Ann. Inst. Statist. Math.
Vol. 41, No. 4, 781–807 (1989)

A PENALTY METHOD FOR NONPARAMETRIC ESTIMATION OF THE INTENSITY FUNCTION OF A COUNTING PROCESS

ANESTIS ANTONIADIS*

Department of Mathematics, U.E.R. des Sciences, University of Saint-Etienne, 23, rue du Dr Paul Michelon, 42100 Saint-Etienne, France

(Received January 12, 1987; revised June 2, 1988)

Abstract. Nonparametric estimators are proposed for the logarithm of the intensity function of some univariate counting processes. An Aalen multiplicative intensity model is specified for our counting process and the estimators are derived by a penalized maximum likelihood method similar to the method introduced by Silverman for probability density estimation. Asymptotic properties of the estimators, such as uniform consistency and normality, are investigated and some illustrative examples from survival theory are analyzed.

Key words and phrases: Penalized likelihood, counting processes, multiplicative intensity, censored data, martingales, Sobolev spaces, kernel estimation.

1. Introduction

This paper deals with the problem of estimating nonparametrically the intensity function of multiplicative intensity counting processes, which constitute an important class of point processes. The importance of such counting processes is that they provide an alternative to the proportional hazard regression model of Cox (1972).

Estimation procedures for the intensity function of point processes have been analyzed in a number of papers, for instance, those written by Ogata (1978), Kutoyants (1979), Lin'kov (1981), Sagalovsky (1983) and Konecny (1984) in the parametric case, Aalen (1975, 1978), Bartoszynski *et al.* (1981), Leadbetter and Wold (1983), Ramlau-Hansen (1983) and Karr (1987) in the nonparametric case, to cite only few. For the multiplicative intensity model, Aalen (1978) provided an estimator for the cumulative hazard function

*This work was conducted while the author was visiting the Department of Mathematics, University of California at Irvine.

rather than the intensity function itself. Under appropriate assumptions, Ramlau-Hansen (1983) used kernel function methods from density estimation to smooth the Aalen estimator and to obtain a kernel estimator of the intensity itself in the multiplicative intensity model. Karr (1987) used Grenander's method of sieves and obtained estimators for the intensity function that are strongly consistent in the L^1 norm. McKeague (1986) used the sieve method to obtain consistent estimators for the parameters of a general semimartingale regression model which contains the point process multiplicative model as a special case. A common feature of these papers is the use of techniques developed for probability density estimation. In spite of this common feature, the means for obtaining detailed results—consistency, asymptotic distribution and the like—depend very much on the specific situation involved.

Our interest in the multiplicative intensity counting process model and the questions addressed here was motivated by the work of Bartoszynski *et al.* (1981). The basic problem they treat is estimation (from i.i.d. realizations) of the intensity function of a nonhomogeneous Poisson process on $\mathbb{R}_+$. The context is metastasis in the growth of malignant tumors. Their techniques are based on penalized maximum likelihood estimation, but questions of consistency or asymptotic normality for their estimators are not considered. Here, we obtain such results for the intensity estimators using methods along the lines of Silverman (1982) for probability density estimation, though some care is needed since the jump times are not i.i.d. random variables. In order to guarantee that the estimators of the intensity are nonnegative, we parametrize the logarithm of the intensity rather than the intensity itself. A penalized maximum likelihood scheme is used to derive the estimators, with its reasonableness justified by virtue of uniform consistency in probability and asymptotic normality. Penalized maximum likelihood methods have also been applied by O'Sullivan (1983) and by Zucker (1986) for nonparametrically estimating a function representing time-dependent covariate effects in a model extending Cox's original model.

The statistical problem we are dealing with is formulated in Section 2, where we also outline some basic results from the theory of counting processes. A survey of this theory intended for applications similar to ours can be found in Gill (1980) or Jacobsen (1982). In Section 3 we present the maximum penalized likelihood procedure for estimating the log intensity and give some conditions for its existence. The remaining sections are devoted to the asymptotic properties of the estimators and to the description of some practical examples. Finally, we would like to point out that our work has been, to a great extent, influenced by Silverman's work on probability density estimation.

2. Formulation of the Aalen model—notation and conventions

Since we are interested in asymptotic properties, we shall consider a sequence of models indexed by $n = 1, 2, \ldots$. As mentioned in the introduction, for each n we will observe an n-component multivariate process

$$N^{(n)} = (N_1^{(n)}, N_2^{(n)}, \ldots, N_n^{(n)}), \tag{2.1}$$

where each component of $(N^{(n)}(t), t \in [0, T])$ is a counting process over the time interval $[0, T]$. The sample paths of $N^{(n)}$ are step functions, zero at time zero, with jumps of size $+1$ only, no two components jumping simultaneously. Thus, multiple events cannot occur. In this model, all properties of stochastic processes are relative to a complete right-continuous filtration $(\mathscr{F}_t^{(n)}: t \in [0, T])$ of sub-σ-algebras on the n-th sample space $(\Omega^{(n)}, \mathscr{F}^{(n)}, P^{(n)})$; $\mathscr{F}_t^{(n)}$ represents all the information available up to time t in the n-th model. We will also assume, as in Aalen (1978), that $E_{P^{(n)}}[N_i^{(n)}(T)] < \infty$ for each n and every i $(i = 1, \ldots, n)$, and that $N^{(n)}$ has a random intensity function $\Lambda^{(n)} = (\lambda_1^{(n)}, \lambda_2^{(n)}, \ldots, \lambda_n^{(n)})$ such that

$$\lambda_i^{(n)}(t) = Y_i^{(n)}(t)\alpha_0(t), \tag{2.2}$$

where α_0 is an unknown nonnegative deterministic function, called the intensity function, while $Y_i^{(n)}$ are nonnegative observable processes on $[0, T]$, adapted to the filtration $\mathscr{F}_t^{(n)}$ and predictable. It will be sufficient for our purpose to note that the adapted processes $Y_i^{(n)}$ are predictable if their sample paths are left-continuous with right-hand limits.

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

$$M_i^{(n)}(t) = N_i^{(n)}(t) - \int_0^t \lambda_i^{(n)}(u)\,du \tag{2.3}$$

are square integrable orthogonal martingales, with dual predictable projections

$$\langle M_i^{(n)}, M_i^{(n)}\rangle(t) = \int_0^t \lambda_i^{(n)}(u)\,du. \tag{2.4}$$

We shall require that α_0 is a continuous smooth function and that $P^{(n)}$-almost all sample paths of $Y_i^{(n)}$ are left-continuous with right-hand limits. Several examples of the multiplicative intensity model described above are given in Aalen (1975, 1978), which indicate its broad scope of applicability.

The basic statistical problem in the multiplicative intensity model (2.2) is to estimate the intensity function α_0 and to discuss asymptotic properties of the estimators as n tends to infinity. Convergence in probability and convergence in distribution will always be relative to the probability measures $P^{(n)}$ parametrized by α_0 on $\Omega^{(n)}$. Our model is nonparametric, meaning that the unknown function α_0 is allowed to range freely in some infinite-dimensional space of functions, which will be specified later. In terms of estimation, we have already mentioned in Section 1 that Ramlau-Hansen has obtained a pointwise consistent kernel estimator for α_0. However, such a procedure may exhibit considerable fluctuations in practice. With milder assumptions on α_0 than ours, McKeague (1986) provided, using the sieve method, an L^2-consistent estimator for the intensity function, but did not discuss the asymptotic distribution of his estimator. Our approach is based on a maximum penalized likelihood method which can be realized as a sort of "dual" of the sieve method (see, e.g., Geman and Hwang (1982)).

3. A penalized maximum likelihood method for estimating the intensity

In this section, we begin with some general consideration of the basic principle we have adopted for estimation and give conditions which guarantee the existence and properties of our estimator. The notation introduced in Section 2 will be employed throughout.

3.1 *The basic principle*

We shall consider in the sequel the stochastic processes on $(\Omega^{(n)}, \mathscr{F}^{(n)}, P^{(n)})$ defined on $[0, T]$ by:

$$\bar{Y}_n(t) = \sum_{i=1}^{n} Y_i^{(n)}(t)$$

and

$$\bar{N}_n(t) = \sum_{i=1}^{n} N_i^{(n)}(t)\,. \tag{3.1}$$

For each n, let $P_0^{(n)}$ be a probability measure on $(\Omega^{(n)}, \mathscr{F}^{(n)})$ that makes the components $N_i^{(n)}$ of $\boldsymbol{N}^{(n)}$ independent homogeneous Poisson processes, each with parameter 1. Then (see, e.g., Aalen (1975), Rebolledo (1978) and Jacobsen (1982)), the distribution of $\boldsymbol{N}^{(n)}$ under $P^{(n)}$ is absolutely continuous with respect to the distribution of $\boldsymbol{N}^{(n)}$ under $P_0^{(n)}$ and, up to a multiplicative random variable constant in α, the likelihood function is given by

$$L_T^{(n)}(\alpha) = \exp\left(-\int_0^T \alpha(s)\bar{Y}_n(s)ds + \int_0^T \ln[\alpha(s)]d\bar{N}_n(s)\right), \tag{3.2}$$

where the stochastic integral appearing in (3.2) is a Lebesgue-Stieltjes integral.

It is well known that the likelihood function $L_T(\alpha)$ is unbounded above, and hence that direct maximum likelihood estimation of the unknown function α_0 is not feasible.

Motivated by the ideas of roughness penalty density estimation (see Tapia and Thompson (1978) for a survey), our purpose in the sequel is to use as an estimate that intensity function which maximizes a penalized modified version of the likelihood function (3.2). To overcome unavoidable nonnegativity constraints on the parameter α, the logarithm of the intensity function will be penalized for roughness rather than the intensity itself. Note that the logarithm of the intensity is a natural quantity to estimate, particularly if the estimates are used for discrimination purposes.

Before stating the estimation algorithm, we formulate a preliminary version of it and provide a heuristic derivation. However, it must be stressed that our estimator is something which is defined and not derived from general principles of statistical inference. That it is a reasonable estimator will appear from its properties established in the next sections.

We will assume that $\theta = \ln(\alpha)$, the unknown parameter, lies in a separable real Hilbert space Θ with norm $\|\cdot\|$. A penalized estimator of θ is obtained by minimization over Θ of a functional of the form

$$A^*_{n,\lambda}(\theta) = l_{n,T}(\theta) + \lambda J(\theta) ,$$

where $l_{n,T}(\theta)$ is the negative log-likelihood function defined by (3.2), and $J(\theta)$ (J: $\Theta \to \mathbb{R}_+$) is a penalty or roughness functional. Smaller values of $l_{n,T}(\theta)$ correspond to "models" θ which are better supported by the data, while smaller values of $J(\theta)$ correspond to more plausible values of θ. The scalar λ ($\lambda > 0$) controls the amount by which the data are smoothed to give the estimate. In our case

$$A^*_{n,\lambda}(\theta) = -\int_0^T \theta(s)\,d\bar{N}_n(s) + \int_0^T e^{\theta(s)}\bar{Y}_n(s)\,ds + \lambda J(\theta) . \tag{3.3}$$

A basic requirement in order to be able to estimate θ on the whole of $[0, T]$ in a meaningful way is that the processes $\bar{Y}_n$ be strictly positive on $[0, T]$. If this is not the case, then given a realization of $\bar{Y}_n$, there will be an infinite number of $\theta(t)$'s that give the same value of $\lambda^{(n)}(t)$. Sometimes, it is clearly possible to make such an assumption, but at other times this may not be possible. Hence, we will introduce a natural modification of the negative likelihood part of (3.3), similar to the one used by Ramlau-Hansen (1983). More precisely, since one may have $\bar{Y}_n(t) = 0$ for some t in $[0, T]$, we will define

$$(3.4)\qquad l^*_{n,\lambda}(\theta) = -\int_0^T \theta(s)\,\frac{\bar{J}_n(s)}{\bar{Y}_n(s)}\,d\bar{N}_n(s) + \int_0^T e^{\theta(s)}\bar{J}_n(s)\,ds\,,$$

where $\bar{J}_n(s) = 1_{[\{\bar{Y}_n(s)>0\}]}$ with $\bar{J}_n(s)[\bar{Y}_n(s)]^{-1} = 0$, whenever $\bar{Y}_n(s) = 0$ (the realizations of the processes $\bar{J}_n(s)$ are left-continuous functions which may jump any time $\bar{Y}_n(s)$ is "updated"). At this point, we should mention that expression (3.2) and the same expression with the log-likelihood function replaced by that in (3.4) do not have the same maximizer with respect to θ. The estimator proposed here deserves consideration for reasons of asymptotic consistency and it is justified by the fact that, asymptotically, it turns out to be a kernel type estimate that adapts locally to the density of the observed jump times. As in Aalen (1975), we will assume in the sequel that the processes $\{\bar{J}_n(s)[\bar{Y}_n(s)]^{-1}, 0 \le s \le T\}$ are $P^{(n)}$-almost surely uniformly bounded, i.e., for each integer n there exists $c_n > 0$ such that $\sup\{\bar{J}_n(s)[\bar{Y}_n(s)]^{-1}, 0 \le s \le T\} < c_n$ with $P^{(n)}$-probability one. This condition ensures that the stochastic integrals above exist. Equation (3.4) provides us information about the random function $\bar{J}_n(s)\theta(s)$ which can only be said to be closely related to $\theta(s)$.

3.2 *Assumptions on the observational model and the penalty functional*

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

ASSUMPTIONS A.

(i) Θ is a real separable Hilbert space of real functions θ: $[0, T] \to \mathbb{R}$ with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$.

(ii) For some real $m > 1$, $\Theta = H_2^m([0, T])$ as sets and they have equivalent norms (H_2^m is the Sobolev space of order m).

(iii) The penalty functional J is defined by $J(\theta) = \|\Pi\theta\|^2$ where Π is a projection operator on Θ with finite-dimensional null space Θ_0 containing the constant functions on $[0, T]$.

(iv) There exist positive constants M_1 and M_2 such that

$$M_1\|\theta\|^2 \le J(\theta) + \|\theta\|^2_{L^2([0,T])} \le M_2\|\theta\|^2\,.$$

(v) The unknown function θ_0 is smooth in the sense of Wahba (1977), that is, θ_0 lies in a Sobolev space on $[0, T]$ of order p, where $p \ge m$.

Recall that, when m is an integer, $H_2^m([0, T])$ is the Hilbert space of functions on $[0, T]$, whose derivatives $f^{(i)}$ up to the order $(m-1)$ exist and are absolutely continuous and such that $f^{(m)}$ is in L^2. The norm of $H_2^m([0, T])$

is $\|f\|_m^2 = \left(\sum_{i=1}^{m} \|f^{(i)}\|^2_{L^2([0,T])} \right)^{1/2}$. When m is a positive number, one can define $H_2^m([0, T])$ using interpolation theory (see, e.g., Adams (1975)). The spaces H_2^p are reproducing kernel Hilbert spaces (R.K.H.S.).

The set of Assumptions A is designed to represent prior notions about the behavior of the unknown function θ_0. A(iii) and A(iv) ensure that the penalty functional J is well behaved, and allow the optimization to be numerically tractable. Very often in practice, the penalty functional $J(\theta)$ is given by $J(\theta) = \int_0^T (D\theta)^2(s)\,ds$ where D: $H_2^m \to L^2([0, T])$ is a differential operator of order m with $\Theta_0 = \ker(D)$ (see, e.g., Cox (1984)). Assumption A(v) formulates the fact that it is probably reasonable to believe that θ_0 is smoother than just being in $H_2^m([0, T])$. In the sequel, following Silverman's terminology, the null family of Π will be called the family of "infinitely smooth" log intensity functions.

The next assumptions pertain to the stochastic part of our model. The conventions of Section 2 are in force. Unless otherwise stated, all limits are taken as n goes to infinity.

ASSUMPTIONS B.

(i) $j_n(s) = E_{P^{(n)}}(\bar{J}_n(s))$ converges uniformly to 1 in $[0, T]$ at a rate $o(n^{-1})$.

(ii) $nE_{P^{(n)}}(\bar{J}_n(s)[\bar{Y}_n(s)]^{-1})$ is uniformly bounded below and above away from 0 and ∞.

(iii) $\{n\bar{J}_n(s)[\bar{Y}_n(s)]^{-1};\ 0 \le s \le T\}$ converges uniformly in probability to a continuous function ζ on $[0, T]$.

The first part of Assumption B(i), i.e., the convergence of $j_n(t)$ to 1 in $[0, T]$ can be regarded as an identifiability criterion. It is easy to construct examples which violate this condition and for which θ_0 is non-identifiable. This condition also ensures asymptotic unbiasedness of our estimators. The rate $o(n^{-1})$ appears rather strong but holds in some important special cases (see examples in Section 6). It is required for the asymptotic normality. Assumptions B(ii) and B(iii) guarantee the stabilization of the variance of the estimators, which is needed to get consistency and asymptotic normality. These conditions are also present, explicitly or implicitly, in the cases studied by Aalen (1975), Jacobsen (1982) and Ramlau-Hansen (1983). In Section 6 we consider the problem of verifying these assumptions in some important practical situations. Under these assumptions, our estimator $\hat{\theta}_n$ of the unknown log intensity function θ_0, if it exists, will be the solution of the following unconstrained optimization problem:

For fixed $\lambda > 0$ and $n \ge 1$, minimize over Θ the functional

$$(3.5) \qquad A_{n,\lambda}(\theta) = -\int_0^T \theta(s)\{\bar{J}_n(s)[\bar{Y}_n(s)]^{-1}\}d\bar{N}_n(s) + \int_0^T e^{\theta(s)}\bar{J}_n(s)ds + \frac{\lambda}{2}J(\theta)\,.$$

Remark. As pointed out by one of the referees, restriction to a fixed (i.e., deterministic) finite interval excludes many practical situations, for which the available data has been randomly censored. The possibility of censoring can be incorporated into our model as follows. Suppose that for each i, the counting process $N_i^{(n)}$ and the baseline stochastic intensity $Y_i^{(n)}$ are observable only up to an $\mathscr{F}_t^{(n)}$ finite stopping time $\tau_i^{(n)}$ and that $(\tau_i^{(n)}, i \geq 1)$ is a stationary sequence of random variables. Define new state and covariate processes (which are observable over the whole interval $[0, T]$) by the stopped processes $\tilde{N}_i^{(n)}(t) = N_i^{(n)}(t \wedge \tau_i^{(n)})$ and $\tilde{Y}_i^{(n)}(t) = Y_i^{(n)}(t \wedge \tau_i^{(n)})$. The censored version of our model is formed by replacing T, $\bar{N}_n$ and $\bar{Y}_n$ by $\tau_n = \mathrm{Max}\ \tau_i^{(n)}$, $\tilde{N}_n$ and $\tilde{Y}_n$ in all previous considerations. In order to use the likelihood approach for estimating θ_0, the restrictions of the distributions of $\tilde{N}_n$ under $P^{(n)}$ and under $P_0^{(n)}$ to the pre-$(T \wedge \tau_n)$-σ-algebra $\mathscr{F}_{\tau_n}$ must be absolutely continuous. Proposition 2.1 of Aven (1986) ensures that the likelihood function exists. The set of Assumptions B should now be checked for the stopped processes $\tilde{N}_n$ and $\tilde{Y}_n$. In some applications it is reasonable to assume that $\tau_i^{(n)}$ is independent of $N_i^{(n)}$ and $Y_i^{(n)}$. In this case, it suffices to check the assumptions for the unstopped processes and have $P^{(n)}(\tau_i^{(n)} > s) > 0$, for all $0 \leq s < T$.

3.3 *Existence of the MPL estimator*

A discussion of the existence of the MPL estimator of the intensity function of a nonhomogeneous Poisson process is given in Section 4 of Bartoszynski *et al.* (1981), drawing on material from Tapia and Thompson (1978). Since, in that work, the penalty functional is a norm (and not a seminorm), the question of existence of our estimator is a little more delicate. The theorem of this section gives a condition for the existence of MPL estimators of the log intensity function and it is similar to Theorem 4.1 of Silverman (1982). However, the context is different and our method of proof cannot be thought of as an extension of the techniques of proof used in the aforementioned paper.

THEOREM 3.1. *Within the notation of this section, the functional $A_{n,\lambda}$, as defined in* (3.5) *above, has a unique minimizer in Θ whenever there exists a minimizer θ_0 of $l_{n,T}^*(\theta)$ in the space of "infinitely smooth" log intensity functions.*

PROOF. Note first that $l_{n,T}^*(\theta)$ is weakly continuous and strictly convex on Θ. Indeed, since, by Assumption A(ii), Θ is a reproducing kernel Hilbert space, we know that pointwise evaluation is a continuous operator. To establish the weak continuity of $l_{n,T}^*(\theta)$ on Θ, we need only to show that for any sequence $\{\theta_k, k \geq 1\}$ converging to θ in Θ,

$$\int_0^T e^{\theta_k(s)} \bar{J}_n(s)\, ds \xrightarrow[k\to\infty]{} \int_0^T e^{\theta(s)} \bar{J}_n(s)\, ds .$$

Since the norm of Θ is equivalent to a Sobolev norm with $m \geq 1$, $\theta_k \to \theta$ uniformly on $[0, T]$. Now, $|\bar{J}_n(s)| \leq 1$ and the exponential function is continuous; thus, the result clearly holds.

A straightforward computation shows that the second Gateaux variation (see Tapia and Thompson (1978)) of $l^*_{n,T}(\theta)$ at θ in the direction η, is given by

$$\ddot{l}^*_{n,T}(\theta)(\eta, \eta) = \int_0^T \eta^2(s) e^{\theta(s)} \bar{J}_n(s)\, ds ,$$

which is $P^{(n)}$-almost surely positive for any η in Θ such that $\|\eta\| > 0$. Therefore, $\ddot{l}^*_{n,T}$ is positive definite relative to Θ and by Proposition 16, Appendix 1 of Tapia and Thompson (1978), $l^*_{n,T}$ is strictly convex in Θ. The last part of the proof is similar to that given by Cox and O'Sullivan (1985) for the maximum penalized likelihood of regression functions. More precisely, let $\lambda > 0$ be given and let θ_1 in Θ. If $B = \{\theta \in \Theta;\, A_{n,\lambda}(\theta) \leq A_{n,\lambda}(\theta_1)\}$ is bounded in Θ, then we are done by the weak lower semicontinuity of $A_{n,\lambda}$ on Θ. Suppose that B is unbounded, that is there exists a sequence $(\theta_k)_{k\geq 1}$ in B such that $\|\theta_k\| \to \infty$ as $k \to \infty$. Since $A_{n,\lambda}(\theta_k)$ is bounded, $\|\Pi(\theta_k)\|$ must be bounded, so, if Id denotes the identity operator on Θ, it must be that $\|(\mathrm{Id} - \Pi)(\theta_k)\| \to \infty$ as $k \to \infty$. Write

$$\theta_k' = \frac{1}{2} (\mathrm{Id} - \Pi)(\theta_k) = \frac{1}{2} \theta_k + \frac{1}{2} (-\Pi)(\theta_k) .$$

By convexity of $l^*_{n,T}$ we have

$$l^*_{n,T}(\theta_k') \leq \frac{1}{2} l^*_{n,T}(\theta_k) + \frac{1}{2} l^*_{n,T}(-\Pi\theta_k) .$$

However, $l^*_{n,T}$ being strictly convex and having a minimizer in $\Theta_0 = \ker(\Pi)$, this minimizer is unique and $l^*_{n,T}$ is coercive on Θ_0, i.e., $l^*_{n,T}(\theta) \to \infty$ as $\|\theta\| \to \infty$. Hence, $l^*_{n,T}(\theta_k') \to \infty$ as $k \to \infty$. But this implies, by the weak continuity of $l^*_{n,T}$ and the boundedness of the sequence $\{\|\Pi(\theta_k)\|\}_{k\geq 1}$ that $l^*_{n,T}(\theta_k) \to \infty$ as k goes to infinity, which contradicts the definition of $(\theta_k)_{k\geq 1}$. □

4. Asymptotic properties

In this section, the consistency and other properties of the MPL estimator introduced above will be studied. Assumptions A and B are in force throughout.

4.1 *Some useful facts*

Note first that, by Assumptions A(ii) and A(v), the intensity function $\alpha_0(s) = \exp(\theta_0(s))$ and its first derivative are functions bounded below and above away from 0 and infinity on $[0, T]$. Following Utreras (1979*a*), we consider the following continuous bilinear forms on $\Theta \times \Theta$:

$$B(u, v) = \langle \Pi u, v \rangle = \langle \Pi u, \Pi v \rangle$$

and

$$A(u, v) = \int_0^T u(t) v(t) \alpha_0(t) dt \,. \tag{4.1}$$

The bilinear form $A(u, v)$ is the scalar product of the weighted space $L^2([0, T], \alpha_0(t)dt)$ and since α_0 is bounded away from zero, $L^2([0, T], dt)$ and $L^2([0, T], \alpha_0(t)dt)$ have equivalent norms. Let $(\mu_i)_{i \geq 0}$ be the eigenvalues of the following variational problem:

$$B(u, v) = \mu A(u, v), \qquad A(u, u) = 1 \,, \tag{4.2}$$

for all v in Θ with corresponding eigenfunctions $\phi_0, \phi_1, \ldots$. Since $B(u, v) = \langle \Pi u, \Pi v \rangle$, $\mu = 0$ is an eigenvalue of (4.2) with corresponding eigenspace the finite-dimensional kernel of the projector Π. Therefore, the eigenvalue 0 has multiplicity $m_0 = \dim [\ker (\Pi)]$ (i.e., $\mu_0 = \mu_1 = \cdots = \mu_{m_0 - 1}$) and a basis $\{\phi_0 = \phi_1 = \cdots = \phi_{m_0 - 1}\}$ of orthonormal eigenfunctions will be chosen in Θ_0 with ϕ_0 constant on $[0, T]$.

On the orthogonal complement of Θ_0 in Θ, B and A are symmetric and positive definite, so the eigenvalues $(\mu_i)_{i \geq m_0}$ are strictly positive and their corresponding eigenfunctions $(\phi_i)_{i \geq \dim(\Theta_0)}$ are orthogonal. Moreover, since α_0 belongs to $H_2^p([0, T])$ and since Assumption A(iv) holds, the eigenvalues associated to (4.2) satisfy the following inequalities:

$$a \cdot i^{2m} \leq \mu_{i + m_0 - 1} \leq b \cdot i^{2m} \qquad i = 1, 2, \ldots \,, \tag{4.3}$$

for some $a, b > 0$ (for a rigorous proof of this, see Utreras (1979*b*)). For each $\lambda > 0$, let B_λ be the symmetric positive definite bilinear form on $\Theta \times \Theta$ defined by

$$B_\lambda(u, v) = \lambda B(u, v) + A(u, v) . \tag{4.4}$$

It is easy to see that B_λ is coercive and defines a norm on Θ which is equivalent to the norm $\|\cdot\|$ by A(iv). We now state and prove, when necessary, some technical results which will be used later.

The notation $f(h) \sim g(h)$ as $h \to 0$ means that, for some constants K and K' and some neighborhood V of zero, $K \leq |f(h)/g(h)| \leq K'$ for all h in V.

LEMMA 4.1. *If $c \geq 0$ and $d \geq 0$ are such that $c < 2 - 2m^{-1}$ and $d \leq 2p$ then, as $\lambda \to 0$,*

$$\sum_{i=0}^{\infty} \mu_i^c (1 + \lambda\mu_i)^{-2} \sim \lambda^{-[(2mc+1)/2m]}$$

and

$$\sum_{i=0}^{\infty} i^d \mu_i^2 (1 + \lambda\mu_i)^{-2} A(\alpha_0, \phi_i)^2 = O(\lambda^{-(d+4m-2p)/2m}) .$$

The proof is elementary (see, e.g., Cox (1984)). One uses the estimates from (4.3) and a standard argument involving the approximation of sums by integrals, for the first assertion of Lemma 4.1. The second part is obtained by applying the dominated convergence theorem.

The next result is the point process analogue of Lemma 5.4 of Silverman (1982). It gives the asymptotic behavior of certain random variables which will occur in the subsequent study of our MPL estimator. Consider the following sequences of random variables, defined on $(\Omega^{(n)}, \mathscr{F}^{(n)}, P^{(n)})$ by

$$\hat{\beta}_k(T) = \int_0^T \phi_k(t) \frac{\bar{J}_n(t)}{\bar{Y}_n(t)} d\bar{N}_n(t)$$

and

$$\beta_k^*(T) = \int_0^T \phi_k(t) \bar{J}_n(t) \alpha_0(t) dt , \tag{4.5}$$

for all integers k and dependence on the sample size n has been dropped for notational convenience.

For two real functions f and g, the notation $f(n) \ll g(n)$ means that there exist a positive constant K and some integer n_0 such that $|f(n)[g(n)]^{-1}| \leq K$ for all $n \geq n_0$. We have then

LEMMA 4.2. *Given n, one has* $E_{P^{(n)}}[\hat{\beta}_k(T)] = E_{P^{(n)}}[\beta_k^*(T)]$ *for all k in* $\mathbb{N}$. *If Assumption* B(i) *is in force with* $\sup_{t \in [0,T]} |j_n(t) - 1| \ll n^{-1}$, *then* $E_{P^{(n)}}[\beta_k^*(T)] \ll n^{-1}$ *and* $E_{P^{(n)}}\left[\beta_0^*(T) - \int_0^T \phi_0(t)\alpha_0(t)dt\right] \ll n^{-1}$. *Moreover, if Assumption* B(ii) *holds, then*

$$E_{P^{(n)}}[\beta_k^*(T)^2] \ll n^{-1} \quad \text{and} \quad E_{P^{(n)}}\left[\left\{\beta_0^*(T) - \int_0^T \phi_0(t)\alpha_0(t)dt\right\}^2\right] \ll n^{-1}.$$

PROOF. Notice first that, by our basic assumptions, each of the processes $\{\bar{J}_n(s)[\bar{Y}_n(s)]^{-1}; 0 \le s \le T\}$ is predictable and uniformly bounded. Thus, the stochastic integrals in (4.5) are well defined for all $n \in \mathbb{N}$. By the theory of stochastic integrals, for each sample size n and each integer k, the processes $\{\hat{\beta}_k(t) - \beta_k^*(t), t \in [0, T]\}$ are mean zero, square integrable martingales with variance process

$$\langle \hat{\beta}_k - \beta_k^* \rangle(t) = \int_0^t \alpha_0(s) \frac{\bar{J}_n(s)}{\bar{Y}_n(s)} \phi_k(s)^2 ds .$$

Therefore, given n and for all k, one has

$$E_{P^{(n)}}[\hat{\beta}_k(T)] = E_{P^{(n)}}[\beta_k^*(T)] = \int_0^T j_n(s)\alpha_0(s)\phi_k(s)ds .$$

By definition of the family $\{\phi_k, k \ge 0\}$ and since ϕ_0 is a constant function on $[0, T]$, for all $k \ge 1$, we have

$$\int_0^T \alpha_0(s)\phi_k(s)ds = 0 .$$

Hence, for all $k \ge 1$,

$$\int_0^T j_n(s)\phi_k(s)\alpha_0(s)ds = \int_0^T (j_n(s) - 1)\phi_k(s)\alpha_0(s)ds$$

and

$$|E(\beta_k^*(T))| \le \int_0^T |\phi_k(s)|\,\alpha_0(s)ds \cdot \sup_{t \in [0,T]} |j_n(t) - 1| \ll n^{-1} ,$$

by our assumption on the asymptotic behavior of the functions j_n. For the particular value $k = 0$,

$$E\left[\beta_0^*(T) - \int_0^T \phi_0(s)\alpha_0(s)ds\right] = \int_0^T (j_n(s) - 1)\phi_0(s)\alpha_0(s)ds ,$$

and the conclusion follows as before.

Following, for example, Jacobsen ((1982), p. 131), one obtains

$$E_{P^{(n)}}\{[\hat\beta_k(T)-\beta_k^*(T)]^2\}=E_{P^{(n)}}\{\langle\hat\beta_k-\beta_k^*\rangle(T)\}$$
$$=\int_0^T\alpha_0(s)E\left(\frac{\bar J_n(s)}{\bar Y_n(s)}\right)\phi_k(s)^2ds\ll n^{-1},$$

by Assumption B(ii). Now, for $k\ge 1$,

$$E_{P^{(n)}}\{[\hat\beta_k(T)]^2\}=\operatorname{Var}[\hat\beta_k(T)]+E^2_{P^{(n)}}\{[\beta_k^*(T)]\}$$

and

$$\operatorname{Var}[\hat\beta_k(T)]=E_{P^{(n)}}\{[\hat\beta_k(T)-E_{P^{(n)}}(\beta_k^*(T))]^2\}$$
$$\le 2E_{P^{(n)}}\{[\hat\beta_k(T)-\beta_k^*(T)]^2\}+2\operatorname{Var}[\beta_k^*(T)].$$

By the Schwarz inequality,

$$\operatorname{Var}[\beta_k^*(T)]=E\left(\left\{\int_0^T\phi_k(s)\alpha_0(s)[\bar J_n(s)-j_n(s)]ds\right\}^2\right)$$
$$\le\int_0^T\alpha_0(s)\phi_k^2(s)ds\cdot\int_0^T\alpha_0(s)E([\bar J_n(s)-j_n(s)]^2)ds,$$

and $E([\bar J_n(s)-j_n(s)]^2)=j_n(s)\cdot(1-j_n(s))$ by definition of $\bar J_n$. Given that $\sup\{j_n(s)\}\le 1$, it follows that $\operatorname{Var}(\beta_k^*(T))\ll n^{-1}$. Regrouping all the previous inequalities, the second assertion of Lemma 4.2 follows. □

4.2 *Approximation of the MPL estimator*

In this subsection, an approximation $\tilde\theta$ to $\hat\theta$ will be defined and studied; the question of the asymptotic closeness of this approximation, in order to obtain consistency results for $\hat\theta$, will be considered later. Our approach here is a point process extension of Section 6 of Silverman (1982).

Given n, define a quadratic form on Θ by

$$Q(\theta)=\frac{1}{2}\,\|\Pi\theta\|^2 \tag{4.6}$$
$$+\int_0^T\left[1+(\theta(s)-\theta_0(s))+\frac{1}{2}(\theta(s)-\theta_0(s))^2\right]\alpha_0(s)ds$$
$$-\int_0^T\theta(s)\frac{\bar J_n(s)}{\bar Y_n(s)}\,d\bar N_n(s).$$

It is straightforward to see that $P^{(n)}$ almost surely, Q is uniformly convex on Θ; thus, Q has a unique minimizer $\tilde{\theta}$ in Θ. Within the notation of Subsection 4.1, up to a constant, $Q(\theta)$ admits the following eigenfunction expansion:

$$\begin{aligned} Q(\theta) &= \frac{\lambda}{2} \sum_{k \geq m_0} \mu_k A(\theta, \phi_k)^2 + A(\theta, \phi_0) \int_0^T \phi_0(s) \alpha_0(s)\, ds + \frac{1}{2} \sum_{k=0}^{\infty} A(\theta, \phi_k)^2 \\ &\quad - \sum_{k=0}^{\infty} A(\theta_0, \phi_k) A(\theta, \phi_k) - \sum_{k=0}^{\infty} A(\theta, \phi_k) \hat{\beta}_k(T) \\ &= \frac{1}{2} \sum_{k \geq m_0} (1 + \lambda \mu_k) A(\theta, \phi_k)^2 + A(\theta, \phi_0) \int_0^T \phi_0(s) \alpha_0(s)\, ds \\ &\quad - \sum_{k=0}^{\infty} [A(\theta_0, \phi_k) + \hat{\beta}_k(T)] A(\theta, \phi_k) . \end{aligned}$$

It follows from this expression that the coefficients $A(\tilde{\theta}, \phi_k)$ satisfy the equations:

$$A(\tilde{\theta}, \phi_0) = \hat{\beta}_0(T) - \int_0^T \phi_0(s) \alpha_0(s)\, ds + A(\theta_0, \phi_0)$$

and

$$A(\tilde{\theta}, \phi_k) = [A(\theta_0, \phi_k) + \hat{\beta}_k(T)](1 + \lambda \mu_k)^{-1}, \qquad k \geq 1 . \tag{4.7}$$

We are now able to state the asymptotic properties of $\tilde{\theta}$ using the results of Subsection 4.2.

THEOREM 4.1. *Defining $\tilde{\theta}$ as above, and using the notation and assumptions of this section, as $\lambda \to 0$ and $n \to \infty$, one has:*

$$E_{P^{(n)}}[A(\tilde{\theta} - \theta_0, \tilde{\theta} - \theta_0)] = O(n^{-1} \lambda^{-1/2m} + \lambda^{p/m})$$

and, given $\delta > 0$,

$$E[\|\tilde{\theta} - \theta_0\|_\infty^2] = o[\lambda^{-\delta}(n^{-1} \lambda^{-1/m} + \lambda^{(2p-1)/2m})] .$$

PROOF. From our notation and equations (4.7), it follows that

$$\begin{aligned} A(\tilde{\theta} - \theta_0, \tilde{\theta} - \theta_0) &= \sum_{k=0}^{\infty} [A(\tilde{\theta}, \phi_k) - A(\theta_0, \phi_k)]^2 \\ &= \left[\hat{\beta}_0(T) - \int_0^T \alpha_0(s) \phi_0(s)\, ds \right]^2 \end{aligned} \tag{4.8}$$

$$
\begin{aligned}
&+\sum_{k\geq 1}\frac{[\hat{\beta}_k(T)-\lambda\mu_k A(\theta_0,\phi_k)]^2}{(1+\lambda\mu_k)^2}\\
&=\left[\hat{\beta}_0(T)-\int_0^T \alpha_0(s)\phi_0(s)\,ds\right]^2\\
&\quad+\sum_{k\geq 1}\frac{\lambda^2\mu_k^2 A(\theta_0,\phi_k)^2}{(1+\lambda\mu_k)^2}\\
&\quad+\sum_{k\geq 1}\frac{\hat{\beta}_k(T)^2}{(1+\lambda\mu_k)^2}-2\lambda\sum_{k\geq 1}\frac{\mu_k A(\theta_0,\phi_k)\hat{\beta}_k(T)}{(1+\lambda\mu_k)^2}\,.
\end{aligned}
$$

Combining the results of Lemmas 4.1 and 4.2, one obtains for the mathematical expectations of the right-hand side of (4.8):

$$
\begin{aligned}
E\{A(\tilde{\theta}-\theta_0,\tilde{\theta}-\theta_0)\}&=O(n^{-1})+\lambda^2 O(\lambda^{(p-2m)/m})+O(n^{-1}\lambda^{-1/2m})\\
&=O(n^{-1}\lambda^{-1/2m}+\lambda^{p/m})\,.
\end{aligned}
$$

Now, for each $\lambda>0$, B_λ defined by (4.4) defines on Θ a norm equivalent to its Sobolev norm. By the Sobolev imbedding theorem (see, e.g., Adams (1975)), for every $\varepsilon>0$, there exists a positive constant C_ε such that

$$
\begin{aligned}
\|\tilde{\theta}-\theta_0\|_\infty^2&\leq C_\varepsilon\sum_{k\geq 1}k^{1+\varepsilon}[A(\tilde{\theta},\phi_k)-A(\theta_0,\phi_k)]^2\\
&=C_\varepsilon\sum_{k\geq 1}k^{1+\varepsilon}\frac{[\hat{\beta}_k(T)-\lambda\mu_k A(\theta_0,\phi_k)]^2}{(1+\lambda\mu_k)^2}\\
&=2C_\varepsilon\left[\sum_{k\geq 1}k^{1+\varepsilon}\frac{\hat{\beta}_k^2(T)}{(1+\lambda\mu_k)^2}+\lambda^2\sum_{k\geq m_0}k^{1+\varepsilon}\frac{\mu_k^2 A(\theta_0,\phi_k)^2}{(1+\lambda\mu_k)^2}\right].
\end{aligned}
$$

Taking $0<\varepsilon<\min(2m\delta, 8m-3, 2p)$ one has

$$
E\left[\sum_{k\geq 1}k^{1+\varepsilon}\frac{\hat{\beta}_k^2(T)}{(1+\lambda\mu_k)^2}\right]\ll n^{-1}\lambda^{-\varepsilon/2m}\lambda^{-1/m}
$$

and

$$
\lambda^2\sum_{k\geq 1}k^{1+\varepsilon}\frac{A(\theta_0,\phi_k)^2\mu_k^2}{(1+\lambda\mu_k)^2}=\lambda^2 o(\lambda^{-[1+\varepsilon+4m-2p]/2m})=o(\lambda^{-[1+\varepsilon-2p]/2m})\,,
$$

by Lemmas 4.1 and 4.2. Thus,

$$
E[\|\tilde{\theta}-\theta_0\|_\infty^2]\ll C_\varepsilon\lambda^{-\varepsilon/2m}(n^{-1}\lambda^{-1/m}+\lambda^{(2p-1)/2m})\,,
$$

and since $\varepsilon < 2m\delta$

$$E[\|\tilde{\theta} - \theta_0\|_\infty^2] = o[\lambda^{-\delta}(n^{-1}\lambda^{-1/m} + \lambda^{(2p-1)/2m})] . \qquad \square$$

Remark. The behavior of the approximant $\tilde{\theta}$ with respect to n is tied directly to the asymptotic behavior of j_n and $(J_n(s)[Y_n(s)]^{-1})$. Any specification as to the consistency and uniform boundedness rates of these quantities will give corresponding rates in Lemma 4.2 and Theorem 4.1.

4.3 *Closeness of the approximation and the main consistency result*

The notation O_P denotes an order of magnitude in probability. The next lemma considers the closeness of $\tilde{\theta}$ to $\hat{\theta}$ and is a consequence of Theorem 4.1.

LEMMA 4.3. *Suppose that the definitions and assumptions of this section are in force and that $\lambda \to 0$ and $n^{m-\delta}\lambda \to \infty$ as $n \to \infty$. Then, for all sufficiently small $\varepsilon > 0$, as $n \to \infty$,*

$$\|\hat{\theta} - \tilde{\theta}\|_\infty = O_P[\lambda^{-\varepsilon/2m}(n^{-1}\lambda^{-1/m} + \lambda^{(2p-1)/2m})] .$$

The result follows by identical calculations as in the proof of Lemma 7.1 of Silverman (1982). The details are straightforward and therefore omitted.

It is now possible to state our main consistency result for the maximum penalized likelihood estimator $\hat{\theta}$ of the true log intensity function θ_0.

THEOREM 4.2. *Suppose that assumptions of Section 3 hold and that the smoothing parameter λ satisfies, for some $\delta > 0$, $\lambda \to 0$ and $n^{m-\delta}\lambda \to \infty$ as $n \to \infty$. Then $\hat{\theta}$ is uniformly consistent as an estimator of θ_0, and in addition, for all $\varepsilon > 0$ sufficiently small, as $n \to \infty$,*

$$\|\hat{\theta} - \theta_0\|_\infty^2 = O_P[\lambda^{-\varepsilon/2m}(n^{-1}\lambda^{-1/m} + \lambda^{(2p-1)/2m})] .$$

PROOF. The proof is obtained by combining the relevant part of Theorem 4.1 with Lemma 4.2. Indeed,

$$\|\hat{\theta} - \theta_0\|_\infty^2 \le 2[\|\tilde{\theta} - \theta_0\|_\infty^2 + \|\hat{\theta} - \tilde{\theta}\|_\infty^2] , \tag{4.9}$$

and by the aforementioned results, it is the $\|\tilde{\theta} - \theta_0\|_\infty^2$ part of the right-hand side of (4.9) which dominates, the other term being negligible. $\square$

We will end this section by describing briefly a possible approach to be employed for the numerical evaluation of the MPLE estimators. It seems reasonable to restrict ourselves to fixed-step size Newton-like schemes, one reason being that these methods can be meshed with existing software.

The maximum penalized likelihood equation for $\hat{\theta}$ is $\Delta A_{n,\lambda}(\hat{\theta}) = 0$. Let $H_{n,\lambda}(\theta)$ be the Hessian of $A_{n,\lambda}$ at θ. Observe that $\Delta A_{n,\lambda}(\theta) \in \Theta^* \approx \Theta$ and that, for every θ in Θ, $H_{n,\lambda}(\theta)$ is a bounded linear operator from Θ into Θ. A Newton-like method for converging to $\hat{\theta}$ is to start with an initial value $\theta^{(0)}$ and iteratively determine $\theta^{(k)}$ by

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - [G_{n,\lambda}(\hat{\theta}^{(k)})]^{-1}\Delta A_{n,\lambda}(\hat{\theta}^{(k)})\,. \tag{4.10}$$

Possible choices for $G_{n,\lambda}(\theta)$ are $G_{n,\lambda}(\theta) = H_{n,\lambda}(\theta)$ (Newton-Raphson method) or $G_{n,\lambda}(\theta) = E[H_{n,\lambda}(\theta)]$ (Fisher's scoring technique). In our context one can easily check that

$$\langle H_{n,\lambda}(\theta)\phi, \psi\rangle = \int_0^T e^{\theta(s)}\phi(s)\psi(s)\bar{J}_n(s)\,ds + \lambda\langle W\phi, \psi\rangle\,.$$

For a numerical implementation it is necessary to evaluate $\Delta A_{n,\lambda}(\theta)$ and $[G_{n,\lambda}(\theta)]^{-1}$. Following the work of Cox and O'Sullivan (1985), it is possible to find for each θ in Θ, eigenvectors spanning Θ such that $\langle G_{n,\lambda}(\theta)\phi_i, \phi_j\rangle = (1+\gamma_i)\delta_{ij}$ where δ_{ij} is Kronecker's delta and $0 \le \gamma_1 \le \gamma_2 \le \cdots$ are the associated eigenvalues of λW. Choosing W to be a differential operator, one then obtains the ϕ_i's and γ_i's as the eigensystem of an elliptic boundary value operator and one can proceed, at least theoretically, to the computation of (4.10). However, a proper numerical analysis of this problem is beyond the scope of this paper and further work needs to be done before we can get a good understanding of the issues and subtleties involved.

5. Asymptotic normality

In this section, the asymptotic normality of our modified penalized maximum likelihood estimator is derived. We shall rely on the theory of weak convergence of stochastic integrals with respect to a martingale and on a particular representation of the approximate estimator $\tilde{\theta}$.

We find it convenient to present first the following result (see, e.g., Rebolledo (1978)) proving that, under certain conditions, martingales converge to normal processes with independent increments. Consider a sequence $\{N_n, n \ge 1\}$ of counting processes on the interval $[0, T]$ with a corresponding sequence of martingales given by

$$M_n(t) = N_n(t) - \int_0^t \Lambda_n(s)\,ds, \qquad t \in [0, T]\,,$$

where $\{\Lambda_n, n \ge 1\}$ is the sequence of intensity processes. Let $\{H_n, n \ge 1\}$ be a sequence of square integrable predictable processes, and define square integrable martingales by

$$\tilde{M}_n(t) = \int_0^t H_n(s)dM_n(s), \qquad t \in [0, T].$$

Since we are only interested in martingales associated with point processes, we do not quote Rebolledo's result (see Rebolledo (1978), Corollary 9) in its full generality.

PROPOSITION 5.1. *Suppose that*

$$\forall \varepsilon > 0, \quad \int_0^t H_n^2(s) I\{|H_n^2(s)| > \varepsilon\} \Lambda_n(s) ds \xrightarrow{P^{(n)}} 0,$$

as $n \to \infty$ (*we then say that* M_n *satisfies the strong asymptotic rarefaction of jumps property of the second kind* ($SARJ2$)) *and that there exists a continuous function* F *on* $[0, T]$, *nondecreasing, with* $F(0) = 0$, *such that*

$$\langle \tilde{M}_n, \tilde{M}_n \rangle(t) = \int_0^t H_n^2(s) \Lambda_n(s) ds \xrightarrow{P^{(n)}} F(t),$$

as $n \to \infty$ *and all* t *in* $[0, T]$. *Then, there exists a Gaussian process* M *with mean* 0 *and covariance function* $F(\min(s, t))$ *on* $[0, T]^2$, *such that* $\tilde{M}_n \xrightarrow{D} M$ *in* $D([0, T])$ *as* $n \to \infty$.

We now reintroduce the sequences $\tilde{\theta}$ and $\hat{\theta}$ of the previous section and we assume that, as $n \to \infty$, $\lambda \to 0$ and obeys the conditions stated in Theorems and Lemmas of Section 4. Under such conditions, $\tilde{\theta} - \hat{\theta}$ converges uniformly to 0 in probability as $n \to \infty$. Therefore, for any s in $[0, T]$, $\tilde{\theta}(s)$ and $\hat{\theta}(s)$ will have the same asymptotic distribution, if such a distribution exists. To derive the result we will use the fact that $\tilde{\theta}$ behaves as a kernel smoothing estimator with a particular kernel.

More precisely, let, for each t in $[0, T]$, Q_t be the functional defined on Θ by

$$Q_t(\theta) = \frac{\lambda}{2} \int_0^T [\theta^{(m)}(s)]^2 ds + \frac{1}{2} \int_0^T \theta^2(s) \alpha_0(s) ds - \theta(t),$$

where from now on we assume that $\|\Pi\theta\|^2 = \int_0^T (\theta^{(m)}(s))^2 ds$. Since α_0 and its first $(m-1)$ derivatives are bounded on $[0, T]$, the functionals Q_t have a unique minimizer in Θ, say $K_\lambda(\cdot, t)$. Moreover, for each $\lambda > 0$, $Q_t(\theta) + \theta(t)$ defines a norm on Θ equivalent to its Sobolev norm. Therefore, as in Cox (1984), $K_\lambda(\cdot, \cdot)$ is the Green's function of a linear elliptic boundary value problem and it is symmetric. In fact, $\{K_\lambda(s,t); (s,t) \in [0,T]^2\}$ is the reproducing kernel of the R.K.H.S. Θ_λ, where Θ_λ denotes the space Θ with the norm $Q_t(\theta) + \theta(t)$. It is straightforward to see that our approximating estimator $\tilde{\theta}$ for the log intensity function θ_0 is implicitly given, for all s in $[0, T]$ by:

$$(5.1) \qquad \tilde{\theta}(s) = \int_0^T K_\lambda(s,t)\theta_0(t)\alpha_0(t)\,dt - \int_0^T K_\lambda(s,t)[\bar{J}_n(t) - 1]\alpha_0(t)\,dt + \int_0^T K_\lambda(s,t)\frac{\bar{J}_n(t)}{\bar{Y}_n(t)}\,dM_n(t)\,.$$

Our consistency result will follow from the particular structure of the kernel K_λ. Let κ^* be the symmetric kernel function on $\mathbb{R}$ given for each $x \geq 0$ by

$$\kappa^*(x) + \kappa(x) = 2^{1/2}\exp(-x/\sqrt{2})\cos(x/\sqrt{2})\,,$$

where κ denotes the solution vanishing at $+\infty$ and $-\infty$ of

$$(5.2) \qquad (-1)^m\kappa^{(2m)} + \kappa = \delta\,,$$

with δ denoting the Dirac delta function. Appealing to Theorem B, Proposition 2 and Remarks in Section 5 of Silverman (1984), there exists a kernel function κ_c on $\mathbb{R}$, depending on both κ^* and κ, such that, for every $t \in [0, T]$ one has

$$(5.3) \qquad \sup_{x \in I_t(\alpha_0)} \left| \lambda^{1/2m}\alpha_0(t)^{-1/2m}K_\lambda(t + \lambda^{1/2m}\alpha_0(t)^{-1/2m}x, t) - \frac{\kappa_c(x)}{\alpha_0(t)} \right| \leq c\{\lambda^{1/2m} + \exp[-\lambda^{-1/2m}\alpha_0(t)^{1/2m}2^{-1/m}\min(t, T-t)]\}\,,$$

the constant c depending only on α_0, where $I_t(\alpha_0)$ is the interval given by

$$I_t(\alpha_0) = [-t\lambda^{-1/2m}\alpha_0(t)^{1/2m}, (T-t)\lambda^{-1/2m}\alpha_0(t)^{1/2m}]\,.$$

We can now state the main result of this section.

THEOREM 5.1. *Let $\tilde{\theta}$, $\hat{\theta}$ be as defined in Section 4 and let Assumptions A and B of Section 3 be in force. Assume also, that, as $n \to \infty$, $\lambda \to 0$ and for some $\delta > 0$, $n^{(2/3)m-\delta}\lambda \to \infty$. Then, for each t in $[0, T]$, $n^{1/2}\lambda^{1/4m}(\hat{\theta}(t) - \theta_0(t))$ converges in distribution to a normal random variable with mean zero and variance*

$$\zeta(t)\alpha_0(t)^{-(2m-1)/2m}\int_{-\infty}^{+\infty}\kappa_c^2(u)\,du\,.$$

PROOF. By Lemma 4.2, for every $\varepsilon > 0$ sufficiently small, we have, as $n \to \infty$,

$$\|\hat{\theta} - \tilde{\theta}\|_\infty = O_P[\lambda^{-\varepsilon/2m}(n^{-1}\lambda^{-1/m} + \lambda^{(2p-1)/2m})].$$

Hence, for all $t \in [0, T]$ and as $n \to \infty$,

$$\sqrt{n}\lambda^{1/4m}(\hat{\theta}(t) - \tilde{\theta}(t)) = O_P[\lambda^{-\varepsilon/2m}(n^{-1/2}\lambda^{-3/4m} + \lambda^{(4p-1)/4m})].$$

Since $n^{(2/3)m-\delta}\lambda \to \infty$ as $n \to \infty$, for ε sufficiently small $n^{1/2}\lambda^{1/4m}(\tilde{\theta}(t) - \hat{\theta}(t))$ converges to 0 in probability as $n \to \infty$. Therefore, $n^{1/2}\lambda^{1/4m}(\tilde{\theta}(t) - \theta_0(t))$ and $n^{1/2}\lambda^{1/4m}(\hat{\theta}(t) - \theta_0(t))$ will have the same asymptotic distribution.

By Lemma 4.2 and the rate at which $\lambda \to 0$, it is clear that $n^{1/2}\lambda^{1/4m}(E[\tilde{\theta}(t)] - \theta_0(t)) \to 0$ as $n \to \infty$. Thus, $n^{1/2}\lambda^{1/4m}(\tilde{\theta}(t) - E[\tilde{\theta}(t)])$ will have the same asymptotic distribution as $n^{1/2}\lambda^{1/4m}(\tilde{\theta}(t) - \theta_0(t))$. Using expression (5.3), the random part of $n^{1/2}\lambda^{1/4m}(\tilde{\theta}(t) - E[\tilde{\theta}(t)])$ is given by

$$\int_0^T \sqrt{n}\lambda^{1/4m} K_\lambda(s, t) \frac{\bar{J}_n(t)}{\bar{Y}_n(t)}\, dM_n(t), \qquad s \in [0, T]. \tag{5.4}$$

Set $H_n(t) = \sqrt{n}\lambda^{1/4m}K_\lambda(s, t)(\bar{J}_n(t)/\bar{Y}_n(t))$ with s fixed in $[0, T]$. Since $K_\lambda(s, t)$ is continuous in t, $\{H_n\}$ defines a sequence of predictable processes. In order to prove our result, it remains to check the conditions of Proposition 5.1 above. We have, for every $\varepsilon > 0$,

$$\begin{aligned}\{|H_n(t)| > \varepsilon\} &= \left\{\left| \sqrt{n}\lambda^{1/4m}K_\lambda(s, t)\frac{\bar{J}_n(t)}{\bar{Y}_n(t)}\right| > \varepsilon\right\}\\ &= \left\{ n\lambda^{1/4m}|K_\lambda(s, t)| \frac{\bar{J}_n(t)}{\bar{Y}_n(t)} > \varepsilon\sqrt{n}\right\}.\end{aligned}$$

Let V_n be defined by

$$V_n(t) = \lambda^{1/4m} \cdot \kappa\left[\frac{(s - t)\alpha_0(t)^{1/2m}}{\lambda^{1/2m}}\right] \cdot \lambda^{-1/2m} \cdot \alpha_0(t)^{1/2m-1} \cdot n\frac{\bar{J}_n(t)}{\bar{Y}_n(t)}.$$

By the uniform approximation (5.3) and its rate with respect to λ, $I\{|H_n(t)| > \varepsilon\}$ and $I\{|V_n(t)| > \varepsilon n^{1/2}\}$ have the same asymptotic behavior. Now, κ_c is uniformly bounded in $\mathbb{R}$ and α_0 is logarithmically uniformly bounded on $[0, T]$. By Assumption B(iii), $n\bar{J}_n\bar{Y}_n^{-1}$ converges uniformly in probability towards a continuous function bounded on $[0, T]$. Hence, $I\{|V_n(t)| > \varepsilon n^{1/2}\} \to 0$ uniformly on $[0, T]$, since $n^{1/2}\lambda^{1/4m} \to \infty$ as $n \to \infty$. By definition of H_n,

$$\int_0^T H_n^2(s) I\{|H_n(s)| > \varepsilon\} \bar{Y}_n(s)\alpha_0(s)\,ds \xrightarrow[P^{(n)}]{} 0\,,$$

as $n \to \infty$ and the SARJ2 is satisfied. Again, by (5.3), one has:

$$\text{(5.5)}\quad \int_0^T H_n^2(t)\Lambda_n(t)\,dt = \int_0^T n\lambda^{1/2m} K_\lambda^2(s,t) \frac{\bar{J}_n(t)}{\bar{Y}_n(t)} \alpha_0(t)\,dt$$

$$= \int_0^T n\lambda^{1/2m} \frac{\kappa^2[(s-t)\lambda^{-1/2m}\alpha_0(t)^{1/2m}]}{\lambda^{1/m}\alpha_0(t)^{-1/m}\alpha_0(t)^2} \alpha_0(t) \frac{\bar{J}_n(t)}{\bar{Y}_n(t)}\,dt$$

$$+ O_P(\lambda^{1/m}) \int_0^T n \frac{\bar{J}_n(t)}{\bar{Y}_n(t)} \alpha_0(t)\,dt\,.$$

The last term on the right-hand side of (5.5) converges to 0 in probability, since $\lambda \to 0$ as $n \to \infty$; therefore expression (5.5) has the same limit in probability as

$$\int_0^T \lambda^{-1/2m} \frac{\kappa_c^2[(s-t)\lambda^{-1/2m}\alpha_0(t)^{1/2m}]}{\alpha_0(t)^{-1/m}\alpha_0(t)^2}\, n \frac{\bar{J}_n(t)}{\bar{Y}_n(t)}\,dt\,.$$

This last expression, by a change of variables, is equal to

$$\text{(5.6)}\qquad \int_{s\lambda^{-1/2m}}^{(s-T)\lambda^{-1/2m}} \frac{n\bar{J}_n(s-u\lambda^{1/2m})}{\bar{Y}_n(s-u\lambda^{1/2m})} \frac{\kappa_c^2[u\alpha_0(s-u\lambda^{1/2m})^{1/2m}]}{\alpha_0(s-u\lambda^{1/2m})^{1-1/m}}\,du\,.$$

Since $n\bar{J}_n\bar{Y}_n^{-1}$ converges uniformly in probability on $[0, T]$ and the limit is a continuous function, and since κ_c^2 and $\alpha_0^{-1+1/m}$ are uniformly continuous and bounded on $[0, T]$, expression (5.6) converges, as $n \to \infty$, to

$$\frac{\zeta(s)}{\alpha_0(s)^{1-1/m}} \int_{-\infty}^{+\infty} \kappa_c^2[u\alpha_0(s)^{1/2m}]\,du = \frac{\zeta(s)}{\alpha_0(s)^{1-1/2m}} \int_{-\infty}^{+\infty} \kappa_c^2(u)\,du\,.$$

Now, Theorem 5.1 follows from Proposition 5.1. □

Remarks.

1. By using a multivariate extension of Rebolledo's theorem, given in Appendix I of the paper by Andersen and Gill (1982), and the Cramer-Wold device, it is not difficult to see that the finite-dimensional distributions of $\{\hat{\theta}(s), s \in [0, T]\}$ are asymptotically multivariate Gaussian.

2. Some rates of uniform consistency have been obtained by Ramlau-Hansen (1983) for the kernel estimator of the unknown intensity; though his results are for different estimators than ours, they appear to be weaker insofar as a comparison is possible (see Ramlau-Hansen (1983), Theorem

4.1.3 and the rates given in the proof of Theorem 5.1 above).

6. Examples

This section describes some typical statistical models for which the penalized likelihood procedure of the previous sections can be applied.

Example 6.1. (Hazard rate for a random censored i.i.d. sample) In statistical analysis of survival data (or failure time data) from a homogeneous population, one is interested in estimating the death intensity (or force of mortality, or hazard function). Nonparametric estimation about hazard rates has been studied by several authors. For recent references, see for example Jacobsen (1982), Yandell (1983), Tanner and Wong (1983) and Ramlau-Hansen (1983), where kernel estimators are defined, extending the results of Rice and Rosenblatt (1976) for kernel estimators in the absence of censoring.

Consider the analysis of a sample of n individuals. The individuals under study may consist of patients at a given hospital suffering from some lethal disease. Let $X_1, X_2, \ldots, X_n$ denote nonnegative independent, identically distributed random lifetimes (times to failure) for the n individuals under study, and independent of the X_i's, let $C_1, C_2, \ldots, C_n$ be the corresponding censoring times. In this kind of data collection, right censoring is inevitable, since in practice one cannot continue the data collection until all individuals are observed to die. Censoring may also occur because some individuals are lost from followup. Suppose that the distribution function F of the X_i's is absolutely continuous with corresponding density f and that H, the continuous distribution function of the C_i's, are such that $F(T) < 1$ and $H(T^-) < 1$. The observed random variables consist of i.i.d. triples (Z_i, δ_i, Y_i) where $Z_i = \min(X_i, C_i)$ is the observable portion of the i-th individual's lifetime, $\delta_i = \boldsymbol{I}_{(X_i \leq C_i)}$ and Y_i is defined by $Y_i(t) = \boldsymbol{I}_{(Z_i \geq t)}$. Now let $N_i(t)$ denote the indicator function of an uncensored failure for individual i prior to time t:

$$N_i(t) = \boldsymbol{I}_{(Z_i \leq t, \delta_i = 1)} ,$$

and suppose that the hazard rate function $\alpha_0(x) = (1 - F(x))^{-1} f(x)$ is a continuous "smooth" function, in the sense that it belongs to a Sobolev space of order m ($m \geq 2$). Hence, no two component processes will jump, with probability one, at the same time. The number of failures at time t or earlier is given by

$$\bar{N}_n(t) = \sum_{i=1}^{n} N_i(t) = \sum_{i=1}^{n} \boldsymbol{I}_{(Z_i \leq t, \delta_i = 1)} ,$$

and has a stochastic intensity process

$$\Lambda_n(t) = \bar{Y}_n(t)\alpha_0(t) ,$$

where

$$\bar{Y}_n(t) = \sum_{i=1}^{n} I_{(X_i \geq t, C_i \geq t)} = [n - \bar{N}_n(t^-)]^+$$

denotes the number of individuals alive just before time t (risk set size at time t^-). Note that $\bar{Y}_n(t)$ is binomially distributed with parameters n and $[(1 - F(t))(1 - H(t^-))]$ (Aalen (1975), Lemma 4.2). Thus, $j_n(t) = E(J_n(t)) = 1 - [(1 - F(t))(1 - H(t^-))]^n$ and j_n converges uniformly to 1 on $[0, T]$ at an exponential rate as $n \to \infty$. Thus, Assumption B(i) of Section 3 holds. Moreover, since

$$0 < \left\{ (1 - H(s^-)) \exp\left[-\int_0^s \alpha_0(u) du \right] \right\} < 1 ,$$

by the Glivenko-Cantelli theorem, we have

$$nJ_n(t)/\bar{Y}_n(t) \to [(1 - H(t^-))(1 - F(t))]^{-1} ,$$

uniformly in probability on $[0, T]$, and Assumption B(iii) is satisfied. It is also straightforward to see that,

$$nE[J_n(t)/\bar{Y}_n(t)] \to [(1 - H(t^-))(1 - F(t))]^{-1} ,$$

as $n \to \infty$; thus, for n sufficiently large, Assumption B(ii) also holds.

The choice of the degree m of smoothness of α_0 is generally dictated by one's prior knowledge of the properties of the hazard function, or by the use one will make of the estimate. With the penalty functional

$$J(\alpha) = \int_0^T [\ddot{\alpha}(u)]^2 du ,$$

all results of the previous sections apply. To apply our maximum penalized estimator in practice, one has to decide only upon a choice for the smoothing parameter λ, while the kernel function smoothing method requires not only a choice for the window size but also the choice of the kernel function.

Example 6.2. (Competing risks model) Apart from the simple survival model, the life history model that has been discussed most frequently in the literature is the competing risks model, where more than one cause of death (failure) is considered. Formally, a competing risks model may be described as a time continuous Markov chain with one transient state labeled 0 and k absorbing states numbered from 1 to k (see, e.g., Aalen (1978)). Let $P_{0i}(s,t)$ be the probability that the process is in state i at time t given that it started at state 0 at time s. We assume that $P_{0i}(s,t)$ is absolutely continuous in both variables and that the intensities, or forces of transition, defined, for $i = 1,\dots,k$, as

$$\alpha_i(s) = \lim_{t \downarrow s} \frac{P_{0i}(s,t)}{t-s},$$

exist and are smooth functions. As mentioned before, if the Markov chain is observed during the time interval $[0, T]$, then transitions from the state 0 to any other state are observed in detail, while $P_{0i}(s,t) = 0$ for every $i \neq 0$. The total force of transition to the set of states $\{1,\dots,k\}$ is denoted by

$$\alpha(s) = \sum_{i=1}^{k} \alpha_i(s).$$

We will assume that, the probability $P_{00}(0,T)$ of not leaving the state $\{0\}$ during the time interval $[0,T]$, which is given by $P_{00}(0,T) = \exp\left(-\int_0^T \alpha(u)\,du\right)$, is strictly positive. We also have, for $i = 1,\dots,k$

$$P_{0i}(0,t) = \int_0^t \alpha_i(s) \exp\left(-\int_0^s \alpha(u)\,du\right) ds.$$

Consider n independent Markov chains of this kind, and assume that each process starts in state 0 at time 0. If we denote the sample paths of the individual processes by $X_j(\cdot)$, it follows that

$$N_n^i(t) = \sum_{j=1}^{n} I\{X_j(t) = i\} \qquad i = 1,\dots,k$$

is the number of processes in state i at time t. Then, each N_n is a counting process with corresponding intensity process

$$\Lambda_n^i(t) = \alpha_i(t) Y_n(t),$$

where

$$Y_n(t) = n - N_n^{\bullet}(t^-)$$

and

$$N_n^{\cdot}(t) = \sum_{i=1}^{k} N_n^i(t) .$$

Fix a value of i. We have $j_n(s) = E(J_n(s)) = P(Y_n(s) > 0) = 1 - [1 - P_{00}(0, T)]^n$. Thus, Assumption B(i) is satisfied. It is also easy to see that, as n goes to infinity and since $P_{00}(0, s) > 0$,

$$n \frac{J_n(s)}{Y_n(s)} = \frac{J_n(s)}{[1 - n^{-1} N_n^{\cdot}(s^-)]} \to \frac{1}{P_{00}(0, s)} .$$

Hence, Assumptions B(ii) and B(iii) hold. If the same smoothness conditions as in the previous example are assumed for each of the α_i's, then each of the intensities is estimable by our MPLE procedure.

7. General remarks

Another approach to the estimation problem addressed in this work would be to extend to the point process setting the results of Klonias (1982, 1984) for density and regression estimators. In these papers, the MPLE problem is studied with a "roughness" penalty imposed on the square root of the density rather than its logarithm, and numerically efficient methods are described.

The computational issues of the penalized likelihood estimator have not been discussed in our work, even though this is an important and difficult numerical analysis problem. When the "sample" size is large, it is quite feasible to use the approximating kernel estimator κ of Section 5 to obtain a good approximation of the MPL estimator. It seems that it may be possible to extend the results of Messer (1986), to provide detailed asymptotic expressions for comparison of the bias and covariance functions of the two estimators. Another important problem is the design of efficient data-based methods for choosing the smoothing parameter, though it should be possible for techniques from other density estimation methods to be adapted for this. It is hoped that these questions will be investigated in a future work.

Acknowledgements

This work was completed while the author was visiting the Department of Mathematics of the University of California at Irvine, whose financial support and hospitality are gratefully acknowledged. The author is very grateful to the referees for their numerous stimulating comments and detailed suggestions which have considerably improved this paper.

REFERENCES

Aalen, O. (1975). Statistical inference for a family of counting processes, Ph. D. Dissertation, University of California, Berkeley.

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

Adams, R. (1975). *Sobolev Spaces*, Academic Press, 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.

Aven, T. (1986). Bayesian inference in a parametric counting process model, *Scand. J. Statist.*, **13**, 87–97.

Bartoszynski, R., Brown, B. W., McBride, C. M. and Thompson, J. R. (1981). Some nonparametric techniques for estimating the intensity function of a cancer related nonstationary Poisson process, *Ann. Statist.*, **9**, 1050–1060.

Cox, D. D. (1984). Multivariate smoothing spline functions, *SIAM J. Numer. Anal.*, **21**, 789–813.

Cox, D. D. and O'Sullivan, F. (1985). Analysis of penalized likelihood-type estimators with applications to generalized smoothing in Sobolev spaces, Tech. Report #51, Department of Statistics, University of California, Berkeley.

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

Geman, S. and Hwang, C. R. (1982). Nonparametric maximum likelihood estimation by the method of sieves, *Ann. Statist.*, **10**, 401–414.

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

Jacobsen, M. (1982). Statistical analysis of counting processes, *Lecture Notes in Statist.*, **12**, Springer, New York.

Karr, A. F. (1987). Maximum likelihood estimation in the multiplicative intensity model, via sieves, *Ann. Statist.*, **15**, 473–490.

Klonias, V. K. (1982). Consistency of two nonparametric penalized likelihood estimators of the probability density function, *Ann. Statist.*, **10**, 811–824.

Klonias, V. K. (1984). On a class of nonparametric density and regression estimators, *Ann. Statist.*, **12**, 1263–1284.

Konecny, F. (1984). Parameter estimation for point processes with partial observations, a filtering approach, *Systems Control Lett.*, **4**, 281–286.

Kutoyants, Y. A. (1979). Local asymptotic normality for processes of Poisson type, *Izv. Akad. Nauk Armyan. SSR, Ser. Mat.*, **14**, 3–20.

Leadbetter, M. R. and Wold, D. (1983). On estimation of point process intensities, *Contribution in Statistics*, (ed. P. K. Sen), North-Holland, New York.

Lin'kov, Y. N. (1981). The asymptotic normality of stochastic integrals and the statistics of discontinuous processes, *Theory Probab. Math. Statist.*, **23**, 97–107.

McKeague, I. W. (1986). Estimation for a semimartingale regression model using the method of sieves, *Ann. Statist.*, **14**, 579–589.

Messer, K. (1986). A comparison of a spline estimate to its "equivalent" kernel estimate, Tech. Report, University of California, Los Angeles.

Ogata, Y. (1978). The asymptotic behavior of maximum likelihood estimators for stationary point processes, *Ann. Inst. Statist. Math.*, **30**, 243–261.

O'Sullivan, F. (1983). The analysis of some penalized likelihood schemes, Tech. Report 726, University of Wisconsin, Madison.

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

Rebolledo, R. (1978). Sur les applications de le theorie des martingales a l'etude statistique d'une famille de processus ponctuels, *Lecture Notes in Math.*, **636**, 27–70, Springer, New York.

Rice, J. and Rosenblatt, M. (1976). Estimation of the log survivor function and hazard function, *Sankhyā Ser. A*, **38**, 60–78.

Sagalovsky, B. (1983). Maximum likelihood and related estimation methods in point processes, Ph. D. Dissertation, University of California, Berkeley.

Silverman, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method, *Ann. Statist.*, **10**, 795–810.

Silverman, B. W. (1984). Spline smoothing: The equivalent variable kernel method, *Ann. Statist.*, **12**, 898–916.

Tanner, M. A. and Wong, W. H. (1983). The estimation of the hazard function from randomly censored data by the kernel method, *Ann. Statist.*, **11**, 989–993.

Tapia, R. A. and Thompson, J. R. (1978). *Nonparametric Probability Density Estimation*, Johns Hopkins University, Baltimore.

Utreras, F. D. (1979*a*). Cross validation techniques for smoothing in one or two dimensions, *Lecture Notes in Math.*, **757**, Springer, New York.

Utreras, F. D. (1979*b*). Utilisation de la methode de validation croisee pour le lissage par fonction spline a une ou deux variables, These d'Etat, University of Grenoble, France.

Wahba, G. (1977). Practical approximate solutions to linear operator equations when the data are noisy, *SIAM J. Numer. Anal.*, **14**, 651–667.

Yandell, B. S. (1983). Nonparametric inference for rates and densities with censored serial data, *Ann. Statist.*, **11**, 1119–1135.

Zucker, D. M. (1986). Survival data regression analysis with time-dependent covariate effects, Ph. D. Dissertation, The Johns Hopkins University, Baltimore.