The Annals of Statistics
1981, Vol. 9, No. 5, 1050–1060

SOME NONPARAMETRIC TECHNIQUES FOR ESTIMATING THE INTENSITY FUNCTION OF A CANCER RELATED NONSTATIONARY POISSON PROCESS[1]

By Robert Bartoszyński, Barry W. Brown, Charles M. McBride and James R. Thompson

Polish Academy of Sciences, M.D. Anderson Hospital and Tumor Institute, and William Marsh Rice University

An attempt is made to model the appearance times of metastases as a nonstationary Poisson process. Three algorithms are developed for this task. The first follows the kernel approach used in probability density estimation by Parzen and Rosenblatt; the second extends the work of Grenander on mortality measurements to a more general censoring scheme appropriate for the present application; the third employs a discrete maximum penalized likelihood approach. We obtain estimates using both stratification and the proportional hazards model. Contrary to customary belief, it seems that the intensity functions associated with the tumor systems under investigation are nonincreasing.

1. Introduction. The kinetic mechanism for the spread of cancer by metastases is not well understood. In general, it is believed that the untreated primary tumor grows roughly exponentially with time, throwing off metastases in random fashion proportional to some monotone increasing function of the primary mass. These metastases grow according to the same kinetic mechanism as that of the primary. As the metastases grow, they themselves are thought to produce metastases, and so on.

For an explication of current considerations in modelling of the metastatic process with twenty-nine references to the clinical and experimental literature see the monograph (Swan, 1977). Liotta et al. (1976) explicitly argue that the number of metastases rapidly increases with time.

If these notions are correct, then one should expect that the appearance times of metastases could be modelled as a nonstationary Poisson process with intensity function $\lambda(t)$, whose functional form can only be conjectured. For example, if the probability of a metastasis being produced is proportional to the total number of tumor cells present in the body, we might expect that $\lambda(t)$ would be roughly exponential. Rather than assuming a functional form for $\lambda(t)$, we have elected to let the data determine its shape.

We shall assume that we have metastasis data on n patients. Each of these patients has his clinical clock set equal to zero at the time of diagnosis and removal of a trunk melanoma. We follow the ith patient, say, for a time T_i. During the time interval $(0, T_i]$ the patient exhibits metastases at times $t_{i1} < t_{i2} < \cdots < t_{im_i}$. If the patient dies at time T_i, then this will be taken as a censoring time for that patient.

Now it is clear that the time of observation is really a concatenation of (at least) two stochastic processes—namely the time at which a new metastasis is thrown off, and the time required for a metastasis to develop to observable size. We shall assume that

Received June, 1979, revised December, 1980.

[1] This work was supported in part by the Polish Academy of Sciences, the United States Public Health Service under CA-11430 and the United States Army Research Office under DAAG29-78-G-0187.

AMS 1970 *subject classifications.* 62G05, 62P10, 65K05.

Key words and phrases. Metastatic process, melanoma, hazard function, maximum penalized likelihood estimation, proportional hazards model.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.* www.jstor.org

observation times of metastases follow a nonstationary Poisson law with intensity function $\lambda(t)$. That is, for any given patient, the probability that no metastasis will be observed from time t_1 to time t_2 will be given by

$$\exp\left[-\int_{t_1}^{t_2}\lambda(\tau)\,d\tau\right].$$

Then, assuming that each patient's intensity of metastatic display is given by the same function, $\lambda(t)$, the joint likelihood given the times of metastasis observations is

$$L(\lambda) = \left[\prod_{i=1}^{n}\prod_{j=1}^{m_i}\lambda(t_{ij})\right]\exp\left[-\sum_{i=1}^{n}\int_0^{T_i}\lambda(\tau)\,d\tau\right].$$

The naive nonparametric maximum likelihood estimate for λ is

$$\lambda_\delta(t) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\delta(t - t_{ij}),$$

where $\delta(\cdot)$ is the Dirac delta function. The estimate $\lambda_\delta(t)$ gives an infinite value to the likelihood, and is generally a poor estimate.

Our task will be to obtain useful nonparametric algorithms for estimating the intensity function λ and then to use them for the analysis of metastasis display intensities for a number of melanoma patients at the M.D. Anderson Hospital and Tumor Institute.

Clevenson and Zidek (1977) develop histogram and kernel estimators assuming $\{\lambda(t)\}$ to be a wide sense stationary stochastic process with specified covariance function and mean trace. Leonard (1978) estimates a Poisson intensity using the relationship (Snyder, 1975) that the conditional distribution of n appearance times given that n observations occur in $(0, T)$ is the same as the distribution of the order statistics of a random sample of size n from a distribution with density

$$f(t) = \lambda(t)\Big/\int_0^T\lambda(\tau)\,d\tau.$$

In obtaining his prior likelihood estimator, Leonard assumes that f'/f possesses the probability structure over differentiable function space of a specified Gaussian process. Rice and Rosenblatt (1976), extending the results of Watson and Leadbetter (1964), use the relationship

$$\lambda(t) = \frac{f(t)}{1 - F(t)},$$

where F is the distribution function of appearance times, to obtain three density estimates for $\lambda(\cdot)$ based on smoothing expressions involving the empirical distribution function F_n.

Because of censoring considerations in the present application and the desirability of being able to impose constraints on λ, we have elected in this paper to use direct estimates. In Section 2 we obtain a kernel estimator for λ. In Section 3 we give an algorithm for maximum likelihood estimation where λ is assumed to be nondecreasing. In Section 4 we discuss maximum penalized likelihood estimation (MPLE) of λ. In Section 5 we discuss a MPLE implementation of Cox's proportional hazards model. Each of our algorithms is designed to require a minimum of input information. All design parameters are data determined. Moreover, no complicated prior information or structure is required for their use. In Section 6 we describe an analysis of melanoma metastasis data, considering only patients who do develop metastases. Contrary to expectations, we find the resulting intensity functions are nonincreasing.

2. A window type estimator for λ. Motivated by Rosenblatt (1956) and Parzen (1962), and following the observation of Silverman (1978) and others that kernel shape is

of marginal importance, we consider the following estimator, using Rosenblatt's step-function kernel,

$$\lambda_w(t) = \frac{\psi(t, h(t))}{h(t)\xi(t)}, \tag{2.1}$$

where

$$\psi(t, h(t)) = \text{number of metastases observed during} \left(t - \frac{h(t)}{2}, t + \frac{h(t)}{2}\right)$$

$$\xi(t) = \text{number of patients still at risk at time } t.$$

In a manner similar to that employed for kernel width selection in probability density estimation in Tapia and Thompson (1978), it can be shown that, if the number of patients censored in $(t - h/2, t + h/2)$ is negligible relative to $\xi(t)$ and if λ is twice differentiable at t, then

$$\text{MSE}[\hat{\lambda}_w(t) \mid \xi(t)] = \frac{\lambda(t)}{h\xi(t)} + \frac{h^2\lambda^4(t)}{4} + O\left(\frac{1}{\xi(t)}\right) + O(h^3). \tag{2.2}$$

Consideration of the two leading terms here suggests the use of

$$h(t) = \left(\frac{2}{\xi(t)}\right)^{1/3} \frac{1}{\lambda(t)}. \tag{2.3}$$

As in the probability density estimation case, $h(t)$ is a function of what is being estimated. Therefore, by analogy with Scott et al. (1977), we use an iterative algorithm to estimate $\lambda(t)$ at each metastasis time t_i. We begin with an initial constant estimate λ_0 equal to the total number of metastases divided by the total duration of the study. Then we iterate as follows: Given a previous estimate of λ, we update h by (2.3) and obtain a new estimate $\hat{\lambda}_w$ from (2.1). The interval $(t - h/2, t + h/2)$ may overlap one or both endpoints of the study in which case the interval is truncated to the endpoint.

A convergence criterion is that the updated $\hat{\lambda}_w$ be close to the previous estimate. However, it is possible for the $\hat{\lambda}_w$'s to cycle over a set of values. In this case, we look for convergence of the Cesaro sum

$$\sum_{j=0}^{N} \hat{\lambda}_{wj}/(N + 1) \tag{2.4}$$

where $\hat{\lambda}_{wj}$ is the estimate of λ in the jth iteration. Our experience has been that terminating after twenty iterations (if the other criteria are not met) does little harm.

In Figure 1, we show the use of this $\hat{\lambda}_w$ algorithm for the case where

$$\begin{aligned} \lambda(t) &= 1 \quad \text{for} \quad t \in (0, 5) \\ &= 0 \quad \text{for} \quad t \notin (0, 5), \end{aligned} \tag{2.5}$$

with 500 simulated patients for whom censoring is applied after the first metastasis or 5, whichever comes first (Curve A).

Because this procedure estimates the hazard function locally whether the Poisson assumption is true or not, it is useful for comparison with global procedures which use this Poisson assumption. Moreover, this algorithm can be performed quite satisfactorily on a hand-held calculator. However, because of the well known tendency of random points to fall in apparent clusters (Feller, 1968, page 161), any procedure for local estimation of λ will exhibit considerable fluctuation (e.g., compare Curve A with Fig. 6 in Rice and Rosenblatt (1976)). The global estimation procedure developed in Section 4 admits of easier generalization and imposition of constraints than does $\hat{\lambda}_w$.

3. A constrained maximum likelihood algorithm. We now consider an estimation procedure similar to that proposed by Grenander (1956).

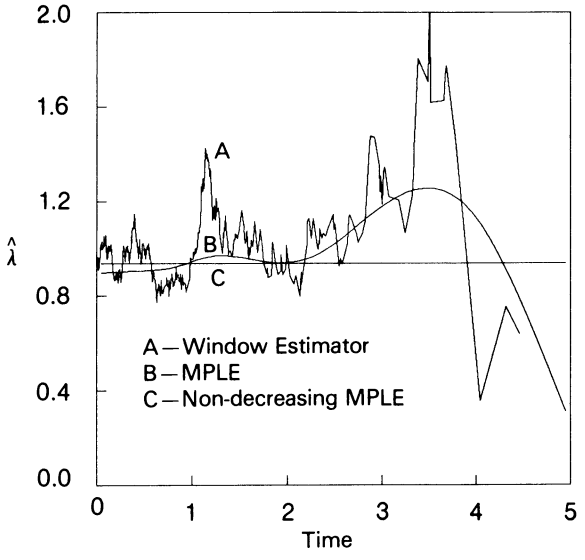

FIG. 1. Estimates of intensity for data generated from Poisson process with $\lambda = 1$.

PROBLEM 1. Let $\lambda(t) \in M(0, T]$, the class of all nonnegative monotone nondecreasing functions on $(0, T]$. We seek to

$$\max_{\lambda \in M(0,T]} \log L(\lambda) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \log \lambda(t_{ij}) - \sum_{i=1}^{n} \int_0^{T_i} \lambda(\tau)\, d\tau.$$

LEMMA 3.1. *A solution to Problem* 1 *must consist of step functions closed on the left with no jumps except at some of the metastasis points.*

PROOF. Increases in $\log L(\lambda)$ can occur only by an increase in $\lambda(\cdot)$ at the metastasis points. Since $\lambda(\cdot)$ is nondecreasing, any increase of $\lambda(\cdot)$ away from these points will make a negative contribution to the log likelihood, via the integral terms, with no attending contribution by the log terms. □

Let us sort the metastasis times into a single list. This gives us the ordering

$$0 < \tau_1 < \tau_2 < \cdots < \tau_s < \tau_{s+1} = T = \max T_i.$$

Let c_j = number of metastases observed at time τ_j,

$$d_j = \sum_{i=1}^{n} \max[0, \min\{(T_i - \tau_j), (\tau_{j+1} - \tau_j)\}] \qquad j = 1, 2, \cdots, s.$$

Clearly for $t < \tau_1$, the optimal value of $\lambda(t)$ is zero. Then as a consequence of Lemma 3.1, we may simplify Problem 1 to the finite dimensional

PROBLEM 1′

$$\max_{\lambda_i \geq 0} \log L(\lambda_1, \lambda_2, \cdots, \lambda_s) = \sum_{i=1}^{s} (c_i \log \lambda_i - d_i \lambda_i). \tag{3.1}$$

Since it leads to a useful algorithm, we take the time to give a constructive proof for the existence and uniqueness of a solution to Problem 1′.

THEOREM 3.1. *The solution to Problem* 1′ *exists and is unique.*

PROOF. Let $z_j = \lambda_j, j = 1, 2, \cdots s$. If $z_1, z_2, \cdots, z_{s-1}$ have already been selected, then z_s is that value of z which is greater than or equal to z_{s-1} and maximizes

$$f_s(z) = c_s \log z - d_s z.$$

The unique local maximum of $f_s(z)$ occurs at $z = c_s/d_s$, so that

$$z_s = \begin{cases} \dfrac{c_s}{d_s} & \text{if } \dfrac{c_s}{d_s} > z_{s-1} \\ z_{s-1} & \text{if } \dfrac{c_s}{d_s} \leq z_{s-1}. \end{cases}$$

Next, let $D_1^*(z_{s-1})$ be the maximum of the last term of the sum in (3.1) given the choice of z_{s-1}, i.e.,

$$D_1^*(z_{s-1}) = \begin{cases} c_s \log z_{s-1} - d_s z_{s-1} & \text{if } z_{s-1} \geq \dfrac{c_s}{d_s} \\ c_s \log \dfrac{c_s}{d_s} - d_s \dfrac{c_s}{d_s} & \text{if } z_{s-1} < \dfrac{c_s}{d_s}. \end{cases}$$

Note that $D_1^*(z_{s-1})$ is at first constant, then decreases monotonically to $-\infty$ as z_{s-1} increases. Suppose we have $z_1, z_2, \cdots, z_{s-2}$. Then z_{s-1} must be that value which is greater than or equal to z_{s-2} and maximizes

$$c_{s-1} \log z_{s-1} - d_{s-1} z_{s-1} + D_1^*(z_{s-1}).$$

Since the derivative is monotone, starting from ∞ at 0 and going to $-\infty$ as $z_{s-1} \to \infty$, the maximum is unique. If $z_{s-1} \leq (c_s)/(d_s)$, the derivative is $(c_{s-1})/(z_{s-1}) - d_{s-1}$, and the maximum occurs at $(c_{s-1})/(d_{s-1})$ if $(c_{s-1})/(d_{s-1}) \leq (c_s)/(d_s)$. If $z_{s-1} > (c_s)/(d_s)$, the derivative equals $(c_{s-1} + c_s)/(z_{s-1}) - (d_{s-1} + d_s)$, and the maximum occurs at $(c_{s-1} + c_s)/(d_{s-1} + d_s)$ if $(c_{s-1})/(d_{s-1}) > (c_s)/(d_s)$.

Let

$$D_2^*(z_{s-2}) = \max_{z_{s-1} \geq z_{s-2}} \{c_{s-1} \log z_{s-1} - d_{s-1} z_{s-1} + D_1^*(z_{s-1})\}.$$

The maximum occurs either at z_{s-2} or at one of the points $(c_{s-1})/(d_{s-1})$, $(c_{s-1} + c_s)/(d_{s-1} + d_s)$ depending on z_{s-2}. Again, $D_2^*(z_{s-2})$ is first constant in z_{s-2} and then decreases to $-\infty$.

Next, assume $D_k^*(z_{s-k})$ is already defined and is first constant in z_{s-k} and then decreases to $-\infty$. We choose z_{s-k} which is greater than or equal to z_{s-k-1} and maximizes

$$c_{s-k} \log z_{s-k} - d_{s-k} z_{s-k} + D_k^*(z_{s-k}).$$

Again, the maximum is unique and

$$D_{k+1}^*(z_{s-k-1}) = \max_{x \geq z_{s-k-1}} \{c_{s-k} \log x - d_{s-k} x + d_k^*(x)\}$$

is of the same form as D_k^*. Proceeding in this way we obtain, finally, the optimal value

$$z_1 = \lambda_1 = \lambda(\tau_1).$$

The problem then returns to the solution in the domain $[\tau_2, T]$ with fixed z_1, etc.

REMARK. We note that Problem 1 without the assumption of monotonicity has no solution. Moreover, the numerical implementation of the maximum penalized likelihood estimation (MPLE) algorithm in the next section includes the present algorithm as a special case. However, relative to the MPLE procedure, the unpenalized algorithm in this section is computationally simple.

4. A maximum penalized likelihood algorithm. To remove nonsmooth estimates

of λ from the feasibility set we shall subtract a penalty term from the log likelihood; see Tapia and Thompson (1978). We shall seek solutions to

PROBLEM 2.

$$\max_{\lambda \in S} J(\lambda, \alpha) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \log \lambda(t_{ij}) - \sum_{i=1}^{n} \int_0^{T_i} \lambda(\tau)\, d\tau - \|\lambda\|^2$$

where S is a closed convex subset of $H^s(0, T) = \{\lambda \mid \lambda^{(j)} \in L^2(0, T); j = 0, 1, \cdots, s \geq 1\}$, $\lambda^{(j)}$ denotes the jth derivative,

$$\|\lambda\|^2 = \sum_{j=0}^{s} \alpha_j \int_0^T (\lambda^{(j)}(\tau))^2\, d\tau \qquad \text{with} \quad \alpha_j > 0 \qquad \text{for} \quad j = 0, 1, \cdots, s$$

and $T = \max\{T_i\}$. It is well known that $H^s(0, T)$ is a Hilbert space. (See Theorem 1.1 of Lions and Magenes (1972)).

DEFINITION. We recall that a Hilbert space H defined on $(0, T)$ is a *reproducing kernel Hilbert space* (RKHS) if pointwise evaluation is a continuous operation, or equivalently, if for every $t \in (0, T)$ there exists M_t such that $|f(t)| \leq M_t \|f\|$ for all $f \in H$.

LEMMA 4.1. *$H^s(0, T)$ is a reproducing kernel Hilbert space.*

PROOF. See Theorem 9.8 of Lions and Magenes (1972).

LEMMA 4.2 *Let S be a closed convex subset of a Hilbert space H. Let $J: H \to R$ be continuous in S, twice Gâteaux differentiable in S with the second Gâteaux variation uniformly negative definite in S. Then J has a unique maximizer in S.*

PROOF. See Theorem 7 of Appendix I of Tapia and Thompson (1978).

THEOREM 4.1 *Problem 2 has a unique solution in $H^s(0, T) \cap \{\lambda \mid \lambda \geq 0\}$.*

PROOF. Clearly $H^s(0, T) \cap \{\lambda \mid \lambda \geq 0\}$ is closed and convex. Since $H^s(0, T)$ is a RKHS, we know that pointwise evaluation is a continuous operation. To establish the continuity of J in $H^s(0, T) \cap \{\lambda \mid \lambda \geq 0\}$, we then need only show that for any sequence such that $\lambda_n \to \lambda_0$ in $H^s(0, T)$ norm, $\int_0^t \lambda_n(\tau)\, d\tau \to \int_0^t \lambda_0(\tau)\, d\tau$. But this is obvious because of the $H^s(0, T)$ norm. Finally, a straightforward computation shows that the second Gâteaux variation of J at λ in the direction η is given by

$$J''(\lambda)(\eta, \eta) = -\sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{\eta^2(t_{ij})}{\lambda^2(t_{ij})} - 2\|\eta\|^2 < -\|\eta\|^2.$$

But then J is uniformly negative definite on $H^s(0, T) \cap \{\lambda \mid \lambda \geq 0\}$. □

To obtain a numerical implementation of the MPLE algorithm, we have used the robust constrained nonlinear optimization routine STEPIT (Chandler, 1975). Because of the rather high storage requirements of STEPIT, we have employed a step function approximation with $(0, T)$ divided into 50 intervals.

COROLLARY 4.1. *If λ is nondecreasing, Problem 2 has a unique solution in this set.*

PROOF. Follows immediately since $H^s(0, T) \cap \{\lambda \mid \lambda > 0, \lambda' \geq 0\}$ is closed and convex.

Because of the naturally occurring term $-\int_0^T \lambda(\tau)\, d\tau$ in $J(\lambda, \alpha)$ we can use $\alpha_0 = 0$. As a practical matter, in our discrete implementation, no instability results when we let $\alpha_1 = 0$ as well. Consequently, we are left with only one design parameter α_2. The method for its selection is to pick a large value of α_2 and to plot the resulting MPLE. Then we successively reduce this α_2 value by negative powers of 10 until high frequency wiggles appear in the

estimate. At this point we return to the previous estimate which did not exhibit the high frequency wiggles. A graphical display of this interactive approach in the probability density estimation case is given in Tapia and Thompson (1978), pages 130–138.

Let us consider again the 500 simulated patients with constant intensity in (2.5). In Figure 1 (Curve B) we examine the MPLE without monotonicity constraints using $\alpha_0 = \alpha_1 = 0$ and $\alpha_2 = 1$. In Figure 1 (Curve C) we show the MPLE using the $\{\alpha_j\}$ above but with the constraint that $\hat{\lambda}$ be nondecreasing on (0, 5).

We observe that the unconstrained MPLE estimates give curves close to smoothed versions of the kernel estimates. In analyzing real data we can use the kernel estimates as rough but robust checks for the MPLEs. We note that the imposition of prior information in the form of constraints on the MPLE solutions is straightforward. Such is not the case with kernel estimators. Moreover, as we shall note in the next section, it is easy to generalize the MPLE to more complicated models.

5. A maximum penalized likelihood solution using the proportional hazards model. In an attempt to estimate an individual intensity function for each patient, we might try to estimate $\lambda_i(t)$ for $i = 1, 2, \cdots, n$ and $t \in (0, T_i)$. We need, in estimating λ_i, to use some mechanism for taking advantage of more than the metastasis record $\{t_{ij}: j = 1, 2, \cdots, m_i\}$. It might be possible to use the entire ensemble $\{t_{ij}: i = 1, 2, \cdots, n; j = 1, \cdots, m_i\}$ if we had some model of similarity of patients. One of the simplest such models is the proportional hazards model of Cox (1972), according to which

$$\lambda_i(t) = \lambda(t) \exp(z_i^T \beta); \qquad i = 1, 2, \cdots, n; \tag{5.1}$$

where $\lambda(t)$ is the same for all patients, z is a k-tuple of risk factors, β is a k-tuple of regression coefficients and

$$z_i^T \beta = \textstyle\sum_{l=1}^{k} \beta_l z_{il}.$$

Accordingly, following our approach in Section 4, we are led to consider

PROBLEM 3. $\max_{(\lambda,\beta)\in S} Q(\lambda, \beta, \alpha, \gamma) = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \log(\lambda(t_{ij})) + \sum_{i=1}^{n} m_i(z_i^T\beta) - \sum_{i=1}^{n} \exp(z_i^T\beta) \int_0^{T_i} \lambda(\tau)\, d\tau - \| (\lambda, \beta) \|^2$, where $z_i \in R_k^+$,

$$\| (\lambda, \beta) \|^2 = \textstyle\sum_{j=0}^{s} \alpha_j \displaystyle\int_0^T \{\lambda^{(j)}(\tau)\}^2\, d\tau + \gamma \textstyle\sum_{i=1}^{k} \beta_i^2 \tag{5.2}$$

with $\alpha_0, \alpha_1, \cdots, \alpha_s, \gamma > 0$, and $S = [H^s(0, T) \cap \{\lambda \mid \lambda \geq 0\}] \times R_k^+$.

REMARK. Note that Problem 3 is *not* separable into two problems, one involving only λ, the other involving only β.

THEOREM 5.1 *Problem 3 has a unique solution in S.*

PROOF. Clearly $H^s(0, T) \times R_k$ is a Hilbert space using the inner product implied by (5.2). Moreover, S is a closed and convex subset of $H^s(0, T) \times R_k$. From Theorem 4.2, we have established the continuity of Q. The second Gâteaux variation at (λ, β) in the direction (η_1, η_2) is given by

$$Q''(\lambda, \beta)((\eta_1, \eta_2), (\eta_1, \eta_2)) = -\sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{\eta_1^2(t_{ij})}{\lambda^2(t_{ij})} - \sum_{i=1}^{n} (z_i^T\eta_2)^2 \exp(z_i^T\beta) \int_0^{T_i} \lambda(\tau)\, d\tau$$

$$-2 \textstyle\sum_{i=1}^{n} (z_i^T\eta_2) \exp(z_i^T\beta) \displaystyle\int_0^{T_i} \eta_1(\tau)\, d\tau - 2\|\eta_1\|^2 - 2\|\eta_2\|^2 < -\|\eta_1, \eta_2\|^2. \tag{5.3}$$

So Q is uniformly negative definite in S. □

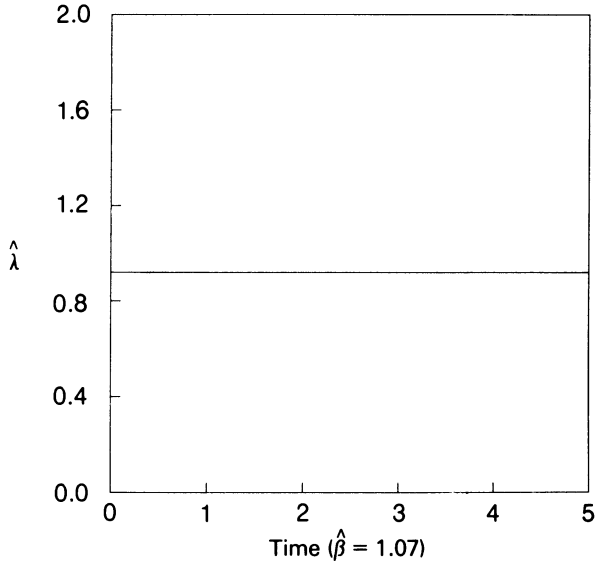

FIG. 2. MPLE of intensity for data generated using proportional hazards model with $\lambda = 1$.

REMARK. We note that it was to establish the uniform negative definiteness of Q that we used the assumption $z_i, \beta \in R_k^+$. In practice, this assumption is probably unnecessarily restrictive, being an artifact of the method of proof. However, because of the form of (5.1), it would not be difficult to arrange for $z_i, \beta \in R_k^+$ in most practical situations. In our numerical work, we have not experienced any instability in our algorithms when we remove the constraint that $z, \beta \in R_k^+$.

COROLLARY 5.1. *If λ is nondecreasing, Problem 3 has a unique solution in this set.*

PROOF. The result follows immediately since $[H^s(0, T) \cap \{\lambda \mid \lambda \geq 0, \lambda' \geq 0\}] \times R_k^+$ is closed and convex.

In Figure 2, we show that MPLE for λ and β for 100 simulated patients where $\lambda(t) = 1$ for $t \in (0, 5)$, $\lambda(t) = 0$ for $t \notin (0, 5)$, $\beta = 1$ and z is $N(1, 1)$. Censoring is after the first metastasis or time $t = 5$, whichever comes first. We have used $\gamma = \alpha_0 = \alpha_1 = 0$ and $\alpha_2 = 1$.

6. Analysis of melanoma metastatic data. We recall that according to the usual notions of the spread of cancer, the metastatic intensity function $\lambda(t)$ should increase with t—perhaps at an exponential rate. We shall examine two sets of data from patients with trunk melanoma treated at the M.D. Anderson Hospital and Tumor Institute (McBride et al., 1976). First we consider 152 patients who were treated for regionally spread melanoma and subsequently developed distant metastases. The time of the first distant metastasis for each was noted. In Figure 3 (Curve A) we note the window estimate (see Section 2) for the intensity function of metastatic display. Although some high frequency wiggles are present, the general pattern of the intensity estimator is clear.

In Figure 3 (Curve C) we show the unpenalized likelihood estimator assuming the intensity is nondecreasing in time (Section 3). The near constancy of the estimated intensity is seen. However, the window estimator in Figure 3 makes us call into question our prior assumption as to the nondecreasing nature of the intensity. The maximum penalized likelihood estimator (Section 4) with $\alpha_0 = \alpha_1 = 0$, $\alpha_2 = 10$ and without any monotonicity assumptions gives us, in Figure 3 (Curve B), effectively a smoothed equivalent of the window estimator.

Next, in Figure 4, we show the MPLE for 192 patients who presented themselves with local melanoma and subsequently displayed distant metastases. There seems to be strong evidence that the intensity is nearly constant, not increasing.

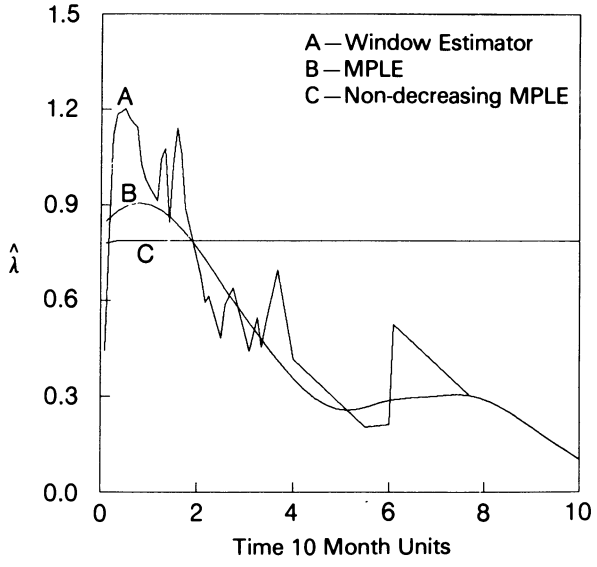

FIG. 3. Estimates of intensity of metastatic display (regional to distant) for 152 patients with melanoma of the trunk.

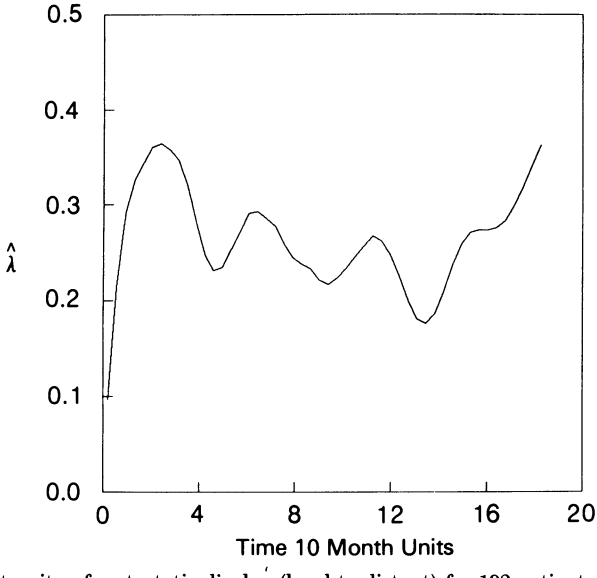

FIG. 4. MPLE of intensity of metastatic display (local to distant) for 192 patients with melanoma of the trunk.

The starting time for events used in Figure 3 is the time of known regional spread of disease. This could be either time of progression for previously treated patients or time of admission for new patients with regionally spread disease. The starting time for events in Figure 4 is time of admission. Both of these times as well as the end event, known distant metastases, are subject to considerable individual variation. However, they represent the best obtainable clinical data and such times are the usual basis for clinical trials of competing treatments.

By definition, the proportional hazards model in (5.1) should be most useful when the ratios of intensities for various strata are in fact constant in time. We have not found natural risk factors in the present investigation which would give us insight beyond those

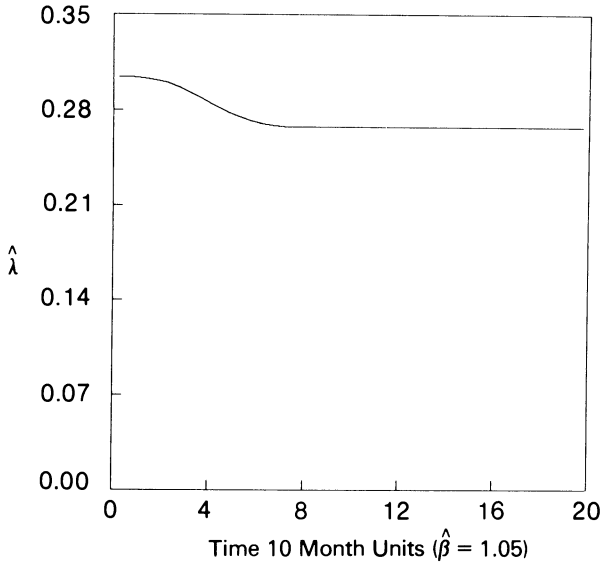

FIG. 5. MPLE of intensity of metastatic display (both local and regional to distant using proportional hazards model) for patients with melanoma of the trunk.

obtained by stratification according to stage of presentation. However, by way of demonstration of the use of the MPLE technique in Problem 3, we consider the estimates obtained using (5.1) in the case where local disease at presentation is coded as $z = 0$ and regional disease is coded as $z = 1$. The estimated (λ, β) is shown in Figure 5.

The nonincreasing nature of our intensity estimates appears to be inconsistent with orthodox notions of metastatic progression. Nor does the nonincreasing intensity of metastatic display appear to be a property only of melanoma. Preliminary results carried out on breast cancer data at the Curie-Sklodowska Institute in Warsaw using our algorithms indicate that, for that neoplastic system as well, the intensity of metastatic display is roughly constant.

Acknowledgement. The authors wish to thank the Editor, the Associate Editor, and the referees for their helpful suggestions.

REFERENCES

CHANDLER, JOHN P. (1975). STEPIT. Distributed by Quantum Chemistry Exchange, Indiana Univ., Bloomington, Indiana.

CLEVENSON, M. LAWRENCE and ZIDEK, JAMES V. (1977). Bayes linear estimators of the intensity function of the nonstationary Poisson process. *J. Amer. Statist. Assoc.* **72** 112–120.

COX, D. R. (1972). Regression models and life tables. *J. Roy. Statist. Soc.* Ser. B **34** 187–200.

FELLER, WILLIAM (1968). *Introduction to Probability Theory and Its Applications.* 3rd Edition. Wiley, New York.

GRENANDER, ULF. (1956). On the theory of mortality measurements, Part II. *Skand. Aktuarietidskr.* **39** 125–153.

LEONARD, TOM (1978). Density estimation, stochastic processes and prior information. *J. Roy. Statist. Soc. B* **40** 113–146.

LIONS, J. L. and MAGENES, E. (1972). *Non-homogeneous Boundary Value Problems and Applications.* Translated from the French by P. Kenneth, Springer, New York.

LIOTTA, LANCE A., SAIDEL, GERALD M. and KLEINERMAN, JEROME (1976). Stochastic model of metastases formation. *Biometrics* **32** 535–550.

MCBRIDE, CHARLES M., SUGARBAKER, EVERETT V., BROWN, BARRY W. (1976). Malignant Melanoma of the Trunk. In *Neoplasms of the Skin and Malignant Melanoma.* Year Book Medical Publishers, Chicago.

PARZEN, EMANUEL (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

RICE, JOHN and ROSENBLATT, MURRAY (1976). Estimation of the log survivor function and hazard function. *Sankhyā A* **38** 60–78.

ROSENBLATT, MURRAY (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27** 832–837.

SCOTT, DAVID W., TAPIA, RICHARD A. and THOMPSON, JAMES R. (1977). Kernel density estimation revisited. *J. Nonlinear Anal.* **1** 339–372.

SILVERMAN, B. W. (1978). Choosing the window width when estimating a density. *Biometrika* **65** 1–11.

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

SWAN, GEORGE W. (1977). *Some Current Mathematical Topics in Cancer Research.* Univ. Microfilm Internat., Ann Arbor.

TAPIA, RICHARD A. and THOMPSON, JAMES R. (1978). *Nonparametric Probability Density Estimation.* Johns Hopkins Univ., Baltimore.

WATSON, G. S. and LEADBETTER, M. R. (1964). Hazard analysis II. *Sankhyā A* **26** 110–116.

ROBERT BARTOSZYŃSKI
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSAW, POLAND

BARRY W. BROWN
DEPARTMENT OF BIOMATHEMATICS
CHARLES M. MCBRIDE
DEPARTMENT OF SURGERY
M.D. ANDERSON HOSPITAL AND TUMOR INSTITUTE
HOUSTON, TEXAS 77030

JAMES R. THOMPSON
DEPARTMENT OF MATHEMATICAL SCIENCES
RICE UNIVERSITY
P.O. BOX 1892
HOUSTON, TEXAS 77001