The Annals of Statistics
1990, Vol. 18, No. 2, 582–602

RANDOM TRUNCATION MODELS AND MARKOV PROCESSES

BY NIELS KEIDING AND RICHARD D. GILL[1]

University of Copenhagen and University of Utrecht

Random left truncation is modelled by the conditional distribution of the random variable X of interest, given that it is larger than the truncating random variable Y; usually X and Y are assumed independent. The present paper is based on a simple reparametrization of the left truncation model as a three-state Markov process. The derivation of a nonparametric estimator of a distribution function under random truncation is then 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 a 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. Although we do not carry through these here, we note that the present framework also allows several generalizations: *censoring* may be incorporated; the independence hypothesis underlying the truncation models may be tested; ties (occurring when the distributions of F and G have discrete components) may be handled.

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. Kaplan and Meier also showed that their "product-limit" estimator was the method of choice under delayed *entry*, or left truncation. Although this portion of their paper has escaped the attention of many later authors, the *practical* use of life table and product-limit methods under left truncation has flourished.

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).

A third apparently independent line of work on this estimator concerns estimation of the distribution of the residual in truncated regression [cf. Bhattacharya, Chernoff and Yang (1983), Tsui, Jewell and Wu (1988) and Bickel and Ritov (1987)].

Received April 1987; revised June 1989.

[1]This work was done while the author was at the Department of Mathematical Statistics, Centre for Mathematics and Computer Science (CWI), Amsterdam.

AMS 1980 *subject classifications*. Primary 62M05, 62G05; secondary 60G55, 60G44.

Key words and phrases. Counting processes, delayed entry, inference in stochastic processes, intensity function, left truncation, nonparametric maximum likelihood, product integral, survival analysis.

Following Woodroofe (1985), our basic setup is that of n i.i.d. replications of the conditional distribution, given $Y < X$, of a pair of independent random variables Y and X with distribution functions G and F, of which nonparametric estimators are sought.

The purpose of this paper is to demonstrate how an embedding of the basic nonparametric estimation problem into a simple Markov process model allows one to exploit the modern techniques of statistical inference in counting processes [Aalen (1978), Aalen and Johansen (1978) and Andersen and Borgan (1985)]. This approach sheds new light on a number of issues in the current literature and also paves the way for several new results.

Section 2 specifies a five-state Markov process and explores its equivalence with observation of (Y, X); some elementary results on conditioning and time-reversal in Markov processes are also recalled. Section 3 then applies the (by now standard) results on statistical inference for counting processes to the present estimation problem. Section 4 shows that the estimators of Section 3 may be interpreted as maximum likelihood estimators. Section 5 contains results on asymptotic normality and Section 6 applies these to obtain the asymptotic distribution of the maximum likelihood estimator of $\alpha = P\{Y \le X\}$. The final section, Section 7, briefly discusses some of the related results in the literature.

An extended version of this paper exists in technical report form (Keiding and Gill, 1988), where some points omitted here may be found.

2. Interpretation of random truncation models in a simple Markov process model. The problem is nonparametric estimation of the distributions G and F of independent, positive random variables Y and X when sampling from the conditional distribution of (X, Y) given $Y \le X$. Define the cumulative hazard functions

$$\Gamma(y) = \int_0^y dG(t)/[1 - G(t-)], \qquad \Phi(x) = \int_0^x dF(s)/[1 - F(s-)].$$

Let $a_G < b_G$ be the essential infimum and supremum of G so that (a_G, b_G) is the interior of the convex support of G; define a_F and b_F similarly.

We assume throughout that Y and X have no common atoms; in particular, $P\{Y = X\} = 0$, so that

$$\alpha = P\{Y \le X\} = P\{Y < X\},$$

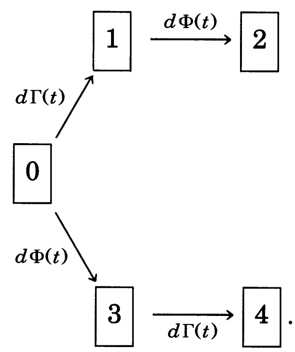
which is assumed to be positive; this is then equivalent to assuming $a_G < b_F$. We suppose also $a_G \le a_F$ and $b_G \le b_F$ to avoid mathematically trivial (though possibly practically rather important) identification problems.

Define a stochastic process $U = \{U(t),\ t \in [0, \infty]\}$ by

$$\begin{aligned} U(t) &= 0 \quad \text{when } t < X \wedge Y \\ U(t) &= 1 \quad \text{when } Y \le t < X \\ U(t) &= 2 \quad \text{when } Y < X \le t \\ U(t) &= 3 \quad \text{when } X \le t < Y \\ U(t) &= 4 \quad \text{when } X \le Y \le t. \end{aligned}$$

It is seen that observation of $\{U(t),\ 0 \le t < \infty\}$ is equivalent to that of (Y, X), and that observation of $\{U(t),\ 0 \le t < \infty\}$ conditional on $U(\infty) = 1$ is equivalent to observation of (Y, X) given $Y < X$.

PROPOSITION 2.1. *U is a Markov process with $U(0) \equiv 0$ and intensities given by the diagram*



In the conditional distribution given $U(\infty) = 2$ (that is, $Y < X$), U is again a Markov process given by the diagram

$$\boxed{0} \xrightarrow{d\Lambda_1(t)} \boxed{1} \xrightarrow{d\Lambda_2(t)} \boxed{2},$$

where $\Lambda_2 = \Phi$, whereas

$$d\Lambda_1(t) = d\Gamma(t)\frac{P_{12}(t,\infty)}{P_{02}(t-,\infty)} = \frac{d\Gamma(t)}{P\{Y < X | X \ge t,\ Y \ge t\}},$$

where $P_{ij}(t, u)$ are the transition probabilities in the original Markov process.

PROOF. Using product-integral formalism, Johansen (1978, 1987) defined finite-state, nonhomogeneous Markov processes from general (not necessarily continuous) intensity measures. That the conditional process given $U(\infty) = 2$ is Markov with the stated intensity measures is well known and easily seen by direct calculation. □

Consider now the time-reversed conditional Markov process $U(t)$ on $[0, \infty]$ with time running backward and $U(\infty) \equiv 2$:

$$\boxed{0} \xleftarrow{d\bar{\Lambda}_1(t)} \boxed{1} \xleftarrow{d\bar{\Lambda}_2(t)} \boxed{2}.$$

The following proposition is a standard result in Markov processes and is easily proved directly.

PROPOSITION 2.2. *Consider a Markov process with states $\{0, 1, 2\}$ defined from intensity measures Λ_1 and Λ_2 as in Proposition 2.1. The intensities of the* *backward Markov process (the "backward intensities") are given by*

$$d\bar{\Lambda}_i(t) = d\Lambda_i(t)\frac{P_{2,i-1}(\infty, t-)}{P_{2,i}(\infty, t)}, \qquad i = 1,2.$$

Define the (left-continuous) *backward cumulative hazard*

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

then an easy calculation from Proposition 2.2 gives $d\bar{\Lambda}_1(t) = d\bar{\Gamma}(t)$, which, of course, also follows directly by symmetry of time.

For the maximum likelihood theory in Section 4, we need the characterization results below, informally formulated as follows: Given a Markov process with transition intensities $d\Lambda_1(t)$ from 0 to 1 and $d\Lambda_2(t)$ from 1 to 2, under which circumstances do there exist distribution functions G and F generating this in the above way, that is, such that $d\Gamma(t)/P\{Y < X | X \geq t, Y \geq t\} = d\Lambda_1(t)$ and $dF(t)/[1 - F(t-)] = d\Lambda_2(t)$, where Y and X are independent with distribution functions G and F?

As preparation, we consider arbitrary integrated intensity measures Λ on $[0,\infty]$; define the *minimal convex support* Σ (which is an open, half-open or closed interval) as the smallest convex set such that $\Lambda(\Sigma^c) = 0$. Define c to be a *termination point* of Λ if either $\Lambda(\{c\}) = 1$ or $\Lambda(c - \varepsilon, c] = \infty$ for all $\varepsilon > 0$, but not both. An intensity measure Λ on $(0,\infty)$ with minimal convex support with endpoints $a < b$ *corresponds to a probability measure* if and only if Λ is finite on $[a, b - \varepsilon)$ for all $\varepsilon > 0$ and it has one and only one termination point, which is the essential supremum b.

PROPOSITION 2.3. *Let $U = (U(s),\ 0 \leq s \leq \infty)$ be a Markov process with state space $\{0,1,2\}$, intensity measures Λ_i: $i - 1 \to i$, $i = 1,2$, all other transitions having zero intensity and $P\{U(0) = 0\} = P\{U(\infty) = 2\} = 1$. Define $\bar{\Lambda}_i$ (the backward intensity measure from i to $i-1$) by $d\bar{\Lambda}_i(t) = d\Lambda_i(t)P\{U(t-) = i-1\}/P\{U(t) = i\}$ and assume that Λ_1, Λ_2 as well as $\bar{\Lambda}_1$, $\bar{\Lambda}_2$ (with time running backward) correspond to probability measures.*

Then there exist distribution functions F and G given by

$$1 - F(x) = \mathscr{P}_0^x(1 - d\Lambda_2), \qquad G(y) = \mathscr{P}_{(y,\infty)}(1 - d\bar{\Lambda}_1),$$

such that the Markov process corresponds to the left truncation model specified by the conditional distribution of independent random variables Y and X on $(0,\infty)$ with distribution functions G and F, given $Y < X$. These are the unique G and F subject to $a_G \leq a_F$, $b_G \leq b_F$.

REMARK. Here and in the following, $\mathscr{P}$ denotes product integral [cf. Gill and Johansen (1989)].

PROOF. The conditions directly imply that F and G are well-defined distribution functions. We need to check that the construction in Proposition 2.1 of a Markov process from these F and G lead us back to the integrated intensities Λ_1 and Λ_2. This can be seen by a direct calculation which we omit here. □

2.1. *The Markov process parametrization* (Λ_1, Λ_2) *and the truncation model parametrization* (G, F). Note that, while $\Lambda_2 = \Phi$ corresponds to the distribution function $F_2 = F$, Λ_1 corresponds to the distribution function of Y given $Y < X$,

$$\begin{aligned} 1 - F_1(y) &= \mathop{\mathscr{P}}_0^y(1 - d\Lambda_1(u)) \\ &= \int_0^y [1 - F(s)]\, dG(s) \Big/ \int_0^\infty [1 - F(s)]\, dG(s) \qquad (2.1) \\ &= \alpha^{-1} \int_0^y [1 - F(s)]\, dG(s), \end{aligned}$$

and that G may also be recovered from F_1 and $F_2 = F$ by the inverse relation

$$G(y) = \alpha \int_0^y [1 - F_2(s)]^{-1}\, dF_1(s),$$

since

$$\alpha = \int_0^\infty [1 - F(s)]\, dG(s) = 1 \Big/ \int_0^\infty [1 - F_2(s)]^{-1}\, dF_1(s).$$

The key point of this paper is the interplay between these two alternative representations: the "random truncation model," specified by G and F, and the Markov process model, specified by F_1 and F_2, as well as the two routes from Λ_1 and Λ_2 to G—that via time reversal and that via α.

3. Estimation. In this section we assume the distributions G and F to be continuous with support $(0, \infty)$; then the corresponding integrated intensities Γ and Φ are also continuous. By Y^*, X^* we denote random variables with the conditional distribution of Y, X given $Y < X$, and $G^*, F^*, \Gamma^*, \Phi^*$ denote marginal distribution functions and integrated intensities in this distribution.

We assume that a sample of n independent identically distributed replications $(Y_1^*, X_1^*), \dots, (Y_n^*, X_n^*)$ of (Y^*, X^*) is observed. Corresponding to (Y_i^*, X_i^*), $i = 1, \dots, n$, we construct (conditional) Markov processes U_i as in Proposition 2.1, which yields the counting processes

$$N_1(t) = \#\{Y_i^* \le t\} = \#\{\text{jumps by } U_1, \dots, U_n \text{ from 0 to 1 in } [0, t]\},$$

$$N_2(t) = \#\{Y_i^* < X_i^* \le t\} = \#\{\text{jumps by } U_1, \dots, U_n \text{ from 1 to 2 in } [0, t]\}.$$

With respect to the self-exciting filtration, the bivariate counting process $N(t) = (N_1(t), N_2(t))$ has compensator $A(t) = (A_1(t), A_2(t))$ given by

$$A_1(t) = \int_0^t V_1(u)\, d\Lambda_1(u), \qquad A_2(t) = \int_0^t V_2(u)\, d\Phi(u),$$

where we have used the fact (Proposition 2.1) that $\Lambda_2 = \Phi$ and where

$$V_1(t) = \#\{Y_i^* \geq t\}, \qquad V_2(t) = \#\{Y_i^* < t \leq X_i^*\};$$

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

3.1. *Estimation of the distribution of X.* According to standard methodology for statistical analysis of counting processes [Aalen (1975), Section 5D, Aalen (1978) and Aalen and Johansen (1978)], we use as estimator of the integrated intensity $\Phi(t)$ the *Nelson–Aalen estimator*,

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

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

$$\tilde{\Phi}(t) = \int_0^t J_2(u)\, d\Phi(u),$$

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

$$\langle \hat{\Phi} - \tilde{\Phi} \rangle(t) = \int_0^t \frac{J_2(u)}{V_2(u)}\, d\Phi(u).$$

These properties imply the unbiasedness result

$$E(\hat{\Phi}(T)) = E(\tilde{\Phi}(T)) \tag{3.1}$$

for any stopping time T (both sides may be ∞) and suggest the estimator

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

or a Greenwood-formula modification, acknowledging the discrete nature of $\hat{\Phi}$,

$$\hat{\tau}_G(t) = \int_0^t \frac{J_2(u)(V_2(u) - 1)}{[V_2(u)]^3}\, dN_2(u),$$

of the mean squared error function $\tau(t) = E[\langle \hat{\Phi} - \tilde{\Phi} \rangle(t)]$. [For a detailed numerical comparison of $\hat{\tau}(t)$ and $\hat{\tau}_G(t)$ in the simple survival analysis situation, see Klein (1988).]

Let us take a concrete look at the process $J_2(u) = I\{V_2(u) > 0\}$. Since we have assumed that $\operatorname{ess\,inf} X = \operatorname{ess\,inf} Y = 0$, 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

$(U_1, Z_1], \dots, (U_k, Z_k]$, $Z_k < X^*_{(n)}$ (it certainly becomes 0 for $u > X^*_{(n)}$). The serious problem is that in this case of "empty inner risk sets," $\Delta\hat{\Phi}(U_i) = \Delta N_2(U_i)/V_2(U_i) = 1$, "using up" the probability mass in the middle of the observation interval.

The estimator of the distribution function F of X or (equivalently) its survival function $1 - F$, is [Aalen and Johansen (1978), Theorem 3.2] the product-limit (or generalized Kaplan–Meier) estimator

$$1 - \hat{F}(t) = \mathop{\mathscr{P}}_{[0,t]}[1 - d\hat{\Phi}(u)] = \prod_{i=1}^{n}[1 - I\{X_i^* \le t\}/V_2(X_i^*)].$$

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

$$1 - \tilde{F}(t) = \mathop{\mathscr{P}}_{[0,t]}[1 - d\tilde{\Phi}(u)],$$

we have that

$$\frac{1 - \hat{F}(t)}{1 - \tilde{F}(t)} - 1 = \int_0^t \frac{1 - \hat{F}(u-)}{1 - \tilde{F}(u)}\, d[\tilde{\Phi}(u) - \hat{\Phi}(u)]$$

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

$$\left\langle \frac{\{1 - \hat{F}\}}{\{1 - \tilde{F}\}} - 1 \right\rangle(t) = \int_0^t \left[\frac{1 - \hat{F}(u-)}{1 - \tilde{F}(u)}\right]^2 d\langle \hat{\Phi} - \tilde{\Phi}\rangle(u)$$

$$= \int_0^t \left[\frac{1 - \hat{F}(u-)}{1 - \tilde{F}(u)}\right]^2 \frac{J_2(u)}{V_2(u)}\, d\Phi(u).$$

Hence, for any bounded stopping time T we get

$$E\left[\frac{1 - \hat{F}(T)}{1 - \tilde{F}(T)}\right] = 1.$$

Taking into account the discrete nature of the estimator $1 - \hat{F}$, the squared variation of $(1 - \hat{F})/(1 - \tilde{F})$ may be estimated by

$$\int_0^t J_2(u)(V_2(u) - 1)V_2(u)^{-3}\, dN_2(u),$$

and it follows that a natural estimate of the covariance function of $1 - \hat{F}$ is given by Greenwood's formula [cf. Meier (1975) and see Klein (1988) for a Monte Carlo study]:

$$\widehat{\operatorname{cov}}(1 - \hat{F}(s), 1 - \hat{F}(t)) = \{1 - \hat{F}(s)\}\{1 - \hat{F}(t)\}$$
$$\times \int_0^{s \wedge t} J_2(u)[V_2(u)\{V_2(u) - 1\}]^{-1}\, dN_2(u).$$

Note that since $d\hat{\Phi}(U_i) = 1$, $i = 1, \dots, k + 1$, and, in particular, $d\hat{\Phi}(U_1) = 1$, the estimator $1 - \hat{F}(t) = 0$ for $t \ge U_1$. 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^*. We shall see in Section 4 that a formal nonparametric maximum likelihood estimator does not exist in this case.

3.2. *Estimation of the distribution of Y.* By reversing time, it is immediate that the backward integrated hazard $\bar{\Gamma}(t)$ may be estimated by a *backward Nelson–Aalen estimator*; similarly, $G(t)$ may be estimated by a generalized backward Kaplan–Meier estimator. (Care should be taken regarding left or right continuity, etc.)

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 estimating at all. In particular, there is no information in the sample on the distribution of Y on $[X_{(n)}^*, \infty)$.

Alternatively, one might start from the Markov process representation of Section 3.1. Then estimators of the integrated intensity Λ_1 and the corresponding distribution function F_1 are immediately given as

$$\hat{\Lambda}_1(y) = \int_0^y \frac{J_1(u)}{V_1(u)}\, dN_1(u) = \sum_{i=1}^n \frac{I\{Y_i^* \le y\}}{\#\{Y_i^* \ge y\}}$$

and the corresponding product-limit estimator $1 - \hat{F}_1$, where $\hat{F}_1$ is nothing but the empirical function of the Y_i^*. Since the martingales $\hat{\Lambda}_1 - \tilde{\Lambda}_1$ and $\hat{\Lambda}_2 - \tilde{\Lambda}_2$ are orthogonal by the general theory of statistical analysis of counting processes, we further have the important property of approximate independence of $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$; this property will be crucial for the asymptotic theory of Section 5.

Since $\hat{F}_1$ estimates

$$F_1(y) = \alpha^{-1} \int_0^y [1 - F(s)]\, dG(s), \tag{3.2}$$

one could then apply the inversion

$$\tilde{G}(y) = \tilde{\alpha} \int_0^y \{1 - \hat{F}(s)\}^{-1} d\hat{F}_1(s), \qquad \tilde{\alpha} = \left[\int_0^\infty \{1 - \hat{F}(s)\}^{-1} d\hat{F}_1(s)\right]^{-1}.$$

However it is not immediate that $\tilde{G}$ equals the simple product-limit estimator $\hat{G}$ of the time-reversal approach and that $\tilde{\alpha} = \hat{\alpha} = \int \hat{G}\, d\hat{F}$. This is, however, a direct consequence of the propositions on Markov processes of Section 2 and the transformation invariance of maximum likelihood estimators, which we will discuss in the next section.

4. Nonparametric maximum likelihood estimation of (G, F). The purpose of this section is to show that $(\hat{G}, \hat{F})$ is the nonparametric maximum likelihood estimator (NPMLE) of (G, F), and that this fact is a direct consequence of the embedding of the left truncation model into the Markov process model, for which results on NPMLE were provided by Johansen (1978). First we discuss the easier result that $\hat{F}$ is a conditional NPMLE of F given $Y_1^*, \dots, Y_n^*$.

4.1. *Conditional nonparametric maximum likelihood estimation of F given* $Y_1^*, \dots, Y_n^*$. As an introduction, consider the factorization of the ("full") likelihood,

$$\text{lik}(G, F) = \alpha^{-n} \prod_i dG(Y_i)\, dF(X_i),$$

into the marginal likelihood of (G, F) based on $(Y_1^*, \dots, Y_n^*)$ and the conditional likelihood of F given $(Y_1^*, \dots, Y_n^*)$,

$$\text{lik}(G, F) = \text{marg.lik}_{\mathbf{Y}^*}(G, F)\text{cond.lik}_{\mathbf{X}^*|\mathbf{Y}^*}(F),$$

where, in particular,

$$\text{cond.lik}_{\mathbf{X}^*|\mathbf{Y}^*}(F) = \prod_{i=1}^{n} \frac{dF(X_i^*)}{1 - F(Y_i^*)}.$$

As in the derivations of the NPMLE for censored data by Kaplan and Meier (1958) and Johansen (1978), it is seen that the candidates for the maximizer $\hat{F}$ of the conditional likelihood must have support $\subseteq \{X_1^*, \dots, X_n^*\}$, and for such F we have the simple combinatorial result

$$\begin{aligned}\text{cond.lik}(F) &= \frac{dF(X_{(1)}^*)}{\left[1 - F(X_{(1)}^* -)\right]^{V_2(X_{(1)}^*)}} \frac{dF(X_{(2)}^*)}{\left[1 - F(X_{(2)}^* -)\right]^{V_2(X_{(2)}^*) - V_2(X_{(1)}^*) + 1}} \\ &\quad \times \cdots \times \frac{dF(X_{(n)}^*)}{\left[1 - F(X_{(n)}^* -)\right]^{V_2(X_{(n)}^*) - V_2(X_{(n-1)}^*) + 1}} \\ &= \prod_{i=1}^{n} d\Phi(X_{(i)}^*)\left[1 - d\Phi\left(X_{(i)}^*\right)\right]^{V_2(X_{(i)}^*) - 1},\end{aligned}$$

using $dF(X_{(i)}) = F(X_{(i)}) - F(X_{(i-1)})$ and the definition $d\Phi = dF/(1 - F_-)$. Recall that a discrete intensity measure Ψ with support contained in n points $a_1 < \cdots < a_n$ corresponds to a probability measure if and only if $0 \le d\Psi(a_i) < 1$, $i = 1, \dots, n-1$, $d\Psi(a_n) = 1$. The maximization problem is then trivial: If and only if $V_2(X_{(i)}) > 1$, $i = 1, \dots, n-1$ (*no empty inner risk sets*), the solution exists and is given by

$$d\hat{\Phi}(X_i) = 1/V_2(X_i),$$

or exactly the Nelson–Aalen estimator $\hat{\Phi}$ of Φ. By transformation invariance of maximum likelihood estimators (the relevant transformation here being the product integral), it follows that the conditional NPMLE of the survivor function $1 - F = \mathscr{P}(1 - d\Phi)$ is the product-limit estimator $1 - \hat{F} = \mathscr{P}(1 - d\hat{\Phi})$ studied in Section 3.

4.2. $\hat{F}$ *is not always the unconditional MLE.* Turning to the full likelihood, the maximization has to be done over (G, F) jointly. Let us first remark that if G cannot vary freely, it is easily seen that the NPMLE $\tilde{F}$ of F may differ from $\hat{F}$. Indeed, Vardi (1985) showed that if G is known,

$$\tilde{F}(X^*_{(i)}) = \sum_{j=1}^{i} G(X^*_{(j)})^{-1} \Big/ \sum_{j=1}^{n} G(X^*_{(j)})^{-1}$$

(a "weighted empirical distribution function"); the asymptotic behavior of this estimator is a special case of Example 4.2 of Gill, Vardi and Wellner (1988). Wang (1989) generalized Vardi's analysis by noticing that if G varies across a parametric family $G = \{G_\theta : \theta \in \Theta\}$, then the NPMLE of F is obtained from $\tilde{F}$ by replacing G by $G_{\hat{\theta}}$, where $\hat{\theta}$ is the MLE derived from the conditional distribution of $\boldsymbol{Y}^*$ given $\boldsymbol{X}^*$. These results strongly suggest that the NPMLE of F in the full model may be similarly obtained by replacing G by $\hat{G}$ in $\tilde{F}$, and this approach was used by Wang (1987). In the next subsection we show that the independent parametrization provided by the Markov process representation of the left truncation model furnishes an immediate answer.

4.3. *NPMLE in the left truncation model.* We embed the class of all left truncation models corresponding to distribution functions G, F on $(0, \infty)$ with $dG\,dF \equiv 0$ in the class of conditional Markov process models as specified in Propositions 2.1 and 2.3. Now we consider estimation of Λ_1 and Λ_2 in this larger class (thus, with Λ_1 and Λ_2 not necessarily corresponding to distribution functions). NPML estimation in such models was studied by Johansen (1978). The likelihood is of the form

$$\prod_i \mathop{\mathscr{P}}_t d\Lambda_i(t)^{dN_i(t)}(1 - d\Lambda_i(t))^{V_i(t) - dN_i(t)}.$$

It follows directly that the NPMLE of Λ_1 and Λ_2 are given by the discrete cumulative intensity functions $\hat{\Lambda}_1$ and $\hat{\Lambda}_2$ as specified in Section 3.

Now from the Markov process

$$\boxed{0} \xrightarrow{\hat{\Lambda}_1} \boxed{1} \xrightarrow{\hat{\Lambda}_2} \boxed{2},$$

one may recover distribution functions F^0 and G^0 corresponding to a left truncation model, if and only if there is no inner jump of size 1 of $\hat{\Lambda}_i (i = 1, 2)$ —this could only happen for $\hat{\Lambda}_2$ because $\hat{\Lambda}_1$ corresponds to an ordinary empirical distribution, with jumps of size j^{-1}, $j = n, \dots, 1$, in that order. When F^0 and G^0 exist, they coincide with $\hat{F}$ and $\hat{G}$ by Proposition 2.3 applied to $\hat{\Lambda}_i$, $i = 1, 2$, and it furthermore follows from the transformation invariance of maximum likelihood estimators and the definition of backward intensities that $\tilde{G}$ of Section 3.2 equals $\hat{G}$.

Finally, to show that an NPMLE in the *left truncation* model does not exist if there are inner jumps of size 1 in the NPMLE for the *Markov* model we now only need to remark that one can then 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.

As a corollary of the NPMLE property of $1 - \hat{F}$ and $\hat{G}$ we remark that the NPMLE of

$$\alpha = P\{Y < X\} = P\{Y \le X\} = \int_0^\infty G(u)\, dF(u) = \int_0^\infty [1 - F(u)]\, dG(u)$$

is

$$\hat{\alpha} = \int_0^\infty \hat{G}(u)\, d\hat{F}(u) = 1 \Big/ \int_0^\infty \{1 - \hat{F}(u)\}^{-1}\, d\hat{F}_1(u).$$

5. Asymptotic results. In this section it is assumed throughout that the distributions of G and F are continuous with support $(0, \infty)$ and integrated hazards Γ and Φ.

As we have seen, in interpreting $\hat{\Phi}$ and $1 - \hat{F}$ 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 $\{t\colon V_2(t) > 0\}$. In deriving asymptotic distribution theory for our estimators, we shall similarly take care of the thin information in the tails by first only estimating

$$\Phi^\varepsilon = \Phi - \Phi(\varepsilon) \quad \text{and} \quad 1 - F^\varepsilon = (1 - F)/\{1 - F(\varepsilon)\}$$

on an interval $[\varepsilon, M]$, this is the content of Section 5.1.

In Section 5.2 we generalize the results to $[0, \infty]$, handling the delicate tightness problems near 0 and ∞ in a similar way as Gill (1983) [cf. Ying (1989)] did for product-limit estimators for right-censored data. These results are most conveniently derived in the original parametrization in terms of $1 - F$ and G with associated product limit estimators $1 - \hat{F}$ and $\hat{G}$.

In contrast, in order to obtain joint convergence results, it is advantageous to use the second parametrization in terms of the conditional Markov process integrated intensities Λ_1 and Λ_2 (with corresponding d.f.s. F_1 and $F_2 = F$) and associated estimators $\hat{\Lambda}_1, \hat{\Lambda}_2, \hat{F}_1, \hat{F}_2$. Because the martingales $\hat{\Lambda}_1 - \tilde{\Lambda}_1$ and $\hat{\Lambda}_2 - \tilde{\Lambda}_2$ are orthogonal and hence asymptotically independent, covariance calculations are easier. We briefly mention the joint convergence results in Section 5.3 and more fully demonstrate the techniques in Section 6 with a derivation of the asymptotic distribution of $\hat{\alpha}$.

5.1. *Convergence on* $[\varepsilon, M]$, $0 < \varepsilon < M < \infty$. Assume that $t = \varepsilon, M$ satisfy $P\{Y < t \le X | Y < X\} > 0$. Let $\hat{\Phi}^\varepsilon$, $\tilde{\Phi}^\varepsilon$, $\hat{F}^\varepsilon$ and $\tilde{F}^\varepsilon$ be defined similarly to Φ^ε and F^ε. Let us also write

$$v_2(t) = E\left[n^{-1} V_2(t)\right] = P\{Y_i^* < t \le X_i^*\} = P\{Y_i^* < t\} - P\{X_i^* < t\}$$

[$= C(t)$ in Woodroofe's (1985) notation]. We have

$$\begin{aligned} v_2(t) &= P\{Y < t \le X, Y < X\}/P\{Y < X\} = G(t)[1 - F(t)]/\alpha \\ &\ge G(\varepsilon)[1 - F(M)]/\alpha > 0 \end{aligned}$$

for $\varepsilon \le t \le M$ by the assumption that Y and X have support $(0, \infty)$.

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, $\tilde{\Phi}^\varepsilon \equiv \Phi^\varepsilon$ and $\tilde{F}^\varepsilon = F^\varepsilon$ on $[\varepsilon, M]$ for all sufficiently large n almost surely.

With these preparations made, consistency of $\hat{\Phi}^\varepsilon$ and $\hat{F}^\varepsilon$ as well as weak convergence of $n^{1/2}(\hat{\Phi}^\varepsilon - \Phi^\varepsilon)$ and/or of $n^{1/2}(\hat{F}^\varepsilon - F^\varepsilon)$ follow immediately from standard results on the Nelson–Aalen and the product-limit estimators in the counting process literature.

PROPOSITION 5.1. *We have*

$$\|\hat{\Phi}^\varepsilon - \Phi^\varepsilon\|_\varepsilon^M \to_P 0 \quad \text{and} \quad \|\hat{F}_X^\varepsilon - F_X\|_\varepsilon^M \to_P 0.$$

PROOF. Apply the inequality of Lenglart (1977) exactly as Gill (1980, 1983) [cf. Andersen and Borgan (1985), Appendix]. □

COROLLARY 5.1. *We have*

$$\|\hat{\Phi} - \Phi\|_0^M \to_P 0 \quad \text{and} \quad \|\hat{F} - F\|_0^M \to_P 0.$$

Corollary 5.1 is obtained easily from Proposition 5.1, using

$$E\{\|\hat{\Phi} - \Phi\|_0^\varepsilon\} \le 2\Phi(\varepsilon)$$

by (3.1).

COROLLARY 5.2. *The event of the existence of $s < t < u \le M$ with $J_2(s) = J_2(u) = 1$, $J_2(t) = 0$ ("empty inner risk sets") is asymptotically negligible.*

It is curious that the probabilistic result of Corollary 5.2 [obtained easily from Corollary 5.1, cf. Woodroofe (1985), page 172] is derived via the proof of consistency of a statistical estimator!

THEOREM 5.1. *Under the stated conditions,*

$$n^{1/2}(\hat{\Phi}^\varepsilon - \Phi^\varepsilon) \to_D W^\varepsilon \quad \text{in} \quad D[\varepsilon, M] \tag{5.1}$$

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

$$\operatorname{var} W^\varepsilon(t) = \int_\varepsilon^t \frac{1}{v_2(s)}\, d\Phi(s); \tag{5.2}$$

we also have

$$n^{1/2}(\hat{F}^\varepsilon - F^\varepsilon) \to_D (1 - F^\varepsilon)W^\varepsilon, \tag{5.3}$$

and $\hat{\Phi}^\varepsilon$, $\hat{F}^\varepsilon$ *converge jointly. Furthermore,* $\int_\varepsilon^{(\cdot)} nV_2(s)^{-2}\, dN_2(s)$ *is a consistent estimator* (*in* $\|\cdot\|_\varepsilon^M$) *of the variance function of* W^ε.

PROOF. We use Rebolledo's (1980) version of the martingale central limit theorem and note that the verification of Rebolledo's conditions is direct in the approach used here, establishing Glivenko–Cantelli convergence for V_2/n. □

5.2. *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 (5.3) at the right-hand endpoint of the time interval was partially carried out by Gill (1983) for the random censorship model under the natural finiteness condition on the asymptotic variance, see Ying (1989) for the final result. The analogous condition in the left truncation model is empty. We shall here illustrate the use of these techniques to study the left-hand endpoint problem.

Since $\hat{S}_X$ and S_X are both close to 1 near $t = 0$, one easily discovers that the extension problem for (5.3) is hardly more difficult than that for (5.1), 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. (5.3)] that

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

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

$$\int_0^\varepsilon G(s)^{-1}\, dF(s) < \infty \tag{5.4}$$

for some (and then all) $\varepsilon > 0$. From now on we assume (5.4) holds. We will have our required result.

$$\begin{aligned} n^{1/2}(\hat{\Phi} - \Phi) &\to_D W && \text{in } D[0,\infty), \\ n^{1/2}(\hat{F} - F) &\to_D (1 - F)W && \text{in } D[0,\infty], \end{aligned} \tag{5.5}$$

where W is W^ε with $\varepsilon = 0$ of Theorem 5.1, if for all $\delta > 0$,

$$\lim_{\varepsilon \downarrow 0} \limsup_{n\to\infty} P\{n^{1/2}\|\tilde{\Phi} - \Phi\|_0^\varepsilon > \delta\} = 0, \tag{5.6}$$

and for all $\delta > 0$,

$$\lim_{\varepsilon \downarrow 0} \limsup_{n\to\infty} P\{n^{1/2}\|\hat{\Phi} - \tilde{\Phi}\|_0^\varepsilon > \delta\} = 0; \tag{5.7}$$

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 (5.7) first.

Since $\hat{\Phi} - \tilde{\Phi}$ is a square integrable martingale, Lenglart's (1977) inequality gives us

$$P\{n^{1/2}\|\hat{\Phi} - \tilde{\Phi}\|_0^\varepsilon > \delta\} \le \eta + P\left\{n\langle\hat{\Phi} - \tilde{\Phi}\rangle(\varepsilon) > \frac{\delta^2}{\eta}\right\}.$$

But

$$n\langle\hat{\Phi} - \tilde{\Phi}\rangle(\varepsilon) = \int_0^\varepsilon \frac{nJ_2(s)}{V_2(s)}\,d\Phi(s) \le 2\int_0^\varepsilon \frac{n+1}{V_2(s)+1}\,d\Phi(s).$$

Since $V_2(s)$ is binomially distributed, $E[(n+1)/\{V_2(s)+1\}] \le P\{Y_i^* < s \le X_i^*\}^{-1}$, so

$$E(n\langle\hat{\Phi} - \tilde{\Phi}\rangle(\varepsilon)) \le 2\alpha\int_0^\varepsilon \frac{d\Phi(s)}{G(s)(1-F(s))} \le \frac{2\alpha}{(1-F(\varepsilon_0))^2}\int_0^\varepsilon \frac{dF(s)}{G(s)}$$

for all $\varepsilon \le \varepsilon_0$. Thus, having assumed $\int_0^\varepsilon dF(s)/G(s) < (\infty)$, we can prove the required result: By Chebyshev's inequality, taking ε arbitrarily small, we can bound $n\langle\hat{\Phi} - \tilde{\Phi}\rangle_\varepsilon$ by an arbitrarily small constant with probability arbitrarily close to 1, uniformly in n; and this establishes (5.7).

As far as (5.6) is concerned, we note that

$$n^{1/2}\|\tilde{\Phi} - \Phi\|_0^\varepsilon = n^{1/2}\int_0^\varepsilon (1 - J_2(s))\,d\Phi(s) = n^{1/2}\Phi(Y_{(1)}^*)$$

with probability $\to 1$ as $n \to \infty$ by Corollary 5.2. It suffices therefore to show $n^{1/2}\Phi(Y_{(1)}^*) \to_P 0$ as $n \to \infty$, which may be proved by a direct calculation (omitted here) using that $\Lambda_1(Y_{(1)}^*)$ is the minimum of n i.i.d. exponential(1) random variables.

5.3. *Joint weak convergence.* By symmetry we can immediately write down weak convergence theorems for $n^{1/2}(\hat{\bar{\Gamma}}^M - \bar{\Gamma}^M)$ and/or $n^{1/2}(\hat{G}^M - G^M)$. To get weak convergence on the whole line we must extend condition (5.4) to

$$\int_0^\infty (1-F)^{-1}\,dG < \infty, \qquad \int_0^\infty G^{-1}\,dF < \infty. \tag{5.8}$$

Joint weak convergence may be proved using compact differentiation [cf. Gill (1989)] and exploiting the functional relationships between Λ_1, Λ_2 on the one hand and F, G on the other, taking special care all the time with the tail problems at $t = 0$ and $t = \infty$. (By transformation invariance of maximum likelihood estimators, the same relationships hold between the estimators.)

Since $\hat{G}$ and $\hat{F}$ will be dependent, the identification of the covariance structure is (as already mentioned) more conveniently based upon the orthogonal and hence asymptotically independent martingales $\hat{\Lambda}_1 - \tilde{\Lambda}_1$ and $\hat{\Lambda}_2 - \tilde{\Lambda}_2$ from the counting process approach.

We shall not carry through this program in detail here, but rather illustrate the techniques by a study of the asymptotic properties of the estimator $\hat{\alpha}$ of $\alpha = P\{Y \le X\}$, in the next section. Woodroofe [(1985), Corollary 4] proved consistency of $\hat{\alpha}$.

6. Asymptotic distribution of $\hat{\alpha}$.

THEOREM 6.1. *Suppose F and G are continuous with common interval of support* $(0, \infty)$. *Then* $\sqrt{n}(\hat{\alpha} - \alpha) \to_D N(0, \sigma^2)$ *with*

$$\sigma^2 = \alpha^3 \int_0^\infty \frac{dG}{1 - F} - \alpha^2 + \alpha^3 \int_0^\infty \left(\frac{1 - G}{1 - F}\right)^2 \frac{dF}{G}$$
$$= \alpha^3 \int_0^\infty \frac{dF}{G} - \alpha^2 + \alpha^3 \int_0^\infty \left(\frac{F}{G}\right)^2 \frac{dG}{1 - F},$$

which is finite if and only if (5.8) *holds.*

PROOF. Suppose (5.8) holds. Since $\sqrt{n}(\hat{F} - F)$ and $\sqrt{n}(\hat{G} - G)$ are each tight in $D[0, \infty]$ with weak limits in $C[0, \infty]$, the same holds for the pair jointly. Any subsequence from $n = 1, 2, \dots$ therefore contains a subsubsequence along which the pair converges weakly to a pair of continuous processes. By the representation $\alpha = \int_0^\infty G\,dF$ and the generalized δ-method [Gill (1989), Theorem 3 and Lemma 3 and Remark] $\sqrt{n}(\hat{\alpha} - \alpha)$ converges in distribution to a finite random variable along such a subsequence. We show that the limit is the same whatever the subsubsequence, thereby proving weak convergence of $\sqrt{n}(\hat{\alpha} - \alpha)$ in general. We split the integral

$$\alpha^{-1} = \int_0^\infty (1 - F_2)^{-1}\, dF_1$$

into an integral over $[0, M]$, to which the generalized δ-method can again be applied, and a remainder term, integrating over (M, ∞). For the remainder term, we have

$$\int_M^\infty (1 - F_2)^{-1}\, dF_1 = \int_M^\infty \frac{(1 - F)^{-1}(1 - F)\, dG}{\alpha} = \alpha^{-1}(1 - G(M))$$

and similarly for the estimators. Thus,

$$\sqrt{n}(\hat{\alpha}^{-1} - \alpha^{-1}) = \sqrt{n}\left[\int_0^M (1 - \hat{F}_2)^{-1}\, d\hat{F}_1 - \int_0^M (1 - F_2)^{-1}\, dF_1\right]$$
$$+ \sqrt{n}\left[\hat{\alpha}^{-1}(1 - \hat{G}(M)) - \alpha^{-1}(1 - G(M))\right]$$
$$= Z_{M,n} + R_{M,n}, \quad \text{say}.$$

Since $\sqrt{n}(\hat{\alpha} - \alpha)$ converges in distribution, $1 - G(M) \to 0$ as $M \to \infty$, and $\sqrt{n}(\hat{G} - G)$ converges in distribution to a *tied down* (Gaussian) process (the limiting process evaluated at time M converges almost surely to 0 as $M \to \infty$), we have easily

$$\lim_{M\uparrow\infty}\ \limsup_{n\to\infty} P(|R_{M,n}| > \varepsilon) = 0$$

for all $\varepsilon > 0$. By the generalized δ-method

$$(6.1)\quad Z_{M,n} = \int_0^M (1-F_2)^{-1}\, d(\sqrt{n}(\hat{F}_1 - F_1)) + \int_0^M \sqrt{n}\,\frac{\hat{F}_2 - F_2}{1-F_2}\,\frac{dF_1}{1-F_2} + o_P(1)$$

as $n\to\infty$, for each $M<\infty$. So $Z_{M,n}$ converges in distribution to a zero-mean normal variate whose variance can be calculated by replacing $\sqrt{n}(\hat{F}_1 - F_1)$ and $\sqrt{n}[(\hat{F}_2 - F_2)/(1-F_2)]$ in (6.1) by the limiting independent Gaussian processes described above. Therefore $Z_{\infty,n} = \sqrt{n}(\hat{\alpha}^{-1} - \alpha^{-1})$ converges in distribution to the normal variate, with necessarily finite variance, obtained by letting $M\to\infty$. Combining these two steps, we find that this variance is a sum of two variances σ_1^2 and σ_2^2 coming from the asymptotically independent terms in (6.1), namely,

$$\sigma_1^2 = \int_0^\infty (1-F_2)^{-2}\, dF_1 - \left(\int_0^\infty (1-F_2)^{-1}\, dF_1\right)^2$$

and

$$\sigma_2^2 = \int_{s=0}^\infty \int_{t=0}^\infty as\operatorname{cov}\left(\sqrt{n}\,\frac{\hat{F}_2(s) - F_2(s)}{1-F_2(s)}, \sqrt{n}\,\frac{\hat{F}_2(t) - F_2(t)}{1-F_2(t)}\right)\frac{F_1(ds)}{1-F_2(s)}\,\frac{F_1(dt)}{1-F_2(t)},$$

which must be finite under (5.8). The double integral is more conveniently evaluated as twice that over $\{0\le t<s<\infty\}$; also use $(1-F_2)^{-1}\,dF_1 = \alpha^{-1}\,dG$. Thus,

$$\sigma_1^2 = \int_0^\infty (1-F)^{-1}\alpha^{-1}\,dG - \alpha^{-2},$$

$$\begin{aligned}
\sigma_2^2 &= 2\iint_{0\le t<s<\infty} \int_{u=0}^{s\wedge t} \frac{F(du)}{\alpha^{-1}G(u)[1-F(u)]^2}\alpha^{-1}G(ds)\alpha^{-1}G(dt)\\
&= 2\alpha^{-1}\iiint_{0\le u<t<s<\infty} \frac{G(ds)G(dt)F(du)}{G(u)[1-F(u)]^2}\\
&= \alpha^{-1}\iint_{0\le u<t<\infty} \frac{2[1-G(t)]G(dt)F(du)}{G(u)[1-F(u)]^2}\\
&= \alpha^{-1}\int_{u=0}^\infty \frac{[1-G(u)]^2F(du)}{[1-F(u)]^2G(u)}.
\end{aligned}$$

Hence,

$$\sigma^2 = \alpha^4(\sigma_1^2 + \sigma_2^2) = \left(\alpha^3\int_0^\infty \frac{dG}{1-F} - \alpha^2\right) + \alpha^3\int_0^\infty \left(\frac{1-G}{1-F}\right)^2\frac{dF}{G}.$$

Note that the first term is positive by, e.g., Jensen's inequality applied to the G-expectation of $(1 - F)^{-1}$. Also note that $\sigma^2 < \infty$ implies $\int_{(0,\infty)}(1 - F)^{-1}\,dG < \infty$, trivially, and $\int_{(0,\infty)}G^{-1}\,dF < \infty$, too, since $(1 - G)/(1 - F)$ is close to 1 near 0, where $G^{-1} \to \infty$. The converse has already been established but is easy to check explicitly. The alternative expression for σ^2, under the assumed conditions, follows from symmetry or by a (rather tedious) exercise in integration by parts. □

6.1. *Chao's asymptotic results.* Chao (1987) [cf. Chao and Lo (1988)] conjectured asymptotic results for $\hat{F}$, $\hat{G}$ and $\hat{\alpha}$ using an influence function approach. By similar techniques as just demonstrated, it can be seen that the asymptotic covariance of $\sqrt{n}(\hat{G} - G)$ and $\sqrt{n}(\hat{F} - F)$ is indeed as given by Chao [(1987), formula (3.2)], except that α^{-1} should be replaced by α in that formula (twice). Chao's expression for the asymptotic variance of $\sqrt{n}(\hat{\alpha} - \alpha)$ differs from ours as shall be seen below.

6.2. *Numerical examples.* To illustrate some of the asymptotic results above, a number of Monte Carlo simulations were performed. A simple example of distributions G and F on $(0, \infty)$ and satisfying conditions (5.8) is G exponential, F gamma(3), from which one may derive $\alpha = 0.875$ and n times the variance of the approximate distribution of $\hat{\alpha}$ as 0.3500. [This result differs from that conjectured by Chao (1987), whose formula for this example yields 0.31.]

Table 1 contains summary data from 10,000 Monte Carlo simulations of n independent samples from the conditional distribution of (Y, X) given $Y < X$ for $n = 5$, 10, 20, 50, 100 and 800. [The random number generator RAN3 of Press, Flannery, Teukolsky and Vetterling (1986) was used on an Olivetti M24 personal computer.] Replications with empty inner risk sets were recorded but could not be included in the averages, which thus represent conditional values, given that there was no empty inner risk set.

TABLE 1

Results from 10,000 Monte Carlo replications of samples of size n from the conditional distribution $(Y, X|Y < X)$ with Y exponential, X gamma(3) and Y and X independent

Sample size, n	Frequency of replications with empty inner risk set	Mean, $\hat{\alpha}$	$n\,\mathrm{Var}(\hat{\alpha})$ (observed)
5	0.0364	0.9014	0.0988
10	0.0091	0.8770	0.1646
20	0.0030	0.8743	0.1869
50	0.0003	0.8744	0.1985
100	0.0003	0.8742	0.2127
800	0.0000	0.8748	0.2400
∞ (theoretical value)	0	0.875	0.3500

Note first that empty inner risk sets occur also for rather large sample size n.

The approximation of $\mathrm{Var}(\hat{\alpha})$ is rather poor, indicating a very slow approach to the limiting distribution in this particular example. It may be seen (not documented here) that the problem is primarily in the (right-hand) tail of the distribution and that the distribution of $\hat{\alpha}$ is heavily skewed to the left, as was to be expected from the restriction $\hat{\alpha} \leq 1$. It is interesting that by calculating the estimator of σ^2 suggested by Theorem 5.2 (just replacing F, G and α by their estimates), a strong negative correlation between $\hat{\alpha}$ and $\hat{\sigma}$ is revealed: The intuitive explanation being that the closer $\hat{\alpha}$ is to 1, the closer we are to full separation between the Y_i and the X_i, in which case $\alpha = P\{Y < X\}$ becomes apparently much easier to estimate. The estimator $\hat{\sigma}^2$ overcompensates for that feature to the extent that the distribution of $\sqrt{n}(\hat{\alpha} - \alpha)/\hat{\sigma}$ becomes skewed to the *right*, but now with about the correct, approximately unit, variance.

As F and G become more different, the approach to the limiting distribution of Theorem 5.2 becomes less slow. Thus, for the case G = exponential, F = gamma(5), one has $\alpha = 0.968706$, σ^2(approximate) $= 0.03255$; 8000 replications of sample size 500 gave no empty inner risk sets and an average $\hat{\alpha}$ of 0.968710, empirical $\sigma^2 = 0.03049$.

Finally, the theory of Section 5.1 did not require the integrability conditions (5.8). Appropriately modified limiting results for the functional

$$\xi_{\varepsilon, M} = \int_{\varepsilon}^{M} \frac{1 - F_2(\varepsilon)}{1 - F_2(x)}\, dF_1(x) = [1 - F(\varepsilon)]\alpha^{-1}[G(M) - G(\alpha)],$$

estimated by

$$\hat{\xi} = \int_{\varepsilon}^{M} \frac{1 - \hat{F}(\varepsilon)}{1 - \hat{F}(x)}\, d\hat{F}_1(x),$$

therefore hold also when (5.8) is violated. Thus, when $F = G$, similar simulation studies (details omitted here) show the approximation of the variance of the distribution to be good for even very small ε and M up to about 5, in the scale given by assuming $F = G$ exponential (1). Further simulation results for cases where (5.8) does not hold suggest that $\sqrt{n}(\hat{\alpha} - \alpha)/\hat{\sigma}(\hat{\alpha})$ can still be usefully applied for statistical inference, cf. the concept of "self norming sums" [LePage, Woodroofe and Zinn (1981)] for nonnormal limiting distributions [cf. Woodroofe (1985), Remark 8].

7. Concluding remarks. In this section we indicate some related work in the literature and suggest possible generalizations where the present framework will be useful.

Recent *biostatistical applications* of the theory of this paper include Keiding, Bayer and Watt-Boolsen (1987) and the AIDS incubation time distribution investigations by Lagakos, Barraj and DeGruttola (1988), Kalbfleisch and

Lawless (1989) and DeGruttola and Lagakos (1989). Similar Markov process models were used to model retrospective observation in epidemiology by Aalen, Borgan, Keiding and Thormann (1980) and Keiding (1986a, b), while Samuelsen (1989) used a Markov process model for double censoring. The empirical phenomenon of *delayed entry* (which here has been modelled using *truncation*, the conditional distribution given that we are beyond entry time) might alternatively be modelled using *filtering*, where the probability measure stays the same, although the individual is not identified until entry [cf. Andersen, Borgan, Gill and Keiding (1988)].

Testing of the *independence hypothesis* underlying the truncation model is naturally achieved by stratification according to entry time (and using any standard nonparametric test) or using entry time as covariate in, for example, a Cox regression model.

Ties require slight extensions in two contexts. In the modelling, a fifth stage is necessary to account for simultaneous occurrence of X and Y. In the martingale theory, the predictable variation process will get an extra term from the discrete component.

The *maximum likelihood results* were also studied by Wang, Jewell and Tsai (1986), who pointed out that the condition of no empty inner risk set forms a concrete example of the rather general Theorem 1 of Vardi (1985).

Asymptotic results for the estimators were studied, using classical methods, by Woodroofe (1985) plus a correction note (Woodroofe, 1987) which employed a martingale argument to deal with the tail problems. Further asymptotic studies were performed by Wang, Jewell and Tsai (1986), who, adding to Woodroofe, identified the covariance structure, and by Davidsen and Jacobsen (1989), who developed weak convergence theory for two-sided stochastic integrals. Although the counting process approach is very well suited for accommodating censoring [cf. e.g., Andersen, Borgan, Gill and Keiding (1988)] and therefore paves the way for extension of our approach to left truncated and right censored data, we do not carry through this program here. Tsai, Jewell and Wang (1987) gave some results in this direction, using a classical approach.

We finally mention that the NPMLE property of our estimators in conjunction with the asymptotic results allow proofs of *efficiency* results using powerful recent results by van der Vaart (1988a, b, 1990); for a sketch of this see Keiding and Gill (1988). The main idea is that joint efficiency of the empirical marginals $\hat{F}^*$ and $\hat{G}^*$ is easy to show directly; $\hat{F}$ and $\hat{G}$ are smooth functionals of this and compact differentiability and a general theorem of van der Vaart (1988a) on preservation of efficiency under differentiable mappings now do the rest.

Acknowledgments. The authors are grateful to W.-M. Huang, W.-Y. Tsai and M.-C. Wang for supplying copies of unpublished work, to Søren Johansen for important contributions, to Ian McKeague for a careful reading of a previous version and to Lars Sommer Hansen for help with some of the calculations.

REFERENCES

AALEN, O. O. (1975). Statistical inference for a family of counting processes. Ph.D. dissertation, Dept. Statist., Univ. California, Berkeley.

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

AALEN, O. O. and JOHANSEN, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scand. J. Statist.* **5** 141–150.

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. *Scand. J. Statist.* **7** 161–171.

ANDERSEN, P. K. and BORGAN, Ø. (1985). Counting process models for life history data: A review (with discussion). *Scand. J. Statist.* **12** 97–158.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. and KEIDING, N. (1988). Censoring, truncation and filtering in statistical models based on counting processes. *Contemporary Math.* **80** 19–60.

BHATTACHARYA, P. K., CHERNOFF, H. and YANG, S. S. (1983). Nonparametric estimation of the slope of a truncated regression. *Ann. Statist.* **11** 505–514.

BICKEL, P. J. and RITOV, J. (1987). Large sample theory of estimation in biased sampling regression models I. Technical Report 115, Dept. Statist., Univ. California, Berkeley.

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

CHAO, M.-T (1987). Influence curves for randomly truncated data. *Biometrika* **74** 426–429.

CHAO, M.-T. and LO, S.-H. (1988). Some representations of the nonparametric maximum likelihood estimators with truncated data. *Ann. Statist.* **16** 661–668.

DAVIDSEN, M. and JACOBSEN, M. (1989). Weak convergence of two-sided stochastic integrals, with an application to models for left truncated survival data. Preprint 2, Inst. Math. Statist., Univ. Copenhagen.

DEGRUTTOLA, V. and LAGAKOS, S. W. (1989). Analysis of doubly censored survival data, with applications to AIDS. *Biometrics* **45** 1–11.

GILL, R. D. (1980). Censoring and Stochastic Integrals. *Mathematical Centre Tracts* **124**. Mathematical Centre, Amsterdam.

GILL, R. D. (1983). Large sample behavior of the product-limit estimator on the whole line. *Ann. Statist.* **11** 49–58.

GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (part 1) (with discussion). *Scand. J. Statist.* **16** 97–128.

GILL, R. D. and JOHANSEN, S. (1989). A survey of product-integration with a view towards application in survival analysis. Technical Report, Univ. Utrecht and Univ. Copenhagen.

GILL, R. D., VARDI, Y. and WELLNER, J. A. (1988). Large sample theory of empirical distributions in biased sampling models. *Ann. Statist.* **16** 1069–1112.

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. *Philos. Trans. Roy. Soc. London* **17** 596–610. [Reprinted in *J. Inst. Act.* **112** (1985) 278–301.]

JOHANSEN, S. (1978). The product limit estimator as maximum likelihood estimator. *Scand. J. Statist.* **5** 195–199.

JOHANSEN, S. (1987). Product integrals and Markov processes. *CWI Newsletter* **12** 3–13. [Originally appeared (1977) as Preprint No. 3, Inst. Math. Statist., Univ. Copenhagen.]

KALBFLEISCH, J. D. and LAWLESS, J. F. (1989). Inference based on retrospective ascertainment: An analysis of the data on transfusion-related AIDS. *J. Amer. Statist. Assoc.* **84** 360–372.

KAPLAN, E. L. and MEIER, P. (1958). Non-parametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481.

KEIDING, N. (1986a). The epidemiological information in a cross-sectional sample: A statistical perspective. Research Report 86/3, Statist. Research Unit, Univ. Copenhagen.

KEIDING, N. (1986b). Discussion of D. R. Brillinger: The natural variability of vital rates and associated statistics. *Biometrics* **42** 721–724.

KEIDING, N. and GILL, R. D. (1988). Random truncation models and Markov processes. Report MS-R8817, Centre for Mathematics and Information Science, Amsterdam.

KEIDING, N., BAYER, T. and WATT-BOOLSEN, S. (1987). Confirmatory analysis of survival data using left truncation of the life times of primary survivors. *Statist. in Medicine* **6** 939–944.

KLEIN, J. P. (1988). Small sample properties of censored data estimators of the cumulative hazard rate, survivor function, and estimators of their variance. Research Report 88/7, Statist. Research Unit, Univ. Copenhagen.

LAGAKOS, S. W., BARRAJ, L. M. and DEGRUTTOLA, V. (1988). Nonparametric analysis of truncated survival data, with application to AIDS. *Biometrika* **75** 515–523.

LENGLART, E. (1977). Relation de domination entre deux processus. *Ann. Inst. H. Poincaré* **13** 171–179.

LEPAGE, R., WOODROOFE, M. and ZINN, J. (1981). Convergence to a stable distribution via order statistics. *Ann. Probab.* **9** 624–632.

LYNDEN-BELL, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices Roy. Astronom. Soc.* **155** 95–118.

MEIER, P. (1975). Estimation of a distribution function from incomplete observations. In *Perspectives in Probability and Statistics* (J. Gani, ed.) 67–87. Applied Probability Trust, Sheffield, England.

PRESS, H. W., FLANNERY, P. B., TEUKOLSKY, A. S. and VETTERLING, T. W. (1986). *Numerical Recipes*. Cambridge Univ. Press.

REBOLLEDO, R. (1980). Central limit theorems for local martingales. *Z. Wahrsch. Verw. Gebiete* **51** 269–286.

SAMUELSEN, S. O. (1989). Asymptotic theory for non-parametric estimators from doubly censored data. *Scand. J. Statist.* **16** 1–21.

TSAI, W.-Y., JEWELL, N. P. and WANG, M.-C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika* **74** 883–886.

TSUI, K.-L., JEWELL, N. P. and WU, C. F. J. (1988). A nonparametric approach to the truncated regression problem. *J. Amer. Statist. Assoc.* **83** 785–792.

VAN DER VAART, A. (1988a). *Statistical Estimation in Large Parameter Spaces. CWI Tract* **44**, Centre for Mathematics and Computer Science, Amsterdam.

VAN DER VAART, A. (1988b). Efficiency and Hadamard differentiability. Technical Report, Free Univ. Amsterdam.

VAN DER VAART, A. (1990). On differentiable functionals. *Ann. Statist.* To appear.

VARDI, Y. (1985). Empirical distributions in selection bias models. *Ann. Statist.* **13** 178–203.

WANG, M.-C. (1987). Product-limit estimates: A generalized maximum likelihood study. *Comm. Statist. Theory Methods* **16** 3117–3132.

WANG, M.-C. (1989). A semi-parametric model for randomly truncated data. *J. Amer. Statist. Assoc.* **84** 742–748.

WANG, M.-C., JEWELL, N. P. and TSAI, W.-Y. (1986). Asymptotic properties of the product limit estimate under random truncation. *Ann. Statist.* **14** 1597–1605.

WOODROOFE, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* **13** 163–177.

WOODROOFE, M. (1987). Correction to "Estimating a distribution function with truncated data." *Ann. Statist.* **15** 883.

YING, Z. (1989). A note on the asymptotic properties of the product-limit estimator on the whole line. *Statist. Probab. Lett.* **7** 311–314.

STATISTICAL RESEARCH UNIT
UNIVERSITY OF COPENHAGEN
BLEGDAMSVEJ 3
DK-2200 COPENHAGEN N
DENMARK

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF UTRECHT
POSTBUS 80.010
3508 TA UTRECHT
THE NETHERLANDS