 Vol. 22: 3–33, 1995

Estimation of a Monotone Density or Monotone Hazard Under Random Censoring

JIAN HUANG and JON A. WELLNER

University of Iowa and University of Washington

ABSTRACT. Consider non-parametric estimation of a decreasing density function f under the random (right) censorship model. Alternatively, consider estimation of a monotone increasing (or decreasing) hazard rate λ based on randomly right censored data. We show that the non-parametric maximum likelihood estimator of the density f (introduced by Laslett, 1982) is asymptotically equivalent to the estimator obtained by differentiating the least concave majorant of the Kaplan–Meier estimator, the non-parametric maximum likelihood estimator of the distribution function F in the larger model without any monotonicity assumption. A similar result is shown to hold for the non-parametric maximum likelihood estimator of an increasing hazard rate λ: the non-parametric maximum likelihood estimator of λ (introduced in the uncensored case by Prakasa Rao, 1970) is asymptotically equivalent to the estimator obtained by differentiation of the greatest convex minorant of the Nelson–Aalen estimator, the non-parametric maximum likelihood estimator of the cumulative hazard function Λ in the larger model without any monotonicity assumption. In proving these asymptotic equivalences, we also establish the asymptotic distributions of the different estimators at a fixed point at which the monotonicity assumption is strictly satisfied.

Key words: Asymptotic equivalence, argmax continuous mapping, asymptotic distribution, density, empirical process, greatest convex minorant, hazard function, Kaplan–Meier estimator, least concave majorant, monotone, Nelson–Aalen estimator, non-parametric maximum likelihood, weak approximation

1. Introduction

Suppose that $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. pairs of random variables where X and Y are independent, X is a failure time with distribution function F on $\mathbb{R}^+ = [0, \infty)$, and Y is a censoring time with distribution function G. Under the random censorship model, the observed data are $(T_1, \delta_1), \ldots, (T_n, \delta_n)$, where $T_i = \min\{X_i, Y_i\}$, $\delta_i = 1_{\{X_i \leqslant Y_i\}}$, $i = 1, 2, \ldots, n$. If F is absolutely continuous with density f, let λ be the hazard (rate) function of X, i.e., $\lambda(x) = f(x)/(1 - F(x))$, and let Λ be the corresponding cumulative hazard function, $\Lambda(x) = \int_0^x \lambda(t)\,dt$.

Our goal is to study non-parametric estimators of the density f and the hazard λ under the following monotonicity assumptions:

(1) f decreasing (non-increasing) on $\mathbb{R}^+$;
(2) λ increasing on $\mathbb{R}^+$;
(3) λ decreasing on $\mathbb{R}^+$.

When there is no censoring, Grenander (1956) showed that the non-parametric maximum likelihood estimator (NPMLE) of a decreasing density f coincides with the left continuous slope of the least concave majorant of the empirical distribution function. Its asymptotic distribution at a fixed point was studied by Prakasa Rao (1969) and Groeneboom (1985). Groeneboom (1985) also derived the limiting distribution of the L_1 distance between the Grenander maximum likelihood estimator and a decreasing density f.

In the case of a monotone decreasing density function f, non-parametric maximum likelihood estimation of f under random censoring has been considered by Laslett (1982) and independently by McNichols & Padgett (1982). Laslett (1982) considered non-parametric maximum likelihood estimation of f under random censoring as a special case of a more general line segment model involving both censoring and biased sampling; see his sects. 2 and 3. Computation of the NPMLE and its asymptotic behavior at a point have recently been considered by Huang & Zhang (1994).

For a monotone hazard function λ, Prakasa Rao (1970) studied the asymptotic distribution of the NPMLE at a fixed point in the case of no censoring, and the estimator was extended to the context of right censoring by Padgett & Wei (1980), but without further development of its asymptotic properties. Mykytyn & Santner (1981) also studied non-parametric maximum likelihood estimation based on monotonicity assumptions concerning the hazard rate λ and several different censoring schemes. They prove strong consistency of the resulting estimator of λ and $1-F$ under random censoring. None of these authors consider the natural estimators based on concave majorants (or convex minorants) of the product limit estimator of F and the Nelson–Aalen estimator of Λ, or the relationship of these estimators to the NPMLEs.

We consider two ways of estimating the density or the hazard under a monotonicity assumption. One way is via the non-parametric maximum likelihood estimator, i.e., maximizing the likelihood function with respect to the density or hazard under monotonicity constraints. We will denote these estimators by $\hat{f}_n$ and $\hat{\lambda}_n$ respectively.

Another approach is based on the product limit estimator $\mathbb{F}_n$ of F or the Nelson–Aalen estimator Λ_n of Λ; see e.g. Lo & Phadia (1992) who follow the framework of Kiefer & Wolfowitz (1976) and Wang (1986). Their focus is on the estimators of F or Λ. To estimate a decreasing density or hazard function, one possibility is to form the least concave majorant (LCM), $\tilde{\mathbb{F}}_n$ (or $\tilde{\Lambda}_n$), of $\mathbb{F}_n$ (or Λ_n), and take the right derivative $\tilde{f}_n$ or $(\tilde{\lambda}_n)$ of $\tilde{\mathbb{F}}_n$ (or $\tilde{\Lambda}_n$) as the estimator of f (or λ). Similarly, to estimate an increasing hazard function, we form the greatest convex minorant (GCM), $\tilde{\Lambda}_n$ of Λ_n, and take the left derivative of $\tilde{\Lambda}_n$ as the estimator of λ.

Our main goal in this paper is to show that these two approaches are asymptotically equivalent. Section 2 contains the statements of our main results. Characterizations of the estimators and a brief discussion of computational methods are given in section 3. Section 4 gives consistency results as well as $\sqrt{n}$-consistency of the non-parametric MLE $\hat{\mathbb{F}}_n$ of F, a result due to Huang & Zhang (1994). Proofs of the main results in section 2 are given in sections 5, 6, and 7 based on an extension of the argmax continuous mapping theorem of Kim & Pollard (1990). In section 8 we compute and compare the estimators graphically for simulated data with sample sizes 100 and 800.

2. Main results

In this section we state our main results. The non-parametric maximum likelihood estimates (NPMLEs) $\hat{f}_n$ and $\hat{\lambda}_n$ of f and λ are characterized in section 3. Proofs will be given in sections 4–7.

Let $H(t)=P\{X\wedge Y\leqslant t\}=1-(1-F(t))(1-G(t))$, $\tau_H=\inf\{t: H(t)=1\}$. We assume G is absolutely continuous.

Theorem 2.1

Suppose f is decreasing on $[0,\infty)$, and $f'(t)$ is continuous in a neighborhood of $t_0\in(0,\tau_H)$, $f(t_0)\neq 0$, $f'(t_0)<0$. Let $\hat{f}_n$ be the NPMLE of f, and let $\tilde{f}_n$ be the left continuous slope of $\tilde{\mathbb{F}}_n$. Then

$$n^{1/3}\left|2\frac{1-G(t_0)}{f(t_0)f'(t_0)}\right|^{1/3}(\hat{f}_n(t_0)-f(t_0))\xrightarrow{d}2Z, \tag{1}$$

where $Z = \operatorname{argmax}(\mathbf{W}(t) - t^2)$, *and* $\mathbf{W}$ *is standard two-sided Brownian motion on* $\mathbb{R}$ *starting at zero. Moreover*

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

so the convergence in distribution in (1) *also holds with* $\hat{f}_n$ *replaced by* $\tilde{f}_n$.

The $\hat{f}_n$ part of the theorem is due to Huang & Zhang (1994). For monotone hazard functions λ, we have the following two theorems.

Theorem 2.2
Suppose $\tau_H = \tau_G < \tau_F$, λ *is increasing on* $[0, \infty)$, *and* $\lambda'(t)$ *is continuous in a neighborhood of* $t_0 \in (0, \tau_H)$, $\lambda(t_0) \neq 0$, $\lambda'(t_0) > 0$. *Let* $\hat{\lambda}_n$ *be the NPMLE of* λ, *and let* $\tilde{\lambda}_n$ *be the right continuous slope of* $\tilde{\Lambda}_n$. *Then*

$$n^{1/3}\left\{2\frac{1 - H(t_0)}{\lambda(t_0)\lambda'(t_0)}\right\}^{1/3}(\hat{\lambda}_n(t_0) - \lambda(t_0)) \xrightarrow{d} 2Z, \tag{2}$$

where Z *is the same as in the previous theorem. Moreover*

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

so the convergence in distribution in (2) *also holds with* $\hat{\lambda}_n$ *replaced by* $\tilde{\lambda}_n$.

Theorem 2.3
Suppose $\tau_H = \tau_G < \tau_F$, λ *is decreasing on* $[0, \infty)$, *and* $\lambda'(t)$ *is continuous in a neighborhood of* $t_0 \in (0, \tau_H)$. $\lambda(t_0) \neq 0$, $\lambda'(t_0) < 0$. *Let* $\hat{\lambda}_n$ *be the NPMLE of* λ, *and let* $\tilde{\lambda}_n$ *be the left continuous slope of* $\tilde{\Lambda}_n$. *Then*

$$n^{1/3}\left|2\frac{1 - H(t_0)}{\lambda(t_0)\lambda'(t_0)}\right|^{1/3}(\hat{\lambda}_n(t_0) - \lambda(t_0)) \xrightarrow{d} 2Z, \tag{3}$$

where Z *is the same as in the previous theorem. Moreover*

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

so the convergence in distribution in (3) *also holds with* $\hat{\lambda}_n$ *replaced by* $\tilde{\lambda}_n$.

Remark 2.1. When specialized to the case of no censoring, theorem 2.1 agrees with the result of Groeneboom (1985), and theorems 2.2 and 2.3 agree with ths. 6.1 and 7.1 of Prakasa Rao (1970). □

Remark 2.2. The distribution of Z has been completely characterized by Groeneboom (1989). □

Remark 2.3. All of the above results assert that $\hat{f}_n$ and $\tilde{f}_n$ (or $\hat{\lambda}_n$ and $\tilde{\lambda}_n$) are asymptotically equivalent at the $n^{1/3}$ scaling level when the true density f is decreasing (or the true hazard λ is increasing or decreasing). Similarly, the consistency results to be established in section 4 assert that $\hat{\mathbb{F}}_n$, $\tilde{\mathbb{F}}_n$, and $\mathbb{F}_n$ are all asymptotically equivalent at the $n^{1/2}$ scaling level where $\hat{\mathbb{F}}_n(t) \equiv \int_0^t \hat{f}_n(s)\,ds$. It would be interesting to study the differences between these estimators. It follows along the lines of Wang (1992) that

$$n^{2/3}(\tilde{\mathbb{F}}_n - \mathbb{F}_n)(t_0) \xrightarrow{d} \left(\frac{2f^2(t_0)}{-f'(t_0)\bar{G}^2(t_0)}\right)^{1/3} C(0)$$

where $C(0)$ is the concave majorant of two-sided Brownian motion minus a parabola at 0. Note that $C(0) \geqslant 0$ a.s. We conjecture that

$$n^{2/3}(\hat{\mathbb{F}}_n - \mathbb{F}_n)(t_0) \xrightarrow{d} D$$

where the random variable D is $\leqslant 0$ with positive probability. Other ways of examining the differences may also be of interest. □

Remark 2.4. Note that the density and hazard rate estimates studied here do not require the choice of a bandwidth parameter. Other methods, including density and hazard rate estimates based on kernel smoothing or splines can be used to exploit additional smoothness assumptions (if they are warranted) to achieve faster convergence rates; e.g. mean square errors of $n^{-4/5}$ in contrast to the $n^{-2/3}$ rates for the estimators studied here. However, the gain in rate comes at the price of additional smoothness hypotheses, the validity of these hypotheses for the true population, and the need to specify a bandwidth parameter. We prefer to use estimation methods based on minimal smoothness hypotheses, and do not yet know how to exploit both monotonicity and smoothness hypotheses in combination. For more on kernel estimates and other smoothing methods we refer the interested reader to Andersen *et al.* (1992, pp. 224–227) and Cox & O'Sullivan (1990).

3. Characterization of the NPMLEs $\hat{f}_n$ and $\hat{\lambda}_n$

If G is absolutely continuous with density g, then the density of (T, δ) with respect to the product of Lebesgue measure on R^+ and counting measure on $\{0, 1\}$ is

$$p(t, \delta) = (f(t)\bar{G}(t))^{\delta}(g(t)\bar{F}(t))^{1-\delta}, \quad 0 \leqslant t < \infty, \delta \in \{0, 1\};$$

here $\bar{G}(t) \equiv 1 - G(t)$ and, similarly, $\bar{F}(t) \equiv 1 - F(t)$. We use this notation throughout for the survival function associated with a distribution function F. The log-likelihood for the observed data is

$$l(\lambda) = \sum_{i=1}^{n} \{\delta_i \log(f(T_i)) + (1 - \delta_i) \log(1 - F(T_i))\} + C(g, G),$$

where $C(g, G)$ is a term not involving f or F. Hence we can treat it as a constant term and will not consider it in the following.

3.1. *Characterization of the NPMLE* $\hat{f}_n$

Let $T_{(1)} \leqslant \cdots \leqslant T_{(n)}$ be the order statistics of $T_1, \ldots, T_n$, and let $\delta_{(i)}$ be the δ corresponding to $T_{(i)}$, i.e., if $T_{(i)} = T_j$, then $\delta_{(i)} = 1\{X_j \leqslant Y_j\}$. Define $T_{(0)} = 0$.

The NPMLE of f under monotonicity constraints is the solution to the following maximization problem: maximize

$$\phi(f) = \sum_{i=1}^{n} [\delta_{(i)} \log f(T_{(i)}) + (1 - \delta_{(i)}) \log(1 - F(T_{(i)}))]$$

over the convex set

$$C = \left\{ z = (z_1, \ldots, z_n): z_1 \geqslant \cdots \geqslant z_n \geqslant 0, \sum_{i=1}^{n} z_i(T_{(i)} - T_{(i-1)}) \leqslant 1 \right\},$$

where $z_i = f(T_{(i)})$, and $F(T_{(i)}) = \Sigma_{j=1}^{i} f(T_{(j)})(T_{(j)} - T_{(j-1)})$. For vectors z such that

$F(T_{(i)}) > 1$, define the objective function to be $-\infty$. In this way, we obtain a concave objective function, defined on the set C.

Define the process by

$$W_f(t) = \int \left[\delta \frac{\{t' \leqslant t\}}{f(t')} + (1-\delta) \frac{\{t' \leqslant t\}}{1 - F(t')}(t - t') - t \right] d\mathbb{P}_n(t', \delta),$$

where $F(t) = \int_0^t f(s)\, ds$, $t \geqslant 0$. Then the following theorem characterizes the NPMLE of f.

Theorem 3.1

The left continuous, decreasing, and non-negative function $\hat{f}_n$ is an NPMLE of the unknown density f if and only if

$$W_{\hat{f}_n}(t) \leqslant 0, \quad \textit{for all } t \geqslant 0, \tag{4}$$

and either

$$\hat{\mathbb{F}}_n(T_{(n)}) < 1 \quad \textit{and} \quad \int \frac{1-\delta}{1 - \hat{\mathbb{F}}_n(t')} d\mathbb{P}_n(t', \delta) = 1, \tag{5}$$

or

$$\hat{\mathbb{F}}_n(T_{(n)}) = 1 \quad \textit{and} \quad \int \frac{1-\delta}{1 - \hat{\mathbb{F}}_n(t')} d\mathbb{P}_n(t', \delta) \leqslant 1. \tag{6}$$

Proof. Sufficiency. Suppose (4) and (5) hold. Then we do not have an active constraint (see e.g. Fletcher, 1987, p. 142, for this terminology). Relation (4) can be written as

$$\sum_{T_{(i)} \leqslant t} \frac{\delta_{(i)}}{\hat{f}_n(T_{(i)})} - \sum_{T_{(j)} \geqslant T_{(i)},\, T_{(i)} \leqslant t} \frac{1 - \delta_{(j)}}{1 - \hat{\mathbb{F}}_n(T_{(j)})}(T_{(i)} - T_{(i-1)}) \leqslant 0, \tag{7}$$

for each t, with equality if t is a point of jump of f. This is because, using the second part of (5),

$$\begin{aligned}
nW_{\hat{f}_n}(t) &= \sum_{T_{(i)} \leqslant t} \frac{\delta_{(i)}}{\hat{f}_n(T_{(i)})} + \sum_{T_{(i)} \leqslant t} \frac{1 - \delta_{(i)}}{1 - \hat{\mathbb{F}}_n(T_{(i)})}(t - T_{(i)}) - nt \\
&= \sum_{T_{(i)} \leqslant t} \frac{\delta_{(i)}}{\hat{f}_n(T_{(i)})} - \sum_{T_{(i)} \leqslant t} \frac{1 - \delta_{(i)}}{1 - \hat{\mathbb{F}}_n(T_{(i)})} T_{(i)} + \sum_{T_{(i)} \leqslant t} \frac{1 - \delta_{(i)}}{1 - \hat{\mathbb{F}}_n(T_{(i)})} t - nt \\
&= \sum_{T_{(i)} \leqslant t} \frac{\delta_{(i)}}{\hat{f}_n(T_{(i)})} - \sum_{T_{(i)} \leqslant t} \frac{1 - \delta_{(i)}}{1 - \hat{\mathbb{F}}_n(T_{(i)})} T_{(i)} - \sum_{T_{(i)} > t} \frac{1 - \delta_{(i)}}{1 - \hat{\mathbb{F}}_n(T_{(i)})} t \\
&= \sum_{T_{(i)} \leqslant t} \frac{\delta_{(i)}}{\hat{f}_n(T_{(i)})} - \sum_{T_{(i)} \leqslant T_{(j)} \leqslant t} \frac{1 - \delta_{(j)}}{1 - \hat{\mathbb{F}}_n(T_{(j)})}(T_{(i)} - T_{(i-1)}) \\
&\quad - \sum_{T_{(i)} \leqslant t,\, T_{(j)} > t} \frac{1 - \delta_{(j)}}{1 - \hat{\mathbb{F}}_n(T_{(j)})}(T_{(i)} - T_{(i-1)}) \\
&= \sum_{T_{(i)} \leqslant t} \frac{\delta_{(i)}}{\hat{f}_n(T_{(i)})} - \sum_{T_{(j)} \geqslant T_{(i)},\, T_{(i)} \leqslant t} \frac{1 - \delta_{(j)}}{1 - \hat{\mathbb{F}}_n(T_{(j)})}(T_{(i)} - T_{(i-1)}).
\end{aligned}$$

Furthermore, the second part of (5) is equivalent to

$$\sum_{i=1}^{n} \left[\delta_{(i)} - \sum_{T_{(j)} \geqslant T_{(i)}} \frac{1 - \delta_{(j)}}{1 - \hat{\mathbb{F}}_n(T_{(j)})} \hat{f}_n(T_{(i)})(T_{(i)} - T_{(i-1)}) \right] = 0, \tag{8}$$

since

$$\sum_{i=1}^{n}\left[1-\frac{1-\delta_{(i)}}{1-\hat{\mathbb{F}}_n(T_{(i)})}\right]=\sum_{i=1}^{n}\left[\delta_{(i)}+1-\delta_{(i)}-\frac{1-\delta_{(i)}}{1-\hat{\mathbb{F}}_n(T_{(i)})}\right]$$
$$=\sum_{i=1}^{n}\left[\delta_{(i)}-\frac{1-\delta_{(i)}}{1-\hat{\mathbb{F}}_n(T_{(i)})}\hat{\mathbb{F}}_n(T_{(i)})\right]$$
$$=\sum_{i=1}^{n}\left[\delta_{(i)}-\sum_{T_{(j)}\geqslant T_{(i)}}\frac{1-\delta_{(j)}}{1-\hat{\mathbb{F}}_n(T_{(j)})}\hat{f}_n(T_{(i)})(T_{(i)}-T_{(i-1)})\right].$$

We show that relations (7) and (8) are sufficient conditions for $\hat{f}_n$ to be a maximizing function. Suppose f_* is another decreasing density function. Since ϕ is concave on the set C, we have

$$\phi(f_*)-\phi(\hat{f}_n)\leqslant\langle\nabla\phi(\hat{f}_n),\hat{f}_n-f_*\rangle,$$

where $\nabla\phi(\hat{f}_n)$ is the vector of partial derivatives

$$\nabla\phi(\hat{f}_n)=\left(\frac{\delta_{(i)}}{\hat{f}_n(T_{(i)})}-\sum_{j\geqslant i}\frac{1-\delta_{(j)}}{1-\hat{\mathbb{F}}_n(T_{(j)})}(T_{(i)}-T_{(i-1)}),\ i=1,\ldots,n\right).$$

Here we also use $\hat{f}_n$ and f_* to denote the vectors $(\hat{f}_n(T_{(1)}),\ldots,\hat{f}_n(T_{(n)}))$ and $(f_*(T_{(1)}),\ldots,f_*(T_{(n)}))$, respectively. In addition, if f satisfies (8), $\langle\nabla\phi(\hat{f}_n),\hat{f}_n\rangle=0$. Next notice that any decreasing f_* can be written as

$$f_*=\sum_{i=1}^{n}\beta_iI_i,$$

where $\beta_i=f_*(T_{(i)})-f_*(T_{(i+1)})$, $i=1,\ldots,n-1$, $\beta_n=f_*(T_{(n)})$, and where I_i is a vector which has ones as its first i components and zeros as its last $n-i$ components. Hence, by (7),

$$\langle\nabla\phi(\hat{f}_n),f_*-\hat{f}_n\rangle=\langle\nabla\phi(\hat{f}_n),f_*\rangle=\sum_{i=1}^{n}\beta_i\langle\nabla\phi(\hat{f}_n),I_i\rangle=\sum_{i=1}^{n}\beta_iW_{\hat{f}_n}(T_{(i)})\leqslant 0.$$

Thus (4) and (5) imply $\hat{f}_n$ maximizes ϕ.

Next suppose that (4) and (6) are satisfied for $\hat{f}_n$. Introducing the constraint by using a Lagrange multiplier, we can take the objective function as

$$\sum_{i=1}^{n}[\delta_{(i)}\log f(T_{(i)})+(1-\delta_{(i)})\log(1-F(T_{(i)}))]-\alpha\left[\sum_{i=1}^{n}f(T_{(i)})(T_{(i)}-T_{(i-1)})-1\right].\qquad(9)$$

Defining α by

$$\alpha=\sum_{i=1}^{n}\left[\delta_{(i)}-\sum_{T_{(j)}\geqslant T_{(i)}}\frac{1-\delta_{(j)}}{1-\hat{\mathbb{F}}_n(T_{(j)})}(T_{(i)}-T_{(i-1)})\right].\qquad(10)$$

Then by (4) and (10), $\hat{f}_n$ satisfies

$$\sum_{T_{(i)}\leqslant T_{(k)}}\frac{\delta_{(i)}}{\hat{f}_n(T_{(i)})}-\sum_{T_{(j)}\geqslant T_{(i)},\,T_{(i)}\leqslant T_{(k)}}\frac{1-\delta_{(j)}}{1-\hat{\mathbb{F}}_n(T_{(j)})}(T_{(i)}-T_{(i-1)})\leqslant\alpha T_{(k)},\qquad(11)$$

for each $T_{(k)}$. This means that f maximizes (9) over the set C. Since the second part of condition (6) implies $\alpha\geqslant 0$, this means that $\hat{f}_n$ also maximizes the log likelihood over the set of functions g satisfying

$$\sum_{i=1}^{n}g(T_{(i)})(T_{(i)}-T_{(i-1)})\leqslant 1.$$

Conversely, suppose $\hat{f}_n$ is a maximization function. If $\hat{\mathbb{F}}_n(T_{(n)}) < 1$, the necessity of the conditions may be proved as in prop. 1.1, p. 39, of Groeneboom & Wellner (1992). So we consider the case $\hat{\mathbb{F}}_n(T_{(n)}) = 1$. Recall the objective function

$$\phi(f) = \sum_{i=1}^{n} \delta_{(i)} \log f(T_{(i)}) + \sum_{i=1}^{n} (1-\delta_{(i)}) \log(1 - F(T_{(i)})).$$

Then by the definition of an NPMLE,

$$0 \geqslant \lim_{\varepsilon \downarrow 0} \varepsilon^{-1}[\phi((1-\varepsilon)\hat{f}_n) - \phi(\hat{f}_n)] = -\sum_{i=1}^{n} \delta_{(i)} + \sum_{i=1}^{n} (1-\delta_{(i)}) \frac{\hat{\mathbb{F}}_n(T_{(i)})}{1-\hat{\mathbb{F}}_n(T_{(i)})}$$

$$= -n + \sum_{i=1}^{n} \frac{1-\delta_{(i)}}{1-\hat{\mathbb{F}}_n(T_{(i)})}.$$

This yields (6). Finally, relation (4) follows from

$$\lim_{\varepsilon \downarrow 0} \varepsilon^{-1}[\phi(f_{\varepsilon,i}) - \phi(\hat{f}_n)] \leqslant 0, \quad i = 1, \ldots, n,$$

where

$$f_{\varepsilon,i}(t) = \begin{cases} (\hat{f}_n(t) + \varepsilon)/(1 + \varepsilon T_{(i)}), & t \leqslant T_{(i)}, \\ \hat{f}_n(t), & t > T_{(i)}. \end{cases} \qquad \square$$

Define a process V_n^F by

$$V_n^F(t) \equiv \int \left[\delta\{t' \leqslant t\} + (1-\delta)\{t' \leqslant t\} \frac{F(t) - F(t'-)}{1 - F(t'-)} \right] d\mathbb{P}_n(t', \delta)$$

$$= \mathbb{H}_n(t) - \int_{[0,t]} \frac{\bar{F}(t)}{\bar{F}(t'-)} \, d\mathbb{H}_n^c(t'). \tag{12}$$

Then an alternative form of (4) is given by

$$V_n^{\hat{\mathbb{F}}_n}(t) \leqslant \hat{\mathbb{F}}_n(t), \quad t \geqslant 0. \tag{13}$$

To see that this is indeed true, let $\tau_1 < \tau_2 < \cdots < \tau_m$ be the jump points of $\hat{f}_n$. Then we have

$$W_{\hat{f}_n}(\tau_i) = 0, \quad i = 1, \ldots, m.$$

For $0 < t \leqslant \tau_1$, we have $\hat{f}_n(t) = \hat{f}_n(\tau_1)$, so

$$\sum_{T_{(i)} \leqslant t} \left[\frac{\delta_{(i)}}{\hat{f}_n(T_{(i)})} + \sum_{T_{(i)} \leqslant t} \frac{1-\delta_{(i)}}{1-\hat{\mathbb{F}}_n(T_{(i)})} (t - T_{(i)}) - nt \right] \leqslant 0$$

is equivalent to

$$\sum_{T_{(i)} \leqslant t} \left[\delta_{(i)} + \sum_{T_{(i)} \leqslant t} \frac{1-\delta_{(i)}}{1-\hat{\mathbb{F}}_n(T_{(i)})} (\hat{\mathbb{F}}_n(t) - \hat{\mathbb{F}}_n(T_{(i)})) \right] \leqslant n\hat{\mathbb{F}}_n(t),$$

i.e.,

$$V_n^{\hat{\mathbb{F}}_n}(t) \leqslant \hat{\mathbb{F}}_n(t).$$

For $\tau_1 < t \leqslant \tau_2$, using $W_{\hat{f}_n}(\tau_1) = 0$, we have

$$W_{\hat{f}_n}(t) - W_{\hat{f}_n}(\tau_1) \leqslant 0$$

is equivalent to

$$V_n^{\hat{\mathbb{F}}_n}(t) - V_n^{\hat{\mathbb{F}}_n}(\tau_1) \leqslant \hat{\mathbb{F}}_n(t) - \hat{\mathbb{F}}_n(\tau_1).$$

This is equivalent to

$$V_n^{\hat{\mathbb{F}}_n}(t) \leqslant \hat{\mathbb{F}}_n(t),$$

since $V_n^{\hat{\mathbb{F}}_n}(\tau_1) = \hat{\mathbb{F}}_n(\tau_1)$. Equation (13) follows by an induction argument. Hence we have the following corollary.

Corollary 3.1

$\hat{f}_n$ is the NPMLE of f_0 if and only if $\hat{\mathbb{F}}_n(t) \equiv \int_0^t \hat{f}_n(s)\, ds$ satisfies

$$\hat{\mathbb{F}}_n(T_{(n)}) \leqslant 1, \quad \int \frac{\delta}{1 - \hat{\mathbb{F}}_n(t')} d\mathbb{P}_n(t', \delta) \leqslant 1, \tag{14}$$

and $\hat{f}_n$ is the left derivative of the least concave majorant (LCM) of $\hat{V}_n^{\hat{\mathbb{F}}_n}(t)$.

Notice that $\hat{f}_n$ is also the left derivative of $\hat{\mathbb{F}}_n$, we also have the following characterization of $\hat{f}_n$.

Corollary 3.2

Let $0 < t_0 < \tau_H = \inf\{t\colon H(t) = 1\}$. Then $\bar{G}(t_0)\hat{f}_n(t)$ is the left derivative of the least concave majorant of $V_n^{\hat{\mathbb{F}}_n}(t) - G(t_0)\hat{\mathbb{F}}_n(t)$.

Proof. Since the LCM of $V_n^{\hat{\mathbb{F}}_n}(t) - G(t_0)\hat{\mathbb{F}}_n(t)$ is the same as the LCM of $V_n^{\hat{F}_n}(t)$ minus $G(t_0)\hat{\mathbb{F}}_n(t)$, thus the slope of the LCM of $V_n^{\hat{\mathbb{F}}_n}(t) - G(t_0)\hat{\mathbb{F}}_n(t)$ is equal to the slope of the LCM of $V_n^{\hat{\mathbb{F}}_n}(t)$ minus the slope of $G(t_0)\hat{\mathbb{F}}_n(t)$, which is $\hat{f}_n(t) - G(t_0)\hat{f}_n(t) = \bar{G}(t_0)\hat{f}_n(t)$. □

We will use corollary 3.2 to derive the asymptotic distribution of $\hat{f}_n(t_0)$ and prove the asymptotic equivalence of the NPMLE $\hat{f}_n$ and the estimator $\tilde{f}_n$ obtained directly from the Kaplan–Meier estimator.

Remark 3.1. Theorem 3.1 or corollary 3.1 suggests a natural iterative concave majorant algorithm for computing the NPMLE $\hat{f}_n$. In our simulations described in section 8, the computation of $\hat{f}_n$ is carried out based on corollary 3.1. It goes as follows:

Step 1. Choose a reasonable initial estimator $F_n^{(0)}$ of F_0, for example, the least concave majorant of the Kaplan–Meier estimator.

Step 2. Compute the left derivative $f_n^{(1)}(t)$ of the least concave majorant of $\hat{V}_n^{F_n^{(0)}}(t)$. If for some t, $f_n^{(1)}(t) < 0$, set $f_n^{(1)}(t) = 0$. If $f_n^{(1)}(T_{(i)}) = 0$ at some uncensored observation $T_{(i)}$, make $f_n^1(T_{(i)})$ slightly bigger than zero.

Step 3. Compute $F_n^{(1)}(t)$ from $f_n^{(1)}(t)$ obtained in Step 2. If $F_n^{(1)}(T_{(i)}) > 1$ for some $T_{(i)}$, set $F_n^{((1))}(T_{(j)}) = 1$, for all $i \leqslant j \leqslant n$.

Step 4. Compute some reasonable norm of the difference between $F_n^{(1)}$ and $F_n^{(0)}$. If the difference is no greater than a prescribed convergence criterion, stop. Otherwise go back to Step 1, with $F_n^{(0)}$ replaced by $F_n^{(1)}$. □

3.2. Characterization of the NPMLE $\hat{\lambda}_n$

Using the definition of a hazard function and $\Lambda(x) = -\log(1 - F(x))$ when F is continuous, we may write

$$l(\lambda) = \sum_{i=1}^{n} \left\{\delta_i \log \frac{f(T_i)}{1 - F(T_i)} + \log(1 - F(T_i))\right\} = \sum_{i=1}^{n} \{\delta_i \log(\lambda(T_i)) - \Lambda(T_i)\}.$$

The NPMLE $\hat{\lambda}_n$ of λ is an increasing step function maximizing the log-likelihood function. If λ is constant on the intervals $[T_{(j-1)}, T_{(j)})$, then it follows that

$$\Lambda(T_{(i)}) = \sum_{j=1}^{i} (T_{(j)} - T_{(j-1)})\lambda(T_{(j)}),$$

where $T_{(0)} = 0$. Then $l(\lambda)$ can be written as

$$\sum_{i=1}^{n} \left\{\delta_{(i)} \log(\lambda(T_{(i)})) - \sum_{j=1}^{i} (T_{(j)} - T_{(j-1)})\lambda(T_{(j)})\right\}$$

$$= \sum_{i=1}^{n} \{\delta_{(i)} \log(\lambda(T_{(i)})) - (n-i+1)(T_{(i)} - T_{(i-1)})\lambda(T_{(i)})\}.$$

Hence the NPMLE is given by the vector $\hat{\lambda}_n = (\lambda_1, \ldots, \lambda_n)^{\mathrm{T}} \in \mathbb{R}^n$, maximizing the function

$$l(z) = \sum_{i=1}^{n} \{\delta_{(i)} \log(z_i) - (n-i+1)(T_{(i)} - T_{(i-1)})z_i\} \tag{15}$$

over the cone $\mathscr{C} = \{z = (z_1, \ldots, z_n) : 0 \leqslant z_1 \leqslant \cdots \leqslant z_n\}$. As in ex. 1.5.7 of Robertson *et al.* (1988), the solution to the maximization problem (15) can be written as

$$\hat{\lambda}_i = \max_{1 \leqslant s \leqslant i} \min_{i \leqslant t \leqslant n} \frac{\sum_{j=s}^{t} \delta_{(j)}}{\sum_{j=s}^{t} (n-j+1)(T_{(j)} - T_{(j-1)})}. \tag{16}$$

Notice that unlike the NPMLE of a monotone density f, here we have an explicit solution for the NPMLE of a monotone hazard.

Now define

$$V_n^{\lambda}(t) = \int \{x \leqslant y \wedge t\}\, d\mathbb{Q}_n(x, y) = \mathbb{H}_n^{uc}(t) \tag{17}$$

and

$$W_n(t) = \int [(x \wedge y)\{x \wedge y \leqslant t\} + t\{x \wedge y > t\}]\, d\mathbb{Q}_n(x, y) = \int_0^t (1 - \mathbb{H}_n(s))\, ds, \tag{18}$$

where $\mathbb{H}_n^{uc}$ is the sub-empirical distribution of the uncensored data, $\mathbb{H}_n$ is the empirical distribution of $T_1, \ldots, T_n$, and $\mathbb{Q}_n$ is the empirical distribution of the (unobserved) (X_i, Y_i) pairs, $i = 1, \ldots, n$. Then

$$V_n^{\lambda}(T_{(i)}) = \frac{1}{n} \sum_{j=1}^{i} \delta_{(j)},$$

$$W_n(T_{(i)}) = \frac{1}{n} \sum_{j=1}^{i} (n-j+1)(T_{(j)} - T_{(j-1)}).$$

Using the graphical representation of the isotonic regression, see e.g. th. 1.2.1 of Robertson *et al.* (1988), we have the following characterization of $\hat{\lambda}_n$.

Theorem 3.2

Let V_n and W_n be defined by (17) *and* (18), *respectively. Then $\hat{\lambda}_n$ is the left derivative of the greatest convex minorant of the cumulative sum diagram, consisting of the points*

$$P_j = (W_n(T_{(j)}), V_n(T_{(j)})), \quad j = 0, 1, \ldots, n;$$

where $P_0 = (0, 0)$ and $T_{(j)}$, $j = 1, \ldots, n$ are the order statistics of T_j, $j = 1, \ldots, n$.

Similarly, for decreasing λ, we have the following characterization for its NPMLE $\hat{\lambda}_n$.

Theorem 3.3
Let V_n and W_n be defined by (17) and (18), respectively. Then $\hat{\lambda}_n$ is the left derivative of the least concave majorant of the cumulative sum diagram consisting of the points

$$P_j = (W_n(T_{(j)}), V_n(T_{(j)})), \quad j = 0, 1, \ldots, n,$$

where $P_0 = (0, 0)$ and $T_{(j)}, j = 1, \ldots, n$ are the order statistics of $T_j, j = 1, \ldots, n$.

4. Consistency

We first prove that the estimators $\tilde{f}_n$ and $\tilde{\lambda}_n$ of f and λ based on the Kaplan–Meier or Nelson–Aalen estimators are consistent. The proof is based on a minor modification of Marshall's lemma. Let $\|\cdot\|_a^b$ denote the supremum norm over the interval $[a, b]$. For concave F or convex Λ, we have the following lemma.

Recall from section 1 that $\tilde{\mathbb{F}}_n$ denotes the least concave majorant of the Kaplan–Meier estimator $\mathbb{F}_n$ and $\tilde{\Lambda}_n$ denotes the greatest convex minorant of the Nelson–Aalen estimator Λ_n.

Lemma 4.1
(A) *If F is concave on $[0, \infty)$ with either $F(0) = 0$ or $F(0) > 0$ and $G\{0\} = 0$, then*

$$\|\tilde{\mathbb{F}}_n - F\|_0^{\tau_H} \leqslant \|\mathbb{F}_n - F\|_0^{\tau_H} \xrightarrow{\text{a.s.}} 0.$$

(B) *If Λ is convex on $[0, \tau_H)$, $\tau_H = \tau_G < \tau_F$, and either $\Lambda(0) = 0$ or $\Lambda(0) > 0$ and $G\{0\} = 0$, then*

$$\|\tilde{\Lambda}_n - \Lambda\|_0^{\tau_H} \leqslant \|\Lambda_n - \Lambda\|_0^{\tau_H} \xrightarrow{\text{a.s.}} 0.$$

Proof. (A) The following proof is an adaptation of the proof given by Robertson *et al.* (1988, p. 329), in the uncensored case. It is also related to Marshall's (1970) lemma.

Let $\varepsilon_n \equiv \|\mathbb{F}_n - F\|_0^{\tau_H}$. First note that concavity of F on $[0, \infty)$ implies continuity of F on $(0, \infty)$. Therefore $F\{\tau_H\} = 0$ and either $F(0) = 0$ or $F(0) > 0$ and $G\{0\} = 0$ imply that F and G have no common jumps. It then follows from Stute & Wang (1993, corol. 1.3, p. 1595), that $\varepsilon_n \xrightarrow{\text{a.s.}} 0$. Now the function $F + \varepsilon_n$ is concave and it majorizes the Kaplan–Meier estimator $\mathbb{F}_n$: $\mathbb{F}_n(x) \leqslant F(x) + \varepsilon_n$ for all $0 \leqslant x \leqslant \tau_H$. Hence it follows from the definition of $\tilde{\mathbb{F}}_n$ that

$$\mathbb{F}_n(x) \leqslant \tilde{\mathbb{F}}_n(x) \leqslant F(x) + \varepsilon_n \quad \text{for all } 0 \leqslant x \leqslant \tau_H,$$

and hence that

$$-\varepsilon_n \leqslant \mathbb{F}_n(x) - F(x) \leqslant \tilde{\mathbb{F}}_n(x) - F(x) \leqslant \varepsilon_n \quad \text{for all } 0 \leqslant x \leqslant \tau_H.$$

This yields the desired conclusion:

$$\|\tilde{\mathbb{F}}_n - F\|_0^{\tau_H} \leqslant \varepsilon_n \equiv \|\mathbb{F}_n - F\|_0^{\tau_H} \xrightarrow{\text{a.s.}} 0.$$

(B) First note that since $\tau_H = \tau_G < \tau_F$, $\Lambda(\tau_H) < \infty$, and Λ_n is a uniformly strongly consistent estimator of Λ on $[0, \tau_H]$: $\|\Lambda_n - \Lambda\|_0^{\tau_H} \xrightarrow{\text{a.s.}} 0$. The rest of the proof is as in the proof of (A). □

Lemma 4.2. (Huang & Zhang, 1994)
If $\hat{\mathbb{F}}_n$ is the NPMLE of F under the assumption that f is non-increasing, $\tilde{\mathbb{F}}_n$ is the least concave majorant of the Kaplan–Meier estimator $\mathbb{F}_n$, and $\bar{H}(\tau) > 0$, then

$$\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_0^{\tau} \leqslant R_n \|\tilde{\mathbb{F}}_n - \mathbb{F}_n\|_0^{\tau}, \tag{19}$$

where

$$R_n \equiv \frac{\mathbb{G}_n(\tau)}{\bar{\mathbb{G}}_n(\tau)} \vee 1 \tag{20}$$

and $\mathbb{G}_n$ *is the Kaplan–Meier estimator of* G.

Proof of lemma 4.2. See Huang & Zhang (1994). □

Remark 4.1. As proved by Lo & Phadia (1992), following the arguments of Kiefer & Wolfowitz (1976) in the uncensored case,

$$\|\tilde{\mathbb{F}}_n - \mathbb{F}_n\|_0^\tau = o_p(n^{-1/2}).$$

Hence it also follows from (4.19) that

$$\|\hat{\mathbb{F}}_n - \mathbb{F}_n\|_0^\tau = o_p(n^{-1/2}),$$

and that

$$\|\hat{\mathbb{F}}_n - F\|_0^\tau = O_p(n^{-1/2}); \tag{21}$$

that is, $\hat{\mathbb{F}}_n$ and $\tilde{\mathbb{F}}_n$ are asymptotically equivalent to $\mathbb{F}_n$ as estimators of F on $[0, \tau]$ with $\tau < \tau_H$. In our proof of theorem 5.2 we will only need (21). □

For convex F or concave Λ, the results are similar. So we will not state them here.

Theorem 4.1

Suppose that F *is concave and* $0 < t_0 < \tau_H$. *Then*

(*a*) $f(t_0+) \leqslant \liminf_{n\to\infty} \tilde{f}_n(t_0) \leqslant \limsup_{n\to\infty} \tilde{f}_n(t_0) \leqslant f(t_0-)$.

(*b*) $f(t_0+) \leqslant \liminf_{n\to\infty} \hat{f}_n(t_0) \leqslant \limsup_{n\to\infty} \hat{f}_n(t_0) \leqslant f(t_0-)$.

Proof. The proof of (a) is completely analogous to that in Robertson *et al.* (1988, p. 330). (b) follows from the results of Huang & Zhang (1994). □

Theorem 4.2

Suppose Λ *is convex and* $0 < t_0 < \tau_H$.

(*a*) $\lambda(t_0+) \leqslant \liminf_{n\to\infty} \tilde{\lambda}_n(t_0) \leqslant \limsup_{n\to\infty} \tilde{\lambda}_n(t_0) \leqslant \lambda(t_0-)$.

(*b*) $\lambda(t_0+) \leqslant \liminf_{n\to\infty} \hat{\lambda}_n(t_0) \leqslant \limsup_{n\to\infty} \hat{\lambda}_n(t_0) \leqslant \lambda(t_0-)$.

Proof. (a) can be proved by the same methods used in Robertson *et al.* (1988, p. 330). (b) is proved by Mykytyn & Santner (1981, th. 4.2, p. 1379). □

We close this section with a consistency result which will be needed in the proof of lemma 7.1, a key step in the proof of theorems 2.1 and 2.2. It is also the first step in proving consistency of the NPMLE $\hat{\lambda}_n$ in the case of an increasing hazard rate λ, via its characterization given in theorem 3.2. Let $V(t) = EV_n(t) = \int_0^t \bar{G}(x) f(x)\,dx = H^{uc}(t)$, and $W(t) = EW_n(t) = \int_0^t \bar{H}(x)\,dx$. Then $dV/dW = \lambda$, or $dV(W^{-1}(z))/dz = \lambda(W^{-1}(z))$ where W^{-1} is the inverse of W. By the assumption that λ is non-decreasing, $V \circ W^{-1}(z)$ is convex.

Lemma 4.3

Suppose Λ *is convex on* $[0, \tau_H]$ *and that* $\tau_H < \infty$. *Let* $\{(W_n(t), \hat{V}_n(t)), 0 \leqslant t \leqslant T_{(n)}\}$ *be the GCM of the cumulative sum diagram* $\{(W_n(t), V_n(t)), 0 \leqslant t \leqslant T_{(n)}\}$ *as in theorem* 3.2. *Then*

$$\|\hat{V}_n - V\|_0^{T_{(n)}} \xrightarrow{\text{a.s.}} 0.$$

Proof. First, by the Glivenko–Cantelli theorem and $\mu_H = W(\tau_H) < \infty$,

$$\| V_n - V \|_0^{\tau_H} \xrightarrow{\text{a.s.}} 0 \quad \text{and} \quad \| W_n - W \|_0^{\tau_H} \xrightarrow{\text{a.s.}} 0. \tag{22}$$

Since $W_n(t)$ is strictly increasing in t, the inverse function W_n^{-1} is well defined on $[0, W_n(T_{(n)})]$ and $W_n^{-1} \circ W_n(t) = t$ for $0 \leqslant t \leqslant T_{(n)}$. We extend W_n and W_n^{-1} to all of $[0, \infty)$ by defining $W_n(t) = W_n(T_{(n)}) + (t - T_{(n)})$ for $t \geqslant T_{(n)}$ so that $W_n^{-1}(z) = T_{(n)} + (t - W_n(T_{(n)}))$ for $t \geqslant W(T_{(n)})$. Similarly extend W to all of $[0, \infty)$ by defining $W(t) = W(\tau_H) + (t - \tau_H)$ for $t \geqslant \tau_H$. Note that $\hat{V}_n \circ W_n^{-1}(z)$ is the GCM of $V_n \circ W_n^{-1}(z)$ on $[0, W_n(T_{(n)})]$. Set

$$\begin{aligned} \varepsilon_n &= \| V_n \circ W_n^{-1} - V \circ W^{-1} \|_0^{W_n(T_{(n)})} \\ &\leqslant \| (V_n - V) \circ W_n^{-1} \|_0^{W_n(T_{(n)})} + \| V \circ W_n^{-1} - V \circ W^{-1} \|_0^{W_n(T_{(n)})} \\ &\leqslant \| V_n - V \|_0^{\tau_H} + \| V \circ W_n^{-1} - V \circ W^{-1} \|_0^{W_n(T_{(n)})} \xrightarrow{\text{a.s.}} 0 \end{aligned}$$

by (4.22), uniform continuity of V on $[0, \tau_H]$, and the fact that (22) implies $\| W_n^{-1} - W^{-1} \|_0^{W_n(T_{(n)})} \xrightarrow{\text{a.s.}} 0$.

Now consider the function $V \circ W^{-1} - \varepsilon_n$. Since Λ is convex on $[0, \tau_H]$, $V \circ W^{-1}$ is convex on $[0, \mu_H]$. Now $V \circ W^{-1} - \varepsilon_n$ minorizes $V_n \circ W_n^{-1}$ on $[0, W_n(T_{(n)})]$, and hence by the definition of $\hat{V}_n \circ W_n^{-1}$,

$$V \circ W^{-1}(z) - \varepsilon_n \leqslant \hat{V}_n \circ W_n^{-1}(z) \leqslant V_n \circ W_n^{-1}(z) \quad \text{for} \quad 0 \leqslant z \leqslant W_n(T_{(n)}).$$

Hence

$$\| \hat{V}_n \circ W_n^{-1} - V \circ W^{-1} \|_0^{W_n(T_{(n)})} \leqslant \varepsilon_n \xrightarrow{\text{a.s.}} 0.$$

This implies

$$\begin{aligned} \| \hat{V}_n - V \|_0^{T_{(n)}} &= \| (\hat{V}_n - V) \circ W_n^{-1} \|_0^{W_n(T_{(n)})} \\ &\leqslant \| \hat{V}_n \circ W_n^{-1} - V \circ W^{-1} \|_0^{W_n(T_{(n)})} + \| V \circ W^{-1} - V \circ W_n^{-1} \|_0^{W_n(T_{(n)})} \\ &\leqslant \varepsilon_n + \| V \circ W^{-1} - V \circ W_n^{-1} \|_0^{W_n(T_{(n)})} \\ &\leqslant \| V_n - V \|_0^{\tau_H} + 2 \| V \circ W^{-1} - V \circ W_n^{-1} \|_0^{W_n(T_{(n)})} \xrightarrow{\text{a.s.}} 0. \end{aligned}$$

□

5. Local convergence theorems

In this section, we first introduce the local processes needed to prove theorems 2.1 and 2.2. We then state all the limit theorems for these local processes, including the fundamental asymptotic equivalencies upon which the asymptotic equivalencies in theorems 2.1 and 2.2 are based.

Our proofs, which rely on the strong approximation results of Burke *et al.* (1988), are given at the end of this section.

Recall the definitions of V_n^λ and W_n given in (17) and (18). Now for fixed $0 < t_0 < \tau$ and $\lambda_0 \equiv \lambda(t_0)$, define local processes $\hat{U}_n^\lambda$ and $\tilde{U}_n^\lambda$ by

$$\begin{aligned} \hat{U}_n^\lambda(t) &= \{ n^{2/3}[V_n^\lambda(t_0 + n^{-1/3}t) - V_n^\lambda(t_0)] - n^{2/3}\lambda_0[W_n(t_0 + n^{-1/3}t) - W_n(t_0)] \} / \bar{H}(t_0), \\ \tilde{U}_n^\lambda(t) &= n^{2/3}\{ \Lambda_n(t_0 + n^{-1/3}t) - \Lambda_n(t_0) - n^{-1/3}\lambda(t_0)t \} \end{aligned} \tag{23}$$

for $t \in R$, with the convention that the processes are zero for $t_0 + n^{-1/3}t < 0$. $\hat{U}_n^\lambda$ will be used to study the estimator $\hat{\lambda}_n(t_0)$ of $\lambda(t_0)$, while $\tilde{U}_n^\lambda$ will be used to study $\tilde{\lambda}_n(t_0)$.

Recall the definitions of V_n^F and characterization of F_n given in (13) and corollary 3.2. For fixed $0 < t_0 < \tau$, define local processes $\hat{U}_n^f$ and $\tilde{U}_n^f$ by

$$\hat{U}_n^f(t) = n^{2/3}[V_n^{F_n}(t_0+n^{-1/3}t) - V_n^{F_n}(t_0) - G(t_0)(F_n(t_0+n^{-1/3}t) - F_n(t_0))] - n^{1/3}\bar{G}(t_0)f(t_0)t,$$

$$\tilde{U}_n^f(t) = n^{2/3}[V_n^{\mathbb{F}_n}(t_0+n^{-1/3}t) - V_n^{\mathbb{F}_n}(t_0) - G(t_0)(\mathbb{F}_n(t_0+n^{-1/3}t) - \mathbb{F}_n(t_0))] - n^{1/3}\bar{G}(t_0)f(t_0)t,$$

for $t \in R$, again with the convention that the processes are zero for $t_0 + n^{-1/3}t < 0$. Here $\mathbb{F}_n$ is the Kaplan–Meier estimator of F and F_n is the non-parametric MLE of F under the assumption of a monotone density f.

Let $\mathbf{B}_{\text{loc}}(\mathbb{R}^d)$ denote the space of all locally bounded real functions on $\mathbb{R}^d$ endowed with the topology of uniform convergence on compacts. Here are the limit theorems for the local processes $\hat{U}_n^\lambda$, $\tilde{U}_n^\lambda$ and $\hat{U}_n^f$, $\tilde{U}_n^f$.

Theorem 5.1
Suppose that $\bar{H}(t_0) > 0$, $\lambda'(t_0) \neq 0$, and $\lambda'(t)$ is continuous in a neighborhood of t_0. Then

$$\|\hat{U}_n^\lambda - \tilde{U}_n^\lambda\|_{-k}^{k} \xrightarrow{p} 0 \tag{24}$$

for any $k > 0$, and both $\hat{U}_n^\lambda$ and $\tilde{U}_n^\lambda$ converge weakly in $\mathbf{B}_{\text{loc}}(\mathbb{R})$ to the process $\tilde{U}^\lambda$ defined by

$$\tilde{U}^\lambda(t) \equiv \mathbf{W}\left(\frac{\lambda(t_0)}{\bar{H}(t_0)}t\right) + \tfrac{1}{2}\lambda'(t_0)t^2 \stackrel{d}{=} \sqrt{\frac{\lambda(t_0)}{\bar{H}(t_0)}}\,\mathbf{W}(t) + \tfrac{1}{2}\lambda'(t_0)t^2$$

for $t \in \mathbb{R}$ where $\mathbf{W}$ is a standard two-sided Brownian motion process.

Theorem 5.2
Suppose that $\bar{H}(t_0) > 0$, $f'(t_0) \neq 0$, $f'(t)$ is continuous in a neighborhood of t_0, and $G(t_0 + t) - G(t_0) = o(t^{1/2})$. Then

$$\|\hat{U}_n^f - \tilde{U}_n^f\|_{-k}^{k} \xrightarrow{p} 0 \tag{25}$$

for any $k > 0$, and both $\hat{U}_n^f$ and $\tilde{U}_n^f$ converge weakly in $\mathbf{B}_{\text{loc}}(\mathbb{R})$ to the process $\tilde{U}^f = \hat{U}^f$ defined by

$$\hat{U}^f(t) \equiv \mathbf{W}(\bar{G}(t_0)f(t_0)t) + \tfrac{1}{2}\bar{G}(t_0)f'(t_0)t^2 \stackrel{d}{=} \sqrt{\bar{G}(t_0)f(t_0)}\,\mathbf{W}(t) + \tfrac{1}{2}\bar{G}(t_0)f'(t_0)t^2$$

for $t \in \mathbb{R}$ where $\mathbf{W}$ is a standard two-sided Brownian motion process.

Our proofs of theorems 5.1 and 5.2 will proceed via the following general result.

Suppose that A is a function which is to be estimated with $a(t_0) = A'(t_0) \neq 0$, $a'(t_0) = A''(t_0) \neq 0$. (Think of A as Λ or F, and a as λ or f.) Suppose that $\mathbb{A}_n$ estimates A and that

$$\mathbb{Z}_n \equiv \sqrt{n}(\mathbb{A}_n - A) \Rightarrow \mathbb{Z}$$

in $D[0, \tau]$ where $\mathbb{Z}(t) = \beta(t)\mathbf{B}(\gamma(t))$ and $\mathbf{B}$ is standard Brownian motion. Moreover suppose there is a construction $\tilde{\mathbb{Z}}_n$, $\tilde{\mathbb{Z}}$ of $\mathbb{Z}_n$, $\mathbb{Z}$ on a common probability space so that

$$\|\tilde{\mathbb{Z}}_n - \tilde{\mathbb{Z}}\| \leqslant b_n \tag{26}$$

where $b_n \to 0$ and $\|x\| \equiv \sup\{|x(t)|: 0 \leqslant t \leqslant \tau\}$. Then we have:

Theorem 5.3

Suppose that $0 < t_0 < \tau$ *and* $a_n \to 0$. *Suppose that* (26) *holds with* $b_n^2/a_n \to 0$, *and that* β *and* γ *are differentiable at* t_0. *Then*

$$\frac{1}{\sqrt{a_n}}\{\mathbb{Z}_n(t_0+ta_n)-\mathbb{Z}_n(t_0)\} \Rightarrow \beta(t_0)\mathbf{W}(\gamma'(t_0)t)$$

in $\mathbf{B}_{\mathrm{loc}}(\mathbb{R})$ *where* $\mathbf{W}$ *is standard two-sided Brownian motion. If, in addition* $a(t_0) \neq 0$, $a'(t_0) \neq 0$, *and* $a'(t)$ *is continuous in a neighborhood of* t_0, *then*

$$n^{2/3}\{\mathbb{A}_n(t_0+n^{-1/3}t)-\mathbb{A}_n(t^0)-n^{-1/3}a(t_0)t\} \Rightarrow \beta(t_0)\mathbf{W}(\gamma'(t_0)t)+\tfrac{1}{2}a'(t_0)t^2$$

in $\mathbf{B}_{\mathrm{loc}}(\mathbb{R})$.

Proof. For any fixed $k > 0$,

$$\frac{1}{\sqrt{a_n}}\|\tilde{\mathbb{Z}}_n(t_0+ta_n)-\tilde{\mathbb{Z}}_n(t_0)-(\tilde{\mathbb{Z}}(t_0+ta_n)-\tilde{\mathbb{Z}}(t_0))\|_{-k}^{k} \leqslant \frac{2b_n}{\sqrt{a_n}} \xrightarrow{\text{a.s.}} 0,$$

so it remains only to show that

$$\frac{1}{\sqrt{a_n}}\{\tilde{\mathbb{Z}}(t_0+ta_n)-\tilde{\mathbb{Z}}(t_0)\} \Rightarrow \beta(t_0)\mathbf{B}(\gamma'(t_0)t) \quad \text{in } C[-k,k].$$

But, for $0 \leqslant t \leqslant k$,

$$\begin{aligned}
&\frac{1}{\sqrt{a_n}}\{\tilde{\mathbb{Z}}(t_0+ta_n)-\tilde{\mathbb{Z}}(t_0)\}\\
&=\frac{1}{\sqrt{a_n}}\{\beta(t_0+a_nt)\mathbf{B}(\gamma(t_0+a_nt))-\beta(t_0)\mathbf{B}(\gamma(t_0))\}\\
&\stackrel{d}{=}\sqrt{a_n}\left\{\frac{\beta(t_0+a_nt)-\beta(t_0)}{a_n}\right\}\mathbf{B}(\gamma(t_0+a_nt))+\beta(t_0)\frac{1}{\sqrt{a_n}}\{\mathbf{B}(\gamma(t_0+a_nt))-\mathbf{B}(\gamma(t_0))\}\\
&\stackrel{d}{=}\sqrt{a_n}\left\{\frac{\beta(t_0+a_nt)-\beta(t_0)}{a_n}\right\}O_p(1)+\beta(t_0)\mathbf{B}\left(\frac{\gamma(t_0+a_nt)-\gamma(t_0)}{a_n}\right)\\
&\Rightarrow 0+\beta(t_0)\mathbf{B}(\gamma'(t_0)t) \stackrel{d}{=} \mathbf{B}(\beta^2(t_0)\gamma'(t_0)t).
\end{aligned}$$

Arguing similarly for $-k \leqslant t \leqslant 0$ completes the proof of the first claim.

To prove the second assertion, write

$$\begin{aligned}
&n^{2/3}\{\mathbb{A}_n(t_0+n^{-1/3}t)-\mathbb{A}_n(t_0)-n^{-1/3}a(t_0)t\}\\
&=n^{2/3}\{\mathbb{A}_n(t_0+n^{-1/3}t)-\mathbb{A}_n(t_0)-A(t_0+n^{-1/3}t)+A(t_0)\}\\
&\quad+n^{2/3}\{A(t_0+n^{-1/3}t)-A(t_0)-n^{-1/3}a(t_0)t\}\\
&\Rightarrow \beta(t_0)\mathbf{W}(\gamma'(t_0)t)+\tfrac{1}{2}a'(t_0)t^2
\end{aligned}$$

by the first part of the theorem with $a_n = n^{-1/3}$ and the hypotheses on a and a' at t_0. □

Proof of theorem 5.1. The convergence of $\tilde{U}_n^{\lambda}$ stated in theorem 5.1 is an easy consequence of theorem 5.3 together with the strong approximation results of Burke *et al.* (1988). An alternative proof could be based on the asymptotic linearity results of Lo & Singh (1986) or Lo *et al.* (1989) together with the empirical process methods of Kim & Pollard (1990).

From Burke *et al.* (1988, corol., p. 53), (5.26) holds for $\mathbb{A}_n = \Lambda_n$ and $A = \Lambda$ with $b_n = (\log n)n^{-1/2}$ where the corresponding limit process $\mathbb{Z}$ has $\beta \equiv 1$ and $\gamma = \int_0^\cdot \bar{H}^{-2}\,dH^{uc} = \int_0^\cdot \bar{F}^{-2}\bar{G}^{-1}\,dF \equiv C$. Since $a_n = n^{-1/3}$, $b_n^2/a_n = (\log n)^2 n^{-2/3} = o(1)$, and hence the convergence of $\tilde{U}_n^\lambda$ asserted in theorem 5.1 follows from theorem 5.3 with $b_n = (\log n)n^{-1/2}$, $\gamma'(t_0) = C'(t_0) = \lambda(t_0)/\bar{H}(t_0)$, and $a'(t_0) = \lambda'(t_0)$.

To complete the proof of theorem 5.1, it remains only to prove (24), since the claimed weak convergence of $\hat{U}_n^\lambda$ follows from (24) and the convergence of $\tilde{U}_n^\lambda$ already proved. First we prove another corollary of theorem 5.3, for a natural local process defined in terms of the Kaplan–Meier estimate for $\mathbb{F}_n$; although this process will not be used in the sequel, it could be used to give an independent treatment of the estimator $\tilde{f}_n$ of f. The corollary further illustrates the use of theorem 5.3.

Corollary 5.1

If $\mathbb{F}_n$ is the Kaplan–Meier estimator of F, $f(t_0) \neq 0$, $f'(t_0) < 0$ and f' is continuous in a neighborhood of t_0, then

$$n^{2/3}\{\mathbb{F}_n(t_0 + n^{-1/3}t) - \mathbb{F}_n(t_0) - n^{-1/3}f(t_0)t\} \Rightarrow \bar{F}(t_0)\mathbf{W}(\lambda(t_0)t/\bar{H}(t_0)) + \tfrac{1}{2}f'(t_0)t^2$$
$$\stackrel{d}{=} \mathbf{W}(f(t_0)t/\bar{G}(t_0)) + \tfrac{1}{2}f'(t_0)t^2.$$

Proof of corollary 5.1. From Burke *et al.* (1988, corol., p. 53), (5.26) holds with $b_n = (\log n)n^{-1/2}$. Since $a_n = n^{-1/3}$, $b_n^2/a_n = (\log n)^2 n^{-2/3} = o(1)$, and hence the corollary follows from theorem 5.1 with $\mathbb{A}_n = \mathbb{F}_n$, $A = F$, $[a, b] = [0, \tau]$ for $\tau < \tau_H$, $\beta = \bar{F}$, $\gamma = \int_0^\cdot \bar{H}^{-2}\,dH^{uc} = \int_0^\cdot \bar{F}^{-2}\bar{G}^{-1}\,dF \equiv C$, and $b_n = (\log n)n^{-1/2}$. □

Proof of theorem 5.1, continued. It suffices to prove (24). First, by standard results from empirical process theory, for any $\alpha > 0$ and $0 < \tau < \tau_H$,

$$n^{1/2-\alpha}\|1/\bar{\mathbb{H}}_n - 1/\bar{H}\|_0^\tau \xrightarrow{\text{a.s.}} 0 \quad \text{and} \quad n^{1/2-\alpha}\|\bar{\mathbb{H}}_n^{uc} - \bar{H}^{uc}\|_0^\tau \xrightarrow{\text{a.s.}} 0.$$

These are in particular true for $1/2 - \alpha = 1/3$. We will use these two facts repeatedly in the following. We prove

$$\|\hat{U}_n^\lambda - \tilde{U}_n^\lambda\|_0^k \xrightarrow{p} 0.$$

The other part can be proved in a completely analogous way. Now the Nelson–Aalen estimator can be expressed as $\Lambda_n(t) = \int_0^t (1 - \mathbb{H}_{n-})^{-1}\,d\mathbb{H}_n^{uc}$, where $\mathbb{H}_{n-}(t) = \mathbb{H}_n(t-)$; see Shorack & Wellner (1986, p. 295). We have, letting $t_n = t_0 + n^{-1/3}t$,

$$\bar{H}(t_0)(\hat{U}_n^\lambda(t) - \tilde{U}_n^\lambda(t)) = -n^{2/3}\int_{t_0}^{t_n}\left(\frac{\bar{H}(t_0)}{\bar{\mathbb{H}}_{n-}} - 1\right)d\mathbb{H}_n^{uc} - n^{2/3}\lambda(t_0)\int_{t_0}^{t_n}(\bar{\mathbb{H}}_n(s) - \bar{H}(t_0))\,ds$$
$$= -I_1 - I_2,$$

say. We can write the first term as

$$I_1 = n^{2/3}\int_{t_0}^{t_n}\left(\frac{\bar{H}(t_0)}{\bar{\mathbb{H}}_{n-}} - \frac{\bar{H}(t_0)}{\bar{H}}\right)d(\mathbb{H}_n^{uc} - H^{uc}) + n^{2/3}\int_{t_0}^{t_n}\left(\frac{\bar{H}(t_0)}{\bar{H}} - 1\right)d(\mathbb{H}_n^{uc} - H^{uc})$$
$$+ n^{2/3}\int_{t_0}^{t_n}\left(\frac{\bar{H}(t_0)}{\bar{\mathbb{H}}_{n-}} - \frac{\bar{H}(t_0)}{\bar{H}}\right)dH^{uc} + n^{2/3}\int_{t_0}^{t_n}\left(\frac{\bar{H}(t_0)}{\bar{H}} - 1\right)dH^{uc}$$
$$\equiv II_1 + II_2 + II_3 + II_4.$$

By lem. 2 of Lo & Singh (1986), for any $\tau \geqslant t_0 + n^{-1/3}k$,

$$\|II_1\|_0^\tau \leqslant 2n^{2/3}\bar{H}(t_0)\left\|\int_0^t \left(\frac{1}{\bar{\mathbb{H}}_{n-}} - \frac{1}{\bar{H}}\right) d(\mathbb{H}_n^{uc} - H^{uc})\right\|_0^\tau \xrightarrow{p} 0.$$

For the II_2 term, we use a maximal inequality. Consider the classes $\mathscr{R}_n$ defined as

$$\mathscr{R}_n = \{n^{1/6}(\bar{H}(t_0)/\bar{H}(s) - 1)1_{[t_0, t_0 + n^{-1/3}t]}(s)\delta: 0 \leqslant t \leqslant k\}.$$

For each n, $\mathscr{R}_n$ is a class of increasing functions, so it is a VC-hull class. Hence the uniform entropy numbers of $\mathscr{R}_n$ grow polynomially of the order $(1/\varepsilon)^r$ with $r < 2$ as $\varepsilon \to 0$. See Dudley (1987) or lem. 2.42 of Van der Vaart & Wellner (1994). This allows us to use the maximal inequality with the envelope

$$R_n(s) = n^{1/6}(\bar{H}(t_0)/\bar{H}(t_0 + n^{-1/3}k) - 1)1_{[t_0, t_0 + n^{-1/3}k]}(s).$$

By Kim & Pollard (1990, inequality 2.1, p. 199), there is a constant C such that

$$E\|II_2\|_0^k \leqslant C\sqrt{ER_n^2}.$$

But,

$$\begin{aligned} ER_n^2 &= n^{1/3}(\bar{H}(t_0)/\bar{H}(t_0 + n^{-1/3}k) - 1)^2(\bar{H}(t_0) - \bar{H}(t_0 + n^{-1/3}k)) \\ &= n^{1/3}(\bar{H}(t_0) - \bar{H}(t_0 + n^{-1/3}k))^3/\bar{H}(t_0 + n^{-1/3}k)^2 \\ &= O(n^{-2/3}). \end{aligned}$$

So $E\|II_2\|_0^k = O(n^{-1/3})$. This implies $\|II_2\|_0^k \xrightarrow{p} 0$.

For the II_3 term, we have

$$\|II_3\|_0^k \leqslant n^{2/3}\bar{H}(t_0)\|1/\bar{\mathbb{H}}_n - 1/\bar{H}\|_0^\tau \cdot \|H^{uc}(t_n) - H^{uc}(t_0)\|_0^k \xrightarrow{p} 0.$$

We will see that the II_4 term combined with a term from I_2 converges to zero. The second term

$$\begin{aligned} I_2 &= n^{2/3}\lambda(t_0)\int_{t_0}^{t_n} (\bar{\mathbb{H}}_n(s) - \bar{H}(t_0))\, ds \\ &= n^{2/3}\lambda(t_0)\int_{t_0}^{t_n} (H(s) - \mathbb{H}_n(s))\, ds + n^{2/3}\lambda(t_0)\int_{t_0}^{t_n} (\bar{H}(s) - \bar{H}(t_0))\, ds \\ &\equiv I_{21} + I_{22}, \end{aligned}$$

say. But

$$\left\|n^{2/3}\int_{t_0}^{t_n} (H(s) - \mathbb{H}_n(s))\, ds\right\|_0^k \leqslant n^{1/3}k\|\mathbb{H}_n - H\|_0^\tau \xrightarrow{p} 0,$$

so $\|I_{21}\|_0^k \xrightarrow{p} 0$. Thus

$$\begin{aligned} \|\bar{H}(t_0)(\hat{U}_n^\lambda - \tilde{U}_n^\lambda)\|_0^k &\leqslant \|II_4 + I_{22}\|_0^k + o_p(1) \\ &\leqslant \left\|n^{2/3}\int_{t_0}^{t_n}\left(\frac{\bar{H}(t_0)}{\bar{H}} - 1\right) dH^{uc} + n^{2/3}\lambda(t_0)\int_{t_0}^{t_n} (\bar{H}(s) - \bar{H}(t_0))\, ds\right\|_0^k + o_p(1) \\ &= \left\|n^{2/3}\int_{t_0}^{t_n} (\bar{H}(t_0) - \bar{H}(s))\lambda(s)\, ds + n^{2/3}\lambda(t_0)\int_{t_0}^{t_n} (\bar{H}(s) - \bar{H}(t_0))\, ds\right\|_0^k + o_p(1) \\ &= n^{2/3}\left\|\int_{t_0}^{t_n} (\bar{H}(t_0) - \bar{H}(s))(\lambda(s) - \lambda(t_0))\, ds\right\|_0^k + o_p(1) \xrightarrow{p} 0. \end{aligned}$$

□

To prove theorem 5.2, we will use lemma 4.2.

Proof of theorem 5.2. We first show that the asserted weak convergence of $\hat{U}_n^f$ holds. To do this we use the definition of V_n^F and the corresponding identity for F:

$$V_n^F(t) = \mathbb{H}_n(t) - \int_0^t \frac{\bar{F}(t)}{\bar{F}(s)}\, d\mathbb{H}_n^c(s)$$

and

$$F(t) = H(t) - \int_0^t \frac{\bar{F}(t)}{\bar{F}(s)}\, dH^c(s).$$

Thus it follows that

$$\begin{aligned}V_n^{\hat{\mathbb{F}}_n}(t) - F(t) &= \mathbb{H}_n(t) - H(t) - \int_0^t \frac{1-\hat{\mathbb{F}}_n(t)}{1-\hat{\mathbb{F}}_n(s)}\, d\mathbb{H}_n^c(s) + \int_0^t \frac{\bar{F}(t)}{\bar{F}(s)}\, dH^c(s)\\ &= (\mathbb{H}_n(t) - H(t)) - \int_0^t \frac{1-\hat{\mathbb{F}}_n(t)}{1-\hat{\mathbb{F}}_n(s)}\, d(\mathbb{H}_n^c - H^c)(s) - \int_0^t \left(\frac{1-\hat{\mathbb{F}}_n(t)}{1-\hat{\mathbb{F}}_n(s)} - \frac{\bar{F}(t)}{\bar{F}(s)}\right)\bar{F}(s)\, dG(s).\end{aligned}$$

Integrating by parts for the second term, we have

$$\int_0^t \frac{1-\hat{\mathbb{F}}_n(t)}{1-\hat{\mathbb{F}}_n(s)}\, d(\mathbb{H}_n^c - H^c)(s) = \mathbb{H}_n^c(t) - H^c(t) - \int_0^t (\mathbb{H}_n^c(s-) - H^c(s-))\frac{1-\hat{\mathbb{F}}_n(t)}{(1-\hat{\mathbb{F}}_n(s))^2}\, d\mathbb{F}_n(s).$$

For the third term, we have

$$\begin{aligned}\int_0^t \left(\frac{1-\hat{\mathbb{F}}_n(t)}{1-\hat{\mathbb{F}}_n(s)} - \frac{\bar{F}(t)}{\bar{F}(s)}\right)\bar{F}(s)\, dG(s) &= \int_0^t \frac{1-\hat{\mathbb{F}}_n(t)}{1-\hat{\mathbb{F}}_n(s)}\bar{F}(s)\, dG(s) - \bar{F}(t)G(t)\\ &= \int_0^t \frac{1-\hat{\mathbb{F}}_n(t)}{1-\hat{\mathbb{F}}_n(s)}(\bar{F}(s) - (1-\hat{\mathbb{F}}_n(s)))\, dG(s) + G(t)(1-\hat{\mathbb{F}}_n(t) - \bar{F}(t)).\end{aligned}$$

Thus

$$V_n^{\hat{\mathbb{F}}_n}(t) = F(t) + (\mathbb{H}_n^{uc}(t) - H^{uc}(t)) + G(t)(\hat{\mathbb{F}}_n(t) - F(t)) + \hat{R}_n(t),$$

where $\hat{R}_n(t) = \hat{R}_{1n}(t) + \hat{R}_{2n}(t)$,

$$\hat{R}_{1n}(t) = -\int_0^t (\mathbb{H}_n^c(s-) - H^c(s-))\frac{1-\hat{\mathbb{F}}_n(t)}{(1-\hat{\mathbb{F}}_n(s))^2}\, d\mathbb{F}_n(s),$$

and

$$\hat{R}_{2n}(t) = -\int_0^t \frac{1-\hat{\mathbb{F}}_n(t)}{1-\hat{\mathbb{F}}_n(s)}(\bar{F}(s) - (1-\hat{\mathbb{F}}_n(s)))\, dG(s).$$

It follows that

$$\begin{aligned}\hat{U}_n^f(t) &= n^{2/3}[F(t_0 + n^{-1/3}t) - F(t_0) - n^{-1/3}tf(t_0)]\bar{G}(t_0)\\ &\quad + n^{2/3}[\mathbb{H}_n^{uc}(t_0 + n^{-1/3}t) - \mathbb{H}_n^{uc}(t_0) - H^{uc}(t_0 + n^{-1/3}t) + H^{uc}(t_0)]\\ &\quad + n^{2/3}[\hat{\mathbb{F}}_n(t_0 + n^{-1/3}t) - F(t_0 + n^{-1/3}t)][G(t_0 + n^{-1/3}t) - G(t_0)]\\ &\quad + n^{2/3}[\hat{R}_n(t_0 + n^{-1/3}t) - \hat{R}_n(t_0)]. \end{aligned} \tag{27}$$

Now for any $k > 0$, by lemma 5.1 and our hypothesis on G,

$$\begin{aligned}&n^{2/3}\|[\hat{\mathbb{F}}_n(t_0 + n^{-1/3}t) - F(t_0 + n^{-1/3}t)][G(t_0 + n^{-1/3}t) - G(t_0)]\|_{-k}^{k}\\ &\quad = n^{2/3}O_p(n^{-1/2})o(n^{-1/6}) = o_p(1).\end{aligned} \tag{28}$$

Furthermore, by lemma 5.1

$$n^{2/3}\|\hat{R}_n(t_0+n^{-1/3}t)-\hat{R}_n(t_0)\|_{-k}^{k} \xrightarrow{p} 0. \tag{29}$$

By the hypotheses on f,

$$n^{2/3}\|F(t_0+n^{-1/3}t)-F(t_0)-n^{-1/3}tf(t_0)-\tfrac{1}{2}n^{-2/3}f'(t_0)t^2\|_{-k}^{k}\to 0. \tag{30}$$

Finally, from Komlós *et al.* (1975), (26) holds for $\mathbb{A}_n=\mathbb{H}_n^{uc}$, $A=H^{uc}$ with $b_n=n^{-1/2}\log n$ where the limit process $\mathbb{Z}(t)\stackrel{d}{=}\beta(t)\mathbf{B}(\gamma(t))$ with $\beta(t)=1-H^{uc}(t)$, $\gamma(t)=H^{uc}(t)/(1-H^{uc}(t))$, and $\gamma'(t)=\bar{G}(t)f(t)/(1-H^{uc}(t))^2$. Hence it follows from theorem 5.3 that

$$\begin{aligned} &n^{2/3}(\mathbb{H}_n^{uc}(t_0+n^{-1/3}t)-\mathbb{H}_n^{uc}(t_0)-H^{uc}(t_0+n^{-1/3}t)+H^{uc}(t_0))\\ &\Rightarrow (1-H_{uc}(t_0))\mathbf{W}\left(\frac{\bar{G}(t_0)f(t_0)}{(1-H^{uc}(t_0))^2}\,t\right)\\ &\stackrel{d}{=}\mathbf{W}(\bar{G}(t_0)f(t_0)t). \end{aligned} \tag{31}$$

Combining (28)–(31) with (27) yields the weak convergence of $\hat{U}_n^f$ as claimed.

To prove the asymptotic equivalence of $\hat{U}_n^f$ and $\tilde{U}_n^f$ asserted in (25), we use the definition of V_n^F and the identity for F again, but now with $\mathbb{F}_n$ replacing $\hat{\mathbb{F}}_n$, to write

$$\begin{aligned} \tilde{U}_n^f(t) &= n^{2/3}[F(t_0+n^{-1/3}t)-F(t_0)-n^{-1/3}tf(t_0)]\bar{G}(t_0)\\ &\quad+n^{2/3}[\mathbb{H}_n^{uc}(t_0+n^{-1/3}t)-\mathbb{H}_n^{uc}(t_0)-H^{uc}(t_0+n^{-1/3}t)+H^{uc}(t_0)]\\ &\quad+n^{2/3}[\mathbb{F}_n(t_0+n^{-1/3}t)-F(t_0+n^{-1/3}t)][G(t_0+n^{-1/3}t)-G(t_0)]\\ &\quad+n^{2/3}[\tilde{R}_n(t_0+n^{-1/3}t)-\tilde{R}_n(t_0)] \end{aligned} \tag{32}$$

where $\tilde{R}_n$ is defined analogously to $\hat{R}_n$, but with $\mathbb{F}_n$ replacing $\hat{\mathbb{F}}_n$. Since (29) continues to hold if $\hat{R}_n$ is replaced by $\tilde{R}_n$, and (28) holds if $\hat{\mathbb{F}}_n$ is replaced by $\mathbb{F}_n$, subtraction of (32) from (27) yields the claimed asymptotic equivalence. □

We close this section with an easy lemma which will be needed in the following section.

Lemma 5.1

Let W_n be defined by (18). *Then, for any $k>0$,*

$$\sup_{-k\leqslant t\leqslant k}|n^{1/3}[W_n(t_0+n^{-1/3}t)-W_n(t_0)]-t\bar{H}(t_0)|\xrightarrow{\text{a.s.}}0.$$

Proof. This follows easily from the Glivenko–Cantelli theorem applied to $\mathbb{H}_n$: from (4) it follows that

$$\begin{aligned} &|n^{1/3}[W_n(t_0+n^{-1/3}t)-W_n(t_0)]-t\bar{H}(t_0)|\\ &=\left|n^{1/3}\int_{t_0}^{t_0+n^{-1/3}t}(1-\mathbb{H}_n(s))\,ds-n^{1/3}\int_{t_0}^{t_0+n^{-1/3}t}(1-H(s))\,ds\right.\\ &\quad\left.+n^{1/3}\int_{t_0}^{t_0+n^{-1/3}t}(1-H(s))\,ds-t\bar{H}(t_0)\right|\\ &\leqslant\|\mathbb{H}_n-H\|\cdot|t|+\|H(t_0+n^{-1/3}t)-H(t_0)\|_{-k}^{k}\cdot|t|, \end{aligned}$$

and hence the claim follows from the Glivenko–Cantelli theorem and continuity of H. □

6. Inverse processes

In this section, we consider some random processes which can be regarded as the inverse processes of the estimators for monotone f or λ, and are easier to deal with analytically. In the following we consider estimation of a monotone increasing hazard rate λ and a monotone decreasing density f.

We first define four processes $\hat{Z}_n^\lambda$, $\tilde{Z}_n^\lambda$ and $\hat{Z}_n^f$, $\tilde{Z}_n^f$ on $\mathbb{R}^2$ by

$$\begin{aligned}
\hat{Z}_n^\lambda(t, a) &= \hat{U}_n^\lambda(t) - a \cdot n^{1/3}[W_n(t_0 + n^{-1/3}t) - W_n(t_0)]/\bar{H}(t_0),\\
\tilde{Z}_n^\lambda(t, a) &= \tilde{U}_n^\lambda(t) - at,\\
\hat{Z}_n^f(t, a) &= \hat{U}_n^f(t) - at,\\
\tilde{Z}_n^f(t, a) &= \tilde{U}_n^f(t) - at
\end{aligned} \tag{33}$$

for $(t, a) \in \mathbb{R}^2$. Let the inverse processes $\hat{S}_n^\lambda$, $\tilde{S}_n^\lambda$ and $\hat{S}_n^f$, $\tilde{S}_n^f$ be defined for $a \geqslant 0$ by

$$\begin{aligned}
\hat{S}_n^\lambda(a) &\equiv \sup\{t \in [0, T_{(n)}]\colon V_n^\lambda(t) - at \text{ is minimal}\},\\
\tilde{S}_n^\lambda(a) &\equiv \sup\{t \in [0, T_{(n)}]\colon \Lambda_n(t) - at \text{ is minimal}\},\\
\hat{S}_n^f(a) &\equiv \sup\{t \in [0, T_{(n)}]\colon V_n^{\hat{\mathbb{F}}_n}(t) - G(t_0)\hat{\mathbb{F}}_n(t) - at \text{ is maximal}\},\\
\tilde{S}_n^f(a) &\equiv \sup\{t \in [0, T_{(n)}]\colon V_n^{\mathbb{F}_n}(t) - G(t_0)\mathbb{F}_n(t) - at \text{ is maximal}\}.
\end{aligned} \tag{34}$$

From the characterizations given in theorem 3.2 and corollary 3.2 in section 3, it follows that

$$\begin{aligned}
\{\hat{\lambda}_n(t) > a\} &= \{\hat{S}_n^\lambda(a) < t\},\\
\{\tilde{\lambda}_n(t) > a\} &= \{\tilde{S}_n^\lambda(a) < t\},\\
\{\bar{G}(t_0)\hat{f}_n(t_0) \leqslant a\} &= \{\hat{S}_n^f(a) \leqslant t_0\},\\
\{\bar{G}(t_0)\tilde{f}_n(t_0) \leqslant a\} &= \{\tilde{S}_n^f(a) \leqslant t_0\}.
\end{aligned} \tag{35}$$

Then with $\hat{Z}_n^\lambda$, $\tilde{Z}_n^\lambda$ and $\hat{Z}_n^f$, $\tilde{Z}_n^f$ as defined in (33), it follows by scaling arguments that

$$\begin{aligned}
\{n^{1/3}(\hat{\lambda}_n - \lambda)(t_0) > a\} &= \{n^{1/3}(\hat{S}_n^\lambda(\lambda(t_0) + n^{-1/3}a) - t_0) < 0\}\\
&= \{\sup[t \in R\colon \hat{Z}_n^\lambda(t, a) \text{ is minimal}] < 0\},\\
\{n^{1/3}(\tilde{\lambda}_n - \lambda)(t_0) > a\} &= \{n^{1/3}(\tilde{S}_n^\lambda(\lambda(t_0) + n^{-1/3}a) - t_0) < 0\}\\
&= \{\sup[t \in R\colon \tilde{Z}_n^\lambda(t, a) \text{ is minimal}] < 0\},\\
\{n^{1/3}\bar{G}(t_0)(\hat{f}_n - f)(t_0) \leqslant a\} &= \{n^{1/3}(\hat{S}_n^f(f(t_0) + n^{-1/3}a) - t_0) \leqslant 0\}\\
&= \{\sup[t \in R\colon \hat{Z}_n^f(t, a) \text{ is maximal}] \leqslant 0\},\\
\{n^{1/3}\bar{G}(t_0)(\tilde{f}_n - f)(t_0) \leqslant a\} &= \{n^{1/3}(\tilde{S}_n^f(f(t_0) + n^{-1/3}a) - t_0) \leqslant 0\}\\
&= \{\sup[t \in R\colon \tilde{Z}_n^f(t, a) \text{ is maximal}] \leqslant 0\}.
\end{aligned} \tag{36}$$

Now our proof of theorems 2.1 and 2.2 will use the following slight extension of the argmax continuous mapping theorem of Kim & Pollard (1990). We let $\mathbf{C}_{\max}(\mathbb{R}^d)$ denote the (separable) subset of continuous functions $x(\cdot)$ in $\mathbf{B}_{\text{loc}}(\mathbb{R}^d)$ which satisfy $x(t) \to -\infty$ as $|t| \to \infty$, and x achieves its maximum at a unique point in $\mathbb{R}^d$. Similarly, $\mathbf{C}_{\min}(\mathbb{R}^d)$ is the subset of continuous functions $x(\cdot)$ which satisfy $x(t) \to \infty$ as $|t| \to \infty$, and x achieves its minimum at a unique point in $\mathbb{R}^d$.

Theorem 6.1

Let $(\mathbb{Y}_n, \mathbb{Z}_n)$ be random maps into $\mathbf{B}_{\text{loc}}(\mathbb{R}^d) \times \mathbf{B}_{\text{loc}}(\mathbb{R}^d)$ and (S_n, T_n) be random maps into $\mathbb{R}^d \times \mathbb{R}^d$ such that:

(i) $(\mathbb{Y}_n, \mathbb{Z}_n) \Rightarrow (\mathbb{Y}, \mathbb{Z})$, $P\{(\mathbb{Y}, \mathbb{Z}) \in \mathbf{C}_{\max}(\mathbb{R}^d) \times \mathbf{C}_{\max}(\mathbb{R}^d)\} = 1$;

(ii) $S_n, T_n = O_p(1)$;

(iii) $\mathbb{Y}_n(S_n) \geqslant \sup_t \mathbb{Y}_n(t) - \alpha_n$, *and* $\mathbb{Z}_n(T_n) \geqslant \sup_t \mathbb{Z}_n(t) - \beta_n$, *where* $\alpha_n, \beta_n = o_p(1)$.

Then $(S_n, T_n) \Rightarrow (S, T) = (\text{argmax}\,(\mathbb{Y}), \text{argmax}\,(\mathbb{Z}))$.

The proof of theorem 6.1 will be deferred to the end of this section. Of course there is a completely analogous argmin continuous mapping theorem with $\mathbf{C}_{\max}$ replaced by $\mathbf{C}_{\min}$ and the "near maxima" condition in (iii) replaced by the corresponding "near minima" condition. Here we will only apply theorem 6.1 to prove theorem 2.1. The proof of theorem 2.2 is similar. Details can be found in Huang & Wellner (1993).

Proof of theorem 2.1. Fix $a, b \in \mathbb{R}$. By theorem 5.1 and lemma 5.2 it follows that

$$(\hat{Z}^f_n(t, a), \tilde{Z}^f_n(t, b)) \Rightarrow (\hat{U}^f(t) - at, \hat{U}^f(t) - bt),$$

where $\hat{U}^f$ is as defined in theorem 5.2. This gives the first hypothesis of theorem 6.1 with $(\mathbb{Y}_n, \mathbb{Z}_n) = (\hat{Z}^f_n, \tilde{Z}^f_n)$ as processes in $\mathbf{B}_{\text{loc}}(\mathbb{R})$. The third hypothesis of theorem 6.1 holds by the identities (36) with

$$(S_n, T_n) = (n^{1/3}(\hat{S}^f_n(f(t_0) + n^{-1/3}a) - t_0), n^{1/3}(\tilde{S}^f_n(f(t_0) + n^{-1/3}a) - t_0)).$$

If we can verify the second condition of theorem 6.1 for this (S_n, T_n), then (the argmax form of) theorem 6.1 implies that

$$(n^{1/3}(\hat{S}^f_n(f(t_0) + n^{-1/3}a) - t_0), n^{1/3}(\tilde{S}^f_n(f(t_0) + n^{-1/3}a) - t_0)) \Rightarrow (\hat{S}^f(a), \hat{S}^\lambda(b)),$$

where

$$\begin{aligned}\hat{S}^f(a) &\equiv \text{argmax}\,\{\hat{Z}^f(t, a)\} = \text{argmax}\,\{\mathbf{W}(\bar{G}(t_0)f(t_0)t) + \tfrac{1}{2}\bar{G}(t_0)f'(t_0)t^2 - at\}\\ &\overset{d}{=} \hat{S}^f(0) + a/\bar{G}(t_0)f'(t_0).\end{aligned}$$

This equality in distribution holds because, with $d \equiv \bar{G}(t_0)f'(t_0)$,

$$\begin{aligned}\hat{S}^f(a) - a/d &= \sup\{t - a/d\colon \sqrt{\bar{G}(t_0)f(t_0)}\mathbf{W}(t) + \tfrac{1}{2}dt^2 - at \text{ is maximal}\}\\ &= \sup\{t\colon \sqrt{\bar{G}(t_0)f(t_0)}\mathbf{W}(t + a/d) + \tfrac{1}{2}d(t + a/d)^2 - a(t + a/d) \text{ is maximal}\}\\ &\overset{d}{=} \sup\{t\colon \sqrt{\bar{G}(t_0)f(t_0)}\mathbf{W}(t) + \tfrac{1}{2}dt^2 \text{ is maximal}\}\\ &= \hat{S}^f(0).\end{aligned}$$

Hence it follows from the identities (36) that

$$\begin{aligned}&P(n^{1/3}\bar{G}(t_0)(\hat{f}_n - f)(t_0) \geqslant a, n^{1/3}\bar{G}(t_0)(\tilde{f}_n - f)(t_0) \geqslant b)\\ &\quad = P(n^{1/3}(\hat{S}^f_n(f(t_0) + n^{-1/3}a) - t_0) \leqslant 0, n^{1/3}(\tilde{S}^f_n(f(t_0) + n^{-1/3}b) - t_0) \leqslant 0)\\ &\quad \to P(\hat{S}^f(0) \leqslant a/\bar{G}(t_0)f'(t_0), \hat{S}^f(0) \leqslant b/\bar{G}(t_0)f'(t_0))\\ &\quad = P(-\bar{G}(t_0)f'(t_0)\hat{S}^f(0) \leqslant a, -\bar{G}(t_0)f'(t_0)\hat{S}^f(0) \leqslant b),\end{aligned}$$

which implies that

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

and hence that

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

This implies the asymptotic equivalence stated in theorem 2.1.

To identify the common limit law, the distribution of $-f'(t_0)\hat{S}^f(0)$, set $R \equiv \bar{G}(t_0)/(f(t_0)f'(t_0))$. Then the asymptotic distribution of $2^{-2/3}R^{1/3}n^{1/3}(\hat{f}_n - f)(t_0)$ is that of $|f'(t_0)|2^{-2/3}R^{1/3}\hat{S}^f(0)$. But by the definition of $\hat{S}^f$,

$$\begin{aligned}
|f'(t_0)|2^{-2/3}R^{1/3}\hat{S}^f(0) &= \sup\{|f'(t_0)|2^{-2/3}R^{1/3}t\colon \sqrt{\bar{G}(t_0)f(t_0)}\mathbf{W}(t) + \tfrac{1}{2}\bar{G}(t_0)f'(t_0)t^2 \text{ is maximal}\}\\
&= \sup\{t\colon \sqrt{\bar{G}(t_0)f(t_0)}\mathbf{W}(|f'(t_0)|^{-1}2^{2/3}R^{-1/3}t) - \tfrac{1}{2}\bar{G}(t_0)|f'(t_0)|^{-1}2^{4/3}R^{-2/3}t^2 \text{ is maximal}\}\\
&= \sup\{t\colon \sqrt{\bar{G}(t_0)f(t_0)}|f'(t_0)|^{-1/2}2^{1/3}R^{-1/6}\mathbf{W}(t) - \tfrac{1}{2}\bar{G}(t_0)|f'(t_0)|^{-1}2^{4/3}R^{-2/3}t^2 \text{ is maximal}\}\\
&= \sup\{t\colon \mathbf{W}(t) - t^2 \text{ is maximal}\}\\
&\equiv Z
\end{aligned}$$

where we have used the scaling property of Brownian motion and the equality

$$\begin{aligned}
\sqrt{\bar{G}(t_0)f(t_0)}|f'(t_0)|^{-1/2}2^{1/3}R^{-1/6} &= \tfrac{1}{2}\bar{G}(t_0)|f'(t_0)|^{-1}2^{4/3}R^{-2/3}\\
&= 2^{1/3}\bar{G}(t_0)^{1/3}f(t_0)^{2/3}|f'(t_0)|^{-1/3}.
\end{aligned}$$

This yields the stated asymptotic distributions.

It remains only to verify the second condition of theorem 6.1; these proofs are grouped together in section 7. □

Proof of theorem 6.1. Invoke the representation theorem for the $(\mathbb{Y}_n, \mathbb{Z}_n)$ processes; see e.g. Kim & Pollard (1990, th. 2.2), or Van der Vaart and Wellner (1994, th. 1.51): Write $(\tilde{\mathbb{Y}}_n(t), \tilde{\mathbb{Z}}_n(t))$ for the composition $(\mathbb{Y}_n(\phi_n(\tilde{\omega}), t), \mathbb{Z}_n(\phi_n(\tilde{\omega}), t))$, $\tilde{S}_n$ for $S_n(\phi_n(\tilde{\omega}))$, $\tilde{T}_n$ for $T_n(\phi_n(\tilde{\omega}))$, and so forth.

We need to prove that $P^*h(S_n, T_n)$ converges to $\tilde{P}h(\tilde{S}, \tilde{T})$ for all h that are bounded, uniformly continuous, real functions on $\mathbb{R}^d \times \mathbb{R}^d$. By the perfectness of ϕ_n,

$$\begin{aligned}
|P^*h(S_n, T_n) - \tilde{P}h(\tilde{S}, \tilde{T})| &= |\tilde{P}^*h(\tilde{S}_n, \tilde{T}_n) - \tilde{P}^*h(\tilde{S}, \tilde{T})|\\
&\leqslant \tilde{P}^*|h(\tilde{S}_n, \tilde{T}_n) - h(\tilde{S}, \tilde{T})|.
\end{aligned}$$

So it is enough to prove that

$$\tilde{P}^*\{|\tilde{S}_n - \tilde{S}| + |\tilde{T}_n - \tilde{T}| > \eta\} \to 0 \quad \text{for each } \eta > 0.$$

This means we only need to prove that

$$\tilde{P}^*\{|\tilde{S}_n - \tilde{S}| > \eta\} \to 0 \quad \text{for each } \eta > 0 \quad \text{and} \quad \tilde{P}^*\{|\tilde{T}_n - \tilde{T}| > \eta\} \to 0 \quad \text{for each } \eta > 0.$$

However, each of the above two convergences follow from the proof of th. 2.7 of Kim & Pollard (1990). □

7. Boundedness in probability of the inverse processes

The key ingredient in the proofs of theorems 2.1 and 2.2 via theorem 6.1 is verification of the condition (ii); i.e. the boundedness in probability of the inverse processes. This is the goal of the present section. We summarize the statements in the following lemma:

Lemma 7.1
Suppose that the hypotheses of theorems 5.1 *and* 5.2 *hold. Then, for each* $\varepsilon > 0$ *and* $M_1 > 0$, *there is an* $M_2 > 0$ *such that*

$$P\left\{\max_{|a|\leqslant M_1} n^{1/3}|\tilde{S}_n^f(f_0+n^{-1/3}a)-t_0|>M_2\right\}<\varepsilon, \tag{37}$$

$$P\left\{\max_{|a|\leqslant M_1} n^{1/3}|\hat{S}_n^f(f_0+n^{-1/3}a)-t_0|>M_2\right\}<\varepsilon, \tag{38}$$

$$P\left\{\max_{|a|\leqslant M_1} n^{1/3}|\tilde{S}_n^\lambda(\lambda_0+n^{-1/3}a)-t_0|>M_2\right\}<\varepsilon, \tag{39}$$

$$P\left\{\max_{|a|\leqslant M_1} n^{1/3}|\hat{S}_n^\lambda(\lambda_0+n^{-1/3}a)-t_0|>M_2\right\}<\varepsilon \tag{40}$$

for n sufficiently large.

Remark 7.1. From (37), it is easy to see that the results of this lemma imply

$$n^{1/3}\max_{|t|\leqslant M_2}|\tilde{f}_n(t_0+n^{-1/3}t)-f(t_0)|=O_p(1),$$

and similarly for $\hat{f}_n$, $\tilde{\lambda}_n$, and $\hat{\lambda}_n$.

Proof of lemma 7.1. We first show the "upper half" of (37); i.e. with the absolute value sign in (37) replaced by $+(\ )$; the part with $-(\ )$ can be proved similarly. Since $\tilde{S}_n^f(a)$ is decreasing in a, we have

$$\begin{aligned}P\left\{\max_{|a|\leqslant M_1} n^{1/3}[\tilde{S}_n^f(f_0+n^{-1/3}a)-t_0]>M_2\right\}&=P\{n^{1/3}[\tilde{S}_n^f(f_0-n^{-1/3}M_1)-t_0]>M_2\}\\&\leqslant P\{\tilde{Z}_n^f(t,-M_1)\geqslant 0 \text{ for some } t\geqslant M_2\},\end{aligned}$$

where $\tilde{Z}_n^f$ is defined by (33). By the continuity of f' in a neighborhood of t_0 and $f'(t_0)<0$, we can find $u_0>0$ such that for any t, $|t-t_0|<u_0$, we have $f'(t)<0$ and $f'(t)$ is close to $f'(t_0)$. By (33), for $n^{-1/3}t\leqslant u_0$, we can write

$$\begin{aligned}\tilde{Z}_n^f(t,-M_1)&=n^{2/3}[F(t_0+n^{-1/3}t)-F(t_0)-n^{-1/3}tf(t_0)]\bar{G}(t_0)+M_1t\\&\quad+n^{2/3}[\mathbb{H}_n^{uc}(t_0+n^{-1/3}t)-\mathbb{H}_n^{uc}(t_0)-H^{uc}(t_0+n^{-1/3}t)+H^{uc}(t_0)]\\&\quad+n^{2/3}[\mathbb{F}_n(t_0+n^{-1/3}t)-F(t_0+n^{-1/3}t)][G(t_0+n^{-1/3}t)-G(t_0)]\\&\quad+n^{2/3}[\tilde{R}_n(t_0+n^{-1/3}t)-\tilde{R}_n(t_0)].\end{aligned}\tag{41}$$

By Taylor expansion

$$n^{2/3}[F(t_0+n^{-1/3}t)-F(t_0)-n^{-1/3}f(t_0)t]\bar{G}(t_0)+M_1t=\tfrac{1}{2}f'(t_n)\bar{G}(t_0)t^2+M_1t,$$

where $|t_n-t_0|<n^{-1/3}t\leqslant u_0$. By the choice of u_0 and $f'(t_0)<0$, we can find $K>0$ such that, for any $t>K$,

$$\tfrac{1}{2}f'(t_n)\bar{G}(t_0)t^2+M_1t<\tfrac{1}{4}f'(t_0)\bar{G}(t_0)t^2$$

for n sufficiently large. By Kim & Pollard (1990, lem. 4.1), for $\alpha=-1/8f'(t_0)\bar{G}(t_0)>0$, and $n^{-1/3}t<u_0$, there exist random variables $A_n=O_p(1)$,

$$\begin{aligned}&n^{2/3}|\mathbb{H}_n^{uc}(t_0+n^{-1/3}t)-\mathbb{H}_n^{uc}(t_0)-(H^{uc}(t_0+n^{-1/3}t)-H^{uc}(t_0))|\\&\quad\leqslant n^{2/3}(\alpha n^{-2/3}t^2+n^{-2/3}A_n^2)=\alpha t^2+A_n^2.\end{aligned}$$

Moreover, the third and fourth terms on the right side of (41) are dominated by

$$o_p(t^{1/2}) + O_p(n^{-1/3}) + O_p(tn^{-1/6}),$$

where the o_p and O_p do not depend on t; to see this, use $\sqrt{n}\|\mathbb{F}_n - F\|_0^\tau = O_p(1)$, the Hölder continuity hypothesis on G at t_0, and the definition of $\tilde{R}_n$ given in the proof of theorem 5.2. It follows that, for $t >$ some $M_2 > K$ we have

$$\begin{aligned}\tilde{Z}_n^f(t, -M_1) &< -\tfrac{1}{8}f'(t_0)\bar{G}(t_0)t^2 + A_n^2 + \tfrac{1}{4}f'(t_0)\bar{G}(t_0)t^2 + o_p(1)\\ &= \tfrac{1}{8}f'(t_0)\bar{G}(t_0)t^2 + A_n^2 + o_p(1)\end{aligned}$$

where the o_p term does not depend on t. We choose M_2 such that $M_2 > K$ and

$$P\{A_n^2 \geqslant -\tfrac{1}{8}f'(t_0)\bar{G}(t_0)M_2^2 + o_p(1)\} \leqslant \varepsilon$$

for n sufficiently large. With this choice of M_2, we have

$$\begin{aligned}&P\{\tilde{Z}_n^f(t - M_1) \geqslant 0, \text{ for some } M_2 \leqslant t \leqslant n^{1/3}u_0\}\\ &\quad\leqslant P\{\tfrac{1}{8}f'(t_0)\bar{G}(t_0)t^2 + A_n^2 + o_p(1) \geqslant 0, \text{ for some } M_2 \leqslant t \leqslant n^{1/3}u_0\}\\ &\quad= P\{A_n^2 \geqslant -\tfrac{1}{8}f'(t_0)\bar{G}(t_0)t^2 + o_p(1) \text{ for some } M_2 \leqslant t \leqslant n^{1/3}u_0\}\\ &\quad\leqslant \varepsilon\end{aligned}$$

for n sufficiently large.

For $n^{-1/3}t > u_0$, we will show that for n sufficiently large with probability close to 1,

$$\tilde{Z}_n^f(t, -M_1) \leqslant \tilde{Z}_n^f(n^{1/3}u_0, -M_1),$$

and hence this term is controlled by the argument given above. Now we use the following self-consistency equation for the Kaplan–Meier estimator $\mathbb{F}_n$:

$$\mathbb{F}_n(t) = \mathbb{H}_n(t) - \int_{[0,\,t]} \frac{\bar{\mathbb{F}}_n(t)}{\bar{\mathbb{F}}_n(s-)}\, d\mathbb{H}_n^c(s); \tag{42}$$

see e.g. Shorack & Wellner (1986, (7.0.11), p. 295, and (7.2.11), p. 304). Notice that, by the definition (3.12) of V_n^F and by the identity (42) it follows that

$$\tilde{Z}_n^f(t, -M_1) = \bar{G}(t_0)n^{2/3}(\mathbb{F}_n(t_0 + n^{-1/3}t) - \mathbb{F}_n(t_0) - n^{-1/3}f(t_0)t) + M_1 t.$$

By lemma 4.1 and the definition of the least concave majorant, we have, with probability one,

$$\mathbb{F}_n(t_0 + u_0) = \tilde{\mathbb{F}}_n(t_0 + u_0) + o(1) \quad \text{and} \quad \mathbb{F}_n(t_0 + n^{-1/3}t) \leqslant \tilde{\mathbb{F}}_n(t_0 + n^{-1/3}t).$$

Hence

$$\begin{aligned}\mathbb{F}_n(t_0 + n^{-1/3}t) - \mathbb{F}_n(t_0 + u_0) &\leqslant \tilde{\mathbb{F}}_n(t_0 + n^{-1/3}t) - \tilde{\mathbb{F}}_n(t_0 + u_0) + o(1)\\ &\leqslant \tilde{f}_n(t_0 + u_0)(n^{-1/3}t - u_0) + o(1).\end{aligned}$$

The last inequality above is from the concavity of $\tilde{\mathbb{F}}_n$. This implies

$$\begin{aligned}&\mathbb{F}_n(t_0 + n^{-1/3}t) - \mathbb{F}_n(t_0 + u_0) - f(t_0)(n^{-1/3}t - u_0) + n^{-1/3}M_1(n^{-1/3}t - u_0)\\ &\quad\leqslant \tilde{f}_n(t_0 + u_0)(n^{-1/3}t - u_0) + (n^{-1/3}M_1 - f(t_0))(n^{-1/3}t - u_0) + o(1)\\ &\quad= (\tilde{f}_n(t_0 + u_0) - f(t_0) + n^{-1/3}M_1)(n^{-1/3}t - u_0) + o(1)\\ &\quad= (f(t_0 + u_0) - f(t_0) + n^{-1/3}M_1 + o(1))(n^{-1/3}t - u_0) + o(1)\\ &\quad\leqslant 0\end{aligned}$$

for n sufficiently large, since f is strictly decreasing at t_0. It therefore follows that

$$\begin{aligned} &P\{\tilde{Z}_n^f(t, -M_1) \geqslant 0 \text{ for some } t > n^{1/3}u_0\} \\ &\leqslant P\{\tilde{Z}_n^f(n^{1/3}u_0, -M_1) \geqslant 0\} \\ &\leqslant P\{\tilde{Z}_n^f(t, -M_1) \geqslant 0 \text{ for some } M_2 \leqslant t \leqslant n^{1/3}u_0\} \\ &< \varepsilon \end{aligned}$$

for n sufficiently large. This completes the proof of (37). □

Yet another proof of this part of lemma 7.1 can be based on the strong approximation results for the product limit estimator $\mathbb{F}_n$ from Burke *et al.* (1988).

The proof of (38) is exactly the same as the proof of (37) for $n^{-1/3}t \leqslant u_0$, but now also using (21) of lemma 4.1 in combination with the definition of $\hat{R}_n$ given in the proof of theorem 5.2 to control the last two terms in the "hat" version of (41). However, for $n^{-1/3}t > u_0$ we can no longer use the identity (42). Instead, note that $\tilde{V}_n^{F_n} = F_n$. Then the argument proceeds as above using the concavity of F_n.

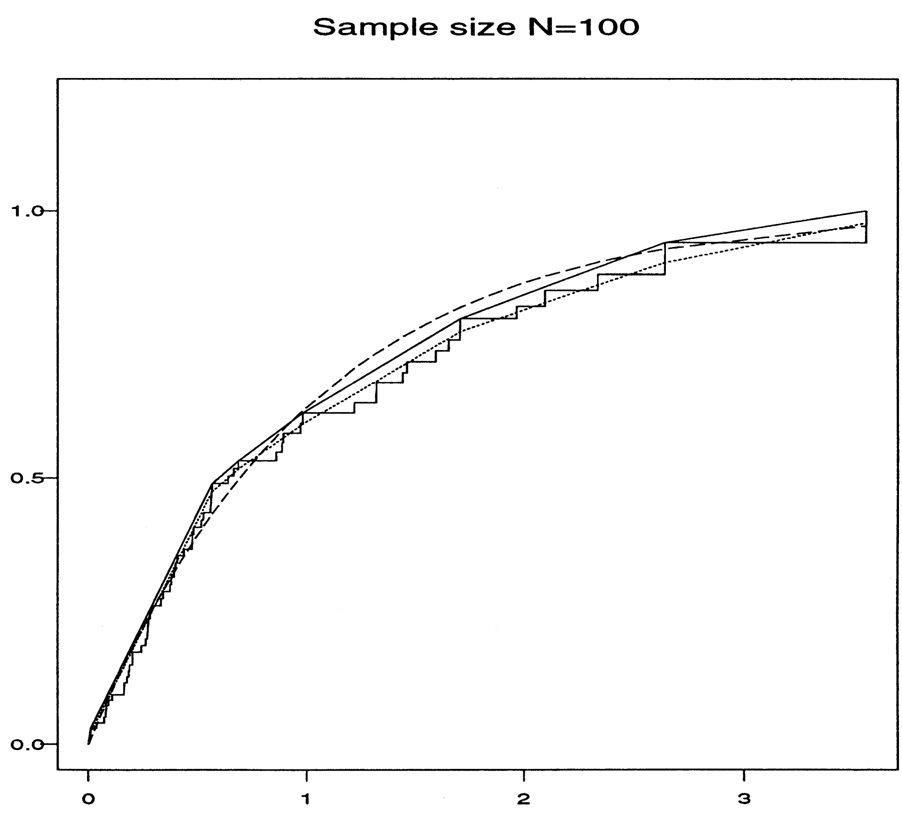

Fig. 1. Estimation of a distribution with decreasing density under censoring. The solid step function is the Kaplan–Meier estimator. The solid line is the concave majorant of the Kaplan–Meier estimator. The dotted line is the NPMLE. The dashed line is the true distribution function $F_0(x) = 1 - \exp(-x)$. The censoring distribution is $G(y) = 1 - \exp(-0.5y)$.

The proof of (39) is exactly the same as the proof of (37), but now we need to use part (B) of lemma 4.1 The proof of (40) is also parallel to the proof of (37). However, we now should use lemma 4.3. The details are given in Huang & Wellner (1993).

8. Illustrations and examples

Now we illustrate the results obtained in sections 2–6 by calculating the estimators $\hat{f}_n$, $\tilde{f}_n$, $\hat{\lambda}_n$, $\tilde{\lambda}_n$ and their corresponding cumulative versions for simulated data.

To compare the NPMLE with the estimators based on the Kaplan–Meier estimator in the estimation of a decreasing density, or the Nelson–Aalen estimator in the estimation of an increasing hazard, and also to see how censoring and sample size effect the estimation, we carried out two sets of simulations.

8.1. Estimation of a decreasing density f

The simulated random samples are generated in S as follows. The failure time $X \sim \exp(1)$, i.e., the exponential distribution with mean one. The censoring distribution is $\exp(1/2)$, i.e., the exponential distribution with mean 2. Then $P(X \leqslant Y) = 2/3$, that is, the uncensored rate is two-thirds. Two different sample sizes are used, $n = 100$ and 800.

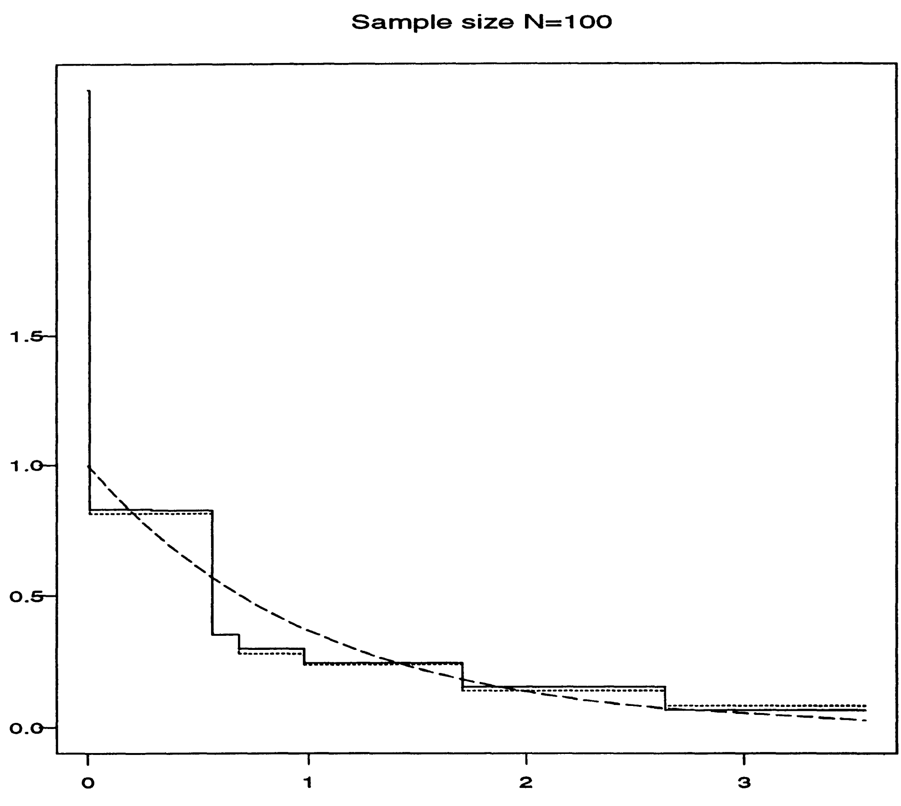

Fig. 2. Estimation of a decreasing density under censoring. The solid line is the slope of the concave majorant of the Kaplan–Meier estimator. The dotted line is the NPMLE. The dashed line is the true density function $f_0(x) = 1 - \exp(-x)$. The censoring distribution is $G(y) = 1 - \exp(-0.5y)$.

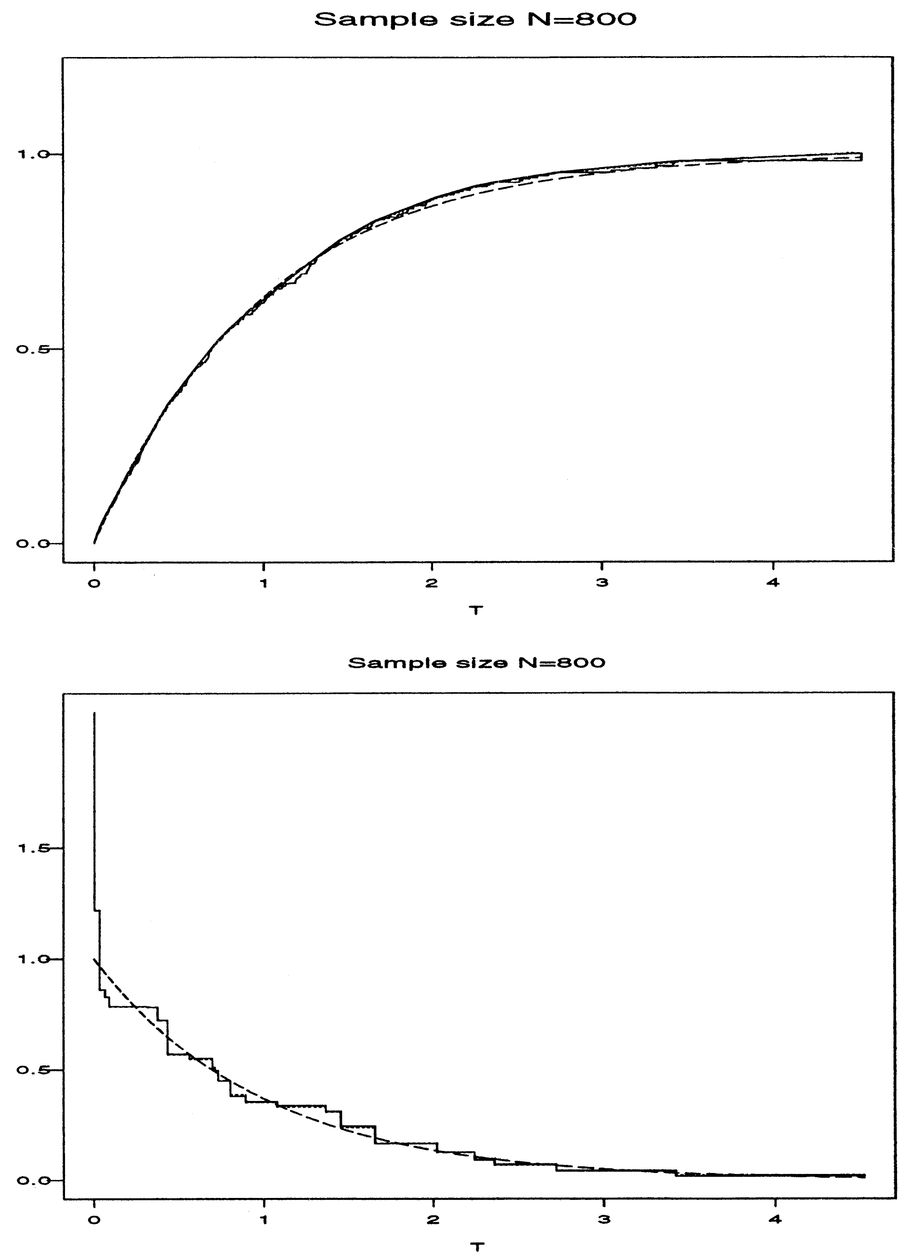

Fig. 3. Estimation of a distribution with decreasing density under censoring. In the top panel, the solid step function is the Kaplan–Meier estimator. The solid line is the concave majorant of the Kaplan–Meier estimator. The dotted line is the NPMLE. The dashed line is the true distribution function $F_0(x) = 1 - \exp(-x)$. The lower panel shows the corresponding slopes of the curves in the top panel with the same line types. The censoring distribution is $G(y) = 1 - \exp(-0.5y)$.

Figure 1 shows the estimators of the distribution function. The step function is the Kaplan–Meier estimator. (In the figures, the step functions are either the Kaplan–Meier estimator of the distribution function, or the Nelson–Aalen estimator of the cumulative hazard function.) The solid line is the least concave majorant of the Kaplan–Meier estimator. The dotted line is the NPMLE of the distribution function with decreasing density. The dashed line is the true distribution function, i.e., $F_0(x) = 1 - \exp(-x)$. The censoring distribution is $G(y) = 1 - \exp(-y/2)$.

Figure 2 shows the estimators of the decreasing density. The solid line is the slope of the concave majorant of the Kaplan–Meier estimator. The dotted line is the NPMLE of the decreasing density. The dashed line is the true density, i.e., $f_0(x) = \exp(-x)$.

We observe that the NPMLE and the estimators based on the Kaplan–Meier estimator are different, although they are very close when sample size is moderately large, e.g., $n = 800$, and with this sample size, $\hat{f}_n$ and $\tilde{f}_n$ almost always have the same jump points. We also notice that both $\hat{f}_n$ and $\tilde{f}_n$ tend to overestimate f at $x = 0$; see Figs. 2 and 3. We believe that both $\hat{f}_n$ and $\tilde{f}_n$ are actually not consistent at zero. For the estimation of a decreasing density without censoring, the NPMLE coincides with the slope of the least concave majorant of the empirical distribution function. Woodroofe & Sun (1993) showed that $\hat{f}_n$ is inconsistent, and in fact that

$$\frac{\hat{f}_n(0)}{f(0+)} \xrightarrow{d} \sup_{1 \leqslant k < \infty} \frac{k}{\Gamma_k},$$

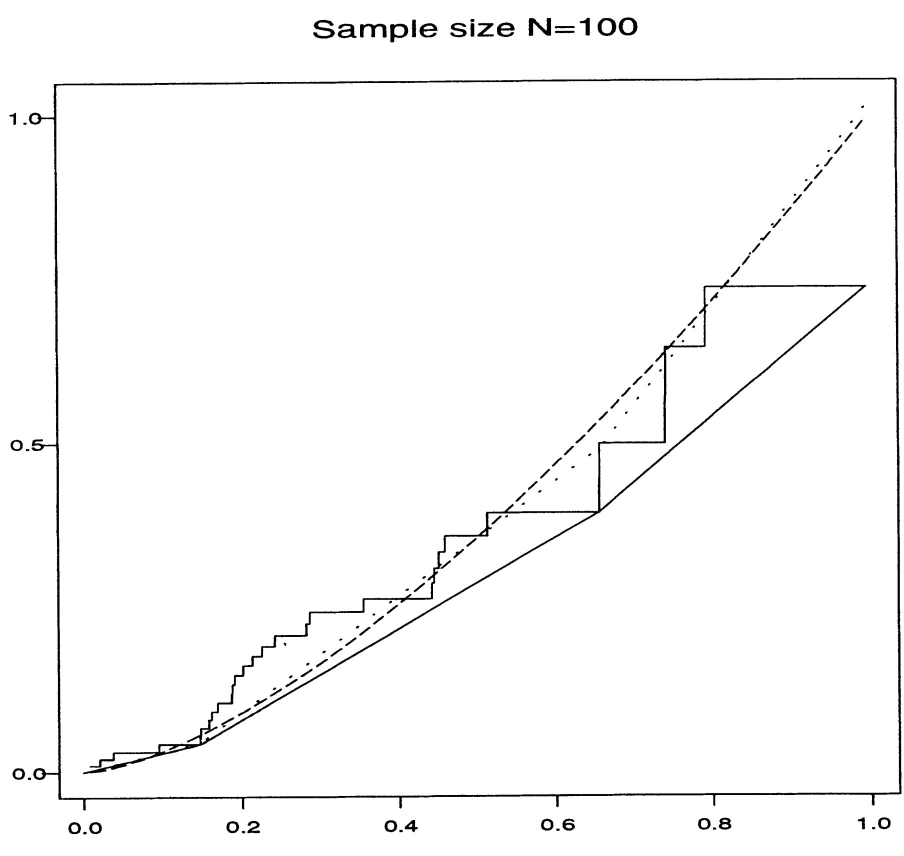

Fig. 4. Estimation of cumulative hazard with increasing hazard rate under censoring. The solid line is the concave majorant of the Nelson–Aalen estimator. The dotted line is the NPMLE. The dashed line is the true cumulative hazard function $\lambda_0(x) = x^{3/2}$. The censoring distribution is uniform(0, 1).

where $\Gamma_1, \Gamma_2, \ldots$ are partial sums of i.i.d. standard exponential random variables. From Pyke (1959), or see Wellner (1977, p. 1008), it follows that $P\{\sup_{1 \leqslant k < \infty} k/\Gamma_k > x\} = 1/x$, for $x \geqslant 1$. Consequently, $P\{\sup_{1 \leqslant k < \infty} k/\Gamma_k > 1\} = 1$.

8.2. Estimation of increasing hazard rate

The simulated random samples are generated in S as follows. The failure times are from a Weibull distribution, i.e., $X \sim 1 - \exp(-x^{3/2})$. The censoring distribution is fixed to be uniform(0, 1). Thus the expected proportion of uncensored observations is $P(X \leqslant Y) = 0.3002\ldots$. Two different sample sizes were used, $n = 100, 800$.

Figure 4 shows the estimators of the cumulative hazard function. The step function is the Nelson–Aalen estimator. The solid line is the greatest convex minorant of the Nelson–Aalen estimator. The dotted line is the NPMLE of the cumulative hazard with increasing hazard rate. The dashed line is the true cumulative hazard curve, i.e., $\Lambda_0(x) = x^{3/2}$. The censoring distribution is uniform(0, 1); $G(y) = y, 0 \leqslant y \leqslant 1$.

Figure 5 shows the estimators of the increasing hazard rate λ_0. The solid line is the slope of the greatest convex minorant of the Nelson–Aalen estimator. The dotted line is the

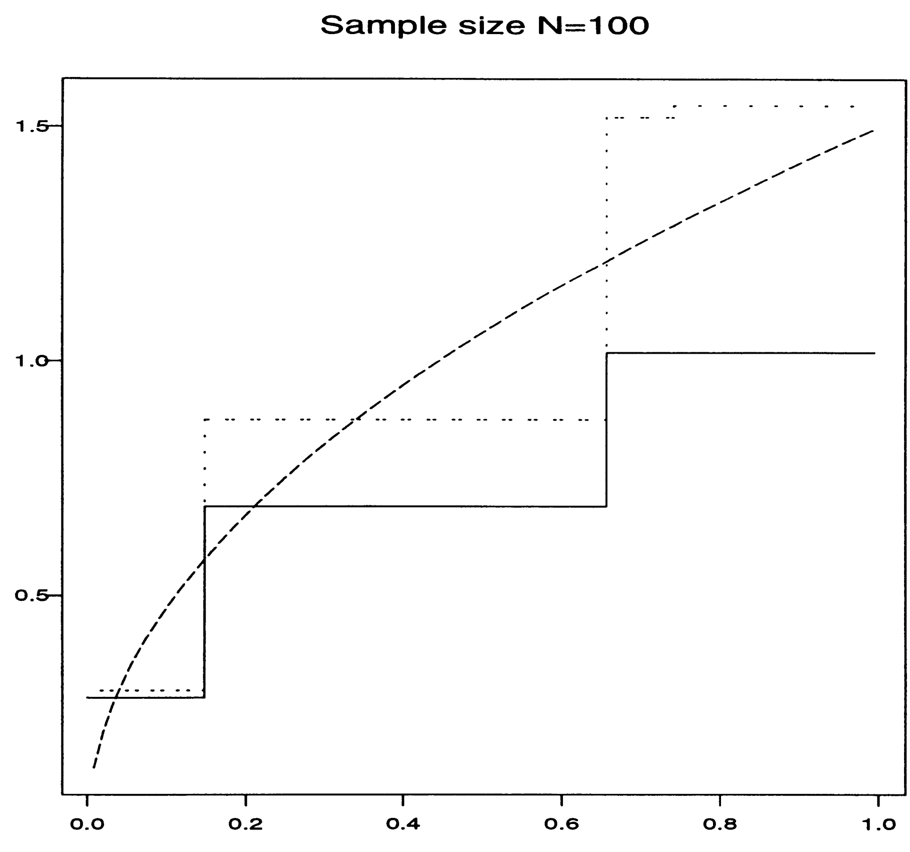

Fig. 5. Estimation of increasing hazard rate under censoring. The solid line is the slope of the concave majorant of the Nelson–Aalen estimator. The dotted line is the NPMLE. The dashed line is the true hazard function $\lambda_0(x) = (3/2)x^{1/2}$. The censoring distribution is uniform(0, 1).

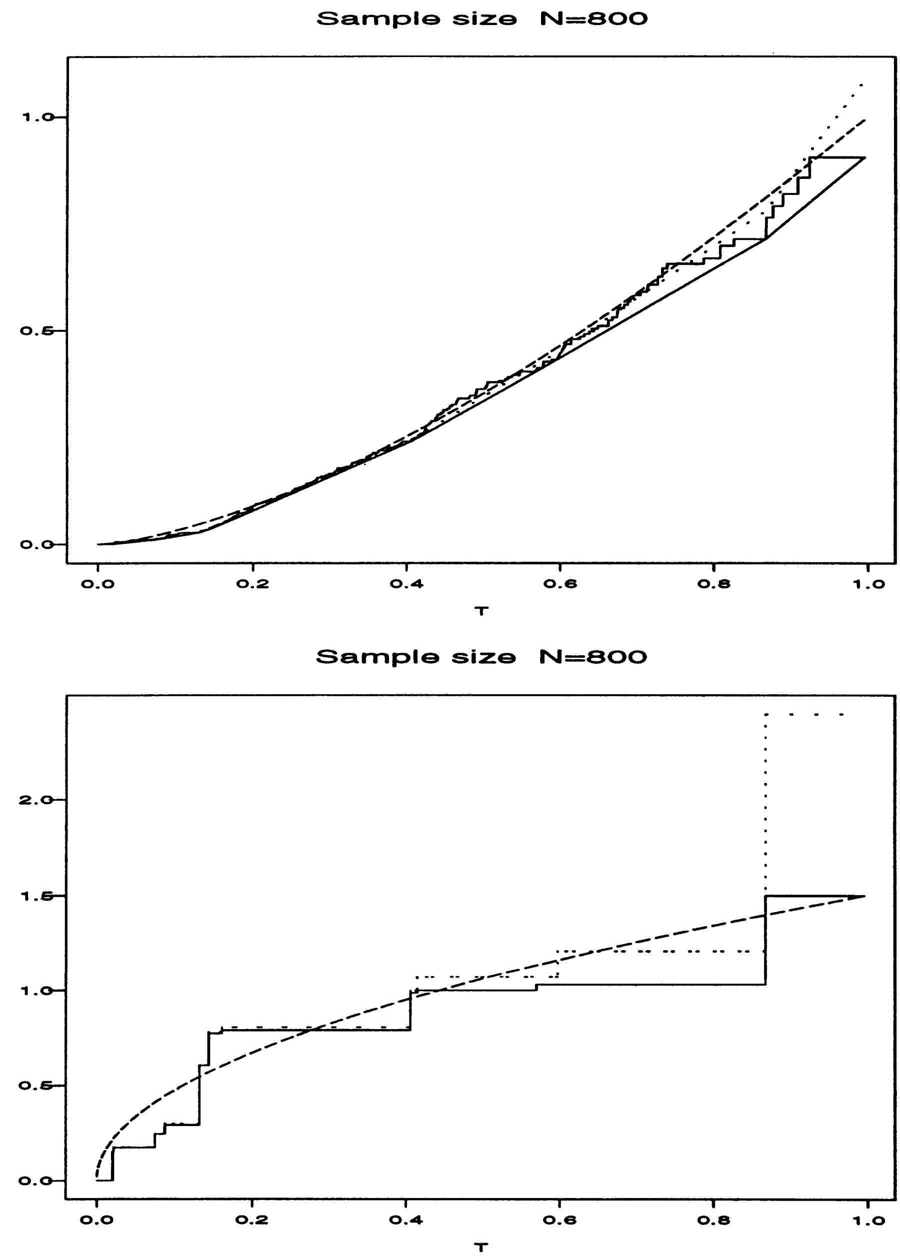

Fig. 6. Estimation of increasing hazard rate under censoring. In the top panel, the solid line is the concave minorant of the Nelson–Aalen estimator. The dotted line is the NPMLE. The dashed line is the true cumulative hazard function $\Lambda_0(x) = x^{3/2}$. The lower panel shows the corresponding slopes of the curves in the top panel with the same line types. The censoring distribution is uniform(0, 1).

NPMLE of the increasing hazard rate. The dashed line is the true hazard rate, i.e., $\lambda_0(x) = (3/2)x^{1/2}$.

Similar to the case of estimating a decreasing density, the NPMLE and the estimators based on the Nelson–Aalen estimator are different, although they are very close when sample size is moderately large, e.g., $n = 800$, and with this moderate large sample size, $\hat{\lambda}_n$ and $\tilde{\lambda}_n$ almost always have the same jump points. We also notice that both $\hat{\lambda}_n$ and $\tilde{\lambda}_n$ behave badly at the right end point. (In our simulations, the right end point is 1.) We believe that both $\hat{\lambda}_n$ and $\tilde{\lambda}_n$ are not consistent at the right extreme point. The problem of the behavior of $\hat{\lambda}_n$ and $\tilde{\lambda}_n$ at the right extreme point has not been worked out yet, although it might be solved along the lines of Woodroofe & Sun (1993) in dealing with $\hat{f}_n(0)$. Notice that when there is no censoring, unlike in the estimation of a decreasing density, the NPMLE of an increasing hazard rate is not the same as the slope of the convex minorant of the Nelson–Aalen estimator.

Acknowledgements

Thanks to a referee for catching an error in an earlier version of lemma 4.1 and for pointing out some additional references; and to Y. Huang and C. H. Zhang for sending us a preprint version of their paper. We also (and especially the second author) owe special thanks to Piet Groeneboom for numerous discussions and suggestions concerning the subject of this paper and related work. Research supported in part by National Science Foundation grant DMS-9108409 and by NIAID grant 2R01 AI291968-04.

References

Andersen, P. K., Borgan, O., Gill, R. D. & Keiding, N. (1992). *Statistical models based on counting processes*. Springer-Verlag, New York.

Burke, M. D., Csörgő, S. & Horvàth, L. (1988). A correction to and improvement of 'Strong approximations of some biometric estimates under random censorship'. *Zeit. Wahrsch. Verw. Geb.* **79,** 51–57.

Cox, D. D. & O'Sullivan, F. (1990). Asymptotic analysis of penalized likelihood and related estimators. *Ann. Statist.* **18,** 1676–1695.

Dudley, R. M. (1987). Universal Donsker classes and metric entropy. *Ann. Probab.* **15,** 1306–1326.

Fletcher, R. (1987). *Practical methods of optimization*. Wiley, Chichester.

Grenander, U. (1956). On the theory of mortality measurement, Part II. *Skand. Actuar.* **39,** 125–153.

Groeneboom, P. (1985). Estimating a monotone density. *Proceedings of the Berkeley conference in honor of Jerzy Neyman and Jack Kiefer*, Vol. II, Lucien M. LeCam and Richard A. Olshen eds. Wadsworth, Monterey; IMS, Hayword.

Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Relat. Fields* **79,** 327–368.

Groeneboom, P. & Wellner, J. A. (1992). *Information bounds and nonparametric likelihood estimation.* Birkhäuser, Basel.

Huang, J. & Wellner, J. A. (1993). Estimation of a monotone density or monotone hazard under random censoring. Technical Report No. 252, Department of Statistics, University of Washington, Seattle.

Huang, Y. & Zhang, C. H. (1994). Estimating a monotone density from censored observations. *Ann. Statist.* **22,** to appear.

Kiefer, J. & Wolfowitz, J. (1976). Asymptotically minimax estimation of concave and convex distribution functions. *Z. Wahrsch. verw. Gebiete* **34,** 73–85.

Kim, J. & Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18,** 191–219.

Komlós, J., Major, P. & Tusnódy, G. (1975). An approximation of partial sums of independent r.v.'s and the sample distribution function. *Z. Wahrsch. verw. Gebiete* **32,** 111–131.

Laslett, G. M. (1982). The survival curve under monotone density constraints with applications to two-dimensional line segment processes. *Biometrika* **69,** 153–160.

Lo, S. H. & Phadia, E. (1992). On estimation of a survival function in reliability theory based on censored data. Preprint, Columbia University.

Lo, S. H., Mack, Y. P. & Wang, J. L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan–Meier estimation. *Probab. Theory Relat. Fields* **80,** 461–473.

Lo, S. H. & Singh, K. (1986). The product-limit estimator and the bootstrap: some asymptotic representations. *Probab. Theory Relat. Fields* **71**, 455–465.

Marshall, A. W. (1970). Discussion of Barlow and Van Zwet's papers. In: *Nonparametric techniques in statistical inference*, M. L. Puri (ed.), pp. 174–176. Cambridge University Press, Cambridge.

McNichols, D. T. & Padgett, W. J. (1982). Maximum likelihood estimation of unimodal and decreasing densities based on arbitrarily right-censored data. *Commun. Statist.—Theor. Meth.* **A11**, 2259–2270.

Mykytyn, S. W. & Santner, T. J. (1981). Maximum likelihood estimation of the survival function based on censored data under hazard rate assumptions. *Commun. Statist.—Theor. Meth.* **A10**, 1369–1387.

Padgett, W. J. & Wei, L. J. (1980). Maximum likelihood estimation of a distribution function with increasing failure rate based on censored observations. *Biometrika* **67**, 470–474.

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankyā Ser. A* **31**, 23–36.

Prakasa Rao, B. L. S. (1970). Estimation for distributions with monotone failure rate. *Ann. Math. Statist.* **41**, 507–519.

Pyke, R. (1959). The supremum and infimum of the Poisson process. *Ann. Math. Statist.* **30**, 568–576.

Robertson, T., Wright, F. T. & Dykstra, R. L. (1988). *Order restricted statistical inference*. Wiley, New York.

Shorack, G. R. & Wellner, J. A. (1986). *Empirical processes with applications to statistics*. Wiley, New York.

Stute, W. & Wang, J.-L. (1993). The strong law under random censorship. *Ann. Statist.* **21**, 1591–1607.

Van der Vaart, A. W. & Wellner, J. A. (1995). *Weak Convergence and Empirical Processes*. To appear. Springer-Verlag, New York.

Wang, J.-L. (1986). Asymptotically minimax estimators for distributions with increasing failure rate. *Ann. Statist.* **14**, 1113–1131.

Wang, Y. (1992). Limit distribution for the concave majorant of empirical distribution. Preprint, University of Missouri.

Wellner, J. A. (1977). Distributions related to linear bounds for the empirical distribution function. *Ann. Statist.* **5**, 1003–1016.

Woodroofe, M. & Sun, J. (1993) A penalized maximum likelihood estimate of $f(0+)$ when f is nonincreasing. *Statist. Sinica* **3**, 501–515.

Received April 1993, in final form May 1994

J. Huang, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa 52242-1419, U.S.A.

A. Wellner, Department of Statistics, GN-22, B313 Padelford Hall, University of Washington, Seattle, Washington 98195, U.S.A.