The Annals of Statistics
2012, Vol. 40, No. 1, 45–72
DOI: 10.1214/11-AOS942

LIKELIHOOD BASED INFERENCE FOR CURRENT STATUS DATA ON A GRID: A BOUNDARY PHENOMENON AND AN ADAPTIVE INFERENCE PROCEDURE

BY RUNLONG TANG[1], MOULINATH BANERJEE[1] AND MICHAEL R. KOSOROK

Princeton University, University of Michigan, Ann Arbor, and University of North Carolina, Chapel Hill

In this paper, we study the nonparametric maximum likelihood estimator for an event 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 n, the number of subjects. The asymptotic behavior falls into three cases depending on γ: regular Gaussian-type asymptotics obtain for $\gamma < 1/3$, nonstandard 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 event 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 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 event time distribution, F, can be estimated by its nonparametric maximum likelihood estimator (NPMLE) at only $n^{1/3}$ rate when the observation time is a continuous random variable. Under mild conditions on F, the limiting distribution of the NPMLE in this setting is the non-Gaussian Chernoff distribution: the distribution of the unique minimizer of $\{W(t) + t^2 : t \in \mathbb{R}\}$, where $W(t)$ is standard two-sided Brownian motion. This is in contrast to data with right-censored event

Received August 2011; revised November 2011.
[1]Supported in part by NSF Grants DMS-07-05288 and DMS-10-07751.
MSC2010 subject classifications. Primary 62G09, 62G20; secondary 62G07.
Key words and phrases. Adaptive procedure, boundary phenomenon, current status model, isotonic regression.

times where F can be estimated nonparametrically at rate $\sqrt{n}$ and is "pathwise norm-differentiable" in the sense of van der Vaart (1991), admitting regular estimators and normal limits. Interestingly, when the observation time distribution has finite support, the NPMLE for F at a point asymptotically simplifies to a binomial random variable and is also $\sqrt{n}$ estimable and regular, with a normal limiting distribution.

An extensive amount of work has been done for inference in the current status model under the assumption of a continuous distribution for the observation time: the classical model considers n subjects whose survival times $T_1, T_2, \ldots, T_n$ are i.i.d. F and whose inspection times $X_1, X_2, \ldots, X_n$ are i.i.d. with some continuous distribution, say G; furthermore, in the absence of covariates, the X_i's and T_i's are considered mutually independent. The observed data are $\{\Delta_i, X_i\}_{i=1}^n$, where $\Delta_i = 1\{T_i \leq X_i\}$, and one is interested in estimating F as n goes to infinity. More specifically, for inference on the value of F at a pre-fixed point of interest under a continuous observation time, see, for example, Groeneboom and Wellner (1992), who establish the convergence of the normalized NPMLE to Chernoff's distribution; Keiding et al. (1996); Wellner and Zhang (2000), who develop pseudo-likelihood estimates of the mean function of a counting process with panel count data, current status data being a special case; Banerjee and Wellner (2001) and Banerjee and Wellner (2005), who develop an asymptotically pivotal likelihood ratio based method; Sen and Banerjee (2007), who extend the results of Wellner and Zhang (2000) to asymptotically pivotal inference for F with mixed-case interval-censoring; and Groeneboom, Jongbloed and Witte (2010) for smoothed isotonic estimation, to name a few.

However, somewhat surprisingly, the problem of making inference on F when the observation times lie on a grid with multiple subjects sharing the same observation time has never been satisfactorily addressed in this rather large literature. This important scenario, which transpires when the inspection times for individuals at risk are evenly spaced, and multiple subjects can be inspected at any inspection time, is *completely precluded* by the assumption of a continuous G, as this does not allow ties among observation times. Consider, for example, a tumorigenicity study where a large number of mice are exposed to some carcinogen at a particular time, and interest centers on the time to development of a tumor. A typical procedure here would be to randomize the mice to be sacrificed over a number of days following exposure; so, one can envisage a protocol of sacrificing a fixed number m of mice at 24 hrs post-exposure, another m mice at 48 hours and so on. The sacrificed mice are then dissected and examined for tumors, thereby leading to current status data on a grid. A pertinent question in this setting is: what is the probability that a mouse develops a tumor by an M-day period after exposure? This involves estimating $F(24M)$, where F is the distribution function of the time to tumor-development. Similar grid-based data can occur with human subjects in clinical settings.

In this paper we provide a clean solution to this problem based on the NPMLE of F which, as is well known, is obtained through isotonic regression [see, e.g., Robertson, Wright and Dykstra (1988)]. The NPMLE of F in the current status model (and more generally in nonparametric monotone function models) has a long history and has been studied extensively. In addition to the attractive feature that it can be computed without specifying a bandwidth, the NPMLE of $F(x_0)$ (where x_0 is a fixed point) attains the best possible convergence rate, namely $n^{1/3}$, in the "classical" current status model with continuous observation times, under the rather mild assumption that F is continuously differentiable in a neighborhood of x_0 and has a nonvanishing derivative at x_0. *This rate cannot be bettered by a smooth estimate under the assumption of a single derivative.* As demonstrated in Groeneboom, Jongbloed and Witte (2010), smoothed monotone estimates of F can achieve a faster $n^{2/5}$ rate under a *twice-differentiability assumption on* F; hence, the faster rate requires additional smoothness. However, as we wish to approach our problem under minimal smoothness assumptions, the isotonic NPMLE is the more natural choice. (Smoothing the NPMLE would introduce an exogenous tuning parameter without providing any benefit from the point of view of the convergence rate.)

The key step, then, is to determine the best asymptotic approximation to use for the NPMLE in the grid-based setting discussed above. If, for example, the number of observation times, K, is far smaller than n, the number of subjects, the problem is essentially a parametric one, and it is reasonable to expect that normal approximations to the MLE will work well. On the other hand, if $K = n$, that is, we have a very fine grid with each subject having their own inspection time, the scenario is similar to the current status model with continuous observation times where no two inspection times coincide, and one may expect a Chernoff approximation to be adequate. However, there is an entire spectrum of situations in between these extremes depending on the size of the grid, K, relative to n, and if n is "neither too large, nor too small relative to K," neither of these two approximations would be reliable.

Some work on the current status model or closely related variants under discrete observation time settings should be noted in this context. 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 (2011) have considered nonparametric inference for (finitely many) 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 toward studying the effect of the trade-off between n and K, that is, *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 $\sqrt{n}$ consistent and asymptotically normal; but as Maathuis and Hudgens (2011) demonstrate in Section 5.1 of their

paper, when the number of distinct observation times is large relative to the sample size, the normal approximations are suspect.

Our approach is to couch the problem in an asymptotic framework where K is allowed to increase with n at rate n^γ for some $0 < \gamma \leq 1$ and study the behavior of the NPMLE at a grid-point. This is achieved by considering the current status model on a regular grid over a compact time interval, say $[a, b]$, with unit spacing $\delta \equiv \delta_n = cn^{-\gamma}$, c being a scale parameter. It will be seen that the limit behavior of the NPMLE depends heavily on the "sparsity parameter" γ, with the Gaussian approximation prevailing for $\gamma < 1/3$ and the Chernoff approximation for $\gamma > 1/3$. When $\gamma = 1/3$, one obtains a discrete analog of the Chernoff distribution which depends on c. Thus, there is an entire family of what we call *boundary distributions*, indexed by c, say $\{F_c : c > 0\}$, by manipulating which, one can approach either the Gaussian or the Chernoff. As c approaches 0, F_c approximates the Chernoff while, as c approaches ∞, it approaches the Gaussian. This property allows us to develop an *adaptive procedure* for setting confidence intervals for the value of F at a grid-point that *obviates the need to know or estimate* γ, the critical parameter in this entire business as it completely dictates the ensuing asymptotics. The adaptive procedure involves *pretending* that the true unknown underlying unknown γ is at the boundary value $1/3$, computing a *surrogate* c, say $\hat{c}$, by equating $(b-a)/K$, the spacing of the grid (which is computable from the data), to $\hat{c}n^{-1/3}$ and using $F_{\hat{c}}$, to approximate the distribution of the appropriately normalized NPLME. The details are given in Section 4. It is seen that this procedure provides asymptotically correct confidence intervals regardless of the true value of γ. Our procedure does involve estimating some nuisance parameters, but this is readily achieved via standard methods.

The rest of the paper is organized as follows. In Section 2, we present the mathematical formulation of the problem and introduce some key notions and characterizations. Section 3 presents the main asymptotic results and their connections to existing work. Section 4 addresses the important question of *adaptive inference* in the current status model: given a time-domain and current status data observed at times on a regular grid of an *unknown level of sparsity* over the domain, how do we make inference on F? Section 5 discusses the implementation of the procedure and presents results from simulation studies, and Section 6 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. The Appendix contains some technical details.

2. Formulation of the problem. Let $\{T_{i,n}\}_{i=1}^n$ be i.i.d. survival times following some unknown distribution F with Lebesgue density f concentrated on the time-domain $[a', b']$ with $0 \leq a' < b' < \infty$ (or supported on $[a', \infty)$ if no such b' exists) and $\{X_{i,n}\}$ be i.i.d. observation times drawn from a discrete probability measure H_n supported on a regular grid on $[a, b]$ with $a' \leq a < b < b'$. 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 $\{Y_{i,n} = 1\{T_{i,n} \le X_{i,n}\}\}$. 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 ith grid point $t_{i,n} = a + i\delta$, the unit spacing $\delta = \delta(n) = cn^{-\gamma}$ (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 each i. 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.

Our interest lies in estimating F at a grid-point. Since we allow the grid to change with n, this will be accomplished by specifying a grid-point with respect to a fixed time $x_0 \in (a, b)$ which does not depend on n and can be viewed as an "anchor-point." Define $t_l = t_{l,n}$ to be the largest grid-point less than or equal to x_0. We devote our interest to $\hat{F}(t_l)$. More specifically, we are interested in the limit distribution of $\hat{F}(t_l) - F(t_l)$ under appropriate normalization? To this end, we start with the characterization of the NPMLE in this model. While this is well known from the current status literature, we include a description tailored for the setting of this paper.

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

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

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 observations 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 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 is given by

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

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

Denote the basic shape-restricted maximizer as

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

From the theory of isotonic regression [see, e.g., Robertson, Wright and Dykstra (1988)], we have

$$\operatorname*{arg\,max}_{F_1\le\cdots\le F_K} l_n(F) = \operatorname*{arg\,min}_{F_1\le\cdots\le F_K} \sum_{i=1}^{K}[(\bar{Z}_i - F_i)^2 N_i].$$

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 on $[a, b]$ as the following right-continuous step function:

$$\hat{F}(t) = \begin{cases} 0, & \text{if } t \in [a, t_1); \\ F_i^\star, & \text{if } t \in [t_i, t_{i+1}),\ i = 1, \ldots, K-1; \\ F_K^\star, & \text{if } t \in [t_K, b]. \end{cases} \tag{2.1}$$

Next, we provide a characterization of $\hat{F}$ as the slope of *the greatest convex minorant* (GCM) of a random processes, which proves useful for deriving the asymptotics for $\gamma \in [1/3, 1]$. Define, for $t \in [a, b]$,

$$G_n(t) = \mathbb{P}_n\{x \le t\}, \qquad V_n(t) = \mathbb{P}_n y\{x \le t\}, \tag{2.2}$$

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

$$\hat{F}(x) = \mathrm{LS}\big[\mathrm{GCM}\{(G_n(t), V_n(t)), t \in [a, b]\}\big](G_n(x)). \tag{2.3}$$

In the above display, GCM means 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 used to refer to the greatest convex minorant of a real-valued function defined on a sub-interval of the real line.

Finally, we introduce a number of random processes that will appear in the asymptotic descriptions of $\hat{F}$.

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

$$X_{\kappa_1,\kappa_2}(h) = \kappa_1 W(h) + \kappa_2 h^2 \qquad \text{for } h \in \mathbb{R}, \tag{2.4}$$

where W is a two-sided Brownian motion with $W(0) = 0$. Let G_{κ_1,κ_2} be the GCM of X_{κ_1,κ_2}. Define, for $h \in \mathbb{R}$,

$$g_{\kappa_1,\kappa_2}(h) = \mathrm{LS}[G_{\kappa_1,\kappa_2}](h). \tag{2.5}$$

The process g_{κ_1,κ_2} will characterize the asymptotic behavior of a localized NPMLE process in the vicinity of t_l for $\gamma > 1/3$, from which the large sample distribution of $\hat{F}(t_l)$ can be deduced.

We also define a three parameter family of processes in discrete time which serve as discrete versions of the continuous-time processes above. For $c, \kappa_1, \kappa_2 > 0$, let

$$\begin{aligned} \mathcal{P}_{c,\kappa_1,\kappa_2}(k) &= (\mathcal{P}_{1,c,\kappa_1,\kappa_2}(k), \mathcal{P}_{2,c,\kappa_1,\kappa_2}(k)) \\ &= \{ck, \kappa_1 W(ck) + \kappa_2 c^2 k(1+k)\}_{k\in\mathbb{Z}}. \end{aligned} \tag{2.6}$$

Define

$$\mathbb{X}_{c,\kappa_1,\kappa_2}(ci) = \mathrm{LS}[\mathrm{GCM}\{\mathcal{P}_{c,\kappa_1,\kappa_2}(k) : k \in \mathbb{Z}\}](ci). \tag{2.7}$$

This slope process will characterize the asymptotic behavior of the NPMLE in the case $\gamma = 1/3$.

3. Asymptotic results. In this section, we state and discuss results on the asymptotic behavior of $\hat{F}(t_l)$ for γ varying in $(0, 1]$. In all that follows, we make the *blanket assumption* that F is once continuously differentiable in a neighborhood of x_0.

3.1. *The case* $\gamma < 1/3$. We start with some technical assumptions:

(A1.1) F has a bounded density f on $[a, b]$, and there exists $f_l > 0$ such that $f(x) > f_l$ for every $x \in [a, b]$.

(A1.2) G has a bounded density g on $[a, b]$, and there exists $g_l > 0$ such that $g(x) \geq g_l$ for every $x \in [a, b]$.

(A1.3) $a' < a$ and $F(a) > 0$.

The above assumptions are referred to collectively as (A1). Letting t_r denote the first grid-point to the right of t_l, we have the following theorem.

THEOREM 3.1. *If* $\gamma \in (0, 1/3)$ *and* (A1) *holds*,

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

where I_2 *is the* 2×2 *identity matrix.*

The proof of this theorem is provided in the supplement to this paper [Tang, Banerjee and Kosorok (2011)]. However, a number of remarks in connection with the above theorem are in order.

REMARK 3.2. From Theorem 3.1, 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.

It can be shown that for $\gamma \in (0, 1/3)$, $N_l/(np_l)$ converges to 1 in probability and that $np_l/cg(x_0)n^{1-\gamma}$ converges to 1. Then the result of Theorem 3.1 can be rewritten as follows:

$$\big(n^{(1-\gamma)/2}\big(\hat{F}(t_l) - F(t_l)\big), n^{(1-\gamma)/2}\big(\hat{F}(t_r) - F(t_r)\big)\big) \xrightarrow{d} \alpha c^{-1/2}N(0, I_2), \tag{3.1}$$

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.

REMARK 3.3. The proof of the above theorem relies heavily on the below proposition which deals with the vector of average responses at the the grid-points: $\{\bar{Z}_i\}_{i=1}^k$. Since $\bar{Z}_i$ is not defined when $N_i = 0$, to avoid ambiguity we set $\bar{Z}_i = 0$ whenever this happens. This can be done without affecting the asymptotic results, since it can be shown that the probability of the event $\{N_i > 0, i = 1, 2, \ldots, K\}$ goes to 1.

PROPOSITION 3.4. *If* $\gamma \in (0, 1/3)$ *and* (A1) *holds, we have*

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

This proposition is established in the supplement, Tang, Banerjee and Kosorok (2011). It says that with probability going to 1, the vector $\{\bar{Z}_i\}_{i=1}^k$ is ordered, and therefore the isotonization algorithm involved in finding the NPMLE of F yields $\{F_i^\star\}_{i=1}^K = \{\bar{Z}_i\}_{i=1}^K$ with probability going to 1. 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 nonlinear algorithm, to the study of the asymptotics of the $\bar{Z}_i$, which are linear statistics and can be handled readily using standard central limit theory. A phenomenon, similar to the one in the above proposition, was observed by Kiefer and Wolfowitz (1976) in connection with estimating the magnitude of the difference between the empirical distribution function and its least concave majorant for an i.i.d. sample from a concave distribution function. See Theorem 1 of their paper and the preceding Lemma 4, which establish the concavity of a piecewise linear estimate of the true distribution obtained by linearly interpolating the restriction of the empirical distribution to a grid with spacings of order slightly larger than $n^{-1/3}$, n being the sample size. A similar result was obtained in Lemma 3.1 of Zhang, Kim and Woodroofe (2001) in connection with isotonic estimation of a decreasing density when the exact observations are not available; rather, the numbers of data-points that fall into equi-spaced bins are observed.

3.2. *The case* $\gamma \in (1/3, 1]$. Our treatment will be condensed since the asymptotics for this case follow the same patterns as when the observation times possess a Lebesgue density. That this ought to be the case is suggested, for example, by Theorem 1 in Wright (1981); see, in particular, the condition on the rate of convergence of the empirical distribution function of the regressors to the true distribution function in the case that $\alpha = 1$ in that theorem, which corresponds to the setting $\gamma > 1/3$ in our problem. Note that the α in the previous sentence refers to notation in Wright (1981) and *should not be confused with* the α defined in this paper.

In order to study the asymptotics of the isotonic regression estimator $\hat{F}(t_l)$, the following localized process will be of interest: for $u \in I_n = [(a - t_l)n^{1/3}, (b -$

$t_l)n^{1/3}]$, define

$$\mathbb{X}_n(u) = n^{1/3}\big(\hat{F}(t_l + un^{-1/3}) - F(t_l)\big). \tag{3.2}$$

Next, define the following normalized processes on I_n:

$$G_n^\star(h) = g(x_0)^{-1}n^{1/3}\big(G_n(t_l + hn^{-1/3}) - G_n(t_l)\big), \tag{3.3}$$

$$\begin{aligned} V_n^\star(h) = g(x_0)^{-1}n^{2/3}\big[&V_n(t_l + hn^{-1/3}) - V_n(t_l) \\ &- F(t_l)\big(G_n(t_l + hn^{-1/3}) - G_n(t_l)\big)\big]. \end{aligned} \tag{3.4}$$

After some straightforward algebra, from (2.3) and (3.2), we have the following technically useful characterization of $\mathbb{X}_n$: for $u \in I_n$,

$$\mathbb{X}_n(u) = \mathrm{LS}[\mathrm{GCM}(G_n^\star(h), V_n^\star(h)), h \in I_n](G_n^\star(u)). \tag{3.5}$$

Let α be defined as Remark 3.2 and $\beta = f(x_0)/2$. We have the following theorem on the distributional convergence of $\mathbb{X}_n$.

THEOREM 3.5 (Weak convergence of $\mathbb{X}_n$). *Suppose F and G are continuously differentiable in a neighborhood of x_0 with derivatives f and g. Assume that $f(x_0) > 0$, $g(x_0) > 0$ and that g is Lipschitz continuous in a neighborhood of x_0. Then, the finite-dimensional marginals of the process $\mathbb{X}_n$ converge weakly to those of the process $g_{\alpha,\beta}$.*

REMARK 3.6. Note that $\mathbb{X}_n(0) = n^{1/3}(\hat{F}(t_l) - F(t_l))$. By Theorem 3.5, it converges in distribution to $g_{\alpha,\beta}(0)$. By the Brownian scaling results on page 1724 of Banerjee and Wellner (2001), for $h \in \mathbb{R}$,

$$g_{\alpha,\beta}(h) \stackrel{d}{=} (\alpha^2\beta)^{1/3} g_{1,1}\big((\beta/\alpha)^{2/3}h\big).$$

Then, by noting that $g_{1,1}(0) \stackrel{d}{=} 2\mathcal{Z}$, we have the following result:

$$n^{1/3}\big(\hat{F}(t_l) - F(t_l)\big) \stackrel{d}{=} \left(\frac{4f(x_0)F(x_0)(1 - F(x_0))}{g(x_0)}\right)^{1/3} \mathcal{Z}. \tag{3.6}$$

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. The proof of this theorem is omitted as it can be established via arguments similar to those in Banerjee (2007) using continuous mapping theorems for slopes of greatest convex minorants.

3.3. *The case* $\gamma = 1/3$. Now, we consider the most interesting boundary case $\gamma = 1/3$. Let the localized process $\mathbb{X}_n(u)$ be defined exactly as in the previous subsection. The order of the grid-spacing δ is now exactly $n^{-1/3}$, which is the order of localization around t_l used to define the process $\mathbb{X}_n$, and it follows that $\mathbb{X}_n$ has potential jumps only at ci for $i \in \mathcal{I}_n = (I_n/c) \cap \mathbb{Z}$, and it suffices to consider $\mathbb{X}_n$ on those ci's. For $i \in \mathcal{I}_n$,

$$\mathbb{X}_n(ci) = n^{1/3}\big(\hat{F}(t_l + cin^{-1/3}) - F(t_l)\big) \tag{3.7}$$

$$= \mathrm{LS}[\mathrm{GCM}\{(G_n^\star(ck), V_n^\star(ck)), k \in \mathcal{I}_n\}](G_n^\star(ci)). \tag{3.8}$$

For simplicity of notation, in 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.

The following theorem gives the limit behavior of $\mathbb{X}_n$.

THEOREM 3.7 (Weak convergence of $\mathbb{X}_n$). *Under the same assumptions as in Theorem 3.5, for each nonnegative integer N, we have*

$$\{\mathbb{X}_n(ci), i \in [-N, N]\} \xrightarrow{d} \{\mathbb{X}_{c,\alpha,\beta}(ci), i \in [-N, N]\}.$$

It follows that $n^{1/3}(\hat{F}(t_l) - F(t_l)) \xrightarrow{d} \mathbb{X}_{c,\alpha,\beta}(0)$.

REMARK 3.8. It is interesting to note the change in the limiting behavior of the NPMLE with varying γ. As noted previously, for $\gamma \in (0, 1/3)$, the grid is sparse enough so that the naive average responses at each inspection time, which provide empirical estimates of F at those corresponding inspection times, are automatically ordered (and therefore the solution to the isotonic regression problem) and there is no "strength borrowed" from nearby inspection times. Consequently, a Gaussian limit is obtained. For $\gamma \geq 1/3$, the grid points are "close enough," so that the naive pointwise averages are no longer the best estimates of F. In fact, owing to the closeness of successive grid-points, the naive averages are no longer ordered, and the PAV pool adjacent violators algorithm (PAVA) leads to a nontrivial solution for the NPMLE which is a highly nonlinear functional of the data, putting us in the setting of nonregular asymptotics. It turns out that for $\gamma \geq 1/3$, the order of the local neighborhoods of t_l that determine the value of $\hat{F}(t_l)$ is $n^{-1/3}$. When $\gamma = 1/3$, the resolution of the grid matches the order of the local neighborhoods, leading in the limit to a process in discrete-time that depends on c. When $\gamma > 1/3$, the number of grid-points in an $n^{-1/3}$ neighborhood of t_l goes to infinity. This eventually washes out the dependence on c and also produces, in the limit, a process in continuous time.

For the rest of this section, we refer to the process $\mathbb{X}_{c,\alpha,\beta}$ simply as $\mathbb{X}$ and the process $\mathcal{P}_{c,\alpha,\beta}$ as $\mathcal{P}_c$.

PROOF–SKETCH OF THEOREM 3.7. The key steps of the proof are as follows. Take an integer $M > N$. Then, the following two claims hold.

CLAIM 1. There exist (integer-valued) random variables $L_n < -M$ and $U_n > M$ which are $O_P(1)$ and satisfy

$$\begin{aligned}&\mathrm{GCM}\{(G_n^\star(ck), V_n^\star(ck)), k \in [L_n, U_n]\}\\ &\quad = \mathrm{GCM}\{(G_n^\star(ck), V_n^\star(ck)), k \in \mathbb{Z}\}|[G_n^\star(cL_n), G_n^\star(cU_n)].\end{aligned}$$

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

$$\mathrm{GCM}\{\mathcal{P}_c(k), k \in [L, U]\} = \mathrm{GCM}\{\mathcal{P}_c(k), k \in \mathbb{Z}\}|[cL, cU].$$

For the proofs of these claims, see Tang, Banerjee and Kosorok (2011). We next need a key approximation lemma, which is a simple extension of Lemma 4.2 in Prakasa Rao (1969).

LEMMA 3.9. *Suppose that for each $\varepsilon > 0$, $\{W_{n\varepsilon}\}$, $\{W_n\}$ and $\{W_\varepsilon\}$ are sequences of random vectors, W is a random vector and that*:

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

Then $W_n \xrightarrow{d} W$, *as* $n \to \infty$.

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

$$P(M_\varepsilon > \max\{|L_n|, U_n, |L|, U\}) > 1 - \varepsilon.$$

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

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

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

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

Then, the following three facts hold:

FACT 1. $\lim_{\varepsilon\to 0} \overline{\lim}_{n\to\infty} \mathbb{P}(A_n) = 0$.

FACT 2. $\lim_{\varepsilon \to 0} \mathbb{P}(A) = 0$.

FACT 3. $\{\mathbb{X}_n^{M_\varepsilon}(ci), i \in [\pm N]\} \xrightarrow{d} \{\mathbb{X}^{M_\varepsilon}(ci), i \in [\pm N]\}$, as $n \to \infty$ for each $\varepsilon > 0$.

Facts 1 and 2 follow since A_n and A are subsets of $\{M_\varepsilon \leq \max\{|L_n|, U_n, |L|, U\}\}$, whose probability is less than ε, Facts 1 and 2 hold. Fact 3 is proved in Tang, Banerjee and Kosorok (2011). A direct application of Lemma 3.9 then leads to the weak convergence that we sought to prove. □

REMARK 3.10. The proofs of Claims 1 and 2 consist of technically important localization arguments. Claim 1 ensures that eventually, with arbitrarily high prespecified probability, the restriction of the greatest convex minorant of the process $(G_n^\star, V_n^\star)$ (which is involved in the construction of $\mathbb{X}_n$) to a bounded domain can be made equal 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, depending on the prespecified probability. It can be proved by using techniques similar to those in Section 6 of Kim and Pollard (1990). Claim 2 ensures that an analogous phenomenon holds for the greatest convex minorant of the process $\mathcal{P}_c$, which is involved in the construction of $\mathbb{X}$. These equalities then translate to the left-derivatives of the GCMs involved, and the proof is completed by invoking a continuous mapping theorem for the GCMs of the restriction of $(G_n^\star, V_n^\star)$ to bounded domains, along with Claims 1 and 2, which enable the use of the approximation lemma adapted from Prakasa Rao (1969).

The basic strategy of the above proof has been invoked time and again in the literature on monotone function estimation. Prakasa Rao (1969) employed this technique to determine the limit distribution of the Grenander estimator at a point, and Brunk (1970) for studying monotone regression. Leurgans (1982) extended these techniques to more general settings which cover weakly dependent data while Anevski and Hössjer (2006) provided a comprehensive and unified treatment of asymptotic inference under order restrictions, applicable to independent as well as short and long range dependent data. This technique was also used in Banerjee (2007) to study the asymptotic distributions of a very general class of monotone response models. It ought to be possible to bring the general techniques of Anevski and Hössjer (2006) to bear upon the boundary case, but we have not investigated that option; our proof-strategy is most closely aligned with the proof of Theorem 2.1 in Banerjee (2007).

3.4. *A brief discussion of the boundary phenomenon.* We refer to the behavior of the NPMLE for $\gamma = 1/3$ as the *boundary phenomenon*. As indicated in the Introduction, the asymptotic distribution for $\gamma = 1/3$ is *different* from both the Gaussian (which comes into play for $\gamma < 1/3$) and the Chernoff (which arises for

$\gamma > 1/3$). This boundary distribution, which depends on the scale parameter, c, can be viewed as an intermediate between the Gaussian and Chernoff, and its degree of proximity to one or the other is dictated by c as we demonstrate in the following section. More importantly, this transition from one distribution to another via the boundary one, has important ramifications for inference in our grid-based problem as also demonstrated in the next section.

The closest result to our boundary phenomenon in the literature appears in the work of Zhang, Kim and Woodroofe (2001) who study the asymptotics of isotonic estimation of a decreasing density with histogram-type data. Thus, the domain of the density is split into a number of *pre-specified bins*, and the statistician knows the number of i.i.d. observations from the density that fall into each bin (with a total of n such observations). The rate at which the number of bins increases relative to n then drives the asymptotics of the NPMLE of the density within the class of decreasing piecewise linear densities, with a distribution similar to $\mathbb{X}(0)$ appearing when this number increases at rate $n^{1/3}$. However, *unlike us*, Zhang, Kim and Woodroofe (2001) do not establish any connections among the different limiting regimes; neither do they offer a prescription for inference when the rate of growth of the bins is unknown as is usually the case in practice.

It is worthwhile contrasting our boundary phenomenon with those observed by some other authors. Anevski and Hössjer (2006) discover a "boundary effect" in their Theorems 5 and 6.1 when dealing with an isotonized version of a kernel estimate (see Section 3.3 of their paper). In the setting of i.i.d. data, when the smoothing bandwidth is chosen to be of order $n^{-1/3}$, the asymptotics of the isotonized kernel estimator are given by the minimizer of *a Gaussian process* (*depending on the kernel*) *with continuous sample paths plus a quadratic drift*, whereas for bandwidths of larger orders than $n^{-1/3}$ normal distributions obtain. A similar phenomenon, in the setting of monotone density estimation, was observed by van der Vaart and van der Laan (2003) in their Theorem 2.2 for an isotonized kernel estimate of a decreasing density while using an $n^{-1/3}$ order bandwidth. Note that these boundary effects are quite different from our boundary phenomenon. *In Anevski and Hossjer's setting*, for example, the underlying regression model is observed on the grid $\{i/n\}$, with one response per grid-point. Kernel estimation with an $n^{-1/3}$ bandwidth smooths the responses over time-neighborhoods of order $n^{-1/3}$ producing a continuous estimator which is then subjected to isotonization. This leads to a limit that is characterized in terms of a process in *continuous time*. *In our setting*, our data are not necessarily observed on an $\{i/n\}$ grid; our grids can be much sparser and for the case $\gamma = 1/3$, multiple responses are available at each grid-point. The NPMLE isotonizes the $\bar{Z}_i$'s; thus, isotonization is preceded by averaging the multiple responses at each time cross-section, but *there is no averaging of responses across time*, in sharp contrast to Anevski and Hossjer's setting. This, in conjunction with the already noted fact at the beginning of this subsection that the grid-resolution when $\gamma = 1/3$ has the same order as the localization involved in constructing the process $\mathbb{X}_n$, leads in our case to a limit distribution for the NPMLE that is characterized as a functional of a process in *discrete time*.

4. Adaptive inference for F at a point. In this section, we develop a procedure for constructing asymptotic confidence intervals for $F(t_l)$ which does not require knowing or estimating the underlying grid resolution controlled by the parameters γ and c. This provides massive advantage from an inferential perspective because the parameter γ critically drives the limit distribution of the NPMLE and mis-specification of γ may result in asymptotically incorrect confidence sets, either due to the use of the wrong limit distribution or due to an incorrect convergence rate, or both.

To this end, we first investigate the relationships among the three different asymptotic limits for $\hat{F}(t_l)$ that were derived in the previous section, for different values of γ. In what follows, we denote $\mathbb{X}_{c,\alpha,\beta}(0)$ by $\mathcal{S}_c$, suppressing the dependence on α, β for notational convenience. The use of the letter $\mathcal{S}$ is to emphasize the characterization of this random variable as the slope of a stochastic process.

Our first result relates the distribution of $\mathcal{S}_c$ to the Gaussian.

THEOREM 4.1. *As $c \to \infty$, $\sqrt{c}\mathcal{S}_c \xrightarrow{d} \alpha Z$, where Z follows the standard normal distribution.*

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

THEOREM 4.2. *As $c \to 0$, $\mathcal{S}_c \xrightarrow{d} g_{\alpha,\beta}(0) \stackrel{d}{=} 2(\alpha^2\beta)^{1/3}\mathcal{Z}$.*

REMARK 4.3. Theorem 4.2 is somewhat easier to visualize heuristically, compared to Theorem 4.1. 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 $\mathcal{P}_c$ is eventually 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 3.6 following Theorem 3.5.

The results of Theorems 4.1 and 4.2 are illustrated next. Suppose the time interval $[a, b]$ is [0,2], $x_0 = 1$ and that F and G are both the uniform distribution on $[0, 2]$. Under these settings, the values of α and β are $\sqrt{2}/4$ and $1/4$, respectively. We generate i.i.d. random samples of $\mathcal{S}_c$ with c being 1, 2, 3, 5 and 10 and the common sample size being 5000. The left panel of Figure 1 compares the empirical cumulative distribution functions (CDF) of $\sqrt{c}\mathcal{S}_c/\alpha$ and the standard Gaussian distribution $N(0, 1)$. It shows clearly that the empirical CDFs move closer to the Gaussian distribution with increasing c and that the empirical CDF of $\sqrt{c}\mathcal{S}_c/\alpha$ with c equal to 3 has already provided a decent approximation to $N(0, 1)$. On the other hand, the right panel of Figure 1 compares the empirical CDFs of $(1/2)(\alpha^2\beta)^{-1/3}\mathcal{S}_c$ and the standard Chernoff distribution $\mathcal{Z}$. Again, the empirical CDFs approach that of $\mathcal{Z}$ with diminishing c, with $c = 1$ providing a close approximation for $\mathcal{Z}$. Note that, while the convergence in this setting is relatively quick

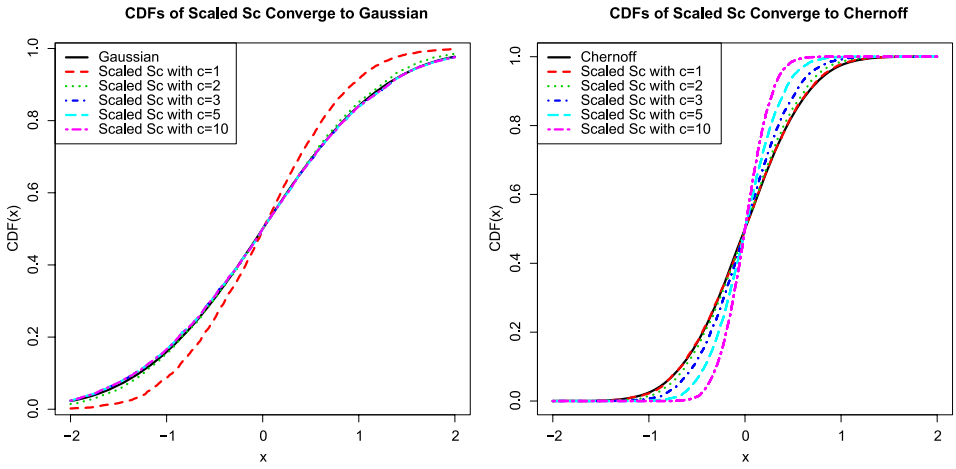

FIG. 1. *The left and right panels show that a sequence of empirical CDFs of the properly scaled* $\mathcal{S}_c$ *converge to the standard Gaussian and Chernoff distributions, respectively. In the left panel, the empirical CDFs with* $c \geq 3$ *almost coincide with the standard Gaussian distribution.*

in the sense that the limiting phenomena manifest themselves at moderate values of c (i.e., neither too large, nor too small), this may not necessarily be the case for other combinations of (α, β), and more extreme values may be required for good enough approximations.

The adaptive inference scheme: We are now in a position to propose our inference scheme. We focus on the so-called "Wald-type" intervals for $F(t_l)$, that is, 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. Let c_0 and γ_0 denote the *true unknown values* of c and γ in the current status model. With $K = K_n$ being the number of grid-points, we have the relation

$$K_n = \lfloor (b-a)/(c_0 n^{-\gamma_0}) \rfloor.$$

Now *pretend* that the true γ is exactly equal to $1/3$. Calculate a surrogate c, say $\hat{c}$, via the relation

$$\lfloor (b-a)/(\hat{c} n^{-1/3}) \rfloor = K_n.$$

Some algebra shows that

$$\hat{c} = \hat{c}_n = c n^{1/3-\gamma_0} + O(n^{1/3-2\gamma_0}) = c n^{1/3-\gamma_0}\big(1 + O(n^{-\gamma_0})\big).$$

Thus, the calculated parameter $\hat{c}$ actually depends on n, and goes to ∞ and 0 for $\gamma_0 \in (0, 1/3)$ and $\gamma_0 \in (1/3, 1]$, respectively.

We propose to use the distribution of $\mathcal{S}_{\hat{c}}$ as an approximation to the distribution of $n^{1/3}(\hat{F}(t_l) - F(t_l))$. Thus, an adaptive approximate $1-\eta$ confidence interval for $F(t_l)$ is given by

$$\big[\hat{F}(t_l) - n^{-1/3} q(\mathcal{S}_{\hat{c}}, 1-\eta/2),\ \hat{F}(t_l) - n^{-1/3} q\big(\mathcal{S}_{\hat{c}}, (\eta/2)\big)\big], \tag{4.1}$$

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

Asymptotic validity of the proposed inference scheme: The above adaptive confidence interval provides the correct asymptotic calibration, *irrespective of the true value of* γ. If γ_0 happens to be $1/3$, then, of course, the adaptive confidence interval is constructed with the correct asymptotic result. If not, consider first the case that $\gamma_0 \in (1/3, 1]$. If *we knew* that $\gamma_0 \in (1/3, 1]$, then, by result (3.6) and the symmetry of $g_{\alpha,\beta}(0)$, the true confidence interval would be

$$\big[\hat{F}(t_l) \pm n^{-1/3} q\big(g_{\alpha,\beta}(0), (1 - \eta/2)\big)\big]. \tag{4.2}$$

Now recall that $\hat{c}$ goes to 0 since $\gamma_0 \in (1/3, 1]$. Thus, by Theorem 4.2, the quantile sequence $q(\mathcal{S}_{\hat{c}}, p)$ converges to $q(g_{\alpha,\beta}(0), p)$, owing to the fact that $g_{\alpha,\beta}(0)$ is a continuous random variable. So, the adaptive confidence interval (4.1) converges to the true one (4.2) obtained when γ_0 is in $(1/3, 1]$.

That the adaptive procedure also works when $\gamma_0 \in (0, 1/3)$ will be shown by using Theorem 4.1. Again, *suppose we know* the value of γ_0. Then, from result (3.1) and the symmetry of the standard normal random variable Z, the confidence interval is given by

$$\big[\hat{F}(t_l) \pm n^{-(1-\gamma_0)/2} \alpha c^{-1/2} q\big(Z, (1 - \eta/2)\big)\big]. \tag{4.3}$$

To show that the adaptive procedure is, again, asymptotically correct, it suffices to show that for every $p \in (0, 1)$, as $n \to \infty$,

$$\frac{n^{-1/3} q(\mathcal{S}_{\hat{c}}, p)}{n^{-(1-\gamma_0)/2} \alpha c^{-1/2} q(Z, p)} = \frac{n^{-1/3} c^{1/2}}{n^{-(1-\gamma_0)/2} \hat{c}^{1/2}} \cdot \frac{\hat{c}^{1/2} q(\mathcal{S}_{\hat{c}}, p)}{\alpha q(Z, p)} = I \cdot II \to 1.$$

Recall that $\hat{c}$ goes to ∞ since $\gamma_0 \in (0, 1/3)$. By Theorem 4.1, we have $II \to 1$ as $n \to \infty$. On the other hand, we can see I simplifies to $(1 + O(n^{-\gamma_0}))^{-1/2}$ and therefore goes to 1. Thus, the adaptive confidence interval (4.1) also converges to the true one (4.3) obtained when γ_0 is known to be in $(0, 1/3)$.

Thus, our procedure *adjusts automatically* to the inherent rate of growth of the number of distinct observation times and that is an extremely desirable property.

We next articulate some practical issues with the adaptive procedure. First, note that $\mathcal{S}_{\hat{c}} = \mathbb{X}_{\hat{c},\alpha,\beta}(0)$, and in practice α and β are unknown, and therefore need to be estimated consistently. We provide simple methods for consistent estimation of these two parameters in the next section. Second, the random variable $\mathbb{X}_{\hat{c},\alpha,\beta}(0)$ does not appear to admit a natural scaling in terms of some canonical random variable: in other words, it cannot be represented as $C(c, \alpha, \beta)J$ where C is an explicit function of c, α, β and J is some fixed well-characterized random variable. Thus, the quantiles of $\mathbb{X}_{\hat{c},\hat{\alpha},\hat{\beta}}$ (where $\hat{\alpha}$ and $\hat{\beta}$ are consistent estimates for the corresponding parameters) need to be calculated by generating many sample paths from the parent process $\mathcal{P}_{\hat{c},\hat{\alpha},\hat{\beta}}$ and computing the left slope of the convex minorant of each

such path at 0. 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 may wonder if resampling the NPMLE would allow adaptation with respect to γ. The problem, however, lies in the fact that while the usual n out of n bootstrap works for the NPMLE when $\gamma \in (0, 1/3)$, it fails under the nonstandard 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 n out of n bootstrap. One could argue that the m out of n bootstrap or subsampling will work irrespective of the value of γ, but the problem that arises here is that these procedures require knowledge of the convergence rate and this is unknown as it depends on the true value of γ.

5. A practical procedure and simulations. In this section, we provide a practical version of the adaptive procedure introduced in Section 4 to construct Wald-type confidence intervals for $F(t_l)$ and assess their performance through simulation studies. The true values of c and γ are denoted by c_0 and γ_0. The process $\mathcal{P}_{c,\alpha,\beta}$ is again abbreviated to $\mathcal{P}_c$.

Recall that in the adaptive procedure, we always specify $\gamma = 1/3$ and compute a surrogate for c_0, namely $\hat{c}$, as a solution of the equation $K = \lfloor (b-a)/\hat{c}n^{-1/3} \rfloor$, where K is the number of grid points. To construct a level $1-2\eta$ confidence interval for $F(t_l)$ for a small positive η, quantiles of $\mathcal{S}_{\hat{c}}$ are needed. Since $\mathcal{S}_c = \mathrm{LS}[\mathrm{GCM}\{\mathcal{P}_c(k), k \in \mathbb{Z}\}](0)$ (c is genetically used), we approximate $\mathcal{S}_c$ with

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

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

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

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

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

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 $\hat{F}(t_l)$, in our simulations we 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 $g(x_0)$, we use the following estimating equation: $(N_{l-j^\star+1} + \cdots + N_{r+j^\star})/n = g(x_0)(t_{r+j^\star} - t_{l-j^\star})$, where $j^\star$ is defined below in the estimation of $f(x_0)$. Since the design

density g is assumed to be continuous in a neighborhood of x_0, and the interval $[t_{l-j^\star}, t_{r+j^\star}]$ is shrinking to x_0, it is reasonable to approximate g over the interval $[t_{l-j^\star}, t_{r+j^\star}]$ with a constant function. Thus, from the above estimating equation, 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 $f(x_0)$. To this end, we estimate $f(t_l)$ using a local linear approximation: 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 by 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. Thus, we first find $j^\star$, the smallest integer such that $\sum_{i=l-j^\star}^{l+j^\star} N_i/n \geq 1/\log n$. Next, we find $i^\star$, the smallest integer larger than $j^\star$ such that $\hat{F}(t_{l-i^\star}) < \hat{F}(t_{l+i^\star})$ and employ a linear approximation over $[t_{l-i^\star}, t_{l+i^\star}]$. More specifically, we compute

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

and 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.

The above procedures provide consistent estimates of $g(x_0)$ and $f(x_0)$ under the assumption of a single derivative for F and G in a neighborhood of x_0, irrespective of the value of γ [since the estimates are obtained by local polynomial fitting over a neighborhood of logarithmic order (in n) around x_0 and such neighborhoods are guaranteed to be asymptotically wider than $n^{-\gamma}$ for any $0 < \gamma \leq 1$]. Two points need to be noted. First, the $1/\log n$ threshold used to determine $j^\star$ in the previous paragraph may need to be changed to a multiple of $1/\log n$, depending on the sample size and the length of the time interval. Second, the locally constant estimate of $g(x_0)$ discussed above could be replaced by a local linear (or quadratic) estimate of g, if the data strongly indicate that G is changing sharply in a neighborhood of x_0.

To evaluate the finite sample performance of the practical adaptive procedure, we also provide simulated confidence intervals of an idealized (theoretical) 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 anchor-point x_0 is 0.5. The pair of grid-parameters (γ, c) takes values $(1/6, 1/6)$, $(1/4, 1/4)$, $(1/3, 1/2)$, $(1/2, 1)$, $(2/3, 2)$ and $(3/4, 3)$. The sample size n ranges 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 $n = 100, 300, 500$. The first panel of Figure 2 shows that both CR(T) and CR(P) are usually close to the nominal level 95% from below and that CR(T) is 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 panel 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*.

We end this section by pointing out that while, for the simulations, we knew the anchor-point x_0 (t_l being the largest grid-point to the left of or equal to x_0), and that we did make use of its value for estimating $F(x_0)$ in our simulations, knowledge of x_0 is not essential to the inference procedure. We could have just estimated $F(x_0)$ by $\hat{F}(t_l)$ [rather than by a convex combination of $\hat{F}(t_l)$ and $\hat{F}(t_r)$ that depends upon x_0] consistently. This is a critical observation, since in a real-life situation what we are provided is current status data on a grid with particular grid points of interest. There is no specification of x_0. To make inference on the value of F at such a grid-point, one can, conceptually, view x_0 as being any point strictly in between the given point and the grid-point immediately after, but its value is not required to construct a confidence interval by the adaptive method. To reiterate, the "anchor-point," x_0 was introduced for developing our theoretical results, but its value can be ignored for the implementation of our method in practice.

6. Concluding discussion. In this paper, we considered maximum likelihood estimation for the event time distribution function, F, at a grid point in the current status model with i.i.d. data and observation times lying on a regular grid. The spacing of the grid δ was specified as $cn^{-\gamma}$ for constants $c > 0$ and $0 < \gamma \leq 1$ in order to incorporate situations where there are systematic ties in observation times,

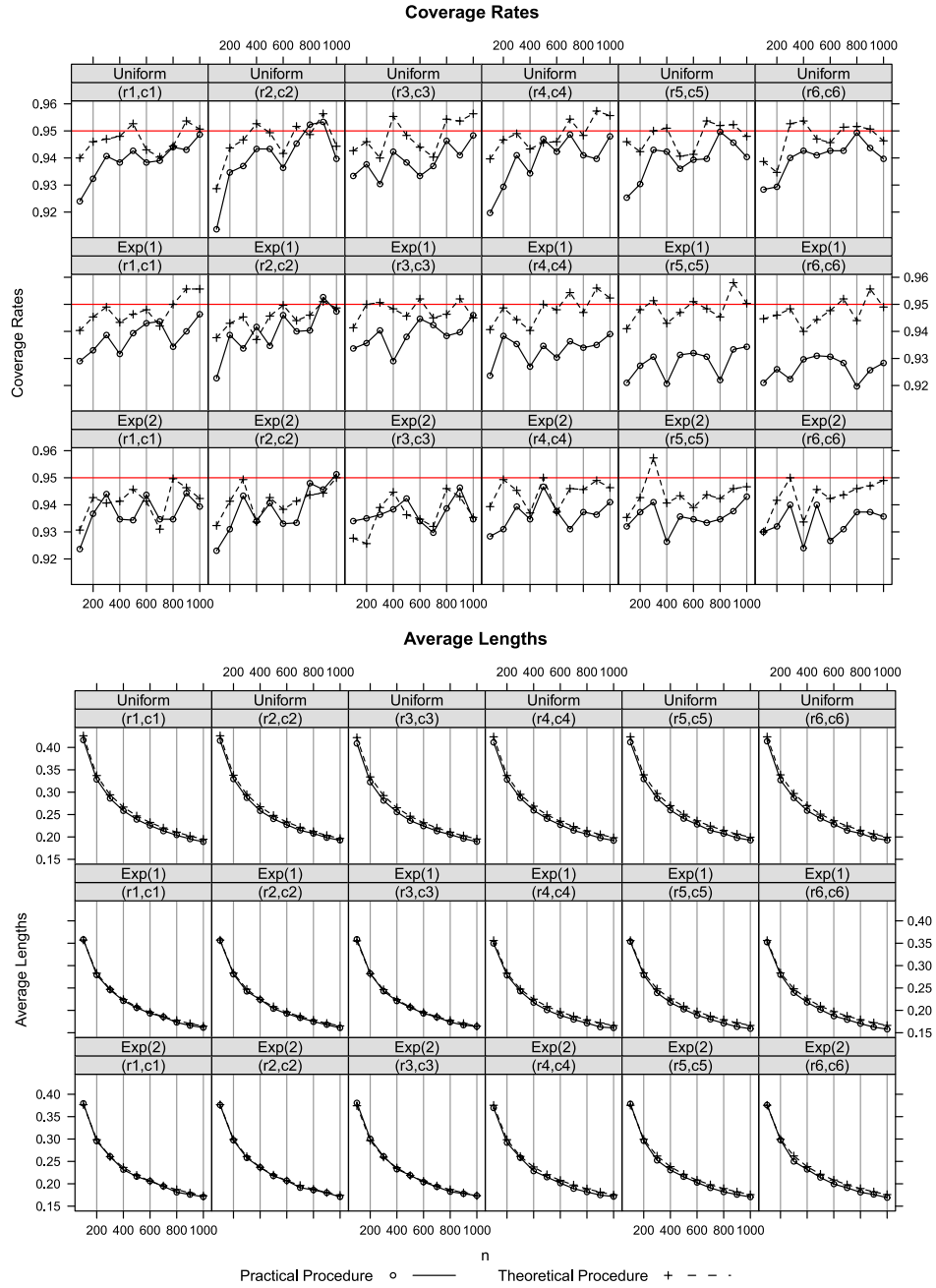

FIG. 2. *A comparison of the coverage rates and average lengths of the practical and theoretical procedures, where* (ri, ci) *for* $i = 1, \ldots, 6$ *are* $(1/6, 1/6)$, $(1/4, 1/4)$, $(1/3, 1/2)$, $(1/2, 1)$, $(2/3, 2)$ *or* $(3/4, 3)$, *respectively. The sample size* n *varies from* 100 *to* 1000 *by* 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)	0.924	0.941	0.943	0.929	0.939	0.939	0.924	0.944	0.934
(1/4, 1/4)	0.914	0.937	0.943	0.923	0.934	0.935	0.923	0.943	0.941
(1/3, 1/2)	0.933	0.930	0.938	0.934	0.940	0.938	0.934	0.936	0.942
(1/2, 1)	0.920	0.941	0.947	0.924	0.935	0.935	0.928	0.939	0.947
(2/3, 2)	0.925	0.943	0.936	0.921	0.931	0.931	0.932	0.941	0.936
(3/4, 3)	0.928	0.940	0.941	0.921	0.922	0.931	0.930	0.940	0.940
CR(T)	$U[0,1]$			**exp(1)**			**exp(2)**		
(1/6, 1/6)	0.940	0.947	0.953	0.940	0.949	0.946	0.931	0.941	0.946
(1/4, 1/4)	0.929	0.947	0.949	0.938	0.945	0.946	0.932	0.949	0.943
(1/3, 1/2)	0.943	0.940	0.948	0.941	0.951	0.946	0.928	0.939	0.936
(1/2, 1)	0.940	0.949	0.946	0.941	0.944	0.950	0.939	0.945	0.950
(2/3, 2)	0.946	0.950	0.941	0.941	0.951	0.947	0.935	0.957	0.943
(3/4, 3)	0.939	0.953	0.947	0.945	0.948	0.944	0.930	0.950	0.946

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)	0.417	0.286	0.239	0.358	0.246	0.206	0.380	0.261	0.216
(1/4, 1/4)	0.415	0.287	0.240	0.356	0.242	0.204	0.376	0.258	0.218
(1/3, 1/2)	0.409	0.281	0.236	0.359	0.243	0.207	0.381	0.258	0.219
(1/2, 1)	0.411	0.287	0.241	0.350	0.243	0.201	0.370	0.258	0.215
(2/3, 2)	0.411	0.286	0.241	0.354	0.239	0.202	0.379	0.253	0.216
(3/4, 3)	0.414	0.287	0.241	0.352	0.239	0.202	0.376	0.250	0.214
AL(T)	$U[0,1]$			**exp(1)**			**exp(2)**		
(1/6, 1/6)	0.426	0.294	0.247	0.357	0.247	0.208	0.377	0.260	0.219
(1/4, 1/4)	0.426	0.295	0.248	0.357	0.247	0.208	0.377	0.261	0.220
(1/3, 1/2)	0.422	0.292	0.246	0.355	0.246	0.208	0.374	0.260	0.219
(1/2, 1)	0.424	0.295	0.249	0.356	0.247	0.209	0.375	0.261	0.220
(2/3, 2)	0.424	0.297	0.251	0.356	0.248	0.209	0.375	0.262	0.221
(3/4, 3)	0.424	0.297	0.251	0.356	0.248	0.209	0.375	0.262	0.221

and the number of distinct observation times can increase with the sample size. The asymptotic properties of the NPMLE were shown to depend on the order of the grid resolution γ and an adaptive procedure, which circumvents the estimation of the unknown γ and c, was proposed for the construction of asymptotically correct confidence intervals for the value of F at a grid-point of interest. We conclude with a description of alternative methods for inference in this problem and potential directions for future research.

Likelihood ratio based inference: An alternative to the Wald-type adaptive confidence intervals proposed in this paper would be to use those obtained via likelihood ratio inversion. More specifically, one could consider testing the null hypothesis H_0 that $F(t_l) = \theta_l$ versus its complement using the likelihood ratio statistics (LRS). When the null hypothesis is true, the LRS converges weakly to χ_1^2 in the limit for $\gamma < 1/3$, to $\mathbb{D}$, the parameter-free limit discovered by Banerjee and Wellner (2001) for $\gamma > 1/3$ and a discrete analog of $\mathbb{D}$ depending on c, α, β, say $\mathcal{M}_{c,\alpha,\beta}$, that can be written in terms of slopes of unconstrained and appropriately constrained convex minorants of the process $\mathcal{P}_{c,\alpha,\beta}$ for $\gamma = 1/3$. Thus, one obtains a boundary distribution for the likelihood ratio statistic as well, and a phenomenon similar to that observed in Section 4 transpires, with the boundary distribution converging to χ_1^2 as $c \to \infty$ and to that of $\mathbb{D}$ as $c \to 0$. An adaptive procedure, which performs an inversion by calibrating the likelihood ratio statistics for testing a family of null hypotheses of the form $F(t_l) = \theta$ for varying θ, using the quantiles of $\mathcal{M}_{\hat{c},\hat{\alpha},\hat{\beta}}$, can also be developed but is computationally more burdensome than the Wald-type intervals. See Tang, Banerjee and Kosorok (2010) for the details.

Smoothed estimators: We recall that all our results have been developed under minimal smoothness assumptions on F: throughout the paper, we assume F to be once continuously differentiable with a nonvanishing derivative around x_0. We used the NPMLE to make inference on F since it can be computed without specifying bandwidths; furthermore, under our minimal assumptions, its pointwise rate of convergence when $\gamma > 1/3$ or when the observation times arise from a continuous distribution cannot be bettered by a smoothed estimator. However, if one makes the assumption of a second derivative at x_0, the kernel-smoothed NPMLE (and related variants) can achieve a convergence rate of $n^{2/5}$ (which is faster than the rate of the NPMLE) using a bandwidth of order $n^{-1/5}$. See Groeneboom, Jongbloed and Witte (2010) where these results are developed and also an earlier paper due to Mammen (1991) dealing with monotone regression. In such a situation, one could envisage using a smoothed version of the NPMLE in this problem with a bandwidth larger than the resolution of the grid, and it is conceivable that an adaptive procedure could be developed along these lines. While this is certainly an interesting and important topic for further exploration, it is outside the scope of this

work, not least owing to the fact that the assumptions underlying such a procedure are different (two derivatives as opposed to one) than those in this paper.

Further possibilities: 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 (2011) deals with the estimation of *competing risks current status data under finitely many risks* with finitely many 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. We suspect that a similar adaptive procedure relying on a boundary phenomenon at $\gamma = 1/3$ can also be developed in this case. Furthermore, one could 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 (2011) 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 and it would not be unreasonable to expect a similar thing to happen for grouped current status data with finitely many competing risks. In fact, an adaptive inference procedure very similar to that in this paper should also work for the problem treated in Zhang, Kim and Woodroofe (2001) and allow inference for the decreasing density of interest without needing to know the rate of growth of the bins.

It is also 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. Furthermore, similar adaptive inference in more complex forms of interval censoring, like Case-2 censoring or mixed-case censoring [see, e.g., Sen and Banerjee (2007) 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 (like convexity, e.g.), though the boundary value of γ as well as the nature of the nonstandard limits will be different and will depend on the "order" of the shape constraint. This will also be a topic of future research.

APPENDIX: PROOFS

PROOF OF THEOREM 4.1. For $k \in \mathbb{Z}$, let

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

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

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

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

$$A_c = \left\{ \frac{h(k)}{ck} < \frac{h(k+1)}{c(k+1)}, k = 1, 2, \ldots \right\},$$

$$B_c = \left\{ \frac{h(-(k-1))}{c(k-1)} < \frac{h(-k)}{ck}, k = 2, 3, \ldots \right\},$$

$$C_c = \left\{ \frac{h(1)}{c} > \frac{-h(-1)}{c} \right\}.$$

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 A.1, $A_c B_c C_c \xrightarrow{P} 1$. 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, $\sqrt{c}\mathcal{S}_c \xrightarrow{d} \alpha Z$. $\square$

LEMMA A.1. *Each of A_c, B_c and C_c in the proof of Theorem 4.1 converges to* 1 *in probability.*

PROOF. It is easy to show C_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 $k \in \mathbb{Z}$,

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

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, e.g., ⟨2⟩ on page 317 of Pollard (2002)]. Then, we have

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

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

PROOF OF THEOREM 4.2. We want to show that $\mathcal{S}_c \xrightarrow{d} g_{\alpha,\beta}(0)$, as $c \to 0$, where $g_{\alpha,\beta}(0) = \mathrm{LS} \circ \mathrm{GCM}\{X_{\alpha,\beta}\}(0) = \mathrm{LS} \circ \mathrm{GCM}\{X_{\alpha,\beta}(t) : t \in \mathbb{R}\}(0)$ and $\mathcal{S}_c = \mathrm{LS} \circ \mathrm{GCM}\{\mathcal{P}_c\}(0) = \mathrm{LS} \circ \mathrm{GCM}\{\mathcal{P}_c(k) : k \in \mathbb{Z}\}(0)$. Since $\mathcal{S}_c = \mathcal{S}'_c + \beta c$, where $\mathcal{S}'_c = \mathrm{LS} \circ \mathrm{GCM}\{\mathcal{P}'_c : k \in \mathbb{Z}\}(0)$ 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, that is, 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 $g^K_{\alpha,\beta}(0) = \mathrm{LS} \circ \mathrm{GCM}\{X_{\alpha,\beta}(t) : t \in [-K, K]\}(0)$ and $\mathcal{S}^K_c = \mathrm{LS} \circ \mathrm{GCM}\{\mathcal{P}_c(t) : t \in [-K, K]\}(0)$.

We will show that, given $\varepsilon > 0$, there exist $M_\varepsilon > 0$ and $c(\varepsilon)$ such that (a) for all $0 < c < c(\varepsilon)$, $P(\mathcal{S}_c^{M_\varepsilon} \neq \mathcal{S}_c) < \varepsilon$ and (b) $P(g_{\alpha,\beta}^{M_\varepsilon}(0) \neq g_{\alpha,\beta}(0)) < \varepsilon$. These immediately imply that both Fact 1: $\lim_{\varepsilon \to 0} \limsup_{c \to 0} P(\mathcal{S}_c^{M_\varepsilon} \neq \mathcal{S}_c) = 0$ and Fact 2: $\lim_{\varepsilon \to 0} P(g_{\alpha,\beta}^{M_\varepsilon}(0) \neq g_{\alpha,\beta}(0)) = 0$ hold. We then show that Fact 3: for each $\varepsilon > 0$, $\mathcal{S}_c^{M_\varepsilon} \xrightarrow{d} g_{\alpha,\beta}^{M_\varepsilon}(0)$ holds as well. Then, by Lemma 3.9, we have the conclusion $\mathcal{S}_c \xrightarrow{d} g_{\alpha,\beta}(0)$. Figure 3 illustrates the following argument.

Let $\tau_{-2} < \tau_{-1} < \tau_1 < \tau_2$ be four consecutive slope changing points of $G_{\alpha,\beta} = \mathrm{GCM}\{X_{\alpha,\beta}\}$ 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

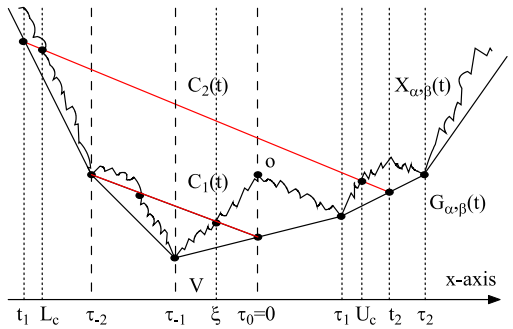

FIG. 3. *An illustration for showing* $\{L_c\}$ *is* $O_P(1)$ *in the proof of Theorem* 4.2.

$\varepsilon > 0$, there exists $M_\varepsilon > 0$ such that $P(-M_\varepsilon < \tau_{-2} < \tau_2 < M_\varepsilon) > 1 - \varepsilon/4$. Note that the event $\{g_{\alpha,\beta}^{M_\varepsilon}(0) = g_{\alpha,\beta}(0)\} \subset \{-M_\varepsilon < \tau_{-2} < \tau_2 < M_\varepsilon\}$, and it follows that $P(g_{\alpha,\beta}^{M_\varepsilon}(0) \neq g_{\alpha,\beta}(0)) < \varepsilon/4 < \varepsilon$. 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 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 nonempty), $\tau_{-1} < \xi < 0$, $C_1(\xi) = X_{\alpha,\beta}(\xi)$ 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 $\delta(\varepsilon) > 0$ such that $P(V > \delta(\varepsilon)) \geq 1 - \varepsilon/4$. Then, the event $E_\varepsilon = \{V > \delta(\varepsilon)\} \cap \{-M_\varepsilon < \tau_{-2}\}$ has probability larger than $1 - \varepsilon/2$. For any $c < c(\varepsilon) =: \delta(\varepsilon)$, 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 $c < c(\varepsilon)$, $P(-M_\varepsilon < L_c) > 1 - \varepsilon/2$. 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), $P(U_c < M_\varepsilon) > 1 - \varepsilon/2$. Hence $P(-M_\varepsilon < L_c < U_c < M_\varepsilon) > 1 - \varepsilon$. On this event, clearly $\mathcal{S}_c^{M_\varepsilon} = \mathcal{S}_c$, and it follows that for all $c < c(\varepsilon)$, $P(\mathcal{S}_c^{M_\varepsilon} \neq \mathcal{S}_c) < \varepsilon$. 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 $[\pm 2M_\varepsilon]$. It follows by elementary analysis that (for almost every ω) on $[\pm M_\varepsilon]$, the process $\mathcal{P}_c$, being the linear interpolant of the points $\{ck, X_{\alpha,\beta}(ck) : -M_\varepsilon \leq ck \leq M_\varepsilon\} \cup \{(-M_\varepsilon, \mathcal{P}_{2c}(-M_\varepsilon)), (M_\varepsilon, \mathcal{P}_{2,c}(M_\varepsilon))\}$, converges uniformly to $X_{\alpha,\beta}$ as $c \to 0$. Thus, the left slope of the GCM of $\{\mathcal{P}_c(t) : t \in [\pm M_\varepsilon]\}$, which is precisely $\mathcal{S}_c^{M_\varepsilon}$, converges to $g_{\alpha,\beta}^{M_\varepsilon}(0)$ since the GCM of the restriction of $X_{\alpha,\beta}$ to $[\pm M_\varepsilon]$ 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. □

Acknowledgments. 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.

SUPPLEMENTARY MATERIAL

More proofs for the current paper "Likelihood based inference for current status data on a grid: A boundary phenomenon and an adaptive inference procedure" (DOI: 10.1214/11-AOS942SUPP; .pdf). The supplementary material contains the details of the proofs of several theorems and lemmas in Sections 3.1 and 3.3 of this paper.

REFERENCES

ABREVAYA, J. and HUANG, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica* **73** 1175–1204. MR2149245

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

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

BANERJEE, M. and WELLNER, J. A. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29** 1699–1731. MR1891743

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

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

GROENEBOOM, P., JONGBLOED, G. and WITTE, B. I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *Ann. Statist.* **38** 352–387. MR2589325

GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. *DMV Seminar* **19**. Birkhäuser, Basel. MR1180321

KEIDING, N., BEGTRUP, K., SCHEIKE, T. H. and HASIBEDER, G. (1996). Estimation from current status data in continuous time. *Lifetime Data Anal.* **2** 119–129.

KIEFER, J. and WOLFOWITZ, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahrsch. Verw. Gebiete* **34** 73–85. MR0397974

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. MR1041391

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

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

MAATHUIS, M. H. and HUDGENS, M. G. (2011). Nonparametric inference for competing risks current status data with continuous, discrete or grouped observation times. *Biometrika* **98** 325–340. MR2806431

MAMMEN, E. (1991). Estimating a smooth monotone regression function. *Ann. Statist.* **19** 724–740. MR1105841

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

PRAKASA RAO, B. L. S. (1969). Estkmation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. MR0267677

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. Wiley, Chichester. MR0961262

SCHICK, A. and YU, Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scand. J. Stat.* **27** 45–55. MR1774042

SEN, B. and BANERJEE, M. (2007). A pseudolikelihood method for analyzing interval censored data. *Biometrika* **94** 71–86. MR2307901

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

TANG, R., BANERJEE, M. and KOSOROK, M. R. (2010). Asymptotics for current status data under varying observation time sparsity. Available at www.stat.lsa.umich.edu/~moulib/csdgriddec23.pdf.

TANG, R., BANERJEE, M. and KOSOROK, M. R. (2011). Supplement to "Likelihood based inference for current status data on a grid: A boundary phenomenon and an adaptive inference procedure." DOI:10.1214/11-AOS942SUPP.

VAN DER VAART, A. (1991). On differentiable functionals. *Ann. Statist.* **19** 178–204. MR1091845

VAN DER VAART, A. W. and VAN DER LAAN, M. J. (2003). Smooth estimation of a monotone density. *Statistics* **37** 189–203. MR1986176

WELLNER, J. A. and ZHANG, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28** 779–814. MR1792787

WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448. MR0606630

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. *Canad. J. Statist.* **26** 619–627. MR1671976

ZHANG, R., KIM, J. and WOODROOFE, M. (2001). Asymptotic analysis of isotonic estimation for grouped data. *J. Statist. Plann. Inference* **98** 107–117. MR1860229

R. TANG
DEPARTMENT OF OPERATIONS RESEARCH
AND FINANCIAL ENGINEERING
PRINCETON UNIVERSITY
214 SHERRERD HALL, CHARLTON STREET
PRINCETON, NEW JERSEY 08544
USA
E-MAIL: runlongt@princeton.edu

M. BANERJEE
DEPARTMENT OF STATISTICS
UNIVERSITY OF MICHIGAN, ANN ARBOR
439 WEST HALL
1085 SOUTH UNIVERSITY
ANN ARBOR, MICHIGAN 48109
USA
E-MAIL: moulib@umich.edu

M. R. KOSOROK
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL
3101 MCGAVRAN-GREENBERG HALL
CB 7420
CHAPEL HILL, NORTH CAROLINA 27599
USA
E-MAIL: kosorok@unc.edu