Optimal Estimation in the Non-parametric Multiplicative Intensity Model

P.W. Millar[1]

Technical Report No. 173
October 1988

[1]Research supported by NSF Grant DMS-84-51753

Department of Statistics
University of California
Berkeley, California

1. Introduction.

Let $\{N_t, F_t, t \geq 0\}$ be a (univariate) point process. The intensity $\{\lambda_t, t \geq 0\}$ of N is assumed multiplicative, in the sense that

$$\lambda_t = \alpha_t Z_t. \tag{1.1}$$

Here Z_t is a non-negative adapted process and $\alpha \in A$, an infinite dimensional collection of a non-negative right continuous (non-random) functions on $[0, \infty)$ satisfying $\int_0^t \alpha_s \, ds < \infty$. The parameter $\alpha \in A$ is unknown, and the statistical problem, *roughly speaking*, is to estimate the integral $\int_0^t \alpha_s \, ds$ on some interval, say $0 \leq t \leq 1$. See section 2 for a more precise description; see, e.g., Jacobsen, 1982, for basic facts about this multiplicative intensity model.

The Aalen estimate of $A(\alpha; t) \equiv \int_0^t \alpha_s \, ds$ is the process

$$\hat{A}(t) = \int_0^t (Z_s)^{-1} I\{Z_{s-} > 0\} N(ds), \quad 0 \leq t \leq 1 \tag{1.2}$$

These estimates are attractive because of their asymptotic normality and their easy computability. There is some work (Jacobsen, 1982, p. 148ff and Karr, 1988) to show that they are similar to maximum likelihood estimates. The present paper shows that they are also quite similar to the empirical cdf as it is used in problems involving iid observations.

The MLE in classical parametric problems, and the empirical cdf, share an asymptotic optimality called the local asymptotic minimax (LAM) property. In the parametric case this property roughly amounts to the assertion that, among all possible estimates of the parameter, the MLE has smallest asymptotic variance. The assertion for the empirical cdf is analogous: among all estimates of the underlying cdf, the empirical cdf has the "smallest asymptotic risk". Thus LAM is an efficiency property. In section 2, we prove that, in an appropriate framework, the Aalen estimators are LAM in a sense very close to that of the empirical cdf.

The MLE's in classical parametric problems and the empirical cdf, share another efficiency property, called a convolution theorem. In the MLE case, this asserts essentially that the asymptotic distribution of the MLE is always "less spread out" than the asymptotic distribution of any other regular estimate of the parameter. A similar result, due to Beran 1977, holds for the empirical cdf. In section 3, we prove under suitable conditions that a convolution theorem holds for $\hat{A}$.

The empirical cdf can be used to construct confidence bands for the underlying cdf. One would like to use $\hat{A}$ similarly to construct confidence bands for $A(\alpha; t) \equiv \int^t \alpha_s \, ds$. Such bands could then be used, e.g., for goodness of fit tests. The construction of such bands in the cdf case is eased considerably by the fact that the Kolmogorov-Smirnov statistic is distribution free, a convenience not shared by the present situation. None the less, in section 5 we provide two methods for constructing confidence bands for $A(\alpha; \cdot)$ which have correct asymptotic level. These bands are also shown to have an asymptotic efficiency property; this development utilizes a kind of LAM property for set valued estimates developed in Beran, Millar, 1985.

In a multiplicative intensity model it is often possible, as shown by Jacobsen, 1982, section 5.3, to construct estimators of the product limit type. The development of this paper automatically provides LAM results, a convolution theorem, and optimal confidence band constructions for these estimates as well. These results follow easily from more general results concerning the estimation of $\xi(A(\alpha; \cdot))$ where ξ is a "differentiable" functional. Our development is designed to show the applicability of our results to the problem of optimally estimating $A(\alpha; \cdot)$ when α is constrained — e.g., assumed to be an increasing function. Section 4 gives two applications of our results to the problem of constrained estimation.

This introduction has emphasized the similarities between the problem of estimating a cdf and that of estimating $A(\alpha; \cdot)$. On the other hand, there are important differences other than mathematical complexity. Perhaps the most interesting difference is that, properly formulated (see section 2), the Aalen estimators in general estimate *random* functions, not deterministic ones. Such an estimation problem cannot fit into the Le Cam theory of experiments, (Le Cam, 1988) and, hence optimality results derived under that theory do not typically extend to this more general framework. The method described here (cf section 2) is to make the randomness in the effective parameter disappear asymptotically; such a phenomenon holds in a number of practical examples. On the other hand, this device is far from being generally satisfactory. Indeed, the development of a *general* LAM theory for optimally estimating random parameters is an important problem which will be discussed elsewhere.

The developments of this paper require several results from the theory of Aalen processes, and a good deal of abstract LAM theory for infinite dimensional parameter sets. To shorten the exposition, we shall refer the reader to the appropriate sections of Jacobsen, 1982, for the former, and to sections in Millar, 1983, for the latter. The recent monograph of Karr, 1986, could also be used for background on Aalen processes.

2. LAM property.

Let $\{N_{n,t} F_{n,t}, 0 \le t \le 1\}$, $n = 1,2,\ldots$, be a sequence of (univariate) Aalen processes; the intensity of N_n is then of the form

$$\lambda_{n,t} = \alpha_t Z_{n,t} \tag{2.1}$$

where $\alpha \in A$, and, for each n, $Z_{n,t}$ satisfies the conditions given on pp. 115-116 of Jacobsen. In many applications, N_n is the sum of n iid copies of a given process, in which case $Z_{n,t}$ is then a sum of iid processes. We shall, however, not make the iid assumption.

The Aalen parameter $\alpha \in A$ does not completely specify the distribution of the process $\{N_{n,t}, t \ge 0\}$. Let β be another parameter with values in a normed space. We *assume* that the pair (α, β) determine the distribution of N_n. The necessity for introducing β, as well as an instance of such a β, are apparent in Example 4.2; see also Illustration 5.1. Let

$$P^n_{\alpha\beta} = \text{law of } \{N_{n,t}, 0 \le t \le 1\} \tag{2.2}$$

when the intensity is given by (2.1). Expectation under $P^n_{\alpha\beta}$ will be denoted by $E^n_{\alpha\beta}$.

This section develops a LAM result in the neighborhood of a pre-selected point (α_0, β_0). In this development, the parameter β can be ignored, so *throughout this section and section 3,* we shall for simplicity *write P^n_α for P^n_{α,β_0}*, and E^n_α for E^n_{α,β_0}. In section 4, the role of β becomes crucial, and so the notation (2.2) will resurface there.

Define for $\alpha \in A$

$$\overset{o}{A}_n(\alpha; t) = \int_0^t \alpha_s I\{Z_{ns} > 0\} ds.$$

The estimation problem is then usually defined as that of estimating the random process $\overset{o}{A}_n(\alpha; t)$ on some interval, which we henceforth take to be $[0,1]$. Under the hypotheses given below, it turns out that $\overset{o}{A}_n(\alpha; t)$ is asymptotically equivalent to

$$A_n(\alpha; t) \equiv \int_0^t \alpha_s I\{E^n_\alpha Z_{ns} > 0\} ds, \tag{2.3}$$

a non-random function, and so we shall deal with $A_n(\alpha; \cdot)$ throughout instead of $\overset{o}{A}_n$. Justification for this appears in the proof of theorem 2.1. The reason usually given for estimating $\overset{o}{A}_n$ instead of $A_n(\alpha; \cdot)$ is that it is impossible to make inference about α on the set of time points s where $Z_{ns} = 0$. The Aalen estimator $\hat{A}_n(\cdot)$ is given by

$$\hat{A}_n(t) = \int_0^t (Z_{n,s-})^{-1} I\{Z_{n,s-} > 0\} N_n(ds). \tag{2.4}$$

To formulate the LAM property, note first that $A_n(\alpha;\cdot)$ and $\hat{A}_n(\cdot)$ both have values in the Banach space $L_\infty([0,1])$, the bounded real functions on $[0,1]$ with supremum norm. Denote the norm of L_∞ by $\|\cdot\|$. Let l be a non-negative subconvex function on L_∞, such as $l(x) = \|x\| \wedge a$, $x \in L_\infty$. Let T_n be an estimator of $A_n(\alpha;\cdot)$ available at stage n; it is assumed that T_n is an L_∞-valued random variable. If $\alpha \in A$, then the risk at α, if T_n is our estimate, is

$$E_\alpha^n l\{a_n(T_n - A_n(\alpha))\} \tag{2.5}$$

Here $\{a_n\}$ is a fixed sequence of numbers, $a_n < a_{n+1}$; in many examples, $a_n = n^{1/2}$. For convenience, assume from now on that l is bounded and continous; this assumption is easily removed by familiar arguments.

To formulate the LAM result, fix $\alpha_0 \in A$ and define $D(n,c) = D(n,c,\alpha_0) = \{\alpha \in A: \|A_n(\alpha) - A_n(\alpha_0)\| \le ca_n^{-1}\}$. Let $\mathbf{T}_n$ denote the collection of estimators of $A_n(\alpha)$ available at stage n.

THEOREM 2.1. (LAM) Assume (2.7) - (2.12) below. Then, if $\hat{A}_n(\cdot)$ is the Aalen estimate,

$$\begin{aligned}&\lim_{c\uparrow\infty}\liminf_{n}\inf_{T\in\mathbf{T}_n}\sup_{\alpha\in D(n,c)} E_\alpha^n l\{a_n(T - A_n(\alpha))\}\\ &= \lim_{c\uparrow\infty}\lim_{n\to\infty}\sup_{\alpha\in D(n,c)} E_\alpha^n l\{a_n(\hat{A}_n - A_n(\alpha))\}\end{aligned}$$

The common value of the limit is characterized in proposition 2.1.

Here are the assumptions for theorem 1, formulated for the fixed α_0 above. The first two assumptions are triangular array variants of those in Jacobsen, sec. 5.2 (except we do not assume a product model); these two assumptions ensure the asymptotic normality of $\hat{A}_n$. To formulate them, let α_n denote a sequence in A such that for some c

$$\sup_{0\le t\le 1}\left|\int_0^t \alpha_{n,s}\,ds - A_n(\alpha_0;t)\right| \le ca_n^{-1}. \tag{2.6}$$

Then we assume:

(2.7) there exists a non-decreasing continuous function Φ (depending on α_0) with $\Phi_0 = 0$, such that for each t, $0 \le t \le 1$

$$\int_0^t \alpha_{n,s}\, a_n^2 (Z_{n,s})^{-1} I\{Z_{n,s} > 0\}ds \to \Phi_t$$

in $P_{\alpha_n}^n$ probability, whenever α_n satisfies (2.6).

(2.8) for all $\varepsilon > 0$, all $t \in [0,1]$, whenever α_n satisfies (2.6),

$$\lim_n E^n_{\alpha_n} \int_0^t \alpha_{ns}\, a_n^2 (Z_{ns})^{-1} I\{0 < Z_{n,s} < a_n \varepsilon^{-1}\} = 0$$

Introduce a third assumption:

(2.9) whenever α_n satisfies (2.6)

$$\lim_n a_n \int_0^1 \alpha_{ns} I\{E^n_{\alpha_n} Z_{ns} > 0\} P^n_{\alpha_n}\{Z_{ns} = 0\}\, ds = 0.$$

This assumption permits us to replace $\overset{o}{A}_n(\alpha_{nj})$ by $A_n(\alpha_n;\cdot)$ in the asymptotic arguments, as described earlier in this section.

Here is the fourth assumption:

(2.10) there is a real function q, on the interval $0 \le s \le 1$, such that

(i) $$a_n^2 \int_0^t \alpha_{ns} (Z_{ns})^{-1} I\{Z_{ns} > 0\} ds \to \int_0^t \alpha_{os} (q_s)^{-1} I_s\, ds \equiv \Phi_t$$

and also

(ii) $$a_n^{-2} \int_0^t \alpha_{ns} Z_{n,s} I\{Z_{ns} > 0\}\, ds \to \int_0^t \alpha_{os}\, q_s I_s\, ds \equiv \Psi_s$$

where $I_s = 1$ if $q_s > 0$, $I_s = 0$ if $q_s = 0$.

The convergences above are in $P^n_{\alpha_n}$ probability. Assumption (2.10i) merely narrows (2.7) a bit. Part (ii) guarantees convergence of (in the sense of Le Cam) certain statistical experiments, and the 'symmetric' nature of the two limits allows one to relate this convergence to that of the Aalen estimator. In case Z_n is the sum of iid copies of Z and $a_n^2 = n$, one gets $q_s = EZ_s$ by the law of large numbers, and so (2.10) holds under modest integrability conditions.

Our fifth assumption is:

(2.11) whenever α_n satisfies (2.6)

$$\lim_n \int_0^1 [\, I\{E^n_{\alpha_n} Z_{ns} > 0\} - I_s\,]^2 \alpha_{os} (q_s)^{-1} I_s\, ds = 0.$$

This assumption is technical: it allows locally the replacement of $A^n_t(\alpha) = \int_0^t \alpha_s I\{E^n_\alpha Z_{ns} > 0\}\, ds$ by $\int_0^t \alpha_s I_s$, after certain preliminary reductions.

The final assumption is that α_0 be a *radial point* of the parameter set A. To describe this concept, let H be the Hilbert space of real functions on $[0,1]$ with the L^2 norm given by the measure $\alpha_0(s) q_s^{-1} I_s\, ds$, so if $h \in H$, $|h|_H^2 = \int h(s)^2 \alpha_0(s)\, [q(s)]^{-1}$

$I(s)\,ds$. In particular $h(s) = h(s)\,I(s)$, as elements of H. Then **A** is radial at $\alpha_0 \in$ **A** if, for each h in a dense set $H_0 \subset H$, the function $\alpha_0(s) + \alpha_0(s)\,a_n^{-1}\,h(s)\,q(s)^{-1}\,I(s)$ belongs to **A** for all sufficiently large n. This property asserts a sense in which α_0 is not a "boundary point" of **A**; it also ensures the "infinite dimensionality" of **A**. Thus, the final assumption is

(2.12) α_0 is radial in **A**.

Remark. Assumption (2.6) can be weakened. An LAM result like Theorem (2.1) can be proved if, in (2.6), only α_n of the form $\alpha_n = \alpha_0 + \alpha_0 hqIa_n^{-1}$ are used.

Having given the basic assumptions, we may now characterize the LAM lower bound in theorem 1.

Proposition 2.1. Under assumptions (2.7) - (2.12), the common value in theorem 1 is

$$El\,(X)$$

where $X = \{X_t, 0 \le t \le 1\}$, $X_t = W \circ \Phi_t$, and $W = \{W_s, s \ge 0\}$ is standard Brownian motion on the line; Φ_t was given in (2.7).

This proposition is immediate from the following

Proof of theorem 1: We first check that the second expression in theorem 1 is equal to $El(X)$, defined in Proposition 2.1. Let $\alpha_n \in$ **A** satisfy (2.6). By (2.7), (2.8) and Rebolledo's CLT (Rebolledo, 1978; see also Jacobsen, p. 163) we find that

$$a_n\,[\,\hat{A}_n(t) - \overset{\circ}{A}_n(\alpha_n;\,t)\,], \quad 0 \le t \le 1$$

converges in $L_\infty[0,1]$ to $\{X_s, 0 \le s \le 1\}$. Next, note that

$$a_n\,\|\,\overset{\circ}{A}_n(\alpha_n;\cdot) - A_n(\alpha_n;\cdot)\,\| \;\to\; 0$$

since this last display equals

$$a_n \int_0^1 \alpha_n(s)\,I\,\{Z_{n,s} = 0\}\,I\{E^n_{\alpha_n} Z_{ns} > 0\}\,ds$$

which goes to zero by (2.9). Thus for every α_n satisfying (2.6)

$$a_n\,[\,\hat{A}_n(\cdot) - A_n(\alpha_n;\cdot)\,] \;\Rightarrow\; X.$$

Since α_n could have been chosen to achieve the supremum over $D(n,c)$, it follows that the second expression in theorem 2.1 is $El(X)$.

To finish the proof, it suffices to show that the first expression in theorem 1 exceeds $El(X)$. Let H be the Hilbert space of real functions on $[0,1]$ introduced before (2.12). Define a mapping $\tau : H \to C[0,1]$ by $(\tau h)(t) = \int_0^t h(s)\alpha_0(s)q^{-1}(s)I_s\,ds$. If τ^* is the adjoint of τ, then integration by parts shows that if $m \in C^*[0,1]$, dual of $C[0,1]$, then $(\tau^* m)(t) = m\{[t,1]\}$; thus, $|\tau^* m|_H^2$

$$= \int_0^1 m[s,1]^2 \alpha_s q_s^{-1} I_s ds$$

$$= \iiint I_{[s,1]}(u) I_{[s,1]}(v) \alpha_{os} q_s^{-1} I_s\, ds m(du) m(dv)$$

$$= \iint \int_0^{u \wedge v} \alpha_{os} q_s^{-1} I_s\, ds\, m(du)\, m(dv)$$

$$= \iint \Phi_{u\wedge v}\, m(du)\, m(dv)$$

by (2.10i). Thus (τ, H, B), $B = \overline{\tau H}$ (closure in $C[0,1]$ of the image of H under τ) is an abstract Wiener space, and the standard normal Q_0^∞ on B is the law of $X = \{X_t, 0 \le t \le 1\}$; see Millar, 1983, Chs V, VI.

Let $\{Q_h^\infty, h \in H\}$ denote the Gaussian shift experiment for (τ, H, B). Then, for example, under Q_0^∞,

$$\log (dQ_n^\infty / dQ_0^\infty)(x) = \int_0^1 h(s)\,dx(s) - 1/2|h|_H^2, \quad x \in B.$$

Next, consider the experiment $\{Q_h^n, h \in H\}$, defined as follows. Q_h^n is the distribution of $\{N_{nt}, 0 \le t \le 1\}$ under $P_{\alpha_n}^n$, when α_n has the form

$$\alpha_{ns} = \alpha_{0s}[1 + h_s q_s^{-1} I_s a_n^{-1}], \quad 0 \le s \le 1. \tag{2.13}$$

We shall argue that the experiments $\{Q_h^n, h \in H\}$ converge, in the sense of Le Cam, to $\{Q_h^\infty, h \in H\}$; see Millar, 1983, ChII for an exposition of this notion of convergence that is easily applicable to the present situation; a deeper development is Le Cam, 1986.

By (2.10ii) and Rebolledo's CLT,

$$a_n^{-1}\{N_{n,t} - \int_0^t \alpha_{0s} Z_{n,s}\,ds;\ 0 \le t \le 1\} \Rightarrow \{Y_t, 0 \le t \le 1\} \tag{2.13}$$

where $Y_t = W \circ \psi_t$, and Ψ was defined in (2.10ii). Let $\alpha_n = \alpha_0 + \alpha_0 h I q^{-1} a_n^{-1} \equiv \alpha_0 + \alpha_0 \alpha_1 a_n^{-1}$. Then using the form of the likelihood ratios for Aalen models (cf Jacobsen, ChIV), we find $\log dQ_h^n / dQ_0^n = \log dP_{\alpha_n}^n / dP_{\alpha_0}^n$

$$= -a_n^{-1} \int_0^1 \alpha_{0s} \alpha_{1s} Z_{ns}\,ds + \int_0^1 \log[1 + a_n^{-1}\alpha_{1s}]\,dN_{n,s}$$

$$= a_n^{-1}\int_0^1 \alpha_{1s}\,[\,dN_{n,s} - \alpha_{0s} Z_{ns}\,ds\,] - 1/2\int_0^1 a_n^{-2}(\alpha_{1s})^2\,dN_{n,s}.$$

Because of (2.13), this converges to

$$\int_0^1 \alpha_{1s}\,dY_s \;-\; 1/2\int_0^1 (\alpha_{1s})^2\,d\Psi_s$$

$$= \int_0^1 h(s)\,[\,q(s)\,]^{-1} I_s dY_s \;-\; 1/2\int_0^1 [\,h(s)/q(s)\,]^2\,d\Psi_s$$

$$= \int_0^1 h(s)\,dX_s \;-\; 1/2\int_0^1 [\,h(s)\,]^2\,d\Phi_s,$$

using, e.g. Doob, 1953, p.   .

Thus the log likelihoods of $\{Q_h^n, h \in H\}$ converge to those of $\{Q_h^\infty, h \in H\}$. Since the likelihoods are asymptotically quadratic in the parameter h, this implies that the experiments converge in the sense of Le Cam.

The form of the LAM lower bound can now be deduced from the Hajèk-Le Cam theorem (Le Cam, 1972; Millar, 1983, chII). Indeed, since α_0 is radial

$$D(n,c) \supset D_0(n,c)$$

where $D_0(n,c)$ consists of all α of the form (2.13) having $\int_0^1 \alpha_{0s}\,|h_s|q_s^{-1} I_s \le c,\ h \in H_0$.

Moreover, for α of the form (2.13), hypothesis (2.11) implies that

$$A_n(\alpha;\cdot) \;=\; A_n(\alpha_0;\cdot) + a_n^{-1}\int \alpha_{0s} h_s q_s^{-1} I_s I\{E_{\alpha_0}^n Z_{ns} > 0\}\,ds$$

$$= A_n(\alpha_0;\cdot) + a_n^{-1}\tau h + o(a_n^{-1}).$$

Therefore

$$\inf_{T\in T_n}\ \sup_{\alpha\in D(n,c)} \int l\,[\,a_n(T - A_n(\alpha;\cdot))\,]\,dP_\alpha^n$$

$$\ge \inf_{T\in T_n}\ \sup_{\alpha\in D_0(n,c)} \int l\,[\,a_n(T - A_n(\alpha;\cdot))\,]\,dP_\alpha^n$$

$$\ge \inf_{T\in T_n}\ \sup_{h:|\tau h|\le c} \int l\,[\,T - \tau h\,]\,dQ_h^n \;+\; o(1).$$

By the asymptotic minimax theorem and the minimax value for a Gaussian experiment (e.g., Millar, 1983, chVI, p. 133), this last expression above is minimized in the limit (as $n \to \infty$ and then $c \uparrow \infty$) by $El(X)$. (A completely detailed proof would use the argument on p. 147 of Millar, 1983, to justify interchanging $\lim_c$ and $\lim_n$.) This completes the proof.

3. Functionals of the integrated Aalen parameter.

Let ξ be a mapping defined on $L_\infty[0,1]$ with values in some Banach space B_2. The task is to estimate $\xi(A_n(\alpha;\cdot))$ (or $\xi(\mathring{A}_n(\alpha;\cdot))$). Under regularity conditions on ξ we show first that the natural estimator $\xi(\hat{A}_n)$, where $\hat{A}_n$ is the Aalen estimate of section 2, is LAM (Proposition 3.1) and efficient in the sense of a convolution theorem (Proposition 3.2). These results are then applied to show that the "product limit" estimators associated with multiplicative intensity models are also LAM and convolution-efficient. The next section presents some illustrations of estimation problems when the Aalen parameter is subject to constraints.

To give the required smoothness property for ξ, fix α_0, and bring in the Hilbert space H of section 1. Again assume that α_0 is radial, and let H_0 be the subset of H given in the definition (cf., (2.12)). Define ξ to be H_0-differentiable at α_0 if, for each $h \in H_0$

(3.1)
$$a_n[\xi(A_n(\alpha_{nh})) - \xi(A_n(\alpha_0))] = \xi' \circ \tau h + o(1)$$

where ξ' is a continuous linear map of $L_\infty[0,1]$ to B_2 (depending on α_0 only), and where α_{nh} is an Aalen parameter of the form

(3.2)
$$\alpha_0 + \alpha_0 h q^{-1} I a_n^{-1}.$$

This differentiability condition is much weaker than Frêchet differentiability; however, the latter will suffice for the examples discussed in the next section. Let $\mathbf{T}_n$ denote all estimators of $\xi(A_n(\alpha))$ available at time n, and let l be bounded and subconvex in B_2.

Proposition 3.1: LAM. Assume the hypotheses of theorem 2.1, and that ξ satisfies the differentiability hypothesis (3.1). Assume also that the range of ξ' is dense in B_2. Then

$$\lim_{c\uparrow\infty} \lim_{n\to\infty} \inf_{T\in\mathbf{T}_n} \sup_{\alpha\in D(n,c)} \int l\,[a_n(T - \xi(A_n(\alpha)))]\,dP_\alpha^n \geq E\,l\,(\xi' \circ X)$$

where D(n,c), X are as in theorem 2.1. If

(3.3)
$$a_n[\xi(\hat{A}_n) - \xi(A_n(\alpha_n))] \Rightarrow \xi' \circ X,$$

under $P_{\alpha_n}^n$ whenever $\{\alpha_n\}$ is an arbitrary sequence such that $\alpha_n \in D(n,c)$, then $\xi(\hat{A}_n)$ is LAM in the sense that

$$\lim_{c\uparrow\infty}\lim_{h\to\infty}\sup_{\alpha\in D(n;c)}\int l\,[a_n(\xi(\hat{A}_n)-\xi(A_n(\alpha)))]\,dP^n_\alpha \;=\; E\,l\,(\xi'\circ X).$$

The proof of proposition 3.1 appears in section 6.

Remarks 3.1. (a) The condition (3.3) is obviously satisfied if ξ is Frêchet differentiable, or even compact differentiable (cf., Reeds, 1976). Condition 3.3 does not follow from the condition (3.1). As is familiar from experience, and as Reeds pointed out at some length, studying the differentiability properties of ξ will often not be the best way to establish (3.3).

(b) The requirement that the range of ξ' be dense in B_2 can be weakened. One can assume this range to be a complemented subspace of B_2, where the associated projection π onto this range has norm ≤ 1. In this case assume in addition that $l\,(x) = g(\|x\|)$ where g is an increasing function, and $\|\;\|$ is the norm of B_2. Then the LAM lower bound becomes $E\,l\,[\pi\circ\xi'\circ X]$, and the LAM estimate, under regularity assumptions on ξ, becomes $\pi\circ\xi(\hat{A}_n)$.

Let us turn next to a convolution theorem. Again fix the radial point α_0 and bring in H_0. Define an estimator T_n of $\xi(A_n(\alpha))$ to be *H_0-regular* if there is a probability G_0 on B_2 such that for every $h \in H_0$:

$$a_n[T_n - \xi(A_n(\alpha_{nh}))] \;\Rightarrow\; G_0, \tag{3.4}$$

convergence in distribution under Q^n_h. Here α_{nh} is defined by (3.2). Let

$$\nu_0 \;\equiv\; \text{distribution of } \xi'\circ X \tag{3.5}$$

where X is defined in proposition (2.1).

Proposition 3.2: convolution. Assume the hypotheses of theorem 2.1, and that ξ satisfies the differentiability hypotheses (3.1). Assume also that the range of ξ' is dense in B_2. Let T_n be an H_0 a regular estimator with limit distribution G_0, as given by (3.4). Then there exists a probability μ on B_2 such that

$$G_0 \;=\; \mu \,*\, \nu_0.$$

If, in addition, (3.3) holds, then $\xi(\hat{A}_n)$ is an H_0 regular estimate, and is efficient in the sense that its μ is unit mass at $0 \in B_2$.

The proof will be given in section 6.

Remarks 3.2. (a) Under the assumptions of Proposition 3.2, $\hat{A}_n$ is a regular estimate of $A(\alpha;\cdot)$, and so is efficient. Thus, the "convolution-efficiency" of $\xi(\hat{A}_n)$ hinges on properties of ξ only — see Remarks 3.1, (a).

(b) If ξ' is not one-to-one, it is possible to get by with an even weaker notion of regularity. Let H_{00} be a subspace of H_0, and define H_{00}-regularity analogously to (3.4). Let ξ be differentiable with respect to H_{00} (i.e., replace H_0 by H_{00} in definition (3.1)). Let $\eta^{\perp}$ be the null space of the mapping $\xi' \circ \tau$, and *assume* $\eta^{\perp} \supset H_{00}$. Then the conclusion of proposition 3.2 continues to hold.

The foregoing results provide a simple way to establish the asymptotic optimality of the so-called product limit estimators. To see this, take $\xi : L_\infty[0,1] \to L_\infty[0,1]$ as follows: if $g \in L_\infty[0,1]$,

(3.6) $$\xi(g)(t) = \exp\{-g(t)\}, \quad t \in [0,1].$$

Then ξ is differentiable (in the sense of Frêchet), with derivative at $g_0 \in L_\infty$ given by

(3.7) $$\begin{aligned} \xi'(g)(t) &= \xi'(g_0; g)(t) \\ &= -g(t)\exp\{-g_0(t)\} \quad t \in [0,1]. \end{aligned}$$

The statistical problem addressed here is the estimation of

$$\xi(A_n(\alpha; \cdot))$$

as an element of $L_\infty([0,1])$. By propositions (3.1), (3.2), an optimal estimate is $\xi(\hat{A}_n)$. On the other hand, a currently popular estimator, based on an extensive history dating at least to 1957, is the product limit estimator defined by

(3.8) $$\hat{G}_n(t) = \prod_{s \le t} [1 - (\Delta N_{ns}/Z_{ns}) I\{Z_{ns} > 0\}].$$

Thanks to work of Jacobsen (1982, section 5.3), it is easy to see in our development that these estimators are, under suitable conditions, also optimal.

To set this up, assume the hypotheses of Theorem 2.1, with α_0 fixed. Define

(3.9) $$G(\alpha_0; t) = \exp\{-\int_0^t \alpha_{0s} I_s \, ds\}$$

and let $\{Y_s, 0 \le s \le 1\}$ be the process

(3.10) $$Y_s = X_s G(\alpha_0; s)$$

where $\{X_s\}$ was defined in section 2. Let now $\mathbf{T}_n$ denote all estimators available at time n for the parameter $\xi(A_n(\alpha; \cdot))$.

Proposition 3.3. Assume the hypotheses of theorem 2.1. Assume in addition the hypotheses of Jacobsen's corollary 5.3.9, in triangular array form as exemplified by (2.10). Assume that Jacobsen's τ_n^* (Jacobsen, p. 181) satisfies $\lim_{n\to\infty} P_{\alpha_n}^n\{\tau_n^* \ge 1\} = 1$. Then $\hat{G}_n$ is *a LAM estimate* on $L_\infty[0,1]$:

$$
\begin{aligned}
&\lim_{c\uparrow\infty}\lim_{n\to\infty}\inf_{T\in T_n}\sup_{\alpha\in D(n,c)}\int l\,[\,a_n\{T_n-\xi(A_n(\alpha))\}\,]\,dP_\alpha^n \\
(3.11)\qquad &= \lim_{c\uparrow\infty}\lim_{n\to\infty}\sup_{\alpha\in D(n,c)}\int l\,[\,a_n\{\hat{G}_n-\xi(A_n(\alpha))\}\,]\,dP_\alpha^n \\
&= El(Y).
\end{aligned}
$$

In addition, a convolution theorem holds: if T^n is any H_0 regular estimate of $\xi(A_n(\alpha))$ (cf. (3.4) above), then

$$
(3.12)\qquad \sqrt{n}\,[\,T_n-\xi(A_n(\alpha_n)\,] \Rightarrow Y' + Y
$$

under $P^n_{\alpha_n}$; here Y' is an $L_\infty[\,0,1\,]$ random variable independent of Y.

Proof of Proposition 3.3. Since in the present case ξ is Frêchet differentiable, we may apply propositions 3.1, 3.2 and the differentiability of ξ to deduce that the LAM lower bound is $El\,[\,\xi'(X)\,]$. The form of ξ' ensures that $\xi'(X)(s) = Y_s$, as defined in (3.10). On the other hand, Jacobsen's result 5.3.9 (when put in triangular array form) ensures that the product limit estimator $\hat{G}_n(s)$ has the limit Y_s (in distribution on $L_\infty[\,0,1\,]$) and this completes the proof.

Remark 3.1. The foregoing results imply that, from the point of view of *asymptotic* optimality, there is no reason to consider the P–L estimators superior to the obvious transformation of the Aalen estimators. The Aalen estimators have in addition, a rather simpler asymptotic theory and they generalized easily to higher dimensions. On the other hand, there appear to be no systematic studies comparing the small sample behavior of these two competing estimates. The P–L estimate (like the empirical cdf) has a characterization in terms of MLE concepts; according to Karr (1988), however, the Aalen estimator also has an MLE interpretation (at least *asymptotically*). In any case, the ML property is not generally regarded as an ''optimality'' property, from the point of view of standard decision theory.

Let us now turn to some examples of the use of this development.

4. Aalen parameters with constraints.

This section addresses the question whether the Aalen estimate of $A_n(\alpha;\cdot)$ loses efficiency when the Aalen parameter α is subject to ''constraints''. This, of course, is a complicated question; we address it here in the context of two substantive examples, where the kind of ''constraint'' often has a transparent physical interpretation. Our goal here is to indicate the scope of the abstract results of previous sections. In these illustrations $a_n = n^{1/2}$.

Example 4.1: hazard functions under a monotone constraint.

Let $x_1, \ldots, x_n$ be i.i.d. non-negative random variables with c.d.f. F. Assume that F has density f, and that $F(1) < 1$. Let $N_{nt} = \sum_{i=1}^{n} I\{x_i \le t\}$; henceforth we delete the subscript n. As is well-known, this point process may be formulated as a multiplicative intensity model with $\alpha(t) = f(t)/\bar{F}(t)$ and $Z_{nt} = n - N_{t-}$. Here $\bar{F}(t) = 1 - F(t)$.

For unknown F let us consider the problem of estimating on $[0,1]$ the *integral* of the *hazard function* $\alpha \equiv f/\bar{F}$, under the assumption that α is known to be an *increasing* function. (In the theory of reliability, this constraint is called "increasing failure rate" and is usually denoted by "IFR".). Thus, in the notations of section 2, $A = \{\text{increasing hazard functions}\}$. Although the elements of A are subject to a constraint, we shall permit the use of any estimator of $A_n(\alpha)$ at all even if it does not precisely have the form $\int \hat{\alpha}$, $\hat{\alpha} \in A$. The Aalen-Nelson estimate for $\overset{o}{A}_n(\alpha, t) \equiv \int_0^t \alpha(s) I\{N_{s-} < n\}\, ds$ is

$$\begin{aligned} \hat{A}_n(t) &= \int_0^t (n - N_{s-})^{-1} I\{N_{s-} < n\}\, N(ds) \\ &= \sum_{k=1}^{N_t \wedge n} [\, n - (k-1)\,]^{-1}. \end{aligned}$$

We shall argue first that $\hat{A}_n(\alpha, t)$ is a LAM estimate of $A_n(\alpha, t)$ at $\alpha_0 \in A$, whenever α_0 is "radial" in A.

To set this up, let us define for $\alpha \in A$, the corresponding cdf F_α by $1 - F_\alpha(t) = \exp\{-\int_0^t \alpha_s\, ds\}$. Fix $\alpha_0 \in A$, and abbreviate F_{α_0} by F_0. By the law of large numbers $\frac{Z_n}{n} \to \bar{F}_0$; since $F_0(1) < 1$ by assumption, one then identifies the function q_s of section 2 as $q_s = \bar{F}_0(s)$, and I_s there is given by $I_s \equiv 1$ $(0 \le s \le 1)$. The hazard function α_0 will therefore be radial if $n^{-1/2}\alpha_0(s)h(s)/\bar{F}_0(s) + \alpha_0(s)$ is an increasing function for a dense set of h in H (cf (2.12)), at least for all sufficiently large n. (The size of n in the previous statement can depend on h). The hazard function $\alpha_0 + n^{-1/2}\alpha_0 h/\bar{F}$ will be increasing if its derivative is positive; this leads to the condition that $n^{-1/2}(h'\bar{F}_0 + hf)\alpha_0/\bar{F}_0 + [\,(n^{-1/2}h/\bar{F}_0) + 1\,]\alpha_0'$ be non-negative. Thus if α_0' is positive, and bounded away from 0 on $[0,1]$, it is easy to see that the hazard function $\alpha_0 + n^{-1/2}\alpha_0 h/\bar{F}_0$ will be increasing for large n, if h and h′ are bounded on $[0,1]$. Such h give a dense subset of H, so α_0 is radial in A provided only that $\inf_{0 \le s \le 1} \alpha_0'(s) > 0$. The triangular array assumptions of section 2 can be checked in the present situation by using familiar triangular array results for empirical processes (cf Beran and Millar 1986, section 4, for example). The mapping Φ, which determines the asymptotic distribution is $\Phi_t \equiv F_0(t)/\bar{F}_0(t)$ (cf (2.7). Since the other assumptions in section 2 are readily verified, we find by theorem 2.1 that the Aalen-Nelson estimator is LAM at each hazard function α_0 in the collection of increasing hazard functions, having derivative bounded away from 0 on compact intervals: i.e. at every α_0 that is strictly increasing and differentiable.

Similar LAM results can be established for other collection A of hazard functions. For example, one may take A to consist of all hazard functions which are decreasing (DFR in reliability terminology) or A could be the collection of all convex hazard functions, and so forth.

Next, consider the functional $\xi: A \to C(R')$ given by $\xi(\alpha)(t) = 1 - \exp\{-\int_0^t \alpha(s)\,ds\}$. If A is regarded as a subset of $L_\infty(R')$, then each $\alpha \in A$ yields an integrated $\overset{\bullet}{\int}\alpha$ which is in $C(R')$. Therefore we may regard ξ as a map of $C(R') \to C(R')$, using the recipe $\xi(c)(t) = 1 - \exp\{-c(t)\}$. If $c_0(t) \equiv \int_0^t \alpha_0(s)\,ds$, then ξ is differentiable at c_0, with derivative $\xi'(c) \equiv \xi'(c_0; c)$ given by (cf., (3.7)):

$$\xi'(c_0; c)(t) = \bar{F}_{\alpha_0}(t)c(t).$$

Of course $\xi(\overset{\bullet}{\int}\alpha) = F_\alpha$. We therefore may estimate F_α, for unknown $\alpha \in A$, where $A = \{\text{increasing hazard functions}\}$, by using $\xi(\hat{A}_n)$, where $\hat{A}_n$ is the Aalen-Nelson estimator. Proposition 3.1 then shows that this estimator is a LAM estimate of $F_0 \equiv F_{\alpha_0}$ at any α_0 radial for A, and that the limit distribution is $\xi'(X)$ where $X = B \circ \Phi$, $\Phi(t) = F_0(t)/\bar{F}_0(t)$. The form of ξ' then guarantees that $\xi'(X)$ is the usual Brownian Bridge: the mean 0 gaussian process with covariance $F_0(s \wedge t) - F_0(s)F_0(t)$. Let $\hat{F}_n$ be the empirical measure of $x_1, \ldots, x_n$; then $\sqrt{n}(\hat{F}_n - F_0)$ has the same asymptotic behavior as $n^{1/2}[\xi(\hat{A}_n) - \xi(A(\alpha_0))]$ and so $\hat{F}_n$ is a LAM estimate of F_{α_0}, $\alpha_0 \in A$ also. For the case $A = \{\text{increasing hazard functions}\}$ this latter result was established, in slightly greater generality, in Millar, 1979.

Example 4.2: Censored data, with constraints. Let $x_1, \ldots, x_n$ be non-negative iid random variables with common cdf F, and let $y_1, \ldots, y_n$ be iid non-negative random variables, independent of $\{x_i, 1 \le i \le n\}$, and having common cdf G. Let $m_j = \min\{x_j, y_j\}$ $\delta_j = I\{x_j \le y_j\}$ and set $N_t = \sum_j I\{m_j \le t, \delta_j = 1\}$. Assume F has a density f, and that F, G have common support that strictly includes the interval $[0,1]$. The foregoing is well known to be an Aalen model with $\alpha(t) = f(t)/\overline{F}(t)$, $Z_{nt} = \sum_{i=1}^{n} I\{m_i \geq t\}$. Since $Z_{nt}/n \to \overline{F}(t) \wedge \overline{G}(t)$, the special functions q, and I_s of section 2 can be identified as $I_s = 1$, $0 \leq s \leq 1$ (since $F(1) \wedge G(1) < 1$) and $q_s = \overline{F}(s) \wedge \overline{G}(s)$.

As in Example 4.1, let us consider the problem of estimating the indefinite integral of the hazard functions α, under the constraint that α be increasing. We shall again argue that the Aalen estimate of the integrated hazard function is LAM at any $\alpha_0 \in A = \{$increasing hazard functions$\}$ which radial for A.

To set this up, fix $\alpha_0 \equiv f_0/\overline{F}_0$ and G_{0s}. For α_0 to be radial in this case, the hazard $\alpha_0 + n^{-1/2}\alpha_0 h/\overline{F}_0 \wedge \overline{G}_0$ should have a positive derivative for a set of h that are dense in the relevant Hilbert space H of section 2. Considerations exactly as in example 4.1 show that α_0 will be radial *in direction h* if h, h′ are bounded, and if $\inf_{0 \leq s \leq 1} \alpha_0'(s) > 0$. Thus, α_0 is radial in A if only $\inf_{0 \leq s \leq 1} \alpha_0'(s) > 0$. Theorem 2.1 then implies that the Aalen estimate of the integrated hazard function is LAM at any $\alpha_0 \in A$ that is radial for A, as claimed.

Proposition 3.3 then shows that the associated product limit estimator for this example is also LAM, even under the constraint that the hazard function be increasing. Here this P–L estimator is, of course, more famously known as the Kaplan-Meier estimate. For this estimator, Wellner (1983) has proved its LAM character (by a different method, and with no constraints on the hazard function); thus the present example extends slightly the work J. Wellner.

5. Confidence Bands.

Previous sections have discussed optimal estimation of $A_n(\alpha;\cdot)$, the integrated Aalen parameter. This section considers construction of optimal confidence bands for $A_n(\alpha;\cdot)$ (on $[0,1]$, say), and for $\exp\{-A_n(\alpha;\cdot)\}$. The main development in this section, carried out in subsections (5.A), (5.B)) constructs these confidence sets using a bootstrap method. Asymptotic "plug-in" methods are briefly described in section(5.C), which in addition contains other complements to the basic development. Proofs are given in section 6.

(5.A). Description of the confidence bands.

To describe the bands based on the bootstrap method, recall from section 2 that $P^n_{\alpha,\beta}$ is the distribution of the basic point process $\{N_{nt}, 0 \leq t \leq 1\}$ which has Aalen parameter α, and nuisance parameter β. Fix η, $0 < \eta < 1$, and suppose the desired confidence level is $1 - \eta$. Let $\hat{A}_n$ be the Aalen estimate of $A_n(\alpha;\cdot)$ and let $\hat{\beta}_n$ be an estimate of β.

Illustration 5.1. To illustrate such $\hat{\beta}_n$, consider the simple censored data model of section 4 (Example 4.2.) Here $\alpha = f/\bar{F}$, and the Aalen estimate is described there. The nuisance parameter β can be taken to be the cdf of the unknown censoring distribution G. Because of the inherent symmetries in this model, the estimate $\hat{\beta}_n$ of β could be taken as the Kaplan-Meier estimate of the censoring distribution G. Obviously, there are other possibilities, as is clear from section 3.4.

Define

$$\hat{C}_{1n} = \{f \in L_\infty[0,1] : n^{1/2}\|f - \hat{A}_n\| \le \hat{r}_{1n}\} \tag{5.1}$$

where the norm is that of $L_\infty[0,1]$ and

$$\hat{r}_{1n} = t_n(\hat{A}_n;\, \hat{\beta}_n;\, \eta) \tag{5.2}$$

and $t_n(A_n(\alpha;\cdot);\, \beta;\, \eta)$ is given by

$$P^n_{\alpha,\beta}\{n^{1/2}\|\hat{A}_n - A(\alpha;\cdot)\| \le t(A_n(\alpha;\cdot);\, \beta;\, \eta)\} \doteq 1 - \eta. \tag{5.3}$$

The random set $\hat{C}_{1n}$ gives, under conditions given in subsection (5.B), an asymptotically optimal, $1 - \eta$ level, confidence band for $A_n(\alpha;\cdot)$ on $[0,1]$.

To build a confidence band for $\exp\{-A(\alpha;\cdot)\}$, proposition 3.3 suggests a construction similar to that of (5.1), but based on product limit estimators. Indeed, if $\hat{G}_n$ is the product limit estimator given in (3.8) set

$$\hat{C}_{2n} = \{f \in L_\infty[0,1] : n^{1/2}\|f - \hat{G}_n\| \le \hat{r}_{2n}\} \tag{5.4}$$

where

$$\hat{r}_{2n} = t_{2n}(\hat{A}_n;\, \hat{\beta}_n;\, \eta) \tag{5.5}$$

and $t_{2n}(A_n(\alpha;\cdot);\, \beta;\, \eta)$ is given by

$$P^n_{\alpha,\beta}\{n^{1/2}\|\hat{G}_n - \exp\{-A(\alpha;\cdot)\}\| \le t_{2n}(A_n(\alpha;\cdot);\, \beta;\, \eta)\} \doteq 1 - \eta. \tag{5.6}$$

The confidence band $\hat{C}_{2n}$ will be asymptotically optimal with the correct (asymptotic) coverage probability.

Application 5.1. The optimal confidence band $\hat{C}_{1n}$ can be used to assess assumptions about the underlying statistical model. For example, to assess the idea that the Aalen parameter be constant, one might check that the band $\hat{C}_{1n}$ contains at least one straight line emanating from 0, and having non-negative slope. If so, then this could be regarded as evidence in favor of the null hypothesis of constant "failure rate" — in the sense that the trustworthy *set estimate* $\hat{C}_{1n}$ contains at least one member of the null hypothesis. The reasoning here is not that of a standard goodness of fit test. A more interesting possibility, is to assess the hypothesis that the Aalen parameter is increasing on $[0,1]$ (say). In examples 4.1, 4.2 this amounts to seeing if "IFR" is a viable possibility. If the Aalen parameter α is increasing then the integrated Aalen parameter would be non-decreasing convex on the intervals $\{s: Z_{ns} > 0\}$ and flat on the intervals $\{s: Z_{ns} = 0\}$. If the confidence band $\hat{C}_{1n}$ contains at least one such piece-wise convex, piece-wise flat increasing function emanating from 0, then the hypothesis that α be increasing would be supported — in the sense given before: the optimal set estimate contains at least one element of the null hypothesis. If n is large, then in our examples, $I\{Z_{ns} > 0\} = 1 \;\; 0 \le s \le 1$, and so one need check here only whether $\hat{C}_{1n}$ contains at least one convex increasing function starting at 0. A closely related, but more traditional method of testing such null hypotheses could be based on minimum distance methods centred at the Aalen estimate. Such methods will be discussed elsewhere.

(5.B). Optimality of the confidence bands.

To describe the optimal nature of the confidence bands $\hat{C}_{1\eta}$ let $C(z,r)$, $z \in L_\infty[0,1]$, $r > 0$ denote the band $\{y \in L_\infty[0,1]: \|y - z\| \le r\}$. Then $C(z,r)$ is a ball in $L_\infty[0,1]$ with centre Z and radius r. To set up a formal decision theoretic framework, let **D** be the collection of all such balls. Then **D** is the decision space. A nonrandomized procedure ($\equiv$ conf. band) is then $C(\hat{z}_n \hat{r}_n)$ where $\hat{z}_n$, $\hat{r}_n$ are functions of the observed data. We restrict attention to those confidence bands $C(\hat{z}_n, \hat{r}_n)$ that have the proper coverage probability:

$$P^n_{\alpha,\beta}\{C(\hat{z}_n, \hat{r}_n) \ni A_n(\alpha;\cdot)\} \;\ge\; 1 - \eta. \tag{5.7}$$

Denote by $\mathbf{D}_{\eta,n}$ the collection of all procedures $C(\hat{z}_n, \hat{r}_n)$ that satisfy (5.7). A confidence band $C(\hat{z}_n, \hat{r}_n)$ will be reasonable if it belongs to $\mathbf{D}_{\eta,n}$, at least approximately, and also is not grossly off centre or excessively wide. To formulate such a condition introduce a loss function l_n, at time n, by:

$$l_n\{C(\hat{z}_n, \hat{r}_n), A_n(\alpha;\cdot)\} \equiv g[n^{1/2} \sup_{y \in C(\hat{z}_n, \hat{r}_n)} \|y - A_n(\alpha;\cdot)\|], \tag{5.8}$$

where g is an increasing function on $[0,\infty)$ which will be assumed bounded and continuous for convenience. The LAM result for confidence bands of $A_n(\alpha;\cdot)$ may now be formulated. Fix α_0, β_0. Let $D^*(n,c) = \{(\alpha,\beta): \;\; \|\alpha - \alpha_0\| \le c/\sqrt{n}$, $\|\beta - \beta_0\| \le c/\sqrt{n}\}$

Proposition 5.1. (LAM lower bound). Assume the hypotheses of theorem 2.1, with $a_n = n^{1/2}$. Then $\lim_{c\uparrow\infty}\lim_{n\to\infty}\inf_{c(\hat{z}_n,\hat{r}_n)\in D_{\eta n}}\sup_{(\alpha,\beta)\in D^*(n,c)}\int l_n[C(\hat{z}_n,\hat{r}_n);A_n(\alpha;\cdot)]$ $dP^n_{\alpha,\beta} \geq Eg[\|X\| + r_{1\eta}]$ where X is given in proposition 2.1 and $r_{1\eta}$ is defined by $P\{\|X\| \leq r_{1\eta}\} = 1 - \eta$.

Remark 5.1. The number $r_{1\eta}$ has, of course, a simple characterization; see Proposition 5.3 below.

The LAM lower bound of Proposition 5.1 required only the hypotheses of theorem 2.1. However, in order that the confidence set $\hat{C}_{1n}$ achieve its lower bound, a slight strengthening of the hypothesis of theorem 2.1 is necessary. The conceptually simpler "plug-in" method of Complement (5.2) will also require strengthening of these hypotheses. Let us therefore introduce the "strong triangular array hypotheses":

(5.8)the hypotheses of theorem (2.1) hold, but with $P^n_{\alpha_n\beta_0}$ replaced by $P^n_{\alpha_n\beta_n}$ where (α_n, β_n) satisfy

$$n^{1/2}|\beta_n - \beta_0| \leq c$$

and

$$n^{1/2}\|A_n(\alpha_n;\cdot) - A(\alpha_0;\cdot)\| \leq c \tag{5.8a}$$

where $A(\alpha_0;t) = \int_0^t \alpha_0(s)I_s[q_s]^{-1}ds$. The escalation to (5.8a) is severe, but appears unavoidable even in the context of Complement (5.2).

Proposition 5.2. (LAM character of $\hat{C}_{1n}$). Assume the hypotheses of Theorem 2.1 in "strong triangular form", as given by (5.8). Then

$$\lim_{c\uparrow\infty}\lim_{n\to\infty}\sup_{(\alpha,\beta)\in D^*(n,c)}\int l_n(\hat{C}_{1n})\,dP^n_{\alpha\beta} = El\,(\|X\| + r_{1\eta}).$$

Moreover, $\hat{r}_{1n} \Rightarrow r_{1\eta}$ under $P^n_{\alpha_n\beta_n}$, $(\alpha_n\beta_n) \in D^*(n,c)$.

The number $r_{1\eta}$, which depends on $A(\alpha_0;\cdot)$, β_0:

$$r_{1\eta} = r_1(\eta; A(\alpha_0;\cdot); \beta_0), \tag{5.9}$$

was given an abstract characterization in Propositions 5.1, 5.2; this abstract description was meant to emphasize the similarities in structure between the confidence bounds here, and those in other statistical applications (cf, Beran and Millar, 1986, for example). Unlike most other non-parametric applications, the numbers $r_{1\eta}$ here have a simple characterization in terms of known distributions. The statistical significance of $r_{1\eta}$ of course is that for large n, the width of the optimal confidence band $\hat{C}_{1n}$, centered at

$\hat{A}_n$, is approximately $2r_{1n}n^{-1/2}$, according to proposition 5.2.

To describe $r_{1\eta}$, recall the function $\Phi(t) \equiv \Phi(A(\alpha_0; \cdot); \beta_0; t)$ given in section 2. If $\{B_s, s \geq 0\}$ is standard Brownian motion, define

$$L(y) = P\{\max_{0 \leq s \leq 1} |B_s| \leq y\}. \tag{5.10}$$

Then $L(y)$ is "known" in, for example, the form of series expansions. Define k_η to be the $1 - \eta$ point of L:

$$L(k_\eta) = 1 - \eta. \tag{5.11}$$

An easy argument (see section 6) then yields:

Proposition 5.3. The number $r_{1\eta}$ is given by

$$r_{1\eta} = k_\eta \Phi(A(\alpha_0; \cdot); \beta_0; 1)^{1/2}$$

(5.C). Complements.

This subsection describes several variants on the ideas of subsections (5.A), (5.B).

Complement 5.1: Estimation of $\exp\{-A_n(\alpha)\}$.

The propositions 5.1, 5.2 are easily extended (using the simple idea of section 3) to the case of the estimation of $\exp\{-A_n(\alpha)\}$ by means of $\hat{C}_{2n}$, the bounds centred at product limit estimators. The number $r_{2\eta} \equiv \lim_n \hat{r}_{2n}$ can be characterized in terms of transformations on Brownian motion, but there is no simple result like proposition 5.3.

Complement 5.2: asymptotic plug-in confidence bands.

The results of subsection 5.B suggest a computationally simpler confidence band of the form:

$$\hat{C}_{an} \equiv \{f \in L_\infty[0,1] : \|f - \hat{A}_n\| \leq k_\eta \Phi(\hat{A}_n, \hat{\beta}_n 1)\} \tag{5.12}$$

where we have used the notation of Proposition 5.3. Under the hypotheses of Proposition 5.2, this confidence band will also be asymptotically optimal in the sense defined in subsection 5.B; the proof is similar to that given for $\hat{C}_{1n}$. The band $\hat{C}_{an}$ is clearly easier to compute than $\hat{C}_{1n}$, since one is not faced with the problem of replicating Aalen processes starting with preliminary estimates. On the other hand, several recent studies in the bootstrap literature show that often a bootstrap confidence set will be "better" than one constructed by "plug-in" methods based on asymptotic formulae. Such analyses depend on "second order" properties, typically involving Edgeworth expansions; see, for example, Abramovitch and Singh (1985), Hall (1986), Diciccio and Romano (1988).

Asymptotic optimality properties such as LAM are "first order" properties, and cannot distinguish between $\hat{C}_{1n}$ and $\hat{C}_{an}$. Since "second order" analysis of Aalen estimates is a completely uncharted field, it is not possible at the present time to decide between $\hat{C}_{1n}$, $\hat{C}_{an}$. On the otherhand, since the trend of research in other statistical areas suggests that $\hat{C}_{1n}$ is often better - and no worse - than $\hat{C}_{an}$, we have featured in this section the more complicated bootstrap method. Finally (cf 5.8a), for the plug in method to give optimal bands, $\Phi(A\,;\,B\,;\,1)$ must be a smooth function of the integrated Aalen parameter A (for the L_∞ norm on A).

Complement 5.3: confidence bands of shape f.

The confidence bands given in subsection 5.A and also in complement 5.2 are based on the notion of a ball in $L_\infty([0,1])$. Obviously, many Banach spaces other than L_∞ could be used here, and also in section 2.3 to express the LAM results. Here is one possibility. Let f be a real function on $[0,1]$, and assume for convenience that $0 < \inf_{0\le t\le 1} f(t) \le \sup_{0\le t\le 1} f(t) < \infty$. Define a norm $\|\ \|_f$ on real functions b: $[0,1] \to R'$ by

$$\|b\|_f = \sup_{0\le t\le 1} |b(t)/f(t)|.$$

One may now repeat the entire development of this section (and preceding ones), replacing the L_∞ norm $\|\ \|$ by $\|\ \|_f$. The resulting confidence bands will then be LAM *with respect to the loss function* determined by $\|\ \|_f$ instead of $\|\ \|$; see subsection 5.B. In this manner we find that the confidence bands "having shape f" as described by Jacobsen, 1982, p. 204, are "optimal". The optimality is relative to the chosen norm; the theory of Beran and Millar (1985) does not, in the form given there, provide comparisons for confidence sets determined by different norms. Note that such optimality results can be extended to any Banach space B consisting of real functions on $[0,1]$, provided mainly that (a) $\hat{A}_n(\cdot)$ is a B-valued random variable and (b) Rebolledo's CLT holds for $\hat{A}_n$ on B. In particular, the stringent conditions on f given above can be greatly relaxed.

Complement 5.4. Implementation of the bootstrap method.

Actual calculation of $\hat{C}_{1n}$ via Monte Carlo methods involves constructing iid copies of a multiplicative intensity process beginning with initial estimates of the *integrated* Aalen parameter and the nuisance parameter β. To see this, let $\hat{A}_n$, $\hat{\beta}_n$ be the estimates of the integrated Aalen parameter and β. *Conditional* on the values of $\hat{A}_n$, $\hat{\beta}_n$ construct $N^*_{n1}, \ldots, N^*_{nn}$ i.i.d. point processes on $[0,1]$ whose integrated Aalen parameter is $\hat{A}_n$, and whose nuisance parameter is $\hat{\beta}$. Next, using N^*_{ni} construct A^*_{ni}, the estimate of the integrated Aalen parameter $\hat{A}_n$, derived by the usual recipes, from N^*_{ni}. Finally, construct the empirical c.d.f. $\hat{F}_n$ of $\{|A^*_{ni} - \hat{A}_n|a_n, 1 \le i \le n\}$, and use as a guess for $\hat{r}_{1n}$ (at level $1-\eta$) the $1-\eta^{th}$ quantile of the $\hat{F}_n$ just defined (or the closest thing to it). The law of large numbers guarantees this will work, at least theoretically.

In the present context, difficulties attend this construction. First, given Aalen parameter α and nuisance parameter β, it is unknown in general how to simulate iid copies of the relevant point process with these parameters. This difficulty is not new, and arises in other areas of bootstrap applications. In important special cases, however, (esp., examples 4.1, 4.2 section 4) one knows methods of effecting such simulations; see Lo and Singh (1986) for discussion of Example 4.2. Indeed a computer intensive methodology for simulating multiplicative intensity models, on anything beyond a case by case bases, is an important open area of research. A second difficulty attending the bootstrap construction centres on the condition that the simulations begin from an estimate of the *integrated* Aalen parameter, and not the parameter itself. The fact that $\hat{A}_n$ estimates a random variable (see section 2) has unknown consequences for the validity of the simulation; the strength of the restriction (5.8a) has already been noted. A further point is worth noting here. The Aalen estimates of the integrated parameter, by definition, have certain measurability properties relative to the given filtration $\{\mathbf{F}_t, t \ge 0\}$. On the other hand, one may wish to take as estimate of α a "smooth" version of the "density" of $\hat{A}_n$; see Ramlau-Hanson 1983 for some possibilities. In particular, one might wish to select a smooth version whose integral fails to have the usual measurability properties. The success of the bootstrap simulations will not be affected by such measurability considerations; the "smooth" estimate $\hat{\alpha}_n$ need only have the property that its integral is subject to the usual CLT.

6. Proofs.

Proof of proposition 5.1. It suffices to establish the lower bound with $D^*(n,c)$ replaced by the $D(n,c)$ of section 2. Take P^n_h in the theorem of section 4, Beran and Millar, 1985, to be the measure Q^n_h described in section 2 of the present paper. Define the ξ *of Beran and Millar, 1985,* to be $\xi(P^n_h) = A_n(\alpha_h)$, where α_h is given in (2.13) above; then ξ' is the identity, and proposition 5.1 follows from theorem 4.5 of Beran and Millar, 1985.

Proof of proposition 5.2. Let $(\alpha_n, \beta_n) \in D^*(n,c)$. Then by the strong triangular array hypotheses of section 5 above, $a_n \| \hat{A}_n - a_n(\alpha_n, \beta_n) \| \Rightarrow \|X\|$ where X was given in section 2, and where the convergence is under $\{P^n_{\alpha_n\beta_n}\}$. It is easy to see that $\|X\|$ has a continuous distribution with strictly increasing cdf; this implies that $t_n(A_n(\alpha_n), \beta_n; \eta)$ converges to r_η. But $\|\hat{A}_n - A(\alpha_0)\| a_n$ is tight and by assumption so is $\|\hat{\beta}_n - \beta_0\| a_n$. It follows that $\hat{r}_{1n} \Rightarrow r_\eta$. Let $(\alpha_n \beta_n) \in D^*(n,c)$ be arbitrary. Then using Beran and Millar, 1985, p. 879, we find that $\lim_{n\to\infty} \int l_n(\hat{C}_{1n}; \alpha_n, \beta_n)\, dP^n_{\alpha_n\beta_n} = \lim_{n\to\infty} E_{\alpha_n,\beta_n} l(a_n \|\hat{A}_n - A(\alpha_n)\| + \hat{r}_{1n}) = E\, l(\|X\| + r_\eta)$. Since $(\alpha_n\beta_n)$ could have been chosen to achieve $\sup_{\alpha,\beta \in D^*(n,c)}$, this completes the proof.

Proof of proposition 5.3. Let $\{B_s, 0 \le s\}$ be standard Brownian motion. Since $\{B_s\}$ is equal in distribution to $t_0^{1/2} B(s/t_0)$, for any $t_0 > 0$, we see that $P\{\sup_{0\le s\le t_0} |B_s| \le y\} = L(y t_0^{-1/2})$, where L was defined in (5.10). If one chooses $t_0 = \Phi(\alpha_0, \beta_0; 1)$ then it is immediate that $r_\eta [\Phi]^{-1/2} = k_\eta$.

Proof of proposition 3.1. As in the proof of theorem 2.1, one may reduce the sup over $D(n,c)$ to a sup over $D_0(n,c)$, to get as a lower bound $\lim_{c\uparrow\infty} \liminf_{n\to\infty} \inf_T \sup_{\alpha \in D_0(n,c)} \int l\,[a_n(T - \xi(A_n(\alpha))]\,dP^n_\alpha$. Because of the differentiability of ξ, if α is given by (2.13), then the argument in l can be replaced by $l\,[T' - \xi' \circ \tau h] + o(1)$ (where $T' = a_n(T - A_n(\alpha_0)))$; this in turn can be replaced, for decision theoretic purposes by $l \circ \xi'[T'' - \tau h]$ where T'' ranges over the decision space T_n of section 2; this uses the hypothesis that ξ' has dense range in B_2. Since $l \circ \xi'$ is subconvex, the result is now immediate from theorem 2.1 An alternative approach, which is more convenient for establishing Remark 3.1b begins with the observation that, if H' is the orthocomplement in H of the null space of the linear map $\xi' \circ \tau$, then $(\xi' \circ \tau, H', B_2)$ is an abstract Wiener space, and the canonical normal on B_2 is the image of Q_0^∞ under ξ'. Using this, plus the evident equivalence of the relevant statistical experiments, one can rework the proof of theorem 2.1 to achieve the greater generality.

The attainment of the lower bound in proposition 3.1 by $\xi(\hat{A}_n)$ is immediate from hypothesis (3.3).

Proof of proposition 3.2. Because of the developments in the proof of theorem 2.1 — in particular the abstract Wiener structure and the convergence, in the sense of Le Cam, of the statistical experiments — the proof of proposition 3.2 is immediate from Millar, 1985, section 4.

References

Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6**, 701-26.

Abramovitch, L. and Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* **13**, *116-132*.

Beran, R.J. (1977). Estimating a distribution function. *Ann. Statist.* **5**, 400-404.

——— (1982). Estimated sampling distributions: the bootstrap and its competitors. *Ann. Statist.* **10**, 212-25.

Beran, R.J. and Millar, P.W. (1985). Asymptotic theory of confidence sets. Proc. Berkeley Conf. Neyman and Kiefer Vol II, Le Cam and Olshen, eds; 865-888. Wadsworth, Inc.

Beran, R.J. and Millar, P.W. (1986). Confidence sets for a multivariate distribution. *Ann Statist.* **14**, 431-443.

Diciccio, T.J. and Romano, J.P. (1988). A review of bootstrap confidence intervals. *J.R.S.S. (B)* **151**, XXX-XXX.

Doob, J.L. (1953). Stochastic Processes. Wiley.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1-26.

Hall, P. (1986). On the bootstrap confidence intervals. *Ann. Statist.* **14**, 1431-1452.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *J.A.S.A.* **53**, 457-481.

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

——— (1986). Point processes and their statistical inference. Dekker, N.Y.

Jacobsen, M. (1982). Statistical analysis of counting processes. Springer Lecture Notes in Statistics, vol. 12. Springer-Verlag, N.Y.

Le Cam, L. (1972). Limits of experiments. Proc. Sixth Berkeley Syposium, Vol. I.

——— (1986). Asymptotic methods in statistical decision theory. Springer series in statistics. Springer-Verlag, Berlin.

Lo, Sha Hwa and Singh, K. (1986). The product limit estimator and the bootstrap. *Prob. Th. Rel. Fields* **71**, 455-65.

Millar, P.W. (1979). Asymptotic minimax theorems for the sample distribution function. *Z. Wahr.* **48**, 233-252.

——— (1983). The Minimax Principle in Asymptotic Statistical Theory. Springer Lecture Notes in Mathematics, vol. 976, pp. 75-265.

——— (1985). Non-parametric applications of an infinite dimensional convolution theorem. *Z. Wahr.* **68**, 545-556.

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

Rebolledo, R. (1978). Sur les applications de la théorie des martingales à l'étude statistique d'une famille des processus stochastique. Lecture notes in Mathematics, 636, 27-70. Springer-Verlag, New York.

Reeds, J. On the definition of von Mises functionals. To appear, Univ. of Chicago Press.