ASYMPTOTIC PROPERTIES OF THE GMLE IN THE CASE 1 INTERVAL-CENSORSHIP MODEL WITH DISCRETE INSPECTION TIMES

By Qiqing Yu[1,2], Anton Schick[2], Linxiong Li[3,4] and George Y. C. Wong[5,6]

October 1996. Revised April 1997.

Abstract. We consider the case 1 interval censorship model in which the survival time has an arbitrary distribution function F_0 and the inspection time has a discrete distribution function G. In such a model one is only able to observe the inspection time and whether the value of the survival time lies before or after the inspection time. We prove the strong consistency of the generalized maximum likelihood estimate (GMLE) of the distribution function F_0 at the support points of G and its asymptotic normality and efficiency at what we call regular points. We also present a consistent estimate of the asymptotic variance at these points. The first result implies uniform strong consistency on $[0, \infty)$ if F_0 is continuous and the support of G is dense in $[0, \infty)$. For arbitrary F_0 and G, Peto (1973) and Turnbull (1976) conjectured that the convergence for the GMLE is at the usual parametric rate $n^{1/2}$. Our asymptotic normality result supports their conjecture under our assumptions. But their conjecture was disproved by Groeneboom and Wellner (1992) who obtained the nonparametric rate $n^{1/3}$ under smoothness assumptions on the F_0 and G.

Short Title: Case 1 Model
AMS 1991 Subject Classification: Primary 62G05; Secondary 62G20.
Key words and phrases: Nonparametric maximum likelihood estimation; consistency; asymptotic normality and efficiency.

[1]Partially supported by NSF Grant DMS-9402561 and DAMD17-94-J-4332.
[2]Dept. of Mathematics, State University of New York at Binghamton, NY 13902, USA.
[3]Partially supported by LEQSF Grant 357-70-4107.
[4]Dept. of Mathematics, University of New Orleans, LA 70148, USA.
[5]Partially supported by DAMD17-94-J-4332.
[6]Strang Cancer Prevention Center, Cornell University Medical School, NY 10021, USA.

1. Introduction

In survival analysis, one frequently is unable to precisely observe the survival time X of interest, but can only assess that it belongs to some random interval. The simplest such model is the so-called case 1 interval censorship model. In this model one is only able to observe a random time Y and whether X lies in the random interval $[0, Y]$ or (Y, ∞). More formally, one observes

$$(Y, \Delta), \quad \text{where} \quad \Delta = I[X \leq Y]. \tag{1.1}$$

Here and below $I[A]$ denotes the indicator function of the event A. The random time Y is called the *inspection time*.

Such data arise in industrial life testing and medical research. Consider for example an animal sacrifice study in which a laboratory animal has to be dissected to check whether a tumor has developed. In this case, X is the onset of tumor and Y is the time of the dissection, and we only can infer at the time of dissection whether the tumor is present or has not yet developed. Other examples are mentioned in Ayer *et al.* (1955), Keiding (1991) and Wang and Gardiner (1996).

We shall assume throughout that the life time X and the inspection time Y are independent and denote their distribution functions by F_0 and G, respectively. Our data consist of n independent copies $(Y_i, \Delta_i) = (Y_i, I[X_i \leq Y_i])$, $i = 1, ..., n$, of (Y, Δ). We consider estimating (characteristics of) the distribution function F_0 based on these data.

Ayer *et al.* (1955) derived the explicit expression of the generalized maximum likelihood estimator (GMLE) of the distribution function F_0. Moreover, they established the weak consistency of the GMLE at continuity points x of F_0 under additional assumptions on G. They also mentioned the strong consistency of the GMLE at each support point of a discrete Y with finitely many values. Using an inequality of theirs we shall generalize this result to arbitrary discrete Y in our Theorem 2.1. From this result we shall derive the uniform strong consistency on the entire line if F_0 is continuous and the support of

Y is dense in the positive half line. Moreover, using Theorem 2.1 of Ayer *et al.* (1955), we shall derive an explicit representation of the GMLE at what we call regular points and conclude the asymptotic normality and efficiency of the GMLE at such points.

Peto (1973) considered the problem of obtaining the GMLE based on interval-censored data using a Newton-Raphson type algorithm. Turnbull (1976) proposed a self-consistent algorithm and showed that it converges to the GMLE $\hat{F}$. Both conjectured that for arbitrary F_0 and G, the GMLE is asymptotically normal at the usual $n^{1/2}$ rate. Thus our results provide a partial justification of their claim for discrete Y. It was, however, shown by Groeneboom and Wellner (1992) that this conjecture is false if F_0 and G satisfy certain smoothness assumptions. Indeed, their Theorem 5.1 establishes that under differentiability assumptions on F_0 and G the convergence is at the slower $n^{1/3}$ rate and the limiting distribution is not normal. Groeneboom and Wellner (1992) also obtained the uniform strong consistency of the GMLE for continuous F_0 and G. A variant of this result was also proved by Wang and Gardiner (1996) using a totally different approach and a slightly different set of assumptions.

The results of Groeneboom and Wellner (1992) give a fairly detailed description for the case of continuous F_0 and G, while ours do so for the case of arbitrary F_0 and discrete G. There are many practical situations in which Y is discrete. In medical research, for example, the data are often recorded as integers (to represent number of days, weeks etc). Motivated by this we assume that the inspection time Y is a discrete random variable with density g. This assumption is used by several authors in survival analysis, see Becker and Melbye (1991) and Finkelstein (1986) among others.

Our paper is organized as follows. We introduce the GMLE in Section 2 and prove its strong consistency. In Section 3 we establish the asymptotic normality and efficiency of the GMLE at what we call regular points. Finally, Section 4 summarizes our work, discusses some of its implications, addresses some questions raised by it and establishes connections with the work of others. In particular, we show by means of an example that our asymptotic normality result fails at nonregular points even though the rate of convergence is still $n^{1/2}$.

2. The consistency of the GMLE

By our assumptions, Y is a discrete random variable with density g. Let $\mathcal{A}$ be the set of possible values of Y, i.e., $\mathcal{A} = \{a \in \mathbb{R} : g(a) > 0\}$. For $a \in \mathcal{A}$, set

$$N_n^-(a) = \frac{1}{n} \sum_{j=1}^n I[X_j \le a, Y_j = a],$$

$$N_n^+(a) = \frac{1}{n} \sum_{j=1}^n I[X_j > a, Y_j = a],$$

$$N_n(a) = \frac{1}{n} \sum_{j=1}^n I[Y_j = a].$$

The generalized likelihood is given by

$$\Lambda_n(F) = \prod_{a \in \mathcal{A}} F(a)^{nN_n^-(a)} (1 - F(a))^{nN_n^+(a)}.$$

In the above we let F range over the set $\mathcal{F}$ of all subdistribution functions. A function F is called a *subdistribution function* if $F = aF_1$ for some distribution function F_1 and some number a in $[0, 1]$. Thus a subdistribution function has all the properties of a distribution function except that its limit at infinity may be less than 1.

Note that $\Lambda_n(F)$ depends on F only through the values of F at the points $a \in \mathcal{A}$ for which $N_n(a) > 0$. Thus there exists no unique maximizer of $\Lambda_n(F)$ in the set $\mathcal{F}$. But there exists a uniquely determined $\mathcal{F}$-valued random element $\hat{F}_n$ which maximizes $\Lambda_n(F)$ and satisfies $\hat{F}_n(b) = \sup\{\hat{F}_n(a) : a \le b, N_n(a) > 0\}$ for each $b \in \mathbb{R}$. Here we interpret the supremum of the empty set as 0. We call $\hat{F}_n$ the GMLE of F_0. It is easy to check that $\hat{F}_n(Y_{(1)}) = 0$ on the event $\{N_n^-(Y_{(1)}) = 0\}$ and $\hat{F}_n(Y_{(n)}) = 1$ on the event $\{N_n^+(Y_{(n)}) = 0\}$, where $Y_{(1)}$ and $Y_{(n)}$ are the smallest and largest among $Y_1, \ldots, Y_n$. For latter use, set

$$\tilde{F}_n(a) = \begin{cases} N_n^-(a)/N_n(a), & \text{if } N_n(a) > 0 \\ 0, & \text{otherwise.} \end{cases}$$

2.1. Theorem. *The GMLE $\hat{F}_n$ satisfies $\hat{F}_n(a) \to F_0(a)$ almost surely for each $a \in \mathcal{A}$.*

PROOF: Using the following inequality given in Ayer *et al.* (1955, p. 644),

$$\sum_{a\in\mathcal{A}} (\hat{F}_n(a) - F_0(a))^2 N_n(a) \le \sum_{a\in\mathcal{A}} (\tilde{F}_n(a) - F_0(a))^2 N_n(a),$$

we get

$$\sum_{a\in\mathcal{A}} (\hat{F}_n(a) - F_0(a))^2 N_n(a) \le \sum_{a\in\mathcal{A}} |N_n(a) - g(a)| + \sum_{a\in\mathcal{A}} (\tilde{F}_n(a) - F_0(a))^2 g(a).$$

It follows from the SLLN that for each $a \in \mathcal{A}$, $N_n(a) \to g(a)$ and $\tilde{F}_n(a) \to F_0(a)$ almost surely. Thus Scheffé's Theorem (see Billingsley (1968, p. 224)) implies

$$\sum_{a\in\mathcal{A}} |N_n(a) - g(a)| \to 0 \quad \text{almost surely}$$

and the Lebesgue Dominating Convergence Theorem implies

$$\sum_{a\in\mathcal{A}} (\tilde{F}_n(a) - F_0(a))^2 g(a) \to 0 \quad \text{almost surely.}$$

It follows that $\sum_{a\in\mathcal{A}} (\hat{F}_n(a) - F_0(a))^2 N_n(a) \to 0$ almost surely. This yields the desired result as $N_n(a)$ is eventually positive with probability 1 for each $a \in \mathcal{A}$. $\square$

The above result was already observed by Ayer *et al.* (1955) in the case when $\mathcal{A}$ is finite. In this case one can even conclude that the GMLE is uniformly strongly consistent on $\mathcal{A}$, i.e., $\sup_{a\in\mathcal{A}} |\hat{F}_n(a) - F_0(a)| \to 0$ almost surely. For countably infinite $\mathcal{A}$, however, additional assumptions are required to conclude this as demonstrated by the following example.

2.2. Example. Suppose $\mathcal{A} = \{y_i : y_i = 1 - 1/i, i \ge 1\}$ and $G(y) = y$ for $y \in \mathcal{A}$. Then the GMLE will not be uniformly strongly consistent on $\mathcal{A}$ if $0 < F(1-) < 1$.

PROOF: Let $\Omega_n = \bigcup_{i=1}^n \bigcap_{j\ne i} \{X_i \le Y_i, Y_j < Y_i\}$. Then $\Omega_n \subset \{N^+(Y_{(n)}) = 0\}$. Since $\hat{F}_n(Y_{(n)}) = 1$ on the event $\{N^+(Y_{(n)}) = 0\}$ as observed prior to Theorem 2.1 and since

$F_0(1-) < 1$, we cannot have uniform strong convergence if $\liminf_{n\to\infty} P(\Omega_n) > 0$. But

$$P(\Omega_n) = nP(\bigcap_{j=2}^{n}\{X_1 \le Y_1, Y_j < Y_1\}) \ge nF_0(y_n)(G(y_n-))^{n-1}P(Y_1 \ge y_n)$$

so that by the choice of $\mathcal{A}$ and G

$$\liminf_{n\to\infty} P(\Omega_n) \ge \liminf_{n\to\infty} F_0(y_n)\Big(1 - \frac{1}{n-1}\Big)^{n-1} = F(1-)/e > 0.$$

Consequently, the GMLE is not uniformly consistent on $\mathcal{A}$. $\square$

We now address the uniform strong consistency.

2.3. Corollary. *Suppose the set $\mathcal{A}$ is closed. Assume that $F_0(a-) = F_0(a)$ for each $a \in \mathcal{A}$ for which there is a strictly increasing sequence of points $\{a_i\}_{i\ge1}$ in $\mathcal{A}$ such that $a_i \uparrow a$. Then the GMLE is uniformly strongly consistent on $\mathcal{A}$.*

Proof: Let m be a positive integer. Let $\mathcal{A}_i = \{a \in \mathcal{A} : x_{i-1} \le a < x_i\}$, $i = 1, \dots, m$, where $x_0 = -\infty$, $x_m = \infty$ and $x_i = \inf\{x : F_0(x) \ge i/m\}$, $i = 1, ..., m-1$. Let $a \in \mathcal{A}$. Then $a \in \mathcal{A}_i$ for some $i = 1, \dots, m$. Since $\mathcal{A}$ is a closed set, $a_i = \inf \mathcal{A}_i$ and $b_i = \sup \mathcal{A}_i$ belong to $\mathcal{A}$. Using the monotonicity of $\hat{F}_n$ and F_0, we find that

$$|\hat{F}_n(a) - F_0(a)| \le \max\{|\hat{F}_n(b_i) - F_0(b_i)|, |\hat{F}_n(a_i) - F_0(a_i)|\} + F_0(b_i) - F_0(a_i).$$

If $b_i < x_i$, then $F_0(b_i) - F_0(a_i) < 1/m$. If $b_i = x_i$, then $F_0(x_i) = F_0(x_i-) = i/m$ and $F_0(b_i) - F_0(a_i) \le 1/m$. This shows that

$$\limsup_{n\to\infty} \sup_{a\in A} |\hat{F}_n(a) - F_0(a)| \le 1/m$$

on the event $\Omega_* = \bigcap_{a\in\mathcal{A}}\{\lim_{n\to\infty} \hat{F}_n(a) = F_0(a)\}$. Since m is arbitrary and $P(\Omega_*) = 1$ by Theorem 2.1, we obtain the desired result. $\square$

In the next corollary, the set $\mathcal{A}$ needs not be closed.

2.4. Corollary. *Assume that* $\mathcal{A} = \{a_i\}_{i\geq 1}$, *where* $a_i < a_{i+1}$ *for all* i. *Let* $\tau = \sup_i a_i$. *If* $F_0(\tau-) = 1$, *then the GMLE is uniformly strongly consistent on* $\mathcal{A}$.

Proof: Let m be a positive integer. Since

$$\sup_{a\in\mathcal{A}} |\hat{F}_n(a) - F_0(a)| \leq \max_{1\leq i\leq m} |\hat{F}_n(a_i) - F_0(a_i)| + 1 - F_0(a_m),$$

it follows from Theorem 2.1 that

$$\limsup_{n\to\infty} \sup_{a\in\mathcal{A}} |\hat{F}_n(a) - F_0(a)| \leq 1 - F(a_m).$$

The desired result follows as m is arbitrary and $F_0(\tau-) = 1$. □

We call a number x a *point of increase of* F_0 if either $F_0(x) < F_0(y)$ for all $y > x$ or $F_0(y) < F_0(x)$ for all $y < x$. Note that, for each α in the interval $(0, 1)$, the left quantile $F_0^{-1}(\alpha) = \inf\{y : F(y) \geq \alpha\}$ is a point of increase of F_0.

2.5. Corollary. *Suppose that* F_0 *is continuous and the closure of* $\mathcal{A}$ *contains the set* S *of all points of increase of* F_0. *Then the GMLE is uniformly strongly consistent, i.e.,* $\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F_0(x)| \to 0$ *almost surely.*

Proof: Let $F_1, F_2, \ldots$ be subdistribution functions such that $F_n(a) \to F_0(a)$ for all $a \in \mathcal{A}$. Let m be a positive integer. Since F_0 is continuous, there are points $x_1 < \cdots < x_m$ in S such that $F_0(x_i) = i/(m+1)$. The continuity of F_0 and the fact that the closure of $\mathcal{A}$ contains S imply that there are points $a_1 < \cdots < a_m$ in $\mathcal{A}$ such that $|F_0(a_i) - F_0(x_i)| \leq 1/m^2$. Using this and the monotonicity of F_0 and F_n we derive that

$$|F_n(x) - F_0(x)| \leq \max_{1\leq i\leq m} |F_n(a_i) - F_0(a_i)| + \frac{3}{m}, \quad x \in \mathbb{R}.$$

This shows that F_n converges to F_0 uniformly.

By the above, the events $\bigcap_{a\in\mathcal{A}} \{\hat{F}_n(a) \to F_0(a)\}$ and $\{\sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F_0(x)| \to 0\}$ are identical and thus have probability 1 by Theorem 2.1. □

3. The asymptotic normality of the GMLE

We shall now discuss asymptotic normality and efficiency of $\hat{F}_n(x)$ for regular points x as defined next. Let $\mathcal{A}_* = \mathcal{A} \cup \{-\infty, \infty\}$. For $x \in \mathbb{R}$, set

$$x_- := \sup\{a \in \mathcal{A}_* : a < x\} \quad \text{and} \quad x_+ := \inf\{a \in \mathcal{A}_* : a > x\}.$$

We say x is a *regular point*, if x belongs to $\mathcal{A}$, x_- and x_+ belong to $\mathcal{A}_*$, $x_- < x < x_+$ and $F_0(x_-) < F_0(x) < F_0(x_+)$. It is worth mentioning that there may be infinitely many regular points. For example, if F_0 is strictly increasing and $\mathcal{A}$ is the set of all positive integers, then every positive integer is a regular point. The conditions imposed on regular points are somewhat similar to the assumption that F_0 and G have positive and continuous derivatives needed in the asymptotic distribution result of the GMLE (see Groeneboom and Wellner (1992)). However, their convergence rate is $n^{1/3}$, while we shall show that the convergence rate is $n^{1/2}$ under our assumptions.

Given a regular point x, $\hat{F}_n(x)$ may or may not be the same as $\tilde{F}_n(x)$ as shown by the following example. Suppose that F is the exponential distribution function and $\mathcal{A} = \{1, 2\}$. Then both 1 and 2 are regular points. If a sample of size 3 consists of observations $\{(1,0), (1,1), (2,1)\}$, then $(\tilde{F}(1), \tilde{F}(2)) = (1/2, 1)$, which is the same as $(\hat{F}(1), \hat{F}(2))$. On the other hand, if a sample of size 3 consists of observations $\{(1,0), (1,1), (2,0)\}$, then $(\tilde{F}(1), \tilde{F}(2)) = (1/2, 0)$, which is not the same as $(\hat{F}(1), \hat{F}(2)) = (1/3, 1/3)$. However, the following lemma shows that the two estimators differ only on a set whose probability tends to zero.

3.1. Lemma. *Suppose x is a regular point. Then $P(\hat{F}_n(x) = \tilde{F}_n(x)) \to 1$.*

PROOF: Assume first that x_- and x_+ belong to $\mathcal{A}$. Let $B_n = \{\hat{F}_n(x_-) < \hat{F}_n(x) < \hat{F}_n(x_+)\}$ and $C_n = \{N_n(x_-) > 0, N_n(x) > 0, N_n(x_+) > 0\}$. It follows from Theorem 2.1 and $F_0(x_-) < F_0(x) < F_0(x_+)$ that $P(B_n) \to 1$, and from the SLLN that $P(C_n) \to 1$. In view of Theorem 2.1 in Ayer *et al.* (1955), we have, on the event $B_n \cap C_n$,

$$\hat{F}_n(x_-) < \tilde{F}_n(x) \le \hat{F}_n(x) \le \tilde{F}_n(x) < \hat{F}_n(x_+).$$

That is, $\hat{F}_n(x) = \tilde{F}_n(x)$. Thus the desired result follows as $P(B_n \cap C_n) \to 1$. This proves the claim when x_- and x_+ belong to $\mathcal{A}$.

If $x_+ \notin \mathcal{A}$ and $x_- \in \mathcal{A}$, then $x_+ = +\infty$ since x is a regular point. Let

$$B_n^+ = \{\hat{F}_n(x_-) < \hat{F}_n(x) < 1\} \text{ and } C_n^+ = \{N_n(x_-) > 0, N_n(x) > 0\}.$$

It follows from Theorem 2.1 and $F_0(x_-) < F_0(x) < 1$ that $P(B_n^+) \to 1$, and from the SLLN that $P(C_n^+) \to 1$. In view of Theorem 2.1 in Ayer *et al.* (1955), we have, on the event $B_n^+ \cap C_n^+$, $\hat{F}_n(x_-) < \tilde{F}_n(x) \le \hat{F}_n(x) \le \tilde{F}_n(x)$. That is, $\hat{F}_n(x) = \tilde{F}_n(x)$. Thus the desired result follows as $P(B_n^+ \cap C_n^+) \to 1$. This proves the claim when x_- but not x_+ belongs to $\mathcal{A}$.

The proofs when x_+ but not x_- belongs to $\mathcal{A}$ is similar and will be omitted. □

The above result shows that $\hat{F}_n(x)$ has the same asymptotic properties as $\tilde{F}_n(x)$. Thus the following result is immediate.

3.2. Theorem. *Let x be a regular point. Then*

$$\hat{F}_n(x) - F_0(x) = \frac{1}{n}\sum_{j=1}^{n} \frac{I[Y_j = x]}{g(x)}(\Delta_j - F_0(x)) + o_p(n^{-1/2}).$$

Consequently, $n^{1/2}(\hat{F}_n(x) - F_0(x))$ is asymptotically normal with mean 0 and variance

$$\frac{F_0(x)(1 - F_0(x))}{g(x)}.$$

This asymptotic variance can be consistently estimated by

$$\frac{\hat{F}_n(x)(1 - \hat{F}_n(x))}{N_n(x)}.$$

Also, if $x_1 < \cdots < x_m$ are regular points, then $n^{1/2}(\hat{F}_n(x_1) - F_0(x_1), \ldots, \hat{F}_n(x_m) - F_0(x_m))$ is asymptotically normal with mean vector 0 and diagonal covariance matrix.

Let us now address efficiency considerations. For this fix a regular point x. It follows from the above theorem that $\hat{F}_n(x)$ has influence function ψ given by

$$\psi(\Delta, Y) = \frac{I[Y = x]}{g(x)}(\Delta - F_0(x)).$$

We shall now show that ψ is the efficient influence function for estimating $F_0(x)$. This will show that $\hat{F}_n(x)$ is a least dispersed regular estimator of $F_0(x)$. The reader unfamiliar with these concepts should consult the monograph by Bickel *et al.* (1993). Let $\mathcal{H}$ be the set of all measurable functions such that $\int h\, dF_0 = 0$ and $\int h^2\, dF_0 < \infty$. For $h \in \mathcal{H}$ define a sequence $F_{n,h}$ of distribution functions by

$$F_{n,h}(t) = \int_{(-\infty,t]} (1 + n^{-1/2} h_n)\, dF_0, \quad t \in \mathbb{R},$$

where $h_n = hI[2|h| \le n^{1/2}] - \int hI[2|h| \le n^{1/2}]\, dF_0$. Then

$$n^{1/2}(F_{n,h}(x) - F_0(x)) \to H(x) = \int_{(-\infty,x]} h\, dF_0.$$

The tangent (or score function) τ_h associated with the perturbed distributions $F_{n,h}$ is given by

$$\tau_h(\Delta, Y) = H(Y)\left(\frac{\Delta}{F_0(Y)} - \frac{1-\Delta}{1-F_0(Y)}\right) = \frac{H(Y)(\Delta - F_0(Y))}{F_0(Y)(1-F_0(Y))}.$$

Finally, it is easy to check that

$$E(\psi(\Delta, Y)\tau_h(\Delta, Y) = H(x).$$

Since this holds for all $h \in \mathcal{H}$ and since ψ is a tangent, i.e., $\psi = \tau_h$ for some $h \in \mathcal{H}$ with $H(Y) = I[Y = x]F_0(x)(1 - F_0(x))/g(x)$, we obtain that ψ is the efficient influence function if G is known. However, ψ is also the efficient influence function if G is unknown as the tangents for G are orthogonal to the tangents $\{\tau_h : h \in \mathcal{H}\}$ for F_0.

4. Concluding Remarks

The main results of our paper are given in Theorems 2.1 and 3.2. Theorem 2.1 gives the strong consistency at each point in $\mathcal{A}$, while Theorem 3.2 obtains asymptotic normality at regular points. Thus $\hat{F}_n(x)$ is both strongly consistent and asymptotically normally distributed at each regular point x. Typically, consistency fails to hold for points of increase that are not in the closure of $\mathcal{A}$. Also, the asymptotic normality result may not hold for nonregular points as the following example shows.

4.1. Example. Assume that $\mathcal{A}$ consists of just four points, namely $a_1 < a_2 < a_3 < a_4$, and that $0 < F(a_1) < F(a_2) = F(a_3) < F(a_4) < 1$. On the event $A_n = \{\tilde{F}_n(a_1) \leq \tilde{F}_n(a_2) \leq \tilde{F}_n(a_3) \leq \tilde{F}_n(a_4)\}$ we have $\hat{F}_n(a_i) = \tilde{F}_n(a_i)$, $i = 1, \ldots, 4$, and on the event $B_n = \{\tilde{F}_n(a_1) \leq \tilde{F}_n(a_2) \leq \tilde{F}_n(a_4), \tilde{F}_n(a_1) \leq \tilde{F}_n(a_3) \leq \tilde{F}_n(a_4), \tilde{F}_n(a_2) > \tilde{F}_n(a_3)\}$, we have $\hat{F}_n(a_i) = \tilde{F}_n(a_i)$ for $i = 1, 4$ and $\hat{F}_n(a_2) = \hat{F}_n(a_3) = \bar{F}_n$ where

$$\bar{F}_n = \frac{N_n^-(a_2) + N_n^-(a_3)}{N_n(a_2) + N_n(a_3)}.$$

It follows from the SLLN that $P(A_n \cup B_n) \to 1$. This shows that the asymptotic distribution of $\sqrt{n}(\hat{F}_n(a_2) - F_0(a_2), \hat{F}_n(a_3) - F_0(a_3))^T$ is the same as that of $\sqrt{n}(\tilde{F}_n^*(a_2) - F_0(a_2), \tilde{F}_n^*(a_3) - F_0(a_3))^T$, where $(\tilde{F}_n^*(a_2), \tilde{F}_n^*(a_3)) = (\tilde{F}_n(a_2), \tilde{F}_n(a_3))$ if $\tilde{F}_n(a_2) \leq \tilde{F}_n(a_3)$ and $\tilde{F}_n^*(a_2) = \tilde{F}_n^*(a_3) = \bar{F}_n$ if $\tilde{F}_n(a_2) > \tilde{F}_n(a_3)$. An application of Slutsky's Theorem yields that the asymptotic distribution of $\sqrt{n}(\tilde{F}_n^*(a_2) - F_0(a_2), \tilde{F}_n^*(a_3) - F_0(a_3))^T$ is the distribution of the bivariate random vector Z^* defined by

$$Z^* = \begin{pmatrix} Z_1^* \\ Z_2^* \end{pmatrix} = \begin{pmatrix} Z_2 \\ Z_3 \end{pmatrix} I[Z_2 \leq Z_3] + \frac{g(a_2)Z_2 + g(a_3)Z_3}{g(a_2) + g(a_3)} \begin{pmatrix} 1 \\ 1 \end{pmatrix} I[Z_2 > Z_3],$$

where Z_2 and Z_3 are independent normal random variables with zero means and variances $F(a_2)(1 - F(a_2))/g(a_2)$ and $F(a_3)(1 - F(a_3))/g(a_3)$, respectively. One can check that the distributions of Z_1^* and Z_2^* are not normal.

The corollaries in Section 2 address uniform strong consistency under different sets of assumptions. Corollary 2.3 implies that the GMLE is uniformly strongly consistent on $\mathcal{A}$ if

F is continuous and $\mathcal{A}$ is closed. Corollary 2.4 gives uniform consistency on $\mathcal{A}$ if this set is generated by an increasing sequence. If F is increasing and $\mathcal{A} \subset \{x \in \mathbb{R} : 0 < F(x) < 1\}$, then the assumptions of Corollary 2.4 imply that each point in $\mathcal{A}$ is regular and thus in view of Theorem 3.2 the asymptotic normality at each point in $\mathcal{A}$.

Corollary 2.5 is of interest from a theoretical point of view as it provides conditions that guarantee the uniform strong consistency on the entire line. From a practical point of view the imposed conditions are rather unrealistic. For example, if F is the uniform distribution on $[0, 1]$, then $\mathcal{A}$ has to contain a dense subset of $[0, 1]$. But distributions G with this property are rarely encountered in practice. Note also that the assumptions of Corollary 2.5 rule out the existence of regular points so that we cannot conclude the asymptotic normality from Theorem 3.2.

It is an open question whether the parametric convergence rate holds at each point in $\mathcal{A}$. Since one can show that $\tilde{F}_n$ has parametric convergence rate at each point in $\mathcal{A}$, we conjecture that the GMLE has the same property although the limit might not be normal as Example 4.1 shows.

Groeneboom and Wellner (1992) showed that the GMLE is uniformly strongly consistent if F_0 and G are continuous and $P_{F_0} << P_G$. The latter means that the probability measure P_{F_0} induced by F_0 is absolutely continuous with respect to the probability measure P_G induced by G. In view of our Corollary 2.5 we expect the uniform strong consistency also if F_0 is continuous and if G is a mixture of a continuous distribution function and a discrete distribution function which satisfies the assumptions in Corollary 2.5.

Groeneboom and Wellner (1992) showed that under the additional assumption that F_0 and G have positive derivatives at a point t_0, the convergence rate of $\hat{F}_n(t_0)$ is $n^{1/3}$. It is an open question whether the rate $n^{1/3}$ is still valid without this additional assumption. It is also not known whether the rate can be improved under additional smoothness assumptions on F_0 and G.

Our parametric convergence rate $n^{1/2}$ in Theorem 3.2 is in contrast to the nonparametric convergence rate $n^{1/3}$ under their assumptions. Our Theorem 3.2 is trivially true under the assumption that both X and Y take on the same finitely many values. In this case, the problem reduces to the estimation of the parameters of a multinomial distribution function, which is a parametric problem giving the usual $n^{1/2}$ convergence rate. This simple fact was noticed without proof by Peto (1973) and Turnbull (1976) as they both conjectured (incorrectly) that the GMLE has a convergence rate $n^{1/2}$ in general. We established the parametric convergence rate of the GMLE for the first time under the assumption that X and Y may take infinitely many values.

5. References

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.

Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and adaptive estimation for semiparametric models.* The Johns Hopkins University Press, Baltimore.

Billingsley, P. (1968). *Convergence of probability measures.* Wiley, New York.

Finkelstein, D.M. (1986). A proportional hazards model for interval-censored failure time data. *Biometrics*, 42, 845-854.

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

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. *Appl. Statist.*, 22, 86-91.

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

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