The Annals of Statistics
1990, Vol. 18, No. 3, 1378–1388

TIME-SEQUENTIAL POINT ESTIMATION THROUGH ESTIMATING EQUATIONS[1]

By I-Shou Chang and Chao A. Hsiung

National Central University and Academia Sinica

Time-sequential point estimation is studied in the model of fully parametric censored data and Cox's regression model. Both are investigated in the context of counting processes through estimating equations defined by martingales. The concept of information and a related inequality developed in estimating function theory by Godambe are adapted to these models. These suggest some optimality criteria for choosing stopping times as well as estimators. These lead naturally to some sequential procedures, which are shown to be asymptotically efficient with respect to the above criteria.

1. Introduction. Let $X_1, X_2 \dots$ be a sequence of i.i.d. random variables with distribution F. Let θ_0 denote a parameter or a functional of F. The classical sequential point estimation is to estimate θ_0 with loss measured by the squared error of estimation and a linear function of the number of observations. In case the estimator is decided, the problem is mainly to study the behavior of the estimator at certain relevant stopping times. With the pioneering work of Robbins (1959), there have been many interesting developments in various directions. [See, for example, Martinsek (1984), Ghosh, Nickerson and Sen (1987) and the references there for some of these developments.]

In the context of life-testing problems, Sen (1980) initiated the study of time-sequential point estimation for the mean θ_0 of an exponential distribution, which is to estimate θ_0 under the situation that the cost of recruitment of subjects into the study and of the follow-up time are to be considered together with the squared error loss of the estimation. Later, Gardiner, Susarla and Van Ryzin (1986) and Aras (1987) extended the work of Sen to the case allowing random withdrawals. Recently, Chang and Hsiung (1988b) studied the problem for a more general parametric model with censored data in the context of counting process theory.

In most of the above works, the estimators are chosen because they enjoy some nice properties such as sufficiency, minimaxity, robustness, etc. Once the estimator is chosen, it is the stopping time, suggested jointly by the estimator and the loss function, that attracts the most attention.

In this paper, we will study the time-sequential point estimation problem from a different perspective. Both the estimators and the relevant stopping times are suggested by the concept of information developed for estimating

Received March 1989; revised August 1989.

[1]Research partially supported by National Science Council of the Republic of China.

AMS 1980 *subject classification*. 62L12.

Key words and phrases. Time-sequential point estimation, Cox's regression, counting processes, estimating equations, martingales, information, stopping times, asymptotic efficiency.

function theory initiated by Godambe (1960, 1984). The estimator is chosen to maximize the information, and the stopping time is chosen to minimize a function $\tilde{R}$, defined to be a sum of the inverse of the information and the expected cost due to follow-up time and recruitment. We note that an analogous treatment in sequential point estimation appeared in Ferreira (1982), which discussed some sequential procedures, without giving their performances.

In our present time-sequential case, we find that this estimating equation approach is quite useful. We will illustrate it in two models of counting processes. The first is a fully parametric model with censored data. The second is Cox's regression model, in which the regression coefficient is to be estimated.

In both cases, estimation is done through estimating equations defined by martingales, which amounts to considering families of estimating functions unbiased at every stopping time. This takes into account progressive censoring schemes. The concept of information and related inequalities in Godambe (1960, 1984) are adapted to this situation, which gives an optimality criterion both for choosing the estimator and deciding the stopping time.

The following describes the major steps in the study for both models. First, the optimal estimator is determined according to certain information inequality. Next, an asymptotic representation for the function $\tilde{R}$ is obtained, which is a function of the unknown parameter and time. This representation suggests a stopping time, which hopefully may give small $\tilde{R}$ value. Finally, we study the $\tilde{R}$ value at this stopping time, establishing the asymptotic efficiency property.

This paper is organized as follows. Section 2 treats the fully parametric model with censored data. We would like to point out that the optimal estimator in this case is the maximum likelihood estimator and hence the problem was studied with classical loss function in Chang and Hsiung (1988b). While our conclusion here proposes the same sequential procedure as the one in Chang and Hsiung (1988b), we do not need many of those stringent conditions on intensity. Section 3 treats the more difficult Cox regression model, using the optimum property of maximum partial likelihood estimation studied in Chang and Hsiung (1990). Because this section parallels Section 2 in many ways, we will only give a condensed account.

2. Parametric case: Optimal estimating equation and asymptotically efficient stopping time.

2.1. *Preliminaries.* Let $(X_1, X_2, \dots)$ be a sequence of i.i.d. nonnegative random variables with distribution function $F(\cdot, \theta_0)$ in $\{F(\cdot, \theta) | \theta \in \Xi\}$, where Ξ is bounded and open in R^1. Let $(C_1, C_2, \dots)$ be a sequence of i.i.d. nonnegative censoring variables with unknown distribution. Assume that the X's and C's are independent and both have continuous densities. Let $\lambda(\cdot, \theta)$ be the intensity function of $F(\cdot, \theta)$. We assume that $\lambda(t, \theta)$ has continuous third derivative in θ, λ is bounded away from 0, and $\partial^2\lambda/\partial\theta^2$ is bounded on

$(0, T_0] \times \Xi$ for some $T_0 \in (0, \infty)$, and all of λ, $\partial\lambda/\partial\theta$ and $\partial^2\lambda/\partial\theta^2$ are continuous on $[0, T_0] \times \Xi$.

For every $n \geq 1$, $t > 0$, let

$$\mathscr{F}_t^{(n)} = \sigma\{[X_i \leq s], [C_i \leq s] | s \leq t, i = 1, 2, \dots, n\},$$
$$N_i(u) = 1_{[X_i, \infty)}(u \wedge C_i), \qquad H_i(u) = 1_{(0, X_i \wedge C_i]}(u).$$

Since we are in the statistical situation that, at time t, the observed variables are $X_i \wedge C_i \wedge t$ and $1_{[X_i \leq C_i \wedge t]}$ for the ith subject, the log-likelihood function of an experiment of size n at time t is

$$C_n(t, \theta) = \sum_{i=1}^{n} \int_0^t \log \lambda(u, \theta)\, dN_i(u) - \sum_{i=1}^{n} \int_0^t H_i(u) \lambda(u, \theta)\, du.$$

The score process and the information process are, respectively,

$$\begin{aligned} U_n(t, \theta) &= \frac{\partial}{\partial\theta} C_n(t, \theta) \\ &= \sum_{i=1}^{n} \int_0^t \frac{\frac{\partial\lambda}{\partial\theta}(u, \theta)}{\lambda(u, \theta)} (dN_i(u) - H_i(u)\lambda(u, \theta)\, du), \\ I_n(t, \theta) &= -\frac{\partial^2}{\partial\theta^2} C_n(t, \theta) \\ &= \int_0^t \tilde{d}(u, \theta)\, d\sum_{i=1}^{n} N_i(u) + \int_0^t \frac{\partial^2\lambda}{\partial\theta^2}(u, \theta) \left(\sum_{i=1}^{n} H_i(u) \right) du, \end{aligned}$$

where

$$\tilde{d}(u, \theta) = \frac{1}{\lambda^2(u, \theta)} \left(\left(\frac{\partial\lambda}{\partial\theta} \right)^2 (u, \theta) - \lambda(u, \theta) \frac{\partial^2\lambda}{\partial\theta^2}(u, \theta) \right).$$

2.2. *Optimal time-sequential estimating equation.* Since we allow the possibility of terminating an experiment at any stopping time, it is more appropriate to consider the estimating function G such that $E_\theta G(T, \theta) = 0$ for every stopping time T and $\theta \in \Xi$. We will formalize this and let $\mathscr{G}^{(n)}$ denote the set of all mean-0, square-integrable right-continuous $(\mathscr{F}_t^{(n)}, P_\theta)$-martingales $G(t, \theta)$ on $[0, T_0]$ for every $\theta \in \Xi$, such that

(a) $\partial G/\partial\theta$ exists and $0 < (E_\theta(\partial G(T, \theta)/\partial\theta))^2 < \infty$ for every stopping time $T \leq T_0$, $\theta \in \Xi$,

(b) some regularity conditions hold to ensure the interchanging of the order of differentiation and integration in the calculation leading to (2.1).

We note that using martingales as estimating functions appeared in Hutton and Nelson (1986), Thavaneswaran and Thompson (1986), Godambe and Heyde (1987) and Chang and Hsiung (1990) etc., after the germinating work of Godambe (1985).

To choose an estimating function from $\mathscr{G}^{(n)}$, we will reason at the guidance of Godambe (1960). Let $G \in \mathscr{G}^{(n)}$. Then, using the fact that $U_n(t,\theta)$ is a $(\mathscr{F}_t^{(n)}, P_\theta)$-martingale [Chang and Hsiung (1988a)] and the argument in Godambe (1960), we get

$$E_\theta\left(\frac{G(T,\theta)}{E_\theta\left(\dfrac{\partial G(T,\theta)}{\partial\theta}\right)}\right)^2 \geq \left(E_\theta U_n^2(T,\theta)\right)^{-1}, \tag{2.1}$$

for every stopping time $T \leq T_0$; and when G is replaced by U_n, (2.1) becomes equality. Godambe (1960) explained that a smaller value of the left of (2.1) is a desirable property of an estimating equation. This leads to the following definition and proposition.

DEFINITION 2.1. An estimating function $G^* \in \mathscr{G}^{(n)}$ is said to be optimum if G^* makes (2.1) an equality. The inverse of the right of (2.1) is called the information about θ at the stopping time T, denoted by $J_n(T,\theta)$.

PROPOSITION 2.2. *The score process $U_n(t,\theta)$ is optimum.*

Khan (1969) obtained results similar to Proposition 2.2 in the classical sequential case, without considering martingale estimating equations.

2.3. *Asymptotically efficient stopping time.* Based on the concept of information in Definition 2.1, we introduce a criterion of optimality for stopping times.

For every $\mathscr{F}_t^{(n)}$-stopping time T, we define

$$\tilde{R}_n(T) = \frac{a}{J_n(T,\theta_0)} + bn + cE_{\theta_0}\sum_{i=1}^{n}(X_i \wedge C_i \wedge T), \tag{2.2}$$

where a, b, c are positive numbers with b interpreted as the per unit cost of recruitment of subjects into the study and c the per unit cost of follow-up time or total time on test expended up to time T. It is assumed that $b = \rho c$ for some $\rho > 0$.

It seems desirable to minimize $\tilde{R}_n(T)$ with appropriate choice of T for a given size n and specified constants a, b, c. Since there is no explicit formula available in general, we shall content ourselves with an asymptotic representation for $\tilde{R}_n(T)$, which enables us to find a stopping time T giving small $\tilde{R}_n$ value.

We will study the asymptotics along a subsequence $n(c)$ such that $cn^2(c)$ converges to some $a^* > 0$ as c tends to 0. Sometimes we shall write n, instead of $n(c)$, to simplify notation.

THEOREM 2.1. *Let T_n be a $\mathscr{F}_t^{(n)}$-stopping time such that $\{T_n\}$ converges in probability to $t_0 \in (\delta, T_0]$, then*

$$\tilde{R}_n(T_n) \sim \frac{1}{n}\left[\frac{a}{\Sigma(t_0)} + \rho a^* + a^* \int_{(0, t_0]} H(u)\, du\right], \tag{2.3}$$

where $H(u) = P[X \wedge C > u]$, $\Sigma(t) = \int_{(0,t]} \phi(u)\, du$, $\phi(u) = H(u)(\partial\lambda/\partial\theta)^2(u, \theta_0)/\lambda(u, \theta_0)$, $\delta = \inf\{t | \Sigma(t) > 0\}$.

PROOF. Since $I_n(t, \theta_0) - U_n^2(t, \theta_0)$ is a martingale, we know that

$$J_n(T_n, \theta_0) = E_{\theta_0}(U_n(T_n, \theta_0))^2 = E_{\theta_0} I_n(T_n, \theta_0).$$

It follows from Theorem 3.3 of Chang and Hsiung (1988a) that $(1/n)I_n(T_n, \theta_0)$ converges to $\int_{(0, t_0]} \phi(u)$ in probability. Since the conditions on λ stated in Section 2.1 entail the uniform integrability of $(1/n)I_n(T_n, \theta_0)$, we know

$$\lim \frac{1}{n} E_{\theta_0} I_n(T_n, \theta_0) = \Sigma(t_0). \tag{2.4}$$

Besides,

$$\lim \frac{1}{n} E_{\theta_0} \sum_{i=1}^{n} (X_i \wedge C_i \wedge T_n) = \int_{(0, t_0]} H(u)\, du. \tag{2.5}$$

With (2.4) and (2.5), we get (2.3) immediately. □

Let

$$\tilde{g}(t) = \frac{a}{\Sigma(t)} + a^* \int_{(0, t]} H(u)\, du \quad \text{on } [\delta, T_0].$$

An ideal choice of $\{T_n\}$ should converge to t^*, a minimum value point of $\tilde{g}$. Since t^* satisfies

$$\left(\Sigma(t^*)\right)^2 = \frac{a}{a^*}\left(\frac{\partial\lambda}{\partial\theta}(t^*, \theta_0)\right)^2 \Big/ \lambda(t^*, \theta_0), \tag{2.6}$$

depending on the unknown parameter θ_0, we are led to consider the stopping time

$$T_c^* = \inf\left\{t > \delta \,\middle|\, \left(\frac{1}{n} I_n(t, \hat{\theta}_n(t))\right)^2 \geq \frac{a}{a^*}\left(\frac{\partial\lambda}{\partial\theta}(t, \hat{\theta}_n(t))\right)^2 \Big/ \lambda(t, \hat{\theta}_n(t))\right\} \wedge T_0, \tag{2.7}$$

where $\hat{\theta}_n(t)$ is the solution of the optimal estimating equation $U_n(t, \theta) = 0$, studied in Chang and Hsiung (1988a).

The asymptotic efficiency of T_c^* in the following theorem can be established by using the same arguments used in Section 4 of Chang and Hsiung (1988b). The proof is hence omitted.

THEOREM 2.2. *Assume $\tilde{g}$ has a unique minimum point $t^* \in (\delta, T_0)$ with $H(t^*) \neq 0$ and its derivative $\tilde{g}'$ is negative on (δ, t^*) and positive on $(t^*, t^* + \xi)$ for some $\xi > 0$. Then T_c^* converges to t^* in probability as c tends to 0. Consequently, $\tilde{R}_n(T_c^*)/\tilde{R}_n(t^*)$ converges to 1 as c tends to 0.*

Thus, the "asymptotically optimal" solution to this time-sequential point estimation problem is $(T_c^*, \hat{\theta}_n(T_c^*))$. Although this is the same as the one discussed in Chang and Hsiung (1988b), we require fewer conditions in this section. There are no conditions on $\tilde{d}$ in the information process I_n, for example.

3. Cox's regression model: Optimal estimating equation and asymptotically efficient stopping time. This section treats time-sequential point estimation of the relative risk β in Cox's model, with the baseline function λ as a nuisance parameter. According to Chang and Hsiung (1990), maximum partial likelihood estimation (MPLE) is the estimation through an optimal estimating equation. Therefore, the task that remains is to establish an asymptotic representation for the function $\tilde{R}$ and investigate the performance of stopping times suggested by this representation.

Although this section is technically more difficult than Section 2, the motivations and rationales behind various concepts and the major steps of the proofs are similar. We will hence omit some of the details.

3.1. *Preliminaries.* Let $(X_1, C_1, Z_1), (X_2, C_2, Z_2), \ldots$ be a sequence of i.i.d. random vectors with $X_i \geq 0$, $C_i \geq 0$. Assume that the conditional intensity of X_i given $Z_i = z$ is of the form

$$\lambda(t)e^{\beta z}, \tag{3.1}$$

where $\lambda(\cdot)$ is a baseline function such that $\lambda \in \Lambda = \{\tilde{\lambda} | \tilde{\lambda}$ is nonnegative continuous function on $[0, T_0]$ and $\int_0^{T_0} \tilde{\lambda}(t)\,dt < \infty\}$, where $T_0 \in (0, \infty)$, and $\beta \in \mathscr{B}$, a compact set in $\mathbb{R}^1$. Assume further that X_i and C_i are conditionally independent given Z_i, and Z_i's are bounded.

The statistical situation we have is to estimate β based on the data $\{(X_i \wedge C_i \wedge t, 1_{[X_i \leq C_i \wedge t]}, Z_i) | i = 1, 2, \ldots, n,\ t \leq T_n\}$ at some stopping time T_n, treating λ as a nuisance parameter, where T_n is a stopping time relative to

$$\mathscr{F}_t^{(n)} = \sigma\Big\{\big(Z_i, X_i \wedge C_i \wedge s, 1_{[X_i \leq C_i \wedge s]}\big) \Big| i = 1, 2, \ldots, n,\ s \leq t\Big\}.$$

Cox (1972, 1975) suggested that inference on β be based on the partial log-likelihood at t:

$$(3.2)\quad C_n(t,\beta) = \sum_{i=1}^{n}\int_0^t \beta Z_i\, dN_i(s) - \sum_{i=1}^{n}\int_0^t \log\left\{\sum_{j=1}^{n} H_j(s)e^{\beta Z_j}\right\} dN_i(s),$$

where $N_i(s) = 1_{[X_i,\infty)}(s \wedge C_i)$, $H_i(s) = 1_{(0, X_i \wedge C_i]}(s)$. In fact, maximum partial likelihood estimator of β at t is the solution of $U_n(t,\beta) = 0$, where $U_n(t,\beta) = (\partial/\partial\beta)C_n(t,\beta)$.

Let

$$S_n^{(l)}(t,\beta) = \frac{1}{n}\sum_{j=1}^{n} H_j(t)Z_j^l e^{\beta Z_j}, \qquad l = 0,1,2,$$

$$V_n(t,\beta) = \frac{S_n^{(2)}(t,\beta)}{S_n^{(0)}(t,\beta)} - \left(\frac{S_n^{(1)}(t,\beta)}{S_n^{(0)}(t,\beta)}\right)^2.$$

Then

$$(3.3)\qquad U_n(t,\beta) = \sum_{i=1}^{n}\int_0^t \left(Z_i - \left(\frac{S_n^{(1)}(s,\beta)}{S_n^{(0)}(s,\beta)}\right)\right) dN_i(s).$$

Let

$$(3.4)\qquad I_n(t,\beta) = -\frac{\partial^2}{\partial\beta^2}C_n(t,\beta) = \sum_{i=1}^{n}\int_0^t V_n(s,\beta)\, dN_i(s).$$

We note that U_n is a $(\mathscr{F}_t^{(n)}, P_{(\beta,\lambda)})$ square-integrable martingale with variation

$$\langle U_n(\cdot,\beta)\rangle_t = \sum_{i=1}^{n}\int_0^t V_n(s,\beta)H_i(s)\lambda(s)\, e^{\beta Z_i}\, ds$$

and $I_n(t,\beta) - \langle U_n(\cdot,\beta)\rangle_t$ is also a square-integrable martingale on $[0, T_0]$.

3.2. *Information and an optimum property of MPLE.* Let $\mathscr{G}^{(n)}$ be the set of all mean-zero, square-integrable right-continuous $(\mathscr{F}_t^{(n)}, P_{(\beta,\lambda)})$-martingales $G(t,\beta)$ on $[0, T_0]$ such that

(a) G is independent of the nuisance parameter λ,
(b) $\partial G/\partial\beta$ exists and $0 < (E_{(\beta,\lambda)}(\partial G(T,\beta)/\partial\beta))^2 < \infty$ for every $\mathscr{F}_t^{(n)}$-stopping time $T \le T_0$ and every $(\beta,\lambda) \in \mathscr{B} \times \Lambda$,
(c) some regularity conditions hold to ensure the interchanging of the order of differentiation and integration in the calculation leading to (3.5).

Let $\mathscr{V}$ be the set of square-integrable $(\mathscr{F}_t^{(n)}, P_{(\beta,\lambda)})$-martingales $V(t;\beta,\lambda)$ such that V is orthogonal to every element G in $\mathscr{G}^{(n)}$; or equivalently, $E_{(\beta,\lambda)}G(T,\beta)V(T,\beta,\lambda) = 0$ for every $\mathscr{F}_t^{(n)}$-stopping time T, and parameter (β,λ).

According to Chang and Hsiung (1990), for every stopping time $T \le T_0$,

$$E_{(\beta,\lambda)}\left(G(T,\beta)/E_{(\beta,\lambda)}\frac{\partial G(T,\beta)}{\partial\beta}\right)^2 \ge \left[\inf_{V\in\mathscr{V}} E_{(\beta,\lambda)}\left(\frac{\partial \log L(T,\beta,\lambda)}{\partial\beta} - V(T,\beta,\lambda)\right)^2\right]^{-1}, \tag{3.5}$$

where $L(T,\beta,\lambda)$ is the likelihood ratio for Cox's model, and that (3.5) becomes an equality when G is the U_n defined in (3.3), and

$$E_{(\beta,\lambda)}\left(U_n(T,\beta)/E_{(\beta,\lambda)}\frac{\partial U_n(T,\beta)}{\partial\beta}\right)^2 = \left(E_{(\beta,\lambda)}U_n^2(T,\beta)\right)^{-1}. \tag{3.6}$$

This is the optimum property of U_n and, consequently, we let

$$J_n(T,\beta) = E_{(\beta,\lambda)}U_n^2(T,\beta) \tag{3.7}$$

denote the information about β, eliminating λ, at stopping time T.

3.3. *Asymptotic representation for $\tilde{R}$.* Based on (3.7), we introduce a criterion of optimality for stopping times as follows. For every $\mathscr{F}_t^{(n)}$-stopping time T, we define

$$\tilde{R}_n(T) = \frac{a}{J_n(T,\beta)} + bn + cE_{(\beta,\lambda)}\sum_{i=1}^n (X_i \wedge C_i \wedge T). \tag{3.8}$$

The following theorem is the counterpart of Theorem 2.1. We shall assume $P\{X \wedge C \ge T_0\} > 0$. The conditions on a, b, c and the subsequence for the asymptotics to be valid are the same as in Theorem 2.1. The arguments for its proof are also similar, except it needs some asymptotic properties of $(1/n)I_n$ which can be found in Andersen and Gill (1982). We shall hence omit its proof.

THEOREM 3.1. *Let T_n be a $\mathscr{F}_t^{(n)}$-stopping time such that $\{T_n\}$ converges in probability to $t_0 \in (\delta, T_0]$, then*

$$\tilde{R}_n(T_n) \sim \frac{1}{n}\left[\frac{a}{\Sigma(t_0)} + \rho a^* + a^*\int_{(0,t_0]} H(u)\,du\right], \tag{3.9}$$

where $H(u) = P[X \wedge C > u]$, $\Sigma(t) = \int_0^t v(u,\beta)s^{(0)}(u,\beta)\lambda(u)\,du$,

$$s^{(l)}(u,\beta) = E_{(\beta,\lambda)}\{H_i(u)Z_i^l e^{\beta Z_i}\}, \qquad l = 0,1,2,$$

$$v(u,\beta) = \frac{s^{(2)}}{s^{(0)}} - \left(\frac{s^{(1)}}{s^{(0)}}\right)^2, \qquad \delta = \inf\{t > 0 | \Sigma(t) > 0\}.$$

3.4. *Asymptotically efficient stopping time.* Let

$$\tilde{g}(t) = \frac{a}{\Sigma(t)} + a^*\int_{(0,t]} H(u)\,du \quad \text{on } [\delta, T_0].$$

An ideal choice of $\{T_n\}$ should converge to t^*, a minimum of $\tilde{g}$. Since t^* satisfies

$$(\Sigma(t^*))^2 = \frac{a}{a^*}\frac{v(t^*,\beta)s^{(0)}(t^*,\beta)\lambda(t^*)}{H(t^*)}, \tag{3.10}$$

we consider the stopping time

$$T_c^* = \inf\left\{t > \delta \Big| \left(\frac{1}{n}I_n(t,\hat{\beta}_n(t))\right)^2 \geq \frac{a}{a^*}\frac{V_n(t,\hat{\beta}_n(t))S_n^{(0)}(t,\hat{\beta}_n(t))\hat{\lambda}_n(t)}{\hat{H}_n(t)}\right\} \wedge T_0 \tag{3.11}$$

where

$$\hat{H}_n(t) = \frac{1}{n}\sum_{i=1}^{n} H_i(t),$$

$\hat{\beta}_n(t)$ is the solution of the optimal estimating equation $U_n(t,\beta) = 0$ and $\hat{\lambda}_n$ is a uniformly consistent estimator of the baseline intensity $\lambda(t)$, to be given below.

Let

$$\hat{\Lambda}_n(s) = \int_0^s \left\{\sum_{i=1}^{n} H_i(u)e^{\hat{\beta}_n(t)Z_i}\right\}^{-1} d\left(\sum_{i=1}^{n} N_i(u)\right), \qquad s \in [0, T_0],$$

be the Aalen–Nelsen type estimator of $\int_0^s \lambda(u)\,du$ at time t. Then the intensity $\lambda(s)$ can be estimated by smoothing $\hat{\Lambda}_n(s)$ with a kernel function [see Ramlau-Hansen (1983)],

$$\hat{\lambda}_n(s) = \frac{1}{\tilde{b}_n}\int_0^{T_0} K\left(\frac{s-u}{\tilde{b}_n}\right) d\hat{\Lambda}_n(u), \qquad s \in \left[T_0\tilde{b}_n, T_0(1-\tilde{b}_n)\right],$$

where K is a suitable kernel function, which is of bounded variation, vanishes outside $[-T_0, T_0]$ and has integral 1. The window $\tilde{b}_n > 0$ goes to 0 and $n\tilde{b}_n^2$ goes to ∞.

The following theorem gives the asymptotic efficiency of T_c^*.

THEOREM 3.2. *Assume that $\tilde{g}$ has a unique minimum point t^* in (δ, T_0) and its derivative $\tilde{g}'$ is negative on (δ, t^*) and positive on $(t^*, t^* + \xi)$ for some $\xi > 0$. Then T_c^* converges to t^* in probability as c tends to 0. Consequently, $\tilde{R}_n(T_c^*)/\tilde{R}_n(t^*)$ converges to 1 as c tends to 0.*

The proof of Theorem 3.2 is omitted because it is similar to that for Theorem 2.2. The only difference is that the following Lemma 3.1 is needed in various places of its proof. Lemma 3.1 strengthens some results in Andersen and Gill (1982) and Ramlau-Hansen (1983). Its proof is also omitted because it is standard and tedious.

LEMMA 3.1.

(i)
$$\sup_{s \in [\delta, T_0]} \left|\hat{\beta}_n(s) - \beta\right| \xrightarrow[n \to \infty]{} 0 \quad \textit{in probability},$$

(ii)
$$\sup_{s \in [\delta, T_0]} \left|\frac{1}{n} I_n(s, \hat{\beta}_n(s)) - \Sigma(s)\right| \xrightarrow[n \to \infty]{} 0 \quad \textit{in probability},$$

(iii)
$$\sup_{s \in [\delta, T_0]} \left|S_n^{(l)}(s, \hat{\beta}_n(s)) - s^{(l)}(s, \beta)\right| \xrightarrow[n \to \infty]{} 0 \quad \textit{in probability},$$

for $l = 0, 1, 2.$

(iv)
$$\sup_{s \in [\delta, T_0]} \left|V_n(s, \hat{\beta}_n(s)) - v(s, \beta)\right| \xrightarrow[n \to \infty]{} 0 \quad \textit{in probability},$$

(v)
$$\sup_{s \in [t_1, t_2]} \left|\hat{\lambda}_n(s) - \lambda(s)\right| \xrightarrow[n \to \infty]{} 0 \quad \textit{in probability},$$

for any $0 < t_1 < t_2 < T_0$.

Acknowledgments. Thanks are due to Dr. M. T. Chao for some helpful conversations. We are also grateful to the associate editor and referee for comments that led to a better presentation.

REFERENCES

ANDERSEN, P. K. and GILL, R. D. (1982). Cox's regression model for counting process: A large sample study. *Ann. Statist.* **10** 1100–1120.

ARAS, G. (1987). Sequential estimation of the mean exponential survival time under random censoring. *J. Statist. Plann. Inference* **16** 147–158.

CHANG, I. S. and HSIUNG, C. A. (1988a). Likelihood process in parametric model of censored data with staggered entry-asymptotic properties and applications. *J. Multivariate Anal.* **24** 31–45.

CHANG, I. S. and HSIUNG, C. A. (1988b). Counting process approach to time-sequential and sequential point estimation with censored data. *Sequential Anal.* **7**(2) 127–148.

CHANG, I. S. and HSIUNG, C. A. (1990). Finite sample optimality of maximum partial likelihood estimation in Cox's model for counting processes. *J. Statist. Plann. Inference.* **25**.

COX, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

COX, D. R. (1975). Partial likelihood. *Biometrika* **62** 269–276.

FERREIRA, P. E. (1982). Sequential estimation through estimating equations in the nuisance parameter case. *Ann. Statist.* **10** 167–173.

GARDINER, J., SUSARLA, V. and VAN RYZIN, J. (1986). Time sequential estimation of the exponential mean under random withdrawals. *Ann. Statist.* **14** 607–618.

GHOSH, M., NICKERSON, D. M. and SEN, P. K. (1987). Sequential shrinkage estimation. *Ann. Statist.* **15** 817–829.

GODAMBE, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31** 1208–1212.

GODAMBE, V. P. (1984). On ancillarity and Fisher information in the presence of a nuisance parameter *Biometrika* **71** 626–629.

GODAMBE, V. P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika* **72** 419–428.

GODAMBE, V. P. and HEYDE, C. C. (1987). Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.* **55** 231–244.

HUTTON, J. E. and NELSON, P. I. (1986). Quasi-likelihood estimation for semimartingales. *Stochastic Process. Appl.* **22** 245–257.

KHAN, R. (1969). Maximum likelihood estimation in sequential experiments. *Sankhyā Ser. A* **31** 49–56.

MARTINSEK, A. T. (1984). Sequential determination of estimator as well as sample size. *Ann. Statist.* **12** 533–550.

RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466.

ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. In *Probability and Statistics: The Harald Cramér Volume* (U. Grenander, ed.) 235–245. Almquist and Wiksell, Uppsala.

SEN, P. K. (1980). On time-sequential point estimation of the mean of an exponential distribution. *Comm. Statist. Theory Methods* **9** 27–38.

THAVANESWARAN, A. and THOMPSON, M. E. (1986). Optimal estimation for semimartingales. *J. Appl. Probab.* **23** 409–417.

DEPARTMENT OF MATHEMATICS
NATIONAL CENTRAL UNIVERSITY
CHUNG-LI, TAIWAN
REPUBLIC OF CHINA

INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
TAIPEI, TAIWAN
REPUBLIC OF CHINA