The Annals of Statistics
1983, Vol. 11, No. 2, 453–466

SMOOTHING COUNTING PROCESS INTENSITIES BY MEANS OF KERNEL FUNCTIONS

BY HENRIK RAMLAU-HANSEN

University of Copenhagen

The kernel function method developed during the last twenty-five years to estimate a probability density function essentially is a way of smoothing the empirical distribution function. This paper shows how one can generalize this method to estimate counting process intensities using kernel functions to smooth the nonparametric Nelson estimator for the cumulative intensity. The properties of the estimator for the intensity itself are investigated, and uniform consistency and asymptotic normality are proved. We also give an illustrative numerical example.

1. Introduction. Let $X_1, \cdots, X_n$ be independent identically distributed random variables with a density f and a distribution function F. The corresponding empirical distribution function is $\hat{F}(t) = (1/n) \sum_{i=1}^{n} I(X_i \le t)$, where $I(A)$ denotes the indicator of the event A. Rosenblatt (1956) suggested that one might estimate the density by

$$\hat{f}(t) = \frac{1}{b} \int_{-\infty}^{\infty} K\left(\frac{t-s}{b}\right) d\hat{F}(s), \tag{1.1}$$

where K is a function with integral 1, called the kernel function, and b is a positive parameter (the window). For a review of existing theory of density estimation by kernel functions, see Bean and Tsokos (1980).

In the present paper we focus on the distributional intensity instead of on the density function. Assume therefore that F is concentrated on $[0, \infty[$, and let the intensity or hazard function be $\alpha(t) = f(t)/\{1 - F(t)\}$ for $t \ge 0$ where $F(t) < 1$. Watson and Leadbetter (1964a, b) studied a kernel estimator for the intensity, given by

$$\hat{\alpha}(t) = (1/b) \sum_{i=1}^{n} K((t - X_{(i)})/b)/(n - i + 1), \tag{1.2}$$

where $X_{(1)}, X_{(2)}, \cdots$ denote the ordered observations. If we introduce $N(t) = n\hat{F}(t)$ and $Y(t) = n - N(t-)$, then (1.2) may be rewritten as

$$\hat{\alpha}(t) = \frac{1}{b} \int_{0}^{\infty} K\left(\frac{t-s}{b}\right) d\hat{\beta}(s), \tag{1.3}$$

where $\hat{\beta}(t) = \int_0^t 1/Y(s)\, dN(s)$ is the nonparametric estimator for the cumulative intensity function $\beta(t) = \int_0^t \alpha(s)\, ds$ introduced by Nelson (1972) and generalized by Aalen (1978), and it is seen that (1.3) essentially is a way of smoothing the increments in $\hat{\beta}(\cdot)$. Recently the estimator (1.3) has also been studied by Rice and Rosenblatt (1976) and Yandell (1981).

In studies of the estimation of densities or intensities by means of kernel functions, most authors have considered the situation of i.i.d. observations. In real life situations in actuarial science, epidemiology, criminology, survival studies, demography, and other fields the observations are often heavily and individually censored, and the theory for the i.i.d. case does not apply. These applications may be further complicated by the existence of several observations for each individual when each life history is modeled as a stochastic process, perhaps as a Markov chain. By applying multiplicative counting processes and

Received December 1981; revised December 1982.
AMS 1980 *subject classifications.* 60G55, 62G05, 62P05.
Key words and phrases. Smoothing, counting processes, kernel functions, intensities.

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

stochastic integrals, Aalen (1978) has recently demonstrated how it is possible to model such situations and develop nonparametric estimators for certain cumulative intensities. The purpose of the present paper is to use kernel functions to smooth the estimator for the cumulative intensity to obtain an estimator of the intensity itself in the multiplicative intensity model. The smoothing method thus developed may then be applied to estimate the hazard rate in survival analysis and to estimate the transition intensities in Markov chain models under very general types of censoring.

In Section 2, we briefly review some general counting process theory as a basis for the multiplicative intensity model, and in Section 3, the generalized kernel estimator for the counting process intensity is defined, and its mean and variance are studied. Uniform consistency and asymptotic normality are proved in Section 4, and Section 5 contains an illustrative numerical example, which shows that the new estimator has considerable practical interest.

In a companion paper we plan to report how a risk function may be used to choose a kernel function and a window. In this paper we assume that the kernel function and the window are given, and that the window tends to 0 when the number of observations tends to infinity.

2. Counting processes. In some fields of science, data frequently consists of counts of the number of transitions between different statuses, such as the number of deaths or failures, the number of disablements or recoveries, or more generally the number of transitions between two states in a Markov chain. The counts may be subject to various kinds of censoring. Even under very general censoring patterns, the number of such transitions observed may be described as a counting process. We will set the scene for our new developments by briefly recounting some elements of the theory of one-dimensional counting processes. When transitions of several types are analysed simultaneously, the theory of multivariate counting processes is the appropriate tool, and the reader is referred to Aalen (1978) and Gill (1980).

Let (Ω, F, P) be a probability space and let $\{F_t, t \in [0, 1]\}$ be an increasing, right-continuous family of sub-sigma algebras of F.

We take F_t to represent the information collected during the period $[0, t]$. A counting process N is a stochastic process on $[0, 1]$, adapted to $\{F_t\}$, where each sample path is a right-continuous step function with $N(0) = 0$ and a finite number of jumps, each of size $+1$. We also assume that $EN(1) < \infty$. Since N is increasing and hence a submartingale it follows from the Doob-Meyer decomposition that $N = A + M$, where A is a predictable increasing process and M is a martingale. We shall assume that there exists a non-negative left-continuous process Λ, adapted to $\{F_t\}$, with right-hand limits such that $A(t) = \int_0^t \Lambda(s)\, ds$. Then, by Aalen (1978),

$$M(t) = N(t) - \int_0^t \Lambda(s)\, ds \tag{2.1}$$

is a square integrable martingale with variance process

$$\langle M \rangle (t) = \int_0^t \Lambda(s)\, ds. \tag{2.2}$$

The process Λ is called the *intensity process* of N and this name is justified by an informal interpretation which takes (2.1) to mean that $E\{dN(t) \mid F_t\} = \Lambda(t+)$ and (2.2) to mean that $\operatorname{Var}\{dN(t) \mid F_t\} = \Lambda(t+)$. This suggests that just after time t, N behaves as a Poisson process with intensity $\Lambda(t+)$.

This paper contributes to the study of the multiplicative intensity model (Aalen, 1978), where it is assumed that Λ can be written in the form

$$\Lambda(t) = \alpha(t) Y(t), \qquad t \in [0, 1], \tag{2.3}$$

where α is an unknown nonstochastic function called the *intensity function*, while Y is an observable stochastic process. The function α and the sample paths of Y are nonnegative and left-continuous with right-hand limits, and Y is assumed to be adapted to $\{F_t\}$. The intensity α is interpreted as the transition intensity on the individual level, and in most applications $Y(t)$ measures the size of the risk population just before time t. Several examples of the multiplicative intensity model may be found in Aalen (1976, 1978) and we give some more below.

In terms of estimation of the intensity function, current nonparametric theory concentrates on the cumulative intensity function $\beta(t) = \int_0^t \alpha(s)\,ds$. Since one can rewrite the combination of (2.1) and (2.3) as $dN(t) = \alpha(t)Y(t)\,dt + dM(t)$, where $dM(t)$ in a sense is noise, a natural estimator of $\beta(t)$ is $\int_0^t dN(s)/Y(s)$. Since one may have $Y(s) = 0$ for some s, it is necessary to modify $\beta(t)$ by defining

$$\beta^*(t) = \int_0^t \alpha(s)J(s)\,ds, \tag{2.4}$$

where $J(s) = I\{Y(s) > 0\}$. When $Y(s) = 0$, we define $J(s)/Y(s) = 0$. We assume that $E\int_0^1 \{J(s)/Y(s)\}\alpha(s)\,ds < \infty$ to ensure that the stochastic integrals below exist. A natural estimator of $\beta^*(t)$ is

$$\hat{\beta}(t) = \int_0^t \left\{\frac{J(s)}{Y(s)}\right\} dN(s), \tag{2.5}$$

and, by the theory of stochastic integrals, it follows that $\hat{\beta} - \beta^*$ is a square integrable martingale with variance process

$$\langle \hat{\beta} - \beta^* \rangle(t) = \int_0^t \left\{\frac{\alpha(s)J(s)}{Y(s)}\right\} ds. \tag{2.6}$$

For a short review of the theory of stochastic integrals, see Aalen (1978, Section 2). By (2.6), the mean square error function becomes

$$\eta(t) = E\{\hat{\beta}(t) - \beta^*(t)\}^2 = E\int_0^t \left\{\frac{\alpha(s)J(s)}{Y(s)}\right\} ds.$$

Let us introduce $\hat{\eta}(t) = \int_0^t \{J(s)/Y^2(s)\}\,dN(s)$. As an illustration of a step which is used throughout the paper we note that by (2.3) and (2.1)

$$\hat{\eta}(t) = \int_0^t \left\{\frac{\alpha(s)J(s)}{Y(s)}\right\} ds + \int_0^t \left\{\frac{J(s)}{Y^2(s)}\right\} dM(s),$$

and the second term on the right side is a zero mean martingale. This implies that $\hat{\eta}(t)$ is an unbiased estimator of $\eta(t)$.

One main advantage of this general formulation of the multiplicative intensity model is that it enables us to take quite general censoring patterns into account. For instance it is possible to model censoring depending on outside random influences in addition to censoring depending on events in the counting process up to the moment in question.

Below we give an example of the so-called random censorship model, where censoring is independent of the counting process.

3. The kernel estimator.

3.1. Definition. Together with the risk population Y, the intensity function α describes how the counting process N develops over time. Generally, behavioural assumptions about real life processes are reflected in formulations about α. For many purposes, therefore, α is the entity of real interest. For technical reasons, however, the now classical theory of the multiplicative counting process model goes no further than the study of $\hat{\beta}$ as an estimator of the cumulative intensity. To come closer to the desired goal, we introduce a kernel estimator for α.

DEFINITION 3.1.1. Let K be a bounded function with integral 1, and let b be a positive parameter. The corresponding kernel estimator for the intensity α is

$$\hat{\alpha}(t) = \frac{1}{b}\int_0^1 K\left(\frac{t-s}{b}\right) d\hat{\beta}(s). \tag{3.1.1}$$

As noted above, this definition is a generalization of the kernel estimator studied by Watson and Leadbetter (1964a, b). If the jump times of N are $T_1, T_2, \cdots$, then

$$\hat{\alpha}(t) = (1/b)\textstyle\sum_{T_i} K((t-T_i)/b)/Y(T_i), \tag{3.1.2}$$

which may be used for computational purposes. By (3.1.2), the sample paths of $\hat{\alpha}$ have properties which are closely related to the kernel function K. If, for example, K is a continuous function, it follows that the sample paths are also continuous. Formula (3.1.2) also shows that the kernel function procedure may be regarded as a kind of continuous moving average with weights integrating to 1, which smooths the occurrence/exposure rates $1/Y(T_i)$ for $i = 1, 2, \cdots$.

3.2. *Mean and variance of the kernel estimator.* To simplify the mathematics, we will assume that the kernel has support within $[-1, 1]$. For practical purposes this is not a real restriction. Let $0 < b < 1/2$ and introduce

$$\alpha^*(t) = \frac{1}{b}\int_0^1 K\left(\frac{t-s}{b}\right) d\beta^*(s) = \int_{-1}^1 K(u)\alpha(t-bu)J(t-bu)\,du, \tag{3.2.1}$$

for $t \in [b, 1-b]$. With a window b it seems natural to try to estimate the intensity only at the points $t \in [b, 1-b]$, since it is only for these values of t that $\hat{\alpha}(t)$ is a real average of the raw estimates given by $\hat{\beta}$. The mean of $\hat{\alpha}$, its variance, and an unbiased estimator of the variance are contained in the following proposition.

PROPOSITION 3.2.1 *Introduce* $K_b(s) = (1/b)K(s/b)$ *and* $j(s) = EJ(s)$. *Then*

$$E\hat{\alpha}(t) = E\alpha^*(t) = K_b * (\alpha j)(t), \tag{3.2.2}$$

where $*$ *denotes ordinary convolution. Furthermore,*

$$\begin{aligned}\sigma^2(t) &= E\{\hat{\alpha}(t) - \alpha^*(t)\}^2 \\ &= \frac{1}{b}\int_{-1}^1 K^2(u)\alpha(t-bu)\, E\left\{\frac{J(t-bu)}{Y(t-bu)}\right\} du\end{aligned} \tag{3.2.3}$$

for $t \in [b, 1-b]$. *An unbiased estimator of* $\sigma^2(t)$ *is*

$$\hat{\sigma}^2(t) = \frac{1}{b^2}\int_0^1 K^2\left(\frac{t-s}{b}\right)\left\{\frac{J(s)}{Y^2(s)}\right\} dN(s) \tag{3.2.4}$$

for $t \in [b, 1-b]$.

PROOF. From the definition of $\hat{\alpha}$ and α^* it follows that

$$\hat{\alpha}(t) - \alpha^*(t) = \frac{1}{b}\int_0^1 K\left(\frac{t-s}{b}\right) d(\hat{\beta} - \beta^*)(s),$$

and (3.2.2) follows since $\hat{\beta} - \beta^*$ is a martingale. By formula (2.6),

$$E\{\hat{\alpha}(t) - \alpha^*(t)\}^2 = \frac{1}{b^2} E \int_0^1 K^2\left(\frac{t-s}{b}\right)\left\{\frac{\alpha(s)J(s)}{Y(s)}\right\} ds$$

$$= \frac{1}{b} \int_{-1}^1 K^2(u)\alpha(t-bu)E\left\{\frac{J(t-bu)}{Y(t-bu)}\right\} du$$

which demonstrates (3.2.3). The final result in the proposition follows from the fact that

$$\hat{\sigma}^2(t) = \frac{1}{b}\int_{-1}^1 K^2(u)\left\{\frac{J(t-bu)}{Y(t-bu)}\right\}\alpha(t-bu)\, du + \frac{1}{b^2}\int_0^1 K^2\left(\frac{t-s}{b}\right)J(s)\, dM(s), \tag{3.2.5}$$

where M is the martingale given by (2.1.). □

If we denote the first term in (3.2.5) by $\sigma^{*2}(t)$ we also see that $E\hat{\sigma}^2(t) = E\sigma^{*2}(t)$, and that

$$E\{\hat{\sigma}^2(t) - \sigma^{*2}(t)\}^2 = \frac{1}{b^3}\int_{-1}^1 K^4(u)\alpha(t-bu)E\left\{\frac{J(t-bu)}{Y(t-bu}\right\} du.$$

In general, (3.2.2) implies that the kernel estimator is not an unbiased estimator of $\alpha(t)$. However, it is possible to show that it is asymptotically unbiased and to derive an asymptotic expression for its variance in the following manner.

If we consider a sequence $\{N_n\}$ of one-dimensional counting processes, each with an intensity process of the form $\Lambda_n(t) = \alpha(t)Y_n(t)$, one may construct a corresponding sequence of kernel estimators, which we denote by $\hat{\alpha}_n(t)$. We use an index "n" to indicate the n-dependence of the counting process. Note that we assume that the kernel is fixed (i.e., independent of n), while the window depends on n, as is usual in the kernel estimation literature. We think of N_n as the relevant counting process when the study population consists of n individuals or items. Then the following proposition holds.

PROPOSITION 3.2.2 *Let $n \to \infty$.* (a) *If the intensity α is continuous at point t and if $j_n = EJ_n \to 1$ uniformly in a neighbourhood of t, then*

$$E\hat{\alpha}_n(t) \to \alpha(t). \tag{3.2.6}$$

(b) *If* (i) *$nE\{J_n(s)/Y_n(s)\} \to 1/\tau(s)$ uniformly in a neighbourhood of t, and* (ii) *α and τ are continuous at the point t, then*

$$\sigma_n^2(t) = E\{\hat{\alpha}_n(t) - \alpha_n^*(t)\}^2 = (nb_n)^{-1}\left\{\frac{\alpha(t)}{\tau(t)}\right\}\int_{-1}^1 K^2(u)\, du + o((nb_n)^{-1}) \tag{3.2.7}$$

as $b_n \to 0$.

PROOF. Relation (3.2.6) follows from (3.2.2) and the fact that the sequence of functions $\{K_{b_n}\}$ is a Dirac sequence when $b_n \to 0$. Relation (3.2.7) follows easily from (3.2.3) and the assumptions made. □

Proposition 3.2.2(a) demonstrates that under the conditions stated, the kernel estimator is asymptotically unbiased. Formula (3.2.7) shows that if $nb_n \to \infty$, then $E\{\hat{\alpha}_n(t) - \alpha_n^*(t)\}^2 \to 0$, and as a consequence $\hat{\alpha}_n(t) - \alpha_n^*(t) \to_P 0$ when $n \to \infty$.

Recently, some authors, Földes et al. (1981) and Yandell (1981), have been interested in estimation by means of kernel functions of the hazard rate for a random censored i.i.d. sample. This situation is covered by our general framework as illustrated in the following example.

EXAMPLE 3.2.3. Consider independent identically distributed death or failure times $X_1, \cdots, X_n$ with values in $[0, \infty[$ and a hazard rate α, where $F(1) = 1 - \exp\{-\int_0^1 \alpha(s)\, ds\} < 1$. Let $T_1, \cdots, T_n$ be corresponding i.i.d. censoring times with distribution H. Assume that the censoring times are independent of the failure times and that $H(1-) < 1$. The number of failures

$$N_n(t) = \sum_{i=1}^n I(X_i \le t, X_i \le T_i)$$

is then a counting process with intensity process $\Lambda_n(t) = \alpha(t) Y_n(t)$, where $Y_n(t) = \sum_{i=1}^n I(X_i \ge t, T_i \ge t)$ denote the number of individuals alive just before time t. The estimator for the cumulative hazard $\beta_n^*(t) = \int_0^t \alpha(s) I\{Y_n(s) > 0\}\, ds$ becomes the ordinary Nelson estimator

$$\hat{\beta}_n(t) = \int_0^t \frac{J_n(s)}{Y_n(s)}\, dN_n(s) = \sum_{X_j \le t} \frac{D_j}{Y_n(X_j)},$$

where D_j is the indicator of death for the jth individual. The corresponding kernel estimator is

$$\hat{\alpha}_n(t) = \frac{1}{b_n} \int_0^1 K\left(\frac{t-s}{b_n}\right) d\hat{\beta}_n(s) = \frac{1}{b_n} \sum_{j=1}^n K\left(\frac{t - X_j}{b_n}\right) \frac{D_j}{Y_n(X_j)}. \tag{3.2.8}$$

Note that $j_n(t) = EJ_n(t) = 1 - [1 - \{1 - F(t)\}\{1 - H(t-)\}]^n \to 1$ uniformly on $[0, 1]$. In connection with an application of Proposition 3.2.2, we see that

$$\begin{aligned} nE\left\{\frac{J_n(s)}{Y_n(s)}\right\} &= E\left\{\frac{n}{Y_n(s)} \,\middle|\, Y_n(s) > 0\right\} j_n(s) \to [\{1 - F(s)\}\{1 - H(s-)\}]^{-1} \\ &= \{1 - H(s-)\}^{-1} \exp\left\{\int_0^s \alpha(u)\, du\right\} \end{aligned} \tag{3.2.9}$$

uniformly on $[0, 1]$, since $Y_n(s)$ is binomially distributed with parameters n and $\{1 - F(s)\}\{1 - H(s-)\}$ (Aalen, 1976, Lemma 4.2). Thus, if we assume that α is continuous on $[0, 1]$, Proposition 3.2.2 applies,

(i) $E\hat{\alpha}_n(t) \to \alpha(t)$ for all $t \in]0, 1[$,

and

(ii) $\sigma_n^2(t) = (nb_n)^{-1} \alpha(t) \{1 - H(t-)\}^{-1} \exp\left\{\int_0^t \alpha(u)\, du\right\} \int_{-1}^1 K^2(u)\, du + o((nb_n)^{-1})$

where $n \to \infty$ and $b_n \to 0$.

3.3 *Estimation of the derivatives of α.* As we have noted, the estimation of an intensity by means of a kernel function and a window is a smoothing method which may be regarded as a continuous moving average. Normally the reason for smoothing an estimator of the intensity is that one believes the true underlying intensity to be a smooth curve. Therefore when one smooths the estimator, it seems natural not only to estimate the intensity itself, but also to try to estimate its slope, its smoothness, etc., as we will now describe. This notion is inspired by the existing statistical theory of moving averages, summarized by Borgan (1979).

The kernel estimator (3.1.1) may formally be rewritten in the form

$$\hat{\alpha} = K_b * \hat{\beta}, \tag{3.3.1}$$

which is a convolution of the function K_b and the measure determined by $\hat{\beta}$. To estimate the νth derivative of the intensity function, it is natural to use the νth derivative of $\hat{\alpha}$. If we assume that the kernel K is absolutely continuous of order ν, then $K, K', \cdots, K^{(\nu-1)}$ exist and are absolutely continuous, and it is possible to differentiate $\hat{\alpha}$ up to order ν. The corresponding estimator then becomes

$$\hat{\alpha}^{(\nu)}(t) = \frac{1}{b^{\nu+1}} \int_0^1 K^{(\nu)}\left(\frac{t-s}{b}\right) d\hat{\beta}(s). \tag{3.3.2}$$

The estimator (3.3.2) has properties similar to those of $\hat{\alpha}$ given in Proposition 3.2.1 and 3.2.2. Some of its small and large sample properties are as follows.

PROPOSITION 3.3.1 (a) *Assume for simplicity that α and $j = EJ$ are members of $C^\nu([0, 1])$, which is the space of ν times differentiable functions with continuous derivatives. Then*

$$E\hat{\alpha}^{(\nu)}(t) = E\alpha^{*(\nu)}(t) = K_b * (\alpha \cdot j)^{(\nu)}(t), \tag{3.3.3}$$

where $\alpha^{*(\nu)}(t) = K_b^{(\nu)} * (\alpha J)(t)$. Furthermore,

$$\sigma_\nu^2(t) = E\{\hat{\alpha}^{(\nu)}(t) - \alpha^{*(\nu)}(t)\}^2 = \frac{1}{b^{2\nu+1}} \int_{-1}^1 K^{(\nu)}(u)^2 E\left\{\frac{J(t-bu)}{Y(t-bu)}\right\}\alpha(t-bu)\, du \tag{3.3.4}$$

for $t \in [b, 1-b]$. This quantity may be estimated by

$$\hat{\sigma}_\nu^2(t) = \frac{1}{b^{2\nu+2}} \int_0^1 K^{(\nu)}\left(\frac{t-s}{b}\right)^2 \left\{\frac{J(s)}{Y^2(s)}\right\} dN(s). \tag{3.3.5}$$

(b) *With the same assumptions as in Proposition 3.2.2, we have*

$$\begin{aligned} \sigma_{n,\nu}^2(t) &= E\{\hat{\alpha}_n^{(\nu)}(t) - \alpha_n^{*(\nu)}(t)\}^2 \\ &= (nb_n^{2\nu+1})^{-1}\left\{\frac{\alpha(t)}{\tau(t)}\right\} \int_{-1}^1 K^{(\nu)}(u)^2\, du + o((nb_n^{2\nu+1})^{-1}) \end{aligned} \tag{3.3.6}$$

when we consider a sequence of counting processes.

The proof is similar to the previous one and is omitted here. One notes that the variance terms $\sigma_{n,\nu}^2$ are asymptotically of completely different sizes when the window tends to zero. This may be interpreted to mean that one ought not to use the same window when estimating the various derivatives.

4. Asymptotic results.

4.1. *Consistency of the kernel estimators.* Under some rather weak conditions, the kernel estimator is a consistent estimator for the underlying intensity. The results given in this section demonstrate both ordinary consistency and mean square uniform consistency.

PROPOSITION 4.1.1. *Consider a sequence $\{N_n\}$ of one-dimensional counting processes exactly as in connection with Proposition 3.2.2. Assume that the conditions in Proposition 3.2.2(b) hold. Then*

$$E\{\hat{\alpha}_n(t) - \alpha(t)\}^2 \to 0$$

when $n \to \infty$, $b_n \to 0$ and $nb_n \to \infty$.

PROOF. By Proposition 3.2.2(b) it follows that $\sigma_n^2(t) \to 0$, so it is sufficient to show that $E\{\alpha_n^*(t) - \alpha(t)\}^2 \to 0$. Since $J_n \to_P 1$ uniformly in a neighbourhood of t, and since

$$\alpha_n^*(t) - \alpha(t) = \int_{-1}^1 K(u)\{\alpha(t - b_n u)J_n(t - b_n u) - \alpha(t)\}\, du, \tag{4.1.1}$$

it follows that $\alpha_n^*(t) - \alpha(t) \to_P 0$, and that there exists a constant $c > 0$, such that $|\alpha_n^*(t) - \alpha(t)| \le c$. Therefore, we conclude that $E\{\alpha_n^*(t) - \alpha(t)\}^2 \to 0$ when $n \to \infty$. □

For a much stronger result about mean square uniform consistency, consider a fixed interval $[z_0, z_1]$ with $0 < z_0 < z_1 < 1$.

THEOREM 4.1.2. *Assume that*
(i) $J_n \to_P 1$ *uniformly on* $[0, 1]$ *when* $n \to \infty$,
(ii) α *is continuous on* $[0, 1]$,
(iii) $n\eta_n(1) = n \int_0^1 E\{J_n(s)/Y_n(s)\}\alpha(s)\,ds$ *is bounded when* $n \to \infty$,
and
(iv) *the kernel K is of bounded variation.*
Then

$$E\{\sup_{t\in[z_0,z_1]}|\hat{\alpha}_n(t) - \alpha(t)|^2\} \to 0 \tag{4.1.2}$$

when $n \to \infty$, $b_n \to 0$ *and* $nb_n^2 \to \infty$.

PROOF. It is sufficient to prove that

$$E\{\sup_t|\hat{\alpha}_n(t) - \alpha_n^*(t)|^2\} \to 0 \tag{4.1.3}$$

and

$$E\{\sup_t|\alpha_n^*(t) - \alpha(t)|^2\} \to 0. \tag{4.1.4}$$

We first prove (4.1.3). From the definitions, it follows that

$$\hat{\alpha}_n(t) - \alpha_n^*(t) = (1/b_n)\int_0^1 K((t-s)/b_n)\,d(\hat{\beta}_n - \beta_n^*)(s).$$

Since K is assumed to be of bounded variation,

$$|\hat{\alpha}_n(t) - \alpha_n^*(t)| \le 2(1/b_n)V(K)\sup_{s\in[0,1]}|\hat{\beta}_n(s) - \beta_n^*(s)|,$$

where $V(K)$ denotes the total variation of K. Therefore, it is sufficient to show that

$$b_n^{-2}E\{\sup_{s\in[0,1]}|\hat{\beta}_n(s) - \beta_n^*(s)|^2\} \to 0.$$

If we apply Doob's inequality to the submartingale $(\hat{\beta}_n - \beta_n^*)^2$, it follows that

$$b_n^{-2}E\{\sup_{s\in[0,1]}|\hat{\beta}_n(s) - \beta_n^*(s)|^2\} \le (nb_n^2)^{-1}4n\eta_n(1).$$

Since $n\eta_n(1)$ is bounded and $nb_n^2 \to \infty$, (4.1.3) is proved.
The proof of (4.1.4) is similar to that of Proposition 4.1.1 and is omitted here. □

Note that in Proposition 4.1.1 we assume that $nb_n \to \infty$, while in Theorem 4.1.2 we assume that $nb_n^2 \to \infty$. Thus, the window must tend towards zero more slowly to obtain uniform consistency than to obtain ordinary consistency.

The assumptions in this theorem are not very restrictive. For example, condition (iii) is often made in the nonparametric estimation theory to ensure that $E\{\sup_{t\in[0,1]}|\hat{\beta}_n(t) - \beta_n^*(t)|^2\} \to 0$ when $n \to \infty$ (Aalen, 1978). As was pointed out by a referee, however, by applying the inequality of Lenglart (1977) instead of Doob's inequality, one may weaken condition (iii) in Theorem 4.1.2 by assuming that $n \int_0^1 \{J_n(s)/Y_n(s)\}\,ds$ is $O(1)$ in probability to obtain uniform consistency instead of mean square uniform consistency.

Similarly, it is possible to show that the kernel estimator of the νth derivative is uniformly consistent. If we consider a sequence of counting processes and the corresponding sequence of estimators $\hat{\alpha}_n^{(\nu)}$ for the νth derivative, the following theorem holds, with a proof similar to that of Theorem 4.1.2.

THEOREM 4.1.3. *Assume that* (i) $J_n \to_P 1$ *uniformly on* $[0, 1]$ *when* $n \to \infty$, (ii) $\alpha \in C^{(\nu)}([0, 1])$, (iii) $n\eta_n(1)$ *is bounded when* $n \to \infty$, *and* (iv) *the* νth *derivative* $K^{(\nu)}$ *is of bounded variation. Then*

$$E\{\sup_{t\in[z_0,z_1]} |\hat{\alpha}_n^{(\nu)}(t) - \alpha^{(\nu)}(t)|^2\} \to 0$$

when $n \to \infty$, $b_n \to 0$ *and* $nb_n^{2\nu+2} \to \infty$.

4.2 *Asymptotic normality.* Parzen (1962) proved that the kernel estimator for a density is asymptotically normal when the number of observations tends to infinity. In the present section, we derive a similar result for the kernel estimators studied in this paper. We cannot use the same technique as Parzen did, since in general our one-dimensional counting process is not based on independent identically distributed observations. Parzen's proof was based on a result about the asymptotic behaviour of triangular arrays of independent observations. We replace that by a result about the asymptotic distribution of a martingale triangular array, proved by Liptser and Shiryayev (1980) and Shiryayev (1981).

Consider a sequence of counting processes (N_n) on $[0, 1]$ with a corresponding sequence of martingales given by

$$M_n(t) = N_n(t) - \int_0^t \Lambda_n(s)\, ds,$$

where $\{\Lambda_n\}$ is the sequence of intensity processes. Let H_n be a sequence of predictable processes where $E \int_0^1 H_n^2(s)\Lambda_n(s)\, ds < \infty$ and introduce $\tilde{M}_n(t) = \int_0^t H_n(s)dM_n(s)$. Then the following proposition is valid.

PROPOSITION 4.2.1. *Suppose that*

(i) $\forall\varepsilon > 0$: $\int_0^1 H_n(s)^2 I(|H_n(s)| > \varepsilon)\Lambda_n(s)\, ds \to_P 0$,

and

(ii) $\int_0^1 H_n^2(s)\Lambda_n(s)\, ds \to_P 1$ *when* $n \to \infty$.

Then $\tilde{M}_n(1) \to_D N(0, 1)$, *where* $N(0, 1)$ *is the standard normal distribution.*

PROOF. Use Liptser and Shiryayev (1980), Corollary 2 and Remark 1. If we apply that result to the sequence $(\tilde{M}_n)$, the proposition follows since the conditions (L_2) and (12) in the paper quoted are equal to our conditions (i) and (ii). □

We now consider a sequence of counting processes where the corresponding kernel estimators may be written in the following form:

$$\hat{\alpha}_n(t) = \alpha_n^*(t) + (1/b_n) \int_0^1 K((t - s)/b_n)J_n(s)/Y_n(s)\, dM_n(s).$$

Fix the value of t. Then

$$(nb_n)^{1/2}\{\hat{\alpha}_n(t) - \alpha_n^*(t)\} = \int_0^1 H_n(s)\, dM_n(s),$$

where

$$H_n(s) = (n/b_n)^{1/2}K((t - s)/b_n)J_n(s)/Y_n(s).$$

We are now able to prove our next result.

THEOREM 4.2.2. *Assume that* (i) $nJ_n/Y_n \to_P 1/\tau$ *uniformly in a neighbourhood of* t

as $n \to \infty$, (ii) *the functions* α *and* τ *are continuous at the point* t. *Then*

$$(nb_n)^{1/2}\{\hat{\alpha}_n(t) - \alpha_n^*(t)\}$$

converges in distribution towards a normal distribution with mean 0 *and variance* $\{\alpha(t)/\tau(t)\} \int_{-1}^{1} K^2(u)\, du$, *when* $n \to \infty$, $b_n \to 0$ *and* $nb_n \to \infty$.

PROOF. We apply Proposition 4.2.1 and verify condition (i) and (ii). First we have $\{|H_n(s)| > \varepsilon\} = \{|K((t-s)/b_n)|\, nJ_n(s)/Y_n(s) > \varepsilon(nb_n)^{1/2}\}$. Since $b_n \to 0$, $nb_n \to \infty$, $nJ_n/Y_n \to_P 1/\tau$ uniformly in a neighbourhood of t, and $1/\tau$ is bounded in this neighbourhood, we see that $I\{|H_n(s)| > \epsilon\} \to_P 0$ uniformly on $[0, 1]$. Therefore, by the definition of $H_n(s)$, $\int_0^1 H_n^2(s) I\{|H_n(s)| > \varepsilon\}\, Y_n(s)\alpha(s)\, ds \to_P 0$ and condition (i) is fulfilled. Since

$$\int_0^1 H_n^2(s)\Lambda_n(s)\, ds = (1/b_n)\int_0^1 K^2((t-s)/b_n)\{\alpha(s) nJ_n(s)/Y_n(s)\}\, ds$$

$$= \int_{-1}^{1} K^2(u)\{nJ_n(t - b_n u)/Y_n(t - b_n u)\}\alpha(t - b_n u)\, du$$

$$\to_P \{\alpha(t)/\tau(t)\}\int_{-1}^{1} K^2(u)\, du,$$

condition (ii) is verified and the theorem is proved. □

In addition to the theorem one may be interested in extra conditions which ensure that $(nb_n)^{1/2}\{\alpha_n^*(t) - \alpha(t)\}$ is asymptotically negligible. Applying the mean value theorem, this occurs if α has a bounded derivative in a neighbourhood of t and $nb_n^3 \to 0$ as $n \to \infty$.

We mention without proof that it also follows from Proposition 4.2.1 and the Cramér-Wold device that $\hat{\alpha}_n(t)$ and $\hat{\alpha}_n(s)$ are asymptotically independent when $s \neq t$, $n \to \infty$, $b_n \to 0$ and $nb_n \to \infty$.

EXAMPLE 4.2.3. At this point, we extend our Example 3.2.3 to a multiple decrement model, which is a time-continuous Markov chain with one transient state labeled 0 and m absorbing states numbered from 1 to m. This model is often used to analyse different causes of decrement in demography and actuarial science. If $m = 2$, it describes the situation in Example 3.2.3 if the number of transitions to state 1 corresponds to failure and the transition to state 2 represents censoring. In other applications, one may be interested in a more general Markov chain where the number of transitions between two different states constitutes a counting process which may be analysed along the ideas below.

Denote the transition probability and transition intensity from state 0 to state i by $P_{0i}(s, t)$ and $\alpha_i(s)$, respectively. Assume that $P_{00}(0, 1) = \exp\{-\int_0^1 \alpha(s) ds\} > 0$, where $\alpha = \sum_{i=1}^m \alpha_i$. Consider n independent Markov chains of this kind, and assume that each process starts at time 0 in state 0. If we denote the sample paths of the individual processes by $S_j(\cdot)$, it follows that

$$N_n^i(t) = \textstyle\sum_{j=1}^n I\{S_j(t) = i\}, \quad i = 1, \cdots, m,$$

is the number of transitions to state i during $[0, t]$. Then each N_n^i is a counting process with corresponding intensity process

$$\Lambda_n^i(t) = \alpha_i(t) Y_n(t), \quad i = 1, \cdots, m,$$

where $Y_n(t) = n - N_n^{\cdot}(t-)$ and $N_n^{\cdot}(t) = \sum_{i=1}^m N_n^i(t)$. The Nelson-Aalen estimator for the cumulative intensity function is

$$\hat{\beta}_n^{\iota}(t) = \int_0^t \{J_n(s)/Y_n(s)\}\, dN_n^{\iota}(s),$$

where $J_n(s) = I(Y_n(t) > 0)$, and a corresponding kernel estimator may be written

$$\hat{\alpha}_n^{\iota}(t) = (1/b_n) \int_0^1 K((t-s)/b_n)\{J_n(s)/Y_n(s)\}\, dN_n^{\iota}(s). \tag{4.2.1}$$

Below we show how it is possible to apply Theorems 4.1.2 and 4.2.2 to this situation. Fix a value of i. Note that since

$$P\{\sup_{t\in[0,1]} |J_n(t) - 1| > \varepsilon\} = \{1 - P_{00}(0, 1)\}^n \to 0,$$

it follows that condition (i) in Theorem 4.1.2 is fulfilled. If we assume that the intensity α_ι is continuous on [0, 1], it follows in a manner similar to (3.2.9) that conditions (ii) and (iii) in Theorem 4.1.2 are satisfied. Thus, if the kernel function is of bounded variation and the intensity is continuous, then the kernel estimator given by (4.2.1) is uniformly consistent in the way described in Theorem 4.1.2.

We then check the conditions in Theorem 4.2.2.

(i) By the Glivenko-Cantelli theorem,

$$nJ_n(s)/Y_n(s) = J_n(s) \Big/ \left\{1 - \frac{1}{n} N_n^{\cdot}(s-)\right\} \to_P 1/P_{00}(0, s)$$

uniformly on [0, 1], since $P_{00}(0, 1) > 0$.

(ii) The function τ is $\tau(s) = P_{00}(0, s) = \exp\{-\int_0^s \alpha(u)\, du\}$, and we assume that each α_ι is continuous.

In summary, we have justified that for a continuous intensity function the kernel estimator given by (4.2.1) is asymptotically normal with a variance equal to

$$(nb_n)^{-1}\alpha_i(t)\exp\left\{\int_0^t \alpha(s)\, ds\right\} \int_{-1}^1 K^2(u)\, du.$$

In a practical situation where one wants to analyse different causes of decrement from an open population, the observations are often censored because of emigration, retirement or other reasons. In that case, a similar result holds if the censoring only depends on outside random mechanisms and if the assumptions in Theorem 4.2.2 are fulfilled.

5. A numerical example. As an illustration of the kernel function smoothing method, we have smoothed the cumulative death rate of 488 patients with cirrhosis of the liver. The figures originate from the Copenhagen Study Group for Liver Diseases, and some results may be found in two reports from the Group (1969, 1974). The purpose of the study was to evaluate the effect of prednisone on the survival of patients with cirrhosis, compared with a suitable control group. By randomisation, the patients with cirrhosis were divided into two groups which received prednisone and placebo tablets, respectively. A preliminary analysis carried out by the Study Group showed that the mortality in the prednisone group was nearly identical to that of the control group (i.e., the difference was not significant). Therefore, the purpose of the present example will be to determine the common mortality in the groups, taken together.

Out of the 488 patients, observation of 196 was censored during the observational period. We will measure time in days elapsed since medication by prednisone or placebo started, which was close to the day when cirrhosis was diagnosed. We assume that the patients behave independently of each other. Then the number $N(t)$ of deaths up to time t is a counting process with intensity process

$$\Lambda(t) = \alpha(t)Y(t),$$

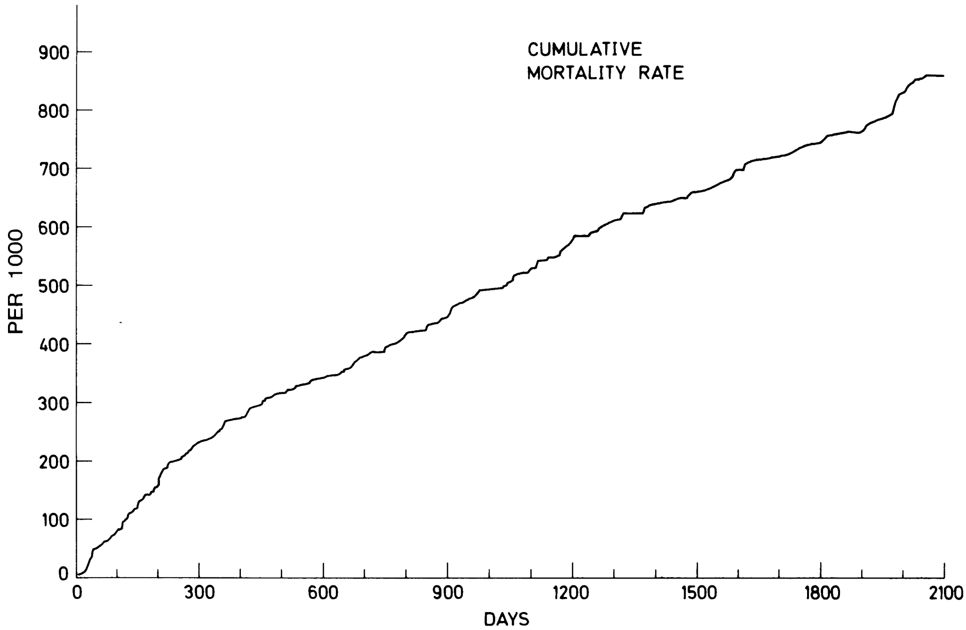

FIG. 1. *Estimated cumulative force of mortality for* 488 *patients from the Copenhagen Study Group for Liver Diseases with cirrhosis of the liver.*

where α is the common hazard rate (force of mortality) and $Y(t)$ is the number of patients alive just before time t. The cumulative mortality rate

$$\hat{\beta}(t) = \int_0^t J(s)/Y(s)\, dN(s)$$

has been plotted in Figure 1 for the first 2100 days of medication. It is difficult to get a precise estimate of α from Figure 1, but it seems that mortality is rather high in the beginning of the period, it decreases up to a duration of some 400 to 500 days, after which it stabilizes at a lower level. These features are illustrated much more lucidly in Figure 2, where we have estimated the force of mortality itself, using $K(x) = 0.75(1 - x^2)$, $|x| \leq 1$, known as Epanechnikov's kernel function. We have also drawn 95% pointwise confidence limits based on asymptotic normality and the variance estimator in (3.2.4). Figure 2 permits us to see that the mortality starts out at a level around 0.8 per 1000 per day, decreases steadily to approximately 0.3 per 1000 per day at a duration of some 500 days, and remains at that level afterwards. These results are in agreement with the previous findings of the Study Group and in accordance with what one would expect from a medical point of view.

The eternal problem of the choice of window size may be attacked in various ways. As in the case of ordinary density estimation, one may derive an asymptotically optimal window or choose a window which minimizes an estimator of the risk function as suggested by Rudemo (1982). Since experience with these methods is still weak, we have simply chosen a window which gives a reasonable picture of the mortality rate. In this connection, it is worth mentioning that if the mortality rate is a linear function over intervals of the form $(t - b, t + b)$, then the kernel estimator will be unbiased. This gives a weak but practical guideline in our choice of window. Since the observations are heavily censored and since the number at risk is also reduced by deaths, we have used a window which increases with time elapsed since medication. We have chosen $b = 100$ for $100 < t \leq 500$ days, $b = 150$ for $500 < t \leq 750$ days, and $b = 200$ for $750 < t \leq 1900$ days. This implies that we are not able to estimate the hazard for $t < 100$ and for $1900 < t \leq 2100$ days.

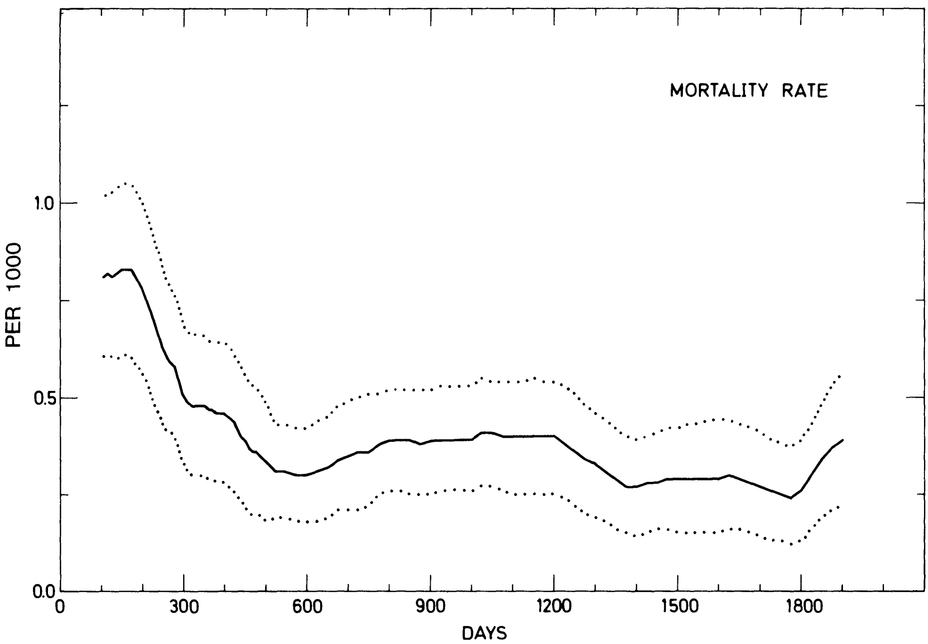

FIG. 2. *Mortality rate for the patients of Figure* 1 *estimated by Epanechnikov's kernel function and a varying window.*

We close this example by pointing out how the analysis can be extended and how the kernel smoothing method may again be put to good use in the extension. To discover why the two groups surprisingly had the same mortality, the investigators have drawn some laboratory, clinical and histological covariates in the analysis, and have reanalysed the data by the method suggested for censored survival data by Cox (1972). For details, see Schlichting et al. (1982). The Cox model specifies the hazard rate $\alpha(t)$ for the survival time X of an individual with a covariate vector z to have the form $\alpha(t; z) = \alpha_0(t)\exp(\phi' z)$. Cox (1972) suggested that ϕ could be estimated through a partial likelihood function

$$L(\phi) = \prod_{i=1}^{n} \{\exp(\phi' z_i)/\textstyle\sum_{j\in R_i} \exp(\phi' z_j)\}^{D_i},$$

where n is the number of possible right censored survival times, $R_i = \{j \mid X_j \geq X_i\}$ and D_i is an indicator for death. If $\hat{\phi}$ denotes the value of ϕ which maximizes the partial likelihood, then

$$\hat{\beta}_0(t) = \textstyle\sum_{X_i \leq t} D_i / \{\sum_{j\in R_i} \exp(\hat{\phi}' z_j)\}$$

was suggested by Breslow (1972, 1974) as an estimator for the cumulative underlying hazard $\beta_0(t) = \int_0^t \alpha_0(s)\, ds$. One may introduce a kernel estimator for the underlying hazard α_0 by defining

$$\hat{\alpha}_0(t) = (1/b) \int_0^\infty K((t-s)/b)\, d\hat{\beta}_0(s),$$

which has the same form as the estimator (3.1.1), and it will surely be possible to derive large sample properties of $\hat{\alpha}_0$ by combining the counting process formulation of the Cox model in Andersen and Gill (1982) together with the theory of the present exposition.

Acknowledgment. I would like to thank Jan M. Hoem and Søren Johansen for their advice throughout my work with this paper, Per Kragh Andersen for useful conversations and for making me aware of the data used in the numerical example, the Copenhagen

Study Group for Liver Diseases for allowing me to use these data, and Richard Gill for his reference to the paper by Liptser and Shiryayev. The comments of an anonymous referee were also helpful. In connection with my work on this project I had the benefit of a fellowship from the Faculty of Social Science of the University of Copenhagen and economic support for Projekt No. 514-20382 from the Danish Social Science Research Council.

REFERENCES

AALEN, O. O. (1976). Nonparametric inference in connection with multiple decrement models. *Scand. J. Statist.* **3** 15–27.

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

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

BEAN, S. J. and C. P. TSOKOS (1980). Developments in non-parametric density estimation. *Int. Statist. Rev.* **48** 267–287.

BORGAN, Ø. (1979). On the theory of moving average graduation. *Scand. Actuarial J.* 83–105.

BRESLOW, N. (1972). Contribution to the discussion of the paper by D. R. Cox. *J. Roy. Statist. Soc. B***34** 187–220.

BRESLOW, N. (1974). Covariance analysis of censored survival data. *Biometrics* **30** 89–99.

THE COPENHAGEN STUDY GROUP FOR LIVER DISEASE (1969). Effect of prednisone on the survival of patients with cirrhosis of the liver. *The Lancet* 119–121.

THE COPENHAGEN STUDY GROUP FOR LIVER DISEASES (1974). Sex, asciies and alcoholism in survival of patients with cirrhosis. *New England J. of Medicine* **291** 271–273.

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

FÖLDES, A., REJTÖ, L., and WINDER, B. B. (1981). Strong consistency properties of nonparametric estimators for randomly censored data. II: Estimation of density and failure rate. *Periodica Math. Hungarica* **12** 15–30.

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

LENGLART, E. (1977). Rélation de domination entre deux processes. *Ann. Inst. Henri Poincaré* **13** 171–179.

LIPTSER, R. SH. and SHIRYAYEV, A. N. (1980). A functional central limit theorem for semimartingales. *Theor. Probab. Appl.* **25** 4, 667–688.

NELSON, W. (1972). Theory and application of hazard plotting for censored failure data. *Technometrics* **14** 945–965.

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

RICE, J. and ROSENBLATT, M. (1976). Estimation of the log survivor function and hazard function. *Sankhya Ser. A* **38** 60–78.

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

RUDEMO, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.* **9** 65–78.

SCHLICHTING, POUL, E. CHRISTENSEN, PER KRAGH ANDERSEN, L. FAUERHOLDT, E. JUUL, H. POULSEN, N. TYGSTRUP, and THE COPENHAGEN STUDY GROUP FOR LIVER DISEASES (1982). Identification of prognostic factors in cirrhosis using Cox-regression model. Unpublished manuscript.

SHIRYAYEV, A. N. (1981). Martingales: Recent developments, results and applications. *Int. Statist. Rev.* **49** 199–234.

WATSON, G. S. and LEADBETTER, M. R. (1964a). Hazard analysis I. *Biometrika* **51** 175–184.

WATSON, G. S. and LEADBETTER, M. R. (1964b). Hazard analysis II. *Sankhya Ser. A* **26** 101–116.

YANDELL, BRAIN S. (1981). Non-parametric inference for rates and densities with censored serial data. Unpublished paper, Biostatistics Program, University of California, Berkeley.

LABORATORY OF ACTUARIAL MATHEMATICS
UNIVERSITY OF COPENHAGEN
UNIVERSITETSPARKEN 5
2100 COPENHAGEN, Ø
DENMARK