Submitted to the Annals of Statistics

ASYMPTOTICS FOR CURRENT STATUS DATA UNDER VARYING OBSERVATION TIME SPARSITY (V10)

BY RUNLONG TANG[*], MOULINATH BANERJEE[*,†] AND MICHAEL R. KOSOROK

University of Michigan-Ann Arbor and University of North Carolina-Chapel Hill

In this paper, we study estimation and hypothesis testing for a failure time distribution function at a point in the current status model with observation times supported on a grid of potentially unknown sparsity and with multiple subjects sharing the same observation time. This is of interest since observation time ties occur frequently with current status data. The grid resolution is specified as $cn^{-\gamma}$ with $c > 0$ being a scaling constant and $\gamma > 0$ regulating the sparsity of the grid relative to the number of subjects (n). The asymptotic behavior falls into three cases depending on γ: regular 'normal–type' asymptotics obtain for $\gamma < 1/3$, non-standard cube-root asymptotics prevail when $\gamma > 1/3$ and $\gamma = 1/3$ serves as a boundary at which the transition happens. The limit distribution at the boundary is different from either of the previous cases and converges weakly to those obtained with $\gamma \in (0, 1/3)$ and $\gamma \in (1/3, \infty)$ as c goes to ∞ and 0, respectively. This weak convergence allows us to develop an adaptive procedure to construct confidence intervals for the value of the failure time distribution at a point of interest *without needing to know or estimate* γ, which is of enormous advantage from the perspective of inference. A simulation study of the adaptive procedure is presented.

1. Introduction. The current status model is one of the most well-studied survival models in statistics. An individual at risk for an event of interest is monitored at a random status or observation time, and an indicator of whether the event has occurred is recorded. An interesting feature of this kind of data is that the underlying survival function for the event time can only be estimated nonparametrically at the $n^{1/3}$ rate when the status time is a continuous random variable. Under mild conditions on the survival distribution, the limiting distribution of the estimator in this setting is the non-Gaussian Chernoff's distribution. This is in contrast to right-censored data where the underlying survival function can be estimated nonparametrically at rate $\sqrt{n}$ under right-censoring and is 'pathwise norm-differentiable' in the sense of van der Vaart (1991), admitting regular estimators and normal limits. Interestingly, when the status time distribution has finite support, the survival

[*]Supported in part by NSF grant DMS-0705288

[†]Supported in part by NSF grants DMS-0705288 and DMS-1007751

AMS 2000 subject classifications: Primary 62G09, 62G20; secondary 62G07

Keywords and phrases: Current Status Model, Discrete Covariate, Isotonic Regression, Adaptive Procedure

function estimator for current status data simplifies to a binomial random variable and is also $\sqrt{n}$ estimable and regular, with a normal limiting distribution. This intriguing change in limiting distribution under different degrees of sparsity of the status time distribution is the focus of this paper.

While the distinction between finite support and continuously distributed status times is clear philosophically, the practical distinction between these two settings is less clear. For example, suppose we observe a large number of ties in status times which are distributed across a large number of distinct times, as can happen when there is only one status time per day or only one status time per week but there are a moderately large number of days or weeks in the study; for example, a group of health care providers inspecting people in a community (at risk for infection, say) over a large number of days may only observe them at a specific time everyday, depending on their schedule, with multiple patients being inspected at the same time. How large should the number of distinct status times relative to n (the total number of indivduals inspected) be for the distribution of the estimates of the survival function to be closer to the Chernoff limit than the Gaussian? Under certain configurations, will the distribution be closer to a limiting distribution that is neither Gaussian nor Chernoff? These kinds of questions are the focus of this paper. Specifically, *we wish to assess the possible, relevant limiting distributions that can arise under different levels of sparsity of the status time distribution as well as develop a statistical procedure that is capable of adapting to whatever asymptotic regime is most suitable for the data*. This technically challenging problem is practically important since the most relevant asymptotic regime is seldom known in practice except for the very extreme settings where either there are only a few distinct status times (in which case a normal approximation should work well) or when there are no ties at all in the status times (the setting of a continuous status time).

The current status model, being the simplest avatar of interval-censoring — one observes the individual at a single time-point and ascertains whether they are infected or not by that time — and therefore best suited to an understanding of the more general phenomenon, has unsurprisingly received much attention in the statistical literature. The model itself goes back to Ayer *et al.* (1955) and was subsequently studied in Turnbull (1976) in a more general framework; asymptotic properties for the nonparametric maximum likelihood estimator (NPMLE) of the survival distribution were first obtained in Groeneboom and Wellner (1992) and involved techniques radically different from those in 'classical' survival analysis, and since then there has been a notable body of work on both the methodological and asymptotic fronts: see, for example, Huang (1996) where current status data under the Cox PH model is studied; Lin, Oakes and Ying (1998) and Shiboski (1998) for work on additive hazards regression and generalized additive models, respectively, for current status data; Sun (1999) for current status data under unequal censoring; Banerjee and Wellner (2001) and Banerjee and Wellner (2005) for a study of pointwise likelihood ratio tests for the survival distribution in the current status model that lead to asymptotically pivotal methods for inference in this model with broader implications for monotone function estimation; Jewell, van der Laan and Henneman (2003), Groeneboom, Maathuis and Wellner (2008a) and Groeneboom, Maathuis and Wellner (2008b) for current status model under competing risks with the last two papers providing a comprehensive description of the highly involved asymptotics that comes into play; and Ma and Kosorok (2005) for some recent semiparametric analysis involving partly linear transformation models with current status data. The above list is only a sample of the enormous body of work involving the current status model and is meant to reflect some of the main directions along which methodological and theoretical research have evolved over the last two decades, and also directions in which the authors have taken a more active interest, but should amply illustrate the wide range of problems that present themselves within the context of this relatively simple model. Interestingly though, the problem of determining the correct asymptotic approximation in current status data with ties and the development of an adaptive inference scheme, as discussed in the first two paragraphs, has not been satisfactorily resolved thus far in the rather large literature.

Denoting the survival distribution by F and with x_0 being a point of interest in the time domain, the goal is to determine how the behavior of the NPMLE of $F(x_0)$ depends on the relative magnitude of the number of distinct observation times, $K(n)$, to the sample size n. This change in the model is important not only because it more closely mirrors applications than the Lebesgue density assumption, as discussed above, but also because it induces surprisingly non-trivial complications in the limiting distribution and consequently in conducting inference. Since F, as will be shown later, is only identifiable at the distinct observation times, one can think about $K(n)$ as the effective number of parameters in the model. We will be investigating the behavior of $\hat{F}$ (the NPMLE of F) under a discrete observation time scheme where $K(n)$, the determinant of the sparsity of the observation time scheme, will be allowed to go to infinity with n at different rates. High rates correspond to dense observation time schemes and low rates to sparse ones. It should be noted here that some authors have indeed studied the current status model or closely related variants under discrete observation time settings. Yu *et al.* (1998) have studied the asymptotic properties of the NPMLE of F in the current status model with discrete observation times and more recently Maathuis and Hudgens (2010) have considered nonparametric inference for competing risks current status data under discrete or grouped observation times. However, these papers consider situations where the observation times are i.i.d. copies from a *fixed discrete distribution* (but not necessarily finitely supported) on the time-domain and are therefore not geared towards studying the effect of the trade-off between n and $K(n)$, i.e. *the effect of the relative sparsity of the number of distinct observation times to the size of the cohort of individuals* on inference for F. In both these papers, the (pointwise) estimates of F are asymptotically normal and $\sqrt{n}$ consistent; however, in situations, where the number of distinct observation times is large relative to the sample size, such normal approximations based on a fixed discrete observation time distribution might be suspect; see, for example, Section 5.1 of Maathuis and Hudgens (2010) where this is illustrated via simulations. We will return to the paper Maathuis and Hudgens (2010) later in our concluding discussion. Zhang, Kim and Woodroofe (2001) consider isotonic regression with grouped data in the monotone density estimation problem. The asymptotic properties of the isotonic estimator are shown to depend on the order of magnitude of the grouping intervals (which are allowed to depend on the sample size) and the corresponding limit distributions determined but no effective inference schemes are developed. More recently, Wang and Shen (2010) have considered monotone regression estimators based on grouped data using B-splines and have studied their asymptotic properties in terms of the number of knots as the sample size increases.

The rest of the paper is organized as follows. In Section 2, we introduce our mathematical formulation of the problem and present an overview of the basic results at a high level. Section 3 discusses the characterizations of the estimators of interest while Section 4 presents the main asymptotic results and major proofs. Section 5 addresses the important question of *adaptive inference* in the current status model: given a time-domain $[a, b]$ and current status data observed at times on a regular grid on $[a, b]$ of an *unknown level of sparsity*, how do we make inference on F? What asymptotic approximations should we use? We provide an answer to this question that *circumvents the need to determine the sparsity of the grid*, and therefore provides a tremendous advantage from the point of view of inference. Section 6 provides simulation results for a practical version of the adaptive inference procedure developed in Section 5. Section 7 concludes with a discussion of the findings of this paper and their implications for monotone regression models in general, as well as more complex forms of interval censoring and interval censoring with competing risks.

2. Formulation of the Problem and Overview of Findings. Let $\{T_{i,n}\}_{i=1}^{n}$ be i.i.d. random survival times following some unknown distribution F with Lebesgue density f concentrated on the time-domain $[a', b']$ with $0 \le a' < b' < \infty$ (or supported on $[a', \infty)$ if no such b' exists) and $\{X_{i,n}\}$ are i.i.d. observation times drawn from a discrete probability measure H_n supported on a regular grid on

$[a, b]$ with $a' \leq a < b < b'$. [1] Also, $T_{i,n}$ and $X_{i,n}$ are assumed to be independent for each i. However, $\{T_{i,n}\}$ are not observed; rather, we observe $\{Y_{i,n} = 1\{T_{i,n} \leq X_{i,n}\}\}$. This puts us in the setting of a binary regression model with $Y_{i,n}|X_{i,n} \sim \text{Bernoulli}(F(X_{i,n}))$. We denote the support of H_n by $\{t_{i,n}\}_{i=1}^K$ where the i-th grid point $t_{i,n} = a + i\delta$, the unit spacing $\delta = \delta(n) = c\,n^{-\gamma}$ (also referred to as the *grid resolution*) with $\gamma \in (0, 1)$ and $c > 0$, and the number of grid points $K = K(n) = \lfloor (b - a)/\delta \rfloor$. On this grid, the distribution H_n is viewed as a discretization of an absolutely continuous distribution G, whose support contains $[a, b]$ and whose Lebesgue density is denoted as g. More specifically, $H_n\{t_{i,n}\} = G(t_{i,n}) - G(t_{i-1,n})$, for $i = 2, 3, \ldots, K - 1$, $H_n\{t_{1,n}\} = G(t_{1,n})$ and $H_n\{t_{K,n}\} = 1 - G(t_{K-1,n})$. For simplicity, these discrete probabilities are denoted as $p_{i,n} = H_n\{t_{i,n}\}$ for $i = 1, 2, \ldots, K$. Let $x_0 \in (a, b)$ be a point around which we are interested in determining the properties of the NPMLE of F as a function of the grid resolution δ or equivalently in terms of (c, γ). In what follows, we refer to the pair $(X_{i,n}, Y_{i,n})$ as (X_i, Y_i), suppressing the dependence on n, but the *triangular array nature* of our observed data should be kept in mind. Similarly, the subscript n is suppressed elsewhere when no confusion will be caused.

Given the point of interest $x_0 \in (a, b)$, define $t_l = t_{l,n}$ to be the largest grid-point less than or equal to x_0 and $t_r = t_{r,n} = t_{l+1,n}$ the smallest grid point larger than x_0. As will be seen, the NPMLE $\hat{F}$ is only identifiable up to its values at the grid-points. To define it on the interval $[a, b]$ one usually resorts to some convention of extension. Our estimate $\hat{F}$ will be taken to be the unique right-continuous step function with potential jumps only at the grid-points that maximizes the likelihood function of the observed data over all piece-wise constant right-continuous distribution functions supported on $[a, b]$.

We start with the estimation of $F(t_l)$ using $\hat{F}(t_l)$. The key features of the effect of sparsity on the asymptotics are best illustrated by focusing on t_l (which can be viewed as a natural surrogate for x_0). We will demonstrate, explicitly, that resolving the inference problem for $F(t_l)$ allows us to resolve the inference problem for $F(x_0)$ with only minor modifications. The reason behind using t_l instead of x_0 primarily stems from the fact that the derivations can be presented in a much cleaner form; to deal with $F(x_0)$ one is forced to resort to a combination of $\hat{F}(t_l)$ and $\hat{F}(t_r)$ at least for sparser grids ($\gamma \leq 1/3$) and while this in itself does not require any significant technical innovation, the derivations tend to become somewhat more cumbersome and tedious. We focus our attention on three key questions: (a) What is the limit distribution of $\hat{F}(t_l) - F(t_l)$ under appropriate normalization? (b) What can

[1] The regularity of the grid corresponds to evenly spaced inspection times which is satisfied in many clinical and engineering applications (patients inspected daily at a clinic or machines inspected routinely every week, for example). It also makes the subsequent derivations simpler to present without compromising the complexity of the intrinsic mathematical issues involved.

we say about the asymptotic properties of the likelihood ratio test statistic (LRS) for testing the hypotheses $H_{0,n} : F(t_l) = \theta_l \leftrightarrow H_{1,n} : F(t_l) \neq \theta_l$ when $H_{0,n}$ holds? (c) How can the methods for making inference on $F(t_l)$ be modified to make inference for $F(x_0)$?

The NPMLE of F, say $\hat{F}$, is characterized by the vector

$$\{\hat{F}(t_i)\}_{i=1}^{K} = \operatorname*{argmin}_{u_1 \leq u_2 \leq \ldots u_K} \sum_{j=1}^{K} (\bar{Z}_i - u_i)^2 N_i,$$

where N_i is the number of X_j's that equal t_i and $\bar{Z}_i$ is the sample mean of all the Y_j's such that $X_j = t_i$, or in the other words the sample proportion of patients inspected at time t_i that are infected by that time. Some mathematics shows that for $\gamma \in (0, 1/3)$, with probability increasing to 1, the $\bar{Z}_i$'s are ordered in i and therefore furnish the minimizer of the optimization problem in the above display. Basically, for $\gamma \in (0, 1/3)$, the grid is sparse enough so that the naive averages at each inspection time, which provide empirical estimates of F at those corresponding inspection times, are automatically ordered and there is no 'strength borrowed' from nearby inspection times. It will be shown that $n^{(1-\gamma)/2}\,(\hat{F}(t_l) - F(t_l))$ converges to a normal distribution. Note that the rate of convergence is always faster than $n^{1/3}$ and we are in the setting of 'standard' asymptotics.

On the contrary, when $\gamma \in (1/3, 1)$, the customary 'non–standard' asymptotics that are prevalent in the usual treatment of the current status model (as in Groeneboom and Wellner (1992) and Banerjee and Wellner (2001)) take over. Now, successive grid points are close by each other, the naive averages are no longer ordered and isotonization algorithms kick in, producing boundary solutions that are highly non-linear functionals of the empirical distribution and the asymptotic distribution turns out to be drastically different: $n^{1/3}(\hat{F}(t_l) - F(t_l))$ converges to a multiple, which depends among other parameters upon c, of Chernoff's distribution, or equivalently a multiple of the slope at 0 of the greatest convex minorant of the stochastic process $\{X(t) = W(t) + t^2 : t \in \mathbb{R}\}$ with $W(t)$ being two sided Brownian motion starting from 0.

The case $\gamma = 1/3$ acts as the boundary scenario at which the transition between standard and non-standard asymptotics happens and interestingly enough, the behavior of $\hat{F}(t_l)$ in this case is different from either of the previous cases. Basically, when $\gamma = 1/3$, the grid points are 'close enough', so that the naive averages are no longer the best estimates of F. On the other hand, the resolution of the grid exactly matches $n^{-1/3}$, the order of the local neighborhoods of the point t_l that are asymptotically relevant to the estimation of F whenever $\gamma \in [1/3, 1)$. As a consequence, the limit distribution of $n^{1/3}(\hat{F}(t_l) - F(t_l))$ is driven, not by a process in continuous time as in the case when $\gamma \in (1/3, 1)$, but a process essentially defined on $c\mathbb{Z}$, where $\mathbb{Z}$ is the set of integers. In fact, the relevant limit process can be written down as the linear interpolant of the set of points $\mathcal{P}_c = \{ck, \alpha W(ck) + \beta c^2 k(1+k) : k \in \mathbb{Z}\}$, where the constants α and β depend on the underlying parameters of the problem, and the limit distribution is that of the left derivative at 0 of the greatest convex minorant of this process, denoted subsequently by $\mathcal{S}_c$. The process $\mathcal{P}_c$ can be viewed as a discretized version of the process X: note that for $\gamma = 1/3$, instead of Brownian motion the random component of the relevant process is the restriction of Brownian motion to the set $c\mathbb{Z}$ and the quadratic drift term t^2, essentially the integral of the function t, is replaced by the sum of the first k integers. As far as the asymptotics of LRS for testing $H_{0,n} : F(t_l) = \theta_l$ are concerned, for the case $\gamma \in (0, 1/3)$ we get the usual χ_1^2 distribution, for $\gamma \in (1/3, 1)$ we obtain the pivotal limit $\mathbb{D}$ of Banerjee and Wellner (2001), the same as when the covariates come from a Lebesgue density, and for $\gamma = 1/3$ we obtain a discrete analogue of $\mathbb{D}$ which can be written as a functional of $\mathcal{P}_c$, though this is no longer pivotal.

In Section 5, we address the rather interesting question of how to make inferences on F from current status data observed on a regular grid in the time-domain by using the results developed in this paper. We argue that irrespective of the inherent resolution parameter γ of the given grid, which is usually unknown in practice, one can always assume γ to be $1/3$, the boundary value, at the cost of adjusting c. Basically, depending on the true inherent grid resolution, an adjusted data-driven c, say $\hat{c}$, is shown to provide the correct asymptotic quantiles — namely those of $\mathcal{S}_{\hat{c}}$ — for the distribution of $n^{1/3}(\hat{F}(t_l) - F(t_l))$. Thus, the 'boundary asymptotics' can be used to approximate both the 'standard' ($\gamma \in (0, 1/3)$) and 'non-standard' ($\gamma \in (1/3, 1)$) situations and provide an effective means of setting confidence intervals for $F(t_l)$ without having to contend with the difficult problem of estimating the inherent resolution of the grid. Similar phenomena are observed with the likelihood ratio statistics. From the point of view of conducting effective inference, we view this as the *key contribution* of our work.

3. Estimation.

In this section, we consider the shape-restricted nonparametric maximum likelihood estimator (NPMLE) of F and the likelihood ratio test statistic (LRS) for testing the value of F at a point of interest. The characterizations of these estimators are well-known from the current status literature but we include a description tailored for the setting of this paper.

The likelihood function of the data $\{(X_i, Y_i)\}$ is given by

$$L_n(F) = \prod_{j=1}^{n} F(X_j)^{Y_j}(1 - F(X_j))^{1-Y_j} p_{\{i: X_j = t_i\}} = \prod_{i=1}^{K} F_i^{Z_i}(1 - F_i)^{N_i - Z_i} p_i^{N_i},$$

where $p_{\{i: X_j = t_i\}}$ denotes the probability that X_j equals a genetic grid point t_i, F_i is an abbreviation for $F(t_i)$, $N_i = \sum_{j=1}^{n}\{X_j = t_i\}$ is the number of the observation at t_i, $Z_i = \sum_{j=1}^{n} Y_j\{X_j = t_i\}$ is the sum of the responses at t_i, $\{\cdot\}$ stands for both a set and its indicator function with the meaning depending on the context of use, and F is generically understood as either a distribution or the vector $(F_1, F_2, \ldots, F_K)$, which sometimes is also written as $\{F_i\}_{i=1}^{K}$. Then, the log-likelihood function $\log(L_n(F))$ is given by

$$l_n(F) = \sum_{i=1}^{K} N_i \log p_i + \sum_{i=1}^{K} \left\{[\bar{Z}_i \log F_i + (1-\bar{Z}_i)\log(1-F_i)]N_i\right\},$$

where $\bar{Z}_i = Z_i/N_i$ is the average of the responses at t_i.

Denote the basic shape-restricted maximizer as

$$\{F_i^\star\}_{i=1}^{K} = \underset{F_1 \le \cdots \le F_K}{\operatorname{argmax}}\; l_n(F).$$

From the theory of isotonic regression (see, for example, Robertson, Wright and Dykstra (1988)), we have

$$\underset{F_1 \le \cdots \le F_K}{\operatorname{argmax}}\; l_n(F) = \underset{F_1 \le \cdots \le F_K}{\operatorname{argmin}} \sum_{i=1}^{K} [(\bar{Z}_i - F_i)^2 N_i].$$

Thus, $\{F_i^\star\}_{i=1}^{K}$ is the weighted isotonic regression of $\{\bar{Z}_i\}_{i=1}^{K}$ with weights $\{N_i\}_{i=1}^{K}$, and exists uniquely. We conventionally define the shape-restricted NPMLE of F as the following right-continuous step function on $[a, b]$:

$$\hat{F}(t) = \begin{cases} 0 & \text{if } t \in [a, t_1); \\ F_i^\star, & \text{if } t \in [t_i, t_{i+1}), i = 1, \cdots, K-1; \\ F_K^\star, & \text{if } t \in [t_K, b]. \end{cases} \tag{3.1}$$

A popular and straightforword method of setting confidence intervals for F at a point of interest $t \in (a, b)$ is first to derive the asymptotic distribution of $\hat{F}(t)$ and then to construct for $F(t)$ the so-called Wald-type confidence intervals of the form $\hat{F}(t)$ plus and minus terms that depend on both the level of confidence and the asymptotic distribution. Another popular but slightly more involved approach to construct confidence intervals is through the inversion of a likelihood ratio test for the value of $F(t)$. More specifically, we first derive the asymptotic null distribution of the likelihood ratio test statistic (LRS) for the following hypothesis testing problem:

$$H_0 : F(t) = \theta \leftrightarrow H_1 : F(t) \neq \theta, \tag{3.2}$$

where $t \in (a, b)$ and $\theta \in (0, 1)$ and then obtain the so-called LR-type confidence intervals via inversion. Note that both t and θ can depend on n. One specific instance of interest in this paper is that $t = t_l$ and $\theta = \theta_l = F(t_l)$. In this case, the hypotheses (3.2) become

$$H_{0,n} : F(t_l) = \theta_l \leftrightarrow H_{1,n} : F(t_l) \neq \theta_l. \tag{3.3}$$

To construct the LRS, we next consider the constrained shape-restricted NPMLE of F under the null hypothesis of (3.3). Define

$$\{F_i^{\star o}\} = \underset{F_1 \leq \cdots \leq F_l = \theta_l \leq F_{l+1} \leq \cdots \leq F_K}{\operatorname{argmax}} l_n(F).$$

It is well known that $\{F_i^{\star o}\}$ is well defined and

$$\{F_i^{\star o}\}_{i=1}^{l-1} = \theta_l \wedge \underset{F_1 \leq \cdots \leq F_{l-1}}{\operatorname{argmin}} \sum_{i=1}^{l-1} [(\bar{Z}_i - F_i)^2 N_i],$$
$$\{F_i^{\star o}\}_{i=r}^{K} = \theta_l \vee \underset{F_r \leq \cdots \leq F_K}{\operatorname{argmin}} \sum_{i=r}^{K} [(\bar{Z}_i - F_i)^2 N_i],$$

where the minimum and maximum operators ($\wedge$ and $\vee$) are interpreted as being taken component-wise. Specifically, $F_l^{\star o} = \theta_l$ as required by the null hypothesis. Thus, the constrained NPMLE of F, similar to the unconstrained one, can be defined as the following right-continuous step function on $[a, b]$:

$$\hat{F}^o(t) = \begin{cases} 0 & t \in [a, t_1); \\ F_i^{\star o}, & t \in [t_i, t_{i+1}), i = 1, \cdots, K-1; \\ F_K^{\star 0}, & t \in [t_K, b]. \end{cases} \tag{3.4}$$

For the details underlying the above characterization of the constrained estimator of F, we refer the readers to Banerjee (2000) and Banerjee and Wellner (2001). Thus, the LRS in given by

$$2 \log \lambda_n = 2[l_n(\hat{F}) - l_n(\hat{F}^o)]. \tag{3.5}$$

An asymptotic $1 - \eta$ confidence interval for θ_l is given by the set of all $0 < \theta < 1$ such that $2 \log \lambda_n(\theta)$, the LRS for testing $H_{0,n} : F(t_l) = \theta$ versus its complement, lies below the $(1 - \eta)$'th quantile of the limit distribution of the LRS under $H_{0,n} : F(t_l) = \theta_l$.

Next, we provide characterizations of $\hat{F}$ and $\hat{F}^o$ as slopes of *the greatest convex minorants* (GCMs) of random processes, which prove useful for deriving the asymptotics for $\gamma \in [1/3, 1)$. First, consider the characterization of $\hat{F}$. Define, for $t \in [a, b]$,

$$G_n(t) = \mathbb{P}_n\{x \leq t\}, \quad V_n(t) = \mathbb{P}_n y\{x \leq t\}, \tag{3.6}$$

where $\mathbb{P}_n$ is the empirical probability measure based on the data $\{(X_i, Y_i)\}$. Then, we have, for each $x \in [a, b]$,

$$\hat{F}(x) = LS \circ GCM \{(G_n(t), V_n(t)), t \in [a, b]\} (G_n(x)) . \tag{3.7}$$

Here, $GCM(\cdot)$ denotes the greatest convex minorant of a set of points in $\mathbb{R}^2$. For any finite collection of points in $\mathbb{R}^2$, its GCM is a continuous piecewise linear convex function and $LS(\cdot)$ denotes the *left slope or derivative function* of a convex function. The term GCM will also be generically used in connection with functions from $\mathbb{R} \mapsto \mathbb{R}$. For such a function H, $GCM(H)$ will denote the greatest convex minorant of H.

Next, consider the characterization of $\hat{F}^o$. Define, for $s \in [a - t_l, b - t_l]$,

$$\tilde{G}_n(s) = G_n(t_l + s) - G_n(t_l), \quad \tilde{V}_n(s) = V_n(t_l + s) - V_n(t_l). \tag{3.8}$$

Then, we have, for $s \in [a - t_l, 0)$,

$$\hat{F}^o_l(t_l + s) = \theta_l \wedge LS \circ GCM \left\{(\tilde{G}_n(u), \tilde{V}_n(u)), u \in [a - t_l, 0)\right\} (\tilde{G}_n(s)) \tag{3.9}$$

and for $s \in [t_r - t_l, b - t_l)$,

$$\hat{F}^o_r(t_l + s) = \theta_l \vee LS \circ GCM \left\{(\tilde{G}_n(u), \tilde{V}_n(u)), u \in [0, b - t_l]\right\} (\tilde{G}_n(s)). \tag{3.10}$$

Therefore, we have, for $s \in [a - t_l, b - t_l]$,

$$\begin{aligned} \hat{F}^o(t_l + s) = {} & \hat{F}^o_l(t_l + s)\{s \in [a - t_l, 0)\} \\ & + \hat{F}^o_r(t_l + s)\{s \in [t_r - t_l, b - t_l]\} + \theta_l\{s \in [0, t_r - t_l)\}. \end{aligned} \tag{3.11}$$

The above characterizations will be further exploited in the next section.

4. Asymptotic Results.

In this section, we primarily study the asymptotics of $\hat{F}$ and $2\log(\lambda_n)$ for three cases with different values of $\gamma \in (0, 1)$. We first consider Case One $\gamma \in (0, 1/3)$, then Case Two $\gamma \in (1/3, 1)$ and finally Case Three $\gamma = 1/3$.

4.1. *Case One* $\gamma \in (0, 1/3)$.

In this subsection, we consider the asymptotics for $\gamma \in (0, 1/3)$. First, we state further technical assumptions :

(A1.1) There exists a constant $f_l > 0$ such that $f(x) > f_l$ for every $x \in [a, b]$.
(A1.2) There exists a constant $g_l > 0$ such that $g(x) \geq g_l$ for every $x \in [a, b]$.
(A1.3) Assume $a' < a$ and $F(a) > 0$.

We denote the above assumptions together as (**A1**). The assumptions (**A1.1**) and (**A1.2**) require that both f and g have lower bounds 'globally'. The assumption (**A1.3**) is technically tailored to establish Proposition 4.9, the key technical tool behind the arguments in this subsection. In fact, the strict assumption (**A1.3**) can be replaced by a weaker one which allows $a' = a$ but requires that there exists $d \in (a, x_0)$ such that $F(d) > 0$. With this weaker assumption, Proposition 4.9 needs some technical adjustments but all the theorems in this subsection still hold without any modification. To emphasize the main idea of the argument, without having to deal with the aforementioned adjustments, we retain the strong assumption (**A1.3**). By (**A1.2**), each discrete probability $p_i \geq g_l\delta$. Denote $m_l = n g_l \delta = g_l c n^{1-\gamma}$, which can be interpreted as the minimum average number of observations at a grid point.
The following lemma shows the consistency of $F_l^\star$, $F_r^\star$ and $\hat{F}(x_0)$.

LEMMA 4.1. *If $\gamma \in (0, 1/3)$ and (**A1**) holds, we have*

$$F_l^\star - F(t_l) \xrightarrow{P} 0, \quad F_r^\star - F(t_r) \xrightarrow{P} 0, \quad \textit{and} \quad \hat{F}(x_0) \xrightarrow{P} F(x_0).$$

The joint limiting distribution of $\hat{F}(t_l)$ and $\hat{F}(t_r)$ is described below.

THEOREM 4.2. *If $\gamma \in (0, 1/3)$ and (**A1**) holds, we have*

$$\left(\sqrt{N_l}(\hat{F}(t_l) - F(t_l)), \sqrt{N_r}(\hat{F}(t_r) - F(t_r))\right) \xrightarrow{d} \sqrt{F(x_0)(1 - F(x_0))} N(0, I_2),$$

where I_2 is the 2×2 identity matrix.

REMARK 4.3. From Theorem 4.2, the quantities $\hat{F}(t_l)$ and $\hat{F}(t_r)$ with proper centering and scaling are asymptotically uncorrelated and independent. In fact, they are essentially the averages of the responses at the two grid points t_l and t_r and are therefore based on responses corresponding to different sets of individuals. Consequently, there is no dependence between them in the long run. Intuitively speaking, $\gamma \in (0, 1/3)$ corresponds to very sparse grids with successive grid points *far enough* so that the responses at different grid points fail to influence each other.

Note that $N_l/(np_l)$ converges to 1 in probability and that $np_l/cg(x_0)n^{1-\gamma}$ converges to 1for $\gamma \in (0, 1/3)$. Then the result of Theorem 4.2 can be rewritten as follows:

$$\left(n^{(1-\gamma)/2}(\hat{F}(t_l) - F(t_l)), n^{(1-\gamma)/2}(\hat{F}(t_r) - F(t_r))\right) \xrightarrow{d} \alpha c^{-\frac{1}{2}} N(0, I_2), \tag{4.1}$$

where $\alpha = \sqrt{F(x_0)\,(1 - F(x_0))/g(x_0)}$. This formulation will be used later and the parameter α will be seen to play a critical role in the asymptotic behavior of $\hat{F}(t_l)$ when $\gamma \in [1/3, 1)$ as well.

Next, we consider the asymptotics of the log-likelihood ratio test statistic (LRS) $2\log\lambda_n$ for testing the hypotheses (3.3).

THEOREM 4.4. *If $\gamma \in (0, 1/3)$ and **(A1)** holds, under the null hypothesis of (3.3), we have $2\log\lambda_n \xrightarrow{d} \chi^2_1$.*

From Theorem 4.2 and Theorem 4.4, we see that the asymptotic distributions are Normal and Chi-Squared, standard limit distributions for parametric problems.

In the following, we discuss testing and estimating $F(x_0)$. Denote $2\log\lambda'_n$ as the LRS based on $\hat{F}$ and $\hat{F}^o$ for the hypotheses (3.2) with $t = x_0$ and $\theta = \theta_0$. Then, this test statistic may not have a proper limit distribution, as shown by the following theorem.

THEOREM 4.5. *Suppose $\gamma \in (0, 1/3)$ and **(A1)** holds. If there exists $\epsilon > 0$ and a subsequence $\{n_k\}$ of $\{n\}$ such that $\min(x_0 - t_l, t_r - x_0)/\delta > \epsilon$, then under the null hypothesis of (3.2) with $t = x_0$ and $\theta = \theta_0$, we have $2\log\lambda'_{n_k} \xrightarrow{P} 0$ as $k \to \infty$.*

An example satisfying the conditions of Theorem 4.5 is as follows. Suppose $a = 0$, $x_0 = 1/2$, $c = 1$, $\gamma = 1/4$ and $n_k = (2k+1)^4$ for $k \in \mathbb{N}$. Then $(x_0 - t_l)/\delta = (t_r - x_0)/\delta = 1/2 > 0$. In fact, this example also concretely explains why we do not make inference about $F(x_0)$ using $\hat{F}(x_0)$. The basic reason is that there may not exist any scaling such that the scaled difference $\hat{F}(x_0) - F(x_0)$ converges to a tight non-degenerate random limit. More specifically, assume the setting of the above example and further suppose $F(t) = t$ for $t \in [0, 1]$. Then, we have $\hat{F}(x_0) - F(x_0) = (\hat{F}(t_l) - t_l) + (t_l - x_0)$. By (4.1), the proper scaler for the first term $\hat{F}(t_l) - t_l$ is $n^{3/8}$ and the limit distribution is normal. However, this scaler times the second term $(x_0 - t_l)$ gives $\sqrt{2k-1}/2$ (along $\{n_k\}$), which diverges to infinity as k goes to infinity.

Inference for $F(x_0)$: It turns out, however, that inference for $F(x_0)$ can still be made provided the estimator $\hat{F}$ is altered slightly. Recall that $\hat{F}$ is defined as the right-continuous step function that assumes the value $F_i^\star$ at t_i. Consider now the piecewise linear interpolant of the $F_i^\star$'s. More precisely, the modified estimator is given by

$$\tilde{F}(t) = \begin{cases} 0, & \text{if } t \in [a, t_1) \\ \frac{t_{i+1}-t}{t_{i+1}-t_i} F_i^\star + \frac{t-t_i}{t_{i+1}-t_i} F_{i+1}^\star, & \text{if } t \in [t_i, t_{i+1}) \text{ and } i = 1, 2, \cdots, K-1; \\ F_K^\star, & \text{if } t \in [t_K, b]. \end{cases}$$

Then, we can estimate $F(x_0)$ by $\tilde{F}(x_0)$, whose consistency is ensured by the following lemma.

LEMMA 4.6. *If $\gamma \in (0, 1/3)$ and **(A1)** holds, we have $\tilde{F}(x_0) \xrightarrow{P} F(x_0)$.*

Define two proportions $p_n = (x_0 - t_l)/(t_r - t_l)$ and $q_n = 1 - p_n$ and a properly normalized random quantity

$$R_{\tilde{F}} = \frac{1}{\sqrt{p_n^2 + q_n^2}} \sqrt{\frac{N_l + N_r}{2}} (\tilde{F}(x_0) - F(x_0)).$$

The asymptotic distribution of $\tilde{F}(x_0)$ is the content of the next theorem.

THEOREM 4.7. *If $\gamma \in (0, 1/3)$, **(A1)** holds, and f'' is bounded in a neighborhood of x_0, we have, for $\gamma \in (1/5, 1/3)$,*

$$R_{\tilde{F}} \xrightarrow{d} \sqrt{F(x_0)(1 - F(x_0))} Z.$$

Further, if $\{n_k\}$ is a subsequence of $\{n\}$ such that p_{n_k} converges to some $p \in (0, 1)$, we have, along this subsequence and for $\gamma = 1/5$,

$$R_{\tilde{F}} \xrightarrow{d} \sqrt{F(x_0)(1 - F(x_0))} Z + \frac{1}{2} \frac{pq}{\sqrt{p^2 + q^2}} g(x_0)^{1/2} c^{5/2} f'(x_0).$$

Furthermore, if $f'(x_0) \neq 0$, we have, for $\gamma \in (0, 1/5)$,

$$R_{\tilde{F}} \xrightarrow{P} \mathit{Sign}(f'(x_0)) \infty,$$

where $q = 1 - p$ and Z follows $N(0, 1)$.

REMARK 4.8. Since N_l/N_r converges in probability to 1, $R_{\tilde{F}}$ is asymptotically equivalent to $(p_n^2 + q_n^2)^{-1/2} \sqrt{N_i} (\tilde{F}(x_0) - F(x_0))$ with $i = l$ or r. Note that the coefficient $(p_n^2 + q_n^2)^{-1/2}$ lies between 1 and $\sqrt{2}$. Comparing the asymptotic results of Theorem 4.7 with $\gamma \in (1/5, 1/3)$ and that of Theorem 4.2 reveals that $\tilde{F}(x_0)$ effectively combines the data around x_0 to successfully achieve smaller asymptotic variance than $\hat{F}(t_l)$ and $\hat{F}(t_r)$.

4.1.1. *Proofs.* The proofs of the above lemmas and theorems rely heavily on some fundamental propositions that we state below and prove in the appendix. In what follows we will need to consider the vector of averages of responses over the grid-points: $\{\overline{Z}_i\}_{i=1}^k$. Since $\overline{Z}_i$ is not defined when $N_i = 0$, to avoid ambiguity we set $\overline{Z}_i = 0$ whenever this happens. This can be done without affecting the asymptotic results, since the probability of the event $\{N_i > 0, i = 1, 2, \ldots, K\}$ goes to 1. See, for example, Lemma A.1 in the appendix where a stronger fact is established.

PROPOSITION 4.9. *If $\gamma \in (0, 1/3)$ and **(A1)** holds, we have*

$$P\left(\bar{Z}_1 \leq \bar{Z}_2 \leq \cdots \leq \bar{Z}_K\right) \to 1.$$

The above proposition says that on a sequence of sets with probability increasing to 1, the vector $\{\overline{Z}_i\}_{i=1}^k$ is ordered and therefore the isotonization algorithm involved in finding the NPMLE of F yields $\{F_i^\star\}_{i=1}^K = \{\overline{Z}_i\}_{i=1}^K$ on these sets. In other words, asymptotically, isotonization has no effect and the naive estimates obtained by averaging the responses at each grid point produce the NPMLE. This lemma is really at the heart of the asymptotic derivations for $\gamma < 1/3$ because it effectively reduces the problem of studying the $F_i^\star$'s which are obtained through a complex non-linear algorithm to the study of the asymptotics of the $\overline{Z}_i$, which are linear statistics and can be handled readily using standard central limit theory. For the following lemmas, we use weaker assumptions (than **(A1)** on f and g):

(A1.1') $f(\cdot)$ has a positive lower bound in a neighborhood of x_0.
(A1.2') $g(\cdot)$ has a positive lower bound in a neighborhood of x_0.

We refer to these two assumptions together as **(A1')**.

PROPOSITION 4.10. *If $\gamma \in (0, 1/3)$ and **(A1')** hold, we have*

$$P\left(\bar{Z}_{l-1} \leq F_l \leq \bar{Z}_r\right) \to 1.$$

Now we state a further weaker assumption on F.

(A1.1") $F(\cdot)$ is positive and continuous in a neighborhood of x_0.

PROPOSITION 4.11. *If $\gamma \in (0, 1)$ and **(A1.1")** and **(A1.2')** hold, we have*

$$\left(\bar{Z}_l - F_l, \bar{Z}_r - F_r\right) \xrightarrow{P} 0,$$

$$\left(\sqrt{N_l}(\bar{Z}_l - F_l), \sqrt{N_r}(\bar{Z}_r - F_r)\right) \xrightarrow{d} \sqrt{F(x_0)(1 - F(x_0))}N(0, I_2).$$

PROOF OF LEMMA 4.1. We only show $\hat{F}(t_l) - F(t_l) \to 0$ in probability. The other two can be shown similarly. We have

$$\hat{F}(t_l) - F(t_l) = (F_l^\star - F_l)B_n + (F_l^\star - F_l)(1 - B_n) =: T_1 + T_2,$$

where $B_n = \{\bar{Z}_1 \leq \cdots \leq \bar{Z}_K\}$ and '=:' means the right-hand side notations are defined by the left-hand terms. By Proposition 4.9, B_n converges to 1 in probability. Thus, T_2 converges to 0 in probability. On the other hand, by Proposition 4.11, $\bar{Z}_l - F_l$ converges to 0 in probability. On B_n, the $\overline{Z}_i$'s are already monotone, and the vector $\{F_i^\star\}$ obtained via isotonization coincides with $\{\overline{Z}_i\}$. We conclude that $T_1 = (\bar{Z}_l - F_l)\, B_n \xrightarrow{P} 0 \cdot 1 = 0$. So, $\hat{F}(x_0)$ is a consistent estimator of $F(x_0)$. $\square$

PROOF OF THEOREM 4.2. Denote $T_{nl} = \sqrt{N_l}(\hat{F}(t_l) - F(t_l))$ and $T_{nr} = \sqrt{N_r}(\hat{F}(t_r) - F(t_r))$. We then have $(T_{nl}, T_{nr}) = S_1 + S_2$, where S_1 and S_2 are $\left(\sqrt{N_l}(\bar{Z}_l - F_l), \sqrt{N_r}(\bar{Z}_r - F_r)\right) B_n$ and $(T_{nl}, T_{nr})(1 - B_n)$, respectively, with B_n as in the preceding proof. By Proposition 4.9, B_n converges to 1 in probability. Then, S_2 converges to 0 in probability. On the other hand, by Proposition 4.11 and Slutsky's Lemma, we have $S_1 \xrightarrow{d} \sqrt{F(x_0)(1 - F(x_0))} N(0, I_2) \cdot 1$. □

PROOF OF THEOREM 4.4. We have

$$2 \log \lambda_n = (2 \log \lambda_n) C_n + (2 \log \lambda_n)(1 - C_n),$$

where $C_n = B_n\{\bar{Z}_{l-1} \le \theta_l \le \bar{Z}_r\}$ with B_n as in the preceding proof. By Proposition 4.9 and Proposition 4.10, C_n converges to 1 in probability under the null hypothesis. Thus, the second term above converges to 0 in probability. Denote the first term above as T_1. Then, we have

$$T_1 = 2\left\{\left[\bar{Z}_l \log\left(\frac{\bar{Z}_l}{\theta_l}\right) + (1 - \bar{Z}_l) \log\left(\frac{1 - \bar{Z}_l}{1 - \theta_l}\right)\right] N_l\right\} C_n =: T_{11} C_n.$$

By routine Taylor expansions of $\log(\bar{Z}_l)$ around θ_l and $\log(1 - \bar{Z}_l)$ around $1 - \theta_l$ up to a third order term and some straightforward algebra, we have

$$T_{11} = \frac{N_l(\bar{Z}_l - \theta_l)^2}{\theta_l\,(1 - \theta_l)} + o_p(1).$$

Note that $\theta_l \to F(x_0)$. Finally, by Proposition 4.11, we have $T_{11} \xrightarrow{d} 1 \cdot \chi_1^2 + 0 = \chi_1^2$, which completes the proof. □

PROOF OF THEOREM 4.5. We have $2 \log \lambda'_{n_k} = 2 \log \lambda'_{n_k} A_{n_k}$ where $A_n = 1 - \{\bar{Z}_1 \le \cdots \le \bar{Z}_K\}\{\bar{Z}_l \le \theta_0 \le \bar{Z}_r\}$, since the likelihood ratio is identically 1 whenever $A_n = 0$. Now, $P(\bar{Z}_l \le \theta_0 \le \bar{Z}_r) \to 1$ under the null hypothesis and the conditions of the theorem. To see this, consider $P(\bar{Z}_l \le \theta_0) = P(n^{(1-\gamma)/2}(\bar{Z}_l - F(t_l)) \le n^{(1-\gamma)/2}(F(x_0) - F(t_l)))$. Now, $n^{(1-\gamma)/2}(\bar{Z}_l - F(t_l))$ is $O_p(1)$ and $n^{(1-\gamma)/2}(F(x_0) - F(t_l))$ is easily seen to be larger than a constant times $n^{(1-\gamma)/2}\,\delta$ along the subsequence $\{n_k\}$. Since δ is of order $n^{-\gamma}$ and $1 - 3\gamma > 0$, it follows that $n^{(1-\gamma)/2}(F(x_0) - F(t_l))$ diverges to ∞ along $\{n_k\}$. So $P(\bar{Z}_l \le \theta_0) \to 1$ along $\{n_k\}$ and a similar argument shows that $P(\theta_0 \le \bar{Z}_r) \to 1$.
This, together with Proposition 4.9, implies that A_{n_k} converges to 0 in probability under the null hypothesis. Therefore, we have $2 \log \lambda'_{n_k} \to 0$ in probability. □

For the proofs of Lemma 4.6 and Theorem 4.7, see the appendix.

4.2. *Case Two* $\gamma \in (1/3, 1)$. We next deal briefly with the case $\gamma \in (1/3, 1)$ before considering the most interesting boundary case $\gamma = 1/3$. Our treatment will be condensed since the asymptotics for this case follow the same patterns as when the observation times possess a Lebesgue density. The notations, however, are developed in detail and will also be used for the boundary case.

For this case, we assume that both f and g are positive and continuously differentiable in a neighborhood of x_0. In order to study the asymptotics of the isotonic regression estimator $\hat{F}(t_l)$ of $F(t_l)$ and the likelihood ratio statistic for testing the hypotheses (3.3), the following localized processes will be of interest: for $u \in I_n = [(a - t_l)n^{1/3}, (b - t_l)n^{1/3}]$, define

$$\mathbb{X}_n(u) = n^{1/3}(\hat{F}(t_l + un^{-1/3}) - F(t_l)), \tag{4.2}$$

$$\mathbb{Y}_n(u) = n^{1/3}(\hat{F}^o(t_l + un^{-1/3}) - F(t_l)). \tag{4.3}$$

Define subsets of I_n as follows: $I_n^l = [(a - t_l)n^{1/3}, 0)$, $I_n^m = [0, (t_r - t_l)n^{1/3})$, $I_n^r = [(t_r - t_l)n^{1/3}, (b - t_l)n^{1/3}]$ and $\tilde{I}_n^r = I_n^r \cup I_n^m = [0, (b - t_l)n^{1/3}]$. Note that $I_n = I_n^l \cup I_n^m \cup I_n^r$ goes to $\mathbb{R}$, I_n^m goes to $\{0\}$, and $\tilde{I}_n^r$ is a left-side enlargement of I_n^r. Next, define the following normalized processes on I_n:

$$G_n^\star(h) = g(x_0)^{-1} n^{1/3} \tilde{G}_n(hn^{-1/3}), \tag{4.4}$$

$$V_n^\star(h) = g(x_0)^{-1} n^{2/3} \left(\tilde{V}_n(hn^{-1/3}) - F(t_l)\tilde{G}_n(hn^{-1/3}) \right), \tag{4.5}$$

where $\tilde{G}_n$ and $\tilde{V}_n$ are defined in (3.8). After some straight forward algebra, from (3.7), (3.11), (4.2), and (4.3), we have the following technically useful characterizations of $\mathbb{X}_n$ and $\mathbb{Y}_n$: for $u \in I_n$,

$$\mathbb{X}_n(u) = LS \circ GCM \left\{ \left(G_n^\star(h), V_n^\star(h) \right), h \in I_n \right\} (G_n^\star(u)); \tag{4.6}$$

(4.7)

$$\mathbb{Y}_n(u) = \begin{cases} 0 \wedge LS \circ GCM \left\{ \left(G_n^\star(h), V_n^\star(h) \right), h \in I_n^l \right\} (G_n^\star(u)), & \text{if } u \in I_n^l, \\ 0, & \text{if } u \in I_n^m, \\ 0 \vee LS \circ GCM \left\{ \left(G_n^\star(h), V_n^\star(h) \right), h \in \tilde{I}_n^r \right\} (G_n^\star(u)), & \text{if } u \in I_n^r. \end{cases}$$

For constants $\kappa_1 > 0$ and $\kappa_2 > 0$, denote

$$\begin{aligned} X_{\kappa_1,\kappa_2}(h) &= \kappa_1 W(h) + \kappa_2 h^2, \text{ for } h \in \mathbb{R}, \\ X_{\kappa_1,\kappa_2}^l(h) &= \kappa_1 W(h) + \kappa_2 h^2, \text{ for } h \in (-\infty, 0], \\ X_{\kappa_1,\kappa_2}^r(h) &= \kappa_1 W(h) + \kappa_2 h^2, \text{ for } h \in [0, \infty), \end{aligned}$$

where W is a two-sided Brownian motion with $W(0) = 0$. Let G_{κ_1,κ_2}, $G^l_{\kappa_1,\kappa_2}$ and $G^r_{\kappa_1,\kappa_2}$ be the greatest convex minorants of X_{κ_1,κ_2}, $X^l_{\kappa_1,\kappa_2}$ and $X^r_{\kappa_1,\kappa_2}$, respectively. Then, define, for $h \in \mathbb{R}$,

$$\begin{aligned} g_{\kappa_1,\kappa_2}(h) &= LS(G_{\kappa_1,\kappa_2})(h), \\ g^o_{\kappa_1,\kappa_2}(h) &= (0 \wedge LS(G^l_{\kappa_1,\kappa_2})(h))\{h \in (-\infty, 0)\} \\ &\quad + (0 \vee LS(G^r_{\kappa_1,\kappa_2})(h))\{h \in (0, \infty)\}. \end{aligned}$$

Denote $\alpha = \sqrt{F(x_0)(1-F(x_0))/g(x_0)}$ and $\beta = f(x_0)/2$. Let $\mathscr{L}^p[-C,C]$ denote the class of Borel measurable real-valued functions defined on $[-C,C]$ that possess a finite p'th moment with respect to Lebesgue measure on $[-C,C]$. Let '$\rightsquigarrow$' denote 'weak convergence' in addition to $\xrightarrow{d}$. With these notations, we have the following theorem on the joint distributional convergence of $(\mathbb{X}_n, \mathbb{Y}_n)$.

THEOREM 4.12 (Weak Convergence of $(\mathbb{X}_n, \mathbb{Y}_n)$). *Suppose f and g are positive and continuously differentiable in a neighborhood of x_0. Then, the finite dimensional marginals of the process $(\mathbb{X}_n, \mathbb{Y}_n)$ converge weakly to those of the process $(g_{\alpha,\beta}, g^o_{\alpha,\beta})$. Furthermore,*

$$\{(\mathbb{X}_n(u), \mathbb{Y}_n(u)), u \in [-C,C]\} \rightsquigarrow \{(g_{\alpha,\beta}(u), g^o_{\alpha,\beta}(u)), u \in [-C,C]\}$$

in $(\mathscr{L}^p[-C,C])^2$ for each $p \geq 1$ and $C > 0$.

REMARK 4.13. Note that $\mathbb{X}_n(0) = n^{1/3}(\hat{F}(t_l) - F(t_l))$. Then, by Theorem 4.12, it converges in distribution to $g_{\alpha,\beta}(0)$. By the Brownian scaling results on Page 1724 of Banerjee and Wellner (2001), we have, for $h \in \mathbb{R}$,

$$(g_{\alpha,\beta}(h), g^0_{\alpha,\beta}(h)) \stackrel{d}{=} (\alpha^2\beta)^{1/3}\,(g_{1,1}((\beta/\alpha)^{2/3}h), g^0_{1,1}((\beta/\alpha)^{2/3}h))$$

Hence, by noting $g_{1,1}(0) \stackrel{d}{=} 2\mathcal{Z}$, we have a concise result:

(4.8)

$$n^{1/3}(\hat{F}(t_l) - F(t_l)) \xrightarrow{d} (\alpha^2\beta)^{1/3} g_{1,1}(0) \stackrel{d}{=} \left(\frac{4f(x_0)F(x_0)(1-F(x_0))}{g(x_0)}\right)^{1/3} \mathcal{Z}.$$

Thus, the limit distribution of $\hat{F}(t_l)$ is exactly the same as one would encounter in the current status model with survival distribution F and the observation times drawn from a Lebesgue density function g.

Inference for $F(x_0)$: Note that $n^{1/3}(\hat{F}(x_0) - F(x_0)) = \mathbb{X}_n(0) + n^{1/3}(F(t_l) - F(x_0))$. Since, under the conditions of Theorem 4.12, $n^{1/3}(F(t_l) - F(x_0))$ converges to 0 for $\gamma \in (1/3, 1)$, we conclude that $n^{1/3}(\hat{F}(x_0) - F(x_0))$ converges in distribution to the right-hand side of (4.8), whose distribution is well-characterized.

The following result states the asymptotic distribution of the LRS $2\log\lambda_n$ for testing the hypotheses (3.3).

THEOREM 4.14 (Weak Convergence of $2\log\lambda_n$). *Under the null hypothesis in (3.3), i.e. $F(t_l)=\theta_l$, we have*

$$2\log\lambda_n \xrightarrow{d} \mathbb{D} = \int_{\mathbb{R}} (g_{1,1}(u)^2 - g_{1,1}^o(u)^2)du.$$

REMARK 4.15. This is the same limit distribution as obtained in Banerjee and Wellner (2001) under a Lebesgue density assumption on the observation times. This distribution also appears in Banerjee (2007) in connection with likelihood ratio tests in general monotone regression models. The proofs of the above theorems are omitted as they can be established via arguments similar to those in Banerjee (2007) using continuous mapping for slopes of greatest convex minorants. Note that the limit distribution of the likelihood ratio statistic is free of nuisance parameters.

REMARK 4.16. Results similar to Theorem 4.12 and Theorem 4.14 hold for the case $\gamma \geq 1$. More specifically, the slight difference is that it is more convenient to consider $F(x_0)$ directly instead of $F(t_l)$ for the case $\gamma \geq 1$. This, together with Theorem 4.12 and Theorem 4.14, means that a discrete random observation time with a dense enough grid, i.e. $\gamma > 1/3$, leads to the same asymptotics as a continuous random observation time.

4.3. *Case Three* $\gamma = 1/3$. Now we consider the most interesting boundary case $\gamma = 1/3$. For this case, we assume as in the previous section that both f and g are is positive and continuously differentiable in a neighborhood of x_0. Let the localized processes $\mathbb{X}_n(u)$ and $\mathbb{Y}_n(u)$ be exactly as in Subsection 4.2 on Case Two $\gamma \in (1/3, 1)$. Since the order of the unit grid spacing δ is exactly $n^{-1/3}$, equal to the order of the increment $un^{-1/3}$ in $\tilde{G}_n$ and $\tilde{V}_n$, $\mathbb{X}_n$ and $\mathbb{Y}_n$ have potential jumps only at ci for $i \in \mathcal{I}_n = (I_n/c) \cap \mathbb{Z}$. Thus, it is equivalent to consider $\mathbb{X}_n$ and $\mathbb{Y}_n$ on those ci's. Define $\mathcal{I}_n^l = (I_n^l/c) \cap \mathbb{Z}$ and $\mathcal{I}_n^m, \mathcal{I}_n^r$ and $\tilde{\mathcal{I}}_n^r$ in analogous fashion. Note that $\mathcal{I}_n$, $\mathcal{I}_n^l$, $\mathcal{I}_n^r$ and $\tilde{\mathcal{I}}_n^r$ go to $\mathbb{Z}$, $-\mathbb{N}$, $\mathbb{N}$ and $\{0\}\cup\mathbb{N}$, respectively, as n goes to infinity and that $\mathcal{I}_n^m$ always equals $\{0\}$. We then have, for $i \in \mathcal{I}_n$,

$$\mathbb{X}_n(ci) = LS \circ GCM\left\{(G_n^\star(ck), V_n^\star(ck)), k \in \mathcal{I}_n\right\}(G_n^\star(ci)), \tag{4.9}$$

and

$$\mathbb{Y}_n(ci) = \begin{cases} 0 \wedge LS \circ GCM\left\{(G_n^\star(ck), V_n^\star(ck)), k \in \mathcal{I}_n^l\right\}(G_n^\star(ci)), & \text{if } i \in \mathcal{I}_n^l; \\ 0, & \text{if } i \in \mathcal{I}_n^m; \\ 0 \vee LS \circ GCM\left\{(G_n^\star(ck), V_n^\star(ck)), k \in \tilde{\mathcal{I}}_n^r\right\}(G_n^\star(ci)), & \text{if } i \in \mathcal{I}_n^r. \end{cases} \tag{4.10}$$

Denote $\mathcal{P}_c(k) = (\mathcal{P}_{1,c}(k), \mathcal{P}_{2,c}(k)) = \big(ck, \alpha W(ck) + \beta c^2 k(1+k)\big)$, for $k \in \mathbb{Z}$, as a discrete process in $\mathbb{R}^2$, where α, β and W are defined as in Subsection 4.2. Based on $\mathcal{P}_c$, we define a discrete process, which will be related to $\mathbb{X}_n$:

$$\mathbb{X}(ci) = LS \circ GCM\left\{\mathcal{P}_c(k), k \in \mathbb{Z}\right\}(ci), \tag{4.11}$$

and a discrete process, which will be related to $\mathbb{Y}_n$:

$$\mathbb{Y}(ci) = \begin{cases} 0 \wedge LS \circ GCM\left\{\mathcal{P}_c(k), k \in -\mathbb{N}\right\}(ci), & \text{if } i \in -\mathbb{N}; \\ 0, & \text{if } i = 0; \\ 0 \vee LS \circ GCM\left\{\mathcal{P}_c(k), k \in \{0\} \cup \mathbb{N}\right\}(ci), & \text{if } i \in \mathbb{N}. \end{cases} \tag{4.12}$$

For simplicity of notation, for the remainder of this section, we will often write an integer interval as a usual interval with two integer endpoints. This will, however, not cause confusion since the interpretation of the interval will be immediate from the context.

For $(\mathbb{X}_n, \mathbb{Y}_n)$, the following joint distributional convergence holds.

THEOREM 4.17 (Weak Convergence of $(\mathbb{X}_n, \mathbb{Y}_n)$). *For each non-negative integer N, we have*

$$\{(\mathbb{X}_n(ci), \mathbb{Y}_n(ci)), i \in [-N, N]\} \rightsquigarrow \{(\mathbb{X}(ci), \mathbb{Y}(ci)), i \in [-N, N]\}.$$

PROOF. For each non-negative integer N, take an integer $M > N$. Then, we have the following two claims:

Claim 1: There exist (integer–valued) random variables $L_n, L_n' < -M$ and $U_n, U_n' > M$ such that L_n, L_n',U_n and U_n' are all $O_P(1)$ and that

$$\begin{aligned} & GCM\left\{(G_n^\star(ck), V_n^\star(ck)), k \in [L_n, U_n]\right\} \\ = \; & GCM\left\{(G_n^\star(ck), V_n^\star(ck)), k \in \mathbb{Z}\right\}|[G_n^\star(cL_n), G_n^\star(cU_n)], \\ & GCM\left\{(G_n^\star(ck), V_n^\star(ck)), k \in [L_n', -1]\right\} \\ = \; & GCM\left\{(G_n^\star(ck), V_n^\star(ck)), k \in -\mathbb{N}\right\}|[G_n^\star(cL_n'), G_n^\star(-c)], \\ & GCM\left\{(G_n^\star(ck), V_n^\star(ck)), k \in [0, U_n']\right\} \\ = \; & GCM\left\{(G_n^\star(ck), V_n^\star(ck)), k \in \{0\} \cup \mathbb{N}\right\}|[G_n^\star(0), G_n^\star(cU_n')]. \end{aligned}$$

Claim 2: There also exist (integer–valued) random variables $L, L' < -M$ and $U, U' > M$ such that L, L', U, U' are $O_P(1)$ and that

$$\begin{aligned} GCM\left\{\mathcal{P}_c(k), k \in [L, U]\right\} &= GCM\left\{\mathcal{P}_c(k), k \in \mathbb{Z}\right\}|[cL, cU], \\ GCM\left\{\mathcal{P}_c(k), k \in [L', -1]\right\} &= GCM\left\{\mathcal{P}_c(k), k \in -\mathbb{N}\right\}|[cL', -c], \\ GCM\left\{\mathcal{P}_c(k), k \in [0, U']\right\} &= GCM\left\{\mathcal{P}_c(k), k \in \{0\} \cup \mathbb{N}\right\}|[0, cU']. \end{aligned}$$

The proofs of Claim 1 and Claim 2 consist of technically important localization arguments. See Lemma 4.23 for the proof of Claim 1 and Lemma A.7 in the appendix for that of Claim 2. Intuitively speaking, Claim 1 ensures that the restriction of the greatest convex minorant of the process $(G_n^\star, V_n^\star)$ (which is involved in the construction of $\mathbb{X}_n$ and $\mathbb{Y}_n$) to a bounded domain can be made equal, eventually, with arbitrarily high probability, to the greatest convex minorant of the restriction of $(G_n^\star, V_n^\star)$ to that domain, provided the domain is chosen appropriately large. A similar fact holds for the greatest convex minorant of the process $\mathcal{P}_c$, which is involved in the construction of $\mathbb{X}$ and $\mathbb{Y}$. These equalities then translate to the left-derivatives of the GCM's involved and the proof is completed by invoking a continuous mapping theorem for the GCM's of the restriction of $(G_n^\star, V_n^\star)$ on bounded domains, along with Claims 1 and 2, which enable the use of a key approximation lemma – a simple extension of Lemma 4.2 in Prakasa Rao (1969) – stated below.

LEMMA 4.18. *Suppose that for each $\epsilon > 0$, $\{W_{n\epsilon}\}$, $\{W_n\}$ and $\{W_\epsilon\}$ are sequences of random vectors, W is a random vector and such that*

1. $\lim_{\epsilon\to 0} \overline{\lim}_{n\to\infty} \mathbb{P}\left(W_{n\epsilon} \neq W_n\right) = 0$,
2. $\lim_{\epsilon\to 0} \mathbb{P}\left(W_\epsilon \neq W\right) = 0$,
3. $W_{n\epsilon} \rightsquigarrow W_\epsilon$, *as* $n \to \infty$ *for each* $\epsilon > 0$.

Then $W_n \rightsquigarrow W$, *as* $n \to \infty$.

From Claims 1 and 2, for every (small) $\epsilon > 0$, there exists an integer M_ϵ large enough such that

$$P\left(M_\epsilon > \max\{|L_n|, |L_n'|, U_n, U_n', |L|, |L'|, U, U'\}\right) > 1 - \epsilon.$$

Denote, for $i \in [-N, N]$,

$$\begin{aligned} \mathbb{X}_n^{M_\epsilon}(ci) &= LS \circ GCM\left\{\left(G_n^\star(ck), V_n^\star(ck)\right), k \in [\pm M_\epsilon]\right\}\left(G_n^\star(ci)\right), \\ \mathbb{X}^{M_\epsilon}(ci) &= LS \circ GCM\left\{\mathcal{P}_c(k), k \in [\pm M_\epsilon]\right\}(ci), \end{aligned}$$

and for $i \in [-N, -1]$, $\{0\}$ and $[1, N]$

$$\mathbb{Y}_n^{M_\epsilon}(ci) = \begin{cases} 0 \wedge LS \circ GCM\left\{\left(G_n^\star(ck), V_n^\star(ck)\right), k \in [-M_\epsilon, -1]\right\}\left(G_n^\star(ci)\right) \\ 0, \\ 0 \vee LS \circ GCM\left\{\left(G_n^\star(ck), V_n^\star(ck)\right), k \in [0, M_\epsilon]\right\}\left(G_n^\star(ci)\right), \end{cases}$$

$$\mathbb{Y}(ci) = \begin{cases} 0 \wedge LS \circ GCM\left\{\mathcal{P}_c(k), k \in [-M_\epsilon, -1]\right\}(ci), & \text{if } i \in [-N, -1]; \\ 0, & \text{if } i = 0; \\ 0 \vee LS \circ GCM\left\{\mathcal{P}_c(k), k \in [0, M_\epsilon]\right\}(ci), & \text{if } i \in [1, N]. \end{cases}$$

Denote $[\pm N] = [-N, N]$ and

$$\begin{aligned} A_n &= \left\{ \left\{ \left(\mathbb{X}_n^{M_\epsilon}(ci), \mathbb{Y}_n^{M_\epsilon}(ci) \right), i \in [\pm N] \right\} \neq \left\{ \left(\mathbb{X}_n(ci), \mathbb{Y}_n(ci) \right), i \in [\pm N] \right\} \right\}, \\ A &= \left\{ \left\{ \left(\mathbb{X}^{M_\epsilon}(ci), \mathbb{Y}^{M_\epsilon}(ci) \right), i \in [\pm N] \right\} \neq \left\{ \left(\mathbb{X}(ci), \mathbb{Y}(ci) \right), i \in [\pm N] \right\} \right\}. \end{aligned}$$

Then, the following three facts hold:

Fact 1: $\lim_{\epsilon \to 0} \overline{\lim}_{n \to \infty} \mathbb{P}(A_n) = 0.$
Fact 2: $\lim_{\epsilon \to 0} \mathbb{P}(A) = 0.$
Fact 3: $\left\{ \left(\mathbb{X}_n^{M_\epsilon}(ci), \mathbb{Y}_n^{M_\epsilon}(ci) \right), i \in [\pm N] \right\} \rightsquigarrow \left\{ \left(\mathbb{X}^{M_\epsilon}(ci), \mathbb{Y}^{M_\epsilon}(ci) \right), i \in [\pm N] \right\}$, as $n \to \infty$ for each $\epsilon > 0$.

Since A_n and A are subsets of $\{M_\epsilon \le \max\{|L_n|, U_n, |L|, U, |L_n'|, U_n', |L'|, U'\}\}$, whose probability is less than ϵ, Facts 1 and 2 hold. Fact 3 will be proved in Lemma 4.24 in Section 4.3.1. Therefore, by Lemma 4.18, we have the finite dimensional weak convergence. □

REMARK 4.19. From Theorem 4.17, we have $n^{1/3}(\hat{F}(t_l) - F(t_l)) = \mathbb{X}_n(0)$ converges weekly to $\mathbb{X}(0) = LS \circ GCM\{\mathcal{P}_c(k), k \in \mathbb{Z}\}(0)$. Note that $\mathbb{X}(0)$ is precisely the random variable $\mathcal{S}_c$ defined in Section 2.

THEOREM 4.20 (Weak Convergence of $2\log\lambda_n$). *Under the null hypothesis of (3.3), i.e. $F(t_l) = \theta_l$, we have*

$$2\log\lambda_n \rightsquigarrow \mathbb{M}_c = \frac{cg(x_0)}{F(x_0)(1 - F(x_0))} \sum_{i=-\infty}^{\infty} \left(\mathbb{X}(ci)^2 - \mathbb{Y}(ci)^2 \right).$$

PROOF. Denote $J_n = \{1 \le i \le n : \hat{F}(X_i) \neq \hat{F}^o(X_i)\}$. Then, we have

$$2\log\lambda_n = 2(I - II) \text{ with } I = \sum_{i \in J_n} \psi(X_i, Y_i; \hat{F}), \quad II = \sum_{i \in J_n} \psi(X_i, Y_i; \hat{F}^o)$$

and $\psi(x, y; F) = y \log F(x) + (1 - y)\log(1 - F(x))$. By Taylor expansions of $\log \hat{F}(X_i)$ and $\log(1 - \hat{F}(X_i))$ around the point $t_l \equiv t_{l,n}$ (with $F_l = F(t_l)$) we have $I = I(1) + I(2) + I(3) + I(4)$, where

$$I(1) = \log F_l \sum_{i \in J_n} Y_i + \log(1 - F_l) \sum_{i \in J_n} (1 - Y_i),$$
$$I(2) = \sum_{i \in J_n} \left(\frac{Y_i}{F_l} - \frac{1 - Y_i}{1 - F_l} \right) \left(\hat{F}(X_i) - F_l \right),$$
$$I(3) = -\frac{1}{2} \sum_{i \in J_n} \left(\frac{Y_i}{F_l^2} + \frac{1 - Y_i}{(1 - F_l)^2} \right) \left(\hat{F}(X_i) - F_l \right)^2,$$
$$I(4) = \frac{1}{3} \sum_{i \in J_n} \left(\frac{Y_i}{F_{li}^{\star 3}} - \frac{1 - Y_i}{(1 - F_{li}^{\star\star})^3} \right) \left(\hat{F}(X_i) - F_l \right)^3,$$

and $F_{li}^{\star}$ and $F_{li}^{\star\star}$ both lie between F_l and $\hat{F}(X_i)$. By noting that $\{X_i : i \in J_n\}$ is the union of several whole "blocks" of $X_{(i)}$'s in the unconstrained isotonic regression and that over each block $\hat{F}$ equals the average of the corresponding block responses, we have, for each nonnegative integer k,

$$\sum_{i \in J_n} (Y_i - \hat{F}(X_i))(\hat{F}(X_i) - F_l)^k = 0. \tag{4.13}$$

Using this fact, we have

$$I(2) = \frac{1}{F_l(1 - F_l)} \sum_{i \in J_n} \left(\hat{F}(X_i) - F_l \right)^2,$$
$$I(3) = -\frac{1}{2} I(2) - \frac{1}{2} \left(\frac{1}{F_l^2} - \frac{1}{(1 - F_l)^2} \right) \sum_{i \in J_n} \left(\hat{F}(X_i) - F_l \right)^3.$$

Thus, I simplifies to

$$I = I(1) + \frac{1}{2F_l(1 - F_l)} \sum_{i \in J_n} \left(\hat{F}(X_i) - F_l \right)^2 + R_1,$$

where R_1 is the remaining term.

Similarly, we have

$$II = II(1) + \frac{1}{2F_l(1 - F_l)} \sum_{i \in J_n} \left(\hat{F}^o(X_i) - F_l \right)^2 + R_2,$$

where $II(1) = I(1)$ and R_2 is the remaining term, by noting that the corresponding equation of (4.13) still holds since $\{X_i : i \in J_n\}$ is the union of whole "blocks" of those two constrained isotonic regressions and a special interval containing t_l

over which $\hat{F}^o$ always equals to F_l. In fact, the special interval consists of t_l and the "truncated" blocks of those two constrained isotonic regressions.

We will show both R_1 and R_2 are $o_P(1)$ in Lemma A.14. Thus, we have:

$$2\log\lambda_n = \frac{1}{F_l(1-F_l)} \sum_{i\in J_n} \left[\left(\hat{F}(X_i) - F_l\right)^2 - \left(\hat{F}^o(X_i) - F_l\right)^2 \right] + o_P(1).$$

Denote $D_n = \{1 \le j \le K : \hat{F}(t_j) \ne \hat{F}^o(t_j)\}$. Then D_n is either an empty set or an integer interval including l. Denote $\tilde{D}_n = D_n - l$ also by $[-L_n, U_n]$. For empty D_n, $[-L_n, U_n]$ is understood as $\emptyset$. A summation with an empty index set is defined to be 0.

Next, for $1 \le j \le K$, let $W'_j = \#\{1 \le i \le n : X_i = t_j\}$, the number of observation times equal to t_j. Then,

$$\begin{aligned}
&\sum_{i\in J_n} \left[\left(\hat{F}(X_i) - F_l\right)^2 - \left(\hat{F}^o(X_i) - F_l\right)^2 \right] \\
&= \sum_{j\in D_n} \left[\left(\hat{F}(t_j) - F_l\right)^2 - \left(\hat{F}^o(t_j) - F_l\right)^2 \right] W'_j \\
&= \sum_{j\in \tilde{D}_n} \left[\left(\hat{F}(t_{l+j}) - F_l\right)^2 - \left(\hat{F}^o(t_{l+j}) - F_l\right)^2 \right] W'_{l+j} \\
&= \sum_{j\in[-L_n,U_n]} \left(\mathbb{X}_n^2(cj) - \mathbb{Y}_n^2(cj)\right) n^{-2/3} W'_{l+j}\,,
\end{aligned}$$

using the definitions of the process $\mathbb{X}_n$ and $\mathbb{Y}_n$. On the other hand,

$$W'_{l+j} = \sum_{i=1}^n \{t_{l+j-1} < X_i \le t_{l+j}\} = g(x_0) n^{2/3} \left(G_n^\star(cj) - G_n^\star(c(j-1))\right),$$

which means $n^{-2/3} W'_{l+j} = g(x_0) W_j$ with $W_j = G_n^\star(cj) - G_n^\star(c(j-1))$. Then, we have:

$$\begin{aligned}
2\log\lambda_n =& \frac{g(x_0)}{F_l(1-F_l)} \sum_{j\in[-L_n,U_n]} \left(\mathbb{X}_n^2(cj) - \mathbb{Y}_n^2(cj)\right) W_j + o_P(1) \\
=& \frac{cg(x_0)}{F_l(1-F_l)} S_n + R_3 + o_P(1),
\end{aligned}$$

where $S_n = \sum_{j\in[-L_n,U_n]} \left(\mathbb{X}_n^2(cj) - \mathbb{Y}_n^2(cj)\right)$ and R_3 is the remaining term.

Similarly, denote $E = \{j \in \mathbb{Z} : \mathbb{X}(cj) \ne \mathbb{Y}(cj)\}$ for the limiting processes $\mathbb{X}$ and $\mathbb{Y}$. Then, E is either an empty set or an integer interval usually including 0.

Denote E also by $[-L, U]$. For empty E, $[-L, U]$ is understood as $\emptyset$ as before. Let $S = \sum_{j \in [-L,U]} \left(\mathbb{X}^2(cj) - \mathbb{Y}^2(cj) \right)$.

We will show $S_n \rightsquigarrow S$ in Lemma A.11. On the other hand, it will be shown that $R_3 = o_P(1)$ in Lemma A.14. Therefore, we complete the proof by noticing $F_l \to F(x_0)$ and applying Slutsky's Lemma. $\square$

REMARK 4.21. The limit distribution $\mathbb{M}_c$ in Theorem 4.20 is, up to a constant, the difference of two discrete stochastic processes summed over the (discrete) time axis and depends on parameters. Owing to the discreteness of the time axis, there is no scaling to filter out the nuisance parameters in $\mathbb{M}_c$ so that we do not have a universal limit distribution for the likelihood ratio statistic in this boundary case. As a comparison, in the case with $\gamma > 1/3$ or with a continuous design density, Brownian scaling ensures that all the nuisance parameters in the expression for the likelihood ratio statistic disappear, leaving us with the universal limit $\mathbb{D}$. For more details on the nature of the Brownian scaling arguments, see, for example, Pages 1724-1725 of Banerjee and Wellner (2001).

Inference for $F(x_0)$: The linearly interpolated estimator that we used to make inference for $F(x_0)$ when $\gamma < 1/3$ cannot be made to work in this situation. Inference on $F(x_0)$ with the linear interpolant can, however, be made if we consider a slightly altered setting for the grid of observation times. This and related issues are discussed towards the end of the next section on adaptive inference for F at a point.

4.3.1. *Technical Details.*

LEMMA 4.22 (Localization Argument). *Claim 1 in the proof of Theorem 4.17 holds.*

PROOF. We only prove the first equality. The others can be established through analogous arguments. Denote $\mathbb{K}_n(x) = GCM \left\{ (G_n(t), V_n(t)), t \in [a, b] \right\}(x)$ for $x \in [0, 1]$. Recall $t_{l+j} = t_l + jcn^{-1/3}$ for $j \in \mathbb{Z}$. Then, it is sufficient to show there exist $x_{L_n} < t_{l-M}$ and $x_{U_n} > t_{l+M}$ such that

(1) both $t_l - x_{L_n}$ and $x_{U_n} - t_l$ are $O_P(n^{-1/3})$, and
(2) $GCM \left\{ (G_n(t), V_n(t)), t \in [x_{L_n}, x_{U_n}] \right\} = \mathbb{K}_n | [G_n(x_{L_n}), G_n(x_{U_n})]$.

Let x_{L_n} and x_{U_n} be the largest grid point less than t_{l-M} and the smallest grid point larger than or equal to t_{l+M}, at which $\hat{F}$ jumps. Thus, the slopes of $\mathbb{K}_n$ change at (and $\mathbb{K}_n$ and CSD agree on) $G_n(x_{L_n})$ and $G_n(x_{U_n})$.

With the above choices of x_{L_n} and x_{U_n}, **(2)** is satisfied. We next show that **(1)** holds. It is sufficient to show both $t_{l-M} - x_{L_n}$ and $x_{U_n} - t_{l+M}$ are $O_P(n^{-1/3})$.

We next show the former and the latter can be established in the same way. Denote $T_n = t_{l-M} - x_{L_n}$ and let S_n be the smallest non-negative value such that $\hat{F}$ changes its value at $t_{l-M} + S_n$. First, by the definition of $\hat{F}$, for all β, we have

$$V_n(t_{l-M} + \beta) \geq \mathbb{K}_n(t_{l-M}) + \hat{F}(t_{l-M}) \left[G_n(t_{l-M} + \beta) - G_{n(t_{l-M})} \right].$$

Denote, for all β,

$$\Gamma_n(\beta) = \left[V_n(t_{l-M} + \beta) - V_n(t_{l-M}) \right] - \hat{F}(t_{l-M}) \left[G_n(t_{l-M} + \beta) - G_n(t_{l-M}) \right].$$

Then, it is easy to see that $\Gamma_n(0) = 0$ and Γ_n achieves its minimum at both $-T_n$ and S_n. Thus, $\Gamma_n(-T_n) \leq 0$.

Letting $\mathbb{P}$ denote the distribution of (Y_{1n}, X_{1n}) (note the suppression of dependence on n) and $\mathbb{P}_n$ the empirical measure of n i.i.d. observations from this distribution, we have via simple algebra,

$$\begin{aligned} \Gamma_n(\beta) =& (\mathbb{P}_n - \mathbb{P}) g_{n1}(x, y; \beta) - \gamma_n (\mathbb{P}_n - \mathbb{P}) g_{n2}(x; \beta) \\ & + \mathbb{P} g_{n1}(x, y; \beta) - \gamma_n \mathbb{P} g_{n2}(x; \beta), \end{aligned}$$

where $\gamma_n = \hat{F}(t_{l-M}) - F(t_{l-M})$, $g_{n1}(x, y; \beta) = (y - F(t_{l-M})) g_{n2}(x; \beta)$ and $g_{n2}(x; \beta) = \{x \leq t_{l-M} + \beta\} - \{x \leq t_{l-M}\}$.

Since T_n, S_n and γ_n are $o_P(1)$ by a standard consistency argument, we can analyze locally. Now, for all sufficiently small $\epsilon > 0$, we can find, a neighborhood $\mathcal{N}$ of 0 (that may depend on ϵ), such that for all sufficiently large n (depending on ϵ), the following facts hold for every $\beta \in \mathcal{N}$:

Fact 1: $\left| \mathbb{P} g_{n1}(x, y; \beta) - (1/2) g(x_0) f(x_0) \beta^2 \right| \leq \epsilon \beta^2 + O(n^{-2/3})$.
Fact 2: $\left| \mathbb{P} g_{n2}(x; \beta) - g(x_0) \beta \right| \leq K \beta^2 + O(n^{-1/3})$.
Fact 3: $\left| (\mathbb{P}_n - \mathbb{P}) g_{n1}(x, y; \beta) \right| \leq \epsilon \beta^2 + O_P(n^{-2/3})$.
Fact 4: $\left| (\mathbb{P}_n - \mathbb{P}) g_{n2}(x; \beta) \right| \leq \epsilon \beta^2 + O_P(n^{-2/3})$.

The O and O_P terms in the above facts are non-negative and can depend on ϵ, but not on $\beta \in \mathcal{N}$. The constant K can again depend on ϵ. For the purpose of continuity, the proofs of the above facts are provided in Lemmas A.3, A.4, A.5, and A.6. Next, choose and fix ϵ_0 small enough such that the above facts hold and furthermore $\epsilon_0 < f(x_0) g(x_0)/8$. The above four facts imply that:

$$\begin{aligned} & \left| \Gamma_n(\beta) - (1/2) f(x_0) g(x_0) \beta^2 + \gamma_n g(x_0) \beta \right| \\ \leq & 3\epsilon_0 \beta^2 + \gamma_n K \beta^2 + O_P(n^{-2/3}) + |\gamma_n| \, O(n^{-1/3}). \end{aligned}$$

Given $\eta > 0$, there exists N_η such that for all $n \geq N_\eta$, we have $\gamma_n K < \epsilon_0$ and $-T_n$ and S_n lie in $\mathcal{N}$ with probability at least $1 - \eta$. Denote this event as E_n. Then, on E_n, $A_n(\beta) \le \Gamma_n(\beta) \le B_n(\beta)$, where

$$A_n(\beta) = C_1\beta^2 - \gamma_n'\beta - O_P(n^{-2/3}) - \left|\gamma_n'\right| O(n^{-1/3}),$$
$$B_n(\beta) = C_2\beta^2 - \gamma_n'\beta + O_P(n^{-2/3}) + \left|\gamma_n'\right| O(n^{-1/3}),$$

and $C_1 = (f(x_0)g(x_0)/2 - 4\epsilon_0) > 0$, $C_2 = (f(x_0)g(x_0)/2 + 4\epsilon_0) > 0$ are constants and $\gamma_n' = \gamma_n g(x_0)$. Note that γ_n' and γ_n have the same stochastic order. From the definition of T_n and S_n,

$$\min_\beta \Gamma_n(\beta) = \max\left\{\Gamma_n(-T_n), \Gamma_n(S_n)\right\} \ge \max\left\{A_n(-T_n), A_n(S_n)\right\}.$$

Since one of $C_1(-T_n)^2 - \gamma_n'(-T_n)$ and $C_1 S_n^2 - \gamma_n' S_n$ must be non-negative, we readily conclude that $\min_\beta \Gamma_n(\beta) \ge -O_P(n^{-2/3}) - |\gamma_n'|O(n^{-1/3})$. Next, note that $B_n(\cdot)$ attains its minimum at $\gamma_n'/2C_2$. Hence, we also have

$$\begin{aligned}\min_\beta \Gamma_n(\beta) &\le \Gamma_n(\gamma_n'/(2C_2)) \le B_n(\gamma_n'/(2C_2)) \\ &\le -(\gamma_n')^2/(4C_2) + O_P(n^{-2/3}) + \left|\gamma_n'\right| O(n^{-1/3}).\end{aligned}$$

We conclude that

$$-O_P(n^{-2/3}) - |\gamma_n'|O(n^{-1/3}) \le -(\gamma_n')^2/(4C_2) + O_P(n^{-2/3}) + \left|\gamma_n'\right| O(n^{-1/3}),$$

which can be written equivalently as

$$(\gamma_n')^2 - |\gamma_n'|O(n^{-1/3}) - O_P(n^{-2/3}) \le 0.$$

This implies $|\gamma_n'| \le O_P(n^{-1/3}) =: \xi_n$ and thus $|\gamma_n'|$ is $O_P(n^{-1/3})$. Further,

$$0 = \Gamma_n(0) \ge \Gamma_n(-T_n) \ge A_n(-T_n) = C_1 T_n^2 + \gamma_n' T_n - O_P(n^{-2/3}) - |\gamma_n'|O(n^{-1/3}),$$

showing that T_n must be less in absolute value than the maximum of the roots of the quadratic equation $C_1 x^2 + \gamma_n' x - O_P(n^{-2/3}) - |\gamma_n'|O(n^{-1/3}) = 0$. This leads to

$$|T_n| \le \frac{\xi_n + \sqrt{\xi_n^2 + 4C_1 O_P(n^{-2/3}) + 4C_1\xi_n O(n^{-1/3})}}{2C_1} =: \tilde{\xi}_n,$$

and $\tilde{\xi}_n$ is again $O_P(n^{-1/3})$. Thus, for $\tau > 0$,

$$P(n^{1/3}T_n > \tau) \le P(E_n^c) + P(\{n^{1/3}T_n > \tau\} \cap E_n) \le P(E_n^c) + P(n^{1/3}\tilde{\xi}_n > \tau),$$

which is less than 2η for all n sufficiently large given τ large enough. Therefore, T_n is $O_P(n^{-1/3})$ and the proof is complete. □

LEMMA 4.23. *Fact 3 in the proof of Theorem 4.17 holds.*

PROOF. We view $\left\{\left(\mathbb{X}_n^{M_\epsilon}(ci), \mathbb{Y}_n^{M_\epsilon}(ci)\right), i \in [\pm N]\right\}$ as the image of a $2(2N+1)$ dimensional vector function Ψ defined on the vector of two processes $\mathcal{W}_n = (\mathcal{G}_{nf}, \mathcal{V}_{nf})$, where $\mathcal{G}_{nf} = \{G_n^\star(ck), k \in [\pm M_\epsilon]\}$ and $\mathcal{V}_{nf} = \{V_n^\star(ck), k \in [\pm M_\epsilon]\}$. More specifically, denote $\Psi = (\{\Psi_i\}_{i=-N}^N, \{\Psi_i^l\}_{i=-N}^{-1}, 0, \{\Psi_i^r\}_{i=1}^N)$, where

$$\Psi_i(\kappa, \lambda) = \max_{-M_\epsilon \le u < i;\, i \le v \le M_\epsilon} \min \{(\lambda(v) - \lambda(u))/(\kappa(v) - \kappa(u))\} \text{ for } i \in [\pm N],$$

$$\Psi_i^l(\kappa_l, \lambda_l) = \max_{-M_\epsilon \le u < i;\, i \le v < 0} \min \{(\lambda(v) - \lambda(u))/(\kappa(v) - \kappa(u))\} \wedge 0 \text{ for } i \in [-N, 1],$$

$$\Psi_i^r(\kappa_r, \lambda_r) = \max_{0 \le u < i;\, i \le v \le M_\epsilon} \min \{(\lambda(v) - \lambda(u))/(\kappa(v) - \kappa(u))\} \vee 0 \text{ for } i \in [1, N];$$

κ and λ are functions defined on $[\pm M_\epsilon]$. Then, $\mathbb{X}_n^{M_\epsilon}(ci) = \Psi_i((G_n^\star(kc), V_n^\star(kc)) : k \in [\pm M_\epsilon])$ for $i \in [\pm N]$, $\mathbb{Y}_n^{M_\epsilon}(ci) = \Psi_i^l((G_n^\star(kc), V_n^\star(kc)) : k \in [\pm M_\epsilon])$ for $i \in [-N, -1]$ and $\mathbb{Y}_n^{M_\epsilon}(ci) = \Psi_i^r((G_n^\star(kc), V_n^\star(kc)) : k \in [\pm M_\epsilon])$ for $i \in [1, N]$. Similarly, $\left\{\left(\mathbb{X}^{M_\epsilon}(ci), \mathbb{Y}^{M_\epsilon}(ci)\right), i \in [\pm N]\right\}$ is the image of the same $2(2N+1)$ dimensional vector function Ψ defined on the vector of two processes $\mathcal{W} = (\mathcal{G}_f, \mathcal{V}_f)$, where $\mathcal{G}_f = \{\mathcal{P}_{1,c}(k), k \in [\pm M_\epsilon]\}$ and $\mathcal{V}_f = \{\mathcal{P}_{2,c}(k), k \in [\pm M_\epsilon]\}$. Letting $\mathcal{H}$ be the set of real-valued functions defined on the integer interval $[\pm M_\epsilon]$, view (κ, λ) as an element of $\mathcal{H}^2$ equipped with the topology of pointwise convergence on each co-ordinate. Note that $\mathcal{W}_n$ and $\mathcal{W}$ are random elements assuming values in $\mathcal{H}^2$. Thus (κ_m, λ_m) is defined to converge to (κ_0, λ_0) if $\kappa_m(j) \to \kappa_0(j)$ for all $j \in [\pm M_\epsilon]$ and the same holds for λ_m and λ_0. Then, Ψ can be viewed as a map from $\mathcal{H}^2$ to $\mathbb{R}^{2(2N+1)}$. The continuity of Ψ follows immediately from the fact that each $\Psi_i, \Psi_i^l, \Psi_i^r$ is continuous: this follows from the fact that the max and min in the definitions of each of these functions are taken with respect to a finite number of elements. We show in Lemmas A.8 and A.9, that

$$\mathcal{G}_{nf} \overset{P^\star}{\to} \mathcal{G}_f \tag{4.14}$$

$$\mathcal{V}_{nf} \rightsquigarrow \mathcal{V}_f \tag{4.15}$$

Then, by Slutsky's lemma, we have $\mathcal{W}_n \rightsquigarrow \mathcal{W}$. The conclusion of the lemma is now a direct outcome of the continuous mapping theorem. □

5. Adaptive Inference for *F* at A Point. The goal in this section is to develop a scheme for constructing confidence intervals for $F(t_l)$ (and later also for $F(x_0)$) without knowledge of the underlying grid resolution controlled by the parameters γ and c, given current status data over an equally spaced grid of observation times on $[a, b]$. To this end, we first investigate the relationships between the three different asymptotic limits for the distribution of $\hat{F}(t_l)$ that were derived in the previous section, under different values of γ. Our first result relates the distribution of $\mathcal{S}_c$ to the standard normal.

THEOREM 5.1. *As $c \to \infty$, we have $\sqrt{c}\mathcal{S}_c \xrightarrow{d} \alpha Z$, where Z follows the standard normal distribution.*

REMARK 5.2. Recall that $\mathcal{S}_c$ does depend on both α and β. So, rigorously, it should be written as $\mathcal{S}_{c,\alpha,\beta}$. The notation $\mathcal{S}_c$, though convenient, is perhaps a bit unfortunate since dependence on other parameters in the problem are ignored. We need to keep this dependence in mind.

Our next result investigates the case where c goes to 0.

THEOREM 5.3. *As $c \to 0$, we have $\mathcal{S}_c \xrightarrow{d} g_{\alpha,\beta}(0) \stackrel{d}{=} 2(\alpha^2\beta)^{1/3}\mathcal{Z}$.*

REMARK 5.4. This result is somewhat easier to visualize at a heuristic level. Recall that $\mathcal{S}_c$ is the left-slope of the GCM of the process $\mathcal{P}_c$ at the point 0, the process itself being defined on the grid $c\mathbb{Z}$. As c goes to 0, the grid becomes finer and the process $\mathcal{P}_c$ will eventually be substituted by its limiting version, namely $X_{\alpha,\beta}$. Thus, in the limit $\mathcal{S}_c$ becomes $g_{\alpha,\beta}(0)$, the left-slope of the GCM of $X_{\alpha,\beta}$ at 0. The representation of this limit in terms of $\mathcal{Z}$ was established in Remark 4.13 following Theorem 4.12.

We are now in a position to propose our inference scheme. We focus on the so-called 'Wald-type' intervals for $F(t_l)$, i.e. intervals of the form $\hat{F}(t_l)$ plus and minus terms depending on the sample size and the large sample distribution of the estimator. Recall that the grid resolution is determined by unknown parameters $\gamma > 0$ and $c > 0$. We now *pretend* that γ is exactly equal to $n^{-1/3}$. This allows us to calculate the value of the corresponding parameter c, say $\hat{c}$, via the relation: $\lfloor (b-a)/(\hat{c}n^{-1/3})\rfloor = K(= \lfloor (b-a)/(cn^{-\gamma})\rfloor)$, where K is the number of grid points. Some algebra shows that

$$\hat{c} = \hat{c}_n = cn^{1/3-\gamma} + O(n^{1/3-2\gamma}) = cn^{1/3-\gamma}(1 + O(n^{-\gamma}))\,.$$

Thus, the calculated parameter $\hat{c}$ actually depends on n, and goes to ∞ and 0 for $\gamma \in (0, 1/3)$ and $\gamma \in (1/3, 1)$, respectively. We propose to use the distribution of $\mathcal{S}_{\hat{c}}$ as an approximation to the distribution of $n^{1/3}(\hat{F}(t_l) - F(t_l))$ for $\gamma \in (0, 1)$. Thus, an adaptive approximate $1 - \eta$ confidence interval for $F(t_l)$ is given by

$$\left[\hat{F}(t_l) - n^{-1/3}\, q(\mathcal{S}_{\hat{c}}, 1-\eta/2), \hat{F}(t_l) - n^{-1/3}\, q(\mathcal{S}_{\hat{c}}, (\eta/2))\right], \tag{5.1}$$

where $\eta > 0$ and $q(X, p)$ stands for the lower pth quantile of a random variable X with $p \in (0, 1)$.

The above adaptive confidence interval provides the correct asymptotic calibration, irrespective of the true value of γ. If γ happens to be $1/3$, then the adaptive confidence interval is constructed with the correct asymptotic result. Usually this situation is rare so that it is important to know the performance of the adaptive confidence interval for $\gamma \neq 1/3$. If *we knew* that $\gamma \in (1/3, 1)$, then, by the result (4.8) and the symmetry of $g_{\alpha,\beta}(0)$, the true confidence interval would be

$$\left[\hat{F}(t_l) \pm n^{-1/3}\, q(g_{\alpha,\beta}(0), (1-\eta/2))\right]. \tag{5.2}$$

Recall that $\hat{c}$ goes to 0 for $\gamma \in (1/3, 1)$. Thus, by Theorem 5.3, the quantile sequence $q(\mathcal{S}_{\hat{c}}, p)$ converges to $q(g_{\alpha,\beta}(0), p)$ since $g_{\alpha,\beta}(0)$ is a continuous random variable. So, the adaptive confidence interval (5.1) converges to the true one (5.2) obtained when γ is in $(1/3, 1)$.

That the adaptive procedure also works with $\gamma \in (0, 1/3)$ will be shown by using Theorem 5.1. When the true γ is known to be less than $1/3$, from the result (4.1) and the symmetry of the standard normal random variable Z, the confidence interval is given by

$$\left[\hat{F}(t_l) \pm n^{-(1-\gamma)/2}\alpha c^{-1/2} q(Z, (1-\eta/2))\right]. \tag{5.3}$$

We then only need to show that, for every $p \in (0, 1)$, as $n \to \infty$,

$$\frac{n^{-1/3} q(\mathcal{S}_{\hat{c}}, p)}{n^{-(1-\gamma)/2}\alpha c^{-1/2} q(Z, p)} = \frac{n^{-1/3} c^{1/2}}{n^{-(1-\gamma)/2}\hat{c}^{1/2}} \cdot \frac{\hat{c}^{1/2} q(\mathcal{S}_{\hat{c}}, p)}{\alpha q(Z, p)} = I \cdot II \to 1.$$

Recall that $\hat{c}$ goes to ∞ for $\gamma \in (0, 1/3)$. By Theorem 5.1, we have $II \to 1$ as $n \to \infty$. On the other hand, we can see I simplies to $(1 + O(n^{-\gamma}))^{-1/2}$ and therefore goes to 1. Thus, the adaptive confidence interval (5.1) also converges to the true one (5.3) obtained when γ is known to be in $(0, 1/3)$.

While the major advantage of our method lies in that it does not require making a call on the degree of sparsity γ — in that respect, it *adjusts automatically* to the inherent rate of growth of the number of distinct observation times to the number of individuals, and that is an extremely desirable property — there are some implementational issues with the method that should be pointed out.

Firstly, note that nuisance parameters, namely $g(x_0)$ and $f(x_0)$, do need to be estimated from the data. Of the two $f(x_0)$ is more difficult to estimate accurately. Estimation of nuisance parameters is however unavoidable using the Wald-type intervals we have been dealing with, *even with γ known*. This nuisance parameter problem is somewhat easier for $\gamma \in (0, 1/3)$, since one at most needs an estimate of $g(x_0)$ in that situation and this is readily available via standard smoothing techniques. For $\gamma \in (1/3, 1)$, both nuisance parameters enter into the limit distribution of $n^{1/3}(\hat{F}(t_l) - F(t_l))$. Of course, with γ *known but not equal to 1/3*, one could have dispensed with the Wald-type intervals altogether, considering instead intervals using LRS inversion. This would have obviated the need to estimate nuisance parameters, since in either case, $\gamma \in (0, 1/3)$ or $\gamma \in (1/3, 1)$, the LRS is asymptotically pivotal with known limit distributions χ_1^2 and $\mathbb{D}$ respectively. Recall that in the boundary scenario, i.e. for $\gamma = 1/3$, the asymptotic distribution of the LRS is no longer pivotal, so the usual advantageous feature of likelihood ratios is absent in this situation.

Secondly, note that unlike the cases $\gamma \in (0, 1/3)$ and $\gamma \in (1/3, 1)$, the limit distribution of $n^{1/3}(\hat{F}(t_l) - F(t_l))$ in the boundary case does not admit a natural scaling in terms of a fixed known distribution; with $\gamma \in (0, 1/3)$, the standard normal plays this role and for $\gamma \in (1/3, 1)$, Chernoff's distribution. This means that the quantiles of the limit distribution in the boundary case cannot be computed by multiplying the quantiles of a generic distribution by an estimated factor and must be generated time and again starting from the parent process $\mathcal{P}_c$ which depends on α and β, or to be more precise, from an estimated parent process generated by using consistent estimates of the unknown parameters α and β. This is, however, not a terribly major issue in these days of fast computing and in our opinion, the mileage obtained in terms of adaptivity more than compensates for the lack of scaling.

Finally, one might wonder if resampling techniques could be used for adaptive estimation. The problem, however, lies in the fact that while the usual n out of n bootstrap works for $\gamma \in (0, 1/3)$, it fails under the non-standard asymptotic regimes that operate for $\gamma \in [1/3, 1)$, as is clear from the work of Abrevaya and Huang (2005), Kosorok (2008) and Sen, Banerjee and Woodroofe (2010). Since γ is unknown, it is impossible to decide whether to use the standard bootstrap or not. One could argue that the m out of n bootstrap or subsampling will work irrespective of the value of γ but, again, the problem with using these methods is that they require knowledge of the convergence rate and this is unknown as it depends on the true value of γ.

Similar to the relationship among $\mathcal{S}_c$, Z and $\mathcal{Z}$, revealed by Theorem 5.1 and Theorem 5.3, we can establish a corresponding relationship among $\mathbb{M}_c$, χ_1^2 and $\mathbb{D}$.

THEOREM 5.5. *As $c \to \infty$, we have $\mathbb{M}_c \xrightarrow{d} \chi_1^2$.*

THEOREM 5.6. *As $c \to 0$, we have $\mathbb{M}_c \xrightarrow{d} \mathbb{D}$.*

From these two theorems, we can similarly develop another adaptive procedure based on the LRS of the boundary case for testing the hypotheses (3.3) to construct the so-called LR-type confidence intervals for $F(t_l)$ without the knowledge of the true value of $\gamma \in (0,1)$. These adaptive LR-type confidence intervals are also asymptotically correct, regardless the value of γ. As pointed out before, the asymptotic distribution of the LRS for the boundary case $\gamma = 1/3$ is no longer pivotal and the same nuisance parameters $g(x_0)$ and $f(x_0)$ need to be estimated in order to construct the adaptive LR-type confidence intervals. This is a reasonable price for not knowing the true value of γ.

Inference for $F(x_0)$: To make adaptive inference on $F(x_0)$ given $x_0 \in (a,b)$, consider a slightly altered setting for the grid. Suppose $p \in [0,1)$ and $q = 1-p$. Let $t_l = x_0 - p\delta$, $t_r = x_0 + q\delta$, $t_{l-i} = t_l - i\delta$, $t_{r+j} = t_r + j\delta$ for $i = 1, \cdots, l-1$ and $j = 1, \cdots, K-r$ with both $t_1 - a$ and $b - t_K$ in $(0, \delta]$. Since the unconstrained and constrained isotonic regressions $\hat{F}$ and $\hat{F}^o$ are well-defined only at the grid-points, the related random quantities $\mathbb{X}_n$ and $\mathbb{Y}_n$ are also well-defined without any modification.

Under the new setting of the grid, first consider the boundary case $\gamma = 1/3$, for which Theorem 4.17 still holds. Define $\tilde{F}(x_0) = q\hat{F}(t_l) + p\hat{F}(t_r)$, which is the linearly interpolated estimator for $F(x_0)$. Then, we have $n^{1/3}(\tilde{F}(x_0) - F(x_0)) = q\mathbb{X}_n(0) + p\mathbb{X}_n(1) + n^{1/3}(F(t_l) - F(x_0))$, which converges weakly to $q\mathbb{X}(0) + p\mathbb{X}(1) - cpf(x_0)$ given f is continuous around x_0. Then, we can construct confidence intervals for $F(x_0)$. However, with our original setting for the grid, $n^{1/3}(\tilde{F}(x_0) - F(x_0))$ might not converge weakly, because the p and q would depend on n without necessarily converging to a limit. To see this, let $a = 0$, $b = 1$, $x_0 = 1/2$, $c = 1$, $n_{1,k} = (2k)^3$ and $n_{2,k} = (2k+1)^3$ for $k \in \mathbb{N}$. Then, along the first subsequence $\{n_{1,k}\}$, $n_{1,k}^{1/3}(\tilde{F}(x_0) - F(x_0))$ converges weakly to $\mathbb{X}_n(0)$. But along the second subsequence $\{n_{2,k}\}$, $n_{2,k}^{1/3}(\tilde{F}(x_0) - F(x_0))$ converges weakly to $(\mathbb{X}_n(0) + \mathbb{X}_n(1))/2 + f(x_0)/2$. Since the limiting distributions along different subsequences are different, $n^{1/3}(\tilde{F}(x_0) - F(x_0))$ does not converge weakly.

Next, consider the case $\gamma \in (1/3, 1)$, for which the above decomposition of $n^{1/3}(\tilde{F}(x_0) - F(x_0))$ still holds. Since $n^{1/3}(F(t_l) - F(x_0))$ converges to 0 and $(\mathbb{X}_n(0), \mathbb{X}_n(1))$ converges weakly to $g_{\alpha,\beta}(0)(1,1)$, we have $n^{1/3}(\tilde{F}(x_0) - F(x_0))$ converges weakly to $g_{\alpha,\beta}(0)$ by noticing $p + q = 1$. On the other hand, similar to the argument for Theorem 5.3, $(\mathbb{X}(0), \mathbb{X}(1))$ converges weakly to $g_{\alpha,\beta}(0)(1,1)$ as c goes to 0. Thus, the limit distribution for $\gamma = 1/3$, $q\mathbb{X}(0) + p\mathbb{X}(1) - cpf(x_0)$, converges weakly to $g_{\alpha,\beta}(0)$, the limit distribution for $\gamma \in (1/3, 1)$, by noticing that $cpf(x_0)$ converges to 0 as c goes to 0. This means that adaptive inference on $F(x_0)$ can be made for large values of γ.

Finally, consider the case $\gamma \in (0, 1/3)$. Similar to the $R_{\hat{F}}$ defined before Theorem 4.7, here we generically denote $R_{\tilde{F}} = (p^2+q^2)^{-1/2}[(N_l+N_r)/2]^{1/2}(\tilde{F}(x_0) - F(x_0))$. Similar to the argument for Theorem 4.7, we have that $R_{\tilde{F}}$ converges weakly to $[F(x_0)(1-F(x_0))]^{1/2}N(0,1)$ for $\gamma \in (1/5, 1/3)$. Since both N_l and N_r

are asymptotically equivalent to $cg(x_0)n^{1-\gamma}$, we conclude that $c^{1/2}n^{(1-\gamma)/2}(\tilde{F}(x_0) - F(x_0))$ converges weakly to $(p^2+q^2)^{1/2}\alpha N(0,1)$. On the limiting distribution side, *another decomposition* of $n^{1/3}(\tilde{F}(x_0)-F(x_0))$ is exploited *for* $\gamma = 1/3$. We have $n^{1/3}(\tilde{F}(x_0) - F(x_0)) = qn^{1/3}(\hat{F}(t_l) - F(t_l)) + pn^{1/3}(\hat{F}(t_r) - F(t_r)) + qn^{1/3}(F(t_l) - F(x_0)) + pn^{1/3}(F(t_r) - F(x_0))$. Similar to the argument for Theorem 4.12, the above first two terms as a vector converges weakly to $(\mathbb{X}(0), \mathbb{X}'(0))$ with $\mathbb{X}'(0) \stackrel{d}{=} \mathbb{X}(0)$; the above last two terms converge to $-pqcf(x_0)$ and $pqcf(x_0)$, respectively. Thus, for $\gamma = 1/3$, $n^{1/3}(\tilde{F}(x_0)-F(x_0))$ converges weakly to $q\mathbb{X}(0)+p\mathbb{X}'(0)$. Now, similar to the argument for Theorem 5.1, we have that, as c goes to ∞, $\sqrt{c}(\mathbb{X}(0), \mathbb{X}'(0))$ converges weakly to $\alpha(Z_1, Z_2)$, where Z_1 and Z_2 are independent standard normal distributions. Thus, $\sqrt{c}(q\mathbb{X}(0) + p\mathbb{X}'(0))$ converges weakly to $(p^2+q^2)^{1/2}\alpha N(0,1)$ as c goes to ∞. This means that adaptive inference on $F(x_0)$ can also be made for values of γ in $(1/5, 1/3)$.

Similar to Theorem 4.7, for $\gamma = 1/5$ and $\gamma \in (0, 1/5)$ with $f'(x_0) \neq 0$, we have $c^{1/2}n^{(1-\gamma)/2}(\tilde{F}(x_0) - F(x_0))$ converges weakly to $(p^2+q^2)^{1/2}\alpha N(0,1) + (1/2)pqc^{5/2}f'(x_0)$ and $Sign(f'(x_0))\infty$, respectively. Thus, the adaptive procedure could still be applied with an adjustment for the bias term $(1/2)pqc^{5/2}f'(x_0)$ for $\gamma = 1/5$ but is not available for $\gamma \in (0, 1/5)$ any more. This basically means that when the grid resolution under the new grid setting is very sparse, the adaptive procedure would lose its effectiveness for inference at x_0. In comparison, in the original grid setting, the adaptive procedure would work *for all values of* $\gamma \in (0, 1/3)$ for inference on F at the point t_l.

5.0.2. *Proofs.* Here, we provide proofs of the main results in this section, apart from the proof of Theorem 5.6 which can be established by an extension of the ideas used in the proof of Theorem 5.3 and is skipped.

PROOF OF THEOREM 5.1. For $k \in \mathbb{Z}$, let

$$\tilde{h}(k) = \alpha\sqrt{c}W(ck) + \beta c^{5/2}k(1+k), \quad h(k) = \alpha c W(k) + \beta c^{5/2}k(1+k).$$

Then, we have $\{\tilde{h}(k), k \in \mathbb{Z}\} \stackrel{d}{=} \{h(k), k \in \mathbb{Z}\}$. Thus,

$$\sqrt{c}\mathcal{S}_c \stackrel{d}{=} LS \circ GCM\left\{(ck, h(k)), k \in \mathbb{Z}\right\}(0).$$

Define $\tilde{\mathcal{S}}_c = \sqrt{c}\,\mathcal{S}_c$. Denote

$$A_c = \left\{\frac{h(k)}{ck} < \frac{h(k+1)}{c(k+1)}, k = 1, 2, \cdots\right\},$$
$$B_c = \left\{\frac{h(-(k-1))}{c(k-1)} < \frac{h(-k)}{ck}, k = 2, 3, \cdots\right\}, \quad C_c = \left\{\frac{h(1)}{c} > \frac{-h(-1)}{c}\right\}.$$

Then, for $\omega \in A_c B_c C_c$, it is easy to see $\tilde{\mathcal{S}}_c = -\alpha W(-1)$. We will show in Lemma 5.7, $A_c B_c C_c \xrightarrow{P} 1$. Thus, $\tilde{\mathcal{S}}_c = \tilde{\mathcal{S}}_c A_c B_c C_c + \tilde{\mathcal{S}}_c(1 - A_c B_c C_c) \xrightarrow{d} -\alpha W(-1) \stackrel{d}{=} \alpha Z$, with $Z \sim N(0,1)$. Therefore, $\sqrt{c}\mathcal{S}_c \xrightarrow{d} \alpha Z$. $\square$

PROOF OF THEOREM 5.5. We have

$$\mathbb{M}_c = \alpha^{-2}\left(\sqrt{c}\mathcal{S}_c\right)^2 + c\alpha^{-2}\sum_{i\neq 0}\left(\mathbb{X}(ci)^2 - \mathbb{Y}(ci)^2\right) =: T + R.$$

By Theorem 5.1, we have $T \xrightarrow{d} Z^2 \sim \chi_1^2$. It suffices to show $R \xrightarrow{P} 0$. Letting A_c, B_c and C_c denote the same quantities as in the proof of Theorem 5.1 and letting

$$D_c = \left\{\frac{h(-1) - h(-2)}{-c + 2c} < 0\right\}, \quad E_c = \left\{\frac{h(1)}{c} > 0\right\},$$

for every $\omega \in A_c B_c C_c D_c E_c$, we have $R = 0$. We will show $A_c B_c C_c D_c E_c \xrightarrow{P} 1$ in Lemma 5.7. Thus, $R \xrightarrow{P} 0$, which completes the proof. $\square$

LEMMA 5.7. *Each of A_c, B_c, C_c, D_c and E_c in the proof of Theorem 5.1 and Theorem 5.5 converges to 1 in probability.*

PROOF. It is easy to show C_c, D_c or E_c converges to 1 in probability. The argument that A_c converges to one in probability is similar to that for B_c and only the former is established here. In order to show $P(A_c) \to 1$, it suffices to show $P(A_c^c) \to 0$. We have, for each $k \in \mathbb{Z}$,

$$\begin{aligned}
& P\left(\frac{h(k)}{ck} \geq \frac{h(k+1)}{c(k+1)}\right) \\
= \;& P\left(\frac{\alpha W(k)}{k} + \beta c^{3/2}(k+1) \geq \frac{\alpha W(k+1)}{k+1} + \beta c^{3/2}(k+2)\right) \\
= \;& P\left(\alpha\left[\frac{W(k)}{k} - \frac{W(k+1)}{k+1}\right] \geq \beta c^{3/2}\right) \\
= \;& P\left(N(0,1) \geq \alpha^{-1}\beta c^{3/2}\sqrt{k(k+1)}\right) \\
\leq \;& 2^{-1}\exp\left\{-2^{-1}\alpha^{-2}\beta^2 c^3 k(k+1)\right\},
\end{aligned}$$

using the fact that $W(k)/k - W(k+1)/(k+1) \sim N(0, (k(k+1))^{-1})$ and the inequality $P(N(0,1) > x) \leq 2^{-1}\exp\{(-2^{-1}x^2)\}$ for $x \geq 0$ (See, for example,

$< 2 >$ on Page 317 of Pollard (2002)). Then, we have

$$P(A_c^c) \le \sum_{k=1}^{\infty} P\left(\frac{h(k)}{ck} \ge \frac{h(k+1)}{c(k+1)}\right) \le \sum_{k=1}^{\infty} 2^{-1}\exp\left\{-2^{-1}\alpha^{-2}\beta^2c^3k^2\right\}$$
$$\le 2^{-1}\int_0^{\infty}\exp\left\{-2^{-1}\alpha^{-2}\beta^2c^3x^2\right\}dx = (\sqrt{2\pi}/4)\alpha\beta^{-1}c^{-3/2} \to 0,$$

as $c \to \infty$. Thus, $P(A_c) \to 1$, which completes the proof. □

PROOF FOR THEOREM 5.3. We want to show that $\mathcal{S}_c \xrightarrow{d} g_{\alpha,\beta}(0)$, as $c \to 0$, where $g_{\alpha,\beta}(0) = LS \circ GCM\{X_{\alpha,\beta}\}(0) = LS \circ GCM\{X_{\alpha,\beta}(t) : t \in \mathbb{R}\}(0)$ and $\mathcal{S}_c = LS \circ GCM\{\mathcal{P}_c\}(0) = LS \circ GCM\{\mathcal{P}_c(k) : k \in \mathbb{Z}\}(0)$. Since $\mathcal{S}_c = \mathcal{S}'_c + \beta c$, where $\mathcal{S}'_c = LS \circ GCM\{\mathcal{P}'_c : k \in \mathbb{Z}\}(0)$ and $\mathcal{P}'_c = \{(ck, \alpha W(ck) + \beta(ck)^2) : k \in \mathbb{Z}\}$, it is sufficient to show $\mathcal{S}'_c \xrightarrow{d} g_{\alpha,\beta}(0)$ as $c \to 0$. To make the notation simple and without causing confusion, in the following we still use $\mathcal{P}_c$ and $\mathcal{S}_c$ to denote $\mathcal{P}'_c$ and $\mathcal{S}'_c$. Also, it will be useful to think of $\mathcal{P}_c$ as a continuous process on $\mathbb{R}$ formed by linearly interpolating the points $\{ck, \mathcal{P}_{2,c}(ck) : k \in \mathbb{Z}\}$, where $\mathcal{P}_{2,c}(ck) = \alpha\, W(ck) + \beta\,(ck)^2 = X_{\alpha,\beta}(ck)$. Note that viewing $\mathcal{P}_c$ in this way keeps the GCM unaltered, i.e. the GCM of this continuous linear interpolated version is the same as that of the set of points $\{ck, \mathcal{P}_{2,c}(ck) : k \in \mathbb{Z}\}$ and the slope-changing points of this piece-wise linear GCM are still grid-points of the form ck.

Let L and U be the largest negative and smallest nonnegative x-axis coordinates of the slope changing points of the GCM of $X_{\alpha,\beta}$. Similarly, let L_c and U_c be the largest negative and smallest nonnegative x-axis coordinates of the slope changing points of the GCM of $\mathcal{P}_c$. For $K > 0$, define $g^K_{\alpha,\beta}(0) = LS \circ GCM\{X_{\alpha,\beta}(t) : t \in [-K, K]\}(0)$ and $\mathcal{S}^K_c = LS \circ GCM\{\mathcal{P}_c(t) : t \in [-K, K]\}(0)$.

We will show that, given $\epsilon > 0$, there exist $M_\epsilon > 0$ and $c(\epsilon)$ such that **(a)** for all $0 < c < c(\epsilon)$, $P(\mathcal{S}^{M_\epsilon}_c \ne \mathcal{S}_c) < \epsilon$ and **(b)** $P(g^{M_\epsilon}_{\alpha,\beta}(0) \ne g_{\alpha,\beta}(0)) < \epsilon$. These immediately imply that both **Fact 1:** $\lim_{\epsilon\to 0}\limsup_{c\to 0} P(\mathcal{S}^{M_\epsilon}_c \ne \mathcal{S}_c) = 0$ and **Fact 2:** $\lim_{\epsilon\to 0} P(g^{M_\epsilon}_{\alpha,\beta}(0) \ne g_{\alpha,\beta}(0)) = 0$ hold. We then show that **Fact 3:** For each $\epsilon > 0$, $\mathcal{S}^{M_\epsilon}_c \xrightarrow{d} g^{M_\epsilon}_{\alpha,\beta}(0)$ holds as well. Then, by Lemma 4.18, we have the conclusion $\mathcal{S}_c \xrightarrow{d} g_{\alpha,\beta}(0)$. Figure 1 illustrates the following argument.

Let $\tau_{-2} < \tau_{-1} < \tau_1 < \tau_2$ be four consecutive slope changing points of $G_{\alpha,\beta} = GCM\{X_{\alpha,\beta}\}$ with τ_{-1} denoting the first slope changing point to the left of 0 and τ_1 the first slope changing point to the right. Since τ_{-2} and τ_2 are $O_P(1)$, given $\epsilon > 0$, there exists $M_\epsilon > 0$ such that $P(-M_\epsilon < \tau_{-2} < \tau_2 < M_\epsilon) > 1 - \epsilon/4$. Note that the event $\{g^{M_\epsilon}_{\alpha,\beta}(0) = g_{\alpha,\beta}(0)\} \subset \{-M_\epsilon < \tau_{-2} < \tau_2 < M_\epsilon\}$ and it follows that $P(g^{M_\epsilon}_{\alpha,\beta}(0) \ne g_{\alpha,\beta}(0)) < \epsilon/4 < \epsilon$. Thus, **(b)** holds.

Next, consider the chord $C_1(t)$ joining $(0, G_{\alpha,\beta}(0))$ and $(\tau_{-2}, G_{\alpha,\beta}(\tau_{-2}))$. By the convexity of $G_{\alpha,\beta}$ over $[\tau_{-2}, 0]$ and $\tau_{-1} \in (\tau_{-2}, 0)$ being a slope changing

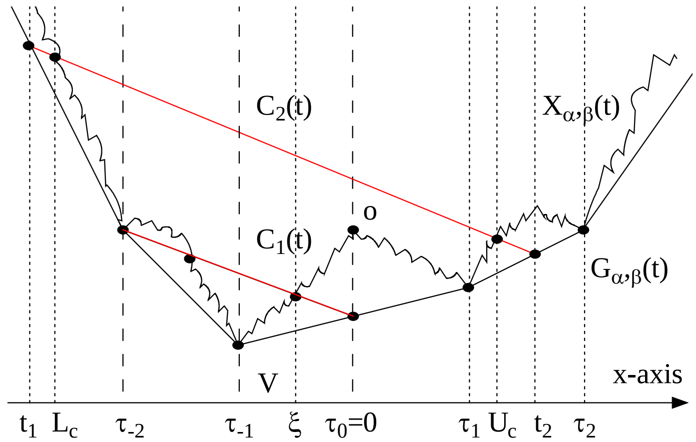

FIG 1. *An illustration for showing* $\{L_c\}$ *is* $O_P(1)$ *in the proof of Theorem* 5.3.

point, $X_{\alpha,\beta}(\tau_{-1}) = G_{\alpha,\beta}(\tau_{-1}) < C(\tau_{-1})$. But $C_1(0) = G_{\alpha,\beta}(0) < X_{\alpha,\beta}(0)$ and it follows by the intermediate value theorem that $\xi = \inf_{\tau_{-1}<t<0} \{t : X_{\alpha,\beta}(t) = C_1(t)\}$ is well-defined (since the set in question is non-empty), $\tau_{-1} < \xi < 0$, $C_1(\xi) = X_{\alpha,\beta}(\xi)$ and on $[\tau_{-1}, \xi)$, $X_{\alpha,\beta}(t) < C_1(t)$. Let $V = \xi - \tau_{-1}$. Since V is a continuous and positive random variable, there exists $\delta(\epsilon) > 0$ such that $P(V > \delta(\epsilon)) \geq 1 - \epsilon/4$. Then, the event $E_\epsilon = \{V > \delta(\epsilon)\} \cap \{-M_\epsilon < \tau_{-2}\}$ has probability larger than $1 - \epsilon/2$. For any $c < c(\epsilon) =: \delta(\epsilon)$, we claim that $L_c \geq \tau_{-2}$ on the event E_ϵ and the argument for this follows below.

If $L_c < \tau_{-2}$, consider the chord $C_2(t)$ connecting two points $(L_c, \mathcal{P}_{2,c}(L_c))$ and $(U_c, \mathcal{P}_{2,c}(U_c))$. This chord must lie strictly above the chord $\{C_1(t) : \tau_{-1} \leq t \leq 0\}$ since it can be viewed as a restriction of a chord connecting two points $(t_1, G_{\alpha,\beta}(t_1))$ and $(t_2, G_{\alpha,\beta}(t_2))$ with $t_1 \leq L_c < \tau_{-1} < 0 \leq U_c \leq t_2$. It then follows that all points of the form $\{ck, \mathcal{P}_{2,c}(ck) = X_{\alpha,\beta}(ck) : ck \in [L_c, U_c]\}$ must lie above $C_2(t)$. But there is at least one $ck^\star$ with $\tau_{-1} < ck^\star < \xi$ and such that $X_{\alpha,\beta}(ck^\star) < C_1(ck^\star) < C_2(ck^\star)$, which furnishes a contradiction.

We conclude that for any $c < c(\epsilon)$, $P(-M_\epsilon < L_c) > 1 - \epsilon/2$. A similar argument to the right-hand side of 0 shows that for the same c's (by the symmetry of two-sided Brownian motion about the origin), $P(U_c < M_\epsilon) > 1 - \epsilon/2$. Hence $P(-M_\epsilon < L_c < U_c < M_\epsilon) > 1 - \epsilon$. On this event, clearly $\mathcal{S}_c^{M_\epsilon} = \mathcal{S}_c$ and it follows that for all $c < c(\epsilon)$, $P(\mathcal{S}_c^{M_\epsilon} \neq \mathcal{S}_c) < \epsilon$. Thus, **(a)** also holds and Facts 1 and 2 are established.

It remains to establish Fact 3. This follows easily. For almost every ω, $X_{\alpha,\beta}(t)$ is uniformly continuous on $[\pm 2\, M_\epsilon]$. It follows by elementary analysis that (for almost every ω) on $[\pm M_\epsilon]$, the process $\mathcal{P}_c$, being the linear interpolant of the points

$\{ck, X_{\alpha,\beta}(ck) : -M_\epsilon \leq ck \leq M_\epsilon\} \cup \{(-M_\epsilon, \mathcal{P}_{2c}(-M_\epsilon)), (M_\epsilon, \mathcal{P}_{2,c}(M_\epsilon))\}$, converges uniformly to $X_{\alpha,\beta}$ as $c \to 0$. Thus, the left slope of the GCM of $\{\mathcal{P}_c(t) : t \in [\pm M_\epsilon]\}$, which is precisely $\mathcal{S}_c^{M_\epsilon}$, converges to $g_{\alpha,\beta}^{M_\epsilon}(0)$ since the GCM of the restriction of $X_{\alpha,\beta}$ to $[\pm M_\epsilon]$ is almost surely differentiable at 0 (see, for example, the Lemma on Page 330 of Robertson, Wright and Dykstra (1988) for a justification of this convergence). □

6. A Practical Procedure and Simulations. In this section, we provide a practical version of the adaptive procedure introduced in Section 5 to construct Wald-type confidence intervals for $F(t_l)$.

Recall that, in the adaptive procedure, we always specify $\gamma = 1/3$ and thus estimate the value of c by a solution $\hat{c}$ of the equation $K = \lfloor (b-a)/\hat{c}n^{-1/3} \rfloor$, where K is the number of grid points. To construct a $1 - 2\eta$ confidence interval for $F(t_l)$, quantiles of $\mathbb{X}_{\hat{c}}(0)$ are needed. Since $\mathbb{X}_c(0) = LS \circ GCM \{T(k), k \in \mathbb{Z}\}(0)$ (c is genetically used), we approximate $\mathbb{X}_c(0)$ with

$$\mathbb{X}_{c,K_a}(0) = LS \circ GCM \{\mathcal{P}_c(k), k \in [-K_a - 1, K_a]\}(0)$$

for some large $K_a \in \mathbb{N}$. Further, since

$$\mathbb{X}_{c,K_a}(0) = LS \circ GCM \{(\mathcal{P}_{1,c}(k)/c, \mathcal{P}_{2,c}(k)/c), k \in [-K_a - 1, K_a]\}(0)$$

where $\mathcal{P}_{1,c}(k)/c = k$ and $\mathcal{P}_{2,c}(k)/c = \alpha W(ck)/c + \beta ck(1+k)$, we have that $\mathbb{X}_{c,K_a}(0)$ is the isotonic regression at $k = 0$ of the data

$$\begin{aligned} & \{(k, \mathcal{P}_{2,c}(k)/c - \mathcal{P}_{2,c}(k-1)/c), k \in [-K_a, K_a]\} \\ = \;& \left\{(k, \alpha Z_k/\sqrt{c} + 2\beta ck, k \in [-K_a, K_a]\right\}, \end{aligned}$$

where $\{Z_k\}_{k=-K_a}^{K_a}$ are i.i.d. from $N(0,1)$, $\alpha = \sqrt{F(x_0)(1-F(x_0))/g(x_0)}$ and $\beta = f(x_0)/2$. To make this adaptive procedure practical, we next consider the estimation of α and β, or equivalently, the estimation of $F(x_0)$, $g(x_0)$ and $f(x_0)$.

First, we consider the estimation of $F(x_0)$ and $g(x_0)$. Although $F(x_0)$ can be consistently estimated by $\hat{F}(t_l)$, it is usually better to estimate $F(x_0)$ by $\rho\hat{F}(t_l) + (1-\rho)\hat{F}(t_r)$ with $\rho = (x_0 - t_l)/(t_r - t_l) \in [0, 1)$. To estimate $g(x_0)$, suppose the design density g is constant within a small interval around x_0, which is chosen to be $[t_{l-j^\star}, t_{r+j^\star}]$, where $j^\star$ is defined below in the estimation of $f(x_0)$. Then, from the estimating equation $(N_{l-j^\star+1} + \cdots + N_{r+j^\star})/n = g(x_0)(t_{r+j^\star} - t_{l-j^\star})$, one simple but consistent estimator of $g(x_0)$ is given by $\hat{g}(x_0) = (N_{l-j^\star+1} + \cdots + N_{r+j^\star})/[n(t_{r+j^\star} - t_{l-j^\star})]$.

Next, we consider the estimation of $f(x_0)$. In fact, we estimate $f(t_l)$ using a local linear approximation. First, we identify a small interval around t_l and then approximate F over this interval by a line, whose slope gives the estimator of

$f(t_l)$. We determine the interval according to the following several requirements. First, the sample proportion p_n in the interval should be larger than the sample proportion at each grid point, which is of order $n^{-\gamma}$ for $\gamma \in (0,1)$. For example, setting p_n be of order $1/\log(n)$ theoretically ensures a sufficiently large interval. Second, for simplicity, we make the interval symmetric around t_l. Third, in order to obtain a positive estimate (since $f(t_l)$ is positive), we symmetrically enlarge the interval satisfying the above two requirements until the values of $\hat{F}$ at the two ends of the interval become different. More specifically, we first find $j^\star$, which is the smallest integer such that $\sum_{i=l-j^\star}^{l+j^\star} N_i/n \geq 1/\log(n)$. Finally, we find $i^\star$, which is the smallest integer larger than $j^\star$ such that $\hat{F}(t_{l-i^\star}) < \hat{F}(t_{l+i^\star})$. After identifying the interval $[t_{l-i^\star}, t_{l+i^\star}]$, we fit a line over this interval by weighted least squares. More specifically, we compute

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{(\beta_0,\beta_1)\in\mathbb{R}^2}{argmax} \left\{ \sum_{i=l-i^\star}^{l+i^\star} \left(\hat{F}(t_i) - \beta_0 - \beta_1 t_i \right)^2 N_i \right\},$$

and then estimate $f(t_l)$ (and $f(x_0)$) by $\hat{\beta}_1$. Once these nuisance parameters have been estimated, the practical adaptive procedure can be implemented.

To evaluate its finite sample performance in simulations, we also provide simulated confidence intervals of an idealized adaptive procedure where the true values of the parameters $F(x_0), g(x_0)$ and $f(x_0)$ are used, but γ is still practically assumed to be $1/3$ and c is taken as the previous $\hat{c}$. These confidence intervals can be considered as the best Wald-type confidence intervals based on the adaptive procedure.

The simulation settings are as follows: The sampling interval $[a,b]$ is $[0,1]$. The design density g is uniform on $[a,b]$. The distribution of T is the uniform distribution over $[a,b]$ or the exponential distribution with $\lambda = 1$ or 2. The point of interest x_0 is 0.5. The pair of two controlling parameters (γ, c) takes values (1/6, 1/6), (1/4, 1/4), (1/3, 1/2), (1/2, 1), (2/3, 2) or (3/4, 3). The sample size n is from 100 to 1000 by 100. When generating the quantiles of $\mathbb{X}_{\hat{c}}(0)$, K_a is set to be 300 and the corresponding iteration number 3000. We are interested in constructing 95% confidence intervals for $F(t_l)$. The iteration number for each simulation is 3000.

Denote the simulated coverage rates and average lengths for the practical procedure as CR(P) and AL(P) and those for the theoretical procedure as CR(T) and AL(T). Figure 2 contains the plots of CR(P), CR(T), AL(P) and AL(T) and Table 1 contains the corresponding numerical values for n being $n_1 = 100$, $n_2 = 300$ or $n_3 = 500$. The first plot of Figure 2 shows that both CR(T) and CR(P) are usually close to the nominal level 95% from below and CR(T) are generally about 1% better than CR(P). This reflects the price of not knowing the true values of the parameters $F(x_0)$, $g(x_0)$ and $f(x_0)$ in the practical procedure. On the other hand, the second plot of Figure 2 shows that the AL(P)s are usually slightly shorter than AL(T)s. This indicates that the practical procedure is slightly more aggressive. As the sample size increases, the coverage rates usually approach the nominal level and the average lengths also become shorter, as expected.

The patterns noted above show up in more extensive simulation studies, not shown here owing to constraints of space. Also, the adaptive procedure is seen to compete well with the asymptotic approximations that one would use for constructing CIs *were* γ *known*. Of course, for extreme values of γ (close to 0 or 1), the likelihood ratio based confidence intervals using the relevant asymptotic approximations (i.e. χ_1^2 in the small γ and $\mathbb{D}$ in the large γ settings) in the known γ case systematically outperform the adaptive ones and also enjoy the advantage of being constructed without nuisance parameter estimation but that is hardly surprising since extreme values of γ correspond to a 'black and white' situation, while moderate γ's correspond to the 'grey' area and pose greater challenges to estimation and it is here that the adaptive procedure is most useful.

7. Conclusions. In this paper, we have considered isotonic nonparametric estimation and hypothesis testing for the survival function at a point in the current status model with i.i.d. data. The design density for the covariate is assumed to be an equally spaced grid distribution with the grid resolution being $\delta = cn^{-\gamma}$ for $c > 0$ and $\gamma > 0$ and incorporates situations where there are systematic ties in the observation times of the entities involved and the number of distinct observation times can increase with the sample size.

The asymptotic properties of the isotonic regression estimator and the likelihood ratio test statistic depend critically on the order of the grid resolution γ. For $\gamma \in (0, 1/3)$, the asymptotic distributions are normal and chi-squared, which are the standard limit distributions in parametric problems with finite number of unknown parameters; for $\gamma \in (1/3, \infty)$, the asymptotic distributions are Chernoff and the so-called $\mathbb{D}$, which are the standard limit distributions in isotonic regression problems with continuous design densities. Thus, when $\gamma \in (0, 1/3)$, the grid is so sparse that the nonparametric problem is essentially reduced to a parametric one. On the other hand, when $\gamma \in (1/3, \infty)$, the grid is dense enough that the observation time can be viewed as an absolutely continuous random variable. For the most interesting boundary case with $\gamma = 1/3$, the grid is, in some sense, neither too sparse nor too dense and the asymptotic distributions are $\mathcal{S}_c$ and $\mathbb{M}_c$, functionals of the unconstrained and constrained GCMs of discrete time stochastic processes which depend on c, the scaling parameter in the grid resolution.

The limit distributions $\mathcal{S}_c$ and $\mathbb{M}_c$ are different from those obtained in the other two cases. However, as c goes to ∞, $\mathcal{S}_c$ and $\mathbb{M}_c$ converge in distribution to the normal and chi-squared distributions; as c goes to 0, $\mathcal{S}_c$ and $\mathbb{M}_c$ converge in distribution

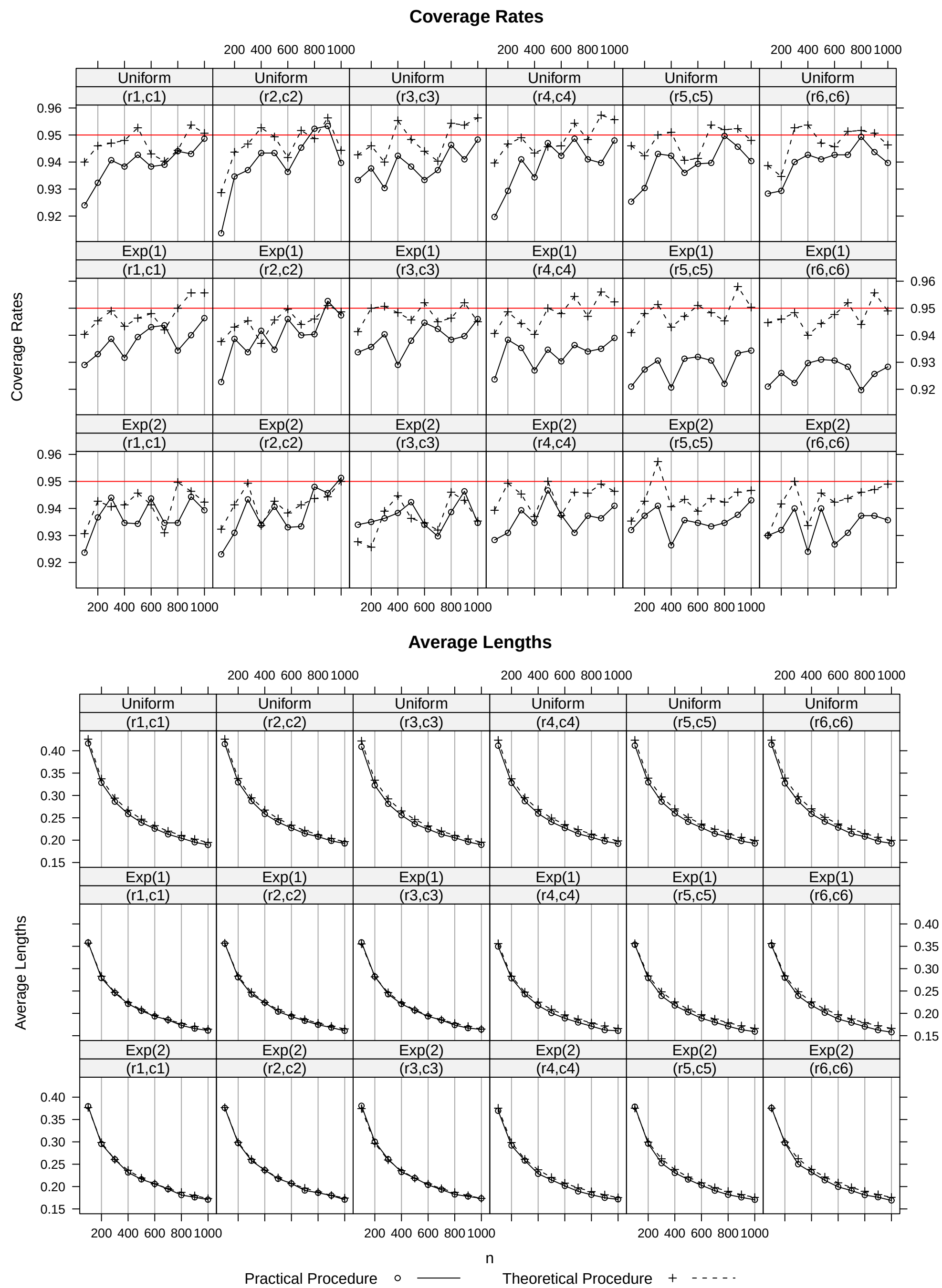

FIG 2. *A comparison of the coverage rates and average lengths of the practical and theoretical procedures, where (ri,ci) for* $i = 1, \cdots, 6$ *are (1/6, 1/6), (1/4, 1/4), (1/3, 1/2), (1/2, 1), (2/3, 2) or (3/4, 3), respectively. The sample size* n *varies from 100 to 1000 by 100.*

TABLE 1

A comparison of the coverage rates and average lengths of the practical procedure with those of the theoretical procedure, where $U[0,1]$ and $\exp(\lambda)$ stand for the uniform distribution over $[0,1]$ and the exponential distributions with the parameter λ, and n_1, n_2 and n_3 are 100, 300, and 500, respectively.

Coverage Rates									
CR(P)	$U[0,1]$			$\exp(1)$			$\exp(2)$		
(γ, c)	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
(1/6, 1/6)	.924	.941	.943	.929	.939	.939	.924	.944	.934
(1/4, 1/4)	.914	.937	.943	.923	.934	.935	.923	.943	.941
(1/3, 1/2)	.933	.930	.938	.934	.940	.938	.934	.936	.942
(1/2, 1)	.920	.941	.947	.924	.935	.935	.928	.939	.947
(2/3, 2)	.925	.943	.936	.921	.931	.931	.932	.941	.936
(3/4, 3)	.928	.940	.941	.921	.922	.931	.930	.940	.940
CR(T)	$U[0,1]$			$\exp(1)$			$\exp(2)$		
(γ, c)	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
(1/6, 1/6)	.940	.947	.953	.940	.949	.946	.931	.941	.946
(1/4, 1/4)	.929	.947	.949	.938	.945	.946	.932	.949	.943
(1/3, 1/2)	.943	.940	.948	.941	.951	.946	.928	.939	.936
(1/2, 1)	.940	.949	.946	.941	.944	.950	.939	.945	.950
(2/3, 2)	.946	.950	.941	.941	.951	.947	.935	.957	.943
(3/4, 3)	.939	.953	.947	.945	.948	.944	.930	.950	.946
Average Lengths									
AL(P)	$U[0,1]$			$\exp(1)$			$\exp(2)$		
(γ, c)	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
(1/6, 1/6)	.417	.286	.239	.358	.246	.206	.380	.261	.216
(1/4, 1/4)	.415	.287	.240	.356	.242	.204	.376	.258	.218
(1/3, 1/2)	.409	.281	.236	.359	.243	.207	.381	.258	.219
(1/2, 1)	.411	.287	.241	.350	.243	.201	.370	.258	.215
(2/3, 2)	.411	.286	.241	.354	.239	.202	.379	.253	.216
(3/4, 3)	.414	.287	.241	.352	.239	.202	.376	.250	.214
AL(T)	$U[0,1]$			$\exp(1)$			$\exp(2)$		
(γ, c)	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
(1/6, 1/6)	.426	.294	.247	.357	.247	.208	.377	.260	.219
(1/4, 1/4)	.426	.295	.248	.357	.247	.208	.377	.261	.220
(1/3, 1/2)	.422	.292	.246	.355	.246	.208	.374	.260	.219
(1/2, 1)	.424	.295	.249	.356	.247	.209	.375	.261	.220
(2/3, 2)	.424	.297	.251	.356	.248	.209	.375	.262	.221
(3/4, 3)	.424	.297	.251	.356	.248	.209	.375	.262	.221

to the Chernoff's and $\mathbb{D}$ distributions. These weak convergence results allow the approximation of the extreme distributions with the boundary ones, by adjusting the value of the scaler c, and lead to an adaptive procedure for statistical inferences which obviates the need to estimate or specify the order γ, and therefore provides a powerful inferential tool. Note that while we have considered grid resolutions of the order $n^{-\gamma}$, the derivations in this paper are easily generalizable to arbitrary grid resolutions that converge to 0 with n. So long as the resolution r_n satisfies $n^{-1/3} = o(r_n)$, standard asymptotics prevail while Chernoff-type asymptotics are obtained when $r_n = o(n^{-1/3})$. However the class of grid resolutions of order $n^{-\gamma}$ for $\gamma \in (0, \infty)$ is rich enough that almost any setting with tied observation times can be viewed as coming from such a scenario.

The results in this paper reveal some new directions for future research. As touched upon in the introduction, some recent related work by Maathuis and Hudgens (2010) deals with the estimation of *competing risks current status data* with discrete (or grouped) observation times. A natural question of interest, then, is what happens if the observation times in their paper are supported on grids of increasing size as considered in this paper for simple current status data. Does $\gamma = 1/3$ also form the boundary between normal and non-normal asymptotics as with the simpler current status model? Can an adaptive procedure of similar vein be devised for current status data with competing risks? One could also consider the problem of grouped current status data (with and without the element of competing risks), where the observation times are not exactly known but grouped into bins. Based on communications with us and preliminary versions of this paper, Maathuis and Hudgens (2010) conjecture that for grouped current status data *without competing risks*, one may expect findings similar to those in this paper, depending on whether the number of groups increases at rate $n^{1/3}$ or at a faster/slower rate. Whether similar phenomena would arise for grouped current status data with competing risks is again, unclear, though $\gamma = 1/3$ is certainly not an un-natural candidate for the transition from normality to non-normality.

Viewed as a regression model, the current status model is a monotone binary regression model and the results obtained here translate almost directly to results on monotone binary regression. In fact, it is fairly clear that the adaptive inference scheme proposed in this paper will apply to monotone regression models with discrete covariates in general. In particular, the very general conditionally parametric response models studied in Banerjee (2007) under the assumption of a continuous covariate can be handled for the discrete covariate case as well by adapting the methods of this paper and the adaptive procedure can be made to work similarly. Furthermore, similar adaptive inference in more complex forms of interval censoring, like Case-2 censoring or mixed-case censoring (see, for example, Sen and Banerjee (2006) and Schick and Yu (2000)), should also be possible in situations where the multiple observation times are discrete-valued. Finally, we conjecture that phenomena similar to those revealed in this paper will appear in nonparametric regression problems with grid-supported covariates under more complex shape constraints (ike convexity, for example), though the boundary value of γ as well as the nature of the non-standard limits will be different and will depend on the 'order' of the shape constraint.

Acknowledgements. We would like to thank Professors Jack Kalbfleisch and Nick Jewell for bringing this problem to our attention. The first author would also like to thank Professor George Michailidis for partial financial support while he was involved with the project.

APPENDIX A: PROOFS IN DETAIL

In this appendix, we provide proofs of some of the technical results.

A.1. Supplementary proofs for Case One $\gamma \in (0, 1/3)$. First, we establish two useful lemmas.

LEMMA A.1. *If $\gamma \in (0,1)$ and (**A1.2**) holds, for every $\eta \in (0,1)$, we have*

$$\mathbb{P}\left(\cap_{i=1}^K \{N_i \geq \eta m_l\}\right) \to 1.$$

PROOF. Denote the event of interest as A_n. Then, $A_n^c = \cup_{i=1}^K \{N_i < \eta m_l\}$ and it is sufficient to show $\mathbb{P}(A_n^c) \to 0$. Let $t = -\log(\eta) > 0$, we have

$$\begin{aligned}\mathbb{P}(N_i < \eta m_l) =&\mathbb{P}(\sum_{j=1}^n \{X_j = t_i\} < \eta m_l) \leq e^{t\eta m_l}\mathbb{P}e^{-t\sum_{j=1}^n\{X_j=t_i\}} \\ \leq& e^{t\eta m_l}\left[1-(1-e^{-t})g_l\delta\right]^n \leq e^{t\eta m_l}e^{-(1-e^{-t})g_l\delta n} \\ \leq& e^{m_l(\eta t - 1 + e^{-t})} = e^{m_l(-\eta\log(\eta)-1+\eta)}.\end{aligned}$$

The third inequality above exploits the fact $1 - x \leq e^{-x}$ for $x > 0$. Note that $-\eta\log(\eta) - 1 + \eta < 0$ for every $\eta \in (0,1)$, $K \sim (b-a)c^{-1}n^\gamma$ and $m_l = g_l c n^{1-\gamma}$. Then, we have

$$\mathbb{P}(A_n^c) \leq \sum_{i=1}^K \mathbb{P}(N_i < \eta m_l) \leq K e^{m_l(-\eta\log(\eta)-1+\eta)} \to 0,$$

which completes the proof. □

Next, we introduce two related binary regression models which yield sufficient statistics whose distributions are identical to the sufficient statistics in the current status model. The first model is as follows. Suppose $\{t_i\}_{i=1}^K$ and X are defined as before. Let $\{X_j\}_{j=1}^n$ be i.i.d. copies of X and $N_i = \sum_{j=1}^n \{X_j = t_i\}$ for $i = 1, 2, \cdots, K$. Given $\{N_i\}_{i=1}^K$, for each i draw an i.i.d. sample $\{Y_{ij}\}_{j=1}^{N_i}$ from $Bernoulli(1, F_i)$. Denote $\bar{Y}_i = N_i^{-1}\sum_{j=1}^{N_i} Y_{ij}$, for each i. The second model is as follows. Suppose $\{t_i\}_{i=1}^K$, X, $\{X_j\}_{j=1}^n$ and $\{N_i\}_{i=1}^K$ are defined as before. Let $\{Y'_{ij} : 1 \le i \le K, 1 \le j \le n\}$ be a family of mutually independent random variables, distributed independently of the variables in the previous sentence, such that for each i, Y'_{ij} follows $Bernoulli(1, F_i)$ for $1 \le j \le n$. Denote $\bar{Y}'_i = N_i^{-1} \sum_{j=1}^n Y'_{ij}\{X_j = t_i\}$ for each i. Then, we have the following equalities in distribution, which will be used in other proofs. Its proof can be shown by straightforward comparing distributions and omitted here.

LEMMA A.2. $\left(\{N_i\}, \{\bar{Z}_i\}\right) \stackrel{d}{=} \left(\{N_i\}, \{\bar{Y}_i\}\right) \stackrel{d}{=} \left(\{N_i\}, \{\bar{Y}'_i\}\right)$

PROOF OF PROPOSITION 4.9. It suffices to show $\mathbb{P}\left(\bar{Y}_1 \le \bar{Y}_2 \le \cdots \le \bar{Y}_K\right) \to 1$ by Lemma A.2, which is equivalent to show $\mathbb{P}\left(\cup_{i=1}^{K-1}\left\{\bar{Y}_i > \bar{Y}_{i+1}\right\}\right) \to 0$. Denote this probability as T. We then have $T = T_1 + T_2$, where

$$T_1 = \mathbb{P}\left(\cup_{i=1}^{K-1}\left\{\bar{Y}_i > \bar{Y}_{i+1}\right\}, \cap_{i=1}^K \{N_i \ge \eta m_l\}\right), \text{ with } \eta > 0$$

and $T_2 = T - T_1$. Since, by Lemma A.1, $T_2 \le \mathbb{P}\left(\cup_{i=1}^K \{N_i < \eta m_l\}\right) \to 0$, it remains to show $T_1 \to 0$.

On one hand, we have $T_1 \le \sum_{i=1}^{K-1} \sum_A S_i \mathbb{P}\left(N_i = n_i, i = 1, 2, \cdots, K\right)$, where

$$S_i = \mathbb{P}\left(\frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} > \frac{1}{n_{i+1}}\sum_{j=1}^{n_{i+1}} Y_{(i+1)j}\right)$$

and $A = \{(n_1, n_2, \cdots, n_K) \in \mathbb{N}^K : \sum_{i=1}^K n_i = n, n_i \ge \eta m_l \text{ for } i = 1, 2\cdots, K\}$. On the other hand, we will show that, for each $i = 1, \cdots, K-1$,

$$S_i \le 2\exp\left\{-\frac{\eta f_l^2}{16}\delta^2 m_l\right\}. \tag{A.1}$$

Then, the result follows by noticing that $\delta^2 m_l \sim c^3 g_l n^{1-3\gamma} \to \infty$ for $\gamma \in (0, 1/3)$.

Next, we show (A.1). Denote $Z_{ij} = Y_{ij} - F_i$, $Z'_{ij} = -Z_{ij}$ for $i = 1, \cdots, K$ and $j = 1, \cdots, N_i$ and $\Delta_i = F_{i+1} - F_i$ for $i = 1, \cdots, K-1$. Then we have

$$S_i \le \mathbb{P}\left(\frac{1}{n_i}\sum_{j=1}^{n_i} Z_{ij} > \frac{\Delta_i}{2}\right) + P\left(\frac{1}{n_{i+1}}\sum_{j=1}^{n_{i+1}} Z'_{(i+1)j} > \frac{\Delta_i}{2}\right) =: S_{i1} + S_{i2}.$$

In a similar way, it can be shown that both S_{i1} and S_{i2} are less than or equal to $\exp\left\{-\eta f_l^2 \delta^2 m_l / 16\right\}$. Next, we only show the inequality on S_{i1}.

For every $t > 0$, we have

$$S_{i1} \leq e^{-\frac{\Delta_i}{2} n_i t} (\mathbb{P} e^{t Z_{i1}})^{n_i} \leq \exp\{-\phi(t) n_i\},$$

where $\phi(t) = t(F_i + F_{i+1})/2 - F_i(e^t - 1)$. Note that the maximizer of $\phi(t)$ is $t_i^\star = \log((F_i + F_{i+1})/2F_i) > 0$ and that the maximum of ϕ is

$$\phi(t_i^\star) = \frac{F_i + F_{i+1}}{2} \log \frac{F_i + F_{i+1}}{2F_i} - \frac{F_{i+1} - F_i}{2}.$$

Thus, we have $S_{i1} \leq \exp\{-\phi(t_i^\star) n_i\}$. For $x \in [t_1, t_{K-1}]$, define

$$h(x) = \frac{F(x) + F(x+\delta)}{2} \log \frac{F(x) + F(x+\delta)}{2F(x)} - \frac{F(x+\delta) - F(x)}{2}.$$

Then, we have $\phi(t^\star) \geq \inf_{x \in [t_1, t_{K-1}]} h(x)$. Since $n_i \geq \eta m_l$, it is sufficient to show

$$\inf_{x \in [t_1, t_{K-1}]} h(x) \geq (f_l \delta)^2/16. \tag{A.2}$$

By the assumption (**A1.1**), there exists $\xi \in (x, x+\delta)$ such that $F(x+\delta) = F(x) + f(\xi)\delta$. Then, we have

$$\begin{aligned}
h(x) &= \left(F(x) + \frac{f(\xi)}{2}\delta\right) \log\left(1 + \frac{f(\xi)}{2F(x)}\delta\right) - \frac{f(\xi)}{2}\delta \\
&= \left(F(x) + \frac{f(\xi)}{2}\delta\right)\left(\frac{f(\xi)}{2F(x)}\delta - \frac{1}{2(1+\zeta)^2}\frac{f(\xi)^2}{4F(x)^2}\delta^2\right) - \frac{f(\xi)}{2}\delta \\
&= \frac{f^2(\xi)}{4F(x)}\delta^2 - \frac{1}{(1+\zeta)^2}\frac{f^2(\xi)}{8F(x)}\delta^2 - \frac{1}{(1+\zeta)^2}\frac{f^3(\xi)}{16F^2(x)}\delta^3 \\
&\geq \frac{f^2(\xi)}{4F(x)}\delta^2 - \frac{f^2(\xi)}{8F(x)}\delta^2 - \frac{f^3(\xi)}{16F^2(x)}\delta^3 \geq \frac{f^2(\xi)}{16F(x)}\delta^2 \geq \frac{f_l^2}{16}\delta^2,
\end{aligned}$$

where the Taylor's expansion of $\log(1+x)$ around 0 is utilized in the second equality and $\zeta \in (0, f(\xi)\delta/(2F(x)))$; the assumptions (**A1.1**) and (**A1.3**) and $F(x) \leq 1$ are exploited for the last two inequalities. □

PROOF OF PROPOSITION 4.10. The proof follows very similar lines as that of Proposition 4.9 and is therefore skipped. □

PROOF OF PROPOSITION 4.11. By Proposition A.2, it suffices to show the results with $\bar{Y}_l'$ and $\bar{Y}_r'$ replacing $\bar{Z}_l$ and $\bar{Z}_r$. We have $\bar{Y}_l' - F_l = T_1/T_2$, where $T_1 = n^{-1}\sum_{j=1}^n p_l^{-1} 1(X_j = t_l)(Y_{lj}' - F_l)$ and $T_2 = n^{-1}\sum_{j=1}^n p_l^{-1} 1(X_j = t_l)$. By Chebyshev's Inequality, it is easy to check that T_1 and T_2 converge to 0 and 1

in probability, respectively. Then $\bar{Z}_l - F_l$ converges to 0 in probability by Slutsky's Lemma. Similarly, $\bar{Z}_r - F_r$ converges to 0 in probability.

Next, we show weak convergence. Denote $Z_{nj} = (Z_{njl}, Z_{njr})$, where $Z_{njl} = p_l^{-1/2} 1(X_j = t_l)(Y'_{lj} - F_l)/\sqrt{F_l(1-F_l)}$ and $Z_{njr} = p_r^{-1/2} 1(X_j = t_r)(Y'_{rj} - F_r)/\sqrt{F_r(1-F_r)}$ for $j = 1, 2, \cdots, n$. Then $\{Z_{nj}\}$ are independent and identically distributed. It is easy to check that the means of Z_{njl} and Z_{njr} are 0, the variances are 1 and the covariance is 0. Then, by a triangular array version of multivariate central limit theorem (see Proposition 2.27 in van der Vaart (1998)), in order to show $n^{-1/2}\sum_{i=1}^n Z_{nj} \xrightarrow{d} N(0, I_2)$, it is sufficient to check the Lindeberg condition: for each $\epsilon > 0$,

$$\sum_{j=1}^n E||n^{-1/2}Z_{nj}||^2\{||n^{-1/2}Z_{nj}|| > \epsilon\} \to 0.$$

Since

$$\sum_{j=1}^n E||n^{-1/2}Z_{nj}||^2\{||n^{-1/2}Z_{nj}|| > \epsilon\} \leq \frac{1}{n\epsilon^2}E||Z_{n1}||^4 = \frac{1}{n\epsilon^2}E[Z_{n1l}^4 + Z_{n1r}^4],$$

$$E[Z_{n1l}^4] = \frac{[F_l^3 + (1-F_l)^3]}{p_l F_l(1-F_l)} \leq \frac{4}{g_l \delta F(x_0)(1-F(x_0))},$$

and similar inequality holds for $E[Z_{n1r}^4]$, we have

$$\sum_{j=1}^n E||n^{-1/2}Z_{nj}||^2\{||n^{-1/2}Z_{nj}|| > \epsilon\} \leq \frac{8}{\epsilon^2 g_l F(x_0)(1-F(x_0))} \cdot \frac{1}{cn^{1-\gamma}} \to 0.$$

Thus, the Lindeberg condition holds.

Denote $T_n = (T_{nl}, T_{nr})$, where $T_{nl} = N_l^{1/2}(\bar{Y}'_l - F_l)$ and $T_{nr} = N_r^{1/2}(\bar{Y}'_r - F_r)$. Then, we have $T_n = \beta(x_0)n^{-1/2}\sum_{i=1}^n Z_{nj} + R_n$, where

$$R_n = \left((\eta_{nl} - \beta(x_0)) \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{njl}, (\alpha_{nr} - \beta(x_0)) \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{njr} \right),$$
$$\beta(x_0) = \sqrt{F(x_0)(1-F(x_0))}, \quad \alpha_{nl} = [F_l(1-F_l)]^{1/2}/[N_l/(np_l)]^{1/2},$$

and α_{nr} is similarly defined. Then, by Slutsky's Lemma, it suffices to show both $\alpha_{nl} - \beta(x_0)$ and $\alpha_{nr} - \beta(x_0)$ converge to 0 in probability. Here we only show the former and the latter can be shown in a similar manner. From the continuity of F at x_0, $[F_l(1-F_l)]^{1/2} \to \beta(x_0)$. Notice that $N_l/(np_l)$ is exactly T_2 and $N_l/(np_l) \to 1$ in probability. This completes the proof. $\square$

PROOF OF LEMMA 4.6. By the definition of $\tilde{F}(x_0)$, we have

$$\begin{aligned}\tilde{F}(x_0) - F(x_0) &= \frac{t_r - x_0}{t_r - t_l}(F_l^\star - F_l) + \frac{x_0 - t_l}{t_r - t_l}(F_r^\star - F_r) \\ &+ \frac{t_r - x_0}{t_r - t_l}(F_l - F(x_0)) + \frac{x_0 - t_l}{t_r - t_l}(F_r - F(x_0)).\end{aligned}$$

By Lemma 4.1, both $F_l^\star - F_l$ and $F_r^\star - F_r$ converge to 0 in probability. For the continuity of f at x_0, both $F_l - F(x_0)$ and $F_r - F(x_0)$ converge to 0. Note that the absolute values of the coefficients are less than 1. Therefore, $\tilde{F}(x_0) - F(x_0)$ converges to 0 in probability. □

PROOF OF THEOREM 4.7. Denote $\tilde{p}_n = p_n/(p_n^2 + q_n^2)^{1/2}$, $\tilde{q}_n = p_n/(p_n^2 + q_n^2)^{1/2}$, $\alpha_l = [p_l F_l(1 - F_l)]^{1/2}$, $\alpha_r = [p_r F_r(1 - F_r)]^{1/2}$, $\xi_{lj} = \tilde{q}_n(Y_{lj} - F_l)\{X_j = t_l\}/\alpha_l$, $\xi_{rj} = \tilde{p}_n(Y_{rj} - F_r)\{X_j = t_r\}/\alpha_r$. Then, we have

$$\frac{1}{\sqrt{p_n^2 + q_n^2}}(\tilde{F}(x_0) - F(x_0)) = [\alpha_l \frac{n^{1/2}}{N_l} T_1 + (R_{n1} + R_{n2})]A_n + \tilde{I}A_n^c + \frac{1}{\sqrt{p_n^2 + q_n^2}} II,$$

where $\tilde{I} = \tilde{q}_n(F_l^\star - F_l) + \tilde{p}_n(F_r^\star - F_r)$, $II = q_n F_l + p_n F_r - F(x_0)$, $T_1 = n^{-1/2}\sum_{j=1}^n(\xi_{lj} + \xi_{rj})$, $R_{n1} = (\alpha_r - \alpha_l)N_r^{-1}\sum_{j=1}^n \xi_{lj}$ and $R_{n2} = \alpha_l(N_r^{-1} - N_l^{-1})\sum_{j=1}^n \xi_{rj}$. When $\gamma \in (0, 1/3)$ and (**A1**) hold, by regular argument, T_1 converges to $N(0,1)$ in distribution; $n^{(1-\gamma)/2}\alpha_l n^{1/2}/N_l$ and $n^{(1-\gamma)/2}(R_{n1} + R_{n2})$ converge to $[F(x_0)(1 - F(x_0))/(cg(x_0))]^{1/2}$ and 0 in probability, respectively.

Since f'' is bounded in a neighborhood of x_0, we have

$$\begin{aligned}F_l &= F(x_0) + f(x_0)(t_l - x_0) + \frac{1}{2}f'(x_0)(t_l - x_0)^2 + o((t_l - x_0)^2), \\ F_r &= F(x_0) + f(x_0)(t_r - x_0) + \frac{1}{2}f'(x_0)(t_r - x_0)^2 + o((t_r - x_0)^2).\end{aligned}$$

Denote $T_2 = (p_n^2 + q_n^2)^{-1/2} n^{(1-\gamma)/2} II$. Then,

$$\begin{aligned}T_2 &= \frac{1}{2\sqrt{p_n^2 + q_n^2}} f'(x_0) n^{(1-\gamma)/2}(t_r - x_0)(x_0 - t_l) \\ &+ o(q_n n^{(1-\gamma)/2}(t_l - x_0)^2 \vee p_n n^{(1-\gamma)/2}(t_r - x_0)^2).\end{aligned}$$

Thus, for $\gamma \in (1/5, 1/3)$, we have $T_2 \to 0$; for $\gamma = 1/5$ and p_n converging to some $p \in (0,1)$, we have $T_2 \to (1/2)pq(p_n^2 + q_n^2)^{-1/2}c^2 f'(x_0)$; Further, for $\gamma \in (0, 1/5)$ and $f'(x_0) \neq 0$, we have $T_2 \to \text{Sign}(f'(x_0))\infty$. Finally, noticing that both np_l/N_l and np_r/N_r converge to 1 in probability and that both p_l and p_r are asymptotically equivalent to $g(x_0)cn^{-\gamma}$ completes the proof. □

A.2. Proofs for Case Three $\gamma = 1/3$. In what follows we will denote the distribution of the vector (X_{1n}, Y_{1n}) by $\mathbb{P}$ (suppressing the dependence on n) and $\mathbb{P}_n$ will denote the empirical measure of the sample, i.e. n i.i.d. observations drawn from $\mathbb{P}$.

LEMMA A.3. *Fact 1 in the proof of Lemma 4.23 holds.*

PROOF. For $\beta = 0$, Fact 1 holds obviously. Next, fix a small interval $\mathcal{N}_0$ about 0 such that f, g, f', g' are all uniformly bounded in this neighborhood. Let $\epsilon > 0$. Consider $\beta > 0, \beta \in \mathcal{N}_0$. We have

$$
\begin{aligned}
\mathbb{P} g_{n1}(x, y; \beta) &= \mathbb{P}_X \mathbb{P}_{Y|X} \left(Y - F(t_{l-M})\right) \{t_{l-M} < X \le t_{l-M} + \beta\} \\
&= \sum_{j=1}^{\lfloor \beta/\delta \rfloor} \left(F(t_{l-M+j}) - F(t_{l-M})\right) \left(G(t_{l-M+j}) - G(t_{l-M+j-1})\right) \\
&= \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j f(t^{1\star}_{l-M+j}) g(t^{2\star}_{l-M+j}) = T_1 + T_2
\end{aligned}
$$

where the mean value theorem is applied twice at the third equality, $t^{1\star}_{l-M+j}$ lies in $[t_{l-M}, t_{l-M+j}]$, $t^{2\star}_{l-M+j}$ lies in $[t_{l-M+j-1}, t_{l-M+j}]$, $T_1 = \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j f(x_0) g(x_0)$ and $T_2 = \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j \left[f(t^{1\star}_{l-M+j}) g(t^{2\star}_{l-M+j}) - f(x_0) g(x_0) \right]$. We will show that for β sufficiently small (depending on ϵ) both $|T_1 - (1/2) f(x_0) g(x_0) \beta^2|$ and $|T_2|$ are dominated by $\epsilon \beta^2 + O(n^{-2/3})$ when n is sufficiently large (depending on ϵ), where the $O(n^{-2/3})$ terms can again depend on ϵ. Then the result follows, as ϵ is arbitrary.

First consider T_1. We have $T_1 - (1/2) f(x_0) g(x_0) \beta^2 = (1/2) f(x_0) g(x_0) T_{11}$ where $T_{11} = \delta^2 \lfloor \beta/\delta \rfloor (1 + \lfloor \beta/\delta \rfloor) - \beta^2$. Since $\lfloor x \rfloor \in (x - 1, x]$ for $x \in \mathbb{R}$, simple algebra gives $|T_{11}| \le \beta\delta$ and consequently

$$
|T_1 - (1/2) f(x_0) g(x_0) \beta^2| \le (f(x_0) g(x_0)/2) \beta \delta \le \epsilon \beta^2 + \tilde{\delta}^2/\epsilon = \epsilon \beta^2 + O(n^{-2/3}),
$$

where $\tilde{\delta} = (f(x_0) g(x_0)/2)\, \delta$ and we use the fact that $\delta = c\, n^{-1/3}$ to get the last expression in the above display. Note that the $O(n^{-2/3})$ term depends on ϵ.

Next consider T_2. We have $|T_2| \le T_{21} + T_{22}$, where

$$
\begin{aligned}
T_{21} &= \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j |f(t^{1\star}_{l-M+j}) - f(x_0)| g(t^{2\star}_{l-M+j}), \\
T_{22} &= \delta^2 \sum_{j=1}^{\lfloor \beta/\delta \rfloor} j |g(t^{2\star}_{l-M+j}) - g(x_0)| f(x_0).
\end{aligned}
$$

Here we only show $T_{21} \leq \epsilon\beta^2 + O(n^{-2/3})$ and $T_{22} \leq \epsilon\beta^2 + O(n^{-2/3})$ can be shown similarly. We have

$$\begin{aligned}
T_{21} &\lesssim \delta^2 \sum_{j=1}^{\lfloor\beta/\delta\rfloor} j|f(t^{1\star}_{l-M+j}) - f(x_0)| = \delta^2 \sum_{j=1}^{\lfloor\beta/\delta\rfloor} j|f'(t^{1\star\star}_{l-M+j})||t^{1\star}_{l-M+j} - x_0| \\
&\lesssim \delta^2 \sum_{j=1}^{\lfloor\beta/\delta\rfloor} j|t^{1\star}_{l-M+j} - x_0| \lesssim \delta^2 \sum_{j=1}^{\lfloor\beta/\delta\rfloor} j\lfloor\beta/\delta\rfloor\delta + \delta^2 \sum_{j=1}^{\lfloor\beta/\delta\rfloor} j(M+1)\delta.
\end{aligned}$$

Denote the above last two terms as T_{211} and T_{212}. The mean value theorem is applied at the second step and $t^{1\star\star}_{l-M+j}$ lies between $t^{1\star}_{l-M+j}$ and x_0. In the first and third steps, the assumption that g and f' are bounded around x_0 is utilized. Denote the constant associated with the $\lesssim$ by $\tilde{K}$. Then $T_{21} \leq T_{211} + T_{212}$ with $T_{211} \leq \tilde{K}\delta^3\lfloor\beta/\delta\rfloor^2(1+\lfloor\beta/\delta\rfloor) \leq \tilde{K}\delta^3(\beta/\delta)^2(\beta/\delta+1) = \tilde{K}\,\beta^3 + \tilde{K}\beta^2\delta \leq \epsilon\beta^2$, by choosing $\beta < \epsilon/\tilde{K}$ and n sufficiently large (depending on ϵ), and $T_{212} \leq \tilde{K}\delta^3\lfloor\beta/\delta\rfloor(1+\lfloor\beta/\delta\rfloor) \leq \tilde{K}\beta^2\delta + \tilde{K}\beta\delta^2 \leq \epsilon\beta^2 + O(n^{-2/3})$, again for n sufficiently large (depending only on ϵ). Thus, $T_{21} \leq \epsilon\beta^2 + O(n^{-2/3})$, which completes the proof. ☐

LEMMA A.4. *Fact 2 in the proof of Lemma 4.23 holds.*

PROOF. For $\beta = 0$, Fact 2 holds obviously. Next, suppose $\beta > 0$ and is restricted to the neighborhood $\mathcal{N}_0$ from the proof of the previous lemma. The case $\beta < 0$ can be handled in the same way. We have

$$\begin{aligned}
\mathbb{P}g_{n2}(x;\beta) &= \mathbb{P}\left\{t_{l-M} < X \leq t_{l-M} + \beta\right\} = \sum_{j=1}^{\lfloor\beta/\delta\rfloor} \left(G(t_{l-M+j}) - G(t_{l-M+j-1})\right) \\
&= \sum_{j=1}^{\lfloor\beta/\delta\rfloor} g(x_0)\delta + \sum_{j=1}^{\lfloor\beta/\delta\rfloor} \left(g(t^{\star}_{l-M+j}) - g(x_0)\right)\delta,
\end{aligned}$$

where the mean value theorem is applied at the last step and $t^{\star}_{l-M+j}$ lies between $[t_{l-M+j-1}$ and $t_{l-M+j}]$. Denote the above last two terms as T_1 and T_2.

We have $|T_1 - g(x_0)\beta| \leq g(x_0)|\delta\lfloor\beta/\delta\rfloor - \beta| \leq g(x_0)\delta = O(n^{-1/3})$. Next, consider T_2. Similar to the argument involving T_{21} in the proof of Lemma A.3, for a constant K' depending only on $\mathcal{N}_0$, we have: we have

$$|T_2| \leq K'\delta \sum_{j=1}^{\lfloor\beta/\delta\rfloor} |t^{\star}_{l-M+j} - x_0| \leq K'\delta \sum_{j=1}^{\lfloor\beta/\delta\rfloor} j\delta + K'\delta \sum_{j=1}^{\lfloor\beta/\delta\rfloor} (M+1)\delta.$$

Denote the above last two terms as S_1 and S_2. We have $S_1 \leq K'\delta^2\lfloor\beta/\delta\rfloor(1+\lfloor\beta/\delta\rfloor) \leq K'\beta^2 + K'\beta\delta \leq K'\beta^2 + O(n^{-1/3})$ and $S_2 \leq K'\delta^2\lfloor\beta/\delta\rfloor \leq K'\beta\delta \leq O(n^{-1/3})$. Thus, $T_2 \leq K'\beta^2 + O(n^{-1/3})$ and Fact 2 now follows. □

The proof of the next lemma follows the lines of the proof of Lemma 4.1 in Kim and Pollard (1990).

LEMMA A.5. *Fact 3 in the proof of Lemma 4.23 holds.*

PROOF. Denote $R_0 > 0$ as a small constant. Let, for each $m > 0$,

$$\begin{aligned}
M_n &= \sup_{|\beta|\leq R_0} n^{2/3}\left(|(\mathbb{P}_n - \mathbb{P})\, g_{n1}(x,y;\beta)| - \epsilon\beta^2\right),\\
E &= \left\{\exists\beta \in [\pm R_0] \text{ s.t. } |(\mathbb{P}_n - \mathbb{P})\, g_{n1}(x,y;\beta)| > \epsilon\beta^2 + n^{-2/3}m^2\right\},\\
A_{n,j} &= \left\{|\beta| \leq R_0 : (j-1)n^{-1/3} \leq |\beta| \leq jn^{-1/3}\right\},\\
B_j &= \left\{\exists\beta \in A_{n,j} \text{ s.t. } n^{2/3}\,|(\mathbb{P}_n - \mathbb{P})\, g_{n1}(x,y;\beta)| > \epsilon(j-1)^2 + m^2\right\}.
\end{aligned}$$

Then, we have $\mathbb{P}\left(M_n > m\right) \leq \mathbb{P}E \leq \sum_{j=1}^{\infty}\mathbb{P}B_j$. By Chebyshev's Inequality,

$$\mathbb{P}B_j \leq n^{4/3}\mathbb{P}\sup_{|\beta|<jn^{-1/3}} |(\mathbb{P}_n - \mathbb{P})\, g_{n1}(x,y;\beta)|^2 \,/\, \left[\epsilon(j-1)^2 + m^2\right]^2.$$

Denote $\mathscr{G}_{n1}(R) = \{g_{n1}(x,y;\beta) : |\beta| \leq R\}$ for $0 < R \leq R_0$. Then one envelope function is $F_{n1}(x,y;R) = \{x \in [t_{l-M} \pm R]\}$ by noticing $|y - F(t_{l-M})| \leq 1$. Straight forward calculation gives $\mathbb{P}F_{n1}^2(R) \leq C_1 R$ for n large enough, where C_1 can be any constant greater than $2g(x_0)$, say $3g(x_0)$. Note that, for every $R \in (0, R_0]$, $\mathscr{G}_{n1}(R)$ is a bounded VC-class of functions with VC-dimension bounded by a constant (independent of n). It follows readily that $J(1, \mathscr{G}_{n1}(R))$ is finite and uniformly bounded in n (see Page 239 of van der Vaart and Wellner (1996)). Theorem 2.14.1 of van der Vaart and Wellner (1996) with $\mathcal{F}$ taken to be $\mathscr{G}_{n1}(jn^{-1/3})$ and $p = 2$ now yields that for some constant C_2 and $C = C_1C_2$

$$\mathbb{P}\sup_{|\beta|<jn^{-1/3}} |(\mathbb{P}_n - \mathbb{P})\, g_{n1}(x,y;\beta)|^2 \leq n^{-1}C_2\mathbb{P}F_{n1}^2(jn^{-1/3}) \leq Cjn^{-4/3}.$$

Thus, we have $\mathbb{P}\left(M_n > m\right) \leq \sum_{j=1}^{\infty}\mathbb{P}B_j \leq \sum_{j=1}^{\infty} Cj/\left[\epsilon(j-1)^2 + m^2\right]^2$. The last term converges for each $m > 0$ and goes to 0 as $m \to \infty$ by the Dominated Convergence Theorem. Therefore $M_n = O_P(1)$, which completes the proof. □

LEMMA A.6. *Fact 4 in the proof of Lemma 4.23 holds.*

PROOF. The proof is the same to that of Lemma A.5 except changing subscripts $n1$ to $n2$. Note that the same envelope function is used. $\square$

LEMMA A.7. *Claim 2 in the proof of Theorem 4.17 holds.*

PROOF. Without loss of generality, here we only prove for the first equality. Denote $\mathbb{K}(i) = GCM\{\mathcal{P}_c(k), k \in \mathbb{Z}\}(ci)$ for $i \in \mathbb{Z}$. Let L and U be the largest integer less than $-M$ and the smallest integer larger than M such that $\mathbb{K}$ change slopes at cL and cU. Then, the first equality holds. Thus, it is sufficient to show both L and U are $O_P(1)$. We show the latter and the former can be shown in the same way.

Denote $A_j = \{\mathbb{K}(M) + \mathbb{X}(M)(j - M) = \mathcal{P}_{2,c}(j)\}$ for $j \geq M$. Then,

$$\{U = +\infty\} \subset \{A_j, i.o. \text{ for } j \geq M\} =: B.$$

On the other hand, we have

$$B \subset \left\{ \limsup_{t \to +\infty} \frac{\alpha W(t) + \beta t(c+t)}{t} = +\infty \right\}^c =: D^c.$$

From $\mathbb{P}\left(\lim_{t \to +\infty} W(t)/t = 0\right) = 1$, the law of large number for a standard Brownian motion, we have $\mathbb{P}(D) = 1$ by noticing $\beta > 0$. Therefore, $\mathbb{P}(U = +\infty) = 0$, which completes the proof. $\square$

LEMMA A.8. *The convergence in probability (4.14) in the proof of Lemma 4.24 holds.*

The proof is fairly straightforward and therefore skipped.

LEMMA A.9. *The weak convergence (4.15) in the proof of Lemma 4.24 holds.*

PROOF. It suffices to show that for each integer $C > 0$,

$$\{V_n^\star(ck), k \in [-C, C]\} \xrightarrow{d} \{\mathcal{P}_{2,c}, k \in [-C, C]\}.$$

We have $V_n^\star(ck) = T_1(ck) + T_2(ck)$ for $k \in [-C, C]$, where

$$T_1(ck) = g(x_0)^{-1} n^{2/3} (\mathbb{P}_n - \mathbb{P})(y - F(t_l)) \left(\left\{ x \leq t_l + kcn^{-1/3} \right\} - \{x \leq t_l\} \right),$$
$$T_2(ck) = g(x_0)^{-1} n^{2/3} \mathbb{P}(y - F(t_l) \left(\left\{ x \leq t_l + kcn^{-1/3} \right\} - \{x \leq t_l\} \right).$$

Then it is sufficient to show

$$\{T_1(ck), k \in [-C, C]\} \xrightarrow{d} \{\alpha W(ck), k \in [-C, C]\},$$
$$\{T_2(ck), k \in [-C, C]\} \to \left\{ \beta c^2 k(1+k), k \in [-C, C] \right\}.$$

It is easy to show that $T_2(ck) \to \beta c^2 k(1+k)$ uniformly for for $k \in [\pm C]$. We, therefore, consider T_1. Set $\xi_{i,n} = (\xi_{i,n,-C}, \cdots, \xi_{i,n,0}, \cdots, \xi_{i,n,C})$ with $\xi_{i,n,k} = g(x_0)^{-1} n^{-1/3} (Y_{i,n} - F(t_l)) \left(\{X_{i,n} \le t_l + kcn^{-1/3}\} - \{X_{i,n} \le t_l\} \right)$ for each $i \in [1,n]$ and $k \in [\pm C]$. Then $\{T_1(ck), k \in [-C, C]\} = \sum_{i=1}^n (\xi_{i,n} - \mathbb{P}\xi_{i,n})$. Note that $\{\xi_{i,n}\}$ is a row independent triangular array and each element is a $2C+1$ dimensional vector. By the triangular version of the central limit theorem (see, for example, Proposition 2.27 of van der Vaart (1998)), it is sufficient to show that $\xi_{i,n}$ has a finite variance for each i, $\sum_{i=1}^n \mathbb{P}||\xi_{i,n}||^2 \{||\xi_{i,n}|| > \epsilon\} \to 0$ for each $\epsilon > 0$, and $\sum_{i=1}^n Cov(\xi_{i,n})$ converges to the covariance matrix of $\{\alpha W(ck), k \in [\pm C]\}$. Since $\mathbb{P}\xi_{i,n,k}^2 \le g(x_0)^{-2} n^{-2/3} \le g(x_0)^{-2}$ by noting both $Y_{i,n}$ and $F(t_l)$ are bounded by 1, we know $\xi_{i,n}$ has finite variance. Further, we have

$$\sum_{i=1}^n \mathbb{P}||\xi_{i,n}||^2 \{||\xi_{i,n}|| > \epsilon\} \le \frac{1}{\epsilon^2} \sum_{i=1}^n \mathbb{P}||\xi_{i,n}||^4$$
$$= \frac{1}{\epsilon^2} \sum_{i=1}^n \mathbb{P} \left(\sum_{k=-C}^{C} \xi_{i,n,k}^2 \right)^2 \le \frac{1}{\epsilon^2 g(x_0)^4} (2C+1)^2 n^{-1/3} \to 0.$$

That $\sum_{i=1}^n Cov(\xi_{i,n,k_1}, \xi_{i,n,k_2})$ converges to $Cov(\alpha W(ck_1, ck_2))$ for $k_1, k_2 \in [\pm C]$ follows by direct calculation. This completes the proof. $\square$

REMARK A.10. In the proofs of Lemma A.8 and Lemma A.9 we exploit the fact that, for the boundary case with $\gamma = 1/3$, we only need to consider finite dimensional random vectors $\{G_n^\star(ck), k \in [\pm C]\}$ and $\{V_n^\star(ck), k \in [\pm C]\}$ for every integer $C > 0$. However, for the case with $\gamma \in (1/3, 1)$, the dimension goes to ∞ as n goes to ∞. Then the usual Chebyshev's inequality and the central limit theorem for a triangular array of random vectors will not work and more powerful tools from empirical processes need to be used.

LEMMA A.11. *In the proof of Theorem 4.20, $S_n \rightsquigarrow S$ holds.*

PROOF. The same truncation technique illustrated in the proof of Theorem 4.17 is utilized here again. More specifically, We will show that, in the following Lemmas A.12 and A.13, the following two claims hold:

Claim 1: Both (nonnegative) L_n and U_n are $O_P(1)$.
Claim 2: Both (nonnegative) L and U are $O_P(1)$.

Then, for each small $\epsilon > 0$, there exists (integer) M_ϵ large enough such that $P(M_\epsilon > \max\{L_n, U_n, L, U\}) > 1 - \epsilon$. Denote

$$S_n^{M_\epsilon} = \sum_{j \in [\pm M_\epsilon]} \left(\mathbb{X}_n^2(cj) - \mathbb{Y}_n^2(cj) \right), \quad S^{M_\epsilon} = \sum_{j \in [\pm M_\epsilon]} \left(\mathbb{X}^2(cj) - \mathbb{Y}^2(cj) \right).$$

Then, we will show the following facts:

Fact 1: $\lim_{\epsilon\to 0} \overline{\lim}_{n\to\infty} \mathbb{P}\left(\mathbb{S}_n^{M_\epsilon} \neq \mathbb{S}_n\right) = 0.$
Fact 2: $\lim_{\epsilon\to 0} \mathbb{P}\left(\mathbb{S}^{M_\epsilon} \neq \mathbb{S}\right) = 0.$
Fact 3: $\mathbb{S}_n^{M_\epsilon} \rightsquigarrow \mathbb{S}^{M_\epsilon}$, as $n \to \infty$ for each $\epsilon > 0$.

Fact 1 and Fact 2 hold since that $\{M_\epsilon > \max\{|L_n|, U_n, |L|, U\}\}$ is a subset of both $\left\{S_n^{M_\epsilon} = S_n\right\}$ and $\left\{S^{M_\epsilon} = S\right\}$. By Theorem 4.17 and the Continuous Mapping Theorem, Fact 3 follows. Therefore, by Lemma 4.18, we have $S_n \rightsquigarrow S$. $\square$

LEMMA A.12. *Claim 1 in Lemma A.11 holds.*

PROOF. Let $-cL_n'$ be the smallest grid point less than 0 such that $\mathbb{X}_n(-cL_n')$ is equal to $\mathbb{X}_n(0)$ and cR_n' the largest grid point larger than or equal to 0 such that $\mathbb{X}_n(cR_n')$ is equal to $\mathbb{X}_n(0)$. Let $-cL_n''$ be the smallest grid point less than 0 such that $\mathbb{X}_n(-cL_n'') > 0$ and cR_n'' the largest grid point greater than 0 such that $\mathbb{X}_n(cR_n'') < 0$. If there does not exist such a grid point $-cL_n''$ or cR_n'', let L_n'' or R_n'' be any nonnegative integer, say 0. Note that L_n', R_n', L_n'' and R_n'' are all nonnegative.

Then, $\mathbb{X}_n$ and $\mathbb{Y}_n$ differ at most over $[-cL_n' - cL_n'', cR_n' + cR_n'']$. From the relationship between $\mathbb{X}_n$ and $\hat{F}$ and that between $\mathbb{Y}_n$ and $\hat{F}^o$, it is clear that $[-L_n, R_n]$ is contained in $[-L_n' - L_n'', R_n' + R_n'']$. Thus, it suffices to show $L_n' + L_n'' = O_P(1)$ and $R_n' + R_n'' = O_P(1)$.

From the proof of Lemma 4.23, we have already shown L_n' and R_n' are $O_P(1)$, just by letting M there be 0. Thus, it remains to show L_n'' and R_n'' are $O_P(1)$. We next show the former and the latter can be derived in the same way.

For each integer $M > 0$, by Theorem 4.17,

$$\limsup_{n\to\infty} \mathbb{P}\left(L_n'' > M\right) \leq \limsup_{n\to\infty} \mathbb{P}\left(\mathbb{X}_n(-cM) > 0\right) = \mathbb{P}\left(\mathbb{X}(-cM) > 0\right).$$

On the other hand, we will show $\lim_{M\to\infty} \mathbb{P}\left(\mathbb{X}(-cM) > 0\right) = 0$ in the proof of the next Lemma A.13. Therefore, L_n'' is $O_P(1)$, which completes the proof. $\square$

LEMMA A.13. *Claim 2 in Lemma A.11 holds.*

PROOF. Define L', R', L'' and R'' similar to those L_n', R_n', L_n'' and R_n'' in the proof of Lemma A.12. More specifically, let $-cL'$ be the smallest grid point less than 0 such that $\mathbb{X}(-cL')$ is equal to $\mathbb{X}(0)$ and cR' the largest grid point larger than or equal to 0 such that $\mathbb{X}(-cR')$ is equal to $\mathbb{X}(0)$. Let $-cL''$ be the smallest grid point less than 0 such that $\mathbb{X}(-cL'') > 0$ and cR'' the largest grid point greater than 0 such that $\mathbb{X}(cR'') < 0$. If there doest not exist such a grid point $-cL''$ or cR'', let L'' or R'' be any nonnegative integer, say 0. Note that L', R', L'' and R'' are nonnegative.

Then, $\mathbb{X}$ and $\mathbb{Y}$ differ at most over $[-cL'-cL'', cR'+cR'']$. Since that $[-L, R]$ is in $[-L'-L'', R'+R'']$, it suffices to show $L'+L'' = O_P(1)$ and $R'+R'' = O_P(1)$.

From the proof of Lemma A.7, we have already known L' and R' are $O_P(1)$, just by letting M there be 0. Thus, it remains to show L'' and R'' are $O_P(1)$. We next show the former, and the latter can be done in the same way.

For each integer $M > 0$, we have

$$\mathbb{P}\left(L'' > M\right) \leq \mathbb{P}\left(\bigcup_{m=M}^{\infty}\left\{\frac{\mathcal{P}_{2,c}(-m)}{-cm} > 0\right\}\right).$$

On the other hand, we have

$$\mathbb{P}\left\{\limsup_{t\to-\infty}\frac{\alpha W(t)+\beta t(c+t)}{t} = -\infty\right\} = 1,$$

which gives

$$\mathbb{P}\left\{\limsup_{t\to-\infty}\frac{\alpha W(t)+\beta t(c+t)}{t} > 0\right\} = 0.$$

Thus, we have

$$\lim_{M\to\infty}\mathbb{P}\left(\bigcup_{m=M}^{\infty}\left\{\frac{\mathcal{P}_{2,c}(-m)}{-cm} > 0\right\}\right) = \mathbb{P}\left(\limsup_{m\to\infty}\frac{\mathcal{P}_{2,c}(-m)}{-cm} > 0\right) = 0,$$

Therefore, L'' is $O_P(1)$. □

LEMMA A.14. *In the proof of Theorem 4.20, the remainder terms R_1, R_2 and R_3 are all $o_P(1)$ holds.*

PROOF. We will only show $R_1 = o_P(1)$ and $R_2 = o_P(1)$ can be established in the same way. Recall that

$$\begin{aligned} R_1 &= -\frac{1}{2}\left(\frac{1}{F_l^2}-\frac{1}{(1-F_l)^2}\right)\sum_{i\in J_n}\left(\hat{F}(X_i)-F_l\right)^3 \\ &+ \frac{1}{3}\sum_{i\in J_n}\left(\frac{Y_i}{F_{li}^{\star 3}}-\frac{1-Y_i}{(1-F_{li}^{\star\star})^3}\right)\left(\hat{F}(X_i)-F_l\right)^3. \end{aligned}$$

First, we have

$$n^{1/3}\sum_{i\in J_n}\left|\hat{F}(X_i)-F_l\right|^3 = \sum_{j\in[L_n,U_n]}\left|\mathbb{X}_n(cj)\right|^3 W_j \rightsquigarrow c\sum_{j\in[L,U]}\left|\mathbb{X}(cj)\right|^3,$$

where the weak convergence holds by the same argument as in Lemma A.11.

Second, it is easy to see that there exists $C > 0$ such that $1/F_l^2 + 1/(1-F_l)^2 < C$ for large n. For example, take $C = 2/F(x_0)^2 + 2/(1 - F(x_0))^2$.

Third, recall that $F_{li}^\star$ and $F_{li}^{\star\star}$ lie between F_l and $\hat{F}(X_i)$ for $i \in J_n$. Note that X_i with $i \in J_n$ can only be t_j for some $j \in D_n$. That is, for each $i \in J_n$, there exists $j \in D_n$ such that $\hat{F}(X_i) = \hat{F}(t_j)$. On the other hand, note L_n and U_n are $O_P(1)$. Then, with arbitrary high probability $D_n \subset [\pm M]$ for M large enough. Since $\hat{F}(t_j)$ converges to $F(x_0)$ in probability for each fixed j and F_l always converges to $F(x_0)$, both $F_{li}^\star$ and $F_{li}^{\star\star}$ converge to $F(x_0)$ in probability for $i \in [\pm M]$ with a fixed M. Thus, we have $\sup_{i \in J_n}(1/F_{li}^{\star 3} + 1/(1 - F_{li}^{\star\star})^3)$ converges to $1/F(x_0)^3 + 1/(1 - F(x_0))^3$ in probability.

Therefore, using the triangular inequality for absolute value, combining the above three results and using Slutsky's Lemma gives $R_1 = o_P(1)$.

Next, consider R_3. We have: $R_3 = g(x_0)(F_l(1 - F_l))^{-1}T_n$, where

$$T_n = \sum_{j \in [-L_n, U_n]} \left(\mathbb{X}_n^2(cj) - \mathbb{Y}_n^2(cj)\right)\left[(G_n^\star(c(j)) - G_n^\star(c(j-1))) - c\right].$$

Then $|T_n| \le S_n \xi_n$ by noticing that $\mathbb{X}_n^2(cj) \ge \mathbb{Y}_n^2(cj)$ for each j, where

$$\xi_n = \sup_{j \in [-L_n, U_n]} \left|(G_n^\star(c(j)) - G_n^\star(c(j-1))) - c\right|.$$

By Lemma A.11, S_n converges weakly. Thus, it suffices to show the nonnegative quantity $\xi_n = o_P(1)$. Denote $D_n = \{[-L_n, U_n] \subset [\pm C]\}$ for $C > 0$. By Lemma A.12, the probability of D_n can be made arbitrarily close to 1 as C becomes large enough. Then, it suffices to show $\xi_n D_n = o_P(1)$. We have

$$\begin{aligned} \xi_n D_n &\le \sup_{j \in [\pm C]} |G_n^\star(cj) - cj| + \sup_{j \in [\pm C]} |G_n^\star(c(j-1)) - c(j-1)| \\ &\le 2 \sup_{j \in [\pm (C+1)]} |G_n^\star(cj) - cj|. \end{aligned}$$

The last term converges to 0 in probability by Lemma A.8, which completes the proof. □

REFERENCES

ABREVAYA, J. and HUANG, J. (2005). On the Bootstrap of the Maximum Score Estimator. *Econometrica* **73** 1175-1204.

AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T. and SILVERMAN, E. (1955). An Empirical Distribution Function for Sampling with Incomplete Information. *The Annals of Mathematical Statistics* **24** 641-647.

BANERJEE, (2000). Likelihood Ratio Inference in Regular and Non-regular Problems PhD thesis, University of Michigan.

BANERJEE, M. (2007). Likelihood based inference for monotone response models. *Annals of Statistics* **35** 931–956.

BANERJEE, M. and WELLNER, J. A. (2001). Likelihood Ratio Tests for Monotone Functions. *Annals of Statistics* **29** 1699 - 1731.

BANERJEE, M. and WELLNER, J. A. (2005). Confidence intervals for current status data. *Scand. J. Statist.* **32** 405–424. MR2204627

GROENEBOOM, P., MAATHUIS, M. H. and WELLNER, J. A. (2008a). Current status data with competing risks: Consistency and rates of convergence of the MLE. *The Annals of Statistics* **36** 1031–1063.

GROENEBOOM, P., MAATHUIS, M. H. and WELLNER, J. A. (2008b). Current status data with competing risks: Limiting distribution of the MLE. *The Annals of Statistics* **36** 1064–1089.

GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhauser-Verlag.

HUANG, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics* **24** 540–568.

JEWELL, N. P., VAN DER LAAN, M. and HENNEMAN, T. (2003). Nonparametric estimation from current status data with competing risks. *Biometrika* **90** 183–197.

KIM, J. and POLLARD, D. (1990). Cube Root Asymptotics. *The Annals of Statistics* **18** 191-219.

KOSOROK, M. R. (2008). Bootstrapping the Grenander estimator. In *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen* 282-292. Institute of Mathematical Statistics: Hayward, CA.

LIN, D. Y., OAKES, D. and YING, Z. (1998). Additive hazards regression with current status data. *Biometrika* **85** 289–298.

MA, S. and KOSOROK, M. R. (2005). Penalized log-likelihood estimation for partly linear transformation models with current status data. *The Annals of Statistics* **33** 2256-2290.

MAATHUIS, M. H. and HUDGENS, M. G. (2010). Nonparametric inference for competing risks current status data with continuous, discrete or grouped observation times. *Biometrika (to appear)*.

POLLARD, D. (2002). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhya Ser. A* **31** 23-36.

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order restricted statistical inference*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. John Wiley & Sons Ltd., Chichester. MR961262

SCHICK, A. and YU, Q. (2000). Consistency of the GMLE with Mixed Case Interval-Censored Data. *Scandinavian Journal of Statistics* **27** p. 45 Length: 11 pages.

SEN, B. and BANERJEE, M. (2006). A pseudo–likelihood method for analyzing interval censored data. *Biometrika*.

SEN, B., BANERJEE, M. and WOODROOFE, M. (2010). Inconsistency of Bootstrap: The Grenander Estimator. *Annals of Statistics* **38** 1953–1977.

SHIBOSKI, S. C. (1998). Generalized Additive Models for Current Status Data. *Lifetime Data Analysis* **4** 1953–50.

SUN, J. (1999). A nonparametric test for current status data with unequal censoring. *Journal of the Royal Statistical Society: Series B* **61** 243–250.

TURNBULL, B. W. (1976). The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. *Journal of the Royal Statistical Society. Series B (Methodological)* **38** 290–295.

VAN DER VAART, A. (1991). On Differentiable Functionals. *The Annals of Statistics* **19** 178-204.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer.

WANG, X. and SHEN, J. (2010). A Class of Grouped Brunk Estimators and Penalized Spline Estimators for Monotone Regression. *Biometrika* **97** 585-601.

YU, Q., SCHICK, A., LI, L. and WONG, G. Y. C. (1998). Asymptotic properties of the GMLE in the case 1 interval-censorship model with discrete inspection times. *The Canadian Journal of Statistics* **26** 619–627.

ZHANG, R., KIM, J. and WOODROOFE, M. (2001). Asymptotic analysis of isotonic estimation for grouped data. *JSPI* **98** 107-117.

ADDRESS OF THE FIRST AND SECOND AUTHORS
USUALLY FEW LINES LONG
E-MAIL: rltang@umich.edu
moulib@umich.edu

ADDRESS OF THE THIRD AUTHOR
USUALLY FEW LINES LONG
USUALLY FEW LINES LONG
E-MAIL: kosork@unc.edu
URL: http://www.foo.com