HHS Public Access
Author manuscript
Ann Stat. Author manuscript; available in PMC 2020 July 23.

Published in final edited form as:
Ann Stat. 2020 April ; 48(2): 1001–1024. doi:10.1214/19-aos1835.

A UNIFIED STUDY OF NONPARAMETRIC INFERENCE FOR MONOTONE FUNCTIONS

TED WESTLING[1], **MARCO CARONE**[2]

[1]Center for Causal Inference, University of Pennsylvania

[2]Department of Biostatistics, University of Washington

Abstract

The problem of nonparametric inference on a monotone function has been extensively studied in many particular cases. Estimators considered have often been of so-called Grenander type, being representable as the left derivative of the greatest convex minorant or least concave majorant of an estimator of a primitive function. In this paper, we provide general conditions for consistency and pointwise convergence in distribution of a class of generalized Grenander-type estimators of a monotone function. This broad class allows the minorization or majoration operation to be performed on a data-dependent transformation of the domain, possibly yielding benefits in practice. Additionally, we provide simpler conditions and more concrete distributional theory in the important case that the primitive estimator and data-dependent transformation function are asymptotically linear. We use our general results in the context of various well-studied problems, and show that we readily recover classical results established separately in each case. More importantly, we show that our results allow us to tackle more challenging problems involving parameters for which the use of flexible learning strategies appears necessary. In particular, we study inference on monotone density and hazard functions using informatively right-censored data, extending the classical work on independent censoring, and on a covariate-marginalized conditional mean function, extending the classical work on monotone regression functions.

Keywords

Cube-root asymptotics; isotonic regression; dependent censoring; g-formula

1. Introduction

1.1. Background

In many scientific settings, investigators are interested in learning about a function known to be monotone, either due to probabilistic constraints or in view of existing scientific knowledge. The statistical treatment of nonparametric monotone function estimation has a

---

tgwest@pennmedicine.upenn.edu.

SUPPLEMENTARY MATERIAL

Supplement: Proofs and simulations (DOI: 10.1214/19-AOS1835SUPP; .pdf). The supplement includes proofs of Theorems 1–7, a heuristic justification for the influence function forms in Theorem 5, additional technical details for the examples of Section 5, and a simulation study that illustrates the large-sample results of Sections 3–4 on Examples 1 and 2 from Section 5.

long and rich history. Early on, Grenander (1956) derived the nonparametric maximum likelihood estimator (NPMLE) of a monotone density function, now commonly referred to as the Grenander estimator. Since then, monotone estimators of many other parameters, including hazard and regression functions, have been proposed and studied.

In the literature, most monotone function estimators have been constructed via empirical risk minimization. Specifically, these are obtained by minimizing the empirical risk over the space of nondecreasing or nonincreasing candidate functions based on an appropriate loss function. The theoretical study of these estimators has often hinged strongly on their characterization as empirical risk minimizers. This is the case, for example, for the asymptotic theory developed by Prakasa Rao (1969) and Prakasa Rao (1970) for the NPMLE of monotone density and hazard functions, respectively, and by Brunk (1970) for the least-squares estimator of a monotone regression function. Kim and Pollard (1990) unified the study of these various estimators by studying the *argmin* process typically driving the pointwise distributional theory of monotone empirical risk minimizers.

Many of the parameters treated in the literature on monotone function estimation can be viewed as an index of the statistical model, in the sense that the model space is in bijection with the product space corresponding to the parameter of interest and an additional variation-independent parameter. In such cases, identifying an appropriate loss function is often easy, and a risk minimization representation is therefore usually available. However, when the parameter of interest is a complex functional of the data-generating mechanism, an appropriate loss function may not be readily available. This occurs often, for example, when identification of the parameter of interest based on the observed data distribution requires adjustment for sampling complications (e.g., informative treatment attribution, missing data or loss to follow-up). It is thus imperative to develop and study estimation methods that do not rely upon risk minimization.

It is a simple fact that the primitive of a nondecreasing function is convex. This observation serves as motivation to consider as an estimator of the function of interest the derivative of the greatest convex minorant (GCM) of an estimator of its primitive function. In the literature on monotone function estimation, many estimators obtained as empirical risk minimizers can alternatively be represented as the left derivative of the GCM of some primitive estimator. This is because the definition of the GCM is intimately tied to the necessary and sufficient conditions for optimization of certain risk functionals over the convex cone of monotone functions (see, e.g., Chapter 2 of Groeneboom and Jongbloed (2014)). In particular, Grenander's NPMLE of a monotone density equals the left derivative of the GCM of the empirical distribution function. In the recent literature, estimators obtained in this fashion have thus been referred to as being of *Grenander-type*. Leurgans (1982) is an early example of a general study of Grenander-type estimators for a class of regression problems.

In a seminal paper, Groeneboom (1985) introduced an approach to studying GCMs based on an inversion operation. This approach has facilitated the theoretical study of certain Grenander-type estimators without the need to utilize their representation as empirical risk minimizers. For example, under the assumption of independent right-censoring, Huang and

Wellner (1995) used this approach to derive large-sample properties of a monotone hazard function estimator obtained by differentiating the GCM of the Nelson–Aalen estimator of the cumulative hazard function. This general strategy was also used by van der Vaart and van der Laan (2006), who derived and studied an estimator of a covariate-marginalized survival curve based on current-status data, including possibly high-dimensional and time-varying covariates. More recently, there has been interest in deriving general results for Grenander-type estimators applicable to a variety of cases. For instance, Anevski and Hössjer (2006) derived pointwise distributional limit results for Grenander-type estimators in a very general setting including, in particular, dependent data. Durot (2007), Durot, Kulikov and Lopuhaä (2012) and Lopuhaä and Musta (2018a) derived limit results for the estimation error of Grenander-type estimators under L_p, supremum and Hellinger norms, respectively. Durot, Groeneboom and Lopuhaä (2013) studied the problem of testing the equality of generic monotone functions with K independent samples. Durot and Lopuhaä (2014), Beare and Fang (2017) and Lopuhaä and Musta (2018b) studied properties of the least concave majorant of an arbitrary estimator of the primitive function of a monotone parameter. The monograph of Groeneboom and Jongbloed (2014) also summarizes certain large-sample properties for these estimators.

1.2. Contribution and organization of the article

In this paper, we wish to address the following three key objectives:

1. to provide a unified framework for studying a large class of nonparametric monotone function estimators that implies classical results but also applies in more complicated, modern applications;
2. to derive tractable sufficient conditions under which estimators in this class are known to be consistent and have a nondegenerate limit distribution upon proper centering and scaling;
3. to illustrate the use of this general framework to construct targeted estimators of monotone parameters that are possibly complex summaries of the observed data distribution, and whose estimation may require the use of data-adaptive estimators of nuisance functions.

Our first goal is to introduce a class of monotone estimators that allow the greatest convex minorization process to be performed on a possibly data-dependent transformation of the domain. For many monotone estimators in the literature, the greatest convex minorization is performed on a transformation of the domain. A strategic domain transformation can lead to significant benefits in practice, including in some cases the elimination of the need to estimate challenging nuisance parameters. Unfortunately, to our knowledge, existing results for general Grenander-type estimators do not apply in a straightforward manner in cases in which a data-dependent transformation of the domain has been used. We will define a class that permits such transformations, and demonstrate both how this class encompasses many existing estimators in the literature and how a transformation can be strategically selected in novel problems.

Our second goal is to derive sufficient conditions on the estimator of the primitive function and domain transformation that imply consistency and pointwise convergence in distribution of the monotone function estimator. As noted above, general results on pointwise convergence in distribution for the class of Grenander-type estimators, applicable in a wide variety of settings, were provided in Anevski and Hössjer (2006). Our work differs from that of Anevski and Hössjer (2006) in a few important ways. First, the role and implications of domain transformations—which, as we show, are often important in practice—were not explicitly considered in Anevski and Hössjer (2006). To our knowledge, the class of generalized Grenander-type estimators we consider in this paper, which allow for domain transformations, has not previously been studied in a unified manner, and hence, general results for this class do not currently exist. Second, in addition to pointwise distributional results, we study weak consistency. Third, in Sections 4 and 5, we pay special attention to parameters for which asymptotically linear estimators of the primitive and transformation functions can be constructed —in such cases, relatively straightforward sufficient conditions can be developed, and the limit distribution has a simpler form. While these results are weaker than those in Section 3 and in Anevski and Hössjer (2006) because they apply only to a special case, they are useful in many settings. We demonstrate the utility of these results for three groups of examples—estimation of monotone density, hazard and regression functions—and show that our results coincide with established results in these settings.

Our third goal is to discuss and illustrate Grenander-type estimation in cases in which nonparametric estimation of the primitive function requires estimation of challenging nuisance parameters. In this sense, our work follows the lead of van der Vaart and van der Laan (2006), whose setting is of this type. More generally, such primitive functions arise frequently, for example, when the observed data unit represents a coarsened version of an ideal data structure, and the coarsening occurs randomly conditional on observed covariates (Heitjan and Rubin (1991)). In our general results, we provide sufficient conditions that can be readily applied to such primitive estimators. To demonstrate the application of our theory in coarsened data structures, we consider extensions of the three classical monotone problems above to more complex settings in which covariates must be accounted for, because either the censoring process or the treatment allocation mechanism are informative, as is typical in observational studies. Specifically, we derive novel estimators of monotone density and hazard functions for use when the survival data are subject to right-censoring that may depend on covariates, and a novel estimator of a monotone dose-response curve for use when the relationship between the exposure and outcome is confounded by recorded covariates. Unlike for their classical analogues, in these more difficult problems, nonparametric estimation of the primitive function involves nuisance functions for which flexible estimation strategies must be employed. As van der Vaart and van der Laan (2006) was able to achieve in a particular problem, our general framework explicitly allows the integration of such strategies while still yielding estimators with a tractable limit theory.

Our paper is organized as follows. In Section 2, we define the class of estimators we consider and briefly introduce our three working examples. In Section 3, we present our most general results for the consistency and convergence in distribution of our class of estimators. We provide refined results, including simpler sufficient conditions and distributional results, for the special case in which the primitive and transformation estimators are asymptotically linear in Section 4. In Section 5, we apply our general theory in three examples, both for classical parameters and for the novel extensions we consider. We provide concluding remarks in Section 6. The proofs of all theorems, additional technical details and results from simulation studies that evaluate the validity of the theory in two examples are provided in the Supplementary Material (Westling and Carone (2020)).

2. Generalized Grenander-type estimators

2.1. Statistical setup and definitions

Throughout, we make use of the following definitions. For intervals $I, J \subseteq R$, define $\ell^\infty(I)$ as the space of bounded, real-valued functions on I, $\mathscr{D}_I \subset \ell^\infty(I)$ as the subset of nondecreasing and càdlàg (right-continuous with left-hand limits) functions on I, and $\mathscr{D}_{I,J} \subset \mathscr{D}_I$ as the further subset of functions whose range is contained in J. The GCM operator $\text{GCM}_I : \ell^\infty(I) \to \ell^\infty(I)$ is defined for any $G \in \ell^\infty(I)$ as the pointwise supremum over all convex functions $H \leq G$ on I. We note that $\text{GCM}_I(G)$ is necessarily convex. For $G \in \mathscr{D}_I$, we denote by G^- the generalized inverse mapping $x \mapsto \inf\{u \in I : G(u) \geq x\}$, and for a left-differentiable G, we denote by $\partial_- G$ the left derivative of G.

We are interested in making inference about an unknown function $\theta_0 \in \mathscr{D}_I$ determined by the true data-generating mechanism P_0 for an interval $I \subseteq \mathbb{R}$. We denote the endpoints of I by $a_I := \inf I$ and $b_I := \sup I$. We define the primitive function Θ_0 of θ_0 pointwise for each $x \in I$ as $\Theta_0(x) := \int_{a_I}^x \theta_0(u)du$, where if $a_I = -\infty$ we assume the integral exists. The results we present in Section 3 apply in contexts with either independent or dependent data. Starting in Section 4, we focus on contexts in which the data consist of independent observations $O_1, \ldots, O_n$ from an unknown distribution P_0 in a nonparametric model $\mathscr{M}$. In such cases, we denote by O a prototypical data unit, $\mathscr{O}(P)$ the support of O under $P \in \mathscr{M}$, and $\mathscr{O} := \cup_{P \in \mathscr{M}} \mathscr{O}(P)$.

In its simplest formulation, a Grenander-type estimator of θ_0 is given by $\partial_-\text{GCM}_I(\Theta_n)$ for some estimator Θ_n of Θ_0. However, as a critical step in unifying classical estimators and constructing procedures with possibly improved properties, we wish to allow the GCM procedure to be performed on a possibly data-dependent transformation of the domain I. To do so, we first define for any interval $J \subseteq \mathbb{R}$ the operator $\text{Iso}_J : \ell^\infty(J) \times \mathscr{D}_{I,J} \to \ell^\infty(I)$ as $\text{Iso}_J(\Psi, \Phi) := (\partial_-\text{GCM}_J(\Psi)) \circ \Phi$ for each $\Psi \in \ell^\infty(I)$ and $\Phi \in \mathscr{D}_{I,J}$. We set $J_0 := [0, u_0]$, with $u_0 \in (0, \infty)$ possibly depending on P_0, and suppose that a domain transform $\Phi_0 \in \mathscr{D}_{I,J_0}$ is chosen. We may then consider the domain-transformed parameter $\psi_0 := \theta_0 \circ \Phi_0^-$, which has primitive Ψ_0 defined pointwise as $\Psi_0(t) := \int_0^t \psi_0(u)du$ for $t \in (0, u_0]$. As with θ_0 and Θ_0, ψ_0 is nondecreasing and Ψ_0 is convex. Thus, $\text{Iso}_{J_0}(\Psi_0, \Phi_0)(x) = \theta_0(x)$ for each $x \in I$ at which θ_0 is left continuous and such that $\Phi_0(u) < \Phi_0(x)$ for all $u < x$. This observation motivates us to consider estimators of θ_0 of the form $\text{Iso}_{J_n}(\Psi_n, \Phi_n)$, where Ψ_n, Φ_n and u_n are estimators of Ψ_0, Φ_0 and u_0, respectively, and we define $J_n := [0, u_n]$. We refer to any such estimator as being of the *generalized Grenander-type*. This class, of course, contains the standard Grenander-type estimators: setting $\Psi_n = \Theta_n$ and $\Phi_n =$ Id for Id the identity mapping yields $\theta_n = \partial_{-}\text{GCM}_I(\Theta_n)$. We note that, in this formulation, we require the domain J_0 over which the GCM is performed to be bounded, but not so for the domain I of θ_0. Additionally, we assume that the left endpoint of J_0 is fixed at 0, while the upper endpoint u_0 may depend on P_0. However, this entails no loss in generality, since if the desired domain is instead $[\ell_0, u_0]$, where now ℓ_0 also depends on P_0, we can define $\bar{u}_0 := u_0 - \ell_0$ and similarly shift Φ_0 by ℓ_0 to obtain the new domain $[0, \bar{u}_0]$.

Defining $\Gamma_0 := \Psi_0 \circ \Phi_0$, we suppose that we have at our disposal estimators Φ_n and Γ_n of Φ_0 and Γ_0, respectively, as well as a weakly consistent estimator u_n of u_0. In this work, we study the properties of a generic generalized Grenander-type estimator θ_n of θ_0 of the form

$$\theta_n := \text{Iso}_{J_n}(\Gamma_n \circ \Phi_n^-, \Phi_n)\,. \tag{1}$$

Our goal is to provide sufficient conditions on the triple (Γ_n, Φ_n, u_n) under which θ_n is consistent, and under which a suitable standardization of θ_n converges in distribution to a nondegenerate limit. As stated above, our only requirement for u_n is that it tends in probability to u_0. Therefore, our focus will be on the pair (Γ_n, Φ_n).

We note that estimators taking form (1) constitute a more restrictive class than the set of all estimators of the form $\text{Iso}_{J_n}(\Psi_n, \Phi_n)$ for arbitrary Ψ_n. Our focus on this slightly less general form is motivated by two reasons. First, as we will see in various examples, Γ_0 often has a simpler form than Ψ_0, and in such cases, it may be significantly easier to verify required regularity conditions for Γ_n and to derive limit distribution properties based on Γ_n rather than Ψ_n. Second, many celebrated monotone estimators in the literature follow this particular form. This can be seen by noting that, if Φ_n is a right-continuous step function with jumps at points $x_1, x_2, \ldots, x_m$, then for each $x \in I$ the estimator $\theta_n(x)$ given in (1) equals the slope at $\Phi_n(x)$ of the greatest convex minorant of the diagram of points $\{(\Phi_n(x_j), \Gamma_n(x_j)) : j = 0, 1, \ldots, m\}$, where $x_0 = a_I$. We highlight well-known examples of estimators of this type below. In brief, we sacrifice a little generality for a substantial gain in the ease of application of our results, both for well known and novel monotone estimators. Nevertheless, conditions on the pair (Ψ_n, Φ_n) under which consistency and distributional results hold for θ_n can be derived similarly.

2.2. Examples

Before proceeding to our main results, we briefly discuss the examples we will use to illustrate how our framework allows us to not only obtain results on classical estimators in the monotone estimation literature directly, but also tackle more complex problems for which no estimators are currently available. These examples will be studied further in Section 5.

EXAMPLE 1 (Monotone density function)—Suppose that T is a univariate positive random variable with nondecreasing density function f_0, and that T is right-censored by an independent random censoring time C. The observed data unit is $O := (Y, \Delta)$, where $Y := \min(T, C)$ and $\Delta := I(T \leq C)$, with distribution P_0 implied by the marginal distributions of T and C. The parameter of interest is $\theta_0 := f_0$, the density function of T with support I. Taking Φ_0 to be the identity function, we get that $\psi_0 = \theta_0$. Here, both Ψ_0 and Θ_0 represent the distribution function F_0 of T, and Φ_0 plays no role. A natural estimator θ_n of θ_0 can be obtained by taking Ψ_n to be the Kaplan–Meier estimator of the distribution function Ψ_0. With $\Phi_n := \text{Id}$, $\Gamma_n := \Psi_n$ and $u_n := \max_i Y_i$ the estimator $\theta_n := \text{Iso}_I(\Gamma_n, \Phi_n)$ is precisely the estimator studied by Huang and Wellner (1995). When $C = +\infty$ with probability one, Ψ_n is the empirical distribution function based on $Y_1, Y_2, \ldots, Y_n$, and θ_n is precisely the Grenander estimator.

In Section 5, we extend estimation of a monotone density function to the setting in which the data are subject to possibly informative right-censoring. Specifically, we only require T and C to be independent conditionally upon a vector W of baseline covariates. We will study the estimator defined by differentiating the GCM of a one-step estimator of Ψ_0. In this context, estimation of Ψ_0 requires estimation of nuisance functions. We will use our general results to provide conditions on the nuisance estimators that imply consistency and distributional results for θ_n.

EXAMPLE 2 (Monotone hazard function)—Suppose now that T is a univariate positive random variable with nondecreasing hazard function λ_0. In this example, we are interested in $\theta_0 := \lambda_0$. Setting $S_0 := 1 - F_0$ to be the survival function of T, we note that $\Gamma_0(u) = \int_0^u f_0(v)/S_0(v)\Phi_0(dv)$, and so, taking Φ_0 to satisfy $\Phi_0(dv) = S_0(v)dv$ makes $\Gamma_0 = F_0$. The restricted mean lifetime function $\Phi_0(u) := \int_0^u S_0(v)dv$ satisfies this condition. Using this transformation, the estimator of the monotone hazard function θ_0 only requires estimation of F_0.

In Section 5, we again extend estimation of a monotone hazard function to allow the data to be subject to possibly informative right-censoring using the same one-step estimator Γ_n of $\Gamma_0 = F_0$ that will be introduced in Example 1 and the data-dependent transformation $\Phi_n(u) := \int_0^u [1 - \Gamma_n(v)]dv$. We will show that, once the simpler details regarding the estimation of a monotone density are established, the asymptotic properties of this estimator of a monotone hazard are obtained essentially for free.

EXAMPLE 3 (Monotone regression function)—As our last example, we study estimation of a nondecreasing regression function. In the simplest setup, the data unit is $O := (Y, A)$ and we are interested in $\theta_0 : x \mapsto E_0(Y \mid A = x)$. Assume without loss of generality that the data are sorted according to the observed values of A. Taking I to be the support of A and Φ_0 to be the marginal distribution function of A, we have that $\psi_0(u) = E_0[Y \mid \Phi_0(A) = u]$ for each $u \in [0, 1]$, and $\Gamma_0(x) = E_0\left[Y I_{(-\infty, x]}(A)\right]$ for each $x \in I$. Thus $\Gamma_n(x) := \frac{1}{n}\sum_{i=1}^n Y_i I_{(-\infty, x]}(A_i)$ and $\Phi_n(x) := \frac{1}{n}\sum_{i=1}^n I_{(-\infty, x]}(A_i)$ are natural nonparametric estimators of $\Gamma_0(x)$ and $\Phi_0(x)$, respectively. Then $\theta_n := \text{Iso}_{[0,1]}(\Gamma_n, \Phi_n)$ is the classical monotone least-squares estimator of θ_0.

In Section 5, we consider an extension of this example to estimation of a covariate-marginalized regression function, for use when the relationship between exposure and outcome of interest is confounded. Specifically, we will consider the data unit $O :=(Y, A, W)$, with W representing a vector of potential confounders, and focus on $\theta_0 : x \mapsto E_0[E_0(Y \mid A = x, W)]$. Under untestable causal identifiability conditions, $\theta_0(x)$ is the mean of the counterfactual outcome $Y(x)$ obtained by setting exposure at level $A = x$. This parameter plays a critical role in causal inference, particularly when the available data are obtained from an observational study and the exposure assignment process may be informative. As before, tackling this more complex parameter will require estimation of certain nuisance functions.

3. General results

We begin with our first set of results on the large-sample properties of θ_n. Our goal is to establish conditions under which consistency and pointwise convergence in distribution hold. First, we provide general results on the consistency of θ_n, both pointwise and uniformly. We note that the results of Anevski and Hössjer (2006), Durot (2007), Durot, Kulikov and Lopuhaä (2012) and Lopuhaä and Musta (2018a) imply conditions for consistency of Grenander-type estimators. However, because the objective of their work is to establish distributional theory for a global discrepancy between the estimated and true function, the conditions they require are stronger than needed for consistency alone. Also, their work is restricted to Grenander-type estimators, without data-dependent transformations of the domain.

Below, we refer to the sets $I_n := \{z \in I : z = \Phi_n^{-}(u), u \in J_n\}$ and $I_{n,\beta} := \left\{x \in I : 0 \le \Phi_0(x - \beta) \le \Phi_0(x + \beta) \le u_n\right\}$ for $\beta \ge 0$.

THEOREM 1 (Weak consistency)

1. *Suppose θ_0 is continuous at $x \in I$ and, for some $\delta > 0$ such that $[x - \delta, x + \delta] \subset \Phi_0^{-1}(J_0)$, Φ_0 is strictly increasing and continuous on $[x - \delta, x + \delta]$. If $\|\Gamma_n - \Gamma_0\|_{\infty, I_n}$, $\|\Phi_n - \Phi_0\|_{\infty, I_n}$ and $\|\Phi_n - \Phi_0\|_{\infty, [x - \delta, x + \delta]}$ tend to zero in probability, then $\theta_n(x) = \theta_0(x) + o_P(1)$.*

2. *Suppose θ_0 and Φ_0 are uniformly continuous on I, and Φ_0 is strictly increasing on I. If $\|\Gamma_n - \Gamma_0\|_{\infty, I_n}$ and $\|\Phi_n - \Phi_0\|_{\infty, I}$ tend to zero in probability, then $\|\theta_n - \theta_0\|_{\infty, I_{n,\beta}} = o_P(1)$ for each fixed $\beta > 0$.*

We note that in part 1 of Theorem 1, we require uniform convergence of Γ_n and Φ_n to obtain a pointwise result for θ_n—this will also be the case for Theorem 2 below. This is because the GCM is a global procedure, and so, the value of $\theta_n(x_1)$ depends on $\Gamma_n(x_2)$ even for x_2 not near x_1. Without uniform consistency of Γ_n, θ_n may indeed fail to be pointwise consistent. Also, we note that in part 1 of Theorem 1, we require that $\Gamma_n - \Gamma_0$ and $\Phi_n - \Phi_0$ tend to zero uniformly over the set I_n. This requirement stems from the fact that θ_n only depends on Γ_n through the composition $\Gamma_n \circ \Phi_n^{-}$, and so, values of Γ_n only matter at points in the range of

Φ_n^-. In part 1, we also require that $\Phi_n - \Phi_0$ tend to zero uniformly in a neighborhood of x, while in part 2, we require that $\Phi_n - \Phi_0$ tend to zero uniformly over I. These requirements allow us to obtain results for x values that are possibly outside I_n for all n. In many applications, it may be the case that $\Gamma_n - \Gamma_0$ and $\Phi_n - \Phi_0$ both tend to zero in probability uniformly over I, which implies convergence over I_n.

The weak conditions required for Theorem 1 are especially important for the extensions of the classical parameters that we consider in Section 5. The estimators we propose require estimating difficult nuisance parameters, such as conditional hazard, density and mean functions. While under mild conditions it is typically possible to construct uniformly consistent estimators of these nuisance parameters, ensuring a given local or uniform rate of convergence often requires additional knowledge about the true function. Thus, Theorem 1 is useful for guaranteeing consistency under weak conditions.

We now provide lower bounds on the convergence rate of θ_n, both pointwise and uniformly, depending on (a) the uniform rates of convergence of Γ_n and Φ_n, and (b) the moduli of continuity of θ_n and Φ_0^-

THEOREM 2 (Rates of convergence)

Let $x \in I$ *be given. Suppose that, for some* $\delta > 0$, $[x - \delta, x + \delta] \subset \Phi_0^{-1}(J_0)$ *and* Φ_n *is strictly increasing and continuous on* $[x - \delta, x + \delta]$. *Let* r_n *be a fixed sequence such that* $r_n\|\Gamma_n - \Gamma_0\|_{\infty, I_n}$, $r_n\|\Phi_n - \Phi_0\|_{\infty, I_n}$ *and* $r_n\|\Phi_n - \Phi_0\|_{\infty, [x - \delta, x + \delta]}$ *are bounded in probability.*

1. *If there exist* $K_1(x), K_2(x) \in [0, \infty)$ *and* $\alpha_1, \alpha_2 \in (0,1]$ *such that* $|\theta_0(u) - \theta_0(x)| \le K_1(x)|u - x|^{\alpha_1}$ *for all* $u \in I$ *and* $|\Phi_0^-(u) - \Phi^-(x)| \le K_2(x)|u - x|^{\alpha_2}$ *for all* $u \in J_0$, *then*

$$r_n^{\frac{\alpha_1\alpha_2}{1 + \alpha_1\alpha_2}}[\theta_n(x) - \theta_0(x)] = O_P(1).$$

2. *If* θ_0 *is constant on* $[x - \delta, x + \delta]$, *then* $r_n[\theta_n(x) - \theta_0(x)] = O_P(1)$.

Let r_n *be a fixed sequence such that* $r_n\|\Gamma_n - \Gamma_0\|_{\infty, I_n}$ *and* $r_n\|\Phi_n - \Phi_0\|_{\infty, I}$ *are bounded in probability, and suppose that* Φ_0 *is strictly increasing on* I.

3. *If there exist* $K_1, K_2 \in [0, \infty)$ *and* $\alpha_1, \alpha_2 \in (0,1]$ *such that* $|\theta_0(u) - \theta_0(v)| \le K_1|u - v|^{\alpha_1}$ *for all* $u, v \in I$ *and* $|\Phi_0^-(u) - \Phi_0^-(v)| \le K_2|u - v|^{\alpha_2}$ *for all* $u, v \in J_0$, *then*

$$r_n^{\frac{\alpha_1\alpha_2}{1 + \alpha_1\alpha_2}}\|\theta_n - \theta_0\|_{\infty, I_{n,\beta_n}} = O_P(1)$$

for any random positive real sequence β_n *such that* $\beta_n r_n^{1/(1 + \alpha_1\alpha_2)} \xrightarrow{P} \infty$.

We note here that the uniform results only cover subintervals of the interval over which the GCM procedure is performed. This should not be surprising given the poor behavior of Grenander-type estimators at the boundary of the GCM interval, as discussed, for example, in Woodroofe and Sun (1993), Kulikov and Lopuhaä (2006) and Balabdaoui et al. (2011). Various boundary corrections have been proposed—applying these in our general framework is an interesting avenue for future work.

We also note that, in Theorem 2, when θ_0 and Φ_0 are locally or globally Lipschitz, then $\alpha_1 = \alpha_2 = 1$ and the resulting rate is $O_P\left(r_n^{-1/2}\right)$, which yields $O_P\left(n^{-1/4}\right)$ when $r_n = n^{1/2}$. This rate is slower than the rate $n^{-1/3}$ that is often achievable for pointwise convergence when θ_0 and Φ_0 are differentiable at x and the primitive estimator converges at rate $n^{-1/2}$, as we discuss below. However, the assumptions in Theorems 2 are significantly weaker than typically required for the $n^{-1/3}$ rate of convergence: they constrain the supremum norm of the estimation error rather than its modulus of continuity, and hold when the true function is Lipschitz but not differentiable. Our results also cover situations in which θ_0 and Φ_0 are in Hölder classes. The rates provided by Theorem 2 should thus be seen as lower bounds on the true rate, for use when less is known about the properties of the estimation error or of the true functions. The distributional results we provide below recover the usual rates under stronger conditions.

For a fixed sequence r_n of positive real numbers, we now study the pointwise convergence in distribution of $r_n\left[\theta_n(x) - \theta_0(x)\right]$ at an interior point $x \in I$ at which Φ_0 has a strictly positive derivative. The rate r_n depends on two interdependent factors. First, we suppose that there exists some $\alpha > 0$ such that $|\theta_0(x+u) - \theta_0(x)| = \pi_0(x)|u|^{\alpha} + o(1)$ as $u \to 0$ for some constant $\pi_0(x) > 0$. Second, writing $\Gamma_{n,0} := \Gamma_n - \Gamma_0$ and $\Phi_{n,0} := \Phi_n - \Phi_0$, we suppose that there exists a sequence of positive real numbers $c_n \to \infty$ such that the appropriately localized process

$$
\begin{aligned}
W_{n,x}: u \mapsto c_n^{\alpha+1}\Big\{&\Gamma_{n,0}\left(x + uc_n^{-1}\right) - \Gamma_{n,0}(x) \\
&- \theta_0(x)\left[\Phi_{n,0}\left(x + uc_n^{-1}\right) - \Phi_{n,0}(x)\right]\Big\}
\end{aligned}
$$

converges weakly. We note that $W_{n,x}$ depends on α. As we formalize below, if $r_n = c_n^{\alpha}$, then $r_n\left[\theta_n(x) - \theta_0(x)\right]$ has a nondegenerate limit distribution under some conditions. We now introduce some of the conditions that we build upon:

(A1) for each $M > 0$, $\{W_{n,x}(u) : |u| \leq M\}$ converges weakly in $\ell^{\infty}[-M, M]$ to a tight limit process $\{W_x(u) : |u| \leq M\}$ with almost surely lower semi-continuous sample paths;

(A2) $\sup \operatorname{argmax}_{u \in \mathbb{R}}\left\{W_x(u) + \pi_0(x)\Phi_0'(x)(\alpha+1)^{-1}|u|^{\alpha+1} + c\Phi_0'(x)u\right\}$ is bounded in probability for every $c \in \mathbb{R}$;

(A3) there exist $\beta \in (1, 1+\alpha)$, $\delta^* > 0$ and a sequence $f_n : \mathbb{R}^+ \to \mathbb{R}^+$ such that $u \mapsto u^{-\beta} f_n(c_n u)$ is decreasing, $f_n(1) = O(1)$, and for all large n and $\delta \leq \delta^*$, $E_0\left[\sup_{|u| \leq c_n\delta}\left|W_{n,x}(u)\right|\right] \leq f_n(c_n\delta)$.

In addition, we introduce conditions on the uniform convergence of estimators Φ_n and Γ_n:

(A4) $c_n E_0[\sup_{|v| < \delta}|\Phi_n(x+v) - \Phi_0(x+v)|] \to 0$ for some δ > 0;

(A5) $\|\Gamma_{n,0} - \theta_0(x) \cdot \Phi_{n,0}\|_{\infty, I_n} \xrightarrow{P} 0.$

THEOREM 3 (Convergence in distribution)

If x is an interior point of I at which Φ_0 *is continuously differentiable with positive derivative and* $\lim_{u \to 0}|\theta_0(x+u) - \theta_0(x)|/|u|^{\alpha} = \pi_0(x)$, *conditions* (A1)–(A5) *imply that*

$$r_n[\theta_n(x) - \theta_0(x)] \xrightarrow{d} \Phi_0'(x)^{-1}\partial_{-}\mathrm{GCM}_{\mathbb{R}}\{M_{x,\alpha}\}(0)$$

with $r_n := c_n^{\alpha}$ *and*

$$M_{x,\alpha}: v \mapsto W_x(v) + \left[\frac{\pi_0(x)\Phi_0'(x)}{\alpha+1}\right]|v|^{\alpha+1}.$$

If also $\alpha = 1, \pi_0(x) = \theta_0'(x)$ *and* W_x *possesses stationary increments, then*

$$r_n[\theta_n(x) - \theta_0(x)] \xrightarrow{d} -\theta_0'(x)\operatorname*{argmin}_{u \in \mathbb{R}}\left\{W_x(u) + \frac{1}{2}\theta_0'(x)\Phi_0'(x)u^2\right\}.$$

Furthermore, if $W_x = [\kappa_0(x)]^{1/2}W_0$ *with* W_0 *a standard two-sided Brownian motion process satisfying* $W_0(0) = 0$, *then*

$$r_n[\theta_n(x) - \theta_0(x)] \xrightarrow{d} \tau_0(x)Z$$

With $\tau_0(x) := \left[4\theta_0'(x)\kappa_0(x)/\Phi_0'(x)^2\right]^{1/3}$ *and* $Z := \mathrm{argmin}_{u \in \mathbb{R}}\left\{W_0(u) + u^2\right\}$.

The latter limit distribution is referred to as a scaled Chernoff distribution, since Z is said to follow the standard Chernoff distribution. This distribution appears prominently in classical results in nonparametric monotone function estimation and has been extensively studied (e.g., Groeneboom and Wellner (2001)). It can also be defined as the distribution of the slope at zero of $\mathrm{GCM}_{\mathbb{R}}\left\{u \mapsto W_0(u) + u^2\right\}$.

Theorem 3 applies in the common setting in which θ_0 is differentiable at x with positive derivative, that is, when $\alpha = 1$. However, as in Wright (1981) and Anevski and Hössjer (2006), Theorem 3 also applies in additional situations, including when θ_0 has $\alpha \in \{2,3,\dots\}$ derivatives at x, with null derivatives of order $j < \alpha$ and positive derivative of order α. Nevertheless, Theorem 3 does not cover situations in which θ_0 is flat in a neighborhood of x. The limit distribution of the Grenander estimator at flat points was studied in Carolan and Dykstra (1999), but it appears that similar results have not been derived for Grenander-type estimators.

We note the similarity of our Theorem 3 to Theorem 2 of Anevski and Hössjer (2006). For the special case in which Φ_0 is the identity transform, the consequents of the two results coincide. Our result explicitly permits alternative transforms. Both results require weak convergence of a stochastic part of the primitive process, and also require the same local rate of growth of θ_0. Additionally, condition (A2) is implied if for every ϵ and δ positive, there exists a finite $m \in (0, +\infty)$ such that $P_0\left(\sup_{|v| \geq m}|W_x(v)||v|^{-\alpha-1} > \epsilon\right) < \delta$, as in Assumption A5 of Anevski and Hössjer (2006). However, the remaining conditions and methods of proof differ. To prove our result, we first generalize the switch relation of Groeneboom (1985) and use it to convert $P_0(r_n[\theta_n(x) - \theta_0(x)] > \eta)$ into the probability that the minimizer of a process involving $W_{n,x}$ falls below some value. After establishing weak convergence of this process, we then use conditions (A2) through (A5) to justify application of the argmin continuous mapping theorem. In contrast, Anevski and Hössjer (2006) establish their result using a direct appeal to convergence in distribution of $\partial_-\text{GCM}_C(Y_n)(0)$ to $\partial_-\text{GCM}_C(Y_0)(0)$, where Y_n is a local limit process and Y_0 its weak limit. They also provide lower-level sufficient conditions for this convergence. It may be possible to establish the consequent of Theorem 3, permitting in particular the use of a nontrivial transformation Φ_0, using Theorem 2 of Anevski and Hössjer (2006) or a suitable generalization thereof. We have specified our sufficient conditions with applications to the setting $\alpha = 1$ and $c_n = n^{1/3}$ in mind, as we discuss at length in the next section.

Suppose that W_x^0 is the limit process that arises when no domain transformation is used in the construction of a generalized Grenander-type estimator, that is, when both Φ_0 and Φ_n are taken to be the identity map. In this case, under (A1)–(A5), Theorem 3 indicates that

$$r_n[\theta_n(x) - \theta_0(x)] \xrightarrow{d} \partial_-\text{GCM}_{\mathbb{R}}\left\{v \mapsto W_x^0(v) + \left[\frac{\pi_0(x)}{\alpha+1}\right]|v|^{\alpha+1}\right\}(0).$$

It is natural to ask how this limit distribution compares to the one obtained using a nontrivial transformation Φ_0. In particular, does using Φ_0 change the pointwise distributional results for θ_n? The answer is of course negative whenever W_x and $\Phi_0'(x)W_x^0$ are equal in distribution, since $\text{GCM}_{\mathbb{R}}$ is a homogeneous operator. A more detailed discussion of this question and lower-level conditions are provided in the next section.

4. Refined results for asymptotically linear primitive and transformation estimators

4.1. Distributional results

In applications of their main result, Anevski and Hössjer (2006) focus primarily on providing lower-level conditions to characterize the relationship between various dependence structures and asymptotic results for monotone regression and density function estimation. Anevski and Soulier (2011), Dedecker, Merlevède and Peligrad (2011) and Bagchi, Banerjee and Stoev (2016) provide additional applications of Anevski and Hössjer (2006) to monotone function estimation with dependent data. Our Theorem 3 could be used, for instance, to relax the common assumption of a uniform design in the analysis of monotone regression estimators. Here, we pursue an alternative direction, focusing instead on providing lower-level conditions for consistency of θ_n and convergence in distribution of $r_n[\theta_n(x) - \theta_0(x)]$ for use in the important setting in which $\alpha = 1$, $r_n = c_n = n^{1/3}$, the data are independent and identically distributed, and Γ_n and Φ_n are asymptotically linear estimators. Such settings arise frequently, for instance, when the primitive and transformation parameters are smooth mappings of the data-generating mechanism.

Below, we write Pf to denote $\int f(o)dP(o)$ for any probability measure P and P-integrable function $f: \mathcal{O} \to \mathbb{R}$. We also use $\mathbb{P}_n$ to denote the empirical distribution of independent observations $O_1, O_2, \ldots, O_n$ from P_0 so that $\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^{n} f(O_i)$ for any $f: \mathcal{O} \to \mathbb{R}$.

Suppose that there exist functions $D_{x,0}^*: \mathcal{O} \to \mathbb{R}$ and $L_{x,0}^*: \mathcal{O} \to \mathbb{R}$ depending on P_0 such that, for each $x \in I$, $P_0 D_{x,0}^* = P_0 L_{x,0}^* = 0$ and both $P_0 D_{x,0}^{*2}$ and $P_0 L_{x,0}^{*2}$ are finite, and

$$\Gamma_n(x) - \Gamma_0(x) = \mathbb{P}_n D_{x,0}^* + H_{x,n}, \qquad \Phi_n(x) - \Phi_0(x) = \mathbb{P}_n L_{x,0}^* + R_{x,n}, \tag{2}$$

where $H_{x,n}$ and $R_{x,n}$ are stochastic remainder terms. If $n^{1/2}\sup_{x \in I}|H_{x,n}|$ and $n^{1/2}\sup_{x \in I}|R_{x,n}|$ tend to zero in probability, we say that Γ_n and Φ_n as *uniformly asymptotically linear* over I as estimators of Γ_0 and Φ_0, respectively. The objects $D_{x,0}^*$ and $L_{x,0}^*$ are referred to as the influence functions of $\Gamma_n(x)$ and $\Phi_n(x)$, respectively, under sampling from P_0.

Assessing consistency and uniform consistency of θ_n is straightforward when display (2) holds. For example, if the classes $\{D_{x,0}^*: x \in I\}$ and $\{L_{x,0}^*: x \in I\}$ are P_0-Donsker, and $n^{1/2}\sup_{x \in I}|H_{x,n}|$ and $n^{1/2}\sup_{x \in I}|R_{x,n}|$ are bounded in probability, then $n^{1/2}\|\Gamma_n - \Gamma_0\|_{\infty, I}$ and $n^{1/2}\|\Phi_n - \Phi_0\|_{\infty, I}$ are both bounded in probability. Thus, Theorems 1 and 2 can be directly applied with $r_n = n^{1/2}$ provided the required conditions on θ_0 and Φ_0 hold. As such, we focus here on deriving a refined version of Theorem 3 for use whenever display (2) holds.

It is reasonable to expect the linear terms $\mathbb{P}_n D_{x,0}^*$ and $\mathbb{P}_n L_{x,0}^*$ to drive the behavior of the standardized difference $r_n[\theta_n(x) - \theta_0(x)]$ in Theorem 3. The natural rate here is $c_n = r_n = n^{1/3}$, for which Kim and Pollard (1990) provide intuition. Our first goal in this section is to provide sufficient conditions for weak convergence of the process $\left\{n^{1/6}\mathbb{G}_n g_{x,n^{-1/3}u} \mid |u| \le M\right\}$, where $\mathbb{G}_n$ is the empirical process $n^{1/2}(\mathbb{P}_n - P_0)$ and we define the localized difference function $g_{x,v} := D_{x+v,0}^* - D_{x,0}^* - \theta_0(x)\left(L_{x+v,0}^* - L_{x,0}^*\right)$. Kim and Pollard (1990) also provide detailed conditions for weak convergence of processes of this type. Building upon their results, we are able to provide simplified sufficient conditions for convergence in distribution of $n^{1/3}[\theta_n(x) - \theta_0(x)]$ when Γ_n and Φ_n are uniformly asymptotically linear estimators.

We begin by introducing some conditions. First, we define $\mathscr{G}_{x,R} := \{g_{x,u} : |u| \le R\}$ and suppose that $\mathscr{G}_R$ has envelope function $G_{x,R}$. The first two conditions concern the size of $\mathscr{G}_{x,R}$ for small R in terms of bracketing or uniform entropy numbers, which for completeness we define here; see van der Vaart and Wellner (1996) for a comprehensive treatment. Denote by $\|G\|_{P,2} = [P(G^2)]^{1/2}$ the $L_2(P)$ norm of a given P-square-integrable function $G: \mathscr{O}(P) \to \mathbb{R}$. The bracketing number $N_{[]}(\varepsilon, \mathscr{G}, L_2(P))$ of a class $\mathscr{G}$ with respect to the $L_2(P)$ norm is the smallest number of ε-brackets needed to cover $\mathscr{G}$, where an ε-bracket is any set of functions $\{f : \ell \le f \le u\}$ with ℓ and u such that $\|\ell - u\|_{P,2} < \varepsilon$. The covering number $N(\varepsilon, \mathscr{G}, L_2(Q))$ of $\mathscr{G}$ with respect to the $L_2(Q)$ norm is the smallest number of ε-balls in $L_2(Q)$ required to cover $\mathscr{G}$. The uniform covering number is the supremum of $N(\varepsilon\|G\|_{2,Q}, \mathscr{G}, L_2(Q))$ over all discrete probability measures Q such that $\|G\|_{2,Q} > 0$, where G is an envelope function for $\mathscr{G}$. We consider conditions on the size of $\mathscr{G}_{x,R}$:

(B1) For some constants $C > 0$ and $V \in [0,2)$, for all $\varepsilon \in (0,1]$ and R small enough, either:

(B1a) $\log N_{[]}(\varepsilon\|G_{x,R}\|_{P_0,2}, \mathscr{G}_{x,R}, L_2(P_0)) \le C\varepsilon^{-V}$ or

(B1b) $\log \sup_Q N(\varepsilon\|G_{x,R}\|_{Q,2}, \mathscr{G}_{x,R}, L_2(Q)) \le C\varepsilon^{-V}$.

(B2) $P_0 G_{x,R}^2 = O(R)$, and $P_0 G_{x,R}^2\{RG_{x,R} > \eta\} = o(R)$ as $R \to 0$ for all $\eta > 0$.

Condition (B1) replaces the notion of *uniform manageability* of the class $\mathscr{G}_{x,R}$ for small R as defined in Kim and Pollard (1990), and condition (B2) corresponds to their condition (vi). Since bounds on the bracketing and uniform entropy numbers have been derived for many common classes of functions, condition (B1) can be readily checked in practice. Together, conditions (B1) and (B2) ensure that $\mathscr{G}_{x,R}$ is a relatively small class, and this helps to establish the weak convergence of the localized process $\{W_{n,x}(u) : |u| \le M\}$.

As in Kim and Pollard (1990), to guarantee that the covariance function of this localized process stabilizes, it suffices that $\delta^{-1}\sup_{|u-v|<\delta} P_0(g_{x,u} - g_{x,v})^2$ be bounded for small enough $\delta > 0$ and that, up to a scaling factor possibly depending on x, $\sigma_{x,\alpha}(u,v) := \alpha^{-1}P_0[(g_{x,\alpha u} - P_0 g_{x,\alpha u})(g_{x,\alpha v} - P_0 g_{x,\alpha v})]$ tend to the covariance function $\sigma^2(u, v)$ of a two-sided Brownian motion as $\alpha \to 0$. Below, we provide simple conditions that imply these two statements for a broad class of settings that includes our examples.

The covariance function of the Gaussian process to which $\{\mathbb{G}_n[D_{t,0}^* - \theta_0(x)L_{t,0}^*] : t\}$ converges weakly is defined pointwise as $\Sigma_0(s,t) := P_0[D_{s,0}^* - \theta_0(x)L_{s,0}^*][D_{t,0}^* - \theta_0(x)L_{t,0}^*]$. The behavior of Σ_0 near (x, x) dictates the covariance of the local limit process W_x, and hence the scale parameter $\kappa_0(x)$. If Σ_0 is differentiable in (s, t) at (x, x), then $\kappa_0(x) = 0$ and θ_n converges at a faster rate, though possibly with an asymptotic bias. When instead Chernoff asymptotics apply, the covariance function can typically be written as

$$\Sigma_0(s,t) = \Sigma_0^*(s,t) + \iint_{-\infty}^{s \wedge t} A_0(s,t,v,w) H_0(dv,w) Q_0(dw) \tag{3}$$

for some functions $\Sigma_0^*: I \times I \to \mathbb{R}$, $A_0: I \times I \times I \times \mathscr{W} \to \mathbb{R}$ and $H_0: I \times \mathscr{W} \to \mathbb{R}$ depending on P_0, where Q_0 is a probability measure induced by P_0 on some measurable space $\mathscr{W}$. In this representation, Σ_0^* is taken to be the differentiable portion of the covariance function, which does not contribute to the scale parameter. The second summand is not differentiable at (x, x) and makes $\sigma_{x,\alpha}(u, v)$ tend to a nonzero limit. We consider cases in which Σ_0^* and H_0 satisfy the following conditions:

(B3) Representation (3) holds, and for some $\delta > 0$, setting $B_\delta(x) := (x - \delta, x + \delta)$, it is also true that:

(B3a) Σ_0^* is symmetric in its arguments and continuously differentiable on $B_\delta(x)$;

(B3b) A_0 is symmetric in its first two arguments, and $s \mapsto A_0(s, t, v, w)$ is differentiable for Q_0-almost every w and each $s, t, v \in B_\delta(x)$, with derivative $A_0'(s,t,v,w)$ continuous in s, t, v each in $B_\delta(x)$ for Q_0-almost every w and satisfying

$$\iint_{-\infty}^{x+\delta} \sup_{s,t \in B_\delta(x)} \left| A_0'(s,t,v,w) \right| H_0(dv,w) Q_0(dw) < \infty;$$

(B3c) $v \mapsto A_0(x, x, v, w)$ is continuous at $v = x$ uniformly in w over the support of Q_0;

(B3d) $v \mapsto H_0(v, w)$ is nondecreasing for all w and differentiable at each $v \in B_\delta(x)$, with derivative $H_0'(v,w)$ continuous at $v = x$ uniformly in w over the support of Q_0.

Representation (3) is deliberately broad to encompass a wide variety of parameters, but in many settings, the covariance function can be considerably simplified, leading then to simpler conditions in (B3). For instance, when W is a vector of covariates over which marginalization is performed to compute the parameter, Q_0 typically plays the role of the marginal distribution of W under P_0. In classical problems in which there is no adjustment for covariates, this feature of representation (3) is not needed and indeed vanishes. In other settings, $A_0(s, t, v, w)$ depends on v and w but not on s and t.

Finally, we must ensure that the stochastic remainder terms $H_{x,n}$ and $R_{x,n}$ arising in (2) do not contribute to the limit distribution. Defining $\widetilde{H}_{u,n} := H_{x+u,n} - H_{x,n}$, $\widetilde{R}_{u,n} := R_{x+u,n} - R_{x,n}$ and $K_n(\delta) := n^{2/3} \sup_{|u| \le \delta n^{-1/3}} \left| \widetilde{H}_{u,n} - \theta_0(x) \widetilde{R}_{u,n} \right|$, we consider the following conditions for the asymptotic negligibility of these remainder terms:

(B4) $K_n(\delta) = o_P(1)$ for each fixed $\delta > 0$;

(B5) for some $\alpha \in (1,2)$, $\delta \mapsto \delta^{-\alpha} E_0\,[K_n(\delta)]$ is decreasing for all δ small enough and n large enough.

Condition (B4) guarantees that the remainder terms do not contribute to the weak convergence of $\{W_{n,x}(u) : |u| \le M\}$, and condition (B5) guarantees that the remainder terms satisfy condition (A3).

Combining the conditions above, we can state the following master theorem for pointwise convergence in distribution when the monotone estimator is based upon asymptotically linear primitive and transformation estimators.

THEOREM 4—*Suppose that, at an interior point $x \in I$, θ_0 is differentiable and Φ_0 is continuously differentiable with positive derivative. Suppose also that Γ_n and Φ_n satisfy display* (2), *and that conditions* (B1)–(B5) *and* (A4)–(A5) *hold* (*with $c_n = n^{1/3}$*). Then it holds that*

$$n^{1/3}[\theta_n(x) - \theta_0(x)] \xrightarrow{d} \tau_0(x) Z,$$

where $\tau_0(x) := \left[4\theta_0'(x)\kappa_0(x)/\Phi_0'(x)^2\right]^{1/3}$ is a scale factor involving $\kappa_0(x) := \int A_0(x, x, x, w) H_0'(x, w) Q_0(dw)$ and Z follows the Chernoff distribution.

4.2. Effect of domain transform on limit distribution

As was done briefly after Theorem 3, it is natural to compare the limit distribution obtained by Theorem 4 when a transformation of the domain is used and when it is not. We will consider $\theta_n := \mathrm{Iso}_I(\Theta_n, \mathrm{Id})$, the estimator obtained by directly isotonizing an estimator Θ_n of the primitive function Θ_0 without use of a domain transformation. Denoting by Φ_0 a candidate nondecreasing transformation function, and letting $\Gamma_0 := \Psi_0 \circ \Phi_0$ be as described in Section 2, we will also consider $\theta_n^* := \mathrm{Iso}_{J_n}(\Gamma_n \circ \Phi_n^-, \Phi_n)$, where Γ_n and Φ_n are estimators of Γ_0 and Φ_0, respectively. Suppose $\Theta_n(x)$, $\Gamma_n(x)$ and $\Phi_n(x)$ are each asymptotically linear estimators of their respective targets with influence functions $M_{x,0}^*$, $D_{x,0}^*$ and $L_{x,0}^*$, respectively, under sampling from P_0.

We wish to compare the scale parameters $\kappa_0(x)$ and $\kappa_0^*(x)$ arising from the use of the distinct estimators $\theta_n(x)$ and $\theta_n^*(x)$. To do so, we can use expression (B3) to examine the covariance obtained in both cases. However, it appears difficult to say much without having more specific forms for the involved influence functions. Unfortunately, it also appears difficult to characterize these influence functions generally since they depend inherently on the parameter of interest θ_0, and we wish to remain agnostic to the form of θ_0. Nevertheless, in our next result, we describe a class of problems, characterized by the generated influence functions and regularity conditions on these, in which domain transformation has no effect on the limit distribution of the generalized Grenander-type estimator.

THEOREM 5—*Suppose conditions* (B1)–(B5) *hold for* (Θ_n, Id) *and* (Γ_n, Φ_n), *and the observed data unit can be partitioned as* $O = (U, Z)$ *with* $U \in \mathbb{R}^+$. *Suppose that the influence functions can be expressed as*

$$M_{x,0}^*: (u, z) \mapsto I_{[0,x]}(u) M_{x,0}^{(1)}(u, z) + M_{x,0}^{(2)}(u, z),$$

$$L_{x,0}^*: (u, z) \mapsto I_{[0,x]}(u) L_{x,0}^{(1)}(u, z) + L_{x,0}^{(2)}(u, z),$$

$$D_{x,0}^*: (u, z) \mapsto I_{[0,x]}(u) \Phi_0'(u) M_{x,0}^{(1)}(u, z) + D_{x,0}^{(2)}(u, z) + \int_0^x \theta_0(v) L_{dv,0}^*(u, z),$$

and satisfy the smoothness conditions stated in the Supplementary Material. *Suppose that the density function* h_0 *of the conditional distribution of U given Z exists and is continuous in a neighborhood of x uniformly over the support of the marginal distribution* $Q_{Z,0}$ *of Z. Then it follows that*

$$\kappa_0(x) = \int \left[M_{x,0}^{(1)}(x, z)\right]^2 h_0(x \mid z) Q_{Z,0}(dz) \quad \text{and}$$

$$\kappa_0^*(x) = \left[\Phi_0'(x)\right]^2 \int \left[M_{x,0}^{(1)}(x, z)\right]^2 h_0(x \mid z) Q_{Z,0}(dz).$$

Consequently, $n^{1/3}[\theta_n(x) - \theta_0(x)]$ and $n^{1/3}[\theta_n^*(x) - \theta_0(x)]$ *have the same limit distribution.*

The forms of $M_{x,0}^*$ and $L_{x,0}^*$ arise naturally in a wide variety of settings because the parameters considered involve a primitive function. The supposed form of $D_{x,0}^*$ may seem restrictive at first glance but is in fact expected given the forms of $M_{x,0}^*$ and $L_{x,0}^*$. A heuristic justification based on the product rule for differentiation is provided in the Supplementary Material. In all of the examples we study in Section 5, the conditions of Theorem 5 apply. This provides justification for why, in each of these examples, the use of a domain transform has no impact on the limit distribution.

We remind the reader that, even if the domain transformation has no impact on the pointwise limit distribution, use of a domain transformation is still of great practical value in many circumstances. In complex problems, an estimator Θ_n may not be readily available for the primitive parameter Θ_0 obtained without the use of a domain transformation. In some cases, Θ_0 may not even be well defined, so that transformation of the domain is unavoidable. Even when Θ_0 is well defined and an estimator Θ_n is available, with the use of a carefully chosen transformation, it may be possible to avoid the need to estimate certain nuisance parameters or to substantially simplify the verification of conditions (B1)–(B5). Examples of these phenomena are presented in Section 5.

4.3. Negligibility of remainder terms

In some applications, the estimators Γ_n and Φ_n may be linear rather than simply asymptotically linear. In such situations, the remainder terms $H_{x,n}$ and $R_{x,n}$ are identically zero, and conditions (B4) and (B5) are trivially satisfied. Otherwise, these conditions must be verified. While in general the exact form of these remainder terms depends upon the specific parameter under consideration and estimators used, it is frequently the case that part of the remainder is an empirical process term arising from the estimation of nuisance functions appearing in the influence functions $D^*_{x,0}$ and $L^*_{x,0}$, as we illustrate below with one particular construction. To facilitate the verification of conditions (B4) and (B5) for these empirical process terms, we outline sufficient conditions in terms of uniform entropy and bracketing numbers.

In this subsection, we assume that $\Gamma_0(x)$ and $\Phi_0(x)$ arise as the evaluation at P_0 of maps from $\mathscr{M}$ to $\mathbb{R}$, and denote by $\Gamma_P(x)$ and $\Phi_P(x)$ the evaluation of these maps at an arbitrary $P \in \mathscr{M}$. Let $\pi = \pi(P)$ be a summary of P, and suppose that $\Gamma_P(x)$, $\Phi_P(x)$ and the nonparametric efficient influence functions of $P \mapsto \Gamma_P(x)$ and $P \mapsto \Phi_P(x)$ *at P* each only depend on P through π. Denote these efficient influence functions by $D^*_x(\pi)$ and $L^*_x(\pi)$, respectively. Since $\mathscr{M}$ is nonparametric, it must be that $D^*_{x,0} = D^*_x(\pi_0)$ and $L^*_{x,0} = L^*_x(\pi_0)$ for $\pi_0 := \pi(P_0)$. To emphasize the fact that $\Gamma_P(x)$ and $\Phi_P(x)$ depend on P only through π, we will use the symbols $\Gamma_\pi(x)$ and $\Phi_\pi(x)$ to refer to $\Gamma_P(x)$ and $\Phi_P(x)$, respectively.

Under regularity conditions, the so-called *one-step estimators*

$$\Gamma_n(x) := \Gamma_{\pi_n}(x) + \mathbb{P}_n D^*_x(\pi_n) \quad \text{and} \quad \Phi_n(x) := \Phi_{\pi_n}(x) + \mathbb{P}_n L^*_x(\pi_n) \tag{4}$$

are asymptotically linear and efficient estimators of $\Gamma_0(x)$ and $\Phi_0(x)$, even when π_n is a data-adaptive (e.g., machine learning) estimator of π_0 (e.g., Pfanzagl (1982)). van der Vaart and van der Laan (2006) pioneered the use of such one-step estimators in the context of nonparametric monotone function estimation. When this one-step construction is used, it can be shown that the remainder terms have the form $H_{x,n} = H_{1,x,n} + H_{2,x,n}$ and $R_{x,n} = R_{1,x,n} + R_{2,x,n}$, where $H_{1,x,n} := (\mathbb{P}_n - P_0)[D^*_x(\pi_n) - D^*_x(\pi_0)]$ and $R_{1,x,n} := (\mathbb{P}_n - P_0)[L^*_x(\pi_n) - L^*_x(\pi_0)]$ are empirical process terms, and $H_{2,x,n}$ and $R_{2,x,n}$ *are* so-called *second-order* remainder terms arising from linearization of the corresponding parameter. Similar representations exist when other constructive approaches, such as gradient-based estimating equations methodology (e.g., Tsiatis (2006), van der Laan and Robins (2003)) and targeted maximum likelihood estimation (e.g., van der Laan and Rose (2011)), are used. As we will see in the examples of Section 5, these second-order terms can usually be shown to be asymptotically negligible provided π_n tends to π_0 fast enough in some appropriate norm. Here, we provide conditions on π_n that ensure that the contribution of $H_{2,x,n} - \theta_0(x)R_{2,x,n}$ to $K_n(\delta)$ satisfies conditions (B4) and (B5).

A key benefit of decomposing the remainder terms as above is that the empirical process terms can be controlled using empirical process theory, a strategy also used in van der Vaart and van der Laan (2006). In particular, we can provide conditions under which $H_{1,x,n}$ and $R_{1,x,n}$ satisfy conditions (B4) and (B5). Defining $g_{x,u}(\pi) := \left[D^*_{x+u}(\pi) - D^*_x(\pi)\right] - \theta_0(x)\left[L^*_{x+u}(\pi) - L^*_x(\pi)\right]$, the relevant contribution of these empirical process terms to $K_n(\delta)$ is

$$K_{1,n}(\delta) := n^{1/6} \sup_{|u| \le \delta} \left|\mathbb{G}_n\left[g_{x,un^{-1/3}}(\pi_n) - g_{x,un^{-1/3}}(\pi_0)\right]\right|.$$

Suppose that π_n falls in a semimetric space $(\mathscr{P}, \rho)$, with probability tending to one, and that $G_{x,\mathscr{P},R}$ is an envelope function for $\mathscr{G}_{x,\mathscr{P},R} := \left\{g_{x,u}(\pi) : |u| \le R, \pi \in \mathscr{P}\right\}$. We consider the following the conditions:

(C1) for some constants $C > 0$ and $V \in [0,2)$, for all $\varepsilon \in (0,1]$ and R small enough, either one of these conditions hold:

(C1a) $\log N_{[]}\left(\varepsilon \|G_{x,\mathscr{P},R}\|_{P_0,2}, \mathscr{G}_{x,\mathscr{P},R}, L_2(P_0)\right) \le C\varepsilon^{-V}$;

(C1b) $\log \sup_Q N\left(\varepsilon \|G_{x,\mathscr{P},R}\|_{Q,2}, \mathscr{G}_{x,\mathscr{P},R}, L_2(Q)\right) \le C\varepsilon^{-V}$;

(C2) $P_0 G^2_{x,\mathscr{P},R} = O(R)$, and for all $\eta > 0$, $P_0 G^2_{x,\mathscr{P},R}\left\{R G_{x,\mathscr{P},R} > \eta\right\} = o(R)$, as $R \to 0$;

(C3) $P_0\left[g_{x,u}(\pi_1) - g_{x,u}(\pi_2)\right]^2 / \rho(\pi_1, \pi_2)^2 = O(|u|)$ uniformly for $\pi_1, \pi_2 \in \mathscr{P}$ and $u \in I$, and $P_0\left[g_{x,u}(\pi) - g_{x,v}(\pi)\right]^2 = O(|u - v|)$ uniformly for $\pi \in \mathscr{P}$;

(C4) there exists some $\bar{\pi} \in \mathscr{P}$ such that $\rho(\pi_n, \bar{\pi}) = o_P(1)$.

Our next result states that, under these conditions, the remainder term $K_{1,n}(\delta)$ stated above is asymptotically negligible in the sense of conditions (B4) and (B5).

THEOREM 6— *Suppose that, with probability tending to one,* $\pi_n \in \mathscr{P}$ *and conditions* (C1)–(C4) *hold. Then,* $K_{1,n}(\delta)$ *satisfies conditions* (B4)–(B5).

We note that conditions (C1) and (C2) together imply conditions (B1) and (B2). As such, if conditions (C1) and (C2) have been verified, there is no need to also verify conditions (B1) and (B2).

5. Applications of the general theory

In this section, we demonstrate the use of our general results for the three examples introduced in Section 2: estimation of monotone density, hazard and regression functions. For each of these functions, we consider various levels of complexity of the relationship between the ideal and observed data units. This allows us to illustrate that our general results (i) coincide with classical results in the simpler cases that have already been studied, and (ii) suggest novel estimation procedures with well-understood inferential properties, even in the context of complex problems that do not appear to have been previously studied. Below, we focus on distributional results for the various estimators considered. In each case, we state the main results in the text, and present additional technical details in the Supplementary Material.

5.1. Example 1: Monotone density function

Let $\theta_0 := f_0$ be the density function of an event time T with support $I := [0, u_0]$, and suppose that f_0 is known to be nondecreasing on I. We will not use any transformation in this example, so we take Φ_0 and Φ_n to be the identity map. Thus, $\psi_0 = \theta_0$ also corresponds to the density function of T, and $\Psi_0 = \Theta_0 = \Gamma_0$ to its distribution function. Below, we consider various data settings that increase in complexity. In the first setting, available observations are subject to independent right-censoring. In the second, the right-censoring mechanism is allowed to be informative—only conditional independence of the event and censoring times given a vector of observed covariates is assumed. The first case has been studied in the literature—for this, we wish to verify that our general results coincide with results already established. The second case is more difficult and does not seem to have been studied before. Our work in this setting not only highlights the generality of the theory in Sections 3 and 4, but also yields novel practical methodology.

5.1.1. Independent censoring—Suppose that C is a positive random variable independent of T, and that the observed data unit is $O = (Y, \Delta)$, where $Y = \min(T, C)$ and $\Delta = I(T \le C)$. The NPMLE of a monotone density function based on independently right-censored data was obtained in Laslett (1982) and McNichols and Padgett (1982), and distributional results were derived in Huang and Zhang (1994). Huang and Wellner (1995) considered an estimator θ_n obtained by differentiating the GCM of the Kaplan–Meier estimator of the distribution function. While this is not the NPMLE, Huang and Wellner (1995) showed that it is asymptotically equivalent to the NPMLE, and it is an attractive estimator because it is simple to construct and reduces to the Grenander estimator if T is fully observed, that is, if $C \ge T$ almost surely.

Since Ψ_0 is the distribution function $F_0 = 1 - S_0$ with S_0 denoting the survival function of T, it is natural to consider $\Psi_n := 1 - S_n$, where S_n is the Kaplan–Meier estimator of S_0. It is well known that $n^{1/2}(S_n - S_0)$ converges weakly in $\ell^\infty([0, \tau])$ to a tight zero-mean Gaussian process as long as $G_0(\tau) > 0$ and $S_0(\tau) < 1$, where G_0 denotes the survival function of C. Denoting by Λ_0 the cumulative hazard function corresponding to S_0, the influence function of the Kaplan–Meier estimator $S_n(x)$ is known to be the nonparametric efficient influence function

$$D^*_{0,x} : (y, \delta) \mapsto S_0(x)\left[-\frac{\delta I_{[0,x]}(y)}{S_0(y)G_0(y)} + \int_0^{y \wedge x} \frac{\Lambda_0(du)}{G_0(u)S_0(u)}\right]$$

and so, the local difference $g_{x,u}(y, \delta)$ can be written as

$$\frac{[S_0(x) - S_0(x+u)]\delta I_{[0,x+u]}(y)}{S_0(y)G_0(y)} - \frac{S_0(x)\delta I_{(x,x+u]}(y)}{S_0(y)G_0(y)} + \int_0^y \frac{I_{(x,x+u]}(v)}{S_0(v)G_0(v)}\Lambda_0(dv).$$

In the Supplementary Material, we verify that condition (B2) is satisfied if S_0 and G_0 are positive in a neighborhood of x, and that condition (B3) is satisfied if θ_0 is positive and continuous in a neighborhood of x. The covariance function is given by $\Sigma_0 : (s, t) \mapsto \int_0^{s \wedge t} S_0(s)S_0(t)/\{S_0(u)G_0(u)\}\Lambda_0(du)$. We then get $\kappa_0(x) = [S_0(x)/G_0(x)]\lambda_0(x) = f_0(x)/G_0(x)$, so that the scale parameter is $\tau_0(x) = [4f_0'(x)f_0(x)/G_0(x)]^{1/3}$. This agrees with the results of Huang and Wellner (1995). In the Supplementary Material, we demonstrate that conditions (B4) and (B5) are also satisfied. In the case of no censoring, $\Sigma_0(s, t)$ simplifies to $\Gamma_0(s \wedge t) - \Gamma_0(s)\Gamma_0(t)$, so that $\Sigma_0^*(s, t) = \Theta_0(s)\Theta_0(t)$, $A_0(s, t, y, w) = \theta_0(y)H_0(y, w) = y$ and $\kappa_0(x) = \theta_0(x)$. This agrees with the classical result of Prakasa Rao (1969) concerning pointwise convergence in distribution of the Grenander estimator.

5.1.2. Conditionally independent censoring—

In many cases, the censoring mechanism may be informative but still independent of the event time process conditionally on a vector of recorded covariates. For simplicity, we only consider the case in which these covariates are defined at baseline, though the case of time-varying covariates can be tackled similarly. The observed data unit is now $O = (Y, \Delta, W)$, and we assume that T and C are independent given W. As long as $P_0(\Delta = 1|W)$ is bounded away from zero almost surely, the survival function S_0 of T can be identified pointwise in terms of the distribution P_0 of O via the product-limit transform

$$S_0(x) = \prod_{t \leq x}\left[1 - \frac{F_{1,0}(dt, w)}{S_{Y,0}(t|w)}\right]Q_0(dw),$$

where $F_{1,0}(t|w) := P_0(Y \leq t, \Delta = 1|W = w)$ is the conditional subdistribution function of Y given $W = w$ corresponding to $\Delta = 1$, $S_{Y,0}(t|w) := P_0(Y \geq t|W = w)$ is the conditional proportion-at-risk at time t given $W = w$, and Q_0 is the marginal distribution of W under P_0. This constitutes an example of coarsening at random, as described in Heitjan and Rubin (1991) and Gill, Van Der Laan and Robins (1997). Estimation of S_0 in the context of conditionally independent censoring has been studied before by Hubbard, van der Laan and Robins (2000), Scharfstein and Robins (2002) and Zeng (2004), among others.

In this context, the nonparametric efficient influence function $D_{0,x}^*$ of $S_0(x)$ has the form $D_{0,x} - S_0(x)$, where $D_{0,x}$ is given by

$$(y, \delta, w) \mapsto -S_0(x|w)\left[\frac{\delta I_{(-\infty, x]}(y)}{S_0(y|w)G_0(y|w)} - \int_0^{y \wedge x}\frac{\Lambda_0(du|w)}{S_0(u|w)G_0(u|w)} - 1\right]$$

with $S_0(x \mid w)$ and $G_0(x \mid w)$ the conditional survival functions of T and C, respectively, at x given $W = w$, and $\Lambda_0(x \mid w)$ is the conditional cumulative hazard function of T at x given $W = w$. A simple one-step estimator of $\Gamma_0(x)$ is given by $\Gamma_n(x) := 1 - \mathbb{P}_n D_{n,x}$, where $D_{n,x}$ is obtained by substituting S_n and G_n for S_0 and G_0, respectively, in $D_{0,x}$. Conditions (B1) and (B2) are satisfied under uniform Lipschitz conditions on S_0 and G_0. As we show in the

Supplementary Material, condition (B3) holds, and we get $\kappa_0(x) = \int [f_0(x|w)/G_0(x|\text{w})] \, Q_0(dw)$, where $f_0(x \mid w)$ is the conditional density of T at x given $W = w$. It follows directly then that the Chernoff scale factor is

$$\tau_0(x) = \left[4 f_0'(x) \int \frac{f_0(x|w)}{G_0(x|w)} Q_0(dw)\right]^{1/3},$$

which reduces to the scale factor of Huang and Wellner (1995) when T and C are independent. In the Supplementary Material, we demonstrate that satisfaction of condition (B4) is highly dependent on the behavior of S_n and G_n. For instance, if $S_n - S_0$ and $G_n - G_0$ uniformly tend to zero in probability at rates faster than $n^{-1/3}$, then conditions (B4) and (B5) are satisfied. This is not a restrictive requirement if W only has few components—in such cases, many nonparametric smoothing-based estimators satisfy such rates. Otherwise, semiparametric estimators building upon additional structure (e.g., additivity on an appropriate scale) could be used. Alternatively, for higher-dimensional W, estimators of the form $S_n(x|w) = \exp\left[-\int_0^x \lambda_n(v|w)dv\right]$ with λ_n an estimator of the conditional hazard λ_0 may be worth considering. For such S_n, we require the product of the convergence rates of $\lambda_n - \lambda_0$ and $G_n - G_0$ to be faster than $n^{-1/3}$. In practice, with a moderate or high-dimensional covariate vector W, it seems desirable to leverage multiple candidate estimators using ensemble learning (e.g., van der Laan, Polley and Hubbard (2007), van der Laan and Rose (2011)). In the Supplementary Material (Westling and Carone (2020)), we conduct a simulation study validating these results using Cox's proportional hazard model for S_n and G_n.

5.2. Example 2: Monotone hazard function

We now consider estimation of $\theta_0 := \lambda_0$, the hazard function of T. The most obvious approach to tackle this problem would be to consider an identity transformation as in the previous example. The primitive function of interest is then the cumulative hazard function Λ_0, which can be expressed as the negative logarithm of the survival function S_0 and estimated naturally using any asymptotically linear estimator of S_0, for example. The conditions of Theorems 3 and 4 can then be directly verified. An alternative, more expeditious approach consists of taking the domain transform Φ_0 to be the restricted mean mapping $u \mapsto \int_0^u S_0(v)dv$. In such cases, Γ_0 is simply the cumulative distribution function F_0, and $u_0 = \int_0^\infty S_0(v)dv$ the mean of T. This particular choice of domain transformation for estimating a monotone hazard function therefore yields the same parameter Γ_0 as for estimating a monotone density with the identity transform. Denoting by S_n the estimator of the survival function S_0 based on the available data, the resulting generalized Grenander-type estimator θ_n is defined by taking $\Gamma_n := 1 - S_n$ and setting Φ_n to be $u \mapsto \int_0^u S_n(v)dv$ over $J_n = [0, u_n]$, where $u_n = \int_0^\infty S_n(v)dv$. As the result below suggests, when this special domain transform is used, we can leverage some of the work performed above in analyzing the Grenander-type estimator of a monotone density function under the various right-censoring schemes considered. We recall that Id denotes the identity function.

THEOREM 7

Suppose that $E_0\left[\sup_{u \in I_n}|S_n(u) - S_0(u)|\right] = o\left(r_n^{-1}\right)$ *and set* $\Gamma_n := 1 - S_n$*. If the pair* (Γ_n, Id) *satisfies conditions* (A1)–(A3), *then the pair* (Γ_n, Φ_n) *with* $\Phi_n: u \mapsto \int_0^u S_n(v)dv$ *necessarily satisfies conditions* (A1)–(A5). *In particular, for* $\theta_n := \mathrm{Iso}_{J_n}(\Gamma_n \circ \Phi_n^-, \Phi_n)$, *this implies that*

$$r_n[\theta_n(x) - \theta_0(x)] \xrightarrow{d} -\theta_0'(x) \operatorname*{argmin}_{u \in \mathbb{R}}\left\{W_x(u) + \frac{1}{2}\theta_0'(x)S_0(x)u^2\right\}.$$

If $W_x = [\kappa_0(x)]^{1/2} W_0$ *for* W_0 *a two-sided standard Brownian motion, then*

$$r_n[\theta_n(x) - \theta_0(x)] \xrightarrow{\mathrm{d}} \tau_0(x)Z,$$

where Z follows the Chernoff distribution and $\tau_0(x) := \left[4\theta_0'(x)\kappa_0(x)/S_0(x)^2\right]^{1/3}$.

Denote by $T_{(j)}$ the jth order statistic of $\{T_1, T_2, \ldots, T_n\}$ and define $T_{(0)} := 0$. When there is no censoring, the choice (Γ_n, Φ_n) prescribed above indicates that Γ_n is the empirical distribution function based on $Y_1, Y_2, \ldots, Y_n$, and Φ_n is defined pointwise as $\Phi_n(x) := \frac{1}{n}\sum_{i=1}^n \min\left(T_{(i)}, x\right)$, which is strictly increasing on $[0, T_{(n)}]$. Therefore, $\theta_n(x)$ is the left derivative at $\Phi_n(x)$ of the GCM of the graph of $\left\{\left(\Phi_n\left(T_{(k)}\right), \Gamma_n\left(T_{(k)}\right)\right): k = 0, 1, \ldots, n\right\} = \left\{\left(\left(\frac{n-k}{n}\right)T_{(k)} + \frac{1}{n}\sum_{i=1}^k T_{(i)}, \frac{k}{n}\right): k = 0, 1, \ldots, n\right\}$. This is the NPMLE of a nondecreasing hazard function with uncensored data; see, for example, Chapter 2.6 of Groeneboom and Jongbloed (2014).

In the Supplementary Material, we verify conditions (A1)–(A3) for each of three right-censoring schemes when $\Theta_n := 1 - S_n$, and Φ_0 and Φ_n are both equal to the identity. Thus, to use Theorem 7, it would suffice to verify that $E_0\left[\sup_{u \in I_n}|S_n(u) - S_0(u)|\right]$ tends to zero faster than $n^{-1/3}$. This is straightforward given the weak convergence of $n^{1/2}(S_n - S_0)$. Thus, the above theorem provides distributional results for monotone hazard function estimators in each right-censoring scheme considered, as summarized below:

i. no censoring: $\tau_0(x) = [4\lambda_0'(x)\lambda_0(x)/S_0(x)]^{1/3}$, which agrees with results from Prakasa Rao (1970);

ii. independent right-censoring: $\tau_0(x) = [\lambda_0'(x)\lambda_0(x)/\{G_0(x)S_0(x)\}]^{1/3}$, which agrees with results from Huang and Wellner (1995);

iii. conditionally independent right-censoring, an important setting that does not seem tohave been previously studied in the literature:

$$\tau_0(x) = \left\{\frac{4\lambda_0'(x)\lambda_0(x)}{G_0(x)S_0(x)}\left[\frac{G_0(x)}{f_0(x)}\int \frac{f_0(x \mid w)}{G_0(x \mid w)}Q_0(dw)\right]\right\}^{1/3}.$$

If either T or C are independent of W, the unadjusted Kaplan–Meier estimator is consistent for the true marginal survival function of T, and so, unadjusted estimators of the density and hazard functions are consistent. In these cases, we may then ask how the asymptotic distributions of the adjusted and unadjusted estimators compare. Since all limit distributions are of the scaled Chernoff type, it suffices to compare the scale factors arising from the different estimators. The second expression in (iii) is helpful to assess the impact of unnecessary covariate adjustment. If C and W are independent, then $G_0(x \mid w) = G_0(x)$ for each w, and so, the scale factors in (ii) and (iii) are identical. If T and W are dependent, so that $f_0(x \mid w) = f_0(x)$ for each w, but C and W are not, then the scale factor in (iii) is generally larger than the scale factor in (ii). In summary, when using an adjusted rather than unadjusted estimator of the hazard function, there may only be a penalty in asymptotic efficiency when adjusting for covariates that C depends on but T does not. The relative loss of efficiency is given by $\left\{\int [G_0(x)/G_0(x \mid w)] Q_0(dw)\right\}^{1/3}$. In the Supplementary Material, we conduct a simulation study validating these results.

5.3. Example 3: Monotone regression function

We finally consider estimation of a monotone regression function. We first focus on the simple case in which the association between the outcome and exposure of interest is not confounded. In such cases, the parameter of interest is the conditional mean of the outcome given exposure level, and the standard least-squares isotonic regression estimators can be used. We show that our general theory covers this classical case. We then consider the case in which the relationship between outcome and exposure is confounded but the confounders of this relationship have been recorded. In this more challenging case, we consider the marginalization (or standardization) of the conditional mean outcome given exposure level and confounders over the marginal confounder distribution. We study this problem using results from Section 4, which allow us to provide theory for a novel estimator proposed for this important case.

5.3.1. No confounding—In the standard least-squares isotonic regression problem, we observe independent replicates of $O := (A, Y)$, where $Y \in \mathbb{R}$ is an outcome and $A \in \mathbb{R}$ is the exposure of interest. We are interested in the conditional mean function $\theta_0 := \mu_0$, where $\mu_0(x) := E_0(Y \mid A = x)$ is the mean outcome at exposure level x. The primitive function of θ_0 can be written as $\Theta_0(t) = E_0\left[Y I_{(-\infty, t]}(A)/f_0(A)\right]$ for each t, where f_0 is the marginal density of A. The corresponding primitive parameter at x is pathwise differentiable with nonparametric efficient influence function $(a, y) \mapsto y I_{(-\infty, x]}(a)/f_0(a) - \Theta_0(x)$. An obvious approach to estimation of θ_0 consists of constructing an asymptotically linear estimator of Θ_0—this involves nonparametric estimation of the nuisance density f_0—and differentiating the GCM of the resulting curve—this involves selecting the interval over which the GCM is calculated.

By using a domain transformation, it is possible to avoid both the need for nonparametric density estimation and the choice of isotonization interval. Let Φ_0 be the marginal distribution function of A. With this transformation, we note that $\Psi_0(t) = E_0\left[Y I_{(-\infty, t]}(\Phi_0(A))\right]$ and $\Gamma_0(t) = E_0\left[Y I_{(-\infty, t]}(A)\right]$ for each t. This suggests taking Φ_n

to be the empirical distribution function based on A_1, A_2,...,A_n and $\Gamma_n(x) := \frac{1}{n}\sum_{i=1}^n Y_i I_{(-\infty, x]}(A_i)$. The resulting estimator $\theta_n(x)$ is precisely the well-known least-squares isotonic regression estimator of $\theta_0(x)$. Since Φ_n is a step function with jumps at the observed values of A, $\theta_n(x)$ is equal to the left-hand slope of the GCM at $\Phi_n(x)$ of the so-called *cusum diagram* $\{(\Phi_n(A_k), \Gamma_n(A_k)) : k = 0, 1, \ldots, n\} = \left\{\left(\frac{k}{n}, \frac{S_k}{n}\right) : k = 0, 1, \ldots, n\right\}$, where we let $A_0 = -\infty$, $S_0 = 0$ and $S_k = \sum_{i=1}^k Y_i$ for $k \geq 1$.

Because both Γ_n and Φ_n are linear estimators, these estimators do not generate second-order remainder terms to analyze. The influence functions of Γ_n and Φ_n are, respectively, $D_{0,x}^* : (a, y) \mapsto yI_{(-\infty, x]}(a) - \Gamma_0(x)$ and $L_{0,x}^* : (a, y) \mapsto I_{(-\infty, x]}(a) - \Phi_0(x)$. In the Supplementary Material, we demonstrate that if in a neighborhood of x, the conditional variance function, defined pointwise as $\sigma_0^2(t) := \mathrm{Var}_0(Y \mid A = t)$, is bounded and continuous, and Φ_0 possesses a positive, continuous density, then Theorem 4 holds with

$$\tau_0(x) = \left[\frac{4\mu_0'(x)\sigma_0^2(x)}{f_0(x)}\right]^{1/3},$$

coinciding with the classical results of Brunk (1970).

5.3.2. Confounding by recorded covariates—We now consider a scenario in which the relationship between outcome Y and exposure A is confounded by a vector W of recorded covariates. The observed data unit is thus $O := (W, A, Y)$. A more relevant estimand in this scenario might be the marginalized regression function $\theta_0 := \nu_0$ with $\nu_0(x)$ defined as $E_0[E_0(Y \mid A = x, W)]$. We note that $\nu_0(x)$ can be interpreted as a causal dose-response curve if (i) W includes all confounders of the relationship between A and Y, and (ii) the probability of observing an individual subject to exposure level x is positive in P_0-almost every stratum defined by W. In many scientific settings, it may be known that the causal dose-response curve is monotone in exposure level.

We again consider transformation by the marginal distribution function of A. In other words, we set $\Phi_0(x) := P_0(A \leq x)$ and take $\Phi_n(x) := \frac{1}{n}\sum_{i=1}^n I_{(-\infty, x]}(A_i)$ for each x. We then have that

$$\Gamma_0(x) = E_0\left[\frac{YI_{(-\infty, x]}(A)}{g_0(A, W)}\right] = \iint I_{(-\infty, x]}(a)\mu_0(a, w)\Phi_0(da)Q_0(dw),$$

where g_0 is the density ratio $(a, w) \mapsto f_0(a \mid w)/f_0(a)$, with $f_0(a \mid w)$ denoting the conditional density function of A at a given $W = w$ and $f_0(a)$ the marginal density function of A at a as before, and μ_0 is the regression function $(a, w) \mapsto E_0(Y \mid = a, W = w)$. While in this case the domain transform does not eliminate the need to estimate nuisance functions, it nevertheless results in a procedure for which there is no need to choose the interval over which the GCM is calculated.

Setting $\eta_0(x, w) := \int I_{(-\infty, x]}(a)\mu_0(a, w)\Phi_0(da)$ for each x and w, the nonparametric efficient influence function of $\Gamma_0(x)$ is

$$(w, a, y) \mapsto I_{(-\infty, x]}(a)\left[\frac{y - \mu_0(a, w)}{g_0(a, w)} + \theta_0(a)\right] + \eta_0(x, w) - 2\Gamma_0(x).$$

Suppose that μ_n and g_n are estimators of μ_0 and g_0, respectively. If the empirical distributions Φ_n and Q_n based on $A_1, A_2, \ldots, A_n$ and $W_1, W_2, \ldots, W_n$, respectively, are used as estimators of Φ_0 and Q_0, it can be shown that

$$\Gamma_n(x) := \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty, x]}(A_i)\left[\frac{Y_i - \mu_n(A_i, W_i)}{g_n(A_i, W_i)} + \frac{1}{n}\sum_{j=1}^{n} \mu_n\left(A_i, W_j\right)\right]$$

is a one-step estimator of $\Gamma_0(x)$, and that it is asymptotically efficient under regularity conditions on the nuisance estimators μ_n and g_n.

Conditions (B1)–(B5) can be verified with routine but tedious work. Here, we focus on condition (B3), which allows us to obtain the scale parameter of the limit distribution, and on condition (B4), which requires that the nuisance estimators converge sufficiently fast. We find that condition (B4) is satisfied if, for some $\epsilon > 0$,

$$\sup_{|x-u| \le \epsilon} E_0[\mu_n(u, W) - \mu_0(u, W)]^2 \sup_{|x-u| \le \epsilon} E_0\left[\frac{g_0(u, W)}{g_n(u, W)} - 1\right]^2 = o_P\left(n^{-1/3}\right),$$

and additional empirical process conditions hold. Turning to condition (B3), under smoothness conditions, $\kappa_0(x) = f_0(x)^2 \int \left[\sigma_0^2(x, w) / f_0(x \mid w)\right] Q_0(dw)$, where $\sigma_0^2 : (a, w) \mapsto \mathrm{Var}_0(Y \mid A = a, W = w)$ denotes the conditional variance function of Y given A and W. We then find that the scale parameter of the limit Chernoff distribution is

$$\tau_0(x) = \left\{4 v_0'(x) \int \left[\frac{\sigma_0^2(x, W)}{f_0(x \mid W)}\right] Q_0(dw)\right\}^{1/3}.$$

The marginalized and marginal regression functions exactly coincide, that is, $\nu_0 = \mu_0$ – if, for example, (i) Y and W are conditionally independent given A, or (ii) A and W are independent. It is natural then to ask how the limit distribution of estimators of these two parameters compare under scenarios (i) and (ii), when the parameters in fact agree with each other. In scenario (i), the scale parameter obtained based on the estimator accounting for potential confounding reduces to

$$\begin{aligned}\tau_{0,\mathrm{red}}(x) &= \left\{4\mu_0'(x)\sigma_0^2(x) \int \frac{Q_0(dw)}{f_0(x \mid w)}\right\}^{1/3} \\ &\ge \left\{\frac{4\mu_0'(x)\sigma_0^2(x)}{\int f_0(x \mid w) Q_0(dw)}\right\}^{1/3} = \left\{\frac{4\mu_0'(x)\sigma_0^2(x)}{f_0(x)}\right\}^{1/3}\end{aligned}$$

by Jensen's inequality. Thus, if Y and W are conditionally independent given A, in which case there is no need to adjust for potential confounders, the marginal isotonic regression estimator has a more concentrated limit distribution than the marginalized isotonic regression estimator. In scenario (ii), the scale parameter of the estimator accounting for potential confounding is

$$\tau_{0,\,\mathrm{red}}(x) = \left\{\frac{4\mu_0'(x)}{f_0(x)}\int \sigma_0^2(x, W)Q_0(dw)\right\}^{1/3} \leq \left\{\frac{4\mu_0'(x)\sigma_0^2(x)}{f_0(x)}\right\}^{1/3}$$

given that $\int \sigma_0^2(x, w)Q_0(dw) \leq \sigma_0^2(x)$ by the law of total variance. Thus, if A and W are independent, the marginal isotonic regression estimator has a less concentrated limit distribution than the marginalized isotonic regression estimator. In both scenarios (i) and (ii), the difference in concentration between the limit distributions of the two estimators varies with the amount of dependence between A and W. We note that these observations are analogous to those obtained in linear regression.

6. Concluding remarks

We have studied a broad class of estimators of monotone functions based on differentiating the greatest convex minorant of a preliminary estimator of a primitive parameter. A novel aspect of the class we have considered is its allowance for the primitive parameter to involve a possibly data-dependent transformation of the domain. The class we have defined is useful because it generalizes classical approaches for simple monotone functions, including density, hazard and regression functions, facilitates the integration of flexible, data-adaptive learning techniques, and allows valid asymptotic statistical inference. We have provided general asymptotic results for estimators in this class and have also derived refined results for the important case wherein the primitive estimator is uniformly asymptotically linear. We have proposed novel estimators of extensions of classical monotone parameters that deal with common sampling complications, and described their large-sample properties using our general results.

Our primary goal in this paper has been to establish general theoretical results that can be applied to study many specific estimators, and as such, there are numerous potential applications of our results. There are also a multitude of useful properties and modifications of Grenander-type estimators that have been studied in the literature and whose extension to our class would be important. For instance, kernel smoothing of a Grenander-type estimator yields a monotone estimator that possesses many of the properties of usual kernel smoothing estimators, including possibly faster convergence to a normal distribution (e.g., Groeneboom, Jongbloed and Witte (2010), Mammen (1991), Mukerjee (1988)). The asymptotic distribution of the supremum norm error of Grenander-type estimators has also been derived (e.g., Durot, Kulikov and Lopuhaä (2012)), and extending this result to our class would refine further our pointwise results. Asymptotic results at the boundaries of the domain and corrections for poor behavior there have been developed and would further enhance the utility of these methods (e.g., Balabdaoui et al. (2011), Woodroofe and Sun (1993), Kulikov and Lopuhaä (2006)).

There have also been various proposals for constructing asymptotically valid pointwise confidence intervals for Grenander-type estimators without the need to compute the complicated scale parameters appearing in their limit distribution. In regular statistical problems, the bootstrap is one of the most widely used such methods; unfortunately, the nonparametric bootstrap is known to fail for Grenander-type estimators (e.g., Kosorok (2008), Sen, Banerjee and Woodroofe (2010)). However, these articles have demonstrated that the m-out-of-n bootstrap can be valid for Grenander-type estimators, and that bootstrapping smoothed versions of Grenander-type estimators can also be an effective strategy for performing inference. Asymptotically pivotal distributions based on likelihood ratios have also been used to avoid the need to estimate nuisance parameters in the limit distribution and to provide a basis for improved finite-sample inference (e.g., Banerjee and Wellner (2001), Banerjee (2005a, 2005b, 2007), Groeneboom and Jongbloed (2015)). Considering these strategies in our setting would be particularly interesting.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank the referees and associate editor for their constructive and insightful comments that helped improve this manuscript. They also thank Antoine Chambaz and Mark van der Laan for stimulating conversations that sparked their interest in this problem, Jon Wellner for sharing insight on the history of this problem and Alex Luedtke and Peter Gilbert for providing feedback early on in this work.

Both authors were supported by NIAID grant 5UM1AI058635.

The second author was supported by the Career Development Fund of the Department of Biostatistics at the University of Washington.

REFERENCES

ANEVSKI D and HÖSSJER O (2006). A general asymptotic scheme for inference under order restrictions. Ann. Statist 34 1874–1930. MR2283721 10.1214/009053606000000443

ANEVSKI D and SOULIER P (2011). Monotone spectral density estimation. Ann. Statist 39 418–438. MR2797852 10.1214/10-AOS804

BAGCHI P, BANERJEE M and STOEV SA (2016). Inference for monotone functions under short- and long-range dependence: Confidence intervals and new universal limits. J. Amer. Statist. Assoc 111 1634–1647. MR3601723 10.1080/01621459.2015.1100622

BALABDAOUI F, JANKOWSKI H, PAVLIDES M, SEREGIN A and WELLNER J (2011). On the Grenander estimator at zero. Statist. Sinica 21 873–899. MR2829859 10.5705/ss.2011.038a

BANERJEE M (2005a). Likelihood ratio tests under local alternatives in regular semiparametric models. Statist. Sinica 15 635–644. MR2233903

BANERJEE M (2005b). Likelihood ratio tests under local and fixed alternatives in monotone function problems. Scand. J. Stat 32 507–525. MR2232340 10.1111/j.1467-9469.2005.00458.x

BANERJEE M (2007). Likelihood based inference for monotone response models. Ann. Statist 35 931–956. MR2341693 10.1214/009053606000001578

BANERJEE M and WELLNER JA (2001). Likelihood ratio tests for monotone functions. Ann. Statist 29 1699–1731. MR1891743 10.1214/aos/1015345959

BEARE BK and FANG Z (2017). Weak convergence of the least concave majorant of estimators for a concave distribution function. Electron. J. Stat 11 3841–3870. MR3714300 10.1214/17-EJS1349

BRUNK HD (1970). Estimation of isotonic regression In Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969) 177–197. Cambridge Univ. Press, London MR0277070

CAROLAN C and DYKSTRA R (1999). Asymptotic behavior of the Grenander estimator at density flat regions. Canad. J. Statist 27 557–566. MR1745821 10.2307/3316111

DEDECKER J, MERLEVÈDE F and PELIGRAD M (2011). Invariance principles for linear processes with application to isotonic regression. Bernoulli 17 88–113. MR2797983 10.3150/10-BEJ273

DUROT C (2007). On the $\mathbb{L}_p$-error of monotonicity constrained estimators. Ann. Statist 35 1080–1104. MR2341699 10.1214/009053606000001497

DUROT C, GROENEBOOM P and LOPUHAÄ HP (2013). Testing equality of functions under monotonicity constraints. J. Nonparametr. Stat 25 939–970. MR3174305 10.1080/10485252.2013.826356

DUROT C, KULIKOV VN and LOPUHAÄ HP (2012). The limit distribution of the L_∞-error of Grenander-type estimators. Ann. Statist 40 1578–1608. MR3015036 10.1214/12-AOS1015

DUROT C and LOPUHAÄ HP (2014). A Kiefer–Wolfowitz type of result in a general setting, with an application to smooth monotone estimation. Electron. J. Stat 8 2479–2513. MR3285873 10.1214/14-EJS958

GILL RD, VAN DER LAAN MJ and ROBINS JM (1997). Coarsening at random: Characterizations, conjectures, counter-examples. In Proceedings of the First Seattle Symposium in Biostatistics (Lin DY, ed.) 255–294. Springer, New York.

GRENANDER U (1956). On the theory of mortality measurement. II. Scand. Actuar. J 39 125–153.

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). Wadsworth Statist./ Probab. Ser 539–555. Wadsworth, Belmont, CA MR0822052

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 10.1017/CBO9781139020893

GROENEBOOM P and JONGBLOED G (2015). Nonparametric confidence intervals for monotone functions. Ann. Statist 43 2019–2054. MR3375875 10.1214/15-AOS1335

GROENEBOOM P, JONGBLOED G and WITTE BI (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. Ann. Statist 38 352–387. MR2589325 10.1214/09-AOS721

GROENEBOOM P and WELLNER JA (2001). Computing Chernoff's distribution. J. Comput. Graph. Statist 10 388–400. MR1939706 10.1198/10618600152627997

HEITJAN DF and RUBIN DB (1991). Ignorability and coarse data. Ann. Statist 19 2244–2253. MR1135174 10.1214/aos/1176348396

HUANG J and WELLNER JA (1995). Estimation of a monotone density or monotone hazard under random censoring. Scand. J. Stat 22 3–33. MR1334065

HUANG Y and ZHANG C-H (1994). Estimating a monotone density from censored observations. Ann. Statist 22 1256–1274. MR1311975 10.1214/aos/1176325628

HUBBARD AE, VAN DER LAAN MJ and ROBINS JM (2000). Nonparametric locally efficient estimation of the treatment specific survival distribution with right censored data and covariates in observational studies In Statistical Models in Epidemiology, the Environment, and Clinical Trials (Minneapolis, MN, 1997). IMA Vol. Math. Appl 116 135–177. Springer, New York MR1731683 10.1007/978-1-4612-1284-3_3

KIM J and POLLARD D (1990). Cube root asymptotics. Ann. Statist 18 191–219. MR1041391 10.1214/aos/1176347498

KOSOROK MR (2008). Bootstrapping the Grenander estimator In Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen (Balakrishnan N, Peña EA and Silvapulle MJ, eds.). Collections 1 282–292. Institute of Mathematical Statistics 10.1214/193940307000000202

KULIKOV VN and LOPUHAÄ HP (2006). The behavior of the NPMLE of a decreasing density near the boundaries of the support. Ann. Statist 34 742–768. MR2283391 10.1214/009053606000000100

LASLETT GM (1982). The survival curve under monotone density constraints with applications to twodimensional line segment processes. Biometrika 69 153–160. MR0655680 10.1093/biomet/69.1.153

LEURGANS S (1982). Asymptotic distributions of slope-of-greatest-convex-minorant estimators. Ann. Statist 10 287–296. MR0642740

LOPUHAÄ HP and MUSTA E (2018a). A central limit theorem for the Hellinger loss of Grenander-type estimators. Stat. Neerl To appear. 10.1111/stan.12153.

LOPUHAÄ HP and MUSTA E (2018b). The distance between a naive cumulative estimator and its least concave majorant. Statist. Probab. Lett 139 119–128. MR3802192 10.1016/j.spl.2018.04.001

MAMMEN E (1991). Estimating a smooth monotone regression function. Ann. Statist 19 724–740. MR1105841 10.1214/aos/1176348117

MCNICHOLS DT and PADGETT WJ (1982). Maximum likelihood estimation of unimodal and decreasing densities based on arbitrarily right-censored data. Comm. Statist. Theory Methods 11 2259–2270. MR0678684 10.1080/03610928208828387

MUKERJEE H (1988). Monotone nonparameteric regression. Ann. Statist 16 741–750. MR0947574 10.1214/aos/1176350832

PFANZAGL J (1982). Contributions to a General Asymptotic Statistical Theory Lecture Notes in Statistics 13 Springer, New York MR0675954

PRAKASA RAO BLS (1969). Estimation of a unimodal density. Sankhya A 31 23–36. MR0267677

PRAKASA RAO BLS (1970). Estimation for distributions with monotone failure rate. Ann. Math. Stat 41 507–519. MR0260133 10.1214/aoms/1177697091

SCHARFSTEIN DO and ROBINS JM (2002). Estimation of the failure time distribution in the presence of informative censoring. Biometrika 89 617–634. MR1929167 10.1093/biomet/89.3.617

SEN B, BANERJEE M and WOODROOFE M (2010). Inconsistency of bootstrap: The Grenander estimator. Ann. Statist 38 1953–1977. MR2676880 10.1214/09-AOS777

TSIATIS AA (2006). Semiparametric Theory and Missing Data Springer Series in Statistics. Springer, New York MR2233926

VAN DER LAAN MJ, POLLEY EC and HUBBARD AE (2007). Super learner. Stat. Appl. Genet. Mol. Biol 6 Art. 25, 23 MR2349918 10.2202/1544-6115.1309

VAN DER LAAN MJ and ROBINS JM (2003). Unified Methods for Censored Longitudinal Data and Causality. Springer, New York.

VAN DER LAAN MJ and ROSE S (2011). Targeted Learning: Causal Inference for Observational and Experimental Data Springer Series in Statistics. Springer, New York MR2867111 10.1007/978-1-4419-9782-1

VAN DER VAART A and VAN DER LAAN MJ (2006). Estimating a survival distribution with current status data and high-dimensional covariates. Int. J. Biostat 2 Art. 9, 42 MR2306498 10.2202/1557-4679.1014

VAN DER VAART AW and WELLNER JA (1996). Weak Convergence and Empirical Processes: With Applications to Statistics Springer Series in Statistics. Springer, New York MR1385671 10.1007/978-1-4757-2545-2

WESTLING T and CARONE M (2020). Supplement to “A unified study of nonparametric inference for monotone functions.” 10.1214/19-AOS1835SUPP.

WOODROOFE M and SUN J (1993). A penalized maximum likelihood estimate of f(0+) when f is nonincreasing. Statist. Sinica 3 501–515. MR1243398

WRIGHT FT (1981). The asymptotic behavior of monotone regression estimates. Ann. Statist 9 443–448. MR0606630

ZENG D (2004). Estimating marginal survival function by adjusting for dependent censoring using many covariates. Ann. Statist 32 1533–1555. MR2089133 10.1214/009053604000000508