The Annals of Statistics
1988, Vol. 16, No. 2, 678–693

COX'S PERIODIC REGRESSION MODEL

BY O. PONS AND E. DE TURCKHEIM

Institut National de la Recherche Agronomique

Cox's regression model has been successfully used for censored survival data. It can be adapted to model a counting process having a periodic underlying intensity. In survival analysis, the asymptotic properties, as studied by Andersen and Gill, correspond to a large number of processes running parallel over the same time interval. Here a single point process is observed over a large number of successive periods. Cox's model can easily be adapted to this situation and conditions are given which ensure the estimators have the classical large sample properties. Proofs use both martingale techniques and theorems for convergence of empirical probability measures. Finally, an example concerning the feeding pattern of domestic rabbits is included.

1. Introduction. Cox's regression model for a sequence of random points $(T_1, T_2, \ldots)$ over the real half line specifies the intensity of the associated counting process N, where $N(t)$ counts observed events up to time t, to have the form

$$\alpha(t) = \lambda_0(t)\exp\{\beta_0 Z(t)\}, \qquad t \geq 0, \tag{1.1}$$

where $\lambda_0(t)$ is an unknown deterministic nonnegative function, $\beta_0 = (\beta_{01}, \ldots, \beta_{0q})$ is a row vector of q real coefficients to be estimated and $Z = (Z_1, \ldots, Z_q)$ is a column vector of q stochastic processes. This model is widely used to describe censored survival data in the presence of explanatory covariates Z. In that case, n independent individuals are observed and each of them has the hazard function $\alpha_i = \lambda_0 \exp\{\beta_0 Z_i\}$, where Z_i is the covariate vector for the ith individual [Cox (1972)]. Defining N_i as the counting process which counts the failure of the ith individual up to time t, and Y_i as the indicator process for this individual being at risk at t^-, N_i has the intensity $\alpha_i = \lambda_0 Y_i \exp\{\beta_0 Z_i\}$. More generally, Andersen and Gill [(1982), referred to as AG] consider an n-dimensional counting process $N^{(n)} = (N_i, \ldots, N_n)$, where N_i has the intensity $\alpha_i = \lambda_0 Y_i \exp\{\beta_0 Z_i\}$, and they study asymptotic properties of the estimators of β_0 and λ_0 when $n \to \infty$. Slud (1984) introduces another modification of the intensity (1.1) to describe the failures of a k-component system. Assuming the covariates are functions of the time spent since the last failure time, he defines a class of renewal Markov processes and derives statistical inference when the observed length of time tends to infinity. This asymptotic point of view is quite usual for statistics of processes [Snyder (1975) and Liptser and Shiryayev (1977)] and we consider here the same framework.

Received January 1986; revised September 1987.
AMS 1980 *subject classifications.* 62F12, 62G05, 62M99, 62P10.
Key words and phrases. Intensity, martingale, weak convergence of stochastic processes, ergodicity, mixing processes, periodic feeding pattern.

In this paper α as given in (1.1) is the intensity of a single point process observed on a long time interval and we assume that the underlying intensity λ_0 is an unknown periodic function with period 1. This assumption enables one to describe point processes having a certain kind of periodicity as, for example, the feeding times of an animal or the times of peaks in hormonal secretion. In the absence of the modifying regressors, that is, if $\beta_l = 0$ for $l = 1, \dots, q$, the process is a Poisson process with nonhomogeneous periodic intensity. The regression part of the intensity takes into account the effect of the past history through predictable processes $Z_1, \dots, Z_q$ and inference on $\beta_1, \dots, \beta_q$ is needed to decide whether the past of the process or other relevant random events modifies the underlying Poisson process. When studying a feeding pattern, the regressors may be, for example, the weights of food intakes as well as their times (short enough to be considered as points).

The periodicity of λ_0 leads to new estimators $\hat{\beta}_n$ and $\hat{\Lambda}_n$ for the regression parameters and the cumulative underlying intensity, and we show that they have the same asymptotic properties as in the classical Cox model. When the processes N and Z are observed on $[0, n]$, the Cox partial likelihood is replaced by

$$W_n(\beta_0) = \prod_{i=1}^{N(n)} \frac{\exp\{\beta_0 Z(T_i)\}}{\sum_{k=0}^{n-1} \exp\{\beta_0 Z(S_i + k)\}}, \tag{1.2}$$

where S_i is the point of the first period $]0, 1]$ which is equivalent to the observed time T_i, modulo 1, $i \leq N(n)$. Writing $\log W_n(\beta_0)$ as

$$\sum_{k=0}^{n-1} \int_0^1 \beta_0 Z(u + k)\, dN(u + k)$$
$$- \int_0^1 \log\left\{ \sum_{k=0}^{n-1} \exp\{\beta_0 Z(u + k)\} \right\} d\left\{ \sum_{j=0}^{n-1} N(u + j) \right\},$$

we can see that it has exactly the same form as the logarithm of the Cox likelihood evaluated at time 1 (AG). However, the process $\sum_{k=0}^{n-1} \exp\{\beta_0 Z(\cdot + k)\}$ is not predictable and, moreover, there is generally no history $\mathscr{F} = (\mathscr{F}_s)_{s \in]0,1]}$ on the probability space such that all the counting processes $N(\cdot + k) - N(k)$ are $\mathscr{F}$-adapted and have an $\mathscr{F}$-intensity, $0 \leq k \leq n - 1$. The proofs of AG have therefore to be modified and stronger conditions are needed.

Section 2 presents a complete specification of the model and general conditions for the asymptotic results. Also, these conditions are compared to those of AG. In Section 3, we establish the asymptotic properties of the estimators, using a martingale convergence theorem [Rebolledo (1978)] to prove weak convergence of variables, and theorems for convergence of probability measures [Billingsley (1968)] when processes are concerned. An example is given (Section 4) where a covariate studied by Slud (1984) is used. Finally the result is generalized to the case of a multiplicative intensity model (Section 5).

2. Definitions and model assumptions. The theory of stochastic integrals and martingales has been fruitfully introduced to study counting processes [Aalen (1978) and Rebolledo (1978)]. It has been used often and we shall apply it throughout this paper as a classical tool. We put our general framework in this background.

We consider an open set B of $\mathbb{R}^p$ with compact closure $\overline{B}$ and a family $\mathscr{L}$ of measures on $\mathbb{R}_+$ having a periodic Radon–Nikodym derivative with respect to Lebesgue measure, with period 1. Let $(\Omega, \mathscr{A})$ be a measurable space and $\mathscr{P} = \{P_{\beta, \Lambda_c}; \beta \in B, \Lambda_c \in \mathscr{L}\}$ be a family of probability measures on $(\Omega, \mathscr{A})$. A history $\mathscr{F} = (\mathscr{F}_t)_{t \geq 0}$ on $(\Omega, \mathscr{A}, \mathscr{P})$ is a right continuous nondecreasing family of P-complete sub-σ-algebra of $\mathscr{A}$ for any P in $\mathscr{P}$. We consider a simple counting process N associated with a real point process $(T_i)_{i \in \mathbb{N}}$ and we assume that N has a P_{β, Λ_c}-predictable compensator $\int_0^{\cdot} e^{\beta Z(t)}\, d\Lambda_c(t)$, where Z is a vector of q $\mathscr{F}$-predictable processes with left continuous sample paths. The processes N and Z are observed on a time interval $[0, n]$, $n \in \mathbb{N}$.

2.1. *Definition of the estimators.* Assuming $\mathscr{F}$ is the internal history of N, maximum likelihood estimators (MLEs) of β_0 and Λ_0 do not exist in $B \times \mathscr{L}$. Following Johansen (1983), an extension of the model may be defined replacing Λ_c by a measure $\Lambda(\cdot) = \Lambda_c(\cdot) + \sum_{u \leq \cdot} \Delta\Lambda(u)$, with Λ_c in $\mathscr{L}$ and $\Delta\Lambda$ having period 1, and replacing N by a marked counting process with integer marks. Because of the periodicity, an empirical distribution-type estimator of Λ_0 has to weigh only the points $S_i \equiv T_i$ modulo 1 and their periodic translates. In this model it follows that MLEs of Λ_0 and β_0 on $[0, n]$ have to maximize

$$V_n(\Lambda, \beta) = \prod_{i=1}^{N(n)} \Delta\Lambda(S_i) e^{\beta Z(T_i)} \prod_{k=0}^{n-1} \exp\{-\Delta\Lambda(S_i) e^{\beta Z(S_i + k)}\};$$

thus $\hat{\Lambda}_n(\cdot) = \sum_{u \leq \cdot} \Delta\hat{\Lambda}_n(u)$ with

$$\Delta\hat{\Lambda}_n(S_i) = \frac{1}{\sum_{k=0}^{n-1} \exp\{\beta Z(S_i + k)\}}, \tag{2.1}$$

and $\hat{\beta}_n$ maximizes $W_n(\beta) = V_n(\hat{\Lambda}_n, \beta)$, which reduces to (1.2).

Though this extension of the model is rather artificial in considering a minimal history and a marked point process when only a simple process is observed, it is interesting for two reasons: First, being a maximum likelihood estimator for some model, $\hat{\beta}_n$ can be expected to have the classical asymptotic properties of MLEs; second, this model gives an immediate method for the computation of $\hat{\beta}_n$ and $\hat{\Lambda}_n$ by analogy with the likelihood of Poisson variables [Whitehead (1980) and Pons and de Turckheim (1988)].

Following Gill (1986), MLEs β_n and Λ_n of β_0 and Λ_0 may also be obtained from an extension of the likelihood in parametric submodels. Let $(P_{\theta, \beta}; \theta \in \mathbb{R}, \beta \in B)$ be a family of probability measures associated with the compensator $\int_0^{\cdot} \{1 + \theta k(t)\} e^{\beta Z(t)}\, d\Lambda_n(t)$ of N, where k is a periodic function with period 1. Then the value of $(\theta, \beta) = (0, \beta_n)$ has to maximize the likelihood of this parametric submodel for any periodic function k. If Λ_n were continuous and $\mathscr{F}$

minimal for N, the Radon–Nikodym derivative $L_n(\theta, \beta)$ of $P_{\theta,\beta}$ with respect to $P_{0,0}$ on $[0, n]$ would be [Jacod (1975)]

$$L_n(\theta, \beta) = \prod_{T_i \le n} \left(\{1 + \theta k(T_i)\} e^{\beta Z(T_i)}\right) \exp \int_0^n \left(1 - \{1 + \theta k(t)\} e^{\beta Z(t)}\right) d\Lambda_n(t).$$

For a general history $\mathscr{F}$ and a general measure Λ_n in

$$\mathscr{D} = \left\{ \Lambda \text{ measure on } \mathbb{R}_+;\ \Lambda(\cdot) = \Lambda_c + \sum_{u \le \cdot} \Delta\Lambda(u),\ \Lambda_c \in \mathscr{L},\ \Delta\Lambda(\cdot) = \Delta\Lambda(\cdot + 1) \right\},$$

we consider the same expression $L_n(\theta, \beta)$ instead of the likelihood associated with a noncontinuous compensator. Maximizing L_n at $(0, \beta_n)$, it follows that $\sum_{T_i \le n} k(T_i) = \int_0^n k e^{\beta_n Z(t)}\, d\Lambda_n$ for any periodic function k; hence, Λ_n is a discrete measure and using the periodicity we get (2.1) and (1.2). It must be noticed that the extension of Johansen defines a model where the likelihood is exactly $L_n(0, \beta)$ with Λ_n in $\mathscr{D}$ when $\mathscr{F}$ is the minimal history.

2.2. *Conditions for convergence.* Some additional notations and definitions are useful: Let $\mathbb{P}$ be the probability associated with the parameters λ_0 and β_0; the expectation and the variance of a variable X with respect to $\mathbb{P}$ are denoted by $\mathbb{E}X$ and $\operatorname{Var} X$; otherwise, they are denoted by $E_\nu X$ and $\operatorname{Var}_\nu X$ when a probability ν is concerned. Let $A = \int_0^{\cdot} \alpha(u)\, du$ be the $\mathbb{P}$-predictable compensator of N, where α is given by (1.1) and let M be the local martingale $N - A$.

The restrictions of N, A and Z to the kth period are the processes N_k^*, A_k^* and Z_k^* defined on $[0,1]$ by

$$\begin{aligned} N_k^*(s) &= N(s + k) - N(k), \\ A_k^*(s) &= A(s + k) - A(k), \qquad (2.2) \\ Z_k^*(s) &= Z(s + k), \qquad 0 \le k \le n - 1. \end{aligned}$$

C is the space of continuous functions on $[0,1]$ and D_- is the space of left continuous functions on $[0,1]$ with right-hand limits. Thus the processes Z_k^* have their sample paths in D_-^q. Let $Z^{\otimes i}$ for $i = 0,1,2$, be 1, Z and the matrix ZZ', respectively. For each β in $\bar{B}$, processes $S_n^{(i)}(\beta)$ are defined on $[0,1]$ by

$$S_n^{(i)}(\beta, s) = \frac{1}{n} \sum_{k=0}^{n-1} Z^{\otimes i}(s + k) \exp\{\beta Z(s + k)\}, \qquad i = 0,1,2, \qquad (2.3)$$

and they are extended to $\mathbb{R}_+$ by periodicity.

We can now set down the following conditions, which will be assumed to hold throughout the proofs.

CONDITION A. The sequence of processes $(Z_k^*)_{k \ge 0}$ is $\mathbb{P}$-ergodic: There exists a probability ν on $(\Omega, \mathscr{A})$ such that for each measurable set B for the Skorohod topology on D_-^q,

$$n^{-1} \sum_{k=0}^{n-1} I(Z_k^* \in B) \to_{\mathbb{P}} \nu(Z_0^* \in B).$$

CONDITION B. The sequence of processes $(Z_k^*)_{k\geq 0}$ is φ-mixing with $\Sigma_{k\geq 0}\varphi(k)^{1/2} < \infty$, where $\sup\{|\mathbb{P}(B/A) - \mathbb{P}(B)|;\ A \in \mathscr{M}_0^i,\ B \in \mathscr{M}_{i+k}^{\infty},\ i \in \mathbb{N}\} \leq \varphi(k)$ and where $\mathscr{M}_0^i$ and $\mathscr{M}_{i+k}^{\infty}$ are the σ-algebras generated by $\{Z_m^*;\ 0 \leq m \leq i\}$ and $\{Z_m^*;\ i+k \leq m\}$, respectively.

Let $s^{(i)}(\beta, s) = E_\nu Z^{\otimes i}(s)\exp\{\beta Z(s)\}$ on $\overline{B} \times [0,1]$ and be extended to $\mathbb{R}_+$ by periodicity for each β in $\overline{B}$, $i = 0, 1, 2$.

CONDITION C. Integrability and regularity conditions:

C1. $\sup_{\beta\in\overline{B}}\sup_{[0,1]}|Z^{\otimes i}\exp\{\beta Z\}|$ is ν-integrable and $\sup_{\mathbb{R}_+}|Z^{\otimes i}\exp\{\beta_0 Z\}|$ is $\mathbb{P}$-integrable, $i = 0, 1, 2$.

C2. $s^{(0)}$ is bounded away from zero on $\overline{B} \times [0,1]$ and the functions $s^{(i)}(\beta_0)$ are continuous on $[0,1]$ for $i = 0, 1$.

C3. $\mathrm{Var}_\nu Z(u)$ is positive definite for each u in a dense subset $\mathscr{U}$ of $[0,1]$ with a positive Λ_0-measure.

C4. Each of $\sup_{k\geq 0}(\mathrm{Var}\exp\{\beta_0 Z_k^*(s)\})$, $\sup_{k\geq 0}\mathrm{Var}(Z_k^*(s)\exp\{\beta_0 Z_k^*(s)\})$ and $n^{1/2}(\mathbb{E}S_n^{(i)}(\beta_0, s) - E_\nu S_n^{(i)}(\beta_0, s))$ is finite for each s in a dense subset of $[0,1]$ including 0 and 1.

C5. Lindeberg condition: For $l = 1, \dots, q$,

$$\limsup_{A\to\infty} E_\nu \int_0^1 Z_l^2 I\{|Z_l| > A\}\exp\{\beta_0 Z\}\, d\Lambda_0 = 0.$$

C6. The sequence of processes $(C_n)_n$ is tight for the uniform topology on $\underline{D}^{q+1}$, where $C_n = (C_n^{(0)}, C_n^{(1)})$ is defined by

$$C_n^{(i)} = n^{1/2}\big(S_n^{(i)} - s^{(i)}\big)(\beta_0), \qquad i = 0, 1. \tag{2.4}$$

These conditions are rather general and we shall see in the examples given in Section 4.1 that Markovian properties imply some of the required probabilistic conditions. Lemmas 2.1–2.3 show how they relate to the conditions of AG.

LEMMA 2.1. *If Conditions* A *and* C1 *hold, then*

$$\sup_{[0,1]}\big|S_n^{(i)}(\beta) - s^{(i)}(\beta)\big| \to_{\mathbb{P}} 0 \quad \textit{for each } \beta \textit{ in } \overline{B}.$$

PROOF. The ergodicity is equivalent to the condition $n^{-1}\Sigma_{k=0}^{n-1} f(Z_k^*) \to_{\mathbb{P}} E_\nu f(Z_0^*)$ for any ν-integrable function $f\colon \underline{D}^q \to \mathbb{R}$. Next, for such a function f, we can extend Rao's theorem (1963) to the ergodic sequence of variables $(f(Z_k^*))_{k\geq 0}$ satisfying the relevant integrability condition

$$E_\nu\Big\{\sup_{s\in[0,1]} f\big(Z_0^*(s)\big)\Big\} < \infty.$$

The required result then follows from C1. □

LEMMA 2.2. *If* C1 *holds, then for each* β *in* $\overline{B}$ *and* s *in* $[0,1]$, $s^{(1)}(\beta, s) = (\partial/\partial\beta)s^{(0)}(\beta, s)$ *and* $s^{(2)}(\beta, s) = (\partial^2/\partial\beta^2)s^{(0)}(\beta, s)$; $s^{(0)}$, $s^{(1)}$ *and* $s^{(2)}$ *are* *bounded on* $\overline{B} \times [0,1]$. *Moreover* C2 *and* C3 *imply that the matrix*

$$\Sigma(\beta) = \int_0^1 \frac{s^{(0)}s^{(2)} - s^{(1)\otimes 2}}{(s^{(0)})^2}(\beta)s^{(0)}(\beta)\,d\Lambda_0 \tag{2.5}$$

is positive definite.

PROOF. The first part of the lemma is a consequence of the dominated convergence theorem. For the second part, note that when C2 holds, the following statements are equivalent to C3:

(i) $\gamma Z(s)$ is not deterministic for each s in $\mathscr{U}$ and γ' in $\mathbb{R}^q_-$.

(ii) $E_\nu\{(\gamma Z(s) + a(s))^2 e^{\beta Z(s)}\} > 0$ for each s in $\mathscr{U}$, γ' in $\mathbb{R}^q$ and each real function $a(s)$.

(iii) Polynomial expansion in $a(s)$ of the previous expression has a negative discriminant function for each s in $\mathscr{U}$ and γ' in $\mathbb{R}^q$.

(iv) $\Sigma(\beta)$ is positive definite. □

LEMMA 2.3. *If* A *and* C1 *hold, then for* $i = 0,1,2$,

$$\sup_{|\beta_0-\beta|\le\rho}\ \sup_{[0,1]} \left|S_n^{(i)}(\beta_0) - S_n^{(i)}(\beta)\right| \to_{\mathbf{P}} 0,$$

when $n \to \infty$ *and* $\rho \to 0$.

PROOF. Let

$$w_\rho^{(i)}(Z) = \sup_{|\beta_0-\beta|\le\rho}\ \sup_{s\in[0,1]} \left|Z^{\otimes i}(s)e^{\beta_0 Z(s)} - Z^{\otimes i}(s)e^{\beta Z(s)}\right|.$$

The variable $\sup_{|\beta_0-\beta|\le\rho}\sup_{[0,1]}|S_n^{(i)}(\beta_0) - S_n^{(i)}(\beta)|$ has bound $n^{-1}\sum_{k=0}^{n-1} w_\rho^{(i)}(Z_k^*)$, which converges to $E_\nu w_\rho^{(i)}(Z_0^*)$. Now $\lim_{\rho\to 0} E_\nu w_\rho^{(i)}(Z_0^*) = 0$, which concludes the proof. □

We can remark that our conditions imply weaker ones than in AG (Lemmas 2.1 and 2.2); however, they are sufficient to prove the same asymptotic properties of the estimators. In particular the result of Lemma 2.3 replaces the continuity in $\overline{B}$, uniformly on $[0,1]$, of the functions $s^{(0)}$, $s^{(1)}$ and $s^{(2)}$ and a stronger convergence of $S_n^{(i)}$ to $s^{(i)}$, uniformly on $\overline{B} \times [0,1]$. The uniqueness of $\hat{\beta}_n$ will be a consequence of the positive definiteness of $\Sigma(\beta_0)$ and, under C2, this is equivalent to the intuitive Condition C3 (Lemma 2.2). Finally our Lindeberg condition is stronger than in AG, but it could be written in a similar form.

3. Asymptotic properties.

3.1. *Consistency and asymptotic normality of* $\hat{\beta}_n$. The estimator $\hat{\beta}_n$ has been chosen to minimize the function

$$K_n(\beta) = n^{-1}\log\{W_n(\beta_0)W_n^{-1}(\beta)\} \tag{3.1}$$

and its asymptotic properties follow from minimum contrast theory [Dacunha-Castelle and Duflo (1986)], which generalizes maximum (partial) likelihood theory.

THEOREM 3.1. *$\hat{\beta}_n$ is a weakly consistent estimator of β_0 in B.*

PROOF. We have to show that $K_n(\beta)$ converges in probability to a continuous function $K(\beta)$ having a unique minimum at β_0 in $\bar{B}$. $K_n(\beta)$ can be written

$$K_n(\beta) = n^{-1}\int_0^n (\beta_0 - \beta) Z\,dN + n^{-1}\int_0^n \left\{\log\frac{S_n^{(0)}(\beta_0)}{S_n^{(0)}(\beta)}\right\} dN.$$

Since the processes $S_n^{(0)}(\beta)$ and $S_n^{(0)}(\beta_0)$ are not predictable, expand $K_n(\beta)$ as

$$n^{-1}\int_0^n \left\{(\beta_0 - \beta)Z - \log\frac{s^{(0)}(\beta_0)}{s^{(0)}(\beta)}\right\} dN$$

$$-n^{-1}\int_0^n \left\{\log\frac{S_n^{(0)}(\beta_0)}{S_n^{(0)}(\beta)} - \log\frac{s^{(0)}(\beta_0)}{s^{(0)}(\beta)}\right\} dN.$$

Next, using the ergodicity condition and Lenglart's (1977) inequality, the first term in this expression converges in probability to

$$K(\beta) = (\beta_0 - \beta)\int_0^1 s^{(1)}(\beta_0)\,d\Lambda_0 - \int_0^1 \log\left\{\frac{s^{(0)}(\beta_0)}{s^{(0)}(\beta)}\right\} s^{(0)}(\beta_0)\,d\Lambda_0$$

and the second term converges to zero [from Lemma 2.1 and convergence in probability of $n^{-1}N(n)$].

K is a continuous function in $\bar{B}$. Its first derivative has the value zero at $\beta = \beta_0$ and its second derivative is $\Sigma(\beta)$, given by (2.5). The consistency of $\hat{\beta}_n$ follows from Lemma 2.2 and a classical convex analysis theorem (cf. AG). □

THEOREM 3.2. *$n^{1/2}(\hat{\beta}_n - \beta_0)$ converges weakly to a Gaussian variable $\mathcal{N}(0, \Sigma(\beta_0)^{-1})$ and $(\partial^2/\partial\beta^2)K_n(\hat{\beta}_n)$ is a weakly consistent estimator of $\Sigma(\beta_0)$.*

PROOF. By a Taylor expansion of $(\partial/\partial\beta)K_n(\hat{\beta}_n)$ around β_0 we obtain $n^{1/2}(\partial/\partial\beta)K_n(\beta_0) + n^{1/2}(\hat{\beta}_n - \beta_0)(\partial^2/\partial\beta^2)K_n(\beta_n^*) = 0$, where β_n^* is between β_0 and $\hat{\beta}_n$. It is therefore sufficient to show

$$(\partial^2/\partial\beta^2)K_n(\beta_n^*) \to_{\mathrm{P}} \Sigma(\beta_0)$$

for any weakly consistent estimator β_n^* of β and

$$n^{1/2}(\partial/\partial\beta)K_n(\beta) \to_{\mathcal{D}} \mathcal{N}(0, \Sigma(\beta_0)).$$

The first point is a consequence of $(\partial^2/\partial\beta^2)K_n(\beta_0) \to_{\mathrm{P}} \Sigma(\beta_0)$, which can be proved as in Theorem 3.1, and of

$$\sup_{|\beta_0 - \beta| \le \rho} |(\partial^2/\partial\beta^2)K_n(\beta) - (\partial^2/\partial\beta^2)K_n(\beta_0)| \to_{\mathrm{P}} 0$$

when $n \to \infty$ and $\rho \to 0$, which follows from Lemma 2.3.

The second point needs a representation of $-n^{1/2}(\partial/\partial\beta)K_n(\beta_0)$ as the difference of two random variables X_n^1 and X_n^2, where

$$X_n^1 = n^{-1/2}\int_0^n\left\{Z - \frac{s^{(1)}}{s^{(0)}}(\beta_0)\right\}dM,$$

$$X_n^2 = n^{-1/2}\int_0^n\left\{\frac{S_n^{(1)}}{S_n^{(0)}} - \frac{s^{(1)}}{s^{(0)}}\right\}(\beta_0)(dN - dA).$$

X_n^1 is the value at 1 of the process

$$Q_n = \int_0^{\cdot} n^{-1/2}\left(Z(ns) - \frac{s^{(1)}}{s^{(0)}}(\beta_0, ns)\right)dM(ns),$$

which is a square integrable local martingale on $[0,1]$ with respect to the history $(\mathscr{F}_{ns})_{s\in[0,1]}$. Using C5 and the convergence in probability of $\langle Q_n, Q_n\rangle(s)$ to $s\Sigma(\beta_0)$, for any s in $[0,1]$ we can apply Rebolledo's convergence theorem and conclude that $X_n^1 \to_{\mathscr{D}} \mathscr{N}(0, \Sigma(\beta_0))$.

In order to show that $X_n^2 \to_{\mathbb{P}} 0$, write $X_n^2 = \int_0^1 B_n(d\overline{N}_n - d\overline{A}_n)$ with

$$\begin{aligned} B_n &= n^{1/2}\big(S_n^{(1)}/S_n^{(0)} - s^{(1)}/s^{(0)}\big)(\beta_0), \\ \overline{N}_n &= n^{-1}\sum_{k=0}^{n-1} N_k^*, \\ \overline{A}_n &= n^{-1}\sum_{k=0}^{n-1} A_k^*. \end{aligned} \tag{3.2}$$

and consider the nondecreasing continuous function $\Phi = \int_0^{\cdot} s^{(0)}(\beta_0)\,d\Lambda_0$. Note that X_n^2 is not a stochastic integral but a difference of two Stieltjes integrals and that $(\overline{N}_n - \overline{A}_n)$ is not a martingale. We will prove that B_n converges weakly to a process B with continuous sample paths and that $\sup_{[0,1]}|\overline{N}_n - \Phi|$ and $\sup_{[0,1]}|\overline{A}_n - \Phi|$ converge in $\mathbb{P}$-probability to zero. Then X_n^2 converges in $\mathbb{P}$-probability to zero as a function of $(B_n, \overline{N}_n, \overline{A}_n)$, which is continuous at the values (B, Φ, Φ), for the Skorohod topology on $D_-^q \times E^2$, where E is the set of nondecreasing right continuous functions on $[0,1]$, with left-hand limits. The process B_n itself may be written as $f(C_n^{(1)}, C_n^{(0)}, S_n^{(0)}(\beta_0))$, where f is the function from D_-^{q+2} to D_-^q defined by

$$f(x, y, z) = \left(x - \frac{s^{(1)}(\beta_0)}{s^{(0)}(\beta_0)}y\right)z$$

for x in D_-^q and y and z in D_-. Under Condition C2, the function f is continuous on C^{q+2}. Since $\sup_{[0,1]}|S_n^{(0)}(\beta_0) - s^{(0)}(\beta_0)| \to_{\mathbb{P}} 0$ and $s^{(0)}(\beta_0)$ is in C, the weak convergence of C_n to a process with continuous sample paths will ensure that B_n also converges weakly to a process with continuous sample paths. Now for any t in a dense subset of $[0,1]$ including 0 and 1, each component of $C_n(t)$ is a sum of n φ-mixing variables of the form $n^{-1/2}\sum_{k=0}^{n-1}(\xi_k - E_\nu\xi_k)$ and it converges weakly from C4 and from a generalization to the nonstationary case of

Theorem 20.1 of Billingsley (1968). From C4 and the Cramér–Wold device, C_n has then convergent finite-dimensional distributions, and by C6 it converges weakly to a process with continuous sample paths. Then the same holds for B_n. It remains to prove that $\sup_{[0,1]}|\overline{A}_n - \Phi| \to_{\mathbf{P}} 0$ and $\sup_{[0,1]}|\overline{N}_n - \Phi| \to_{\mathbf{P}} 0$; this is a consequence of Lemma 2.1 for $\overline{A}_n$ and of Lenglart's inequality for the local martingale $n^{-1}\int_0^{\cdot} I\{u \in]0, s]\}\, dM(u)$ which yields $|\overline{N}_n(s) - \overline{A}_n(s)| \to_{\mathbf{P}} 0$ at any point s in $[0,1]$. Now the continuity of Φ gives the required result. □

3.2. *Asymptotic distribution of* $n^{1/2}(\hat{\Lambda}_n - \Lambda_0)$. The asymptotic properties of $(\hat{\Lambda}_n - \Lambda_0)$ require further technical results: We have to establish weak convergence of a process written as a Stieltjes integral with respect to $(\overline{N}_n - \overline{A}_n)$. Though $(\overline{N}_n - \overline{A}_n)$ is not a local martingale, we can prove that the process

$$\mu_n = n^{1/2}(\overline{N}_n - \overline{A}_n) \tag{3.3}$$

converges weakly to a Gaussian process with zero mean, independent increments and continuous sample paths. The convergence of its finite-dimensional distributions uses Rebolledo's convergence theorem and its C-tightness needs an adaptation of Billingsley's method for the empirical process associated with a φ-mixing sequence of variables [Billingsley (1968), Section 22]. A stronger φ-mixing condition is necessary to bound the fourth moment of μ_n. We now assume

CONDITION D. $(N_k^*, Z_k^*)_{k \geq 0}$ is a φ-mixing sequence of processes with

$$\sum_{k=0}^{n-1} (k+1)^{l-2}\varphi(k)^{1/l} < \infty \quad \text{for } l = 2,3,4$$

and we prove some preliminary results.

LEMMA 3.1. *If λ_0 is bounded, there exist constants K_1 and K_2 such that for any s and t in* $[0,1]$,

$$\mathbb{E}(\mu_n(t) - \mu_n(s))^4 \leq K_1(t-s)^2 + K_2\frac{|t-s|}{n}.$$

PROOF. For any $s \leq t$ in $[0,1]$, $\mu_n(t) - \mu_n(s) = n^{-1/2}\sum_{k=0}^{n-1}\int_s^t dM(\cdot + k)$ and $(\int_s^t dM(\cdot + k))_{k \geq 0}$ is a sequence of φ-mixing variables with zero mean. From Doukhan and Portal's Theorem 2.5 (1983) and Condition D, we get, with a constant c_4,

$$\begin{aligned} &\mathbb{E}(\mu_n(t) - \mu_n(s))^4 \\ &\quad \leq c_4\left(\sup_{k\geq 0}\left\{\mathbb{E}\left|\int_s^t dM(u+k)\right|^2\right\}^2 + n\sup_{k \geq 0}\mathbb{E}\left|\int_s^t dM(u+k)\right|^4\right). \end{aligned} \tag{3.4}$$

We have

$$\mathbb{E}\left|\int_s^t dM(u+k)\right|^2 = \mathbb{E}\left|\int_s^t dA(u+k)\right|,$$

which is bounded by

$$|t-s|\sup_{[0,1]}|\lambda_0|\mathbb{E}\left\{\sup_{\mathbb{R}_+} e^{\beta_0 Z}\right\}$$

for any integer k.

Considering $\int_s^t dM(u+k)$ as the value at point 1 of the square integrable local martingale $\int_0^{\cdot} I\{u \in]s,t]\}\, dM(u+k)$ and using the Burkholder–Davis–Gundy (1972) inequality for local martingales, we also have, for a constant a_4,

$$\begin{aligned}\mathbb{E}\left|\int_s^t dM(u+k)\right|^4 &\le a_4 \mathbb{E}\left|\int_0^1 I\{u\in]s,t]\}\, dN(u+k)\right|^2\\ &= a_4\mathbb{E}\left|\int_s^t dM(u+k) + \int_s^t dA(u+k)\right|^2\\ &\le 2a_4\left(\mathbb{E}\left|\int_s^t dA(u+k)\right| + \mathbb{E}\left|\int_s^t dA(u+k)\right|^2\right),\end{aligned}$$

which is bounded by $|t-s|$, except for a multiplicative constant, from C4. By introducing these bounds in (3.4), we get the required result. □

LEMMA 3.2. *If λ_0 is bounded, the sequence of processes $(\mu_n)_n$ is C-tight.*

PROOF. Let $w(\mu_n, \delta)$ be the δ-modulus of continuity of μ_n,

$$a(\eta) = \eta^{-1}\sup_{[0,1]}|\lambda_0|\mathbb{E}\left\{\sup_{\mathbb{R}_+} e^{\beta_0 Z}\right\}\quad \text{for } \eta > 0,$$

and

$$\Omega(p,a) = \left\{w(\bar{A}_n, p) \le ap\right\}\quad \text{for } p \text{ in }]0,1[\text{ and } a \text{ in } \mathbb{R}_+.$$

Since $\mu_n(0)$ converges weakly (Rebolledo's theorem) and $\mathbb{P}\{\Omega(p, a(\eta))\} \ge 1-\eta$ for any p and η, it is sufficient to prove that for any nonnegative ε and η, there exist δ and p in $]0,1[$ such that, when n is large enough,

$$\mathbb{P}\left\{\Omega(p,a(\eta)) \cap \left\{\sup_{s\le t\le s+\delta}|\mu_n(t)-\mu_n(s)| > \varepsilon\right\}\right\} \le \delta\eta$$

for each s in $[0,1]$ [Billingsley (1968) Theorem 8.3].

Now, this can be shown just as Billingsley does [(1968) Section 22] because of Lemma 3.1 and the inequality

$$|\mu_n(t)-\mu_n(s)| \le |\mu_n(s+p)-\mu_n(s)| + apn^{1/2},\qquad s\le t\le s+p,$$

which is true on $\Omega(p, a(\eta))$. □

THEOREM 3.3. *If λ_0 is bounded and D holds, then the process*

$$n^{1/2}(\hat{\Lambda}_n - \Lambda_0) + n^{1/2}(\hat{\beta}_n-\beta_0)\int_0^{\cdot} s^{(1)}(\beta_0)\{s^{(0)}(\beta_0)\}^{-1}\, d\Lambda_0$$

converges weakly to a Gaussian process with continuous sample paths, zero mean and covariance $\int_0^{s\wedge t}\{s^{(0)}(\beta_0)\}^{-1}\, d\Lambda_0$ at points s and t.

PROOF. As $\hat{\Lambda}_n(s) = \int_0^s \{S_n^{(0)}(\hat{\beta}_n)\}^{-1} d\overline{N}_n$, the process $n^{1/2}(\hat{\Lambda}_n - \Lambda_0)$ can be written as a sum of three terms,

$$n^{1/2}\big(\hat{\Lambda}_n(s) - \Lambda_0(s)\big) = n^{1/2}\left(\int_0^s \frac{d\overline{N}_n - d\overline{A}_n}{S_n^{(0)}(\hat{\beta}_n)} + \int_0^s \frac{d\overline{A}_n}{S_n^{(0)}(\hat{\beta}_n)} - \int_0^s d\Lambda_0\right).$$

Taylor expanding $(S_n^{(0)}(\hat{\beta}_n))^{-1}$ around β_0, the last two terms are replaced by

$$n^{1/2}\big(\beta_0 - \hat{\beta}_n\big)\int_0^s \frac{S_n^{(1)}}{S_n^{(0)2}}(\tilde{\beta}_n) S_n^{(0)}(\beta_0)\, d\Lambda_0,$$

where $\tilde{\beta}_n$ is between β_0 and $\hat{\beta}_n$. Then by ergodicity and weak convergence of $n^{1/2}(\hat{\beta}_n - \beta_0)$, the processes

$$\text{(3.5)} \qquad \begin{gathered} n^{1/2}\big(\hat{\Lambda}_n - \Lambda_0\big) + n^{1/2}\big(\hat{\beta}_n - \beta_0\big)\int_0^{\cdot} s^{(1)}(\beta_0)\{s^{(0)}(\beta_0)\}^{-1} d\Lambda_0, \\ n^{1/2}\int_0^{\cdot}\{S_n^{(0)}(\hat{\beta}_n)\}^{-1}\big(d\overline{N}_n - d\overline{A}_n\big) \end{gathered}$$

have the same limiting distribution if they converge. The expression (3.5) can again be split into two terms,

$$n^{1/2}\int_0^{\cdot}\Big(\{S_n^{(0)}(\hat{\beta}_n)\}^{-1} - \{s^{(0)}(\beta_0)\}^{-1}\Big)\big(d\overline{N}_n - d\overline{A}_n\big),$$

which converges to zero [expand $\{S_n^{(0)}(\hat{\beta}_n)\}^{-1}$ around β_0 and use $\sup_{[0,1]}|\overline{N}_n - \overline{A}_n| \to_{\mathbf{P}} 0$, Lemma 2.3 and weak convergence of C_n as in Theorem 3.2] and

$$\text{(3.6)} \qquad R_n = \int_0^{\cdot}\{s^{(0)}(\beta_0)\}^{-1} d\mu_n.$$

Now $(R_n)_n$ is C-tight because $s^{(0)}(\beta_0)$ is bounded away from zero and by Lemma 3.2. Moreover the convergence of its finite-dimensional distributions can be shown by considering $R_n(s)$ as the value at $t = 1$ of

$$G_n(t) = \int_0^t \frac{J_s(nu)\, dM(nu)}{n^{1/2} s^{(0)}(\beta_0, nu)},$$

where $J_s(u) = \sum_{k \geq 0} I\{u\colon k < u \leq k + s\}$. G_n is a square integrable local martingale on $[0,1]$, with respect to $(\mathscr{F}_{ns})_{s \in [0,1]}$, and it converges weakly to a Gaussian process with continuous sample paths and variance $\sigma_s(t) = t\int_0^s \{s^{(0)}(\beta_0)\}^{-1} d\Lambda_0$ (Rebolledo's theorem). It follows that R_n converges weakly to a Gaussian process having the required properties. □

Convergence of $n^{1/2}(\hat{\Lambda}_n - \Lambda_0)$ is then a simple consequence of the convergence of the process

$$\left(R_n, n^{1/2}\big(\hat{\beta}_n - \beta_0\big)\int_0^{\cdot} \frac{s^{(1)}}{s^{(0)}}(\beta_0)\, d\Lambda_0\right)$$

with respect to the Skorohod topology on D_+^{q+1} since the second component has continuous sample paths. The covariance of its limiting distribution follows from the asymptotic null correlation of the variables $R_n(s)$ and $(d/d\beta)K_n$ for the probability ν,

$$\lim E_\nu\left\{R_n(s)n^{1/2}\frac{dK_n}{d\beta}(\beta_0)\right\} = \lim E_\nu\left\{\int_0^n \frac{J_s}{s^{(0)}(\beta_0)}\left(Z - \frac{s^{(1)}}{s^{(0)}}(\beta_0)\right) dA\right\} = 0.$$

The computation of the asymptotic covariance of $n^{1/2}(\hat{\Lambda}_n - \Lambda_0)$ is then straightforward and is, at points s and t,

$$\int_0^{s\wedge t}\frac{d\Lambda_0}{s^{(0)}(\beta_0)} + \left(\int_0^s \frac{s^{(1)}}{s^{(0)}}(\beta_0)\, d\Lambda_0\right)' \Sigma(\beta_0)^{-1}\int_0^t \frac{s^{(1)}}{s^{(0)}}(\beta_0)\, d\Lambda_0.$$

4. Special cases and an example.

4.1. *Some special cases.* The case of bounded processes Z_l is common in practical situations and Conditions C1, C2, C4 and C5 are then nearly fulfilled. The condition of full-rank randomness of Z is also natural: If the different processes $Z_1,\dots,Z_q$ are linearly dependent some of them should be dropped as noninformative. To actually compute $\hat{\beta}_n$, we need a condition corresponding to the empirical version of C3: The matrix $(Z_l(U_j))_{l,j}$ is of rank q on the observed sample path, where the U_j's are all periodic translates of the observed points lying in the observed time interval.

In the general case, when the processes Z_l take into account more information than the history of N, the ergodic and mixing properties of $(Z_k^*)_{k\ge 0}$ have to look plausible. However, when the processes Z_l are known functions of N itself, the intensity (1.1) defines the distribution of (N, Z). We therefore have to prove that these conditions are fulfilled.

More generally, suppose that $Z(t)$ depends on the points of the process N observed in the time interval $[t-1, t[$ and on marks associated with these points in the following way: If N has m points $T_{l+1},\dots,T_{l+m}$ on $[t-1,t[$,

$$Z(t) = g_m(t, T_{l+1},\dots,T_{l+m}, X_{l+1},\dots,X_{l+m}),$$

where X_j is a mark associated with T_j and g_m is an invariant function for a common translation of its first $(m+1)$ arguments, $m \ge 1$, g_0 being a constant of $\mathbb{R}^q$.

In that case and under a certain condition on the distribution of the marks, the sequences of processes $(Z_k^*)_{k\ge 0}$ and $(N_k^*, Z_k^*)_{k\ge 0}$ are homogeneous Markov chains; they have a recurrent point when $\sup_{k\ge 0}\int_k^{k+1}\alpha(u)\,du$ is bounded [Pons and de Turckheim (1986)]. They are therefore Doeblin recurrent chains, so that Conditions A, B and D hold and $\lim n^{1/2}\sum_{k=0}^{n-1}(\mathbb{E}S_n^{(i)}(\beta_0, s) - E_\nu S_n^{(i)}(\beta_0, s)) = 0$ for any s in $[0,1]$. As a particular case, let Z_l be the time elapsed since the last but $(l-1)$th event of N and let Z_l be set to 1 when this time is longer than one period: For $l = 1,\dots,p$,

(4.1) $$Z_l(t) = (t - T_{i-l+1}) \wedge 1 \quad \text{when } T_i < t \le T_{i+1}.$$

Another process can be defined as the mark X_i observed at time T_i, and we bound it by a fixed value x_0 so that the required conditions are fulfilled: For $l = p+1,\dots,2p$,

(4.2) $$Z_l(t) = (X_{i-l+1} \wedge x_0) I\{t - T_{i-l+1} \le 1\} \quad \text{when } T_i < t \le T_{i+1}.$$

For the process $Z = (Z_l)_{l \le 2p}$ defined by (4.1) and (4.2), the functions $s^{(i)}(\beta_0)$ are continuous on $[0,1]$, $i = 0, 1$. Moreover the C-tightness of the related sequence $(C_n)_n$ can still be established by a modification of Billingsley's method for empirical distribution function of φ-mixing variables: The processes Z_l are bounded and have no more jumps than the counting process N. We give elsewhere general conditions that guarantee Condition C6 and we apply them to the processes defined by (4.1) and (4.2) and to other processes [Pons and de Turckheim (1986)].

4.2. *Application.* The feeding times of rabbits have been recorded during seven days, as well as the weight of each food intake. The point process of these feeding times could be modelled by a Poisson process with a nonhomogeneous one day periodic intensity [Jolivet, Reyne and Teyssier (1983)], but that would suppose the process has no memory. The possibility of a dependence of the feeding times on the preceding times and quantities of food is considered using Cox's periodic model with the regressors (4.1) and (4.2), where the measuring unit for the T_i's is one day. For $p = 6$ they define the models M60 (regressors $Z_1, \dots, Z_6$), M06 (regressors $Z_7, \dots, Z_{12}$) and M66 (regressors $Z_1, \dots, Z_{12}$); for $p = 1$ the corresponding models are M10, M01 and M11. These models have been fitted to the data of an 18 week old domestic rabbit that proceeded to 219 food intakes in seven days.

Table 1 shows the values of the likelihood ratio statistic $\mathrm{LR} = -2\log(W_n(\hat{\beta})/W_n(\hat{\hat{\beta}}))$ to test a nested model M_0 with estimate $\hat{\hat{\beta}}$ in a model M with estimate $\hat{\beta}$; it is asymptotically distributed as a χ^2 variable under M_0. The following points may be noted:

(i) Each of the fitted models is better at level 0.001 than the Poisson model: The past summarized by the chosen processes has an effect on the distribution of the eventual intakes.

(ii) Comparing M11 to M66, M10 to M60 or M01 to M06 we see that the second to sixth preceding intakes do not add significant information to the first one.

TABLE 1
Likelihood ratio statistics and corresponding degrees of freedom for different nested models

	LR (df) for submodels					
Model	**Poisson**	**M10**	**M01**	**M11**	**M60**	**M06**
M10	47.6 (1)					
M01	15.7 (1)					
M11	93.9 (2)	46.3 (1)	78.2 (1)			
M60	51.1 (6)	3.9 (5)				
M06	17.0 (6)		1.3 (5)			
M66	104.0 (12)			10.1 (10)	52.6 (6)	87.0 (6)

TABLE 2
Estimates in models M11, M10 and M01

Covariate Z_l	$\hat{\beta}_l$	$\{\hat{\mathrm{V}}\mathrm{ar}(\hat{\beta}_l)\}^{1/2}$	Wald statistics
Z_1	0.0270	0.0034	7.84
	(0.0185)	(0.0029)	(6.28)
Z_2	−0.0023	0.0003	−6.53
	(−0.0012)	(0.0003)	(−3.96)

(iii) The weight of the preceding intake does add information to its time (comparison of M10 to M11) and vice versa.

(iv) Comparing M01 to M10 we see that M10 corresponds to a better fit since the likelihood ratio statistic with respect to the Poisson model may be understood as a goodness-of-fit criterion in the class of models with the intensity $\alpha(t) = \lambda_0(t)F(Z(t))$, where F is any nonnegative function and Z is a given maximal choice of processes to summarize the past [Pons and de Turckheim (1988)].

Table 2 gives the estimates of β_1 and β_2 in model M11 and, in brackets, in M10 and M01, respectively. The signs are the expected ones: The more has been eaten or the shorter the time since the previous intake, the lower the probability of observing the next intake. Though they keep the same sign in the submodels, $\hat{\beta}_1$ and $\hat{\beta}_2$ change when the other regressor is omitted: Their absolute value is underestimated when a relevant regressor is not considered. This agrees with the result of Bretagnolle and Huber (1985) in the case of dichotomous time independent regressors for survival analysis models. Because of this instability, the coefficients are difficult to interpret just as in the classical linear regression model with nonorthogonal regressors. Thus the easiest use of these models consists in comparing models taking different durations into account or summarizing the past by different processes [Pons and de Turckheim (1988)].

Finally, the estimate of Λ_0 is also available from the same program that computes $\hat{\beta}_n$, but here it is not interpretable since the regressors also have some kind of periodicity: The periodicity expressed by λ_0 is only a part of the model periodicity. It is a situation analogous to that of confounding effects in analysis of variance models and λ_0 should only be considered as a nuisance parameter. The estimation of Λ_0 could have a biological meaning in the case of discrete regressors as, for example, a regressor taking the value 0 or 1 according to the health of the rabbit.

5. Extensions. The particular exponential form of the Cox model has no important role and the previous results can be extended to a more general class of models, where $e^{\beta_0 Z}$ is replaced by $r(\beta_0 Z)$ with a real known function r having relevant properties [cf. Prentice and Self (1983)]. The case of a multiplicative intensity model with a periodic underlying intensity can also be considered:

Assume now that on $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F})$ a real counting process N has the intensity

$$\alpha(t) = \lambda_0(t)Z(t), \qquad t \geq 0,$$

where λ_0 is a periodic function with period 1 and Z a real predictable process. Using the previous arguments, the cumulative underlying intensity $\Lambda_0 = \int_0^{\cdot}\lambda_0(u)\,du$ is estimated by $\hat{\Lambda}_n$, defined as

$$\hat{\Lambda}_n(s) = \int_0^s n\left\{\sum_{k=0}^{n-1} Z(u+k)\right\}^{-1} d\bar{N}_n(u), \qquad 0 \leq s \leq 1.$$

Assume that Z is left continuous. Considering Z_k^*, the restriction of Z to the kth period and $S_n^{(0)} = (1/n)\sum_{k=0}^{n-1} Z_k^*$, we get $n^{1/2}(\hat{\Lambda}_n - \Lambda_0) = \int_0^{\cdot}\{S_n^{(0)}\}^{-1}\,d\mu_n$ with the nonpredictable process $S_n^{(0)}$ and μ_n [defined by (3.3)] which is not a martingale.

Assume Conditions A, C and D hold, where $s^{(0)} = E_\nu Z$ and where Conditions C are reduced to:

C1′. $\sup_{[0,1]}|Z|$ is ν-integrable and $\sup_{\mathbb{R}_+}|Z|$ is $\mathbb{P}$-integrable.
C2′. $s^{(0)}$ is bounded away from zero and is continuous on $[0,1]$.
C4′. $\sup_{k\geq 0}(\operatorname{Var} Z_k^*(s))$ and $n^{1/2}(\mathbb{E}S_n^{(0)}(s) - E_\nu S_n^{(0)}(s))$ are finite for each s in a dense subset of $[0,1]$ including 0 and 1.
C6′. The sequence of processes $(n^{1/2}(S_n^{(0)} - s^{(0)}))_n$ is tight for the uniform topology on D_-.

Then Theorem 3.3 becomes: The process $n^{1/2}(\hat{\Lambda}_n - \Lambda_0)$ converges weakly to a Gaussian process with zero mean, continuous sample paths and covariance $\int_0^{s\wedge t}\{s^{(0)}(u)\}^{-1}\lambda_0(u)\,du$ at points s and t.

This result is the same as Aalen's classical one (1978) for the case of martingale-based estimation of the underlying intensity though his method is not available here: $\lambda_0 S_n^{(0)}$ is not the intensity of $(1/n)\sum_{k=0}^{n-1} N_k^*$.

Acknowledgment. The authors are grateful for the referees' comments which led to an improved presentation of the paper.

REFERENCES

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

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

BAKER, R. J. and NELDER, J. A. (1978). The GLIM system. Release No. 3, Numerical Algorithms Group, Oxford.

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

BRETAGNOLLE, J. and HUBER, C. (1985). Effet de l'omission de covariables dans le modèle de Cox. In *Séminaire Huber, Lellouch, Prum* 1–20. Université Paris V, Paris.

BURKHOLDER, D., DAVIS, B. and GUNDY, R. (1972). Integral inequalities for convex functions of operators on martingales. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 223–240. Univ. California Press.

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

DACUNHA-CASTELLE, D. and DUFLO, M. (1986). *Probability and Statistics* **2**. Springer, Berlin.

DOUKHAN, P. and PORTAL, F. (1983). Principes d'invariance faibles, avec vitesse, dans un cadre mélangeant. Report No. 83T28, Université Paris-Sud, Orsay.

GILL, R. D. (1986). Non- and semi-parametric maximum likelihood estimators and the von Mises method. Part 1. CWI Report MS-R 8604, Mathematical Center, Amsterdam.

JACOD, J. (1975). Multivariate point processes: Predictable projections, Radon–Nikokym derivatives, representation of martingales. *Z. Wahrsch. verw. Gebiete* **31** 235–253.

JOHANSEN, S. (1983). An extension of Cox's regression model. *Internat. Statist. Rev.* **51** 165–174.

JOLIVET, E., REYNE, Y. and TEYSSIER, J. (1983). Approche méthodologique de la répartition nycthémérale des prises d'aliment chez le lapin domestique en croissance. *Reproduction Nutrition Développement* **23** 13–24.

LENGLART, E. (1977). Relation de domination entre deux processus. *Ann. Inst. H. Poincaré Probab. Statist.* **13** 171–179.

LIPTSER, R. S. and SHIRYAYEV, A. N. (1978). *Statistics of Random Processes* **2**. Springer, Berlin.

PONS, O. and DE TURCKHEIM, E. (1986). Cox's periodic regression model: The Markovian case. Rapport Technique de Biométrie 86-10, Institut National de la Recherche Agronomique, Paris.

PONS, O. and DE TURCKHEIM, E. (1988). Modèle de régression de Cox périodique et étude d'un comportement alimentaire. *Cahiers de Biométrie* **1**. Institut National de la Recherche Agronomique, Paris.

PRENTICE, R. L. and SELF, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Statist.* **11** 804–813.

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

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

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

SLUD, E. (1984). Multivariate dependent renewal processes. *Adv. in Appl. Probab.* **16** 347–362.

SNYDER, D. L. (1975). *Random Point Processes*. Wiley, New York.

WHITEHEAD, J. (1980). Fitting Cox's regression model to survival data using GLIM. *Appl. Statist.* **29** 268–275.

INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE
DÉPARTEMENT DE BIOMÉTRIE
DOMAINE DE VILVERT
78350 JOUY-EN-JOSAS
FRANCE