ESTIMATION OF THE MEDIAN SURVIVAL TIME UNDER RANDOM CENSORSHIP*

Joseph C. Gardiner

Michigan State University

V. Susarla

State University of New York - Binghamton

and

John Van Ryzin

Columbia University

An estimator of the median survival time m is constructed from survival data subject to random right censorship. For a broad class of nonparametric survival and censoring distributions, the asymptotic theory of the estimator is derived including its consistency in p-th mean, asymptotic normality, and a.s. convergence.

1. Introduction.

In several longitudinal studies the median and mean survival times are considered important summary statistics describing the survival experience of the sample under observation. The mean survival time is a commonly used statistic in the case of no censoring. This is due to its ease of computation and the considerable literature available on its properties. However, its competitor, the median survival time may be preferred with censored survival data because it is less sensitive to large observations and to the censoring pattern. The purpose of this article is to introduce an estimator of the median survival time which has applications in a variety of situations encountered in

---

*Research sponsored in part by a grant form the National Institutes of Health under Grant 2R01-GM 28405 and in part by the Office of Naval Research under Contract N00014-79-C-0522.

AMS 1980 subject classifications. 62G05, 62G15, 62F12.

Key words and phrases. Median survival time estimator, asymptotic normality, consistency in p-th mean, censored survival data.

survival analyses. Furthermore, we derive and study its asymptotic properties.

We envisage the usual random censorship model, that is, the survival variable X may be deterred from complete observation by the action of a competing censoring time variable Y, which is independent of X. Thus, one has available only the datum (Z,δ), where $Z = \min(X,Y)$ and δ identifies whether Z is a true survival time ($\delta = 1$) or a censoring time ($\delta = 0$). The underlying distributions of X and Y are left unspecified.

Given a sample of n independent observations $\{(Z_i,\delta_i) : 1 \leq i \leq n\}$ with each (Z_i,δ_i) having the distribution of (Z,δ), we shall estimate the survival function $F(t) = P[X > t]$ by the product limit estimator given by

$$(1.1) \qquad F_n(t) = \Pi_{i=1}^n \left\{\frac{N(Z_i)}{1 + N(Z_i)}\right\}^{[Z_i \leq t,\ \delta_i=1]}, \quad t \geq 0,$$

where $N(t) = \sum_{i=1}^n [Z_i > t]$ and $[A]$ is the indicator function of the event A. The median $m = m(F)$ of F is then naturally estimated by $m(F_n)$, where $m(F_n) = \inf\{t > 0 : F_n(t) \leq \frac{1}{2}\}$. Since F_n is a step function that decreases at uncensored observations, $m(F_n)$ coincides with one of the order statistics of $Z_1,\ldots,Z_n$ corresponding to an uncensored observation. Under some mild conditions on F and the censoring distribution $G(.) = P[Y > .]$, it can be shown that

$$(1.2) \qquad L\{n^{1/2}(m(F_n) - m)\} \to N(0, \sigma^2/4),$$

where

$$(1.3) \qquad \sigma^2 = \sigma^2(F,G) = f^{-2}(m)C(m),$$

f being the density of $(1 - F)$ and $C(t) = \int_0^t F^{-2}G^{-1}f$. Notice that an estimator of (1.3) can be constructed from the observations once a suitable estimator of f is available.

In some medical applications, Peto et al. (1977) have advised that a confidence interval estimate of m may be more appropriate than a point estimate since considerable variation could exist in the estimated median at 50% survival. The main purpose of this article is to develop an estimate $\hat{m}$ (of m), its large sample properties, and confidence intervals for m.

We remark here that previous investigators have avoided the problem of estimation of the variance (1.3) by examining direct estimates of the distribution of $m(F_n)$. (See Reid (1981)). Brookmeyer and Crowley (1982) and Emerson (1982) describe a method of constructing a confidence interval for m by utilizing an appropriate analog, for the case of censored data $\{(Z_i,\delta_i) : 1 < i < n\}$, of the traditional sign test statistic. This modified test statistic can be given in terms of the product limit estimator (1.1). Since

$$(1.4) \qquad L\{n^{1/2}(F_n(m) - \tfrac{1}{2})/F_n(m)\} \to N(0,C(m))$$

and, since with m fixed, C(m) can be estimated by $n\Sigma_{i=1}^{n}\delta_i[Z_i < m]N^{-2}(Z_i)$, a confidence region R_α for m with confidence coefficient $1 - \alpha$ is immediately obtained by inversion, that is

$$(1.5) \qquad R_\alpha = \{x : |F_n(x) - \tfrac{1}{2}|^2 < \chi^2(\alpha)F_n^2(x)\textstyle\sum_{i=1}^{n} \delta_i[Z_i < x]N^{-2}(Z_i)\},$$

where $\chi^2(\alpha)$ is the $(1 - \alpha)$th quantile of the chi-square distribution with one degree of freedom. The problem with (1.5) is that it need not be an interval. Simulation results have been reported in Brookmeyer and Crowley (1982) indicating that R_α may be a one-sided interval, especially if there is heavy censoring. The reader is referred to their paper cited above for further details. The methodology which we have developed avoids this difficulty by giving a confidence interval for m.

There are certain drawbacks associated with the papers to date. In the case of the papers by Sander (1975) or Reid (1981), fixed sample size confidence intervals are not possible even for large n since no estimate of σ^2 in (1.3) has been suggested. The papers by Brookmeyer and Crowley (1982) and Emerson (1982), although they contain confidence regions for m with the possible difficulties noted above, do not contain an explicit estimate of m. This paper provides an estimator $\hat{m}$ in explicit form which has all the necessary fixed large sample size properties. Furthermore, these properties can be used to develop certain desirable sequential properties of $\hat{m}$. This will be done in a subsequent paper.

The substantive material of this paper is divided into several sections. In Section 2 we introduce the basic notation and assumptions used throughout the entire paper together with our definition of an estimator $\hat{m}$ of m based on a sample of size n. Section 3 states the theorems concerning our estimator $\hat{m}$ and gives a consistent estimator of σ^2 in (1.3). It also contains some remarks on these theorems. Section 4 contains proofs of the theorems stated in Section 3.

2. Preliminary notions and definitions.

Let $\{X_i; i \geq 1\}$ be a sequence of nonnegative independent and identically distributed (i.i.d.) random variables (rv) representing the survival times, and let $\{Y_i; i \geq 1\}$, be the corresponding sequence of iid rv's of censoring times. We assume censoring is noninformative; that is, censoring and survival times are independent. The observable variables are $\{(Z_i, \delta_i) : i \geq 1\}$, where

$$Z_i = X_i \wedge Y_i \text{ and } \delta_i = [X_i \leq Y_i]. \tag{2.1}$$

Here [A] denotes the indicator of the event A and $X \wedge Y$ denotes min(X,Y). If F, G denote the survival (right hand tail) distribution functions of X, Y, then Z has survival distribution H and subdistributions $\tilde{H}$, $\tilde{\tilde{H}}$ given by

$$H = FG \; ; \; \tilde{\tilde{H}} = P[Z \leq \cdot, \delta = 1] \; ; \; \tilde{H} = 1 - H - \tilde{\tilde{H}}. \tag{2.2}$$

The median corresponding to the distribution (1 - F) is defined as

$$m(F) = \inf\{t > 0,\ F(t) \leq \tfrac{1}{2}\} \tag{2.3}$$

and will be abbreviated by the symbol m.

Given the sample of observations $\{(Z_i, \delta_i) : 1 \leq i \leq n\}$, we estimate F by the product limit estimator F_n of (1.1). The corresponding product limit estimator G_n of G is obtained by interchanging the roles of the censoring and survival times. It follows that

$$F_n G_n = N/n \tag{2.4}$$

where $N = nH_n$ and H_n is the ordinary empirical survival distribution of $\{Z_i\ ;\ 1 \leq i \leq n\}$. Thus we always have

$$m(F_n) \geq m(H_n). \tag{2.5}$$

In general there is no nontrivial upper bound on $m(F_n)$. As described above the ultimate objective is to obtain certain asymptotically efficient procedures for the estimation of m. For this purpose the convergence of the moments of our estimate and of those of the asymptotic variance σ^2 in (1.3) together with appropriate rates of convergence are needed.

In order to facilitate such results, we consider $\hat{m}$ defined by

$$\hat{m} = m(F_n) \wedge a_n, \tag{2.6}$$

where a_n is a known sequence of constants diverging to ∞ such that $\overline{\lim}\ a_n/n^\beta < 0$, for some $\beta > 0$, and $m(F_n)$ is the median of the estimator F_n in (1.1). For most studies in which lifetimes are known to be no larger than T, one can take $a_n \equiv T$. To obtain r-th moment convergence for every $r > 0$ (which implies a.s. convergence), the asymptotic distribution of $\hat{m}$, etc., we need some local conditions on F and G. Hence, we assume

[A1] F is continuous at m with F' continuous and negative at m,

[A2] G continuous at m and $G(m) > 0$,

and that, additionally, F and G do not have any common points of discontinuity in $(0,m)$. Without further reference, we assume that there is a Δ_o (> 0) such that $H(m + \Delta_o) > 0$, and that the conditions above are satisfied.

3. Results.

The estimator $\hat{m}$ defined by (2.6) satisfies the following theorems. Their proofs are deferred to Sections 4 and 5.

THEOREM 3.1. With σ^2 defined by (1.3), $\sqrt{n}(\hat{m} - m)$ converges in law to the normal distribution with mean zero and variance σ^2.

THEOREM 3.2. For each $p > 0$, $||\hat{m} - m||_p = O(n^{-1/2})$ where the constant in the order could depend on p.

The asymptotic distribution in Theorem 3.1 involves σ^2. To find approximate confidence intervals for m, a consistent estimator of σ^2 is desired. For this, we let k be a real valued function such that k vanishes off $(0,1)$, $\int k = 1$, and its derivative is bounded. Then define $\hat{f}$ by

$$-\varepsilon_n \hat{f}(\cdot) = \int k((t - \cdot)/\varepsilon_n) dF_n(t), \tag{3.1}$$

where ε_n converges to zero as $n \to \infty$. Since $C(t) = \int_o^t H^{-2} d\tilde{H}$, we estimate $\sigma^2 = C(m)/f^2(m)$ by

$$\hat{\sigma}^2 = \int_0^{\hat{m}} (H_n^{-2}) d\tilde{H}_n / (\hat{f}(\hat{m}))^2, \tag{3.2}$$

where $nH_n(\cdot) = \sum_{i=1}^n [Z_i > \cdot]$ and $n\tilde{H}_n(\cdot) = \sum_{i=1}^n \delta_i[Z_i \leq \cdot]$ and $\hat{m}$ is the estimator of m defined by (2.6).

THEOREM 3.3. $\hat{\sigma}^2$ is a consistent estimator of σ^2 if $n\varepsilon_n^2 \to \infty$.

REMARK 3.1. The asymptotic distribution given by Theorem 3.1 is the same as the asymptotic distribution given by Sander (1975) or Reid (1981). However, their results do not contain the moment covergence results for their estimator as Theorem 3.2 does for our $\hat{m}$. In fact the conclusion of Theorem 3.2 can be strengthened in the following way. One can show that $E[|\hat{m} - m|^p] = c_p/n^{p/2} + o(n^{-p/2})$ for some constant c_p. Note that c_2 comes out to be σ^2 as in (1.3). Results of this type are needed to obtain sequential estimators of m. The proof of the above expansion is quite involved and will appear in a separate paper.

REMARK 3.2. Since a_n diverges to infinity, $\hat{m}$ does not differ from $m(F_n)$ with high probability and $m(F_n)$ is simply an order statistic of $\{Z_i\}_{i=1}^n$. Since F_n decreases only at uncensored observations, $m(F_n)$ is necessarily an uncensored observation.

REMARK 3.3. Here, an estimator of f(m) was based on the kernel function k as in Blum and Susarla (1980). However, one could also consider the histogram type estimators such as those studied by Liu and Van Ryzin (1985).

REMARK 3.4. With appropriate modifications of the assumptions [A1], [A2], etc., we can obtain estimators $\hat{m}_\alpha$ of the α-th quantile $(0 < \alpha < 1)$ of F with $\hat{m}_\alpha$ satisfying conclusions similar to Theorems 3.1, 3.2, and 3.3. In fact, some results for the process $\{\hat{m}_\alpha : \alpha_0 < \alpha < \alpha_1\}$, $0 < \alpha_0 < \alpha_1 < 1$, are possible. However, these will not be pursued here.

The following theorem proved in Section 4 is used repeatedly in the proofs of Theorems 3.1 - 3.3.

THEOREM 3.4. For each $t > 0$,

$$F_n(t) - F(t) = \{n^{-1}\sum \xi(Z_i,\delta_i,t) + R_n(t)\}F(t),$$

where

(3.3) $$\xi(Z_i,\delta_i,t) = C(Z_i \wedge t) - \delta_i[Z_i \leq t]H^{-1}(Z_i) \text{ and}$$

(3.4) $$\sup_{t \leq c}||R_n(t)||_p = O(n^{-1}), \text{ if } p \geq 2 \text{ and } H(c) > 0.$$

4. Proofs.

We begin with a description of notation used throughout this section: i ranges from 1 through n; the range of summation is over i from 1 through n. Arguments will not be exhibited if they are clear from the context. c_1, c_2, etc. denote constants independent of n. We assume Δ (> 0) is such that $F'(m + \Delta_1)\, H(m + \Delta_1) < 0$ for $0 \leq \Delta_1 < \Delta$. Such a Δ is guaranteed to exist by [A1] and [A2]. All limits are as $n \to \infty$.

Proof of Theorem 3.1. Since $\sqrt{n}(m(F_n) - m)$ is shown to have an asymptotically normal distribution with zero mean and variance σ^2 by Sander (1975), it is enough to show that $\sqrt{n}(\hat{m} - m(F_n)) \to 0$ in probability. For this, let $\varepsilon > 0$. Then, $P(\sqrt{n}|\hat{m} - m(F_n)| > \varepsilon) \leq P(m(F_n) > a_n + \varepsilon/\sqrt{n}) \leq P(m(F_n) > m + \Delta)$, where the last inequality holds for sufficiently large n since $a_n \to \infty$. Now $P(m(F_n) > m + \Delta) \leq P(F_n(m + \Delta) - F(m + \Delta) > \eta)$ where $2\eta = 1 - 2F(m + \Delta)$. This last probability is at most $P(|n^{-1}\sum \xi_i(m + \Delta)| > \eta/2) + P(|R_n(m + \Delta)| > \eta/2)$.

Now the conclusion of Theorem 3.4 gives that $P(|R_n(m + \Delta)| > \eta/2)$ converges to zero. Since ξ_i are i.i.d. bounded random variables, it also follows that $P(|n^{-1}\sum \xi_i| > \eta/2)$ converges to zero, completing the proof of the theorem.

Proof of Theorem 3.2. To prove Theorem 3.2, first write $n^{r/2}||\hat{m} - m||_r^r = \int_0^\infty P(\sqrt{n}|\hat{m} - m| > t)dt^r$. Since $\hat{m} \leq a_n$ by definition, the range of the integral will be at most $\sqrt{n}\, a_n$. We break the resulting integral into two parts, one on $(0,\Delta\sqrt{n})$, and the other on $(\Delta\sqrt{n},\sqrt{n}\, a_n)$. For the integral on the latter range, it is observed that this integral will be at most $(\sqrt{n}\, a_n)^r\ P(|\hat{m} - m| > \Delta)$. So it is enough to show that $P(|\hat{m} - m| > \delta) = n^{-\gamma}$ for any positive γ, since it is assumed that $\overline{\lim}\ a_n/n^\beta < \infty$ for some β. We deal only with $P(\hat{m} > m + \Delta)$; the other case being similar. Note that

$$P(\hat{m} > m + \Delta) \leq P(m(F_n) > m + \Delta) \leq P(n^{-1}\sum \xi_i(m + \Delta) > \eta/2) + P(R_n(m + \Delta) > \eta/2),$$

where $2\eta = 1 - 2F(m + \Delta)$. The second probability on the r.h.s. converges to zero like $O(n^{-s})$ for any $s > 0$ by Theorem 3.4 and the first probability there converges to zero like $O(n^{-s})$ again for any $s > 0$ since ξ_i are i.i.d. bounded random variables. Hence, $(\sqrt{n}\, a_n)^r\ P(\hat{m} - m > \Delta) \to 0$. A similar argument also shows that $(\sqrt{n}\, a_n)^r\ P(\hat{m} - m < -\Delta) \to 0$.

We are now left with $\int_0^{\Delta\sqrt{n}} P(\sqrt{n}|\hat{m} - m| > t)dt^r$. It suffices to show that $\int_0^{\Delta\sqrt{n}} P(\sqrt{n}(\hat{m} - m) > t)dt^r$ is finite. The finiteness of the other integral can be shown in a similar fashion. Rewrite the integral as $P(\hat{m} > m + t/\sqrt{n})$ which is at most $P(m(F_n) > m + t/\sqrt{n})$ which in turn can be bounded by

$P(F_n(m + t/\sqrt{n}) - F(m + t/\sqrt{n}) > F(m) - F(m + t/\sqrt{n}))$. Since $t/\sqrt{n} \leq \Delta$, this last inequality is at most $P(F_n(m + t/\sqrt{n}) - F(m + t/\sqrt{n}) > tc_1/\sqrt{n})$ for some $c_1 > 0$ because f is continuous and positive at m. Hence

$\int_0^{\Delta\sqrt{n}} P(\sqrt{n}(\hat{m} - m) > t)dt^r \leq n^{r/2}\int_0^\Delta P(F_n(m + u) - F(m + u) > c_1u)du^r$. It is enough to consider the last integral on the range $(c_2/\sqrt{n}, \Delta)$ to obtain the result.

Now by Theorem 3.4, $P(F_n(m + u) - F(m + u) > c_1u)$ is dominated by $P(|n^{-1}\sum \xi_i(m + u)| > c_1u/2) + P(|R_n(m + u)| > c_1u/2)$ where the second term is $u^{-s}O(n^{-s})$ for any $s > 0$ by Theorem 3.4. Thus,

$n^{r/2}\int_{c_2/\sqrt{n}}^\Delta P(|R_n(m + u)| > c_1u/2)du^r = O(1)$. Since $\xi_i(m + u)$ are i.i.d. bounded random variables, we have by Hoeffding's inequality (1963),

$P(|n^{-1}\sum \xi_i(m + u)| > c_1u/2) \leq d_4 \cdot \exp(-d_3nu^2)$, where d_3, d_4 are constants independent of n and u.

Hence, $e^{r/2}\int_0^\Delta P(|n^{-1}\sum \xi_i(m + u)| > c_1u/2)du^r = O(1)$, and this completes the proof.

Proof of Theorem 3.3. We show that $\hat{C}(\hat{m}) \to C(m)$ and $\hat{f}(\hat{m}) \to f(m)$, both in probability. Let $\varepsilon > 0$ and $\Delta > 0$ be such that $\int_m^{m+\Delta} d\tilde{H}/H^2 < \varepsilon$ and $\int_{m-\Delta}^m H^{-2} d\tilde{H} < \varepsilon$. Since $\hat{m} \to m$ in probability, we consider the $\hat{C}(\hat{m}) - C(m)$ on the set

$[|\hat{m} - m| < \delta] = A$. On A, $|\hat{C}(\hat{m}) - C(m)| < \int_m^{m+\Delta} \hat{H}^{-2} d\hat{\tilde{H}} + \int_{m-\Delta}^m \hat{H}^{-2} d\hat{\tilde{H}} +$

$|\int_0^m (\hat{H}^{-2} - H^{-2}) d\hat{\tilde{H}}| + |\int_0^m (\hat{\tilde{H}} - \tilde{H}) dH^{-2}| + |H^{-2}(m)|\ |\hat{\tilde{H}}(m) - \tilde{H}(m)|$. Hence $P(A \cap (|\hat{C}(\hat{m}) - C(m)| > 5\varepsilon))$ will converge to zero because the first two quantities on the r.h.s. of the above inequality converge in probability to $\int_m^{m+\Delta} H^{-2} d\tilde{H}$ and $\int_{m-\Delta}^m H^{-2} d\tilde{H}$ with both of these limits less than ε, the third quantity converges to zero by the strong law of large numbers since $H(m) > 0$ and the last two quantities can be shown to converge to zero by using the fact that

$\sup\{||\hat{\tilde{H}}(x) - \tilde{H}(x)||_p : 0 < x < m\} = 0(n^{-1/2})$.

Now consider $\hat{f}(\hat{m}) - f(m)$. Write $\hat{f}(\hat{m}) - f(m)$ as $\{\hat{f}(\hat{m}) - \hat{f}(m)\} + \{\hat{f}(m) - f(m)\} = I + II$. By the mean value theorem, $I = (\hat{m} - m)\hat{f}'(m^*)$ with the consequence that $|I| < |\hat{m} - m|/\varepsilon_n^2 = 0(\varepsilon_n^{-2} n^{-1/2})$ by Theorem 3.2. So $I = 0(1)$ by our assumption on ε_n. For II, we have $\hat{f}(m) - f(m) = [-\varepsilon_n^{-1} \int k((x - m)/\varepsilon_n) d(\hat{F}(x) - F(x))] - [\varepsilon_n^{-1} \int k((x - m)/\varepsilon_n) dF(x) + f(m)]$. By standard arguments for kernels (see, e.g., Parzen (1962)), the second nonstochastic term converges to zero since f is continuous at m. Since k has support in (0,1), integration by parts gives that the first term in absolute value is at most $\varepsilon_n^{-1} \sup\{|\hat{F}(x) - F(x)| : m < x < m + \Delta\}$ for sufficently large n. But by the results available in the literature on the Kaplan-Meier process $Z = n^{1/2} (\hat{F}-F)F^{-1}$ on $(m, m+\Delta)$ (see, e.g., Gill, 1983, Theorem 1.1) this bound is $0_p(\varepsilon_n^{-1} n^{-1/2})$. This becomes $o_p(1)$ by our assumption on ε_n, which completes the proof.

Proof of Theorem 3.4. First write $F_n - F$ in the form $\{\exp(\ln(F_n/F)) - 1\} F$ and then via a two-term Taylor expansion we get

$$F_n - F = \{\ln\{F_n/F\} + \frac{1}{2} e^{c*}(\ln(F_n/F))^2\}F, \tag{4.1}$$

where $c*$ lies between 0 and $\ln(F_n/F)$. From (1.1) we have

$$\ln F_n(t) = \sum[Z_i < t, \delta_i = 1]\ln\{N(Z_i)(1 + N(Z_i))^{-1}\}.$$

Expanding the logarithm term, we can write

$$\ln F_n(t) = -\sum[Z_i < t, \delta_i = 1]\{(1 + N(Z_i))^{-1} + r(Z_i)\}, \tag{4.2}$$

where $r(Z_i) = (1 + N(Z_i))^{-2}\sum_{j=0}^{\infty} (j + 2)^{-1}(1 + N(Z_j))^{-j}$. Write $n(1 + N)^{-1} = \{n(1 + N)^{-1} - 2H^{-1} + H_nH^{-2}\} + 2H^{-1} - H_nH^{-2}$ and note that

$$n^{-1}\sum[Z_i < t, \delta_i = 1]H_n(Z_i)H^{-2}(Z_i) = n^{-1}(n - 1)U_n(t),$$

where $U_n(t)$ is a U-statistic of degree 2 with kernel

$$\Phi_t((Z_1,\delta_1),(Z_2,\delta_2)) = \frac{1}{2}\{\delta_1[Z_1 < t, Z_2 > Z_1]H^{-2}(Z_1) + \delta_2[Z_2 < t, Z_1 > Z_2]H^{-2}(Z_2)\}. \tag{4.3}$$

Hence, (4.2) takes the form

$$\ln F_n(t) = -2n^{-1}\sum \delta_i[Z_i < t]H^{-1}(Z_i) + n^{-1}(n - 1)U_n(t) + r_1(t),$$

where

$$r_1(t) = n^{-1}\sum \delta_i[Z_i< t]\{n(1 + N(Z_i))^{-1} - 2H^{-1}(Z_i) - H_n(Z_i)H^{-2}(Z_i) + nr(Z_i)\}. \tag{4.4}$$

Finally, replace the U-statistic U_n by its projection $\tilde{U}_n$, namely

$$\tilde{U}_n = 2n^{-1}\sum \{E(\Phi|(Z_i,\delta_i))\} + \ln F(t).$$

By direct computation, with $C(t) = \int_0^t H^{-2}d\tilde{\tilde{H}}$, we have

$$E(\Phi|(Z_i,\delta_i)) = \frac{1}{2}\{\delta_i[Z_i < t]H^{-1}(Z_i) + C(Z_i \wedge t)\}.$$

Collecting our results this leads to

$$\ln\{F_n(t)/F(t)\} = n^{-1}\sum\{C(Z_i \wedge t) - \delta_i[Z_i < t]H^{-1}(Z_i)\} + (U_n - \tilde{U}_n) - n^{-1}U_n - r_1(t)$$

Therefore, from (4.2) this yields

$$F_n(t) - F(t) = n^{-1}\sum \xi(Z_i,\delta_i,t) + R_n(t),$$

where $R_n(t) = (U_n - \tilde{U}_n) - n^{-1}U_n - r_1(t) + \frac{1}{2}e^{c*}\{\ln(F_n(t)/F(t))\}$. To complete the proof of the lemma we treat separately each of the terms in $R_n(t)$. Clearly $\sup_{t<c} \Phi_t < \Phi_d$ and so $||\sup_{t<c} U_n(t)||_p = O(1)$ for any $p > 0$. Also note that in (4.2), $r(Z_i) < N^{-2}(Z_i)$ and so

$$(4.5) \qquad ||\sup_{t<c} \textstyle\sum_{i=1}^n r(Z_i)\delta_i[Z_i < t]||_p < n||n^{-1}\sum N^{-2}(Z_i)\delta_i[Z_i < c]||_p.$$

On applying the Hölder inequality with $p > 1$ and then evaluating the expectation by first conditioning with respect to (δ_i,Z_i), we get that the right-hand side of (4.5) is bounded by

$$n(n^{-1}\sum E\{\delta_i[Z_i < c]E\{N^{-2p}(Z_i)|(\delta_i,Z_i)\}\})^{1/p}.$$

The inner conditional expectation is an inverse moment of a left-truncated binomial random variable with parameters $(n, H(Z_i))$ and therefore it is bounded by $C_p(n,H(Z_i))^{-2p}$. Arranging our results, we finally get (suppressing constants), for $p > 1$

$$(4.6)\qquad \|\sup_{t\leq c} \textstyle\sum r(Z_i)\delta_i[Z_i \leq t]\|_p \leq n^{-1}\int_0^t H^{-2}d\tilde{\tilde{H}} = O(n^{-1}).$$

To handle the other terms in $r_1(t)$ of (4.4) use the fact that

$$|n(1 + N)^{-1} - (2H^{-1} - H^{-2}H_n)| \leq 2N^{-1}H^{-1} + nN^{-1}H^{-2}|H_n - H|^2 + 2N^{-1}H^{-2}|H_n - H|.$$

Following the same arguments for each term as those leading to (4.6), we get, for $p > 1$,

$$(4.7)\qquad \|\sup_{t\leq c} r_1(t)\|_p = O(n^{-1}).$$

Observe that $\exp(c^*)\ F(t) \leq 1$ and that

$$\begin{aligned}(4.8)\qquad |\ln\{F_n(t)/F(t)\}| \leq\ & \{n^{-1}\textstyle\sum \delta_i[Z_i \leq t]H^{-1}(Z_i) + \ln F(t)\} \\ & + n^{-1}\textstyle\sum \delta_i[Z_i \leq t]\{n(1 + N(Z_i))^{-1} - H^{-1}(Z_i)\} \\ & + \textstyle\sum \delta_i[Z_i \leq t]N^{-2}(Z_i)\end{aligned}$$

For the first term in (4.8) apply the Marcinkiewicz-Zygmund inequality. With $p > 1$ and suppressing constants, we get

$$\|n^{-1}\textstyle\sum \delta_i[Z_i \leq t]H^{-1}(Z_i) + \ln F(t)\|_p \leq \|n^{-1}(\textstyle\sum \delta_i[Z_i \leq t]H^{-2}(Z_i))^{1/2}\|_p.$$

Applying the Hölder inequality with $p \geq 2$ yields a bound for the right-hand side as $(\int_0^t H^{-p}d\tilde{\tilde{H}})n^{-1/2}$. The remaining terms in (4.8) are handled as before. Thus we have obtained, with $p > 1$

$$(4.9)\quad \sup_{t\leq c}\|(\ln(F_n(t)/F(t)))^2 e^{c^*}F(t)\|_p \leq \sup_{t\leq c}(\|\ln(F_n(t)/F(t))\|_{2p})^2.$$

Finally to handle $(U_n(t) - \tilde{U}_n(t))$ we employ the arguments in Serfling (1980, Lemma B, page 186). We get, with $p \geq 2$

$$\sup \| U_n(t) - \tilde{U}_n(t) \|_p = O(n^{-1}). \tag{4.10}$$

From (4.6) - (4.10), (3.2) of Theorem 3.4 holds. This completes the proof.

REFERENCES

Blum, J. and Susarla, V. (1980). Maximal deviation theory of density and failure rate function estimates based on censored data. Multivariate Analysis - V, P.R. Krishniah, Ed., North-Holland, New York, 213-222.

Brookmeyer, R. and Crowley, J. (1982). A confidence interval for the median survival time. Biometrics **38** 29-41.

Emerson, R. (1982). Nonparametric confidence intervals for the median in the presence of right censoring. Biometrics **38** 17-27.

Gill, R. (1983). Large sample behavior of the product-limit estimator on the whole line. Ann. Statist. **11** 49-58.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. J. Amer. Statist. Assoc. **36** 13-30.

Liu, R. and Van Ryzin, J. (1985). A histogram estimator of the hazard rate with censored data. Ann. Statist. **13** 592-605.

Parzen, E. (1962). On the estimation of a probability density function and mode. Ann. Math. Statist. **33** 1065-1076.

Peto, R., Mike, M.C., Armitage, P., Breslow, N.E, Cox, D.R., Howard, S.V., Mantel, N., McPherson, K., Peto, J. and Smith, P.G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and Examples. Br. J. Cancer **35** 1-39.

Reid, N. (1981). Influence functions for censored data. Ann. Statist. **9** 78-92.

Sander, J. (1975). Asymptotic normality of linear combinations of functions of order statistics with censored data. Tech. Report No. 8, Div. Biostatist., Stanford Univ.

Serfling, R.J. (1980). Approximation Theorems of Mathematical Statistics. John Wiley & Sons, New York.