The Annals of Statistics
1983, Vol. 11, No. 1, 39–48

THE ASYMPTOTIC JOINT DISTRIBUTION OF REGRESSION AND SURVIVAL PARAMETER ESTIMATES IN THE COX REGRESSION MODEL[1]

BY KENT R. BAILEY

National Heart, Lung and Blood Institute

In this paper it is shown that the Cox likelihood (Cox, 1972) may be treated as a standard likelihood, in the sense that its maximizer $\hat{\beta}$ is asymptotically normally distributed with asymptotic covariance matrix equal to $-\{E\partial^2 \log L(\beta)/\partial\beta\partial\beta'\}^{-1}$. In the process, an asymptotic representation of the score function is obtained in terms of functions of the independent observations. This representation may have some uses in itself such as: (1) providing a kind of residual for each observation, censored or uncensored, thereby indicating the relative influence of the observations, and (2) providing some information about the applicability of the asymptotics in a particular small sample.

The asymptotic joint distribution of $\hat{\beta}$ and of the cumulative hazard function estimator $\hat{\Lambda}_0(t)$ is also derived via a representation of the latter involving an independent increments process. Bailey (1982) shows that the "joint likelihood function" of the regression parameters β and of the cumulative hazard jump parameters $\{\Lambda_i\}$ can be used in a natural way to obtain consistent estimates of these joint asymptotic covariances in the case of no ties. This justifies, to some extent, use of the general ML method for joint estimation of β and $\Lambda_0(t)$.

1. Introduction. The Cox regression model (Cox, 1972) has had much use in the analysis of censored survival data in the presence of explanatory covariates. The model is that there is some common unknown function $\lambda_0(t)$, which gives the shape of the hazard function for all individuals. An individual whose covariate vector is given by z has a hazard function proportional to $\lambda_0(t)$ with proportionality factor $\exp(\beta' z)$, where β is an unknown vector of regression parameters. The problem is to estimate β (and possibly $\lambda_0(t)$ as well).

Under these assumptions, and assuming we observe n individuals and the minimum of their survival or censoring time, Cox derived what he later called the "partial likelihood" function of β, and suggested estimating β by maximizing this function. He suggested that the usual asymptotic properties of standard parametric likelihood functions based on i.i.d. observations would pertain to this partial likelihood. The likelihood can be written

$$L(\beta) = \prod_{i=1}^{k} \exp(\beta' Z_{(i)})/[\textstyle\sum_{\mathscr{R}_{(i)}} \exp(\beta' Z_j)]. \tag{1.1}$$

Here, $Z_{(i)}$ is the covariate of the individual failing at the ith epoch and $\mathscr{R}_{(i)}$ the risk set at that epoch. Thus the claim was that if $\hat{\beta}$ maximized (1.1), then $n^{1/2}(\hat{\beta} - \beta)$ would be asymptotically distributed as $N(0, I^{-1}(\beta))$, where $I(\beta)$ is the Fisher information corresponding to (1.1). However, not until recently has there been any formal justification for this assertion.

Tsiatis (1981) proves asymptotic normality in the setting of random sampling from individuals with a given distribution of covariates, and random censorship. In this paper, another proof is provided in the slightly more general context of a fixed (arbitrary)

Received April 1980; revised August 1982.

[1] Research was supported by National Science Foundation Grant MCS 76-81435 while the author was at the University of Chicago Statistics Department.

AMS 1970 *subject classifications.* Primary 62E20; secondary 62F10.

Key words and phrases. Asymptotic representation, Cox regression, Hajek projection, independent increments, maximum likelihood, proportional hazards.

sequence of covariates and censoring times. This may be an important distinction in the context of clinical trials where one may not wish to assume the subjects are randomly sampled from any population. Another difference between this work and that of Tsiatis is the method of proof. Tsiatis uses weak convergence of some associated processes in proving normality of $\hat{\beta}$. Our proof involves a projection technique leading to an asymptotic representation of the score function as a sum of independent nonidentically distributed random variables. This representation has interest in its own right, since it indicates the influence of censored and of uncensored observations.

Also of interest is the asymptotic distribution of the survival function estimator, or equivalently of the cumulative hazard estimator. Tsiatis also finds the asymptotic distribution of the cumulative hazard estimates including the joint normality of $[\hat{\beta}, \hat{\Lambda}_0(t)]$ and weak convergence of $\hat{\Lambda}_0(t)$ to a normal process. Again the present work is in the slightly different context of fixed covariates, and uses a different method of proof, namely an asymptotic representation of the hazard function estimator as the sum of an independent increments process and another term which is the product of $(\hat{\beta} - \beta)$ and a fixed function of time. This representation [see (3.17)] has intrinsic interest, since it shows how simple the structure of the joint estimator is.

2. Asymptotic distribution of $\hat{\beta}$. In this section it is verified that the maximizer $\hat{\beta}$ of (1.1) obeys the same asymptotic theory as if (1.1) were a standard parametric likelihood based on i.i.d. observations. This will be done assuming an arbitrary fixed sequence of covariate values and censoring times. This is an important case, since it arises in the context of continuous recruitment in clinical trials, and since it generalizes the case of random censoring, and random covariates.

The method of proof is to provide an asymptotic representation of $\hat{\beta}$ in terms of a sum of independent (but not identically distributed) random variables. This is done by a projection method due to Hajek applied to the first derivative of the logarithm of (1.1). Thus the first derivative is shown to be asymptotically equivalent to a sum of independent random variables satisfying the conditions of the Liapunov Central Limit Theorem.

2.1 *Assumptions and results.* Let $\{(Z_1, \tau_1), (Z_2, \tau_2), \cdots\}$ be an arbitrary (nonrandom) sequence of covariate vectors and censoring times associated with a sequence of individuals. Further assume that $T_1, T_2, \cdots$ are mutually independent survival times with distribution as implied by the Cox model $S(t \mid Z) = \{S_0(t)\}^{\exp(\beta' Z)}$. The parameter β is assumed unknown. Assume S_0 arbitrary, except that it is differentiable with hazard function λ_0. We observe $\{(T_i^*, \delta_i) : i = 1, \cdots, n\}$ where $T_i^* \equiv \min(T_i, \tau_i)$, and $\delta_i \equiv I\{T_i < \tau_i\}$. Let us define $U(\beta) \equiv \partial \log L(\beta)/\partial\beta$ and $V(\beta) \equiv -\partial^2 \log L(\beta)/\partial\beta\partial\beta'$. A direct calculation shows that

$$U(\beta) = \textstyle\sum_{i=1}^n (Z_i - \bar{Z}_i)\delta_i \equiv \sum_{i=1}^n U_i(\beta), \quad \text{and} \quad V(\beta) = \sum_{i=1}^n V_i\delta_i, \tag{2.1}$$

where

$$\bar{Z}_i \equiv \textstyle\sum_{\mathscr{R}_i} Z_j \exp(\beta' Z_j) / \sum_{\mathscr{R}_i} \exp(\beta' Z_j),$$

$$V_i \equiv \textstyle\sum_{\mathscr{R}_i} Z_j Z_j' \exp(\beta' Z_j) / \sum_{\mathscr{R}_i} \exp(\beta' Z_j) - \bar{Z}_i \bar{Z}_i', \tag{2.2}$$

and

$$\mathscr{R}_i \equiv \{j \mid T_j^* > T_i^*\}.$$

Let us impose the following conditions on the Z's and τ's.

ASSUMPTION A1. $|Z_i| < M < \infty$ for all i and some constant M.

ASSUMPTION A2. $n^{-1} x' \{EV(\beta)\} x > c > 0$ for all $|x| = 1$ and all sufficiently large n.

Assumption 1 is quite reasonable and very convenient for the purpose of moment calculations. Assumption 2 says that the expected "information matrix" increases proportionally with n, another reasonable condition.

Under the above assumptions, we have

THEOREM 1. (a) $\Pr\{\hat{\beta}$ *exists and is unique*$\} \to 1$ *as* $n \to \infty$, (b) $|\hat{\beta} - \beta| \to_P 0$, *and* (c) $\{V(\hat{\beta})\}^{1/2}(\hat{\beta} - \beta) \to N(0, I)$.

Theorem 1 says that the maximizer of (1.1) may be treated just as if it were a standard parametric MLE.

2.2 *Preliminaries.* In the course of the proof, two systems of indices will be used: (1) the original system, corresponding to the sequence of independent observations, and (2) a system corresponding to the ordered times of death, in which the subscripts will be given in parentheses. Thus Z_i will refer to the ith covariate vector in the sequence, whereas $Z_{(i)}$ will refer to the covariate vector for the individual dying at the ith time of death. Note that this will depend on n, which will be suppressed in the notation. The dependence of U and V and other quantities on β will also be suppressed where convenient.

The functions $U(\beta)$ and $V(\beta)$ can now be expressed as follows:

$$U(\beta) = \textstyle\sum_{i=1}^n U_i(\beta) = \sum_{i=1}^k U_{(i)}(\beta), \tag{2.3}$$

where k is the number of deaths. Similarly,

$$V(\beta) = \textstyle\sum_{i=1}^n \delta_i V_i(\beta) = \sum_{i=1}^k V_{(i)}(\beta). \tag{2.4}$$

As Cox (1972) noted, since $U_{(i)}$ and $-V_{(i)}$ are first and second derivatives of the log of the conditional likelihood of the data given a death from risk set $R_{(i)}$, it follows immediately that

$$\begin{aligned} E(U_{(i)}) &= E\{E(U_{(i)} \mid R_{(i)})\} = 0 \\ \operatorname{Cov}(U_{(i)}) &= E\{\operatorname{Cov}(U_{(i)} \mid R_{(i)})\} + \operatorname{Cov}(0) = E(V_{(i)}), \quad \text{and} \\ E(U_{(j)} \mid U_{(i)}) &= E\{E(U_{(j)} \mid U_{(i)}, R_{(j)}) \mid U_{(i)}\} = E(0) = 0, \quad \text{for} \quad j > i. \end{aligned} \tag{2.5}$$

Therefore, summing (2.5) over all deaths, it follows that

$$E\{U(\beta)\} = 0, \quad \text{and} \quad \operatorname{Cov}\{U(\beta)\} = E\{V(\beta)\}. \tag{2.6}$$

It will first be shown (after Lemma 2.3, a technical lemma) that

LEMMA 2.1. $n^{-1}|V(\beta) - EV(\beta)| \to_P 0$.

Lemma 2.1 will be shown to imply the existence, uniqueness, and consistency of $\hat{\beta}$, and to imply the existence of a Taylor series representation:

$$V^*(\hat{\beta} - \beta) = U(\beta), \tag{2.7}$$

where the ith row of V^* is that of V evaluated at some point β_i^* intermediate between β and $\hat{\beta}$. V^* is not necessarily positive definite. Nevertheless, the consistency of $\hat{\beta}$ and the equicontinuity of the components of V near β imply that $n^{-1}V^*$ converges to $n^{-1}V(\beta)$, and therefore to $n^{-1}EV(\beta)$. Therefore, it will suffice to show that

LEMMA 2.2. $\{EV(\beta)\}^{-1/2}U(\beta) \to N(0, I)$.

There is a key technical lemma which is very useful in the proof of these lemmas, involving approximation of random sums by their expectations. First let us define

$$\begin{aligned} m(t) &\equiv \textstyle\sum_{j=1}^n I\{T_j^* > t\}, \quad \text{and} \\ \mu(t) &\equiv \textstyle\sum_{j=1}^n I\{\tau_j > t\} S(t \mid Z_j) = E\{m(t)\}, \end{aligned} \tag{2.8}$$

the observed and expected number at risk at time t. Let $d(z)$ be any continuous nonnegative function of z bounded away from zero. Let $a(z)$ be a bounded function. Let $d(t) \equiv \sum I\{T_j^* > t\}d(Z_j)$, and let $\bar{d}(t)$ be $E\{d(t)\}$, and similarly for $a(t)$ and $\bar{a}(t)$. The notation $Y = O_u[f(x)]$ is used to mean that $|Y|$ is bounded by a constant multiple of $f(x)$ uniformly in x. Finally, the notation $f^{-1}(x)$ is used to mean $1/f(x)$. The following lemma is merely an application of the delta method.

LEMMA 2.3. *Conditional on $T_i^* > t$, and for $j = 1, 2, \cdots,$*

(a) $E[d^{-1}(t)] = \bar{d}^{-1}(t) + O_u[\min(\mu^{-2}(t), 1)]$

(b) $\mathrm{Var}[d^{-1}(t)] = O_u[\min(\mu^{-3}(t), 1)]$

(c) $E[a^j(t)/d^j(t)] = [\bar{a}(t)/\bar{d}(t)]^j + O_u[\min(\mu^{-1}(t), 1)]$

(d) $\mathrm{Var}[a^j(t)/d^j(t)] = O_u[\min(\mu^{-1}(t), 1)]$

PROOF OF (a). Suppressing the dependence on t, d^{-1} may be expanded as

$$d^{-1} = \bar{d}^{-1} - (d - \bar{d})\bar{d}^{-2} + (d - \bar{d})^2(d^*)^{-3}, \tag{2.9}$$

where d^* is intermediate between d and $\bar{d}$. Now $d^*(t) > cm(t)$ for some c, by assumption, and therefore

$$|E(d^{-1}| T_i^* > t) - \bar{d}^{-1}| < \bar{d}^{-2}|E(d | T_i^* > t) - \bar{d}| + c^{-3}E\{(d - \bar{d})^2/m^3\}. \tag{2.10}$$

It is easy to see that $E(d | T_i^* > t) - \bar{d} = O(1)$. It follows that the first RHS term of (2.10) is $O_u[\mu^{-2}(t)]$. Let $A \equiv (d - \bar{d})^2$ and let $B \equiv m^{-3}$. By Cauchy-Schwarz, $E(AB) < E^{1/2}(A^2)E^{1/2}(B^2)$. It is clear that $E(A^2) = O_u[\mu^2(t)]$. Similarly, $E(m^{-6}(t) | T_i^* > t] = O_u[\mu^{-6}(t)]$. Part (a) follows immediately. The other proofs are parallel.

2.3 *Proof of Lemma* 2.1. Next it is shown that $\mathrm{cov}\{V(\beta)\} = o(n^2)$. For simplicity. Let us consider a single element of V, say v. The variance of v is bounded by

$$\mathrm{Var}(v) < \textstyle\sum_{i=1}^n \sum_{j=1}^n |\mathrm{cov}(v_i, v_j)|. \tag{2.11}$$

Conditioning on the value of T_i^* gives

$$\mathrm{Cov}(v_i, v_j) = E\{\mathrm{Cov}(v_i, v_j | T_i^*)\} + \mathrm{Cov}\{E(v_i | T_i^*), E(v_j | T_i^*)\}. \tag{2.12}$$

By Lemma 2.3, conditioning on T_i^*, $\mathrm{var}(v_i | T_i^*) = O_u\{\min[\mu^{-1}(T_i^*), 1]\}$. Thus,

$$\mathrm{Cov}(v_i, v_j | T_i^*) = O_u\{\min[\mu^{-1/2}(T_i^*), 1]\}. \tag{2.13}$$

For the second term in (2.12), note that v_j depends on T_i^* only through the presence or absence of the ith individual in R_j. Hence $|E(v_j | T_i^*) - E(v_j)| = O\{E[m^{-1}(T_j^*)]\}$, which implies that

$$\mathrm{Cov}[E(v_i | T_i^*), E(v_j | T_i^*)] = O\{E[m^{-1}(T_j^*)]\}. \tag{2.14}$$

It is now claimed that the expectation of (2.13), summed over all i and j, is $O(n^{3/2})$. If the index i is drawn at random uniformly from $\{1, \cdots, n\}$, then $n^{-1}\mu(T_i^*)$ is a random variable which is stochastically larger than a $U[0, 1]$ random variable. Substitution of a $U[0, 1]$ random variable into (2.13), taking expectations and summing over i and j verifies the claim. A similar argument can be made that $n^{-1}m(T_j^*)$ can be replaced by a $U[0, 1]$ random variables when (2.14) is summed over all i and j. This implies that the sum is $O(n\log n)$. Therefore, (2.11) is $o(n^2)$, as claimed. Lemma 2.1 follows by A2 and Chebychev's inequality.

The existence and uniqueness of $\hat{\beta}$ can be shown as follows. Lemma 2.1 and A2 imply that $V(\beta)$ is positive definite with probability approaching 1. Since, for any fixed b, $V(b)$ is a sum of weighted covariance matrices with weights bounded away from zero and bounded above, it follows that $V(\cdot)$ is positive definite for all β if it is positive definite for any. Therefore, with probability approaching, 1, $\log(L)$ is strictly concave. Therefore, $\hat{\beta}$ exists, is unique, and has the representation (2.7). To prove consistency, consider the distances $r_n = n^{-1/2+\varepsilon}$, $0 < \varepsilon < ½$. Consider the events E_n^1 and E_n^2, where

$$E_n^1 \equiv \{|\hat{\beta} - \beta| > r_n\}, \quad \text{and} \quad E_n^2 \equiv \{\log[L(\beta + r_n u_n)] > \log[L(\beta)]\}, \tag{2.15}$$

where $u_n = (\hat{\beta} - \beta)/|\hat{\beta} - \beta|$. By log concavity of L, $E_n^1 \subset E_n^2$, and therefore $\Pr\{E_n^1\} < \Pr\{E_n^2\}$. But E_n^2 can be expanded as

$$E_n^2 = \{2n^{-1/2-\varepsilon} u_n' U(\beta) > n^{-1} u_n' V(\beta^*) u_n\}, \tag{2.16}$$

where β^* is intermediate between β and $\beta + r_n u_n$. Since $\operatorname{cov}(U) = EV = O(n)$, the LHS in brackets converges to zero in probability. By assumption A2, since $\beta^* \to \beta$ by construction, and by the fact that $n^{-1}V(\cdot)$ is equicontinuous near β (having bounded derivatives), the RHS in brackets is bounded away from zero for all sufficiently large n. Therefore $\Pr(E_n^2) \to 0$, and $\hat{\beta}$ is consistent.

By Lemma 2.1, the consistency of $\hat{\beta}$, and the equicontinuity of the components of V near β, it is clear that V^*, $V(\hat{\beta})$, and $EV(\beta)$ are asymptotically interchangeable in the representation (2.7).

2.4 *Asymptotic representation of U.* Turning to Lemma 2.2, it was noted earlier that $\operatorname{cov}(U) = EV$, so the only issue is the normality of $U(\beta)$. As Cox pointed out, $U(\beta)$ is the sum of uncorrelated but not independent vectors $U_{(i)}$. Let us therefore define $\hat{U}_i \equiv E\{U(\beta) \mid (Z_i, T_i^*, \delta_i)\}$. We may think of $\hat{U}_i$ as the projection of U onto the space of square integrable functions of the ith independent observation. Clearly $\{\hat{U}_i\}$ form a set of independent random vectors. Let $\hat{U} \equiv \sum \hat{U}_i$. By Hajek's projection lemma [Hajek, 1968, Lemma 4.1],

$$\operatorname{cov}(\hat{U}) + \operatorname{cov}(U - \hat{U}) = \operatorname{cov}(U). \tag{2.17}$$

Thus if $\operatorname{cov}(U) - \operatorname{cov}(\hat{U})$ can be shown to be $o(n)$, it will follow that

LEMMA 2.4. $n^{-1/2}(U - \hat{U}) \to_P 0$.

This lemma says that $\hat{U}$ is an asymptotic representation of U in terms of independent random vectors. To prove it, let us compute $\hat{U}_i$ directly. First note that

$$\hat{U}_i = E(U_i \mid Z_i, T_i^*, \delta_i) + E(\textstyle\sum_{j \neq i} U_j \mid Z_i, T_i^*, \delta_i). \tag{2.18}$$

By Lemma 2.3, with $a(z) = z\exp(\beta' z)$ and $d(z) = \exp(\beta' z)$,

$$E(\bar{Z}_i \mid T_i^*, Z_i) = g(T_i^*) + O_u\{\min[\mu^{-1}(T_i^*), 1]\}, \quad \text{where} \tag{2.19}$$
$$g(t) \equiv \frac{\sum_{j=1}^n Z_j \exp(\beta' Z_j) S(t \mid Z_j) I\{t < \tau_j\}}{\sum_{j=1}^n \exp(\beta' Z_j) S(t \mid Z_j) I\{t < \tau_j\}.}$$

To approximate the second RHS term in (2.18), let us define $\mathscr{R}_{ji}'$ as $\mathscr{R}_j$ minus the ith individual, if present. With $\bar{Z}_{ji}'$ defined analogously, $\bar{Z}_{ji}'$ can be expanded as

$$\bar{Z}_{ji}' = \bar{Z}_j - I\{i \in \mathscr{R}_j\}(D_{ji}')^{-1}(Z_i - \bar{Z}_{ji}')\exp(\beta' Z_i)[1 - \exp(\beta' Z_i)/D_j], \tag{2.20}$$

where $D_j \equiv \sum_{\mathscr{R}_j} \exp(\beta' Z_k)$, and D_{ji}' is relative to $\mathscr{R}_{ji}'$. Clearly $(\bar{Z}_{ji}', D_{ji}')$ is independent of (Z_i, T_i^*, δ_i), so that $E\{\sum_{j \neq i}(Z_j - \bar{Z}_{ji}')\delta_j \mid (Z_i, T_i^*, \delta_i)\} = 0$. Therefore, if we let $I_{ij} \equiv I\{i \in \mathscr{R}_j\}$, the second RHS term in (2.18) may be expanded as

$$E(\textstyle\sum_{j \neq i} U_j \mid Z_i, T_i^*, \delta_i) = \tag{2.21}$$
$$-E\{\textstyle\sum_{j \neq i}[I_{ij}(D_{ji}')^{-1}(Z_i - \bar{Z}_{ji}')\exp(\beta' Z_i)(1 + O_u[m^{-1}(T_j^*)]] \mid (Z_i, T_i^*, \delta_i)\}.$$

Conditioning on T_j in (2.22), applying Lemma 2.3, and integrating with respect to the distribution of T_j, denoted by $dF(s\,|\,Z_j)$, gives

$$\text{(2.22)}\quad \begin{aligned}\text{RHS} &= -\exp(\beta' Z_i) \textstyle\sum_{j\neq i} \displaystyle\int_0^{T_i^*} [I\{\tau_j > s\}\bar{D}^{-1}(s)\{Z_i - g(s)\}]dF(s\,|\,Z_j)\\ &\quad -\exp(\beta' Z_i) \textstyle\sum_{j\neq i} \displaystyle\int_0^{T_i^*} [I\{\tau_j > s\}\; O_u\{\min[\mu^{-2}(s), 1]\}]dF(s\,|\,Z_j),\end{aligned}$$

where $\bar{D}(s) \equiv \sum I\{s < \tau_j\}\exp(\beta' Z_j)S(s\,|\,Z_j)$. Since $\sum I\{\tau_j > s\}dF(s\,|\,Z_j) = \lambda_0(s)\bar{D}(s)\,ds$, the integral in (2.22) simplifies to

$$\text{(2.23)}\quad \text{RHS} = -\exp(\beta' Z_i) \int_0^{T_i^*} \{Z_i - g(s)\}\lambda_0(s)\,ds + O_u\{\min[\mu^{-1}(T_i^*), 1]\}.$$

The bound on the remainder term follows from the fact that the measure $d\mu(s)$ dominates the measure $\sum I\{\tau_j > s\}\; dF(s\,|\,Z_j)$. Combining (2.1), (2.18), (2.19), (2.21), and (2.23), the term $\hat{U}_i$ can be written as

$$\text{(2.24)}\quad \begin{aligned}\hat{U}_i &= \{Z_i - g(T_i^*)\}\delta_i - \int_0^{T_i^*} \{Z_i - g(s)\}\lambda(s\,|\,Z_i)\,ds + O_u\{\min[\mu^{-1}(T_i^*), 1]\}\\ &\equiv A_i - B_i + r_i.\end{aligned}$$

With this representation in hand, let us compute the moments of $\hat{U}_i$. By an argument like the one after (2.14), $E(\sum r_i r_i') = O(1)$. By the independence of r_i and r_j,

$$\text{(2.25)}\quad \operatorname{cov}(\textstyle\sum_{i=1}^n r_i) = O(1).$$

Next consider the moments of A_i and B_i. Let $dF^*(s\,|\,Z_j, \tau_j)$ be the distribution of T_j^* and $f(s\,|\,Z_j)$ the density of T_j. By a direct calculation,

$$\text{(2.26)}\quad \begin{aligned}E(B_i) &= \int_0^\infty \left[\int_0^s \{Z_i - g(r)\}\lambda(r\,|\,Z_i)\,dr\right] dF^*(s\,|\,Z_i, \tau_i)\\ &= \int_0^\infty \{Z_i - g(r)\}f(r\,|\,Z_i)S^{-1}(r\,|\,Z_i)\left\{\int_r^\infty dF^*(s\,|\,Z_i, \tau_i)\right\} dr\\ &= \int_0^{T_i^*} \{Z_i - g(r)\}f(r\,|\,Z_i)\,dr = E(A_i).\end{aligned}$$

The third line follows because, for $r < \tau_i$, $\int_r^\infty dF^*(s\,|\,Z_i, \tau_i) = S(r\,|\,Z_i)$. A similar calculation shows that

$$\text{(2.27)}\quad E(A_i B_i') = E(B_i B_i')/2,$$

and therefore, $\operatorname{cov}(A_i - B_i) = E(A_i A_i')$. Now $E(A_i) \neq 0$. Nevertheless, by the independence of $(A_i - B_i)$ and $(A_j - B_j)$, it follows that

$$\text{(2.28)}\quad \begin{aligned}\operatorname{cov}\{\textstyle\sum_{i=1}^n (A_i - B_i)\} &= \textstyle\sum_{i=1}^n \operatorname{cov}(A_i - B_i) = \sum_{i=1}^n E(A_i A_i')\\ &= E(\textstyle\sum_{i=1}^n A_i A_i') = E\{\sum_{i=1}^k A_{(i)} A_{(i)}'\}.\end{aligned}$$

To evaluate the last expectation, condition first on the ith event occurring, and on $T_{(i)}^*$ and $\mathscr{R}_{(i)}$ to get

$$\text{(2.29)}\quad E(A_{(i)}A_{(i)}') = E(V_{(i)}) + E[\{\bar{Z}_{(i)} - g(T_{(i)}^*)\}\{\bar{Z}_{(i)} - g(T_{(i)}^*)\}'].$$

The first RHS term in (2.29), summed over i, is $EV(\beta)$. The sum of second terms can be rewritten as $\sum E[\{\bar{Z}_i - g(T_i^*)\}\{\bar{Z}_i - g(T_i^*)\}'\delta_i]$. In this form, Lemma 2.3 can be applied and the sum can be seen to be $O(\log n)$, by using the usual $U[0, 1]$ comparison. It follows that

$$E\left(\sum_{i=1}^{k} A_{(i)}A'_{(i)}\right) = E(V) + o(n). \tag{2.30}$$

Combining (2.6), (2.17), (2.24), (2.25), (2.28), (2.30), and Chebychev's inequality gives Lemma 2.4.

2.5 *Asymptotic Normality of* $\hat{U}$. The only issue remaining is asymptotic normality. Consider an arbitrary linear combination $x'\,\hat{U}$ with $|x| = 1$. Let $M_3 = (\sum E\,|x'\,\hat{U}_i|^3)^{1/3}$, and let $M_2 = (\sum E\,|x'\,\hat{U}_i|^2)^{1/2}$. It will be shown that $\lim M_3/M_2 = 0$. Since $\hat{U}_i$ and $\hat{U}_j$ are independent, $M_2 = E(|x'\,\hat{U}|^2)^{1/2}$. The preceding proof, plus assumption A2 imply that $M_2 > cn^{1/2}$ for some c and all sufficiently large n. The quantities $x'\,A_i$ and $x'\,r_i$ are uniformly $O(1)$. The term $x'B_i$ is $O[\int_0^{T_i^*} \lambda(s\,|\,Z_i)\,ds] = O[-\log S(T_i^*\,|\,Z_i)]$. Thus $x'B_i$ is dominated by a unit exponential random variable, which implies that $E[\,|x'\,\hat{U}_i|^3] = O(1)$. Therefore $M_3 = O(n^{1/3})$, $\hat{U}$ satisfies the Liapunov CLT conditions, and Lemma 2.2 and Theorem 1 follow.

2.6 *Remarks.* Having proved Theorem 1, it is worth examining the representation (2.24), which in effect gives us the "regression" of the score function $U(\beta)$ on each independent observation. In particular, note that $A_i - B_i$ can be considered (asymptotically) as a residual for the ith observation. This vector can be estimated from the sample by an empirical analogue. This has potential diagnostic value in assessing the influence of various points on $\hat{\beta}$. Of course, it must be born in mind that the $\hat{U}_i$'s do not have identical covariance matrices. Nevertheless, these too can be estimated from the data. In fact one can write

$$\operatorname{cov}(A_i - B_i | Z_i, \tau_i) = E(A_i A_i' \,|\, Z_i, \tau_i) = \int_0^{\tau_i} \{Z_i - g(s)\}\{Z_i - g(s)\}'\,dF(s\,|\,Z_i). \tag{2.31}$$

In order to estimate (2.31), τ_i must be known, which is typically the case in clinical trials. In that case, the joint MLE of $(\beta, S(s\,|\,Z_i))$ as given in Bailey (1980) or in Prentice and Gloeckler (1978) can be substituted (along with $\bar{Z}(s)$ for $g(s)$) into (2.31) to obtain an estimate of $\operatorname{cov}(A_i - B_i)$. This requires substantial computation, since for each Z_i, one integrates with respect to a different measure $\hat{S}(s\,|\,Z_i)$. Clearly, whether we standardize or not, the asymptotic distribution of these "residuals" is not normal, depending so directly on δ_i. However, there is enough continuity to suggest that a normal approximation might be useful in an informal way.

3. The distribution of $\hat{\Lambda}_0(t)$. In this section the asymptotic distribution of $\hat{\Lambda}_0(t)$ is derived, where

$$\hat{\Lambda}_0(t) \equiv \sum_{i=1}^{k(t)} \hat{D}_{(i)}^{-1}, \quad \text{with} \tag{3.1}$$
$$k(t) \equiv \operatorname{card}\{i\,|\,T_i^* \le t,\, \delta_i = 1\}, \qquad \hat{D}_{(i)} \equiv \sum_{\mathscr{R}(i)} \exp(\hat{\beta}'\,Z_j).$$

The quantity $\hat{\Lambda}_0(t)$ is often used to estimate the cumulative hazard function $\Lambda_0(t) = \int^t \lambda_0(s)\,ds$. In the interest of simplicity, assume no censoring, although the modifications needed for censoring will be suggested. Further assume that attention is restricted to $[0, T]$, where $S_0(T) = \delta \to 0$. This implies that, for any $\varepsilon > 0$,

$$\Pr\{k(T) < n(1 - \delta + \varepsilon)\} \to_P 0. \tag{3.2}$$

Given these assumptions, the following theorem holds.

THEOREM 2. $C^{-1/2}(t)((\hat{\beta} - \beta), \{\hat{\Lambda}_0(t) - \Lambda_0(t)\}) \to N(0, I)$, *where*

$$C(t) \equiv \begin{bmatrix} \{EV(\beta)\}^{-1} & \{EV(\beta)\}^{-1}\Gamma(t) \\ \Gamma'(t)\{EV(\beta)\}^{-1} & \psi(t) + \Gamma'(t)\{EV(\beta)\}^{-1}\Gamma(t) \end{bmatrix} \tag{3.3}$$

with

$$\Gamma(t) \equiv \int_0^t g(s)\lambda_0(s)\, ds, \quad \textit{and} \quad \psi(t) \equiv \int_0^t \bar{D}^{-2}(s)\, d\mu(s).$$

Since there is no censoring, $d\mu(s) = \sum dF(s \mid Z_j)$. *Furthermore, note that, since the sequence* $\{Z_i\}$ *is arbitrary,* $C(t)$ *depends on n, so Theorem 2 must be stated with the C dependence on the left.*

PROOF. The basic idea of the proof is to expand $\hat{\Lambda}_0(t)$ about the true parameter value. This gives

$$\hat{\Lambda}_0(t) = \textstyle\sum_{i=1}^{k(t)} \hat{D}_{(i)}^{-1} = \sum_{i=1}^{k(t)} D_{(i)}^{-1} - \{\sum_{i=1}^{k(t)} D_{(i)}^{-1} \bar{Z}_{(i)}\}'(\hat{\beta} - \beta) + \frac{1}{2}(\hat{\beta} - \beta)' \rho^*(t)(\hat{\beta} - \beta), \tag{3.4}$$

where $\rho^*(t) \equiv \sum_{i=1}^{k(t)} D_{(i)}^{*-1}(V_{(i)}^* - \bar{Z}_{(i)}^* \bar{Z}_{(i)}^{*\prime})$, and the asterisks indicate evaluation at an intermediate value β^*. Let us define

$$\tilde{\Lambda}_0(t) \equiv \textstyle\sum_{i=1}^{k(t)} D_{(i)}^{-1}, \qquad \eta(t) \equiv \tilde{\Lambda}_0(t) - \Lambda_0(t), \qquad \tilde{\Gamma}(t) \equiv \sum_{i=1}^{k(t)} \bar{Z}_{(i)} D_{(i)}^{-1}.$$

The proof of Theorem 2 consists in showing that

LEMMA 3.1. $\psi^{-1/2}(t)\eta(t) \to N(0, 1)$,

LEMMA 3.2. $\tilde{\Gamma}(t) - \Gamma(t) \to_P 0$,

and

LEMMA 3.3. $n^{-1/2} \rho^*(t) \to_P 0$.

Lemma 3.3 follows from the observation that the summands of $\rho^*(t)$ are $O_u[m^{-1}(t)]$, and therefore, by (3.2), for $t \le T$, $\sup \rho^*(t) = O_p(1)$.

To prove Lemma 3.1, note that $\eta(t)$ can be expressed as the sum

$$\begin{aligned} \eta(t) &= \textstyle\sum_{i=1}^{k(t)} \left\{ D_{(i)}^{-1} - \displaystyle\int_{T_{(i-1)}}^{T_{(i)}} \lambda_0(s)\, ds \right\} - \int_{T_{k(t)}}^{t} \lambda_0(s)\, ds \\ &= -\textstyle\sum_{i=1}^{k(t)} D_{(i)}^{-1} \left\{ \displaystyle\int_{T_{(i-1)}}^{T_{(i)}} D_{(i)}\lambda_0(s)\, ds - 1 \right\} - \int_{T_{k(t)}}^{t} \lambda_0(s)\, ds \\ &= -\textstyle\sum_{i=1}^{k(t)} D_{(i)}^{-1}(e_i - 1) - \displaystyle\int_{T_{k(t)}}^{t} \lambda_0(s)\, ds, \end{aligned} \tag{3.5}$$

where $\{e_i\}$ are i.i.d. unit exponential random variables. The second term in (3.5) is of order $O[m^{-1}(t)e_{k(t)}]$. From (3.2) and the fact that $\max_i e_i = O_p(\log n)$, the supremum over $t \le T$ of this quantity is $O_p(n^{-1} \log n)$. Thus, apart from a term $o_p(1)$, $\eta(t)$ has the form of a weighted sum of a random number of independent centered unit exponentials, where the weights are also random. If we could replace both the weights and the random index $k(t)$ by their expectations, then Lemma 3.1 would follow. Indeed, one has weak convergence of $\eta(t)$ to an independent increments process with covariance function $\operatorname{cov}\{\eta(s), \eta(t)\} = \psi(\min(s, t))$.

The random weights $D_{(i)}^{-1}$ can be taken care of as follows. The variance of $D_{(i)}$ is of order

$O(n)$. This can be seen by writing

$$D_{(i)} = \sum_{j=1}^{n} \exp(\beta' Z_j) I\{T_j \geq T_{(i)}\},$$

and noting that the terms in the sum are $O(1)$ and slightly negatively correlated. Note that for $i < n(1 - \delta + \varepsilon)$, $D_{(i)} > cn$, for some positive c. Therefore, $\operatorname{var}(D_{(i)}^{-1}) = O(n^{-3})$, uniformly in $i < n(1 - \delta + \varepsilon)$. Let $\varepsilon_{(i)} \equiv D_{(i)}^{-1} - E(D_{(i)}^{-1})$. Since e_i is independent of $D_{(i)}$, and of the past up to $T_{(i-1)}$, it follows that $E\{\varepsilon_{(i)}^2 (e_i - 1)^2\} = O_u(n^{-3})$, and that

$$\operatorname{cov}\{\varepsilon_{(j)}(e_j - 1), \varepsilon_{(i)}(e_i - 1)\} = 0, \qquad i \neq j. \tag{3.6}$$

Therefore, the sequence of partial sums $\sum \varepsilon_{(i)}(e_i - 1)$ forms a martingale, which, along with (3.2), implies that

$$\sup_{t \leq T} n^{1/2} \left| \sum_{i=1}^{k(t)} \varepsilon_{(i)}(e_i - 1) \right| \to_P 0. \tag{3.7}$$

If $\bar{D}_{(i)} \equiv E(D_{(i)})$, then $\bar{D}_{(i)}^{-1}$ can be substituted for $E(D_{(i)}^{-1})$ in (3.7).

The random index $k(t)$ can be handled as follows. Let $\hat{t}$ be defined as the smallest value such that $k(\hat{t}) = [Ek(t)]$. Partition $[0, T]$ into subintervals $[t_i, t_{i+1}]$, where t_i is defined by $Ek(t_i) = \nu_i = in^{1/2+\alpha}$, $\alpha > 0$. The uniform convergence in probability of $n^{-1/2}\{k(t) - Ek(t)\}$ to 0 implies that $\hat{t}$ and t will be in the same or adjacent subintervals for all t in $[0, T]$ with probability approaching 1. The triangle inequality gives

$$\sup_{t<T} n^{1/2} \left| \sum_{k=1}^{k(t)} \bar{D}_{(k)}^{-1}(e_k - 1) - \sum_{k=1}^{Ek(t)} \bar{D}_{(k)}^{-1}(e_k - 1) \right| < 3 \max_i \xi_i, \quad \text{where} \tag{3.8}$$

$$\xi_i \equiv \max_{\nu_i < j < \nu_{i+1}} \left| n^{1/2} \sum_{\ell=\nu_i}^{j} \bar{D}_{(\ell)}^{-1}(e_\ell - 1) \right|. \tag{3.9}$$

The fourth moment version of Kolmogorov's inequality gives

$$\Pr\{\xi_i > \varepsilon\} < \varepsilon^{-4} E\{n^{1/2} \textstyle\sum_{\nu_i}^{\nu_{i+1}} \bar{D}_{(j)}^{-1}(e_j - 1)\}^4 = \varepsilon^{-4} O_u(n^{-1+2\alpha}). \tag{3.10}$$

Taking maxima over i gives

$$\Pr(\max_i \xi_i > \varepsilon) = \varepsilon^{-4} O(n^{1/2-\alpha}) O(n^{-1+2\alpha}) = \varepsilon^{-4} O(n^{-1/2+\alpha}). \tag{3.11}$$

If ε is taken to be $n^{-1/8+\alpha/4+\rho}$, with $0 < \alpha/4 + \rho < 1/8$, then

$$\Pr(\max_i \xi_i > n^{\alpha/4+\rho-1/8}) = O(n^{-4\rho}), \tag{3.12}$$

and so $\max_i \xi_i \to_P 0$. Thus there is a uniform representation of $\eta(t)$,

$$\sup_{0 \leq r \leq T} n^{1/2}\{\eta(t) - \bar{\eta}(t)\} \to_P 0, \tag{3.13}$$

where $\bar{\eta}(t) \equiv \sum_{i=1}^{Ek(t)} \bar{D}_{(i)}^{-1}(e_i - 1)$. It follows immediately that $n^{1/2}\eta(t)$ is essentially an independent increments process with covariance function $\hat{\psi}(t) = n \sum_{i=1}^{Ek(t)} \bar{D}_{(i)}^{-2}$. This converges to $\psi(t)$, which can be shown as follows. Notice that

$$n \sum_{i=1}^{n} D^{-2}(T_i) I\{T_i \leq t\} = n \sum_{i=1}^{k(t)} D_{(i)}^{-2}. \tag{3.14}$$

The LHS of (3.14) can be approximated by

$$\left| n \sum_{i=1}^{n} D^{-2}(T_i) I\{T_i \leq t\} - n \sum_{i=1}^{n} \bar{D}^{-2}(T_i) I\{T_i \leq t\} \right| \to_P 0. \tag{3.15}$$

This follows from a variance calculation based on Lemma 2.3. The approximating sum converges to $\psi(t)$ by the strong law of large numbers. The RHS of (3.14) converges to $n \sum_{i=1}^{Ek(t)} \bar{D}_{(i)}^{-2}$ by variance calculations based on $\operatorname{var}(D_{(i)}^{-2}) = O(n^{-5})$. Thus the equivalent sums in (3.14) converge to $\psi(t)$ and $\hat{\psi}(t)$. It also follows from (3.13) that $\eta(t)$ and $\hat{\beta}$ are asymptotically independent, since clearly the representation $\bar{\eta}(t)$ is independent of $\hat{\beta}$.

Turning to Lemma 3.2, $\tilde{\Gamma}(t) - \Gamma(t)$ can be expressed as

$$\tilde{\Gamma}(t) - \Gamma(t) = \sum_{i=1}^{k(t)} \bar{Z}_{(i)} D_{(i)}^{-1} - \int_0^t g(s) \lambda_0(s)\, ds$$

$$(3.16) \qquad = \sum_{i=1}^{k(t)} D_{(i)}^{-1}\{\bar{Z}_{(i)} - g(T_{(i)})\} - \sum_{i=1}^{k(t)} D_{(i)}^{-1} g(T_{(i)})(e_i - 1) + \sum_{i=1}^{k(t)} O(e_i D_{(i)}^{-1})\{g(T_{(i)}) - g(T_{(i-1)})\} - \int_{T_{k(t)}}^{t} g(s)\lambda_0(s)\,ds,$$

using an integration by parts to obtain the second and third terms. The first term converges uniformly to zero by a Glivenko-Cantelli argument applied to the process $\bar{Z}(s) - g(s)$. For the second term, use a martingale argument. For the third term, note that $g'(s) = O_u(\lambda_0(s))$, so that $g(T_{(i)}) - g(T_{(i-1)}) = O_u(e_i/D_{(i)})$. Therefore, the third term is $O_p(n^{-1}\log n)$, uniformly in t. The fourth term is also uniformly $O_p(n^{-1}\log n)$. This establishes Lemma 3.2, and the uniform representation

$$(3.17) \qquad n^{1/2}\{\hat{\Lambda}_0(t) - \Lambda_0(t)\} = n^{1/2}\bar{\eta}(t) + n^{1/2}(\hat{\beta} - \beta)\Gamma(t) + o_p(1),$$

where $\bar{\eta}(t)$ is an independent increments process independent of $\hat{\beta}$ with covariance function which is asymptotically equivalent to ψ, and where $\Gamma(t)$ is the nonrandom function given in (3.3). Theorem 2 follows immediately. The uniformity of the representation (3.17) implies weak convergence to a normal process, by using the standard case given in Billingsley (1968, pages 68–70].

If censoring is allowed, the technical details become more cumbersome, but the idea is identical. In that case one has to allow for the fact that the ith event may not occur. If attention is limited to an interval and to such i that Pr{the ith event occurs} $\to 1$, and if the definition of $\eta(t)$ is modified appropriately, the same basic argument can be used. One nice feature of this representation is that it shows the essential simplicity of the joint estimate of $[\beta, S_0(t)]$ in the Cox model.

Acknowledgments. The author is grateful to Paul Meier for suggesting the original problem, and for many helpful discussions. The referees, by their many detailed criticisms and suggestions, contributed substantially to the accuracy and clarity of the proofs, and are gratefully acknowledged.

REFERENCES

BAILEY, K. (1979). The general ML approach to the Cox regression model. Unpublished Ph.D. thesis.

BAILEY, K. (1980). A comparison of the Cox estimator and the general ML estimators of β in the Cox model. Unpublished manuscript.

BILLINGSLEY (1968). *Convergence of Probability Measure.* Wiley. New York.

BILLINGSLEY (1979). *Probability and Measure.* Wiley, New York.

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

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

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

HAJEK, J. (1968). Asymptotic normality of simple linear rank statistics under alternatives. *Ann. Math. Statist.* **39** 325–346.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1973). Marginal likelihoods based under Cox's regression and life model. *Biometrika* **39** 267–278.

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–581.

LEHMANN, E. L. (1953). The power of rank tests. *Ann. Math. Statist.* **24** 23–43.

PRENTICE, R. L., and GLOECKLER, L. A. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* **34** 57–67.

SPROTT, D. and KALBFLEISCH, J. D. (1969). Examples of likelihoods and comparison with point estimates and large sample approximations. *J. Amer. Statist. Assoc.* **64** 468–484.

STIGLER, S. M. (1974). Linear functions of order statistics with smooth weight functions. *Ann. Statist.* **4** 676–693.

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

MATHEMATICAL AND APPLIED STATISTICS BRANCH
DIVISION OF HEART AND VASCULAR DISEASES
NATIONAL HEART, LUNG AND BLOOD INSTITUTE
7550 WISCONSIN AVENUE
BETHESDA, MD 20205