Journal of the Korean
Statistical Society
Vol. 26, No. 1, 1997

Convergence of Score process in the Cox Proportional Hazards Model†

Jinsoo Hwang[1]

Abstract

We study the asymptotic behavior of the maximum partial likelihood estimator in the Cox proportional hazards model in the presence of nuisance parameters when the entry of patients is staggered. When entry of patients is simultaneous and there is only one regression parameter in the Cox model, the efficient score process of the partial likelihood is martingale and converges weakly to a time-changed Brownian motion. Our problem is to get a similar result in the presence of nuisance parameters when entry of patient is staggered.

Key Words : Score process; Staggered-entry; Maximum partial likelihood estimator.

†This work supported by a research grant from Korea Science and Engineering Foundation, 96-0701-01-01-3.

[1]Department of Statistics, Inha University, Inchon, 402-751, Korea.

1. INTRODUCTION

We have seen a rapid growth in the field of survival analysis with censored data since Cox(1972,1975) introduced the proposed proportional hazards model. Cox proportional hazards model assumes that

$$\lambda_i(s) = \lambda_0(s)\exp(\beta' z_i),$$

where $\beta = (\theta, \phi_1, \dots, \phi_{p-1})'$ is a vector of unknown parameters, and $\lambda_0(s)$ is an unknown baseline hazard function which is common to all patients. It may sometimes be the case that the first coordinate θ of β is of primary interest. For instance, the first coordinate of z_i might be a treatment indicator, with $z_i = 0$ if patient i is receiving treatment A and $z_i = 1$ if patient i is receiving treatment B. Then the relative efficacy of the treatments would depend on the value of θ. The other coordinates of z_i might be other covariates affecting survival time(for example sex, race, or blood pressure), but their effects on survival might be of secondary interest. In this case, the other coordinates $\phi_1, \dots, \phi_{p-1}$ of β might be regarded as nuisance parameters. This is the sort of situation we have in mind in this paper, and our goal is to probabilistically describe how the natural estimator of θ behaves over time as information accumulates during the course of a staggered-entry clinical trial.

Let $\hat{\beta}(t)$ be the maximum partial likelihood estimator of β at time t. We are going to show that $(\hat{\theta}(t) - \theta)/V\hat{a}r(\hat{\theta}(t))$ is approximately standard Brownian motion when $1/V\hat{a}r(\hat{\theta}(t))$ is used as a time scale. Here, $1/V\hat{a}r(\hat{\theta}(t))$ is the observed Fisher information for θ in the partial likelihood, evaluated at $\hat{\beta}(t)$. The basic condition needed to make the theorem true is that individual patients are uniformly negligible, although shortcomings in our methods of proof force us to make some additional technical assumptions.

2. NOTATIONS AND FORMULATION OF THE MODEL

The clinical trials discussed in this paper will have the following form. First, we have possibly infinite sequence of entry times $0 \le y_1 \le y_2 \le \dots$ with the property that there are only finitely many entry times in each interval $[0, t]$. For each entry time y_i, there is a censoring time $c_i \in [0, \infty)$ and a p-dimensional covariate process $\{z_i(s), s \ge 0\}$ assumed to be measurable; real-valued function bounded in absolute value by a fixed constant B. Let

$\Lambda(s), s \geq 0$ be a continuous, nondecreasing baseline cumulative hazard function which satisfies $\Lambda(0) = 0$. Patient i entering at calendar time y_i has a survival time X_i which has survival function

$$Pr\{X_i \geq s\} = \exp\{-\int_0^s e^{\beta' z_i(u)} d\Lambda(u)\}.$$

Thus, the cumulative hazard function Λ_i of X_i satisfies

$$d\Lambda_i(s) = e^{\beta' z_i(s)} d\Lambda(s).$$

The survival times are assumed to be independent of each other. Note that the survival times are the only random quantities in the model. When the clinical trial is conducted, the i-th patient is on test during the calendar time interval $[y_i, y_i + X_i \wedge c_i]$. At time $y_i + X_i \wedge c_i$ we observe the death of the i-th patient if $X_i \leq c_i$ and otherwise observe that he is censored. At any time t there is in effect a second censoring variable $(t - y_i)^+$ in the sense that the time on test of patient i prior to t is $X_i \wedge c_i \wedge (t - y_i)^+$. We shall refer to X_i, c_i, and $(t - y_i)^+$ as "age" variables: the age of the i-th patient at death, at censoring, and at time t, respectively. Define the set of patients at risk at age s and time t by

$$R(t, s) = \{i : s \leq X_i \wedge c_i \wedge (t - y_i)^+\}.$$

Let $N_i(t, s)$ be an indicator function for the event that patient i was observed to die before or at time t and at an age less than or equal to s, so that

$$N_i(t, s) = I_{\{X_i \leq s \wedge c_i \wedge (t - y_i)^+\}}.$$

Let $A_i(t, s)$ be the amount of hazard to which patient i has been exposed before time t and age s, so that

$$A_i(t, s) = \Lambda_i\{s \wedge X_i \wedge c_i \wedge (t - y_i)^+\}.$$

For $s \geq 0$, let $\mathcal{F}_s$ be the σ-algebra generated by $I_{\{X_i \leq s\}}$ and $X_i I_{\{X_i \leq s\}}$. Thus, $\mathcal{F}_s$ is the σ-algebra of events of age less than or equal to s. It is easy to show(Sellke and Siegmund(1983)) that $M_i(t, s)$ defined by

$$M_i(t, s) = N_i(t, s) - A_i(t, s)$$

is an $\mathcal{F}_s$-martingale in s for each fixed t.

The Cox(1975) log partial likelihood for β at time t can be expressed as

$$l(t,\beta) = \sum \int_0^t \left[\beta' z_i(u) - \log\left\{ \sum_{j\in R(t,u)} e^{\beta' z_j(u)} \right\} \right] N_i(t, du). \quad (2.1)$$

Differentiating (2.1) with respect to $\beta = (\theta, \phi_1, \ldots, \phi_{p-1})'$ gives the efficient score vector $\dot{l}(t,\beta)$ at time t whose components given by

$$l_\theta(t,\beta) = \sum \int_0^t \{ z_{\theta i}(u) - \mu_\theta(t,u)\} N_i(t, du) \quad (2.2)$$

$$l_{\phi_j}(t,\beta) = \sum \int_0^t \{ z_{\phi_j i}(u) - \mu_{\phi_j}(t,u)\} N_i(t, du) \quad (2.3)$$

where

$$\mu_\theta(t,s) = \frac{\sum_{j\in R(t,s)} z_{\theta j}(s) e^{\beta' z_j(s)}}{\sum_{j\in R(t,s)} e^{\beta' z_j(s)}}$$

and the $\mu_{\phi_j}(t,s)$ are defined similarly($z_{\theta i}(s)$ denotes the covariate coordinate corresponding to parameter θ for patient i at age s.). So

$$\dot{l}(t,\beta) = (l_\theta(t,\beta), l_{\phi_1}(t,\beta), \ldots, l_{\phi_{p-1}}(t,\beta))'.$$

Minus the second derivative of (2.1) with respect to θ and ϕ gives the observed Fisher information matrix $i(t,\beta)$ as follows.

$$i(t,\beta) = \begin{bmatrix} -l_{\theta\theta}(t,\beta) & -l'_{\theta\phi}(t,\beta) \\ -l_{\theta\phi}(t,\beta) & -l_{\phi\phi}(t,\beta) \end{bmatrix}$$

where

$$l'_{\theta\phi}(t,\beta) = l_{\theta\phi_k}(t,\beta)_{(k=1,\ldots,p-1)}$$

$$l_{\phi\phi}(t,\beta) = (l_{\phi_l\phi_m})_{(l,m=1,\ldots,p-1)}$$

and

$$-l_{\theta\theta}(t,\beta) = \sum_i \int_0^t \frac{\sum_{j\in R(t,s)} \{z_{\theta j}(u) - \mu_\theta(t,u)\}^2 e^{\beta' z_j(s)}}{\sum_{j\in R(t,s)} e^{\beta' z_j(s)}} N_i(t, du).$$

Other components of $i(t,\beta)$ defined similarly. Let $I(t,\beta)$ be the expected Fisher information matrix, i.e., $I(t,\beta) = E\{i(t,\beta)\}$.

To study the asymptotic behavior of the score process, we study the behavior in two-dimensional time of

$$l_\theta(t,s,\beta) = \sum_i \int_0^s \{z_{\theta i}(u) - \mu_\theta(t,u)\} N_i(t,du)$$

$$l_\phi(t,s,\beta) = \sum_i \int_0^s \{z_{\phi i}(u) - \mu_\phi(t,u)\} N_i(t,du).$$

One can think of $l_\theta(t,s,\beta)$ and $l_\phi(t,s,\beta)$ as the efficient scores at time t of events of age less than or equal to s. Likewise, we can define the observed Fisher information matrix at time t and age s in two-dimensional time scale.

By simple algebra we have

$$l_\theta(t,s,\beta) = \sum_i \int_0^s \{z_{\theta i}(u) - \mu_\theta(t,u)\} M_i(t,du)$$

$$l_\phi(t,s,\beta) = \sum_i \int_0^s \{z_{\phi i}(u) - \mu_\phi(t,u)\} M_i(t,du)$$

Since the bounded integrands in the above are $\mathcal{F}_s$-predictable in u and since we are integrating with respect to martingales, it follows that $l_\theta(t,s,\beta)$ and all the coordinates of $l_\phi(t,s,\beta)$ are $\mathcal{F}_s$-martingales for each fixed t by standard theory of stochastic integration. The predictable variance process of $l_\theta(t,s,\beta)$ is given by

$$\begin{aligned} \langle l_\theta(t,\cdot,\beta)\rangle(s) &= \sum_i \int_0^s \{z_{\theta i}(u) - \mu_\theta(t,u)\}^2 A_i(t,du) \\ &= \sum_i \int_0^s \{z_{\theta i}^2(u) - \mu_\theta^2(t,u)\} A_i(t,du) \end{aligned}$$

and the predictable variance process matrix of $l_\phi(t,s,\beta)$ is given similarly.

3. MAIN THEOREM

Suppose that, for each $n = 1, 2, \ldots$, we have a stochastic clinical trial of the sort described in the previous section. The configuration of entry times, the covariate processes, the baseline hazard function, and the parameter β are all allowed to differ from trial to trial. The subscript n will indicate quantities associated with the n-th clinical trial, though this subscript may be omitted when there is no danger of ambiguity. Suppose that the first coordinate of $\beta = (\theta, \phi_1, \ldots, \phi_{p-1})'$ is of primary interest, with the other coordinates being regarded as nuisance parameters. This would typically be the case when the first coordinate is a treatment indicator in a clinical trial comparing two treatment. Let $1/l^{\theta\theta}(t,s,\beta)$ be the leading element of the inverse of the observed Fisher information matrix, so that

$$l^{\theta\theta}(t,s,\beta) = l_{\theta\theta}(t,s,\beta) - l_{\theta\phi}(t,s,\beta) l_{\phi\phi}^{-1}(t,s,\beta) l_{\theta\phi}(t,s,\beta).$$

Sellke and Siegmund(1983) use the reciprocal of the asymptotic variance of the estimator as a clock time, which is the Fisher information for the parameter. Likewise, since the asymptotic variance of $\hat{\theta}(t)$ in the multidimensional parameter case is $-1/l^{\theta\theta}(t,\beta) = -1/l^{\theta\theta}(t,t,\beta)$, we are going to use $-l^{\theta\theta}(t,\beta)$ as our clock time. Of course, in practice one must use an esimator of β, such as the maximum partial likelihood estimator $\hat{\beta}(t)$, to define the clock time. See Grambsch(1983) or Lai and Siegmund(1983) for discussion of the use of Fisher information as a means of rescaling time. Define for $v \geq 0$

$$\tau_n(v) = \inf\{t : -l^{\theta\theta}(t,\beta) \geq nv\}$$

and

$$T_n(v) = \inf\{t : -I^{\theta\theta}(t,\beta) \geq nv\}$$

where

$$I^{\theta\theta}(t,\beta) = El_{\theta\theta}(t,\beta) - El'_{\theta\phi}(t,\beta)\{El_{\phi\phi}(t,\beta)\}^{-1} El_{\theta\phi}(t,\beta)$$

and $\phi = (\phi_1, \ldots, \phi_{p-1})'$. Modulo the scaling by n, $T_n(v)$ is the calendar time in the n-th trial at which the "expected information for θ clock" reaches level v. Note that $\tau_n(v)$ is a random process, while $T_n(v)$ is deterministic. Before we proceed, here are the assumptions we need throughout this paper.

- A.1 All covariate process $z_i^{(n)}(s)$ in all clinical trials are bounded in absolute value by a fixed constant B.
- A.2 There exist a constant K such that $\|\beta^{(n)}\| < K$ for any n.
- A.3 There exist a fixed small constant $\delta > 0$ such that $T_n(1+2\delta) < \infty$ for all n.
- A.4 For some fixed constant $v_* > 0$, $\lambda_{min}(T_n(v_*)) \geq O(n^{3/5})$, where $\lambda_{min}(t)$ is the minimum eigenvalue of the Fisher information matrix $I(t,\beta)$.
- A.5 There exists a fixed constant $b \in R$ such that the number of patients entering before time $T_n(1+2\delta)$ is $o(n^b)$ as $n \to \infty$ for $\delta > 0$.

Let $v > 0$ and define

$$\hat{\tau}_n(v) = \inf\{t : -l^{\theta\theta}(t, \hat{\beta}(t)) \geq nv\}$$

Theorem 1. Let v_{**} be a constant with $0 < v_* < v_{**} < 1$. As $n \to \infty$, then under A.1-A.5

$$n^{1/2} v[\hat{\theta}(\hat{\tau}_n(v)) - \theta] \xrightarrow{\mathcal{D}} W(v).$$

where $W(\cdot)$ is standard Brownian motion and the convergence is in $D[v_{**}, 1]$ with the Skorokhod topology.

Theorem 1 is the main result of this paper. It says that properly normalized difference process $\{\hat{\theta}(t) - \theta\}$ behaves asymptotically like standard Brownian motion in an information-based time clock. Let's define a "reduced" score process for θ

$$l^\theta(t, \beta) = [l_\theta(\cdot) - El_{\theta\phi}(\cdot)(El_{\phi\phi}(\cdot))^{-1} l_\phi(\cdot)](t, \beta).$$

We can think of the above "reduced" score process as "a-directional score" where $a = (1, -El_{\theta\phi}(t, \beta) El^{-1}_{\phi\phi}(t, \beta))_{1\times p}$. That is

$$l^\theta(t, \beta) = a' \dot{l}(t, \beta).$$

By Taylor expansion argument we can easily show that the "reduced" score process and properly normalized $\hat{\theta}(t) - \theta$ process asymptotically equivalent. We can also define

$$l^\theta(t, s, \beta) = [l_\theta(\cdot) - El_{\theta\phi}(\cdot)(El_{\phi\phi}(\cdot))^{-1} l_\phi(\cdot)](t, s, \beta)$$

which can be considered as a "reduced" score for θ at time t of events of age less than or equal to s. For $v > 0$, define

$$S_n(v, s) = n^{1/2} l^\theta(T_n(v), s, \beta). \tag{3.1}$$

The main process of our interest is $S_n(v, \infty) = n^{1/2} l^\theta(T_n(v), \beta)$.

Theorem 2 For $v \in [v_*, 1]$, under the assumption A.1-A.5,

$$S_n(v, \infty) \xrightarrow{\mathcal{D}} W(v)$$

where $W(\cdot)$ is standard Brownian motion and the convergence is in $D[v_*, 1]$ with the Skorokhod topology.

Theorem 1 follows from Theorem 2 by a Taylor expansion argument, together with the consistency of $\hat{\beta}(t)$(Hwang(1995)). Theorem 2 uses the expected information for θ, $I^{\theta\theta}(t, \beta)$ to scale time, but the proof needs the fact that the predictable variance process $\left\langle l^{\theta}(t,\cdot,\beta)\right\rangle(t) = \left\langle a'\ddot{l}(t,\beta)\right\rangle$ is approximately equal to its expectation $I^{\theta\theta}(t,\beta)$. Finally, neither $I^{\theta\theta}(t,\beta)$ nor $\left\langle l^{\theta}(t,\cdot,\beta)\right\rangle(t)$ is observable, and we need to show that the observable time scale given by the observed Fisher information $-l^{\theta\theta}(t,\hat{\beta})$ is asymptotically equivalent to the other two. Otherwise the results are not statistically interesting. The following lemma states the asymptotic equivalence whose proof given by Hwang(1997).

Lemma 3.1. For $0 < \epsilon < 1/4$ and $a \in R^p$,
(a)

$$\Pr\{\sup_a \left|\frac{\left\langle a'\ddot{l}(t)\right\rangle - I_a(t)}{I_a(t)}\right| > 10p^2 n^{-1/4+\epsilon}, \text{ for some } t \in [T_n(v_*), T_n(1+\delta)]\} = o(1)$$

(b)

$$\Pr\{\sup_a \left|\frac{i_a(t) - I_a(t)}{I_a(t)}\right| > 10p^2 n^{-1/4+\epsilon}, \text{ for some } t \in [T_n(v_*), T_n(1+\delta)]\} = o(1)$$

where $I_a(t) = a'I(t,\beta)a$, $i_a(t) = a'(i(t,\beta))a$, and $\left\langle a'\ddot{l}(t)\right\rangle = a'\left\langle \ddot{l}(t)\right\rangle a$

$$\left\langle \ddot{l}(t)\right\rangle = \begin{bmatrix} \langle l_\theta(t,\beta), l_\theta(t,\beta)\rangle & \langle l_\theta(t,\beta), l_\phi(t,\beta)\rangle \\ \langle l_\theta(t,\beta), l_\phi(t,\beta)\rangle & \langle l_\phi(t,\beta), l_\phi(t,\beta)\rangle \end{bmatrix}$$

$$I(t,\beta) = \begin{bmatrix} -El_{\theta\theta}(t,\beta) & -El'_{\theta\phi}(t,\beta) \\ -El_{\theta\phi}(t,\beta) & -El_{\phi\phi}(t,\beta) \end{bmatrix} = \begin{bmatrix} I_{\theta\theta}(t) & I_{\theta\phi}(t) \\ I_{\theta\phi}(t) & I_{\phi\phi}(t) \end{bmatrix}$$

3.1 Finite dimensional convergence of score process

The score process $S_n(v,s)$ is a linear combination of $\mathcal{F}_s$-martingales, so $S_n(v,s)$ is also a $\mathcal{F}_s$-martingale. To apply the martingale central limit theorem to our score process, first we need to show that the maximum jump size of our martingale is getting smaller as $n \to \infty$.

If we set $a = (1, -I_{\theta\phi}(t)\{I_{\phi\phi}(t)\}^{-1})'$, then by simple calculus and assumption A.4 we have

$$I_{\theta\phi}(t)\{I_{\phi\phi}(t)\}^{-1} = o(n^{1/5}).$$

So the maximum jump size of $S_n(v,s)$ is bounded by $2Bn^{-1/2}(1 + o(n^{1/5}))$ which goes to zero as n tends to ∞. Also by Lemma 3.1 $\langle S_n(v,\cdot,\beta)\rangle(\infty) \to v$ as $n \to \infty$, uniformly in $v \in [v_*, 1+\delta]$. So for a fixed time point v, by the martingale CLT, $S_n(v,\cdot)$ converges weakly to normal distributions with mean zero and variance v. Now for the two dimensional distribution, fix v_1 and v_2 with $0 \le v_1 \le v_2 \le 1+\delta$.

Lemma 3.2. For v_1 and v_2, as $n \to \infty$

$$\langle S_n(v_1,\cdot,\beta), S_n(v_2,\cdot,\beta)\rangle(\infty) \xrightarrow{p} v_1.$$

Proof. Let

$$a'(v) = (1, -I_{\theta\phi}(T_n(v),\beta) I^{-1}_{\phi\phi}(T_n(v),\beta))$$

then

$$\langle S_n(v_1,\cdot,\beta), S_n(v_2,\cdot,\beta)\rangle(\infty) = n^{-1} a'(v_1)\left\langle \dot{l}(T_n(v_1))\right\rangle a(v_2) \tag{3.2}$$

where $\left\langle \dot{l}(T_n(v_1))\right\rangle$ is the predictable variance matrix of marginal score process evaluated at $T_n(v_1)$. And if we let $B(v) = I^{-1}(v,\beta)\left\langle \dot{l}(v)\right\rangle$, then we can decompose the right side of (3.2) as follows.

$$n^{-1} a'_1 \left\langle \dot{l}(v_1)\right\rangle a_2 = n^{-1}\{I^{\theta\theta}(v_1)(B_{11} - B_{12} I^{-1}_{\phi\phi}(v_2) I_{\theta\phi}(v_2)\}.$$

Using Lemma 3.1, we have $B_{11} = 1 + o_p(n^{\epsilon-1/4})$ and $B_{12} = o_p(n^{\epsilon-1/4})$. Therefore we get the result.

Lemma 3.2 is just another statement asymptotic orthogonality property of $S_n(v,\cdot)$ process. The orthogonal structure of $l_\theta(t,s,\beta)$ and $l_\phi(t,s,\beta)$ can be established easily by showing the following claim1. The proof is straightforward algebra based on Gill(1980,page 10).

Claim 1. For fixed $0 \le t_1 \le t_2 \le t_3$, the $\mathcal{F}_s$-martingales

$$l_\theta(t_1,s,\beta) \text{ and } l_{\phi_j}(t_1,s,\beta), j = 1,\ldots,p-1$$

are all orthogonal to the $\mathcal{F}_s$-martingales

$$l_\theta(t_3,s,\beta) - l_\theta(t_2,s,\beta), \text{ and } l_{\phi_k}(t_3,s,\beta) - l_{\phi_k}(t_2,s,\beta), k = 1,\ldots,p-1.$$

By previous lemma 3.1 and 3.2 the predictable variance process of the linear combination of martinagle for $c_1, c_2 \in R$

$$c_1 S_n(v_1,s) + c_2\{S_n(v_2,s) - S_n(v_1,s)\} \tag{3.3}$$

is

$$c_1^2 \langle S_n(v_1,\cdot)\rangle (s) + c_2^2 \langle S_n(v_2,\cdot) - S_n(v_1,\cdot)\rangle (s). \tag{3.4}$$

Again by orthogonality, we have

$$\langle S_n(v_2,\cdot) - S_n(v_1,\cdot)\rangle (s) \approx \langle S_n(v_2,\cdot)\rangle (s) - \langle S_n(v_1,\cdot)\rangle (s). \tag{3.5}$$

The lemma 3.1 implies that

$$\langle S_n(v,\cdot)\rangle (\infty) \xrightarrow{p} v$$

as $n \to \infty$, uniformly in $v \in [v_*, 1+\delta]$. By (3.5), the value of (3.4) at $s = \infty$ converges in probability to

$$c_1^2 v_1 + c_2^2 (v_2 - v_1) \tag{3.6}$$

The jumps of the martingale (3.3) are bounded in size by

$$4Bn^{-1/2}(1 + o(n^{1/5})\{|c_1| + |c_2|\})$$

which converges to zero as $n \to \infty$. By the martingale central limit theorem of Rebolledo(1980),

$$c_1 S_n(v_1,\infty) + c_2\{S_n(v_2,\infty) - S_n(v_1,\infty)\}$$

converges in distribution to a normal random variable with mean zero and variance (3.6). Since this holds for all $c_1, c_2 \in R$. Applying Cramér-Wold device gives us two dimensional distribution of the process $S_n(v,\cdot,\beta)$. The same argument works for general finite dimensional distributions.

3.2 Tightness of Score process

Now for tightness, define $v_i = in^{-1/5}, 0 < i < n^{1/5}(1+2\delta)$ and also define $\mathcal{F}_s$-stopping time $s^* = s_n^*$ by,

$$s^* = \inf\{s > 0 : \sup_i [\langle S_n(v_{i+1},\cdot)\rangle (s) - \langle S_n(v_i,\cdot)\rangle (s)] \geq 2n^{-1/5}\}.$$

By lemma 3.1, it is easy to see that $\Pr\{s^* = \infty\} \to 1$ as $n \to \infty$. Note

$$\langle S_n(v_{i+1},\cdot)\rangle (s) - \langle S_n(v_i,\cdot)\rangle (s) = \langle S_n(v_{i+1},\cdot) - S_n(v_i,\cdot)\rangle (s)$$

are decresing in s. In order to apply the tightness condition in Billingsley(1968), first we are going to interpolate linearly $S_n(v,s)$ as follows. For $s \geq 0$, define

$$\tilde{S}_n(v_i,s) = S_n(v_i, s \wedge s^*)$$

and for $v_i \le v \le v_{i+1}$

$$\tilde{S}_n(v,s) = \tilde{S}_n(v_i,s) + \frac{v - v_i}{v_{i+1} - v_i}\{\tilde{S}_n(v_{i+1},s) - \tilde{S}_n(v_i,s)\}.$$

By applying martingales $\tilde{S}_n(v,\infty) - S_n(v,\infty)$ and $S_n(v,\infty) - \tilde{S}_n(v,\infty)$, respectively to a continuous version of Freedman(1975)'s tail inequality, we can get the following claim.

Claim 2. As $n \to \infty$, $\sup_{v_* \le v \le 1+\delta} |\tilde{S}_n(v,\infty) - S_n(v,\infty)| \xrightarrow{p} 0$.

Proof. Let's take polynomially many points(n^{8b}) between $[0, T_n(1+2\delta)]$ as follows.

$$0 = t_0 < t_1 < \ldots < t_{n^{8b}} = T_n(1+2\delta)$$

Let $\tilde{v}_j = T_n(t_j)$. Since within any of the tiny intervals no more than one death occurs and no more than one risk set of a previously observed death is changed within the interval, therefore it suffices to consider only the $\tilde{v}_j$ points which are less than $1+\delta$. Now we know

$$\{s^* = \infty\} = \{\sup_i \langle S_n(v_{i+1},\infty) - S_n(v_i,\infty)\rangle \le 2n^{-1/5}\}$$

so, for $v_i \le \tilde{v}_j < v_{i+1}$

$$\Pr\{|S_n(\tilde{v}_j,\infty) - S_n(v_i,\infty)| > n^{-1/20}, s^* = \infty\} \tag{3.7}$$

$$\le \exp\left[-\frac{n^{-1/10}}{2B\sqrt{1/\lambda_{min}}n^{-1/20} + 4n^{-1/5}}\right] = o(n^{-8b})$$

and

$$\Pr\{|\tilde{S}_n(v_{i+1},\infty) - S_n(v_i,\infty)| > n^{-1/20}, s^* = \infty\} = o(n^{-8b}) \tag{3.8}$$

as long as $\lambda_{min} \ge O(n^{1/10+\epsilon})$. Since $\tilde{S}_n(v,\infty)$ is obtained by linearly interpolating $S_n(v,\infty)$ provided $s^* = \infty$, it follows from (3.7) and (3.8) that, for all $\tilde{v}_j \le 1+\delta$,

$$\Pr\{|\tilde{S}_n(\tilde{v}_j,\infty) - S_n(\tilde{v}_j,\infty)| > 2n^{-1/20}, s^* = \infty\} = o(n^{-8b}).$$

Since there are at most n^{8b} points $\tilde{v}_j$ to consider, the claim folows.

By above claim 2, it is now sufficient to prove $\tilde{S}_n(\cdot,\infty) \xrightarrow{\mathcal{D}} W(\cdot)$ in $D[v_*, 1+\delta]$. Again by claim 1, we already have convergence of finite dimensional distributions of $\tilde{S}_n(v,s)$. The weak convergence will follow if we can establish the tightness criterion (12.51) on Billingsely(1968). In this case, take $F(t) = c^{1/2}t, \gamma = 4$ and $\alpha = 2$ in (12.51), so that the tightness criterion to be verified is

$$E\{|\tilde{S}_n(v',\infty) - \tilde{S}_n(v,\infty)|^4\} \le c_0(v'-v)^2 \tag{3.9}$$

where, c_0 is some fixed positive constant and $0 \le v < v' \le 1+\delta$. The key tool verifying (3.9) will be a Burkholder-Davis-Gundy inequality for continuous-time, square integrable martingale. It follows from the definition of $\tilde{S}_n(v,s)$ and s^* that

$$\left\langle \tilde{S}_n(v',\cdot) - \tilde{S}_n(v,\cdot)\right\rangle(\infty) \le 2(v'-v)$$

for $0 \le v < v' \le 1+\delta$. Let $D(v',v)$ be the maximum possible jump size of the $\mathcal{F}_s$-martingale $\tilde{S}_n(v',s) - \tilde{S}_n(v,s)$. Since the jumps of the $\mathcal{F}_s$-martingale, $S_n(v,s) - S_n(v',s)$ is bounded by $2B\sqrt{1/\lambda_{min}}$. And $\tilde{S}_n(v,s)$ is defined by linear interpolation of $S_n(v,s)$ and $v_{j+1} - v_j \approx n^{-1/5}$. Then

$$D(v,v') \le 2B\sqrt{1/\lambda_{min}}\{1 \wedge \frac{v'-v}{n^{-1/5}}\}.$$

If $v'-v \ge 4B^2/\lambda_{min}$,

$$D(v,v') \le 2B\sqrt{1/\lambda_{min}} \le (v'-v)^{1/2}, \text{ for large } n.$$

If $v'-v \le 4B^2/\lambda_{min}$,

$$D(v,v') \le 2Bn^{1/5}\sqrt{1/\lambda_{min}}(v'-v) \le (4B^2/\lambda_{min})n^{1/5}(v'-v)^{1/2} \le (v'-v)^{1/2}$$

for sufficiently large n. Therefore we get,

$$\begin{aligned} E\{\tilde{S}_n(v',\infty) - \tilde{S}_n(v,\infty)|^4\} &\le c[\{3(v'-v)^{1/2}\}^4] \\ &\le 81c(v'-v)^2 \end{aligned}$$

which proves (3.9) with $c_0 = 81c$. This establish the tightness.

4. DISCUSSION

Our main theorem is stated in terms of a sequence of "clinical trials" indexed by n. As one proceeds along the sequence of clinical trials, individual patients become increasingly negligible, modulo the rescaling by n.

There are considerable similarlity between our main theorem and standard invariance principles for random walks with independent increments. Letting $X_1, X_2, \ldots$ be a martingale difference sequence with respect to a filtration $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \ldots$, one gets that the $\mathcal{F}_k$-martingale $S_k = \sum_1^k X_i$ behaves approximately like standard Brownian motion in clock time $\sum_1^k var(X_i|\mathcal{F}_{i-1})$, providing that the increments X_i are uniformly negligible. Note that the clock time $\sum_1^k var(X_i|\mathcal{F}_{i-1})$ can grow in a random way, since the variance increment at time k may depend on earlier X_i values. A somewhat analogous weakning of assumptions in our case would allow the covariates of later patients in a staggered-entry trial to depend on the survival times of earlier patients. This kind of dependence would obviously be present if patients are adaptively allocated to treatment groups as the staggered-entry trial progress. We are confident that our main theorem continues to be true in the presence of this sort of dependence, subject to mild restrictions. However, our methods of proof depend heavily on the assumption that patients are independent with respect to their covariates and survival times, given the values of all covariates and censoring times.

REFERENCES

(1) Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley, New York.

(2) Cox, D.R.(1972), Regression Models and life tables(with discussion), *Journal of Royal Statistical Society Series B,* **34,** 187-220.

(3) Cox, D.R.(1975), Partial Likelihood, *Biometrika,* **62,** 269-276.

(4) Freedman, D. (1975), On tail probabilities for martingales. *Annals of Probability,* **3**, 110-118.

(5) Gill, R. (1980), *Censoring and Stochastic Integrals,* Mathematisch Centrum 124, Amsterdam.

(6) Grambsch, P.(1983), Sequential sampling based on observed Fisher information to gurantee the accuracy of the maximum likelihood estimator, *Annals of Statistics*, **11,** 68-77.

(7) Hwang, J.S.(1995), On uniform consistency of Cox-Regression parameter, *Bulletin of the Institute for Basic Science, Inha University*, **16,** 85-88.

(8) Hwang, J.S.(1997), Uniform asymptotic equivalence between observed Fisher information number and expected Fisher information number, *Bulletin of the Institute for Basic Science, Inha University.*

(9) Lai, T.L., Siegmund, D.O.(1983), Fixed accuracy estimation of a autoregressive parameters, *Annals of Statistics*, **11**, 478-485.

(10) Rebolledo, R. (1980), Central limit theorem for local martingales, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandt Gebiete*, **51**, 269-286.

(11) Sellke and Siegmund, D.O. (1983), Sequential Analysis of the proportional hazards model, *Biometrika*, **70**, 315-326.