The Annals of Statistics
1982, Vol. 10, No. 4, 1100–1120

COX'S REGRESSION MODEL FOR COUNTING PROCESSES: A LARGE SAMPLE STUDY

BY P. K. ANDERSEN AND R. D. GILL

Statistical Research Unit, Copenhagen; and Mathematical Centre, Amsterdam

The Cox regression model for censored survival data specifies that covariates have a proportional effect on the hazard function of the life-time distribution of an individual. In this paper we discuss how this model can be extended to a model where covariate processes have a proportional effect on the intensity process of a multivariate counting process. This permits a statistical regression analysis of the intensity of a recurrent event allowing for complicated censoring patterns and time dependent covariates. Furthermore, this formulation gives rise to proofs with very simple structure using martingale techniques for the asymptotic properties of the estimators from such a model. Finally an example of a statistical analysis is included.

1. Introduction. The Cox-model for censored survival data (Cox, 1972) specifies the hazard rate or intensity of failure $\lambda(t) = \lim_{h \downarrow 0} \mathscr{P}[T \leq t + h \mid T > t]$ for the survival time T of an individual with covariate vector z which may depend on the time t to have the form

$$\lambda(t; z) = \lambda_0(t)\exp\{\beta_0' z(t)\}, \quad t \geq 0. \tag{1.1}$$

Here β_0 is a p-vector of unknown regression coefficients and $\lambda_0(t)$, the underlying hazard, is an unknown and unspecified nonnegative function. The statistical problem is the one of estimating β_0 and the function λ_0 on the basis of, say, n possibly right censored survival times $T_1, \cdots, T_n$ and the corresponding covariate vectors $z_1, \cdots, z_n$, where z_i is observed on $[0, T_i]$.

Cox (1972) suggested that inference on β_0 be based on the function

$$L(\beta) = \prod_{i=1}^{n} \left\{ \frac{e^{\beta' z_i(T_i)}}{\sum_{j \in \mathscr{R}_i} e^{\beta' z_j(T_i)}} \right\}^{\delta_i} \tag{1.2}$$

where $\mathscr{R}_i = \{j : T_j \geq T_i\}$ and $1 - \delta_i$ is an indicator for censoring. In a later paper (Cox, 1975) he derived (1.2) as a partial likelihood function. Letting $\hat{\beta}$ be the value that maximizes (1.2), then the continuous estimator obtained by linear interpolation between failure times of

$$\hat{\Lambda}(t) = \sum_{T_i \leq t} \frac{\delta_i}{\sum_{j \in \mathscr{R}_i} e^{\hat{\beta}' z_j(T_i)}} \tag{1.3}$$

for the underlying cumulative hazard $\Lambda_0(t) = \int_0^t \lambda_0(s)\, ds$ was suggested by Breslow (1972, 1974). In a recent paper (Johansen, 1983) it was demonstrated that L given by (1.2) is a likelihood profile in the sense that $L(\beta) = \max_\Lambda L(\beta, \Lambda)$ where $L(\beta, \Lambda)$ is a joint likelihood for the unknown parameters β_0 and Λ_0. Also the value of Λ that maximizes $L(\hat{\beta}, \Lambda)$ is exactly $\hat{\Lambda}$ given by (1.3). This joint likelihood was derived by extending the Cox-model (1.1) to a model allowing for multiple jumps and furthermore allowing (1.1) to be the intensity of a recurrent event.

In this paper we consider the large sample properties of a counting process model with

Received September 1981; revised May 1982.

AMS 1980 *subject classifications.* 62F12, 62G05, 62M99, 62P10.

Key words and phrases. Censoring; intensity; martingale; survival analysis; time dependent covariates.

intensity given by (1.1), i.e. we include the possibility of the event considered being recurrent, but not more than one event may happen at a given time.

We have had several motivations for undertaking this study. First of all, much effort has been spent on deriving the asymptotic properties of the estimators $\hat{\beta}$ and $\hat{\Lambda}$ in the Cox-model, see Cox (1975), Liu and Crowley (1978), Tsiatis (1978a, 1978b, 1981a), Link (1979), Bailey (1979) and Næs (1982). Asymptotic distribution theory for the score function test statistic based on Cox's likelihood is given by Tsiatis (1981b) and Sen (1981). We found that the martingale theory which emerges very naturally from the counting process formulation of (1.1) and which is also the starting point of Næs (1982) could be used very efficiently to give proofs whose basic ideas are very simple. Næs (1982) and Sen (1981) each use *discrete time* martingale theory which to our minds does not fit so naturally in this set-up. Secondly we found that the assumption of the covariates being bounded made by all of the above mentioned authors except Tsiatis (1981a) and Sen (1981) (who on the other hand only considered time-independent covariates) is too restrictive and should be avoided. Finally in a practical example (see Andersen and Rasmussen (1982)) concerning admissions to psychiatric hospitals for women giving birth, the results were needed for the more general case of describing the effect of covariate measurements on the intensity of a recurrent phenomenon. This example will be discussed in more detail below.

We conclude that it is useful to formulate the Cox-model in the more general set-up of multivariate counting processes of Aalen (1978). This will be done in Section 2 where we also outline the basic ideas of the proofs of asymptotic normality and consistency of $\hat{\beta}$ and of weak convergence of $\hat{\Lambda}(\cdot) - \Lambda_0(\cdot)$. In Section 3 we give the basic assumptions and the technical details for proving the results rigorously and in Section 4 we consider the problem of actually verifying these conditions in the special case where the counting processes and the covariate vectors are assumed to be independent and identically distributed. In that final section we also return to the practical example.

Regression models for counting processes were also considered by Aalen (1980). He parameterized the intensity process itself linearly rather than the logarithm as we do, and thus the standard martingale central limit theory applied rather immediately when deriving the asymptotic properties. On the other hand, this approach does not guarantee the estimator of the intensity to be non-negative and hence some posterior smoothing of the estimate has to be performed before applying the results from an analysis.

Methods for *checking* the assumptions of the Cox-model are given by Andersen (1982).

2. Counting process formulation of the Cox-model and its properties. In this section we shall formulate the model (1.1) in the framework of multivariate counting processes and sketch how proofs of the asymptotic properties of $\hat{\beta}$ and $\hat{\Lambda}$ may be carried out. For simplicity we shall be working on the time interval [0, 1] and we refer to Section 4 for a discussion on how to extend the results to processes on $[0, \infty)$.

We shall use basic results from the theory of multivariate counting processes, stochastic integrals and local martingales without further comment. A survey of this theory intended for similar applications to ours can be found in Gill (1980). The survey in Aalen (1978) is also very useful, though it does not include the concept of *local* martingales. Using this concept allows us to avoid making superfluous integrability conditions. Apart from this background theory, our basic tools are the inequality of Lenglart (1977) and the martingale central limit theorems of Rebolledo (1978, 1980) which we state in specialized forms adapted to our needs in Appendix I.

2.1. *Formulation of the model.* Since we are interested in asymptotic properties, we shall in fact consider a sequence of models, indexed by $n = 1, 2, \cdots$. Also, as mentioned in Section 1, we shall generalize from the possibly censored observation of the lifetimes of n individuals to the observation (in the nth model) of an n-component multivariate counting process $N^{(n)} = (N_1^{(n)}, \cdots, N_n^{(n)})$, where $N_i^{(n)}$ counts observed events in the life of the ith individual, $i = 1, \cdots, n$, over the time interval [0, 1]. So the sample paths of $N_1^{(n)}, \cdots, N_n^{(n)}$

are step functions, zero at time zero, with jumps of size +1 only, no two component processes jumping at the same time. Unlike Aalen (1978) we need not assume that $\mathscr{E}N_i^{(n)}(1) < \infty$, only that $N_i^{(n)}(1)$ is almost surely finite. For discussion of how the usual special models for censoring can be treated in this set-up, see Aalen (1978), Gill (1980) or Andersen et al. (1982).

In our model, properties of stochastic processes, such as being a local martingale or a predictable process, are relative to a right-continuous nondecreasing family $(\mathscr{F}_t^{(n)} : t \in [0, 1])$ of sub σ-algebras on the nth sample space $(\Omega^{(n)}, \mathscr{F}^{(n)}, \mathscr{P}^{(n)})$; $\mathscr{F}_t^{(n)}$ represents everything that happens up to time t (in the nth model).

Our basic assumption is that for each n, $N^{(n)}$ has random intensity process $\lambda^{(n)} = (\lambda_1^{(n)}, \cdots, \lambda_n^{(n)})$ such that

$$\lambda_i^{(n)}(t) = Y_i^{(n)}(t)\lambda_0(t)\exp\{\beta_0' Z_i^{(n)}(t)\}. \tag{2.1}$$

Here β_0 is a fixed column vector of p coefficients, λ_0 a fixed underlying hazard function, and $Y_i^{(n)}$ is a predictable process taking values in $\{0, 1\}$ indicating (by the value 1) when the ith individual is under observation (so in particular, $N_i^{(n)}$ only jumps when $Y_i^{(n)} = 1$). Finally $Z_i^{(n)} = (Z_{i1}^{(n)}, \cdots, Z_{ip}^{(n)})'$ is a column vector of p covariate processes for the ith individual. We suppose that $Z_i^{(n)}$ is predictable and locally bounded (which is the case for instance if $Z_i^{(n)}$ is left continuous with right-hand limits and adapted).

By stating that $N^{(n)}$ has intensity process $\lambda^{(n)}$ we mean that the processes $M_i^{(n)}$ defined by

$$M_i^{(n)}(t) = N_i^{(n)}(t) - \int_0^t \lambda_i^{(n)}(u)\, du, \quad i = 1, \cdots, n, \quad t \in [0, 1] \tag{2.2}$$

are local martingales on the time interval [0, 1]. As a consequence, they are in fact local square integrable martingales, with

$$\langle M_i^{(n)}, M_i^{(n)} \rangle(t) = \int_0^t \lambda_i^{(n)}(u)\, du \quad \text{and} \quad \langle M_i^{(n)}, M_j^{(n)} \rangle = 0, \quad i \neq j, \tag{2.3}$$

i.e. $M_i^{(n)}$ and $M_j^{(n)}$ are orthogonal when $i \neq j$. Under various regularity conditions which do not concern us here, these facts are equivalent to the following generalization of (1.1):

$$\lim_{h \downarrow 0} \frac{1}{h} \mathscr{P}[N_i^{(n)}(t+h) - N_i^{(n)}(t) = 1 \mid \mathscr{F}_t^{(n)}] = \lambda_i^{(n)}(t+). \tag{2.4}$$

See e.g. Dolivo (1974, Theorem 2.5.1), Aalen (1978, Section 3.2) and Gill (1980, Section 2.3). For an application to censored survival data see Gill (1980, Theorem 3.1.1.). One could start with (2.4) plus regularity conditions as the basic model; however we prefer to take the local martingale property of $M^{(n)}$ in (2.2) as primary, and only mention the "intensity" property (2.4) as a motivation for this more abstract looking model.

In the following we shall everywhere drop the superscript (n). Only β_0 and λ_0 are the same in all models (i.e. for each n). Convergence in probability $(\to_{\mathscr{P}})$ and convergence in distribution $(\to_{\mathscr{D}})$ are always relative to the probability measures $\mathscr{P}^{(n)}$ parameterized by β_0 and λ_0.

2.2. *Asymptotic normality of $\hat{\beta}$.* As demonstrated by Johansen (1983), Cox's likelihood (1.2) is a reasonable basis for the estimation of the regression parameter vector β_0 in our more general set-up too. Let $C(\beta, t)$ be the logarithm of the Cox likelihood evaluated at time t, so that according to (1.2) and (2.1) we have

$$C(\beta, t) = \sum_{i=1}^n \int_0^t \beta' Z_i(s)\, dN_i(s) - \int_0^t \log\{\textstyle\sum_{i=1}^n Y_i(s) e^{\beta' Z_i(s)}\}\, d\bar{N}(s),$$

where $\bar{N} = \sum_{i=1}^n N_i$. Then we have that $C(\beta, 1) = \log L(\beta)$, and the estimator $\hat{\beta}$ is defined

as the solution to the likelihood equation $(\partial/\partial\beta)C(\beta, 1) = 0$, where the vector of derivatives $U(\beta, t)$ of $C(\beta, t)$ w.r.t. β has the form

$$U(\beta, t) = \sum_{i=1}^n \int_0^t Z_i(s)\, dN_i(s) - \int_0^t \frac{\sum_{i=1}^n Y_i(s)Z_i(s)e^{\beta' Z_i(s)}}{\sum_{i=1}^n Y_i(s)e^{\beta' Z_i(s)}}\, d\bar{N}(s).$$

From (2.2) it is immediately seen that

$$U(\beta_0, t) = \sum_{i=1}^n \int_0^t Z_i(s)\, dM_i(s) - \int_0^t \frac{\sum_{i=1}^n Y_i(s)Z_i(s)e^{\beta_0' Z_i(s)}}{\sum_{i=1}^n Y_i(s)e^{\beta_0' Z_i(s)}}\, d\bar{M}(s),$$

where $\bar{M} = \sum_{i=1}^n M_i$ is a local martingale. Taylor expanding $U(\beta, 1)$ around β_0, we get

$$U(\beta, 1) - U(\beta_0, 1) = -\mathscr{I}(\beta^*, 1)(\beta - \beta_0), \tag{2.5}$$

where β^* is on the line segment between β and β_0 and the positive semidefinite matrix

$$\mathscr{I}(\beta, t) = \int_0^t \left(\frac{\sum_{i=1}^n Y_i(s)Z_i(s)^{\otimes 2}e^{\beta' Z_i(s)}}{\sum_{i=1}^n Y_i(s)e^{\beta' Z_i(s)}} - \left(\frac{\sum_{i=1}^n Y_i(s)Z_i(s)e^{\beta' Z_i(s)}}{\sum_{i=1}^n Y_i(s)e^{\beta' Z_i(s)}} \right)^{\otimes 2} \right) d\bar{N}(s)$$

is minus the second derivative of $C(\beta, t)$ w.r.t. β. (For a column vector a we denote by $a^{\otimes 2}$ the matrix aa', cf. Section 3.) Inserting $\hat{\beta}$ in (2.5) we get

$$n^{-1/2}U(\beta_0, 1) = \{n^{-1}\mathscr{I}(\beta^*, 1)\}n^{1/2}(\hat{\beta} - \beta_0), \tag{2.6}$$

since by definition $U(\hat{\beta}, 1) = 0$.

To prove asymptotic normality of $n^{1/2}(\hat{\beta} - \beta_0)$ it now suffices to prove weak convergence to a Gaussian process of the local martingale $n^{-1/2}U(\beta_0, \cdot)$ and to prove convergence in probability to a non-singular matrix of $n^{-1}\mathscr{I}(\beta^*, 1)$. For the weak convergence we utilize the central limit theorems for local martingales given by Rebolledo (1978, 1980). As to the convergence in probability of $n^{-1}\mathscr{I}(\beta^*, 1)$, it suffices to prove that $\hat{\beta}$ is consistent and that $n^{-1}\mathscr{I}(\beta^*, 1)$ converges in probability for any β^* such that $\beta^* \to_{\mathscr{P}} \beta_0$. The fact that $U(\beta_0, \cdot)$ is a local martingale implies that the discrete time process obtained from it by only registering its values at jump times of $\bar{N}$ is also a local martingale (and under integrability conditions, a martingale). This fact is derived by Sen (1981) in an i.i.d. set-up by a permutation approach and used to obtain weak convergence results.

2.3. *Consistency of* $\hat{\beta}$. To prove consistency of $\hat{\beta}$ consider the process

$$\begin{aligned} X(\beta, t) &= n^{-1}(C(\beta, t) - C(\beta_0, t)) \\ &= n^{-1}\left[\sum_{i=1}^n \int_0^t (\beta - \beta_0)' Z_i(s)\, dN_i(s) - \int_0^t \log\left\{ \frac{\sum_{i=1}^n Y_i(s)e^{\beta' Z_i(s)}}{\sum_{i=1}^n Y_i(s)e^{\beta_0' Z_i(s)}} \right\} d\bar{N}(s) \right]. \end{aligned} \tag{2.7}$$

Then $X(\beta, 1)$ is a concave function with a (with probability tending to 1) unique maximum at $\beta = \hat{\beta}$ by definition of $\hat{\beta}$. Using the inequality of Lenglart (1977) (see also Appendix I) it can be proved that $X(\beta, 1)$ converges in probability to a function of β which is concave with a unique maximum at β_0. A fairly simple argument using convex function theory then shows that $\hat{\beta} \to_{\mathscr{P}} \beta_0$.

2.4. *Asymptotic distribution of* $\hat{\Lambda}$. Formulated by means of counting processes, the estimator $\hat{\Lambda}$ given by (1.3) has the form

$$\hat{\Lambda}(t) = \int_0^t \frac{d\bar{N}(s)}{\sum_{i=1}^n Y_i(s)e^{\hat{\beta}' Z_i(s)}},$$

and hence

$$n^{1/2}\{\hat{\Lambda}(t) - \Lambda_0(t)\} = n^{1/2}\int_0^t \left\{\frac{1}{\sum_{i=1}^n Y_i(s)e^{\hat{\beta}'Z_i(s)}} - \frac{1}{\sum_{i=1}^n Y_i(s)e^{\beta_0'Z_i(s)}}\right\} d\bar{N}(s)$$

$$+ n^{1/2}\left\{\int_0^t \frac{d\bar{N}(s)}{\sum_{i=1}^n Y_i(s)e^{\beta_0'Z_i(s)}} - \Lambda_0^*(t)\right\} \tag{2.8}$$

$$+ n^{1/2}\{\Lambda_0^*(t) - \Lambda_0(t)\},$$

where

$$\Lambda_0^*(t) = \int_0^t \lambda_0(s)I\{\textstyle\sum_{i=1}^n Y_i(s) > 0\}ds.$$

Here the third term is asymptotically negligible; the second term is a local martingale, namely

$$W(t) = \int_0^t \frac{n^{1/2}d\bar{M}(s)}{\sum_{i=1}^n Y_i(s)e^{\beta_0'Z_i(s)}};$$

and a Taylor expansion of the first term yields the quantity $H(\beta^*, t)'\, n^{1/2}(\hat{\beta} - \beta_0)$, where the vector H is given by

$$H(\beta, t) = -\int_0^t \frac{\sum_{i=1}^n Y_i(s)Z_i(s)e^{\beta'Z_i(s)}}{\{\sum_{i=1}^n Y_i(s)e^{\beta'Z_i(s)}\}^2}\, d\bar{N}(s), \tag{2.9}$$

and β^* is on the line segment between $\hat{\beta}$ and β_0. The asymptotic distribution of $n^{1/2}\{\hat{\Lambda}(\cdot) - \Lambda_0(\cdot)\}$ now follows by finding the asymptotic distribution of the local martingale $W(t)$ and by proving convergence in probability of $H(\beta^*, t)$ for any β^* such that $\beta^* \to_p \beta_0$, and by finally noting that $W(\cdot)$ is orthogonal to $U(\beta_0, \cdot)$:

$$\left\langle U(\beta_0, \cdot), \int \frac{d\bar{M}}{\sum_{i=1}^n Y_i e^{\beta_0'Z_i}} \right\rangle = 0.$$

From this last fact it follows that $W(t)$ and $H(\beta^*, t)'\, n^{1/2}(\hat{\beta} - \beta_0)$ are asymptotically independent since $\hat{\beta}$ is a function of $U(\beta_0, 1)$ (cf. (2.6)). Hence the desired asymptotic distribution can be derived using 2.2.

3. Asymptotic properties. In this section the notation is the same as in Section 2. In particular, this means we are dropping a superfix (n) almost everywhere; only β_0 and λ_0 are fixed (i.e. independent of n). Unless otherwise stated, all limits are taken as $n \to \infty$. Suppose $a = (a_1, \cdots, a_p)'$ and $b = (b_1, \cdots, b_p)'$ are p-vectors, then we write $a \otimes b$ for the $p \times p$ matrix ab' with (i, j)th element $a_i b_j$. Also we write $a^{\otimes 2}$ for the matrix $a \otimes a$. For a matrix A or vector a, $\|A\| = \sup_{i,j} |a_{ij}|$ and $\|a\| = \sup_i |a_i|$. For a vector a, $|a| = (\sum a_i^2)^{1/2} = (a'a)^{1/2}$.

Some further important definitions are:

$$S^{(0)}(\beta, t) = \frac{1}{n}\textstyle\sum_{\ell=1}^n Y_\ell(t)e^{\beta'Z_\ell(t)}$$

$$S^{(1)}(\beta, t) = \frac{1}{n}\textstyle\sum_{\ell=1}^n Z_\ell(t)Y_\ell(t)e^{\beta'Z_\ell(t)},$$

$$S^{(2)}(\beta, t) = \frac{1}{n}\textstyle\sum_{\ell=1}^n Z_\ell(t)^{\otimes 2}Y_\ell(t)e^{\beta'Z_\ell(t)},$$

$$E(\beta, t) = \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)},$$

and

$$V(\beta, t) = \frac{S^{(2)}(\beta, t)}{S^{(0)}(\beta, t)} - E(\beta, t)^{\otimes 2}.$$

Note that $S^{(0)}$ is a scalar, $S^{(1)}$ and E are p-vectors and $S^{(2)}$ and V are $p \times p$ matrices.

These quantities can be interpreted as follows. Suppose at time t, we select an individual i out of those individuals under observation (i.e. with $Y_i(t) = 1$) with probabilities proportional to $e^{\beta' Z_i(t)}$. Then $E(\beta, t)$ and $V(\beta, t)$ are the expectation and variance respectively of the covariate vector $Z_i(t)$ of the individual selected. $S^{(0)}$, $S^{(1)}$ and $S^{(2)}$ are roughly to be interpreted as a norming factor, a sum and a sum of squares respectively.

The following list of conditions will be assumed to hold throughout this section. There are a number of redundancies in them, and not all conditions are needed for every result, but in this way we hope to avoid too many technical distractions in the theorems and their proofs. Further discussion of the conditions is deferred till Section 4.

CONDITIONS.

A. (Finite interval). $\int_0^1 \lambda_0(t)\, dt < \infty$.

B. (Asymptotic stability). There exists a neighbourhood $\mathscr{B}$ of β_0 and scalar, vector and matrix functions $s^{(0)}$, $s^{(1)}$ and $s^{(2)}$ defined on $\mathscr{B} \times [0, 1]$ such that for $j = 0, 1, 2$

$$\sup_{t\in[0,1],\beta\in\mathscr{B}} \| S^{(j)}(\beta, t) - s^{(j)}(\beta, t) \| \to_{\mathscr{P}} 0.$$

C. (Lindeberg condition). There exists $\delta > 0$ such that

$$n^{-1/2} \sup_{i,t} | Z_i(t) |\, Y_i(t) I\{\beta_0' Z_i(t) > -\delta\, | Z_i(t) |\} \to_{\mathscr{P}} 0.$$

D. (Asymptotic regularity conditions). Let $\mathscr{B}$, $s^{(0)}$, $s^{(1)}$ and $s^{(2)}$ be as in Condition B and define $e = s^{(1)}/s^{(0)}$ and $v = s^{(2)}/s^{(0)} - e^{\otimes 2}$. For all $\beta \in \mathscr{B}$, $t \in [0, 1]$:

$$s^{(1)}(\beta, t) = \frac{\partial}{\partial\beta} s^{(0)}(\beta, t), \quad s^{(2)}(\beta, t) = \frac{\partial^2}{\partial\beta^2} s^{(0)}(\beta, t),$$

$s^{(0)}(\cdot, t)$, $s^{(1)}(\cdot, t)$ and $s^{(2)}(\cdot, t)$ are continuous functions of $\beta \in \mathscr{B}$, uniformly in $t \in [0, 1]$, $s^{(0)}$, $s^{(1)}$ and $s^{(2)}$ are bounded on $\mathscr{B} \times [0, 1]$; $s^{(0)}$ is bounded away from zero on $\mathscr{B} \times [0, 1]$, and the matrix

$$\Sigma = \int_0^1 v(\beta_0, t) s^{(0)}(\beta_0, t) \lambda_0(t)\, dt$$

is positive definite.

Note that the partial derivative conditions on $s^{(0)}$, $s^{(1)}$ and $s^{(2)}$ are satisfied by $S^{(0)}$, $S^{(1)}$ and $S^{(2)}$; and that Σ is automatically positive semidefinite. Furthermore the interval [0, 1] in the conditions may everywhere be replaced by the set $\{t : \lambda_0(t) > 0\}$.

LEMMA 3.1. (*Consistency of* $\hat{\beta}$). $\hat{\beta} \to_{\mathscr{P}} \beta_0$.

PROOF. (See Section 2.3.) Consider the processes $X(\beta, \cdot)$ given by (2.7) and

$$A(\beta, t) = n^{-1}\left[\sum_{i=1}^n \int_0^t (\beta - \beta_0)' Z_i(u) \lambda_i(u) du - \int_0^t \log\left\{\frac{S^{(0)}(\beta, u)}{S^{(0)}(\beta_0, u)}\right\} \bar{\lambda}(u)\, du\right]$$

$$= \int_0^t \left[(\beta - \beta_0)' S^{(1)}(\beta_0, u) - \log\left\{\frac{S^{(0)}(\beta, u)}{S^{(0)}(\beta_0, u)}\right\} S^{(0)}(\beta_0, u)\right] \lambda_0(u)\, du,$$

where $\bar{\lambda} = \sum_{i=1}^n \lambda_i$.

Then for each β, $X(\beta, \cdot) - A(\beta, \cdot)$ is a local square integrable martingale with

$$\langle X(\beta, \cdot) - A(\beta, \cdot), X(\beta, \cdot) - A(\beta, \cdot)\rangle = B(\beta, \cdot),$$

say, where

$$B(\beta, t) = n^{-2} \sum_{i=1}^n \int_0^t \left[(\beta - \beta_0)' Z_i(u) - \log\left\{\frac{S^{(0)}(\beta, u)}{S^{(0)}(\beta_0, u)}\right\}\right]^2 \lambda_i(u)\, du$$

$$= n^{-1} \int_0^t \left((\beta - \beta_0)' S^{(2)}(\beta_0, u)(\beta - \beta_0) - 2(\beta - \beta_0)' S^{(1)}(\beta_0, u) \log\left\{ \frac{S^{(0)}(\beta, u)}{S^{(0)}(\beta_0, u)} \right\} \right.$$
$$\left. + \left[\log\left\{ \frac{S^{(0)}(\beta, u)}{S^{(0)}(\beta_0, u)} \right\} \right]^2 S^{(0)}(\beta_0, u) \right) \lambda_0(u)\, du.$$

By Conditions A, B and D it follows that for each $\beta \in \mathscr{B}$,

$$A(\beta, 1) \to_{\mathscr{P}} \int_0^1 \left[(\beta - \beta_0)' s^{(1)}(\beta_0, u) - \log\left\{ \frac{s^{(0)}(\beta, u)}{s^{(0)}(\beta_0, u)} \right\} s^{(0)}(\beta_0, u) \right] \lambda_0(u)\, du,$$

while $nB(\beta, 1)$ converges in probability to some finite quantity (depending on β). Therefore by the inequality of Lenglart (I.2) we see that $X(\beta, 1)$ converges in probability to the same limit as $A(\beta, 1)$ for each $\beta \in \mathscr{B}$.

Now by the boundedness conditions in D we may evaluate the first and second derivatives of this limiting function of β by taking partial derivatives inside the integral (cf. Bartle, 1966, Corollary 5.9); these derivatives are therefore also by D equal to

$$\int_0^1 \left\{ s^{(1)}(\beta_0, u) - s^{(1)}(\beta, u) \frac{s^{(0)}(\beta_0, u)}{s^{(0)}(\beta, u)} \right\} \lambda_0(u)\, du$$
$$= \int_0^1 \{ e(\beta_0, u) - e(\beta, u) \} s^{(0)}(\beta_0, u) \lambda_0(u)\, du$$

and

$$\int_0^1 \left\{ -s^{(2)}(\beta, u) \frac{s^{(0)}(\beta_0, u)}{s^{(0)}(\beta, u)} + s^{(1)}(\beta, u)^{\otimes 2} \frac{s^{(0)}(\beta_0, u)}{s^{(0)}(\beta, u)^2} \right\} \lambda_0(u)\, du$$
$$= - \int_0^1 v(\beta, u) s^{(0)}(\beta_0, u) \lambda_0(u)\, du$$

respectively.

The first derivative is zero at $\beta = \beta_0$; the second is minus a positive semidefinite matrix; and at $\beta = \beta_0$ is minus a positive definite matrix. Thus for each $\beta \in \mathscr{B}$, $X(\beta, 1)$ converges in probability to a concave function of β with a unique maximum at $\beta = \beta_0$. Since $\hat{\beta}$ maximizes the random concave function $X(\beta, 1)$, it follows by some convex analysis (see Appendix 2) that $\hat{\beta} \to_{\mathscr{P}} \beta_0$. □

THEOREM 3.2. (*Asymptotic normality of* $\hat{\beta}$). $n^{1/2}(\hat{\beta} - \beta_0) \to_{\mathscr{D}} \mathscr{N}(0, \Sigma^{-1})$.

PROOF. (see Section 2.2.) We have two tasks here: first to show that

$$n^{-1/2} U(\beta_0, 1) \to_{\mathscr{D}} \mathscr{N}(0, \Sigma)$$

and second to show that

$$n^{-1} \mathscr{I}(\beta^*, 1) \to_{\mathscr{P}} \Sigma$$

for any *random* $\beta^* = \beta^{*(n)}$ such that $\beta^* \to_{\mathscr{P}} \beta_0$.

For the first part we use the fact

$$n^{-1/2} U(\beta_0, t) = n^{-1/2} \sum_{i=1}^n \int_0^t Z_i(u) dM_i(u) - n^{-1/2} \int_0^t E(\beta_0, u)\, d\bar{M}(u)$$
$$= \sum_{i=1}^n \int_0^t n^{-1/2} \{ Z_i(u) - E(\beta_0, u) \}\, dM_i(u).$$

We apply Theorem I.2 of Appendix I with

$$H_{i\ell}(t) = n^{-1/2}\{Z_\ell(t) - E(\beta_0, t)\}_i.$$

To verify condition (I.3), we note that

$$\int_0^t \sum_{\ell=1}^n H_{i\ell}(u)H_{j\ell}(u)\lambda_\ell(u)\, du$$

$$= \left(\int_0^t \left\{S^{(2)}(\beta_0, u) - \frac{S^{(1)}(\beta_0, u)^{\otimes 2}}{S^{(0)}(\beta_0, u)}\right\}\lambda_0(u)du\right)_{ij} \to_{\mathscr{P}} \left(\int_0^t v(\beta_0, u)s^{(0)}(\beta_0, u)\lambda_0(u)\, du\right)_{ij}$$

by Conditions A, B and D.

To verify (I.4) note first that by the simple inequality

$$|a - b|^2 I\{|a - b| > \varepsilon\} \le 4|a|^2 I\left\{|a| > \frac{\varepsilon}{2}\right\} + 4|b|^2 I\left\{|b| > \frac{\varepsilon}{2}\right\}$$

it is sufficient to verify

(i) $$\int_0^1 |E(\beta_0, t)|^2 I\{n^{-1/2}|E(\beta_0, t)| > \varepsilon\} \sum_{\ell=1}^n \frac{1}{n} Y_\ell(t)e^{\beta_0' Z_\ell(t)}\lambda_0(t)\, dt \to_{\mathscr{P}} 0,$$

(iia) $$\int_0^1 \frac{1}{n}\sum_{\ell=1}^n |Z_\ell(t)|^2 I\{n^{-1/2}|Z_\ell(t)| > \varepsilon, \beta_0' Z_\ell(t) \le -\delta|Z_\ell(t)|\} Y_\ell(t)e^{\beta_0' Z_\ell(t)}\lambda_0(t)\, dt \to_{\mathscr{P}} 0$$

and

(iib) $$\int_0^1 \frac{1}{n}\sum_{\ell=1}^n |Z_\ell(t)|^2 I\{n^{-1/2}|Z_\ell(t)| > \varepsilon, \beta_0' Z_\ell(t) > -\delta|Z_\ell(t)|\} Y_\ell(t)e^{\beta_0' Z_\ell(t)}\lambda_0(t)\, dt \to_{\mathscr{P}} 0.$$

Convergence of $E(\beta_0, \cdot)$ and $S^{(0)}(\beta_0, \cdot)$ and finiteness of $\int_0^1 \lambda_0(t)\, dt$ deals with (i) immediately. For (iib), we note that

$$\mathscr{P}[\exists \ell, t: n^{-1/2}|Z_\ell(t)| > \varepsilon, \beta_0' Z_\ell(t) > -\delta|Z_\ell(t)|, Y_\ell(t) = 1] \to 0$$

by Condition C. Finally, the quantity on the left hand side of (iia) is bounded by

$$\int_0^1 \frac{1}{n}\sum_{\ell=1}^n |Z_\ell(t)|^2 e^{-\delta|Z_\ell(t)|} I\{|Z_\ell(t)| > n^{1/2}\varepsilon\}\lambda_0(t)\, dt.$$

But $x^2 e^{-\delta x} \to 0$ as $x \to +\infty$. So for any $\eta > 0$, for large enough n this quantity is bounded by $\eta \int_0^1 \lambda_0(t)\, dt$.

This shows that $n^{-1/2}U(\beta_0, \cdot)$ converges weakly to a certain continuous Gaussian process. Since this process evaluated at time $t = 1$ has covariance matrix Σ, the first part of the proof is complete.

For the second part of the proof note that

$$n^{-1}\mathscr{I}(\beta^*, 1) = \int_0^1 V(\beta^*, t)\frac{d\bar{N}(t)}{n}$$

and that

$$\Sigma = \int_0^1 v(\beta_0, t)s^{(0)}(\beta_0, t)\lambda_0(t)\, dt.$$

Hence

$$\|n^{-1}\mathscr{I}(\beta^*, 1) - \Sigma\| \leq \left\| \int_0^1 \{V(\beta^*, t) - v(\beta^*, t)\} \frac{d\bar{N}}{n}(t) \right\|$$

$$+ \left\| \int_0^1 \{v(\beta^*, t) - v(\beta_0, t)\} \frac{d\bar{N}}{n}(t) \right\| + \left\| \int_0^1 v(\beta_0, t) \left\{ \frac{d\bar{N}}{n}(t) - \frac{\bar{\lambda}(t)}{n}\, dt \right\} \right\| \tag{3.1}$$

$$+ \left\| \int_0^1 v(\beta_0, t)\{S^{(0)}(\beta_0, t) - s^{(0)}(\beta_0, t)\}\lambda_0(t)\, dt \right\|.$$

First we show that $\lim_{c\uparrow\infty} \lim_{n\to\infty} \mathscr{P}\left[\frac{\bar{N}(1)}{n} > c\right] = 0$.

By consequence (I.1) of Lenglart's inequality,

$$\mathscr{P}\left[\frac{\bar{N}(1)}{n} > c\right] \leq \frac{\delta}{c} + \mathscr{P}\left[\int_0^1 S^{(0)}(\beta_0, t)\lambda_0(t)\, dt > \delta\right]. \tag{3.2}$$

For $\delta > \int_0^1 s^{(0)}(\beta_0, t)\lambda_0(t)\, dt$ the latter probability tends to zero as $n \to \infty$; and the required result now follows easily. Next, by B and the boundedness conditions in D, it follows that

$$\sup_{t\in[0,1],\beta\in\mathscr{B}} \| V(\beta, t) - v(\beta, t) \| \to_{\mathscr{P}} 0.$$

Hence $\beta^* \to_{\mathscr{P}} \beta_0$ and (3.2) imply that the first term on the right hand side of (3.1) converges in probability to zero.

Again, (3.2) together with the continuity in β, uniformly in t, in Condition D implies that the second term on the right hand side of (3.1) is also asymptotically negligible.

For the third term we use consequence (I.2) of the inequality of Lenglart. We have

$$\mathscr{P}\left[\left| \int_0^1 v_{ij}(\beta_0, t) \frac{d\bar{M}}{n}(t) \right| > \delta\right] \leq \eta/\delta^2 + \mathscr{P}\left[\frac{1}{n}\int_0^1 \{v_{ij}(\beta_0, t)\}^2 S^{(0)}(\beta_0, t)\, \lambda_0(t)\, dt > \eta\right].$$

Thus condition B plus the boundedness conditions in A and D shows that this term disappears too.

Finally the fourth term on the right hand side of (3.1) converges in probability to zero by directly applying Conditions A, B and D.

Note that this proof actually yields the stronger result

$$\sup_t \left\| \frac{1}{n}\mathscr{I}(\beta^*, t) - \Sigma(t) \right\| \to_{\mathscr{P}} 0,$$

where $\Sigma(t) = \int_0^t v(\beta_0, u)s^{(0)}(\beta_0, u)\lambda_0(u)\, du$. □

COROLLARY 3.3. *(Consistency of estimator of asymptotic covariance matrix of* $n^{1/2}(\hat{\beta} - \beta_0)$*).* $n^{-1}\mathscr{I}(\hat{\beta}, 1) \to_{\mathscr{P}} \Sigma$.

PROOF. See the last part of the proof of Theorem 3.2. □

THEOREM 3.4. *(Weak convergence of* $n^{1/2}(\hat{\Lambda} - \Lambda_0)$*).* $n^{1/2}(\hat{\beta} - \beta_0)$ *and the process equal in the point t to*

$$n^{1/2}\{\hat{\Lambda}(t) - \Lambda_0(t)\} + n^{1/2}(\hat{\beta} - \beta_0)' \int_0^t e(\beta_0, u)\lambda_0(u)\, du$$

are asymptotically independent, the latter being asymptotically distributed as a Gaussian martingale with variance function

$$\int_0^t \frac{\lambda_0(u)}{s^{(0)}(\beta_0, u)}\, du.$$

PROOF. (See Section 2.4.) Note first that by Condition B and the boundedness condition in Condition D,

$$\mathscr{P}(\Lambda_0^* = \Lambda_0 \quad \text{on} \quad [0, 1]) \to 1.$$

So we need not consider the term $n^{1/2}(\Lambda^*_0 - \Lambda_0)$ in the equality (2.8).

By precisely the same arguments as those we used to deal with $\mathscr{I}(\beta^*, 1)$ in the preceding proof, we can now show that

$$\sup_{t\in[0,1]} \left\| H(\beta^*, t) + \int_0^t e(\beta_0, u)\lambda_0(u)\, du \right\| \to_{\mathscr{P}} 0$$

for any β^* such that $\beta^* \to_{\mathscr{P}} \beta_0$.

It remains to apply Theorem I.2 to

$$n^{-1/2}U(\beta_0, \cdot) \quad \text{and} \quad \int_0^{\cdot} \frac{n^{-1/2}d\bar{M}(u)}{S^{(0)}(\beta_0, u)}$$

jointly.

By orthogonality of these two local square integrable martingales we need only consider verifying (I.3) and (I.4) for the second of the two; i.e. we take $p = 1$ and $H_{1\ell}(t) = n^{-1/2}/S^{(0)}(\beta_0, t)$.

But

$$\int_0^t \sum_{\ell=1}^n H_{1\ell}(u)^2\lambda_\ell(u)du = \int_0^t \frac{I\{\sum_{\ell=1}^n Y_\ell(u) > 0\}}{S^{(0)}(\beta_0, u)}\lambda_0(u)du \to_{\mathscr{P}} \int_0^t \frac{\lambda_0(u)}{s^{(0)}(\beta_0, u)}\, du$$

(giving (I.3)), and

$$\int_0^1 \sum_{\ell=1}^n H_{1\ell}(t)^2\lambda_\ell(t)I\{|H_{1\ell}(t) > \varepsilon|\}\, dt$$

is zero on the complement of the event $\{S^{(0)}(\beta_0, t) \le n^{1/2}\varepsilon \text{ for all } t\}$. So by Conditions B and D, (I.4) holds too. □

COROLLARY 3.5. (*Consistency of estimator of limiting covariance function of* $n^{1/2}(\hat{\Lambda} - \Lambda_0)$.)

$$\sup_{t\in[0,1]} \left\| H(\hat{\beta}, t) + \int_0^t e(\beta_0, u)\lambda_0(u)\, du \right\| \to_{\mathscr{P}} 0$$

where H is given by (2.9).

PROOF. It was indicated how this could be proved at the beginning of the proof of Theorem 3.4. □

4. Some special cases. Before considering some special cases of our model in detail, we shall give a general discussion of our conditions.

A. (Finite interval). From a practical point of view, this condition is hard to justify. One would like to use all the observations on the whole line $[0, \infty)$, and since in general we will

have $\int_0^\infty \lambda_0(t)\,dt = \infty$, the infinite interval case cannot be derived from the finite interval case by a simple mapping. Also in general we will have $s^{(0)}(t) \to 0$ as $t \to \infty$, so Condition D also prevents easy extension to $[0, \infty)$.

An identical problem arises when other statistical methods for analysing censored data are described from the point of view of counting processes, see Gill (1980) or Andersen et al. (1982). In those cases some extra conditions have to be made ensuring that the contribution to the test statistics from the data on $[\tau, \infty)$ can be made arbitrarily small, uniformly in n, by taking τ large enough. In Theorem 4.2 we shall give such an extension in the special case of bounded covariates, only sketching the proof though.

Our finite interval condition is also present, explicitly or implicitly, in the cases studied by Tsiatis (1981a), Næs (1982) and Bailey (1979).

B. (Asymptotic stability). Up to the uniformity in β, and to a lesser extent, in t, these conditions speak for themselves. Such conditions have been proposed in a heuristic proof of consistency of $\hat{\beta}$ by Oakes (1981). It may be noted that uniformity in β is not required for consistency of $\hat{\beta}$ or for asymptotic normality of $n^{-1/2}\partial \log L(\beta) / \partial\beta \,|_{\beta=\beta_0}$. However it is to be expected that some kind of convergence uniform in β will be needed to ensure convergence in probability of $n^{-1}\partial^2 \log L(\beta)/\partial\beta^2 \,|_{\beta=\hat{\beta}}$.

C. (Lindeberg condition). This condition appears at first sight complicated, but in some important special cases it is very easy to verify. For instance, if the covariates are bounded, the condition is completely empty; if they are bounded by random variables having a bounded rth moment for some $r > 2$ it is also easy to verify. We shall see presently that in the special case of i.i.d. observations it is implied by a natural second moment condition. Finally the condition simplifies somewhat in the one-dimensional case $p = 1$; it is then equivalent to

$$n^{-1/2} \sup_{i,t} Z_i(t) Y_i(t) \to_{\mathscr{P}} 0 \qquad \text{if} \quad \beta_0 > 0$$

$$n^{-1/2} \sup_{i,t} -Z_i(t) Y_i(t) \to_{\mathscr{P}} 0 \qquad \text{if} \quad \beta_0 < 0$$

$$n^{-1/2} \sup_{i,t} |Z_i(t)|\, Y_i(t) \to_{\mathscr{P}} 0 \qquad \text{if} \quad \beta_0 = 0.$$

D. (Asymptotic regularity conditions). These conditions on boundedness, continuity, and interchanging of orders of various limiting operations do not require any discussion.

Finally some miscellaneous remarks. Note that the function $\lambda_0(t)$ itself is never needed in the theorems or their proofs; we could replace $\lambda_0(t)dt$ everywhere with $d\Lambda_0(t)$, where Λ_0 is assumed continuous, nondecreasing and $0 = \Lambda_0(0) < \Lambda_0(1) < \infty$. Next, a suitable choice of covariates yields as score test many of the standard $(p + 1)$-sample tests in the literature of censored data, see Lustbader (1980) and Oakes (1981) for a discussion of this point in the case $p = 1$. Since we only need to work under β_0 for this statistic, the uniformity in β in the conditions may be relaxed. Finally, we have not discussed the asymptotic distribution of the generalized likelihood ratio test. However, it is clear that our methods will give the expected results under the same conditions, see Rao (1973, Section 6e).

The literature so far only contains rigorous treatments of the Cox model in what are essentially i.i.d. cases, and in order to show how powerful our methods are, we shall show here how Conditions A to D are satisfied in such cases. Four approaches deserve special attention. Næs (1982) employs martingale techniques in the model where $(N_i^{(n)}, Y_i^{(n)}, Z_i^{(n)})$, $i = 1, \cdots, n$ are i.i.d. replicates of (N, Y, Z) say. He works on a finite interval, with bounded covariates. Tsiatis (1981a) works also in an i.i.d. set up with Z 1-dimensional and time independent though possibly random. He also has a finite interval condition and further a second moment condition $\mathscr{E}(Z^2 e^{2\beta Z}) < \infty$ for β in a neighbourhood of β_0. This is stronger than what would be a natural condition here, $\mathscr{E}(Z^2 e^{\beta Z}) < \infty$ for β in a neighbourhood of β_0. Liu and Crowley (1978) discuss a situation in which, in each of a finite number

of independent strata, $(N_i^{(n)}, Y_i^{(n)}, Z_i^{(n)})$ are i.i.d. and $Z_i^{(n)}$ is time independent and non-random. They do not make a finite interval assumption; however their proof is extremely complicated. Finally Bailey (1979) assumes fixed censoring imposed on an i.i.d. situation: i.e. one observes i.i.d. replicates of (N, Y, Z) on intervals $[0, t_i^{(n)}]$, $i = 1, \cdots, n$. All these authors work in the original life testing situation considered by Cox: i.e. $N_i^{(n)}$ makes at most one jump. Various independence assumptions are also made. In each case our counting process model is applicable.

We shall consider the i.i.d. case in detail: $(N_i^{(n)}, Y_i^{(n)}, Z_i^{(n)})$ are i.i.d. replicates of (N, Y, Z). We suppose that Y and Z are left continuous processes with right hand limits, which will allow us to apply laws of large numbers for the space $D[0, 1]$ (after reversing the time axis!). It will be obvious (taking the results of Appendix 3 into account) how to make either or both of the following extensions: i.i.d. case within each of a finite number of independent strata, a positive limiting fraction of observations in each stratum as $n \to \infty$; and the case where $(N_i^{(n)}, Y_i^{(n)}, Z_i^{(n)})$ are observations of independent replicates of (N, Y, Z) on fixed intervals $[0, t_i^{(n)}]$, where the distribution of the $t_i^{(n)}$'s converges as $n \to \infty$. Thus all the cases considered by the above authors are covered.

THEOREM 4.1. *In the* i.i.d. *case with Z and Y left continuous with right hand limits, Conditions A to D are satisfied if*:

$$\int_0^1 \lambda_0(t)\, dt < \infty, \tag{4.1}$$

there exists a neighbourhood $\mathscr{B}$ of β_0 such that

$$\mathscr{E}\{\sup_{t\in[0,1],\beta\in\mathscr{B}} Y(t)\,|\,Z(t)\,|^2 e^{\beta' Z(t)}\} < \infty, \tag{4.2}$$

$$\mathscr{P}\{Y(t) = 1 \forall t \in [0, 1]\} > 0, \tag{4.3}$$

and

(4.4) Σ *is positive definite, where $s^{(0)}$, $s^{(1)}$ and $s^{(2)}$ are now defined by $s^{(0)}(\beta, t) = \mathscr{E}\{Y(t)e^{\beta' Z(t)}\}$, $s^{(1)}(\beta, t) = \mathscr{E}\{Y(t)Z(t)e^{\beta' Z(t)}\}$, $s^{(2)}(\beta, t) = \mathscr{E}\{Y(t)Z(t)^{\otimes 2}e^{\beta' Z(t)}\}$.*

PROOF. By (4.2) we also have

$$\mathscr{E}\{\sup_{t\in[0,1],\beta\in\mathscr{B}} Y(t)\,|\,Z(t)\,|\,e^{\beta' Z(t)}\} < \infty$$

and

$$\mathscr{E}\{\sup_{t\in[0,1],\,\beta\in\mathscr{B}} Y(t)e^{\beta' Z(t)}\} < \infty.$$

By dominated convergence $s^{(0)}$, $s^{(1)}$ and $s^{(2)}$ are continuous functions of $\beta \in \mathscr{B}$ for each $t \in [0, 1]$, uniformly in $t \in [0, 1]$. They are also bounded on $\mathscr{B} \times [0, 1]$ and, by (4.3), $s^{(0)}$ is bounded away from zero on $\mathscr{B} \times [0, 1]$. Without loss of generality we may take $\mathscr{B}$ to be compact. We can consider $Y(t)e^{\beta' Z(t)}$ as a random element of $D[0, 1]$, where the elements of $D[0, 1]$ take values not in $\mathbb{R}$ but in the Banach space of continuous functions on $\mathscr{B}$ endowed with the supremum norm. Then by Theorem III.1, Condition B holds for $S^{(0)}$; the same argument works for $S^{(1)}$ and $S^{(2)}$.

This leaves Condition C (the Lindeberg condition) to verify. First note that if $X_1, X_2, \cdots$ are i.i.d. random variables with $\mathscr{E}(X_i^2) < \infty$, then by the central limit theorem it holds that $\sup_{i=1,\ldots,n} n^{-1/2}\,|X_i - \mathscr{E}X_i| \to_{\mathscr{P}} 0$ as $n \to \infty$, which implies that $\sup_{i=1,\ldots,n} n^{-1/2}\,|X_i| \to_{\mathscr{P}} 0$ as $n \to \infty$. Thus in the i.i.d. case, Condition C holds if there exists $\delta > 0$ such that

$$\mathscr{E}[\sup_{t\in[0,1]} Y(t)|\,Z(t)|^2 I\{\beta_0' Z(t) > -\delta\,|\,Z(t)|\}] < \infty.$$

Now choose δ such that the closed cube of side 2δ, centre β_0, is contained in $\mathscr{B}$. Then

$$\beta_0' Z(t) > -\delta\,|\,Z(t)| \Rightarrow \exists \beta \in \mathscr{B} \quad \text{such that} \quad \beta' Z(t) > 0;$$

simply choose the jth coordinate of β to be $(\beta_0)_j + \delta$ if $(Z(t))_j \geq 0$, $(\beta_0)_j - \delta$ if $(Z(t))_j < 0$. Thus

$$\sup_{\beta \in \mathscr{B}} Y(t) |\, Z(t)|^2 e^{\beta' Z(t)} \geq Y(t) |\, Z(t)|^2 I\{\beta_0' Z(t) > -\delta \mid Z(t)|\};$$

and Condition C holds by (4.2). □

Next we shall see how in the i.i.d. case with bounded covariates, one can easily extend the results to the infinite interval $[0, \infty)$. As an example we shall sketch the proof of an extended version of Theorem 3.2, but also Lemma 3.1, Corollary 3.3, Theorem 3.4 and Corollary 3.5 generalize in the obvious fashion. Σ is now defined by integration over $[0, \infty)$. The condition of bounded covariates can probably be replaced by conditions similar to (4.2).

THEOREM 4.2. *Suppose in the* i.i.d. *case, given in Theorem* 4.1, *Z is bounded and $\mathscr{E} N(\infty) < \infty$. Then if $\hat{\beta}$ is Cox's estimator of β based on the observations on* $[0, \infty)$ *instead of only on* $[0, 1]$,

$$n^{1/2}(\hat{\beta} - \beta_0) \to_{\mathscr{D}} \mathscr{N}(0, \Sigma^{-1}),$$

provided only that Σ is positive definite, and that for each $\tau < \infty$, $\mathscr{P}(Y(t) = 1 \ \forall t \leq \tau) > 0$.

PROOF. Since $\mathscr{E} N(\infty) < \infty$, we also have

$$\mathscr{E} \int_0^\infty Y(s) e^{\beta_0' Z(s)} \lambda_0(s)\, ds < \infty.$$

So by boundedness of Z and the condition on Y, it must be that $\int_0^\tau \lambda_0(s)\, ds < \infty$ for each $\tau < \infty$. Thus the conditions of Theorem 4.1 are satisfied on the interval $[0, \tau]$ for each $\tau < \infty$.

To extend to $[0, \infty]$ we must show that throughout the proofs, the contribution from (τ, ∞) can be made arbitrarily small, uniformly in n, by choosing τ large enough, see Gill (1980) Theorems 4.2.1 and 4.3.1 or Andersen et al. (1982) Theorem 3.1 for examples of this technique. We illustrate it here by considering one of the simplest such cases where this must be done. This is in the proof of Theorem 3.2 where we must show that in particular

$$\lim_{\tau \uparrow \infty} \limsup_{n \to \infty} \mathscr{P}\left\{ \int_\tau^\infty \| S^{(2)}(\beta_0, t) \| \lambda_0(t)\, dt > \varepsilon \right\} = 0$$

for any $\varepsilon > 0$. By boundedness of Z we may consider proving instead

$$\lim_{\tau \uparrow \infty} \limsup_{n \to \infty} \mathscr{P}\left\{ \int_\tau^\infty \frac{1}{n} \sum_{i=1}^n Y_i(t) e^{\beta_0' Z_i(t)} \lambda_0(t)\, dt > \varepsilon \right\} = 0.$$

Now

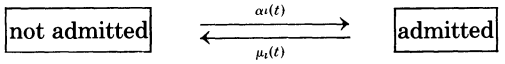
$$\mathscr{P}\left\{ \int_\tau^\infty \frac{1}{n} \sum_{i=1}^n Y_i(t) e^{\beta_0' Z_i(t)} \lambda_0(t)\, dt > \varepsilon \right\} \leq \mathscr{E}\left\{ \int_\tau^\infty \frac{1}{n} \sum_{i=1}^n Y_i(t) e^{\beta_0' Z_i(t)} \lambda_0(t)\, dt \right\} \Big/ \varepsilon$$

$$= \mathscr{E}\left\{ \frac{1}{n} \sum_{i=1}^n N_i(\infty) - N_i(\tau) \right\} \Big/ \varepsilon \to 0 \text{ as } \tau \uparrow \infty.$$

The quantity $\sup_{\beta \in \mathscr{B}} \{\mathscr{I}(\beta, \infty) - \mathscr{I}(\beta, \tau)\}$ in the proof of Theorem 3.2 may be dealt with similarly. □

EXAMPLE. Finally we shall consider an example concerning admissions to psychiatric hospitals among women giving birth (Andersen and Rasmussen, 1982). In that study it was investigated who among the about $n = 70{,}000$ Danish women giving birth to a child in 1975 had been admitted (possibly more than once) to a psychiatric hospital in the period from 1 October 1973 to 31 December 1976 and the dates of admission and discharge respectively were registered. Moreover, information on such demographic factors as age, marital status and parity (= number of children born before 1975) was available. Due to the fact that the exact date of birth was known only for the women who were actually admitted during the time span considered, reliable information on admissions was only available in the time interval ranging from −15 months = −456 days to 12 months = 366 days relative to the date of birth, and hence that interval is the relevant one to consider.

Let $Y_i(t) = 0$ if women i ($i = 1, \cdots, n$) is resident in a psychiatric hospital at time t relative to the date of birth (-456 days $\leq t \leq 366$ days) and let $Y_i(t) = 1$ otherwise; let $N_i(t)$ be the number of admissions for woman i in the interval [−456 days, t]. For each woman $i = 1, \cdots, n$ we consider the two state Markov process model:



where $\alpha_i(t)$ and $\mu_i(t)$ are the forces of transition. It follows that $N_i(t)$ is a counting process with intensity process $\lambda_i(t) = \alpha_i(t)Y_i(t)$ (cf. Aalen, 1978, page 709). In the following we assume that this intensity process has the form (2.1), i.e.

$$\lambda_i(t) = \lambda_0(t)e^{\beta' Z_i(t)}Y_i(t), \quad i = 1, \cdots, n,$$

where the information on demographic variables and admissions prior to time t for woman i is collected in the vector $Z_i(t)$.

First we consider a model (Model I) where only the parity of the woman and the time relative to the date of birth are assumed to influence the probability of being admitted to a psychiatric hospital and we define the (time independent) covariates

$$Z_{i1} = \begin{cases} 1 & \text{if woman } i \text{ has parity 0,} \\ 0 & \text{otherwise,} \end{cases}$$

$$Z_{i2} = \begin{cases} 1 & \text{if woman } i \text{ has parity 2,} \\ 0 & \text{otherwise,} \end{cases}$$

$$Z_{i3} = \begin{cases} 1 & \text{if woman } i \text{ has parity} \geq 3, \\ 0 & \text{otherwise,} \end{cases}$$

$\lambda_0(t)$ being the force of transition for women with parity 1. The estimated regression coefficients in Model I are given in Table 1 (numbers in brackets) together with their estimated standard errors and correlations.

From Table 1 it seems that the intensity of being admitted is much larger when parity exceeds 2 but also the women with 2 children seem to have a somewhat increased intensity compared to those with parity 1 or 0. The Wald test statistic for the global null hypothesis $(\beta_1, \beta_2, \beta_3) = 0$ takes the highly significant value 18.34 with 3 degrees of freedom in fairly close agreement with the value of the likelihood ratio test statistic 16.22.

As mentioned above, the counting process generalizations of the usual nonparametric k sample tests for censored survival data may be obtained as score tests by appropriate choices of stochastic covariates. In the present example, the generalized log-rank test statistic takes the value 18.6 and the same value is attained by the Breslow generalisation of the Kruskal-Wallis test (see Andersen et al., 1982). We conclude that parity has a highly significant influence on the intensity of being admitted to a psychiatric hospital.

Consider next the model (Model II) obtained by introducing the age of the women in the model by means of the two covariates

$$Z_{i4} = \begin{cases} 1 & \text{if woman } i \text{ is} \leq 18 \text{ years old,} \\ 0 & \text{otherwise,} \end{cases} \qquad Z_{i5} = \begin{cases} 1 & \text{if woman } i \text{ is} > 34 \text{ years old,} \\ 0 & \text{otherwise.} \end{cases}$$

TABLE 1
Estimates in Markov process models (Models I, II*)*

i	covariate Z_i	$\hat{\beta}_i$	$\{\text{var}(\hat{\beta}_i)\}^{1/2}$	estimated correlations				
				1	2	3	4	5
1	parity 0	0.094	0.010	1				
		(0.090)	(0.099)	(1)				
2	parity 2	0.202	0.131	0.42	1			
		(0.253)	(0.130)	(0.43)	(1)			
3	parity ≥ 3	0.458	0.168	0.32	0.30	1		
		(0.641)	(0.157)	(0.36)	(0.27)	(1)		
4	age ≤ 18 years	0.116	0.238	−0.16	0.01	0.00	1	
5	age > 34 years	0.601	0.162	0.04	−0.13	−0.35	0.02	1

TABLE 2
Estimates in semi-Markov process model (Model III*)*

i	covariate Z_i	$\hat{\beta}_i$	$\{\text{var}(\hat{\beta}_i)\}^{1/2}$	estimated correlations					
				1	2	3	4	5	6
1	parity 0	0.036	0.102	1					
2	parity 2	0.252	0.132	0.42	1				
3	parity ≥ 3	0.473	0.167	0.32	0.30	1			
4	age ≤ 18 years	0.113	0.251	−0.16	0.02	0.01	1		
5	age > 34 years	0.473	0.160	0.03	−0.13	−0.33	0.01	1	
6	admission during latest month	6.13	0.133	−0.03	0.02	−0.01	0.02	0.01	1

The estimates for Model II are given in Table 1. By comparing $\hat{\beta}_5$ with its estimated standard error it seems that the intensity of being admitted is increased for the older women compared with the younger ones. Furthermore the influence of parity has diminished compared to Model I; this is of course due to the positive correlation between parity and age, which is reflected for example in the fairly large negative estimated correlation coefficients: −0.16 between $\hat{\beta}_1$ and $\hat{\beta}_4$ and −0.35 between $\hat{\beta}_3$ and $\hat{\beta}_5$. The Wald test statistic for no influence of age when parity is included in the model takes the value 14.01 with 2 degrees of freedom compared to the value 12.28 of the likelihood ratio test statistic. The Wald test statistic for no influence of parity when age is included in the model takes the value 8.04 with 3 degrees of freedom, and we conclude that both age and parity have a significant influence on the intensity of being admitted in spite of the positive correlation between these two covariates.

Finally we consider a model (Model III) where the psychiatric past of the women is introduced by means of the time dependent covariate

$$Z_{i6}(t) = \begin{cases} 1 & \text{if woman } i \text{ has been resident in a psychiatric hospital} \\ & \text{during the month } [t - 30 \text{ days}, t), \\ 0 & \text{otherwise.} \end{cases}$$

Since the probability of being admitted in this model depends on the time since latest admission, Model III is a semi-Markov model rather than a Markov model as Models I and II. The estimates in Model III are given in Table 2.

From Table 2 we notice first the marked influence of prior admissions reflected by the value $\hat{\beta}_6 = 6.13$ with an estimated standard error of 0.133. The estimated effects of age and

parity are practically unchanged compared with Model II, and we conclude that the oldest women and women with high parity, i.e. women who have terminated their "birth career," have the highest intensity of being admitted to a psychiatric hospital in connection with another pregnancy which is being carried to term. Furthermore prior admissions increase the risk of being admitted again.

APPENDIX I.

All stochastic processes are defined on the time interval [0, 1].

THEOREM I.1. *(Two applications of the inequality of Lenglart).*

(a) *Let N be a univariate counting process with intensity process λ. Then for all δ, $\eta > 0$*

$$\mathscr{P}\{N(1) > \eta\} \leq \frac{\delta}{\eta} + \mathscr{P}\left\{\int_0^1 \lambda(t)\, dt > \delta\right\}. \tag{I.1}$$

(b) *Let W be a local square integrable martingale. Then for all δ, $\eta > 0$*

$$\mathscr{P}\{\sup_{t\in[0,1]} |\, W(t)| > \eta\} \leq \frac{\delta}{\eta^2} + \mathscr{P}\{\langle W, W\rangle(1) > \delta\}. \tag{I.2}$$

THEOREM I.2. *(Application of Rebolledo's Central Limit Theorem for local square integrable martingales). For each $n = 1, 2, \cdots$ let $N^{(n)}$ be a multivariate counting process with n components. Let $H^{(n)}$ be a $p \times n$ ($p \geq 1$ is fixed) matrix of locally bounded predictable processes. Suppose that $N^{(n)}$ has an intensity process $\lambda^{(n)}$, and define local square integrable martingales $W^{(n)} = (W_1^{(n)}, \cdots, W_p^{(n)})$ by*

$$W_i^{(n)}(t) = \int_0^t \sum_{\ell=1}^n H_{i\ell}^{(n)}(u)\{dN_\ell^{(n)}(u) - \lambda_\ell^{(n)}(u)\, du\}.$$

Let A be a $p \times p$ matrix of continuous functions on $[0, 1]$ which form the covariance functions of a continuous p-variate Gaussian martingale $W^{(\infty)}$, with $W^{(\infty)}(0) = 0$; i.e. $\mathrm{Cov}(W_i^{(\infty)}(t), W_j^{(\infty)}(u)) = A_{ij}(t \wedge u)$ for all i, j, t and u. Suppose that for all i, j and t

$$\langle W_i^{(n)}, W_j^{(n)}\rangle(t) = \int_0^t \sum_{\ell=1}^n H_{i\ell}^{(n)}(s) H_{j\ell}^{(n)}(s) \lambda_\ell^{(n)}(s)\, ds \to_{\mathscr{P}} A_{ij}(t) \tag{I.3}$$

as $n \to \infty$ and that for all i and $\varepsilon > 0$

$$\int_0^1 \sum_{\ell=1}^n H_{i\ell}^{(n)}(t)^2 \lambda_\ell^{(n)}(t) I\{|\, H_{i\ell}^{(n)}(t)| > \varepsilon\}\, dt \to_{\mathscr{P}} 0 \quad \text{as } n \to \infty. \tag{I.4}$$

Then $W^{(n)} \to_{\mathscr{D}} W^{(\infty)}$ as $n \to \infty$ in $D([0, 1]^p)$.

These two results have each in different ways been slightly extended with respect to the originals. In the first place Theorem I.1 is only a direct application of Lenglart's inequality when $\mathscr{E}\{N(1)\} < \infty$ in (I.1) or $\mathscr{E}\{W(1)^2\} < \infty$ in (I.2). In our situation we only know that a sequence $T_1 \leq T_2 \leq \cdots \leq 1$ of stopping times exists, $\mathscr{P}(T_i = 1) \to 1$ as $i \to \infty$, such that $\mathscr{E}\{N(T_i)\} < \infty$ or $\mathscr{E}\{W(T_i)^2\} < \infty$ for all i. So the inequality of Lenglart does directly apply to the "stopped" processes, $N(t \wedge T_i)$ and $W(t \wedge T_i)$. Letting $i \to \infty$ then gives our versions.

Theorem I.2 has been extended by making the original univariate theorem into a p-variate theorem. This extension can be done by standard Cramer-Wold type arguments, see for instance Aalen (1977, Lemma A.1) for a similar extension worked out in detail.

APPENDIX II.

Pointwise convergence in probability of random concave functions implies uniform convergence on compact subspaces.

The "almost sure" version of this theorem is a direct consequence of Rockafellar (1970, Theorem 10.8). However for an "in probability" result we must be more careful. The following "diagonalization method" was pointed out by T. Brown.

THEOREM II.1. *Let E be an open convex subset of $\mathbb{R}^p$ and let $F_1, F_2, \cdots$, be a sequence of random concave functions on E such that $\forall x \in E$, $F_n(x) \to_{\mathscr{P}} f(x)$ as $n \to \infty$ where f is some real function on E. Then f is also concave and for all compact $A \subset E$,*

$$\sup_{x \in A} | F_n(x) - f(x)| \to_{\mathscr{P}} 0 \text{ as } n \to \infty.$$

PROOF. Concavity of f is obvious. Next let $x_1, x_2 \cdots$ be a countable dense set of points in E. Since $F_n(x_1) \to_{\mathscr{P}} f(x_1)$ as $n \to \infty$ there exists a subsequence along which convergence holds almost surely. Along this subsequence $F_n(x_2) \to_{\mathscr{P}} f(x_2)$ so a further subsubsequence exists along which also $F_n(x_2) \to_{\text{a.s.}} f(x_2)$. Repeating the argument, along a $(\text{sub})^k$ sequence, $F_n(x_j) \to_{\text{a.s.}} f(x_j)$ for $j = 1, \cdots, k$. Now consider the new subsequence formed by taking the first element of the first subsequence, the second of the second, etc. Along the new subsequence we must have $F_n(x_j) \to_{\text{a.s.}} f(x_j)$ for each $j = 1, 2, \cdots$.

By Rockafellar (1970, Theorem 10.8) it now follows that

$$\sup_{x \in A} | F_n(x) - f(x)| \to_{\text{a.s.}} 0 \text{ along this subsequence.}$$

We have shown more generally how, from any subsequence, a further subsequence can be extracted along which $\sup_{x \in A} | F_n(x) - f(x)| \to_{\text{a.s.}} 0$. It now follows that

$$\sup_{x \in A} | F_n(x) - f(x)| \to_{\mathscr{P}} 0 \text{ along the whole sequence.} \quad \square$$

COROLLARY II.2. *Suppose f has a unique maximum at $\hat{x} \in E$. Let $\hat{X}_n$ maximize F_n. Then $\hat{X}_n \to_{\mathscr{P}} \hat{x}$ as $n \to \infty$.*

PROOF. The proof, a simple $\varepsilon - \delta$ argument, is left to the reader. $\square$

APPENDIX III.

Extension of SLLN for $D[0, 1]$.

Let $X; X_1, X_2, \cdots$ be i.i.d. random elements of $D[0, 1]$ with $\mathscr{E}\|X\| = \mathscr{E} \sup_{t \in [0,1]} |X(t)| < \infty$. Then by Theorem 1 of R. Ranga Rao (1963) we have almost surely

$$\left\| \frac{1}{n} \sum_{i=1}^{n} X_i - \mathscr{E}X \right\| \to 0 \quad \text{as } n \to \infty.$$

We need to extend this result in two directions. Firstly we must allow the random elements of $D[0, 1]$ to be random functions not from $[0, 1]$ to $\mathbb{R}$ but from $[0, 1]$ to the space of continuous real functions on $\mathscr{B}$, where $\mathscr{B}$ is a compact neighbourhood of $\beta_0 \in \mathbb{R}^p$. If we endow this space of functions with the supremum norm it becomes a separable Banach space, and that will be all the structure we need.

Secondly we must allow for censoring. To tie in with the usual right continuity convention for $D[0, 1]$, we shall consider *left* censoring: X_i is only observed on an interval $[t_i, 1]$, or more generally, in a triangular array scheme, on $[t_i^{(n)}, 1]$ or $[T_i^{(n)}, 1]$ for fixed or random times $t_i^{(n)}$ or $T_i^{(n)}$ respectively.

THEOREM III.1. *Let $X; X_1, X_2, \cdots$ be* i.i.d. *random elements of $D_E[0, 1]$ (endowed with the Skorohod topology) where the elements of $D_E[0, 1]$ are right continuous functions on $[0, 1]$ with left hand limits taking values in a separable Banach space E (rather than the usual $\mathbb{R}$). Suppose that $\mathscr{E}\|X\| = \mathscr{E} \sup_{t \in [0,1]} \|X(t)\| < \infty$.*

For each, n, let $t_1^{(n)} \geq \cdots \geq t_n^{(n)}$ *be fixed time instants in* $[0, 1]$. *Let* $y_i^{(n)} = I_{[t_i^{(n)}, 1]}$ *and suppose there exists a distribution function* y *such that, on* $[0, 1]$,

$$\left\| \frac{1}{n} \sum_{i=1}^n y_i^{(n)} - y \right\| \to 0 \text{ as } n \to \infty.$$

Then

$$\left\| \frac{1}{n} \sum_{i=1}^n X_i y_i^{(n)} - \mathscr{E} Xy \right\| \to 0 \text{ almost surely as } n \to \infty.$$

PROOF. Note that $D_E[0, 1]$, just like $D_R[0, 1]$, is separable and complete with respect to the Skorohod d_0-metric (see Billingsley (1968, Section 14), replacing $|\cdot|$ where appropriate with $\|\cdot\|$). Thus any random element of $D_E[0, 1]$ is tight (Billingsley, 1968, Theorem 1.4). Also X *is* a random element of $D_E[0, 1]$, provided that its sample paths have the correct properties and that $X(t)$ is a random vector for each $t \in [0, 1]$. The characterizations of compact sets of $D_R[0, 1]$ given by Billingsley (1968, Theorems 14.3 and 14.4) do not carry over directly to $D_E[0, 1]$, since in E a closed, bounded set is not necessarily compact. However it can be verified that the conditions given are still *necessary* for compactness, even if not *sufficient* any more.

We shall make use of the following three properties of $D_E[0, 1]$ corresponding to properties used by Rao (1963) in the case $E = \mathbb{R}$:

(i) $\forall \varepsilon > 0$, $\forall$ compact $K \subset D_E[0, 1]$, $\exists \delta > 0$ such that $x \in K$ and $\alpha \leq t < \beta \leq \delta \Rightarrow \| x(t) - x(\alpha) \| \leq \| x(\beta-) - x(\alpha) \| + \varepsilon$. (Billingsley, 1968, Theorem 14.4, necessary condition for compactness; see above.)

(ii) $\mathscr{E}[\| X \|] < \infty \Rightarrow \forall \delta > 0\ \exists\ 0 = t_0 < t_1 < \cdots < t_N = 1$ such that, for all j, $| t_{j+1} - t_j | < \delta$ and $\mathscr{E}[\| X(t_{j+1} -) - X(t_j) \|] \leq \varepsilon$. (Proof as in Rao, 1963, Lemma 2.)

(iii) $\mathscr{E}[\| X \|] < \infty \Rightarrow \mathscr{E}[\| X \| I\{X \notin K\}]$ can be made arbitrarily small by suitable choice of compact $K \subset D_E[0, 1]$. (Tightness of random elements of $D_E[0, 1]$, see above.)

Given compact $K \subset D_E[0, 1]$ and $\varepsilon > 0$, choose δ by (i). There exists a finite partition of $[0, 1)$ such that $y(\beta -) - y(\alpha) < \varepsilon$ for each interval $[\alpha, \beta)$ in the partition. By applying (ii) for each interval in the partition separately, we can find $0 = t_0 < t_1 < \cdots < t_N = 1$ such that $| t_{j+1} - t_j | < \delta$, $| y(t_{j+1}) - y(t_j) | < \varepsilon$ and $\mathscr{E}\{\| X(t_{j+1} -) - X(t_j) \|\} < \varepsilon$ for each $j = 0, \cdots, N - 1$.

Since trivially $\| \mathscr{E} X((1/n) \sum_{i=1}^n y_i^{(n)} - y) \| \to 0$ as $n \to \infty$, it suffices to consider $\| (1/n) \sum_{i=1}^n (X_i - \mathscr{E} X) y_i^{(n)} \|$. For any $t \in [0, 1)$, let $[\alpha, \beta)$ be the $[t_j, t_{j+1})$ interval containing t. Then

$$\left\| \frac{1}{n} \sum_{i=1}^n (X_i(t) - \mathscr{E} X(t)) y_i^{(n)}(t) \right\| \leq \varepsilon_n^K(t) + \frac{1}{n} \sum_{i=1}^n \| X_i \| I\{X_i \notin K\} + \mathscr{E}[\| X \| I\{X \notin K\}]$$

where

$$\begin{aligned}
\varepsilon_n^K(t) &= \left\| \frac{1}{n} \sum_{i=1}^n (X_i(t) I\{X_i \in K\} - \mathscr{E}[X(t) I\{X \in K\}]) y_i^{(n)}(t) \right\| \\
&\leq \left\| \frac{1}{n} \sum_{i=1}^n (X_i(\alpha) I\{X_i \in K\} - \mathscr{E}[X(\alpha) I\{X \in K\}]) y_i^n(t) \right\| \\
&\quad + \frac{1}{n} \sum_{i=1}^n \| X_i(\beta-) - X_i(\alpha) \| + \mathscr{E}\{\| X(\beta-) - X(\alpha) \|\} + 2\varepsilon \\
&\leq \left\| \frac{1}{n} \sum_{i=1}^n (X_i(\alpha) I\{X_i \in K\} - \mathscr{E}[X(\alpha) I\{X \in K\}]) y_i^{(n)}(\alpha) \right\| \\
&\quad + \frac{1}{n} \sum_{i=1}^n \| X_i(\alpha) \| \{ y_i^{(n)}(\beta-) - y_i^{(n)}(\alpha) \} \\
&\quad + \frac{1}{n} \sum_{i=1}^n \mathscr{E} \| X(\alpha) \| \{ y_i^{(n)}(\beta-) - y_i^{(n)}(\alpha) \} + \frac{1}{n} \sum_{i=1}^n \| X_i(\beta-) - X_i(\alpha) \| + 3\varepsilon.
\end{aligned}$$

Now suppose U; U_1, U_2, $\cdots$ are i.i.d. separable Banach space valued random variables with $\mathscr{E}[\| U \|] < \infty$. By the SLLN for separable Banach space (Mourier, 1953) we have $(1/n) \sum_{i=1}^{n} U_i \to \mathscr{E} U$ almost surely as $n \to \infty$.

We want to show that we now also have

$$\frac{1}{n} \sum_{i=1}^{n} U_i y_i^{(n)}(\alpha) \to \mathscr{E} U y(\alpha) \text{ almost surely as } n \to \infty.$$

Define $k(n, \alpha) = \max\{i : t_i^{(n)} \geq \alpha\}$. Then

$$\frac{1}{n} \sum_{i=1}^{n} U_i y_i^{(n)}(\alpha) = \frac{k(n, \alpha)}{n} \frac{\sum_{i=1}^{k(n,\alpha)} U_i}{k(n, \alpha)},$$

where $k(n, \alpha)/n \to y(\alpha)$ as $n \to \infty$. If $k(n, \alpha) \to \infty$ as $n \to \infty$ then the required result obviously holds. Also if $k(n, \alpha)$ remains bounded as $n \to \infty$ (which implies, but is not implied by, $y(\alpha) = 0$) then the required result again holds. Since from any subsequence (n_k) we can always select a further subsequence along which either $k(n, \alpha) \to \infty$ or $k(n, \alpha)$ is bounded as $n \to \infty$, the result is true generally.

Applying this result to $\varepsilon_n^K(t)$ we see that

$$\limsup_{n\to\infty} \sup_{t \in [\alpha, \beta)} \varepsilon_n^K(t) \leq 2 \mathscr{E} \| X(\alpha) \| \{ y(\beta-) - y(\alpha) \} + 4\varepsilon \leq 2\varepsilon \mathscr{E} \| X \| + 4\varepsilon$$

and hence, again applying the SLLN, (and treating $t = 1$ separately),

$$\limsup_{n\to\infty} \sup_{t \in [0,1]} \left\| \frac{1}{n} \sum_{i=1}^{n} \{ X_i(t) - \mathscr{E} X(t) \} y_i^{(n)}(t) \right\|$$

$$\leq 2\varepsilon \mathscr{E} \| X \| + 4\varepsilon + 2 \mathscr{E} [\| X \| I\{X \notin K\}].$$

Since ε and K were arbitrary, by (iii) the theorem has been proved. □

COROLLARY III.2. *Suppose that for each n, $X_1^{(n)}$, $\cdots$, $X_n^{(n)}$ are* i.i.d. *elements of $D_E[0, 1]$ with the same distribution as X; suppose that $\mathscr{E} \| X \| < \infty$. Suppose also that for each n, $T_1^{(n)}$, $\cdots$, $T_n^{(n)}$ are independent censoring times in $[0, 1]$, independent also of the $X_i^{(n)}$'s; suppose that their average distribution function*

$$\frac{1}{n} \sum_{i=1}^{n} \mathscr{P}\{T_i^{(n)} \leq t\}$$

converges uniformly in t to some distribution function y. Then

$$\left\| \frac{1}{n} \sum_{i=1}^{n} X_i^{(n)} I_{[T_i^{(n)}, 1]} - \mathscr{E} X y \right\| \to_{\mathscr{P}} 0 \quad \text{as} \quad n \to \infty.$$

PROOF. By Van Zuijlen (1978, Theorem 2.1 and Corollary 3.1) we have $(1/n) \sum_{i=1}^{n} I_{[T_i^{(n)}, 1]}$ converges in the supremum norm to y, in probability, as $n \to \infty$. Thus by a Skorohod-Dudley construction (see Wichura, 1970) we can construct a new probability space on which are defined $\tilde{T}_i^{(n)}$, $i = 1, \cdots, n$, $n = 1, 2, \cdots$, and $\tilde{X}_1, \tilde{X}_2, \cdots$ such that

(i) $\tilde{X}_1, \tilde{X}_2, \cdots$ is independent of $\tilde{T}_i^{(n)}$, $i = 1, \cdots, n$, $n = 1, 2, \cdots$;

(ii) $\tilde{T}_1^{(n)} \geq \tilde{T}_2^{(n)} \geq \cdots \geq \tilde{T}_n^{(n)}$ almost surely;

(iii) $\frac{1}{n} \sum_{i=1}^{n} I_{[\tilde{T}_i^{(n)}, 1]}$ converges almost surely uniformly to y; and

(iv) $\frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i I_{[\tilde{T}_i^{(n)}, 1]} =_{\mathscr{D}} \frac{1}{n} \sum_{i=1}^{n} X_i^{(n)} I_{[T_i^{(n)}, 1]}$.

On this new probability space we can apply Theorem III.1 (or rather its proof, since in the theorem the $T_i^{(n)}$'s were supposed to be nonrandom; however this played no part in the

proof), to give

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i I_{[\tilde{T}_i^{(n)},1]} - \mathscr{E}Xy \right\| \to 0 \text{ almost surely.}$$

Therefore we also have

$$\left\| \frac{1}{n} \sum_{i=1}^{n} X_i^{(n)} I_{[T_i^{(n)},1]} - \mathscr{E}Xy \right\| \to 0 \text{ in probability.} \qquad \square$$

REFERENCES

AALEN, O. O. (1977). Weak convergence of stochastic integrals related to counting processes. *Z. Wahrsch. verw. Gebiete* **38** 261–277. Correction: 1979, vol. **48** 347.

AALEN, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

AALEN, O. O. (1980). A model for nonparametric regression analysis of counting processes. *Springer Lect. Notes in Statist.* **2** 1–25.

ANDERSEN, P. K. (1982). Testing goodness of fit of Cox's regression and life model. *Biometrics* **38** 67–77.

ANDERSEN, P. K. and RASMUSSEN, N. K. (1982). Admissions to psychiatric hospitals among women giving birth and women having induced abortions, a statistical analysis of a counting process model. Research Report from Statistical Research Unit, Danish Medical and Social Science Research Councils.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1982). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data (with discussion). *Int. Statist. Rev.* **50** (to appear).

BAILEY, K. R. (1979). The general maximum likelihood approach to the Cox regression model. Ph.D. dissertation, University of Chicago, Chicago, Illinois.

BARTLE, R. G. (1966). *The Elements of Integration.* Wiley, New York.

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

BRESLOW, N. (1972). Contribution to the discussion of the paper by D.R. Cox. *J. Roy. Statist. Soc. B.* **34** 187–220.

BRESLOW, N. (1974). Covariance analysis of censored survival data. *Biometrics* **30** 89–99.

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

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

DOLIVO, F. G. (1974) Counting processes and integrated conditional rates: a martingale approach with application to detection theory. Ph.D. thesis, University of Michigan.

GILL, R. D. (1980). Censoring and stochastic integrals. *Mathematical Centre Tracts 124.* Mathematisch Centrum Amsterdam.

JOHANSEN, S. (1983). An extension of Cox's regression model. *Int. Statist. Rev.* **51** (to appear).

LENGLART, E. (1977). Relation de Domination entre deux Processus. *Ann. Inst. H. Poincaré* **13** 171–179.

LINK, C. L. (1979). Confidence intervals for the survival function using Cox's proportional hazard model with covariates. Tech. Rep. **45**, Division of Biostatistics, Stanford University.

LIU, P. Y. and CROWLEY, J. (1978). Large sample theory of the m.l.e. based on Cox's regression model for survival data. Tech. Rep. **1,** Wisconsin Clinical Cancer Centre, Biostatistics, University of Wisconsin-Madison.

LUSTBADER, E. D. (1980). Time-dependent covariates in survival analysis. *Biometrika* **67** 697–698.

MOURIER, E. (1953). Eléments aléatoires dans un espace de Banach. *Ann. Inst. H. Poincaré* **13** 161–244.

NÆS, T. (1982). The asymptotic distribution of the estimator for the regression parameter in Cox's regression model. *Scand. J. Statist.* **9** 107–115.

OAKES, D. (1981). Survival times: aspects of partial likelihood (with discussion). *Int. Statist. Rev.* **49** 235–264.

RAO, C. R. (1973). *Linear statistical inference and and its applications,* (2nd ed.). Wiley, New York.

RAO, R. R. (1963). The law of large numbers for D[0, 1]-valued random variables. *Theor. Probab. Appl.* **8** 70–74.

REBOLLEDO, R. (1978). Sur les applications de la théorie des martingales a l'étude statistique d'une famille de processus ponctuels. *Springer Lect. Notes in Mathematics* **636** 27–70.

REBOLLEDO, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. verw. Gebiete* **51** 269–286.

ROCKAFELLAR, R. T. (1970). *Convex analysis.* Princeton University Press, Princeton.

SEN, P. K. (1981). The Cox regression model, invariance principles for some induced quantile processes and some repeated significance tests. *Ann. Statist.* **9** 109–121.

TSIATIS, A. A. (1978a). A heuristic estimate of the asymptotic variance of the survival probability in Cox's regression model. Tech. Rep. **524**, Dept. of Stat., University of Wisconsin-Madison.

TSIATIS, A. A. (1978b). A large sample study of the estimate for the integrated hazard function in Cox's regression model for survival data. Tech Rep. **526**, Dept. of Stat., University of Wisconsin-Madison.

TSIATIS, A. A. (1981a). A large sample study of Cox's regression model. *Ann. Statist.* **9** 93–108.

TSIATIS, A. A. (1981b). The asymptotic distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika* **68** 311–315.

WICHURA, M. J. (1970). On the construction of almost uniformly convergent random variables with given weakly convergent laws. *Ann. Math. Statist.* **41** 284–291.

VAN ZUIJLEN, M. C. A. (1978). Properties of the empirical distribution function for independent nonidentically distributed random variables. *Ann. Probability* **6** 250–266.

STATISTICAL RESEARCH UNIT,
DANISH MEDICAL AND SOCIAL SCIENCE
RESEARCH COUNCILS
5, UNIVERSITETSPARKEN,
DK-2100 COPENHAGEN Ø
DENMARK.

MATHEMATISCH CENTRUM
KRUISLAAN 413 1098 SJ AMSTERDAM
POSTBUS 4079
1009 AB AMSTERDAM
THE NETHERLANDS