JOURNAL OF MULTIVARIATE ANALYSIS 50, 17–29(1994)

# The Kaplan–Meier Estimate for Dependent Failure Time Observations

Z. YING*

*University of Illinois*

AND

L. J. WEI†

*Harvard University*

In some long term medical follow-up studies, a series of dependent and possibly censored failure times may be observed. Suppose that these failure times were generated from the same distribution function, and inferences about it are of our main interest. In this article, we show that under rather weak conditions for the dependence among the observations, the Kaplan–Meier estimator is still consistent and asymptotically normal. For a special dependent case in which highly stratified data are observed, a valid estimate for the limiting variance of the Kaplan–Meier estimate is also provided. Our proposal is illustrated with an examply. © 1994 Academic Press, Inc.

## 1. INTRODUCTION

In the analysis of survival data it is often quite useful to summarize the survival experience of particular groups of study patients using the sample survival function, for example, the Kaplan and Meier (1958) (KM) estimator. If the failure time observations in the sample are assumed to be mutually independent, the KM estimator is consistent and asymptotically

Received January 2, 1992; revised

AMS 1980 subject classifications: primary 62G05, 62P10.

Keywords and phrases: censored data, Kaplan–Meier estimator, weak convergence, $\phi$-mixing, stratified data.

* Partially supported by NSF Grants DMS-90-01726 and DMS-92-07730 and NSA Grant MDA-904-92-H-3058.

† Partially supported by NIH Grants AI95030 and CA56844.



0047-259X/94 $6.00

normal (see Breslow and Crowley, 1974; Gill, 1980). Furthermore, for each study time point $t$, Greenwood's formula (Cox and Oakes, 1984, p. 50) provides a valid estimate for the asymptotic variance of the KM estimate at $t$. For a large sample, inferences about the survival experience can then be made based on this asymptotic theory.

However, it is not clear that these large sample properties for the KM estimate still hold when observations are dependent. For example, in the Diabetes Control and Complications Trial, DCCT (The DCCT Research Group, 1986), patients are randomly assigned to receive either experimental or standard therapy. The purpose of the study is to assess the relationship between glycemic control and the development or progression of early vascular complications in persons with insulin-dependent diabetes mellitus. Experimental therapy involves the use of an intensive insulin regimen designed to maintain near normal glycemic levels in the absence of severe hypoglycemia. Standard treatment is designed to maintain subjects free of clinical symptoms related to hyper- or hypoglycemia while receiving up to two insulin injections daily. In the prevention study of this trial, one of the principal outcomes is the initial appearance of background retinopathy for individual eyes of the study patients. Therefore, for patients in each treatment group, it is natural to treat each eye as a sample unit and use the KM estimate based on event times from individual eyes to summarize their "failure" experience. However, the event times of the two eyes of each patient are expected to be correlated. It is important to know if the resulting KM estimate is still consistent and asymptotically normal. Moreover, Greenwood's formula may not be valid in this case.

In this article, we study large sample properties of the KM estimator for cases in which the underlying failure times may be dependent. The conditions under which the KM estimate is consistent and asymptotically normal are rather mild. For highly stratified data, for example, in the DCCT, these conditions are trivially satisfied. A valid estimate of the asymptotic variance for the KM estimate in the stratified case is also provided. Since the DCCT is an ongoing trial and its data are not yet available to the public, our new proposal is illustrated with a small-scaled data set from an animal study.

## 2. The Consistency of the Kaplan–Meier Estimate

Let $X_1, ..., X_n$ be a sequence of failure times which may not be mutually independent, but have a common continuous marginal distribution function $F$. Let $C_1, ..., C_n$ be the corresponding censoring times which are assumed to be independent of the $X_i$ and may be regarded as nonrandom constants. The observations consist of $Z_i = \min(X_i, C_i)$ and

$\Delta_i = I_{(X_i \leqslant C_i)}$, $i = 1, ..., n$, where $I_{(\cdot)}$ is the indicator function. The Kaplan–Meier estimator for $1 - F$ is

$$1 - \hat{F}_n(t) = \prod_{i : Z_i \leqslant t} \left(1 - \frac{\Delta_i}{Y_n(Z_i)}\right),$$

where $Y_n(t) = \sum_{i=1}^n I_{(Z_i \geqslant t)}$. When there is no censoring, convergence (in probability) of $\hat{F}_n(t) = n^{-1} \sum_{i=1}^n I_{(X_i \leqslant t)}$ to $F(t)$ should hold if the correlation between $I_{(X_k \leqslant t)}$ and $I_{(X_{k+i} \leqslant t)}$ decreases to zero as $i$ increases to infinity. Now, let

$$N_n(t) = \sum_{i=1}^n I_{(Z_i \leqslant t, \Delta_i = 1)}, \quad G_n(t) = n^{-1} \sum_{i=1}^n I_{(C_i > t)},$$

$$\text{and} \qquad \hat{\Lambda}_n(t) = \int_0^t \frac{dN_n(s)}{Y_n(s)}. \tag{2.1}$$

Note that $\hat{\Lambda}_n$ is the usual so-called Nelson estimator for the underlying cumulative hazard function $\Lambda(t) = -\log(1 - F(t))$. Also, let

$$\tau_F = \sup\{t : F(t) < 1\}, \quad \text{and} \quad \tau_G = \sup\{t : \liminf_n G_n(t) > 0\}. \tag{2.2}$$

Then under some mild conditions on $X$'s, $\hat{F}_n$ is still consistent. This is summarized in

THEOREM 1. *Let $\phi(1)$, $\phi(2)$, ... be a sequence of nonnegative numbers such that*

$$|\text{cov}\,(I_{(X_k \leqslant s)}, I_{(X_{k+i} \leqslant t)})| \leqslant \phi(i), \tag{2.3}$$

*for all $k$, $i$, $s$, and $t$.*

(i) *If $\phi(i) \to 0$ as $i \to \infty$, then for $\tau < \min(\tau_F, \tau_G)$,*

$$\sup_{0 \leqslant t \leqslant \tau} |\hat{F}_n(t) - F(t)| \to 0, \qquad \textit{in probability, as } n \to \infty. \tag{2.4}$$

(ii) *If $\sum_{i=1}^{\infty} \phi(i) < \infty$, then for $\tau < \min(\tau_F, \tau_G)$,*

$$\sup_{0 \leqslant t \leqslant \tau} |\hat{F}_n(t) - F(t)| \to 0 \qquad \textit{a.s., as } n \to \infty. \tag{2.5}$$

(iii) *If $\tau_F \leqslant \tau_G$ and $F(\tau_F -) = 1$, then under the assumptions of* (i), (2.4) *holds for $\tau = \tau_F$, while under the assumptions of* (ii), (2.5) *holds for $\tau = \tau_F$.*

*Proof.* See the Appendix.

Theorem 1 indicates that if for a fixed $k$ the dependency between $X_k$ and $X_{k+i}$ becomes weaker and weaker as $i$ increases, then $\hat{F}_n$ is consistent. For highly stratified data, for example, the paired failure time observations in the DCCT, these conditions in (i) and (ii) are trivially satisfied.

## 3. The Large-Sample Distribution for the KM Estimator

To obtain the asymptotic distribution of the KM estimate for the dependent case, a stronger condition on $\{X_i, i \geq 1\}$ than those in (i) and (ii) of Theorem 1 is needed. A sequence $\{X_i, i \geq 1\}$ is called $\phi$-mixing for a sequence of nonnegative constants $\{\phi(i), i \geq 0\}$, where $\phi(i) \to 0$ as $i \to \infty$, if for any $k$, any set $A$ generated by $X_j, j \leq k$ and any $B$ generated from $X_j, j \geq k+i$,

$$|P(B \mid A) - P(B)| \leq \phi(i). \tag{3.1}$$

An example of $\phi$-mixing sequence of random variables is $m$-dependent sequence. The $\{X_i, i \geq 1\}$ is an $m$-dependent sequence if $\{X_1, ..., X_n\}$ and $\{X_{n+j+1}, X_{n+j+2}, ...\}$ are independent classes of random variables for each $n \geq 1$ and $j \geq m$. Naturally, the highly stratified failure times can be easily ordered as an $m$-dependent sequence.

Now, let

$$M_i(t) = I_{(Z_i \leq t,\, \Delta_i = 1)} - \int_0^t I_{(Z_i \geq s)}\, d\Lambda(s),$$

and $M^{(n)}(t) = \sum_{i=1}^n M_i(t)$. Then for $t \geq 0$

$$\frac{n^{1/2}(\hat{F}_n(t) - F(t))}{1 - F(t)} = \int_0^t \frac{1 - \hat{F}_n(s-)}{1 - F(s)} \frac{d[n^{1/2} M^{(n)}(s)]}{n^{-1} Y_n(s)} \tag{3.2}$$

(see Gill, 1980, 3.2.13). If $X$'s are mutually independent, $M^{(n)}(s)$ is a martingale in $s$. The limiting distribution of $\hat{F}_n$ can be derived from Rebolledo's martingale central limit theorem (Gill, 1980). For dependent cases, we summarize the results in

Theorem 2. *Suppose that $\{X_i; i \geq 1\}$ is $\phi$-mixing in the sense of* (3.1) *with $\sum \phi(i)\, i^2 < \infty$. Assume that*

$$H(s, t) = \lim_{n \to \infty} \frac{1}{n} \sum_{1 \leq i, j \leq n} E\{M_i(s)\, M_j(t)\} \tag{3.3}$$

*exists for $s, t \leq \tau < \min(\tau_F, \tau_G)$ and that $\lim_{n \to \infty} G_n(t) = G(t)$ for $t \leq \tau < \min(\tau_F, \tau_G)$. Then $n^{1/2}(\hat{F}_n - F)/(1 - F)$ converges weakly in $\mathscr{D}[0, \tau]$ to a* *zero-mean Gaussian process $W$ with $W(0)=0$. The covariance function of $W$ is*

$$E[W(s)\,W(t)] = \int_0^s \int_0^t \frac{dH(u, v)}{G(u)[1-F(u)]\,G(v)[1-F(v)]}. \tag{3.4}$$

*This implies that $n^{1/2}(\hat{F}_n - F)$ converges weakly to $(1-F)\,W$.*

*Proof.* See the Appendix.

Note that if $\{X_i\}$ is stationary, then (3.3) is satisfied. If $X$'s are independent, then $E\{M_i(s)\,M_j(t)\}=0$ for $i \neq j$ and $E\{M_i(s)\,M_i(t)\} = \int_0^{t \wedge s} EI_{(C_i \geq u)}\,dF(u)$. It follows then that $E[W(s)\,W(t)] = \int_0^{s \wedge t} dF(u)/[(1-F(u))^2\,G(u)]$. Furthermore, for stratified observations, the convergence of (3.3) holds under some mild conditions and a consistent estimator for (3.4) is presented in the next section.

## 4. The KM Estimator with Hightly Stratified Observations

Consider now a special case in which survival times are highly stratiefied. Let the observations in the $i$th stratum be denoted by $\{X_{ij}, j=1, ..., K_i\}$, $i=1, ..., m$. The corresponding "failure" indicator $\Delta$ is denoted by $\Delta_{ij}$. Assume that the depencency exists only among individual observations within each stratum. Also assume that the size of each stratum is small relative to $m$. Let $M_{ij}(t) = I_{(Z_{ij} \leq t,\, \Delta_{ij}=1)} - \int_0^t I_{(Z_{ij} \geq s)}\,d\Lambda(s)$. Then, if $K=\max\{K_i\}=o(m)$, the resulting KM estimator $\hat{F}_n$ for $F$ is consistent based on the proof of Theorem 1.

Furthermore, suppose that the corresponding $G_n(t)$ converges for $0 \leq t \leq \tau$, and that $K$ is bounded. Then, from Theorem 2, $n^{1/2}(\hat{F}_n - F)/(1-F)$ converges weakly to a Gaussian process $W$ with mean 0 and covariance function (3.4). For the present case, it is easy to see that $H(s, t)$ can be consistently estimated by

$$\hat{H}_n(s, t) = \frac{1}{n} \sum_{i=1}^{m} \sum_{j=1}^{K_i} \sum_{l=1}^{K_i} \left\{ I_{(Z_{ij} \leq s,\, \Delta_{ij}=1)} - \int_0^s I_{(Z_{ij} \geq u)}\,d\hat{\Lambda}_n(u) \right\}$$
$$\times \left\{ I_{(Z_{il} \leq t,\, \Delta_{il}=1)} - \int_0^t I_{(Z_{il} \geq u)}\,d\hat{\Lambda}_n(u) \right\}.$$

It follows that a valid estimator for the asymptotic variance of $\hat{F}_n(t)$ is

$$\hat{V}(t) = n(1-\hat{F}_n(t))^2 \int_0^t \int_0^t \frac{d\hat{H}_n(u, v)}{Y_n(u)\,Y_n(v)}.$$

Unfortunately the data from the DCCT are not available to the public. We use a smaller data set from a tumorigenesis in a litter-matched experiment (see Mantel and Ciminera, 1977). In this study, there are two controls in each litter. Table I gives the weeks of tumor appearance. The + indicates the week of death prior to any tumor. In the experiment conducted, all rats were sacrified at the end of 104 weeks. For each triplet $(X_1, X_2, X_3)$ in the table, $X_1$ is the observation from a drug-treated rat and $X_2$ and $X_3$ are the responses from the corresponding litter-matched controls. Suppose that one is interested in estimating the common marginal distribution $F$ of the tumor appearance time for the controls. With those 100 controls in Table I, the KM estimates $1-\hat{F}^*$ and the corresponding standard error estimates based on $\hat{V}$ at various time points are reported in Table II. For comparisons, we also report the standard error estimates from the canned statistical package BMDP by ignoring the litter effect. For this example, the two variance estimates are not drastically different. This may be due to a weak litter effect. On the other hand, if the failure times are highly positively correlated in each stratum, one would expect that our variance estimate tends to be much larger than the standard estimate with independent observations.

TABLE I

Time (in weeks) to Tumor Appearance in a Litter-Matched Tumorigenesis Experiment

| Drug treated | Control 1 | Control 2 |
|---|---|---|
| 101.0+ | 49.0 | 104.0+ |
| 104.0+ | 102.0+ | 104.0+ |
| 104.0+ | 104.0+ | 104.0+ |
| 77.0+ | 97.0+ | 79.0+ |
| 89.0+ | 104.0+ | 104.0+ |
| 88.0 | 96.0 | 104.0+ |
| 104.0 | 94.0+ | 77.0 |
| 96.0 | 104.0+ | 104.0+ |
| 82.0+ | 77.0+ | 104.0+ |
| 70.0 | 104.0+ | 77.0+ |
| 89.0 | 91.0+ | 90.0+ |
| 91.0+ | 70.0+ | 92.0+ |
| 39.0 | 45.0+ | 50.0 |
| 103.0 | 69.0+ | 91.0+ |
| 93.0+ | 104.0+ | 103.0+ |
| 85.0+ | 72.0+ | 104.0+ |
| 104.0+ | 63.0+ | 104.0+ |
| 104.0+ | 104.0+ | 74.0+ |
| 81.0+ | 104.0+ | 69.0+ |

*Table continued*

TABLE I (*continued*)

| Drug treated | Control 1 | Control 2 |
|---|---|---|
| 67.0 | 104.0+ | 68.0 |
| 104.0+ | 104.0+ | 104.0+ |
| 104.0+ | 104.0+ | 104.0+ |
| 104.0+ | 83.0+ | 40.0 |
| 87.0+ | 104.0+ | 104.0+ |
| 104.0+ | 104.0+ | 104.0+ |
| 89.0+ | 104.0+ | 104.0+ |
| 78.0+ | 104.0+ | 104.0+ |
| 104.0+ | 81.0 | 64.0 |
| 86.0 | 55.0 | 94.0+ |
| 34.0 | 104.0+ | 54.0 |
| 76.0+ | 87.0+ | 74.0+ |
| 103.0 | 73.0 | 84.0 |
| 102.0 | 104.0+ | 80.0+ |
| 80.0 | 104.0+ | 73.0+ |
| 45.0 | 79.0+ | 104.0+ |
| 94.0 | 104.0+ | 104.0+ |
| 104.0+ | 104.0+ | 104.0+ |
| 104.0+ | 101.0 | 94.0+ |
| 76.0+ | 84.0 | 78.0 |
| 80.0 | 81.0 | 76.0+ |
| 72.0 | 95.0+ | 104.0+ |
| 73.0 | 104.0+ | 66.0 |
| 92.0 | 104.0+ | 102.0 |
| 104.0+ | 98.0+ | 73.0+ |
| 55.0+ | 104.0+ | 104.0+ |
| 49.0+ | 83.0+ | 77.0+ |
| 89.0 | 104.0+ | 104.0+ |
| 88.0+ | 79.0+ | 99.0+ |
| 103.0 | 91.0+ | 104.0+ |
| 104.0+ | 104.0+ | 79.0 |

TABLE II

Distribution of Tumor Appearance Time for the Controls

| Study time $t$ (weeks) | KM estimate | Standard error | |
|---|---|---|---|
| | | New | BMDP |
| 70 | 0.9190 | 0.026 | 0.028 |
| 80 | 0.8733 | 0.032 | 0.034 |
| 90 | 0.8227 | 0.046 | 0.041 |
| 100 | 0.8074 | 0.047 | 0.043 |

## 5. Remarks

The consistency and asymptotic normality for the KM estimate presented in this article hold under rather mild conditions. However, it seems rather difficult, if not impossible, to obtain a general estimate for the limiting variance of the KM estimate for the dependence case. A valid variance estimate is provided for the highly stratified observations in Section 4. Now, if the sequence $\{(X_i, C_i), i \geq 1\}$ is stationary and $\phi$-mixing with respect to the $\phi$ sequence in Theorem 2, then a valid variance estimate for $n^{1/2}\hat{F}_n(t)$ is

$$\frac{(1-\hat{F}_n(t))^2}{n}\left\{\sum_{|i-j| \leq n^{1/3}} D(i, t)\, D(j, t)\right\},$$

where

$$D(i, t) = \frac{\Delta_i I_{(Z_i \leq t)}}{n^{-1} Y_n(Z_i)} - \sum_{l=1}^{n} \frac{I_{(Z_l \leq t)} I_{(Z_i \geq Z_l)} \Delta_l}{n^{-1} Y_n^2(Z_l)}.$$

This estimate is useful for a survival study when there may be a seasonal time trend among the observations.

## Appendix

*Proof of Theorem* 1. We first show that

$$\sup_{0 \leq t \leq \tau} |\hat{\Lambda}(t) - \Lambda(t)| = o_p(1) \tag{A.1}$$

hods under the assumptions of Theorem 1(i), while

$$\sup_{0 \leq t \leq \tau} |\hat{\Lambda}(t) - \Lambda(t)| = o(1) \qquad \text{a.s.} \tag{A.2}$$

holds under the assumptions of (ii).

From (2.3) and the assumption that $\phi(\mathrm{i}) \to 0$, it is easy to see that for every $t \leq \tau$,

$$E\left[\frac{1}{n}(Y_n(t) - EY_n(t))\right]^2 \to 0, \qquad \text{as} \quad n \to \infty,$$

which, together with the Chebyshev inequality, implies

$$n^{-1} Y_n(t) - G_n(t)(1 - F(t)) = o_p(1). \tag{A.3}$$

Likewise,

$$n^{-1} N_n(t) - \int_0^t G_n(s)\, dF(s) = o_p(1). \tag{A.4}$$

From (A.3) and (A.4) we next show that $\hat{\Lambda}_n(t) - \Lambda(t) = o_p(1)$ for every $t \leqslant \tau$. Because $F$ is continuous, so is $\int_0^u G_n(s)\, dF(s)$. Thus for every $n$ and $r$, we can find $0 = u_0 < u_1 < \cdots < u_r = t$ such that $\int_{u_{i-1}}^{u_i} G_n(s)\, dF(s) = \eta \leqslant r^{-1}$. Note that $\eta$ here depends on $n$ and $r$. Furthermore, for fixed $r$, (A.3) and (A.4) imply

$$\max_{0 \leqslant i \leqslant r} \left\{ |n^{-1} Y_n(u_i) - G_n(u_i)(1 - F(u_i))| + \left| n^{-1} N_n(u_i) - \int_0^{u_i} G_n(s)\, dF(s) \right| \right\} = o_p(1) \tag{A.5}$$

Now

$$\begin{aligned} \hat{\Lambda}_n(t) = \int_0^t \frac{dN_n(s)}{Y_n(s)} &\leqslant \sum_{i=1}^r \frac{N_n(u_i) - N_n(u_{i-1})}{Y_n(u_i)} \\ &= \sum_{i=1}^r \frac{EN_n(u_i) - EN_n(u_{i-1})}{EY_n(u_i)} + o_p(1) \\ &= \sum_{i=1}^r \frac{\eta}{n^{-1}\, EY_n(u_i)} + o_p(1). \end{aligned} \tag{A.6}$$

Similarly,

$$\hat{\Lambda}_n(t) \geqslant \sum_{i=1}^r \frac{\eta}{n^{-1} EY_n(u_{i-1})} + o_p(1) \tag{A.7}$$

But $\sum_{i=1}^r n\eta / EY_n(u_i) \geqslant \Lambda(t) \geqslant \sum_{i=1}^r n\eta / EY_n(u_{i-1})$ and $\sum_{i=1}^r n\eta / EY_n(u_i) - \sum_{i=1}^r n\eta / EY_n(u_{i-1}) \leqslant r^{-1}\, n / EY_n(\tau)$, which can be made arbitrarily small since $r$ is arbitrary. These along with (A.6) and (A.7) entail $\hat{\Lambda}_n(t) - \Lambda(t) = o_p(1)$ for each fixed $t$. Now for any $\varepsilon > 0$, we can choose, because of continuity of $\Lambda, k$ and $0 = t_0 < t_1 < \cdots < t_k = \tau$ such that $\Lambda(t_i) - \Lambda(t_{i-1}) \leqslant \varepsilon$. A simple algebra shows that

$$\begin{aligned} \sup_{0 \leqslant t \leqslant \tau} |\hat{\Lambda}_n(t) - \Lambda(t)| &\leqslant \max_{1 \leqslant i \leqslant k} |\hat{\Lambda}_n(t_i) - \Lambda(t_i)| + \max_{1 \leqslant i \leqslant k} |\Lambda(t_i) - \Lambda(t_{i-1})| \\ &\leqslant o_p(1) + \varepsilon, \end{aligned}$$

which implies (A.1) since $\varepsilon$ is arbitrary.

For (A.2), we first show that

$$n^{-1}\,|Y_n(t)-EY_n(t)|=o(1) \qquad \text{a.s.} \tag{A.8}$$

for every $t \leqslant \tau$. Let $\xi_i = I_{(Z_i \geqslant t)} - EI_{(Z_i \geqslant t)}$. By Kronecker's lemma, (A.8) is implied by the almost sure convergence of the partial sum $S_n = \sum_{i=1}^n \xi_i/i$. The first step to prove the convergence is to show that the subsequence $S_{2^k}$ converges a.s. By the mixing condition,

$$\begin{aligned} E(S_{n+k}-S_n)^2 &\leqslant 2 \sum_{i=n+1}^{n+k} \sum_{j=i}^{n} \frac{\phi(j-i)}{ij} \\ &\leqslant 2 \sum_{j=0}^{\infty} \phi(j) \sum_{i=n+1}^{n+k} \frac{1}{i^2} \leqslant \frac{2}{n} \sum_{j=0}^{\infty} \phi(j), \end{aligned}$$

which implies that $S_n$ is an $L^2$ Cauchy sequence. Thus there exists a random variable $S$ such that $E(S_n - S)^2 \to 0$. Again, by (A.9)

$$E(S-S_{2^k})^2 = \lim_{n\to\infty} E(S_{2^k+n}-S_{2^k})^2 \leqslant \frac{1}{2^{k-1}} \sum_{j=0}^{\infty} \phi(j). \tag{A.10}$$

Therefore, by a simple application of the Chebyshev inequality and the Borel–Cantelli lemma, $S_{2^k} \to S$ a.s. From the method of subsequences (cf. Stout, 1974, Lemma 2.3.1) in order to prove $S_n \to S$, it suffices to show

$$\max_{2^k < n \leqslant 2^{k+1}} |S_n - S_{2^k}| \to 0 \qquad \text{a.s.} \tag{A.11}$$

as $k \to \infty$. Using a dyadic expansion method (cf. Stout, 1974, pp. 16–18) and the inequality (A.9), we get

$$E\{\max_{2^k < n \leqslant 2^{k+1}} (S_n - S_{2^k})^2\} \leqslant \frac{(k+1)^2}{2^{k-1}} \sum_{j=0}^{\infty} \phi(j) \tag{A.12}$$

From (A.12), the Chebyshev inequality and the Borel–Cantelli lemma follows (A.11). Hence (A.8) holds. Similarly, we can show that

$$n^{-1}N_n(t) - \int_0^t G_n(s)\,dF(s) = o(1) \qquad \text{a.s.} \tag{A.13}$$

From (A.8) and (A.13), we can proceed in exactly the same way as in proving (A.1) except replacing "$o_p(1)$" by "$o(1)$ a.s." to show (A.2). The details for this are omitted.

By (3.2.12) of Gill (1980), when $Y_n(t) > 0$,

$$\frac{\hat{F}_n(t)-F(t)}{1-F(t)} = \int_0^t \frac{1-\hat{F}_n(s-)}{1-F(s)}\, d[\hat{\Lambda}_n(s) - \Lambda(s)]. \tag{A.14}$$

Since $1-\hat{F}_n(t-)\geqslant 1-\hat{F}_n(\tau-)\geqslant n^{-1}Y_n(\tau)$,

$$\int_0^\tau \left| d\frac{1-\hat{F}_n(s-)}{1-F(s)} \right| = O_p(1) \tag{A.15}$$

under the assumptions of (i), while

$$\int_0^\tau \left| d\frac{1-\hat{F}_n(s-)}{1-F(s)} \right| = O(1) \qquad \text{a.s.} \tag{A.16}$$

under the assumptions of (ii). Thus (2.4) follows from (A.1), (A.14), and (A.15), whereas (2.5) follows from (A.2), (A.14), and (A.16).

Finally, part (iii) follows from (i), (ii), and the fact that $F(\tau)\uparrow 1$ as $\tau\uparrow\tau_F\wedge\tau_G=\tau_F$. ∎

*Proof of Theorem* 2. From (3.2) it is clear that a key step in proving Theorem 2 is to establish a weak convergence result for $n^{-1/2}M^{(n)}$. We will show later that $n^{-1/2}M^{(n)}$ converges weakly in $\mathscr{D}[0,\tau]$ to a Gaussian process $W_1$ with $W_1(0)=0$, $EW_1(t)=0$, and

$$E[W_1(s)\,W_1(t)]=H(s,t). \tag{A.17}$$

Since $\hat{F}_n\to F$ by Theorem 1, it follows from (3.2), (A.1), and (A.17) that

$$(\hat{F}_n, n^{-1}Y_n, n^{-1/2}M^{(n)}) \xrightarrow{\mathscr{D}[0,\tau]} (F, (1-F)\,G, W_1). \tag{A.18}$$

Therefore, by the Skorokhod–Dudley–Wichura theorem (Shorack and Wellner, 1986, p. 47), there exists a special construction $(\hat{F}_n^*, n^{-1}Y_n^*, n^{-1/2}M_n^*)$, which has the same distribution as $(\hat{F}_n, n^{-1}Y_n, n^{-1/2}M^{(n)})$ and which converges to $(F,(1-F)\,G, W_1^*)$ a.s., where $W_1^*$ has the same probability distribution as $W_1$. The convergence of $n^{-1/2}M_n^*$ to $W_1^*$ is also in uniform topology since, with probability one, any sample path of the latter is continuous. Since $\tau<\tau_F\vee\tau_G$, it is clear that

$$\limsup_{n\to\infty}\int_0^\tau \left| d\left\{\frac{1-\hat{F}_n^*(u-)}{1-F()}\,\frac{n}{Y_n^*(u)}\right\}\right| <\infty. \tag{A.19}$$

Moreover, $\sup_{0\leqslant t\leqslant\tau}|n^{-1/2}M_n(t)-W_1^*(t)|\to 0$ a.s. This, together with (A.19) and integration by parts, implies that

$$\sup_{0\leqslant t\leqslant\tau}\left|\int_0^t \frac{1-\hat{F}_n^*(u-)}{1-F(u)}\,\frac{dn^{-1/2}M_n^*(u)}{n^{-1}Y_n^*(u)} - \int_0^t \frac{1-\hat{F}_n^*(u-)}{1-F(u)}\,\frac{dW_1^*(u)}{n^{-1}Y_n^*(u)}\right| \to 0 \qquad \text{a.s.} \tag{A.20}$$

Likewise, applying integration by parts again, we get

$$\sup_{0\leqslant t\leqslant\tau}\left|\int_0^t \frac{1-\hat{F}_n^*(u-)}{1-F(u)}\frac{dW_1^*(u)}{n^{-1}Y_n^*(u)} - \int_0^t \frac{dW_1^*(u)}{(1-F(u))\,G(u)}\right| \to 0 \quad \text{a.s.} \tag{A.21}$$

From (A.20) and (A.21) we have

$$\sup_{0\leqslant t\leqslant\tau}\left|\int_0^t \frac{1-\hat{F}_n^*(u-)}{1-F(u)}\frac{dn^{-1/2}M_n^*(u)}{n^{-1}Y_n^*(u)} - \int_0^t \frac{dW_1^*(u)}{(1-F(u))\,G(u)}\right| \to 0 \quad \text{a.s.} \tag{A.22}$$

In view of (A.22) and (3.2), theorem 2 follows.

It remains to show that $n^{-1/2}M^{(n)}$ converges weakly to $W_1$. Recall that $M^{(n)}(t)=\sum_{i=1}^n M_i(t)$. Since $F$ is continuous, we can, without loss of generality, assume that the $X_i$ are $U(0, 1)$ random variables. Applying the same argument as that in the proof of Lemma 22.1 of Billingsley (1968, p. 195) it can be shown that

$$E[M^{(n)}(t)-M^{(n)}(s)]^4 \leqslant K_1\,[n^2 \sup_i E^2(M_i(t)-M_i(s))^2 + n \sup_i E(M_i(t)-M_i(s))^2] \tag{A.23}$$

for some constant $K_1>0$ and all $t, s\in[0,\tau]$, nothing that $\sum i^2\,\phi(i)<\infty$. Since each $X_i$ is uniformly distributed, $E[M_i(t)-M_i(s)]^2\leqslant|t-s|$, which together with (A.23) gives

$$E[M^{(n)}(t)-M^{(n)}(s)]^4 \leqslant K[n^2(t-s)^2+n\,|t-s|] \tag{A.24}$$

for some $K>0$ and all $s, t\leqslant\tau$. From (A.24) and exactly the same argument as that for (22.14)–(22.21) of Billingsley (1968, pp. 198–199) we conclude that, for every $\varepsilon>0$ and $\eta>0$, there exists $\delta>0$ such that

$$P\left\{\sup_{|t-s|\leqslant\delta,\,0\leqslant s,\,t\leqslant\tau}\left|\frac{1}{\sqrt{n}}(M^{(n)}(t)-M^{(n)}(s))\right|\geqslant\varepsilon\right\}\leqslant\eta \tag{A.25}$$

for all sufficiently large $n$. Therefore $\{n^{-1/2}M^{(n)}\}$ is tight. From (A.25), it also follows that any limiting distribution of $\{n^{-1/2}M^{(n)}\}$ is in $C[0,\tau]$. Moreover, from the $\phi$-mixing condition, for any $0\leqslant t_1<\cdots<t_p\leqslant\tau$, and $0<s_1<\cdots<s_q<1$, $\{M^{([ns_1])}(t_i), i=1, ..., p\}$, $\{M^{([ns_2])}(t_i)-M^{([ns_1])}(t_i), i=1, ..., p\}$, ..., $\{M^{([ns_q])}(t_i)-M^{([ns_{q-1}])}(t_i), i=1, ..., p\}$, where $[a]$ denotes the largest integer $\leqslant a$, are symptotically independent. Therefore, any limiting distribution of $(n^{-1/2}M^{(n)}(t_1), ..., n^{-1/2}M^{(n)}(t_p))$ is a $p$-variate normal distribution. On the other hand, from (A.23), it follows that for fixed $s, t \in [0, \tau]$, $n^{-1}M^{(n)}(s)\,M^{(n)}(t)$ is uniformly integrable. This and (3.3) imply that the finite dimensional distributions of $n^{-1/2}M^{(n)}$ converge to those of $W_1$. The desired conclusion follows from the tightness of $\{n^{-1/2}M^{(n)}\}$. ∎

## ACKNOWLEDGMENT

The authors thank the referee for many constructive comments.

## REFERENCES

1. BILLINGSLEY, P. (1968). *Weak Covergence of Probability Measures.* Wiley, New York.
2. BRESLOW, N., AND CROWLEY, J. (1974). A large sample study of the life table and product-limit estimates under random censorship. *Ann. Statist.* **2** 437–453.
3. COX, D. R., AND OAKES, D. (1984). *Analysis of Survival Data.* Chapman and Hall, London.
4. THE DCCT RESEARCH GROUP (1986). The Diabetes Control and Complications Trial (DCCT): the design and methodological considerations for feasibility phase. *Diabetes* **35** 530–545.
5. GILL, R. D. (1980). *Censoring and Stochastic Integrals.* Mathematical Centre Tracts, Vol. 124, Mathematisch Centrum, Amsterdam.
6. KAPLAN, E. L., AND MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.
7. MANTEL, N., AND CIMINERA, J. L. (1979). Use of logrank scores in the analysis of litter-matched data on time to tumor appearance. *Cancer Res.* **39** 4308-4315.
8. SHORACK, G. R., AND WELLNER, J. A. (1986) *Emperical Processes with Applications to Statistics.* Wiley, New York.
9. STOUT, W. F., (1974) *Almost Sure Convergence.* Academic Press, New York.