Martin Jacobsen

Maximum-Likelihood Estimation in the Multiplicative Intensity Model

Preprint
August
1982
4

Institute of Mathematical Statistics
University of Copenhagen

Martin Jacobsen

MAXIMUM-LIKELIHOOD ESTIMATION IN THE MULTIPLICATIVE INTENSITY MODEL

Preprint 1982 No. 4

INSTITUTE OF MATHEMATICAL STATISTIC
UNIVERSITY OF COPENHAGEN

August 1982

SUMMARY

The non-parametric multiplicative intensity model for simple point processes on the line is considered. By introducing a suitable topology on the space of all point process paths, a topological extension of the model is found, accomodating for point processes with multiple points. In the extended model the maximum-likelihood estimator is determined and shown to be interpretable in terms of a multiple point process. Finally, the maximum-likelihood estimator and the Aalen estimator are compared. The results are illustrated by examples involving a model for Poisson processes, a model for censored survival data and the Cox regression model.

Key words Multiplicative intensity model, generalized maximum-likelihood estimator, Aalen estimator, Kaplan-Meier estimator, Cox regression model, point processes with multiple points.

1. INTRODUCTION

In a basic paper Aalen (1978) introduced the multiplicative intensity model for counting processes, defined a class of martingale estimators for the unknown integrated intensity functions that serve as parameters in the model, and showed that under suitable conditions, these estimators are well behaved asymptotically.

The models are non-parametric in the sense that the unknown intensities may be chosen arbitrarily from a class of functions, rather than a finite-dimensional parameter set. This means that e.g. standard maximum-likelihood methods do not apply, and that creating the Aalen estimator is a matter of definition, not of derivation.

One gap in the Aalen theory is that it is difficult to interpret the estimator: while any integrated intensity from the model is a continuous function, the estimator turns out a step function. An exact analogue (a special case) is provided by considering i.i.d. observations from an unknown distribution, supposed to be continuous, yet estimating that distribution by the discrete empirical distribution.

Of course, in the situation with i.i.d. observations, it is well known how to view the empirical distribution as a maximum-likelihood estimator. More generally, for i.i.d. observations with right censoring, it is known that the Kaplan-Meier estimator is the maximum-likelihood estimator. But in both cases, the derivation of the maximum-likelihood estimator and its interpretation is made possible only by considering an extension of the original model, namely the distribution generating the data is allowed to be completely arbitrary, in particular it may be discrete.

This paper proposes an extension of the multiplicative intensity model. To accomodate for analogues of the discrete distributions appearing in the extended models for i.i.d. data, it is necessary to introduce counting processes with multiple jumps. The space of paths W^* for such processes is equipped with a suitable topology, and the desired extension of the Aalen model P then arises by considering the weak closure of P when using on the space of probabilities on W^*, the weak topology derived from the topology on W^*. In the extended model, the maximum-likelihood estimator is then found by maximizing the probability of the observation.

It emerges that the Aalen estimator is not in general the maximum-likelihood estimator. But it is shown that the difference is often negligeable.

The extension used here is based on topological considerations. Alternatively one might try to find an algebraic extension, but there does not seem to be any natural way of doing this. One such extension has been proposed by Johansen (1981) and is discussed in Section 5 below.

In Section 2, the necessary terminology is introduced. In Section 3 the structure of counting processes with multiple jumps is recapitulated and in Section 4 the extension of a simple Markov process model is determined. The most important results appear in Section 5, where the extension of the general Aalen model is studied under assumptions that make it possible to apply locally in time the results from Section 4. The methods are illustrated by means of the model for i.i.d. censored observations mentioned above, and by the Cox regression model. Finally, in Section 6, a comparison is made between the Aalen estimator and the maximum-likelihood estimator.

2. COUNTING PROCESSES AND THE MULTIPLICATIVE INTENSITY MODEL

By definition a one-dimensional counting process $N = (N_t)_{t \geq 0}$ is an integer-valued stochastic process satisfying $N_0 = 0$ and with right-continuous non-decreasing sample paths, increasing only by jumps of size 1.

We shall here use the canonical realisation of such a process, i.e. we shall denote by W the space of all right-continuous paths $w: [0,\infty) \to \mathbb{N}_0 = \{0,1,\ldots\}$ with $w(0) = 0$, $w(s) \leq w(t)$ for $s \leq t$ and $\Delta w(t) = 0$ or 1 for all t with $\Delta w(t) = w(t) - w(t-)$ the jump of w at t; we shall define N_t by $N_t(w) = w(t)$ and then simply consider probabilities on W.

The measurable structure to be used on W consists of the σ-algebra $\mathcal{F}$ spanned by all $(N_t)_{t \geq 0}$, and, for every t, the pre-t σ-algebra $\mathcal{F}_t$ spanned by $(N_s)_{s \leq t}$. Further, for $t > 0$ we define $\mathcal{F}_{t-}$ as the smallest σ-algebra containing all $\mathcal{F}_s$ for $s < t$. (Note that we do not assume $\mathcal{F}$ or $\mathcal{F}_t$ to have been completed with respect to one or more probabilities, as is the custom in the general theory of processes).

With this setup, according to the terminology from Jacobsen (1982), a stable canonical counting process is a probability on $(W,\mathcal{F})$. Here 'stable' refers to the assumption that only finitely many jumps are allowed in finite time. With P such a process, we write PU rather than EU for the P-expectation of U.

A real-valued process $Z = (Z_t)_{t \geq 0}$ defined on $(W,\mathcal{F})$ is measurable if $(t,w) \to Z_t(w)$ is measurable, adapted if each Z_t is $\mathcal{F}_t$-measurable and predictable if Z is measurable and each Z_t is $\mathcal{F}_{t-}$-measurable.

Suppose now given a probability P on $(W,\mathcal{F})$ such that $PN_t < \infty$ for all t, and suppose that the integrated intensity for N with respect to P may be written

$$\Lambda_t = \int_0^t \lambda_{s-} \, ds \ ,$$

where $\lambda_- = (\lambda_{t-})_{t>0}$, the intensity, is a left-continuous, adapted non-negative process with right-limits, (in particular λ_- is predictable). Thus Λ is continuous, predictable and $M = N - \Lambda$ is a P-martingale. Recall that λ_- or Λ determines P: given a process λ_- which is the intensity for some P, that P is unique. With $\tau_k = \inf\{t : N_t = k\}$, P is completely specified by the distribution of τ_1 and, for every $k \geqq 1$, the conditional distribution of τ_{k+1} given $\xi_k = (\tau_1, \ldots, \tau_k)$. The assumptions about P imply that we may write

$$P(\tau_{k+1} > t \mid \xi_k) = \exp(- \int_{\tau_k}^t \mu_{k,\xi_k}(s) ds)$$

on the set $(\tau_k < \infty)$, where each μ-function is right-continuous with left-limits and

$$\lambda_{t-} = \mu_{N_{t-}, \xi_{N_{t-}}}(t-) \ .$$

The multiplicative intensity model arises by specifying that

$$\lambda_{t-} = \alpha(t-) Z_{t-} \ , \tag{2.1}$$

where α, the unknown parameter, is a suitable non-negative, right-continuous, left-limit function and Z is a given adapted, non-negative process such that Z_t is right-continuous in t with left-limits. To ensure that λ_- given by (2.1) is the intensity for a stable counting process, some further conditions must be imposed on α, Z. As in Jacobsen (1982) we shall assume that $\int_0^t \alpha < \infty$ for all t and for the moment that

$$Z_t \leqq c + d \ N_t \tag{2.2}$$

for some constants c and d.

With these definitions, the model is characterized exclusively by the process Z. In the terminology of Jacobsen (1982) the model is the full Aalen model for Z.

The Aalen estimator (Aalen (1978)) of the integrated intensity

$$\beta_t = \int_0^t ds\, \alpha(s) 1_{(Z_s > 0)}$$

is

$$\overline{\beta}_t = \int_{(0,t]} N(ds) \frac{1}{Z_{s-}} 1_{(Z_{s-} > 0)} , \qquad (2.3)$$

the prototype of what may be called martingale unbiased estimators or m-estimators (Rebolledo (1978)), i.e. $\hat{\beta} - \beta$ is a martingale.

Apart from being very simple to compute, the usefulness of the Aalen estimator lies in its good asymptotic properties as demonstrated by Aalen (1978). (The asymptotic theory is based on what in Jacobsen (1982) is called the product Aalen model, where information about α comes from r independent processes. These, perhaps the most important of the multiplicative intensity models, are discussed in Section 5 below).

One problem with the Aalen estimator is that β_t, which is continuous in t, is estimated by the step function $\overline{\beta}_t$, i.e. the estimator is not range-preserving. This makes it difficult to interpret the estimator: there is no obvious guess of what counting process from the model generated the data.

This paper proposes an alternative to the Aalen estimator, which is the maximum-likelihood estimator in an extension of the multiplicative intensity model.

In a recent paper Johansen (1981) has proposed a different extension and derived the Aalen estimator as maximum-likelihood estimator. To this author, Johansen's extension appears unsatisfactory for various reasons, as will be argued at the end of Section 5.

Write P_α for the process (probability) with intensity (2.1). Assuming the process to be observed on the interval $[0,t]$, i.e. considering the restriction to $\mathcal{F}_t$ of P_α, the likelihood function for the Aalen model is proportional to

$$\ell_t = \exp(-\int_0^t ds\, \alpha(s) Z_s) \prod_{k=1}^{N_t} \alpha(\tau_k -) Z_{\tau_k -} , \tag{2.4}$$

cf. Jacobsen (1982, Theorem 1.6.1). Except in the trivial case where no jumps are observed at all, we have $\sup_\alpha \ell_t = \infty$, a value that is gradually attained by choosing α's with $\int_0^t \alpha$ small and the α-value at each observed jump time big. (Recall from Jacobsen (1982 , Proposition 1.4.8 (e)) that $\lambda_{\tau_k -} > 0$ a.s. on $(\tau_k < \infty)$, so that $Z_{\tau_k -} > 0$).

Thus there is no maximum-likelihood estimator in the Aalen model. Instead we shall derive the estimator to be proposed here, by the following principle, applicable in much greater generality (see Kiefer and Wolfowitz (1956)).

Suppose X is a random element taking values in a metric space S, and suppose given a statistical model for observation of X, i.e. a family $\mathcal{P}$ of probabilities on S, the possible distributions for X. Consider the situation where $\mathcal{P}$ is dominated, so that the likelihood function L is well defined, but assume that $\sup_{\mathcal{P}} L = \infty$, so there is no maximum-likelihood estimator. Then instead of $\mathcal{P}$ consider the extended model $\overline{\mathcal{P}}$ obtained as the weak closure of the set $\mathcal{P}$, where the space of probabilities on S is equipped with the weak topology matching the given topology on S. The new model $\overline{\mathcal{P}}$ may not be dominated, but following Kiefer and Wolfowitz (1956), a maximum-likelihood estimator may still be defined and will often exist. In particular, if to every possible value $x \in S$ of X there corresponds a $P \in \overline{\mathcal{P}}$ with $P(\{x\}) > 0$, the maximum-likelihood estimator for the unknown distribution of X, is any probability $\hat{P} \in \overline{\mathcal{P}}$ which maximizes the probability of the observation, i.e.

$$\hat{P}(\{x\}) = \sup_{\overline{\mathcal{P}}} P(\{x\}) \ .$$

In order to apply this principle to the Aalen model, we must define a topology on W, in fact we shall introduce a larger space W^*, the path-space for counting processes with multiple jumps. The reason for this is the following: as remarked above, ℓ_t given by (2.4) gets large by choosing α's with $\int_0^t \alpha$ small and with big values at the observed jump times. If these are denoted by t_j, it is clear that with such an α, P_α will be a counting process with all jumps close to some t_j, but with more than one jump possible in a small neighborhood of t_j. Performing a limit operation via a sequence of peaked α's with small integrals, leads in a natural way to the inclusion of paths that may have jumps > 1. The topology to be used on this space, is defined in the next section.

It should be pointed out, that typically the maximum-likelihood estimator found in the extended model $\overline{\mathcal{P}}$, will belong to $\overline{\mathcal{P}} \setminus \mathcal{P}$ (because usually $P(\{x\}) = 0$ for $P \in \mathcal{P}$, $x \in S$). For the estimator to have good asymptotic properties (consistency for instance), it is therefore important that $\overline{\mathcal{P}} \setminus \mathcal{P}$ be dense in $\mathcal{P}$. This will be true for the estimator we shall determine in Section 5.

3. THE MULTIPLE JUMP SPACE W^*

Let W^* be the space of increasing right-continuous paths $w:[0,\infty) \to \mathbb{N}_0$ with $w(0) = 0$, $w(s) \leqq w(t)$ for $s < t$ and $\Delta w(t) \in \mathbb{N}_0$ for all t. Thus W is the subset

$$W = \{w \in W^* : \Delta w(t) \leqq 1 \quad \text{for all } t\}$$

of W^*. On W^* we shall write $N_t^*(w) = w(t)$ and denote the relevant σ-algebras by $\mathcal{F}^*, \mathcal{F}_t^*, \mathcal{F}_{t-}^*$.

Each $w \in W^*$ may be identified with a positive integer-valued measure $\kappa = \kappa_w$ on $[0,\infty)$ given by $\kappa\{0\} = 0$ and

$$\kappa(s,t] = w(t) - w(s) \quad \text{or} \quad \kappa(B) = \sum_{t \in B} \Delta w(t)$$

for $0 \leqq s \leqq t$ and $B \subset [0,\infty)$ a Borel set respectively.

Equipping the space of all such measures with the vague topology so that $\kappa_n \to \kappa$ iff

$$\int f d\kappa_n \to \int f d\kappa$$

for all bounded continuous $f : [0,\infty) \to \mathbb{R}$ with compact support, we obtain the <u>vague topology on W^*</u>. Obviously $w_n \to w$ iff

$$w_n(t) \to w(t) \quad \text{for all } t \text{ with } \Delta w(t) = 0 .$$

If for instance $w_n \in W$ with k jumps situated at $t_{1,n} < \ldots < t_{k,n}$ and if $t_{j,n} \to t_0 > 0$ for all j, then $w_n \to w \in W^*$, where

$$w(t) = k \, 1_{[t_0,\infty)}(t) .$$

This agrees with the ideas sketched at the end of Section 2. Notice that w_n does not converge to w in the Skorohod topology, which is finer than the vague topology and too fine for our purpose.

With the vague topology on W^*, consider the matching weak topology on the space of probability measures on W^*, so that $P_n \Rightarrow P$ iff

$$\int \phi dP_n \to \int \phi dP$$

for all bounded vaguely continuous $\phi : W^* \to \mathbb{R}$.

The following result is a special case of Kallenberg (1975, Theorem 4.2).

3.1. Proposition Let $(P_n)_{n \geq 1}, P$ be probabilities on W^*. In order that $P_n \Rightarrow P$ it is sufficient that for some $\varepsilon > 0$ and for some dense set $D \subseteq [0,\infty)$ of points t with $P(\Delta N^*_t = 0) = 1$,

$$P_n(N^*_{u_k} - N^*_{t_k} = x_k, k = 1,\dots,r) \to P(N^*_{u_k} - N^*_{t_k} = x_k, k = 1,\dots,r)$$

for all $r \in \mathbb{N}$, all $t_1,\dots,t_r,u_1,\dots,u_r \in D$ with $0 < u_k - t_k < \varepsilon$ and all $x_k \in \mathbb{N}_0$. □

An alternative criterion may be obtained as follows. With the convention $\inf \emptyset = \infty$, define for $k \in \mathbb{N}$,

$$\sigma_k = \inf\{t : N^*_t \geq k\} .$$

In particular $\sigma_k = \tau_k$ on W. Consider the mapping

$$w \overset{\psi}{\to} (\sigma_1(w), \sigma_2(w), \dots)$$

from W^* onto the space S of sequences $(s_k)_{k \geq 1}$ with $0 < s_k \leq \infty$, $s_1 \leq s_2 \leq \dots$. On S introduce the topology for coordinatewise convergence, so that $(s_{n,k})_{k \geq 1} \to (s_k)_{k \geq 1}$ iff for every k, $s_{n,k} \to s_k$ as a limit in $(0,\infty]$. The proof of the following result is absolutely straightforward and is omitted.

3.2 Lemma The mapping ψ is a homeomorphism from W^* onto S. □

With P a probability on W^*, write $Q = \psi(P)$ for the P-distribution of $(\sigma_k)_{k \geq 1}$, and $Q^{(k)}$ for the P-distribution of $(\sigma_1, \ldots, \sigma_k)$.

3.3. <u>Proposition</u> Let $(P_n)_{n \geq 1}$, P be probabilities on W^*. In order that $P_n \Rightarrow P$ it is necessary and sufficient that $Q_n^{(k)} \Rightarrow Q^{(k)}$ for every $k \in \mathbb{N}$.

<u>Proof</u> By Lemma 3.2 $P_n \Rightarrow P$ iff $Q_n \Rightarrow Q$. But $Q_n \Rightarrow Q$ iff $Q_n^{(k)} \Rightarrow Q^{(k)}$ for each k and the sequence $(Q_n)_{n \geq 1}$ is tight. So we only have to show that if $Q_n^{(k)} \Rightarrow Q^{(k)}$ for all k, then (Q_n) is tight. Since in particular $Q_n^{(1)} \Rightarrow Q^{(1)}$, to every $\varepsilon > 0$ there is a compact subset K_1 of $(0,\infty]$ with $Q_n^{(1)}(K_1) \geq 1 - \varepsilon$ for all n. But then, since $s_1 \leq s_2 \leq \ldots \leq \infty$ for $(s_k) \in S$,

$$K = \{(s_k)_{k \geq 1} \in S : s_1 \in K_1\}$$

is a compact subset of S with $Q_n(K) \geq 1 - \varepsilon$ for all n . □

We shall conclude this section with a brief discussion of the structure of the probabilities on W^*. Such a probability P is a canonical counting process with multiple jumps and, as shown by Jacobsen (1982, Section 2.5), may be viewed as a multivariate canonical counting process with infinite type-set $E = \mathbb{N}$ corresponding to the possible jump sizes. In particular, if $\tau_k^*(w)$ denotes the time of the k'th jump for $w \in W^*$ and $Y_k(w) = \Delta N_{\tau_k^*}(w)$ the size of that jump (defined only if $\tau_k^*(w) < \infty$), then with $\xi_k^* = (\tau_1^*, \ldots, \tau_k^*, Y_1, \ldots, Y_k)$, P is specified by the conditional probabilities

$$P(\tau_{k+1}^* > t \mid \xi_k^*), \qquad P(Y_{k+1} = y \mid \xi_k^*, \tau_{k+1}^*)$$

determined on the sets $(\tau_k^* < \infty)$ and $(\tau_{k+1}^* < \infty)$ respectively.

We shall mainly be interested in the case where P is <u>purely discrete</u>. Recall from Jacobsen (1982) that a probability Pr on $(0,\infty]$ with survivor function G is purely discrete if there is a countable subset D of $(0,\infty]$ with $Pr(D) = 1$ and all points in $D \cap (0,t^\dagger)$ isolated, where $t^\dagger =$

$\inf\{t : Pr(t,\infty] = 0\}$. With this in mind, P is purely discrete if the conditional distribution of τ^*_{k+1} given ξ^*_k is purely discrete for all possible values of ξ^*_k.

The intensity function for Pr is $\mu(t) = Pr\{t\}/Pr[t,\infty]$ and $G(t) = \prod_{0<s\leqq t} (1-\mu(s))$. Thus, if P is purely discrete we may write

$$P(\tau^*_{k+1} > t \mid \xi^*_k) = \prod_{\tau_k < s \leqq t} (1-\mu_{\xi^*_k}(s)) .$$

Introducing $N^y_t = \sum_{s\leqq t} 1_{(\Delta N^*_s = y)}$, $\overline{N}_t = \sum_{y\geqq 1} N^y_t$ as the number of jumps of size y and the total number of jumps on $[0,t]$ respectively, P is characterized by its (predictable) intensity process $\lambda = (\lambda^y)_{y\geqq 1}$, where

$$\lambda^y_t = P(\Delta N^y_t = 1 \mid \mathcal{F}^*_{t-})$$

$$= P(\Delta N^*_t = y \mid \mathcal{F}^*_{t-}) .$$

Writing

$$P(Y^*_{k+1} = y \mid \xi^*_k, \tau^*_{k+1}) = \pi_{\xi^*_k}(\tau^*_{k+1}, y)$$

we have

$$\lambda^y_t = \mu_{\xi^*_{\overline{N}_{t-}}}(t)\,\pi_{\xi^*_{\overline{N}_{t-}}}(t,y) ,$$

and with $\overline{\lambda} = \sum_{y\geqq 1} \lambda^y$ the total intensity,

$$\overline{\lambda}_t = \mu_{\xi^*_{\overline{N}_{t-}}}(t) .$$

For every $t \geqq 0$, P is concentrated on a countable collection of $\mathcal{F}^*_t$-atoms. With

$$A = (\xi^*_k = (t_1,\dots,t_k;\ y_1,\dots,y_k)\ ,\ \overline{N}_t = k)$$

an arbitrary F_t^*-atom, where $0 < t_1 < \ldots < t_k \leqq t$, $y_1, \ldots, y_k \in \mathbb{N}$, and $\lambda_s^y(A)$, $\overline{\lambda}_s(A)$ the constant values of λ_s^y and $\overline{\lambda}_s$ on A for $s \leqq t$,

$$P(A) = \prod_{\substack{0<s\leqq t \\ s\neq t_j}} (1 - \overline{\lambda}_s(A)) \prod_{j=1}^{k} \lambda_{t_j}^{y_j}(A) \ , \tag{3.4}$$

an expression that is very important for what follows.

For details about the preceding, see Jacobsen (1982, Section 2.5).

About the notation in the sequel: suppose an expression involves a probability on W^* which is concentrated on W; to emphasize this fact, the * symbol is omitted from the expression, which is legitimate since e.g. the restriction to W of τ_k^* is the τ_k of Section 2.

4. EXTENSION OF A MARKOV PROCESS MODEL

In this section we shall discuss maximum-likelihood estimation in a particularly simple Aalen model, namely we shall assume that the intensity (2.1) has the form

$$\lambda_{t-} = \alpha(t-)a_{N_{t-}} , \tag{4.1}$$

where $a_0, a_1, \ldots$ are given non-negative constants with $\Sigma 1/a_k = \infty$.

The process with intensity (4.1) is a Markov process. It is well known that with $\int_0^t \alpha < \infty$ for all t, the condition $\Sigma 1/a_k = \infty$ (which is less restrictive than (2.2)) is necessary and sufficient for this Markov process to have only finitely many jumps in finite time.

We shall first introduce a class of functions which will play a vital role in the theory developed in this section and the next.

For the time being, let $a_k \geq 0$ but do not assume $\Sigma 1/a_k = \infty$. Also let $0 < \beta \leq 1$. Define

$$\pi_0(a_0;\beta) = \beta^{a_0} \tag{4.2}$$

and define recursively

$$\pi_k(a_0,\ldots,a_k;\beta) = \begin{cases} \dfrac{a_{k-1}}{a_{k-1}-a_k} \left(\pi_{k-1}(a_0,\ldots,a_{k-2},a_k;\beta) - \pi_{k-1}(a_0,\ldots,a_{k-1};\beta)\right) & \text{if } a_{k-1} \neq a_k \\[2ex] -\, a_{k-1} \dfrac{\partial}{\partial a_{k-1}} \pi_{k-1}(a_0,\ldots,a_{k-1};\beta) & \text{if } a_{k-1} = a_k \end{cases} . \tag{4.3}$$

For $\beta = 0$ define

$$\pi_k(a_0,\ldots,a_k;0) = \begin{cases} 1 & \text{if } a_0,\ldots,a_{k-1} > 0,\ a_k = 0 \\ 0 & \text{otherwise}. \end{cases} \tag{4.4}$$

The main properties of the functions π_k may be summarized as follows.

<u>4.5 Lemma</u> (a) For $0 < \beta \leqq 1$, the function π_k is jointly continuous in $a_0,\ldots,a_k;\beta$. For $a_0,\ldots,a_k$ fixed $\pi_k(a_0,\ldots,a_k;\beta)$ is continuous at $\beta = 0$.

(b) For $0 < \beta \leqq 1$ and no two a_j equal

$$\pi_k(a_0,\ldots,a_k;\beta) = a_{k-1}\cdots a_0 \sum_{\nu=0}^{k} \frac{\beta^{a_\nu}}{\prod_{\substack{j=0 \\ j\neq\nu}}^{k} (a_j - a_\nu)} .$$

(c) $0 \leqq \pi_k \leqq 1$ always, and $\pi_k(a_0,\ldots,a_k;\beta) = 0$ if $a_0\cdots a_{k-1} = 0$ or if $k \geqq 1$ and $\beta = 1$.

(d) For $0 < \beta \leqq 1$

$$\sum_{k\geqq 0} \pi_k(a_0,\ldots,a_k;\beta) = 1$$

provided $\sum_{k\geqq 0} 1/a_k = \infty$.

<u>Sketch of proof</u> (a) From (4.3) it follows that $\pi_k(a_0,\ldots,a_k;\beta)$ is $a_{k-1}\cdots a_0$ times the k'th order divided difference of the function $f(x) = \beta^x$ evaluated at $a_0,\ldots,a_k$, (see e.g. Nørlund (1954, pp. 8-10)). Since for $\beta > 0$, f is infinitely often differentiable, the first assertion follows from the theory of divided differences.

(b) Follows immediately from Nørlund (1954, p.8, equation (19)).

(c) The first and third part is a consequence of the next proposition. The second is immediate from (b) and the continuity of π_k.

(d) This is another corollary to Proposition 4.6. □

The relevance of the functions π_k is clear from the next result, due to Feller (1940, p.513). Let P_α denote the canonical counting process with intensity (4.1), assuming $\int_0^t \alpha < \infty$, $\Sigma\, 1/a_k = \infty$.

<u>4.6 Proposition</u> Subject to P_α, the process $(N_t)_{t \geqq 0}$ is a Markov process such that $P_\alpha(N_t < \infty) = 1$ for all $t \geqq 0$. The transition probabilities $p_{ij}(s,t) = P_\alpha(N_t = j \mid N_s = i)$ are given by

$$p_{ij}(s,t) = \pi_{j-i}(a_i,\ldots,a_j;\ \exp(-\int_s^t \alpha(u)du)) \tag{4.7}$$

for $i \leqq j$, $s \leqq t$. □

If $a_k = 0$ for some k, the <u>absorbing case</u>, we define the absorbing state to be k_0, the smallest k such that $a_k = 0$. Then $P_\alpha(\sup_t N_t \leqq k_0) = 1$.

Suppose that $\int_0^\infty \alpha = \infty$. In the absorbing case, k_0 is reached in finite time, so taking $i = 0$ in (4.7) and letting $t \to \infty$ it follows that

$$\lim_{\beta \downarrow\downarrow 0} \pi_k(a_0,\ldots,a_k;\beta) = \begin{cases} 1 & \text{if } k = k_0 \\ 0 & \text{otherwise}. \end{cases} \tag{4.8}$$

In the non-absorbing case, $P_\alpha(\sup_t N_t = \infty) = 1$ when $\int_0^\infty \alpha = \infty$ so that the same limit is 0 always.

Let now $\mathcal{P} = \{P_\alpha\}$ denote the model of processes with intensities of the form (4.1) for a given sequence (a_k) with $\Sigma\, 1/a_k = \infty$ and $\int_0^t \alpha < \infty$ for all t.

If $a_0 = 0$, $\mathcal{P}$ only comprises the process which is identically 0. We therefore assume that $a_0 > 0$ from now on.

We shall determine $\overline{\mathcal{P}}$, the weak closure of $\mathcal{P}$. Write

$$G_\alpha(t) = \exp(-\int_0^t \alpha) .$$

For G a survivor function for a probability on $(0,\infty]$, denote by C_G the points of continuity for G in $(0, \infty)$.

<u>4.9 Lemma</u> Let $P_{\alpha_n} \in \mathcal{P}$ for $n \geqq 1$. If $P_{\alpha_n} \Rightarrow P$ for some probability P on W^*, then G_{α_n} converges weakly to some G, i.e. there is a survivor function G for a probability on $(0,\infty]$ such that

$$G_{\alpha_n}(t) \to G(t)$$

for all $t \in C_G$. In addition, $G(t) > 0$ for all t in the non-absorbing case.

<u>Proof</u> By Proposition 3.3,

$$P_{\alpha_n}(\sigma_1 > t) \to P(\sigma_1 > t) := (G(t))^{a_0}$$

for all t with $P(\sigma_1 = t) = 0$. But

$$P_{\alpha_n}(\sigma_1 > t) = P_{\alpha_n}(\tau_1 > t) = (G_{\alpha_n}(t))^{a_0} ,$$

and since $a_0 > 0$, the first assertion of the lemma follows.

To show that $G(t) > 0$ always if all $a_k > 0$, suppose that $G(t_0) = 0$. Choose $t > t_0$ such that $P(\sigma_k = t) = 0$ for all $k \in \mathbb{N}_0$. We have

$$P(N_t^* < k) = P(\sigma_k > t) = \lim_{n\to\infty} P_{\alpha_n}(\sigma_k > t)$$

for every k. But

$$P_{\alpha_n}(\sigma_k > t) = P_{\alpha_n}(N_t < k) = \sum_{0 \leqq j < k} p_{0j}^{(\alpha_n)}(0,t)$$

with $p^{(\alpha_n)}$ the transition probabilities for P_{α_n}. Since $G_{\alpha_n}(t) \to 0$, it follows from (4.7), part (a) of Lemma 4.5 and (4.4) that $P_{\alpha_n}(\sigma_k > t) \to 0$. Thus $P(N_t^* < k) = 0$ for all k violating the fact that $P(N_t^* < \infty) = 1$.

□

4.10 Theorem Let the survivor function G and the sequence (α_n) satisfy

$$\lim_{n\to\infty} G_{\alpha_n}(t) = G(t)$$

for $t \in C_G$ and assume in the non-absorbing case that $G(t) > 0$ for all t.

(a) The quantities

$$p_{ij}(s,t) = \pi_{j-i}(a_i,\ldots,a_j;\ G(t)/G(s)) \tag{4.11}$$

defined for $i \leqq j \in \mathbb{N}_0$, $s \leqq t$ in the non-absorbing case and for $i \leqq j \leqq k_0$, $s \leqq t$ with $G(s) > 0$, together with

$$p_{ij}(s,t) = \delta_{ij} \tag{4.12}$$

for $i \leqq j \leqq k_0$, $s \leqq t$ with $G(s) = 0$ in the absorbing case, are the transition probabilities for a unique Markov probability P on W^* with state-space $\mathbb{N}_0$ in the non-absorbing and $\{1,\ldots,k_0\}$ in the absorbing case.

(b) $P_{\alpha_n} \Rightarrow P$ for $n \to \infty$.

Proof (a) With $p^{(\alpha_n)}$ the transition probabilities for P_{α_n}, it is clear from (4.7) and Lemma 4.5 (a) that

$$p_{ij}^{(\alpha_n)}(s,t) \to p_{ij}(s,t) \tag{4.13}$$

for all $s \leqq t \in C_G$ with $G(s) > 0$. Therefore the Chapman-Kolmogorov equations being valid for $p^{(\alpha_n)}$, hold for p at time points in C_G, i.e.

$$\sum_{i\leqq \ell \leqq j} p_{i\ell}(s,u)p_{\ell j}(u,t) = p_{ij}(s,t) \tag{4.14}$$

for $s \leqq u \leqq t \in C_G$ with $G(u) > 0$. But (4.11) shows that p_{ij} is right-continuous in s,t, hence (4.14) is true for all $s \leqq u \leqq t$ with $G(u) > 0$. It is now easy to extend (4.14) to all $s \leqq u \leqq t$ using (4.12) and the fact that by (4.11),

$p_{ij}(s,t)$ is constant in t if $G(s) > 0$, $G(t) = 0$.

To establish the existence of the Markov probability P it remains to show that $\sum_{j \geq i} p_{ij}(s,t) = 1$ in the non-absorbing and that $\sum_{j \leq k_0} p_{ij}(s,t) = 1$ for $i \leq k_0$ in the absorbing case. (The fact that the Markov chain with transition probabilities p may be realized with right-continuous paths, i.e. as a probability on W^*, is standard, see e.g. Jacobsen (1972)). But this is evident in the absorbing case and follows from (4.11) and Lemma 4.5 (d) in the non-absorbing case since $G(t) > 0$.

(b) It is easy to see that $P(\Delta N_t^* = 0)$ for $t \in C_G$, so by Proposition 3.3 it suffices to show that

$$P_{\alpha_n}(\sigma_1 > s_1, \ldots, \sigma_k > s_k) \to P(\sigma_1 > s_1, \ldots, \sigma_k > s_k) \tag{4.15}$$

for all k and all $s_1 \leq \ldots \leq s_k \in C_G$. But expressing (4.15) in terms of the transition probabilities for P_{α_n} and P, the convergence follows readily using (4.13). □

Thus the closure $\overline{P}$ of the model P comprises the Markov probabilities P described in part (a) of the theorem. As we shall see presently, for maximum-likelihood estimation the important part of the extension $\overline{P} \setminus P$ consists of the P determined by purely discrete G.

Suppose now a Markov chain (with unknown G) arising from $\overline{P}$ is observed on $[0,t]$. With k jumps occurring at $0 < t_1 < \ldots < t_k \leq t$ and y_k the size of the k'th jump, the probability of the observation becomes

$$P(\tau_\ell^* = t_\ell,\ \Delta N_{\tau_\ell}^* = y_\ell,\ \ell = 1,\ldots,k, \tau_{k+1}^* > t)$$

$$= \Big(\prod_{\ell=1}^{k} p_{x_{\ell-1}x_{\ell-1}}(t_{\ell-1}, t_\ell-) p_{x_{\ell-1}x_\ell}(t_\ell-, t_\ell)\Big) p_{x_\ell x_\ell}(t_\ell, t) \tag{4.16}$$

where $x_\ell = y_1 + \ldots + y_\ell$ and $x_0 = 0$, $t_0 = 0$. Here

$$p_{ii}(s,u-) = \lim_{u'\uparrow\uparrow u} p_{ii}(s,u') = (G(u-)/G(s))^{a_i} , \tag{4.17}$$

$$p_{ij}(s-,s) = \lim_{s'\uparrow\uparrow s} p_{ij}(s',u) = \pi_{j-i}(a_i,\ldots,a_j;\ G(s)/G(s-)) \tag{4.18}$$

assuming $G(s) > 0$ for (4.17) and $G(s-) > 0$ for (4.18).

<u>4.19 Theorem</u> (a) The likelihood function in the model $\overline{P}$ for observation of the Markov chain on $[0,t]$ is

$$L_t(G) = \Big(\prod_{k=1}^{\overline{N}_t} p_{N_{\tau^*_{k-1}} N_{\tau^*_{k-1}}}(\tau^*_{k-1},\tau^*_k-) p_{N_{\tau^*_{k-1}} N_{\tau^*_k}}(\tau^*_k-,\tau^*_k)\Big) p_{N^*_t N^*_t}(\tau^*_k,t) ,$$

with the $p_{ij}(s,t)$ given by (4.11) and (4.12).

(b) The maximum-likelihood estimator $\hat{G}$ of G is unique and is a purely discrete survivor function with atoms at $\tau^*_1,\ldots,\tau^*_{\overline{N}_t}$ only. More specifically, if $\hat{\mu}$ is the intensity function for $\hat{G}$, then $\hat{\mu}(s) = 0$ for $s \neq \tau^*_k$ while $\hat{\mu}(\tau^*_k)$ is the value of $0 \leqq \mu \leqq 1$ maximizing

$$\pi_{\Delta N^*_{\tau^*_k}}(a_{N^*_{\tau^*_{k-1}}},\ldots,a_{N^*_{\tau^*_k}};\ 1-\mu) . \tag{4.20}$$

(c) In the special case where all $\Delta N^*_{\tau^*_k} = 1$, $\hat{\mu}$ is given by $\hat{\mu}(s) = 0$ for $s \neq \tau^*_k$ and

$$\hat{\mu}(\tau_k) = \begin{cases} 1 - \left(\dfrac{a_k}{a_{k-1}}\right)^{1/(a_{k-1}-a_k)} & \text{if } a_{k-1} \neq a_k \\[2ex] 1 - \exp\left(-\dfrac{1}{a_{k-1}}\right) & \text{if } a_{k-1} = a_k . \end{cases} \tag{4.21}$$

<u>Proof</u> (a) The likelihood function is just (4.16) rewritten.

(b) and (c). Consider first the non-absorbing case. Since it is seen that $L_t(G) = 0$ if $G(t) = 0$, we must have $\hat{G}(t) > 0$. But then (4.17) shows that $\hat{G}$

must be constant on the intervals $[0,\tau_1^*),(\tau_1^*,\tau_2^*),\ldots,(\tau_{\overline{N}_t}^*,t]$ in order that the factors in L_t involving transitions p_{ii} attain their maximal value 1. The last assertion of (b) follows from (4.18), except for the uniqueness of the maximum, which we shall not argue. Part (c) is likewise immediate from (4.18), using that $N^*_{\tau_k^*}=k$ because all $\Delta N^*_{\tau_j^*}=1$, and that

$$\pi_1(a_{k-1},a_k;\beta) = \begin{cases} \dfrac{a_{k-1}}{a_k-a_{k-1}}\left(\beta^{a_{k-1}}-\beta^{a_k}\right) & \text{if } a_{k-1}\neq a_k \\ -\,a_{k-1}\,\beta^{a_{k-1}}\log\beta & \text{if } a_{k-1}=a_k\,. \end{cases} \tag{4.22}$$

The same reasoning applies in the absorbing case if one works strictly to the the left of the time where k_0 is reached. If that happens at τ_j^*, then (4.20) is < 1 if $\mu<1$ (use the interpretation of π provided in Proposition 4.6), and by (4.4) = 1 if $\mu=1$. Thus $\hat{\mu}(\tau_j^*)=1$ so that $\hat{G}(\tau_j^*)=0$ forcing $\hat{G}\equiv 0$ to the right of τ_j^*. □

Remarks $\hat{G}$ is a survivor function of the type permitted by the model $\overline{P}$, i.e. $\hat{G}(s)>0$ for all s in the non-absorbing case. This is clear from (4.20) because the quantity there vanishes if $\mu=1$ when all $a_k>0$.

Notice also that the expression in (4.22) is well defined: it is only possible to observe the first k jumps to have size 1 if $a_{k-1}>0$.

□

Consider a Markov chain P from $\overline{P}$ generated by a purely discrete G with intensity function μ. Viewed as a counting process with multiple jumps, the intensity process for P is $\lambda = (\lambda^y)_{y\geq 1}$ where

$$\lambda_t^y = P(\Delta N_t^* = y\,|\,F_{t-}^*) = p_{N_{t-}^*,\,N_{t-}^*+y}(t-,t)\ ,$$

and $\overline{\lambda}=\sum_y \lambda^y$ is given by

$$\bar{\lambda}_t = 1 - (1 - \mu(t))^{a_{N_{t-}}} .$$

With these expressions, (3.4) provides an alternative derivation of $L_t(G)$ for G purely discrete.

<u>4.23 Example</u> Let $a_k = 1$ for all k. The model $\bar{P}$ consists of Poisson processes, and if P is determined by the survivor function G, then

$$p_{ij}(s,t) = \frac{1}{(j-i)!} (- \log(G(t)/G(s)))^{j-i} \frac{G(t)}{G(s)}$$

in agreement with

$$\pi_k(1,\dots,1;\beta) = \frac{\beta}{k!} (-\log \beta)^k . \qquad (4.24)$$

In point process terminology, P is the Poisson point process with intensity measure Λ on $(0,\infty]$ given by $\Lambda(0,t] = PN_t^* = -\log G(t)$.

For P_α belonging to the Aalen model P, $\Lambda(0,t] = \int_0^t \alpha$. That the extension $\bar{P} \setminus P$ comprises Poisson processes with multiple points is quite reasonable.

The maximum-likelihood estimator for G can be found explicitly. With $Y_k = \Delta N^*_{\tau_k^*}$ we must maximize

$$\frac{1-\mu}{Y_k!} (-\log(1-\mu))^{Y_k} ,$$

cf. (4.20) and (4.24). This gives

$$\hat{\mu}(\tau_k^*) = 1 - e^{-Y_k} ,$$

which together with $\hat{\mu}(s) = 0$ for $s \neq \tau_k^*$, specifies $\hat{G}$.

The intensity measure $\hat{\Lambda}$ for the Poisson process $\hat{P}$ determined by $\hat{G}$ is

$$
\begin{aligned}
\hat{\Lambda}(0,t] &= -\log \hat{G}(t) \\
&= -\log \prod_{0<s\leq t} (1-\hat{\mu}(s)) \\
&= N_t^*
\end{aligned}
$$

as should be expected. □

5. EXTENSION OF THE GENERAL AALEN MODEL

Consider the Aalen model $\mathcal{P}$ of processes with intensities

$$\lambda_{t-} = \alpha(t-)\ Z_{t-} \tag{5.1}$$

as in (2.1), and let P_α be the process with this intensity, assuming $\int_0^t \alpha < \infty$ for all t.

As a positive adapted right-continuous process $Z = (Z_t)$ has the following representation: for every $k \geqq 0$, $t_1 < \ldots < t_k$ there is a function $z_{k,t_1 \ldots t_k} : [t_k, \infty) \to [0,\infty)$ such that

$$Z_t(w) = z_{k,t_1 \ldots t_k}(t) \qquad (t \geqq t_k)$$

for every $w \in W$ belonging to the $\mathcal{F}_t$-atom $(\tau_1 = t_1, \ldots, \tau_k = t_k, N_t = k)$.

In the Markov chain case considered in the previous section, we have $z_{k,t_1 \ldots t_k}(t) = a_k$.

For most of the remainder of this section we shall make the following assumptions.

A.1 There exists $(a_k)_{k \geqq 0}$ with $\Sigma\, 1/a_k = \infty$ such that $z_{k,t_1 \ldots t_k}(t) \leqq a_k$ for all $k, t_1 < \ldots < t_k \leqq t$.

A.2 For every k, $z_{k,t_1 \ldots t_k}(t)$ is jointly continuous in $0 < t_1 < \ldots < t_k \leqq t$ and extends to a continuous function of $t_1, \ldots, t_k, t$ on the domain $0 < t_1 \leqq \ldots \leqq t_k \leqq t$.

A.3 Either $z_{k,t_1 \ldots t_k}(t) > 0$ for all $k, t_1 \leqq \ldots \leqq t_k \leqq t$ (the non-absorbing case) or else, for every t there exists $k_0(t)$ finite and decreasing in t, such that $z_{k,t_1 \ldots t_k}(t) = 0$ for $k \geqq k_0(t)$ and all $t_1 \leqq \ldots \leqq t_k \leqq t$ while $z_{k,t_1 \ldots t_k}(t) > 0$ for $k < k_0(t)$ and all $t_1 \leqq \ldots \leqq t_k \leqq t$ (the absorbing case).

Assumption A.1, which may be restated as $Z_t \leqq a_{N_t}$, guarantees that the counting process with intensity (5.1) exists. It is satisfied in particular if (2.2) holds. A.2 may be relaxed a little as will be mentioned later. A.3 implies that any process P_α can only jump after time t if $N_t < k_0(t)$, i.e. $N_s = N_t$ for all $s \geqq t$ P_α-a.s on $(N_t \geqq k_0(t))$. (But it is quite possible that $N_t > k_0(t)$).

<u>5.2 Lemma</u> Let G be a survivor function and suppose that $G_{\alpha_n}(t) \to G(t)$ for $t \in C_G$. Let $f : [0,\infty) \to [0,\infty)$ be continuous. Then for $t \in C_G$

$$\exp(-\int_0^t f(s)\alpha_n(s)ds) \to \exp(-\int_0^t f(s)(-\log G)(ds)) . \tag{5.3}$$

□

The proof is easy and is omitted. If G is purely discrete with intensity function μ, (5.3) reads

$$\exp(-\int_0^t f(s)\alpha_n(s)ds) \to \prod_{0<s\leqq t} (1-\mu(s))^{f(s)} . \tag{5.4}$$

We have the following analogue of Lemma 4.9.

<u>5.5 Lemma</u> Let $P_{\alpha_n} \in \mathcal{P}$ for $n \geqq 1$. If $P_{\alpha_n} \Rightarrow P$ for some probability P on W^*, then $G_{\alpha_n}(t) \to G(t)$ for all $t \in C_G$ with $z_0(t) > 0$.

<u>Proof</u> We have $P_{\alpha_n}(\sigma_1 > t) \to P(\sigma_1 > t)$ if $P(\sigma_1 = t) = 0$. But

$$P_{\alpha_n}(\sigma_1 > t) = \exp(-\int_0^t z_0(s)\alpha_n(s)ds) ,$$

and since by A.3, $z_0(s) > 0$ for $s \leqq t$ if $z_0(t) > 0$, the conclusion follows from Lemma 5.2 with $f(s) = 1/z_0(s)$, which is continuous by A.2. □

We shall not determine in detail the weak closure $\overline{\mathcal{P}}$ of $\mathcal{P}$. With Lemma 5.5 as motivation, we shall consider sequences (α_n) such that

$$G_{\alpha_n}(t) \to G(t) \qquad (t \in C_G) \tag{5.6}$$

for some survivor function G. Further, with reference to the results of Section 4, we shall only consider the case where G is purely discrete. Without further argument we then claim that the partial extension determined this way, always contains the maximum-likelihood estimator.

Let G be purely discrete with intensity function μ so that

$$G(t) = \prod_{0<s\leqq t} (1-\mu(s)) \; .$$

If $t^{\dagger} = \inf\{t : G(t) = 0\} < \infty$, the definition of μ to the right of $t^{\dagger}$ is immaterial. We shall adopt the convention that $\mu(t) = 0$ for $t > t^{\dagger}$.

<u>5.7 Theorem</u> Suppose that assumptions A.1-A.3 hold, let (α_n) be given such that (5.6) is satisfied with G purely discrete with intensity function μ. Suppose finally in the non-absorbing case that $G(t) > 0$ for all t.

(a) The collection $\lambda = (\lambda^y)_{y\geqq 1}$ of predictable processes on the multiple jump path-space W^* defined by

$$\lambda_t^y = \pi_y(z_{x_k}(t),\ldots,z_{x_k+y}(t);\; 1-\mu(t)) \tag{5.8}$$

on the F_{t-}^*-atom

$$(\tau_1^* = t_1, Y_1 = y_1, \ldots, \tau_k^* = t_k, Y_k = y_k, \tau_{k+1}^* \geqq t) \; ,$$

is the intensity process for a unique purely discrete multiple jump process P_μ. Here $t_1 < \ldots < t_k < t$, $x_k = y_1 + \ldots + y_k$ and

$$z_{x_k+\ell}(t) = z_{x_k+\ell, t_1\ldots t_1, \ldots, t_k \ldots t_k, t\ldots t}(t)$$

with t_j repeated y_j times, t repeated ℓ times.

(b) $P_{\alpha_n} \Rightarrow P_\mu$ as $n \to \infty$.

Proof (a) By Lemma 4.5 (d) and A.1

$$\overline{\lambda}_t = \sum_{y \geqq 1} \lambda_t^y = 1 - (1-\mu(t))^{z_{x_k}(t)} \leqq 1 \tag{5.9}$$

if $\mu(t) < 1$.

By assumption, this is always true in the non-absorbing case. We can therefore construct a 'counting process' with multiple jumps, having λ for intensity process. The only problem is that this process may reach ∞ in finite time. But this possibility is excluded by assumption A.1: replacing z_k by the upper bound a_k, yields a Markov process that moves more quickly than (is dominated stochastically by) the one we have constructed, and as we know from Section 4, the Markov process always stays finite.

In the absorbing case, the same reasoning applies except if $\mu(t) = 1$. The process is well defined and finite to the left of t. At time t, (4.4) shows that no jump occurs if $x_k \geqq k_0(t)$, while otherwise a jump to $k_0(t)$ is forced. The convention $\mu(s) = 0$ for $s > t$ now shows that the definition of λ fits if the process is prolonged beyond t by not moving any more.

(b) From (5.9) it is seen that $\overline{\lambda}_t = 0$ whenever $\mu(t) = 0$ so that the process P_μ can only jump at times t that are atoms for G. From Proposition 3.1 it therefore follows easily that to show $P_{\alpha_n} \Rightarrow P$, it is sufficient to show (5.10) and (5.11): for any $k \geqq 1$, $0 < t_1 < \ldots < t_k, y_1, \ldots, y_k \geqq 1$

$$P(\tau_1^* = t_1, Y_1 = y_1, \ldots, \tau_k^* = t_k, Y_k = y_k) \tag{5.10}$$

$$= \lim_{h_1 \downarrow\downarrow 0, \ldots, h_k \downarrow\downarrow 0} \lim_{n \to \infty} P_{\alpha_n}(t_j - h_j < \tau_{x_{j-1}+1} < \ldots < \tau_{x_j} \leqq t_j + h_j < \tau_{x_j+1}, j=1,\ldots,k)$$

where $x_0 = 0$, $x_j = y_1 + \ldots + y_j$; and for $s < t$ such that $[s,t]$ contains no atoms for G

$$\lim_{n\to\infty} P_{\alpha_n}(N_t - N_s = 0) = 1 \ . \tag{5.11}$$

The critical part of the proof consists in establishing (5.10). Introduce for $h_1,\ldots,h_k > 0$ fixed with $t_{j-1} + h_{j-1} < t_j - h_j$

$$A_j = (t_j - h_j < \tau_{x_{j-1}+1} < \ldots < \tau_{x_j} \leqq t_j + h_j) \ .$$

The probability p_n on the right of (5.10) is

$$p_n = P_{\alpha_n}(A_j(t_j + h_j < \tau_{x_j+1}) \ ; \ j = 1,\ldots,k) \ ,$$

which we shall evaluate by conditioning on the σ-algebra G generated by $\tau_1,\ldots,\tau_{x_{k-1}}$ and the event $B = (t_k - h_k < \tau_{x_{k-1}+1})$. Inside B, $P_{\alpha_n}(\cdot|G)$ may be viewed as a counting process on the interval $(t_k - h_k,\infty)$, where the distribution of the time of the first jump $\tau_{x_{k-1}+1}$ has intensity function

$$1_{(t_k-h_k,\infty)}(u_1) z_{x_{k-1},\tau_1\ldots\tau_{x_{k-1}}}(u_1)\alpha_n(u_1) \ ,$$

given further that this jump occurs at u_1, the intensity function governing the distribution of the second jump is

$$z_{x_{k-1}+1,\tau_1 \ \ldots \ \tau_{x_{k-1}},u_1}(u_2)\alpha_n(u_2)$$

and so on.

Now introduce for $0 \leqq \ell \leqq y_k$

$$\overline{z}_{x_{k-1}+\ell} = \sup z_{x_{k-1}+\ell,s_1\ldots s_{x_{k-1}},u_1\ldots u_\ell}(u_{\ell+1}) \ ,$$

where the sup extends over values of s_i and u_i such that

$$t_j - h_j < s_{x_{j-1}+1} < \ldots < s_{x_j} \leqq t_j + h_j \tag{5.12}$$

for $1 \leqq j \leqq k-1$ and

$$t_k - h_k < u_1 < \dots < u_{\ell+1} \leqq t_k + h_k \quad .$$

Let $\underline{z}_{x_{k-1}+\ell}$ be the corresponding inf.

Next write

$$A_k(t_k + h_k < \tau_{x_k+1}) = A_k \setminus C$$

where $C = (t_k - h_k < \tau_{x_{k-1}+1} < \dots < \tau_{x_k} < \tau_{x_k+1} \leqq t_k + h_k)$. Then $P_{\alpha_n}(A_k | G)$ may be bounded above (below) by the probability that the faster (slower) moving counting process on $(t_k - h_k, \infty)$ with jumps generated by the successive intensity functions $\overline{z}_{x_{k-1}} \alpha_n(u_1), \dots, \overline{z}_{x_k - 1} \alpha_n(u_{y_k})$ (intensity functions $\underline{z}_{x_{k-1}} \alpha_n(u_1), \dots, \underline{z}_{x_k - 1} \alpha_n(u_{y_k})$) have at least y_k jumps before time $t_k + h_k$. Since this new process is Markov, the results of Section 4 provide expressions for the two bounds. Bounds for $P_{\alpha_n}(C|G)$ are obtained in the same fashion, requiring that the Markov process have at least $y_k + 1$ jumps before $t_k + h_k$. Since $P_{\alpha_n}(A_k \setminus C | G)$ is determined by z_x for $x \leqq x_k$, we may create $x_k + 1$ an absorbing state for the Markov process, making it simple to compute the probability of at least y_k , respectively $y_k + 1$ jumps. The end result is

$$\begin{aligned}
&\pi_{y_k}(\underline{z}_{x_{k-1}}, \dots, \underline{z}_{x_k}; \gamma_n) + \pi_{y_k+1}(\underline{z}_{x_{k-1}}, \dots, \underline{z}_{x_k}, 0; \gamma_n) \\
&\qquad\qquad - \pi_{y_k+1}(\overline{z}_{x_{k-1}}, \dots, \overline{z}_{x_k}, 0; \gamma_n) \\
&\leqq P_{\alpha_n}(A_k(t_k + h_k < \tau_{x_k+1}) | G) \\
&\leqq \pi_{y_k}(\overline{z}_{x_{k-1}}, \dots, \overline{z}_{x_k}; \gamma_n) + \pi_{y_k+1}(\overline{z}_{x_{k-1}}, \dots, \overline{z}_{x_k}, 0; \gamma_n) \\
&\qquad\qquad - \pi_{y_k+1}(\underline{z}_{x_{k-1}}, \dots, \underline{z}_{x_k}, 0; \gamma_n)
\end{aligned}$$

where $\gamma_n = \exp(-\int_{t_k - h_k}^{t_k + h_k} \alpha_n)$. These constant bounds are valid on the set

$$D = [\bigcap_{j=1}^{k-1} A_j(t_j + h_j < \tau_{x_j+1})] \cap (t_k - h_k < \tau_{x_{k-1}+1}) ,$$

so writing $\underline{r}_k$, $\overline{r}_k$ for the lower and upper bound respectively we get

$$\underline{r}_k P_{\alpha_n}(D) \leqq p_n \leqq \overline{r}_k P_{\alpha_n}(D) .$$

But

$$P_{\alpha_n}(D) = P_{\alpha_n}(\exp(- \int_{t_{k-1}+h_{k-1}}^{t_k-h_k} z_{x_{k-1},\tau_1\ldots\tau_{x_{k-1}}}(u)\alpha_n(u)du ; E_{k-1})$$

where $E_{k-1} = (A_j(t_j + h_j < \tau_{x_j+1}); j = 1,\ldots,k-1)$. Thus

$$\underline{r}_k \exp(- \int_{t_{k-1}+h_{k-1}}^{t_k-h_k} \overline{z}_{x_{k-1}}(u)\alpha_n(u)du) P_{\alpha_n}(E_{k-1}) \leqq \tag{5.13}$$

$$\leqq p_n \leqq \overline{r}_k \exp(- \int_{t_{k-1}+h_{k-1}}^{t_k-h_k} \underline{z}_{x_{k-1}}(u)\alpha_n(u)du) P_{\alpha_n}(E_{k-1}) ,$$

where

$$\overline{z}_{x_{k-1}}(u) = \sup z_{x_{k-1},s_1\ldots s_{x_{k-1}}}(u) ,$$

the sup extending over the same values of s_i as in (5.12), and where $\underline{z}_{x_{k-1}}(u)$ is the corresponding inf.

The inequalities (5.13) bound p_n, the probability of the intersection of the events $A_j(t_j + h_j < \tau_{x_j+1})$ for $j = 1,\ldots,k$, by a factor times the probability of the first $k-1$ of these events. Hence, proceeding by induction, p_n may be bounded above and below by a product of k factors, each containing a r_j and an exponential integral as in (5.13).

By assumption A.2, $\overline{z}_{x_{k-1}}(u)$ and $\underline{z}_{x_{k-1}}(u)$ are continuous in u. Therefore, if $t_{k-1} + h_{k-1} \in C_G$, $t_k + h_k \in C_G$, by Lemma 5.2 and (5.4) e.g.

$$\lim_{n\to\infty} \exp\left(-\int_{t_{k-1}+h_{k-1}}^{t_k+h_k} \bar{z}_{x_{k-1}}(u)\alpha_n(u)\,du\right) = \prod_{t_{k-1}+h_{k-1}<u<t_k+h_k} (1-\mu(u))^{\bar{z}_{x_{k-1}}(u)} .$$

Referring again to A.2 and using the continuity properties of the π-functions, it is now clear that

$$\lim_{h_1\downarrow\downarrow 0,\ldots,h_k\downarrow\downarrow 0}\ \lim_{n\to\infty} p_n$$

$$= \prod_{j=1}^{k} \left[\pi_{y_j}(z_{x_{j-1}}(t_j),\ldots,z_{x_j}(t_j);1-\mu(t_j)) \prod_{t_{j-1}<s<t_j} (1-\mu(s))^{z_{x_{j-1}}(s)}\right]$$

with $t_0 = 0$, $x_0 = 0$. For the proof of (5.10) it remains to identify this expression with $P(\tau_j^* = t_j, Y_j = y_j, 1 \leqq j \leqq k)$ which is straightforward from the definition of λ^y, (5.9) and (3.4).

For the proof of (5.11), define

$$\bar{G}(t) = \lim_{s\downarrow\downarrow t}\ \sup_n\ \exp\left(-\int_0^s \alpha_n(u)\,du\right) ,$$

which is a survivor function.

Consider the non-absorbing case. By assumption, $\bar{G}(s) > 0$ for all s, so the Markov probability Q on W^* with transition probabilities

$$p_{ij}(s,t) = \pi_{j-i}(a_i,\ldots,a_j;\ \bar{G}(t)/\bar{G}(s))$$

is well defined. (The a_i are the constants from assumption A.1). The definition of $\bar{G}$ ensures that for all t, N_t^* under Q is stochastically larger than $N_t^* = N_t$ under P_{α_n} for any n.

Suppose that G has no atoms in $[s,t]$. Given $\varepsilon > 0$, let n be so large that $\exp(-\int_s^t \alpha_n) > 1-\varepsilon$. Then

$$P_{\alpha_n}(N_t - N_s = 0) = \sum_{x=0}^{\infty} P_{\alpha_n}(\exp(-\int_s^t z_{x,\tau_1\ldots\tau_x}(u)\,\alpha_n(u)\,du\,;\ N_s = x)$$

$$\geqq \sum_{x=0}^{\infty} P_{\alpha_n}(\exp(-a_x \int_s^t \alpha_n);\ N_s = x)$$

$$\geqq P_{\alpha_n}(1-\varepsilon)^{a_{N_s}}$$

$$\geqq P_{\alpha_n}(1-\varepsilon)^{b_{N_s}}$$

$$\geqq Q(1-\varepsilon)^{b_{N_s^*}} \quad ,$$

where $b_k = a_0 \vee \ldots \vee a_k$ and we have used stochastic domination for the last step. As $\varepsilon \downarrow 0$, $Q(1-\varepsilon)^{b_{N_s^*}} \to 1$ and (5.11) follows.

In the absorbing case, the same argument applies for $s < t$ such that $G(s) = G(t) > 0$. If $G(s) = 0$, we have

$$P_{\alpha_n}(N_t - N_s = 0) \geqq P_{\alpha_n}(N_s \geqq \kappa_0(s)) = 1 - P_{\alpha_n}(N_s < \kappa_0(s))$$

by A.3 and the following remark. For $k < \kappa_0(s)$ define

$$c_k = \inf z_{k,s_1\ldots s_k}(s) \quad ,$$

the inf extending over s_i with $0 \leqq s_1 \leqq \ldots \leqq s_k \leqq s$. By A.2 and A.3, $c_k > 0$. The process P_{α_n} moves faster than the Markov process Q_{α_n} with transitions

$$q_{ij}(u,v) = \pi_{j-i}(c_i,\ldots,c_j;\ \exp(-\int_u^v \alpha_n)) \quad ,$$

and therefore

$$P_{\alpha_n}(N_s < \kappa_0(s)) \leqq \sum_{x<\kappa_0(s)} \pi_x(c_0,\ldots,c_x;\ \exp(-\int_0^s \alpha_n)) \quad ,$$

and here each term tends to 0 by Lemma 4.5 (a) and (4.4) because

$\exp(-\int_0^s \alpha_n) \to 0$. □

Consider the model $\overline{P}$ obtained by closing P . For observation of the process on $[0,t]$, the likelihood function is 0 for probabilities from P , while by (3.4) it equals

$$L_t(\mu) = \Big(\prod_{\substack{0<s\leq t \\ s\neq\tau_1^*,\ldots,\tau^*_{\overline{N}_t}}} (1-\overline{\lambda}_s)\Big) \prod_{k=1}^{\overline{N}_t} \lambda_{\tau_k^*}^{Y_k} \tag{5.14}$$

for the member of the extended model given by Theorem 5.7 (a).

If the observation comes from the original model, only simple jumps are observed and the maximum-likelihood estimator may be found explicitly. Since the restriction to W (from W^*) of the simple jump intensity λ_t^1 satisfies

$$\lambda_{\tau_k}^1 = \pi_1(Z_{\tau_k-}, Z_{\tau_k}; 1-\mu(\tau_k))$$

the restriction to W of L_t becomes, cf. (4.22)

$$L_t(\mu) = \Big(\prod_{\substack{0<s\leq t \\ s\neq\tau_1,\ldots,\tau_{N_t}}} (1-\mu(s))^{Z_s}\Big) \prod_{k=1}^{N_t} \frac{Z_{\tau_k-}}{Z_{\tau_k}-Z_{\tau_k-}}\Big((1-\mu(\tau_k))^{Z_{\tau_k-}} - (1-\mu(\tau_k))^{Z_{\tau_k}}\Big)$$

(5.15)

and the following result follows quite easily. That the estimator corresponds to one of the processes from Theorem 5.7 follows from A.1 - A.3. (Note that by A.2, the exponent Z_s in (5.15) may be replaced by Z_{s-}).

<u>5.16 Theorem</u> (a) The maximum-likelihood estimator $\hat{P}$ in the model $\overline{P}$ for observation on $[0,t]$, is a purely discrete process of the type given in Theorem 5.7. The intensity function $\hat{\mu}$ determining $\hat{P}$ is given by $\hat{\mu}(s) = 0$ if $s \neq \tau_k^*$ while $\hat{\mu}(\tau_k^*)$ for $k = 1,\ldots,\overline{N}_t$ is the value of $\mu(\tau_k^*)$ which maximizes the observed value of $\lambda_{\tau_k^*}^{Y_k}$ as given by (5.8).

(b) If only simple jumps are observed, then $\hat{\mu}(s) = 0$ for $s \neq \tau_k^* = \tau_k$ and

$$\hat{\mu}(\tau_k) = \begin{cases} 1 - \left(\frac{Z_{\tau_k}}{Z_{\tau_k -}}\right)^{1/(Z_{\tau_k -} - Z_{\tau_k})} & \text{if } Z_{\tau_k -} \neq Z_{\tau_k} , \\ 1 - \exp(-\frac{1}{Z_{\tau_k -}}) & \text{if } Z_{\tau_k -} = Z_{\tau_k} , \end{cases}$$

for $k = 1,\ldots,N_t = \overline{N}_t$. □

We shall generalize Theorem 5.16 to the situation where r independent processes are observed, all having intensities of the form (5.1) with a common unknown α. Thus the intensity of the i'th subprocess is

$$\lambda^i_{t-} = \alpha(t-)Z^i_{t-} \quad . \tag{5.17}$$

We assume that all Z^i satisfy assumptions A.1 - A.3.

Formally we have a product model $P = \{P_\alpha\}$ as defined in Jacobsen (1982, Section 4.2), i.e.

$$P_\alpha = P^1_\alpha \otimes \ldots \otimes P^r_\alpha \quad ,$$

with P^i_α the process with intensity (5.17). If we consider the limiting procedure (5.6) with G purely discrete, then each $P^i_{\alpha_n}$ converges to some purely discrete P^i and evidently then $P_{\alpha_n} \Rightarrow P^1 \otimes \ldots \otimes P^r$, (using of course the product topology on $W^* \times \ldots \times W^*$). This gives a partial extension of the product model P which is rich enough to define the maximum-likelihood estimator. The likelihood function is the product of r factors of the type (5.14).

We shall determine the estimator in the case where the product process is observed on $[0,t]$, and where only simple jumps are observed with no two subprocesses jumping simultaneously. (Of course this is the case if the data is generated by the original model $\mathcal{P}$). Write

$$\tilde{Z}_t = \sum_{i=1}^{r} Z_t^i , \quad \tilde{N}_t = \sum_{i=1} N_t^i ,$$

let $\tilde{\tau}_k$ denote the time of the k'th jump for all the subprocesses combined and put $I_k = i$ if subprocess i jumps at $\tilde{\tau}_k$. Finally, let $\overline{\mathcal{P}}$ denote the weak closure of $\mathcal{P}$.

<u>5.18 Theorem</u> Suppose the product process is followed on $[0,t]$ and that only simple jumps are observed with no two subprocesses jumping simultaneously. The maximum-likelihood estimator $\hat{P}$ in the model $\overline{\mathcal{P}}$ is then the product $\hat{P} = \hat{P}^1 \otimes \ldots \otimes \hat{P}^r$ of purely discrete processes of the type from Theorem 5.7, and the intensity for $\hat{P}^i$ is of the form (5.8) for each i with z replaced by z^i and $\mu = \hat{\mu}$ not depending on i.

The estimator $\hat{\mu}$ for μ is given by $\hat{\mu}(s) = 0$ if $s \neq \tilde{\tau}_1, \ldots, \tilde{\tau}_{\tilde{N}_t}$ and

$$\hat{\mu}(\tilde{\tau}_k) = \begin{cases} 1 - \left(\dfrac{\tilde{Z}_{\tilde{\tau}_k}}{\tilde{Z}_{\tilde{\tau}_k -}} \right)^{1/(Z^{I_k}_{\tilde{\tau}_k -} - Z^{I_k}_{\tilde{\tau}_k})} & \text{if } Z^{I_k}_{\tilde{\tau}_k -} \neq Z^{I_k}_{\tilde{\tau}_k} , \\ 1 - \exp\left(-\dfrac{1}{\tilde{Z}_{\tilde{\tau}_k -}}\right) & \text{if } Z^{I_k}_{\tilde{\tau}_k -} = Z^{I_k}_{\tilde{\tau}_k} \end{cases}$$

for $k = 1, \ldots, \tilde{N}_t$.

<u>Proof</u> If $I_k = i$ the part of the likelihood involving $\mu(\tilde{\tau}_k)$ comprises the factor

$$\pi_1(Z^i_{\tilde{\tau}_k -} , Z^i_{\tilde{\tau}_k} ; 1 - \mu(\tilde{\tau}_k))$$

from subprocesses i, and the factors

$$(1 - \mu(\tilde{\tau}_k))^{Z^j_{\tilde{\tau}_k}}$$

from subprocesses $j, j \neq i$. The result follows readily from this, using that since only process i jumps at $\tilde{\tau}_k$, for $j \neq i$, Z_s^j is continuous at $s = \tilde{\tau}_k$ by A.2. □

If $\tilde{Z}^{I_k}_{\tau_k -} \neq \tilde{Z}^{I_k}_{\tau_k}$, the expression for $\hat{\mu}(\tilde{\tau}_k)$ may be written

$$\hat{\mu}(\tilde{\tau}_k) = 1 - \left(1 + \frac{\Delta_k}{\tilde{\tilde{Z}}_{\tau_k -}}\right)^{-1/\Delta_k} \tag{5.19}$$

where $\Delta_k = \Delta\tilde{\tilde{Z}}_{\tau_k} = \Delta\tilde{Z}^{I_k}_{\tau_k}$. We have used that only subprocess I_k jumps at time $\tilde{\tau}_k$, together with the continuity assumption A.2.

Throughout this section we have assumed A.1 - A.3. The continuity requirements in A.2 may however be relaxed. Under the limit (5.6) with G purely discrete, the conclusion in Theorem 5.7 (b) still holds provided all $z_{n,t_1 \ldots t_n}(t)$ are continuous in neighborhoods of points $(t_1^0, \ldots, t_n^0, t^0)$ with the t_k^0 and t^0 atoms for G, and possess continuous extensions in these neighborhoods as required globally in A.2. It follows in particular, that if the original Aalen model or product Aalen model is the true one, then the maximum-likelihood estimator may be found as in Theorem 5.16 (b) or Theorem 5.18 and has the same interpretation, provided for instance that there is a set $\mathcal{O} \subset [0,\infty)$ of Lebesgue measure 0 such that $z_{n,t_1 \ldots t_n}(t)$ is continuous at all $(t_1, \ldots, t_n, t)$ with $t_k \notin \mathcal{O}$, $t \notin \mathcal{O}$.

5.20 Example Let for $1 \leq i \leq r$, $0 < u_i \leq \infty$ be given and assume that

$$Z^i_{t-} = 1_{(\tau_1^i \wedge u_i \geq t)}$$

with τ_1^i the time of the first jump for the i'th process. The corresponding product model arises when observing i.i.d. lifetimes with right-censoring at $u_1, \ldots, u_r$.

Obviously for every i, $z_n^i \equiv 0$ except for $n = 0$, and $z_0^i(t) = 1_{[0,u_i)}(t)$ with a discontinuity at u_i. However, the remark above applies if we take $\mathcal{O} = \{u_1,\ldots,u_r\}$.

Let μ be a purely discrete intensity function and suppose that $\mu(u_i) = 0$ for all i. Consider the process $P_\mu = P_\mu^1 \otimes \ldots \otimes P_\mu^r$ from the extended product model with each P^i given by Theorem 5.7 (a). We find that $\lambda_t^{i,y} \equiv 0$ for $y \geqq 2$ while by (5.8) $\lambda_t^{i,1} \neq 0$ only on $(\tau_1^{i*} \geqq t)$ and on that set equals

$$\lambda_t^{i,1} = \pi_1(1_{[0,u_i)}(t), 0;\ 1 - \mu(t))$$

$$= 1_{[0,u_i)}(t)\mu(t) \quad .$$

The conclusion is that P_μ is a model for i.i.d. lifetimes with survivor function $G(t) = \prod_{s \leqq t} (1 - \mu(s))$, right-censored at $u_1,\ldots,u_r$.

If the original product model is the true one, the maximum-likelihood estimator may be found using Theorem 5.18. Now

$$\tilde{Z}_{t-} = \sum_{i=1}^{r} 1_{(\tau_1^i \wedge u_i \geqq t)}$$

is the number R_{t-} of individuals at risk (i.e. not dead and not censored) immediately before t. Since a jump at t is possible only if $\tilde{Z}_{t-} > 0$, we always have $\tilde{Z}_{\tilde{\tau}_k -}^{I_k} - \tilde{Z}_{\tilde{\tau}_k}^{I_k} = 1$ and find that $\hat{\mu}(s) = 0$ except at the observed times of death $\tilde{\tau}_k$ where

$$\hat{\mu}(\tilde{\tau}_k) = \frac{1}{R_{\tilde{\tau}_k -}} \quad .$$

The maximum-likelihood estimator $\hat{P}_\mu$ for P_μ is therefore the process generating i.i.d. lifetimes right-censored at $u_1,\ldots,u_r$ with survivor function

$$\hat{G}(t) = \prod_{k:\tilde{\tau}_k \leq t} \left(1 - \frac{1}{R_{\tilde{\tau}_k -}}\right),$$

i.e. the Kaplan-Meier estimator.

In the special case with no censoring (all $u_i = \infty$), $R_{\tilde{\tau}_k -} = r - k + 1$ and $\hat{G}$ is the survivor function for the empirical distribution attaching mass $1/r$ to each $\tilde{\tau}_k$. □

<u>5.21 Example</u> Suppose that for $i = 1,\dots,r$

$$Z^i_{t-} = 1_{(\tau^i_1 \geq t)} e^{(\beta, v_i(t))} \tag{5.22}$$

with $\beta \in \mathbb{R}^p$ and each $v_i(t)$ a p-vector of given continuous covariate functions.

For β given, the corresponding product model is the counting process formulation of the Cox regression model (Cox (1972)) in its simplest form. (See Andersen and Gill (1981) or Jacobsen (1982, Section 4.5). Thus r independent lifetimes are observed with survivor function

$$G^i(t) = \exp\left(- \int_0^t \alpha(s) e^{(\beta, v_i(s))} ds\right) \tag{5.23}$$

for individual i.

We have $z^i_n \equiv 0$ for $n \geqq 1$ and

$$z^i_0(t) = e^{(\beta, v_i(t))},$$

so A.1 - A.3 are satisfied, and for β fixed we may find the maximum-likelihood estimator in the extended product model, using Theorem 5.18. With no two deaths observed to occur simultaneously the result is that

$$\hat{\mu}(\tilde{\tau}_k) = 1 - \left(1 - \frac{e^{(\beta, v_{I_k}(\tilde{\tau}_k))}}{\sum_{i \in R_k} e^{(\beta, v_i(\tilde{\tau}_k))}}\right)^{e^{-(\beta, v_{I_k}(\tilde{\tau}_k))}} ,$$

writing $R_k = \{i : \tau_1^i \geqq \tilde{\tau}_k\}$ for the population at risk immediately before $\tilde{\tau}_k$. (In particular $I_k \in R_k$). The interpretation of $\hat{\mu}$ is that we believe the data to be generated by $\hat{P} = \hat{P}^1 \otimes \ldots \otimes \hat{P}^r$ where the intensity for $\hat{P}^i$ is specified by (see Theorem 5.7 (a)), $\hat{\lambda}_t^{i,y} \equiv 0$ for $y \geqq 2$ while

$$\hat{\lambda}_t^{i,1} = 1_{(\tau_1^{i*} \geqq t)} \, \pi_1(e^{(\beta, v_i(t))}, 0; \; 1 - \hat{\mu}(t))$$

$$= 1_{(\tau_1^{i*} \geqq t)} \, [1 - (1 - \hat{\mu}(t))^{e^{(\beta, v_i(t))}}] .$$

Thus $\hat{P}$ generates independent lifetimes with survivor functions

$$\hat{G}^i(t) = \prod_{0 < s \leqq t} (1 - \hat{\mu}(s))^{e^{(\beta, v_i(s))}} .$$

In the special case with $v_i(t) = v_i$ not depending on t we have

$$\hat{G}^i(t) = (\hat{G}_0(t))^{e^{(\beta, v_i(t))}} ,$$

where $\hat{G}_0$ estimates $\exp(-\int_0^t \alpha)$, the survivor function arising for $\beta = 0$.

We may now proceed to estimate β by inserting $\hat{\mu}$ in the expression for the likelihood function and maximizing the resulting likelihood profile as a function of β. (We have considered an extension of the product model specified by (5.22) for each value of β. We ought to have considered the extension obtained by allowing $\alpha = \alpha_n$ and $\beta = \beta_n$ to vary simultaneously, but this would not give anything new). For observation on $[0,t]$, $L_t(\mu) = L_t(\mu,\beta)$ is a product of r factors of the form (5.15). We get (since process i jumps only once)

$$L_t(\mu,\beta) = \prod_{i=1}^{r} \Big(\prod_{\substack{0<s\leq t \\ s<\tau_1^i}} (1-\mu(s))^{e^{(\beta,v_i(s))}} \Big)\Big[1-(1-\mu(\tau_1^i))^{e^{(\beta,v_i(\tau_1^i))}}\Big]$$

and rearranging the factors, inserting $\mu = \hat{\mu}$ and writing

$$e(i) = e^{(\beta,v_i(\tau_1^i))}, \qquad \Sigma_k = \sum_{i \in R_k} e(i)$$

this yields

$$L_t(\hat{\mu},\beta) = \prod_{k=1}^{\tilde{N}_t} \left(1 - \frac{e(I_k)}{\Sigma_k}\right)^{\Sigma_k/e(I_k) - 1} \frac{e(I_k)}{\Sigma_k} .$$

The factors $e(I_k)/\Sigma_k$ multiply to give Cox's partial likelihood (Cox (1972), (1975), Oakes (1981)) commonly used for estimating β. We see that the partial likelihood does not give the maximum-likelihood estimator for β, although in practice the difference may well turn out to be negligeable.

The estimators derived in this example were first obtained by Bailey (1979). □

Johansen (1981) (see also his discussion of Oakes (1981)) has proposed a different extension of the multiplicative intensity model and derived the Aalen estimator (2.3) as maximum-likelihood estimator. We shall conclude this section with a discussion of his proposal.

Consider the model (5.1). The purely discrete members of Johansen's extension have multiple jumps with intensity $\lambda = (\lambda^y)_{y \geqq 1}$ given by

$$\lambda_t^y = \frac{1}{y!} (Z_{t-} \nu(t))^y \exp(- Z_{t-} \nu(t)) \tag{5.24}$$

with $\nu(t) \geqq 0$ always and > 0 only at a sequence of isolated points.

There are two main reasons why we consider this extension unsatisfactory: 1) the extension does not respect the structure of the original model; 2) the extension is not properly defined by (5.24).

To argue the first point, suppose that $Z_{t-} = 1_{(\tau_1 \geqq t)}$, so that the process with intensity (5.1) only has one jump of size one and

$$P_\alpha(\tau_1 > t) = \exp(- \int_0^t \alpha) .$$

The process P_ν with intensity (5.24) has one jump such that

$$P_\nu(\tau_1^* > t) = \exp(- \sum_{0<s\leqq t} \nu(s)) ,$$

but this jump may be arbitrarily large:

$$P_\nu(Y_1 = y \mid \tau_1^* = t) = \frac{1}{y!} (\nu(t))^y e^{-\nu(t)} .$$

Thus a model for observing one positive random variable (lifetime) is replaced by a model for observing a lifetime <u>and</u> a jump size, which is not natural. (The topological considerations behind the extension defined in this paper, prevents similar things from happening here).

As for the second point made above, look at (5.24). Given the model (5.1), the process Z is only defined on the space W of simple jumps, whereas for (5.24) to define the intensity of a process with multiple jumps, Z must be extended to W^*. It is not at all obvious how this should be done, indeed a major effort in this paper has been devoted to the problem of finding the proper replacement for Z when switching from the simple to the multiple jump case.

Thus (5.24) only specifies the intensity on W, and although this is enough to find the maximum-likelihood estimator when only simple jumps are observed, it is not sufficient for interpretation of the estimator as a probability in the extended model. And to obtain such an interpretation, might be a good reason for considering an extension in the first place.

6. COMPARING THE AALEN ESTIMATOR AND THE MAXIMUM-LIKELIHOOD ESTIMATOR

For the product model (5.17), the Aalen estimator of

$$\beta_t = \int_0^t ds\ \alpha(s)\ 1_{(\tilde{Z}_s > 0)}$$

is defined similarly to (2.3):

$$\bar{\beta}_t = \int_{[0,t]} \tilde{N}(ds) \frac{1}{\tilde{Z}_{s-}} 1_{(\tilde{Z}_{s-} > 0)} .$$

Based on this, the natural estimator of the exponential

$$G_\alpha(t) = \exp\left(-\int_0^t \alpha\right)$$

is

$$\bar{G}(t) = \exp(-\bar{\beta}_t) ,$$

but in fact one often uses instead the product-limit estimator determined by $\bar{\beta}$,

$$\check{G}(t) = \prod_{0<s\leqq t} (1 - \Delta\bar{\beta}_s) ,$$

because this gives for instance the Kaplan-Meier estimator for the model from Example 5.20. ($\check{G}$ only makes sense as long as $\Delta\bar{\beta}_s \leqq 1$, see Section 5.3 of Jacobsen (1982)).

Supposing the model (5.17) to be true, the maximum-likelihood estimator is determined by $\hat{\mu}$ given as in Theorem 5.18. Based on this, the maximum-likelihood estimator of G_α is the product-limit estimator determined by $\hat{\mu}$,

$$\hat{G}_\alpha(t) = \prod_{0<s\leqq t} (1 - \hat{\mu}(s)) ,$$

so the maximum-likelihood estimator of $\int_0^t \alpha$ becomes

$$\hat{\beta}_t = - \log \hat{G}_\alpha(t) .$$

Instead of this one may consider the accumulated maximum-likelihood estimator

$$\check{\beta}_t = \sum_{0<s\leq t} \hat{\mu}(s) .$$

It is seen from (5.19) that formally the Aalen estimator $\bar{\beta}$ and $\check{\beta}$ agree if

$$\Delta \tilde{Z}_{\tilde{\tau}_k} = - 1$$

for all k, as is the case in Example 5.20. (If $\Delta\tilde{Z}_{\tilde{\tau}_k} = 1$ one gets

$$\check{\beta}_t = \int_{(0,t]} \tilde{N}(ds) \frac{1}{\tilde{Z}_s}$$

which resembles $\bar{\beta}_t$, but of course $\check{\beta}_t < \bar{\beta}_t$ on $(\tilde{\tau}_1 \leq t)$).

Asymptotic results (as the number r of subprocesses tends to ∞) for the distribution of $\bar{\beta}$ were given by Aalen (1978), see also Chapter 5 of Jacobsen (1982). The results state that under certain conditions, including integrability conditions we have not used here, the sequence of processes

$$M_r = a_r(\bar{\beta} - \beta) ,$$

where the a_r are constants, $a_r \to \infty$, converges in distribution to a mean-zero Gauss-process with covariance function

$$R(s,t) = \Phi(s) \wedge \Phi(t) ,$$

where Φ is increasing and continuous, i.e. the limit process has independent increments. (Here the distribution of M_r is viewed as a probability on the Skorohod space $D[0,\infty)$. Further the distribution of M_r is considered with respect to $P_\alpha = P_{\alpha,r}$ for some α which is fixed as $r \to \infty$. In particular Φ depends on α. In the notation we have suppressed r when writing $\overline{\beta}$, β. Of course $\tilde{N}$, $\tilde{Z}$ depend on r, and β depends on α also).

Under the relevant integrability assumptions, M_r is for each α and r a $P_{\alpha,r}$-martingale and one of the essential conditions for the weak convergence is that for all α fixed, with respect to the $P_{\alpha,r}$,

$$\langle M_r \rangle_t \to \Phi(t) \quad \text{in probability} \tag{6.1}$$

for each t, with $\langle M_r \rangle$ the quadratic characteristic of M_r.

Subject to some extra assumptions, it may also be shown that the processes

$$a_r(\check{G}(t) - G_\alpha(t)), \qquad a_r(\overline{G}(t) - G_\alpha(t))$$

both converge in distribution to the mean-zero Gauss-process with covariance function

$$V(s,t) = \Phi(s)\, G_\alpha(s)\, G_\alpha(t) \qquad (s \leqq t),$$

see Jacobsen (1982, Section 5.3).

We shall now show that $\check{\beta}$ is close to $\overline{\beta}$ and that $\hat{G}$ is close to $\overline{G}$.

We shall assume the following:

(i) $P_{\alpha,r}\tilde{N}_t < \infty$ for all α, r, t;

(ii) $\langle M_r \rangle_t \to \Phi(t)$ in probability for all α, t;

(iii) there is a constant $c > 0$ such that $|\Delta\tilde{Z}_t(w)| \leqq c$ for all t, w;

(iv) $P_{\alpha,r}(m_r(t) \leqq 2c) \to 0$ where $m_r(t) = \inf_{s \leqq t} \tilde{Z}_s$;

(v) $a_r^2 \int_0^t ds\, \alpha(s)\, P_{\alpha,r} \frac{\tilde{Z}_s}{(\tilde{Z}_s - c)^4} 1_{(\tilde{Z}_s > 2c)} \to 0$ for all α, t.

Condition (i) ensures that

$$\tilde{M}_t = \tilde{N}_t - \int_0^t ds\, \alpha(s)\, \tilde{Z}_s$$

is a $P_{\alpha,r}$-martingale with $P_{\alpha,r}\tilde{M}_t^2 < \infty$. (ii) is just (6.1) repeated, and because

$$\langle M_r \rangle_t = a_r^2 \int_0^t ds\, \alpha(s) \frac{1}{\tilde{Z}_s} 1_{(\tilde{Z}_s > 0)}, \tag{6.2}$$

is typically satisfied when for each s, $\tilde{Z}_s/a_r^2$ converges. This makes (iv) and (v) reasonable, in particular it is plausible that the quantity in (v) is often of the order of magnitude a_r^{-4}. Condition (iii) is a genuine assumption, satisfied in Examples 4.23, 5.20 and 5.21.

6.3 Theorem Suppose (i)-(v) hold. Then

(a) $\sup_{s \leqq t} a_r |\check{\beta}_s - \bar{\beta}_s| \to 0$ in probability for all α, t;

(b) $\sup_{s \leqq t} a_r \left|\frac{\hat{G}(s)}{\bar{G}(s)} - 1\right| \to 0$ in probability for all α, t.

Proof (a). Using a Taylor expansion in (5.19) we get, writing $\tilde{Z}_{k-} = \tilde{Z}_{\tau_k -}$

$$\hat{\mu}(\tilde{\tau}_k) = \Delta\bar{\beta}_{\tilde{\tau}_k} - \frac{1 + \Delta_k}{2\tilde{Z}_{k-}^2} (1 + \theta)^{-1/\Delta_k - 2},$$

where θ is between 0 and $\Delta_k/\tilde{Z}_{k-}$ if $\Delta_k \neq 0$. In all cases (also if $\Delta_k = 0$) we get

$$|\hat{\mu}(\tilde{\tau}_k) - \Delta\bar{\beta}_{\tilde{\tau}_k}| \, 1_{(\tilde{Z}_{k-}>2c)} \leqq \frac{1+c}{2(\tilde{Z}_{k-}-c)^2} 1_{(\tilde{Z}_{k-}>2c)}$$

using (iii). Consequently, on $(m_r(t) > 2c)$,

$$D_t := \sup_{s\leqq t} a_r |\overset{\vee}{\beta}_s - \bar{\beta}_s| \leqq \frac{1+c}{2} \int_{(0,t]} \tilde{N}(ds) \frac{a_r}{(\tilde{Z}_{s-}-c)^2} .$$

Still working on $(m_r(t) > 2c)$, the integral on the right may be written

$$a_r \int_{(0,t]} \tilde{M}(ds) \frac{1}{(\tilde{Z}_{s-}-c)^2} 1_{(\tilde{Z}_{s-}>2c)} + a_r \int_0^t ds\, \alpha(s) \frac{\tilde{Z}_s}{(\tilde{Z}_s-c)^2} 1_{(\tilde{Z}_s>2c)} \qquad (6.4).$$

Considering the square of the first term and taking expectation with respect to $P_{\alpha,r}$ yields

$$a_r^2 P_{\alpha,r} \int_0^t \langle\tilde{M}_r\rangle (ds) \frac{1}{(\tilde{Z}_s-c)^4} 1_{(\tilde{Z}_s>2c)}$$

$$= a_r^2 P_{\alpha,r} \int_0^t ds\, \alpha(s) \frac{\tilde{Z}_s}{(\tilde{Z}_s-c)^4} 1_{(\tilde{Z}_s>2c)}$$

which tends to 0 by (v). Rewriting the second term in (6.4) as (cf. (6.2))

$$a_r \int_0^t ds\, \alpha(s) \frac{1}{\tilde{Z}_s} \frac{\tilde{Z}_s^2}{(\tilde{Z}_s-c)^2} 1_{(\tilde{Z}_s>2c)}$$

and observing that

$$\frac{\tilde{Z}_s^2}{(\tilde{Z}_s-c)^2} 1_{(\tilde{Z}_s>2c)}$$

is uniformly bounded, it follows from (ii) and the fact that $a_r \to \infty$, that the term vanishes in probability as $r \to \infty$. We have thus shown that $D_t\, 1_{(m_r(t)>2c)}$ converges to 0 in probability, and by (iv) this implies (a).

(b). By (5.19)

$$\log\,(1 - \hat{\mu}(\tilde{\tau}_k)) = \frac{1}{\Delta_k}\,\log\,(1 + \frac{\Delta_k}{\tilde{Z}_{k-}})$$

$$= -\,\Delta\,\overline{\beta}_{\tilde{\tau}_k} + \frac{\Delta_k}{2\tilde{Z}^2_{k-}}\,\frac{1}{(1+\theta)^2}$$

with θ between 0 and $\Delta_k/\tilde{Z}_k$. Thus on $(m_r(t) > 2c)$, for some $K > 0$,

$$\sup_{s\leqq t} a_r \left|\frac{\hat{G}(s)}{\overline{G}(s)} - 1\right| \leqq a_r(e^{KD_t/a_r} - 1) \leqq K\,D_t\,e^{KD_t/a_r},$$

which by (a) tends to 0 in probability. □

The maximum-likelihood estimator of $\int_0^t \alpha$ is of course $-\log \hat{G}(t)$, and it would appear natural to compare this with $\overline{\beta}(t)$. Since however $\hat{G}$ may vanish, there may be probability > 0 that $-\log \hat{G}(t) = \infty$, and no analogue of Theorem 6.3 (a) is available.

7. CONCLUDING REMARKS

With the extension of the product Aalen model proposed here, it has been possible to find a maximum-likelihood estimator, and also to interpret this estimator in terms of a well defined counting process with multiple jumps. Furthermore, evidence has been given that certain functionals derived from the estimator are asymptotically equivalent to the Aalen estimator and its negative exponential.

The most general Aalen model considered here was obtained as a product of one-dimensional processes. But the existing theory covers products of multivariate counting processes as well, and it would be of interest to discuss extensions of these models. This however may turn out to be quite difficult, one reason being that the structure of multivariate Markov counting processes is a great deal more complex than that of the one-dimensional processes treated in Section 4.

ACKNOWLEDGEMENTS

This work was inspired by a seminar held at the Statistical Research Unit in Copenhagen during the autumn of 1980. Among the participants I would like to thank Per Kragh Andersen, Richard Gill and Niels Keiding for stimulating discussions. But especially I thank Søren Johansen for many long conversations on the topics presented here and in Johansen (1981).

REFERENCES

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

Andersen, P.K. & Gill, R.D. (1981). Cox's regression model for counting processes: a large sample study. Research report 81/6. Statistical Research Unit, Danish Medical and Social Science Research Councils.

Bailey, K.R. (1979). The general maximum-likelihood approach to the Cox regression model. Ph.D. dissertation, University of Chicago, Chicago, Illinois.

Cox, D.R. (1972). Regression models and lifetables. J.Roy.Statist.Soc.Ser. B 34, 187-220. (With discussion).

Cox, D.R. (1975). Partial likelihood. Biometrika 62, 269-276.

Feller, W. (1940). On the integro-differential equations of purely discontinuous Markoff processes. Trans.Amer.Math.Soc. 48, 488-515. Errata. Ibid. (1945), 58, 474.

Jacobsen, M. (1972). A characterization of minimal Markov jump processes. Z. Wahrsch.Verw. Gebiete, 23 , 32-46.

Jacobsen, M. (1982). Statistical analysis of counting processes. Lecture notes in statistics, Springer, New York.

Johansen, S. (1981). An extension of Cox's regression model. Preprint 11, Institute of Mathematical Statistics, University of Copenhagen.

Kallenberg, O. (1975). Random measures. Akademie-Verlag, Berlin.

Kiefer, J. & Wolfowitz, J. (1956). Consistency of the maximum-likelihood estimator in the presence of infinitely many incidence parameters. Ann. Math.Statist. 27, 887-906.

Nørlund, N.E. (1954). Vorlesungen über Differenzenrechnung. Chelsea, New York.

Oakes, D. (1981). Survival times: aspects of partial likelihood. Internat. Statist.Rev. 49, 235-264. (With discussion).

Rebolledo, R. (1978). Sur les applications de la théorie des martingales à l'étude statistique d'une famille de processus ponctuels. Proceedings, Journées de statistique des processus stochastique, Grenoble 1977. D. Dacunha-Castelle, B.v. Cutsem (eds). Lecture notes in mathematics 636, pp. 27-70, Springer, New York.

PREPRINTS 1981

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Johansen, Søren: Asymptotic Inference in Random Coefficient Regression Models.

No. 2 Asmussen, Søren: On the Role of a Certain Eigenvalue in Estimating the Growth Rate of a Branching Process.

No. 3 Lauritzen, Steffen.: Time Series Analysis in 1880. A Discussion of Contributions made by T.N. Thiele.

No. 4 Asmussen, Søren: Conditioned Limit Theorems Relating a Random Walk to its Associate, with Applications to Risk Reserve Processes and the GI/G/1 Queue.

No. 5 Johansen, Søren: The Statistical Analysis of a Markov Branching Process.

No. 6 Asmussen, Søren: Time - Dependent Approximations in some Queueing Systems with Imbedded Markov Chains Related to Random Walks.

No. 7 Skovgaard, Ib M.: A Second - Order Investigation of Asymptotic Ancillarity.

No. 8 Rootzén, Holger: The Rate of Extremes of Stationary Normal Sequences.

No. 9 Skovgaard, Ib M.: Large Deviation Approximations for Maximum Likelihood Estimators.

No. 10 Jensen, Ulla Funck: A Stochastic Projection Model with Implications for Multistate Demography and Manpower Analysis.

No. 11 Johansen, Søren: An Extension of Cox's Regression Model.

No. 12 Skovgaard, Ib M.: Edgeworth Expansions in Statistics.

PREPRINTS 1982

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, 2100 COPENHAGEN Ø, DENMARK.

No. 1 Holmgaard, Simon and Yu, Song Yu: Gaussian Markov Random Fields Applied to Image Segmentation.

No. 2 Andersson, Steen A., Brøns, Hans K. and Jensen, Søren Tolver: Distribution of Eigenvalues in Multivariate Statistical Analysis.

No. 3 Tjur, Tue: Variance Component Models in Orthogonal Designs.

No. 4 Jacobsen, Martin: Maximum-Likelihood Estimation in the Multiplicative Intensity Model.