Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

N. Keiding, R.D. Gill

Random truncation models and Markov processes

Department of Mathematical Statistics Report MS-R8702 February

The Centre for Mathematics and Computer Science is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

Random Truncation Models and Markov Processes

Niels Keiding
Statistical Research Unit, University of Copenhagen,
Blegdamsvej 3, DK-2200 Copenhagen N, Denmark

Richard D. Gill
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The life table with delayed entry is a concept as old as the life table itself. The product-limit estimator is the continuous-time version of the life table, and its generalization to allow for delayed entry, or left truncation, is currently being repeatedly rediscovered in the mathematical-statistical literature even though it is a well-established biostatistical tool.
The present note shows how the derivation of a nonparametric estimator of a distribution function under random truncation is a special case of results on the statistical theory of counting processes by Aalen and Johansen. This framework also clarifies the status of the estimator as nonparametric maximum likelihood estimator, and consistency, asymptotic normality and efficiency may be derived directly as special cases of Aalen and Johansen's general theorems and later work.

1980 Mathematics Subject Classification: Primary: 62G05, Secondary: 62M05.
Key Words & Phrases: counting processes, delayed entry, inference in stochastic processes, intensity function, left truncation, nonparametric maximum likelihood, product integral, survival analysis.
Note: The authors are grateful to Wei-Yann Tsai, Stefan Wellek and Benjamin B. Winter for supplying copies of unpublished work and to Søren Johansen for interesting discussions.

Report MS-R8702
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

1. INTRODUCTION

As has been known since Halley (1693), the construction of a life table involves following persons from an entrance age to an exit age and registering whether exit is due to death or end of observation for other reasons (censoring, in modern terminology). Kaplan and Meier (1958) initiated the modern mathematical-statistical analysis of the life-table in continuous time, or equivalently, the nonparametric estimation of a distribution function from right-censored observations.

Although the practical use of life table methods with delayed entry has flourished, a theoretical-statistical analysis of the parallel problem of the life table with delayed entry (in continuous time), that is nonparametric estimation of a distribution function under left truncation, has been scant. An important and original contribution was given by Hyde (1977) who however concentrated on hypothesis testing problems; Hyde's (1980) later contribution contains a product-limit estimator, but it is not clear from the context whether it is in fact properly corrected for left truncation. Cox and Oakes (1984, p. 177f.) followed the biostatistical tradition by referring to the product-limit estimator for left-truncated data as an obvious and well-known generalization.

The general theory of statistical models based on counting processes by Aalen (1975, 1978, Example 2) explicitly allowed for delayed entry and right censoring in a general context including the present one, as pointed out in a survey article for the medical research community by Kjøller, Larsen and Mortensen (1978), cf. Aalen (1980). Practical application to survival for diabetics, explicitly based on this approach, was given by Deckert, Poulsen and Larsen (1978), while Green and Hougaard (1984) fitted a Cox regression model to left-truncated survival data using age as basic time variable.

A different empirical motivation for the study of nonparametric estimation under random truncation comes from astronomy, as recently summarized by Woodroofe (1985). In fact, a heuristic maximum likelihood argument for the product-limit estimator under random truncation was given by Lynden-Bell (1971).

The purpose of this note is to demonstrate how an embedding of the basic nonparametric estimation problem into a simple Markov process model allows the product-limit estimator to be derived as a direct consequence of results by Aalen and Johansen (1978), thereby providing an alternative interpretation of Woodroofe's analysis. The consistency and asymptotic normality results of Woodroofe (1985) and Wang, Jewell & Tsai (1986) are shown to be consequences of Aalen and Johansen's limit theorems. Finally, the question whether the product-limit estimator is a (marginal) nonparametric maximum likelihood estimator, discussed by Wang et al. (1986) and by the same authors in further unpublished work, may be answered in the present framework by referring to the results by Johansen (1978). We indicate briefly that efficiency questions may be studied using the maximum likelihood property and a functional version of the δ-method (Reeds, 1976; Gill, 1986; van der Vaart, 1987).

In a final section we show how the present framework also covers a concept of 'noninformative left truncation' recently given a precise mathematical formulation by Wellek (1986) as well as an estimation problem in steady-state renewal processes studied by Winter and Földes (1986).

To keep the present note reasonably brief, no results on right censoring are given. However, the basic tools as given by Aalen (1975, 1978), Aalen and Johansen (1978) and Johansen (1978) explicitly include right censoring, so that this generalization is also covered by the present framework.

2. INTERPRETATION OF RANDOM TRUNCATION MODELS IN A SIMPLE MARKOV PROCESS MODEL

Woodroofe (1985) surveyed the problem of estimating the distributions of independent, positive random variables Y and X when sampling from the conditional distribution of (Y,X) given $Y \leq X$. We are in this note interested in the structure of the problem and (unlike Woodroofe) specialize to the case of absolutely continuous X and Y. Then $Y \leq X$ is almost surely equivalent to $Y < X$, which we shall use for ease of presentation. Non-absolutely continuous X and Y could be treated by similar methods, cf. Gill (1980).

Keeping as closely as possible to Woodroofe's notation, assume that Y and X are independent, positive random variables with distribution functions G and F, densities g and f, cumulative hazard functions Γ and Φ and hazard functions

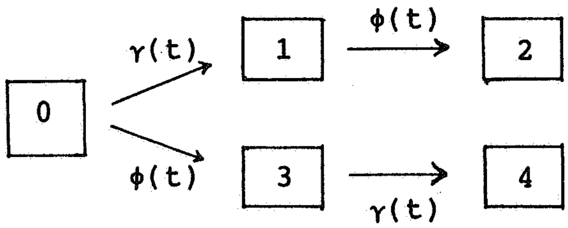
$$d\Gamma(y)/dy = \gamma(y) = g(y)/[1-G(y)]$$

and

$$d\Phi(x)dx = \phi(x) = f(x)/[1-F(x)].$$

We restrict in this note attention to the most interesting (and most difficult) situation where Y and X have common support, without loss of generality taken to be $[0,\infty)$; it follows that $\alpha = P\{Y<X\}>0$.

Define a Markov process $U(t)$ on $[0,\infty)$ by $U(0) \equiv 0$ and the transition intensities specified in the diagram



Then observation of (Y,X) given $Y<X$ is equivalent to observing $U(t)$ in the conditional distribution given $U(\infty)=2$ ("ultimate absorption in 2").

A practical application of a slightly more general Markov process model, containing nonparametric estimation with delayed entry, was given by Aalen, Borgan, Keiding and Thormann (1980).

The following is a standard result for Markov processes; an explicit formulation was given by Hoem (1969).

<u>Proposition 2.1</u> In the conditional distribution given $U(\infty)=2$, the stochastic process $\{U(t),\ t\geq 0\}$ is a Markov process with transition intensity from 0 to 1 given by

$$\lambda_1(t) = \gamma(t)\,\frac{P_{12}(t,\infty)}{P_{02}(t,\infty)} = \frac{\gamma(t)}{P\{X>Y|X>t,Y>t\}}$$

and from 1 to 2 given by

$$\lambda_2(t) = \phi(t) \frac{P_{22}(t,\infty)}{P_{12}(t,\infty)} = \phi(t),$$

where $P_{ij}(t,u)$ are the transition probabilities in the original Markov process. Define $\Lambda_i(t) = \int_0^t \lambda_i(u)du$, $i=1,2$.

2a. Estimation of the Distribution of X.

It is seen that $\phi(t)$ is directly identified as the transition intensity $\lambda_2(t)$, and its nonparametric estimation may therefore be derived directly from the counting process treatment given by Aalen (1975, Section 5D; 1978) and Aalen and Johansen (1978).

Corresponding to (Y,X) define a two-dimensional counting process by

$$\begin{aligned} N_1(t) &= \#\{\text{jumps from 0 to 1 in } [0,t]\} \\ &= I\{Y \leq t\} \end{aligned}$$

$$\begin{aligned} N_2(t) &= \#\{\text{jumps from 1 to 2 in } [0,t]\} \\ &= I\{Y < X \leq t\}. \end{aligned}$$

With respect to the self-exciting filtration, and under the conditional probability given $Y<X$, the bivariate counting process $N(t) = (N_1(t), N_2(t))$ has compensator $A(t) = (A_1(t), A_2(t))$ given by

$$dA_1(t) = \lambda_1(t)V_1(t)dt = \lambda_1(t)I\{Y>t\}dt$$

$$dA_2(t) = \phi(t)V_2(t)dt = \phi(t)I\{X>Y,X>t\}dt.$$

In the following we study independent identically distributed replications $(Y_1,X_1),\ldots,(Y_n,X_n)$ of (Y,X) conditional on $Y<X$. (The variables Y_i,X_i with subscript will from now on be interpreted in this conditional distribution, the variables Y,X without subscript in the original distribution where Y and X are independent).

Thus from now on

$$N_1(t) = \#\{Y_i \leq t\}$$
$$N_2(t) = \#\{Y_i<X_i \leq t\}$$
$$V_1(t) = \#\{Y_i \geq t\}$$
$$V_2(t) = \#\{Y_i<t \leq X_i\},$$

and define also $J_i(t) = I\{V_i(t)>0\}$, $i=1,2$.

As estimator of the integrated intensity $\Phi(t)$ we derive the <u>Nelson-Aalen estimator</u>

$$\hat{\Phi}(t) = \int_0^t \frac{J_2(u)}{V_2(u)} dN_2(u) = \sum_{i=1}^{n} \frac{I\{X_i \leq t\}}{V_2(X_i)} .$$

It is then a basic result in the statistical analysis of counting processes that, defining

$$\Phi^*(t) = \int_0^t J_2(u)\phi(u)du,$$

the process $\hat{\Phi}(t) - \Phi^*(t)$ is a mean zero, square integrable martingale with predictable variation process given by

$$\langle\hat{\Phi} - \Phi^*\rangle(t) = \int_0^t \frac{J_2(u)}{V_2(u)} \phi(u)du.$$

These properties imply the unbiasedness result

$$E(\hat{\Phi}(T)) = E(\Phi^*(T)) \tag{2.1}$$

for any stopping time T and suggest the estimator

$$\hat{\tau}(t) = \int_0^t \frac{J_2(u)}{[V_2(u)]^2} dN_2(u)$$

of the mean squared error function $\tau(t) = E[<\hat{\Phi} - \Phi^*>(t)]$.

Let us take a concrete look at the process $J_2(u) = I\{V_2(u)>0\}$. Since we have assumed that ess inf X = ess inf Y=0, we will with probability one have $Y_{(1)}>0$, $Y_{(1)}$ as usual denoting the smallest Y, so that $V_2(u) = 0$ on a proper interval $[0,Y_{(1)}]$. It may happen that $V_2(u) = 0$ on further intervals $(U_1,Z_1],\ldots,(U_k,Z_k]$, and it certainly becomes 0 for $u>X_{(n)} = U_{k+1}$. The serious problem is that $\Delta\hat{\Phi}(U_1) = \Delta N_2(U_i)/V_2(U_i) = 1$, "using up" the probability mass in the middle of the observation interval.

Interpreting $Z_0 = Y_{(1)}$ the unbiasedness relation (2.1) specializes to

$$E[\int_0^t I\{u\in \bigcup_{i=0}^{k} (Z_i,U_{i+1}]\} (\frac{dN_2(u)}{V_2(u)} - \phi(u)du)] = 0.$$

In particular, the integrated intensity estimate $\hat{\Phi}(t) - \hat{\Phi}(u)$ over the subinterval $(u,t]$ has a useful interpretation as long as $u<t$ both belong to the closure of an "observation interval" $(Z_i,U_{i+1}]$ given by $V_2(t)$ constantly positive: for such (u,t) we have

$$\hat{\Phi}(t) - \hat{\Phi}(u) = \int_u^t dN_2(u)/V_2(u),$$

$$\Phi^*(t) - \Phi^*(u) = \int_u^t \phi(u)du = \Phi(t) - \Phi(u).$$

It is however still difficult to formulate applicable unbiasedness results, because there are no restrictions on the location of the intervals with no risk set $(V_2=0)$.

The interest in the literature has focused on estimating not the integrated intensity of X but rather its distribution function F or (equivalently) its survivor function S_X. The formal Aalen and Johansen (1978, Theorem 3.2) answer is to use the product-limit (or generalized Kaplan-Meier) estimator

$$\hat{S}_X(t) = \prod_{[0,t]} [1-d\hat{\Phi}(u)]$$

where the product integral is the finite product

$$\prod_{i=1}^{n} \left(1-\frac{I\{X_i \leq t\}}{V_2(X_i)}\right).$$

Unbiasedness and mean square error results derive from the fact that defining

$$S_X^*(t) = \prod_{[0,t]} [1-d\Phi^*(u)]$$

$$= \begin{cases} \dfrac{S_X(U_1)}{S_X(Z_0)} \cdots \dfrac{S_X(U_i)}{S_X(Z_{i-1})} \dfrac{S_X(t)}{S_X(Z_i)} & \text{for } t \in (Z_i, U_{i+1}] \\ \\ \dfrac{S_X(U_1)}{S_X(Z_0)} \cdots \dfrac{S_X(U_i)}{S_X(Z_{i-1})} & \text{for } t \in (U_i, Z_i] \end{cases}$$

we have that

$$\frac{\hat{S}_X(t)}{S_X^*(t)} - 1 = \int_0^t \frac{\hat{S}_X(u-)}{S_X^*(u)} \, d[\hat{\Phi}(u) - \Phi^*(u)]$$

is a zero-mean, locally square integrable martingale with predictable squared variation process given by

$$\langle \hat{S}_X(t)/S_X^*(t) - 1 \rangle$$

$$= \int_0^t [\hat{S}_X(u-)/S_X^*(u)]^2 \quad d\langle \hat{\Phi} - \Phi^* \rangle(u)$$

$$= \int_0^t [\hat{S}_X(u-)/S_X^*(u)]^2 \quad \frac{J_2(u)}{V_2(u)} \quad \phi(u)du.$$

Hence for any bounded stopping time T

$$E[\hat{S}_X(T)/S_X^*(T)] = 1$$

and the squared variation of $\hat{S}/S^*$ may be estimated by

$$\int_0^t J_2(u)\, V_2^{-2}(u)\, dN_2(u).$$

It follows that a natural estimate of the covariance function of $\hat{S}_X$ is given by Major Greenwood's formula

$$\text{cov}\,(\hat{S}_X(s),\hat{S}_X(t)) = \hat{S}_X(s)\hat{S}_X(t) \int_0^{s\wedge t} J_2(u)\, V_2^{-2}(u) dN_2(u).$$

Note that since $d\hat{\Phi}(U_i) = 1$, $i=1,\ldots,k+1$, and in particular $d\hat{\Phi}(U_1) = 1$, the estimator $\hat{S}_X(t) = 0$ for $t \geq U_1$. This is a serious problem if there exist values of Y_j (and hence X_j) larger than U_1 because the estimator of the distribution of X will then be supported by a proper subset consisting of the smaller observed X_j. Woodroofe (1985, p. 168) recognized the problem and suggested an ad hoc mending. We shall see in Section 4 below that the formal nonparametric maximum likelihood estimator in this case does not exist. We prefer to forgo the desire to estimate S_X on the whole of $[0,\infty)$ but rather limit ourselves to quote estimated conditional probabilities $P\{X>u|X>z\}$ for u and z belonging to the same observation interval: $Z_i \leq z < u \leq U_{i+1}$ for some i. Perhaps the hazard (or cumulative hazard) is more appropriate than the distribution function for communicating the results of the estimation since the hazard at any time x is the same for all conditional distributions of X given $X > x_0$, $x_0 < x$.

For $t \in (Z_0,U_1]$ (which may be the only observation interval) it should be remarked that

$$S_X^*(t) = \prod_{[0,t]} (1-d\Phi^*(t)) = \prod_{(Z_0,t]} (1-d\Phi(t)) = S_X(t)/S_X(Z_0)$$

which motivates the interpretation of $\hat{S}_X(t)$ as estimating not $P\{X>t\}$ but rather the conditional probability $P\{X>t|X>Z_0\}$.

In the example in Table 1 of Woodroofe (1985), $z_0 = y_{(1)} = .0136$ and the estimator F_n might be interpreted as that of the conditional distribution of X given $X > .0136$, which is in this case the uniform distribution on $[.0136,1]$. The maximum likelihood estimator of the mean was quoted by Woodroofe as .5192, which is closer to that of the conditional distribution (.5098) than to the marginal mean of .5.

2b. Estimation of the Distribution of Y

Whereas the conditional Markov process defined at the beginning of the section allowed direct estimation of the hazard ϕ of X, ϕ being identified as the transition intensity from 1 to 2, no similar simple approach is possible regarding the hazard γ of Y. Instead it is useful to consider the backwards intensities $\bar{\lambda}_i(t)$ from state $i+1$ to state i in the timereversed conditional Markov process $U(t)$ on $[0,\infty]$ with time running backwards:

$$U(\infty) \equiv 2$$

$$P\{U(t-h) = i-1 \mid U(t) = i,\ Y<X\} = \overline{\lambda}_i(t)h + o(h),\ i=1,2.$$

(Note that $U(t)=i \Rightarrow Y<X$ for $i=1,2$ so that the conditioning event could be replaced by $\{U(t)=i\}$). The following proposition is a standard result in Markov processes and is easily proved directly.

<u>Proposition 2.2</u> The backwards intensities are given by

$$\overline{\lambda}_i(t) = \lambda_i(t) \frac{P\{U(t) = i-1 \mid X>Y\}}{P\{U(t) = i \mid X>Y\}},\ i = 1,2.$$

Proposition 2.2 implies in particular that

$$\overline{\lambda}_1(t) = \frac{\gamma(t)}{P\{Y>X \mid Y>t, X>t\}} \frac{P\{Y>t,\ X>t \mid Y<X\}}{P\{Y\leq t,\ X>t \mid Y>X\}}$$

$$= \gamma(t) \frac{P\{Y>t,\ X>t\}}{P\{Y<X,\ Y\leq t, X>t\}}$$

$$= \gamma(t) \frac{P\{Y>t,\ X>t\}}{P\{Y\leq t,\ X>t\}}$$

$$= \gamma(t) \frac{1 - G(t)}{G(t)}$$

by the independence of X and Y under P. Define the <u>backwards hazard</u>

$$\bar{\gamma}(t) = g(t)/G(t)$$

with the interpretation

$$P\{Y \geq t-h \mid Y<t\} = \bar{\gamma}(t)h + o(h),$$

then we have shown that $\bar{\lambda}_1(t) = \bar{\gamma}(t)$, which of course also follows directly by symmetry of time.

The result is that $\bar{\gamma}(t)$ or its integral $\bar{\Gamma}(t)$ may be estimated by reversing time and using a <u>backwards Nelson-Aalen estimator</u>; similarly $G(t)$ may be estimated by a generalized backwards Kaplan-Meier estimator. These results are also immediate by symmetry of time since the original problem contains no special preference for one particular direction of time.

The various complications are exactly as for the estimation of the distribution of X, and moreover, there are complications in estimating both distributions or not at all.

In particular, there is no information in the sample on the distribution of Y on $[X_{(n)}, \infty)$. A useful formulation of this is to state that only the conditional distribution of Y on an interval $[0,y)$, $y \leq X_{(n)}$ may be estimated.

Completing the example of Woodroofe (1985) for the distribution of Y, one obtains a maximum likelihood estimate of the mean as .414, which is closer to the mean (.472) of the conditional distribution given $Y < x_{(10)} = .9441$ than to the marginal mean of .5.

2c. Woodroofe's Inversion Formulae.

Woodroofe (1985, Theorem 1) derived formulae to show that under suitable conditions, the marginal distribution functions G and F are identifiable from the conditional joint distribution function

$$H_*(x,y) = P\{X \leq x,\ Y \leq y \mid Y < X\} = \alpha^{-1} \int_0^x G(y \wedge z)\, dF(z)$$

which has marginals

$$F_*(x) = H_*(x,\infty) = P\{X \leq x \mid Y < X\}$$

and

$$G_*(y) = H_*(\infty,y) = P\{Y \leq y \mid Y < X\};$$

here

$$\alpha = P\{Y<X\} = \int_0^x G(z)dF(z) = \int_0^x [1 - F(z)]dG(z).$$

Using our notation Γ and Φ, Woodroofe proved

$$\Phi(x) = \int_0^x dF_*(z)/C(z), \quad 0 \leq x < \infty$$

and

$$\overline{\Gamma}(y) = \int_y^\infty dG(z)/G(z) = \int_y^\infty dG_*(z)/C(z), \quad 0 \leq y < \infty$$

with

$$C(z) = G_*(z) - F_*(z), \quad 0 \leqq z < \infty.$$

Since

$$C(z) = P\{Y\leq z<X|Y<X\} = P\{U(z)=1|Y<X\},$$

it is seen that Woodroofe's formulae are contained in the Markov process framework introduced above, and in fact allow attractive interpretations there.

3. ASYMPTOTIC RESULTS

3a. Convergence on $[\varepsilon,M]$, $0<\varepsilon<M<\infty$.

As we have seen, in interpreting $\hat{\Phi}$ and $\hat{S}_X$ in a practical situation, it is rather important to take account of the fact that $d\Phi$ can really only be estimated on the interval or intervals $\{t:V_2(t)>0\}$. In sketching how the counting process formulation of the left-truncation problem can be used in a very direct way to derive asymptotic distribution theory for our estimators, we shall similarly take care of this problem by first only estimating

$$\Phi^{\varepsilon} = \Phi - \Phi(\varepsilon)$$

and

$$S_X^{\varepsilon} = S_X/S_X(\varepsilon)$$

on an interval $[\varepsilon,M]$ whose endpoints $t=\varepsilon,M$ satisfy $P\{Y<t\leq X|Y<X\} > 0$.

Let $\hat{\Phi}^{\varepsilon}$, $\Phi^{*\varepsilon}$, $\hat{S}_X^{\varepsilon}$ and $S_X^{*\varepsilon}$ be defined similarly to Φ^{ε} and S_X^{ε}, and recall our notational conventions; Y and X are independent random variables with distribution functions G and F; (Y_i,X_i) for $i=1,\ldots,n$ denote independent replicas of (Y,X) <u>conditional</u> on $Y<X$. Thus $P\{Y_i<X_i\}=1$ while $P\{Y<X\}=\alpha<1$. Let us also write

$$v_2(t) = E(n^{-1}V_2(t)) = P\{Y_i<t\leq X_i\} = P\{Y_i<t\}-P\{X_i<t\} = C(t),$$

in Woodroofe's notation. We have

$$v_2(t) = P\{Y<t\leq X,Y<X\}/P\{Y<X\} = G(t)[1-F(t)]/\alpha$$

$$\geq G(\varepsilon)[1-F(M)]/\alpha > 0$$

for $\varepsilon\leq t\leq M$ by the assumption that Y and X have support $[0,\infty)$, so that $0<G(t)<1$, $0<F(t)<1$ for $0<t<\infty$.

Now $n^{-1}\,V_2$ is the difference between two empirical distribution functions, so by the Glivenko-Cantelli theorem

$$||\ n^{-1}V_2 - v_2\ ||_{\varepsilon}^{M} \to 0 \quad \text{a.s.}$$

as $n \to \infty$, where $||\cdot||_{\varepsilon}^{M}$ denotes the supremum norm over $[\varepsilon,M]$. Thus by boundedness away from zero of v_2

we also have

$$|| v_2^{-1} - nV_2^{-1} ||_\varepsilon^M \to 0 \quad \text{a.s.}$$

as $n \to \infty$, and $J_2 \equiv 1$ on $[\varepsilon,M]$ for all sufficiently large n a.s. Thus $\Phi^{*\varepsilon} \equiv \Phi^\varepsilon$ and $S_X^{*\varepsilon} \equiv S_X^\varepsilon$ on $[\varepsilon,M]$ for all sufficiently large n almost surely.

With these preparations made, weak convergence of $n^{\frac{1}{2}}(\hat{\Phi}^\varepsilon - \Phi^\varepsilon)$ and/or of $n^{\frac{1}{2}}(\hat{S}_X{}^\varepsilon - S_X{}^\varepsilon)$ follows immediately from standard results on the Nelson-Aalen and the product-limit estimators in the counting process literature.

Briefly, Aalen (1975, Theorem 8.2), cf. Aalen (1978, Theorem 6.4) proved weak convergence of the Nelson-Aalen estimator (in a general model containing the present one) using martingale central limit theory; Aalen and Johansen (1978, Theorem 4.6) treated the product-limit estimator in a general Markov process model (containing ours). These early results relied on uniform integrability of the random variables $nJ_2(t)/V_2(t)$ over $n=1,2,...$ and $t \in [\varepsilon,M]$, which is true but requires some calculation, see Aalen (1976, Proof of Lemma 4.2 in Appendix). Indeed, later developments in the theory of stochastic integrals have also made Aalen and Johansen's assumption (4.1) unnecessary.

The approach used here, establishing Glivenko-Cantelli convergence for V_2/n, leads to an easier verification of the conditions used by Rebolledo (1980) for his version of the martingale central limit theorem. This approach was used by Gill (1980, Section 4.2; 1983) for the product limit estimator in a model of random censorship (though the proof is valid in our situation too (Gill, 1980, Section 6)), and by Andersen and Borgan (1985, Appendix) for the Nelson-Aalen estimator in a general model, including the present one.

Theorem 3.1 Under the stated conditions,

$$n^{\frac{1}{2}}(\hat{\Phi}^{\varepsilon} - \Phi^{\varepsilon}) \xrightarrow{D} W^{\varepsilon} \quad \text{in} \quad D[\varepsilon, M] \tag{3.1}$$

as $n \to \infty$, where W^{ε} is a Gaussian martingale with zero mean and variance function

$$\operatorname{var} W^{\varepsilon}(t) = \int_{\varepsilon}^{t} \frac{\phi(s)}{v_2(s)} \, ds; \tag{3.2}$$

we also have (in fact, jointly)

$$n^{\frac{1}{2}}(\hat{S}_X{}^{\varepsilon} - S_X{}^{\varepsilon}) \xrightarrow{D} -S_X{}^{\varepsilon} . W^{\varepsilon}. \tag{3.3}$$

Furthermore, $\int_{\varepsilon}^{(\cdot)} n \, v_2(s)^{-2} d\, N_2(s)$ is a consistent estimator (in $||\cdot||_{\varepsilon}^{M}$) of the variance function of W^{ε}.

Corollary 3.1 We have

$|| \hat{\Phi} - \Phi ||_0^M \xrightarrow{P} 0$ and $|| \hat{S}_X - S_X ||_0^M \xrightarrow{P} 0$.

Corollary 3.2 The event of the existence of $s<t<u\leq M$ with $J_2(s) = J_2(u) = 1$, $J_2(t) = 0$, is asymptotically negligible.

Proofs of the corollaries are obtained easily, using $v_2(\varepsilon)>0$ for all $\varepsilon>0$, cf. Woodroofe (1985, p. 172).

It is curious that the probabilistic result of Corollary 3.2 is derived via the proof of consistency of a statistical estimator!

One should note that the counting process framework allows a direct identification from the martingale central limit theorem of the asymptotic covariance structure of the estimators, which was already suggested by the small sample arguments of Section 2a; thus no heavy calculations as used by Wang et al. (1986) are necessary.

3b. Convergence on $[0,\infty]$.

Since $v_2(t)>0$ for all $t \in (0,\infty)$ one can ask whether or not these results can be extended to yield weak convergence in $D[0,M]$ or $D[\varepsilon,\infty]$ or even $D[0,\infty]$, cf. Woodroofe (1985, Section 6; 1987). The extension of (3.3) at the righthand endpoint of the time interval was carried out by Gill (1983) (for the random censorship model) under natural additional conditions.

The analogous conditions in the left truncation model are automatically satisfied. We shall use the same techniques in order to study the lefthand endpoint problem, which may be of greater practical importance.

Since $\hat{S}_X$ and S_X are both close to 1 near $t=0$, one easily discovers that the extension problem for (3.3) is hardly more difficult than that for (3.1), on which we will concentrate. Also there is no hope of making an extension unless the limiting process can be extended too; for this we need to assume (cf. (3.2)) that

$$\int_0^{\varepsilon} \frac{\phi(s)}{v_2(s)}\, ds < \infty.$$

Now

$$\int_0^{\varepsilon} \frac{\phi(s)}{v_2(s)}\, ds = \alpha \int_0^{\varepsilon} \frac{dF(s)}{G(s)(1-F(s))^2} .$$

Since $F(s) \to 0$ as $s \to 0$, we have finiteness if and only if

$$\int_0^{\varepsilon} \frac{dF(s)}{G(s)} < \infty \tag{3.4}$$

for some (and then all) $\varepsilon>0$. From now on we assume (3.4) holds. We will have our required result

$$\left.\begin{array}{ll} n^{\frac{1}{2}}(\hat{\Phi} - \Phi) \xrightarrow{D} W & \text{in } D[0,M] \\ n^{\frac{1}{2}}(\tilde{S}_X - S_X) \xrightarrow{D} -S_X.W & \text{in } D[0,M] \end{array}\right\} \quad (3.5)$$

where W is W^ε with $\varepsilon=0$ of Theorem 3.1 if for all $\delta>0$

$$\lim_{\varepsilon\downarrow 0} \limsup_{n\to\infty} P\{n^{\frac{1}{2}} \, || \, \Phi^* - \Phi \, ||_0^\varepsilon > \delta\}=0 \qquad (3.6)$$

and for all $\delta>0$

$$\lim_{\varepsilon\downarrow 0} \limsup_{n\to\infty} P\{n^{\frac{1}{2}} \, || \, \hat{\Phi} - \Phi^* \, ||_0^\varepsilon > \delta\}=0 \qquad (3.7)$$

see Billingsley (1968; Theorem 4.2) for the basic idea here and Gill (1983; Proof of Theorem 2.1) for a similar application. We look at the easier term (3.7) first.

Now since $\hat{\Phi} - \Phi^*$ is a square integrable martingale, Lenglart's (1977) inequality gives us

$$P\{n^{\frac{1}{2}} \, || \, \hat{\Phi} - \Phi^* \, ||_0^\varepsilon > \delta\} \leq \eta + P\{n\langle\hat{\Phi} - \Phi^*\rangle(\varepsilon) > \frac{\delta^2}{\eta}\}.$$

But

$$n\langle\Phi - \Phi^*\rangle(\varepsilon) = \int_0^\varepsilon \frac{nJ_2(s)}{V_2(s)} \phi(s)ds$$

$$\leqq \int_0^\varepsilon \frac{nJ_2(s)}{\#\{i:Y_i<s\}} \phi(s)ds$$

$$\leqq \int_0^\varepsilon \frac{\phi(s)ds}{\beta P(Y_i<s)}$$

with probability $1-o(1)$ as $\beta\downarrow 0$, uniformly in n, by Wellner (1978, Remark 1 (ii)). Furthermore

$$\int_0^\varepsilon \frac{\phi(s)ds}{P(Y<s|Y<X)} = \alpha \int_0^\varepsilon \frac{\phi(s)ds}{P(Y<s,X>Y)}$$

$$\leqq \quad \alpha \int_0^\varepsilon \frac{\phi(s)ds}{G(s)(1-F(s))} \;.$$

Thus having assumed $\int_0^\varepsilon dF(s)/G(s) < \infty$, we can prove the required result: for taking ε sufficiently small, we can bound $n\langle\hat{\Phi} - \Phi^*\rangle_\varepsilon$ by an arbitrarily small constant with probability arbitrarily close to 1, uniformly in n; and this establishes (3.7).

As far as (3.6) is concerned, we note that

$$n^{\frac{1}{2}} \; || \; \Phi^* - \Phi \; ||_0^\varepsilon = n^{\frac{1}{2}} \int_0^\varepsilon (1-J_2(s))\phi(s)ds = n^{\frac{1}{2}}\Phi(Y_{(1)})$$

with probability $\to 1$ as $n \to \infty$ by Corollary 3.2. It suffices therefore to show $n^{\frac{1}{2}}\Phi(Y_{(1)}) \overset{P}{\to} 0$ as $n \to \infty$. Now $\Lambda_1(Y_{(1)})$ is the minimum of n i.i.d. exponential (1) random variables, hence

$$P\{n^{\frac{1}{2}}\Phi(Y_{(1)})>c\} = P\{\Lambda_1(Y_{(1)})>\Lambda_1(\Phi^{-1}(n^{-\frac{1}{2}}c)\} = e^{-n\Lambda_1(\Phi^{-1}(n^{-\frac{1}{2}}c))}.$$

So putting $\delta = n^{-\frac{1}{2}}c$, it suffices to prove

$$\delta^{-2}\Lambda_1(\Phi^{-1}(\delta)) \to \infty \quad \text{as} \quad \delta\downarrow 0$$

or (putting $\delta = \Phi(\varepsilon)$),

$$\Phi(\varepsilon)^{-2}\Lambda_1(\varepsilon) \to \infty \quad \text{as} \quad \varepsilon\downarrow 0$$

or

$$\Phi(\varepsilon)^2/\Lambda_1(\varepsilon) \to 0 \quad \text{as} \quad \varepsilon\downarrow 0.$$

Now one easily verifies that

$$\int_0^{\varepsilon} dF/G < \infty \quad \forall\ \varepsilon > 0$$

$$\Leftrightarrow \int_0^{\varepsilon} d\Phi/\Gamma < \infty \quad \forall\ \varepsilon > 0$$

$$\Leftrightarrow \int_0^{\varepsilon} d\Phi/\Lambda_1 < \infty \quad \forall\ \varepsilon > 0$$

$$\Rightarrow \int_0^{\varepsilon} d\Phi/\Lambda_1 \to 0 \quad \text{as} \quad \varepsilon\downarrow 0.$$

But $\int_0^{\varepsilon} d\Phi/\Lambda_1 \geqq \Phi(\varepsilon)/\Lambda_1(\varepsilon)$. So we certainly have

$$\Phi(\varepsilon)\ \frac{\Phi(\varepsilon)}{\Lambda_1(\varepsilon)} \to 0 \quad \text{as} \quad \varepsilon\downarrow 0,$$

as required.

3c. Remarks on Joint Weak Convergence

Finally we make some remarks on the estimation of Γ and G, possibly jointly with Φ and F.

By symmetry we can immediately write down weak convergence theorems for $n^{\frac{1}{2}}(\hat{\bar{\Gamma}}^M - \bar{\Gamma}^M)$ and/or $n^{\frac{1}{2}}(\hat{G}^M - G^M)$ and under further conditions drop the "M". A joint weak convergence result, e.g. for $n^{1/2}(\hat{S}_X^\varepsilon - S_X^\varepsilon)$ and $n^{1/2}(\hat{G}^M - G^M)$ is a little trickier. What can be argued is the following.

We certainly do have joint weak convergence of $n^{1/2}[n^{-1}V_1(\varepsilon) - v_1(\varepsilon)]$, $n^{1/2}[n^{-1}V_2(\varepsilon) - v_2(\varepsilon)]$, $n^{1/2}(\hat{\Lambda}_1^\varepsilon - \Lambda_1^\varepsilon)$ and $n^{1/2}(\hat{\Lambda}_2^\varepsilon - \Lambda_2^\varepsilon)$ in $R^2 x(D[\varepsilon,M])^2$ to a bivariate normal distribution and an independent pair of independent continuous Gaussian martingales.

Now $\hat{S}_X^\varepsilon = \Pi(1-d\hat{\Lambda}_2^\varepsilon)$ while $\hat{G}^M$ is a rather more complicated functional of $V_1(\varepsilon)/n$, $V_2(\varepsilon)/n$, Λ_1^ε and Λ_2^ε. However this functional is built up in a simple way of integrals of one empirical process with respect to another, products or ratios of empirical processes, and product integrals. By the compact differentiability of all these mappings and the functional version of the δ-method (see Reeds (1976), Gill (1986) and Gill and Johansen (1987)) weak convergence carries over to $(n^{1/2}(\hat{G}^M - G^M), n^{1/2}(\hat{S}_X^\varepsilon - S_X^\varepsilon))$ jointly. However identification of the covariance structure is not a pleasant task. It is clear that there will be dependence (if G is known, S_X can be better estimated; see Vardi (1985)).

When the extra conditions $\int_0^\infty (1-F)^{-1} dG < \infty$, $\int_0^\infty G^{-1} dF < \infty$ hold, the previously obtained extension results show that we have joint weak convergence of $n^{1/2}(\hat{F}-F)$ and $n^{1/2}(\hat{G}-G)$ in $(D[0,\infty])^2$. Another compact differentiability calculation leads to asymptotic normality of $n^{1/2}(\hat{\alpha}-\alpha)$ where $\hat{\alpha} = \int(1-\hat{F})d\hat{G}$.

We now sketch an alternative approach which leads more directly to fairly simple formulae for the variances and covariances of all these objects. For $i=1,2$ define $1-\hat{F}_i(t) = \prod_0^t (1-d\hat{\Lambda}_i)$ and similarly for F_i. Of course $\hat{F}_2=\hat{F}$ and $F_2=F$, while $\hat{F}_1$ is the empirical d.f. of the Y_i's and F_1 is their true d.f.,

$$F_1(t) = \frac{\int_0^t (1-F(s))dG(s)}{\int_0^\infty (1-F(s))dG(s)} .$$

By the simultaneous representations of $\hat{F}_i-F_i$ $(i=1,2)$ in terms of the orthogonal martingales $M_i = N_i - \int_0^{(\cdot)} Y_i d\Lambda_i$ and the same martingale Central limit theorem and extension results as before we can prove joint weak convergence in $(D[0,\infty])^2$ of $n^{\frac{1}{2}}(\hat{F}_i-F_i)$, $i=1,2$, to two independent processes $(1-F_i).W_i$ where W_i is a zero mean Gaussian martingale with $\text{var } W_i(t)=\int_0^t (v_i(s))^{-1}d\Lambda_i(s)$. In fact $(1-F_1).W_1$ has the same distribution as $B^0 o F_1$, where B^0 is a Brownian bridge on $[0,1]$. Since $dF_1 \propto (1-F)dG$ we have $dG \propto (1-F)^{-1}dF_1$ or

$$G(t) = \frac{\int_0^t (1-F(s))^{-1}dF_1(s)}{\int_0^\infty (1-F(s))^{-1}dF_1(s)} .$$

Also $\alpha = \int_0^\infty (1-F(s))dG(s) = [\int_0^\infty (1-F_2(s))^{-1}dF_1(s)]^{-1}$. By the invariance properties of maximum likelihood estimators (see Section 4a) the same relationships hold between $\hat{F}$, $\hat{G}$, $\hat{F}_1$, $\hat{F}_2$ and $\hat{\alpha}$. These rather simple expressions together with the simple form of the asymptotic covariance structure of $\hat{F}_1$ and $\hat{F}_2$ enable one to write down the asymptotic covariance structure of $\hat{F}$, $\hat{G}$ and $\hat{\alpha}$ rather easily.

One could consider using this alternative route to actually prove the weak convergence of $\hat{F}$ and $\hat{G}$, rather that the route we took earlier. However, one must then solve the nontrivial problem of extending weak convergence of

$$\{n^{\frac{1}{2}}[\int_0^t (1-\hat{F}(s))^{-1}d\hat{F}_1(s)-\int_0^t (1-F(s))^{-1}dF_1(s)],\ t\ [0,M]\}$$

from $D[0,M]$ to $D[0,\infty]$. This will require some careful calculations perhaps on the lines of Woodroofe (1987).

4. NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATION OF (G,F).

The embedding of the estimation problem into the simple Markov process in Section 2 also offers a framework for the interpretation of the pair of generalized product limit estimators of the distribution functions of Y and X as maximum likelihood estimator. Any absolutely continuous distribution functions G and F on $[0,\infty]$ will lead to a "conditional" Markov process $U(t)$

$$\boxed{0} \xrightarrow{\lambda_1(t)} \boxed{1} \xrightarrow{\lambda_2(t)} \boxed{2}$$

with $\lambda_2(t) = \phi(t)$,

$$\lambda_1(t) = \frac{\gamma(t)}{P\{Y<X \mid Y>t, X>t\}}$$

and initial and final conditions $U(0) = 0$ and $U(\infty) = 2$ a.s. given $Y<X$. We also assume that the supports of Y and X are both equal to $[0,\infty)$; this implies $F(x)<1 \Leftrightarrow \Phi(x)<\infty$ on $[0,\infty)$, that is for the Markov process that $P\{U(t) = i \mid Y>X\}>0$ for $i = 0,1,2$ and all $t \in (0,\infty)$. In terms of the integrated intensities $\Lambda_1(t) = \int_0^t \lambda_i(u)du$ and backwards integrated intensities $\bar{\Lambda}_i(t) = \int_t^\infty \bar{\lambda}_i(u)du$ we have

<u>Conditions 4.1</u> For $i = 1,2$ and $0<a<\infty$

$$0<\Lambda_i(a)<\infty,\ \Lambda_i(\infty)=\infty,\ 0<\overline{\Lambda}_i(a)<\infty,\ \overline{\Lambda}_i(0)=\infty.$$

One may, conversely, start with a Markov process $U(t)$ with intensities λ_i satisfying Conditions 4.1 and ask whether a left truncation model may be defined from it. That is, can we define the distribution of independent random variables Y,X with intensities γ and ϕ so that $U(t)$ is the conditional Markov chain constructed in Section 2, in particular such that $\overline{\gamma}=\overline{\lambda}_1$ and $\phi=\lambda_2$?

This is indeed possible: define $\overline{\gamma} = \overline{\lambda}_1$, $\phi = \lambda_2$ (these are intensities of proper probability distributions with support $[0,\infty)$ by Conditions 4.1); we only need to show

<u>Proposition 4.1</u> We have

$$\lambda_1(t) = \frac{\gamma(t)}{P\{Y<X|Y>t,X>t\}},\quad 0\leq t<\infty .$$

Proof. By the definition of the relevant quantities

$$\frac{\gamma(t)}{P\{Y<X|Y>t,X>t\}} = \frac{g(t)\ /\ S_y(t)}{\int_t^\infty \frac{g(u)}{S_y(t)} \frac{S_X(u)}{S_X(t)}\,du}$$

$$= \frac{g(t)\ /\ G(t)}{\int_t^\infty \frac{g(u)}{S_y(t)} \frac{G(u)}{G(t)} \frac{S_X(u)}{S_X(t)}\,du}$$

$$= \frac{\bar{\gamma}(t)}{\int_t^\infty \gamma(u)\ e^{-\int_t^u \bar{\gamma}(v)dv}\ e^{-\int_t^u \phi(v)dv} du}$$

and according to the definition of $\bar{\gamma}$ and ϕ from the Markov process, this may be written as

$$\frac{\bar{\gamma}(t)}{\int_t^\infty \bar{\gamma}(u) \frac{P\{U(u)=1|U(t)=1\}}{P\{U(t)=1|U(u)=1\}}\,du}$$

from which one may again use elementary reductions to obtain

$$\frac{\bar{\gamma}(t)P\{U(t)=1\}}{\int_t^\infty \bar{\gamma}(u)P\{U(u)=1\}du} = \frac{\gamma(t)P\{U(t)=0\}}{\int_t^\infty \gamma(u)P\{U(u)=0\}du}$$

where the denominator is the probability that the transition $0 \to 1$ will happen in $[t,\infty)$; this equals $P\{U(t)=0\}$ because of our assumption that $P\{U(\infty)=0\}=0$, and this completes the proof.

In order to rigorously discuss nonparametric maximum likelihood estimation, the left truncation model has to be extended to include also discrete distributions of (Y,X). We shall for ease of exposition assume that $dF(u)dG(u)=0$ for all u.

Define corresponding (right continuous) integrated intensity functions as

$$\Phi(t) = \int_{[0,t]} \frac{dF(u)}{1-F(u-)}$$

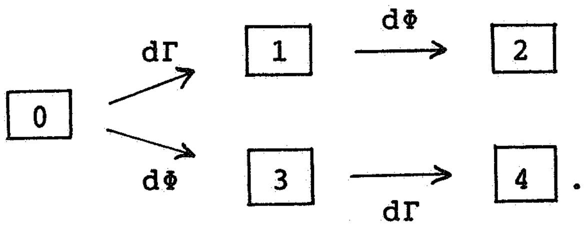
and similarly for $\Gamma(t)$. Also define the backwards integrated intensity function

$$\bar{\Gamma}(t) = \int_{[t,\infty)} \frac{dG(u)}{G(u)}$$

and notice that

$$d\bar{\Gamma}(t) = d\Gamma(t) \frac{1-G(t-)}{G(t)} .$$

From the functions Γ and Φ we may first define a Markov process $U(t)$ by $U(0)\equiv 0$ and the diagram



This process has the properties $P\{U(0)=0\}=1$, $P\{U(\infty)=2\}$ $= P\{Y<X\} = 1-P\{Y>X\} = 1-P\{U(\infty)=4\}$. Observation of (Y,X) given $Y<X$ is equivalent to observing the conditional Markov process given $U(\infty)=2$, which has intensity from 0 to 1 given by

$$d\Lambda_1(t) = d\Gamma(t)\,\frac{P_{12}(t,\infty)}{P_{02}[t,\infty)} = \frac{d\Gamma(t)}{P\{Y<X|Y\geqq t,X>t\}} \qquad (4.5)$$

and from 1 to 2 given by

$$d\Lambda_2(t) = d\Phi(t)\,\frac{P_{22}(t,\infty)}{P_{12}[t,\infty)} = d\Phi(t).$$

The backwards intensity from 1 to 0 in the conditional process may be seen by the same calculation as in Section 2b to be identical to the backwards intensity of Y : $\bar{\Lambda}_1 = \bar{\Gamma}$.

Let the supports of Y and X be $[s_1,t_1]$ and $[s_2,t_2]$ respectively (so that $d\Lambda_i(t)=0$ for $t \notin [s_i,t_i]$); it then follows from (4.5) that $d\Lambda_1(t)=0$, $t>t_2$ and therefore by inversion of time also that $d\bar{\Lambda}_2(t)=0$, $t<s_1$.

4a. Maximum Likelihood Estimation in a Markov Process.

Following the approach of Johansen (1978), we now define an extension of the conditional Markov process model (2.4) by allowing the integrated intensities $\Lambda_i(t) = \int_0^t \lambda_i(u)du$, $i=1,2$, to be arbitrary increasing right-continuous functions with jumps ≤ 1 (but no simultaneous jumps of $\Lambda_1(t)$ and $\Lambda_2(t)$) with the property that there exist times t_i with $d\Lambda_i(t_i)=1$, $\Lambda_i(u)=\Lambda_i(t_i)$ for $u>t_i$. We still assume $P\{U(0)=0\}=P\{U(\infty)=2\}$ $=1$, which implies that $t_1<t_2$: if the process has to end in state 2, all transitions $0\to1$ must happen before the last possible transition $1\to2$. Let $s_1=\sup\{s|P\{U(s)=0\}=1\}$ $=\sup\{s|\Lambda_1(s)=0\}$.

Before turning to estimation in such a model, we need to investigate whether any such Markov process will correspond to a left truncation model. That is, is it possible to recover integrated intensities Γ and Φ which correspond to the marginal distributions of independent random variables Y and X ?

If it is possible, Φ would have to be given by Λ_2 and the backwards integrated intensity $\bar{\Gamma}$ would have to be given by the backwards integrated intensity $\bar{\Lambda}_1$.

If there exists an <u>inner jump of size 1</u> of Λ_2, that is if there are $s<t<u$ such that $d\Lambda_2(s)>0$, $d\Lambda_2(t)=1$ and $d\Lambda_2(u)>0$, then Λ_2 does not correspond to a probability distribution. In this case there is no hope of interpreting Λ_2 as the cumulative intensity function Φ corresponding to X.

Hence assume that there are no inner jumps of size 1 of Λ_2 and define $\Gamma = \Lambda_1$ and $\Phi = \Lambda_2$. With these definitions, we certainly have $\Lambda_2 = \Phi$ and need to prove

$$d\Lambda_1(t) = \frac{d\Gamma(t)}{P\{Y<X|Y\geq t, X>t\}}, \quad s_1<t\leqq t_1.$$

<u>Proof</u>. Reasoning as for (4.1) we have

$$d\Lambda_1(t) = \frac{d\Lambda_1(t)P\{U(t-)=0\}}{P\{\text{the transition } 0\to 1 \text{ happens in } [t,\infty)\}}$$

$$= \frac{d\Lambda_1(t)P\{U(t-)=0\}}{\int_{[t,\infty)} P\{U(u-)=0\}d\Lambda_1(u)}$$

$$= \frac{d\bar{\Lambda}_1(t)P\{U(t+)=1\}}{\int_{[t,\infty)} P\{U(u+)=1\}d\Lambda_1(u)} \quad .$$

Because of the assumption of no inner jumps of Λ_2 we have

$$P\{U(u+)=1\}/P\{U(t+)=1\}$$

$$= P\{U(u+)=1|U(t+)=1\}/P\{U(t+)=1|U(u+)=1\}$$

for $s_1<u<t_1$, and the proof can be completed as for (4.1).

In this extended model the maximum likelihood estimator $(\hat{\Lambda}_1,\hat{\Lambda}_2)$ of (Λ_1,Λ_2) may be derived following arguments similar to Johansen's. By transformation invariance of maximum likelihood estimators, it follows that $(\hat{\overline{\Lambda}}_1,\hat{\Lambda}_2)$ is the maximum likelihood estimator of $(\overline{\Lambda}_1,\Lambda_2)$, which is identical to $(\overline{\Gamma},\Phi)$ under Assumptions 4.1. It follows from Johansen (1978) that $\hat{\Lambda}_2$ is exactly the Nelson-Aalen estimator $\hat{\Phi}$ defined in Section 2, and therefore, by inversion of time, $\hat{\overline{\Lambda}}_1$ is exactly $\hat{\overline{\Gamma}}$. This shows that $(\hat{\overline{\Gamma}},\hat{\Phi})$ is indeed the nonparametric maximum likelihood estimator of $(\overline{\Gamma},\Phi)$.

Further, it is seen that $\hat{\overline{\Gamma}}$ and $\hat{\Phi}$ are of the form specified in the beginning of this section. Provided they contain <u>no inner jumps of size</u> 1, they may be interpreted as integrated intensities of probability distributions, and the corresponding product-limit estimators $(\hat{G},\hat{S}_X)$ are therefore NPMLE of (G,S_X) in the smaller "left truncation" model.

To show that <u>an NPMLE in the left truncation model exists</u> if and only if there are no inner jumps of size 1 in the NPMLE for the Markov model we now only need to remark that if there are inner jumps of size 1, one can make the ("discrete") likelihood function in the left truncation model arbitrarily close to the maximum likelihood in the Markov model, without however being able to achieve this value.

4b. Remarks on Efficiency.

The identification of the left truncation model as a Markov process model may be exploited to prove asymptotic efficiency results, and we shall in this section briefly indicate how to proceed. (Direct calculations of the efficiency of the product limit estimator under random truncation, defined on a compact subinterval ($[\varepsilon,M]$ in our notation) were recently given by Huang and Tsai (1986).)

Since the Nelson-Aalen estimator in a univariate counting process both has a nice asymptotic distribution theory and is an NPMLE, one should expect it also to be efficient in the sense of the modern extensions of the Hajek-LeCam convolution and asymptotic minimax theorems to semi-parametric models; see e.g. Begun, Hall, Huang and Wellner (1983). A result of such a nature by Helgeland & Hjort (to appear) was announced by Hjort in

§5 of his discussion of Andersen and Borgan (1985). Hjort only mentioned a univariate result. However in a multivariate counting process model, and in particular Aalen and Johansen's (1978) Markov process model, a multivariate version of this theorem can be proved in just the same way, since each intensity λ_i contributes an additive term to the log likelihood of exactly the same form.

It should also be emphasized that this model is one in which the usual locally asymptotic normality and information calculations are very simple, since the likelihood

$$\prod_i \left[\prod_t (\log \lambda_i(t))^{dN_i(t)} \exp(-\int V_i \lambda_i)\right]$$

is so simple too.

However, the problems of empty risk sets near 0 and ∞ which arise in establishing asymptotic normality also give serious technical difficulties in an efficiency theory.

We will take an easy way out as already introduced at the end of Section 3. Suppose we only consider the data collected on a time interval $[\varepsilon, M]$, that is, times of transitions and numbers at risk. The distribution of this data has parameters $v_1(\varepsilon)$, $v_2(\varepsilon)$, Λ_1^ε and Λ_2^ε, from which Φ^ε, $\bar{\Gamma}^M$ and hence G^M, S_X^ε can be calculated in the usual way. These parameters vary freely in $(0,1)^2 \times$ {increasing absolutely continuous functions on $[\varepsilon,M]$, zero at $\varepsilon\}^2$ by our results on equivalence of left-truncation and Markov models. Now by the natural extension of Helgeland & Hjort's efficiency results,

$$(V_1(\varepsilon)/n, V_2(\varepsilon)/n, \hat{\Lambda}_1^{\varepsilon}, \hat{\Lambda}_2^{\varepsilon})$$

is a jointly asymptotically efficient estimator of $(v_1(\varepsilon), v_2(\varepsilon), \Lambda_1^{\varepsilon}, \Lambda_2^{\varepsilon})$. Furthermore G^M and S_X^{ε} (as indicated in Section 3) are compactly differentiable functions of these parameters, and therefore by van der Vaart (1987; Theorem 4.1) also jointly efficiently estimated by the same functions of the efficient estimators: precisely $(\hat{G}^M, \hat{S}_X^{\varepsilon})$.

5. TWO RELATED ESTIMATION PROBLEMS

5a. Noninformative Truncation

In the previous sections of this note we have stuck to the random left truncation model as discussed by Woodroofe (1985) and postulating the existence of a pair of latent independent random variables (Y,X) from which only the conditional distribution given $Y<X$ is observable. In particular in biostatistical applications a certain uneasiness is prevalent about the assumption of such latent variables, in much similar fashion as for the competing-risk version of the random right censoring model, where a pair of independent random variables (X,Y) is postulated as underlying the observation of $(\min(X,Y), I\{X\leq Y\})$.

A recent note by Wellek (1986) made this specific by considering a random variable T "of interest", left-truncated by the "truncation mechanism", that is random variable, S, meaning that T is only observed if $T>S$. Wellek defined S as a noninformative truncation mechanism if for $0\leq s\leq t$

$$P\{T>t|S=s\} = P\{T>t|S=0\}/P\{T>s|S=0\}.$$

Define a stochastic process U on $[0,\infty)$ by

$$U(t) = 0 \quad \text{if} \quad t \leq S$$
$$U(t) = 1 \quad \text{if} \quad S < t \leq T$$
$$U(t) = 2 \quad \text{if} \quad T < t .$$

Then U is a Markov process if and only if S is a noninformative truncation mechanism, and is in fact exactly the model of this paper.

Verbal versions of the concept of a noninformative truncation mechanism were given by Aalen (1980):

"One has to require that the patients who enter the study after time zero are homogeneous with those who have been in the study from the beginning. In the case of a survival study this means that the mortality of patients who enter the study after time zero must be the same as that of the patients already in the study. In other words the fact that the entry is delayed must not be related to the development of the disease."

and by Hyde (1980):

"In the analysis that follows we will be assuming that if we know the age of a person who entered Channing House, then knowing the person's entry age will provide no additional information about prospects for survival".

5b. An Estimation Problem of Winter and Földes

Recently Winter and Földes (1986) considered the following estimation problem. Consider n independent renewal processes in equilibrium with underlying distribution function H, which we shall assume absolutely continuous with density h and support $[0,\infty)$. Corresponding to a fixed time, say 0, the forward and backward recurrence times S_i and R_i are observed; then $Q_i = R_i + S_i$ is a length-biased observation corresponding to the distribution function H. We quote the following distributional results: let χ be the expectation of H,

$$\chi = \int_0^\infty [1-H(u)]du,$$

then the joint distribution of (R,S) has density $\chi^{-1}h(r+s)$, the marginal distributions of R and S are equal with density $\chi^{-1}[1-H(r)]$, and the marginal distribution of $Q = R + S$ has density $\chi^{-1}qh(q)$, the length-biased density corresponding to h.

Winter and Földes considered (a slight modification of) the ordinary product-limit estimator based on the forward recurrence times $S_1,\ldots,S_n$ and showed that it is strongly consistent for the underlying survivor function $1-H$. We shall demonstrate how the derivation of this estimator follows immediately from the Markov process framework considered in this note.

First notice that the conditional distribution of $Q = R + S$ given that $R = r$ has density

$$\frac{\chi^{-1}h(q)}{\chi^{-1}[1-H(r)]}, \quad r \leqq q < \infty$$

that is, intensity (hazard) $h(q)/[1-H(q)]$, which is just the hazard corresponding to the underlying distribution H. Now define for each i (the i is suppressed in the notation) a stochastic process U on $[0,\infty)$ with state space $\{0,1,2\}$ by

$$U(t) = \begin{cases} 0, & 0 \leqq t < R \\ 1, & R \leqq t < R + S \\ 2, & R + S \leqq t\,. \end{cases}$$

We have

$$P\{U(t+h)=2 \mid U(u), 0 \leqq u \leqq t\} = o(h)$$

for $U(t)=0$, and for $U(t)=1$ (that is, $R \leqq t < R+S$) this is

$$P\{R+S \leqq t+h \mid R, R+S>t\} = \frac{h(t)}{1-H(t)}\, h+o(h)$$

by the above result on the hazard of $R+S \mid R$. That this depends only on t but not on R proves that U is a Markov process

$$\boxed{0} \xrightarrow{\alpha} \boxed{1} \xrightarrow{\beta} \boxed{2}$$

with intensities

$$\alpha(t) = [1-H(t)]/\int_t^\infty [1-H(r)]dr$$

(the marginal hazard of R, equal to the residual mean lifetime function of the underlying distribution H) and

$$\beta(t) = h(t)/[1-H(t)].$$

The Markov process framework of Section 2 indicates that the Nelson-Aalen and product limit estimators based on $S_1,\ldots,S_n$ are natural estimators of the integrated intensity $B(t) = \int_0^t \beta(q)dq$ respectively the survivor function $1-H$ of the underlying distribution, and consistency and asymptotic normality may be obtained as shown in Section 3.

Note that the backwards intensity

$$\bar{\alpha}(t) = \alpha(t)\,\frac{P\{U(t)=0\}}{P\{U(t)=1\}}$$

$$= \alpha(t)\,\frac{P\{R>t\}}{P\{R\leq t<R+S\}}$$

$$= \alpha(t) \frac{\chi^{-1} \int_t^\infty [1-H(r)]dr}{\chi^{-1} \int_0^t \int_{t-r}^\infty h(r+s)dsdr}$$

$$= \frac{1-H(t)}{\int_t^\infty [1-H(r)]dr} \quad \frac{\int_t^\infty [1-H(r)]dr}{\int_0^t [1-H(t)]dr} = \frac{1}{t} ,$$

the intensity of a uniform distribution on some interval $[0,A]$. Since it has been assumed that R has support $[0,\infty)$, this shows that the present model may not be interpreted as a left truncation model, which would require that $\bar{\alpha}(t)$ corresponded to a probability distribution on $[0,\infty)$.

The fact that $\bar{\alpha}(t)$ is uniform corresponds to Winter and Földes' statement that (R,S) contain no more information than $R+S$ about H. This might already have been gleaned from the likelihood function based on observation of $(R_1,S_1),\ldots,(R_n,S_n)$, which is

$$\chi^{-n} \prod_{i=1}^{n} h(r_i+s_i)$$

from which the NPMLE of H is readily derived as

$$\hat{H}(t) = \sum_{i=1}^{n} \frac{I\{R_i+S_i \leq t\}}{R_i+S_i} \Big/ \sum_{i=1}^{n} \frac{1}{R_i+S_i} ,$$

that is the Cox-Vardi estimator in the terminology of Winter and Földes (Cox 1969, Vardi 1985).

It follows that the estimators based on the forward recurrence times $S_1,\dots,S_n$ are not NPMLE. The difference between the situation here and that of Section 3 is that not only the intensity $\beta(t)$, but also $\alpha(t)$ depends only on the estimand H. In Section 3 λ_1 depended on both parameters γ and ϕ in such a way that even when ϕ was fixed, λ_1 could vary freely by varying γ.

5c. Asymptotic Results for the Winter-Földes Estimation Problem

Weak convergence of the Winter-Földes estimator is immediate from our results in Section 3. In particular, in order to achieve the extension to convergence on $[0,M]$ it should be required that

$$\int_0^{\varepsilon} [\phi(s)/v_2(s)]\, ds < \infty$$

in the terminology of Section 3c, and using $\phi(t) = \beta(t)$ and

$$v_2(t) = P\{U(t)=1\}$$

$$= \int_0^t \frac{1-H(s)}{X} \frac{1-H(t)}{1-H(s)} ds$$

$$= \frac{t}{X} [1-H(t)] \quad ,$$

the integrability condition translates into

$$\int_0^\varepsilon t^{-1} h(t)dt < \infty \; ,$$

or finiteness of $E(X^{-1})$ where X has the underlying ("length-unbiased") interarrival time distribution H. It may easily be seen from Gill and Wellner (1986) that the same condition is needed to ensure weak convergence of the Cox-Vardi estimator.

REFERENCES

AALEN, O.O. (1975), "Statistical Inference for a Family of Counting Processes," Ph.D. Dissertation, Department of Statistics, University of California, Berkeley.

AALEN, O.O. (1976), "Nonparametric Inference in Connection with Multiple Decrement Models," Scandinavian Journal of Statistics, 3, 15-27.

AALEN, O.O. (1978), "Nonparametric Inference for a Family of Counting Processes," Annals of Statistics, 6, 701-726.

AALEN, O.O. (1980), "Evaluation of Prognosis," Danish Medical Bulletin, 27, 167-168.

AALEN, O.O., BORGAN, Ø., KEIDING, N., and THORMANN, J. (1980), "Interaction Between Life History Events: Nonparametric Analysis of Prospective and Retrospective Data in the Presence of Censoring," Scandinavian Journal of Statistics, 7, 161-171.

AALEN, O.O. and JOHANSEN, S. (1978), "An Empirical Transition Matrix for Nonhomogeneous Markov Chains Based on Censored Observations," Scandinavian Journal of Statistics, 5, 141-150.

ANDERSEN, P.K. and BORGAN, Ø. (1985), "Counting Process Models for Life History Data: A Review" (with discussion), Scandinavian Journal of Statistics, 12, 97-158.

BEGUN, J.M., HALL, W.J., HUANG, W.M. and WELLNER, J.A. (1983), "Information and Asymptotic Efficiency in Parametric-Nonparametric Models," Annals of Statistics, 11, 432-452.

BILLINGSLEY, P. (1968), "Convergence of Probability Measures," New York: Wiley.

COX, D.R. (1969), "Some Sampling Problems in Technology," in New Developments in Survey Sampling, eds. N.L. Johnson and H. Smith, Jr. New York: Wiley, pp. 506-527.

COX, D.R. and OAKES, D. (1984), "Analysis of Survival Data," London: Chapman and Hall.

DECKERT, T., POULSEN, J.E. and LARSEN, M. (1978), "Prognosis of Diabetics with Diabetes Onset before the Age of Thirtyone," Diabetologia, 14, 363-370.

GILL, R.D. (1980), "Censoring and Stochastic Integrals," Mathematical Centre Tracts, No. 124, Amsterdam: The Mathematical Centre.

GILL, R.D. (1983), "Large Sample Behavior of the Product-Limit Estimator on the Whole Line," Annals of Statistics, 11, 49-58.

GILL, R.D. (1986), "Non- and Semi-Parametric Maximum Likelihood Estimators and the von Mises Method (part I)," Report MS-R8604, Centre for Mathematics and Computer Science, Department of Mathematical Statistics.

GILL, R.D. and JOHANSEN, S. (1987), "Product-Integrals and Counting Processes," (in preparation).

GILL, R.D. and WELLNER, J.A. (1986), "Large Sample Theory of Empirical Distributions in Biased Sampling Models," Technical Report MS-R8603, Centre for Mathematics and Computer Science, Department of Mathematical Statistics.

GREEN, A. and HOUGAARD, P. (1984), "Epidemiological Studies of Diabetes Mellitus in Denmark: 5. Mortality and Causes of Death among Insulin-Treated Diabetics," Diabetologia, 26, 190-194.

HALLEY, E. (1693), "An Estimate of the Degrees of the Mortality of Mankind Drawn from Curious Tables of the Births and Funerals at the City of Breslaw," Philosophical Transactions of the Royal Society of London, 17, 596-610.

HOEM, J.M. (1969), "Purged and Partial Markov Chains," Skandinavisk Aktuarietidskrift, 52, 147-155.

HUANG, W.-M. and TSAI, W.-Y. (1986), "Asymptotic Optimality of the Nonparametric Conditional Maximum Likelihood Estimator Under Random Truncation," Technical Report, Lehigh University and Brookhaven National Laboratory.

HYDE, J. (1977), "Testing Survival Under Right Censoring and Left Truncation," Biometrika, 64, 225-230.

HYDE, J. (1980), "Survival Analysis With Incomplete Observations," in Biostatistics Casebook, eds. R.G. Miller, Jr., B. Efron, B.W. Brown, Jr. and L.E. Moses, New York: Wiley, pp. 31-46.

JOHANSEN, S. (1978), "The Product Limit Estimator as Maximum Likelihood Estimator," Scandinavian Journal of Statistics, 5, 195-199.

KAPLAN, E.L. and MEIER, P. (1958), "Non-parametric Estimation From Incomplete Observations," Journal of the American Statistical Association, 53, 457-481.

KJØLLER, E., MORTENSEN, L.S. and LARSEN, M.E. (1978), "Evaluation of Prognosis," Danish Medical Bulletin, 25, 238-242.

LENGLART, E. (1977), "Relation de Domination Entre Deux Processus," Annales de l'Institut Henri Poincaré, 13, 171-179.

LYNDEN-BELL, D. (1971), "A Method of Allowing for Known Observational Selection in Small Samples Applied to 3CR Quasars," Monthly Notices of the Royal Astronomical Society, 155, 95-118.

REBOLLEDO, R. (1980), "Central Limit Theorems for Local Martingales," Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 51, 269-286.

REEDS, J.A., III (1976), "On the Definition of von Mises Functionals," Ph.D. Thesis, Department of Statistics, Harvard University.

VAN DER VAART, A. (1987), "Statistical Estimation in Models With Large Parameter Space," Ph.D. Thesis, University of Leiden; to appear as CWI Tract, Center for Mathematics and Computer Science, Amsterdam.

VARDI, Y. (1985), "Empirical Distributions in Selection Bias Models," Annals of Statistics, 13, 178-203.

WANG, M.-C., JEWELL, N.P. and TSAI, W.-Y. (1986), "Asymptotic Properties of the Product Limit Estimate under Random Truncation," Annals of Statistics, 14 (to appear).

WELLEK, S. (1986), "A Nonparametric Model for Product-Limit Estimation Under Right Censoring and Left Truncation," Technical Report, Institut für Medizinische Statistik und Dokumentation der Universität Mainz.

WELLNER, J.A. (1978), "Limit Theory for the Ratio of the Empirical Distribution Function to the True Distribution Function," Zeitschrift für Wahrscheinlickeitstheorie und Verwandte Gebiete, 45, 73-88.

WINTER, B.B. and FÖLDES, A. (1986), "Product-Limit Estimators for Use With Length-Biased Data," Technical Report, Department of Mathematics, University of Ottawa.

WOODROOFE, M. (1985), "Estimating a Distribution Function With Truncated Data," Annals of Statistics, 13, 163-177.

WOODROOFE, M. (1987), Correction to "Estimating a Distribution Function With Truncated Data," Annals of Statistics, 15 (to appear).