NONPARAMETRIC ESTIMATION OF A CONVEX BATHTUB-SHAPED HAZARD FUNCTION

HANNA K. JANKOWSKI AND JON A. WELLNER

Abstract. In this paper we study the nonparametric MLE and LSE of a convex hazard function. Our estimators are shown to be consistent and to converge at rate $n^{2/5}$. Moreover we establish the pointwise asymptotic distribution theory of both estimators under the assumption that the true hazard function is positive with positive second derivative at the fixed point. The same problems for a convex hazard function under right censoring and for the Poisson process with a convex rate are also considered briefly.

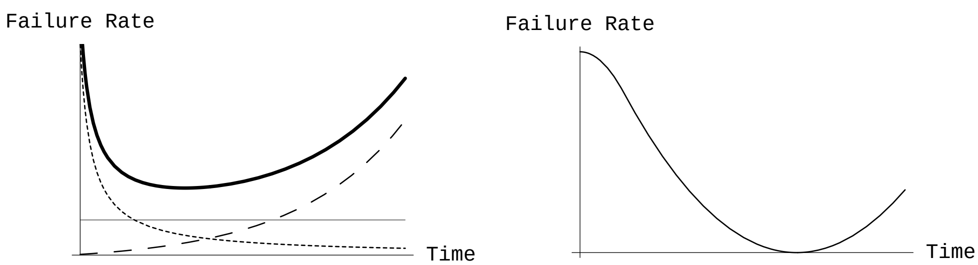
1. Introduction

Information on the behavior of time to a random event is of much interest in many fields. The random event could be failure of a material or machine, death, an earthquake, or infection by a disease, to name but a few examples. Frequently, this type of data is called lifetime data, and it is natural to assume that it takes values in $[0, \infty)$. If the lifetime distribution F, has a density f, then a key quantity of interest is the hazard (or failure) rate $h(t) = f(t)/(1 - F(t))$. Heuristically, $h(t)dt$ is the probability that, given survival until time t, the event will occur in the next dt amount of time. The hazard function is also known as the force of mortality in actuarial science, or the intensity function in extreme value theory.

There are many parametric families which receive attention in lifetime analysis, the exponential (with a constant hazard rate) being arguably the best known and most thoroughly studied. However, in practice it is often not desirable to assume a particular parametric model. On the other hand, certain shape restrictions arise quite naturally in this context. In this work, we are particularly interested in the family of hazard functions which are *bathtub* or U-shaped. That is, there exists a $t_0 \in [0, \infty]$, such that h is decreasing for $0 < t < t_0$ and h is increasing for $t_0 < t < \infty$. Some authors, e.g. Lai, Xie and Murthy (2001), insist that a bathtub shaped hazard be strictly increasing or decreasing for all t, and hence rule out intervals of constant h. Other authors, e.g. Savits (2003) and Marshall and Olkin (2007), pages 120-133, permit intervals of constancy. The literatures of reliability theory and demography contain extensive discussion of the general notion of bathtub shaped hazard functions: see e.g. Singpurwalla (2006), pages 72 - 74, and the review articles of Lai, Xie and Murthy (2001) and Rajarshi and Rajarshi (1988).

Date: October 2, 2008, corrected version.

Heuristically, bathtub shaped hazards correspond to lifetime distributions with high initial hazard (or infant mortality), lower and often rather constant hazard during the middle of life, and then increasing hazard of failure (or wear out) as aging proceeds. The observed failure rate is then a mixture of these three types of failure, as seen in Figure 1(a).



(a) (b)

FIGURE 1. **(a)**. Example when the observed failure (bold) is a mixture of the infant mortality (short-dashed), constant, and wear-out (long-dashed) failure rates. **(b)**. Example of a non-convex failure rate.

Accurate information on the failure rate is of vital importance to practitioners. A reliability engineer can delay the release of a product to the public until it has survived through the infant mortality phase (this practice is known as "burn-in"; see e.g. Lynn and Singpurwalla (1997)). On the other hand, the hazard rate of significant seismic events (within a window of time) can be used to determine when an earthquake alert needs to be issued (cf. Ellis (1985); La Rocca (2008)).

Our focus here is on nonparametric estimation of the subclass of bathtub shaped hazard functions which are also convex. Although there clearly exist many bathtub shaped hazards functions which are not convex (see e.g. Figure 1(b)), many of the frequently proposed parametric families of bathtub hazard functions are also convex. We feel that the additional assumption of convexity of the hazard function is frequently natural and appealing, and it is one way of ensuring that the resulting hazard function is continuous, a condition which is often not satisfied by nonparametric approaches (see the following discussion). We know of no previous attempt to study nonparametric estimation of a convex hazard function with unknown point of minimum or "antimode" as we will call it throughout the rest of this paper.

The nonparametric maximum likelihood estimator (NPMLE) of an increasing hazard function was first derived by Grenander (1956). The estimator may be found explictily via the following "graphical" representation. Define the function $A(j) = \int_0^{X_{(j)}} \mathbb{S}_n(s)ds$ for $j = 1, \ldots, n$, where $X_{(j)}$ are the observed order statistics and $\mathbb{S}_n(x)$ is the empirical survival function. Then the MLE is the reciprocal of the derivative of the least concave majorant of A (see e.g. Robertson, Wright and Dykstra (1988), page 342). In particular, the MLE in this setting is a piecewise constant function (and not smooth).

Consistency of the MLE was established by Marshall and Proschan (1965). Asymptotic distribution theory of the the MLE at a fixed point was treated by Prakasa Rao (1970), where it is shown that the estimator converges at rate $n^{1/3}$. This is analogous to similar results obtained for monotone density estimators, which are well known to converge more slowly than the $\sqrt{n}$-rate typical of regular parametric models. These developments were extended to incorporate censoring by Mykytyn and Santner (1981); see also Huang and Zhang (1994) and Huang and Wellner (1995).

Estimation of the hazard rate with decreasing rate is analogous to that of an increasing rate, as is the estimation of a bathtub failure rate with known antimode. The first consideration of bathtub hazards with no information on the antimode appears in Bray, Crawford and Proschan (1967a,b). These authors derived nonparametric maximum likelihood estimators in the class of all bathtub shaped (or inverted bathtub shaped) hazards, without the additional restriction of convexity, and this was extended to incorporate right censoring in a brief treatment by Mykytyn and Santner (1981). Although the NPMLE here has no explicit formula, it may be found exactly via a simple search algorithm, and is again a piecewise constant function. In both cases, the estimator is shown to be consistent, but further asymptotics are not considered. However, we believe that these will also converge at the $n^{1/3}$ rate typical to MLEs under monotonicity constraints. Banerjee (2007) establishes pointwise confidence intervals for nonparametric estimators of monotone or U-shaped hazard functions. The technique used to do this is based on the likelihood ratio methods introduced in a related problem by Banerjee and Wellner (2001, 2005). Reboul (2005) adapted the methods for unimodal density estimation of Birgé (1997) to the situation of hazard estimation under shape constraints.

There is also an extensive literature on nonparametric Bayes approaches. To our best knowledge, the first consideration of nonparametric Bayes methods begin with Dykstra and Laud (1981) and Padgett and Wei (1981). Essentially, the hazard rate is written as

$$h(t|\mu) = \int k(t,x)d\mu(x), \tag{1.1}$$

where k is a predetermined kernel function, and a prior is chosen for μ. The estimator is then the pointwise posterior mean. A variety of kernels and priors have been considered, and for each possible combination different computational methods are necessary (from MCMC and MC sampling, to exact formulas in certain cases). A nice review of this material appears, for instance, in the recent papers of Ho (2007) and La Rocca (2008).

The choice of kernel in (1.1) clearly imposes shape restrictions on the estimator. For example, $k(t,x) = 1\{t \le x\}$ yields the class of nondecreasing hazard rates. This formulation limits the possibilities to functions with known antimode. To our knowledge, the only consideration of a general bathtub hazard occurs in Ho (2007), where

$$h(t|\mu,\theta) = \int \Big[1\{t-\theta \le u < 0\} + 1\{0 < u \le t-\theta\} \Big]\, d\mu(u),$$

and a prior is placed on both μ *and* θ. However, due to the choice of prior in Ho (2007), the resulting estimates are again piecewise constant.

Let $C \subset \mathbb{R}_+ = [0,\infty)$ be convex. Recall that a function $h : C \mapsto \mathbb{R}$ is convex (on C) if it satisfies

$$h(\lambda x + (1-\lambda)y) \le \lambda h(x) + (1-\lambda)h(y), \quad 0 < \lambda < 1$$

for all $x, y \in C$. Equivalently, a function is convex if its *epigraph*

$$\{(x,\mu) : x \in C, \mu \in \mathbb{R}, \mu \ge f(x)\}$$

is a convex set in $\mathbb{R}^2$ (see e.g. Rockafellar (1970), Section 4). Thus, a convex function on C may be extended to a convex function on $\mathbb{R}_+$ by setting $h(x) = +\infty$ for $x \in \mathbb{R}_+ \cap C^c$.

We consider here two nonparametric estimators of a convex hazard function: the maximum likelihood estimator, and the least squares estimator. The latter is found on $[0,T]$ for a fixed $T < \infty$, by considering the closest function to the empirical hazard function $\mathbb{H}_n$ in the sense that

$$\text{LSE } = \text{argmin}_{h \ge 0 \ \text{ convex}} \left\{ \frac{1}{2} \int_0^T h^2(t)dt - \int_0^T h(t) d\mathbb{H}_n(t) \right\}.$$

To define the MLE, we first consider the likelihood expressed in terms of the hazard rate

$$\mathcal{L}(h) = \prod_{i=1}^{n} h(X_i) \exp\left\{-H(X_i)\right\}.$$

This can be made arbitrarily large by increasing the value of $h(X_{(n)})$. We therefore maximize the modified likelihood

$$\mathcal{L}^{mod}(h) = \prod_{i=1}^{n-1} h(X_i) \exp\left\{-H(X_i)\right\} \times \exp\left\{-H(X_{(n)})\right\}. \qquad (1.2)$$

and set $\widehat{h}_n(X_{(n)})$ to be arbitrarily large (i.e. $\widehat{h}_n(X_{(n)}) = \infty$) to find the MLE. That is, the MLE on $[0, X_{(n)})$ is found by maximizing $\mathcal{L}^{mod}(h)$, and it is set to $+\infty$ for all $x \ge X_{(n)}$. This is the same approach as taken in Grenander (1956) page 142. Equivalently, one could assume that $h \le M$ (i.e. $h(X_{(n)}) = M$ for M sufficiently large), and then let $M \to \infty$ (see e.g. Robertson et al. (1988), page 338).

To illustrate our proposed estimators, consider the distribution with density given by

$$f(t) = \frac{1+2b}{2A\sqrt{b^2 + (1+2b)t/A}}, \quad \text{on } 0 \le t \le A.$$

This distribution was derived in Haupt and Schäbe (1997) as a relatively simple model with bathtub-shaped hazards, which also has an adequate ability to model lifetime behavior. For simplicity, we will call this the H-S distribution. It has bathtub shaped hazard function h for $-1/3 < b < 1$, and has *convex* hazards for all values of b in the parameter space ($b > -1/2$). In Figure 2, we present an example of the LSE and MLE for a simulation from this distribution with a sample size of 100. For the LSE estimator we set T to be 0.9. In general, we find that setting T too large, such as $T = X_{(n)}$, often does not provide good estimates, especially if the data is sparse in the tail.

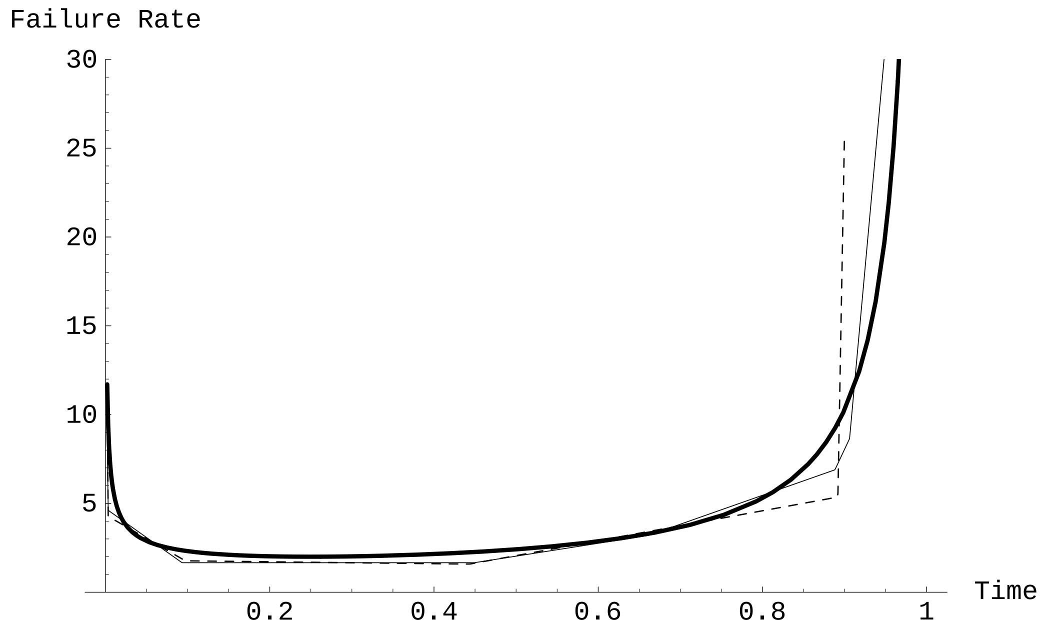


FIGURE 2. Estimation of the H-S hazard with $b = 0, A = 1$ for a sample size of 100. Bold = true hazard, solid = MLE, dashed = LSE ($T = 0.9$).

We also applied our estimators to the earthquake data of the *Appennino Abruzzese* region of Italy (Region 923) recently considered by La Rocca (2008, 2007), who studies Bayesian estimation methods. The data comes from the Gruppo di Lavoro CPTI (2004) catalogue [di Lavoro MPS (2004)]. It consists of 46 inter-quake times, for Region 923, occurring after the year 1650, and with moment magnitude greater than 5.1 (details on the justification of these criteria is available in La Rocca (2008), p.14). Figure 3 shows the resulting estimators.

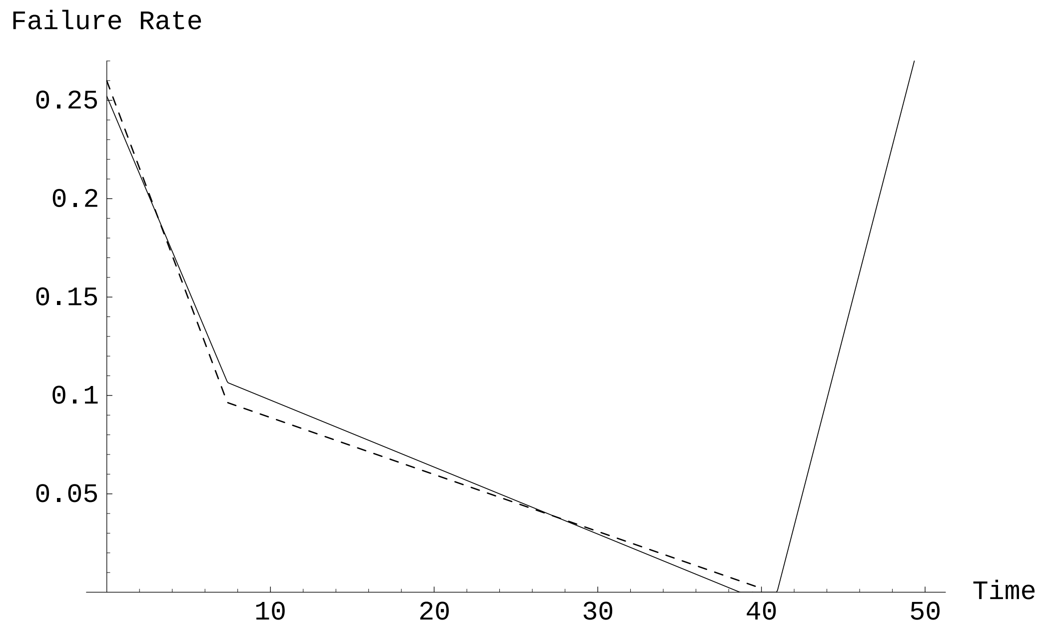


FIGURE 3. Estimation of the Earthquake hazard from CPTI04 data. Solid = MLE, dashed = LSE ($T = 40$).

The main results of this paper are the characterizations and asymptotic behavior of the nonparametric MLE and LSE of a convex hazard function. Unlike most other nonparametric estimators of a bathtub failure rate, our estimators are continuous and piecewise linear. We also show that the estimators are consistent, and establish a local rate of convergence of $n^{2/5}$ (under certain natural assumptions). Although we give a characterization of the MLE and LSE, the final form of the estimators is not explicit. We therefore propose an algorithm (based on the support reduction algorithm of Groeneboom, Jongbloed and Wellner (2007)) to find the estimators. This algorithm is discussed in a separate report, Jankowski and Wellner (2008).

The development of our theory is an extension of the results of Groeneboom, Jongbloed and Wellner (2001b), who study the estimation of a convex *and* decreasing density function. The general treatment is quite similar, however our setting has several additional complications and difficulties (most of these are caused by removing the assumption that the function in question is monotone). Since there have been so few developments in this direction in the intervening time, we feel that it may be helpful to give many of the details in this new case even though the general pattern is similar.

Here is an outline of the paper: We begin with a summary of our main results. Section 3 is dedicated to the proof of characterizations, and existence and uniqueness of the estimators. Consistency of the two estimators is proved in Section 4. Section 5 establishes lower bounds for the pointwise risk of any estimator in our problem. Rates of convergence are established for the MLE and LSE in Section 6, while Section 7 gives our main results concerning the limiting distributions of the estimators at a fixed point. In Section 8, we summarize similar results for the problem of estimating the hazard function with censoring and for estimating a bathtub-shaped (or U-shaped) hazard. Because of the similarity of the problem, we also give a brief treatment here of nonparametric estimation of the intensity function of a Poisson process on $[0, T]$ for some T under the assumption that it is convex. This is done in Section 9. Lastly, in Section 10, we consider estimation of the antimode, and of the local inverses of the hazard function.

2. Summary of Main Results

Suppose we observe $X_1, \ldots, X_n$ i.i.d. from a distribution F_0 with density f_0 and hazard rate $h_0(t)$. We suppose that F_0 is concentrated on $[0, \infty)$; thus $F_0(0) = 0$ and $F_0(\infty) = 1$. We denote the true cumulative hazard function by $H_0(t) = \int_0^t h_0(s)ds$, and the true survival function as $S_0(t)$. Let $\mathbb{F}_n$ denote the empirical distribution function of $X_1, \ldots, X_n$, and let $\mathbb{H}_n$ denote the empirical hazard function: thus

$$\mathbb{F}_n(t) = \frac{1}{n}\sum_{i=1}^n 1_{[0,t]}(X_i), \qquad \mathbb{H}_n(t) = \int_{[0,t]} \frac{1}{1-\mathbb{F}_n(s-)} d\mathbb{F}_n(s).$$

Also, $\mathbb{S}_n(t) = 1 - \mathbb{F}_n(t)$, and we let $0 < X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ denote the order statistics corresponding to $X_1, \ldots, X_n$.

As discussed above, the MLE, $\widehat{h}_n$ on the set $[0, X_{(n)})$ is found by maximizing the modified likelihood (1.2), and setting $\widehat{h}_n(x) = +\infty$ for $x \geq X_{(n)}$. Thus, we need to minimize the criterion function

$$\varphi_n(h) = \int_0^\infty \left\{ H(t) - \log h(t)\mathbb{I}(t \neq X_{(n)}) \right\} d\mathbb{F}_n(t)$$

over the class of nonnegative convex functions on $[0, X_{(n)})$. We will call this space of functions $\mathcal{K}_+$. The MLE, $\widehat{h}_n = \mathrm{argmin}\varphi_n(h)$, is described as follows.

In Proposition 3.2, we show that $\widehat{h}_n$ is piecewise linear. Therefore, it may be expressed as

$$\widehat{h}_n(t) = \widehat{a} + \sum_{j=1}^k \widehat{\nu}_j(\tau_j - t)_+ + \sum_{j=1}^m \widehat{\mu}_j(t - \eta_j)_+, \tag{2.1}$$

where $\widehat{\nu}_j, \widehat{a}, \widehat{\mu}_j \geq 0$. We let τ_j denote the points of change of slope of $\widehat{h}_n$ where $\widehat{h}_n$ is decreasing, and let $\eta_j > 0$ denote the points of change of slope where $\widehat{h}_n$ is increasing. For simplicity we assume that these are ordered. Also, we have $\tau_k \leq \eta_1$. As seen in the next lemma, the τ_j's and η_j's correspond to "points of touch" or equality of processes defined on the one hand in terms of $\widehat{h}_n$ and the data, and on the other hand just in terms of the data. We therefore also refer to them as "touch points" repeatedly in the remainder of the paper.

Lemma 2.1. *Let* $\widetilde{\mathbb{F}}_n(t) = (1/n)\sum_{i=1}^{n-1} \mathbb{I}_{[0,t]}(X_{(i)})$. *A function* $\widehat{h}_n$ *minimizes* φ_n *over* $\mathcal{K}_+$ *(and hence is the MLE) if and only if:*

$$\begin{aligned}
\int_{[0,x]} \frac{x-t}{\widehat{h}_n(t)} d\widetilde{\mathbb{F}}_n(t) &\leq \frac{x^2}{2} - \int_{[0,x]} \frac{(x-t)^2}{2} d\widetilde{\mathbb{F}}_n(t) = \int_0^x \int_0^t \mathbb{S}_n(s)\,ds\,dt, \quad (2.2)\\
&\text{for all } x \geq 0 \quad \text{with} \quad \text{equality at } \tau_i \text{ for } i = 1, \ldots, k;\\
\int_{[x,\infty)} \frac{t-x}{\widehat{h}_n(t)} d\widetilde{\mathbb{F}}_n(t) &\leq \int_{[x,\infty)} \frac{(t-x)^2}{2} d\mathbb{F}_n(t) = \int_x^\infty \int_t^\infty \mathbb{S}_n(s)\,ds\,dt, \quad (2.3)\\
&\text{for all } x \geq 0 \quad \text{with} \quad \text{equality at } \eta_j \text{ for } j = 1, \ldots, m;\\
\int_{[0,\infty)} \frac{1}{\widehat{h}_n(t)} d\widetilde{\mathbb{F}}_n(t) &\leq \int_{[0,\infty)} t\,d\mathbb{F}_n(t) = \int_0^\infty \mathbb{S}_n(t)\,dt, \quad (2.4)\\
\int_{[0,\infty)} \widehat{H}_n(t)\,d\mathbb{F}_n(t) &= 1 - 1/n. \quad (2.5)
\end{aligned}$$

Moreover, the minimizer $\widehat{h}_n$ *satisfies*

$$\int_0^x \widehat{h}_n(t)\mathbb{S}_n(t)\,dt = \mathbb{F}_n(x), \quad (2.6)$$

for $x \in \{\tau_1, \ldots, \tau_k, \eta_1, \ldots, \eta_m\}$.

Remark 2.2. *Note that* (2.4) *implies that if* $\widehat{h}_n$ *is decreasing on* $[0, X_{(n-1)}]$, *then it must be strictly positive there. Also, as we assume* apriori *that* $\widehat{h}_n(X_{(n)}) = \infty$ *we may rewrite the left-hand side terms in* (2.2), (2.3) *and* (2.4) *via*

$$\begin{aligned}
\int_A \frac{x-t}{\widehat{h}_n(t)}\, d\widetilde{\mathbb{F}}_n(t) &= \int_A \frac{x-t}{\widehat{h}_n(t)}\, \mathbb{I}(t \neq X_{(n)})d\mathbb{F}_n(t) = \int_A \frac{x-t}{\widehat{h}_n(t)}\, d\mathbb{F}_n(t)\\
\int_A \frac{1}{\widehat{h}_n(t)}\, d\widetilde{\mathbb{F}}_n(t) &= \int_A \frac{1}{\widehat{h}_n(t)}\, \mathbb{I}(t \neq X_{(n)})d\mathbb{F}_n(t) = \int_A \frac{1}{\widehat{h}_n(t)}\, d\mathbb{F}_n(t),
\end{aligned}$$

for any set A. *We will use this latter formulation from here onwards.*

As in Groeneboom, Jongbloed and Wellner (2001b), a "least squares" type of estimator will also play a role in our development here. One version of such an estimator on a compact sub-interval $[0, T]$ with T fixed is as follows: a least squares estimator of the convex function h_0 is defined to be the convex function $\widetilde{h}_n$ on $[0, T]$ which minimizes

$$\psi_n(h) = \frac{1}{2}\int_0^T h^2(t)\,dt - \int_0^T h(t)\,d\mathbb{H}_n(t)$$

over the class $\mathcal{K}_T$, the class of nonnegative convex functions on $[0, T]$. Heuristically, $\widetilde{h}_n$ is the function which is "closest" to $d\mathbb{H}_n(t)$: for if the latter existed (i.e. $d\mathbb{H}_n(t) = \bar{h}_n(t)dt$), then

$$\widetilde{h}_n = \text{argmin}\left\{\int_0^T (h - \bar{h}_n)^2 dt\right\}.$$

For $\widetilde{h}_n = \text{argmin}\psi_n(h)$, define $\widetilde{H}_n(t) = \int_0^t \widetilde{h}_n(s)ds$, and $\widetilde{\mathcal{H}}_n(t) = \int_0^t \widetilde{H}_n(s)ds$. Also, let $\mathbb{Y}_n(t) = \int_0^t \mathbb{H}_n(s)ds$.

Lemma 2.3 (Characterization of LSE)**.** *The function $\widetilde{h}_n$ minimizes $\psi_n(h)$ over $\mathcal{K}_T$ if and only if it satisfies*

$$\widetilde{H}_n(T) = \mathbb{H}_n(T), \tag{2.7}$$

$$\widetilde{\mathcal{H}}_n(T) = \mathbb{Y}_n(T), \tag{2.8}$$

$$\widetilde{\mathcal{H}}_n(t) \geq \mathbb{Y}_n(t) \text{ for all } t \in [0, T], \tag{2.9}$$

$$\int_0^T (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(t) d\,\widetilde{h}'_n(t) = 0. \tag{2.10}$$

The last statement is the same as: $\widetilde{\mathcal{H}}_n(\tau) = \mathbb{Y}_n(\tau)$ for all changes of slope τ of $\widetilde{h}_n$.

Notice that the function $\widetilde{\mathcal{H}}_n(t) - \mathbb{Y}_n(t)$ is nonnegative for all $t \in [0, T]$, and it is minimized at the points where $\widetilde{h}_n$ has a change of slope. Hence we have:

Corollary 2.4. *$\widetilde{H}_n(\tau) = \mathbb{H}_n(\tau)$ for all changes of slope τ of $\widetilde{h}_n$.*

Remark 2.5 (Notation)**.** *For any function h, we write $H(t) = \int_0^t h(s)ds$, and $\mathcal{H}(t) = \int_0^t H(s)ds$.*

The above lemma shows that the LSE is given by the second derivative of a functional of the process $\mathbb{Y}_n$. This functional (a.k.a. the "invelope", a term coined in Groeneboom et al. (2001a)) is a cubic spline, which sits above the process $\mathbb{Y}_n$. This is in direct analogy to the Grenander estimator of a non-increasing density, which is the first derivative of the least concave majorant of $\mathbb{F}_n$. In this latter case, the MLE and LSE may be shown to be identical, a coincidence which does not hold for either convex densities (as in Groeneboom et al. (2001a)) or here.

In Section 3, we prove the above characterizations. We also show that the estimators are piecewise linear, and, most importantly, that they exist and are unique. In Section 4, we show that the estimators are consistent: see Theorems 4.1 and 4.4 for the exact statements of consistency. Asymptotic minimax risk lower bounds are obtained in Section 5.

The main results of our paper concern the asymptotic behavior of these estimators at a fixed point x_0. To describe these results we introduce several processes from Groeneboom et al. (2001a).

Definition 2.6. *Let $W(s)$ denote a standard two-sided Brownian motion, with $W(0) = 0$, and define $Y(t) = \int_0^t W(s)ds + t^4$. The function $\{\mathcal{I}(t) : t \in \mathbb{R}\}$, the invelope of the process $\{Y(t) : t \in \mathbb{R}\}$, is defined as follows:*

- *The function $\mathcal{I}$ is above the function Y: $\mathcal{I}(t) \geq Y(t)$ for all $t \in \mathbb{R}$.* (2.11)
- *The function $\mathcal{I}$ has a convex second derivative.* (2.12)
- *The function $\mathcal{I}$ satisfies $\int_{\mathbb{R}}\{\mathcal{I}(t) - Y(t)\}d\mathcal{I}^{(3)}(t) = 0$.* (2.13)

It was shown in Groeneboom et al. (2001a) that the process $\mathcal{I}$ exists and is almost surely uniquely defined. Moreover, with probability one, $\mathcal{I}$ is three times differentiable at $t = 0$.

The asymptotic behavior of all of our estimators may be described in terms of the derivatives of the invelope $\mathcal{I}$ at zero. Notably, both the MLE and LSE have the *same* asymptotic behavior.

Theorem 2.7. *Suppose that h_0 is convex and that $x_0 > 0$ is a point which satisfies $0 < h_0(x_0) < \infty$, $h_0''(x_0) > 0$, and that $h_0''(\cdot)$ is continuous in a neighborhood of x_0 (also, $x_0 < T$ for the LSE). Then the nonparametric maximum likelihood estimator and least squares estimator are asymptotically equivalent in the following sense: if $\bar{h}_n = \widehat{h}_n$ or $\widetilde{h}_n$, then*

$$\begin{pmatrix} n^{2/5}(\bar{h}_n(x_0) - h_0(x_0)) \\ n^{1/5}(\bar{h}_n'(x_0) - h_0'(x_0)) \end{pmatrix} \to_d \begin{pmatrix} c_1\ \mathcal{I}^{(2)}(0) \\ c_2\ \mathcal{I}^{(3)}(0) \end{pmatrix}$$

where $\mathcal{I}^{(2)}(0)$ and $\mathcal{I}^{(3)}(0)$ are the second and third derivatives at 0 of the invelope of $Y(t) \equiv \int_0^t W(s)ds + t^4$, and where

$$c_1 = \left(\frac{h_0^2(x_0)h_0''(x_0)}{24S_0^2(x_0)}\right)^{1/5}, \qquad c_2 = \left(\frac{h_0(x_0)h_0''(x_0)^3}{24^3 S_0(x_0)}\right)^{1/5}.$$

This result is proved in Section 7, using the rate of convergence results of Section 6.

A natural next question is to consider the asymptotic behavior at a point if the true hazard function violates the assumptions described above. We conjecture that if $h_0(x_0) = 0$ while h_0'' is continuous and strictly positive then a faster rate of convergence than $n^{2/5}$ will be achieved. On the other hand, if $h_0''(x) = 0$ at x_0, then we believe that the convergence rate $n^{1/2}$ will hold, although these are still open problems. Indeed, at this time, we have no intuition as to the convergence rate if h_0' is discontinuous at x_0, although unpublished work of Cai and Low (2007) suggests that the rate is $n^{1/3}$ at such points.

3. Characterizations, Uniqueness and Existence of the Estimators

Because the MLE is somewhat more difficult than the LSE in terms of its characterization, we first prove existence and uniqueness of the LSE and provide a proof of the characterization given in Lemma 2.3. We then do the same for the MLE.

3.1. **LSE.**

Proposition 3.1 (Existence and Uniqueness of LSE). *Suppose that $\mathbb{H}_n(T) < \infty$. Then the minimizer of ψ_n over the set $\mathcal{K}_T$ exists and is unique.*

Moreover, the minimizer $\widetilde{h}_n$ is piecewise linear. It has at most one change of slope between jumps of $\mathbb{H}_n$, except perhaps in one such interval, where, if the estimator touches zero, it may have two changes of slope (and it is zero between these two changes). Also, between zero and the first jump of $\mathbb{H}_n$, the minimizer may have at most one change of slope, but this happens only if it touches zero, and in this case the estimator is increasing and equal to zero before the first slope change. The same is true for the interval from the last change of slope of $\mathbb{H}_n$ to T, except here if the estimator changes slope, it is decreasing, and it is zero after the change.

Proof. The linearity of the minimizer is straightforward. Let $\{X_{(i)}\}_{i=1}^N$ denote the ordered points of jump of $\mathbb{H}_n$ such that $X_{(N)} \leq T$; here N is defined so that $X_{(N)} \leq T < X_{(N+1)}$. Fix any $h \in \mathcal{K}$, and consider any $g \in \mathcal{K}$ smaller than h, such that $g(X_{(i)}) = h(X_{(i)})$ for all $i = 1, \ldots, N$. For such an g we have that $\psi_n(h) \geq \psi_n(g)$ if and only if $\int h^2 dt \geq \int g^2 dt$. Thus, we will select g so that this term is minimized. Clearly this will be accomplished if we select g to be the smallest possible g, with $h \geq g \geq 0$. With this in mind, it is not difficult to see that such a g (and hence the minimizer of ψ_n) must satisfy the properties described in the second paragraph of the proposition. (For example, suppose that h is differentiable. Then, we choose g to satisfy $g'(X_i) = h'(X_i)$ for $i = 1, \ldots, N$.)

We will next show that the minimizer of $\psi_n(h)$ over $\mathcal{K}$ must lie in the compact set

$$\{h : h \in \mathcal{K}, 0 \leq h \leq B\}$$

for some constant B. Since the function ψ_n is strictly convex on $\mathcal{K}$ it will follow that there exists a unique minimizer.

Next, note that for any two functions h and g, we obtain, using integration by parts,

$$\begin{aligned}
\psi_n(h) - \psi_n(g) &= \frac{1}{2}\int_0^T (h(t) - g(t))^2 dt \\
&\quad + [h-g](T)(G - \mathbb{H}_n)(T) - [h-g]'(T)(\mathcal{G} - \mathbb{Y}_n)(T) \\
&\quad + \int_0^T (\mathcal{G} - \mathbb{Y}_n)(s) d[h-g]'(s). \qquad (3.1)
\end{aligned}$$

From this we see that if g satisfies

(1) $(G - \mathbb{H}_n)(T) = 0$,
(2) $(\mathcal{G} - \mathbb{Y}_n)(T) = 0$,
(3) $(\mathcal{G} - \mathbb{Y}_n)(t) \geq 0$ for all t, and
(4) $\int_0^T (\mathcal{G} - \mathbb{Y}_n)(s) d\, g'(s) = 0$,

then we must necessarily have that $\psi_n(h) - \psi_n(g) \geq 0$, for any $h \in \mathcal{K}$. Thus we may restrict our search for the minimizer of ψ_n over $\mathcal{K}$ to the functions which satisfy the above conditions. In particular, this implies that if a function h is a candidate to be the minimizer it must be that $H(T) = \mathbb{H}_n(T)$.

Next, consider an h such that $H(T) = \mathbb{H}_n(T)$. Note that we may write $h = h_+ + h_-$, where h_+ is increasing and h_- is decreasing. It follows that for all x

$$\begin{array}{rl} \mathbb{H}_n(T) \geq & \int_0^x h_-(t)dt \geq h_-(x)x \\ \mathbb{H}_n(T) \geq & \int_x^T h_+(t)dt \geq h_+(x)(T-x). \end{array} \tag{3.2}$$

Hence,

$$h(x) \leq \mathbb{H}_n(T) \left\{ \frac{1}{x} + \frac{1}{T-x} \right\}. \tag{3.3}$$

Because of convexity of h it follows that h must be bounded on the set $(X_{(1)}, X_{(N)})$, and it remains to argue the same for $[0, X_{(1)}]$ and $[X_{(N)}, T]$.

Consider the set $[0, X_{(1)}]$. Clearly if $h'(X_{(1)}) \geq 0$ then we are done. Assume then that $h'(X_{(1)}) < 0$. For $0 \leq t \leq X_{(1)}$, let $g(t) = h(X_{(1)}) + h'(X_{(1)})(t - X_{(1)})$. To simplify notation we denote $h_1 = h(X_{(1)})$ and $h_2 = h'(X_{(1)})$, and $X_{(1)} = x$. It follows from (3.3) that

$$\psi_n(h) \geq \frac{1}{2} \int_0^x (h_1 + h_2(t-x))^2 dt - h_1 - D,$$

where $D = \sum_{k=2}^n 2\mathbb{H}_n(T)/\min\{X_{(k)}, T - X_{(k)}\}$. The right side of the display is equal to

$$\frac{1}{2} h_1^2 x - h_1 h_2 \frac{x^2}{2} + h_2^2 \frac{x^3}{6} - h_1 - D.$$

By using the inequality $at^2 - bt \geq -b^2/(4a)$ for $a > 0$, $b \in \mathbb{R}$, with the choices $t = h_2$, $a = x^3/6$, $b = h_1 x^2/2$, it follows that

$$\psi_n(h) \geq \frac{1}{8} h_1^2 - h_1 - D.$$

Similarly by the same inequality with $t = h_1$, $b = x/2$, and $b = 1 + x^2 h_2/2$, we find that

$$\psi_n(h) \geq \frac{x^3}{24} h_2^2 - \frac{x}{2} h_2 - D - \frac{1}{2x}.$$

These two inequalities imply that both $|h_1|$ and $|h_2|$ must be finite, if h is to minimize $\psi_n(h)$. This in turn implies that such an h must be finite on $[0, X_{(1)}]$. The argument may be repeated for $[X_{(N)}, T]$. Hence we have shown that the minimizer of $\psi_n(h)$ over $\mathcal{K}$ must lie in the set $\{h : h \in \mathcal{K}, 0 \leq h \leq B\}$ for some constant B, as required.

□

Proof of Lemma 2.3. We first show that conditions (2.7)-(2.10) are sufficient. For $\widetilde{h}_n$ to be the minimizer we must show that

$$\psi_n(h) - \psi_n(\widetilde{h}_n) \geq 0,$$

for any h. By (3.1), $\psi_n(h) - \psi_n(\widetilde{h}_n)$ is bounded below by

$$(h - \widetilde{h}_n)(T)(\widetilde{H}_n - \mathbb{H}_n)(T) - (h - \widetilde{h}_n)'(T)(\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(T) + \int_0^T (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(s) d(h - \widetilde{h}_n)'(s).$$

From conditions (2.7), (2.8) and (2.10) this is equal to

$$\int_0^T (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(s) dh'(s).$$

However, since h is convex, it follows from condition (2.9) that this is non-negative, as required. We next show that the conditions are necessary.

Assume then that $\widetilde{h}_n$ minimizes ψ_n. Setting $h = \widetilde{h}_n + \epsilon\gamma$, we obtain that

$$\begin{aligned} \partial_\gamma \psi_n(\widetilde{h}_n) &\equiv \lim_{\epsilon \to 0} \frac{\psi_n(\widetilde{h}_n + \epsilon\gamma) - \psi_n(\widetilde{h}_n)}{\epsilon} \\ &= \gamma(T)(\widetilde{H}_n - \mathbb{H}_n)(T) - \gamma'(T)(\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(T) \\ &\quad + \int_0^T (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(s) d\gamma'(s). \end{aligned} \tag{3.4}$$

If $\widetilde{h}_n + \epsilon\gamma$ is in $\mathcal{K}_T$ for sufficiently small ϵ then $\partial_\gamma \psi_n(\widetilde{h}_n) \geq 0$. If, however, $\widetilde{h}_n \pm \epsilon\gamma$ is in $\mathcal{K}_T$ for sufficiently small ϵ then, $\partial_\gamma \psi_n(\widetilde{h}_n) \geq 0$ and $-\partial_\gamma \psi_n(\widetilde{h}_n) = \partial_{-\gamma} \psi_n(\widetilde{h}_n) \geq 0$, implying $\partial_\gamma \psi_n(\widetilde{h}_n) = 0$.

Choosing, respectively, $\gamma(t) \equiv 1, t, (t-y)_+, (y-t)_+$ and plugging these into (3.4) gives

$$(\widetilde{H}_n - \mathbb{H}_n)(T) \geq 0 \tag{3.5}$$

$$T(\widetilde{H}_n - \mathbb{H}_n)(T) - (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(T) \geq 0 \tag{3.6}$$

$$(T - y)(\widetilde{H}_n - \mathbb{H}_n)(T) - \int_y^T (\widetilde{H}_n - \mathbb{H}_n)(t) dt \geq 0 \tag{3.7}$$

$$(\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(y) \geq 0, \tag{3.8}$$

for any $y \in [0, T]$. The last statement is, of course, condition (2.9) of the characterization.

Consider the form of the function $\widetilde{h}_n$. This may be written as

$$\widetilde{h}_n(t) = \widetilde{a} + \sum_{j=1}^k \widetilde{\nu}_j (\tau_j - t)_+ + \sum_{j=1}^m \widetilde{\mu}_j (t - \eta_j)_+,$$

where $\widetilde{\nu}_j, \widetilde{a}, \widetilde{\mu}_j \geq 0$. We let τ_j denote the changes of slope where $\widetilde{h}_n$ is decreasing, and $\eta_j > 0$ denote the changes of slope where $\widetilde{h}_n$ is increasing. For simplicity we assume that these are ordered. Also, we have $\tau_k \leq \eta_1$.

Choose $\gamma(t) = \widetilde{h}_n, (\tau_j - t)_+, (t - \eta_j)_+$. For any such γ, if ϵ is small enough $\widetilde{h}_n \pm \epsilon\gamma$ is in $\mathcal{K}_T$. It thus follows that

$$0 = \widetilde{h}_n(T)(\widetilde{H}_n - \mathbb{H}_n)(T) - \widetilde{h}'_n(T)(\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(T) + \int_0^T (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(s) d\, \widetilde{h}'_n(s), \quad (3.9)$$

as well as

$$(\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(\tau_i) = 0 \quad (3.10)$$
$$(\widetilde{H}_n - \mathbb{H}_n)(\tau_i) = 0 \quad (3.11)$$
$$(\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(\eta_j) = (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(T) - (T - \eta_j)(\widetilde{H}_n - \mathbb{H}_n)(T) \quad (3.12)$$
$$(\widetilde{H}_n - \mathbb{H}_n)(\eta_j) = (\widetilde{H}_n - \mathbb{H}_n)(T). \quad (3.13)$$

We have used the same argument as in Corollary 2.4 to obtain the second and fourth equalities from (3.10), (3.8) and (3.12), (3.7) respectively.

Our next goal will be to show (2.7). In the characterization, two possibilities exist: either $\tilde{a} > 0$ or $\tilde{a} = 0$. In the former case, we know that $\widetilde{h}_n(t) \pm \epsilon\gamma(t)$ is in $\mathcal{K}_T$ for ϵ small enough and $\gamma \equiv 1$, and hence $\partial_\gamma \psi_n(\widetilde{h}_n) = 0$, implies condition (2.7).

Assume then that $\tilde{a} = 0$. This implies that $\widetilde{H}_n(\eta_1) = \widetilde{H}_n(\tau_k)$. From (3.8), we know that $(\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(\eta_1) \geq 0$, and (3.10), (3.11) imply that

$$\begin{aligned} (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(\eta_1) &= \int_{\tau_k}^{\eta_1} (\widetilde{H}_n - \mathbb{H}_n)(s) ds \\ &= \mathbb{H}_n(\tau_k)(\eta_1 - \tau_k) - \int_{\tau_k}^{\eta_1} \mathbb{H}_n(s) ds \leq 0. \end{aligned}$$

It thus follows that

$$(\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(\eta_1) = 0. \quad (3.14)$$

In particular, there must be no observations between τ_k and η_1 in this setting.

We now calculate, using (3.9),

$$\begin{aligned}
0 &= \widetilde{h}_n(T)(\widetilde{H}_n - \mathbb{H}_n)(T) - \widetilde{h}_n'(T)(\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(T) + \int_0^T (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(s)d\widetilde{h}_n'(s) \\
&\overset{(3.14)}{=} \widetilde{h}_n(T)(\widetilde{H}_n - \mathbb{H}_n)(T) - \widetilde{h}_n'(T)(\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(T) + \int_{\eta_1}^T (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(s)d\widetilde{h}_n'(s) \\
&\overset{(3.12)}{=} \widetilde{h}_n(T)(\widetilde{H}_n - \mathbb{H}_n)(T) - \widetilde{h}_n'(T)(\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(T) \\
&\qquad + \int_{\eta_1}^T [(\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(T) - (T-s)(\widetilde{H}_n - \mathbb{H}_n)(T)]d\widetilde{h}_n'(s) \\
&= \widetilde{h}_n'(\eta_1)\left[T(\widetilde{H} - \mathbb{H}_n)(T) - (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(T)\right] + \widetilde{h}_n(\eta_1)(\widetilde{H} - \mathbb{H}_n)(T) \\
&\geq \widetilde{h}_n'(\eta_1)\left[T(\widetilde{H} - \mathbb{H}_n)(T) - (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(T)\right] \geq 0,
\end{aligned}$$

by (3.5), (3.6), and the definition of η_1.

If $\widetilde{h}_n'(\eta_1) > 0$ where $\widetilde{h}_n'$ denotes the right derivative of $\widetilde{h}_n$, it follows that

$$T(\widetilde{H} - \mathbb{H}_n)(T) - (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(T) = 0.$$

Now, from (3.12) and (3.14) it follows that

$$0 = (T - \eta_1)(\widetilde{H}_n - \mathbb{H}_n)(T) - \int_{\eta_1}^T (\widetilde{H}_n - \mathbb{H}_n)(s)ds = -\eta_1(\widetilde{H}_n - \mathbb{H}_n)(T).$$

We thus obtain condition (2.7), for $\widetilde{h}_n'(\eta_1) > 0$.

If $\widetilde{h}_n'(\eta_1) = 0$, then $\widetilde{h}_n$ must be purely decreasing and also $\widetilde{h}_n(t) = 0$ for $t \in [\tau_k, T]$. This implies that

$$\begin{aligned}
(\widetilde{H}_n - \mathbb{H}_n)(T) &= \widetilde{H}_n(\tau_k) - \mathbb{H}_n(T) \\
&\overset{(3.11)}{=} \mathbb{H}_n(\tau_k) - \mathbb{H}_n(T) \leq 0.
\end{aligned}$$

Condition (2.7) now follows from (3.5). Note that again this implies that there must be no observations in the interval $[\tau_k, T]$.

We have thus obtained (2.7). Plugging it into (3.6) and using (3.8) implies condition (2.8). Lastly, using conditions (2.7) and (2.8) in (3.12), implies that

$$(\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(\eta_j) = 0.$$

Together with (3.10) this yields condition (2.10).

□

3.2. **MLE.** In this section, we show that the MLE of the hazard function exists, and is unique. We also prove the characterization of the MLE given in Lemma 2.1.

Proposition 3.2. *The function $\widehat{h}_n$ which minimizes φ_n over $\mathcal{K}_+$ is piecewise linear. It has at most one change of slope between jumps of $\mathbb{H}_n$, except perhaps in one such interval, where, if the estimator touches zero, it may have two changes of slope (it is zero between these two changes). Also, between zero and the first jump of $\mathbb{H}_n$, the minimizer may have at most one change of slope, but this happens only if it touches zero, and in this case the estimator is increasing, and equal to zero before the first change of slope. Between $X_{(n-1)}$ and $X_{(n)}$, the minimizer will also have at most one change of slope, and this only in the case if it is decreasing on $[X_{(n-1)}, X_{(n)})$, and equal to zero after the change.*

Proof. Consider any convex function h, and choose a g such that $h(X_i) = g(X_i)$ for $i = 1, \ldots, n-1$, and $h \geq g \geq 0$. It follows that $\varphi_n(h) - \varphi_n(g) \geq 0$ if and only if $H(X_i) \geq G(X_i)$ for $i = 1, \ldots, n$. Hence, the smaller we make g on $[0, X_{(n)})$, the smaller $\varphi_n(g)$ will become. The specific linear form of g now follows by arguing as in the proof of Proposition 3.1 for the LSE. □

We shall next provide a proof of the characterization of the MLE. Before we do this though, we provide a useful corollary of Lemma 2.1.

Corollary 3.3. *Let $\{\tau_i\}_{i=1}^k$ and $\{\eta_j\}_{j=1}^m$ denote the change points of $\widehat{h}_n$ as in (2.1). It follows that*

$$\int_0^{\tau_i} \frac{1}{\widehat{h}_n(t)} d\mathbb{F}_n(t) = \int_0^{\tau_i} \mathbb{S}_n(u) du, \tag{3.15}$$

$$\int_{\eta_j}^{\infty} \frac{1}{\widehat{h}_n(t)} d\mathbb{F}_n(t) = \int_{\eta_j}^{\infty} \mathbb{S}_n(u) du, \tag{3.16}$$

for $i = 1, \ldots, k$ and $j = 1, \ldots, m$.

Proof of Corollary 3.3. The function

$$\phi(x) \equiv \int_0^x \frac{x-t}{\widehat{h}_n(t)} d\mathbb{F}_n(t) - \int_0^x \int_0^t \mathbb{S}_n(s) ds dt$$

is maximized at τ_i, for $i = 1, \ldots, k$. Since it is also differentiable, (3.15) follows. A similar argument proves (3.16). □

Proof of Lemma 2.1. Consider any nonnegative convex function h. It follows that there exists a nonnegative constant a, and nonnegative measures ν and μ (indeed, these measures have supports with intersection containing at most one point), such that

$$h(t) = a + \int_0^{\infty} (x-t)_+ d\nu(x) + \int_0^{\infty} (t-x)_+ d\mu(x).$$

For any function $\hat{h}$ in $\mathcal{K}$ we calculate

$$\varphi_n(h) - \varphi_n(\widehat{h}) \geq \int_0^\infty \left\{ H(t) - \widehat{H}(t) + \left(1 - \frac{h(t)}{\widehat{h}(t)}\right) \mathbb{I}(t \neq X_{(n)}) \right\} d\mathbb{F}_n(t)$$

since $-\log x \geq 1 - x$. Plugging in the explicit form of h from above, we find that the right hand side is equal to

$$a \left\{ \int_{[0,\infty)} \left(t - \frac{1}{\widehat{h}(t)} \mathbb{I}(t \neq X_{(n)}) \right) d\mathbb{F}_n(t) \right\} + \left\{ \frac{n-1}{n} - \int_0^\infty \widehat{H}(t) d\mathbb{F}_n(t) \right\}$$
$$+ \int_0^\infty \left\{ \int_0^x \int_0^t \mathbb{S}_n(s) ds dt - \int_0^x \frac{x-t}{\widehat{h}(t)} \mathbb{I}(t \neq X_{(n)}) d\mathbb{F}_n(t) \right\} d\nu(x)$$
$$+ \int_0^\infty \left\{ \int_x^\infty \int_t^\infty \mathbb{S}_n(s) ds dt - \int_0^x \frac{t-x}{\widehat{h}(t)} \mathbb{I}(t \neq X_{(n)}) d\mathbb{F}_n(t) \right\} d\mu(x).$$

This is nonnegative if $\widehat{h}$ is a function which satisfies conditions (2.2)-(2.5). It follows that these conditions are sufficient to describe a minimizer of φ_n.

We next show that these conditions are necessary. To do this, we first define the directional derivative

$$\partial_\gamma \varphi_n(h) \equiv \lim_{\epsilon \to 0} \frac{\varphi_n(h + \epsilon\gamma) - \varphi_n(h)}{\epsilon} = \int_0^\infty \left\{ \Gamma(t) - \frac{\gamma(t)}{h(t)} \mathbb{I}(t \neq X_{(n)}) \right\} d\mathbb{F}_n(t). \tag{3.17}$$

If $\widehat{h}_n$ minimizes φ_n, then for any γ such that $\widehat{h}_n + \epsilon\gamma$ is in $\mathcal{K}_+$ for sufficiently small ϵ we must have $\partial_\gamma \varphi_n(\widehat{h}_n) \geq 0$. If, however, $\widehat{h}_n \pm \epsilon\gamma$ is in $\mathcal{K}_+$ for sufficiently small ϵ then, $\partial_\gamma \varphi_n(\widehat{h}_n) = 0$.

Choosing, respectively, $\gamma(t) \equiv 1, (t-y)_+, (y-t)_+$ then $\widehat{h}_n + \epsilon\gamma$ is in $\mathcal{K}_+$, and we obtain the inequalities in conditions (2.2)-(2.4). Since $(1 \pm \epsilon)\widehat{h}_n$ is also in $\mathcal{K}_+$, for sufficiently small ϵ, we obtain (2.5). Choosing, $\gamma = (\tau_i - t)_+, (t - \eta_j)_+$, yields the equalities in (2.2) and (2.3), since each of these functions $\widehat{h}_n \pm \epsilon\gamma$ is in $\mathcal{K}_+$.

Lastly, we prove (2.6). For any τ_i, define

$$\gamma(t) = \begin{cases} \widehat{h}_n(t) - \widehat{h}_n(\tau_i) & \text{for } t \in [0, \tau_i] \\ 0 & \text{otherwise.} \end{cases}$$

Since $(1 \pm \epsilon)\gamma$ is also in $\mathcal{K}_+$, it follows that $\partial_\gamma \varphi_n(\widehat{h}_n) = 0$ and hence

$$0 = \left\{ \int_0^{\tau_i} \widehat{H}_n(t) d\mathbb{F}_n(t) - \mathbb{F}_n(\tau_i) + \widehat{H}_n(\tau_i)\mathbb{S}_n(\tau_i) \right\}$$
$$+ \widehat{h}_n(\tau_i) \left\{ \int_0^{\tau_i} \frac{1}{\widehat{h}_n(t)} d\mathbb{F}_n(t) - \int_0^{\tau_i} t d\mathbb{F}_n(t) - \tau_i \mathbb{S}_n(\tau_i) \right\}.$$

Integration by parts and Corollary 3.3 yield (2.6) for $x = \tau_i$. The case when $x = \eta_j$, is obtained in a similar manner, but using

$$\gamma(t) = \begin{cases} 0 & \text{for } t \in [0, \eta_j] \\ \widehat{h}_n(t) - \widehat{h}_n(\eta_j) & \text{otherwise.} \end{cases}$$

Here, we also use (2.5). □

Corollary 3.4. *Suppose that $\widehat{h}_n$ is strictly positive, and recall the formulation given in* (2.1). *Then we also have that*

$$\int_0^{\eta_1} \frac{\eta_1 - t}{\widehat{h}_n(t)} d\mathbb{F}_n(t) = \int_0^{\eta_1} \int_0^s \mathbb{S}_n(u) du ds, \tag{3.18}$$

$$\int_{\tau_k}^{\infty} \frac{t - \tau_k}{\widehat{h}_n(t)} d\mathbb{F}_n(t) = \int_{\tau_k}^{\infty} \int_t^{\infty} \mathbb{S}_n(s) ds dt, \tag{3.19}$$

$$\int_0^{\eta_1} \frac{1}{\widehat{h}_n(t)} d\mathbb{F}_n(t) = \int_0^{\eta_1} \mathbb{S}_n(u) du, \tag{3.20}$$

$$\int_{\tau_k}^{\infty} \frac{1}{\widehat{h}_n(t)} d\mathbb{F}_n(t) = \int_{\tau_k}^{\infty} \mathbb{S}_n(u) du. \tag{3.21}$$

Proof. The first two equalities follow by noting that if $\widehat{h}_n$ is strictly positive, then for ϵ sufficiently small, $\widehat{h}_n \pm \epsilon\gamma$ is in $\mathcal{K}_+$ for $\gamma(t) = (t - \tau_k)_+, (\eta_1 - t)_+$. Arguing as for Corollary 3.3 proves the remaining lines. □

This corollary allows us to extend the equalities of the characterization of the MLE to some extra change points. The significance of this will become clear in Sections 6 and 7, where we consider asymptotics of the estimator.

Proposition 3.5. *There exists a unique minimizer $\widehat{h}_n$ of φ_n over $\mathcal{K}_+$.*

Proof. We will show that a minimizer exists by reducing the search to bounded positive convex functions on a compact domain. As this is a compact set, a minimizer of φ_n exists. However, since φ_n is not strictly convex, we will later need to argue that the minimizer is unique.

We must first handle the issue of a compact domain: As we assume apriori that $\widehat{h}_n(X_{(n)}) = \infty$, then we are really looking for the minimizer of the the modified negative of the loglikelihood with domain $[0, X_{(n)})$. However, we have also argued that the minimizer must have the specific functional form as described in Proposition 3.2. Therefore, it is sufficient to reduce the domain to $[0, X_{(n-1)} + \delta]$, for any $\delta > 0$, since $\widehat{h}_n$ is then extended linearly beyond $X_{(n-1)} + \delta$ in a unique manner. Therefore, it will be sufficient to show that we may reduce the search to functions bounded on $[0, X_{(n-1)}]$, with a derivative at $X_{(n-1)}$ which is bounded above.

Recall that the minimizer must satisfy

$$\int_0^\infty H(t)d\mathbb{F}_n(t) = 1 - 1/n, \tag{3.22}$$

and we may hence reduce our search to the class of functions which satisfies this condition. For any such h, write $h = h_+ + h_-$, where h_+ is increasing and h_- is decreasing. It follows that for any x

$$1 \geq \int_0^\infty H(t)d\mathbb{F}_n(t) = \int_0^\infty h(t)\mathbb{S}_n(t)dt \geq h_-(x)\int_0^x \mathbb{S}_n(t)dt.$$

A similar bound for h_+ yields

$$h(x) \leq \frac{1}{\int_0^x \mathbb{S}_n(t)dt} + \frac{1}{\int_x^\infty \mathbb{S}_n(t)dt} \equiv M_n(x)$$

for all x in $(0, X_{(n)})$. Thus we know that $h(x)$ must be bounded for $x \in (0, X_{(n-1)}]$.

To show that h is also bounded at zero, we need to show that $h'(X_{(1)})$ is bounded from below. Assuming that it is negative, we may write for $0 < x \leq X_{(1)}$

$$h(X_{(1)}) + h'(X_{(1)})(x - X_{(1)}) = h(x) \leq M_n(x).$$

Thus, fix $x^* > 0$ and less than $X_{(1)}$; we obtain that

$$h'(X_{(1)}) \geq \frac{M_n(x^*) - h(X_{(1)})}{x^* - X_{(1)}},$$

from which it follows that h must be bounded on the set $(0, X_{(n-1)}]$.

By (2.5), we also have that

$$n \geq H(X_{(n)}) \geq \int_{X_{(n-1)}}^{X_{(n-1)}+\delta} h(t)dt = \int_{X_{(n-1)}}^{X_{(n-1)}+\delta} \left\{h(X_{(n-1)}) + h'(X_{(n-1)})(t - X_{(n-1)})\right\} dt,$$

if h is increasing on $[X_{(n-1)}, X_{(n)})$. This implies that $h'(X_{(n-1)})$ is bounded above, completing the proof.

We now show uniqueness. Suppose that h_1 and h_2 both minimize φ_n. It follows from the arguments above that they must be piecewise linear, with at most one change of slope between successive order statistics $X_{(i)}$ (with at most one exceptional interval). Also, we have shown that $\int_0^\infty H_1(t)d\mathbb{F}_n(t) = \int_0^\infty H_2(t)d\mathbb{F}_n(t)$. Therefore $\varphi_n(h_1)$ and $\varphi_n(h_2)$ differ only in the term $-\int_0^\infty \log h_i(t)\mathbb{I}(t \neq X_{(n)})d\mathbb{F}_n(t)$. However, this term is strictly convex, and it follows that $h_1(X_{(i)}) = h_2(X_{(i)})$ for all $i = 1, \ldots, n-1$.

Let $\bar{h} = (h_1 + h_2)/2$. By linearity, we have that $\varphi_n(h_1) = \varphi(h_2) = \varphi(\bar{h})$, which implies that $\bar{h}$ is also a minimizer. However, the only way that this is possible is if $\bar{h}$ also satisfies the conditions of Proposition 3.2. This implies the following:

(1) Either both h_1 and h_2 are increasing and $h_1(0) = h_2(0) = 0$. In this case, they must have the same locations of their changes of slope, as otherwise $\bar{h}$ violates Proposition 3.2.

(2) Otherwise, (1) does not hold. In this case though, by the same argument as above, if h_1 and h_2 have at least one change of slope in an interval between observations (or between zero and $X_{(1)}$), then these locations of change of slope must be equal.

If the first case holds, then it is not difficult to see that $h_1 \equiv h_2$ on $[0, X_{(n)}]$, as $h_1(t) = h_1(t) = 0$ on $[0, \tau_1]$, and $h_1(X_i) = h_2(X_i)$ for all observation points.

In the second case, we use a different argument. We know that neither h_1 nor h_2 have touch points before $X_{(1)}$. Let t^* denote the first touch point of h_1, and (without loss of generality) assume that the first touch point of h_2 is greater than t^*. Hence, by (2.6),

$$h_1(X_{(1)}) = h_2(X_{(1)}), \qquad \int_0^{t^*} h_1(t)dt = \mathbb{F}_n(t^*).$$

Next, notice that $\bar{h} = (h_1 + h_2)/2$ and h_2 are also minimizers of the MLE criterion function φ_n. Notice that $\bar{h}$ also has a touch point at t^*, and that $\bar{h}(X_{(1)}) = h_2(X_{(1)})$.

Now continue averaging $\bar{h}$ with h_2: this yields the functions

$$\bar{h}_l = 2^{-l}(h_1 - h_2) + h_2,$$

which satisfy

$$\bar{h}_l(X_{(1)}) \ = h_2(X_{(1)}), \qquad \textstyle\int_0^{t^*} \bar{h}_l(t)dt = \mathbb{F}_n(t^*).$$

Since $\bar{h}_l \to h_2$ pointwise, it follows from the dominated convergence theorem that

$$\int_0^{t^*} h_2(t)dt = \mathbb{F}_n(t^*).$$

Therefore, since h_1 and h_2 are both linear on $[0, t^*]$,with

$$h_1(X_{(1)}) = h_2(X_{(1)}), \qquad \int_0^{t^*} h_1(t)dt = \int_0^{t^*} h_2(t)dt,$$

it follows that they both must have the same value *and slope* at $X_{(1)}$; i.e. both $h_1(X_{(1)}) = h_2(X_{(2)})$ and $h_1'(X_{(1)}) = h_2'(X_{(2)})$ hold.

Now write

$$h_1(t) = a_1 + b_1 t + \sum_{i=1}^{m_1-1} \nu_{i,1}(t - t_{i,1})_+,$$
$$h_2(t) = a_2 + b_2 t + \sum_{i=1}^{m_2-1} \nu_{i,2}(t - t_{i,2})_+,$$

where $X_{(1)} < t_{1,j} < t_{2,j} < \cdots < t_{m_j-1,j} < X_{(n)}$, $j = 1, 2$, and where $h_1(X_{(i)}) = h_2(X_{(i)})$ for $i = 1, \ldots, n$. We also assume that $\nu_{i,j} > 0$ for $i = 1, \ldots, m_j - 1$, $j = 1, 2$.

This implies in particular that $h_j(t) = a_j + b_j t$ for $t \le t_{1,j}$, $j = 1, 2$, and since $X_{(1)} < t_{1,j}$, $j = 1, 2$,

$$h_1(X_{(1)}) = h_2(X_{(1)}).$$

Thus $a_1 + b_1 X_{(1)} = a_2 + b_2 X_{(1)}$. From the argument above

$$b_1 = h_1'(X_{(1)}) = h_2'(X_{(1)}) = b_2.$$

We conclude that $a_1 = a_2$ and $b_1 = b_2$ so that $h_1(t) = h_2(t)$ for $0 \le t \le t^*$. It also follows that $t_{1,1} = t_{1,2}$.

Repeating this argument on the interval $[t^*, t^{**}]$ with $t^{**} = \min\{t_{2,1}, t_{2,2}\}$ shows that $\nu_{1,1} = \nu_{1,2}$ or $t_{2,1} = t_{2,2}$. Proceeding by induction yields $\nu_{j,1} = \nu_{j,2}$ and $t_{j+1,1} = t_{j+1,2}$ for $j = 1, \ldots, m_1 - 1 = m_2 - 1$, and hence uniqueness. $\square$

4. Consistency

Theorem 4.1. *(Consistency of the MLE). Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables with convex hazard function and corresponding distribution function F_0. Let $T_0 \equiv T_0(F_0) \equiv \inf\{t : \ F_0(t) = 1\}$. Then the MLE $\widehat{h}_n(t)$ is consistent for all $t \in (0, T_0)$. Also for all $\delta > 0$,*

$$\sup_{\delta \le t \le T_0 - \delta} |\widehat{h}_n(t) - h(t)| \to 0 \qquad \textit{almost surely}$$

if $T_0 < \infty$. If $T_0 = \infty$, the above statement holds with $T_0 - \delta$ replaced by any $K < \infty$.

Proof. We first show that $\widehat{h}_n$ is bounded appropriately so that we can select convergent subsequences.

For any convex hazard function h we can write

$$h = h_\downarrow + h_\uparrow \tag{4.1}$$

where $h_\downarrow$ is nonincreasing and $h_\uparrow$ is nondecreasing.

Now write $\widehat{h}_n = \widehat{h}_{n,\downarrow} + \widehat{h}_{n,\uparrow}$. Then from (2.5)

$$1 \ge \int_0^\infty \widehat{h}_n(t)\mathbb{S}_n(t)dt \ge \int_0^x \widehat{h}_{n,\downarrow}(t)\mathbb{S}_n(t)dt \ge \widehat{h}_{n,\downarrow}(x)\int_0^x \mathbb{S}_n(t)dt.$$

This yields

$$\widehat{h}_{n,\downarrow}(x) \le \frac{1}{\int_0^x \mathbb{S}_n(t)dt} \tag{4.2}$$

where the right side is almost surely bounded, and, in fact, converges almost surely to $1/\int_0^x S_0(t)dt < \infty$ for all $x > 0$.

Similarly, for $x \in (\mathrm{supp}(F_0))^\circ$, and fixed $\delta > 0$

$$1 \ \ge \ \int_0^\infty \widehat{h}_{n,\uparrow}(t)\mathbb{S}_n(t)dt \ge \widehat{h}_{n,\uparrow}(x)\int_x^{x+\delta} \mathbb{S}_n(t)dt.$$

This yields

$$\widehat{h}_{n,\uparrow}(x) \leq \frac{1}{\int_x^{x+\delta} \mathbb{S}_n(t)dt} \tag{4.3}$$

where the right side is almost surely bounded for $x \in (\mathrm{supp}(F_0))^\circ$, and converges almost surely to $1/\int_x^{x+\delta} S_0(t)dt < \infty$.

Remark 4.2. *Indeed, using similar arguments, one may show that if the first moment of F_0 is infinite, then h_0 must be nonincreasing in the tail. If h_0 is nondecreasing in the tail, it is straightforward to show that F_0 must have a finite first moment.*

Now we take $\gamma = h_0$ in the directional derivative, (3.17); it follows that

$$0 \leq \lim_{\epsilon \downarrow 0} \frac{\varphi_n(\widehat{h}_n + \epsilon h_0) - \varphi_n(\widehat{h}_n)}{\epsilon} = \int_0^\infty \left\{ H_0(t) - \frac{h_0(t)}{\widehat{h}_n(t)} \right\} d\mathbb{F}_n(t),$$

noting that $\widehat{h}_n(X_{(n)}) = \infty$, and hence,

$$\int_0^\infty \frac{h_0(t)}{\widehat{h}_n(t)} d\mathbb{F}_n(t) \leq \int_0^\infty H_0(t) d\mathbb{F}_n(t) \to_{a.s.} \int_0^\infty H_0(t) dF_0(t) = 1.$$

Fix any $0 < a < b < \infty$ such that $a, b \in (\mathrm{supp}(F_0))^\circ$. It follows that $\lim_n X_{(n)} > b$ with probability one (this can be shown using the Borel-Cantelli theorem). Also, $\sup |\mathbb{F}_n(t) - F_0(t)| \to_{a.s.} 0$ by the Glivenko-Cantelli lemma. Both of these events occur on the set Ω, with $P(\Omega) = 1$. Fix $\omega \in \Omega$. We will show that $\widehat{h}_n \to h_0$ for such an ω.

Let $\{n'\}$ denote any subsequence of $\{n\}$. By the bounds in (4.2) and (4.3) (which are finite for our choice of ω), using a classical diagonalization argument and the continuity of convex functions, we may extract a further subsequence $\{n''\}$ such that $\widehat{h}_{n''} \to \widehat{h}$ pointwise on $[a, b]$, where the limit $\widehat{h}$ must be convex. We denote the subsequence as $\{n\}$ to simplify notation.

From Fatou's lemma, it follows that

$$\begin{aligned} \int_a^b \frac{h_0^2(t)}{\widehat{h}_n(t)} S_0(t) dt &= \int_a^b \frac{h_0(t)}{\widehat{h}_n(t)} f_0(t) dt \leq \liminf_n \int_a^b \frac{h_0(t)}{\widehat{h}_n(t)} d\mathbb{F}_n(t) \\ &\leq \limsup_n \int_0^\infty \frac{h_0(t)}{\widehat{h}_n(t)} d\mathbb{F}_n(t) \leq \lim_n \int_0^\infty H_0(t) d\mathbb{F}_n(t) \leq 1. \end{aligned}$$

Note that this implies that if $\widehat{h}(t) = 0$ then $h_0(t) = 0$. By (2.5) and integration by parts, we see that $1 \geq \int_{[0, X_{(n)})} \widehat{h}_n(t) \mathbb{S}_n(t) dt$. Therefore, applying Fatou's lemma again,

$$1 \geq \int_a^b \widehat{h}(t) S_0(t) dt.$$

It also follows that,

$$\begin{aligned} 0 &\leq \int_a^b \frac{(\widehat{h}(t)-h_0(t))^2}{\widehat{h}(t)} S_0(t)dt \\ &= \int_a^b \widehat{h}(t)S_0(t)dt - 2\int_a^b h_0(t)S_0(t)dt + \int_a^b \frac{h_0^2(t)}{\widehat{h}(t)} S_0(t)dt. \\ &\leq 2 - 2\int_a^b h_0(t)S_0(t)dt. \end{aligned}$$

Define $\widehat{h} = h_0$ for $t \notin [a, b]$, which allows us to let both $a, b \to \infty$ in the above display. Since $\int_0^\infty h_0(t)S_0(t)dt = 1$, it follows that

$$\int_0^\infty \frac{(\widehat{h}(t)-h_0(t))^2}{\widehat{h}(t)} S_0(t)dt = 0,$$

and this implies that $\widehat{h}(t) = h_0(t)$ for all $t \in [a, b]$.

We have thus shown that every subsequence $\{\widehat{h}_n(x)\}$ has a further subsequence which converges to the true hazard function $h_0(x)$ pointwise, for all $x \in (\text{supp}F_0)^\circ$. It follows that $\{\widehat{h}_n\}$ converges to h_0 pointwise. By Theorem 10.8, page 90, Rockafellar (1970), this implies that the claimed uniform convergence on $[a, b]$ also holds. As this happens for any $\omega \in \Omega$, and $P(\Omega) = 1$, we have shown the result. □

Corollary 4.3. *Suppose that h_0'' is continuous and strictly positive at x_0. It follows that there exist touchpoints $\tau_n \leq x_0 \leq \eta_n$ such that $\tau_n, \eta_n \to x_0$ in probability.*

Proof. Let η_n, τ_n be touchpoints such that $\tau_n \leq x_0 \leq \eta_n$. If τ_n does not exist then set $\tau_n = 0$, and $\eta_n = \infty$ otherwise. Suppose that it is not the case that $\tau_n, \eta_n \to_p x_0$. Then it follows from Theorem 4.1 that there exists an interval $I = [a, b]$ such that $x_0 \in I, |I| > 0$, and $\limsup_n \tau_n \leq a, \liminf_n \eta_n \geq b$ almost surely, and lastly $\lim \widehat{h}_n(t) \to_{a.s.} h_0(t)$ on I. However, this implies that $h_0(t)$ is linear on I, which is a contradiction. □

Theorem 4.4. *(Consistency of the LSE). Suppose that $H_0(T) < \infty$, and $h_0 \in \mathcal{K}_T$. Then the LSE $\widetilde{h}_n$ described in Section 3 is consistent: the estimator $\widetilde{h}_n(t)$ converges to $h_0(t)$ for $t \in (0, T)$ with probability one. Also, for all $\delta > 0$, we have*

$$\sup_{t\in[\delta,T-\delta]} |\widetilde{h}_n(t) - h_0(t)| \to_{a.s.} 0.$$

Remark 4.5. *Using the same argument as in Groeneboom et al. (2001b) in Remark on page 1673, it follows that if h_0 is decreasing near zero, then $\widehat{h}_n$ is not consistent at zero. The same holds for $\widetilde{h}_n$. If h_0 is increasing near T, then $\widetilde{h}_n$ is not consistent at T. Lastly, if h_0 is increasing in the tail, then $\widehat{h}_n$ is not consistent in the tail.*

Sketch of proof. The proof of the statement near zero is *exactly* the same as in Groeneboom et al. (2001b) (since $h(0) = f(0)$). For the LSE at T, we obtain from (3.7) that

$$(T - X_{(N)})\widetilde{h}_n(T) \geq 2\mathbb{H}_n(T),$$

where $N = N(n, \omega)$ is the index of the largest observation less than or equal to T. Now, $\mathbb{H}_n(T)/(T - X_{(N)})$ will blow up as n gets large.

For the MLE, we get from (2.3) that

$$\widehat{h}_n(X_{(n)}) \geq \frac{n}{X_{(n)} - X_{(n-1)}},$$

and as n grows, the right hand side of the above display converges to a random variable (using the same arguments as in page 1673 of Groeneboom et al. (2001b)), Z, such that $P(Z > c) > 0$ for all $c > 0$. □

Proof of Theorem 4.4. Let Ω denote the set such that

$$\|\mathbb{H}_n - H_0\|_0^T \equiv \sup_{0 \leq t \leq T} |\mathbb{H}_n(t) - H_0(t)| \to 0 \tag{4.4}$$

holds; note that $P(\Omega) = 1$: see e.g. Shorack and Wellner (1986), Theorem 7.3.1, page 304.

Let $\{n'\}$ denote any subsequence of $\{n\}$. Our goal will be to show that for any such subsequence there exists a further subsequence $\{n''\}$ such that $\widetilde{h}_{n''}$ converges (pointwise) to h_0 for our chosen ω. We do this in two steps. First, we consider the case where $h_0(0)$ and $h_0(T)$ are both finite. This gives us that $\int_0^T h_0^2(t)dt < \infty$, and we may use the least squares criterion function to obtain the result. The argument becomes more delicate when $\int_0^T h_0^2(t)dt$ may be infinite, and this is handled in the second case.

Suppose then that $h_0(0), h_0(T) < \infty$. Fix any $\delta > 0$, and let τ_n be the last changepoint before δ (0 if no such point exists). We will next show that $\widetilde{h}_n(\tau_n)$ and $\widetilde{h}'_n(\tau_n)$ are uniformly bounded.

To do this we write $\widetilde{h}_n = \widetilde{h}_{n,\downarrow} + \widetilde{h}_{n,\uparrow}$, the sum of its increasing and decreasing components. From (3.2) and (2.7) it follows that

$$\begin{aligned} &\text{if } \tau_n \geq \delta/2, && \widetilde{h}_{n,\downarrow}(\tau_n) \leq \widetilde{h}_{n,\downarrow}\left(\tfrac{\delta}{2}\right) \leq 2\,\tfrac{\mathbb{H}_n(T)}{\delta}, \\ &\text{and if } \tau_n < \delta/2, && \mathbb{H}_n(T) \geq \textstyle\int_{\tau_n}^{\delta} \widetilde{h}_{n,\downarrow}(t)dt \geq \tfrac{\delta}{4}(\widetilde{h}_{n,\downarrow}(\tau_n) + \widetilde{h}_{n,\downarrow}(\delta)). \end{aligned}$$

Since (4.4) holds and $H_0(T) < \infty$, it follows that $\widetilde{h}_{n,\downarrow}(\tau_n)$ is bounded. Similarly, $\widetilde{h}_{n,\uparrow}(\tau_n) \leq \widetilde{h}_{n,\uparrow}(\delta) \leq \mathbb{H}_n(T)/(T - \delta)$, and hence $\widetilde{h}_n(\tau_n)$ is bounded by a constant depending on $H_0(T)$ and δ.

To bound the derivative, notice that (the right derivative) $\widetilde{h}'_n(\tau_n)$ is the same as the left derivative of $\widetilde{h}_n$ at δ. Writing the latter as $\widetilde{h}'_n(\delta)$, from convexity it follows that

$$\mathbb{H}_n(T) \geq \int_0^\delta \widetilde{h}_n(\delta) + \widetilde{h}_n'(\delta)\, t\, dt.$$

We thus conclude that (again, the right derivative) $|\widetilde{h}_n'(\tau_n)|$ is also bounded uniformly in n.

Similarly, let η_n denote the first changepoint after $T-\delta$. By an identical argument, we also obtain that the left derivative $\widetilde{h}_n'(\eta_n)$ and $\widetilde{h}_n(\eta_n)$ are bounded uniformly in n.

Let $\{n'\}$ be any subsequence of $\{n\}$. Because of the above bounds on $\widetilde{h}_n$, we may select a subsequence $\{n''\}$ of $\{n'\}$ such that $\tau_{n''} \to \tau$, $\eta_{n''} \to \eta$, and $\widetilde{h}_{n''}(t) \to \widetilde{h}(t)$ for all $t \in (\tau, \eta)$. Here $\widetilde{h}$ is convex, $\tau \leq \delta$, and $\eta \geq T - \delta$. We will next show that $\widetilde{h} \equiv h_0$ on (τ, η). To simplify notation we denote the subsequence $\{n''\}$ simply as $\{n\}$.

Let $\psi_n^{a,b}(h) = \frac{1}{2}\int_a^b h^2(t)dt - \int_a^b h d\mathbb{H}_n$. Repeating the integration by parts argument used in (3.1) we obtain that

$$\begin{aligned}
&\psi_n^{a,b}(h_0) - \psi_n^{a,b}(\widetilde{h}_n) \\
&\quad\geq\ [h_0 - \widetilde{h}_n](t)(\widetilde{H}_n - \mathbb{H})(t)\Big|_a^b - [h_0 - \widetilde{h}_n]'(t)(\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(t)\Big|_a^b \\
&\qquad + \int_0^T (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n)(s) d[h_0 - \widetilde{h}_n]'(s). \qquad (4.5)
\end{aligned}$$

From the characterization of $\widetilde{h}_n(t)$, (2.7)-(2.10), it follows that the right side of the last display is non-negative and hence

$$\begin{aligned}
&\psi_n^{\tau_n,\eta_n}(h_0) - \psi_n^{\tau_n,\eta_n}(\widetilde{h}_n) \\
&= \frac{1}{2}\int_{\tau_n}^{\eta_n} h_0^2(t)dt - \int_{\tau_n}^{\eta_n} h_0(t)d\mathbb{H}_n(t) - \frac{1}{2}\int_{\tau_n}^{\eta_n} \widetilde{h}_n^2(t)dt + \int_{\tau_n}^{\eta_n} \widetilde{h}_n(t)d\mathbb{H}_n(t) \geq 0. \quad (4.6)
\end{aligned}$$

Recall that $\widetilde{h}_n$ converges pointwise and is bounded above uniformly in n. Moreover, (4.4) holds, and $h_0 \in \mathcal{K}_+$. These facts allow us to take limits in the above display (using e.g. Proposition 18 on p. 270 of Royden (1988)) to obtain that

$$\int_\tau^\eta (\widetilde{h}(t) - h_0(t))^2 dt \leq 0,$$

from which the desired result follows.

Since δ may be chosen arbitrarily small in the above argument, we have thus shown that $\widetilde{h}_n(t) \to h_0(t)$ pointwise on $(0, T)$. By convexity, we obtain uniform convergence on closed subsets.

It remains to extend the argument to the case when $h_0(0), h_0(T)$ are possibly not finite. As we mentioned previously, this could possibly imply that $\int_0^T h_0^2(t)dt = \infty$, which would cause a problem in the above argument. We give the details assuming that $h_0(0) < \infty$, and $h_0(T) = \infty$, as the argument is the same on the other side (and in particular given in Groeneboom et al. (2001b)). The idea of the argument is as follows: if $h_0(T) = \infty$ then for all $x < T$ there exists a y, where $x < y < T$, and a sequence of changepoints, $\{\eta_n\}$, such that $\eta_n \to y$ (possibly along a subsequence). Also, for any $y < T$, $h_0(y) < \infty$. This allows us to repeat the argument above to show that

$$\int_\tau^y (\widetilde{h}(t) - h_0(t))^2 dt \leq 0.$$

The reason that these points exist is that we select the point y such that h_0 is strictly convex at y, and hence $\widetilde{h}_n$ must have changepoints converging to y.

We next give the remaining details. Choose any $\delta > 0$. Since $h_0(T) = \infty$, $h_0(t)$ cannot be purely linear on $[\delta, T]$. That is, there exists a point $y \in (T - \delta, T)$ such that λ_0 is strictly convex at y.

Let ϵ_n and η_n be the changepoints of $\widetilde{h}_n$ closest to y such that ϵ_n is smaller than y and η_n is greater than y (note, if no such points exist we select 0, and T respectively for ϵ_n and η_n). Since the points are contained inside $[0, T]$, there exists a subsequence (which we denote by $\{n\}$) such that $\epsilon_n \to \epsilon$ and $\eta_n \to \eta$.

By definition, $0 \leq \epsilon \leq y \leq \eta \leq T$. We next show that the only possibility is that $\epsilon = y = \eta$. By way of contradiction, suppose then that $\epsilon < \eta$.

We first show that $\widetilde{h}_n(\eta_n)$ is bounded. To do this we write (as before), $\widetilde{h}_n = \widetilde{h}_{n,\downarrow} + \widetilde{h}_{n,\uparrow}$, the sum of its increasing and decreasing components. From (3.2) and (2.7) it follows that

$$\begin{aligned} & & \widetilde{h}_{n,\downarrow}(\eta_n) &\leq \widetilde{h}_{n,\downarrow}(T - \delta) \leq \tfrac{\mathbb{H}_n(T)}{T-\delta}, \\ &\text{if } \eta_n \leq T - \delta/2 & \widetilde{h}_{n,\uparrow}(\eta_n) &\leq \widetilde{h}_{n,\uparrow} + \left(T - \tfrac{\delta}{2}\right) \leq 2\,\tfrac{\mathbb{H}_n(T)}{\delta}, \\ &\text{and if } \eta_n > T - \delta/2 & \mathbb{H}_n(T) &\geq \textstyle\int_{T-\delta}^{\eta_n} \widetilde{h}_{n,\uparrow}(t)dt \geq \tfrac{\delta}{4}(\widetilde{h}_{n,\uparrow}(\eta_n) + \widetilde{h}_{n,\uparrow}(T - \delta)). \end{aligned}$$

Since (4.4) holds and $H_0(T) < \infty$, it follows that $\widetilde{h}_n(\eta_n)$ is bounded.

Let $\{t_n\}_{n=0}^m$ denote the changepoints of $\widetilde{h}_n$, with $t_0 = 0$ and $t_m = T$. From (2.7)-(2.10) it follows that $\int_0^{t_k} \widetilde{h}_n(t)dt = \mathbb{H}_n(t_k)$ for all $k = 0, \ldots m$. Since $\widetilde{h}_n$ is piecewise linear, we may also write

$$\int_0^{t_k} \widetilde{h}_n(t)dt = \sum_{i=0}^{k-1} \frac{\widetilde{h}_n(t_{i+1}) + \widetilde{h}_n(t_i)}{2}(t_{i+1} - t_i).$$

Hence by Corollary 2.4 we have

$$\frac{1}{2}\left(\widetilde{h}_n(\epsilon_n) + \widetilde{h}_n(\eta_n)\right) = \frac{\mathbb{H}_n(\eta_n) - \mathbb{H}_n(\epsilon_n)}{\eta_n - \epsilon_n}.$$

Hence, $\widetilde{h}_n(\epsilon_n)$ must also be bounded uniformly in n. The same is true of the left derivative of $\widetilde{h}'_n(\eta_n)$ (which is equal to the right derivative of $\widetilde{h}'_n(\epsilon_n)$ by definition of ϵ_n and η_n). Thus, there exists a further subsequence (again denoted by $\{n\}$), so that $\widetilde{h}_n(t) \to \widetilde{h}(t)$ for t in (ϵ, τ).

By integration by parts, it follows from (2.7)-(2.10), that for any $x \in [0,T]$

$$\int_{\epsilon_n}^{x}(x-t)\widetilde{h}_n(t)dt \geq \int_{\epsilon_n}^{x}(x-t)d\mathbb{H}_n(t).$$

Letting $n \to \infty$, this implies that

$$\int_{\epsilon}^{x}(x-t)\widetilde{h}(t)dt \geq \int_{\epsilon}^{x}(x-t)h_0(t)dt.$$

Choosing $x = \eta$ in the above, shows that it is not possible that $\eta = T$, as $h_0(T) = \infty$, while $\widetilde{h}(T)$ is bounded (because $\widetilde{h}_n(\eta_n)$ and $\widetilde{h}'_n(\eta_n)$ are bounded).

If, on the other hand, $\eta < T$, then, since $h_0(\eta) < \infty$, we may use (4.5) with $(a,b) = (\epsilon_n, \tau_n)$ (as in (4.6)) to obtain that

$$\int_{\epsilon}^{\eta}(\widetilde{h}(t) - h_0(t))^2 dt \leq 0.$$

However, this implies that $\widetilde{h}(t) = h_0(t)$ on $[\epsilon, \eta]$, where $\widetilde{h}$ is linear and h_0 is not. The result follows. □

Notice that we have in fact proved the following result.

Corollary 4.6. *Suppose that h_0'' is continuous and strictly positive at x_0. It follows that there exist touch points (changes of slope of $\tilde{h}_n$) $\tau_n \leq x_0 \leq \eta_n$ such that $\tau_n, \eta_n \to_{a.s.} x_0$.*

From the above we also obtain consistency of the derivatives of both the MLE and LSE. This follows from the following result.

Lemma 4.7. *Suppose that $\bar{h}_n$ is a sequence of functions in $\mathcal{K}_+$ (or $\mathcal{K}_T$) satisfying $\sup_{a \leq x \leq b} |\bar{h}_n(t) - h_0(t)| = 0$ with probability one. Then (also with probability one) for all $x \in (a,b)$*

$$-\infty < h_0'(x^-) \leq \liminf_{n\to\infty} \bar{h}_n'(x^-) \leq \limsup_{n\to\infty} \bar{h}_n'(x^+) \leq h_0'(x^+) < \infty.$$

Proof. Let $\epsilon > 0$. Since $\bar{h}_n \in \mathcal{K}_+$

$$\frac{\bar{h}_n(x-\epsilon) - \bar{h}_n(x)}{-\epsilon} \leq \bar{h}_n'(x^-) \leq \bar{h}_n'(x^+) \leq \frac{\bar{h}_n(x+\epsilon) - \bar{h}_n(x)}{\epsilon}.$$

Letting $n \to \infty$ and then $\epsilon \to 0$ proves the result. □

Here is the immediate corollary for the our two sequences of estimators:

Corollary 4.8. *Suppose that $x \in (a,b)$ and with $\overline{h}_n$ the MLE $\widehat{h}_n$ or the LSE $\widetilde{h}_n$ we have $\sup_{a \leq t \leq b} |\overline{h}_n(t) - h_0(t)| \to_{a.s.} 0$. Then $\overline{h}_n'(x) \to_{a.s.} h_0'(x)$ at all continuity points x of h_0'.*

5. Asymptotic lower bounds for the minimax risk

Define the class of densities $\mathcal{C}$ by

$$\begin{aligned}\mathcal{C} \;=\; \Big\{ & f : [0,\infty) \to [0,\infty) : \int_0^\infty f(x)dx = 1, \\ & h(x) = f(x)/(1-F(x)) \text{ is convex}, \quad h(x) > 0 \text{ for all } x > 0 \}\end{aligned}$$

We want to derive asymptotic lower bounds for the local minimax risks for estimating the convex hazard function h and its derivative at a fixed point. The L_1- minimax risk for estimating a functional T of f_0 based on a sample $X_1, \ldots, X_n$ of size n from f_0 which is known to be in a subset $\mathcal{C}_n$ of $\mathcal{C}$ is defined by

$$MMR_1(n, T, \mathcal{C}_n) = \inf_{T_n} \sup_{f \in \mathcal{C}_n} E_f |T_n - Tf|. \tag{5.1}$$

where the infimum ranges over all possible measurable functions $T_n = t_n(X_1, \ldots, X_n)$ mapping $\mathbb{R}^n$ to $\mathbb{R}$. The shrinking classes $\mathcal{C}_n$ used here are Hellinger balls centered at f_0:

$$\mathcal{C}_{n,\tau} = \left\{ f \in \mathcal{C} : \; H^2(f, f_0) = \frac{1}{2} \int_0^\infty \left(\sqrt{f(z)} - \sqrt{f_0(z)} \right)^2 dz \le \tau/n \right\}.$$

Consider estimation of

$$T_1(f) = \frac{f(x_0)}{1 - F(x_0)} = h(x_0), \qquad T_2(f) = h'(x_0). \tag{5.2}$$

Let $f_0 \in \mathcal{C}$ and $x_0 > 0$ be fixed such that h_0 is twice continuously differentiable at x_0. Define, for $\epsilon > 0$, the functions h_ϵ as follows:

$$h_\epsilon(z) = \begin{cases} h_0(x_0 - \epsilon c_\epsilon) + (z - x_0 + \epsilon c_\epsilon) h_0'(x_0 - \epsilon c_\epsilon), & z \in [x_0 - \epsilon c_\epsilon, x_0 - \epsilon], \\ h_0(x_0 + \epsilon) + (z - x_0 - \epsilon) h_0'(x_0 + \epsilon), & z \in [x_0 - \epsilon, x_0 + \epsilon], \\ h_0(z), & \text{otherwise.} \end{cases}$$

Here c_ϵ is chosen so that h_ϵ is continuous at $x_0 - \epsilon$. Using continuity of h_ϵ and a second order expansion of h_0 it follows that $c_\epsilon = 3 + o(1)$ as $\epsilon \to 0$. Now define f_ϵ by

$$f_\epsilon(z) = \exp(-H_\epsilon(z)) h_\epsilon(z)$$

where $H_\epsilon(z) \equiv \int_0^z h_\epsilon(u)du$. It follows easily that

$$T_1(f_\epsilon) - T_1(f_0) = \frac{1}{2} h_0''(x_0)\epsilon^2 + o(\epsilon^2), \tag{5.3}$$

$$T_2(f_\epsilon) - T_2(f_0) = h_0''(x_0)\epsilon + o(\epsilon). \tag{5.4}$$

Furthermore, the following lemma holds.

Lemma 5.1. *Under the above assumptions*

$$H^2(f_\epsilon, f_0) \;=\; \frac{2}{5} \frac{h_0''(x_0)^2 (1 - F(x_0))}{h_0(x_0)} \epsilon^5 + o(\epsilon^5) \equiv \nu_0 \epsilon^5 + o(\epsilon^5).$$

Proof. The lemma will follow from Lemma 3.2 of Jongbloed (1995) if we show that

$$\int \frac{(f_\epsilon(x) - f_0(x))^2}{f_0(x)} dx = \frac{16}{5} \frac{h_0''(x_0)^2(1 - F(x_0))}{h_0(x_0)} \epsilon^5 + o(\epsilon^5).$$

Thus we write

$$\begin{aligned}
&\int \frac{(f_\epsilon(x) - f_0(x))^2}{f_0(x)} dx \\
&\quad = \int \frac{(h_\epsilon(x) \exp(-H_\epsilon(x)) - h_0(x) \exp(-H_0(x)))^2}{h_0(x) \exp(-H_0(x))} dx \\
&\quad = \int \{(h_\epsilon(x) - h_0(x)) \exp(-H_\epsilon(x))) \\
&\qquad\quad + h_0(x)(\exp(-H_\epsilon(x)) - \exp(-H_0(x)))\}^2 \frac{1}{h_0(x) \exp(-H_0(x))} dx \qquad (5.5)
\end{aligned}$$

where

$$H_\epsilon(z) = \begin{cases} H_0(z), & z \le x_0 - \epsilon c_\epsilon, \\[1ex] H_0(x_0 - \epsilon c_\epsilon) + h_0(x_0 - \epsilon c_\epsilon)(z - x_0 + \epsilon c_\epsilon) \\ \quad + \frac{1}{2} h_0'(x_0 - \epsilon c_\epsilon)(z - x_0 + \epsilon c_\epsilon)^2, & z \in [x_0 - \epsilon c_\epsilon, x_0 - \epsilon], \\[1ex] H_0(x_0 - \epsilon c_\epsilon) + h_0(x_0 - \epsilon c_\epsilon)(c_\epsilon - 1)\epsilon \\ \quad + \frac{1}{2} h_0'(x_0 - \epsilon c_\epsilon)(c_\epsilon - 1)^2 \epsilon^2 \\ \quad + h_0(x_0 + \epsilon)(z - x_0 + \epsilon) \\ \quad + \frac{1}{2} h_0'(x_0 + \epsilon)[(z - x_0 - \epsilon)^2 - (2\epsilon)^2], & z \in [x_0 - \epsilon, x_0 + \epsilon], \\[1ex] H_0(x_0 - \epsilon c_\epsilon) + h_0(x_0 - \epsilon c_\epsilon)(c_\epsilon - 1)\epsilon \\ \quad + \frac{1}{2} h_0'(x_0 - \epsilon c_\epsilon)(c_\epsilon - 1)^2 \epsilon^2 \\ \quad + h_0(x_0 + \epsilon)(2\epsilon) + \frac{1}{2} h_0'(x_0 + \epsilon)[-(2\epsilon)^2] \\ \quad + H_0(z) - H_0(x_0 + \epsilon), & z \in [x_0 + \epsilon, \infty). \end{cases}$$

It follows that $\sup_z |H_\epsilon(z) - H_0(z)| = O(\epsilon^3)$, so the second term of (5.5) contributes a term of order $O(\epsilon^6)$. Calculations similar to those of Jongbloed (1995) (see also Jongbloed (2000)) and Groeneboom et al. (2001b)) complete the proof of the lemma. □

Combining (5.3) and (5.4) with the lemma, and writing $S_0(x) = 1 - F_0(x)$, it follows that

$$|T_1(f_{(\epsilon/\nu_0)^{1/5}}) - T_1(f_0)| \ge \left(\frac{h_0(x_0)\sqrt{h_0''(x_0)}}{S_0(x_0) 8\sqrt{2}} \right)^{2/5} \epsilon^{2/5}(1 + o(1)),$$

and

$$|T_2(f_{(\epsilon/\nu_0)^{1/5}}) - T_2(f_0)| \geq \left(\frac{5h_0(x_0)h_0''(x_0)^3}{2S_0(x_0)}\right)^{1/5} \epsilon^{1/5}(1+o(1)).$$

From these calculations together with Lemma 5.1 of Groeneboom et al. (2001b), we have the following theorem.

Theorem 5.2. *(Minimax risk lower bound). For the functionals T_1 and T_2 as defined in (5.2), and with $MMR_1(n, T, \mathcal{C}_{n,\tau})$ as defined in (5.1),*

$$\sup_{\tau>0} \limsup_{n\to\infty} n^{2/5} MMR_1(n, T_1, \mathcal{C}_{n,\tau}) \geq \frac{1}{4}\left(\frac{h_0(x_0)\sqrt{h_0''(x_0)}}{S_0(x_0)e8\sqrt{2}}\right)^{2/5}$$

and

$$\sup_{\tau>0} \limsup_{n\to\infty} n^{1/5} MMR_1(n, T_2, \mathcal{C}_{n,\tau}) \geq \frac{1}{4}\left(\frac{1}{4e}\frac{h_0(x_0)h_0''(x_0)^3}{2S_0(x_0)}\right)^{1/5}.$$

6. Rates of Convergence

This section contains several technical results which allow us to identify the *local* rates of convergence for our estimators. For the LSE, fix a point $x_0 \in (0, T)$ where $T < F_0^{-1}(1)$. For the MLE, fix a point $x_0 \in (\mathrm{supp} f_0)^\circ$. Throughout this section we assume that $h_0''(\cdot)$ is continuous and strictly positive in a neighborhood of x_0, and that $h(x_0) > 0$.

6.1. **U Function Estimates.** We begin by defining two key processes: for $0 < x \leq y$, define the functions

$$U_n^{lse}(x,y) = \int_x^y \left\{z - \frac{1}{2}(x+y)\right\} d\left(\mathbb{H}_n - H_0\right)(z),$$

and

$$U_n^{mle}(x,y) = \int_x^y \left\{\frac{z - \frac{1}{2}(x+y)}{\widehat{h}_n(z)}\right\} d\left(\mathbb{F}_n - F_0\right)(z).$$

Lemma 6.1. *Let $x_0 \in (0,T)$ and assume that $H_0(T) < \infty$. Then for each $\epsilon > 0$ there exist constants $\delta, c_0, n_0 > 0$ and (positive) random variables M_n of order $O_p(1)$ such that for each $|x - x_0| < \delta$*

$$|U_n^{lse}(x,y)| \leq \epsilon(y-x)^4 + n^{-4/5}M_n, \qquad 0 \leq y - x \leq c_0. \tag{6.1}$$

for all $n \geq n_0$. The same inequality holds for $U_n^{lse}(x,y)$ replaced with $U_n^{mle}(x,y)$.

Proof. The difficulty here in comparison to the proofs in Groeneboom, Jongbloed and Wellner (2001b) is that $\mathbb{H}_n(x) = \int_{[0,x]}(1-\mathbb{F}_n(s-))^{-1}d\mathbb{F}_n(s)$ is not a linear function of the empirical distribution $\mathbb{F}_n$, so extra work is involved in handling the random denominator.

We begin by writing

$$\begin{aligned} U_n^{lse}(x,y) &= \int_{[x,y]} \frac{f_{x,y}(z)}{\mathbb{S}_n(z-)} d(\mathbb{F}_n(z) - F_0(z)) \\ &\quad + \int_{[x,y]} \frac{[\mathbb{S}_n(z-) - S_0(z-)]}{\mathbb{S}_n(z-)S_0(z-)} f_{z,y}(z) dF_0(z) \\ &\equiv U_n^{(1)}(x,y) + U_n^{(2)}(x,y) \end{aligned}$$

where

$$f_{x,y}(z) \equiv (z-(x+y)/2)1_{[x,y]}(z) = (z-x)1_{[x,y]}(z) - \frac{1}{2}(y-x)1_{[x,y]}(z).$$

To handle $U_n^{(1)}$, choose δ and c_0 so that $\gamma \equiv S_0(x_0+\delta+c_0)/2 > 0$, and consider the class of functions

$$\mathcal{F}_{x,R}^{(1)} \equiv \left\{ \begin{array}{c} z \mapsto \frac{f_{x,y}(z)}{S(z-)} : \ x \le y \le x+R, \ S \text{ right continuous,} \\ \text{and nonincreasing with } \quad \|S - S_0\|_{x_0-\delta}^{x_0+\delta+c_0} \le \gamma \end{array} \right\}.$$

Then $\mathcal{F}_{x,R}^{(1)}$ has envelope function

$$F_{x,R}^{(1)}(z) = \frac{1}{\gamma}\left\{(z-x)1_{[x,x+R]}(z) + \frac{1}{2}R1_{[x,x+R]}(z)\right\}$$

with

$$E\left\{[F_{x,R}^{(1)}]^2\right\} = \frac{1}{\gamma^2}\int_{[x,x+R]}[(z-x)+R/2]^2 f_0(z)dz \le \frac{13}{12\gamma^2}\|f_0\|_{x_0-\delta}^{x_0+\delta}R^3.$$

Since $\log N_{[]}(\epsilon, \mathcal{F}_{x,R}^{(1)}, L_2(P)) \le K/\epsilon$ for some constant K by van der Vaart and Wellner (1996), Theorem 2.7.5, page 164, and a straightforward bracketing argument, it follows from van der Vaart and Wellner (1996), Theorems 2.14.2 and 2.14.5, pages 240 and 244, that

$$E\left\{\left(\sup_{f\in\mathcal{F}_{x,R}^{(1)}} |(\mathbb{P}_n - P_0)(f)|\right)^2\right\} \le \frac{1}{n}K'E\{[F_{x,R}^{(1)}(X_1)]^2\} = O(n^{-1}R^3). \tag{6.2}$$

To control $U_n^{(2)}$, note that it can be rewritten as

$$(\mathbb{P}_n - P_0)(g_{x,y,\mathbb{S}_n})$$

where

$$g_{x,y,S}(u) \equiv \int_{[x,y]} \frac{f_{x,y}(z)}{S(z-)S_0(z-)} 1_{[z,\infty)}(u) f_0(z)dz.$$

This leads to consideration of the class of functions

$$\mathcal{F}_{x,R}^{(2)} \equiv \left\{ \begin{array}{c} u \mapsto g_{x,y,S}(u) : \ x \le y \le x+R, \ S \text{ right continuous,} \\ \text{and nonincreasing with} \quad \|S - S_0\|_{x_0-\delta}^{x_0+\delta+c_0} \le \gamma \end{array} \right\}.$$

For this class of functions we can take the envelope function to be

$$F_{x,R}^{(2)}(u) = \frac{(\|f_0\|_{x_0-\delta}^{x_0+\delta})}{\gamma^2} \{R(u-x)\} 1_{[x,x+R]}(u),$$

with

$$E\{[F_{x,R}^{(2)}(X_1)]^2\} = \frac{(\|f_0\|_{x_0-\delta}^{x_0+\delta})^2}{2\gamma^4} R^5.$$

Since $\log N_{[\,]}(\epsilon, \mathcal{F}_{x,R}^{(2)}, L_2(P)) \le K/\epsilon$ for some constant K by van der Vaart and Wellner (1996), Theorem 2.7.5, page 159, and a straightforward bracketing argument, it then follows from van der Vaart and Wellner (1996), Theorems 2.14.2 and 2.14.5, pages 240 and 244, that

$$E\left\{\left(\sup_{f \in \mathcal{F}_{x,R}^{(2)}} |(\mathbb{P}_n - P_0)(f)|\right)^2\right\} \le \frac{1}{n} K E\{[F_{x,R}^{(1)}(X_1)]^2\} = O(n^{-1}R^5). \tag{6.3}$$

Let $G_n \equiv \{\|\mathbb{S}_n - S_0\|_{x_0-\delta}^{x_0+\delta+c_0} \le \gamma\}$, and note that

$$P(G_n^c) = P(\|\mathbb{S}_n - S_0\|_{x_0-\delta}^{x_0+\delta+c_0} > \gamma) \le P(\|\mathbb{S}_n - S_0\| > \gamma) \le 2e^{-2n\gamma^2}$$

by Massart's sharpening of the DKW inequality (Massart (1990)).

Now define $M_n(\omega)$ as the infimum (possibly $+\infty$) of those values such that (6.1) holds. Define $A(n,j)$ to be the set $\big[(j-1)n^{-1/5}, jn^{-1/5}\big)$. Then for m constant

$$\begin{aligned} & P(M_n > m) \\ & \quad \le \ P([M_n > m] \cap G_n) + P(G_n^c) \\ & \quad \le \ P\left([\exists u : |U_n^{lse}(x,x+u)| > \epsilon u^4 + n^{-4/5} m] \cap G_n\right) + 2e^{-2n\gamma^2} \\ & \quad \le \ \sum_{j \ge 1} P\left([\exists u \in A(n,j) : n^{4/5} \left|U_n^{lse}(x,x+u)\right| > \epsilon(j-1)^4 + m] \cap G_n\right) + 2e^{-2n\gamma^2} \end{aligned}$$

The jth summand is hence bounded by

$$\begin{aligned} & n^{8/5} E\left[\sup_{u \in A(n,j)} |U_n^{lse}(x,x+u)|^2 1_{G_n}\right] / \left[m + \epsilon(j-1)^4\right]^2 \\ & \quad \le \ C_1 \frac{j^3}{[m \ + \epsilon(j-1)^4]^2} + C_2 n^{-2/5} \frac{j^5}{[m \ + \epsilon(j-1)^4]^2} \end{aligned}$$

due to (6.2) and (6.3). Thus it follows that

$$\limsup_{n\to\infty} P(M_n > m) \le C_1 \sum_{j=1}^{\infty} \frac{j^3}{[m \ + \epsilon(j-1)^4]^2}$$

where the sum in the bound is finite and converges to zero as $m \to \infty$. This completes the proof of the claim for U_n^{lse}.

The same type of proof with appropriate modifications works for U_n^{mle}: note that

$$U_n^{mle} = (\mathbb{P}_n - P_0)(g_{x,y,\widehat{h}_n})$$

where

$$g_{x,y,h}(z) \equiv \frac{f_{x,y}(z)}{h(z)} 1_{[x,y]}(z),$$

and, in view of the consistency established in Theorem 4.1, $\widehat{h}_n$ is a convex function uniformly close to h_0 on neighborhoods of x_0. This leads to consideration of the class of functions

$$\mathcal{F}_{x,R} \equiv \left\{ \begin{array}{c} z \mapsto g_{x,y,h}(z): \ x \le y \le x+R, \ h \text{ convex}, \\ \|h - h_0\|_{x_0-\delta}^{x_0+\delta+c_0} \le \gamma \end{array} \right\}.$$

with $\gamma \equiv \inf_{x_0-\delta \le x \le x_0+\delta+c_0} h_0(x)/2$, and where we now take $G_n \equiv \{\|\widehat{h}_n - h_0\|_{x_0-\delta}^{x_0+\delta+c_0} \le \gamma\}$. The class $\mathcal{F}_{x,R}$ has an envelope function of the same form as the envelope $F_{x,R}^{(1)}$ in (6.1) with the new definition of γ, and hence the same second moment bound holds:

$$E\left\{[F_{x,R}]^2\right\} = \frac{1}{\gamma^2} \int_{[x,x+R]} [(z-x) + R/2]^2 f_0(z)dz \le \frac{13}{12\gamma^2} \|f_0\|_{x_0-\delta}^{x_0+\delta} R^3.$$

Furthermore, $\log N_{[]}(\epsilon, \mathcal{F}_{x,R}, L_2(P_0)) \le K/\epsilon^{1/2}$ for some constant K by van der Vaart and Wellner (1996), Theorem 2.7.10, page 159, and a straightforward bracketing argument. It then follows from van der Vaart and Wellner (1996), Theorems 2.14.2 and 2.14.5, pages 240 and 244, that

$$E\left\{ \left(\sup_{f \in \mathcal{F}_{x,R}} |(\mathbb{P}_n - P_0)(f)| \right)^2 \right\} \le \frac{1}{n} K' E\{[F_{x,R}(X_1)]^2\} = O(n^{-1}R^3). \tag{6.4}$$

The remainder of the argument is the same as for the LSE. □

Define

$$V_n(x,y) = \int_x^y \left\{ z - \frac{x+y}{2} \right\} (\mathbb{S}_n(z) - S_0(z))dz.$$

Lemma 6.2. *Let $x_0 \in (suppF_0)^\circ$. Then for each $\epsilon > 0$ there exist constants $\delta, c_0 > 0$ and (positive) random variables M_n of order $O_p(1)$ such that for each $|x - x_0| < \delta$*

$$|V_n(x,y)| \le \epsilon n^{-1/5}(y-x)^4 + n^{-1}M_n, \qquad 0 \le y - x \le c_0. \tag{6.5}$$

Proof. This follows from the same argument used for $U_n^{(2)}(x,y)$ in the proof of Lemma 6.1, but now much more easily since there is no troublesome denominator term and no density term to complicate the calculation. In this case we compute

$$V_n(x,y) = (\mathbb{P}_n - P_0)(g_{x,y})$$

where

$$g_{x,y}(z) = \frac{1}{2}(z-x)(z-y)1_{[x,y]}(z) = \frac{1}{2}\left\{(z-x)^2 - (y-x)(z-x)\right\}1_{[x,y]}(z).$$

Thus we consider the class of functions

$$\mathcal{G}_{x,R} \equiv \{z \mapsto g_{x,y}(z): \ x \le y \le x+R\},$$

a VC-subgraph class with envelope function

$$G_{x,R}(z) = \frac{1}{2}(z-x)^2 + \frac{R}{2}(z-x)1_{[x,x+R]}(z),$$

so that

$$E\{[G_{x,R}^2(X_1)]^2\} \le C\|f_0\|_{x_0-\delta}^{x_0+\delta}R^5.$$

Thus by van der Vaart and Wellner (1996), Theorem 2.14.1, page 239,

$$E\left\{\left(\sup_{f\in\mathcal{G}_{x,R}}|(\mathbb{P}_n - P_0)(f)|\right)^2\right\} \le \frac{1}{n}KE\{[G_{x,R}(X_1)]^2\} = O(n^{-1}R^5). \tag{6.6}$$

Now let $M_n(\omega)$ be the infimum of those values (possibly infinity) such that (6.5) holds. Then for $m > 0$,

$$\begin{aligned}
P(M_n > m) &\le P\left(\exists u : |V_n(x,x+u)| > \epsilon n^{-1/5}u^4 + n^{-1}m\right) \\
&\le \sum_{j\ge 1} P\left(\exists u \in A(n,j) : n\,|V_n(x,x+u)| > \epsilon(j-1)^4 + m\right) \\
&\le \sum_{j\ge 1} n^2 E\left\{\sup_{u\in A(n,j)}|V_n(x,x+u)|^2\right\}\Big/[m+\epsilon(j-1)^4]^2 \\
&\le C\sum_{j\ge 1}\frac{j^5}{[m+\epsilon(j-1)^4]^2}
\end{aligned}$$

where the right side converges to 0 as $m \to \infty$. □

6.2. **Midpoint Properties.**

Lemma 6.3. *Let $x_0 > 0$ be a point at which h_0 has a continuous and strictly positive second derivative, and $h(x_0) > 0$. Let ξ_n be any sequence of numbers converging to x_0 and define τ_n and η_n to be the largest touchpoint of $\widehat{h}_n$ smaller than ξ_n and the smallest touchpoint larger than ξ_n respectively. Then*

$$\eta_n - \tau_n = O_p(n^{-1/5})$$

for the MLE of the hazard rate. The same conclusion holds for the LS estimator (with $x_0 < T$).

Proof. Define m_n to be the midpoint of $[\tau_n, \eta_n]$, $m_n = (\tau_n + \eta_n)/2$.

We first consider the MLE, as this is the more difficult of the two cases. By Theorem 4.1, we know that $\widehat{h}_n$ is finite and positive near x_0 for large enough n. Also, it is either strictly increasing or strictly decreasing in a neighborhood of x_0 (as in Figure 6.2 (b)), or is locally flat as the picture in Figure 6.2 (a).

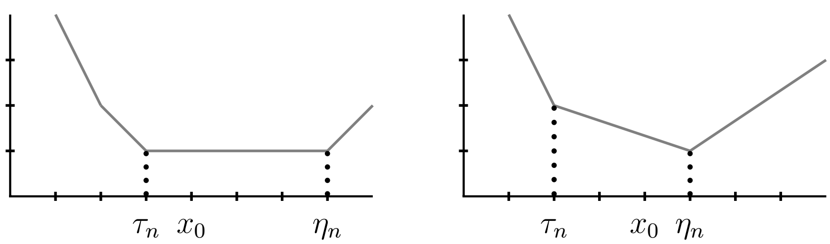


Figure 6.2 (a) Figure 6.2 (b)

If $\widehat{h}_n$ is decreasing between τ_n and η_n, then (3.15) and (2.2) with equality at both η_n, τ_n hold. If $\widehat{h}_n$ is increasing instead, then (3.16) and (2.3) with equality at both η_n, τ_n hold. There is only the potential for a problem in the situation shown in Figure 6.2 (a). However, since $\widehat{h}_n$ is strictly positive, by Corollary 3.4 we can extend the necessary equalities to this case as well. Therefore, we may only consider two cases, either $\widehat{h}_n$ is nonincreasing or nondecreasing on $[\tau_n, \eta_n]$.

We first assume that $\widehat{h}_n$ is nonincreasing on $[\tau_n, \eta_n]$. Define

$$\widehat{\mathcal{H}}_{n,\downarrow}(z) = \int_0^z \frac{z-t}{\widehat{h}_n(t)} d\mathbb{F}_n(t) \tag{6.7}$$

$$\mathbb{A}_{n,\downarrow}(z) = \int_0^z \mathbb{S}_n(t)dt. \tag{6.8}$$

We may then calculate

$$\widehat{\mathcal{H}}_{n,\downarrow}(m_n) = \int_{m_n}^{\eta_n} \frac{x-m_n}{\widehat{h}_n(x)} d\mathbb{F}_n(x) + \widehat{\mathcal{H}}_{n,\downarrow}(\eta_n) - (\eta_n - m_n)\widehat{\mathcal{H}}'_{n,\downarrow}(\eta_n)$$

and

$$\widehat{\mathcal{H}}_{n,\downarrow}(m_n) = \int_{\tau_n}^{m_n} \frac{m_n - x}{\widehat{h}_n(x)} d\mathbb{F}_n(x) + \widehat{\mathcal{H}}_{n,\downarrow}(\tau_n) + (m_n - \tau_n)\widehat{\mathcal{H}}'_{n,\downarrow}(\tau_n).$$

From (2.2) we know that $2\widehat{\mathcal{H}}_{n,\downarrow}(m_n) \leq 2\int_0^{m_n} \mathbb{A}_{n,\downarrow}(t)dt$. Using the equality in (2.2) and (3.15) allow us to rewrite this as $0 \geq L_{1,\downarrow} + L_{2,\downarrow}$, where $L_{1,\downarrow}$ is equal to

$$\begin{aligned} & \int_{m_n}^{\eta_n} \frac{x - m_n}{\widehat{h}_n(x)} d\mathbb{F}_n(x) + \int_{\tau_n}^{m_n} \frac{m_n - x}{\widehat{h}_n(x)} d\mathbb{F}_n(x) - \frac{\eta_n - \tau_n}{4}\left\{\widehat{\mathcal{H}}'_{n,\downarrow}(\eta_n) - \widehat{\mathcal{H}}'_{n,\downarrow}(\tau_n)\right\} \\ = & \int_{m_n}^{\eta_n} \frac{x - \frac{1}{2}(\eta_n + m_n)}{\widehat{h}_n(x)} d\mathbb{F}_n(x) + \int_{\tau_n}^{m_n} \frac{\frac{1}{2}(\tau_n + m_n) - x}{\widehat{h}_n(x)} d\mathbb{F}_n(x). \end{aligned}$$

and

$$\begin{aligned} L_{2,\downarrow} &= \int_{m_n}^{\eta_n} \mathbb{A}_{n,\downarrow}(x)dx - \int_{\tau_n}^{m_n} \mathbb{A}_{n,\downarrow}(x)dx - \frac{1}{4}(\eta_n - \tau_n)\left\{\mathbb{A}_{n,\downarrow}(\eta_n) - \mathbb{A}_{n,\downarrow}(\tau_n)\right\} \\ &= -\left\{\int_{m_n}^{\eta_n} \left\{x - \frac{1}{2}(\eta_n + m_n)\right\}\mathbb{S}_n(x)dx + \int_{\tau_n}^{m_n}\left\{\frac{1}{2}(\tau_n + m_n) - x\right\}\mathbb{S}_n(x)dx\right\}, \end{aligned}$$

by integration by parts.

Now suppose that $\widehat{h}_n$ is nondecreasing on $[\tau_n, \eta_n]$. Define

$$\widehat{\mathcal{H}}_{n,\uparrow}(z) = \int_z^\infty \frac{t - z}{\widehat{h}_n(t)} d\mathbb{F}_n(t) \tag{6.9}$$

$$\mathbb{A}_{n,\uparrow}(z) = \int_z^\infty \mathbb{S}_n(t)dt. \tag{6.10}$$

We again calculate, using the equality of (2.3) and (3.16),

$$\widehat{\mathcal{H}}_{n,\uparrow}(m_n) = \int_{m_n}^{\eta_n} \frac{x - m_n}{\widehat{h}_n(x)} d\mathbb{F}_n(x) + \widehat{\mathcal{H}}_{n,\uparrow}(\eta_n) + (\eta_n - m_n)\widehat{\mathcal{H}}'_{n,\uparrow}(\eta_n)$$

and

$$\widehat{\mathcal{H}}_{n,\uparrow}(m_n) = \int_{\tau_n}^{m_n} \frac{m_n - x}{\widehat{h}_n(x)} d\mathbb{F}_n(x) + \widehat{\mathcal{H}}_{n,\uparrow}(\tau_n) - (m_n - \tau_n)\widehat{\mathcal{H}}'_{n,\uparrow}(\tau_n).$$

From (2.3) we know that $2\widehat{\mathcal{H}}_{n,\uparrow}(m_n) \leq 2\int_{m_n}^\infty \mathbb{A}_{n,\uparrow}(t)dt$. Using the equality of (2.3) and (3.16) we rewrite this as $0 \geq L_{1,\uparrow} + L_{2,\uparrow}$, where $L_{1,\uparrow} = L_{1,\downarrow}$ and $L_{2,\uparrow}$ is calculated to be

$$-\left\{\int_{\tau_n}^{\eta_n} \mathbb{A}_{n,\uparrow}(x)dx - 2\int_{\tau_n}^{m_n} \mathbb{A}_{n,\uparrow}(x)dx - \frac{1}{4}(\eta_n - \tau_n)\left\{\mathbb{A}_{n,\uparrow}(\eta_n) - \mathbb{A}_{n,\uparrow}(\tau_n)\right\}\right\}.$$

By integration by parts, this is equal to $L_{2,\downarrow}$. Therefore the two cases both satisfy the inequality $0 \geq L_{1,\downarrow} + L_{2,\downarrow}$.

Now replace $\mathbb{F}_n$ by the true F_0 in the definition of $L_{1,\downarrow}$ to obtain

$$\begin{aligned}
L^0_{1,\downarrow} &\equiv \int_{m_n}^{\eta_n} \frac{x-\frac{1}{2}(\eta_n+m_n)}{\widehat{h}_n(x)} dF_0(x) + \int_{\tau_n}^{m_n} \frac{\frac{1}{2}(\tau_n+m_n)-x}{\widehat{h}_n(x)} dF_0(x) \\
&= \int_{m_n}^{\eta_n} \left\{x-\frac{1}{2}(\eta_n+m_n)\right\}\left\{\frac{1}{\widehat{h}_n(x)}-\frac{1}{h_0(x)}\right\} dF_0(x) \\
&\quad + \int_{\tau_n}^{m_n} \left\{\frac{1}{2}(\tau_n+m_n)-x\right\}\left\{\frac{1}{\widehat{h}_n(x)}-\frac{1}{h_0(x)}\right\} dF_0(x) - L^0_{2,\downarrow},
\end{aligned}$$

where

$$L^0_{2,\downarrow} = - \int_{m_n}^{\eta_n} \left\{x-\frac{1}{2}(\eta_n+m_n)\right\} S_0(x)dx - \int_{\tau_n}^{m_n} \left\{\frac{1}{2}(\tau_n+m_n)-x\right\} S_0(x)dx.$$

Using a Taylor expansion of order 2 about the point m_n, we next get that

$$\begin{aligned}
L^0_{1,\downarrow}+L^0_{2,\downarrow} &= \int_{m_n}^{\eta_n} \left\{x-\frac{1}{2}(\eta_n+m_n)\right\}\left\{\frac{1}{\widehat{h}_n(x)}-\frac{1}{h_0(x)}\right\} f_0(x)dx \\
&\quad + \int_{\tau_n}^{m_n} \left\{\frac{1}{2}(\tau_n+m_n)-x\right\}\left\{\frac{1}{\widehat{h}_n(x)}-\frac{1}{h_0(x)}\right\} f_0(x)dx \\
&= \frac{1}{192}\left\{\left(\frac{1}{\widehat{h}_n(\cdot)}-\frac{1}{h_0(\cdot)}\right) f_0(\cdot)\right\}''(x_0)(\eta_n-\tau_n)^4 + o((\eta_n-\tau_n)^4) \\
&= \frac{1}{192}\left\{\frac{h_0''(x_0)}{h_0^2(x_0)} f_0(x_0)\right\}(\eta_n-\tau_n)^4 + o((\eta_n-\tau_n)^4),
\end{aligned}$$

since both $\widehat{h}_n$ and $\widehat{h}_n'$ are consistent by Theorem 4.1, $\widehat{h}_n''(x) = 0$ on (τ_n, η_n), and because $\tau_n - \eta_n = o_p(1)$ by Corollary 4.3.

Therefore, by Lemmas 6.1 and 6.2 and the above calculations, we may write

$$\begin{aligned}
0 &\geq L_{1,\downarrow}+L_{2,\downarrow} \\
&= L^0_{1,\downarrow}+L^0_{2,\downarrow}+(L_{1,\downarrow}-L^0_{1,\downarrow})+(L_{2,\downarrow}-L^0_{2,\downarrow}) \\
&\geq L^0_{1,\downarrow}+L^0_{2,\downarrow}-\epsilon(\eta_n-\tau_n)^4 - O_p(n^{-4/5}) - \epsilon n^{-1/5}(\eta_n-\tau_n)^4 - O_p(n^{-1}) \\
&= \frac{1}{192}\left\{\frac{h_0''(x_0)}{h_0^2(x_0)} f(x_0) - 192\epsilon\right\}(\eta_n-\tau_n)^4 + o\left((\eta_n-\tau_n)^4\right) - O_p(n^{-4/5}).
\end{aligned}$$

We choose ϵ sufficiently small (so that the leading term in the last line of the above display is positive), and hence conclude that $(\eta_n - \tau_n) = O_p(n^{-1/5})$.

Now for the rate in the case of the LSE. As before, let $\mathbb{H}_n$ denote the empirical hazard function, $\widetilde{H}_n(t) = \int_0^t \widetilde{h}_n(s)ds$, and $\widetilde{\mathcal{H}}_n(t) = \int_0^t \widetilde{H}_n(s)ds$. It follows from Lemma

2.3 that

$$\widetilde{\mathcal{H}}_n(m_n) \geq \mathbb{Y}_n(m_n),$$

where $\mathbb{Y}_n(t) = \int_0^t \mathbb{H}_n(s)ds$. Using the equality in (2.9) and Corollary 2.4 we calculate

$$\widetilde{\mathcal{H}}_n(m_n) = \frac{1}{2}\{\mathbb{Y}_n(\tau_n) + \mathbb{Y}_n(\eta_n)\} - \frac{1}{8}\{\mathbb{H}_n(\eta_n) - \mathbb{H}_n(\tau_n)\}(\eta_n - \tau_n).$$

Let $Y_0(t) = \int_0^t H_0(s)ds$. Then

$$\begin{aligned}
&2\widetilde{\mathcal{H}}_0(m_n) - 2Y_0(m_n) \\
&= \int_{\tau_n}^{m_n} \left\{x - \frac{1}{2}(\tau_n + m_n)\right\} h_0(x)dx + \int_{m_n}^{\eta_n} \left\{\frac{1}{2}(\tau_n + m_n) - x\right\} h_0(x)dx \\
&= -\frac{1}{192} h_0''(m_n)(\eta_n - \tau_n)^4 + o(\eta_n - \tau_n)^4 \\
&= -\frac{1}{192} h_0''(x_0)(\eta_n - \tau_n)^4 + o(\eta_n - \tau_n)^4,
\end{aligned}$$

using Corollary 4.6 in the last line.

From Lemma 6.1 it therefore follows that

$$\begin{aligned}
0 &\leq 2\widetilde{\mathcal{H}}_n(m_n) - 2\mathbb{Y}_n(m_n) \\
&= 2\widetilde{\mathcal{H}}_0(m_n) - 2Y_0(m_n) + \int_{\tau_n}^{m_n} \left\{x - \frac{1}{2}(\tau_n + m_n)\right\} d\{\mathbb{H}_n - H_0\}(x) \\
&\qquad + \int_{m_n}^{\eta_n} \left\{\frac{1}{2}(\tau_n + m_n) - x\right\} d\{\mathbb{H}_n - H_0\}(x) \\
&= -\left(\frac{1}{192} h_0''(x_0) - \epsilon\right)(\eta_n - \tau_n)^4 + o(\eta_n - \tau_n)^4 + O_p(n^{-4/5})
\end{aligned}$$

Choosing ϵ sufficiently small proves the result. □

6.3. **Some heuristics.** In this section we briefly describe some heuristics which hopefully shed light on the use of the function $U_n^{\#}(x, y)$ for # =MLE or LSE. We do this only for the LSE case. Let $\tau_n \leq \eta_n$ be two touch points. Since $\widetilde{h}_n(t)$ is piecewise linear, and in particular linear between the touch points, we note that $\widetilde{\mathcal{H}}_n(t)$ is a cubic spline. Thus, using the two conditions from the equality in (2.7) and (2.4) evaluated at τ_n and η_n we may calculate $\widetilde{H}_n(t)$ explicitly. This gives

$$(\widetilde{H}_n)'''(t) = \frac{12}{(\eta_n - \tau_n)^3} \int_{\tau_n}^{\eta_n} \{t - m_n\} \, d\mathbb{H}_n(t),$$

for $t \in [\tau_n, \eta_n]$. Also, expanding the function h_0 around $m_n = (\eta_n + \tau_n)/2$ to second order shows that

$$\int_{\tau_n}^{\eta_n} \{t - m_n\}\, dH_0(t) \;=\; h_0'(m_n)\frac{1}{12}(\eta_n - \tau_n)^3 + O(\eta_n - \tau_n)^4.$$

This yields that for $t \in [\tau_n, \eta_n]$

$$\widetilde{h}_n'(t) - h_0'(t) \;=\; \frac{12}{(\eta_n - \tau_n)^3} U_n^{lse}(\tau_n, \eta_n) + O(\eta_n - \tau_n).$$

Since $\eta_n - \tau_n = O_p(n^{-1/5})$, we apply Lemma 6.1 and, if we could also assume that $(\eta_n - \tau_n)^{-1} = O_p(n^{-1/5})$, it would follow that

$$\begin{aligned} |\widetilde{h}_n'(t) - h_0'(t)| &\leq \frac{12}{(\eta_n - \tau_n)^3}\left\{\epsilon(\eta_n - \tau_n)^4 + M_n n^{-4/5}\right\} + O_p(n^{-1/5}) \\ &\approx 12\epsilon(\eta_n - \tau_n) + O_p(n^{-1/5}) = O_p(n^{-1/5}), \end{aligned}$$

for $t \in [\tau_n, \eta_n]$, which gives us the correct rates of convergence for our estimator. In the next section, we make these ideas rigorous.

6.4. **From Here to Tightness.**

Lemma 6.4. *Let ξ_n be a sequence converging to x_0. Then for any $\epsilon > 0$ there exists and $M > 1$ and a $c > 0$ such that, with probability greater than $1 - \epsilon$ we have that there exist change points $\tau_n < \xi_n < \eta_n$ of $\widehat{h}_n$ such that*

$$\inf_{t \in [\tau_n, \eta_n]} |\widehat{h}_n(t) - h_0(t)| < cn^{-2/5}$$

for all n sufficiently large. The same statement holds for $\widetilde{h}_n$.

Proof. Fix $\epsilon > 0$. From Lemma 6.3 it follows that there exist touchpoints η_n and τ_n and an $M > 1$ such that $\xi_n - Mn^{-1/5} \leq \tau_n \leq \xi_n - n^{-1/5} \leq \xi_n + n^{-1/5} \leq \eta_n \leq \xi_n + Mn^{-1/5}$.

Fix $c > 0$ and consider the event

$$\inf_{t \in [\tau_n, \eta_n]} |\widehat{h}_n(t) - h_0(t)| \geq cn^{-2/5}. \tag{6.11}$$

First, assume that $\widehat{h}_n$ is nonincreasing on $[\tau_n, \eta_n]$. On this set, we have that

$$\left| \int_{\tau_n}^{\eta_n} (\eta_n - t)\frac{\widehat{h}_n(t) - h_0(t)}{\widehat{h}_n(t)} S_0(t)dt \right| \geq Bcn^{-2/5}(\eta_n - \tau_n)^2 \geq Bcn^{-4/5},$$

where B is some constant depending on x_0. Using the definitions (6.7) and (6.8), as well as the equality in Condition (2.2) with (3.15), it follows that

$$\begin{aligned} 0 &= \widehat{\mathcal{H}}_{n,\downarrow}(\eta_n) - \int_0^{\eta_n} \mathbb{A}_{n,\downarrow}(t)dt - \widehat{\mathcal{H}}_{n,\downarrow}(\tau_n) + \int_0^{\tau_n} \mathbb{A}_{n,\downarrow}(t)dt \\ &\quad - (\widehat{\mathcal{H}}'_{n,\downarrow}(\tau_n) - \mathbb{A}_{n,\downarrow}(\tau_n))(\eta_n - \tau_n) \\ &= \int_{\tau_n}^{\eta_n} \frac{\eta_n - t}{\widehat{h}_n(t)} d\mathbb{F}_n(t) - \int_{\tau_n}^{\eta_n} (\eta_n - t)\mathbb{S}_n(t)dt \\ &= \int_{\tau_n}^{\eta_n} (\eta_n - t)\frac{\widehat{h}_n(t) - h_0(t)}{\widehat{h}_n(t)} S_0(t)dt + \int_{\tau_n}^{\eta_n} \frac{\eta_n - t}{\widehat{h}_n(t)} d\breve{\mathbb{F}}_n(t) - \int_{\tau_n}^{\eta_n} (\eta_n - t)\breve{\mathbb{S}}_n(t)dt, \end{aligned}$$

where $\breve{\mathbb{F}}_n(t) = \mathbb{F}_n(t) - F_0(t)$ and $\breve{\mathbb{S}}_n(t) = \mathbb{S}_n(t) - S_0(t)$. By the assumption on h_0 and x_0 and arguments similar to those of Lemmas 6.1 and 6.2, it follows that

$$\int_{\tau_n}^{\eta_n} (\eta_n - t)\frac{\widehat{h}_n(t) - h_0(t)}{\widehat{h}_n(t)} S_0(t)dt = O_p(n^{-4/5}),$$

which is a contradiction to (6.11) if c is chosen large enough.

Next, suppose that $\widehat{h}_n$ is nondecreasing on $[\tau_n, \eta_n]$. Using the definitions (6.10) and (6.9), as well as the equality in Condition (2.3) with (3.16), it follows that

$$\begin{aligned} 0 &= \widehat{\mathcal{H}}_{n,\uparrow}(\eta_n) - \int_{\eta_n}^{\infty} \mathbb{A}_{n,\uparrow}(t)dt - \widehat{\mathcal{H}}_{n,\uparrow}(\tau_n) \\ &\quad + \int_{\tau_n}^{\infty} \mathbb{A}_{n,\uparrow}(t)dt + (\widehat{\mathcal{H}}'_{n,\uparrow}(\eta_n) - \mathbb{A}_{n,\uparrow}(\eta_n))(\eta_n - \tau_n) \\ &= \int_{\tau_n}^{\eta_n} \frac{\tau_n - t}{\widehat{h}_n(t)} d\mathbb{F}_n(t) - \int_{\tau_n}^{\eta_n} (\tau_n - t)\mathbb{S}_n(t)dt. \end{aligned}$$

The same argument as above now proves the result.

Now for the LSE. We begin by assuming that (6.11) holds with $\widehat{h}_n$ replaced with $\widetilde{h}_n$, by way of contradiction. Using the equality of Condition (2.7) and Corollary 2.4, we calculate

$$\begin{aligned} 0 &= \int_{\tau_n}^{\eta_n} (\eta_n - t)d(\widetilde{H}_n - \mathbb{H}_n)(t) \\ &= \int_{\tau_n}^{\eta_n} (\eta_n - t)(\widetilde{h}_n(t) - h_0(t))dt - \int_{\tau_n}^{\eta_n} (\eta_n - t)d\{\mathbb{H}_n - H_0\}(t) \end{aligned}$$

implying as above that

$$\int_{\tau_n}^{\eta_n} (\eta_n - t)(\widetilde{h}_n(t) - h_0(t))dt = O_p(n^{-4/5}),$$

by a similar argument to Lemma 6.1, which contradicts the assumption, and the result follows.

□

The next result is the actual statement of tightness in this setting. The results follows from the previous lemmas, and make extensive use of the underlying convexity.

Proposition 6.5. *Under the assumptions of this section, we have that for each* $M > 0$

$$\sup_{|t|\leq M} |\widehat{h}_n(x_0 + n^{-1/5}t) - h_0(x_0) - n^{-1/5}th_0'(x_0)| = O_p(n^{-2/5}) \tag{6.12}$$

and

$$\sup_{|t|\leq M} |\widehat{h}_n'(x_0 + n^{-1/5}t) - h_0'(x_0)| = O_p(n^{-1/5}). \tag{6.13}$$

The same statement holds for $\widehat{h}_n$ *replaced with* $\widetilde{h}_n$.

Proof. The proof of this result is exactly the same in both cases, as it depends only on Lemmas 6.3 and 6.4, and the convexity of both the estimators and the true function. We therefore write the result only for $\widehat{h}_n$.

We begin with the proof of (6.13). Fix $M > 0$ and $\epsilon > 0$. Define $\eta_{n,1}$ to be the first point of touch after $x_0 + Mn^{-1/5}$, $\eta_{n,2}$ to be the first point of touch after $\eta_{n,1} + n^{-1/5}$, and $\eta_{n,3}$ to be the first point of touch after $\eta_{n,2} + n^{-1/5}$. Define the points $\tau_{n,i}$ for $i = 1, 2, 3$ similarly, but working to the left of x_0. That is, $\tau_{n,1}$ is the first touch point smaller than $x_0 - Mn^{-1/5}$ and so forth. By Lemma 6.4, there exist points $\xi_{n,i} \in (\eta_{n,i}, \eta_{n,i+1})$ and $\zeta_{n,i} \in (\tau_{n,i}, \tau_{n,i+1})$ for $i = 1, 2$, and a constant $c > 0$, such that with probability at least $1 - \epsilon$ we have that

$$|\widehat{h}_n(\xi_{n,i}) - h_0(\xi_{n,i})| \leq cn^{-2/5},$$

and similarly with $\xi_{n,i}$ replaced with $\zeta_{n,i}$ for $i = 1, 2$.

In what follows, if $\widehat{h}_n'(t)$ does not exist, we take the right derivative at t. From the convexity of $\widehat{h}_n$, it follows that for any $t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]$

$$\begin{aligned} \widehat{h}_n'(t) \leq \widehat{h}_n'(\xi_{n,1}) &\leq \frac{\widehat{h}_n(\xi_{n,2}) - \widehat{h}_n(\xi_{n,1})}{\xi_{n,2} - \xi_{n,1}} \\ &\leq \frac{h_0(\xi_{n,2}) - h_0(\xi_{n,1}) + 2cn^{-2/5}}{\xi_{n,2} - \xi_{n,1}} \\ &\leq h_0'(\xi_{n,2}) + 2cn^{-1/5}, \end{aligned}$$

since $\xi_{n,2} - \xi_{n,1} \geq n^{-1/5}$. Because of the continuity of $h_0''(\cdot)$ near x_0 we may replace $h_0'(\xi_{n,2})$ in the above display with $h_0'(x_0) + \tilde{c}n^{-1/5}$, for some new constant $\tilde{c}$. The result follows. A similar argument shows the lower bound.

Now for (6.12). By Lemma 6.3, there exists a constant $K > M$ such that there exist two touch points in $[x_0 + Mn^{-1/5}, x_0 + Kn^{-1/5}]$, $n^{-1/5}$ apart with probability $1 - \epsilon$. The same occurs in the interval $[x_0 - Mn^{-1/5}, x_0 - Kn^{-1/5}]$. From Lemma 6.4, it follows that there exists points $\xi_n \in [x_0 + Mn^{-1/5}, x_0 + Kn^{-1/5}]$ and $\zeta_n \in [x_0 - Mn^{-1/5}, x_0 - Kn^{-1/5}]$ such that

$$|\widehat{h}_n(\xi_n) - h_0(\xi_n)| \le cn^{-2/5},$$

and also for ζ_n with probability at least $1-\epsilon$ and sufficiently large n. Lastly, we have already shown that there exists a c' such that with probability at least $1-\epsilon$

$$\sup_{t\in[x_0-Kn^{-1/5},x_0+Kn^{-1/5}]} |\widehat{h}_n'(t) - h_0'(x_0)| \le c'n^{-1/5}.$$

Therefore, with probability at least $1-3\epsilon$, we have that for any $t \in [x_0 - Mn^{-1/5}, x_0 + Mn^{-1/5}]$ and sufficiently large n

$$\begin{aligned}
\widehat{h}_n(t) &\ge \widehat{h}_n(\xi_n) + \widehat{h}_n'(\xi_n)(t-\xi_n) \\
&\ge h_0(\xi_n) - cn^{-2/5} + (h_0'(x_0) - c'n^{-1/5})(t-\xi_n) \\
&= h_0(x_0) + h'(x_0)(\xi_n - x_0) + \frac{1}{2}h''(x_0^*)(\xi_n - x_0)^2 + (h_0'(x_0) - c'n^{-1/5})(t-\xi_n) - cn^{-2/5} \\
&= h_0(x_0) + h'(x_0)(t - x_0) + \frac{1}{2}h''(x_0^*)(\xi_n - x_0)^2 - c'n^{-1/5}(t-\xi_n) - cn^{-2/5} \\
&\ge h_0(x_0) + h'(x_0)(t-x_0) - Bn^{-2/5}.
\end{aligned}$$

for some constant $B > 0$.

For the upper bound, we have that

$$\begin{aligned}
\widehat{h}_n(t) &\le \widehat{h}_n(\zeta_n) + \frac{\widehat{h}_n(\xi_n) - \widehat{h}_n(\zeta_n)}{\xi_n - \zeta_n}(t-\zeta_n) \\
&\le h_0(\zeta_n) + cn^{-2/5} + \frac{h_0(\xi_n) - h_0(\zeta_n)}{\xi_n - \zeta_n}(t-\zeta_n) + 2cn^{-1/5}(t-\zeta_n) \\
&= h_0(x_0) + h_0'(x_0)(\zeta_n - x_0) + \frac{1}{2}h_0''(x_0^{**})(\zeta_n - x_0)^2 \\
&\qquad + h_0'(x_0)(t-\zeta_n) + 3cn^{-2/5} + \left\{ \frac{h_0(\xi_n) - h_0(\zeta_n)}{\xi_n - \zeta_n} - h_0'(x_0) \right\}(t-\zeta_n) \\
&\le h_0(x_0) + h_0'(x_0)(t-x_0) + B'n^{-2/5},
\end{aligned}$$

for some $B' > 0$, by the smoothness properties of h_0 near x_0. This proves the result. □

7. Limit distribution theory for the estimators at a fixed point

This section is dedicated to the analysis of the limiting distribution of the estimators at a fixed point x_0, and hence the proof of Theorem 2.7. As in the previous section, we assume here that $h_0''(\cdot)$ is continuous and strictly positive in a neighborhood of x_0, and that $h_0(x_0) > 0$.

The asymptotics are described by the invelope $\mathcal{I}(\cdot)$ of the "driving" process $Y(\cdot)$. Our goal will then be to identify the two processes, one which will converge to the invelope, and another which converges to the driving process Y in the limit problem.

The proof is split into two cases corresponding to the LSE and the MLE respectively. We handle the LSE first, since it is easier.

For any interval $[a,b] \subset \mathbb{R}$, let $D[a,b]$ denote the space of cadlag functions from $[a,b]$ into $\mathbb{R}$ endowed with the Skorohod topology. Similarly, $C[a,b]$ denotes the space of continuous functions endowed with the uniform topology.

7.1. **Proof of Theorem 2.7 for the LSE.**

Driving process for the LSE. Define

$$\mathbb{B}_n(t) \equiv \sqrt{n}(\mathbb{H}_n(t) - H_0(t)) \tag{7.1}$$

From Shorack and Wellner (1986), Chapter 7, Theorem 7.4.1, page 307, we know that for $t \in (0, T_0)$ with $T_0 \equiv T_0(F_0) \equiv \inf\{x : \; F(x) = 1\}$, $\mathbb{B}_n(t) \Rightarrow B(C(t))$ in $D[0,M]$ for $M < T_0$, where B denotes a standard Brownian motion on $[0,\infty)$ and

$$C(t) = \frac{F_0(t)}{1 - F_0(t)} = \frac{1}{1 - F_0(t)} - 1. \tag{7.2}$$

Since $\mathbb{Y}_n(t) = \int_{[0,t]} \mathbb{H}_n(s)ds$, we define $x_n(t) = x_0 + n^{-1/5}t$ and

$$\widetilde{\mathbb{Y}}_n^{loc}(t) \equiv n^{4/5} \int_{x_0}^{x_n(t)} \left\{ \mathbb{H}_n(v) - \mathbb{H}_n(x_0) - \int_{x_0}^{v} (h_0(x_0) + (u - x_0)h_0'(x_0))du \right\} dv. \tag{7.3}$$

Thus we can rewrite $\widetilde{\mathbb{Y}}_n^{loc}$ as follows:

$$\begin{aligned}
\widetilde{\mathbb{Y}}_n^{loc}(t) &= n^{4/5} \int_{x_0}^{x_n(t)} \{\mathbb{H}_n(v) - \mathbb{H}_n(x_0) - (H_0(v) - H_0(x_0))\}\, dv \\
&\quad + n^{4/5} \int_{x_0}^{x_n(t)} \left\{ H_0(v) - H_0(x_0) - \int_{x_0}^{v} (h_0(x_0) + (u - x_0)h_0'(x_0))du \right\} dv \\
&= n^{3/10} \int_{x_0}^{x_n(t)} \{\mathbb{B}_n(v) - \mathbb{B}_n(x_0)\}\, dv + \; n^{4/5} \int_{x_0}^{x_n(t)} \frac{1}{6} h_0''(x_0^*)(v - x_0)^3 dv + o(1) \\
&= n^{3/10} \int_{x_0}^{x_0 + n^{-1/5}t} \{\mathbb{B}_n(v) - \mathbb{B}_n(x_0)\}\, dv + \frac{1}{24} h_0''(x_0)t^4 + o(1).
\end{aligned}$$

Thus it follows easily that

$$\widetilde{\mathbb{Y}}_n^{loc}(t) \Rightarrow \sqrt{C'(x_0)} \int_0^t W(s)ds + \frac{1}{24} h_0''(x_0)t^4 \qquad \text{in} \;\; D[-M, M]$$

for each fixed $0 < M < \infty$ where W is a two-sided Brownian motion process starting at 0 and

$$C'(t) = \frac{f_0(t)}{(1 - F_0(t))^2} = \frac{h_0(t)}{1 - F_0(t)}.$$

Now for the derivative. Let $\breve{\mathbb{H}}_n(t) = \mathbb{H}_n(t) - H_0(t)$.

$$\begin{aligned}
(\widetilde{\mathbb{Y}}_n^{loc})'(t) &= n^{3/5}\left\{\mathbb{H}_n(x_n(t)) - \mathbb{H}_n(x_0) - \int_{x_0}^{x_n(t)} (h_0(x_0) + (u - x_0)h_0'(x_0))du\right\} \\
&= n^{3/5}\left\{\breve{\mathbb{H}}_n(x_0 + n^{-1/5}t) - \breve{\mathbb{H}}_n(x_0)\right. \\
&\qquad \left. + H_0(x_n(t)) - H_0(x_0) - \int_{x_0}^{x_n(t)} (h_0(x_0) + (u - x_0)h_0'(x_0))du\right\} \\
&= n^{1/10}\left\{\mathbb{B}_n(x_0 + n^{-1/5}t) - \mathbb{B}_n(x_0)\right\} + \frac{1}{3!}h_0''(x_0)t^3 + o(1)
\end{aligned}$$

and hence

$$(\widetilde{\mathbb{Y}}_n^{loc})'(t) \Rightarrow \sqrt{C'(x_0)}W(t) + \frac{1}{3!}h_0''(x_0)t^3,$$

again in $D[-M, M]$.

Now that we have identified the limiting behavior of the "driving" process $\widetilde{\mathbb{Y}}_n^{loc}(t)$, and its derivative, we turn to finding a process, $\widetilde{\mathbb{I}}_n^{loc}(t)$, which will serve as the invelope process for $\widetilde{\mathbb{Y}}_n^{loc}(t)$.

As above, let $\mathbb{H}_n$ denote the empirical hazard function, and let $\widetilde{H}_n(t) = \int_0^t \widetilde{h}_n(s)ds$, and $\widetilde{\mathcal{H}}_n(t) = \int_0^t \widetilde{H}_n(s)ds$. Then define

$$\begin{aligned}
\widetilde{\mathbb{I}}_n^{loc}(t) &= n^{4/5}\int_{x_0}^{x_0+n^{-1/5}t}\int_{x_0}^{v}\left\{\widetilde{h}_n(u) - h_0(x_0) - (u - x_0)h_0'(x_0)\right\}dudv \\
&\quad + \widetilde{A}_n t + \widetilde{B}_n,
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{A}_n &= n^{3/5}\left\{\widetilde{H}_n(x_0) - \mathbb{H}_n(x_0)\right\} \\
\widetilde{B}_n &= n^{4/5}\left\{\widetilde{\mathcal{H}}_n(x_0) - \int_0^{x_0}\mathbb{H}_n(v)dv\right\}.
\end{aligned}$$

We will show that this process converges to an appropriately scaled version of $\mathcal{I}$ above, and that its derivatives describe the limiting behavior of our estimators as above.

Define the vector

$$\widetilde{\mathbb{Z}}_n(t) = (\widetilde{\mathbb{Y}}_n^{loc}(t), (\widetilde{\mathbb{Y}}_n^{loc})'(t), \widetilde{\mathbb{I}}_n^{loc}(t), (\widetilde{\mathbb{I}}_n^{loc})'(t), (\widetilde{\mathbb{I}}_n^{loc})''(t), (\widetilde{\mathbb{I}}_n^{loc})'''(t)), \tag{7.4}$$

and fix $M > 0$. We will show that $\widetilde{\mathbb{Z}}_n$ is tight in the product space

$$E[-M, M] \equiv C[-M, M] \times D[-M, M] \times C[-M, M]^3 \times D[-M, M]. \tag{7.5}$$

This will be done last. We first assume that $\widetilde{\mathbb{Z}}_n$ has a weak limit, and we identify its unique limit. The two arguments together prove that $\widetilde{\mathbb{Z}}_n$ has the appropriate limiting distribution.

Identifying the Limit of the LSE. We will first show that $\widetilde{\mathbb{I}}_n^{loc}(t)$ is the invelope of $\widetilde{\mathbb{Y}}_n^{loc}(t)$. That is, we show that these processes satisfy the conditions (2.11)-(2.13) of the definition of the process $\mathcal{I}$. We also argue that these conditions pass to the limiting processes.

For condition (2.11), calculate

$$\begin{aligned}\widetilde{\mathbb{I}}_n^{loc}(t) - \widetilde{\mathbb{Y}}_n^{loc}(t) &= n^{4/5}\int_{x_0}^{x_0+n^{-1/5}t} \widetilde{H}_n(u) - \mathbb{H}_n(u)dv + \widetilde{A}_n t + \widetilde{B}_n \\ &= n^{4/5}\left\{\widetilde{\mathcal{H}}_n(x_0+n^{-1/5}t) - \int_0^{x_0+n^{-1/5}t} \mathbb{H}_n(s)ds\right\} \geq 0\end{aligned}$$

by (2.9).

Next, we calculate the derivatives of $\widetilde{\mathbb{I}}_n^{loc}(t)$.

$$\begin{aligned}(\widetilde{\mathbb{I}}_n^{loc})'(t) &= n^{3/5}\int_{x_0}^{x_0+n^{-1/5}t}\left\{\widetilde{h}_n(u) - h_0(x_0) - (u-x_0)h_0'(x_0)\right\}du + \widetilde{A}_n, \\ (\widetilde{\mathbb{I}}_n^{loc})''(t) &= n^{2/5}\left\{\widetilde{h}_n(x_0+n^{-1/5}t) - h_0(x_0) - (n^{-1/5}t)h_0'(x_0)\right\}, \\ (\widetilde{\mathbb{I}}_n^{loc})'''(t) &= n^{1/5}\left\{\widetilde{h}_n'(x_0+n^{-1/5}t) - h_0'(x_0)\right\}.\end{aligned}$$

Clearly, $(\widetilde{\mathbb{I}}_n^{loc})''(t)$ is convex, and differentiable at 0. In the space $E[-M,M]$, these conditions also pass to the limit. It remains to show that

$$\int_{-c}^{c}\left\{\widetilde{\mathbb{I}}_n^{loc}(t) - \widetilde{\mathbb{Y}}_n^{loc}(t)\right\} d\,(\widetilde{\mathbb{I}}_n^{loc})'''(t) = 0.$$

for any choice of $c > 0$ such that $0 \leq x_0 - n^{1/5}c \leq x_0 + n^{1/5}c \leq T$. But $(\widetilde{\mathbb{I}}_n^{loc})'''(t)$ has changepoints only where $\widetilde{h}_n'(x_0+n^{-1/5}t)$ has changepoints. Now, let $t = \tau$ denote such a changepoint. By our previous calculations, we have that

$$\widetilde{\mathbb{I}}_n^{loc}(t) - \widetilde{\mathbb{Y}}_n^{loc}(t) = n^{4/5}\left\{\widetilde{\mathcal{H}}_n(x_0+n^{-1/5}t) - \mathbb{Y}_n(x_0+n^{-1/5}t)\right\},$$

and hence $\widetilde{\mathbb{I}}_n^{loc}(\tau) - \widetilde{\mathbb{Y}}_n^{loc}(\tau) = 0$ by (2.9) and (2.10). Hence, $\widetilde{\mathbb{I}}_n^{loc}(t)$ is an invelope.

Lastly we show that on this space the conditions (2.11)-(2.13) are maintained under limits. This is clear for conditions (2.11) and (2.12). For the last condition, we show that the continuous mapping theorem applies. This follows since for any element $z = \{z_1, z_2, z_3, z_4, z_5, z_6\} \in E[-M,M]$,

$$\psi(z) = \int_{-M}^{M}(z_3 - z_1)dz_6$$

is continuous in z, for z_6 increasing. Since $(\widetilde{\mathbb{I}}_n^{loc})''(t)$ is convex, we deduce that $(\widetilde{\mathbb{I}}_n^{loc})'''(t)$ is increasing. This shows that the only possible limit of $\widetilde{\mathbb{I}}_n^{loc}(t)$ is the process $\mathcal{I}$.

Now, it is easy to see that the second and third derivatives of $\widetilde{\mathbb{I}}_n^{loc}$ evaluated at $t = 0$ are equal to

$$\begin{pmatrix} n^{2/5}(\widetilde{h}_n(x_0) - h_0(x_0)) \\ n^{1/5}(\widetilde{h}_n'(x_0) - h_0'(x_0)) \end{pmatrix},$$

as desired. Also, recall that

$$\widetilde{\mathbb{Y}}_n^{loc}(t) \Rightarrow \sqrt{C'(x_0)} \int_0^t W(s)ds + \frac{1}{24} h_0''(x_0) t^4 \equiv k_1 \int_0^t W(s)ds + k_2 t^4.$$

Now we use the scaling properties of the process Y involved in the definition of the "invelope process"; recall Definition 2.6. For any $a, b > 0$, $bY(at) \stackrel{d}{=} a^{3/2} b \int_0^t W(s)ds + a^4 b t^4$. Therefore, choose a, b so that $a^4 b = k_2$, and $a^{3/2} b = k_1$. It follows that

$$\widetilde{\mathbb{Y}}_n^{loc}(t) \Rightarrow bY(at).$$

Applying this re-scaling to all processes shows that

$$(\widetilde{\mathbb{I}}_n^{loc})''(0) \Rightarrow ba^2 \mathcal{I}''(0) \quad \text{and} \quad (\widetilde{\mathbb{I}}_n^{loc})'''(0) \Rightarrow ba^3 \mathcal{I}'''(0).$$

It is now straightforward to calculate the correct constants.

Tightness for the LSE. We already know that both $\widetilde{\mathbb{Y}}_n^{loc}(t)$ and $(\widetilde{\mathbb{Y}}_n^{loc})'(t)$ are tight in $C[-M, M]$ and $D[-M, M]$ respectively. Proposition 6.5 says that $(\widetilde{\mathbb{I}}_n^{loc})''(t)$ and $(\widetilde{\mathbb{I}}_n^{loc})'''(t)$ are tight in $C[-M, M]$. It remains to argue the same for $(\widetilde{\mathbb{I}}_n^{loc})'(t)$ and $\widetilde{\mathbb{I}}_n^{loc}(t)$. However, this will follow by Proposition 6.5 if we can show that both $\widetilde{A}_n$ and $\widetilde{A}_n t + \widetilde{B}_n$ are tight.

Let τ_n be the largest touchpoint smaller than x_0. Using Corollary 2.4 we have

$$\begin{aligned} \widetilde{A}_n &= n^{3/5} \left\{ \widetilde{H}_n(x_0) - \mathbb{H}_n(x_0) \right\} - n^{3/5} \left\{ \widetilde{H}_n(\tau_n) - \mathbb{H}_n(\tau_n) \right\} \\ &= n^{3/5} \left\{ \int_{\tau_n}^{x_0} \widetilde{h}_n(u) - h_0(x_0) - h_0'(x_0)(u - x_0) du \right\} \\ &\quad - n^{3/5} \left\{ \int_{\tau_n}^{x_0} h_0(u) - h_0(x_0) - h_0'(x_0)(u - x_0) du \right\} \\ &\quad\quad - n^{3/5} \int_{\tau_n}^{x_0} d\left\{ \mathbb{H}_n - H_0 \right\}(u) \\ &= n^{3/5} \left\{ \int_{\tau_n}^{x_0} \widetilde{h}_n(u) - h_0(x_0) - h_0'(x_0)(u - x_0) du \right\} \\ &\quad - n^{3/5} \left\{ h_0''(x_0) + o(1) \right\} (x_0 - \tau_n)^3 - n^{3/5} \int_{\tau_n}^{x_0} d\left\{ \mathbb{H}_n - H_0 \right\}(u). \end{aligned}$$

By Proposition 6.5 and Lemma 6.3 the first two terms are tight in $C[-M, M]$. Arguments similar to those used in the proof of Lemma 6.1 show that if $b \leq 1$ then

$$\limsup_{n\to\infty} P\left(\sup_{0\leq u\leq b}\left|\int_{x_0-u}^{x_0} d\left\{\mathbb{H}_n - H_0\right\}(u)\right| \geq M\sqrt{\frac{b}{n}}\right) \to 0 \qquad \text{as } \; M\to\infty.$$

Since $\tau_n - x_0 = O_p(n^{-1/5})$ by Lemma 6.3, this implies that $\widetilde{A}_n$ is tight in $D[-M, M]$. This of course implies that $\widetilde{B}_n$ is tight in $C[-M, M]$. This, in turn, implies that $\widetilde{\mathbb{Z}}_n$ is tight in the space $E[-M, M]$. This completes the proof of Theorem 2.7 for the LSE.

7.2. **Proof of Theorem 2.7 for the MLE.**

The proof here is quite similar to that of the LSE section. There are however, several additional technical difficulties which arise.

One of the main changes is that we now need to consider two separate cases. Note that at x_0 (where $h''(x_0) > 0$), we have three possibilities

(1) $h_0'(x_0) > 0$ By continuity, $h_0'(x) > 0$ in a neighborhood of x_0. It follows from the consistency of the MLE derivatives, that $\widehat{h}_n' > 0$ for sufficiently large n, and hence all touch points to consider are of the "increasing" kind.
(2) $h_0'(x_0) < 0$ By the same argument, all touch points are decreasing.
(3) $h_0'(x_0) = 0$ This is the tricky case. However, since $h(x_0) > 0$, by Corollary 3.4 we know that there is always at least one touch point which satisfies both the nonincreasing and nondecreasing properties, the limiting process may be "stitched" together in an appropriate manner.

Therefore it will be sufficient to prove the asymptotic results for both types of touch points. We also note that because we work in a neighborhood of x_0 such that $h(x_0) > 0$, we may assume that $\widehat{h}_n$ is always well-defined (i.e. finite).

Nonincreasing.

Driving process for the MLE, nonincreasing case. In this case the driving process is slightly different than that for the LSE. It is

$$\begin{aligned}
&\widehat{\mathbb{Y}}_{n,\downarrow}^{loc}(t) \\
&= n^{4/5}\frac{h_0(x_0)}{S_0(x_0)}\int_{x_0}^{x_0+n^{-1/5}t}\int_{x_0}^{v}\left\{\frac{h_0(u)-h_0(x_0)-(u-x_0)h_0'(x_0)}{\widehat{h}_n(u)}\right\}\mathbb{S}_n(u)du dv \\
&\quad + n^{4/5}\frac{h_0(x_0)}{S_0(x_0)}\int_{x_0}^{x_0+n^{-1/5}t}\int_{x_0}^{v}\frac{\mathbb{S}_n(u)}{\widehat{h}_n(u)}d\{\mathbb{H}_n^*(u)-H_0(u)\}dv,
\end{aligned}$$

where $d\mathbb{H}_n^*(u) = \frac{\mathbb{S}_n(u^-)}{\mathbb{S}_n(u)} d\mathbb{H}_n(u)$. The derivative is

$$
\begin{aligned}
(\widehat{\mathbb{Y}}_{n,\downarrow}^{loc})'(t) &= n^{3/5}\frac{h_0(x_0)}{S_0(x_0)}\int_{x_0}^{x_0+n^{-1/5}t}\left\{\frac{h_0(u)-h_0(x_0)-(u-x_0)h_0'(x_0)}{\widehat{h}_n(u)}\right\}\mathbb{S}_n(u)du \\
&\quad + n^{3/5}\frac{h_0(x_0)}{S_0(x_0)}\int_{x_0}^{x_0+n^{-1/5}t}\frac{\mathbb{S}_n(u)}{\widehat{h}_n(u)}d\{\mathbb{H}_n^*(u)-H_0(u)\}.
\end{aligned}
$$

Note that by consistency of $\widehat{h}_n$ and since $\sup_t |\mathbb{S}_n(t)-S_0(t)| \to 0$ a.s., for any $M>0$

$$
\lim_n \sup_{|t|\le M} |\widehat{\mathbb{Y}}_{n,\downarrow}^{loc}(t) - \widetilde{\mathbb{Y}}_n^{loc}(t)| = \lim_n \sup_{|t|\le M} |(\widehat{\mathbb{Y}}_{n,\downarrow}^{loc})'(t) - (\widetilde{\mathbb{Y}}_n^{loc})'(t)| = 0 \quad \text{a.s.} \tag{7.6}
$$

Identifying the Limit for the MLE, nonincreasing case. Recall definitions (6.8) and (6.7). The (interim) invelope here is defined by

$$
\begin{aligned}
\widehat{\mathbb{I}}_{n,\downarrow}^{loc}(t) &= n^{4/5}\frac{h_0(x_0)}{S_0(x_0)}\int_{x_0}^{x_0+n^{-1/5}t}\int_{x_0}^{v}\left\{\frac{\widehat{h}_n(u)-h_0(x_0)-(u-x_0)h_0'(x_0)}{\widehat{h}_n(u)}\right\}\mathbb{S}_n(u)dudv \\
&\quad +\widehat{A}_{n,\downarrow}t+\widehat{B}_{n,\downarrow},
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{A}_{n,\downarrow} &= -n^{3/5}\frac{h_0(x_0)}{S_0(x_0)}\left\{\widehat{\mathcal{H}}_{n,\downarrow}'(x_0)-\mathbb{A}_{n,\downarrow}(x_0)\right\} \\
\widehat{B}_{n,\downarrow} &= -n^{4/5}\frac{h_0(x_0)}{S_0(x_0)}\left\{\widehat{\mathcal{H}}_{n,\downarrow}(x_0)-\int_0^{x_0}\mathbb{A}_{n,\downarrow}(v)dv\right\}.
\end{aligned}
$$

Next, calculate

$$
\begin{aligned}
&\widehat{\mathbb{I}}_{n,\downarrow}^{loc}(t)-\widehat{\mathbb{Y}}_{n,\downarrow}^{loc}(t) \\
&\quad = n^{4/5}\frac{h_0(x_0)}{S_0(x_0)}\int_{x_0}^{x_0+n^{-1/5}t}\int_{x_0}^{v}\left\{\mathbb{S}_n(u)du-\frac{\mathbb{S}_n(u)}{\widehat{h}_n(u)}d\mathbb{H}_n^*(u)\right\}dv+\widehat{A}_{n,\downarrow}t+\widehat{B}_{n,\downarrow} \\
&\quad = n^{4/5}\frac{h_0(x_0)}{S_0(x_0)}\left\{\int_0^{x_0+n^{-1/5}t}\mathbb{A}_{n,\downarrow}(v)dv-\widehat{\mathcal{H}}_{n,\downarrow}(x_0+n^{-1/5}t)\right\}\ge 0,
\end{aligned} \tag{7.7}
$$

with equality at the (nonincreasing) touchpoints of $\widehat{h}_n$, using (2.2).

Notice that because of the presence of $\mathbb{S}_n(v)$ in its definition, $\widehat{\mathbb{I}}_{n,\downarrow}^{loc}(t)$ is not three times differentiable. We therefore define

$$
\begin{aligned}
\widehat{\mathbb{I}}_{n,\downarrow}^{*,loc}(t) &= n^{4/5}\frac{h_0(x_0)}{S_0(x_0)}\int_{x_0}^{x_n(t)}\int_{x_0}^{v}\left\{\frac{\widehat{h}_n(u)-h_0(x_0)-(u-x_0)h_0'(x_0)}{\widehat{h}_n(u)}\right\}S_0(u)dudv \\
&\quad +\widehat{A}_{n,\downarrow}t+\widehat{B}_{n,\downarrow},
\end{aligned}
$$

where, again, $x_n(t) = x_0 + n^{-1/5}t$, and for any $M > 0$,

$$\lim_n \sup_{|t|\le M} |\widehat{\mathbb{I}}^{loc}_{n,\downarrow}(t) - \widehat{\mathbb{I}}^{*,loc}_{n,\downarrow}(t)| = 0, \qquad (7.8)$$

due to Proposition 6.5.

Now, fix $M > 0$, and define

$$\widehat{\mathbb{Z}}_{n,\downarrow}(t) = (\widehat{\mathbb{Y}}^{loc}_{n,\downarrow}(t), (\widehat{\mathbb{Y}}_{n,\downarrow}n^{loc})'(t), \widehat{\mathbb{I}}^{*,loc}_{n,\downarrow}(t), (\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow})'(t), (\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow})''(t), (\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow})'''(t)). \qquad (7.9)$$

We will show that $\widehat{\mathbb{Z}}_{n,\downarrow}(t)$ is tight in $E[-M, M]$ (defined in (7.5)). However, we first argue that $\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow}(t)$ is the invelope of $\widehat{\mathbb{Y}}^{loc}_{n,\downarrow}(t)$. We have already seen that these properties pass to the limit. However, in this case, $\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow}(t)$ satisfies (2.11)-(2.13) asymptotically.

First, from (7.7) and (7.8) it follows that $\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow}(t)$ satisfies invelope condition (2.11) in the limit. Next, the derivatives of $\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow}(t)$ are calculated as follows:

$$\begin{aligned}
(\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow})'(t) &= n^{3/5}\frac{h_0(x_0)}{S_0(x_0)}\int_{x_0}^{x_n(t)}\left\{\frac{\widehat{h}_n(u) - h_0(x_0) - (u - x_0)h_0'(x_0)}{\widehat{h}_n(u)}\right\}S_0(u)du + \widehat{A}_{n,\downarrow} \\
(\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow})''(t) &= n^{2/5}\frac{h_0(x_0)}{S_0(x_0)}\left\{\frac{\widehat{h}_n(x_n(t)) - h_0(x_0) - n^{-1/5}t\,h_0'(x_0)}{\widehat{h}_n(x_n(t))}\right\}S_0(x_0 + n^{-1/5}t)
\end{aligned}$$

Due to Theorem 4.1 and Proposition 6.5, we have that

$$\lim_n \sup_{|t|\le M}\left|(\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow})''(t) - n^{2/5}\left[\widehat{h}_n(x_0 + n^{-1/5}t) - h_0(x_0) - n^{-1/5}t\,h_0'(x_0)\right]\right| = 0, \qquad (7.10)$$

where $n^{2/5}[\widehat{h}_n(x_0 + n^{-1/5}t) - h_0(x_0) - n^{-1/5}t\,h_0'(x_0)]$ is convex, and hence the limit of $(\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow})''(t)$ will be convex.

Let $B_n(t) = (h_0(x_0)/S_0(x_0)) \times (S_0(t)/\widehat{h}_n(t))$; then

$$B_n'(t) = \frac{h_0(x_0)}{S_0(x_0)}\left[\frac{S_0'(t)\widehat{h}_n(t) - S_0(t)\widehat{h}_n'(t)}{\widehat{h}_n^2(t)}\right]$$

and

$$\begin{aligned}
dB_n'(t) = \frac{h_0(x_0)}{S_0(x_0)}\Bigg[&-\frac{2}{\widehat{h}_n^3(t)}\widehat{h}_n'(t)[S_0'(t)\widehat{h}_n(t) - S_0(t)\widehat{h}_n'(t)]dt \\
&+\frac{1}{\widehat{h}_n^2(t)}[S_0''(t)\widehat{h}_n(t)]dt - \frac{S_0(t)}{\widehat{h}_n^2(t)}d\widehat{h}_n'(t)\Bigg].
\end{aligned}$$

We may then write

$$\begin{aligned}
(\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow})'''(t) &= n^{1/5}[\widehat{h}_n'(x_n(t)) - h_0'(x_0)]B_n(x_n(t)) \\
&\quad + n^{1/5}[\widehat{h}_n(x_n(t)) - h_0(x_0) - n^{-1/5}t\,h_0'(x_0)] \times B_n'(x_n(t)),
\end{aligned}$$

Notice that $\sup_{|t|\leq M} |1 - B_n(x_0 + n^{-1/5}t)| \rightarrow_{a.s.} 0$, with $\lim_n B_n'(x_0 + n^{-1/5}t)$ bounded. Therefore, from Proposition 6.5 it follows

$$\lim_n \sup_{|t|\leq M} \left| (\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow})'''(t) - n^{1/5}[\widehat{h}_n'(x_0 + n^{-1/5}t) - h_0'(x_0)] \right| = 0, \tag{7.11}$$

where $n^{1/5}[\widehat{h}_n'(x_0 + n^{-1/5}t) - h_0'(x_0)]$ is piecewise linear, with jumps at the touchpoints of $\widehat{h}_n$. By consistency of $\widehat{h}_n$, we have

$$\begin{aligned} d(\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow})'''(t) &= B_n(x_0 + n^{-1/5}t) d\hat{g}_n(t) \\ &\quad + 2[\widehat{h}_n'(x_0 + n^{-1/5}t) - h_0'(x_0)] B_n'(x_0 + n^{-1/5}t) dt \\ &\quad + n^{1/5}[\widehat{h}_n(x_0 + n^{-1/5}t) - h_0(x_0) - n^{-1/5}t\, h_0'(x_0)] dB_n'(x_0 + n^{-1/5}t) \\ &= \left\{ B_n(x_0 + n^{-1/5}t) + O_p^*(n^{-2/5}) \right\} d\hat{g}_n(t) + O_p^*(n^{-1/5}) dt, \end{aligned}$$

where $\hat{g}_n(t) = n^{1/5}[\widehat{h}_n'(x_0 + n^{-1/5}t) - h_0'(x_0)]$. We say that a process $X_n(t)$ is $O_p^*(1)$ if $\sup_{|t|\leq M} |X_n(t)|$ is $O_p(1)$.

Next, arguing as for the LSE, we may show that for all $c > 0$

$$0 = \int_{-c}^{c} (\widehat{\mathbb{I}}^{loc}_{n,\downarrow}(t) - \widehat{\mathbb{Y}}^{loc}_{n,\downarrow}(t)) d\hat{g}_n(t)$$

and hence

$$\begin{aligned} \int_{-c}^{c} (\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow}(t) - \widehat{\mathbb{Y}}^{loc}_{n,\downarrow}(t)) d(\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow})'''(t) &= \int_{-c}^{c} (\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow}(t) - \widehat{\mathbb{Y}}^{loc}_{n,\downarrow}(t)) d[(\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow})''' - \hat{g}_n](t) \\ &\quad + \int_{-c}^{c} (\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow}(t) - \widehat{\mathbb{I}}^{loc}_{n,\downarrow}(t)) d\hat{g}_n(t) = o_p(1), \end{aligned}$$

using Proposition 6.5,(7.8), and the fact that $\hat{g}_n$ is increasing.

This shows that $\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow}(t)$ satisfies the invelope conditions (2.11)-(2.13) asymptotically. From (7.6), it follows that $\widehat{\mathbb{I}}^{*,loc}_{n,\downarrow}(t)$ has the (appropriately re-scaled) process $\mathcal{I}$ as its only possible limit. From (7.10) and (7.11), and the same re-scaling argument as for the LSE, the limits of the MLE estimators are identified.

Tightness for the MLE, nonincreasing case. The tightness arguments here are the same as for the LSE case, and we therefore omit the details. The only "new" calculation is shown below.

Let τ_n be the largest touchpoint smaller than x_0. By (3.15) we have

$$
\begin{aligned}
&-\frac{S_0(x_0)}{h_0(x_0)}\widehat{A}_{n,\downarrow} \\
&= n^{3/5}\left\{\widehat{\mathcal{H}}_n'(x_0) - \int_0^{x_0}\mathbb{S}_n(u)du\right\} - n^{3/5}\left\{\widehat{\mathcal{H}}_n'(\tau_n) - \int_0^{\tau_n}\mathbb{S}_n(u)du\right\} \qquad (7.12)\\
&= -n^{3/5}\left\{\int_{\tau_n}^{x_0}\frac{\widehat{h}_n(u)-h_0(u)}{\widehat{h}_n(u)}S_0(u)du\right\} + n^{3/5}\int_{\tau_n}^{x_0}\frac{1}{\widehat{h}_n(u)}d\left\{\mathbb{F}_n - F_0\right\}(u) \\
&\qquad + n^{3/5}\int_{\tau_n}^{x_0}(\mathbb{S}_n(u)-S_0(u)du) \\
&= -n^{3/5}\left\{\int_{\tau_n}^{x_0}\frac{\widehat{h}_n(u)-h_0(x_0)-h_0'(x_0)(u-x_0)}{\widehat{h}_n(u)}S_0(u)du\right\} \\
&\qquad + n^{3/5}\left\{\int_{\tau_n}^{x_0}\frac{h_0(u)-h_0(x_0)-h_0'(x_0)(u-x_0)}{\widehat{h}_n(u)}S_0(u)du\right\} \\
&\qquad + n^{3/5}\int_{\tau_n}^{x_0}\frac{1}{\widehat{h}_n(u)}d\left\{\mathbb{F}_n - F_0\right\}(u) + n^{3/5}\int_{\tau_n}^{x_0}\mathbb{S}_n(u)-S_0(u)du.
\end{aligned}
$$

Nondecreasing.

Driving process for the MLE, nondecreasing case.

$$
\begin{aligned}
\widehat{\mathbb{Y}}^{loc}_{n,\uparrow}(-t) &= n^{4/5}\frac{h_0(x_0)}{S_0(x_0)}\int_{x_n(-t)}^{x_0}\int_v^{x_0}\left\{\frac{h_0(u)-h_0(x_0)-(u-x_0)h_0'(x_0)}{\widehat{h}_n(u)}\right\}\mathbb{S}_n(u)dudv \\
&\quad + n^{4/5}\frac{h_0(x_0)}{S_0(x_0)}\int_{x_n(-t)}^{x_0}\int_v^{x_0}\frac{\mathbb{S}_n(u)}{\widehat{h}_n(u)}d\{\mathbb{H}_n^*(u)-H_0(u)\}dv.
\end{aligned}
$$

Define

$$
\widetilde{\mathbb{Y}}^{loc}_{n,\uparrow}(-t) = n^{4/5}\int_{x_n(-t)}^{x_0}\left\{\mathbb{H}_n(x_0)-\mathbb{H}_n(v)-\int_v^{x_0}(h_0(x_0)+(u-x_0)h_0'(x_0))du\right\}dv.
$$

Note that by consistency of $\widehat{h}_n$ and as $\sup_t|\mathbb{S}_n(t)-S_0(t)|\to 0$ a.s., we have that for any $M>0$

$$
\sup_{|t|\le M}|\widehat{\mathbb{Y}}^{loc}_{n,\uparrow}(t)-\widetilde{\mathbb{Y}}^{loc}_{n,\uparrow}(t)| \vee \sup_{|t|\le M}|(\widehat{\mathbb{Y}}^{loc}_{n,\uparrow})'(t)-(\widetilde{\mathbb{Y}}^{loc}_{n,\uparrow})'(t)| \to_{a.s.} 0.
$$

Also, using the same arguments as in Subsection 7.1, we have

$$
\lim_n \widetilde{\mathbb{Y}}^{loc}_{n,\uparrow}(\cdot) \stackrel{d}{=} \lim_n \widetilde{\mathbb{Y}}^{loc}_{n}(\cdot) \text{ and } \lim_n (\widetilde{\mathbb{Y}}^{loc}_{n,\uparrow})'(\cdot) \stackrel{d}{=} \lim_n (\widetilde{\mathbb{Y}}^{loc}_{n})'(\cdot),
$$

that is, they have the same weak limit.

Identifying the Limit, MLE nondecreasing case. Recall definitions (6.10) and (6.9). Let

$$\begin{aligned}\widehat{\mathbb{I}}^{loc}_{n,\uparrow}(-t) &= n^{4/5}\frac{h_0(x_0)}{S_0(x_0)}\int_{x_n(-t)}^{x_0}\int_v^{x_0}\left\{\frac{\widehat{h}_n(u)-h_0(x_0)-(u-x_0)h_0'(x_0)}{\widehat{h}_n(u)}\right\}\mathbb{S}_n(u)dudv\\ &\quad +\widehat{A}_{n,\uparrow}t+\widehat{B}_{n,\uparrow},\end{aligned}$$

where

$$\begin{aligned}\widehat{A}_{n,\uparrow} &= -n^{3/5}\frac{h_0(x_0)}{S_0(x_0)}\left\{\widehat{\mathcal{H}}'_{n,\uparrow}(x_0)-\mathbb{A}_{n,\uparrow}(x_0)\right\}\\ \widehat{B}_{n,\uparrow} &= -n^{4/5}\frac{h_0(x_0)}{S_0(x_0)}\left\{\widehat{\mathcal{H}}_{n,\uparrow}(x_0)-\int_0^{x_0}\mathbb{A}_{n,\uparrow}(v)dv\right\}.\end{aligned}$$

And we easily calculate

$$\widehat{\mathbb{I}}^{loc}_{n,\uparrow}(-t)-\widehat{\mathbb{Y}}^{loc}_{n,\uparrow}(-t) = n^{4/5}\frac{h_0(x_0)}{S_0(x_0)}\left\{\int_{x_0-n^{-1/5}t}^{\infty}\mathbb{A}_{n,\uparrow}(v)dv-\widehat{\mathcal{H}}_{n,\uparrow}(x_0-n^{-1/5}t)\right\}\geq 0,$$

with equality at the (nondecreasing) touchpoints of $\widehat{h}_n$, using (2.3). The rest of the proof proceeds in a similar manner. The only difference is that to prove tightness of the process $\widehat{A}_{n,\uparrow}$ we argue as (7.12), but with η_n the smallest touchpoint *greater than* x_0.

This completes the proof of Theorem 2.7 for the MLE.

8. Some Further Results

Here we collect several further results without proof. In particular: (i) we give characterizations of the MLE and LSE in the presence of right-censored data, and state the analogue of Theorem 2.7 in this case; (ii) we give characterizations of the MLE and LSE for estimation of a bathtub-shaped (i.e. $U-$shaped) hazard, and again state the analogue of Theorem 2.7.

8.1. **Estimating the Hazard with Right Censoring.** The model for random right censoring is described as follows. Let (X,Y) be two independent random variables with respective cumulative distribution functions F and G. Let h denote the hazard function for the random variable X. The data we observe is described as (T,Δ), where $T=\min\{X,Y\}$ and $\Delta = 1\{X\leq Y\}$. Observing i.i.d. observations from this model, the likelihood is (proportional to)

$$\Pi_{i=1}^n h(T_i)^{\Delta_i}e^{-H(T_i)}.$$

Let $0\leq T_{(1)}\leq T_{(1)}\leq\cdots\leq T_{(n)}$ denote the ordered T_j's, and let $\Delta_{(1)},\Delta_{(2)},\cdots,\Delta_{(n)}$ denote the corresponding Δ's. The same difficulties with the largest observation occur in this problem as with uncensored data if $\Delta_{(n)}=1$. That is, if $\Delta_{(n)}=1$, the MLE is found by minimizing $\varphi_n(h)$ over convex functions on $[0,T_{(n)})$ and $\widehat{h}_n(x)=\infty$ for

$x \geq T_{(n)}$. However, if $\Delta_{(n)} = 0$, then the censoring of the largest observation gives us the necessary control over the hazard function and the MLE is found by minimizing $\varphi_n(h)$ over convex functions on $[0, T_{(n)}]$; in this case no information is given for the values of $\widehat{h}_n(x)$ for $x > T_{(n)}$. Thus we define the MLE as the minimizer of the criterion function

$$\varphi_n(h) = \int_0^\infty H(t) d\mathbb{F}_n^T(t) - \int_0^\infty \log h(t) d\widetilde{\mathbb{G}}_n(t),$$

over the class $\mathcal{K}_+^{cens}$, where

$$\mathbb{F}_n^T(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{I}_{[0,t]}(T_i), \quad \mathbb{G}_n(t) = \frac{1}{n}\sum_{i=1}^n \Delta_i \mathbb{I}_{[0,t]}(T_i),$$
$$\text{and} \quad \widetilde{\mathbb{G}}_n(t) = \frac{1}{n}\sum_{i=1}^{n-1} \Delta_{(i)} \mathbb{I}_{[0,t]}(T_{(i)}),$$

and $\mathcal{K}_+^{cens}$ is the space of convex functions on $[0, T_{(n)})$ if $\Delta_{(n)} = 1$ and $[0, T_{(n)}]$ if $\Delta_{(n)} = 0$.

Note that $\mathbb{G}_n$ is the *sub*distribution function of the uncensored T_i's, and $\widetilde{\mathbb{G}}_n$ is the version of $\mathbb{G}_n$ which deletes $T_{(n)}$ if it corresponds to an uncensored observation (with $\Delta_{(n)} = 1$). Lastly, let $\mathbb{H}_n^{NA}$ denote the Nelson-Aalen estimator of the (cumulative) hazard function H,

$$\mathbb{H}_n^{NA}(t) = \int_{[0,t]} \frac{1}{1 - \mathbb{F}_n^T(s-)} d\mathbb{G}_n(s).$$

Fix a constant $M > 0$ with $P(T > M) > 0$. The LSE of h is the minimizer of

$$\psi_n(h) = \frac{1}{2}\int_0^M h^2(t) dt - \int_0^M h(t) d\mathbb{H}_n^{NA}(t).$$

First we state the generalizations of Lemmas 2.1 and 2.3 to this censored version of the problem.

Lemma 8.1. *Let $\{\tau_i, i = 1, \ldots, k\}$ denote all of the change points of $\widehat{h}_n$, where $\widehat{h}_n$ is non-increasing. Let $\{\eta_j, j = 1, \ldots, m\}$ denote all of the change points of $\widehat{h}_n$, where $\widehat{h}_n$ is non-decreasing.*

Then $\widehat{h}_n$ minimizes φ_n over $\mathcal{K}_+^{cens}$ if and only if:

$$\begin{aligned}
\int_0^x \frac{x-t}{\widehat{h}_n(t)} d\widetilde{\mathbb{G}}_n(t) &\leq \int_0^x \int_0^t (1-\mathbb{F}_n^T)(s) ds dt, && (8.1)\\
&\text{for all } x \geq 0 \quad \text{with} \quad \text{equality at } \tau_i \text{ for } i=1,\ldots,k;\\
\int_x^\infty \frac{t-x}{\widehat{h}_n(t)} d\widetilde{\mathbb{G}}_n(t) &\leq \int_x^\infty \int_t^\infty (1-\mathbb{F}_n^T)(s) ds dt, && (8.2)\\
&\text{for all } x \geq 0 \quad \text{with} \quad \text{equality at } \eta_j \text{ for } j=1,\ldots,m;\\
\int_0^\infty \frac{1}{\widehat{h}_n(t)} d\widetilde{\mathbb{G}}_n(t) &\leq \int_0^\infty (1-\mathbb{F}_n^T)(t) dt, && (8.3)\\
\int_0^\infty \widehat{H}_n(t) d\mathbb{F}_n^T(t) &= \frac{1}{n}\sum_{i=1}^{n-1} \Delta_{(i)}. && (8.4)
\end{aligned}$$

Moreover, the minimizer $\widehat{h}_n$ satisfies

$$\int_0^x \widehat{h}_n(t)(1-\mathbb{F}_n^T)(t) dt = \widetilde{\mathbb{G}}_n(x), \tag{8.5}$$

for $x \in \{\tau_1, \ldots, \tau_k, \eta_1, \ldots, \eta_m\}$.

Lemma 8.2. *Let $\mathbb{Y}_n^{NA}(t) = \int_0^t \mathbb{H}_n^{NA}(s) ds$. The function $\widetilde{h}_n$ minimizes $\psi_n(h)$ over $\mathcal{K}_M$ if and only if it satisfies*

$$\widetilde{H}_n(M) = \mathbb{H}_n^{NA}(M), \tag{8.6}$$

$$\widetilde{\mathcal{H}}_n(M) = \mathbb{Y}_n^{NA}(M), \tag{8.7}$$

$$\widetilde{\mathcal{H}}_n(t) \geq \mathbb{Y}_n^{NA}(t) \text{ for all } t \in [0, M], \tag{8.8}$$

$$\int_0^T (\widetilde{\mathcal{H}}_n - \mathbb{Y}_n^{NA})(t) d\, \widetilde{h}_n'(t) = 0. \tag{8.9}$$

The last statement is the same as: $\widetilde{\mathcal{H}}_n(\tau) = \mathbb{Y}_n^{NA}(\tau)$ for all changes of slope τ of $\widetilde{h}_n$.

The Maximum Likelihood and Least Squares estimators continue to be consistent on the interior of the support of F_0 in this setting of right-censored data. Moreover, Theorem 2.7 also holds with some changes in the constants:

Theorem 8.3. *Suppose that h_0 is convex, and that the censoring distribution function G_0, and $x_0 \in (0, M)$ satisfy $0 < h_0(x_0) < \infty$, $h_0''(x_0) > 0$, $G_0(x_0) < 1$, and that $h_0''(\cdot)$ is continuous in a neighborhood of x_0. Then the nonparametric MLE and LSE are asymptotically equivalent: for $\bar{h}_n = \widehat{h}_n$ or $\widetilde{h}_n$, then*

$$\begin{pmatrix} n^{2/5}(\bar{h}_n(x_0) - h_0(x_0)) \\ n^{1/5}(\bar{h}_n'(x_0) - h_0'(x_0)) \end{pmatrix} \to_d \begin{pmatrix} c_1\, \mathcal{I}^{(2)}(0) \\ c_2\, \mathcal{I}^{(3)}(0) \end{pmatrix}$$

where $\mathcal{I}^{(2)}(0)$ and $\mathcal{I}^{(3)}(0)$ are the second and third derivatives at 0 of the invelope of $Y(t) \equiv \int_0^t W(s)ds + t^4$, and where

$$c_1 = \left(\frac{h_0^2(x_0) h_0''(x_0)}{24 S_0^2(x_0)(1 - G_0(x_0))^2} \right)^{1/5}, \qquad c_2 = \left(\frac{h_0(x_0) h_0''(x_0)^3}{24^3 S_0(x_0)(1 - G_0(x_0))} \right)^{1/5}.$$

8.2. **Characterizations of the MLE and LSE of a U-Shaped hazard.** If we do not insist on convexity of the hazard function h, but still assume that the hazard is bathtub (or U-shaped), then we find ourselves in the same framework as Bray, Crawford and Proschan (1967b) or Banerjee (2007). In this section we derive the characterizations for the MLE and LSE, as in Lemmas 2.1 and 2.3, but under the "pure" bathtub assumptions.

We first consider the maximum likelihood problem. Here, we must first find the MLE constrained to having an antimode at t_0, and then maximize over all possible values of t_0 to find the overall MLE. Notice that the likelihood for any $t_0 < X_{(n)}$

$$\mathcal{L}(h) = \prod_{i=1}^n h(X_i) \exp\{-H(X_i)\},$$

can be made arbitrarily large by increasing the value of $h(X_{(n)})$. As before, we set $h(X_{(n)})$ in these cases to be arbitrarily large, and maximize the modified likelihood

$$\mathcal{L}^{mod}(h) = \prod_{i=1}^{n-1} h(X_i) \exp\{-H(X_i)\} \times \exp\left\{-H(X_{(n)})\right\}.$$

Hence our goal will be to minimize the criterion function

$$\varphi_n(h) = \int_0^\infty \left\{ H(t) - \log h(t) \mathbb{I}(t \neq X_{(n)}) \right\} d\mathbb{F}_n(t).$$

Let $\mathcal{U}_+$ denote the space of positive bathtub shaped hazard functions defined on $[0, X_{(n)}]$, and $\mathcal{U}_+(t_0)$ denote those functions in $\mathcal{U}_+$ that have an antimode at t_0. To find the MLE, we will therefore minimize φ_n over $h \in \mathcal{U}_+(t_0)$ and then minimize this over t_0. We need only consider $t_0 < X_{(n)}$.

Lemma 8.4. *The constrained MLE over $\mathcal{U}_+(t_0)$ exists and is unique. It is a piecewise constant, upper semi-continuous function, with jumps occurring only at data points. Also, if t_0 does not fall on a data point, then the MLE $\widehat{h}_n(t_0) = 0$. Moreover, the constrained MLE, $\widehat{h}_n$, is characterized by the following set of equations.*

Let $\{\tau_i, i = 1, \ldots, k\}$ denote all of the change points of $\widehat{h}_n$, where $\widehat{h}_n$ is non-increasing. Let $\{\eta_j, j = 1, \ldots, m\}$ denote all of the change points of $\widehat{h}_n$, where $\widehat{h}_n$ is non-decreasing. Then $\widehat{h}_n$ minimizes φ_n over the space of bathtub functions with *antimode at t_0 if and only if:*

$$\begin{aligned}
\int_{[0,\infty)} \widehat{H}_n(t) d\mathbb{F}_n(t) &= 1 - 1/n, & (8.10)\\
\int_{[0,\infty)} \frac{1}{\widehat{h}_n(t)} d\mathbb{F}_n(t) &\leq \int_0^\infty t d\mathbb{F}_n(t) = \int_0^\infty \mathbb{S}_n(t)dt, & (8.11)\\
\int_{[0,x]} \frac{1}{\widehat{h}_n(t)} d\mathbb{F}_n(t) &\leq x - \int_0^x (x-t) d\mathbb{F}_n(t) = \int_0^x \mathbb{S}_n(u)du, \ \forall\, x < t_0 & (8.12)\\
\int_{[x,\infty)} \frac{1}{\widehat{h}_n(t)} d\mathbb{F}_n(t) &\leq \int_x^\infty (t-x) d\mathbb{F}_n(t) = \int_x^\infty \mathbb{S}_n(u)du, \forall\, x > t_0 & (8.13)
\end{aligned}$$

with equality in (8.12) *at $x = \tau_i$, $i = 1, \ldots, k$, and equality in* (8.13) *at $x = \eta_j$ for $j = 1, \ldots, m$. Moreover, the minimizer $\widehat{h}_n$ satisfies*

$$\begin{aligned}
\int_0^{\tau_i} \widehat{h}_n(t)\mathbb{S}_n(t)dt &= \mathbb{F}_n(\tau_i), \quad i = 1, \ldots, k, & (8.14)\\
\int_0^{\eta_j} \widehat{h}_n(t)\mathbb{S}_n(t)dt &= \mathbb{F}_n(\eta_j^-) \quad j = 1, \ldots, m. & (8.15)
\end{aligned}$$

The unconstrained MLE over $\mathcal{U}_+$ is found by maximizing the likelihood over all possible antimodes t_0, with t_0 falling in between the data points. Indeed, it is sufficient to consider the n choices of intervals for t_0: $[0, X_{(1)}), (X_{(1)}, X_{(2)}), \ldots, (X_{(n-1)}, X_{(n)})$.

Remark 8.5. *We note that* (8.10) *was also observed to hold by Grenander (1956), page 143.*

Notice that (8.13) implies that $\widehat{h}_n(X_{(n)}) = \infty$. Also, if the values τ_i, η_j are known, then (2.6) may be used to calculate the values of $\widehat{h}_n$ directly. Since the MLE is piecewise constant, let us denote these constants as $h_1, h_2, \ldots$, and we therefore obtain that

$$h_1 = \frac{\mathbb{F}_n(\tau_1)}{\int_0^{\tau_1} \mathbb{S}_n(t)dt}, \qquad h_2 = \frac{\mathbb{F}_n(\tau_2) - \mathbb{F}_n(\tau_1)}{\int_{\tau_1}^{\tau_2} \mathbb{S}_n(t)dt},$$

and so forth. This is directly related to the more well-known calculation of the MLE via least concave majorants (see e.g. Robertson et al. (1988), page 342).

The least squares estimator is found by minimizing the criterion function

$$\psi_n(h) = \frac{1}{2}\int_0^T h^2 dt - \int_0^T h d\mathbb{H}_n(t).$$

We again adopt the approach of minimizing first over the space of bathtub-shaped functions on $[0, T]$ with antimode at t_0, $\mathcal{U}_T(t_0)$, and then minimizing over the possible choices of antimode. We will denote the space of non-negative bathtub-shaped functions on $[0, T]$ as $\mathcal{U}_T$.

Lemma 8.6. *The constrained LSE over $\mathcal{U}_+(t_0)$ exists and is unique. It is a piecewise constant, upper semi-continuous function, with jumps occurring only at data points. Also, when t_0 does not fall on a data point, then the LSE $\widetilde{h}_n(t_0) = 0$.*

The function $\widetilde{h}_n$ minimizes $\psi_n(h)$ over $\mathcal{U}_T(t_0)$, for t_0 not falling on a data point, if and only if it satisfies,

$$\widetilde{H}_n(T) = \mathbb{H}_n(T), \tag{8.16}$$

$$\widetilde{H}_n(t) \geq \mathbb{H}_n(t), \qquad \text{for all } \ 0 \leq t < t_0 \tag{8.17}$$

$$\widetilde{H}_n(t) \leq \mathbb{H}_n(t), \qquad \text{for all } \ t_0 < t \leq T \tag{8.18}$$

$$\int_0^T (\widetilde{H}_n - \mathbb{H}_n)(t) d\widetilde{h}_n(t) = 0 \tag{8.19}$$

for some $t_0 \in [0, T]$. The last statement is the same as: $\widetilde{H}_n(\tau) = \mathbb{H}_n(\tau)$ for all jumps τ of $\widetilde{h}_n$. Thus $\widetilde{h}_n$ is the (left) derivative of the least concave majorant of $\mathbb{H}_n$ on $[0, t_0]$, and the (right) derivative of the greatest convex minorant of $\mathbb{H}_n$ on $(t_0, T]$.

To find the LSE over $\mathcal{U}_T$, we minimize ψ_n over all possible choices of $t_0 \in [0, X_{(n)})$, where t_0 does not equal a data point. Indeed, it is sufficient to consider the N choices of intervals for t_0: $[0, X_{(1)}), (X_{(1)}, X_{(2)}), \ldots, (X_{(N-1)}, X_{(N)})$ and $(X_{(N)}, T)$, where $X_{(N)}$ is the largest data point less than or equal to T (if $X_{(N)} = T$ then $(X_{(N)}, T)$ is empty).

The maximum likelihood and least squares estimators of bathtub-shaped hazards continue to be consistent on the interior of the support of F_0 assuming that the true hazard, h_0, is also bathtub-shaped. Moreover, the results of Prakasa Rao (1970) continue to hold with only a minor adjustment to the constants (as noted by Banerjee (2007)):

Theorem 8.7. *Suppose that h_0 is bathtub shaped and that $x_0 \in (0, \infty)$ satisfy $0 < h_0(x_0) < \infty$, $h_0'(x_0) \neq 0$, and h_0' is continuous in a neighborhood of x_0. Then the nonparametric MLE and LSE are equivalent: if $\bar{h}_n = \widehat{h}_n$ or $\widetilde{h}_n$, then*

$$n^{1/3}(\bar{h}_n(x_0) - h_0(x_0)) \rightarrow_d c_1 \mathbb{Z}$$

where $\mathbb{Z}$ is the (left)-derivative of the greatest convex minorant of $W(h) + h^2$ at 0 where $W(h)$ is a two-sided Brownian motion started from 0 and where $c_1 = (h_0(x_0)|h_0'(x_0)|/(2(1 - F_0(x_0))))^{1/3}$.

8.3. **Proofs of Characterization Lemmas.**

Proof of Lemma 8.4. Our goal is first to minimize the function φ_n over the space of positive bathtub-shaped functions on $[0, X_{(n)}]$ with an antimode at t_0. We will do this in a series of steps: functional form, characterization, existence and uniqueness last.

Suppose then that $\widehat{h}_n$ is known for all the data points, $X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}$. Then to maximize the likelihood we must minimize the integral $\widehat{H}_n(X_{(i)})$ for all $i = 1, \ldots, n$. This is achieved by extending $h(X_{(i)})$, $i = 1, \ldots, n-1$ to the entire domain $[0, X_{(n)})$ in such a way so that $\widehat{h}_n$ is piecewise constant and upper semi-continuous. It follows that if a jump occurs, it must do so at a data point. In addition, if $t_0 \neq X_{(j)}$, for some j, then clearly we need $\widehat{h}_n(t_0) = 0$ to minimize $\widehat{H}_n$.

Next we consider the characterization (8.10)-(8.13). Notice that (8.13) implies that

$$0 \leq \frac{1}{n\widehat{h}_n(X_{(n)})} = \int_{[X_{(n)},\infty)} \frac{1}{\widehat{h}_n(t)} d\mathbb{F}_n(t) \leq \int_{X_{(n)}}^{\infty} \mathbb{S}_n(u)du = 0.$$

This implies that $\widehat{h}_n(X_{(n)}) = \infty$, as desired. Also, this implies that (8.13) may be re-written as

$$\int_{[x,X_{(n)})} \frac{1}{\widehat{h}_n(t)} d\mathbb{F}_n(t) \leq \int_x^{X_{(n)}} \mathbb{S}_n(u)du \tag{8.20}$$

Consider then any nonnegative convex function h in $\mathcal{U}_+(t_0)$. It follows that there exists a nonnegative constant a, and nonnegative measures ν and μ, such that

$$h(t) \;=\; a + \int_{[0,t_0)} \mathbb{I}(t \leq x)d\nu(x) + \int_{(t_0,\infty)} \mathbb{I}(t \geq x)d\mu(x). \tag{8.21}$$

For any function $\hat{h}_n$ in $\mathcal{U}_+(t_0)$ we calculate

$$\varphi_n(h) - \varphi_n(\widehat{h}_n) \;\geq\; \int_{[0,\infty)} \left\{ H(t) - \widehat{H}_n(t) + \left(1 - \frac{h(t)}{\widehat{h}_n(t)}\right) \mathbb{I}(t \neq X_{(n)}) \right\} d\mathbb{F}_n(t)$$

since $-\log x \geq 1 - x$. Plugging in the explicit form of h from above, we find that the right hand side is equal to

$$\begin{aligned} a &\left\{ \int_{[0,\infty)} \left(t - \frac{1}{\widehat{h}_n(t)} \, \mathbb{I}(t \neq X_{(n)}) \right) d\mathbb{F}_n(t) \right\} + \left\{ \frac{n-1}{n} - \int_{[0,\infty)} \widehat{H}_n(t) d\mathbb{F}_n(t) \right\} \\ &+ \int_{[0,t_0)} \left\{ \int_0^x \mathbb{S}_n(t)dt - \int_{[0,x]} \frac{1}{\widehat{h}_n(t)} \, \mathbb{I}(t \neq X_{(n)}) d\mathbb{F}_n(t) \right\} d\nu(x) \\ &\quad + \int_{(t_0,\infty)} \left\{ \int_x^{\infty} \mathbb{S}_n(t)dt - \int_{[x,\infty)} \frac{1}{\widehat{h}_n(t)} \, \mathbb{I}(t \neq X_{(n)}) d\mathbb{F}_n(t) \right\} d\mu(x). \end{aligned}$$

This is nonnegative if $\widehat{h}_n$ is a function which satisfies conditions (8.10)-(8.13), and hence also (8.20). We also note that in (8.21), we could replace the decomposition functions $\mathbb{I}(t \leq x)$ with $\mathbb{I}(t < x)$ and similarly for the increasing elbow functions (we could indeed take any combination thereof). We note, without going into the details, that these different decompositions also yield that $\varphi(h) - \varphi(\widehat{h}_n) \geq 0$ by an identical argument, for any $\widehat{h}_n$ satisfying (8.10)-(8.13). It follows that these conditions are sufficient to describe a minimizer of φ_n over $\mathcal{U}_+(t_0)$.

We next show that the conditions are necessary. To do this, we first define the directional derivative

$$\partial_\gamma \varphi_n(h) \equiv \lim_{\epsilon \to 0} \frac{\varphi_n(h+\epsilon\gamma) - \varphi_n(h)}{\epsilon} = \int_0^\infty \left\{ \Gamma(t) - \frac{\gamma(t)}{h(t)} \mathbb{I}(t \neq X_{(n)}) \right\} d\mathbb{F}_n(t).$$

If $\widehat{h}_n$ minimizes φ_n, then for any γ such that $\widehat{h}_n + \epsilon\gamma$ is in $\mathcal{U}_+(t_0)$ for sufficiently small ϵ we must have $\partial_\gamma \varphi_n(\widehat{h}_n) \geq 0$. If, however, $\widehat{h}_n \pm \epsilon\gamma$ is in $\mathcal{U}_+(t_0)$ for sufficiently small ϵ then, $\partial_\gamma \varphi_n(\widehat{h}_n) = 0$. Choosing, respectively, $\gamma(t) \equiv 1$, $\mathbb{I}(t \leq x)$ for $x < t_0$, $\mathbb{I}(t \geq x)$ for $x > t_0$ then $\widehat{h}_n + \epsilon\gamma$ is in $\mathcal{U}_+(t_0)$, and we obtain the inequalities (8.11)-(8.13).

To obtain the equalities, we note that $\widehat{h}_n \pm \gamma$ is in $\mathcal{U}_+(t_0)$ for $\gamma = \widehat{h}_n, \mathbb{I}(t \leq \tau_i)$ with $i = 1, \ldots, k$ and $\mathbb{I}(t \geq \eta_j)$ for $j = 1, \ldots, m$. These give (8.10) and the equalities in (8.12) and (8.13).

Lastly, we prove (8.14) and (8.15). For any τ_i, define

$$\gamma(t) = \begin{cases} \widehat{h}_n(t) & \text{for } t \in [0, \tau_i] \\ 0 & \text{otherwise.} \end{cases}$$

Since $(1 \pm \epsilon)\gamma$ is also in $\mathcal{U}_+(t_0)$, it follows that $\partial_\gamma \varphi_n(\widehat{h}_n) = 0$ and hence

$$0 = \left\{ \int_0^{\tau_i} \widehat{H}_n(t) d\mathbb{F}_n(t) - \mathbb{F}_n(\tau_i) + \widehat{H}_n(\tau_i)\mathbb{S}_n(\tau_i) \right\}.$$

Applying Fubini to the first term on the right-hand side gives (8.14) for $x = \tau_i$. (8.15) is obtained in a similar manner, but using

$$\gamma(t) = \begin{cases} 0 & \text{for } t \in [0, \eta_j) \\ \widehat{h}_n(t) & \text{otherwise.} \end{cases}$$

Thus,

$$\begin{aligned} 0 &= \int_{(\eta_j, \infty)} \widehat{H}_n(t) d\mathbb{F}_n(t) - \widehat{H}_n(\eta_j)\mathbb{S}_n(\eta_j) - \left(1 - \frac{1}{n} - \mathbb{F}_n(\eta_j^-)\right) \\ &\overset{\text{Fubini}}{=} \int_{\eta_j}^\infty \widehat{h}_n(s)\mathbb{S}_n(s) ds - \left(1 - \frac{1}{n}\right) + \mathbb{F}_n(\eta_j^-) \\ &\overset{(8.10)}{=} -\int_0^{\eta_j} \widehat{h}_n(s)\mathbb{S}_n(s) ds + \mathbb{F}_n(\eta_j^-). \end{aligned}$$

If we adopt the bounded approach (i.e. assume that $h \leq M$), then existence of the MLE is immediate. As we did not do this, a little more fiddling is necessary. Because the MLE must be piecewise constant and upper semi-continuous, it is enough to find the MLE only on the domain $[X_{(1)}, X_{(n-1)}]$. If we can show that we can reduce our search to functions bounded on this domain, which is a compact set, then existence will follow because the criterion function φ_n is convex. Also, φ_n is strictly convex on the data points $X_{(1)}, \ldots, X_{(n-1)}$, and because of the functional form of the constrained MLE, this gives us uniqueness of the entire function.

Recall that the minimizer must satisfy (8.10), and hence we may reduce our search to functions which satisfy this condition. For any such h, write $h = h_+ + h_-$, where h_+ is increasing and h_- is decreasing. It follows that for any x

$$1 \;\; \geq \;\; \int_0^\infty H(t) d\mathbb{F}_n(t) = \int_0^\infty h(t)\mathbb{S}_n(t)dt \geq h_-(x)\int_0^x \mathbb{S}_n(t)dt.$$

A similar bound for h_+ yields

$$h(x) \leq \frac{1}{\int_0^x \mathbb{S}_n(t)dt} + \frac{1}{\int_x^\infty \mathbb{S}_n(t)dt} \equiv M_n(x)$$

for all x in $(0, X_{(n)})$. Thus we know that $h(x)$ must be bounded for $x \in [X_{(1)}, X_{(n-1)}]$, as desired.

Lastly, we address the question of finding the unconstrained MLE. To do this, we need to check the likelihood evaluated at all of the constrained MLEs with antimode at t_0. We claim that we can reduce the search to those t_0 which do not fall on a data point, and because of the functional form of the constrained MLE, we need only check the intervals $[0, X_{(1)}), (X_{(1)}, X_{(2)}), \ldots, (X_{(n-1)}, X_{(n)})$. This is not difficult to see. Suppose that $t_0 = X_{(j)}$, for some $j \neq n$, and the resulting MLE is $\widehat{h}_n^j$. Then the likelihood can be increased by setting $\widehat{h}_n^j = 0$ on either $(X_{(j-1)}, X_{(j)})$ or $(X_{(j)}, X_{(j+1)})$. Therefore, the unconstrained MLE cannot have an antimode falling on a data point. $\square$

Proof of Lemma 8.6. We begin by showing that the LSE must be piecewise constant and upper semi-continuous. Since we are minimizing the criterion function

$$\varphi(h) = \int_0^T h^2 dt - 2\int_0^T h d\mathbb{H}_n(t),$$

over the space of U-shaped functions h. Suppose then that h and g are two bathtub shaped functions such that $h(X_i) = g(X_i)$ on all the data points. Then $\varphi(h)$ and $\varphi(g)$ differ only in the value of the first term. Therefore, if h is the smallest possible positive bathtub shaped function for $x \neq X_i$, then it will have a smaller criterion function than g. This will tell us the shape of the minimizer.

First of all, clearly h needs to be piecewise constant. Next, suppose then that for some m, h is decreasing before $X_{(m)}$ and increasing after $X_{(m+1)}$, then the smallest h we can pick is zero on $(X_{(m)}, X_{(m+1)})$. Also, this implies that h should be left-continuous before $X_{(m)}$ and right-increasing after $X_{(m+1)}$. Note that h will actually never be equal to zero on any of the observations points this way.

This shows that the overall LSE will have an antimode which does not fall on a data point. Therefore, we consider constrained LSEs with antimodes in between the data.

We next show that the conditions (8.16)-(8.19) are sufficient. That is, we show that any function which satisfies the conditions has a smaller criterion function than any other function in $\mathcal{U}_+(t_0)$. Suppose then that $\widetilde{h}_n$ satisfies (8.16)-(8.19), and consider any $h \in \mathcal{U}_+(t_0)$. As before, h may be decomposed as

$$h(t) = a + \int_0^{t_0} \mathbb{I}(t \leq x) d\nu(x) + \int_{t_0}^{T} \mathbb{I}(t \geq x) d\mu(x).$$

The inequalities in $\mathbb{I}(t \leq x)$ may be replaced with strict inequalities, and the proof will be the same. We calculate

$$\begin{aligned}
\varphi(h) - \varphi(\widetilde{h}_n) \quad &= \quad \frac{1}{2}\int_0^T (h - \widetilde{h}_n)^2 dt + \int_0^T (h - \widetilde{h}_n) d(\widetilde{H}_n - \mathbb{H}_n) \\
&\geq \quad \int_0^T (h - \widetilde{h}_n) d(\widetilde{H}_n - \mathbb{H}_n) \\
&= \quad (h - \widetilde{h}_n)(T)(\widetilde{H}_n - \mathbb{H}_n)(T) - \int_0^T (\widetilde{H}_n - \mathbb{H}_n) d(h - \widetilde{h}_n) \\
&\overset{(8.16),(8.19)}{=} \quad - \int_0^T (\widetilde{H}_n - \mathbb{H}_n) dh \\
&= \quad \int_0^{t_0} (\widetilde{H}_n - \mathbb{H}_n)(x) d\nu(x) - \int_{t_0}^{T} (\widetilde{H}_n - \mathbb{H}_n)(x) d\mu(x) \\
&\overset{(8.17),(8.18)}{\geq} \quad 0.
\end{aligned}$$

This argument shows that the conditions (8.16)-(8.19) are sufficient to describe the LSE. It also allows us to reduce our search to those functions, which in particular satisfy condition (8.16). Our next goal is to show that that the LSE over $\mathcal{U}_T(t_0)$ exists and is unique. This will follow if we can show that we can reduce our search to bounded functions on $[X_{(1)}, X_{(N)}]$ (arguing as for the MLE), since the criterion function ψ_n is *stricly* convex. Here we let N denote the largest integer such that $X_{(N)} \leq T$.

Therefore consider an h such that $H(T) = \mathbb{H}_n(T)$. Note that we may write $h = h_+ + h_-$, where h_+ is increasing and h_- is decreasing. It follows that for all x

$$\begin{aligned}
\mathbb{H}_n(T) \geq& \quad \int_0^x h_-(t)dt \geq h_-(x)x \\
\mathbb{H}_n(T) \geq& \quad \int_x^T h_+(t)dt \geq h_+(x)(T-x).
\end{aligned}$$

Hence,

$$h(x) \leq \mathbb{H}_n(T) \left\{ \frac{1}{x} + \frac{1}{T-x} \right\}.$$

It follows that h must be bounded on the set $[X_{(1)}, X_{(N)}]$, as required.

Next, we show that the conditions (8.16)-(8.19) are necessary. The proof as, always, depends on the directional derivative. In this case, we may calculate it as

$$\begin{aligned}\nabla_g \varphi(h) &= \int_0^T ghdt - \int_0^T gd\mathbb{H}_n(t) \\ &= \int_0^T gd(H - \mathbb{H}_n)(t) \\ &= g(T)(H - \mathbb{H}_n)(T) - \int_0^T (H - \mathbb{H}_n)(t)dg(t).\end{aligned}$$

The next question to ask is which basis functions can we add to h, and still keep it monotone. The answer is $\mathbb{I}(t \le x)$ for $x \in [0, t_0)$, and $\mathbb{I}(t \ge x)$ for $x \in (t_0, T]$, as well as $\gamma(t) = 1, \mathbb{I}(t < \eta_j), \mathbb{I}(t > \tau_i)$. For these choices of γ, we must have $\nabla_\gamma \varphi(\widetilde{h}_n) \ge 0$. This yields the inequalities

- $\widetilde{H}_n(x) - \mathbb{H}_n(x) \ge 0$ for $x \in [0, t_0)$
- $(\widetilde{H}_n(T) - \mathbb{H}_n(T)) - (\widetilde{H}_n(x) - \mathbb{H}_n(x)) \ge 0$ for $x \in (t_0, T]$
- $(\widetilde{H}_n(T) - \mathbb{H}_n(T)) \ge 0$
- $(\widetilde{H}_n(\eta_j) - \mathbb{H}_n(\eta_j)) \ge 0$
- $(\widetilde{H}_n(T) - \mathbb{H}_n(T)) - (\widetilde{H}_n(\tau_i) - \mathbb{H}_n(\tau_i)) \ge 0.$

We also ask which which basis functions can we add and subtract to $\widetilde{h}_n$, and still keep it monotone. The answer is $\mathbb{I}(t \le \tau_i)$, $\mathbb{I}(t \ge \eta_j)$ and $\widetilde{h}_n$ resulting in the equalities

- $\widetilde{H}_n(\tau_i) - \mathbb{H}_n(\tau_i) = 0$
- $(\widetilde{H}_n(T) - \mathbb{H}_n(T)) - (\widetilde{H}_n(\eta_j) - \mathbb{H}_n(\eta_j)) = 0$
- $\widetilde{h}_n(T)(\widetilde{H}_n - \mathbb{H}_n)(T) - \int_0^T (\widetilde{H}_n - \mathbb{H}_n)(t)d\widetilde{h}_n(t) = 0.$

To complete the proof it remains to show that $\widetilde{H}_n(T) = \mathbb{H}_n(T)$. However, since $\widetilde{h}_n(t_0) = 0$, it must be that $\widetilde{H}_n(\eta_1) = \widetilde{H}_n(\tau_k)$. From the inequalities for η_1, we have that $\widetilde{H}_n(\eta_1) \ge \mathbb{H}_n(\eta_1)$. But $\widetilde{H}_n(\eta_1) = \widetilde{H}_n(\tau_k) = \mathbb{H}_n(\tau_k)$, and hence we have $\mathbb{H}_n(\eta_1) \ge \mathbb{H}_n(\tau_k) = \widetilde{H}_n(\eta_1) \ge \mathbb{H}_n(\eta_1)$. It follows that $\widetilde{H}_n(\eta_1) = \mathbb{H}_n(\eta_1)$, and plugging this into the equality for η_1 yields the result. Therefore, we obtain the necessity of the conditions. □

9. Estimating the Rate of a Poisson Process.

Let $\{\mathbb{N}(t) : 0 \le t \le T\}$ be a Poisson process with rate function $\lambda(t)$ and cumulative intensity function (or mean function) $\Lambda(t)$. We assume that λ is convex on $[0, T]$. Denote the measure of the Poisson process (on [0,T]) as P^λ if the rate function is λ, and P is $\lambda \equiv 1$. The Radon-Nikodym derivative of P^λ with respect to P is

$$\frac{dP^\lambda}{dP} = \Pi \lambda(\tau_i) e^{-\Lambda(T)},$$

where the product is taken over all arrival times of $\mathbb{N}(t)$ on $[0, T]$. Hence, the maximum likelihood estimator of $\lambda(t)$ is the function which minimizes the negative log-likelihood given by

$$\varphi(\lambda) = -\int_0^T \log \lambda(t) d\mathbb{N}(t) + \int_0^T \lambda(t) dt. \tag{9.1}$$

We minimize this over $\mathcal{K}_T$, the set of all nonegative convex functions on $[0, T]$. The least squares estimator is found by minimizing

$$\psi(\lambda) = \frac{1}{2}\int_0^T \lambda^2(t) dt - \int_0^T \lambda(t) d\mathbb{N}(t) \tag{9.2}$$

over $\mathcal{K}_T$. Using the same arguments as in the previous sections, we may show that both of these estimators exist and are unique. Here are versions of the characterization Lemmas 2.1 and 2.3 for this Poisson process version of the problem.

Lemma 9.1 (Characterization of Poisson MLE). *Let $\{\tau_i, i = 1, \ldots, k\}$ denote all of the change points of $\widehat{\lambda}$, where $\widehat{\lambda}$ is non-increasing. Let $\{\eta_j, j = 1, \ldots, m\}$ denote all of the change points of $\widehat{\lambda}$, where $\widehat{\lambda}$ is non-decreasing.*

Then $\widehat{\lambda}$ minimizes φ over $\mathcal{K}_T$ if and only if the following conditions hold:

$$\begin{aligned}
\int_0^x \frac{x-t}{\widehat{\lambda}(t)} d\mathbb{N}(t) &\leq \frac{x^2}{2} \qquad (9.3)\\
\text{for all } x \geq 0 \quad \text{with} &\quad \text{equality at } \tau_i \text{ for } i = 1, \ldots, k;\\
\int_x^T \frac{t-x}{\widehat{\lambda}(t)} d\mathbb{N}(t) &\leq \frac{(T-x)^2}{2} \qquad (9.4)\\
\text{for all } x \geq 0 \quad \text{with} &\quad \text{equality at } \eta_j \text{ for } j = 1, \ldots, m;\\
\int_0^T \frac{1}{\widehat{\lambda}(t)} d\mathbb{N}(t) &\leq T; \qquad (9.5)\\
\widehat{\Lambda}_n(T) &= \mathbb{N}(T). \qquad (9.6)
\end{aligned}$$

Moreover,

$$\widehat{\Lambda}(x) = \mathbb{N}(x) \text{ for } x = \tau_i, \eta_j \quad i = 1, \ldots, k, j = 1, \ldots, m. \tag{9.7}$$

Let $\mathbb{Y}(t) \equiv \int_0^t \mathbb{N}(s) ds$, and $\widetilde{\mathcal{L}}(t) = \int_0^t \widetilde{\Lambda}(s) ds$.

Lemma 9.2 (Characterization of Poisson LSE)**.** *The function $\widetilde{\lambda}$ minimizes $\psi(\lambda)$ over $\mathcal{K}_T$ if and only if it satisfies*

$$\widetilde{\Lambda}(T) = \mathbb{N}(T), \tag{9.8}$$

$$\widetilde{\mathcal{L}}(T) = \mathbb{Y}(T), \tag{9.9}$$

$$\widetilde{\mathcal{L}}(t) \geq \mathbb{Y}(t) \text{ for all } t \in [0, T], \tag{9.10}$$

$$\int_0^T (\widetilde{\mathcal{L}} - \mathbb{Y}_n)(t) d\widetilde{\lambda}'(t) = 0. \tag{9.11}$$

The last statement is the same as: $\widetilde{\mathcal{L}}(\tau) = \mathbb{Y}(\tau)$ for all changes of slope τ of $\widetilde{\lambda}$.

To study the asymptotics of these estimators, assume that we observe n independent and identically distributed Poisson processes $\{\mathbb{N}_i(\cdot)\}_{i=1}^n$ with the same distribution as $\mathbb{N}$ above. Let $\mathbb{C}_n(\cdot)$ denote the (pointwise) average of these. Here, the maximum likelihood estimator of λ is the function $\widehat{\lambda}_n$ which minimizes

$$\varphi(\lambda) = -\int_0^T \log \lambda(t) d\mathbb{C}_n(t) + \int_0^T \lambda(t) dt.$$

The least squares estimator, $\widetilde{\lambda}_n$, is found by minimizing

$$\psi(\lambda) = \frac{1}{2}\int_0^T \lambda^2(t)dt - \int_0^T \lambda(t) d\mathbb{C}_n(t)$$

over $\mathcal{K}_T$. The characterization is of course the same, except with $\mathbb{N}$ replaced by $\mathbb{C}_n$ in Lemmas 9.1 and 9.2. Let λ_0 denote the true rate function and assume that $\Lambda_0(T) < \infty$. Then the LSE $\widetilde{\lambda}_n$ is consistent: $\widetilde{\lambda}_n(t)$ converges to $\lambda_0(t)$ for $t \in (0, T)$ with probability one. Also, for all $\delta > 0$

$$\sup_{t\in[\delta,T-\delta]} |\widetilde{\lambda}_n(t) - \lambda_0(t)| \rightarrow_{a.s.} 0.$$

The same is true of the MLE $\widehat{\lambda}_n$. Lastly, their asymptotics may again be described in terms of the invelope as for hazard estimators:

Theorem 9.3. *Suppose that λ_0 is convex and that x_0 is a point such that $0 < \lambda(x_0) < \infty$, $\lambda_0''(x_0) > 0$, and that $\lambda_0''(\cdot)$ is continuous in a neighborhood of x_0 (also that $x_0 < T$ for the LSE). Then the nonparametric maximum likelihood estimator and least squares estimator are asymptotically equivalent in the following sense: if $\bar{\lambda}_n = \widehat{\lambda}_n$ or $\widetilde{\lambda}_n$, then*

$$\begin{pmatrix} n^{2/5}(\bar{\lambda}_n(x_0) - \lambda_0(x_0)) \\ n^{1/5}(\bar{\lambda}_n'(x_0) - \lambda_0'(x_0)) \end{pmatrix} \rightarrow_d \begin{pmatrix} c_1\, \mathcal{I}^{(2)}(0) \\ c_2\, \mathcal{I}^{(3)}(0) \end{pmatrix}$$

where $\mathcal{I}^{(2)}(0)$ and $\mathcal{I}^{(3)}(0)$ are the second and third derivatives at 0 of the invelope of $Y(t) \equiv \int_0^t W(s)ds + t^4$, and where

$$c_1 = \left(\frac{\lambda_0^2(x_0)\lambda_0''(x_0)}{24}\right)^{1/5}, \qquad c_2 = \left(\frac{\lambda_0(x_0)\lambda_0''(x_0)^3}{24^3}\right)^{1/5}.$$

10. Consistency of antimode and (local) inverse hazards

The following results are stated for the convex hazard estimation problem. However, they remain valid for both the hazard under right censoring, and for the Poisson intensity estimators.

Theorem 10.1. *In the case of the LSE let $S = [0,T]$, while for the MLE let $S = \operatorname{supp} f_0$. Suppose that h_0 has a unique minimum located in S°. Then the sample antimode converges in probability to the true antimode. That is*

$$\operatorname{argmin} \widehat{h}_n \to_p \operatorname{argmin} h_0,$$

and the same holds for the LSE.

Proof. Suppose that h_0 has a unique minimum in S°. Then there exists a compact set $K = [a,b] \subset (0,T)^\circ$ such that $\inf_S h_0 = \inf_K h_0$.

We also have that for all $x \in K$, and any functions f, g.

$$\begin{aligned} \inf_{x \in K} f(x) \le f(x) &= f(x) - g(x) + g(x) \\ &\le \sup_{x \in K} |f(x) - g(x)| + g(x). \end{aligned}$$

It follows that

$$\inf_{x \in K} f(x) - \inf_{x \in K} g(x) \le \sup_{x \in K} |f(x) - g(x)|.$$

By symmetry of the argument, it follows

$$|\inf_{x \in K} f(x) - \inf_{x \in K} g(x)| \le \sup_{x \in K} |f(x) - g(x)|. \tag{10.1}$$

There exists a $\delta > 0$ such that

$$\inf_{x \in [0,a]} h_0(x) = h_0(a) \ge \delta + \inf h_0(x).$$

There also exists an $n_0 > 0$ such that for all $n \ge n_0$

$$|\widehat{h}_n(a) - h_0(a)| < \delta/16 < \delta/2, \quad \text{and} \quad |\widehat{h}_n(m) - h_0(m)| \le \delta/16$$

where $m = \operatorname{argmin} h_0(x)$. Therefore,

$$\begin{aligned} \widehat{h}_n(a) &> h_0(a) - \delta/2 \\ &\ge \delta/2 + h_0(m). \end{aligned}$$

But then for all $n \geq n_0$

$$\begin{aligned} h_0(m) + \delta/4 &\leq h_0(a) \\ \Rightarrow \quad \widehat{h}_n(m) - \delta/16 + \delta/4 &\leq \widehat{h}_n(a) + \delta/16 \\ \Rightarrow \quad \widehat{h}_n(m) + \delta/8 \leq \widehat{h}_n(a). \end{aligned}$$

By a similar argument $\widehat{h}_n(m) + \delta/8 \leq \widehat{h}_n(b)$, and hence $\operatorname{argmin} \widehat{h}_n(x) \in K$. By 10.1 and consistency of $\widehat{h}_n$, it hence follows that

$$\inf \widehat{h}_n(x) \to \inf h_0(x) \text{ a.s.}$$

The same argument may be used to show that

$$\inf_F \widehat{h}_n(x) \to \inf_F h_0(x) \text{ a.s.},$$

for any closed set $F \subset S^\circ$.

We next show that $\widehat{m}_n = \operatorname{argmin} \widehat{h}_n$ converges to m a.s. To do this, consider any open set G such that $m \in G$. Since m is unique, there exists a $\delta > 0$ such that

$$\inf h_0(x) + \delta \leq \inf_{x \in G^c} h_0(x).$$

It therefore follows that there exists an n_0 such that for all $n \geq n_0$

$$\inf \widehat{h}_n(x) - \delta/4 + \delta \leq \inf_{x \in G^c} \widehat{h}_n(x) + \delta/4.$$

This of course implies that $\widehat{m}_n \in G$ for all $n \geq n_0$. □

Motivated by the notion of burn-in (see e.g. Lynn and Singpurwalla (1997)) and earthquake alerts (e.g. Ellis (1985); La Rocca (2008)), we also consider convergence of estimators of the (local) inverse of the hazard function h_0. That is, if h_0 is decreasing near zero, we may be interested in estimating the first time that h_0 hits a level α. Likewise, if h_0 is increasing on some interval, we may be interested in the last time that h_0 hits a level α.

In what follows, we discuss the MLE $\widehat{h}_n$, however, the statements and results are equally valid for the LSE $\widetilde{h}_n$. Also, to simplify notation, we assume that h_0 is strictly decreasing, and hence we study the estimation of $h_0^{-1}(\alpha)$. The results are clearly valid in the more general setting.

Proposition 10.2. *Let $x_0 = h_0^{-1}(\alpha)$. Suppose that $x_0 \in S^\circ$, and that h_0 is strictly decreasing near x_0. Then $\widehat{x}_n = \widehat{h}_n^{-1}(\alpha)$ converges to x_0 almost surely.*

Proof. We first argue that the estimator $\widehat{x}_n$ makes sense, at least for large enough n. This follows from the consistency of the derivatives of the estimators (Corollary 4.8), and convexity.

Now, fix $\delta > 0$, and note that there exists an $\epsilon > 0$ such that

$$h_0(x_0 - \delta) \geq \alpha + \epsilon \geq \alpha - \epsilon \geq h_0(x_0 + \delta).$$

From consistency of $\widehat{h}_n$, there exists an $n_0 = n_0(\omega)$ such that for all $n \geq n_0$ and all ω in a set with probability 1

$$\begin{aligned}\widehat{h}_n(x_0 - \delta) &\geq h_0(x_0 - \delta) - \epsilon/2 \\ &\geq \alpha + \epsilon/2,\end{aligned}$$

and similarly

$$\alpha - \epsilon/2 \geq \widehat{h}_n(x_0 + \delta),$$

almost surely. Therefore,

$$\widehat{h}_n(x_0 - \delta) - \epsilon/2 \geq \alpha \geq \widehat{h}_n(x_0 + \delta) + \epsilon/2.$$

Hence, for any $\delta > 0$, $\widehat{x}_n \in (x_0 - \delta, x_0 + \delta)$ almost surely, proving the result. □

References

Banerjee, M. (2007). Estimating monotone, unimodal and U-shaped failure rates using asymptotic pivots. *Statistica Sinica* To appear.

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

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

Birgé, L. (1997). Estimation of unimodal densities without smoothness assumptions. *Ann. Statist.* **25** 970–981.

Bray, T., Crawford, G. and Proschan, F. (1967a). Maximum likelihood estimation of a failure rate function initially increasing and then decreasing. In *Proceedings 1967 Spring Seminar on Reliability, Maintainability, IEEE.*

Bray, T., Crawford, G. and Proschan, F. (1967b). Maximum likelihood estimation of a U-shaped failure rate function. *Mathematical Note* **534**. Mathematics Research Laboratory, Boeing Scientific Research Laboratories, Seattle, WA; available at `http://www.stat.washington.edu/jaw/RESEARCH/OLD-PAPERS-OTHERS/UMLE.pdf`.

di Lavoro MPS, G. (2004). Catalogo parametrico dei terremoti italiani, versione 2004 (cpti04). INGV, Bologna, http://emidius.mi.ingv.it/CPTI. In Italian.

Dykstra, R. L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9** 356–367.

Ellis, S. (1985). An optimal statistical decision rule for calling earthquake alerts. *Earthquake Prediction Research* **3** 1–10.

Grenander, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153 (1957).

Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2001a). A canonical process for estimation of convex functions: the "invelope" of integrated Brownian motion $+t^4$. *Ann. Statist.* **29** 1620–1652.

Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2001b). Estimation of a convex function: characterizations and asymptotic theory. *Ann. Statist.* **29** 1653–1698.

Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2007). The support reduction algorithm for computing nonparametric function estimates in mixture models. Technical report, Vrije Universiteit, Amsterdam.

Haupt, E. and Schäbe, H. (1997). The TTT transformation and a new bathtub distribution model. *J. Statist. Plann. Inference* **60** 229–240.

Ho, M.-W. (2007). A bayes method for a bathtub failure rate via two s-paths. Technical report, National University of Singapore. Available at http://arxiv.org/abs/0707.2257v1.

Huang, J. and Wellner, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scand. J. Statist.* **22** 3–33.

Huang, Y. and Zhang, C.-H. (1994). Estimating a monotone density from censored observations. *Ann. Statist.* **22** 1256–1274.

Jankowski, H. and Wellner, J. A. (2008). Nonparametric estimation of convex hazard functions and the support reduction algorithm. Tech. rep., University of Washington, Department of Statistics. In preparation.

Jongbloed, G. (1995). *Three Statistical Inverse Problems*. Ph.D. thesis, Delft University of Technology.

Jongbloed, G. (2000). Minimax lower bounds and moduli of continuity. *Statist. Probab. Lett.* **50** 279–284.

La Rocca, L. (2007). *BayHaz: R functions for Bayesian hazard rate estimation*. R package version 0.0-6.
URL `http://www.r-project.org, http://www-dimat.unipv.it/luca`

La Rocca, L. (2008). Bayesian non-parametric estimation of smooth hazard rates for seismic hazard assessment. *Scandinavian Journal of Statistics* **35** 524 – 539.

Lai, C. D., Xie, M. and Murthy, N. P. (2001). Bathtub-shaped failure rate life distributions. In *Advances in Reliability* (N. Balakrishnan and C. R. Rao, eds.), vol. 20 of *Handbook of Statistics*. North-Holland Publishing Co., Amsterdam, 69–104.

Lynn, N. and Singpurwalla, N. (1997). Comment: Burn-in makes us feel good. *Statistical Science* **12** 13–19.

Marshall, A. W. and Olkin, I. (2007). *Life Distributions, Structure of Nonparametric, Semiparametric, and Parametric Families*. Springer-Verlag, New York. Springer Series in Statistics.

Marshall, A. W. and Proschan, F. (1965). Maximum likelihood estimation for distributions with monotone failure rate. *Ann. Math. Statist* **36** 69–77.

Massart, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283.

Mykytyn, S. W. and Santner, T. J. (1981). Maximum likelihood estimation of the survival function based on censored data under hazard rate assumptions. *Comm. Statist. A—Theory Methods* **10** 1369–1387.

Padgett, W. J. and Wei, L. G. (1981). A Bayesian nonparametric estimator of survival probability assuming increasing failure rate. *Comm. Statist. A—Theory Methods* **10** 49–63.

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

Rajarshi, S. and Rajarshi, M. B. (1988). Bathtub distributions: A review. *Communications in Statistics: Theory and Methods* **17** 2597–2621.

Reboul, L. (2005). Estimation of a function under shape restrictions. Applications to reliability. *Ann. Statist.* **33** 1330–1356.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order restricted statistical inference*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Ltd., Chichester.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J.

Royden, H. L. (1988). *Real Analysis*. 3rd ed. Macmillan Publishing Company, New York.

Savits, T. H. (2003). Preservation of generalized bathtub-shaped functions. *J. Appl. Probab.* **40** 473–484.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.

Singpurwalla, N. D. (2006). *Reliability and Risk: A Bayesian Perspective*. Wiley Series in Probability and Statistics, John Wiley & Sons Ltd., Chichester.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications in Statistics*. Springer, New York.

Department of Mathematics and Statistics, York University
e-mail: hkj@mathstat.yorku.ca

Department of Statistics, University of Washington
e-mail: jaw@stat.washington.edu