Journal of Multivariate Analysis **67**, 318–348 (1998)
Article No. MV981769

Kaplan–Meier Estimator under Association

Zongwu Cai

University of North Carolina at Charlotte

and

George G. Roussas

University of California, Davis

Received April 24, 1997; revised February 11, 1998

View metadata, citation and similar papers at core.ac.uk

Consider a long term study, where a series of possibly censored failure times is observed. Suppose the failure times have a common marginal distribution function F, but they exhibit a mode of dependence characterized by positive or negative association. Under suitable regularity conditions, it is shown that the Kaplan–Meier estimator $\tilde{F}_n$ of F is uniformly strongly consistent; rates for the convergence are also provided. Similar results are established for the empirical cumulative hazard rate function involved. Furthermore, a stochastic process generated by $\tilde{F}_n$ is shown to be weakly convergent to an appropriate Gaussian process. Finally, an estimator of the limiting variance of the Kaplan–Meier estimator is proposed and it is shown to be weakly convergent. © 1998 Academic Press

Key Words and Phrases: censored data; Kaplan–Meier estimator; negative association; positive association; strong consistency; variance estimator; weak convergence.

1. INTRODUCTION, ASSUMPTIONS, AND STATEMENT OF MAIN RESULTS

Let $T_1, ..., T_n$ be a sequence of the true survival times for n individuals in a life table. These random variables (r.v.s) are not assumed to be mutually independent (see Assumption (K1) below for the kind of dependence stipulated); it is assumed, however, that they have a common unknown continuous marginal distribution function (d.f.) $F(x)=P(T_i \leqslant x)$, and that $F(0)=0$. Let the r.v.s. T_i be censored on the right by the r.v.s. Y_i, so that one observes only

$$Z_i = T_i \wedge Y_i \qquad \text{and} \qquad \delta_i = I(T_i \leqslant Y_i), \tag{1.1}$$

0047-259X/98 $25.00

Copyright © 1998 by Academic Press
All rights of reproduction in any form reserved.

where $\wedge$ denotes minimum and $I(\cdot)$ is the indicator r.v. of the event specified in the parenthesis. In this random censorship model, the censoring times Y_i, $i=1, ..., n$, are assumed to be independent identically distributed (i.i.d.) r.v.s. with d.f. $G(y)=P(Y_i \leqslant y)$ such that $G(0)=0$; they are also assumed to be independent of the T_i's. The problem at hand is that of drawing nonparametric inference about F, based on the censored observations (Z_i, δ_i), $i=1, ..., n$. For this purpose, for any $t \geqslant 0$, define three stochastic processes on $[0, \infty)$ as follows:

$$N_n(t)=\sum_{i=1}^{n} I(Z_i \leqslant t, \delta_i=1)=\sum_{i=1}^{n} I(T_i \leqslant t \wedge Y_i), \tag{1.2}$$

the number of uncensored observations less than or equal to t;

$$Y_n(t)=\sum_{i=1}^{n} I(Z_i \geqslant t), \tag{1.3}$$

the number of censored or uncensored observations greater than or equal to t; and

$$M_n(t)=Y_n(-\infty)-Y_n(t+)-N_n(t)=\sum_{i=1}^{n} I(Z_i \leqslant t, \delta_i=0), \tag{1.4}$$

the number of observations censored at a value less than or equal to t. The Kaplan–Meier (K–M) estimator $\tilde{F}_n$ of F (Kaplan and Meier, 1958), based on the censored data (Z_i, δ_i), $i=1, ..., n$, is defined through the relation

$$1-\tilde{F}_n(t)=\prod_{s \leqslant t}\left(1-\frac{dN_n(s)}{Y_n(s)}\right), \tag{1.5}$$

where $dN_n(s)=N_n(s)-N_n(s-)$. As is known (see, for example, Gill, 1980), for a d.f. F on $[0, \infty)$, the cumulative hazard function $\Lambda(t)$ is given by:

$$\Lambda(t)=\int_0^t \frac{dF(s)}{1-F(s-)}, \tag{1.6}$$

and $\Lambda(t)=-\log(1-F(t))$ for the case that F is continuous. The empirical cumulative hazard function $\hat{\Lambda}_n(t)$ is taken to be

$$\hat{\Lambda}_n(t)=\int_0^t \frac{dN_n(s)}{Y_n(s)}, \tag{1.7}$$

which is referred to in literature as Nelson estimator of $\Lambda(t)$.

For the case that the failure time observations are mutually independent, the K–M estimator $\tilde{F}_n(t)$ has been studied extensively by many investigators during the last three decades. For example, uniform consistency, weak convergence and other asymptotic properties were obtained by Breslow and Crowley (1974), Peterson (1977), Gill (1980, 1983), Wang (1987), Stute and Wang (1993), and Stute (1994) among others. However, there are preciously few results available for the case that these observations exhibit some kind of dependence. For instance, for positively associated r.v.s., Bagai and Prakasa Rao (1991) discussed the strong consistency and asymptotic normality of the empirical survival function; weak convergence of the empirical process was obtained by Yu (1993). However, these authors focused on the uncensored observations. Voelkel and Crowley (1984) used an approach, based on semi-Markov processes, to establish a reasonable model in Cancer Research Clinical Trials assuming that each patient may either remain in an initial state, or progress, or respond and then possibly relapse. Ying and Wei (1994) explored consistency and asymptotic normality of $\tilde{F}_n(t)$ in the ϕ-mixing context. An application of the right censoring model was also given for a special dependent case, in which survival times are highly stratified.

In this paper, we study the large sample properties of the K–M estimator $\tilde{F}_n(t)$ for the case in which the underlying failure times are assumed to be positively or negatively associated (see Definition 1.1 below). More precisely, the main results obtained in this work are as follows. The K–M estimator $\tilde{F}_n$, defined through (1.5), is shown to be uniformly strongly consistent under either positive or negative association (see Theorem 1.2); rates of convergence are also provided (see Theorem 1.4). The proofs of these theorems are facilitated by first establishing similar results for the empirical cumulative hazard function $\hat{\Lambda}_n(t)$ defined by (1.7) as an estimate of the cumulative hazard function $\Lambda(t)$, given in (1.6) (see Theorems 1.1 and 1.3). Next, consider the stochastic process generated by the estimator $\tilde{F}_n(t)$ and defined by (1.12). It is then shown that this process converges weakly to a suitable Gaussian process (see Theorem 1.5) with specified covariance structure, given by (4.12), and (4.7). Finally, a valid estimate is constructed for the variance of the Gaussian process just mentioned; this is the content of Theorem 1.6.

The definition of the underlying dependence considered here is as follows.

Definition 1.1. The r.v.s. $\{X_j; 1 \leqslant j \leqslant n\}$ are said to be *positively associated* (PA), if for every $G, H\colon \mathfrak{R}^n \to \mathfrak{R}$, which are coordinatewise nondecreasing, and for which $E[G^2(X_j, 1 \leqslant j \leqslant n)] < \infty$, $E[H^2(X_j, 1 \leqslant j \leqslant n)] < \infty$, it holds that:

$$\mathrm{Cov}[G(X_i, 1 \leqslant i \leqslant n), H(X_j, 1 \leqslant j \leqslant n)] \geqslant 0.$$

The above r.v.s. are said to be *negatively associated* (NA), if for every non-empty proper subset A of $\{1, 2, ..., n\}$ and for every $G: \mathfrak{R}^{\# A} \to \mathfrak{R}$, $H: \mathfrak{R}^{\# A^c} \to \mathfrak{R}$, which are nondecreasing as above, and for which $E[G^2(X_j, j \in A)] < \infty$, $E[H^2(X_j, j \in A^c)] < \infty$, it holds that:

$$\text{Cov}[G(X_i, i \in A), H(X_j, j \in A^c)] \leqslant 0;$$

here and in the sequel, $\# A$ denotes the cardinality of A. Infinitely many r.v.s. are said to be PA(NA), if any finite subset of them is a set of PA(NA) r.v.s.

Positive association has found extensive applications in systems reliability and various problems in statistical mechanics. See, for example, Harris (1960), Esary *et al.* (1967), Fortuin *et al.* (1971), Barlow and Proschan (1975), and Newman (1980, 1990). On negative dependence, in the paper by Ebrahimi and Ghosh (1981), it is argued (see Section 5) that this kind of dependence holds in many multivariate distributions, and has an impact on certain reliability problems. Block *et al.* (1982) undertook a systematic study of negative dependence. The concept of NA was introduced in Joag-Dev and Proschan (1983). They compared it with other proposed concepts of negative dependence, and justified the claim that NA possesses certain advantages over competing notions of negative dependence. The authors emphasized that NA is not simply dual to PA, but it differs from it in important respects. They also derived, as a byproduct of their main results, that many well-known multivariate distributions are NA. The papers by Joag-Dev (1983), and Brindley and Thompson (1972) are also relevant references. Bozorgnia *et al.* (1993) derives a wealth of results regarding limiting theorems for negatively dependent r.v.s, in general, and NA r.v.s in particular, including Weak and Strong Laws of Large Numbers, and Roussas (1994) established the asymptotic normality for random fields under either PA or NA. Perhaps, the significance of NA may lie, however, in the perception that NA is the appropriate modeling for several species competing for the same limited resources.

The additional assumptions under which the main results of this paper are derived are gathered below for easy reference.

ASSUMPTIONS. (K1) $\{T_j; j \geqslant 1\}$ *is a stationary sequence of* (*positively or negatively*) *associated r.v.s. with marginal d.f.* F, *having a bounded density and finite second moment.*

(K2) *The censoring time variables* $\{Y_j; j \geqslant 1\}$ *are i.i.d. r.v.s with bounded density, and are independent of* $\{T_j; j \geqslant 1\}$.

(K3) $\sum_{j=2}^{\infty} j^{-2} \sum_{i=1}^{j-1} |\text{Cov}(T_i, T_j)|^{1/3} < \infty$.

(K4) $\sum_{j=2}^{n} j^2 \, |\mathrm{Cov}(T_1, T_j)|^{1/3} = O(n^{\tau_0})$, *for some* $0 \leqslant \tau_0 < 1$.

(K5) $\sum_{j=n+1}^{\infty} |\mathrm{Cov}(T_1, T_j)|^{1/3} = O(n^{-(r-2)/2})$, *for some* $r > 2$.

For the d.f.s F and G, define (the possibly infinite) times τ_F and τ_G by:

$$\tau_F = \inf\{y: F(y) = 1\}, \qquad \text{and} \qquad \tau_G = \inf\{y: G(y) = 1\}. \tag{1.8}$$

Then for the marginal d.f. H of the Z_i's, it holds (see, for example, Stute and Wang, 1993):

$$\tau_H = \tau_F \wedge \tau_G. \tag{1.9}$$

At this point, it is mentioned that all limits are taken as $n \to \infty$, unless otherwise stated, and proceed with the statement of the main results of this paper.

THEOREM 1.1. *Suppose that Assumptions* (K1)–(K2) *hold. Then, for any* $0 < \tau < \tau_H$, *where* τ_H *is given by* (1.9), *we have*:

(i) *If the r.v.s.* $\{T_j;\, j \geqslant 1\}$ *are PA*, *and Assumption* (K3) *is satisfied, it holds*:

$$\sup_{0 \leqslant t \leqslant \tau} |\hat{\Lambda}_n(t) - \Lambda(t)| \to 0 \quad a.s. \tag{1.10}$$

(ii) *If the r.v.s* $\{T_j;\, j \geqslant 1\}$ *are NA*, *then*, (1.10) *holds.*

THEOREM 1.2. *Suppose that Assumptions* (K1)–(K2) *hold. Then, under the additional assumptions either in part* (i) *or part* (ii) *of Theorem* 1.1, *it holds*:

$$\sup_{0 \leqslant t \leqslant \tau_H} |\tilde{F}_n(t) - F(t)| \to 0 \quad a.s., \qquad and$$

$$\sup_{0 \leqslant t \leqslant Z_{n:n}} |\tilde{F}_n(t) - F(t)| \to 0 \quad a.s., \tag{1.11}$$

where $Z_{n:n} = \max_{i \leqslant n} Z_i$.

THEOREM 1.3. *Let the r.v.s* $\{T_j;\, j \geqslant 1\}$ *be either PA or NA*, *and suppose that Assumptions* (K1), (K2) *and* (K5) *hold. Then, for any* $0 < \tau < \tau_H$, *we have*:

$$\sup_{0 \leqslant t \leqslant \tau} |\hat{\Lambda}_n(t) - \Lambda(t)| = o(n^{-\theta}) \quad a.s.,$$

where $0<\theta<(r-2)/(2r+2+\rho)$ *for any* $\rho>0$, *and* r *is given in Assumption* (K5).

THEOREM 1.4. *Let the r.v.s* $\{T_j;\, j\geqslant 1\}$ *be either PA or NA, and suppose that Assumptions* (K1), (K2), *and* (K5) *hold. Then, for any* $0<\tau<\tau_H$, *we have*:

$$\sup_{0\leqslant t\leqslant\tau} |\tilde{F}_n(t)-F(t)|=o(n^{-\theta}) \quad a.s.,$$

where θ *is defined in Theorem* 1.3.

Next, let $\tilde{Z}_n(t)$ be defined by

$$\tilde{Z}_n(t)=\frac{\sqrt{n}[\tilde{F}_n(t)-F(t)]}{1-F(t)}, \tag{1.12}$$

and let $W(\cdot)$ be a zero-mean Gaussian process in $D[0,\tau]$ for some τ such that $\tau<\tau_H$ and $F(\tau)<1$, and $W(0)=0$; the covariance structure of $W(\cdot)$ is given by (4.12) and (4.7). Then the following result is true.

THEOREM 1.5. *Let* $W(\cdot)$ *be the zero-mean Gaussian process with covariance function defined by* (4.12) *and* (4.7). *Then, under Assumptions* (K1), (K2) *and* (K4), *and the additional condition that* $\{T_j;\, j\geqslant 1\}$ *are either PA or NA, the process* $\tilde{Z}_n(\cdot)$ *defined in* (1.2) *converges weekly as follows*:

$$\tilde{Z}_n(\cdot)\xrightarrow{\mathscr{D}} W(\cdot) \qquad in \quad D[0,\tau],$$

for any $\tau<\tau_H$ *such that* $F(\tau)<1$. *This implies that* $\sqrt{n}(\tilde{F}_n-F)$ *converges weakly to* $(1-F)\,W$.

Remark 1.1. The asymptotic behavior of the stochastic process under consideration here was studied by Yu (1993) for the uncensored case under PA. The author proved that, under suitable conditions, two versions of the process converge weakly to a zero-mean Gaussian process (see Theorem 2.2 and Corollary 1 in Yu, 1993, p. 360). However, there are differences in two points between what Yu (1993) obtained and what is derived here. First, the space $D[0,1]$ in Yu (1993) is replaced by the space $D[0,\tau]$ $(0<\tau<\tau_H)$ here, and second the covariance structures of the two limiting processes in the two papers are not identical. Both differences are due to the random censorship assumed here. Incidentally, Theorem 1.5 here also holds for the NA case as well, not treated in Yu (1993). Finally, the basic assumption (2.5) in Yu (1993) implies the corresponding assumption here stated in (K4) for some $\tau_0>0$. Indeed, setting for convenience $c_j=|\mathrm{Cov}(T_1,T_j)|$, the left-hand side of (K4) is

$$\begin{aligned}
\sum_{j=2}^{n} j^2 c_j^{1/3} &= \sum_{j=2}^{n} j^{(13/2+\nu)/3} c_j^{1/3} j^{-(1+2\nu)/6} \\
&\leqslant \left(\sum_{j=2}^{n} j^{13/2+\nu} c_j\right)^{1/3} \left(\sum_{j=2}^{n} j^{-(1+2\nu)/4}\right)^{2/3} \\
&\qquad \text{(by the Cauchy–Schwarz Inequality)} \\
&\leqslant C\left(\sum_{j=2}^{n} j^{-(1+2\nu)}/4\right)^{2/3} \qquad \text{(by (2.5) in Yu (1993))} \\
&\leqslant C((1/(-\gamma+1))(n^{-\gamma+1}-1))^{2/3} \\
&\leqslant C((1/(-\gamma+1))\, n^{-\gamma+1})^{2/3}, \qquad \text{where} \quad \gamma = (1+2\nu)/4 \\
&= 55Cn^{(3-2\nu)/6} = Cn^{\tau_0};
\end{aligned}$$

that is, $\sum_{j=2}^{n} j^2\, |\mathrm{Cov}(T_1, T_j)|^{1/3} = O(n^{\tau_0})$, as asserted.

Finally, the result below provides for a consistent estimate of the variance $\sigma^2(t, t)$ of the limiting process $W(\cdot)$, given by (4.7) (for $s = t$). More precisely, we have:

THEOREM 1.6. *Let* $\sigma^2(t, t)$ *and* $\hat{V}_n(t)$ *be given by* (4.7) (*with* $s = t$) *and* (5.4), *respectively, and suppose Assumptions* (K1), (K2), *and* (K5) *hold. Then, if the r.v.s* $\{T_j;\, j \geqslant 1\}$ *are either PA or NA, the quantity* $n\hat{V}_n(t)$ *is a weakly consistent estimate of* $\sigma^2(t, t)$.

The proofs of Theorems 1.1 through 1.4 are given in Section 3. Theorem 1.5 is established in Section 4 along with some auxiliary results. Finally, a justification of Theorem 1.6 is presented in the final section of the paper, Section 5.

2. SOME PRELIMINARY RESULTS

In this section, some preliminary results are discussed to be used in the proofs of the theorems stated.

LEMMA 2.1. *Let* $\{X_j;\, j \geqslant 1\}$ *be a stationary sequence of r.v.s. Then*:

(i) *If the r.v.s are PA, having finite variance and satisfying the condition*:

$$\sum_{j=2}^{\infty} j^{-2} \sum_{i=1}^{j-1} \mathrm{Cov}(X_i, X_j) < \infty, \tag{2.1}$$

it follows that:

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - EX_i) \to 0 \quad a.s. \tag{2.2}$$

(ii) *If the r.v.s are NA with finite first moment, the convergence in* (2.2) *holds true.*

Proof. Case (i) is discussed in Birkel (1989), Theorem 1, and case (ii) is found in Theorem 5.1 of Bozorgnia *et al.* (1993). ∎

LEMMA 2.2. *Let* $\underline{U}=(U_1, ..., U_n)'$ *and* $\underline{W}=(W_1, ..., W_m)'$ *be two independent random vectors. Then*:

(i) *If* $X_i \stackrel{d}{=} g_i(U_i, \underline{W})$, $i=1, ..., n$, *where* g_i *is nondecreasing in each* w_j, $1 \leqslant j \leqslant m$, *for fixed* u_i, *and* $\underline{W}$ *are PA, then so are* $\underline{X}=(X_1, ..., X_n)'$.

(*ii*) *For* $m=n$, *if* $X_i \stackrel{d}{=} g_i(U_i, W_i)$, $i=1, ..., n$, *and* g_i *is nondecreasing in* w_i, *for fixed* u_i, *and* $\underline{W}$ *are NA, then* $\underline{X}$ *are also NA.*

Proof. Case (i) is discussed in Egeland (1992), Theorem 2.2. We will present the proof of (ii). Let A be any nonempty subset of $\{1, ..., n\}$, and let $H_1: \mathfrak{R}^{\#A} \to \mathfrak{R}$, $H_2: \mathfrak{R}^{\#A^c} \to \mathfrak{R}$ be coordinatewise nondecreasing, and suppose that:

$$E[H_1^2(X_j, j \in A)] < \infty, \qquad \text{and} \qquad E[H_2^2(X_j, j \in A^c)] < \infty. \tag{2.3}$$

Also, let F_U denote the joint d.f. of $\underline{U}$. Then it follows by the independence of $\underline{U}$ and $\underline{W}$ that:

$$\begin{aligned} &\operatorname{Cov}[H_1(X_i, i \in A), H_2(X_j, j \in A^c)] \\ &\quad = \int_{\mathfrak{R}^n} \operatorname{Cov}[H_1(g_i(u_i, W_i), i \in A), H_2(g_j(u_j, W_j), j \in A^c)] \\ &\quad \times dF_U(u_1, ..., u_n). \end{aligned}$$

For fixed u_j, $1 \leqslant j \leqslant n$, let $H_1^*(W_i, i \in A) = H_1(g_i(u_i, W_i), i \in A)$ and $H_2^*(W_j, j \in A^c) = H_2(g_j(u_j, W_j), j \in A^c)$. Clearly, $H_1^*: \mathfrak{R}^{\#A} \to \mathfrak{R}$ and $H_2^*: \mathfrak{R}^{\#A^c} \to \mathfrak{R}$ are coordinatewise nondecreasing and satisfy (2.3). Therefore, for fixed u_j, $1 \leqslant j \leqslant n$, it follows by negative association that:

$$\operatorname{Cov}[H_1^*(W_i, i \in A), H_2^*(W_j, j \in A^c)] \leqslant 0,$$

which implies that:

$$\operatorname{Cov}[H_1(X_i, i \in A), H_2(X_j, j \in A^c)] \leqslant 0.$$

This inequality completes the proof of the lemma. ∎

The following lemma extends Theorem 2 in Birkel (1988) for the PA case to the NA case. This theorem is stated as part (i) of the lemma below.

LEMMA 2.3. *Let* $\{X_j;\ j \in \mathcal{N}\}$ *be a sequence of r.v.s satisfying the requirements*: $E(X_j)=0$ *and* $|X_j| \leqslant C < \infty$ *for* $j \in \mathcal{N}$, *the set of all positive integers. Then*:

(i) *If the r.v.s are PA and satisfy the condition*:

$$\sup_{k \in \mathcal{N}} \sum_{j:\ |j-k| \geqslant n} |\mathrm{Cov}(X_j, X_k)| = O(n^{-(r-2)/2}) \tag{2.4}$$

for some $r>2$, *it follows that, for all* $n \in \mathcal{N}$, *there exists a constant* $B>0$, *not depending on* n, *such that*:

$$\sup_{m \in \mathcal{N} \cup \{0\}} E\left|\sum_{j=m+1}^{m+n} X_j\right|^r \leqslant Bn^{r/2}. \tag{2.5}$$

(ii) *If the r.v.s are NA, the inequality in* (2.5) *holds true for any* $r \geqslant 1$.

Proof. As already mentioned, case (i) is discussed in Birkel (1988), Theorem 2. We will present the proof of (ii). Indeed, it follows by Proposition 3.1 in Roussas (1996), the Hoeffding inequality for NA r.v.s, that, for any $r \geqslant 1$ and $m \geqslant 0$:

$$E\left|\sum_{j=m+1}^{m+n} X_j\right|^r = rn^r \int_0^\infty t^{r-1} P\left(\left|\sum_{j=m+1}^{m+n} X_j\right| > nt\right) dt$$

$$\leqslant 2rn^r \int_0^\infty t^{r-1} e^{-t^2 n/2C}\, dt = Bn^{r/2},$$

where $B = r(2C)^{r/2}\, \Gamma(r/2)$. This completes the proof of the lemma. ∎

Recall that F and G are the d.f.s of the T_i's and the Y_j's, respectively, and set

$$F_*(t) = P(Z_1 \leqslant t, \delta_1 = 1) = P(T_1 \leqslant t \wedge Y_1), \tag{2.6}$$

where Z_1 and δ_1 are given in (1.1). Then we have:

$$F_*(t) = \int_0^\infty F(t \wedge z)\, dG(z) = \int_0^t [1-G(z)]\, dF(z) = \int_0^t [1-G(z)]\, dF(z). \tag{2.7}$$

Next, let

$$\alpha_0 = \alpha_0(F, G) = P(T_1 \leqslant Y_1) = \int_0^\infty F(z)\, dG(z) = \int_0^\infty [1 - G(z)]\, dF(z),$$

and assume that $\alpha_0 > 0$. Clearly, $F_*(t)/\alpha_0$ is the conditional d.f. of Z_1, given $\delta_1 = 1$. Define τ_{F_*} by

$$\tau_{F_*} = \inf\{t; F_*(t) = \alpha_0\}. \tag{2.8}$$

Then it is easily seen that $\tau_{F_*} = \tau_F \wedge \tau_G$, where τ_F and τ_G are given in (1.8). It follows that $\tau_H = \tau_{F_*}$. Finally, with $N_n(t)$ and $Y_n(t)$ given in (1.2) and (1.3), respectively, set:

$$\bar{N}_n(t) = N_n(t)/n, \qquad \text{and} \qquad \bar{Y}_n(t) = Y_n(t)/n. \tag{2.9}$$

Then we have:

Proposition 2.1. *Suppose that Assumptions* (K1)–(K2) *hold. Then*:

(i) *If the r.v.s* $\{T_j;\ j \geqslant 1\}$ *are PA, and Assumption* (K3) *is fulfilled, it holds*:

$$\sup_{0 \leqslant t \leqslant \tau_H} |\bar{Y}_n(t) - [1 - H(t)]| \to 0 \quad a.s., \qquad and$$

$$\sup_{0 \leqslant t \leqslant \tau_H} |\bar{N}_n(t) - F_*(t)| \to 0 \quad a.s. \tag{2.10}$$

(ii) *If the r.v.s* $\{T_j;\ j \geqslant 1\}$ *are NA, it follows that* (2.10) *holds true.*

Proof. (i) By Lemma 2.2(i), we have that the r.v.s $\{Z_i;\ 1 \leqslant i \leqslant n\}$ are PA, which implies that the r.v.s $\{I(Z_i \geqslant t);\ 1 \leqslant j \leqslant n\}$ are also PA, for each fixed t. By Corollary A.3 in Roussas (1991), there exists $M_0 > 0$ such that, for all $i \neq j$,

$$\begin{aligned} \operatorname{Cov}[I(Z_i \geqslant t), I(Z_j \geqslant t)] &= P(Z_i < t, Z_j < t) - H^2(t) \\ &\leqslant M_0 [\operatorname{Cov}(Z_i, Z_j)]^{1/3}. \end{aligned}$$

Next, we will find an upper bound for $\operatorname{Cov}(Z_i, Z_j)$ for all $i \neq j$. To this end, by independence of $\{T_i\}$ and $\{Y_j\}$, we have:

$$\begin{aligned} \operatorname{Cov}(Z_i, Z_j) &= \int_{\mathfrak{R}^2} [E(T_i \wedge y_1)(T_j \wedge y_2) - (EZ_1)^2]\, dG(y_1)\, dG(y_2) \\ &= \int_{\mathfrak{R}^2} \operatorname{Cov}(T_i \wedge y_1, T_j \wedge y_2)\, dG(y_1)\, dG(y_2). \end{aligned} \tag{2.11}$$

Let $f_y(x)=\min\{x, y\}$ for fixed $y>0$. Then $f_y(x)$ is a nondecreasing function of x. Therefore, it follows from the Hoeffding equality (see, for example, Lemma 2 in Lehmann, 1966, and (2.12) in Yu, 1993) that, for fixed $y_1>0$ and $y_2>0$:

$$
\begin{aligned}
\operatorname{Cov}(T_i \wedge y_1, T_j \wedge y_2)
&= \operatorname{Cov}[f_{y_1}(T_i), f_{y_2}(T_j)] \\
&= \int_0^{y_1} \int_0^{y_2} [P(T_i \leqslant r, T_j \leqslant s) - P(T_i \leqslant r)\, P(T_j \leqslant s)]\, dr\, ds \\
&\leqslant \int_0^{\infty} \int_0^{\infty} [P(T_i \leqslant r, T_j \leqslant s) - P(T_i \leqslant r)\, P(T_j \leqslant s)]\, dr\, ds = \operatorname{Cov}(T_i, T_j).
\end{aligned}
\tag{2.12}
$$

Hence, substitution of (2.12) into (2.11) gives an upper bound for Cov-(Z_i, Z_j); namely,

$$\operatorname{Cov}(Z_i, Z_j) \leqslant \operatorname{Cov}(T_i, T_j), \tag{2.13}$$

which implies that:

$$\operatorname{Cov}(I(Z_i \geqslant t), I(Z_j \geqslant t)) \leqslant M_0[\operatorname{Cov}(T_i, T_j)]^{1/3}, \tag{2.14}$$

for some constant $M_0>0$, not depending on i, j and t. Clearly, Assumption (K3) implies that the r.v.s $\{I(Z_j \geqslant t);\ j \geqslant 1\}$ satisfy (2.1) for each fixed t. Thus, an application of Lemma 2.1(i) yields, for all $0 \leqslant t \leqslant \tau_H$,

$$\bar{Y}_n(t) \to P(Z_1 \geqslant t) = 1 - H(t) = \bar{H}(t) \quad a.s.$$

Since both $1-\bar{Y}_n(t)$ and $H(t)$ are d.f.s, we have by employing the same arguments as those used in the proof of the Glivenko–Cantelli theorem (see, for example, Theorem 1, pp. 127–128 in Tucker, 1967),

$$\sup_{0 \leqslant t \leqslant \tau_H} |\bar{Y}_n(t) - \bar{H}(t)| \to 0 \quad a.s., \tag{2.15}$$

which is what the first relation in (2.10) asserts. Next,

$$1 - \bar{N}_n(t) = \frac{1}{n} \sum_{j=1}^{n} I(T_j > t \wedge Y_j),$$

and, for each fixed t, the r.v.s $\{I(T_j > t \wedge Y_j);\ 1 \leqslant j \leqslant n\}$ are PA by Lemma 2.2(i). Therefore, an application of Corollary A.3 in Roussas (1991) again gives:

$$
\begin{aligned}
&\operatorname{Cov}[I(T_i > t \wedge Y_i), I(T_j > t \wedge Y_j)] \\
&\quad = \operatorname{Cov}[I(T_i \leqslant t \wedge Y_i), I(T_j \leqslant t \wedge Y_j)] \\
&\quad = \int_0^\infty \int_0^\infty \operatorname{Cov}[I(T_i \leqslant t \wedge y_1), I(T_j \leqslant t \wedge y_2)]\, dG(y_1)\, dG(y_2) \\
&\quad \leqslant M_0[\operatorname{Cov}(T_i, T_j)]^{1/3}. \qquad (2.16)
\end{aligned}
$$

Clearly, Assumption (K3) implies that the r.v.s $\{I(T_j > t \wedge Y_j);\, j \geqslant 1\}$ satisfy (2.1), for each fixed t. Hence, we have that, for all $0 \leqslant t \leqslant \tau_H$, and with $F_*(t)$ defined by (2.6):

$$\bar{N}_n(t) \to P(T_1 \leqslant t \wedge Y_1) = F_*(t) \quad a.s.$$

Then arguments similar to those used in establishing (2.15) yield:

$$\sup_{0 \leqslant t \leqslant \tau_H} |\bar{N}_n(t) - F_*(t)| \to 0 \quad a.s.$$

This completes the proof of part (i).

(ii) Follows by an argument similar to the one used in the proof of part (i) and by utilizing Lemma 2.1(ii) and Lemma 2.2(ii). ∎

The results established in this section are employed throughout the paper.

3. STRONG UNIFORM CONSISTENCY WITH RATES: PROOFS OF THEOREMS 1.1–1.4

Proof of Theorem 1.1. By Lemma 2 in Gill (1981), we have that:

$$\sup_{0 \leqslant t \leqslant \tau} |\hat{\varLambda}_n(t) - \varLambda(t)| \leqslant \frac{2\rho_\tau(\bar{N}_n, F_*)}{\bar{Y}_n(\tau)} + \frac{\rho_\tau(\bar{Y}_n, \bar{H})[\bar{N}_n(\tau) + \rho_\tau(\bar{N}_n, F_*)]}{\bar{Y}_n(\tau)[\bar{Y}_n(\tau) - \rho_\tau(\bar{Y}_n, \bar{H})]}, \tag{3.1}$$

where ρ_τ is defined by $\rho_\tau(F, G) = \sup\{|F(x) - G(x)|;\, x \leqslant \tau\}$ for any two functions F and G. An application of Proposition 2.1 concludes the proof of the theorem. ∎

Suppose G_1 is a bounded, nondecreasing, and right-continuous function on $\Re$ such that $G_1(-\infty) = 0$. Let $\mathscr{G}$ be the set of all such functions. For any pair $(G_1, G_2) \in \mathscr{G}$, define (see relation (6) in Gill, 1981):

$$\Phi(G_1, G_2)(t) = \prod_{s \leqslant t} \left(1 - \frac{dG_1(s)}{\bar{G}(s)}\right) \exp\left\{-\int_{-\infty}^{t} \frac{dG_{1c}(s)}{\bar{G}(s)}\right\},$$

where $G = G_1 + G_2$, $\bar{G}(t) = \bar{G}_1(t) + \bar{G}_2(t) = G_1(\infty) - G_1(t) + G_2(\infty) - G_2(t)$, and G_{1c} is the continuous part of G_1. Note that $\Phi(G_1, G_2)$ is a right-continuous, nonnegative, and nonincreasing function on $\Re$ with $\Phi(G_1, G_2)(-\infty) = 1$ (see Gill, 1981, p. 856). Let

$$L_1(t) = P(Z_1 \leqslant t, \delta_1 = 1), \qquad \bar{L}_1(t) = P(Z_1 > t, \delta_1 = 1),$$
$$L_0(t) = P(Z_1 \leqslant t, \delta_1 = 0), \quad \text{and} \quad \bar{L}_0(t) = P(Z_1 > t, \delta_1 = 0).$$

Then:

$$\bar{H}(t) = \bar{L}_1(t) + \bar{L}_0(t), \quad \text{and} \quad L_1(t) = F_*(t).$$

Clearly, $(L_1, L_0) \in \mathscr{G}$ and

$$\Phi(L_1, L_0)(t) = \prod_{s \leqslant t} \left(1 - \frac{dL_1(s)}{\bar{H}(s)}\right) \exp\left\{-\int_0^t \frac{dL_{1c}(s)}{\bar{H}(s)}\right\}.$$

If F_* is continuous, then $\Phi(L_1, L_0)$ becomes $1 - F(t)$. Since $(N_n, M_n) \in \mathscr{G}$, it follows by relation (7) in Gill (1981) that:

$$\Phi(\bar{N}_n, M_n/n)(t) = 1 - \tilde{F}_n(t),$$

where M_n is defined in (1.4). By Proposition 2.1,

$$\sup_{0 \leqslant t \leqslant \tau_H} |M_n(t)/n - L_0(t)| \to 0 \quad a.s.$$

Now using Lemma 2 in Gill (1981), we have:

$$\sup_{0 \leqslant t \leqslant \tau} |\tilde{F}_n(t) - F(t)| \to 0 \quad a.s., \tag{3.2}$$

for any $\tau \leqslant \tau_H$ for which $F(\tau) < 1$. We may now proceed with the proof of the next theorem.

Proof of Theorem 1.2. For the proof of the first relation in (1.11), in view of (3.2), it suffices to consider the case $F(\tau_H) = 1$. For an arbitrary $0 < \varepsilon < 1$, choose $\tau < \tau_H$ such that $1 - \varepsilon < F(\tau) < 1$. Then for $t \in [\tau, \tau_H]$, we have:

$$\tilde{F}_n(\tau-) \leqslant \tilde{F}_n(t) \leqslant 1, \quad \text{and} \quad 1 - \varepsilon < F(\tau-) = F(\tau) \leqslant F(t) < 1.$$

Therefore,

$$\sup_{\tau \leqslant t \leqslant \tau_H} |\tilde{F}_n(t) - F(t)| < \max\{\varepsilon, 1 - \tilde{F}_n(\tau-)\}.$$

By combining this inequality with (3.2), and using the fact that ε is arbitrary, it follows that the first relation in (1.11) holds. Since $H(\tau_H)=1$, then $Z_{n:n}=\max_{i\leqslant n} Z_i<\tau_H$. Hence, the second relation in (1.11) is a consequence of the first one. The proof of the theorem is completed. ∎

Proof of Theorem 1.3. For the PA case, Assumption (K5) implies (2.4). Then, apply Lemma 2.3(i), Remark 1.3 and Corollary 2.1 in Roussas (1991) to conclude that:

$$\sup_{0\leqslant t\leqslant\tau} |\bar{Y}_n(t)-\bar{H}(t)|=o(n^{-\theta}) \quad a.s., \qquad \text{and}$$

$$\sup_{0\leqslant t\leqslant\tau} |\bar{N}_n(t)-F_*(t)|=o(n^{-\theta}) \quad a.s. \tag{3.3}$$

For the NA case, applying Lemma 2.3(ii) and the same arguments as those used in the proof of Lemma 2.3 in Roussas (1991), we see that Corollary 2.1 in Roussas (1991) still holds true. This implies that (3.3) is true both for the PA and the NA case. Therefore, the theorem follows from (3.1). ∎

Proof of Theorem 1.4. For any $0<\tau<\tau_H$, $H(Z_{n:n})\to 1$ a.s. by Theorem 1.2, so that $0<\tau<Z_{n:n}$ for sufficiently large n. Therefore, Lemma 1 in Breslow and Crowley (1974) gives:

$$0<-\log(1-\tilde{F}_n(t))-\hat{\Lambda}_n(t)<\frac{n-Y_n(t)}{nY_n(t)}.$$

This implies that:

$$\rho_\tau(-\log(1-\tilde{F}_n), \hat{\Lambda}_n)\leqslant\frac{n-Y_n(\tau)}{nY_n(\tau)}\leqslant C(\tau)/n \tag{3.4}$$

for sufficiently large n, where $0<C(\tau)<\infty$, independent of n. Using Taylor expansion, we have:

$$\begin{aligned}\tilde{F}_n(t)-F(t)&=1-F(t)-[1-\tilde{F}_n(t)]=e^{-\Lambda(t)}-e^{\log(1-\tilde{F}_n(t))}\\&=[e^{-\Lambda(t)}-e^{-\hat{\Lambda}_n(t)}]+[e^{-\hat{\Lambda}_n(t)}-e^{\log(1-\tilde{F}_n(t))}]\\&=e^{-\hat{\Lambda}_n^*(t)}[\hat{\Lambda}_n(t)-\Lambda(t)]+e^{-\hat{\Lambda}_n^{**}(t)}[-\log(1-\tilde{F}_n(t))-\hat{\Lambda}_n(t)],\end{aligned} \tag{3.5}$$

where

$$\rho_\tau(\hat{\Lambda}_n^*, \Lambda)\leqslant\rho_\tau(\hat{\Lambda}_n, \Lambda),$$

and, from (3.4),

$$\rho_\tau(\hat{\Lambda}_n^{**}, \hat{\Lambda}_n) \leqslant \rho_\tau(-\log(1-\tilde{F}_n), \hat{\Lambda}_n) \leqslant C(\tau)/n.$$

Therefore, it follows from (3.5), (3.4) and Theorem 1.3 that:

$$\rho_\tau(\tilde{F}_n, F) \leqslant O\left(\frac{1}{n}\right) + O(\rho_\tau(\hat{\Lambda}_n, \Lambda)) = o(n^{-\theta}) \quad a.s.$$

This completes the proof of the theorem. ▮

4. WEAK CONVERGENCE: PROOF OF THEOREM 1.5

This section is devoted to showing that the process $\tilde{Z}_n(t)$, defined by (1.2), converges weakly to a suitable process. To this end, for any $t \geqslant 0$, let

$$\tilde{M}_n(t) = N_n(t) - \int_0^t Y_n(s)\, d\Lambda(s).$$

By (3.2.13) in Gill (1980), we have:

$$\frac{\tilde{F}_n(t) - F(t)}{1-F(t)} = \int_0^t \frac{1-\tilde{F}_n(s-)}{1-F(s)} \frac{d\tilde{M}_n(s)}{Y_n(s)}.$$

Then, by (1.12):

$$\tilde{Z}_n(t) = \int_0^t \frac{1-\tilde{F}_n(s-)}{1-F(s)} \frac{d(n^{-1/2}\tilde{M}_n(s))}{n^{-1}Y_n(s)}. \tag{4.1}$$

In order to show that $\tilde{Z}_n(\cdot)$ converges weakly to a zero-mean Gaussian process $W(\cdot)$ in $D[0, \tau]$ for some τ such that $\tau < \tau_H$ and $F(\tau) < 1$, with $W(0) = 0$, it suffices to show that $n^{-1/2}\tilde{M}_n$ converges weakly in $D[0, \tau]$ to a Gaussian process W_1 with $W_1(0) = 0$, $EW_1(t) = 0$ and whose covariance has the following structure:

$$E[W_1(s)\, W_1(t)] = M(s, t), \tag{4.2}$$

where $M(s, t)$ is defined by

$$\begin{aligned} M(s, t) = {} & M_{11}(s, t) - \int_0^s M_{12}(t, v)\, d\Lambda(v) - \int_0^t M_{12}(s, u)\, d\Lambda(u) \\ & + \int_0^s \int_0^t M_{22}(u, v)\, d\Lambda(u)\, d\Lambda(v); \end{aligned} \tag{4.3}$$

here

$$M_{11}(s,t)=\operatorname{Cov}(I(T_1\leqslant s\wedge Y_1), I(T_1\leqslant t\wedge Y_1))$$
$$+\sum_{j=2}^{\infty}[\operatorname{Cov}(I(T_1\leqslant s\wedge Y_1), I(T_j\leqslant t\wedge Y_j))$$
$$+\operatorname{Cov}(I(T_1\leqslant s\wedge Y_1), I(T_j\leqslant t\wedge Y_j))], \tag{4.4}$$
$$M_{12}(s,t)=\operatorname{Cov}(I(T_1\wedge Y_1\leqslant s), I(T_1\leqslant t\wedge Y_1))$$
$$+\sum_{j=2}^{\infty}[\operatorname{Cov}(I(T_1\leqslant s\wedge Y_1), I(T_j\wedge Y_j\leqslant t))$$
$$+\operatorname{Cov}(I(T_j\wedge Y_j\leqslant s), I(T_1\leqslant t\wedge Y_1))], \tag{4.5}$$

and

$$M_{22}(s,t)=\operatorname{Cov}(I(T_1\wedge Y_1\leqslant s), I(T_1\wedge Y_1\leqslant t))$$
$$+\sum_{j=2}^{\infty}[\operatorname{Cov}(I(T_1\wedge Y_1\leqslant s), I(T_j\wedge Y_j\leqslant t))$$
$$+\operatorname{Cov}(I(T_j\wedge Y_j\leqslant s), I(T_1\wedge Y_1\leqslant t))]. \tag{4.6}$$

This is so on account of the proof of Theorem 2 in Ying and Wei (1994). Define the following two stochastic processes:

$$\tilde{N}_n(t)=\frac{1}{\sqrt{n}}[N_n(t)-nF_*(t)], \qquad \text{and} \qquad \tilde{Y}_n(t)=\frac{1}{\sqrt{n}}[Y_n(t)-n\bar{H}(t)].$$

Then

$$n^{-1/2}\tilde{M}_n(t)=\tilde{N}_n(t)-\int_0^t \tilde{Y}_n(u)\,d\Lambda(u).$$

Define the function $\sigma^2(s,t)$ on $\Re_+^2$ into $\Re_+=[0,\infty)$ in the following manner:

$$\sigma^2(s,t)=\int_0^s\int_0^t \frac{dM(u,v)}{\bar{H}(u)\,\bar{H}(v)}. \tag{4.7}$$

As has already been mentioned, in order to prove Theorem 1.5, it suffices to show that the process $n^{-1/2}\tilde{M}_n(\cdot)$ converges weakly in $D[0,\tau]$. To this end, we need to show that $(\tilde{N}_n, \tilde{Y}_n)$ converges weakly to (B_1, B_2) for suitable Gaussian processes B_1 and B_2. This will follow by establishing the following two premises: First, that all finite-dimensional distributions of

$(\tilde{N}_n, \tilde{Y}_n)$ converge weakly to the appropriate multidimensional normal distribution, and second, that the processes are tight. The first premise is stated here as Proposition 4.1 and proved later. The rest of this section, revolves around the justification of the second premise. As it will be explained below, this effort amounts to obtaining suitable bounds for the moments $E[\tilde{N}_n(t)-\tilde{N}_n(s)]^4$ and $E[\tilde{Y}_n(t)-\tilde{Y}_n(s)]^4$, for any $t, s \in [0, \tau]$. The relevant result is stated below as a proposition. All results in this section are derived under the assumptions made in Theorem 1.5, unless otherwise stated.

PROPOSITION 4.1. *For any integers* $k \geqslant 1$, $l \geqslant 1$ *and any* $t_1, ..., t_k, s_1, ..., s_l \in [0, \tau]$, *we have*:

$$(\tilde{N}_n(t_1), ..., \tilde{N}_n(t_k), \tilde{Y}_n(s_1), ..., \tilde{Y}_n(s_l))$$

$$\xrightarrow{\mathscr{D}} (B_1(t_1), ..., B_1(t_k), B_2(s_1), ..., B_2(s_l)),$$

where B_1 *and* B_2 *are two zero-mean Gaussian processes having the following covariance structures*:

$$E[B_1(s)\, B_1(t)] = M_{11}(s, t), \qquad E[B_1(s)\, B_2(t)] = M_{12}(s, t), \tag{4.8}$$

and

$$E[B_2(s)\, B_2(t)] = M_{22}(s, t), \tag{4.9}$$

where M_{11}, M_{12} *and* M_{22} *are defined in* (4.4)–(4.6).

PROPOSITION 4.2. *The stochastic processes* $\{\tilde{N}_n(t);\ 0 \leqslant t \leqslant \tau\}$ *and* $\{\tilde{Y}_n(t);\ 0 \leqslant t \leqslant \tau\}$ *are tight.*

Suppose temporarily that these propositions have been established and proceed with the proof of the theorem.

Proof of Theorem 1.5. By Propositions 4.1 and 4.2,

$$(\tilde{N}_n(\cdot), \tilde{Y}_n(\cdot)) \xrightarrow{\mathscr{D}[0, \tau]} (B_1(\cdot), B_2(\cdot)).$$

Let $\phi(\tilde{N}_n, \tilde{Y}_n)(t) = \tilde{N}_n(t) - \int_0^t \tilde{Y}_n(u)\, d\Lambda(u)$. Then, by the continuity mapping theorem (see Theorem 5.1 in Billingsley, 1968, p. 30), we have:

$$\frac{1}{\sqrt{n}} \tilde{M}_n(\cdot) = \phi(\tilde{N}_n, \tilde{Y}_n)(\cdot) \xrightarrow{\mathscr{D}[0, \tau]} W_1(\cdot), \tag{4.10}$$

where

$$W_1(t) = \phi(B_1, B_2)(t) = B_1(t) - \int_0^t B_2(u)\, d\Lambda(u).$$

A simple computation establishes that $E[W_1(s)\, W_1(t)] = M(s, t)$, which is (4.2). By Theorem 1.2 and relation (4.10), it follows from Theorem 4.4 in Billingsley (1968) that:

$$(\tilde{F}_n, n^{-1} Y_n, n^{-1/2} \tilde{M}_n) \xrightarrow{\mathscr{D}[0,\tau]} (F, \bar{H}, W_1).$$

Therefore, by the Skorokhod–Dudley–Wichura Theorem (see Shorack and Wellner, 1986, p. 47), there exists a special construction $(\tilde{F}_n^*, n^{-1} Y_n^*, n^{-1/2} \tilde{M}_n^*)$, which has the same distribution as $(\tilde{F}_n, n^{-1} Y_n, n^{-1/2} \tilde{M}_n)$ and which converges to $(F, \bar{H}, W_1^*)$ almost surely, where W_1^* has the same probability distribution as W_1. From this point on, a repetition of the arguments in Ying and Wei (1994) (see pp. 27–28) leads to the following result:

$$\sup_{0 \leqslant t \leqslant \tau} \left| \int_0^t \frac{1 - \tilde{F}_n^*(s-)}{1 - F(s)} \frac{d(n^{-1/2} \tilde{M}_n^*(s))}{n^{-1} Y_n^*(s)} - \int_0^t \frac{dW_1^*(s)}{\bar{H}(s)} \right| \to 0 \quad a.s.,$$

which implies that:

$$\sup_{0 \leqslant t \leqslant \tau} \left| \tilde{Z}_n^*(t) - \int_0^t \frac{dW_1^*(s)}{\bar{H}(s)} \right| \to 0 \quad a.s.. \tag{4.11}$$

Let

$$W(t) = \int_0^t \frac{dW_1(u)}{\bar{H}(u)}.$$

Then a simple computation yields the covariance structure of $W(\cdot)$; namely, for any $s, t \in [0, \tau]$,

$$E[W(s)\, W(t)] = \sigma^2(s, t), \tag{4.12}$$

where $\sigma^2(s, t)$ is defined in (4.7). Therefore, (4.11) gives:

$$\tilde{Z}_n(\cdot) \xrightarrow{\mathscr{D}} W(\cdot) \qquad \text{in } D[0, \tau],$$

and the theorem follows. ∎

Remark 4.1. It is worth pointing out here that the covariance structure defined by (4.12) and (4.7) was also arrived at by Ying and Wei (1994) by imposing a ϕ-mixing condition on the survival times rather than PA (see relation (3.4) in Ying and Wei, 1994). That the two covariance structures are, indeed, the same can be seen by noting that $\Delta(t)$ here (relation (1.6) and the comment following it) equals $\Lambda(t)$ defined two lines after relation (2.1) in Ying and Wei (1994). Also, their δ_i is denoted by Δ_i here. It follows that $M^{(n)}(t)$ in Ying and Wei (1994) (defined two lines above relation (3.2)) is the same as $\tilde{M}_n(t)$ here. The process $\tilde{Z}_n(t)$ defined by (1.12) here is the same as that defined by (3.2) in Ying and Wei (1994) (where there is a misprint: $n^{1/2}M^{(n)}(s)$ should be $n^{-1/2}M^{(n)}(s)$). Furthermore, the stipulated limit $H(s, t)$ in (3.3) of Ying and Wei (1994) is our $M(s, t)$ defined right after relation (4.2). Finally, with F, G, and H standing for the d.f.s of the r.v.s T_1, Y_1 and $Z_1 = T_1 \wedge Y_1$, respectively, we have $\bar{H}(u) = P(Z_1 > u) = P(T_1 > u)\, P(Y_1 > u)$ (by the independent assumption of the T_i's and Y_i's), and this is equal to $\bar{G}(u)[1 - F(u)]$. The comparison is concluded by noting that our $\bar{G}(u)$ is denoted by $G(u)$ in Ying and Wei (1994) (see relation (2.1) and formulation of Theorem 2 in Ying and Wei, 1994).

In the proof of Proposition 4.1, we need the concept of weakly positive (negative) association, which is now defined.

DEFINITION 4.1. Let $\{\underline{X}_1, \underline{X}_2, ..., \underline{X}_m\}$ be $\mathfrak{R}^d$-valued random vectors. They are said to be *weakly positively associated* (WPA), if whenever π is a permutation of $\{1, 2, ..., m\}$, $1 \leqslant k < m$, and $f\colon \mathfrak{R}^{kd} \to \mathfrak{R}$, $g\colon \mathfrak{R}^{(m-k)\,d} \to \mathfrak{R}$ are coordinatewise nondecreasing, then

$$\operatorname{Cov}[f(\underline{X}_{\pi(1)}, ..., \underline{X}_{\pi(k)}), g(\underline{X}_{\pi(k+1)}, ..., \underline{X}_{\pi(m)})] \geqslant 0,$$

if the covariance is defined. The above r.v.s are said to be *weakly negatively associated* (WNA), if the above inequality is reversed. An infinite family of $\mathfrak{R}^d$-valued random vectors is WP(N)A, if every finite subfamily is WP(N)A.

We may now proceed with the proof of the proposition.

Proof of Proposition 4.1. Without loss of generality, it suffices to show that, for any t_1, t_2, s_1, $s_2 \in [0, \tau]$,

$$(\tilde{N}_n(t_1), \tilde{N}_n(t_2), \tilde{Y}_n(s_1), \tilde{Y}_n(s_2)) \xrightarrow{\mathscr{D}} (B_1(t_1), B_1(t_2), B_2(s_1), B_2(s_2)).$$

By the Cramér–Wold device, it suffice to show that, for any a_1, a_2, b_1, $b_2 \in \mathfrak{R}$,

$$a_1 \tilde{N}_n(t_1) + a_2 \tilde{N}_n(t_2) + b_1 \tilde{Y}_n(s_1) + b_2 \tilde{Y}_n(s_2) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \xi_j \xrightarrow{\mathscr{D}} N(0, \underline{a}'\Sigma\underline{a}),$$

where $\underline{a} = (-a_1, -a_2, b_1, b_2)'$,

$$\begin{aligned}\xi_j = {} & a_1 I(T_j \leqslant t_1 \wedge Y_j) + a_2 I(T_j \leqslant t_2 \wedge Y_j) \\ & + b_1 I(T_j \wedge Y_j \geqslant s_1) + b_2 I(T_j \wedge Y_j \geqslant s_2) \\ & - [a_1 F_*(t_1) + a_2 F_*(t_2) + b_1 \bar{H}(s_1) + b_2 \bar{H}(s_2)],\end{aligned}$$

and

$$\Sigma = \begin{pmatrix} M_{11}(t_1, t_1) & M_{11}(t_1, t_2) & M_{12}(t_1, s_1) & M_{12}(t_1, s_2) \\ M_{11}(t_2, t_1) & M_{22}(t_2, t_2) & M_{12}(t_2, s_1) & M_{12}(t_2, s_2) \\ M_{12}(s_1, t_1) & M_{12}(s_1, t_2) & M_{22}(s_1, s_1) & M_{22}(s_1, s_2) \\ M_{12}(s_2, t_1) & M_{12}(s_2, t_2) & M_{22}(s_2, s_1) & M_{22}(s_2, s_2) \end{pmatrix};$$

the quantities M_{11}, M_{12} and M_{22} are defined in (4.4)–(4.6). For fixed t_1, t_2, s_1 and s_2, let

$$\underline{\eta}_j = \underline{\eta}_j(T_j, Y_j) = \begin{pmatrix} -I(T_j \leqslant t_1 \wedge Y_j) \\ -I(T_j \leqslant t_2 \wedge Y_j) \\ I(T_j \wedge Y_j \geqslant s_1) \\ I(T_j \wedge Y_j \geqslant s_2) \end{pmatrix}, \qquad j \geqslant 1.$$

Then, $\underline{\eta}_j$ is a nondecreasing function of T_j for fixed Y_j, $j \geqslant 1$, and

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \xi_j = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \underline{a}'(\underline{\eta}_j - E\underline{\eta}_j).$$

Now, we wish to show that $\{\underline{\eta}_j;\, j \geqslant 1\}$ are WP(N)A corresponding to the positive (negative) association of $\{T_j;\, j \geqslant 1\}$. To this end, let π be a permutation of $\{1, 2, ..., n\}$, and for any $1 \leqslant k < n$, let $f_1\colon \Re^{4k} \to \Re$, $f_2\colon \Re^{4(n-k)} \to \Re$ be coordinatewise nondecreasing functions; also, let $A = \{\pi(1), ..., \pi(k)\}$, $A^c = \{\pi(k+1), ..., \pi(n)\}$, and $f_1^*(T_j, Y_j;\, j \in A) = f_1(\underline{\eta}_j;\, j \in A)$, $f_2^*(T_j, Y_j;\, j \in A^c) = f_2(\underline{\eta}_j;\, j \in A^c)$. Then, by Lemma 2.2, $\{f_1^*(T_j, Y_j;\, j \in A);\, f_2^*(T_j, Y_j;\, j \in A^c)\}$ are P(N)A, according to the positive (negative) association of $\{T_j;\, j \geqslant 1\}$. Therefore,

$$\begin{aligned} & \mathrm{Cov}(f_1(\underline{\eta}_{\pi(1)}, ..., \underline{\eta}_{\pi(k)}), f_2(\underline{\eta}_{\pi(k+1)}, ..., \underline{\eta}_{\pi(n)})) \\ & \quad = \mathrm{Cov}(f_1^*(T_j, Y_j;\, j \in A), f_2^*(T_j, Y_j;\, j \in A^c)) \\ & \quad = \begin{cases} \geqslant 0, & \text{if } \{T_j;\, j \geqslant 1\} \text{ are PA}, \\ \leqslant 0, & \text{if } \{T_j;\, j \geqslant 1\} \text{ are NA}. \end{cases} \end{aligned}$$

This implies that $\{\underline{\eta}_j;\, 1 \leqslant j \leqslant n\}$ are WP(N)A. Clearly, (2.14), (2.16) and Assumption (K4) yield:

$$E\,\|\underline{\eta}_1\|^2 + 2 \sum_{j=2}^{\infty} \sum_{i=1}^{4} |\mathrm{Cov}(\underline{\eta}_1^{(i)}, \underline{\eta}_j^{(i)})| < \infty.$$

Now, for the WPA case, by Theorem 2 in Burton *et al.* (1986), we have:

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\underline{\eta}_j - E\underline{\eta}_j) \xrightarrow{\mathscr{D}} N_4(0, \Sigma).$$

Therefore,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \xi_j \xrightarrow{\mathscr{D}} N(0, \underline{a}'\Sigma\underline{a}). \tag{4.13}$$

For the WNA case, it is not difficult to extend Theorem 2 in Burton *et al.* (1986) to the WNA case. It follows then that (4.13) holds also for the WNA case. Hence, there exist two zero-mean Gaussian processes B_1 and B_2 such that $B_1(0) = 0$, $B_2(0) = 0$, having covariance structures as specified in (4.8) and (4.9), and such that

$$(\tilde{N}_n(t_1), \tilde{N}_n(t_2), \tilde{Y}_n(s_1), \tilde{Y}_n(s_2)) \xrightarrow{\mathscr{D}} (B_1(t_1), B_1(t_2), B_2(s_1), B_2(s_2)).$$

This completes the proof of the proposition. ∎

The following two lemmas will be needed before we embark on the proof of Proposition 4.2.

LEMMA 4.1. *There exists a constant $C > 0$, not depending on n, such that, for any $0 \leqslant t_1 \leqslant t_2 \leqslant \tau$:*

$$E[\tilde{Y}_n(t_2) - \tilde{Y}_n(t_1)]^4 \leqslant C(n^{\tau_0 - 1} + (H(t_2) - H(t_1))^{6/5}),$$

where τ_0 is given in (K4).

Proof. By (2.13), we can use exactly the same arguments as those employed in proving Lemma 4.4 in Yu (1993) in order to justify the lemma. ∎

Next, we wish to obtain an upper bound for $E[\tilde{N}_n(t) - \tilde{N}_n(s)]^4$, both for the PA and the NA case. This is stated and proved below as a lemma.

LEMMA 4.2. *There exists a constant* $C>0$, *not depending on* n, *such that, for any* $0\leqslant t_1\leqslant t_2\leqslant \tau$:

$$E[\tilde{N}_n(t_2)-\tilde{N}_n(t_1)]^4\leqslant C(n^{\tau_0-1}+(F(t_2)-F(t_1))^{6/5}), \tag{4.14}$$

where τ_0 *is given in* (K4).

Proof. Let T_0 be an independent copy of T_1, and let Y_0 be an independent copy of Y_1. Let $\{T_j^0;\, j\geqslant 0\}$, $\{Y_j^0;\, j\geqslant 0\}$ be an independent copy of $\{T_j;\, j\geqslant 0\}$ and $\{Y_j;\, j\geqslant 0\}$, respectively, and be mutually independent. Then, by Jensen's inequality for conditional expectations, we have:

$$\begin{aligned}
&E[\tilde{N}_n(t_2)-\tilde{N}(t_1)]^4\\
&\quad=\frac{1}{n^2}E\left\{\sum_{j=1}^n [I(t_1\wedge Y_j<T_j\leqslant t_2\wedge Y_j)-EI(t_1\wedge Y_j<T_j\leqslant t_2\wedge Y_j)]\right\}^4\\
&\quad=\frac{1}{n^2}E\left\{E\left[\sum_{j=1}^n (I(t_1\wedge Y_j<T_j\leqslant t_2\wedge Y_j)\right.\right.\\
&\qquad\left.\left.-I(t_1\wedge Y_j^0<T_j^0\leqslant t_2\wedge Y_j^0))\right]\,\middle|\, T_i,\, Y_i;\, 1\leqslant i\leqslant n\right\}^4\\
&\quad\leqslant\frac{1}{n^2}E\left(\sum_{j=1}^n [I(t_1\wedge Y_j<T_j\leqslant t_2\wedge Y_j)-I(t_1\wedge Y_j^0<T_j^0\leqslant t_2\wedge Y_j^0)]\right)^4.
\end{aligned}$$

By means of relation (22.2) in Billingsley (1968, p. 196), we have that:

$$\begin{aligned}
&E[\tilde{N}_n(t_2)-\tilde{N}_n(t_1)]^4\\
&\qquad\leqslant\frac{4!}{n^2}\sum|E\Gamma(I(t_1\wedge Y_0<T_0\leqslant t_2\wedge Y_0),\, I(t_1\wedge Y_i<T_i\leqslant t_2\wedge Y_i),\\
&\qquad\quad I(t_1\wedge Y_{i+j}<T_{i+j}\leqslant t_2\wedge Y_{i+j}),\\
&\qquad\quad I(t_1\wedge Y_{i+j+k}<T_{i+j+k}\leqslant t_2\wedge Y_{i+j+k}))|,
\end{aligned}$$

where the summation is over $Q=\{(i,j,k)\colon 0\leqslant i+j+k\leqslant n\}$, and the function Γ is defined by

$$\Gamma(\xi_j,\, j=1,\ldots,4)=\tfrac{1}{2}E\prod_{j=1}^{4}(\xi_j-\eta_j),$$

for any random variables $\xi_1,\ldots,\xi_4$ and an independent copy $(\eta_1,\ldots,\eta_4)$ of $(\xi_1,\ldots,\xi_4)$; this expression is a special case of relation (2.12) in Yu (1993). Now using the independence of $\{T_i\}$ and $\{Y_j\}$, we have that:

$$\begin{aligned}&E\Gamma[I(t_1 \wedge Y_0 < T_0 \leqslant t_2 \wedge Y_0), I(t_1 \wedge Y_i < T_i \leqslant t_2 \wedge Y_i),\\
&\qquad I(t_1 \wedge Y_{i+j} < T_{i+j} \leqslant t_2 \wedge Y_{i+j}),\\
&\qquad I(t_1 \wedge Y_{i+j+k} < T_{i+j+k} \leqslant t_2 \wedge Y_{i+j+k})]\\
&\quad = \int_{\mathfrak{R}^4_+} E\Gamma[I(t_1 \wedge y_0 < T_0 \leqslant t_2 \wedge y_0), I(t_1 \wedge y_i < T_i \leqslant t_2 \wedge y_i),\\
&\qquad I(t_1 \wedge y_{i+j} < T_{i+j} \leqslant t_2 \wedge y_{i+j}),\\
&\qquad I(t_1 \wedge y_{i+j+k} < T_{i+j+k} \leqslant t_2 \wedge y_{i+j+k})]\\
&\qquad \times dG(y_0)\, dG(y_i)\, dG(y_{i+j})\, dG(y_{i+j+k}).\end{aligned}$$

Next, observe that, for any $y \geqslant 0$,

$$EI(t_1 \wedge y < T_i \leqslant t_2 \wedge y) = F(t_2 \wedge y) - F(t_1 \wedge y) \leqslant F(t_2) - F(t_1).$$

Then, employ the approach used in the proof of Lemma 4.4 in Yu (1993) in order to obtain:

$$E[\tilde{N}_n(t) - \tilde{N}_n(s)]^4 \leqslant C[n^{\tau_0 - 1} + (F(t_2) - F(t_1))^{6/5}] \tag{4.15}$$

for the PA case. For the NA case, use Lemma 3.2 in Cai and Roussas (1995), which is the appropriate version of Lemma 4.2 in Yu (1993), and the approach mentioned above in order to arrive at (4.15). Therefore, (4.15) is true, both for the PA and the NA case. This completes the proof of the lemma. ∎

Proof of Proposition 4.2. Once inequality (4.14) has been established, a repetition of the arguments in Yu (1993) (see pp. 363–364) leads to the following inequality. For every ε and $\eta > 0$, there is a $\delta > 0$ such that, for all sufficiently large n:

$$P[\sup\{|\tilde{N}_n(t) - \tilde{N}_n(s)| \geqslant 4\varepsilon;\ s \leqslant t \leqslant s + \delta\}] \leqslant \eta\delta. \tag{4.16}$$

However, inequality (4.16) ensures tightness of the process $\{\tilde{N}_n(t);\ 0 \leqslant t \leqslant \tau\}$ by relation (15.22) and Theorem 15.6 in Billingsley (1968, pp. 128–129). Since the process $\{\tilde{Y}_n(t);\ t \in [0, \tau]\}$ is likewise tight by Lemma 4.1, the proof of the proposition is concluded. ∎

5. VARIANCE ESTIMATION: PROOF OF THEOREM 1.6

If the r.v.s $\{T_j;\ j \geqslant 1\}$ are independent, then for each time point t, Greenwood's formula (see, for example, Cox and Oakes, 1984, p. 50) provides a valid estimate for the asymptotic variance of the K-M estimator at t. However, as Ying and Wei (1994) pointed out, Greenwood's formula may not be valid for dependent situations (see Ying and Wei, 1994, p. 18). These authors have analyzed a real data set from a tumorigenesis in a litter-matched experiment. All experimental rates are sacrificed at the end of 104 weeks. Let X_1 be the observation from a drug-treated rat, and let X_2 and X_3 be the responses from the litter-matched controls. The main interest here lies in estimating the common marginal d.f. F of the tumor appearance time (in weeks) for the controls. It is argued that, if the failure times are highly positively correlated in each stratum of highly stratified survival times, one would expect that the variance estimate provided by the authors would tend to be much larger than the standard estimate with independent observations. This did not materialize in the specified data set due, perhaps, to weak litter effect (see Ying and Wei, 1994, p. 22). The valid estimate given by Ying and Wei (1994) is the expression at the bottom of page 21 of their paper. The estimate proposed here in (5.4) is similar to that of Ying and Wei (1994). They are both constructed in the same way and by utilizing the same entities, including suitable estimates of the quantities $H(u, v)$ and $M(u, v)$ mentioned in Remark 4.1. The persuasive argument, advanced by Ying and Wei (1994) that one would expect larger variance estimates under highly positive correlation than under independence, is quite appropriate for the case of PA discussed here. The estimate given in (5.4) is also related to the variance estimate derived in Künsch (1989) by using jackknife procedure, as well as to Carlestein's (1986) variance estimate based on nonoverlapping blocks.

The present section is devoted to studying a valid estimate of the asymptotic variance of $\tilde{F}_n(t)$. It follows from Theorem 1.5, that the limiting variance of $\sqrt{(n)}\,\tilde{F}_n(t)$ is $\sigma^2(t, t)$, defined in (4.7); namely,

$$\sigma^2(t, t) = \int_0^t \int_0^t \frac{dM(u, v)}{\bar{H}(u)\,\bar{H}(v)}.$$

One way to construct an estimate of $\sigma^2(t, t)$ is to estimate $M(u, v)$ first. For this purpose, consider the following consistent estimate of $M(u, v)$,

$$\hat{M}_n(u, v) = \frac{1}{n-l+1} \sum_{j=0}^{n-l} \tilde{S}_j(l, u)\,\tilde{S}_j(l, v), \tag{5.1}$$

where, for $0 \leqslant j \leqslant n-l$,

$$\tilde{S}_j(l, u) = \frac{1}{\sqrt{l}} \sum_{k=j+1}^{j+l} \left[\delta_k I(Z_k \leqslant u) - \int_0^u I(Z_k \geqslant z)\, d\hat{\Lambda}_n(z) \right], \tag{5.2}$$

and $1 \leqslant l \leqslant n$ is such that:

$$l = l_n \to \infty, \qquad \text{and} \qquad l_n = o(n^{\theta}), \tag{5.3}$$

where θ is given in Theorem 1.3. Therefore, the proposed consistent estimator of the asymptotic variance of $\sqrt{n}\,\tilde{F}_n(t)$ is $n\hat{V}(t)$, where

$$\hat{V}(t) = n[1 - \tilde{F}_n(t)]^2 \int_0^t \int_0^t \frac{d\hat{M}_n(u, v)}{Y_n(u)\, Y_n(v)}. \tag{5.4}$$

Remark 5.1. Theorem 1.3 implies that l_n is related to the decay rate of the covariance between T_1 and T_n. The size of l_n or the selection of parameter θ is in particular a tractable problem under additional information on the size of covariance and should be studied for each individual problem separately, see Künsch (1989) and Peligrad and Shao (1995) for details. As a rule of thumb, one may take $\theta = 1/3$.

Substituting (5.1) and (5.2) into (5.4), and simplifying, we have:

$$\hat{V}(t) = \frac{n(1 - \tilde{F}_n(t))^2}{n - l + 1} \sum_{j=0}^{n-l} \left(\frac{D_j(l, t)}{\sqrt{l}} \right)^2,$$

where, for $0 \leqslant j \leqslant n - l$,

$$D_j(l, t) = \sum_{i=j+1}^{j+l} \left[\frac{\delta_i I(Z_i \leqslant t)}{Y_n(Z_i)} - \sum_{k=1}^{n} \frac{\delta_k I(Z_k \leqslant Z_i \wedge t)}{Y_n^2(Z_k)} \right].$$

In order to show that $n\hat{V}(t)$ is a consistent estimate of $\sigma^2(t, t)$, it suffices to show, by Theorem 1.2 and Proposition 2.1, that $\hat{M}_n(u, v)$ is a consistent estimate of $M(u, v)$, for any $u, v \in [0, \tau]$. For this purpose and for $0 \leqslant j \leqslant n - l$, let

$$S_j(l, u) = \frac{1}{\sqrt{l}} \sum_{k=j+1}^{j+l} \left[\delta_k I(Z_k \leqslant u) - \int_0^u I(Z_k \geqslant z)\, d\Lambda(z) \right],$$

and

$$M_n(u, v) = \frac{1}{n - l + 1} \sum_{j=0}^{n-l} S_j(l, u)\, S_j(l, v). \tag{5.5}$$

Then, it follows from (5.2), Theorem 1.3 and (5.3) that, for all $0 \leqslant j \leqslant n-l$ and $u \in [0, \tau]$,

$$|S_j(l, u) - \tilde{S}_j(l, u)| \leqslant \frac{1}{\sqrt{l}} \sum_{k=j+1}^{j+l} |\hat{\Lambda}_n(u \wedge Z_k) - \Lambda(u \wedge Z_k)|$$
$$= o(n^{-\theta} \sqrt{l_n}) \quad a.s.; \tag{5.6}$$

also,

$$S_j(l, u) = O(\sqrt{l_n}) \quad a.s. \qquad \text{and} \qquad \tilde{S}_j(l, u) = O(\sqrt{l_n}) \quad a.s..$$

Therefore, in view of (5.3), it suffices to show that $M_n(u, v)$ converges to $M(u, v)$ in L_2 in order to establish that so does $\hat{M}_n(u, v)$. Set:

$$\xi_k(u) = I(T_k > u \wedge Y_k) + \Lambda(T_k \wedge u \wedge Y_k), \qquad \text{and}$$
$$A_j(l, u) = \frac{1}{\sqrt{l}} \sum_{k=j+1}^{j+l} \xi_k(u).$$

Then $\xi_k(u)$ is a nondecreasing, nonnegative and bounded function of T_k, $j+1 \leqslant k \leqslant j+l$, $j = 0, ..., n-l$, for any fixed u. Therefore, it follows from Lemma 2.2, that $\{\xi_k(u); k \geqslant 1\}$ is either PA or NA, according to the positive or negative association of $\{T_k; k \geqslant 1\}$. A simple calculation yields:

$$S_j(l, u) = \sqrt{l} - \frac{1}{\sqrt{l}} \sum_{k=j+1}^{j+l} \xi_k(u) = \sqrt{l} - A_j(l, u),$$

and

$$M_n(u, v) = l - \frac{\sqrt{l}}{n-l+1} \sum_{j=0}^{n-l} [A_j(l, u) + A_j(l, v)]$$
$$+ \frac{1}{n-l+1} \sum_{j=0}^{n-l} A_j(l, u)\, A_j(l, v).$$

It is also easily seen that:

$$\operatorname{Var}(M_n(u, v)) \leqslant \frac{4 l_n}{(n-l+1)^2} \left[\operatorname{Var}\left(\sum_{j=0}^{n-l} A_j(l, u) \right) + \operatorname{Var}\left(\sum_{j=0}^{n-l} A_j(l, v) \right) \right]$$
$$+ \frac{2}{(n-l+1)^2} \operatorname{Var}\left(\sum_{j=0}^{n-l} A_j(l, u)\, A_j(l, v) \right). \tag{5.7}$$

For the NA case, it follows that, for $|i-j|>l$, $\operatorname{Cov}(A_i(l,u), A_j(l,u)) \leqslant 0$ and $\operatorname{Cov}(A_i(l,u)\, A_i(l,v), A_j(l,u)\, A_j(l,v)) \leqslant 0$. Therefore, it follows, by the Cauchy–Schwarz inequality, that:

$$\begin{aligned}
\operatorname{Var}&\left(\sum_{j=0}^{n-l} A_j(l,u)\right)\\
&=(n-l+1)\operatorname{Var}(A_0(l,u))+2\sum_{i=1}^{l}(n-l-i+1)\operatorname{Cov}(A_0(l,u), A_i(l,u))\\
&\quad+2\sum_{i=l+1}^{n-l}(n-l-i+1)\operatorname{Cov}(A_0(l,u), A_i(l,u))\\
&\leqslant(n-l+1)\operatorname{Var}(\xi_1(u))+2\sum_{i=1}^{l}(n-l-i+1)\operatorname{Var}(A_0(l,u))=O(l_n n).
\end{aligned}$$

That is,

$$\operatorname{Var}\left(\sum_{j=0}^{n-l} A_j(l,u)\right)=O(l_n n). \tag{5.8}$$

Similarly,

$$\operatorname{Var}\left(\sum_{j=0}^{n-l} A_j(l,v)\right)=O(l_n n). \tag{5.9}$$

An argument similar to the one used in the proof of (5.8) yields

$$\begin{aligned}
\operatorname{Var}&\left(\sum_{j=0}^{n-l} A_j(l,u)\, A_j(l,v)\right)\\
&\leqslant(n-l+1)\operatorname{Var}(A_0(l,u)\, A_0(l,v))\\
&\quad+2(n-l+1)\sum_{m=1}^{l}\operatorname{Cov}(A_0(l,u)\, A_0(l,v), A_m(l,u)\, A_m(l,v))\\
&\leqslant(n-l+1)(2l+1)\operatorname{Var}(A_0(l,u)\, A_0(l,v)),
\end{aligned} \tag{5.10}$$

whereas it is easily seen that

$$\begin{aligned}
\operatorname{Var}&(A_0(l,u)\, A_0(l,v))\\
&\leqslant 2\operatorname{Var}[(A_0(l,u)-EA_n(l,u))\, A_0(l,v)]+2(EA_0(l,u))^2\operatorname{Var}(A_0(l,v))\\
&\leqslant 2E[(A_0(l,u)-EA_n(l,u))\, A_0(l,v)]^2+O(l_n)\\
&=O(l_n)\operatorname{Var}(A_0(l,u))+O(l_n)=O(l_n).
\end{aligned} \tag{5.11)}$$

Substituting (5.11) into (5.10), we obtain

$$\operatorname{Var}\left(\sum_{j=0}^{n-l} A_j(l,u)\, A_j(l,v)\right)=O(l_n^2 n). \tag{5.12}$$

Substitution of (5.8), (5.9) and (5.12) into (5.7) leads to

$$\operatorname{Var}(M_n(u,v))=O\left(\frac{l_n^2}{n-l_n+1}\right). \tag{5.13}$$

Next, we consider the PA case. Since $\{\xi_k(u); k\geqslant 1\}$ are PA, so are $\{A_j(l,u); 0\leqslant j\leqslant n-l\}$, for any fixed u. By the c_r-inequality and Assumption (K5), we have

$$\begin{aligned}
&l_n \operatorname{Var}\left(\sum_{j=0}^{n-l} A_j(l,u)\right)\\
&\quad=\operatorname{Var}\left(\sum_{j=1}^{l-1} j\xi_j(u)+l\sum_{j=l}^{n-l+1}\xi_j(u)+\sum_{j=n-l+2}^{n}(n-j+1)\,\xi_j(u)\right)\\
&\quad\leqslant 8\operatorname{Var}\left(\sum_{j=1}^{l-1} j\xi_j(u)\right)+4l^2\operatorname{Var}\left(\sum_{j=1}^{n-2l+1}\xi_j(u)\right)=O(l_n^2 n),
\end{aligned}$$

so that

$$\operatorname{Var}\left(\sum_{j=0}^{n-l} A_j(l,u)\right)=O(l_n n). \tag{5.14}$$

Similarly,

$$\operatorname{Var}\left(\sum_{j=0}^{n-l} A_j(l,v)\right)=O(l_n n). \tag{5.15}$$

Employing the same arguments as those used in the proof of (5.11) and Assumption (K5) again, we obtain

$$\operatorname{Var}(A_0(l,u)\, A_0(l,v))=O(l_n).$$

Thus,

$$\begin{aligned}
\operatorname{Var}&\left(\sum_{j=0}^{n-l} A_j(l,u)\,A_j(l,v)\right)\\
&=(n-l+1)\operatorname{Var}(A_0(l,u)\,A_0(l,v))\\
&\quad+2\sum_{m=1}^{2l}(n-l+1-m)\operatorname{Cov}(A_0(l,u)\,A_0(l,v),A_m(l,u)\,A_m(l,v))\\
&\quad+2\sum_{m=2l+1}^{n-l}(n-l+1-m)\operatorname{Cov}(A_0(l,u)\,A_0(l,v),A_m(l,u)\,A_m(l,v))\\
&\leqslant(4l+1)(n-l+1)\operatorname{Var}(A_0(l,u)\,A_0(l,v))\\
&\quad+2\sum_{m=2l+1}^{n-l}(n-l+1-m)\operatorname{Cov}(A_0(l,u)\,A_0(l,v),A_m(l,u)\,A_m(l,v))\\
&=O(l_n^2 n)+2\sum_{m=2l+1}^{n-l}(n-l+1-m)\\
&\quad\times\operatorname{Cov}(A_0(l,u)\,A_0(l,v),A_m(l,u)\,A_m(l,v)).
\end{aligned}$$

That is,

$$\begin{aligned}
\operatorname{Var}&\left(\sum_{j=0}^{n-l} A_j(l,u)\,A_j(l,v)\right)\\
&=O(l_n^2 n)+2\sum_{m=2l+1}^{n-l}(n-l+1-m)\\
&\quad\times\operatorname{Cov}(A_0(l,u)\,A_0(l,v),A_m(l,u)\,A_m(l,v)).
\end{aligned}\tag{5.16}$$

By a simple manipulation and Assumption (K5), the second term on the right-hand side of (5.16) becomes

$$\sum_{m=2l+1}^{n-l}(n-l+1-m)\operatorname{Cov}(A_0(l,u)\,A_0(l,v),A_m(l,u)\,A_m(l,v))=O(l_n^2 n).$$

Hence

$$\operatorname{Var}\left(\sum_{j=0}^{n-l} A_j(l,u)\,A_j(l,v)\right)=O(l_n^2 n).\tag{5.17}$$

Substituting (5.14), (5.15) and (5.17) into (5.7), we have:

$$\operatorname{Var}(M_n(u,v))=O\left(\frac{l_n^2}{n-l+1}\right).\tag{5.18}$$

It is easily seen that $EM_n(u, v) \to M(u, v)$. This, in conjunction with (5.3), (5.13) and (5.18), implies that $M_n(u, v)$ converges to $M(u, v)$ in L_2. Then $\hat{M}_n(u, v)$ converges to $M(u, v)$ in L_2. This completes the proof of the theorem. ∎

ACKNOWLEDGMENT

The authors are indebted to an anonymous referee whose truly constructive comments helped to improve an earlier version of this paper. This research was supported in part by a research grant to the second author from the University of California, Davis.

REFERENCES

I. Bagai and B. L. S. Prakasa Rao, Estimation of the survival function for stationary associated processes, *Statist. Probab. Lett.* **12** (1991), 385–391.

R. E. Barlow and F. Proschan, "Statistical Theory of Reliability and Life Testing: Probability Models," Holt, Rinehart and Winston, New York, 1978.

P. Billingsley, "Convergence of Probability Measures," Wiley, New York, 1968.

T. Birkel, Moment bounds for associated sequences, *Ann. Probab.* **16** (1988), 1184–1193.

T. Birkel, A note on the strong law of large numbers for positively dependent random variables, *Statist. Probab. Lett.* **7** (1989), 17–20.

H. W. Block, T. H. Savits, and M. Shaked, Some concepts of negative dependence, *Ann. Probab.* **10** (1982), 765–772.

A. Bozorgnia, R. F. Patterson, and R. L. Taylor, Limit theorems for negatively dependent random variables, unpublished manuscript, 1993.

N. Breslow and J. Crowley, A large sample study of the life table and product limit estimators under random censorship, *Ann. Statist.* **2** (1974), 437–453.

E. C. Brindley, Jr. and W. A. Thompson, Jr., Dependence and aging aspects of multivariate survival, *J. Amer. Statist. Assoc.* **67** (1972), 822–830.

R. Burton, A. R. Dabrowski, and H. Dehling, An invariance principle for weakly associated random vectors, *Stochastic Process. Appl.* **23** (1986), 301–306.

Z. W. Cai and G. G. Roussas, Weak convergence for smooth estimate of a distribution function under association, Technical Report, No. 307, Division of Statistics, University of California, Davis, 1995. Also, *Stochastic Anal. Appl.* **17** (1999).

E. Carlstein, The use of subseries values for estimating the variance of a general statistic from a stationary sequence, *Ann. Statist.* **14** (1986), 1171–1179.

D. R. Cox and D. Oakes, "Analysis of Survival Data," Chapman and Hall, London, 1984.

N. Ebrahimi and M. Ghosh, Multivariate negative dependence, *Commun. Statist. Theory Meth.* **10** (1981), 307–337.

T. Egeland, Mixing dependence and system reliability, *Scand. J. Statist.* **19** (1992), 173–183.

J. D. Easry, F. Proschan, and D. W. Walkup, Association of random variables, with applications, *Ann. Math. Statist.* **38** (1967), 1466–1474.

C. M. Fortuin, P. W. Kastelyn, and J. Ginibre, Correlation inequalities on some partially ordered sets, *Commun. Math. Phys.* **22** (1971), 89–103.

R. D. Gill, "Censoring and Stochastic Integrals", Mathematical Centre tracts, No. 124, Mathematisch Centrum, Amsterdam, 1980.

R. D. Gill, Testing with replacement and the product limit estimator, *Ann. Statist.* **9** (1981), 853–860.

R. D. Gill, Large sample behavior of the product limit estimator on the whole line, *Ann. Statist.* **11** (1983), 49–56.

T. E. Harris, A lower bound for the critical probability in a certain percolation process, *Proc. Camb. Phil. Soc.* **56** (1960), 13–20.

K. Joag-Dev, Independence via uncorrelatedness under certain dependence structures, *Ann. Probab.* **11** (1983), 1037–1041.

K. Joag-Dev and F. Proschan, Negative association of random variables, with applications, *Ann. Statist.* **11** (1983), 286–295.

E. L. Kaplan and P. Meier, Nonparametric estimation from incomplete observations, *J. Amer. Statist. Assoc.* **53** (1958), 457–481.

H. R. Künsch, The jackknife and the bootstrap for general stationary observations, *Ann. Statist.* **17** (1989), 1217–1241.

E. Lehmann, Some concepts of dependence, *Ann. Math. Statist.* **37** (1966), 1137–1153.

C. M. Newman, Normal fluctuations and the FKG inequalities, *Commun. Math. Phys.* **74** (1980), 119–128.

C. M. Newman, Ising models and dependent percolation, *in* "Topics in Statistical Dependence" (H. W. Block, A. R. Sampson, and T. H. Savits, Eds.), IMS Lecture Notes-Monograph Series, Vol. 16, pp. 395–401, Hayward, CA, 1990.

M. Peligrad and Q. Shao, Estimation of variance of partial sums of an associated sequence of random variables, *Stochastic Process. Appl.* **56** (1995), 307–319.

A. V. Peterson, Expressing the Kaplan–Meier estimator as a function of empirical subsurvival functions, *J. Amer. Statist. Assoc.* **72** (1977), 854–858.

G. G. Roussas, Kernel estimates under association: Strong uniform consistency, *Statist. Probab. Lett.* **12** (1991), 393–403.

G. G. Roussas, Asymptotic normality of random fields of positively or negatively associated processes, *J. Multivariate Anal.* **50** (1994), 152–173.

G. G. Roussas, Exponential probability inequalities with some applications, *in* "Statistics, Probability and Game Theory: Papers in Honor of David Blackwell" (T. S. Ferguson, L. S. Shapley, and J. B. Macqueen, Eds.), IMS Lecture Notes-Monograph Series, Vol. 30, Hayward, CA, 1996.

G. R. Shorack and J. A. Wellner, "Empirical Processes with Applications to Statistics," Wiley, New York, 1986.

W. Stute and J. L. Wang, A strong law under random censorship, *Ann. Statist.* **21** (1993), 1591–1607.

W. Stute, Convergence of Kaplan–Meier estimator in weighted sup-norms, *Statist. Probab. Lett.* **20** (1994), 219–223.

H. Tucker, "A Graduate Course in Probability," Academic Press, New York, 1967.

J. Voelkel and J. Crowley, Nonparametric inference for a class of semi-Markov process with censored observation, *Ann. Statist.* **12** (1984), 142–160.

J. G. Wang, A note on the uniform consistency of the Kaplan–Meier estimator, *Ann. Statist.* **15** (1987), 1313–1316.

Z. Ying and L. J. Wei, The Kaplan–Meier estimate for dependent failure time observations, *J. Multivariate Anal.* **50** (1994), 17–29.

H. Yu, A Glivenko–Cantelli lemma and weak convergence for empirical processes of associated sequences, *Probab. Theor. Relat. Fields* **95** (1993), 357–370.