fast enough for our later development, L_2 -rate of convergence of $\hat{f}_{i,n}$, and hence that of \hat{g}_n ; see Lemma 5.4 for details. We next implement the approach from Huang and Wellner (1995a), by using Karush–Kuhn–Tucker conditions, we can “insert” any function γ such that $\sum_{i=1}^n \frac{\partial \psi_n}{\partial z_i}(\hat{z})\gamma(\hat{g}_n(T_i)) = 0$ (Lemma 5.5). Using suitable choices of γ , together with the L_2 -rate of convergence of \hat{g}_n and the Donsker class property of a class of modified version of \hat{g}_n , we obtain the key equation (5.5) in order to deduce the desired \sqrt{n} -convergence of $\int w_i \hat{g}_n$ (Lemma 5.6). The form of (5.5) motivates us to consider a linear system by suitably choosing a number of γ 's for which s linearly independent equations involving $\int w_i \hat{g}_n$ will then result. Direct matrix inversion immediately gives $\sqrt{n}(\int w_i \hat{g}_n - \int w_i g_0) = O_p(1)$ as desired (Proposition 5.8). To the best of our knowledge, methods for handling s linear functionals of this kind simultaneously are rare in the literature.

Next, we then show that $\sum_{j=1}^n g_0(T_j)c_{ij} - \int w_i g_0 = O_p(n^{-1/2})$ in Lemma 5.9. Hence, as a direct consequence of Proposition 5.8 and Lemma 5.9, we obtain $\sum_{j=1}^n g_0(T_j)c_{ij} - \int w_i \hat{g}_n = O_p(n^{-1/2})$ as claimed before.

LEMMA 5.1. *Suppose that $W_1, \dots, W_n \stackrel{\text{i.i.d.}}{\sim} F$ with a density function f supported on a finite interval $[c, d] \subset \mathbb{R}$ so that f is bounded above and below from*

zero on $[c, d]$: there exists $0 < m_f \leq M_f < \infty$ such that $m_f \leq f(x) \leq M_f$ for all $x \in [c, d]$. Let $W_{(h)}$ be the h th order statistic of W_1, \dots, W_n . Then:

- (i) $\max_{h=1, \dots, n} \frac{n(W_{(h)}-c)}{h} = O_p(1)$;
- (ii) $(\min_{h=1, \dots, n} \frac{n(W_{(h)}-c)}{h})^{-1} = O_p(1)$.

PROOF. See the details provided in Section 10 of the Supplementary Material [Chan et al. (2018)]. \square

- COROLLARY 5.2. (i) $\max_{h=1, \dots, n} \frac{n(T_n-T_{n-h})}{h} = O_p(1)$;
- (ii) $(\min_{h=1, \dots, n} \frac{n(T_h-a)}{h})^{-1} = O_p(1)$.

PROOF. See the details provided in Section 10 of the Supplementary Material [Chan et al. (2018)]. \square

LEMMA 5.3. The monotone MLE \hat{g}_n is bounded above in probability, that is,

$$\hat{g}_n(a) \triangleq \hat{g}_n(a+) = \hat{z}_1 = O_p(1).$$

Remark: we also have $\sup_{x \in [a, b]} \hat{f}_{i,n}(x) = O_p(1)$ for $i = 1, \dots, s$.

PROOF. In the Supplementary Material [Chan et al. (2018)], we show that

$$(5.2) \quad \hat{z}_1 \leq \hat{z}_n \frac{M^2 \max_{h=1, \dots, n} \frac{n(T_n-T_{n-h})}{h}}{m^2 \min_{j=1, \dots, n} \frac{n(T_j-a)}{j}} + \frac{1}{T_{n-1} - a}.$$

Note that since $\hat{z}_n = \hat{g}_n(T_n)$, $T_n \xrightarrow{a.s.} b$, and Proposition 4.1(iii) implies that $\hat{g}_n(b - \delta) \leq \hat{g}_n(a + \delta) = O(1)$ for a small enough $\delta > 0$, we have $\hat{z}_n = O_p(1)$. As $T_{n-1} \xrightarrow{a.s.} b$, $\frac{1}{T_{n-1}-a} = O_p(1)$. In light of Corollary 5.2, we have $\hat{z}_1 = O_p(1)$. \square

To prepare for the proof of Lemma 5.6, we provide the first batch of rough estimates of the rate of convergence.

LEMMA 5.4. For each $i = 1, \dots, s$:

- (i) $h(\hat{f}_{i,n}, f_i) = O_p(n^{-1/4})$;
- (ii) $|\int w_i \hat{g}_n - \int w_i g_0| = O_p(n^{-1/4})$;
- (iii) $\int_a^b (\hat{f}_{i,n}(x) - f_i(x))^2 dx = O_p(n^{-1/2})$;
- (iv) $\int_a^b (\hat{g}_n(x) - g_0(x))^2 dx = O_p(n^{-1/2})$.

PROOF. See the details provided in Section 10 of the Supplementary Material [Chan et al. (2018)]. \square

LEMMA 5.5. For any function γ , we have

$$(5.3) \quad \sum_{j=1}^n \left(-\frac{1}{n \hat{g}_n(T_j)} + \sum_{i=1}^s \frac{n_i}{n} \frac{c_{ij}}{f w_i \hat{g}_n} \right) \gamma(\hat{g}_n(T_j)) = 0.$$

PROOF. See the details provided in Section 10 of the Supplementary Material [Chan et al. (2018)]. \square

Next, we choose a particular form of function γ to facilitate the subsequent proof. For some constants α_i 's and β , define $\gamma^{\alpha, \beta} : (0, \infty) \times [\frac{\lambda_1}{2}, 1] \times \dots \times [\frac{\lambda_s}{2}, 1] \rightarrow \mathbb{R}$ by

$$(5.4) \quad \gamma^{\alpha, \beta}(x, r_1, \dots, r_s) \triangleq \begin{cases} \frac{x[\sum_{i=1}^s \alpha_i f_i(b) + \beta g_0(b)]}{\sum_{j=1}^s r_j f_j(b)} & \text{for } 0 < x < g_0(b), \\ \frac{x[\sum_{i=1}^s \alpha_i f_i(g_0^{-1}(x)) + \beta x]}{\sum_{j=1}^s r_j f_j(g_0^{-1}(x))} & \text{for } g_0(b) \leq x \leq g_0(a), \\ \frac{x[\sum_{i=1}^s \alpha_i f_i(a) + \beta g_0(a)]}{\sum_{j=1}^s r_j f_j(a)} & \text{for } x > g_0(a). \end{cases}$$

Note that $0 < \frac{\lambda_j}{2} \leq r_j \leq 1$ for each j , therefore, the denominator in the definition of $\gamma^{\alpha, \beta}$ for all values of x is bounded away from zero with a lower bound $\sum_{j=1}^s \frac{\lambda_j}{2} f_j(b) > 0$. In addition, as $w_i \circ g_0^{-1}$ is assumed to be Lipschitz and bounded above and below, it is straightforward to see that:

(i) The mapping $(x, r_1, \dots, r_s) \mapsto \gamma^{\alpha, \beta}(x, r_1, \dots, r_s)/x$ is bounded and Lipschitz continuous on $(0, \infty) \times [\frac{\lambda_1}{2}, 1] \times \dots \times [\frac{\lambda_s}{2}, 1]$;

(ii) The families of functions $\{x \mapsto \gamma^{\alpha, \beta}(x, r_1, \dots, r_s)/x : r_j \in [\frac{\lambda_j}{2}, 1], j = 1, \dots, s\}$ and $\{x \mapsto \gamma^{\alpha, \beta}(x, r_1, \dots, r_s) : r_j \in [\frac{\lambda_j}{2}, 1], j = 1, \dots, s\}$ are uniformly and globally Lipschitz such that for any $x_1, x_2 \in (0, \infty)$,

$$\sup_{r_j \in [\frac{\lambda_j}{2}, 1], j=1, \dots, s} |\gamma^{\alpha, \beta}(x_1, r_1, \dots, r_s)/x_1 - \gamma^{\alpha, \beta}(x_2, r_1, \dots, r_s)/x_2| \leq C|x_1 - x_2|,$$

and

$$\sup_{r_j \in [\frac{\lambda_j}{2}, 1], j=1, \dots, s} |\gamma^{\alpha, \beta}(x_1, r_1, \dots, r_s) - \gamma^{\alpha, \beta}(x_2, r_1, \dots, r_s)| \leq C|x_1 - x_2|,$$

where C is a constant that depends only on $\lambda_1, \dots, \lambda_s$.

Denote $\gamma_n^{\alpha, \beta}(\cdot) \triangleq \gamma^{\alpha, \beta}(\cdot, \frac{n_1}{n}, \dots, \frac{n_s}{n})$. Note that for small n , $\gamma_n^{\alpha, \beta}$ is not well defined if some $\frac{n_i}{n} < \frac{\lambda_i}{2}$. However, since $\frac{n_i}{n} \rightarrow \lambda_i$ as $n \rightarrow \infty$, $\gamma_n^{\alpha, \beta}$ is then well

defined for large enough n and this observation is used in the following Lemma 5.6. Note that we are only concerned with the asymptotic behaviors of \hat{g}_n and the linear functional $\int w_i \hat{g}_n$, and so it suffices to consider $\gamma^{\alpha,\beta}$ for large enough n .

LEMMA 5.6. *For all large enough n , $\gamma_n^{\alpha,\beta}$ is a well-defined function and*

$$(5.5) \quad \sum_{i=1}^s \frac{n_i}{n} \frac{1}{(\int w_i g_0)^2} \int_a^b w_i(x) \gamma_n^{\alpha,\beta}(g_0(x)) dx \int w_i(g_0 - \hat{g}_n) \\ = \int_a^b \left(\sum_{i=1}^s \frac{n_i}{n} \frac{w_i(x)}{\int w_i g_0} \right) \frac{\gamma_n^{\alpha,\beta}(g_0(x))}{g_0(x)} (g_0(x) - \hat{g}_n(x)) dx + O_p(n^{-1/2}).$$

PROOF. For notational simplicity, we shall write γ_n for $\gamma_n^{\alpha,\beta}$ in this proof. Consider all enough large n such that $\frac{n_i}{n} \geq \lambda_i/2 > 0$ for each i so that $\gamma_n^{\alpha,\beta}$ is well defined. From (5.3), by separating the terms in the first sum into s -sample, we have

$$(5.6) \quad \sum_{i=1}^s \frac{n_i}{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \left(-\frac{\gamma_n(\hat{g}_n(X_{ij}))}{\hat{g}_n(X_{ij})} \right) + \sum_{i=1}^s \frac{n_i}{n} \frac{\sum_{j=1}^{n_i} c_{ij} \gamma_n(\hat{g}_n(T_j))}{\int w_i \hat{g}_n} = 0.$$

Note that

$$\left| \int_a^b w_i(x) \gamma_n(\hat{g}_n(x)) dx - \sum_{j=1}^{n_i} c_{ij} \gamma_n(\hat{g}_n(T_j)) \right| \\ = \left| \int_{T_n}^b w_j(x) \gamma_n(\hat{g}_n(x)) dx \right| \\ \leq M \int_{T_n}^b \frac{\hat{g}_n(x) (\sum_{i=1}^s |\alpha_i| \sup_{x \in [a,b]} f_i(x) + |\beta| g_0(a))}{\sum_{j=1}^s \frac{\lambda_j}{2} \inf_{x \in [a,b]} f_j(x)} dx \\ \leq MC'(b - T_n) \hat{g}_n(0+) = O_p(n^{-1}),$$

where $C' \triangleq \frac{\sum_{i=1}^s |\alpha_i| \sup_{x \in [a,b]} f_i(x) + |\beta| g_0(a)}{\sum_{j=1}^s \frac{\lambda_j}{2} \inf_{x \in [a,b]} f_j(x)}$ and the last equality follows from the fact that $b - T_n = O_p(n^{-1})$ and Lemma 5.3. To see that $b - T_n = O_p(n^{-1})$, let $T_j^1, j = 1, \dots, n_1$, be the order statistics from the first sample $X_{1j}, j = 1, \dots, n_1$. Observe that, for some universal constant L ,

$$\mathbb{P}(n_1(b - T_{n_1}^1) > \tilde{M}) = \left(1 - \mathbb{P}\left(X_{11} \in \left[b - \frac{\tilde{M}}{n_1}, b\right]\right) \right)^{n_1} \leq \left(1 - \frac{L\tilde{M}}{n_1} \right)^{n_1}.$$

Hence, $b - T_n \leq b - T_{n_1}^1 = O_p(n_1^{-1}) = O_p(n^{-1})$. Also, note that $1/\int w_i \hat{g}_n = O_p(1)$ by Proposition 4.1(ii) together with the latest result that $|\int_a^b w_i(x) \times$

$\gamma_n(\hat{g}_n(x)) dx - \sum_{j=1}^n c_{ij} \gamma_n(\hat{g}_n(T_j))| = O_p(n^{-1})$, by writing (5.6) in terms of empirical measures \mathbb{F}_i , we obtain

$$\sum_{i=1}^s \frac{n_i}{n} \frac{\int_a^b w_i(x) \gamma_n(\hat{g}_n(x))}{\int w_i \hat{g}_n} dx = \sum_{i=1}^s \frac{n_i}{n} \int_a^b \frac{\gamma_n(\hat{g}_n(x))}{\hat{g}_n(x)} d\mathbb{F}_{i,n_i}(x) + O_p(n^{-1}).$$

By applying a change of variable formula $f_i(x) dx = dF_i(x)$,

$$\begin{aligned} \sum_{i=1}^s \frac{n_i}{n} \int_a^b \frac{\gamma_n(\hat{g}_n(x))}{g_0(x)} dF_i(x) & \frac{\int w_i g_0}{\int w_i \hat{g}_n} \\ & = \sum_{i=1}^s \frac{n_i}{n} \int_a^b \frac{\gamma_n(\hat{g}_n(x))}{\hat{g}_n(x)} d\mathbb{F}_{i,n_i}(x) + O_p(n^{-1}). \end{aligned}$$

By telescoping and rearranging the terms, we obtain

$$(5.7) \quad \sum_{i=1}^s \frac{n_i}{n} \int_a^b \frac{\gamma_n(\hat{g}_n(x))}{g_0(x)} dF_i(x) \frac{\int w_i g_0 - \int w_i \hat{g}_n}{\int w_i \hat{g}_n} = A_1 + A_2 + O_p(n^{-1}),$$

where

$$\begin{aligned} A_1 & \triangleq \sum_{i=1}^s \frac{n_i}{n} \int_a^b \frac{\gamma_n(\hat{g}_n(x))}{\hat{g}_n(x)} d(\mathbb{F}_{i,n_i} - F_i)(x), \\ A_2 & \triangleq \sum_{i=1}^s \frac{n_i}{n} \int_a^b \left(\frac{\gamma_n(\hat{g}_n(x))}{\hat{g}_n(x)} - \frac{\gamma_n(\hat{g}_n(x))}{g_0(x)} \right) dF_i(x). \end{aligned}$$

(i) Let $\mathcal{F} \triangleq \{f : \mathbb{R} \rightarrow \mathbb{R} : f \text{ is decreasing and } 0 \leq f \leq g_0(a)\}$, and $\mathcal{F}_i \triangleq \{f : \mathbb{R} \rightarrow \mathbb{R} : f \equiv C, C \in [\frac{\lambda_i}{2}, 1]\}$, for each $i = 1, \dots, s$, which is a class of constant functions. For A_1 , note that $\frac{\gamma_n(\hat{g}_n(x))}{\hat{g}_n(x)} = \frac{\gamma_n(\tilde{g}_n(x))}{\tilde{g}_n(x)}$ if we define

$$\tilde{g}_n(x) \triangleq \begin{cases} g_0(b) & \text{if } \hat{g}_n(x) < g_0(b), \\ \hat{g}_n(x) & \text{if } g_0(b) \leq \hat{g}_n(x) \leq g_0(a), \\ g_0(a) & \text{if } \hat{g}_n(x) > g_0(a). \end{cases}$$

Clearly, $\{\tilde{g}_n\} \subset \mathcal{F}$. Define

$$(5.8) \quad \Gamma^* \triangleq \left\{ \frac{\gamma^{\alpha,\beta}(f_d, f_1^r, \dots, f_s^r)}{f_d} : f_d \in \mathcal{F}, f_i^r \in \mathcal{F}_i, i = 1, \dots, s \right\}.$$

As the mapping $(x, r_1, \dots, r_s) \mapsto \gamma^{\alpha,\beta}(x, r_1, \dots, r_s)/x$ is uniformly bounded and globally Lipschitz continuous on $(0, \infty) \times [\frac{\lambda_1}{2}, 1] \times \dots \times [\frac{\lambda_s}{2}, 1]$, and the families of functions $\mathcal{F}, \mathcal{F}_1, \dots, \mathcal{F}_s$ are Donsker classes [see Example 2.6.21 in van der Vaart and Wellner (1996)], where all functions in each family are also uniformly bounded, all the conditions in Theorem 2.10.6 in van der Vaart and Wellner (1996) are satisfied and we can conclude that Γ^* is also a Donsker class.

Let $\Gamma \triangleq \{\frac{\gamma_n(\hat{g}_n)}{\hat{g}_n}\}_{n=1}^\infty$. Clearly $\Gamma \subset \Gamma^*$ and so A_1 is $O_p(n^{-1/2})$ as s is finite.

(ii) For A_2 , by telescoping and rearranging the terms

$$\begin{aligned}
 & \sum_{i=1}^s \frac{n_i}{n} \int_a^b \left(\frac{\gamma_n(\hat{g}_n(x))}{g_n(x)} - \frac{\gamma_n(\hat{g}_n(x))}{g_0(x)} \right) dF_i(x) \\
 (5.9) \quad &= \sum_{i=1}^s \frac{n_i}{n} \int_a^b \frac{\gamma_n(\hat{g}_n(x))}{\hat{g}_n(x)g_0(x)} (g_0(x) - \hat{g}_n(x)) \frac{w_i(x)g_0(x)}{\int w_i g_0} dx \\
 &= \sum_{i=1}^s \frac{n_i}{n} \frac{1}{\int w_i g_0} \int_a^b w_i(x) \frac{\gamma_n(\hat{g}_n(x))}{\hat{g}_n(x)} (g_0(x) - \hat{g}_n(x)) dx \\
 &= I_1 + I_2,
 \end{aligned}$$

where

$$\begin{aligned}
 I_1 &\triangleq \sum_{i=1}^s \frac{n_i}{n} \frac{1}{\int w_i g_0} \int_a^b w_i(x) \left(\frac{\gamma_n(\hat{g}_n(x))}{\hat{g}_n(x)} - \frac{\gamma_n(g_0(x))}{g_0(x)} \right) (g_0(x) - \hat{g}_n(x)) dx, \\
 I_2 &\triangleq \sum_{i=1}^s \frac{n_i}{n} \frac{1}{\int w_i g_0} \int_a^b w_i(x) \frac{\gamma_n(g_0(x))}{g_0(x)} (g_0(x) - \hat{g}_n(x)) dx.
 \end{aligned}$$

For I_1 , note that

$$\begin{aligned}
 |I_1| &\leq \sum_{i=1}^s \frac{n_i}{n} \frac{1}{\int w_i g_0} \int_a^b w_i(x) \left| \frac{\gamma_n(\hat{g}_n(x))}{\hat{g}_n(x)} - \frac{\gamma_n(g_0(x))}{g_0(x)} \right| |g_0(x) - \hat{g}_n(x)| dx \\
 &\leq \sum_{i=1}^s \frac{n_i}{n} \frac{MC}{\int w_i g_0} \int_a^b (g_0(x) - \hat{g}_n(x))^2 dx = O_p(n^{-1/2}),
 \end{aligned}$$

by the boundedness of w_i , the Lipschitz continuity of $x \mapsto \gamma_n(x)/x$ and Lemma 5.4. Hence, by (5.7), (5.9), the fact that $A_1 = O_p(n^{-1/2})$ and $I_1 = O_p(n^{-1/2})$,

$$\begin{aligned}
 & \sum_{i=1}^s \frac{n_i}{n} \int_a^b \frac{\gamma_n(\hat{g}_n(x))}{g_0(x)} dF_i(x) \frac{\int w_i g_0 - \int w_i \hat{g}_n}{\int w_i \hat{g}_n} \\
 &= \sum_{i=1}^s \frac{n_i}{n} \frac{1}{\int w_i g_0} \int_a^b w_i(x) \frac{\gamma_n(g_0(x))}{g_0(x)} (g_0(x) - \hat{g}_n(x)) dx + O_p(n^{-1/2}).
 \end{aligned}$$

Changing the variable $dF_i(x) = f_i(x) dx$ on the left-hand side, we obtain

$$\begin{aligned}
 & \sum_{i=1}^s \frac{n_i}{n} \frac{1}{\int w_i g_0 \int w_i \hat{g}_n} \int_a^b w_i(x) \gamma_n(\hat{g}_n(x)) dx \int w_i (g_0 - \hat{g}_n) \\
 &= \int_a^b \left(\sum_{i=1}^s \frac{n_i}{n} \frac{w_i(x)}{\int w_i g_0} \right) \frac{\gamma_n(g_0(x))}{g_0(x)} (g_0(x) - \hat{g}_n(x)) dx + O_p(n^{-1/2}).
 \end{aligned}$$

Telescoping the denominator on the left-hand side and using the fact that $\int w_i(g_0 - \hat{g}_n) = O_p(n^{-1/4})$ from Lemma 5.4(ii), and noting that $\int_a^b w_i(x)\gamma_n(\hat{g}_n(x)) dx \leq MC'(b - a)\hat{g}_n(0+) = O_p(1)$, and $\frac{1}{\int w_i \hat{g}_n} = O_p(1)$ from Proposition 4.1(ii),

$$\begin{aligned}
 (5.10) \quad & \sum_{i=1}^s \frac{n_i}{n} \frac{1}{(\int w_i g_0)^2} \int_a^b w_i(x)\gamma_n(\hat{g}_n(x)) dx \int w_i(g_0 - \hat{g}_n) \\
 &= \int_a^b \left(\sum_{i=1}^s \frac{n_i}{n} \frac{w_i(x)}{\int w_i g_0} \right) \frac{\gamma_n(g_0(x))}{g_0(x)} (g_0(x) - \hat{g}_n(x)) dx + O_p(n^{-1/2}).
 \end{aligned}$$

Fix $\varepsilon > 0$. As $x \mapsto \gamma_n(x)$ is Lipschitz,

$$\begin{aligned}
 & \left| \int_a^b w_i(x)\gamma_n(\hat{g}_n(x)) dx - \int_a^b w_i(x)\gamma_n(g_0(x)) dx \right| \\
 & \leq M \int_a^b |\gamma_n(\hat{g}_n(x)) - \gamma_n(g_0(x))| dx \\
 & \leq MC \int_a^b |\hat{g}_n(x) - g_0(x)| dx \\
 & \leq MC(b - a)^{1/2} \left(\int_a^b (\hat{g}_n(x) - g_0(x))^2 dx \right)^{1/2} \\
 & = O_p(n^{-1/4}),
 \end{aligned}$$

where the last equality follows from Lemma 5.4(iv). In light of this result, without changing the overall order, we can replace $\gamma_n(\hat{g}_n(x))$ by $\gamma_n(g_0(x))$ in the first integral on the left-hand side of (5.10) to obtain (5.5). \square

Denote $S_n \triangleq \sum_{i=1}^s \frac{n_i}{n} f_i \geq m$, Λ_n be the $s \times s$ diagonal matrix with elements $\frac{n_i}{n}$ in order and

$$(5.11) \quad W_n \triangleq \int (f_1 \quad \dots \quad f_s)^\top \frac{1}{S_n} (f_1 \quad \dots \quad f_s).$$

Also define S , W and Λ to be the limiting versions of S_n , W_n and Λ_n , respectively, such that $S \triangleq \sum_{i=1}^s \lambda_i f_i$,

$$W \triangleq \int (f_1 \quad \dots \quad f_s)^\top \frac{1}{S} (f_1 \quad \dots \quad f_s),$$

and Λ is the diagonal matrix $\text{diag}(\lambda_1, \dots, \lambda_s)$.

LEMMA 5.7. (i) For all large enough n , the rank of the matrix $\Lambda_n W_n \Lambda_n - \Lambda_n$ is $s - 1$.

(ii) The rank of the matrix $\Lambda W \Lambda - \Lambda$ is also $s - 1$.

PROOF. We shall only prove part (i), as part (ii) is completely analogous. Consider all large n such that $\frac{n_i}{n} \geq \frac{\lambda_i}{2} > 0$ for each i . Using Theorem 5.19 in [Perlis \(1991\)](#), it suffices to show that $\Lambda_n \bar{W}_n \Lambda_n - \Lambda_n$ is singular and has a $(s - 1) \times (s - 1)$ nonsingular principal submatrix. Denote $\mathbf{1} \triangleq (1, \dots, 1)^\top$ and

$$\bar{W}_n \triangleq \Lambda_n W_n \Lambda_n = \int \begin{pmatrix} \frac{n_1}{n} f_1 & \dots & \frac{n_s}{n} f_s \end{pmatrix}^\top \frac{1}{S_n} \begin{pmatrix} \frac{n_1}{n} f_1 & \dots & \frac{n_s}{n} f_s \end{pmatrix}.$$

Direct computation gives $(\bar{W}_n - \Lambda_n)\mathbf{1} = \mathbf{0}$. Therefore, $\bar{W}_n - \Lambda_n$ has eigenvalue $\mathbf{0}$, implying that it is singular. Now, let $\Lambda_{n,b} \triangleq \text{diag}(\frac{n_1}{n}, \dots, \frac{n_{s-1}}{n})$ and

$$\bar{W}_{n,b} \triangleq \int \begin{pmatrix} \frac{n_1}{n} f_1 & \dots & \frac{n_{s-1}}{n} f_{s-1} \end{pmatrix}^\top \frac{1}{S_n} \begin{pmatrix} \frac{n_1}{n} f_1 & \dots & \frac{n_{s-1}}{n} f_{s-1} \end{pmatrix}.$$

Clearly, $\bar{W}_{n,b} - \Lambda_{n,b}$ is a principal submatrix of $\bar{W}_n - \Lambda_n$. For any $\mathbf{0} \neq \mathbf{u} \in \mathbb{R}^{s-1}$, $\mathbf{u}^\top \Lambda_{n,b} \mathbf{u} = \sum_{i=1}^{s-1} u_i^2 \frac{n_i}{n}$. Applying the Cauchy–Schwarz inequality to $(\sum_{i=1}^{s-1} \frac{n_i}{n} u_i f_i(x))^2$, we obtain

$$\begin{aligned} \mathbf{u}^\top \bar{W}_{n,b} \mathbf{u} &= \int_a^b \frac{(\sum_{i=1}^{s-1} \frac{n_i}{n} u_i f_i(x))^2}{\sum_{k=1}^s \frac{n_k}{n} f_k(x)} dx \\ &\leq \int_a^b \frac{(\sum_{i=1}^{s-1} \frac{n_i}{n} u_i^2 f_i(x)) (\sum_{i=1}^{s-1} \frac{n_i}{n} f_i(x))}{\sum_{k=1}^s \frac{n_k}{n} f_k(x)} dx \\ &< \int_a^b \left(\sum_{i=1}^{s-1} \frac{n_i}{n} u_i^2 f_i(x) \right) \cdot 1 dx \\ &= \sum_{i=1}^{s-1} \frac{n_i}{n} u_i^2 \int_a^b f_i(x) dx = \mathbf{u}^\top \Lambda_{n,b} \mathbf{u}, \end{aligned}$$

where the strict inequality follows as $\sum_{i=1}^{s-1} \frac{n_i}{n} u_i^2 f_i \neq 0$ whenever $\mathbf{u} \neq \mathbf{0}$ and $\sum_{i=1}^{s-1} \frac{n_i}{n} f_i(x) < \sum_{k=1}^s \frac{n_k}{n} f_k(x)$ for all $x \in [a, b]$ as $\frac{n_s}{n} > 0$. Therefore, $\mathbf{u}^\top (\bar{W}_{n,b} - \Lambda_{n,b}) \mathbf{u} < 0$ whenever $\mathbf{u} \neq \mathbf{0}$. Hence, $\bar{W}_{n,b} - \Lambda_{n,b}$ is negative definite and so it is of full rank, $s - 1$. \square

PROPOSITION 5.8. For each $i = 1, \dots, s$, we have

$$(5.12) \quad \int w_i \hat{g}_n - \int w_i g_0 = O_p(n^{-1/2}).$$

PROOF. It suffices to consider all large n such that $\frac{n_i}{n} \geq \lambda_i/2 > 0$ for each i . For a sequence of matrix B_n , we denote B_n converges to B in the usual matrix norm by $B_n \rightarrow B$. Let $\boldsymbol{\eta}_n \triangleq (\frac{\int w_1(g_0 - \hat{g}_n)}{\int w_1 g_0}, \dots, \frac{\int w_s(g_0 - \hat{g}_n)}{\int w_s g_0})^\top$. We split the proof into three steps:

(i) We first provide some useful identities. For some constants $\alpha_{ki}, k = 1, \dots, s, i = 1, \dots, s$, define $\gamma_{n,k}(\cdot) \triangleq \gamma^{\alpha_{k,0}}(\cdot, \frac{n_1}{n}, \dots, \frac{n_s}{n})$ for each $k = 1, \dots, s$. That is,

$$\gamma_{n,k}(x) = \begin{cases} \frac{x \sum_{i=1}^s \alpha_{ki} f_i(b)}{\sum_{j=1}^s \frac{n_j}{n} f_j(b)} & \text{for } 0 < x < g_0(b), \\ \frac{x \sum_{i=1}^s \alpha_{ki} f_i(g_0^{-1}(x))}{\sum_{j=1}^s \frac{n_j}{n} f_j(g_0^{-1}(x))} & \text{for } g_0(b) \leq x \leq g_0(a), \\ \frac{x \sum_{i=1}^s \alpha_{ki} f_i(a)}{\sum_{j=1}^s \frac{n_j}{n} f_j(a)} & \text{for } x > g_0(a). \end{cases}$$

Recall that $S_n = \sum_{k=1}^s \frac{n_k}{n} f_k$. Clearly, we have some systematic representations for the terms in (5.5) as follows:

$$(5.13) \quad \int_a^b w_i(x) \gamma_{n,k}(g_0(x)) dx = \alpha_{k1} \int_a^b \frac{w_i(x) g_0(x) f_1(x)}{S_n(x)} dx + \dots + \alpha_{ks} \int_a^b \frac{w_i(x) g_0(x) f_s(x)}{S_n(x)} dx,$$

and for $x \in [a, b]$,

$$(5.14) \quad \left(\sum_{i=1}^s \frac{n_i}{n} \frac{w_i(x)}{f w_i g_0} \right) \frac{\gamma_{n,k}(g_0(x))}{g_0(x)} = \alpha_{k1} \frac{w_1(x)}{f w_1 g_0} + \dots + \alpha_{ks} \frac{w_s(x)}{f w_s g_0}.$$

Note that $\int w_i g_0$ and n_i/n are nonzero. Hence, we can find scalar $d_{ki}^{(n)}$'s such that

$$(5.15) \quad \int_a^b w_i(x) \gamma_{n,k}(g_0(x)) dx = d_{ki}^{(n)} \frac{\int w_i g_0}{n_i/n} \quad \text{for } i, k = 1, \dots, s.$$

Define $A \triangleq (\alpha_{ki})_{s \times s}, D_n \triangleq (d_{ki}^{(n)})_{s \times s}$. From (5.13) and (5.15), we have

$$(5.16) \quad W_n A^\top = \Lambda_n^{-1} D_n^\top,$$

where $W_n = W_n^\top$ is defined in (5.11) and $\Lambda_n = \text{diag}(\frac{n_1}{n}, \dots, \frac{n_s}{n})$.

(ii) Let $V_n \triangleq D_n - A$. Combining (5.5), (5.14) and (5.15), we have

$$V_n \eta_n = \left(O_p(n^{-1/2}) \quad \dots \quad O_p(n^{-1/2}) \right)^\top.$$

If V_n were nonsingular, the result would follow immediately; however, using (5.16), for a nonsingular choice of $A, V_n = A \Lambda_n^{-1} (\Lambda_n W_n \Lambda_n - \Lambda_n)$ exactly has rank $s - 1$ by Lemma 5.7, therefore, V_n is rank deficient. On the other hand, in the proof of Lemma 5.7, we know that the principal submatrix obtained by deleting the last row and last column of $\Lambda_n W_n \Lambda_n - \Lambda_n$ is nonsingular. This implies that the first $s - 1$ rows of $\Lambda_n W_n \Lambda_n - \Lambda_n$ are linearly independent. Choose A to be the identity matrix, then since Λ_n is diagonal, it follows that the first $s - 1$ rows of V_n are also linearly independent. Denote $v_{n,1}^\top, \dots, v_{n,s-1}^\top$ to be the first $s - 1$ rows

of V_n . To complete the proof, we construct a vector that is linearly independent of $(\mathbf{v}_{n,1}^\top, \dots, \mathbf{v}_{n,s-1}^\top)$.

(iii) Let $\gamma_{n,0}(\cdot) \triangleq \gamma^{0,1}(\cdot, \frac{n_1}{n}, \dots, \frac{n_s}{n})$. That is,

$$(5.17) \quad \gamma_{n,0}(x) = \begin{cases} \frac{xg_0(b)}{\sum_{j=1}^s \frac{n_j}{n} f_j(b)} & \text{for } 0 < x < g_0(b), \\ \frac{x^2}{\sum_{j=1}^s \frac{n_j}{n} f_j(g_0^{-1}(x))} & \text{for } g_0(b) \leq x \leq g_0(a), \\ \frac{xg_0(a)}{\sum_{j=1}^s \frac{n_j}{n} f_j(a)} & \text{for } x > g_0(a). \end{cases}$$

Clearly,

$$(5.18) \quad \int_a^b w_i(x) \gamma_{n,0}(g_0(x)) dx = \int_a^b \frac{w_i(x) g_0^2(x)}{S_n(x)} dx,$$

and for $x \in [a, b]$,

$$(5.19) \quad \left(\sum_{i=1}^s \frac{n_i}{n} \frac{w_i(x)}{\int w_i g_0} \right) \frac{\gamma_{n,0}(g_0(x))}{g_0(x)} \equiv 1.$$

Let $\mathbf{q}_n \triangleq (\frac{n_1}{n} \int \frac{f_1 g_0}{S_n}, \dots, \frac{n_s}{n} \int \frac{f_s g_0}{S_n})^\top$. Following (5.18), (5.19) and the fact that $\int g_0(x) dx = \int \hat{g}_n(x) dx = 1$, (5.5) becomes $\mathbf{q}_n^\top \boldsymbol{\eta}_n = O_p(n^{-1/2})$. We finally claim that $\mathbf{v}_{n,1}, \dots, \mathbf{v}_{n,s-1}, \mathbf{q}_n$ form s linearly independent vectors in \mathbb{R}^s ; indeed, if

$$c_1 \mathbf{v}_{n,1} + \dots + c_{s-1} \mathbf{v}_{n,s-1} + c_s \mathbf{q}_n = \mathbf{0},$$

then, in particular, we have

$$c_1 \langle \mathbf{v}_{n,1}, \mathbf{1} \rangle + \dots + c_{s-1} \langle \mathbf{v}_{n,s-1}, \mathbf{1} \rangle + c_s \langle \mathbf{q}_n, \mathbf{1} \rangle = 0,$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product on Euclidean space. In the proof of Lemma 5.7, we know that $(\Lambda_n W_n \Lambda_n - \Lambda_n) \mathbf{1} = \mathbf{0}$. As $V_n = A \Lambda_n^{-1} (\Lambda_n W_n \Lambda_n - \Lambda_n)$, hence as row vectors of V_n , $\langle \mathbf{v}_{n,i}, \mathbf{1} \rangle = 0$ for each $i = 1, \dots, s - 1$. Finally, as $\langle \mathbf{q}_n, \mathbf{1} \rangle = \sum_{i=1}^s \frac{n_i}{n} \int_a^b \frac{f_i g_0}{S_n} \geq \sum_{i=1}^s \frac{\lambda_i}{2} \frac{m}{M} > 0$, we know $c_s = 0$. Since $\mathbf{v}_{n,1}, \dots, \mathbf{v}_{n,s-1}$ are linearly independent, we know that $c_1 = \dots = c_{s-1} = 0$. Hence, $\mathbf{v}_{n,1}, \dots, \mathbf{v}_{n,s-1}, \mathbf{q}_n$ are linearly independent. Now, consider

$$(5.20) \quad (\mathbf{v}_{n,1} \ \dots \ \mathbf{v}_{n,s-1} \ \mathbf{q}_n)^\top \boldsymbol{\eta}_n = \left(O_p(n^{-1/2}) \ \dots \ O_p(n^{-1/2}) \right)^\top.$$

Note that $V_n = A W_n \Lambda_n - A$. As the integrands in W_n are bounded, $W_n \rightarrow W$ by bounded convergence theorem. Also, $\Lambda_n \rightarrow \Lambda \triangleq \text{diag}(\lambda_1, \dots, \lambda_s)$. Therefore, $V_n \rightarrow A \Lambda (\Lambda^{-1} W \Lambda - \Lambda)$. Similar to the discussion of V_n , we see that the first $s - 1$ rows of V are linearly independent. Moreover, $\mathbf{q}_n \rightarrow \mathbf{q} \triangleq (\lambda_1 \int \frac{f_1 g_0}{S}, \dots, \lambda_s \int \frac{f_s g_0}{S})^\top$. Using the same argument for the linear independence of $\{\mathbf{v}_{n,1}, \dots, \mathbf{v}_{n,s-1}, \mathbf{q}_n\}$, we see that $\{v_1, \dots, v_{s-1}, \mathbf{q}\}$ is also linearly independent.

In addition, $(\mathbf{v}_{n,1} \cdots \mathbf{v}_{n,s-1} \mathbf{q}_n)^\top \rightarrow (v_1, \dots, v_{s-1}, q)^\top$. By their nonsingularities, the convergence of the inverses is warranted:

$$((\mathbf{v}_{n,1} \cdots \mathbf{v}_{n,s-1} \mathbf{q}_n)^\top)^{-1} \rightarrow ((v_1, \dots, v_{s-1}, q)^\top)^{-1}.$$

As a finitely linear combination of terms of the common order $O_p(n^{-1/2})$, each component of $\boldsymbol{\eta}_n$ preserves the same order of convergence. Hence, the result follows. \square

LEMMA 5.9. *For each $i = 1, \dots, s$, we have*

$$\left| \sum_{j=1}^n g_0(T_j)c_{ij} - \int w_i g_0 \right| = O_p(n^{-1/2}).$$

PROOF. See the details provided in Section 10 of the Supplementary Material [Chan et al. (2018)]. \square

COROLLARY 5.10. *For each $i = 1, \dots, s$, we have*

$$\left| \sum_{j=1}^n g_0(T_j)c_{ij} - \int w_i \hat{g}_n \right| = O_p(n^{-1/2}).$$

PROOF. The claim follows directly from Proposition 5.8 and Lemma 5.9. \square

5.2. *Rate of convergence of the monotone MLE \hat{g}_n .* To establish the asymptotic equivalence of \tilde{G}_{n,\hat{g}_n} and \tilde{G}_{n,g_0} and that of \tilde{U}_{n,\hat{g}_n} and \tilde{U}_{n,g_0} , the local consistency of \hat{g}_n at an appropriate rate, particularly of order around $1/3$ if not too less (see Proposition 5.13), is necessary in addition to the result $\sqrt{n}(\sum_{j=1}^n g_0(T_j)c_{ij} - \int w_i \hat{g}_n) = O_p(1)$. Essentially, for each target point t_0 , we want to show that certain events have arbitrarily small probability. To this end, we first aim to show that there is an arbitrarily high probability that \hat{g}_n has a jump in an interval of the form $(t_0 - R, t_0 - Cn^{-1/3})$ for large enough n . This result can be ensured by the uniform consistency of \hat{g}_n in Proposition 4.1(iv).

Then, by considering a sample point for which \hat{g}_n has a jump in that interval, the key inequality in proving Proposition 5.13 is (10.23) in the Supplementary Material [Chan et al. (2018)], which is similar to the one in the proof of Lemma 2.1 in Banerjee (2007) and to (5.20) in Groeneboom and Wellner (1992). All of these inequalities could actually be obtained by considering Karush–Kuhn–Tucker optimality conditions; also see (8.2) and (8.3) in the Supplementary Material [Chan et al. (2018)]. The inequalities in Banerjee (2007) and Groeneboom and Wellner (1992) involve expressions that contain estimators $\hat{\psi}_n(z)$ and $\hat{F}_n(T_i)$, which can be bounded by the same expression with $\psi(z_0)$ and $F_0(t_0)$ in place of $\hat{\psi}_n(z)$ and $\hat{F}_n(T_i)$, respectively. However, we can only replace $\hat{g}_n(T_j)$ by $g_0(t_0)$ in the first

term of the corresponding expression in (10.23) as there is no direct comparison between the magnitudes of $\int w_i \hat{g}_n$ and $\int w_i g_0$, in which the former appears in the second term of that expression. Nevertheless, by using the \sqrt{n} -convergence of $\int w_i \hat{g}_n$ as explained before, we are still able to show that the event considered in (10.24) has an arbitrarily small probability; see Lemma 5.11. The proof of the local consistency of \hat{g}_n with $n^{-1/3}$ -rate then follows from the arguments as that developed in Groeneboom and Wellner (1992) and Banerjee (2007).

LEMMA 5.11. *For any $\varepsilon > 0$, there exist $C_0 > 0$ and $R_0 > 0$ such that for any $C \geq C_0$ and $0 < R \leq R_0$, we have*

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

for all sufficiently large n ; here, $I_n \triangleq (t_0 - R, t_0 - Cn^{-1/3}]$.

PROOF. See the details provided in Section 10 of the Supplementary Material [Chan et al. (2018)]. □

LEMMA 5.12. *For any $\varepsilon > 0$, there exist $C_1 > 0$ and $R_1 > 0$ such that for any $C \geq C_1$ and $R \leq R_1$, we have*

$$\mathbb{P}\left(\inf_{t \in \tilde{I}_n} \sum_{j: t_0 - 2Cn^{-1/3} \leq T_j < t} \left(-\frac{1}{ng_0(t_0 - 2Cn^{-1/3})} + \sum_{i=1}^s \frac{n_i}{n} \frac{c_{ij}}{\int w_i \hat{g}_n}\right) \leq 0\right) \leq \varepsilon,$$

for all sufficiently large n ; here $\tilde{I}_n \triangleq [t_0 - Cn^{-1/3}, t_0 + R)$.

PROOF. See the details provided in Section 10 of the Supplementary Material [Chan et al. (2018)]. □

PROPOSITION 5.13. *For any $K_1 > 0$, we have*

$$\sup_{h \in [-K_1, K_1]} |\hat{g}_n(t_0 + hn^{-1/3}) - g_0(t_0)| = O_p(n^{-1/3}).$$

PROOF. See the details provided in Section 10 of the Supplementary Material [Chan et al. (2018)]. □

LEMMA 5.14. *Given $\varepsilon > 0$ and $\tilde{C} > 0$, there exist $\tilde{D} > 0$ and $\tilde{R} > 0$ such that for large enough n ,*

$$\mathbb{P}\left(\sup_{t \in \bar{I}_n} \sum_{j: t \leq T_j < t_0 - 2\tilde{C}n^{-1/3}} \left(-\frac{1}{ng_0(t_0 - 2\tilde{C}n^{-1/3})} + \sum_{i=1}^s \frac{n_i}{n} \frac{c_{ij}}{\int w_i \hat{g}_n}\right) \geq 0\right) \leq \varepsilon;$$

here $\bar{I}_n \triangleq (t_0 - \tilde{R}, t_0 - (2\tilde{C} + \tilde{D})n^{-1/3}]$.

PROOF. See the details provided in Section 10 of the Supplementary Material [Chan et al. (2018)]. \square

LEMMA 5.15. Fix $K_2 > 0$. Consider an interval of the form $[t_0 - K_2 n^{-1/3}, t_0 + K_2 n^{-1/3}]$. Let τ_n^- and τ_n^+ be the two points corresponding to the last change of slope of $\hat{g}_n \leq t_0 - K_2 n^{-1/3}$ and the first change of slope $\hat{g}_n \geq t_0 + K_2 n^{-1/3}$, respectively. Then $\tau_n^- - t_0 = O_p(n^{-1/3})$ and $\tau_n^+ - t_0 = O_p(n^{-1/3})$.

PROOF. See the details provided in Section 10 of the Supplementary Material [Chan et al. (2018)]. \square

6. Asymptotic distribution of the monotone MLE. Let $\mathbb{B}_{\text{loc}}(\mathbb{R})$ denote the space of all locally bounded real functions on \mathbb{R} endowed with the topology of uniform convergence on compacta.

6.1. *Asymptotic distributions of \tilde{G}_{n,\hat{g}_n} and \tilde{U}_{n,\hat{g}_n} .* In Lemmas 6.1 and 6.2, we first establish the asymptotic equivalence of \tilde{G}_{n,\hat{g}_n} and \tilde{G}_{n,g_0} and the asymptotic equivalence of \tilde{U}_{n,\hat{g}_n} and \tilde{U}_{n,g_0} , respectively.

LEMMA 6.1. For every $K > 0$,

$$\sup_{t \in [-K, K]} |\tilde{G}_{n,g_0}(t) - \tilde{G}_{n,\hat{g}_n}(t)| \xrightarrow{\mathbb{P}} 0.$$

PROOF. See the details provided in Section 11 in the Supplementary Material [Chan et al. (2018)]. \square

LEMMA 6.2. For every $K > 0$,

$$\sup_{t \in [-K, K]} |\tilde{U}_{n,g_0}(t) - \tilde{U}_{n,\hat{g}_n}(t)| \xrightarrow{\mathbb{P}} 0.$$

PROOF. See the details provided in Section 11 in the Supplementary Material [Chan et al. (2018)]. \square

In showing the asymptotic equivalence of \tilde{U}_{n,\hat{g}_n} and \tilde{U}_{n,g_0} in Lemma 6.2, the choice of $\frac{1}{\hat{z}_j^2}$ instead of $\frac{\partial^2 \psi_n}{\partial z_j^2}(\hat{z})$ for the diagonal elements of the positive definite matrix used in the proof of Proposition 3.2 will lead to incomplete cancellation of terms when applying Taylor’s expansion theorem. Therefore, we require the results in Section 5 in order to complete the proof. In addition, the function $\phi(x, \cdot)$, which is the negative log-likelihood of a single datum, as previously mentioned, appears in $\tilde{B}_{n,\hat{\psi}_n}$ and $\tilde{B}_{n,\psi}$ of Lemma 2.4 in Banerjee (2007)

as a function of one variable; while our presently proposed ψ_n , the negative log-likelihood function of all the data, appears in \tilde{U}_{n,\hat{g}_n} and \tilde{U}_{n,g_0} as a function of n variables. This is because unlike the monotone response model considered in Banerjee (2007), where separability of the arguments in the log-likelihood function can be achieved, such a mathematical simplification does not appear in our present problem. Indeed, in Banerjee (2007), the negative of the log-likelihood function is $\sum_{i=1}^n \phi(X_i, \psi(Z_i))$ while the corresponding expression in our case $\psi_n(z_1, \dots, z_n) = -\sum_{j=1}^n \log z_j + \sum_{i=1}^s n_i \log(\sum_{k=1}^n z_k c_{ik})$. As a result, unlike in the proof of Lemma 2.4 in Banerjee (2007), where algebraic cancellation of some terms could be ensured when applying Taylor’s series expansion theorem, we are actually confronted with more terms, where the determination of their orders could not be resolved by applying any common approach available in the literature. More specifically, the minimal possible order $o_p(n^{-1/3})$ of the term $\sum_{k=1}^n g_0(T_k) c_{ik} - \int w_i \hat{g}_n$ would be needed. Corollary 5.10 fills this important gap.

In light of the asymptotic equivalence of \tilde{G}_{n,\hat{g}_n} and \tilde{G}_{n,g_0} and that of \tilde{U}_{n,\hat{g}_n} and \tilde{U}_{n,g_0} , it suffices to find out the asymptotic distributions of \tilde{G}_{n,g_0} and \tilde{U}_{n,g_0} , which will be established in the following Lemmas 6.3 to 6.5, respectively, by using arguments as developed in van der Vaart and Wellner (1996).

LEMMA 6.3. *The process $\tilde{G}_{n,g_0}(t)$ converges uniformly in probability to the identity function on any compact interval $[-K, K]$, for any $K > 0$.*

PROOF. See the details provided in Section 11 in the Supplementary Material [Chan et al. (2018)]. □

Define

$$q_{n,t}(x) \triangleq n^{1/6} \left(\frac{1}{g_0(x)} + \frac{g_0(x) - g_0(t_0)}{g_0^2(x)} \right) [1(x \leq t_0 + tn^{-1/3}) - 1(x \leq t_0)].$$

Simple calculation leads that $\lambda \tilde{U}_{n,g_0}(t) = A_1(t) + A_2(t) + A_3(t)$, where

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

Let $l^\infty[-K, K]$ be the space of uniformly bounded functions on $[-K, K]$ equipped with the topology of uniform convergence. Let $W(t)$, $t \in \mathbb{R}$, be the standard two-sided Brownian motion with $W(0) = 0$.

LEMMA 6.4. *The process $A_1(t)$ converges weakly, in $\mathbb{B}_{\text{loc}}(\mathbb{R})$, to the process $\lambda^{1/2}W(t)$.*

PROOF. See the details provided in Section 11 in the Supplementary Material [Chan et al. (2018)]. \square

LEMMA 6.5. *\tilde{U}_{n,g_0} converges weakly in $\mathbb{B}_{\text{loc}}(\mathbb{R})$ to the process U defined by*

$$U(t) \triangleq \frac{1}{\lambda^{1/2}}W(t) - \frac{|g'(t_0)|}{2}t^2, \quad t \in \mathbb{R}.$$

PROOF. See the details provided in Section 11 in the Supplementary Material [Chan et al. (2018)]. \square

6.2. *Main theorem.* Let $\mathcal{L}_{\text{loc}}^2(\mathbb{R}) \triangleq \{\phi : \int_{-c}^c \phi^2(t) dt < \infty \text{ for all } c > 0\}$, with the topology of L^2 -convergence on compacta. For $\alpha, \beta > 0$, define the process $\mathbb{X}_{\alpha,\beta}(t) \triangleq \alpha W(t) - \beta t^2$, $t \in \mathbb{R}$. Let $G_{\alpha,\beta}$ denote the LCM of $\mathbb{X}_{\alpha,\beta}$ and $g_{\alpha,\beta}$ the left derivative of $G_{\alpha,\beta}$. Denote also $\mathbb{X}_n(t) \triangleq n^{1/3}(\hat{g}_n(t_0 + tn^{-1/3}) - g_0(t_0))$, $a^* \triangleq \lambda^{-1/2}$ and $b^* \triangleq |g'_0(t_0)|/2$.

Our main theorem, Theorem 6.6, can be proven using continuous-mapping arguments for slopes of least concave majorant estimators as illustrated by Banerjee (2007). The main ingredients of the proof will be the asymptotic distributions of \tilde{G}_{n,\hat{g}_n} and \tilde{U}_{n,\hat{g}_n} discussed in Section 6.1. Finally, the asymptotic distribution of \hat{g}_n at $t_0 \in (a, b)$ is a direct consequence of Theorem 6.6 and follows the argument as in Banerjee (2007); see Theorem 1.1.

THEOREM 6.6. *Under Assumptions 2.1, $\mathbb{X}_n(t) \xrightarrow{\mathbb{D}} g_{a^*,b^*}(t)$ finite dimensionally and also in the space $\mathcal{L}_{\text{loc}}^2(\mathbb{R})$.*

PROOF. See the details provided in Section 11 in the Supplementary Material [Chan et al. (2018)]. \square

As in equations (6.7)–(6.9) in Banerjee and Wellner (2001), it is easy to see that

$$g_{a^*,b^*}(t) \stackrel{\mathbb{D}}{=} a^*(b^*/a^*)^{1/3} \cdot g_{1,1}((b^*/a^*)^{2/3}t),$$

as a processes indexed by $t \in \mathbb{R}$. Using switch relationship, we also know that $g_{1,1}(0) \stackrel{\mathbb{D}}{=} 2 \arg \max_t \{W(t) - t^2\}$. Hence, we obtain Theorem 1.1.

7. Discussion.

7.1. *Comparison with Groeneboom (1996).* A comprehensive piece of work that also deals with nonseparated log-likelihoods in monotone function models is Groeneboom (1996), which studied the estimation of survival functions under case-2 interval censoring. Here we provide a detailed comparison between the approach in Groeneboom (1996) and ours:

1. The indicator function $1_{[0,\cdot)}$ in Lemma 4.4 and g in Corollary 4.3 in Groeneboom (1996) play similar roles as w_i in our Proposition 5.8 in the sense that they appear in certain linear functionals whose rates of convergence are demanded for establishing the asymptotic distribution of the corresponding estimators at a fixed interior point. The $1_{[0,\cdot)}$ (resp., g) in Groeneboom (1996) are quite arbitrary and the corresponding statement in Lemma 4.4 (resp., Corollary 4.3) is valid uniformly in t (resp., g in a suitable class of functions). On the other hand, as long as Assumptions 2.1 are satisfied, our Proposition 5.8 holds. Therefore, there is also flexibility for the choices of w_i but the problem setting requires them to be fixed at the first place. In principle, $\sqrt{n}(\int w \hat{g}_n - \int w g_0) = O_p(1)$ should also hold uniformly in w over a suitable class of functions; however, this is not of our primary concern in this article and, therefore, we do not discuss in details.

2. The proof leading to Lemma 4.4 in Groeneboom (1996) and that for our Proposition 5.8 are very different. In his case, due to a missing data structure, integral equation (3.8) is constructed by considering the score operator and its adjoint, but we do not have such a missing data structure in our present problem. Instead, we make use of the Karush–Kuhn–Tucker conditions to prove Lemma 5.5 and its more useful corollary Lemma 5.6. In particular, Groeneboom (1996) uses (4.29) to define $\tilde{\theta}_{t,F}$ that links the linear functional of interest as indicated in (4.36). This is similar in spirit to our $\gamma_n^{\alpha,\beta}$ in Lemma 5.6 and $\gamma_{n,k}$ in Lemma 5.8. While $\tilde{\theta}_{t,F}$ serves as a transformation of $1_{[0,t)}$ and g through the solution to the integral equation (4.29); our $\gamma_{n,k}$ in Lemma 5.8 is defined so that we can form an invertible matrix transformation to recover the rate of convergence of the linear functional $\int w_i \hat{g}_n$ due to our multiple samples mechanism.

3. Moreover, in the proof of Lemma 4.4 in Groeneboom (1996), two crucial estimates are required, namely (4.39) being obtained using the L_2 -rate of convergence of \hat{F}_n and (4.42) being a consequence of Donsker class of the set of functions involving $\tilde{\theta}_{t,F}$. In our case, the corresponding results would be Lemma 5.4(iv) for (4.39) and the fact that a class of functions related to $\gamma_n^{\alpha,\beta}$ is a Donsker class; see (5.8), for (4.42). In particular, in deriving the (squared) L_2 -rate of convergence of \hat{g}_n from its rate of convergence in Hellinger distance in Lemma 5.4, the crucial fact that $\hat{g}_n(a+) = O_p(1)$ is required (Lemma 5.3).

4. After proving Lemma 4.4 and Corollary 4.3 in Groeneboom (1996), the remaining steps to derive the asymptotic distribution of the estimator at a fixed interior point are relatively similar to other monotone-constrained estimation problems

such as that in [Groeneboom and Wellner \(1992\)](#) and [Banerjee \(2007\)](#). In particular, the proofs of Lemma 4.5, (4.56) and (4.57) are analogous to our proofs of Lemmas 5.11 and 5.12; Lemma 4.6 in [Groeneboom \(1996\)](#) is analogous to Proposition 5.13; Theorem 4.4 in [Groeneboom \(1996\)](#) corresponds to the derivations in our Section 6.

7.2. Concluding remarks. In this article, we study nonparametric estimation of a decreasing density function g_0 in a s -sample biased sampling model, and provide the existence, uniqueness, self-characterization, consistency, rates of convergence and asymptotic distribution of the maximum likelihood estimator at a fixed interior point. The major challenges come from nonseparability and a lack of strictly positive second-order derivatives of the negative of the log-likelihood function. We have developed notable arguments to establish the tightness of the monotone MLE and the rate of convergence of the linear functionals of the estimator, which are key ingredients to complete the proof of asymptotic distribution.

The self-characterization of the monotone MLE suggests an iterative algorithm to compute the MLE. An initial estimator can be obtained as the slope of the least concave majorant of the distribution function estimator of [Vardi \(1985\)](#), and an update of the estimator is defined as the solution of the right-hand side (3.5). These updated values will then serve as the initial values for the next iteration and the procedure will continue iteratively until convergence.

Kernel smoothing is an alternative approach to density estimation. In comparison to kernel smoothing, which typically requires selection of a bandwidth parameter, an advantage of the monotone MLE is that it can be defined and computed unambiguously without introducing a smoothing parameter. The price paid is the monotonicity assumption of the density function. In one-sample estimation with an unbiased sample, [Jankowski \(2014\)](#) recently developed the asymptotic distribution of the Grenander estimator under misspecification of the monotone density assumption, and the extension to s -sample biased sampling models will be studied in the future.

Acknowledgments. The authors thank Professor Ed George, an Associate Editor and a reviewer for their helpful suggestions that improve various aspects of the manuscript. We also thank Professor Jon Wellner for his insightful comments and warm encouragement, Professor El Barmi's inspiring comments to the preliminary version and Professor Ian McKeague's comments to the revised manuscript. The second author acknowledges the financial supports from the Chinese University of Hong Kong, the UGC of HKSAR and Columbia University; the present work constitutes part of his research study leading to his MPhil. thesis in CUHK. The fourth author, Phillip Yam also expresses his gratitude to the financial support of Department of Statistics in Columbia University in the City of New York over the period when he was a visiting faculty member.

SUPPLEMENTARY MATERIAL

Supplement to “Estimation of a monotone density in s -sample biased sampling models” (DOI: [10.1214/17-AOS1614SUPP](https://doi.org/10.1214/17-AOS1614SUPP); .pdf). In the supplementary paper, we provide the proofs for Propositions 3.1, 3.2, 4.1 and 5.13, Lemmas 5.1, 5.2, 5.4, 5.5, 5.9, 5.11, 5.12, 5.14, 5.15, 6.1, 6.2, 6.3, 6.4 and 6.5, Theorems 1.1 and 6.6. In addition, we also state and prove the fact that the function $\tilde{\mathcal{L}}_n$ defined in (3.3) is concave in \mathbf{p} in Proposition 8.1, and hence establishes the unique existence of \hat{g}_n in Proposition 8.2.

REFERENCES

- BANERJEE, M. (2007). Likelihood based inference for monotone response models. *Ann. Statist.* **35** 931–956. [MR2341693](#)
- BANERJEE, M. and WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29** 1699–1731. [MR1891743](#)
- BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M. and BRUNK, H. D. (1972). *Statistical Inference Under Order Restrictions*. Wiley, New York.
- CHAN, K. C. G. and WANG, M.-C. (2012). Estimating incident population distribution from prevalent data. *Biometrics* **68** 521–531. [MR2959619](#)
- CHAN, K. C. G., LING, H. K., SIT, T. and YAM, S. C. P. (2018). Supplement to “Estimation of a monotone density in s -sample biased sampling models.” DOI:[10.1214/17-AOS1614SUPP](https://doi.org/10.1214/17-AOS1614SUPP).
- COOK, R. C. and MARTIN, F. B. (1974). A model for quadrat sampling with visibility bias. *J. Amer. Statist. Assoc.* **69** 345–349.
- COX, D. R. (1968). Some sampling problems in technology. In *New Developments in Survey Sampling* (N. L. Johnson and H. Smith, eds.) 506–527. Wiley, New York.
- DAVIDOV, O. and ILIOPOULOS, G. (2009). On the existence and uniqueness of the NPMLE in biased sampling models. *J. Statist. Plann. Inference* **139** 176–183. [MR2473996](#)
- DRUMMER, T. D. and McDONALD, L. L. (1987). Size bias in line transect sampling. *Biometrics* **43** 13–21.
- DÜMBGEN, L., WELLNER, J. A. and WOLFF, M. (2016). A law of the iterated logarithm for Grenander’s estimator. *Stochastic Process. Appl.* **126** 3854–3864. [MR3565482](#)
- EL BARMÍ, H. and NELSON, P. I. (2002). A note on estimating a non-increasing density in the presence of selection bias. *J. Statist. Plann. Inference* **107** 353–364. [MR1927774](#)
- GILL, R. D., VARDI, Y. and WELLNER, J. A. (1988). Large sample theory of empirical distributions in biased sampling models. *Ann. Statist.* **16** 1069–1112. [MR0959189](#)
- GRENANDER, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153. [MR0093415](#)
- GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II* (Berkeley, Calif., 1983) 539–555. Wadsworth, Belmont, CA. [MR0822052](#)
- GROENEBOOM, P. (1996). Lectures on inverse problems. In *Lectures on Probability Theory and Statistics (Saint-Flour, 1994)*. *Lecture Notes in Math.* **1648** 67–164. Springer, Berlin. [MR1600884](#)
- GROENEBOOM, P., HOOGHIEMSTRA, G. and LOPUHAÁ, H. P. (1999). Asymptotic normality of the L_1 error of the Grenander estimator. *Ann. Statist.* **27** 1316–1347. [MR1740109](#)
- GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation Under Shape Constraints: Estimators, Algorithms and Asymptotics*. *Cambridge Series in Statistical and Probabilistic Mathematics* **38**. Cambridge Univ. Press, New York. [MR3445293](#)
- GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. *DMV Seminar* **19**. Birkhäuser, Basel. [MR1180321](#)

- HAUSMAN, J. A. and WISE, D. A. (1981). Stratification on endogenous variables and estimation: The Gary income maintenance experiment. In *Structural Analysis of Discrete Data with Econometric Applications* (C. Manski and D. McFadden, eds.) 365–391. MIT Press, Cambridge, MA.
- HUANG, J. and WELLNER, J. A. (1995a). Asymptotic normality of the NPMLE of linear functionals for interval censored data, case 1. *Stat. Neerl.* **49** 153–163. [MR1345376](#)
- HUANG, J. and WELLNER, J. A. (1995b). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Stat.* **22** 3–33. [MR1334065](#)
- IMBENS, G. W. and LANCASTER, T. (1996). Efficient estimation and stratified sampling. *J. Econometrics* **74** 289–318. [MR1411601](#)
- JANKOWSKI, H. (2014). Convergence of linear functionals of the Grenander estimator under misspecification. *Ann. Statist.* **42** 625–653. [MR3210981](#)
- KANG, Q., NELSON, P. I. and VAHL, C. I. (2010). Parameter estimation from an outcome-dependent enriched sample using weighted likelihood method. *Statist. Sinica* **20** 1529–1550. [MR2777335](#)
- PATIL, G. P. (1984). Studies in statistical ecology involving weighted distributions. In *Statistics: Applications and New Directions (Calcutta, 1981)* 478–503. Indian Statist. Inst., Calcutta. [MR0786158](#)
- PATIL, G. P. and RAO, C. R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. *Biometrics* **34** 179–189. [MR0507202](#)
- PERLIS, S. (1991). *Theory of Matrices*. Dover, New York. [MR1215486](#)
- PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhya, Ser. A* **31** 23–36. [MR0267677](#)
- SMITH, W. and PARNES, M. (1994). Mean streets: The median of a size-biased sample and the population mean. *Amer. Statist.* 106–10.
- VAN DE GEER, S. (2000). *Empirical Processes in M-Estimation*. Cambridge Univ. Press, Cambridge.
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. [MR1385671](#)
- VARDI, Y. (1985). Empirical distributions in selection bias models. *Ann. Statist.* **13** 178–205. [MR0773161](#)
- WANG, M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *J. Amer. Statist. Assoc.* **86** 130–143. [MR1137104](#)
- WANG, M.-C. (1992). The analysis of retrospectively ascertained data in the presence of reporting delays. *J. Amer. Statist. Assoc.* **87** 397–406.
- WANG, X. and ZHOU, H. (2006). A semiparametric empirical likelihood method for biased sampling schemes with auxiliary covariates. *Biometrics* **62** 1149–1160. [MR2307440](#)
- WOODROOFE, M. and SUN, J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is nonincreasing. *Statist. Sinica* **3** 501–515. [MR1243398](#)

K. C. G. CHAN
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195
USA
E-MAIL: kegchan@uw.edu

H. K. LING
DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
1255 AMSTERDAM AVENUE
NEW YORK, NEW YORK 10027
USA
E-MAIL: hokkan.ling@columbia.edu

T. SIT
S. C. P. YAM
DEPARTMENT OF STATISTICS
CHINESE UNIVERSITY OF HONG KONG
SHATIN, NT
HONG KONG SAR
E-MAIL: tonysit@sta.cuhk.edu.hk
scpyam@sta.cuhk.edu.hk