CONSISTENCY OF THE GMLE WITH MIXED CASE INTERVAL-CENSORED DATA

By Anton Schick and Qiqing Yu

Binghamton University

April 1997. Revised December 1997, Revised July 1998

Abstract. In this paper we consider an interval censorship model in which the endpoints of the censoring intervals are determined by a two stage experiment. In the first stage the value k of a random integer is selected; in the second stage the endpoints are determined by a case k interval censorship model. We prove the strong consistency in the $L_1(\mu)$-topology of the nonparametric maximum likelihood estimate of the underlying survival function for a measure μ which is derived from the distributions of the endpoints. This consistency result yields strong consistency for the topologies of weak convergence, pointwise convergence and uniform convergence under additional assumptions. These results improve and generalize existing ones in the literature.

Short Title: Interval censorship model.
AMS 1991 Subject Classification: Primary 62G05; Secondary 62G20.
Key words and phrases: Nonparametric maximum likelihood estimation; current status data; case k interval-censorship model.

1. Introduction

In industrial life testing and medical research, one is frequently unable to observe the random variable X of interest directly, but can observe a pair (L, R) of extended random variables such that

$$-\infty \leq L < X \leq R \leq \infty.$$

For example consider an animal study in which a mouse has to be dissected to check whether a tumor has developed. At the time of dissection we can only infer whether the tumor is present, or has not yet developed. Thus, if we let X denote the onset of tumor and Y the time of the dissection, then the corresponding pair (L, R) is given by

$$(L, R) = \begin{cases} (-\infty, Y), & X \leq Y, \\ (Y, \infty), & X > Y. \end{cases}$$

If X and Y are independent, then this model is called the case 1 interval censorship model (Groeneboom and Wellner (1992)) and the data pair (L, R) is usually replaced by the current status data

$(Y, I[X \leq Y])$, where $I[A]$ is the indicator function of the set A. Examples of the current status data are mentioned in Ayer et al. (1955), Keiding (1991) and Wang and Gardiner (1996).

Another interval censorship model is the case 2 model considered by Groeneboom and Wellner (1992). Consider an experiment with two inspection times U and V such that $U < V$ and (U, V) is independent of X. One can only determine whether X occurs before time U, between times U and V or after time V. More formally, one observes the random vector $(U, V, I[X \leq U], I[U < X \leq V])$. In this model

$$(L, R) = \begin{cases} (-\infty, U), & X \leq U, \\ (U, V), & U < X \leq V, \\ (V, \infty), & X > V. \end{cases}$$

Note that (L, R) is a function of the random vector $(U, V, I[X \leq U], I[U < X \leq V])$. However, V cannot be recovered from the pair (L, R) on the event $\{X \leq U\}$. Thus the pair (L, R) carries less information than the vector $(U, V, I[X \leq U], I[U < X \leq V])$.

The case 1 and case 2 models are special cases of the case k model (Wellner, 1995). In this model there are k inspection times $Y_1 < \cdots < Y_k$ which are independent of X, and one observes into which of the random intervals $(-\infty, Y_1], \ldots, (Y_k, \infty)$ the random variable X belongs. Note that the case k model for $k > 2$ can be formally reduced to a case 2 model with U and V functions of X and the inspection times $Y_1, \ldots, Y_k$. The resulting U and V are then no longer independent of X violating a key assumption used in deriving consistency results for the case 2 model.

While the case 1 model gives a good description of the animal study mentioned above, a data set from a case k model ($k \geq 2$) is difficult to find in medical research since it is very unlikely that every patient under study has exactly the same number of visits. Finkelstein and Wolfe (1985) presented a closely related type of interval-censored data in comparing two different treatments for breast cancer patients. The censoring intervals arose in the follow-up studies for patients treated with radiotherapy and chemotherapy. The failure time X is the time until cosmetic deterioration as determined by the appearance of breast retraction. Each patient had several follow-ups and the number of follow-ups differed from patient to patient. One only knows that the failure time occurred either before the first follow-up, or after the last follow-up or between two consecutive follow-ups. Other examples of such type of interval-censored data can be found in AIDS studies (Becker and Melbye (1991); Aragon and Eberly (1992)).

In this paper we assume that the pair (L, R) is generated as a mixture of case k models. This formulation encompasses the various case k models and the data setting occurring in Finkelstein and Wolfe (1985). A precise definition of this mixture model is given in Section 2.

Let F_0 denote the unknown distribution function of X. This distribution function is commonly estimated by the generalized maximum likelihood estimate (GMLE). Ayer et al. (1955) derived an explicit expression of the GMLE for the case 1 model. However, in general the GMLE does not have an explicit solution. In deriving a numerical solution for the GMLE, Peto (1973) used the Newton-Raphson algorithm; Turnbull (1976) proposed a self-consistent algorithm; Groeneboom and Wellner (1992) proposed an iterative convex minorant algorithm. A detailed discussion of some computational aspects is given in Wellner and Zhan (1997).

Various consistency results are available for the GMLE. In the case 1 model, Ayer et al. (1955) proved the weak consistency of the GMLE at continuity points of F_0 under additional assumptions on G, the distribution function of Y. The uniform strong consistency of the GMLE has been established by Groeneboom and Wellner (1992), van de Geer (1993, Example 3.3a), Wang and Gardiner (1996) and Yu et al. (1998a) for continuous F_0 using various assumptions and techniques. In the case 2 model, the uniform strong consistency of the GMLE has been established by Groeneboom and Wellner (1992), van de Geer (1993, Example 3.3b), and Yu et al. (1998b) for continuous F_0.

In Section 2 we shall obtain the strong $L_1(\mu)$-consistency of the GMLE for our mixture of case k models for some measure μ. This result shows that the $L_1(\mu)$-topology is the appropriate topology as it gives consistency without additional assumptions in the case k models. Convergence in stronger topologies such as the topologies of weak convergence and uniform convergence requires additional conditions. This is pursued in Section 3. In the process we also point out some erroneous consistency claims in the literature. The proof of the $L_1(\mu)$-consistency is given in Section 4. It exploits the special structure of the likelihood for this model and does not require any advanced theory. Section 5 collects various other proofs.

2. Main Results

We begin by giving a precise definition of our model. This is done by describing how the endpoints L and R are generated. Let K be a positive random integer and $\mathbf{Y} = \{Y_{k,j} : k = 1, 2, \ldots, j = 1, \ldots, k\}$ be an array of random variables such that $Y_{k,1} < \cdots < Y_{k,k}$. Assume throughout that $(K, \mathbf{Y})$ and X are independent. On the event $\{K = k\}$, let (L, R) denote the endpoints of that random interval among $(-\infty, Y_{k,1}], (Y_{k,1}, Y_{k,2}], \ldots, (Y_{k,k}, \infty)$ which contains X. We refer to this model as the *mixed case* model as it can be viewed as a mixture of the various case k models.

In some clinical studies, an examination is performed at the start of the study and follow-ups are scheduled one at a time till the end of the study. This can be modeled by taking $Y_{k,j} = \sum_{i=1}^{j-1} \xi_i$ and $K = \sup\{k \geq 1 : \sum_{i=1}^{k-1} \xi_i \leq \tau\}$, where $\xi_1, \xi_2, \ldots$ denote the (positive) inter-follow-up times and τ is the length of the study. In this case K may not be bounded. For example, if the inter-follow-up times are independent with a common exponential distribution, then $K - 1$ is a Poisson random variable; thus K is unbounded, yet $E(K) < \infty$. In general, if the inter-follow-up times are independent and identically distributed, then $E(K) < \infty$.

To define the GMLE, let $(L_1, R_1), \ldots, (L_n, R_n)$ be independent copies of the pair (L, R) defined above and define the generalized likelihood function Λ_n by

$$\Lambda_n(F) = \prod_{j=1}^{n} [F(R_j) - F(L_j)], \quad F \in \mathcal{F},$$

where $\mathcal{F}$ is the collection of all nondecreasing functions F from $[-\infty, +\infty]$ into $[0, 1]$ with $F(-\infty) = 0$ and $F(+\infty) = 1$. We think of F_0 as a member of $\mathcal{F}$. Note that $\Lambda_n(F)$ depends on F only through the values of F at the points L_j or R_j, $j = 1, \ldots, n$. Thus there exists no unique maximizer of $\Lambda_n(F)$ over the set $\mathcal{F}$. However, there exists a unique maximizer $\hat{F}_n$ over the set $\mathcal{F}$ which is right continuous and piecewise constant with possible discontinuities only at the observed values of L_j and R_j, $j = 1, \ldots, n$. We call this maximizer $\hat{F}_n$ the GMLE of F_0.

Define a measure μ on the Borel σ-field $\mathcal{B}$ on $\mathbb{R}$ by

$$\mu(B) = \sum_{k=1}^{\infty} P(K = k) \sum_{j=1}^{k} P(Y_{k,j} \in B \mid K = k), \qquad B \in \mathcal{B}.$$

We are now ready to state our main result, namely the (strong) $L_1(\mu)$ consistency of the GMLE.

2.1. Theorem. *Let $E(K) < \infty$. Then $\int |\hat{F}_n - F_0| \, d\mu \to 0$ almost surely.*

The condition $E(K) < \infty$ implies the finiteness of the measure μ and of the expectation $E[\log(F_0(R) - F_0(L))]$. These two latter conditions play an important role in our proof given in Section 4.

One referee pointed out that results of van de Geer's (1993) (namely her Lemma 1.1 and Theorem 3.1) may be used to prove a result very similar to our Theorem 2.1 with the help of some inequalities suggested by this referee. This alternative proof leads to $L_1(\tilde{\mu})$-consistency for some finite measure $\tilde{\mu}$ that is equivalent to our measure μ and does not require the finiteness of $E(K)$. Actually, such a result implies our result in view of the following simple lemma which we state without a proof.

2.2. Lemma. *Let μ_1 and μ_2 be two finite measures and $g, g_1, g_2, \ldots$ be measurable functions into $[0, 1]$. Suppose that μ_2 is absolutely continuous with respect to μ_1. Then $\int |g_n - g| \, d\mu_1 \to 0$ implies $\int |g_n - g| \, d\mu_2 \to 0$.*

We have decided to present our original proof since it is direct and elementary and since $E(K) < \infty$ is a rather mild assumption that is typically satisfied in applications.

In the remainder of this section we mention some corollaries of Theorem 2.1. The first one is of interest when the inspection times are discrete. It follows from the fact that $\mu(\{a\})|\hat{F}_n(a) - F_0(a)| \leq \int |\hat{F}_n - F_0| \, d\mu$ for every $a \in \mathbb{R}$ and generalizes the consistency results given in Yu et al. (1998a,b) for the case 1 and case 2 models with discrete inspection times.

2.3. Corollary. *Let $E(K) < \infty$. Then $\hat{F}_n(a) \to F_0(a)$ almost surely for each point a with $\mu(\{a\}) > 0$.*

In the next corollary we state results for a measure ν that depends on the distribution of L and R and is easier to interpret than μ. We take ν to be the sum of the marginal distributions of L and R:

$$\nu(B) = P(L \in B) + P(R \in B), \quad B \in \mathcal{B}.$$

In view of the set inclusion

$$\{L \in B\} \cup \{R \in B\} \subset \bigcup_{k=1}^{\infty} \bigcup_{i=1}^{k} \{K = k, Y_{k,i} \in B\},$$

we have $\nu(B) \le 2\mu(B)$. Thus we immediately get the following corollary.

2.4. Corollary. *Let $E(K) < \infty$. Then the following are true.*

(1) $\int |\hat{F}_n - F_0| \, d\nu \to 0$ *almost surely.*
(2) $\hat{F}_n(a) \to F_0(a)$ *almost surely for each point a with $\nu(\{a\}) > 0$.*

3. Other Consistency Results

In this section we shall show that under additional assumptions strong $L_1(\mu)$-consistency implies strong consistency in other topologies such as the topologies of weak convergence, pointwise convergence and uniform convergence. Throughout we always assume that $E(K)$ is finite so that μ is a finite measure and $P(\Omega_\mu) = 1$ by Theorem 2.1, where

$$\Omega_\mu = \{\lim_{n\to\infty} \int |\hat{F}_n - F_0| \, d\mu = 0\}.$$

Although the results of this section are formulated for the measure μ defined in the previous section, they are true for any finite measure for which the GMLE is strongly L_1-consistent as only the finiteness of μ and $P(\Omega_\mu) = 1$ are used in their proofs. These proofs are deferred to Section 5.

Let a be a real number. We call a a *support point of* μ if $\mu((a-\epsilon, a+\epsilon)) > 0$ for every $\epsilon > 0$. We call a *regular* if $\mu((a-\epsilon, a]) > 0$ and $\mu([a, a+\epsilon)) > 0$ for all $\epsilon > 0$. We call a *strongly regular* if $\mu((a-\epsilon, a)) > 0$ and $\mu([a, a+\epsilon)) > 0$ for all $\epsilon > 0$. We call a a *point of increase of* F_0 if $F_0(a+\epsilon) - F_0(a-\epsilon) > 0$ for each $\epsilon > 0$.

In view of the inequality $\nu \le 2\mu$, sufficient conditions for the first three of the above concepts are obtained by replacing μ be ν. As these sufficient conditions are in terms of the distribution of L and R, they are easier to interpret and thus more meaningful from an applied point of view.

Ayer et al. (1955) established the weak consistency of the GMLE at regular continuity points of F_0 in the case 1 model. Our first proposition gives a strong consistency result for regular continuity points in our more general model.

3.1. Proposition. *For each $\omega \in \Omega_\mu$ and each regular continuity point a of F_0, $\hat{F}_n(a;\omega) \to F_0(a)$.*

The next two propositions address weak convergence on an open interval and on the entire line.

3.2. Proposition. *Suppose every point in an open interval (a, b) is a support point of μ. Then $\hat{F}_n(x;\omega) \to F_0(x)$ for every continuity point x of F_0 in (a, b) and every $\omega \in \Omega_\mu$. If also $F_0(a) = 0$ and $F_0(b-) = 1$, then $\hat{F}_n(x;\omega) \to F_0(x)$ for all continuity points x of F_0 and all $\omega \in \Omega_\mu$.*

3.3. Proposition. *If every point of increase of F_0 is strongly regular, then $\hat{F}_n(x;\omega) \to F_0(x)$ for all continuity points of F_0 and all $\omega \in \Omega_\mu$.*

Combining these propositions with Corollary 2.3 yields the following results on pointwise convergence on open intervals and on the entire line.

3.4. Corollary. *Suppose every point x in an open interval (a, b) is a support point of μ and satisfies $\mu(\{x\}) > 0$ if x is a discontinuity point of F_0. Then $\hat{F}_n(x;\omega) \to F_0(x)$ for every x in (a, b) and every $\omega \in \Omega_\mu$. Moreover, if $F_0(a) = 0$ and $F_0(b-) = 1$, then $\hat{F}_n(x;\omega) \to F_0(x)$ for all $x \in \mathbb{R}$ and all $\omega \in \Omega_\mu$.*

3.5. Corollary. *If every point of increase of F_0 is strongly regular and if $\mu(\{a\}) > 0$ for each discontinuity point a of F_0, then $\hat{F}_n(x;\omega) \to F_0(x)$ for all $x \in \mathbb{R}$ and all $\omega \in \Omega_\mu$.*

The next proposition addresses uniform convergence.

3.6. Proposition. *Suppose that F_0 is continuous and that, for all $a < b$, $0 < F_0(a) < F_0(b) < 1$ implies $\mu((a, b)) > 0$. Then the GMLE is uniformly strongly consistent, i.e.,*

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F_0(x)| \to 0 \quad \text{a.s..}$$

This proposition generalizes the strong uniform consistency results given by Groeneboom and Wellner (1992) for the case 1 and 2 models. In the case 1 model they require that F_0 and G, the distribution function of Y, are continuous and that the probability measure μ_{F_0} induced by F_0 is absolutely continuous with respect to μ ($\mu_{F_0} << \mu$). Proposition 3.6 does not require the continuity of G and weakens the absolute continuity requirement. In the case 2 model Groeneboom and Wellner assume that F_0 is continuous and that the joint distribution of U and V has a Lebesgue density g such that $g(u, v) > 0$ if $0 < F_0(u) < F_0(v) < 1$. Their assumption implies that the measure μ has a Lebesgue density which is positive on the set $\{t : 0 < F_0(t) < 1\}$ and therefore implies that $\mu((a, b)) > 0$ if $0 < F_0(a) < F_0(b) < 1$. Consequently, Proposition 3.6 improves and generalizes their result.

Proposition 3.6 also generalizes the strong uniform consistency results given by van de Geer (1993) for the case 1 and 2 models under the assumption that F_0 is continuous and $\mu_{F_0} << \mu$. The latter implies that $\mu((a, b)) > 0$ if $0 < F_0(a) < F_0(b) < 1$. However, if μ is discrete, its support is dense in $(0, +\infty)$, and F_0 is exponential, then the assumption in Proposition 3.6 is satisfied, but $\mu_{F_0} << \mu$ is not true.

In clinical follow-ups, the studies typically last for a certain period of time, say $[\tau_1, \tau_2]$. It is often that $F_0(\tau_2) < 1$ in which case the conditions in Proposition 3.6 are not satisfied. In this regard, Gentleman and Geyer (1994) claimed a vague convergence result in their Theorem 2 and Huang (1996) claimed a uniform strong consistency result in his Theorem 3.1. Both of their results as stated imply the uniform strong consistency of the GMLE on $[\tau_1, \tau_2]$ in the case 1 model, if F_0

is continuous and the inspection time Y is uniformly distributed on $[\tau_1, \tau_2]$. The following example shows that this is not true.

3.7. Example. Consider current status data $(Y_1, I[X_1 \le Y_1]), \ldots, (Y_n, I[X_n \le Y_n])$, where the survival times $X_1, \ldots, X_n$ are uniformly distributed on $[0, 3]$ and the inspection times $Y_1, \ldots, Y_n$ are uniformly distributed on $[1, 2]$. Then F_0 is the uniform distribution function on $[0, 3]$ and μ is the uniform distribution on $[1, 2]$. Note that on the event $\bigcup_{j=1}^n \{X_j > 2 > Y_j, Y_j < Y_i, i = 1, \ldots, n, i \ne j\}$ we have $\hat{F}_n(1) = 0$, and on the event $\bigcup_{j=1}^n \{X_j \le 1 \le Y_j, Y_j > Y_i, i = 1, \ldots, n, i \ne j\}$ we have $\hat{F}_n(2) = \hat{F}_n(2-) = 1$. Both events have probability $1/3$. Since $F_0(1) = 1/3$ and $F_0(2) = F_0(2-) = 2/3$, we see that $\hat{F}_n(x)$ does not converge to $F_0(x)$ almost surely for $x = 1, 2$ and $\hat{F}_n(2-)$ does not converge to $F_0(2-)$ almost surely. This shows that pointwise convergence on the closed interval $[\tau_1, \tau_2]$ to a continuous F_0 is not implied by the condition: $\mu([a, b]) > 0$ for all a and b such that $\tau_1 \le a < b \le \tau_2$.

The following proposition indicates how to fix the assumptions.

3.8. Proposition. *Suppose the following four conditions hold for real numbers $\tau_1 < \tau_2$.*

(1) *F_0 is continuous at every point in the interval $(\tau_1, \tau_2]$;*
(2) *either $\mu(\{\tau_1\}) > 0$ or $F_0(\tau_1) = 0$;*
(3) *either $\mu(\{\tau_2\}) > 0$ or $F_0(\tau_2-) = 1$;*
(4) *for all a and b in (τ_1, τ_2), $0 < F_0(a) < F_0(b) < 1$ implies $\mu((a, b)) > 0$.*

Then the GMLE is uniformly strongly consistent on $[\tau_1, \tau_2]$, i.e.,

$$\sup_{x \in [\tau_1, \tau_2]} |\hat{F}_n(x) - F_0(x)| \to 0 \quad \text{a.s..}$$

4. Proof of Theorem 2.1

Recall that L may take the value $-\infty$ and R the value $+\infty$. The normalized log-likelihood is

$$\mathcal{L}_n(F) = \frac{1}{n} \sum_{j=1}^n \log\, [F(R_j) - F(L_j)], \quad F \in \mathcal{F}.$$

By the strong law of large numbers (SLLN), $\mathcal{L}_n(F)$ converges almost surely to its mean

$$\mathcal{L}(F) = E(\log\, [F(R) - F(L)]) = \sum_{k=1}^\infty P(K = k) E(h_{F,k}(Y_{k,1}, \ldots, Y_{k,k}) \mid K = k),$$

where

$$h_{F,k}(y_1, \ldots, y_k) = \sum_{j=0}^k (F_0(y_{j+1}) - F_0(y_j)) \log(F(y_{j+1}) - F(y_j)),$$

for $-\infty = y_0 < y_1 < \cdots < y_k < y_{k+1} = \infty$. Here and below we interpret $0 \log 0 = 0$ and $\log 0 = -\infty$.

It is easy to check that, for each positive integer k and real numbers $y_1 < \cdots < y_k$, the expression $h_{F,k}(y_1, \ldots, y_k)$ is maximized by a function $F \in \mathcal{F}$ if and only if $F(y_j) = F_0(y_j)$ for $j = 1, \ldots, k$. Since $\sup\{|p \log p| : 0 \leq p \leq 1\} < 1$, $|h_{F_0,k}|$ is bounded by k. Since K has finite expectation, we see that $\mathcal{L}(F_0)$ is finite. Hence F_0 maximizes $\mathcal{L}(\cdot)$ over the set $\mathcal{F}$ and any other function $F \in \mathcal{F}$ that maximizes $\mathcal{L}(\cdot)$ satisfies that $F = F_0$ a.e. μ.

Let $\{F_n\}$ be a sequence in $\mathcal{F}$. By a pointwise limit of this sequence we mean an $F \in \mathcal{F}$ such that $F_{n'}(x) \to F(x)$ for all $x \in \mathbb{R}$ and some subsequence $\{n'\}$. Helly's selection theorem (Rudin (1976), pg 167) guarantees the existence of pointwise limits. Let now Ω' be the set of all sample points ω for which the sequence $\{\hat{F}_n(\cdot;\omega)\}$ has only pointwise limits F such that $\mathcal{L}(F) \geq \mathcal{L}(F_0)$. In view of the above discussion, for each $\omega \in \Omega'$, all the limit points of $\{\hat{F}_n(\cdot;\omega)\}$ equal F_0 a.e. μ and this gives that $\int |\hat{F}_n(x;\omega) - F_0(x)|\, d\mu(x) \to 0$. Thus the desired result follows if we show that Ω' has probability 1. Let $\hat{Q}_n$ denote the empirical estimator of Q, the distribution of (L, R). By the SLLN, $\Omega_0 = \{\mathcal{L}_n(F_0) \to \mathcal{L}(F_0)\}$ has probability 1, and so does $\Omega_U = \{\hat{Q}_n(U) \to Q(U)\}$ for every Borel subset U of $\Delta = \{(l, r) : -\infty \leq l < r \leq \infty\}$. Thus we are done if we show that Ω' contains the intersection Ω_* of Ω_0 and $\bigcap_{U \in \mathcal{U}} \Omega_U$ for some countable collection $\mathcal{U}$ of Borel subsets of Δ.

Let α be a positive integer. Then there are finitely many extended real numbers

$$-\infty = q_0 < q_1 < q_2 < \cdots < q_\beta = \infty$$

such that $\mu((q_{i-1}, q_i)) < 2^{-\alpha}$ for $i = 1, \ldots, \beta$. Now form the sets $U_0, \ldots, U_{2\beta}$ by setting $U_{2i-1} = (q_{i-1}, q_i)$ for $i = 1, \ldots, \beta$, and $U_{2i} = [q_i, q_i]$ for $i = 0, \ldots, \beta$. Let $\mathcal{U}_\alpha$ denote the collection of all nonempty sets of the form $U_{ij} = \Delta \cap (U_i \times U_j)$ for $0 \leq i \leq j \leq 2\beta$. We shall take $\mathcal{U} = \bigcup_\alpha \mathcal{U}_\alpha$.

Let now ω belong to Ω_*. Let F_n denote the distribution function defined by $F_n(x) = \hat{F}_n(x;\omega)$ and Q_n the measure defined by $Q_n(A) = \hat{Q}_n(A;\omega)$. Let F be a pointwise limit of $\{F_n\}$. For simplicity in notation we shall assume that $F_n(x) \to F(x)$ for all $x \in \mathbb{R}$. We shall show that

$$\mathcal{L}(F_0) \leq \liminf_{n\to\infty} \mathcal{L}_n(\hat{F}_n)(\omega) \leq \limsup_{n\to\infty} \mathcal{L}_n(\hat{F}_n)(\omega) \leq \mathcal{L}(F).$$

The first inequality follows from $\mathcal{L}_n(\hat{F}_n)(\omega) \geq \mathcal{L}_n(F_0)(\omega)$, a consequence of the definition of the GMLE, and the fact that $\mathcal{L}_n(F_0)(\omega) \to \mathcal{L}(F_0)$ by the choice of ω. Thus we only need to establish the last inequality. For this note that $\mathcal{L}_n(\hat{F}_n)(\omega)$ can be expressed as

$$\int_\Delta \log\,[F_n(r) - F_n(l)]\, dQ_n(l, r).$$

The desired inequality is thus equivalent to

$$\limsup_{n\to\infty} \int_\Delta \log\,[F_n(r) - F_n(l)]\, dQ_n(l, r) \leq \int_\Delta \log\,[F(r) - F(l)]\, dQ(l, r). \tag{4.1}$$

Now fix a positive integer α and a negative integer q. Then

$$\int_{\Delta} \log\,[F_n(r) - F_n(l)]\, dQ_n(l,r) \le \int_{\Delta} q \vee \log\,[F_n(r) - F_n(l)]\, dQ_n(l,r)$$
$$\le \sum_{U \in \mathcal{U}_\alpha} M_n(U) Q_n(U),$$

where

$$M_n(U) = \sup_{(l,r)\in \bar{U}} q \vee \log\,[F_n(r) - F_n(l)]$$

and $\bar{U}$ is the closure of U. It is easy to check that $M_n(U) = q \vee \log\,[F_n(r_U) - F_n(l_U)]$, where $r_U = \sup\{r : (l,r) \in U\}$ and $l_U = \inf\{l : (l,r) \in U\}$. Thus

$$M_n(U) \to M(U) := q \vee \log\,[F(r_U) - F(l_U)] = \sup_{(l,r)\in \bar{U}} q \vee \log\,[F(r) - F(l)].$$

Also, by the choice of ω, $Q_n(U) \to Q(U)$ for all $U \in \mathcal{U}_\alpha$. Therefore we can conclude that

$$\sum_{U \in \mathcal{U}_\alpha} M_n(U) Q_n(U) \to \sum_{U \in \mathcal{U}_\alpha} M(U) Q(U).$$

Let now

$$m(U) = \inf_{(l,r)\in \bar{U}} q \vee \log\,[F(r) - F(l)], \quad U \in \mathcal{U}_\alpha.$$

Using the bound

$$|q \vee \log(x) - q \vee \log(y)| \le e^{-q}|x - y|, \quad 0 \le x, y \le 1,$$

it is easy to verify that

$$M(U) - m(U) \le e^{-q} \sup_{(l,r)\in \bar{U}} [F(r_U) - F(r) + F(l) - F(l_U)], \quad U \in \mathcal{U}_\alpha.$$

This shows the following.

(1) If $U = \Delta \cap [(q_{i-1}, q_i) \times (q_{j-1}, q_j)]$, then $M(U) - m(U) > 2/\alpha$ implies either $F(q_i) - F(q_{i-1}) > e^q/\alpha$ or $F(q_j) - F(q_{j-1}) > e^q/\alpha$;

(2) if $U = \Delta \cap [[q_i, q_i] \times (q_{j-1}, q_j)]$, then $M(U) - m(U) > 2/\alpha$ implies $F(q_j) - F(q_{j-1}) > e^q/\alpha$;

(3) if $U = \Delta \cap [(q_{i-1}, q_i) \times [q_j, q_j]]$, then $M(U) - m(U) > 2/\alpha$ implies $F(q_i) - F(q_{i-1}) > e^q/\alpha$.

Of course, if U contains only one point, then $M(U) - m(U) = 0$. Using this, we derive

$$\sum_{U \in \mathcal{U}_\alpha} (M(U) - m(U))Q(U) \le \frac{2}{\alpha} + |q| \sum_{U \in \mathcal{U}_\alpha} Q(U) I[(M(U) - m(U)) > 2/\alpha]$$
$$\le \frac{2}{\alpha} + |q| \sum_{i=1}^{\beta} P(q_{i-1} < L < q_i) I[F(q_i) - F(q_{i-1}) > e^q/\alpha]$$
$$+ |q| \sum_{j=1}^{\beta} P(q_{j-1} < R < q_j) I[F(q_j) - F(q_{j-1}) > e^q/\alpha]$$
$$\le \frac{2}{\alpha} + |q|(1 + \alpha e^{-q}) 2^{1-\alpha}.$$

In the last step we use the facts that

$$P(q_{i-1} < L < q_i) + P(q_{i-1} < R < q_i) \leq 2\mu((q_{i-1}, q_i)) \leq 2^{1-\alpha}$$

and that at most $1 + \alpha e^{-q}$ among the terms $F(q_1) - F(q_0), \ldots, F(q_\beta) - F(q_{\beta-1})$ exceed e^q/α.

Combining the above shows that

$$\begin{aligned}
&\limsup_{n\to\infty} \int_\Delta \log\left[F_n(r) - F_n(l)\right] dQ_n(l,r) \\
&\qquad \leq \int_\Delta q \vee \log\left[F(r) - F(l)\right] dQ(l,r) + \frac{2}{\alpha} + |q|(1 + \alpha e^{-q})2^{1-\alpha}.
\end{aligned}$$

The desired inequality (4.1) follows from this by first letting $\alpha \to \infty$ and then $q \to -\infty$.

5. Proof of the Propositions

Fix $\omega \in \Omega_\mu$. Abbreviate $\hat{F}_n(\cdot;\omega)$ by F_n. Let F be a pointwise limit of F_n. Without loss of generality, assume that $\lim_{n\to\infty} F_n(x) = F(x)$ for all x. Set

$$D = \{x \in \mathbb{R} : F(x) \neq F_0(x)\}.$$

Since $\int |F_n - F_0|\, d\mu \to 0$ and μ is a finite measure in view of the assumption $E(K) < \infty$, we have $\mu(D) = 0$.

Proof of Proposition 3.1: We need to show that D does not contain regular continuity points of F_0. Let x_0 be a continuity point of F_0. If x_0 belongs to D, then $F(x_0) \neq F_0(x_0)$ and the continuity of F_0 at x_0 and the monotonicity of F and F_0 yield that there exists a positive ϵ such that either $(x_0 - \epsilon, x_0]$ or $[x_0, x_0 + \epsilon)$ is contained in D. Thus either $\mu((x_0 - \epsilon, x_0]) = 0$ or $\mu([x_0, x_0 + \epsilon)] = 0$, and x_0 is not regular. □

Proof of Proposition 3.2: Let x_0 be a continuity point of F_0 which is also an interior point of S, the set of support points of μ. Then x_0 does not belong to D; otherwise, there exist, for each $\epsilon > 0$, support points x_1 and x_2 of μ and a positive η such that $(x_1 - \eta, x_1 + \eta)$ is contained in $(x_0 - \epsilon, x_0]$ and $(x_2 - \eta, x_2 + \eta)$ is contained in $[x_0, x_0 + \epsilon)$ and this leads to the contradiction $\mu(D) > 0$. This shows that $F(x) = F_0(x)$ for all continuity points x of F_0 that belong to the interior of S and proves the first part of Proposition 3.2. The second part follows from the first part and the monotonicity of F and F_0. □

Proof of Proposition 3.3: Suppose every point of increase of F_0 is strongly regular. We shall show that D does not contain continuity points of F_0. Let x_0 be a continuity point of F_0. If x_0 is a point of increase of F_0, then it is strongly regular and hence regular and cannot belong to D by Proposition 3.1. Suppose now x_0 is not a point of increase of F_0. Then again x_0 cannot belong to D. Otherwise, either $F(x_0) > F_0(x_0)$ or $F(x_0) < F_0(x_0)$ and we shall show that each leads to the contradiction $\mu(D) > 0$. In the first case, $b := \sup\{x : F_0(x) = F_0(x_0)\}$ is a point of increase of F_0,

$b > x_0$ and $F(b-) \geq F(x_0) > F_0(x_0) = F_0(b-)$; thus $[x_0, b) \subset D$ and, since b is strongly regular by our assumption, $\mu(D) \geq \mu((x_0, b)) > 0$. In the second case, $a := \inf\{x \in \mathbb{R} : F_0(x) = F_0(x_0)\}$ is a point of increase of F_0, $a < x_0$ and $F(a) \leq F(x_0) < F_0(x_0) = F_0(a)$; thus $[a, x_0) \subset D$ and, since a is strongly regular by our assumption, $\mu(D) \geq \mu([a, x_0)) > 0$. This shows that D does not contain continuity points of F_0, which is the desired result of Proposition 3.3. □

Proof of Proposition 3.6: Make the assumptions of Proposition 3.6. Then D is empty; otherwise, we can use the continuity of F_0 to construct an open interval, that contains a point of increase of F_0 and is contained in D, and arrive at the contradiction $\mu(D) > 0$. Since D is empty, F_n converges to F_0 pointwise and hence uniformly as F_0 is continuous. This proves Proposition 3.6. □

Proof of Proposition 3.8: We shall only give the proof in the case $\mu(\{\tau_1\}) > 0$ and $F_0(\tau_2-) = 1$. We shall show that $D \cap [\tau_1, \tau_2] = 0$. This implies that $F_n(x) \to F_0(x)$ for all $x \in [\tau_1, \tau_2]$, and, by the continuity assumption on F_0, this convergence is even uniform on $[\tau_1, \tau_2]$.

It follows from Corollary 2.3 that $F(\tau_1) = F_0(\tau_1)$. This gives the desired result if $F_0(\tau_1) = 1$. Thus assume from now on that $F_0(\tau_1) < 1$. We are left to show that $D_1 = D \cap (\tau_1, \tau_2]$ is empty. If D_1 were not empty, we could use the continuity assumption on F_0, the monotonicity of F_0 and F and $F(\tau_1) = F_0(\tau_1) < F_0(\tau_2-) = 1$ to show that D_1 contains an open interval (a, b) such that $0 < F_0(a) < F_0(b) < 1$ and $\tau_1 < a < b < \tau_2$ and arrive at the contradiction $\mu(D) \geq \mu((a, b)) > 0$.

Acknowledgement. We thank the referees, the Associate Editor and Professor Tjøstheim for helpful remarks. Special thanks go to one referee for an elaborate report that suggested an alternative proof, provided additional references and raised interesting questions.

6. References

Aragon, J. and Eberly, D. (1992). On convergence of convex minorant algorithms for distribution estimation with interval-censored data. *J. of Computational and Graphical Statistics*, 1, 129-140.

Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T. and Silverman, E. (1955). An empirical distribution function for sampling incomplete information. *Ann. Math. Statist.*, 26, 641-647.

Becker, N.G. and Melbye, M. (1991). Use of a log-linear model to compute the empirical survival curve from interval-censored data, with application to data on tests for HIV positivity. *Austral. J. Statist.*, 33, 125-133.

Finkelstein, D.M. and Wolfe, R.A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics*, 41, 933-945.

Gentleman, R. and Geyer, C. J. (1994). Maximum likelihood for interval censored data: Consistency and computation. *Biometrika*, 81, 618-623.

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

Huang, J. (1996). Efficient estimation for proportional hazards models with interval censoring. *Ann. Statist.*, 24, 540-568.

Keiding. N. (1991). Age-specific incidence and prevalence: A statistical perspective (with discussion). *J. Roy. Statist. Soc. Ser. A*, 154, 371-412.

Peto, R. (1973). Experimental survival curve for interval-censored data. *Applied Statistics*, 22, 86-91.

Rudin, W. (1976). Principles of mathematical analysis. McGraw-Hill. New York.

Turnbull, B. W. (1976). The empirical distribution function with arbitrary grouped, censored and truncated data. *J. Royal Statist. Soc. Ser. B*, 38, 290-295.

van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.*, 21, 14-44.

Wang, Z. and Gardiner, J. C. (1996). A class of estimators of the survival function from interval-censored data. *Ann. Statist.*, 24, 647-658.

Wellner, J. A. (1995). Interval censoring case 2: alternative hypotheses. In *Analysis of censored data*, Proceedings of the workshop on analysis of censored data, December 28, 1994-January 1, 1995, University of Pune, Pune, India. *IMS Lecture Notes, Monograph Series.* **27**, 271 - 291. H. L. Koul and J. V. Deshpande, editors.

Wellner, J. A. and Zhan, Y. (1997). A hybrid algorithm for computation of the NPMLE from censored data. *JASA*, **92**, 945-959.

Yu, Q, Schick, A., Li, L. and Wong, G.Y.C. (1998a): Asymptotic properties of the GMLE in the case 1 interval-censorship model with discrete inspection times. To appear in *Canad. J. Statist.*

Yu, Q, Schick, A., Li, L. and Wong, G.Y.C. (1998b): Asymptotic properties of the GMLE with case 2 interval-censored data. *Prob. & Statist. Lett.*, **37**, 223-228.

Anton Schick
Department of Mathematical Sciences
Binghamton University
Binghamton, NY 13902-6000
E-mail: anton@math.binghamton.edu

Qiqing Yu
Department of Mathematical Sciences
Binghamton University
Binghamton, NY 13902-6000
E-mail: qyu@math.binghamton.edu