The Annals of Statistics
1983, Vol. 11, No. 3, 804–813

ASYMPTOTIC DISTRIBUTION THEORY FOR COX-TYPE REGRESSION MODELS WITH GENERAL RELATIVE RISK FORM[1]

BY ROSS L. PRENTICE AND STEVEN G. SELF

The Fred Hutchinson Cancer Research Center and University of Washington, and the Johns Hopkins University

The theory and application of the Cox (1972) failure time regression model has, almost without exception, assumed an exponential form for the dependence of the hazard function on regression variables. Other regression forms may be more natural or descriptive in some applications. For example, a linear relative risk regression model provides a convenient framework for studying epidemiologic risk factor interactions. This note uses the counting process formulation of Andersen and Gill (1982) to develop asymptotic distribution theory for a class of intensity function regression models in which the usual exponential regression form is relaxed to an arbitrary non-negative twice differentiable form. Some stability and regularity conditions, beyond those of Andersen and Gill, are required to show the consistency of the observed information matrix, which in general need not be positive semidefinite.

1. Introduction. In a recent paper Andersen and Gill (1982), hereafter referred to as AG, apply some powerful results on the asymptotic behavior of stochastic integrals with respect to martingales in order to derive sufficient conditions for the consistency and asymptotic normality of the Cox (1972, 1975) maximum partial likelihood estimator. Their results are general enough to allow locally bounded stochastic time-dependent covariates and to include certain multivariate failure time problems.

In the notation of AG, $(\Omega, \mathscr{F}, P)$ is a complete probability space and $\{\mathscr{F}_t, t \in [0, 1]\}$ is an increasing right-continuous family of sub σ-algebras of $\mathscr{F}$ that includes failure time and covariate histories to time t, and censoring histories to t^+. The multivariate counting process $N = (N_1, \cdots, N_n)$ is such that N_i counts failures on the ith subject at times $t \in [0, 1]$ at which the subject is under observation. N_i is required to have totally inaccessible jump times and $N_i(1)$ is almost surely finite. The censoring process $Y = (Y_1, \cdots, Y_n)$ is defined so that $Y_i(t) = 1$ if the ith subject is under observation at time t and $Y_i(t) = 0$ otherwise. Most questions of interest concern the relationship between failure rate and the histories of some "basic" covariate process. Let $Z' = (Z_1, \cdots, Z_n)$ denote covariate processes such that $Z_i'(t) = \{Z_{i1}(t), \cdots, Z_{ip}(t)\}$ consists of data-analyst-defined functions of the basic covariate histories or the counting process histories up to time t (or even functions of the censoring process). Of course, N, Y and Z are assumed adapted to $\{\mathscr{F}_t, t \in [0, 1]\}$.

This counting process formulation permits each N_i to be uniquely decomposed into the sum of its cumulative intensity process Λ_i and a local square integrable martingale M_i, so that

$$N_i(t) = \Lambda_i(t) + M_i(t) \tag{1.1}$$

for all (t, i). The increasing process Λ_i is, for convenience, taken to be absolutely continuous, giving

$$\Lambda_i(t) = \int_0^t \lambda_i(u)\, du.$$

Received May 1982; revised April 1983.

[1] This work was supported in part by NIH grants GM-28314 and GM-24472.

AMS 1970 *subject classification.* Primary 62E20; secondary 62G05, 60G15.

Key words and phrases. Censoring, counting process, Cox-regression, failure-time data, intensity process, martingale, partial likelihood, relative risk, time-dependent covariates.

The intensity process $\lambda = (\lambda_1, \cdots, \lambda_n)$ under some regularity (e.g., each λ_i bounded by an integrable random variable, Aalen, 1978) can be written

$$\lambda_i(t) = \lim_{s\uparrow t}\lim_{h\downarrow 0} h^{-1}P[N_i(s+h) - N_i(s) = 1 \mid \mathscr{F}_s]. \tag{1.2}$$

In fact, under usual assumptions concerning the independence of the censoring mechanism and the independence of failure times on distinct study subjects λ_i will have a standard "hazard function" interpretation.

AG model the intensity process, in the manner of Cox (1972), by setting

$$\lambda_i(t) = Y_i(t)\lambda_0(t)\exp\{\beta_0' Z_i(t)\}, \quad i = 1, \cdots, n. \tag{1.3}$$

In order that martingale convergence result apply, Y_i and Z_i are required to have sample paths that are left-continuous with right-hand limits (and so to be predictable and locally bounded). Note that $Z_i(t)$ may involve functions of covariate measurements at time $s < t$ but not at time t itself (see Self and Prentice, 1982 for an elaboration of this point and for a presentation of possible underlying partial likelihood functions).

Under these conditions the martingale property (1.1) is shown by AG to yield general and conceptually simple derivations of the asymptotic distribution theory for the maximum partial likelihood estimator $\hat{\beta}$ and related random variables. The estimator $\hat{\beta}$ is defined as a solution to $\partial \log L(\beta, 1)/\partial\beta = 0$ where

$$\log L(\beta, t) = \sum_{i=1}^n \int_0^t \beta' Z_i(s)\, dN_i(s) - \int_0^t \log[\textstyle\sum_{l=1}^n Y_l(s)\exp\{\beta' Z_l(s)\}]\, d\bar{N}(s), \tag{1.4}$$

and $\bar{N} = N_1 + \cdots + N_n$.

In this note, corresponding asymptotic distribution theory is developed for a more general class of intensity function regression models given by

$$\lambda_i(t) = Y_i(t)\lambda_0(t) r\{\beta_0' Z_i(t)\} \tag{1.5}$$

all (t, i), where $r: \mathscr{R} \to \mathscr{R}$ is some fixed twice continuously differentiable function. An additional condition is necessary to ensure the positivity of $r\{\beta' Z_i(t)\}$ for β in some neighborhood of β_0, for each $i = 1, \cdots, n$. A sufficient requirement is simply $r(w) > 0$, $w \in \mathscr{R}$, but this condition is too strong to permit inclusion of regression model forms of particular interest, such as the linear form $r(\cdot) = 1 + (\cdot)$. A weaker positivity requirement for r is introduced below (Condition F). The partial likelihood argument leading to (1.4) (see Self and Prentice, 1982) is unaffected by this relaxation of intensity model form and yields a maximum partial likelihood estimator $\hat{\beta}$ that is a solution to $\partial \log L(\beta, 1)/\partial\beta = 0$, where, now,

$$\begin{aligned} \log L(\beta, t) = {} & \sum_{i=1}^n \int_0^t \log r\{\beta' Z_i(s)\}\, dN_i(s) \\ & - \int_0^t \log[\textstyle\sum_{l=1}^n Y_l(s) r\{\beta' Z_l(s)\}]\, d\bar{N}(s). \end{aligned} \tag{1.6}$$

A number of authors have noted that an exponential form for the hazard (intensity) function's dependence on covariates may not be the best choice in specific applications. Cox (1981) points to this topic as one meriting careful attention. The linear intensity function model, $r(\cdot) = 1 + (\cdot)$, provides particular motivation for the generalization (1.5). For example, a linear (polynomial) relationship between the incidence rates for certain types of cancer and radiation exposure level is often taken to be a consequence of biological models for the effect of radiation on individual cells. Daffer, et al. (1980) develop asymptotic distribution theory for the special case of a single fixed (time-independent) exposure variable and a polynomial relative risk function having non-negative coefficients. A linear relative risk model also provides a natural framework within which to assess departures from an additive relative risk model when two or more risk factors are studied in relation to the incidence of a disease (e.g., Thomas, 1982; Prentice, et al, 1983). Aalen (1980) considers asymptotic distribution theory for a less flexible model in which the entire intensity process (rather than just the relative risk) is modeled as a linear function of covariates.

2. Asymptotic distribution theory. Asymptotic distribution theory for $\hat{\beta}$ from (1.6) involves the limiting behavior of $\log L(\beta, t) - \log L(\beta_0, t)$ and its first and second derivatives with respect to β and the limiting behavior of the "score" statistic variance matrix. The following processes arise in their consideration:

$$S^{(0)}(\beta, t) = n^{-1} \sum Y_l(t) r\{\beta' Z_l(t)\}$$

$$S^{(1)}(\beta, t) = \partial S^{(0)}(\beta, t)/\partial\beta = n^{-1} \sum Y_l(t) Z_l(t) r^{(1)}\{\beta' Z_l(t)\}$$

$$S^{(2)}(\beta, t) = n^{-1} \sum Y_l(t) Z_l(t)^{\otimes 2} u^{(1)}\{\beta' Z_l(t)\}^2 r\{\beta' Z_l(t)\}$$

$$S^{(3)}(\beta, t) = \partial^2 S^{(0)}(\beta, t)/\partial\beta^2 = n^{-1} \sum Y_l(t) Z_l(t)^{\otimes 2} r^{(2)}\{\beta' Z_l(t)\}$$

$$S^{(4)}(\beta, t) = n^{-1} \sum Y_l(t)[u\{\beta' Z_l(t)\} - u\{\beta_0' Z_l(t)\}] r\{\beta_0' Z_l(t)\}$$

$$S^{(5)}(\beta, t) = \partial S^{(4)}(\beta, t)/\partial\beta = n^{-1} \sum Y_l(t) Z_l(t) u^{(1)}\{\beta' Z_l(t)\} r\{\beta_0' Z_l(t)\}$$

$$S^{(6)}(\beta, t) = \partial^2 S^{(4)}(\beta, t)/\partial\beta^2 = n^{-1} \sum Y_l(t) Z_l(t)^{\otimes 2} u^{(2)}\{\beta' Z_l(t)\} r\{\beta_0' Z_l(t)\}.$$

In these expressions all summations are over $l = 1, \cdots, n$, for any vector $a' = (a_1, \cdots, a_p)$, $a^{\otimes 2}$ denotes the $p \times p$ matrix with (i, j) entry $a_i a_j$, $r^{(1)}(x) = dr(x)/dx$, $r^{(2)}(x) = dr^{(1)}(x)/dx$, $u(x) = \log r(x)$, $u^{(1)}(x) = du(x)/dx = r^{(1)}(x)/r(x)$, and $u^{(2)}(x) = du^{(1)}(x)/dx = r^{(2)}(x)/r(x) - \{r^{(1)}(x)/r(x)\}^2$. Note that $S^{(5)}(\beta_0, t) = S^{(1)}(\beta_0, t)$ and $S^{(6)}(\beta_0, t) = S^{(3)}(\beta_0, t) - S^{(2)}(\beta_0, t)$. Note also that under the exponential regression form, given by $r(\cdot) = \exp(\cdot)$, one has $S^{(3)}(\beta, t) = S^{(2)}(\beta, t)$, $S^{(4)}(\beta, t) = (\beta - \beta_0) S^{(1)}(\beta_0, t)$, $S^{(5)}(\beta, t) \equiv S^{(1)}(\beta_0, t)$ and $S^{(6)}(\beta, t) \equiv 0$, while under the linear form, $r(\cdot) = 1 + (\cdot)$, $S^{(3)}(\beta, t) \equiv 0$.

Also set

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

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

Note that $E(\beta_0, t)$ and $V(\beta_0, t)$ can be thought of as the expected covariate vector at time t and corresponding covariance matrix for a study subject failing at t, given $\mathscr{F}_{t^-}$ and given the fact that a failure occurs at t. Note also that

$$V(\beta, t) = \sum Y_l(t) r\{\beta' Z_l(t)\} S^{(0)}(\beta, t)^{-1} [Z_l(t) u^{(1)}\{\beta' Z_l(t)\} - E(\beta, t)]^{\otimes 2}$$

so that $V(\beta, t)$ is positive semidefinite.

The following conditions for the asymptotic normality of $n^{1/2}(\hat{\beta} - \beta_0)$ generalize those given in AG:

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

B. (Asymptotic stability). There exists a neighborhood $\mathscr{B}$ of β_0 ($\mathscr{B} \subset \mathscr{B}_0$ as defined below) and functions $s^{(0)}, \cdots, s^{(6)}$ defined on $\mathscr{B} \times [0, 1]$ such that for $j = 0, \cdots, 6$

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

C. (Lindeberg condition)

$$\int_0^1 n^{-1} \sum_{l=1}^n [Z_{li}(t) u^{(1)}\{\beta_0' Z_l(t)\} - E(\beta_0, t)_i]^2 Y_l(t) r\{\beta_0' Z_l(t)\}$$

$$I\{n^{-1/2} | Z_{li}(t) u^{(1)}\{\beta_0' Z_l(t)\} - E(\beta_0, t)_i | > \varepsilon\} \lambda_0(t)\, dt \to_P 0$$

for any $\varepsilon > 0$, $i = 1, \cdots, p$.

D. (Asymptotic regularity conditions). Set $e = s^{(1)}/s^{(0)}$ and $v = s^{(2)}/s^{(0)} - e^{\otimes 2}$. For each $j = 0, \cdots, 6$, $s^{(j)}(\cdot, t)$ are continuous functions of $\beta \in \mathscr{B}$, uniformly in $t \in [0, 1]$. Also, $s^{(j)}$, $j = 0, \cdots, 6$ are bounded on $\mathscr{B} \times [0, 1]$, $s^{(0)}$ is bounded away from zero and the matrix

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

is positive definite. Also, $s^{(0)}(\beta, t)$ and $s^{(4)}(\beta, t)$ are assumed to be twice differentiable with respect to β on $\mathscr{B} \times [0, 1]$.

E. (Asymptotic stability of observed information matrix).

$$\sup_{\beta \in \mathscr{B}} \int_0^1 n^{-2} \textstyle\sum_{i=1}^n \| Z_i(t) \|^4 u^{(2)} \{\beta' Z_i(t)\}^2 Y_i(t) r\{\beta_0' Z_i(t)\} \lambda_0(t)\, dt \to_P 0.$$

F. (Regression function positivity). There exists a neighborhood $\mathscr{B}_0$ of β_0 such that, for $\beta \in \mathscr{B}_0$, $r\{\beta' Z_i\}$ is locally bounded away from zero for all $i = 1, \cdots, n$.

In these conditions $\| \cdot \|$ refers to the supremum norm and convergence properties involve $n \to \infty$. In the special case $r(\cdot) = \exp(\cdot)$, these conditions reduce precisely to those given by AG with the exception of the Lindeberg condition C. As it stands C is precisely the condition required to apply the Rebolledo central limit theorem in the form given in Appendix I of AG. It can be interpreted as requiring the contribution from $[t, t + dt)$ to the variance of standardized score statistic at β_0 to be asymptotically trivial for any $t \in [0, 1]$. Condition C can be simplified slightly as in AG (page 1107). Upon applying their inequality and using conditions A and B, condition C is implied by

$$\int_0^1 n^{-1} \textstyle\sum_{l=1}^n Y_l(t) | Z_{li}(t) u^{(1)} \{\beta_0' Z_l(t)\} |^2 r\{\beta_0' Z_l(t)\} I\{n^{-1/2} | Z_{li}(t) u^{(1)} \{\beta_0' Z_l(t)\} | > \varepsilon\} \lambda_0(t)\, dt \to_p 0 \qquad (i = 1, \cdots, p)$$

which, in view of condition A and condition B in respect to $S^{(2)}$, is implied by:

C′. (alternate Lindeberg condition)

$$n^{-1/2} \sup_{i,l,t} | Z_{li}(t) u^{(1)} \{\beta_0' Z_l(t)\} | \to_p 0$$

In view of the n^{-2} factor, condition E is rather weak. It ensures that the variance of the observed information matrix converges to a zero matrix. As with C, it could be replaced by a similar, but more restrictive, condition in which the supremum of the integrand over $t \in [0, 1]$ converges in probability to zero. Condition E is vacuous in the special case $r(\cdot) = \exp(\cdot)$. Condition F is required for such processes as $\log r\{\beta' Z_i\}$ and $r^{-1}\{\beta' Z_i\}$ be locally bounded for $\beta \in \mathscr{B}$, $i = 1, \cdots, n$. Condition F will trivially hold if $r(w) > 0$ for all $w \in \mathscr{R}$. Note that it follows easily from B and D that $s^{(1)}(\beta, t) = \partial s^{(0)}(\beta, t)/\partial \beta$, $s^{(3)}(\beta, t) = \partial^2 s^{(0)}(\beta, t)/\partial \beta^2$, $s^{(5)}(\beta, t) = \partial s^{(4)}(\beta, t)/\partial \beta$, $s^{(6)}(\beta, t) = \partial^2 s^{(4)}(\beta, t)/\partial \beta^2$, $s^{(5)}(\beta_0, t) = s^{(1)}(\beta_0, t)$ and $s^{(6)}(\beta_0, t) = s^{(3)}(\beta_0, t) - s^{(2)}(\beta_0, t)$ for all $t \in [0, 1]$.

LEMMA 2.1. (*Consistency of* $\hat{\beta}$) $\hat{\beta} \to_P \beta_0$.

PROOF. For $\beta \in \mathscr{B}$ set

$$X(\beta, t) = n^{-1}\{\log L(\beta, t) - \log L(\beta_0, t)\}$$

and define

$$A(\beta, t) = \int_0^t [S^{(4)}(\beta, w) - \log\{S^{(0)}(\beta, w)/S^{(0)}(\beta_0, w)\} S^{(0)}(\beta_0, w)] \lambda_0(w)\, dw.$$

Using (1.1), (1.5) and (1.6) one can write

$$X(\beta, t) - A(\beta, t) = \int_0^t n^{-1} \sum_{i=1}^n \log[r\{\beta' Z_i(w)\}/r\{\beta_0' Z_i(w)\}]\, dM_i(w)$$

$$- \int_0^t \log\{S^{(0)}(\beta, w)/S^{(0)}(\beta_0, w)\}\, d\bar{M}(w),$$

where $\bar{M} = M_1 + \cdots + M_n$. The predictableness of each Z_i, the continuity of r and condition F ensures $\log r\{\beta' Z_i\}$ and $\log S^{(0)}(\beta, \cdot)$ to be predictable and locally bounded for each $\beta \in \mathscr{B}_0$. It follows that $X(\beta, \cdot) - A(\beta, \cdot)$ is a local square integrable martingale with variance process $B(\beta, \cdot)$ given by

$$B(\beta, t) = \int_0^t n^{-2} \sum_{i=1}^n [u\{\beta' Z_i(w)\} - u\{\beta_0' Z_i(w)\} - \log\{S^{(0)}(\beta, w)/S^{(0)}(\beta_0, w)\}]^2 \lambda_i(w)\, dw.$$

We would like to show that $B(\beta, 1) \to_P 0$ so that $X(\beta, 1)$ and $A(\beta, 1)$ will be shown to converge in probability to the same limit. Expanding the squared term in the above expression and substituting from (1.5) gives

$$B(\beta, 1) = \int_0^1 n^{-2} \sum_{i=1}^n Y_i(w)[u\{\beta' Z_i(w)\} - u\{\beta_0' Z_i(w)\}]^2 r\{\beta_0' Z_i(w)\}\lambda_0(w)\, dw$$

$$-2 \int_0^1 n^{-2} \sum_{i=1}^n Y_i(w)[u\{\beta' Z_i(w)\} - u\{\beta_0' Z_i(w)\}]$$

$$\cdot \log\{S^{(0)}(\beta, w)/S^{(0)}(\beta_0, w)\} r\{\beta_0' Z_i(w)\}\lambda_0(w)\, dw$$

$$+ n^{-1} \int_0^1 \log^2\{S^{(0)}(\beta, w)/S^{(0)}(\beta_0, w)\} S^{(0)}(\beta_0, w)\lambda_0(w)\, dw.$$

The final integral in this expression converges in probability to zero in view of conditions B and D on $S^{(0)}$ along with the finite interval condition A. The middle integral will converge to zero if the first does upon applying the Schwarz inequality and using the convergence just noted for the third integral. It remains to show that the first integral in the expression for $B(\beta, 1)$ converges to zero.

A Taylor expansion about β_0 gives

$$u\{\beta' Z_i(w)\} - u\{\beta_0' Z_i(w)\}$$

$$= (\beta - \beta_0)' Z_i(w) u^{(1)}\{\beta_0' Z_i(w)\} + \tfrac{1}{2}(\beta - \beta_0)' Z_i(w)^{\otimes 2} u^{(2)}\{\beta_*' Z_i(w)\}(\beta - \beta_0),$$

where $\beta_* = \beta_*(w)$ is between β and β_0. Upon substitution the first integral in $B(\beta, 1)$ equals

$$(\beta - \beta_0)'\left[\int_0^1 n^{-2} \sum_{i=1}^n Y_i(w) Z_i(w)^{\otimes 2} u^{(1)}\{\beta_0' Z_i(w)\}^2 r\{\beta_0' Z_i(w)\}\lambda_0(w)\, dw\right](\beta - \beta_0)$$

$$+ (\beta - \beta_0)'\left[\int_0^1 n^{-2} \sum_{i=1}^n Y_i(w) Z_i(w) u^{(1)}\{\beta_0' Z_i(w)\}\right.$$

$$\left. \cdot (\beta - \beta_0)' Z_i(w)^{\otimes 2} u^{(2)}\{\beta_*' Z_i(w)\} r\{\beta_0' Z_i(w)\}\lambda_0(w)\, dw\right](\beta - \beta_0)$$

$$\tfrac{1}{4}(\beta - \beta_0)'\left[\int_0^1 n^{-2} \sum_{i=1}^n Y_i(w) Z_i(w)^{\otimes 2} (\beta - \beta_0)(\beta - \beta_0)' Z_i(w)^{\otimes 2} u^{(2)}\{\beta_*' Z_i(w)\}^2\right.$$

$$\left. \cdot r\{\beta_0' Z_i(w)\}\lambda_0(w)\, dw\right](\beta - \beta_0).$$

The first term in this expression converges in probability to zero on the basis of condition A and conditions B and D in respect to $S^{(5)}$. The middle term will converge in probability to zero if the final term does, again on the basis of the Schwarz inequality. The final term is equal to or less than

$$\tfrac{1}{4}\int_0^1 n^{-2}\textstyle\sum_{i=1}^n Y_i(w)p^4\|\beta-\beta_0\|^4\|Z_i(w)\|^4 u^{(2)}\{\beta_*' Z_i(w)\}^2 r\{\beta_0' Z_i(w)\}\,\lambda_0(w)\,dw,$$

which converges in probability to zero on the basis of E. It now follows that $B(\beta, 1) \to_P 0$ so that by an inequality of Lenglart (Appendix 1 in AG) $X(\beta, 1)$ converges in probability to the same limit as does $A(\beta, 1)$, for each $\beta \in \mathscr{B}$.

From conditions A, B and D

$$A(\beta, 1) \to_P \int_0^1 [s^{(4)}(\beta, w) - \log\{s^{(0)}(\beta, w)/s^{(0)}(\beta_0, w)\}]s^{(0)}(\beta_0, w)\lambda_0(w)\,dw.$$

Following the arguments of AG, the boundedness conditions in D now allow one to take first and second derivatives of $A(\beta, 1)$ with respect to β by differentiating under the integral sign. The first derivative vector is, therefore, equal to

$$\int_0^1 [s^{(5)}(\beta, w) - \{s^{(1)}(\beta, w)/s^{(0)}(\beta, w)\}s^{(0)}(\beta_0, w)]\lambda_0(w)\,dw$$

which is equal to zero at $\beta = \beta_0$ since $s^{(5)}(\beta_0, w) = s^{(1)}(\beta_0, w)$. The negative of the second derivative matrix can be written

$$-\int_0^1 \left[s^{(6)}(\beta, w) - \left\{\frac{s^{(3)}(\beta, w)}{s^{(0)}(\beta, w)} - \frac{s^{(1)}(\beta, w)^{\otimes 2}}{s^{(0)}(\beta, w)^2}\right\} s^{(0)}(\beta_0, w)\right]\lambda_0(w)\,dw, \tag{2.1}$$

which at $\beta = \beta_0$ equals

$$\int_0^1 \left[\frac{s^{(2)}(\beta_0, w)}{s^{(0)}(\beta_0, w)} - \frac{s^{(1)}(\beta_0, w)^{\otimes 2}}{s^{(0)}(\beta_0, w)^2}\right] s^{(0)}(\beta_0, w)\lambda_0(w)\,dw = \Sigma$$

(by virtue of $s^{(6)}(\beta_0, w) = s^{(3)}(\beta_0, w) - s^{(2)}(\beta_0, w)$) which is positive definite by condition D. The continuity conditions D on $s^{(0)}$, $s^{(1)}$, $s^{(3)}$ and $s^{(6)}$ for $\beta \in \mathscr{B}$, uniform in $t \in [0, 1]$, condition A and the positive definiteness of (2.1) at $\beta = \beta_0$ imply the existence of a neighborhood $\mathscr{B}_1 \subset \mathscr{B}$ of β_0 such that (2.1) is positive definite for $\beta \in \mathscr{B}_1$. It follows that for $\beta \in \mathscr{B}_1$, $X(\beta, 1)$ converges in probability to a concave function of β with unique maximum at $\beta = \beta_0$. Since $\hat\beta$ maximizes $X(\beta, 1)$ one can argue exactly as in Andersen and Gill, Appendix 2 to show $\hat\beta \to_P \beta_0$.

THEOREM 2.1. (*Asymptotic normality of* $\hat\beta$).

$$n^{1/2}(\hat\beta - \beta_0) \to_D N(0, \Sigma^{-1}).$$

PROOF. A Taylor expansion of $\partial \log L(\beta, 1)/\partial\beta$ about β_0, evaluated at $\hat\beta$, gives

$$n^{-1/2}\partial \log L(\beta_0, 1)/\partial\beta_0 = \{-n^{-1}\partial^2 \log L(\beta_*)/\partial\beta_*^2\}n^{1/2}(\hat\beta - \beta_0), \tag{2.2}$$

where β_* is between $\hat\beta$ and β_0, whence it is sufficient to show

$$n^{-1/2}\partial \log L(\beta_0, 1)/\partial\beta_0 \to_D N(0, \Sigma)$$

and

$$-n^{-1}\partial^2 \log L(\beta_*, 1)/\partial\beta_*^2 \to_P \Sigma$$

for any random β_* such that $\beta_* \to_P \beta_0$.

Using (1.1) one can write

$$n^{-1/2}\partial \log L(\beta_0, t)/\partial\beta_0 = \int_0^t \textstyle\sum_{i=1}^n n^{-1/2}[Z_i(w)u^{(1)}\{\beta_0' Z_i(w)\} - E(\beta_0, w)]\,dM_i(w).$$

Upon setting $H_i(t) = n^{-1/2}[Z_i(t)u^{(1)}\{\beta_0'Z_i(t)\} - E(\beta_0, t)]$, $i = 1, \cdots, n$ it is only necessary to note that H_i is predictable and locally bounded (using condition F, the local boundedness of Z_i and the continuity of $u^{(1)}$) and that

$$\int_0^t \textstyle\sum_{i=1}^n H_i(w)^{\otimes 2}\lambda_i(w)\, dw = \int_0^t [S^{(2)}(\beta_0, w) - S^{(1)}(\beta_0, w)^{\otimes 2}]\lambda_0(w)\, dw$$

$$= \int_0^t V(\beta_0, w)S^{(0)}(\beta_0, w)\lambda_0(w)\, dw$$

$$\to_P \int_0^t v(\beta_0, w)s^{(0)}(\beta_0, w)\lambda_0(w)\, dw$$

(using conditions A, B and D) in order to apply the Rebolledo central limit theorem to the local square integrable martingale $n^{-1/2}\partial \log L(\beta_0, t)/\partial\beta_0$ (see Andersen and Gill, Appendix 1). The desired distributional result for $n^{-1/2}\partial \log L(\beta_0, 1)/\partial\beta_0$ then follows upon noting that the limiting variance function $\int_0^t v(\beta_0, w)s^{(0)}(\beta_0, w)\lambda_0(w)\, dw$ evaluated at $t = 1$ is Σ.

Now consider convergence in probability for the observed information matrix. For $\beta \in \mathscr{B}$ one can write

$$- n^{-1}\partial^2\log L(\beta, t)/\partial\beta^2 = \int_0^t n^{-1} \textstyle\sum_{i=1}^n [U(\beta, w) - Z_i(w)^{\otimes 2}u^{(2)}\{\beta' Z_i(w)\}]\, dN_i(w)$$

where $U(\beta, w) = S^{(3)}(\beta, w)/S^{(0)}(\beta, w) - S^{(1)}(\beta, w)^{\otimes 2}/S^{(0)}(\beta, w)^2$. Define

$$C(\beta, t) = \int_i^t n^{-1} \textstyle\sum_{i=1}^n [U(\beta, w) - Z_i(w)^{\otimes 2}u^{(2)}\{\beta' Z_i(w)\}]\lambda_i(w)\, dw.$$

Then $- \partial^2\log L(\beta,\cdot)/\partial\beta^2 - C(\beta, \cdot)$ is a $(p \times p)$ local square integrable martingale (the continuity of $r^{(2)}$ is required here) with variance process $D(\beta, \cdot)$ given by

$$D(\beta, t) = \int_0^t n^{-2} \textstyle\sum_{i=1}^n [U(\beta, w) - Z_i(w)^{\otimes 2}u^{(2)}\{\beta' Z_i(w)\}]^2\lambda_i(w)\, dw$$

$$= \int_0^t n^{-1}U(\beta, w)^{\otimes 2}S^{(0)}(\beta_0, w)\lambda_0(w)\, dw - 2\int_0^t n^{-1}U(\beta, w)S^{(6)}(\beta, w)\lambda_0(w)\, dw$$

$$+ \int_0^t n^{-2} \textstyle\sum_{i=1}^n Z_i(w)^{\otimes 2}Z_i(w)^{\otimes 2}u^{(2)}\{\beta' Z_i(w)\}^2 Y_i(w) r\{\beta_0' Z_i(w)\}\lambda_0(w)\, dw.$$

In the expression for $D(\beta, 1)$ the first two integrals converge to a zero matrix by virtue of the stability, regularity and boundedness conditions B and D and condition A while the final integral converges to a zero matrix on the basis of condition E. It follows that $D(\beta, 1)$ converges in probability to a zero matrix so that $-n^{-1}\partial^2\log L(\beta, 1)/\partial\beta^2$ and $C(\beta, 1)$ converge in probability to the same matrix for $\beta \in \mathscr{B}$. By conditions A, B and D

$$C(\beta, 1) \to_P \int_0^1 \{s^{(3)}(\beta, w)s^{(0)}(\beta_0, w)/s^{(0)}(\beta, w)$$

$$- s^{(1)}(\beta, w)^{\otimes 2}s^{(0)}(\beta_0, w)/s^{(0)}(\beta, w)^2 - s^{(6)}(\beta, w)\}\lambda_0(w)\, dw.$$

Hence $\beta_* \to_P \beta_0$ together with the continuity in β of $s^{(0)}$, $s^{(1)}$, $s^{(3)}$ and $s^{(6)}$ uniform in t (condition D) along with condition A imply

$$C(\beta_*, 1) \to_P \int_0^1 \{s^{(3)}(\beta_0, w) - s^{(1)}(\beta_0, w)^{\otimes 2}/s^{(0)}(\beta_0, w) - s^{(6)}(\beta_0, w)\}\lambda_0(w)\, dw = \Sigma,$$

since $s^{(6)}(\beta_0, w) = s^{(3)}(\beta_0, w) - s^{(2)}(\beta_0, w)$, concluding the proof.

In order to apply Theorem 2.1 it will usually be necessary to insert an estimator for Σ. The last part of the above proof shows $-\partial^2 \log L(\hat{\beta}, 1)/\partial\hat{\beta}^2$ to be a consistent estimator, but this estimator need not even be positive semidefinite. In view of the definition of Σ (see condition D) we may instead consider $\hat{\Sigma}(\hat{\beta})$ as a variance estimator, where

$$\hat{\Sigma}(\beta) = n^{-1} \int_0^1 V(\beta, t)\, d\bar{N}(t).$$

As noted above $\hat{\Sigma}(\beta)$ is positive semidefinite. A very simple martingale argument, of the type given above for $-n^{-1}\partial^2 \log L(\hat{\beta}, 1)/\partial\hat{\beta}^2$, shows $\hat{\Sigma}(\hat{\beta})$ to be a consistent estimator of Σ; in fact, it is only this result that requires the asymptotic stability and regularity conditions (B and D) on $S^{(2)}$. It is interesting to note that the variance estimator $\hat{\Sigma}(\beta_0)$ can be obtained by applying a finite population variance argument to the contributions at each failure time to the score statistic $n^{-1/2}\partial \log L(\beta_0, 1)/\partial\beta_0$. Under $\beta_0 = 0$ the score statistic standardized by $\hat{\Sigma}(0)$ is precisely equal to that from the exponential special case $\{r(\cdot) = \exp(\cdot)\}$ for any intensity function form r. Note that if $\beta = 0$ it will be necessary that $r'(0) \neq 0$ in order that Σ can be positive definite as is assumed in condition D.

Consider now estimation of the cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)\, du$, $t \in [0, 1]$. A natural candidate for such estimation is

$$\hat{\Lambda}(t) = \int_0^t \left(\textstyle\sum_{i=1}^n Y_i(w) r\{\hat{\beta}' Z_i(w)\}\right)^{-1} d\bar{N}(w).$$

THEOREM 2.2. (*Weak convergence of* $n^{1/2}(\hat{\Lambda} - \Lambda_0)$). $n^{1/2}(\hat{\beta} - \beta_0)$ *and the process given at* $t \in [0, 1]$ *by*

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

are asymptotically independent, the latter converging in distribution to a Gaussian martingale with variance function

$$\int_0^t \{s^{(0)}(\beta_0, w)\}^{-1}\lambda_0(w)\, dw.$$

PROOF. The proof goes through in a virtually identical manner to that given in AG for the special case $r(\cdot) = \exp(\cdot)$.

From Theorem 2.2 it follows, for example, that at any $t \in [0, 1]$, $n^{1/2}\{\hat{\Lambda}(t) - \Lambda_0(t)\}$ is asymptotically normal with mean 0 and variance

$$\int_0^t \{s^{(0)}(\beta_0, w)\}^{-1}\lambda_0(w)\, dw + \left(\int_0^t e(\beta_0, w)\lambda_0(w)\, dw\right)' \Sigma \left(\int_0^t e(\beta_0, w)\lambda_0(w)\, dw\right)$$

which is readily shown to be consistently estimated by

$$n^{-1} \int_0^t \{S^{(0)}(\hat{\beta}, w)\}^{-2}\, d\bar{N}(w) + \left(n^{-1} \int_0^t \frac{S^{(1)}(\hat{\beta}, w)}{S^{(0)}(\hat{\beta}, w)^2}\, d\bar{N}(w)\right)^1 \hat{\Sigma}(\hat{\beta}) \left(n^{-1} \int_0^t \frac{S^{(1)}(\hat{\beta}, w)}{S^{(0)}(\hat{\beta}, w)^2}\, d\bar{N}(w)\right).$$

3. Discussion. A referee has suggested that Conditions A to F be considered in the context of some simple failure time data problem with a linear intensity process. For this purpose consider a non-negative scalar covariate Z with monotone increasing sample paths and suppose $r(\cdot) = 1 + (\cdot)$ and that $\beta_0 > 0$. Suppose also that the observations (N_i, Y_i, Z_i)

are i.i.d. For example, such a model may describe the relationship between a cancer incidence rate function and the cumulative exposure to a carcinogenic substance. Conditions A through F are implied by

$$\int_0^1 \lambda_0(t)\, dt < \infty, \quad P\{Y(t) = 1\ \forall\ t \in [0, 1]\} > 0, \quad E\{Z(T)^4\} < \infty$$

and

$$\int_0^1 \operatorname{Var}\{Z(t) \mid Y(t) = 1\}\lambda_0(t)\, dt > 0,$$

where $T = \sup\{t: Y(t) = 1\}$. Conditions B, D and E may be verified using the same arguments as in AG Theorem 4.1 in conjunction with the monotonicity of the sample paths of Z. The positivity of Σ follows by noting that $V(\beta_0, t)$ may be bounded below by

$$E\left\{ Y(t) \left(Z(t) - \frac{E[Y(t)Z(t)r\{\beta_0 Z(t)\}]}{S^{(0)}(\beta_0, t)} \right)^2 \frac{r\{\beta_0 Z(t)\}}{S^{(0)}(\beta_0, t)} \right\}$$

which is positive if $\operatorname{Var}\{Z(t) \mid Y(t) = 1\} > 0$. To verify Condition C it is sufficient to show

$$\int_0^1 n^{-1} \sum_{i=1}^n Y_i(t) \frac{S^{(1)}(\beta_0, t)}{S^{(0)}(\beta_0, t)} I\left\{ n^{-1/2} \frac{S^{(1)}(\beta_0, t)}{S^{(0)}(\beta_0, t)} > \varepsilon \right\} \lambda_0(t)\, dt \to_P 0$$

and

$$\int_0^1 n^{-1} \sum_{i=1}^n Y_i(t) \frac{Z_i^2(t)}{1 + \beta_0 Z_i(t)} I\left\{ n^{-1/2} \frac{Z_i(t)}{1 + \beta_0 Z_i(t)} > \varepsilon \right\} \lambda_0(t)\, dt \to_p 0.$$

The first limit follows immediately from convergence of $S^{(1)}(\beta_0, \cdot)$ and $S^{(0)}(\beta_0, \cdot)$ and the finite interval assumption. The second limit follows similarly from moment conditions on Z and the assumption $\beta_0 > 0$. Finally, Condition F follows immediately from range restrictions on β_0 and $Z(t)$.

Careful attention may be required in covariate modeling, to near violations of the relative risk positivity Condition F. Such near violations may cause data on individual study subjects to substantially impact the parameter space and the likelihood function. In such circumstances enormous sample sizes may be required for the asymptotic distribution theory to apply. For example, consider a single fixed covariate $Z(t) = z$ and linear relative risk function $r(z) = 1 + \beta z$ with $\beta_0 < 0$. Only z values less than $-\beta_0^{-1}$ are consistent with the model while a value near $-\beta_0^{-1}$ may very substantially affect the likelihood function shape. In such circumstances covariate transformations or a change of model form $r(\cdot)$ would usually be appropriate. Corresponding to specific choices of $r(\cdot)$ it would be useful to study parameter transformations, $\theta = g(\beta)$, that will reduce the sample sizes necessary to suitably apply the above distribution theory.

REFERENCES

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 analysis of counting processes. In *Lecture Notes in Statistics* 2. Springer-Verlag, New York.

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

COX, D. R. (1972). Regression models with life tables (with discussion). *J. Roy. Statist. Soc.* **B34** 187–220.

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

COX, D. R. (1981). Discussion of paper by D. Oakes, entitled "Survival Times: Aspects of Partial Likelihood". *Internat. Statist. Rev.* **49** 235–264.

DAFFER, P. Z., CRUMP, K. S. and MASTERMAN, M. D. (1980). Asymptotic theory for analyzing dose-response survival data with application to low-dose extrapolation problem. *Math. Biosciences* **50** 207–230.

PRENTICE, R. L., YOSHIMOTO, Y. and MASON, M. (1983). Cigarette smoking and radiation exposure in relation to cancer mortality in Hiroshima and Nagasaki. *J. National Cancer Institute* **70** 611–622.

SELF, S. G. and PRENTICE, R. L. (1982). Commentary on Andersen and Gill's distribution theory for the Cox regression model. *Ann. Statist.* **10** 1121–1124.

THOMAS, D. C. (1982). General relative-risk models for survival time and matched case-control analysis. *Biometrics* **37** 673–686.

DIVISION OF PUBLIC HEALTH SCIENCES
THE FRED HUTCHINSON CANCER RESEARCH CENTER
1124 COLUMBIA STREET
SEATTLE, WASHINGTON 98104

DEPARTMENT OF BIOSTATISTICS
SCHOOL OF HYGIENE AND PUBLIC HEALTH
JOHNS HOPKINS UNIVERSITY
615 N. WOLFE STREET
BALTIMORE, MARYLAND 21205