Martin Jacobsen

Right Censoring and the Kaplan-Meier and Nelson-Aalen Estimators

Preprint
November
1986

6

Institute of Mathematical Statistics
University of Copenhagen

Martin Jacobsen

RIGHT CENSORING AND THE KAPLAN - MEIER AND NELSON - AALEN ESTIMATORS

Preprint 1986 No. 6

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

November 1986

Abstract.

Statistical models are considered for (partial) observation of independent, identically distributed failure times, subjected to censoring from the right. A number of conditions on the censoring pattern are studied, which allows the use of for example the Kaplan - Meier estimator as estimator of the survivor function for the failure time distribution. All the concitions and results presented are based on the counting process description of survival data.

AMS 1980 subject classification. Primary 62 G 05,
secondary 62 M 09, 60 G 55.

Key words and phrases. Survival analysis, counting process, intensity process, multiplicative intensity model, martingale estimator.

1. Introduction.

When Kaplan and Meier [9] introduced the product limit estimator for an unknown survivor function based on observation of a sample of failure times subjected to right censoring, they primarily had in mind the situation where the censoring times are stochastically independent of the failure times. They included however ([9], Section 3.2) a brief discussion of what may happen if this independence is not valid, and in particular they stress the dangers of using the product limit estimator in such cases.

Of course later a host of nonparametric models have been introduced, with dependence between failures and censorings, where it is still natural to use the Kaplan - Meier estimator or its twin, the Nelson - Aalen estimator, when estimating the integrated hazard rather than the survivor function (Nelson [11], Aalen [1]). Discussions of these models may be found in Kalbfleisch and Prentice ([8], Chapters 3 and 5) and Gill ([3], Chapter 3). The common structure pertaining to all these models is that the dependence between failures and censorings must be such that, phrased quite informally

(S) "past observations do not affect the probabilities of future failures"

where "observations" mean observed failures and observed censorings.

A breakthrough in the conception of models for censored survival data came with Aalen's [1] formulation in terms of counting processes and his demonstration that the classical models had the multiplicative intensity structure introduced by him. Using counting processes and their compensators (intensity processes), one is provided with an ideal tool for formulating rigorously what is meant by the informal statement (S) above. Gill [3] in particular used this framework for his study of censoring patterns, and it is also the backbone of the present paper.

Still, there is scope for numerous formulations of what must be the essential structure of the censoring mechanisms if one is to apply the Kaplan - Meier and Nelson - Aalen estimators. The main purpose of this paper is to present explicit conditions on the distributional properties of the censoring times themselves, assuming always that the failure times are independent and identically distributed. By way of comparison, the condition introduced by Gill [3] is more implicit and in the vein of requiring censorings to occur in such a way that the lifetime distributions for the population still at risk, are as they would be without any censoring.

In Section 2 we introduce the counting process setup used throughout the paper and with reference to Aalen's work [1], introduce the fundamental multiplicative intensity structure. Also here the questions concerning the structure of censoring patterns are posed, that are then answered in Section 3, which contains the main results of the paper. Naturally we are dealing with nonparametric models, with an arbitrary unknown failure time distribution. However, all results in Section 3 are really about the structure of each probability belonging to the model, rather than the model as a whole. In Section 4 we discuss models where not only for instance the Kaplan - Meier estimator can be used according to the criterion from Section 2, but where also this estimator exploits all essential information about the unknown failure time distribution. This discussion in Section 4 is related to the concept of information versus noninformative censorings (see [8], p. 121) and also the recent work of Arjas and Haara [2] on innovation versus noninnovation. Finally, Section 5 reviews briefly the asymptotic theory for the Nelson - Aalen and Kaplan - Meier estimators.

2. The counting process description of right censored data.

Consider the usual setup for observing independent and identically distributed failure times (lifetimes) subject to censoring from the right. More precisely, let $X_1,\ldots,X_n$ be i.i.d. and let $U_1,\ldots,U_n$ be the censoring times with all X_i, U_i strictly positive. Then the observations consist of the pairs

$$(T_i\,,\delta_i) \qquad (i=1,\ldots,n)$$

where $T_i = X_i \wedge U_i$ and $\delta_i = 1_{(X_i \leq U_i)}$ is an indicator showing whether T_i is the failure time X_i $(\delta_i = 1)$ or the censoring time U_i $(\delta_i = 0)$.

The statistical problem is to estimate the unknown distribution of the X_i on the basis of the observations (T_i,δ_i) alone.

For all censoring patterns to be considered in this paper, it will be assumed that the distribution of the observations must be compatible with the assumption that the X_i are independent and identically distributed. In doing this we are making at least some assumptions about the distribution of the unobserved lifetimes. On the other hand, the unobserved censoring times (corresponding to i with $\delta_i = 1$), we shall consider irrelevant, and to avoid any confusion about what they might or might not have been, we shall henceforth make the following assumption, which we list together with the above condition on the X_i:

(D) The failure times $X_1,\ldots,X_n$ and censoring times $U_1,\ldots,U_n$ are strictly positive, possibly infinite, random variables such that $U_i = \infty$ whenever $U_i \geq X_i$, and where the joint distribution of all X_i and U_i satisfies that the X_i are independent and identically distributed.

Note that with assumption D in force, we may write

$$(\delta_i = 1) = (U_i = \infty) .$$

From now on we shall assume the unknown distribution of the X_i to be absolutely continuous with unknown hazard μ , i.e.

$$P(X_i > t) = \exp(- \int_0^t \mu(s)ds) .$$

For simplicity we write G_μ or just G for this survivor function and $F_\mu = 1 - G_\mu$ for the distribution function. Also, for $s \leq t$ we write

$$G_\mu(t|s) = \frac{G_\mu(t)}{G_\mu(s)} = \exp(- \int_s^t \mu(u)du)$$

for the conditional survivor function $P(X_i > t \mid X_i > s)$ and denote by Q_μ or Q the joint distribution of $X_1,\ldots,X_n$:

$$Q_\mu(B) = P((X_1,\ldots,X_n) \in B) .$$

It will be convenient for us to assume that $\int_0^t \mu(u)du < \infty$ for all t . (This is not essential, but if $t^\dagger = \inf\{t : \int_0^t \mu = \infty\} < \infty$ much of what is said below, will be valid only on the timeinterval $[0,t^\dagger)$). We shall however not assume that $\int_0^\infty \mu = \infty$, so the X_i are allowed to take the conditions $0 < U_i < X_i$ or $U_i = \infty$ no restrictions whatever are placed upon the possible values for the censoring times, in particular two or ono of them may coinside or coinside also with one of the failure times. (Of course with a continuous distribution for the X_i , all finite X_i are distinct). Notice that the model with i.i.d. failure times and no censoring is obtained by defining $U_i = \infty$ for all i .

The statistical problem to be discussed is that of estimating the integrated hazard $\int_0^t \mu$ or the survivor function G_μ . More specifically we shall discuss censoring patterns that allows one to use the Nelson - Aalen estimator as estimator of the integrated hazard and the Kaplan - Meier estimator as estimator of G_μ .

Recall that with $\tilde{N}$ the counting process

$$\tilde{N}(t) = \sum_{i=1}^{n} 1_{(X_i \leq t, \delta_i = 1)} , \tag{2.1}$$

and $|R(t-)|$ the number of individuals at risk immediately before t ,

$$|R(t-)| = \sum_{i=1}^{n} 1_{(T_i \geq t)} , \tag{2.2}$$

the Nelson - Aalen estimator (Nelson [11], Aalen [1]) is given by the Stochastic integral

$$\hat{\beta}(t) = \int_{(0,t]} \frac{1}{|R(s-)|} \tilde{N}(ds) \tag{2.3}$$

and the Kaplan - Meier estimator [9] by the product integral

$$\hat{G}(t) = \prod_{0 < s \leq t} (1 - \hat{\beta}(ds)) = \prod_{0 < s \leq t} \left(1 - \frac{\tilde{N}(ds)}{|R(s-)|}\right) . \tag{2.4}$$

The simplest situation where it is reasonable to use the estimators (2.3), (2.4), is the model for <u>random censorship</u> which is usually formulated as follows: the X_i are i.i.d. with hazard μ and the U_i are mutually independent and independent of $(X_1,\ldots,X_n)$. With our basic assumption (D) we can of course not use this description. Our specification of the random censorship model appears in Example 3.30 below.

On the other hand it is easy to construct formal censoring patterns, where it is absurd to use the estimators (2.3), (2.4): if $U_i < X_i$ for all i , (2.3), (2.4) degenerate since no failures are observed. As a concrete example, consider $U_i = \frac{1}{2} X_i$, in which case it is of course obvious which estimators should replace (2.3), (2.4). Also, the censorings in an explicit manner anticipate future failures, and this is precisely what must not happen if (2.3), (2.4) are to make sense.

We shall now describe how the observations $(T_1,\delta_1),\dots,(T_n,\delta_n)$ may be viewed as a multivariate counting process and how the distribution of the observations is given in terms of the corresponding intensity process (compensator).

This counting process approach was initiated by Aalen's [1] definition of the multiplicative intensity model and his observation that some relevant models for right censoring appear as special cases. This in turn led to Gill's work [3], which is a main reference for counting processes and censoring and the main reference for this paper.

Before describing in detail the counting process approach we remind the reader of the following simple facts.

Suppose Pr is a probability on $(0,\infty]$ with survivor function $G(t) = Pr((t,\infty])$. Let $t^\dagger = \inf\{t : G(t) = 0\}$ denote the termination point for G and write ν for the hazard measure

$$\nu(dt) = \frac{1}{G(t-)} Pr(dt) \tag{2.5}$$

with $G(t-) = \lim_{s\to t, s<t} G(s)$. Then ν is a measure on $(0,t^\dagger]$, well-defined if we put the mass at $t^\dagger$ equal to $\Delta\nu(t^\dagger) = 0$ if $G(t^\dagger-) = 0$. (Here as elsewhere, if κ is a measure, $\Delta\kappa(t)$ denotes the point mass $\kappa(\{t\})$, while if κ is a function, rightcontinuous with left limits $\Delta\kappa(t) = \kappa(t) - \kappa(t-)$). Further G (or Pr) is determined by $t^\dagger$ and ν,

$$G(t) = \prod_{0<s\leq t} (1-\nu(ds))$$

for $t \leq t^\dagger$, where the product integral on the right equals

$$\exp(-\nu^c((0,t])) \prod_{0<s\leq t} (1-\Delta\nu(s))$$

with ν^c the continuous part of ν and the product the contribution from the discrete part, so that the factor $1-\Delta\nu(s)$ is $\neq 1$ only if

s is one of the at most countably many atoms for ν (or Pr).

With Pr absolutely continuous with hazard μ , of course $\nu(dt) = \mu(t)dt$, and if Pr is discrete, ν is also discrete with point masses $\Delta\nu(s) = -\Delta G(s)/G(s-)$. In particular

2.6. Example. The discrete distribution with survivor function $\hat{G}$ has hazard measure $\hat{\nu}$ given by $\hat{\nu}((0,t]) = \hat{\beta}(t)$, see (2.3) and (2.4). Hence the Kaplan - Meier and Nelson - Aalen estimators are equivalent in the sense that they describe two different characteristics of the same distribution. □

Consider now a collection of failure times $X_1,\dots,X_n$ and censoring times $U_1,\dots,U_n$ satisfying (D). To give the counting process description of the observations (T_i,δ_i) and their distribution, we shall use the method given e.g. in Chapter 2 of Jacobsen [5].

To observe $(T_i,\delta_i)_{1 \le i \le n}$ is equivalent to observing the occurence in time of a sequence of events. The possible events consist in either a failure simultaneously with a number of censorings or in the occurence of one or more censorings (but no failure). We shall give each event a name (mark) and collect the names in the type set (mark space)

$$E = \{(i,A) : 1 \le i \le n \ , \ A \subseteq \{1,\dots,n\} \setminus \{i\}\}$$
$$\cup \ \{(c,B) : \emptyset \ne B \subseteq \{1,\dots,n\}\}$$

with (i,A) the name of the event 'failure for i , all $j \in A$ censored' and (c,B) the name of the event 'no failures, all $j \in B$ censored'. In particular, $(i,\emptyset)$ is the name of the event 'failure for i , no censorings'. For each $y \in E$ we define $K(y)$, the set of individuals involved in y , by

(2.7) $$K(y) = \begin{cases} \{i\} \cup A & \text{if} \quad y = (i,A) \\ B & \text{if} \quad y = (c,B) \end{cases} .$$

In finite time a random number $(\leq n)$ and a random selection of events are observed. Let τ_k be the time of occurence of the k'th event and let Y_k be the name of the event. If precisely m events, $0 \leq m \leq n$, are observed on $(0,\infty)$, we have $0 < \tau_1 < \ldots < \tau_m < \infty$ and define $\tau_{m+1} = \ldots = \tau_n = \infty$ and leave $Y_{m+1}, \ldots, Y_n$ unspecified.

Clearly, observing all (T_i, δ_i) is equivalent to observing all τ_k and the Y_k corresponding to k with $\tau_k < \infty$. Thus the distribution of the observations is specified by giving the joint distribution of all (τ_k, Y_k), and following Jacobsen [5], we shall do this by specifying for $k \geq 0$ the conditional probabilities

$$G_{k,\xi_k}(t) = P(\tau_{k+1} > t \mid \mathcal{F}_{\tau_k}) \quad \text{on} \quad (\tau_k < \infty) ,$$

$$\pi_{k,\xi_k}(t,y) = P(Y_{k+1} = y \mid \mathcal{F}_{\tau_{k+1}-}) \quad \text{on} \quad (\tau_{k+1} = t) .$$

Here $\mathcal{F}_{\tau_k}$ is the σ-algebra generated by $\xi_k = (\tau_1, \ldots, \tau_k, Y_1, \ldots, Y_k)$ with $\mathcal{F}_{\tau_0} = \mathcal{F}_0 = \{\emptyset, \Omega\}$, and $\mathcal{F}_{\tau_{k+1}-}$ is the σ-algebra spanned by (ξ_k, τ_{k+1}). Note that G_{k,ξ_k} is a survivor function on $(\tau_k, \infty]$ and that $\sum_y \pi_{k,\xi_k}(t,y) = 1$.

Letting ν_{k,ξ_k} be the hazard measure for the probability on $(\tau_k, \infty]$ with survivor function G_{k,ξ_k} (cf. (2.5)), introduce for $y \in E$ the random <u>intensity measure</u>.

(2.8) $$\Lambda^y(dt) = \nu_{k,\xi_k}(dt)\, \pi_{k,\xi_k}(t,y)$$

on $(\tau_k < t \leq \tau_{k+1})$ and define the <u>intensity process</u> (compensator) $\Lambda = (\Lambda^y)_{y \in E}$ by

$$\Lambda^y(t) = \Lambda^y((0,t]) .$$

Also let N^y be the counting process which at time t equals 1 if the event with name y has occurred in $(0,t]$ and 0 otherwise. Then $N = (N^y)_{y \in E}$ is a multivariate counting process such that with $\mathcal{F}_t$ the σ-algebra generated by $(N(s))_{s \leq t}$, the predictable compensator for N^y with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$ is Λ^y, i.e. Λ^y is predictable and

$$M^y = N^y - \Lambda^y$$

is an $\mathcal{F}_t$-martingale for each y, cf. Jacod [6] or, for the special case of absolutely continuous G_{k,ξ_k}, Jacobsen [5], Section 2.2. Also the distribution of N is uniquely determined by the intensity process Λ as may be seen from the following expression (see [5], Proposition 2.5.13): introduce the total intensity

$$\overline{\Lambda} = \sum_{y \in E} \Lambda^y$$

and consider the infinitesimal event that on $(0,t]$ precisely m jumps occur at times in $dt_1, \ldots, dt_m$, the k'th jump occurring in component y_k. Then

(2.9) $$P(\tau_1 \in dt_1, Y_1 = y_1, \ldots, \tau_m \in dt_m, Y_m = y_m, \tau_{m+1} > t) =$$

$$= \prod_{\substack{0 < s \leq t \\ s \neq t_k}} (1 - \overline{\Lambda}(ds,w)) \prod_{k=1}^{m} \Lambda^{y_k}(dt_k,w) ,$$

where w is any sample path for N which on $[0,t]$ jumps at the timepoints $t_1, \ldots, t_m$ in components $y_1, \ldots, y_m$. (Because Λ is predictable, the infinitesimal neighbourhoods dt_k should be thought of as intervals $(t_k - dt_k, t_k]$ to the <u>left</u> of t_k).

<u>Note</u>. Adding up infinitesimal probabilities like (2.9) shows in particular that for any $F \in \mathcal{F}_t$, $P(F)$ is determined by the behaviour of Λ on F.

We shall call the counting process N recording the observations (T_i, δ_i) a failure - censoring process (FC - process).

Remark. When stating that a process $Z = (Z(t))_{t \geq 0}$ is predictable (like each Λ^y), we mean that for every $t, Z(t)$ is F_{t-} - measurable with $F_{0-} = F_0$ and for $t > 0$, F_{t-} the σ - algebra generated by $(N(s))_{s<t}$, cf. Jacobsen [5], Section 2.5. Any process which is measurable and predictable in this path - algebraic sense, is also predictable in the sense of the general theory of processes.

As a multivariate counting process, N has a special structure. Not only does each component have at most one jump, but a jump in one component precludes jumps in a host of other components. More specifically, recalling the definition (2.7) of $K(y)$, since the occurrence of the event y means that all individuals $j \in K(y)$ are removed, if y happens, no event $y' \neq y$ involving a $j \in K(y)$ can ever occur. This special structure of N is reflected in a special structure of Λ , see (2.10) below.

Some counting processes derived from N are of special interest. For $i = 1, \ldots, n$ introduce

$$N^i = \sum_A N^{(i,A)} ,$$

the sum extending over all $A \subseteq \{1, \ldots, n\} \setminus \{i\}$. Then N^i registers the failure of i and

$$\tilde{N} = \sum_{i=1}^{n} N^i$$

registers the total number of failures, cf. (2.1). The compensators for $N^i, \tilde{N}$ are

$$\Lambda^i = \sum_A \Lambda^{(i,A)} , \quad \tilde{\Lambda} = \sum_{i=1}^{n} \Lambda^i .$$

It is also useful to introduce the observed lifetimes

$$X_i^* = \inf\{t : N^i(t) = 1\} ,$$

where we use the standard convention $\inf \emptyset = \infty$. The corresponding description of the censoring times is that

$$U_i = \inf\{t : N^{(j,A)}(t) = 1 \quad \text{for some} \quad j \neq i \ , \ A \ni i \quad \text{or} \quad N^{(c,B)}(t) = 1 \quad \text{for some} \quad B \ni i\}$$

as is seen referring to (D).

At each time point t , the collection $\{1,\ldots,n\}$ of individuals split into three disjoint sets, the <u>risk set</u> $R(t-)$, the <u>censoring set</u> $C(t-)$ and the <u>failure set</u> $D(t-)$, where

$$R(t-) = \{i : U_i \geq t \ , \ X_i^* \geq t\} \ ,$$

$$C(t-) = \{i : U_i < t\} \ ,$$

$$D(t-) = \{i : X_i^* < t\} \ .$$

Each of these random sets is F_{t-} -measurable, for instance $R(t-)$ is the set of individuals at risk immediately before t . Occasionally we shall use

$$R(t+) = \{i : U_i > t \ , \ X_i^* > t\}$$

and the analogues $C(t+)$, $D(t+)$. For $t = \infty$ we define the set of individuals always at risk $R(\infty)$, the set of censored individuals $C(\infty)$ and the set of failed individuals $D(\infty)$,

$$R(\infty) = \{i : U_i = X_i^* = \infty\} \ ,$$

$$C(\infty) = \{i : U_i < \infty\} \ ,$$

$$D(\infty) = \{i : X_i^* < \infty\} \ .$$

Observing N on $[0,t)$ is equivalent to keeping track of $R(s-)$, $C(s-)$, $D(s-)$ for $0 < s \leq t$.

Recalling the special structure of the sample paths for N described above, it should be clear that the intensity process Λ always satisfies the following condition: for every $y \in E$ and $s < t$,

$$(2.10) \qquad \Lambda^y(t) - \Lambda^y(s) = (\Lambda^y(t) - \Lambda^y(s))\,1_{(K(y) \subseteq R(s-))} \,,$$

i.e. the intensity measure on $[s,\infty)$ is >0 only if all individuals involved in y are at risk immediately before s.

From now on we shall consider the canonical version of the counting process N, i.e. we introduce W as the space of all possible paths for N and define $N(t,w) = w(t)$ for any $w \in W$. Thus from the description above, W consists of all functions $w:[0,\infty) \to \{0,1\}^E$ such that each component $w^y:[0,\infty) \to \{0,1\}$ is rightcontinuous with $w^y(0) = 0$ and either identically 0 or with 1 jump from 0 to 1, and also the collection (w^y) satisfies that no two components jump simultaneously and whenever w^y jumps, all $w^{y'}$ with $y' \neq y$ and $K(y') \cap K(y) \neq \emptyset$ are identically 0.

An observation is now a path $w \in W$, and all observable objects such as $\tau_k, Y_k, X_i^*, U_i, \Lambda, R, C, D$ are defined on W. The measurable structure is given by $\mathcal{F} = \sigma(N(t))_{t \geq 0}$ with $\mathcal{F}_t, \mathcal{F}_{t-}, \mathcal{F}_{\tau_k}$ sub σ-algebras defined exactly as before. Note that τ_k, X_i^*, U_i are all stopping times with respect to the filtration $(\mathcal{F}_t)$.

<u>2.11. Definition</u>. A (canonical) <u>failure-censoring process</u> is a probability on $(W,\mathcal{F})$.

□

Henceforth we reserve the letter P for denoting FC-processes. A FC-process is given by its compensator Λ, and it is useful to summarize the structure of those compensators that yield FC-processes.

<u>2.12. Fact</u>. Let $\Lambda = (\Lambda^y)_{y \in E}$ be a collection of processes $\Lambda^y:[0,\infty) \to [0,\infty)$ defined on W and consider the following conditions: for all $y \in E$

(i) Λ^y is rightcontinuous and increasing with $\Lambda^y(0) = 0$,

(ii) $\Lambda^y(t)$ is $\mathcal{F}_{t-}$ - measurable for all t ,

(iii) $\Delta \bar{\Lambda}(t) \leq 1$ for all t ,

(iv) $\Lambda^y(t) - \Lambda^y(s) = (\Lambda^y(t) - \Lambda^y(s)) 1_{(K(y) \subseteq R(s-))}$
for all $s \leq t$.

Then

(a) For any FC - process P there is a version of its compensator that satisfies (i) - (iv);

(b) Any predictable Λ satisfying (i) - (iv) is the compensator for a uniquely determined FC - process P .

□

The first three conditions are satisfied by all compensators, while (iv) is (2.10) repeated.

With the concept of canonical FC - processes introduced we can define statistical models for the observation N by giving a family of compensators satisfying (i) - (iv). But at the same time the model should allow for the observed failure times to be interpreted as coming from an i.i.d. sample with some hazard μ . Therefore we must discuss not only a model for the distribution of N , but a model for the joint distribution of X and N , where $X = (X_1,\dots,X_n)$ is the vector of all failure times (observed or unobserved).

Thus, formally the joint model for (X,N) should of course have the property that the marginal distribution of X makes the X_i i.i.d. μ with μ arbitrary. In addition we shall now introduce a condition, first used by Gill [3], on the marginal distribution of N and which we consider the minimal requirement on a model for which the Nelson - Aalen estimator is a sensible estimator of the integrated hazard, or, equivalently, the Kaplan - Meier estimator is a sensible estimator of G_μ . Define

$$I_i(t) = 1_{(i \in R(t-))}, \; I(t) = 1_{(R(t-) \neq \emptyset)}$$

and suppose that for $i = 1, \ldots, n$

(2.13) $$\Lambda^i(dt) = \mu(t) \; I_i(t)dt \; .$$

Then with respect to P, the FC-process with compensator Λ, the processes $M^i = N^i - \Lambda^i$ are orthogonal martingales, and the Nelson-Aalen estimator is a martingale estimator of the integrated hazard, i.e.

$$\hat{\beta}(t) - \int_0^t \mu(s) \; I(s)ds$$

is a P-martingale, where $\hat{\beta}$ is defined by (2.3). Condition (2.13) is crucial for all that follows. We shall refer to it as the <u>martingale condition</u>.

With (2.13) and the preceding remarks in mind, our main purpose is to discuss the following three problems:

I With the X_i i.i.d. μ, what kind of structure must be imposed on the censoring pattern for the marginal distribution of N to satisfy (2.13) ?

II. (The embedding problem). Supposing the distribution of N to satisfy (2.13) for some μ, is it always possible to obtain this distribution as the N-marginal distribution of a pair (X,N), where the X_i are i.i.d. μ ?

III. What must be the structure of statistical models for the distribution of N, satisfying (2.13) for all μ, in order that no essential information about μ is lost when using the Nelson-Aalen estimator ?

When answering I and II, it is enough to fix an arbitrary μ at a time, while III involves the structure of the complete model obtained when μ varies.

Gill [3] gave several examples of censoring patterns, where (2.13) is satisfied including type II progressive censorship, where at the time of the k'th observed failure, a given number r_k of individuals selected at random from the i still at risk, are censored.

Another example (Williams and Lagakos [12] and Kalbfleisch and MacKay [7]) treats the case where the pairs (X_i, U_i) are i.i.d. (In those two papers our convention about the U_i from (D) is not used, but that is immaterial for (2.14) below). In this example, because of the independence, (2.13) holds iff it holds separately for each i, and with our notation the constant sum condition from [12] as reformulated by Kalbfleisch and MacKay [7] then reads

$$(2.14) \qquad P(T_i \in dt\, ,\, \delta_i = 1 \mid T_i \geq t) = \mu(t)dt \ .$$

In particular, this is a condition on the observations only, and since $(T_i \geq t) = (i \in R(t-))$ it is immediate that (2.14) is precisely (2.13).

As stressed above, it is necessary to study the joint distribution of X and N, and we shall now introduce the notation and formal apparatus needed to do this.

Let $L \subset (0,\infty]^n$ denote the space of vectors of possible failure times, i.e. vectors $x = (x_1, \ldots, x_n)$ with $0 < x_i \leq \infty$ and such that no two finite x_i are equal. Also write $X_i(x) = x_i$, define $\mathcal{H} = \sigma(X_i)_{1 \leq i \leq n}$ and let $\mathcal{H}_t = \sigma(X_i 1_{(X_i \leq t)})_{1 \leq i \leq n}$ be the σ-algebra measuring all failure times $\leq t$, whether observed or not. Note that $(X_i > t) \in \mathcal{H}_t$.

A realization of all failure times and the observations is a point $\omega = (x,w) \in \overline{\Omega} := L \times W$ such that x and w are <u>compatible</u> , i.e. ω belongs to Ω defined as the space of pairs $(x,w) \in \overline{\Omega}$ with $x_i = X_i^*(w)$ for $i \in D(\infty,w)$, $x_i > U_i(w)$ for $i \in C(\infty,w)$ and $x_i = \infty$

for $i \in R(\infty,w)$. (See p.11 for the definition of $D(\infty)$, $C(\infty)$, $R(\infty)$).

On Ω we use the σ-algebra $\mathcal{G} = \Omega \cap (\mathcal{H} \otimes \mathcal{F})$ with the filtration $\mathcal{G}_t = \Omega \cap (\mathcal{H}_t \otimes \mathcal{F}_t)$.

With this setup the joint distribution of (X,N) is a probability on $(\Omega,\mathcal{G})$, which we denote by $\mathbb{P}$. (Recall that P (Q_μ) is the notation for the marginal distribution of N (X)).

We shall say that $x \in L$, $w \in W$ are <u>t-compatible</u> if $x_i = X_i^*(w)$ for $i \in D(t-,w)$, $x_i > U_i(w)$ for $i \in C(t-,w)$ and $x_i \geq t$ if $i \in R(t-,w)$. Thus x and w are t-compatible simply if x is a vector of failure times consistent with the observation of w on $[0,t)$.

Given $x \in L$, denote by W_x the space of w compatible with x : $W_x = \{w \in W : (x,w) \in \Omega\}$.

Any function defined on W (or L) may be viewed as a function on Ω , e.g. define $N^y(t,(x,w)) = N^y(t,w)$ and $X_i(x,w) = x_i$. And a set $F \in \mathcal{F}$ may be viewed as the set $F = \{(x,w) \in \Omega : w \in F\} \in \mathcal{G}$, a set $H \in \mathcal{H}$ as the set $H = \{(x,w) \in \Omega : x \in H\}$. Then also $\mathcal{H}_t, \mathcal{F}_t$ may be considered sub σ-algebras of $\mathcal{G}$ and then $\mathcal{G}_t = \mathcal{H}_t \vee \mathcal{F}_t$, the smallest σ-algebra containing both $\mathcal{H}_t$ and $\mathcal{F}_t$.

We shall construct probabilities on Ω by letting the X_i be i.i.d. μ (the distribution of X is $Q = Q_\mu$), and then consider the conditional distribution of the counting process N given X .

So for every $x \in L$, let P_x be a probability on $(W,\mathcal{F})$ with $P_x(W_x) = 1$ and such that $x \to P_x(F)$ is measurable for all $F \in \mathcal{F}$. Then

$$\mathbb{P} = \int P_x \, Q(dx) \tag{2.15}$$

defines a probability on Ω . By standard results about regular conditional probabilities, any $\mathbb{P}$ making the X_i i.i.d. with hazard μ has this structure. A useful variation of (2.15) is the following: if $f \geq 0$ defined on Ω is measurable, then

(2.16) $$\int f \, d\mathbb{P} = \int Q(dx) \int f(x,w) \, P_x(dw) \; .$$

Each P_x is a FC-process, hence is specified by its intensity process Λ_x, formally defined on all of W. In order that $P_x(W_x) = 1$, Λ_x must satisfy (i) - (iv) from Fact 2.12 plus some extra conditions, that we now list:

For all $\emptyset \neq B \subseteq \{1,\dots,n\}$ and all i

(2.17) $$\Delta \Lambda_x^{(c,B)}(x_i) = 0 \quad \text{on} \quad (i \in R(x_i-))$$

and for all $(i,A) \in E$, the intensity measure $\Lambda_x^{(i,A)}$ is concentrated at x_i with

(2.18) $$\Delta \Lambda_x^{(i,A)}(x_i) = \Delta \Lambda_x^{(i,A)}(x_i) \, 1_{(\{i\} \cup A \subseteq R(x_i-))} \; .$$

Finally,

(2.19) $$\Delta \Lambda_x^{i}(x_i) = 1 \quad \text{on} \quad (i \in R(x_i-))$$

because, with respect to P_x, i must fail at time x_i if still at risk.

It is not necessary for Λ_x to satisfy (2.17) - (2.19) on all of W, it is enough that these conditions hold on W_x:

<u>2.20. Lemma</u>. Let $x \in L$ and let $\Lambda_x = (\Lambda_x^y)_{y \in E}$ be the intensity for some FC-process P_x. If (2.17) - (2.19) hold for the restriction of Λ_x to W_x, then $P_x(W_x) = 1$.

<u>Sketch of proof</u>. Fix $x = (x_1,\dots,x_n) \in L$ and let $x_{(1)} = \min x_i$. Consider the pathfragment w_0 for N, which has no jumps on the interval $[0,x_{(1)})$. From the definition of W_x, it is clear there exists $w \in W_x$ with this behaviour on $[0,x_{(1)})$, hence (2.17)-(2.19) apply when evaluating $\Lambda_x(t,w_0)$ for $0 \leq t \leq x_{(1)}$. Since by (2.19), $\Delta \overline{\Lambda}_x(x_{(1)},w_0) = 1$, we have $P_x(\tau_1 \leq x_{(1)}) = 1$, and then because each

Λ_x^i is concentrated at x_i , it follows that P_x - a.s. on $(\tau_1 < x_{(1)})$, Y_1 is of the form (c,B) , while because of (2.19), P_x - a.s. on $(\tau_1 = x_{(1)})$, Y_1 is of the form (j,A) with j determined by $x_{(1)} = x_j$. Thus the first jump time τ_1 and the type of the first jump Y_1 generated by P_x are compatible with x . Proceeding by induction one shows similarly that if $(\tau_1, Y_1), \ldots,$ (τ_k, Y_k) are compatible with x , $P_x(\cdot \mid \mathcal{F}_{\tau_k})$ generates (τ_{k+1}, Y_{k+1}) compatible with x , and that only the restriction of Λ_x to W_x is needed to determine the conditional probability.

□

3. Main results - problems I and II.

We shall begin by quoting Gill's [3] answer to question I. With our notation Gill's basic condition, which is a condition on the joint distribution $\mathbb{P}$ of (X,N) may be phrased as follows:

(G) For any $t \geq 0$, given $\mathcal{F}_t$ the X_i for $i \in R(t+)$ are i.i.d. on $(t,\infty]$ with hazard μ.

3.1. Theorem. (Gill [3], Theorem 3.1). If the joint distribution of (X,N) is such that (G) is satisfied, then the martingale condition (2.13) holds.

□

If (G) is true, taking $t = 0$ shows all X_i to be i.i.d. with hazard μ. Also, given $\mathcal{F}_t$ with $t > 0$, in particular the risk set $R(t+)$ is known and for each $i \in R(t+)$ we have $X_i > t$. So the condition states that given $\mathcal{F}_t$, the X_i for $i \in R(t+)$ are i.i.d., each following the same distribution as X_i given the event $(X_i > t)$.

We shall present a new condition (C) below. To formulate it we use the construction of a probability $\mathbb{P}$ on Ω described in the previous section: Start with the X_i i.i.d. μ, then use P_x, the conditional distribution of N given $X = x$. (C) will be phrased as a condition on the intensity process Λ_x for P_x, $x \in L$ arbitrary.

From a probabilistic point of view, using conditional probabilities is certainly the most natural way of constructing the joint law $\mathbb{P}$, when the marginal distribution of X is prescribed. But from a statistical point of view the idea of conditioning on all failure times, whether observed or not, appears quite unnatural. However, if the conditional distribution of N given X depends on the failure times only through what is observed about them, not only does this make sense statistically, but as we shall see, it also captures the structure we are looking for. This then is the essence of condition (C).

This condition is explicitly a condition on the censoring pattern, where (G) is more implicit, stating that in a suitable sense, the censorings should leave the failure time distributions unchanged.

(C) For any $t>0$ and $w \in W$, $\Lambda_x(t,w)$ is the same for all $x \in L$ which are t-compatible with w and satisfy $x_i > t$ for $i \in R(t-,w)$.

Recall that the definition of t-compatibility between x and w only involves the behaviour of w on $[0,t)$, which, since Λ_x is predictable, also determines $\Lambda_x(t,w)$. Because of (2.19) it is essential that in (C) a x t-compatible with w is required to satisfy $x_i > t$ for $i \in R(t-,w)$ rather than $x_i \geq t$. (See also Lemma 3.5 below).

Because each $\Lambda_x(t,w)$ is rightcontinuous in t, (C) is equivalent to the following, seemingly stronger condition.

(C') For any $t>0$ and $w \in W$, the function $s \to \Lambda_x(s,w)$ on $[0,t]$ is the same for all $x \in L$ which are t-compatible with w and satisfy $x_i > t$ for $i \in R(t-,w)$.

Indeed, given t, w let $x, x' \in L$ satisfy the requirements in (C') or (C). Using (C) (with t replaced by an arbitrary $s \leq t$), it is seen that $\Lambda_x(s,w) = \Lambda_{x'}(s,w)$ for all $s \leq t$ except possibly $s = x_j$ or $s = x'_j$. Now use right continuity to complete the argument that (C) implies (C').

It may now also be seen that (C) implies the following:

(C") For any $t>0$, $w \in W$ and $i \in R(t-,w)$, $\Lambda_x(t,w)$ is the same for all $x \in L$ which are t-compatible with w and satisfy that $x_i = t$ and $x_j > t$ for all $j \in R(t-,w)$, $j \neq i$.

To argue this, fix t, w, i and let $x, x' \in L$ satisfy the requirements in (C"). Because w is piecewise constant and all x_j, x'_j for $j \neq i$ are $\neq t$, condition (C') applies with t replaced by $t + \delta$ for $\delta > 0$ sufficiently small.

In the sequel, when referring to (C), we mean either of (C), (C'), (C").

<u>Note</u>. Of course, since (C) is a condition on conditional probabilities, it may be relaxed, allowing for exceptional sets of w's and x's. For instance one may always ignore x belonging to some given set A with $\mathbb{P}(X \in A) = 0$. Also, it must be remembered that Λ_x as a function on W (or W_x), is determined only up to P_x-indistinguishability.

Apart from conditions (G) and (C) we shall introduce a third condition on the structure of $\mathbb{P}$. Recall that $\mathcal{H}_t = \sigma(X_i 1_{(X_i \leq t)})_{1 \leq i \leq n}$ and introduce $\mathcal{H}^t = \sigma(X_i 1_{(X_i > t)})_{1 \leq i \leq n}$, the σ-algebra spanned by those X_i greater than t.

(M) For any $t \geq 0$, the σ-algebras $\mathcal{F}_t$ and $\mathcal{H}^t$ are conditionally independent given $\mathcal{H}_t$.

If (M) holds, then for $t_i \geq t$,

$$\mathbb{P}(X_i > t_i, i \in S | \mathcal{G}_t) = \prod_{i \in S} G_\mu(t_i | t) \tag{3.2}$$

on the set $(\{i : X_i > t\} = S) \in \mathcal{H}_t$: by the conditional independence, $\mathbb{P}(H | \mathcal{G}_t) = \mathbb{P}(H | \mathcal{H}_t)$ for $H \in \mathcal{H}^t$, and (3.2) follows because the X_i are i.i.d. μ. (3.2) should be compared to (G). (The notation $G_\mu(\cdot | \cdot)$ used in (3.2) was introduced early in Section 2).

Writing the conditional independence as

$$\mathbb{P}(F | \mathcal{H}) = \mathbb{P}(F | \mathcal{H}_t), \quad F \in \mathcal{F}_t \tag{3.3}$$

we can give yet another version of (M). It is standard terminology to call a random time τ a <u>randomized stopping time</u> for the filtration $(\mathcal{H}_t)$, if for any t, $\mathbb{P}(\tau > t | \mathcal{H}) = \mathbb{P}(\tau > t | \mathcal{H}_t)$. Taking $\tau = U_i$, (3.3) shows that each censoring time is a randomized stopping time for the filtration induced by the failure times. Furthermore, conditioning on $\mathcal{H}$ (i.e. X) leaves only the U_i as random and hence (3.3) holds for all $F \in \mathcal{F}_t$ iff it holds for all $F \in \mathcal{U}_t := \sigma(U_i 1_{(U_i \leq t)})_{1 \leq i \leq n}$. Thus (M) is equivalent to the statement that $(U_1,\ldots,U_n)$ is a multivariate randomized stopping time if by this (non-standard) statement, we mean that (3.3) holds for all t and all $F \in \mathcal{U}_t$. Finally, note that since $\mathbb{P}(\cdot|\mathcal{H}) = P_x$ on $(X = x)$, (3.3) is a condition on the P_x.

<u>3.4. Proposition</u>. (C) $\Rightarrow$ (M) $\Rightarrow$ (G) and neither implication can be reversed.

<u>Proof</u>. Suppose (C) holds, let $t > 0$ and let $x, x' \in L$ satisfy that for each i either $x_i = x'_i$ or $x_i > t$, $x'_i > t$. In particular, any w t-compatible with x is also t-compatible with x' and from (C) it follows via (C'), (C") that for all such w, $\Lambda_x(s,w) = \Lambda_{x'}(s,w)$ for $s \leq t$. In other words, the intensity processes for P_x and $P_{x'}$ agree on $[0,t]$ and hence $P_x \equiv P_{x'}$ when restricted to $\mathcal{F}_t$. But conditioning on $\mathcal{H}_t$ amounts precisely to specifying for each i the value of X_i if $X_i \leq t$ and the event $(X_i > t)$ otherwise. Thus we have shown that (3.3) and hence (M) hold.

That (M) implies (G) is easy to see. For an arbitrary subset R of $\{1,\ldots,n\}$ we must show that for $t_i \geq t$, $i \in R$

$$\mathbb{P}(X_i > t_i, i \in R | \mathcal{F}_t) = \prod_{i \in R} G_\mu(t_i | t)$$

on the set $F = (R(t+) = R) \in \mathcal{F}_t$. Conditioning first on $\mathcal{G}_t$ and noting that on F , $R \subseteq \{i : X_i > t\}$, this follows from (3.2).

That the implications cannot be reversed will be shown in Example 3.31 below.

□

We shall later use the following consequence of (M).

3.5. Lemma. Let $t > 0$ and let $x, x' \in L$ satisfy that for all i either $x_i = x'_i$ or $x_i, x'_i \geq t$. If (M) holds, then $P_x \equiv P_{x'}$ on $\mathcal{F}_{t-}$.

Proof. Because for any $s < t$, $x_i = x'_i$ or $x_i, x'_i > s$, condition (M) tells us that $P_x \equiv P_{x'}$ on $\mathcal{F}_s$. Since $\mathcal{F}_{t-}$ is the σ-algebra spanned by $(\mathcal{F}_s)_{s<t}$, the lemma follows.

□

Our main result gives the structure of the intensity process Λ for N , when (C) is satisfied, and in particular, it follows that the martingale condition (2.13) holds. Of course this fact alone follows directly from Gill's theorem and Proposition 3.4. Part of the justification for introducing the restrictive condition (C) lies in the more detailed information provided by Theorem 3.6 and the solution to the embedding problem given in Theorem 3.21 below.

In the sequel, if $x \in L$, $t > 0$, we write $x_{|i,t}$ for the vector $(x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_n)$. Note that $x_{|i,t} \in L$ except if $t = x_j$ for some $j \neq i$.

3.6. Theorem. Let $\mathbb{P}$ be a probability on Ω such that the X_i are i.i.d. with hazard μ and the intensities Λ_x for the conditional probabilities P_x satisfy (C). The compensator for the counting process N is then indistinguishable from $\Lambda = (\Lambda^y)_{y \in E}$, where

$$(3.7)\qquad \Lambda^{(i,A)}(dt) = \mu(t)\, \Delta\Lambda^{(i,A)}_{X|i,t}(t)\, I_i(dt)\,,$$

$$(3.8)\qquad \Lambda^{(c,B)} = \Lambda^{(c,B)}_{X}$$

for $i = 1,\dots,n$, $A \subseteq \{1,\dots,n\}\setminus\{i\}$, $\emptyset \neq B \subseteq \{1,\dots,n\}$. In particular Λ as given by (3.7) and (3.8) and viewed as a function of $(x,w) \in \Omega$ depends on w alone, each $M^y = N^y - \Lambda^y$ is a (P, F_t)-martingale and (2.13) holds,

$$(3.9)\qquad \Lambda^{i}(dt) = \mu(t)\, I_i(t)\, dt\,.$$

Remarks. The reader is reminded that two processes V_1 and V_2 are indistinguishable if almost surely $V_1(t) = V_2(t)$ simultaneously for all t. The qualification indistinguishable is required because, as we shall see $\Lambda^{(c,B)}_{X}(t)$ is not predictable in the strict algebraic sense we are using here (cf. the remark following (2.9) above). Equation (3.7) is best understood recalling that the measure $\Lambda^{(i,A)}_{X}$ is concentrated at x_i. Also note that because of (2.18), the factor $I_i(t)$ may be omitted from (3.7).

Proof. We begin by showing that Λ, which formally depends on both x and w, is determined by w alone, i.e. that whenever $(x,w) \in \Omega$, $(x',w) \in \Omega$, the paths for Λ evaluated at (x,w) and (x',w) agree. But x and x' are both compatible with w, therefore $x|i,t$ and $x'|i,t$ are t-compatible with w for all t such that $I_i(t,w) = 1$, and from (C) it then follows easily that

$$\Delta\Lambda^{(i,A)}_{x|i,t}(t,w)\, I_i(t,w) = \Delta\Lambda^{(i,A)}_{x'|i,t}(t,w)\, I_i(t,w)$$

$$\Lambda^{(c,B)}_{x}(t,w) = \Lambda^{(c,B)}_{x'}(t,w)$$

for all t not equal to some x_j or x'_j. Clearly then the Lebesgue integrals defining $\Lambda^{(i,A)}$ agree, when evaluated for (x,w) and

(x',w) . That the censoring components $\Lambda^{(c,B)}$ agree, even at the exceptional time points $t = x_j$ or x'_j , follows most easily by rightcontinuity, but is also a consequence of some of the basic properties of the Λ_x , viz. Fact 2.12 (iv) and (2.17).

Thus we may write $\Lambda(t,w) = \Lambda(t,(x,w))$, and shall now proceed to show that the process $\Lambda^{(i,A)}$ is predictable. Following Jacobsen [5], we do this by showing that for all t , if $w,w' \in W$ satisfy $w \underset{t-}{\sim} w'$, i.e. $w(s) = w'(s)$ for all $s \in [0,t)$, then $\Lambda^{(i,A)}(t,w) = \Lambda^{(i,A)}(t,w')$. But

$$\Lambda^{(i,A)}(t) = \int_0^t \mu(s)\,\Delta\,\Lambda^{(i,A)}_{X|i,s}(s)\, I_i(s)\,ds$$

and it follows from the fact that any Λ_x is predictable that the integrands evaluated for w,w' are the same, except possibly at finitely many timepoints s . Thus $\Lambda^{(i,A)}$ is predictable.

The next step consists in showing that a modified version of $\Lambda^{(c,B)}$ is predictable. Fix some $\delta > 0$ and define

$$(3.10) \qquad \tilde{\Lambda}^{(c,B)} = \Lambda_X^{(c,B)} + \sum_{i \in D(\infty)} \Delta\,\Lambda^{(c,B)}_{X|i,X_i+\delta}(X_i)\,\varepsilon_{X_i} ,$$

so the measure $\tilde{\Lambda}^{(c,B)}$ differs from $\Lambda^{(c,B)}$ only by point masses at the observed failure times X_i , $i \in D(\infty)$. First, using (C) it is easy to see that each term in the sum evaluated at $(x,w) \in \Omega$ depends on w only. Next, let $w \underset{t-}{\sim} w'$ and choose arbitrary $x,x' \in L$ such that $(x,w) \in \Omega$, $(x',w') \in \Omega$, in particular x,x' are t-compatible with both w and w' .

If for all $i \in R(t-,w) = R(t-,w') = R$ we have $x_i = x'_i$ or $x_i, x'_i > t$ it is immediate from (C") that $\Lambda^{(c,B)}(t,w) = \Lambda^{(c,B)}(t,w')$. Therefore suppose that e.g. for some $i \in R$, $t = x_i < x'_i$, that is x_i is the observed failure time for i based on the path w , while for w' , the failure time $x'_i > t$. Then $\tilde{\Lambda}^{(c,B)}(s,w) = \tilde{\Lambda}^{(c,B)}(s,w')$ for

$s < t$ and it remains to show that

$$\Delta \tilde{\Lambda}^{(c,B)}(t,w) = \Delta \tilde{\Lambda}^{(c,B)}(t,w') .$$

By (2.17), $\Delta \Lambda_x^{(c,B)}(t,w) = 0$, so the left hand side equals

$$\Delta \Lambda_{x|i,t+\delta}^{(c,B)}(t,w) . \tag{3.11}$$

To evaluate the right hand side (RS), we must distinguish between two cases and shall use that for all $j \in R$ with $j \neq i$ we have $x_j > t$. The first case is that for all such j, $x'_j > t$. Then (RS) equals $\Delta \Lambda_{x'}^{(c,B)}(t,w')$ and (C) shows this to be the same as (3.11). For the second case, assume $x'_j = t$ for some $j \in R$, $j \neq i$. Then (RS) becomes $\Delta \Lambda_{x'|j,t+\delta}^{(c,B)}(t,w')$ and again by (C) this equals (3.11).

The last assertion (3.9) in the theorem follows immediately from (3.7) and (2.19). So the remainder of the proof is concerned with showing that each M^y is a martingale and that $\tilde{\Lambda}^{(c,B)}$ and $\Lambda^{(c,B)}$ are indistinguishable.

We shall show that each M^y is a $(\mathbb{P}, G_t)$-martingale, which certainly renders M^y a (P, F_t)-martingale. Recalling that $G = \Omega \cap (H_t \otimes F_t)$ the martingale property amounts to

$$\mathbb{P}(N^y(t) - N^y(s);G) = \mathbb{P}(\Lambda^y(t) - \Lambda^y(s);G) \tag{3.12}$$

for all $y \in E$, $s \leq t$, $G = \Omega \cap (H \times F)$ with $H \in H_s$, $F \in F_s$. But because $M^y_x = N^y - \Lambda^y_x$ is a (P_x, F_t)-martingale, using (2.16) we see that

$$\begin{aligned} \mathbb{P}(N^y(t) - N^y(s);G) &= \int_H P_x(N^y(t) - N^y(s);F)Q(dx) \\ &= \int_H P_x(\Lambda^y_x(t) - \Lambda^y_x(s);F)Q(dx) \\ &= \mathbb{P}(\Lambda^y_X(t) - \Lambda^y_X(s);G) . \end{aligned} \tag{3.13}$$

Considering the censoring intensities first, let $y = (c,B)$. That $\tilde{\Lambda}^{(c,B)}$ is the compensator for $N^{(c,B)}$ will follow from (3.13) if we show that $\tilde{\Lambda}^{(c,B)}$ is indistinguishable from $\Lambda_X^{(c,B)}$, i.e. looking at (3.10) it is enough to show that for all i,

$$\text{(3.14)} \qquad \mathbb{P}(\Delta \Lambda^{(c,B)}_{X|i,X_i+\delta}(X_i) ; X_i < \infty) = 0 .$$

For every $x \in L$, the function

$$w \to \Delta \Lambda^{(c,B)}_{x|i,x_i+\delta}(x_i,w)$$

is F_{x_i-} -measurable. Hence by Lemma 3.17 below the integral from (3.14) equals

$$\mathbb{P}(\int_0^{X_i} \mu(t)\, \Delta \Lambda^{(c,B)}_{X|i,t+\delta}(t)dt) .$$

By (C), $\Delta \Lambda^{(c,B)}_{X|i,t+\delta}(t) = \Delta \Lambda^{(c,B)}_X(t)$ for $t < X_i$. Since the measure $\Lambda^{(c,B)}_X$ has at most countably many atoms, the Lebesgue integral is 0 and (3.14) follows.

To prove the martingale property (3.12) for y of the form (i,A), note that

$$w \to \Delta \Lambda^{(i,A)}_x(x_i,w)\, 1_G(x,w)\, 1_{(s,t]}(x_i)$$

is F_{x_i-} - measurable. So by Lemma 3.17

$$\mathbb{P}(\Delta \Lambda^{(i,A)}_X(X_i) ; \; G, s < X_i \le t)$$

$$= \mathbb{P}(\int_0^{X_i} \mu(u)\, \Delta \Lambda^{(i,A)}_{X|i,u}(u)\, 1_G(X|i,u ; N)\, 1_{(s,t]}(u)du) .$$

Now

$$1_G(X|i,u ; N) = 1_H(X|i,u)\, 1_F(N)$$

and since the Lebesgue integral extends over $u \in (s,t]$ only and $H \in H_s$, it is seen that this indicator is constant in u, and hence the Lebesgue integral may be written

$$\text{(3.15)} \qquad (\int_s^{t \wedge X_i} \mu(u)\, \Delta \Lambda^{(i,A)}_{X|i,u}(u)du)\, 1_{(X_i > s,G)} \cdot$$

Here we may throw in for free the factor $I_i(u)$ in the integrand, and since $I_i(u)=1$ implies $X_i \geq u$, it is clear that (3.15) reduces to

$$(\Lambda^{(i,A)}(t) - \Lambda^{(i,A)}(s))\,1_G\ ,$$

and we have shown that

$$(3.16)\qquad \mathbb{P}(\Delta\Lambda_X^{(i,A)}(X_i)\ ;G,s<X_i\leq t)$$
$$= \mathbb{P}(\Lambda^{(i,A)}(t) - \Lambda^{(i,A)}(s)\ ;G)\ .$$

But because $\Lambda_X^{(i,A)}$ is concentrated at X_i,

$$\Delta\Lambda_X^{(i,A)}(X_i)\,1_{(s<X_i\leq t)} = \Lambda_X^{(i,A)}(t) - \Lambda_X^{(i,A)}(s)\ .$$

Inserting this in (3.16) and comparing with (3.13), (3.12) follows.

□

In the proof the following observation was used.

<u>3.17. Lemma</u>. Suppose that $\mathbb{P}$ satisfies (M) and let $f_i = f_i(X,N) : \Omega \to \mathbb{R}$ be $\mathbb{P}$-integrable. If for every $x \in L$, $w \to f_i(x,w)$ is $\mathcal{F}_{x_i-}$-measurable on W_x, then

$$\mathbb{P}(f_i\ ;X_i<\infty) = \mathbb{P}\left(\int_0^{X_i} \mu(t) f_i(X|i,t\ ;N)\,dt\right)\ .$$

<u>Note</u>. In understanding the lemma and the applications already given, the following point should be made: if $(x,w)\in\Omega$ it need not of course be true that $(x|i,t\ ;w)\in\Omega$ and hence, at first sight, it is not clear whether and how $f(x|i,t\ ;w)$ is defined for the all but finitely many t with $x|i,t \in L$. But since x and w are t-compatible, also $x|i,t$ and w are t-compatible for $t\leq x_i$, in particular for each t a $w^{(t)}$ may be found with $w^{(t)} \underset{t-}{\sim} w$ and such that $(x|i,t\ ;w^{(t)})\in\Omega$. The measurability assumption on f_i then shows that $f_i(x|i,t\ ;w^{(t)})$ does not depend on the choice of

$w^{(t)}$ and the common value is the one to be used as the value of $f_i(x|i,t\,;w)$.

Proof. By (2.16)

$$\mathbb{P}(f_i\,;X_i<\infty) = \int_{x_i<\infty} P_x f_i(x,N)Q(dx)\,.$$

Because $f(x,\cdot)$ is $\mathcal{F}_{x_i-}$-measurable, Lemma 3.5 implies that

$$P_{x|i,t}\, f_i(x,N) = P_x f_i(x,N)$$

for all $t \geq x_i$. Hence

$$\mathbb{P}(f_i\,;X_i<\infty) = \int_{x_i<\infty} Q(dx)\, G^{-1}(x_i) \int_{[x_i,\infty]} P_{x|i,t}\, f_i(x,N)\, F(dt)\,.$$

(Recall that $F = F_\mu = 1-G$ is the distribution function for X_i). Relabelling x_i into t, t into x_i, using that Q is a product and Fubini, reduces this to

$$\int Q(dx) \int_0^{x_i} \mu(t)\, P_x f_i(x|i,t\,;N)dt$$
$$= \mathbb{P}\left(\int_0^{X_i} \mu(t)\, f_i(X|i,t\,;N)dt\right)$$

by another application of (2.16) and Fubini.

□

Viewing probabilities on Ω as the joint distribution of the random elements X and N, we have so far determined these probabilities from the distribution of X and the conditional distribution of N given X. The next result, which provides the basis for our solution to the embedding problem, goes the other way, yielding in particular in part (b) the structure of the conditional distribution of X given N for probabilities satisfying (C). Proposition 3.18 should also be compared to (G) and (3.2).

3.18. Proposition. Suppose that the X_i are i.i.d. μ and that $\mathbb{P}$ satisfies (C).

(a) For any t, given $\mathcal{F}_t$, the failure times $(X_i)_{i \in C(t+) \cup R(t+)}$ are independent such that for $i \in C(t+)$, X_i has hazard $\hat{\mu}$ on $(U_i, \infty]$ and for $i \in R(t+)$, X_i has hazard μ on $(t, \infty]$.

(b) Given $\mathcal{F}$, the failure times $(X_i)_{i \in C(\infty)}$ are independent such that X_i has hazard μ on $(U_i, \infty]$.

Proof. Since (b) is a consequence of (a) for $t \to \infty$, we only prove (a). Let C, R be disjoint subsets of $\{1, \ldots, n\}$. We must show that for all such C, R and all $F \in \mathcal{F}_t \cap (C(t+) = C,\ R(t+) = R)$, $x_i > 0$ for $i \in C$, $x_i > t$ for $i \in R$,

(3.19) $$\mathbb{P}(X_i > x_i,\ i \in C \cup R,\ F) = \mathbb{P}\Big(\prod_{i \in C} G(U_i \vee x_i | U_i) \prod_{i \in R} G(x_i | t)\ ; F\Big).$$

On F, $D(t+) = D = \{1, \ldots, n\} \setminus (C \cup R)$. Defining $A = \{x' \in L : x'_i > x_i,\ i \in C \cup R,\ x'_i \le t,\ i \in D\}$ and writing $a_{kp} = \frac{k}{2^p} t$, the left hand side becomes, when approximating the values of the U_i for $i \in C$,

$$\lim_{p \to \infty} \sum_{k_i} \mathbb{P}(a_{k_i - 1, p} < U_i \le a_{k_i, p},\ i \in C,\ X \in A,\ F),$$

the sum extending over indices $k_i = 1, \ldots, 2^p$ for $i \in C$. Because $U_i < \infty$ forces $X_i > U_i$ we can in the limit add the condition $X_i > a_{k_i, p}$ and using (2.16) the expression then becomes

(3.20) $$\lim_{p \to \infty} \sum_{k_i} \int_{A,\ x'_i > a_{k_i, p},\ i \in C} P_{x'}(a_{k_i - 1, p} < U_i \le a_{k_i, p},\ i \in C,\ F)\, Q(dx').$$

On the domain of integration, $x'_i > x_i \vee a_{k_i, p}$ for $i \in C$ and $x'_i > x_i > t$ for $i \in R$. The integrand is the $P_{x'}$-probability of a set $B \in \mathcal{F}_t$, hence, by the note following (2.9), it is determined from the behaviour of $\Lambda_{x'}$ on this set B. But the constraints imposed on U_i for $i \in C$ on the set B and the fact that $B \subseteq (R(t+) = R)$ together with (C) shows $\Lambda_{x'} \equiv \Lambda_z$ on B for any $z \in L$ such that $z_i > a_{k_i,p}$ for $i \in C$, $z_i > t$ for $i \in R$ and $z_i = x'_i$ for $i \in D$. Hence we may replace the integrand $P_{x'}(B)$ by

$$\prod_{i \in C} \bar{G}^{1}(a_{k_i,p}) \prod_{i \in R} \bar{G}^{1}(t) \int_{\substack{z_i > a_{k_i,p},\, i \in C \\ z_i > t,\, i \in R}} P_z(B) \prod_{i \in C \cup R} F(dz_i)$$

where $z_i = x'_i$ for $i \in D$. Inserting this in (3.20), relabelling x'_i into z_i and z_i into x'_i for $i \in C \cup R$, using Fubini and the product structure of Q, we arrive at a new expression for (3.20),

$$\lim_p \sum_{k_i} \int_{A_0} P_{x'}(B) \prod_{i \in C} G(a_{k_i,p} \vee x_i | a_{k_i,p}) \prod_{i \in R} G(x_i | t) Q(dx')$$

where $A_0 = \{x' \in L : x'_i > a_{k_i,p}, i \in C, x'_i > t, i \in R, x'_i \le t, i \in D\}$.

Because of (2.16) this becomes

$$\lim_p \sum_{k_i} \mathbb{P}(a_{k_i-1,p} < U_i \le a_{k_i,p}, i \in C, X \in A_0, F)$$
$$\prod_{i \in C} G(a_{k_i,p} \vee x_i | a_{k_i,p}) \prod_{i \in R} G(x_i | t).$$

Since

$$\lim_p \sum_{k_i} 1_{(a_{k_i-1,p} < U_i \le a_{k_i,p}, i \in C, X \in A_0, F)}$$
$$\prod_{i \in C} G(a_{k_i,p} \vee x_i | a_{k_i,p}) \prod_{i \in R} G(x_i | t)$$
$$= 1_F \prod_{i \in C} G(U_i \vee x_i | U_i) \prod_{i \in R} G(x_i | t),$$

dominated convergence provides the last step towards (3.19). □

The solution to the embedding problem is provided by the next result.

3.21. Theorem. Let Λ be the intensity for a FC - process P and suppose that for $i = 1,\ldots,n$, $t > 0$

$$\Lambda^i(dt) = \mu(t)\, I_i(dt)\, dt \tag{3.22}$$

with μ some hazard function. Then there is exactly one probability $\mathbb{P}$ on Ω which satisfies (C) and is such that the X_i are i.i.d. μ and N has intensity Λ .

Proof. Proposition 3.18 shows that there can be at most one $\mathbb{P}$ meeting the requirements, and part (b) even tells us what it must look like. It does not appear to be easy to show that this candidate makes the X_i i.i.d. μ and satisfies (C), so we shall use a different approach. Given Λ satisfying (3.22) we shall solve (3.7),(3.8) for the conditional intensities Λ_x , show that they satisfy (C) and then define $\mathbb{P}$ by assuming the X_i to be i.i.d. μ with Λ_x the intensity for the conditional distribution of N given $X = x$. From Theorem 3.6 it will follow that N has intensity Λ , and the proof will be complete.

With Λ given such that (3.22) holds, solving (3.7),(3.8) for the Λ_x suggests that

$$\Lambda_x(c,B) = \Lambda(c,B) \tag{3.23}$$

and that the point mass for $\Lambda_x^{(i,A)}$ ought to be

$$\Delta\Lambda_x^{(i,A)}(x_i) = \frac{d\Lambda^{(i,A)}}{d\Lambda^i}(x_i)\, I_i(x_i) \quad , \tag{3.24}$$

where $\frac{d\Lambda^{(i,A)}}{d\Lambda^i}$ is the (pathwise) Radon - Nikodym derivative of $\Lambda^{(i,A)}$ with respect to Λ^i .

The Λ_x must satisfy (2.17)-(2.19). Defining $\Lambda_x^{(c,B)}$ by (3.23) gives a problem with (2.17), while of course (3.24) as it stands is useless, since globally the derivative is uniquely determined only Λ^i almost everywhere, with Λ^i absolutely continuous, and we are interested in its value at one particular point.

To solve the first of these problems, modify $\Lambda^{(c,B)}$ and define

$$\tilde{\Lambda}^{(c,B)} = \Lambda^{(c,B)} - \sum_{i \in D(\infty)} \Delta \Lambda^{(c,B)}(X^*_i) \, \varepsilon_{X^*_i} ,$$

i.e. the discontinuities at the observed failure times are removed. We claim that if P is the FC - process with intensity Λ, then $\tilde{\Lambda}^{(c,B)}$ is P - indistinguishable from $\Lambda^{(c,B)}$, i.e. the $\tilde{\Lambda}^{(c,B)}$ are also censoring intensities for P. This assertion amounts to showing that

$$(3.25) \qquad P(\Delta \Lambda^{(c,B)}(X^*_i) ; X^*_i < \infty) = 0$$

for all i. But if finite, X^*_i is a jump time τ_n for N and the jump is of the form (i,A). And

$$P(\Delta \Lambda^{(c,B)}(X^*_i) ; \tau_n = X^*_i < \infty , \; Y_n = (i,A))$$

$$= P(\Delta \Lambda^{(c,B)}(\tau_n) ; \tau_n < \infty , i \in R(\tau_n -) , Y_n = (i,A))$$

$$= P(\pi_{n-1,\xi_{n-1}}(\tau_n,(i,A)) \, \Delta \Lambda^{(c,B)}(\tau_n) ; \tau_n < \infty , i \in R(\tau_n -))$$

as is seen conditioning on $\mathcal{F}_{\tau_n -}$. Now

$$\Delta \Lambda^{(c,B)}(\tau_n) \leq \Delta \bar{\Lambda}(\tau_n) = \Delta \nu_{n-1,\xi_{n-1}}(\tau_n)$$

so that

$$\pi_{n-1,\xi_{n-1}}(\tau_n,(i,A)) \, \Delta \Lambda^{(c,B)}(\tau_n)$$

$$\leq \pi_{n-1,\xi_{n-1}}(\tau_n,(i,A)) \, \Delta \nu_{n-1,\xi_{n-1}}(\tau_n) = \Delta \Lambda^i(\tau_n)$$

which is 0 since Λ^i is absolutely continuous by assumption, and (3.25) follows.

As the final definition of $\Lambda_x^{(c,B)}$ we now use

$$\Lambda_x^{(c,B)} = \tilde{\Lambda}^{(c,B)} . \tag{3.26}$$

To pick out the correct value of the derivative in (3.24) we proceed as follows: by assumption, for any $w \in W$ the measures $\Lambda^{(i,A)}(\cdot,w)$ and $\Lambda^{i}(\cdot,w)$ are absolutely continuous. Hence $t \to \Lambda^{(i,A)}(t,w)$, $t \to \Lambda^{i}(t,w)$ are differentiable almost everywhere (with respect to Lebesgue measure ℓ), and defining for any $x \in L$,

$$\Delta\Lambda_x^{(i,A)}(x_i,w) = \liminf_{k\to\infty} \frac{\Lambda^{(i,A)}(x_i,w)-\Lambda^{(i,A)}(x_i-\frac{1}{k},w)}{\Lambda^{i}(x_i,w)-\Lambda^{i}(x_i-\frac{1}{k},w)} I_i(x_i,w) \tag{3.27}$$

if $A \neq \emptyset$, with the limit 0 if the denominator is 0 for k large, and

$$\Delta\Lambda_x^{(i,\emptyset)}(x_i,w) = (1 - \sum_{A\neq\emptyset} \Delta\Lambda_x^{(i,A)}(x_i,w))\, I_i(x_i,w) , \tag{3.28}$$

it is clear that with w,i fixed, for ℓ- almost all x_i , (3.24) holds for all A .

Adding the requirement that $\Lambda_x^{(i,A)}$ be concentrated at x_i , (3.26)-(3.28) provides an explicit definition of Λ_x on all of W . It is immediate that each Λ_x is predictable and satisfies (i) - (iv) of Fact 2.12, and also (2.17)-(2.19), so that each Λ_x is the intensity for a FC - process with paths in W_x . Since $\Lambda_x^{(c,B)}$ does not depend on x , the censoring intensities obviously satisfy (C). To check that (C) holds for $\Lambda_x^{(i,A)}$, fix t,w and consider $x,x' \in L$ both t - compatible with w , and such that $x_j,x'_j > t$ for all $j \in R(t-,w)$. Then $\Lambda_x^{(i,A)}(\cdot,w)$ and $\Lambda_{x'}^{(i,A)}(\cdot,w)$ are both 0 on $[0,t]$, unless, say, $x_i \leq t$ and $i \in R(x_i-,w)$. Here $x_i = t$ is impossible by assumption, and then x,x' t - compatible with w forces $x'_i = x_i$. Since $\Lambda_x^{(i,A)}$ $(\Lambda_{x'}^{(i,A)})$ depends on x through x_i (x'_i) only (see (3.27) , (3.28)) , $\Lambda_x^{(i,A)}(t,w) = \Lambda_{x'}^{(i,A)}(t,w)$ follows.

Now consider the FC - process N constructed from X_i which are i.i.d. μ and with the intensity for N given $X = x$ equal to Λ_x . The intensity for N is given by Theorem 3.6, and we must show that the $\Lambda^{(c,B)}$ in (3.8) equals $\tilde{\Lambda}^{(c,B)}$ which is just (3.26), and that the $\Lambda^{(i,A)}$ from (3.7) equal the $\Lambda^{(i,A)}$ for the P we started off with. Here we give the proof if $A \neq \emptyset$, the case $A = \emptyset$ being an easy consequence. But by (3.27)

$$\Delta \Lambda^{(i,A)}_{X|i,t}(t) = \liminf_{k \to \infty} \frac{\Lambda^{(i,A)}(t) - \Lambda^{(i,A)}(t-\frac{1}{k})}{\Lambda^{i}(t) - \Lambda^{i}(t-\frac{1}{k})} I_i(t) .$$

Fixing a value $X = x$ and a path w , as was noted above, for $\Lambda^{i}(\cdot,w)$ almost all t

$$\Delta \Lambda^{(i,A)}_{X|i,t}(t,w) = \frac{d \Lambda^{(i,A)}}{d \Lambda^{i}} (t,w) \; I_i(t,w) .$$

Using the assumption (3.22), then

$$\int_0^s \mu(t) \, \Delta \Lambda^{(i,A)}_{X|i,t}(t) \, I_i(t) \, dt$$

$$= \int_0^s \frac{d \Lambda^{(i,A)}}{d \Lambda^{i}} (t) \, \Lambda^{i}(dt) = \Lambda^{(i,A)}(t) .$$

□

For the sake of completeness, we include a proof of Theorem 3.1, different from Gill's original argument.

<u>Proof of Theorem 3.1</u>. We must show that if $\mathbb{P}$ satisfies (G), and P as usual is the marginal distribution of N , then for $s < t$, $F \in \mathcal{F}_s$, $i = 1,\ldots,n$

$$P(N^{i}(t) - N^{i}(s) \, ; F) = P(\int_s^t \mu(u) \, I_i(u) du \, ; \, F) .$$

The integrand on the left is 0 unless $s < X^*_i \leq t$. The idea is to use discrete approximations to the value of $X^*_i = X_i$. Introducing $a_{k,p} = s + (t-s)\frac{k}{2^p}$ we find

$$P(N^i(t) - N^i(s) ; F) = \lim_{p\to\infty} \sum_{k=1}^{2^p} \mathbb{P}(a_{k-1,p} < X_i \le a_{k,p}, F, U_i = \infty)$$

$$= \lim \sum \mathbb{P}(a_{k-1,p} < X_i \le a_{k,p}, F, i \in R(a_{k-1,p}+)) ,$$

and conditioning on $F_{a_{k-1,p}}$ and using (b), this becomes

$$\lim \sum P(\int_{a_{k-1,p}}^{a_{k,p}} \mu(u) G(u|a_{k-1,p}) I_i(a_{k-1,p}+) du ; F) .$$

Taking the summation inside P, the integrand becomes

$$\int_s^t \sum 1_{(a_{k-1,p}, a_{k,p}]}(u) \mu(u) G(u|a_{k-1,p}) I_i(a_{k-1,p}+) du ,$$

which by dominated convergence tends to

$$\int_s^t \mu(u) I_i(u) du$$

as $p \to \infty$. Another application of dominated convergence completes the proof. □

The basis for all results in this section is the crucial condition (C). In view of its importance, we shall now indicate how an equivalent version of (C) may be obtained, which also shows how to simulate FC-processes with the martingale property (2.13).

Any observation of N starts with a number of censorings preceding the first observed failure. As long as only censorings occur, by (C) the specific values of all the failure times are immaterial except of course that they are known to exceed the corresponding observed censorings or the right endpoint of this initial interval of observation. This means that to begin with, the censorings are independent of the failure times, and we arrive at the following simulation procedure, which is updated at each observed failure:

Step 0. Generate X_i which are i.i.d. μ .

Step 1. Generate a vector $(U_1^{(1)},\ldots,U_n^{(1)})$ of possible censoring times, $0 < U_i^{(1)} \leq \infty$, stochastically independent of X . Find the smallest X_i , X_{i_1} say, such that $X_{i_1} \leq U_{i_1}^{(1)}$. This X_{i_1} is to be the first observed failure time. On $[0,X_{i_1})$ at times $U_j = U_j^{(1)}$ those j are censored for which $U_j^{(1)} < X_{i_1}$. Call C_1 the set of these j . All i equal to a $j \in C_1$ and i_1 have now been removed, leaving a set R_1 of individuals. The unused $U_j^{(1)}$, $j \in i_1 \cup R_1$ are discarded.

Step 2. Generate a vector $(U_i^{(2)})_{i \in R_1}$ of possible censoring times, $X_{i_1} \leq U_i^{(2)} \leq \infty$, using a distribution depending on i_1 , X_{i_1} , C_1 , $(U_j)_{j \in C_1}$, but independent of $(X_i)_{i \in R_1}$. Find the smallest X_i for $i \in R_1$, X_{i_2} say, such that $X_{i_2} \leq U_{i_2}^{(2)}$. This X_{i_2} is to be the second observed failure time. On $[X_{i_1}, X_{i_2})$ at times $U_j = U_j^{(2)}$ those $j \in R_1$ are censored for which $U_j^{(2)} < X_{i_2}$. Call C_2 the set of these j . All i equal to a $j \in C_2$ and i_2 have now been removed from R_1 , leaving a set R_2 of individuals. The unused $U_j^{(2)}$, $j \in i_2 \cup R_2$ are discarded.

It should be clear how the simulation procedes. The k'th vector of possible censoring times $(U_i^{(k)})_{i \in R_{k-1}}$ are chosen from a distribution depending on $i_1,\ldots,i_{k-1}$, $X_{i_1},\ldots,X_{i_{k-1}}$, $C_1,\ldots,C_{k-1}$, $(U_j)_{j \in C_1 \cup \ldots \cup C_{k-1}}$, i.e. everything observed on $[0,X_{i_{k-1}})$ and i_{k-1} , $X_{i_{k-1}}$, but independent of $(X_i)_{i \in R_{k-1}}$.

3.29. Example. Consider the scheme for progressive type II censorship (mentioned in Section 2 following problems I - III), where at the time of the k'th observed failure a fixed number r_k of individuals are chosen at random from those still at risk and censored concurrently with the failure. Thus $X_{(1)} = \min X_i$ is the first observed failure time, $|R_1| = |R(X_{(1)}+)| = n-1-r_1$ and the size of the risk set just after the k'th observed failure is $|R_k| = n-k-(r_1+\ldots+r_k)$.

Choosing X_i i.i.d. μ, the conditional intensities Λ_x are specified by $\Lambda_x^{(c,B)} \equiv 0$ and

$$\Delta\Lambda_x^{(i,A)}(x_i) = \frac{1}{\binom{|R(x_i-)|-1}{r_{\tilde{N}(x_i-)+1}}} 1_{(i \cup A \subseteq R(x_i-))}$$

for all subsets A of $\{1,\ldots,n\} \setminus \{i\}$ of cardinality $r_{\tilde{N}(x_i-)+1}$ and $=0$ otherwise. It is immediate to verify that the Λ_x satisfy (C).

Alternatively, going through the first two steps of the simulation procedure, it is also clear that all $U_i^{(1)} \equiv \infty$, while given that the first failure occurs for i_1 at x_{i_1}, r_1 of the $j \neq i_1$ are selected at random and $U_j^{(2)} \equiv x_{i_1}$ for these j, $U_j^{(2)} \equiv \infty$ for the remaining j. □

3.30. Example. Suppose $n=1$. It is easy to characterize all pairs $X = X_1$, $U = U_1$ of one failure time, one censoring time, for which the derived FC-process satisfies the martingale condition (2.13), even without using our convention (D) that $U = \infty$ when $X \leq U$. Equivalently, see the discussion in connection with (2.14), this characterizes all pairs (X,U) satisfying Lagakos' constant sum condition.

On some probability space, set up the random variable X together with two other random variables $U', V > 0$ such that X has hazard μ, X is independent of U' and $V \geq X$. Defining

$$U = \begin{cases} U' & \text{if} \quad U' < X \\ V & \text{if} \quad U' \geq X , \end{cases}$$

the FC-process

$$N^{(1,\emptyset)}(t) = 1_{(X \leq t, X \leq U)} , N^{(c,1)}(t) = 1_{(U \leq t, U < X)}$$

determined from (X,U) satisfies (2.13).

Having (D) satisfied corresponds to taking $V \equiv \infty$.

Condition (C) is equivalent to the following, writing $x = x_1$,

$$\Delta \Lambda_x^{(1,\emptyset)}(x) = 1_{(1 \in R(x-))} ; \Lambda_x^{(c,1)}(dt) = \nu(dt) 1_{(1 \in R(t-))}$$

with ν the hazard measure for the distribution of U'.

□

We shall conclude this section with the counterexamples showing that the two implications in Proposition 3.4 cannot be reversed.

3.31. Example. Suppose $n = 2$ with X_1, X_2 i.i.d. μ and consider Λ_x with only the components $\Lambda_x^{(i,\emptyset)}$ for $i = 1,2$ and $\Lambda_x^{(c,i)}$ for $i = 1,2$ not identically 0 and of the form, with $x = (x_1, x_2)$ and $a \neq 0$ given

(3.32) $$\Lambda_x^{(i,\emptyset)}(dt) = \varepsilon_{x_i}(dt) 1_{(i \in R(t-))} , \quad i = 1,2$$

(3.33) $$\Lambda_x^{(c,1)}(dt) = dt \, 1_{(1 \in R(t-), x_1 > t)} ,$$

(3.34) $$\Lambda_x^{(c,2)}(dt) = \varepsilon_{x_1+a}(dt) 1_{(N^{(c,1)}(t-)=1, 2 \in R(t-), x_2 > t)} .$$

Here (3.32) comes from (2.19), while (3.33) shows that individual 1 is censored at an exponential time, independent of everything else, provided this time is $< x_1$. Finally (3.34) shows that individual 2 can only be censored at time $x_1 + a$ (provided of course that $x_1 + a > 0$), and then only if the failure time x_1 is not observed.

This dependence on an unobserved x_1 shows that the Λ_x do not satisfy (C). We shall now argue that for all $a \neq 0$, (G) holds, while (M) holds if $a > 0$ but not when $a < 0$.

As already used previously, (M) essentially amounts to the following: for any $t > 0$, if $x, x' \in L$ are such that either $x_i = x'_i$ or $x_i, x'_i > t$, then $\Lambda_x \equiv \Lambda_{x'}$ on $[0,t]$. By inspection it is clear that (3.32)-(3.34) implies this and (M) if $a > 0$, but not if $a < 0$, not even when allowing for a Q_μ-null set of exceptional x, x'.

Now suppose that $a < 0$. To show that (G) is satisfied, consider the following condition on the conditional intensities:

(G') For any $t > 0$ and any $w \in W$, $\Lambda_x(t,w) = \Lambda_{x'}(t,w)$ whenever x, x' are t-compatible with w and satisfy that for all i either $x_i = x'_i$ or $i \in R(t+,w)$ and $x_i, x'_i > t$.

Trivially (3.32)-(3.34) implies (G'): the only problem is the $(c,2)$-intensity, where the dependence on x_1 is ruled out since the intensity vanishes if $1 \in R(t+)$.

The proof that (3.32)-(3.34) implies (G) is completed by observing that (G') ⇒ (G). This may be argued along the lines of the proof of Proposition 3.18. We omit the details.

Tedious but straightforward calculations yield the following expression for the marginal intensity Λ for N:

$$\Lambda^{(i,\emptyset)}(dt) = \mu(t) I_i(t) dt , \quad i = 1,2$$

$$\Lambda^{(c,1)}(dt) = I_1(t) dt ,$$

$$(3.35) \qquad \Lambda^{(c,2)}(dt) = \begin{cases} \mu(t-a) I_2(t) 1_{(U_1 < t-a)} dt & \text{if} \quad a > 0 \\ \mu(t-a) \dfrac{G(t-a)}{G(t-a) + G(\ddot{U}_1) - G(\ddot{U}_1 - a)} I_2(t) 1_{(U_1 < t)} dt & \\ & \text{if} \quad a < 0 . \end{cases}$$

By Theorem 3.21 there is a unique embedding of this FC - process such that (C) holds for the joint distribution of X and N. Performing this embedding leads to conditional intensities given by (3.32), (3.33) and replacing (3.34) by letting $\Lambda_x^{(c,2)}$ be given by the right hand side of (3.35), cf. (3.8).

□

4. Statistical models - problem III.

We shall discuss models of FC - processes obtained as families of marginal distributions of N from families of joint distributions of (X,N), where each such distribution renders the X_i i.i.d. with some hazard μ and satisfies condition (C).

The model is specified by allowing μ to be arbitrary, and for each μ considering for all $x \in L$ a family of conditional intensities Λ_x, all of which obey (C). The marginal models for N constructed this way, we shall call FC(C) - models.

With the models built in this manner, it is guaranteed that (2.13) holds for all μ, but as stressed earlier, that is only a minimal requirement for the classical estimators to apply and a property of each distribution in the model rather than a property of the model as a whole.

Our purpose now is to solve problem III, i.e. to discuss models where no essential information about μ is lost when using e.g. the Nelson - Aalen estimator.

We shall first consider the structure of the likelihood function. Suppose N is observed on $[0,t]$, and that the likelihood $L(t)$ has the form

$$(4.1) \qquad L(t) \propto \exp(-\int_0^t \mu(s)\,|R(s-)|\,ds) \prod_{i:X_i^* \le t} \mu(X_i^*)$$

apart from factors that do not depend on the unknown hazard μ. This form of L certainly arises if there are no censorings, and is also known to be valid in other cases, see (5.2) of Kalbfleisch and Prentice [8], Lagakos [10].

The second ingredient we shall need is the concept of non - innovation introduced by Arjas and Haara [2], condition (A). Here we shall consider all failures as innovative and all censorings as non - innovative.

From our point of view, a quite natural condition on FC(C) - models is that the family of conditional intensities should not depend on μ . With this and the preceding remarks in mind, we shall show the following result.

4.2. Theorem. For a FC(C) - model the following three conditions are equivalent:

(i) For all x , the family of conditional intensities Λ_x may be chosen not to depend on μ .

(ii) For all t , the likelihood function for observation of N on $[0,t]$ is proportional to

(4.3)
$$\exp(-\int_0^t \mu(s)\,|R(s-)|\,ds) \prod_{i:X^*_i \leq t} \mu(X^*_i) \ .$$

(iii) All censorings are non - innovative in the sense of Arjas and Haara.

Remark. The qualification 'may be chosen' in (i), is the necessary safeguard against μ - dependent choices of Λ_x for an exceptional set of x - values.

Proof. The likelihood function for observation of N on $[0,t]$ is (cf. (2.9))

(4.4)
$$\prod_{\substack{0 < s \leq t \\ s \neq \tau_k}} (1 - \overline{\Lambda}(ds)) \prod_{k=1}^{\overline{N}(t)} \Lambda^{Y_k}(d\tau_k)$$

with $d\tau_k$ an infinitesimal neighbourhood to the left of and including τ_k.

Since we are dealing with a FC(C) - model, Theorem 3.6 applies and yields an expression for the likelihood in terms of the Λ_x .

In particular

$$\bar{\Lambda}(ds) = \mu(s)|R(s-)|ds + \sum_B \Lambda_X^{(c,B)}(ds) ,$$

with the first term absolutely continuous. Recalling the description in Section 2 of the product integral, it emerges that

$$(4.5) \qquad \prod_{\substack{0<s\leq t \\ s\neq\tau_k}} (1-\bar{\Lambda}(ds)) = \exp(-\int_0^t \mu(s)|R(s-)|ds) \prod_{0<s\leq t} (1-\sum_B \Lambda_X^{(c,B)}(ds)) .$$

By (3.8) and (3.7), the contribution to the likelihood from the observed events take the form

$$(4.6) \qquad \Lambda_X^{(c,B)}(d\tau_k)$$

for pure censorings and, with $\tau_k = X^*_i$

$$(4.7) \qquad \mu(X^*_i)\,\Delta\,\Lambda^{(i,A)}_{X|i,X^*_i}(X^*_i)\,d\,X^*_i$$

for the observation of a failure.

Combining (4.4)-(4.7) it is seen that the likelihood may be written as (4.3) times a factor determined exclusively by the Λ_X, and that this factor does not depend on μ iff (i) holds. Thus (i) and (ii) are equivalent.

In [2] the type set (mark space) E is split into two parts E' and E''. In our case

$$E' = \{(i,A) : 1\leq i\leq n , A\subseteq\{1,\dots,n\}\setminus\{i\}\} ,$$

$$E'' = \{(c,B) : \emptyset\neq B\subseteq\{1,\dots,n\}\} .$$

Recognizing that because the intensities $\Lambda^{(i,A)}$ are absolutely continuous the ρ_t defined on p.198 of [2] vanishes, it is easy to see that condition (A) of [2] for the censorings to be non-innovative amounts to the condition that for all probabilities in the model, the following is true: for all i,A, $d\Lambda^{(i,A)}/d\Lambda^i$ and for all B , $\Lambda^{(c,B)}$ must not depend on μ . But by (3.7) and (3.8)

$$\frac{d\Lambda^{(i,A)}}{d\Lambda^{i}}(s) = \Delta\Lambda^{(i,A)}_{X|i,s}(s)\; I_i(s) \,,$$

$$\Lambda^{(c,B)} = \Lambda^{(c,B)}_{X} \,,$$

and it should be clear that (i) and (iii) are equivalent.

□

5. Asymptotic theory.

In the preceding sections, n , the number of individuals has been fixed. We shall now review the asymptotic distribution results available as $n \to \infty$. For this we shall rely on the martingale property (2.13), but need not specifically work with FC(C) - models.

As always we assume the unknown hazard function μ to satisfy $\int_0^t \mu < \infty$ for all t . For each value of n , there is a FC - process N_n involving n individuals such that for $i = 1,\dots,n$, the compensator for N_n^i is

$$\Lambda_n^i(t) = \int_0^t \mu(s)\, I_{n,i}(s)\, ds \ ,$$

with $I_{n,i}(s)$ the indicator that i is at risk just before s .

Then $\tilde{N}_n = \sum_{i=1}^{n} N_n^i$ has compensator

$$\tilde{\Lambda}_n(t) = \int_0^t \mu(s)\, \rho_n(s)\, ds \tag{5.1}$$

with $\rho_n(s) = |R_n(s-)| = \sum_{i=1}^{n} I_{n,i}(s)$ the size of the risk set just before s .

Thus, based on data from n individuals,

$$\hat{\beta}_n(t) = \int_{(0,t]} \frac{1}{\rho_n(s)}\, \tilde{N}_n(ds)$$

is the Nelson - Aalen estimator of $\int_0^t \mu$ and

$$\hat{G}_n(t) = \prod_{0 < s \leq t} \left(1 - \frac{\tilde{N}_n(ds)}{\rho_n(s)}\right)$$

is the corresponding Kaplan - Meier estimator.

In particular $\hat{\beta}_n - \beta_n$ is a martingale, where

$$\beta_n(t) = \int_0^t \mu(s)\, 1_{(\rho_n(s) > 0)}\, ds$$

and

$$<\hat{\beta}_n - \beta_n>(t) = \int_0^t \mu(s) \frac{1}{\rho_n(s)}\, 1_{(\rho_n(s) > 0)}\, ds \ . \tag{5.2}$$

Also, if we define

$$\hat{\omega}^2_n(t) = \int_{(0,t]} \frac{n}{\rho^2_n(s)} \tilde{N}_n(ds) ,$$

$\hat{\omega}^2_n - n < \hat{\beta}_n - \beta_n >$ is a martingale.

Because of (5.1) we are dealing with a multiplicative intensity model for observation of N_n. But the model has a particularly simple structure because the process ρ_n is decreasing and integer valued, which makes it possible to obtain asymptotic results under much less restrictive conditions than are required for the general multiplicative intensity models.

All processes we encounter have sample paths in the Shorokhod space $D[0,\infty)$. So below, convergence in distribution of a sequence of processes means weak convergence of the distributions, when using the Shorokhod $D[0,\infty)$ topology on the space of sample paths.

5.3. Theorem. Suppose that with μ the true value of the hazard for the distribution of the failure times, it holds that for F_μ - almost all $t \in [0,\infty)$ there is a constant $\theta(t) > 0$ such that

(5.4) $$H_n(t) := \frac{1}{n} \rho_n(t) \xrightarrow[n\to\infty]{} \theta(t)$$

in probability. Suppose also that for every t there is a constant $K(t) > 0$ such that

(5.5) $$P(H_n(t) > K(t)) \xrightarrow[n\to\infty]{} 1 .$$

Then the following statements are valid:

(i) The sequence of processes $(\sqrt{n}\,(\hat{\beta}_n - \beta_n))$ converges in distribution to the mean zero, continuous Gaussian process which has independent increments and variance function

$$\sigma^2(t) = \int_0^t \frac{1}{\theta(s)} \mu(s) ds .$$

(ii) The sequence of processes $(\sqrt{n}\,(\hat{G}_n - G_\mu))$ converges in distribution to the mean zero, continuous Gaussian process with covariance function

$$r(s,t) = \sigma^2(s)G_\mu(s)G_\mu(t) \qquad (s \le t) .$$

(iii) The sequence $(\hat{\beta}_n)$ is a uniformly consistent estimator of $\int_0^{\cdot} \mu$ in the sense that for all t

$$P(\sup_{s \le t} (\hat{\beta}_n(s) - \beta_n(s))^2) \xrightarrow[n\to\infty]{} 0 .$$

If in addition it holds that

$$(5.6) \qquad \int_0^t \mu(s)\, P\left(\frac{n^2}{\rho_n^3(s)}\, 1_{(\rho_n(s) > 0)}\right) ds \xrightarrow[n\to\infty]{} 0 ,$$

then also

(iv) The sequence $(\hat{\omega}_n^2)$ is a uniformly consistent estimator of σ^2 in the sense that for all t

$$\sup_{s \le t} |\hat{\omega}_n^2(s) - \sigma^2(s)| \xrightarrow[n\to\infty]{} 0$$

in probability.

Remarks. The main condition is (5.4). It automatically implies (5.5) except in the rather special case, where $F_\mu((t,\infty)) = 0$ for some t. In that case, because $\int_0^t \mu < \infty$ for all t, F_μ necessarily has an atom at ∞. Instead of (5.4), (5.5) as they stand, one may of course assume that the convergence (5.4) obtains for all rather than just F_μ - almost all t. But as Example 5.13 will show, that is too restrictive.

Notation. For each n, there is a probability P_n relating to the n'th FC-process. In the statement of the theorem, we have written $P = P_n$, and shown the dependence on n by indexing the random variables involved.

<u>Proof</u>. The methods used for proving results like (i) - (iv) are by now standard. Thus, by inspecting e.g. the relevant parts of Chapter 5 of Jacobsen [5], the reader should recognize that (i) follows if it is shown that for all t and all $\varepsilon > 0$,

$$(5.7) \qquad \int_0^t \frac{n}{\rho_n(s)} 1_{\left(0 < \rho_n(s) < \frac{\sqrt{n}}{\varepsilon}\right)}(s)\,ds \xrightarrow[n \to \infty]{} 0$$

in probability, and that for all t

$$(5.8) \qquad n < \hat{\beta}_n - \beta_n > (t) \xrightarrow[n \to \infty]{} \sigma^2(t)$$

in probability. (iii) follows if it is shown in addition that for all t

$$(5.9) \qquad \int_0^t \mu(s)\, P\left(\frac{1}{\rho_n(s)} 1_{(\rho_n(s) > 0)}\right) ds \xrightarrow[n \to \infty]{} 0 \ ,$$

while (ii) apart from (5.7)-(5.9) also requires that

$$(5.10) \qquad \tau_n^* \xrightarrow[n \to \infty]{} \infty$$

in probability, where $\tau_n^* = \inf\{t : \rho_n(t) = 0\}$.

Finally, (iv) is a direct application of Proposition 5.2.24 and Exercise 5.E.1 in [5].

So we must show that (5.4) and (5.5) implies (5.7)-(5.10).

Because $(\tau_n^* < t) \subseteq (H_n(t) = 0)$, (5.10) is immediate from (5.5). To show (5.7), it is because of (5.5) enough to consider the behaviour of the integral on the set $H_n(t) > K(t)$. But then, for all $s \leq t$, $\rho_n(s) \geq \rho_n(t) > n K(t)$ and the integral in (5.7) vanishes for n so large that $n K(t) > \sqrt{n}/\varepsilon$.

The most important step is to verify (5.8). Referring to (5.2) it is seen that (5.8) follows from showing that for all $t > 0$

$$\int_0^t |H_n^{-1}(s) 1_{(\rho_n(s) > 0)} - \gamma(s)| \mu(s) ds \xrightarrow[n \to \infty]{} 0$$

in probability, where $\gamma = \theta^{-1}$. Essentially this amounts to interchanging the order of integration and taking the limit in probability. A most useful tool for doing this is Gill's [4] concept of convergence, boundedly in probability.

Let D denote the set of timepoints t for which the convergence (5.4) holds. Thus, introducing

$$V_n(s) = |H_n^{-1}(s) 1_{(\rho_n(s) > 0)} - \gamma(s)| 1_D(s) \mu(s)$$

for every s,

(5.11) $V_n(s) \longrightarrow 0$

in probability. Since D has full F_μ - measure, it suffices to show that for all t

(5.12) $\int_0^t V_n(s) ds \longrightarrow 0$

in probability. But for this, according to [4] it is enough that the sequence (V_n) of processes converges to 0, boundedly in probability, i.e. apart from the pointwise convergence (5.11), for all $t, \delta > 0$ there exists a function $k_{\delta,t} : [0,t] \to [0,\infty)$ with $\int_0^t k_{\delta,t}(s) ds < \infty$ such that

$$\liminf_{n \to \infty} P(V_n(s) \leqq k_{\delta,t}(s) \quad \text{for all} \quad s \in [0,t]) \geqq 1 - \delta.$$

But the function γ is defined and finite on D and for $t \in D$, $\gamma(t) \leqq K^{-1}(t)$. Since both $H_n^{-1}(s)$ and $\gamma(s)$ are increasing, it is clear from (5.5), that we may use

$$k_{\delta,t}(s) = K^{-1}(t) \mu(s)$$

and the proof is complete. □

<u>5.13. Example</u>. We shall determine the limit θ from (5.14) in a case involving progressive type II censorship. To simplify we shall only allow censorings at one timepoint.

For each n, let n_0, r_1 be integers such that as $n \to \infty$

$$\frac{n_0}{n} \to p \, , \; \frac{r_1}{n} \to \pi$$

where $0 < p, \pi < 1$, $p + \pi < 1$. With $X_1, \ldots, X_n$ the failure times, let $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ denote the order statistics. Then all $X_{(1)}, \ldots, X_{(n_0)}$ are observed and at time $X_{(n_0)}$, r_1 individuals among those still at risk are selected at random and censored, and the remaining $n - n_0 - r_1$ among $X_{(n_0+1)}, \ldots, X_{(n)}$ are all observed.

Let $[x_p^-, x_p^+]$ be the interval of p-fractiles for F_μ, i.e.

$$[x_p^-, x_p^+] = \{x : F_\mu(x) = p\}$$

if $F_\mu(x) = p$ for some x, and $= \{\infty\}$ in the case of a defective F_μ with $F_\mu(x) < p$ for all x. Then, for any $a^- < x_p^-$, $a^+ > x_p^+$,

$$P(X_{(n_0)} \in (a^-, a^+)) \xrightarrow[n \to \infty]{} 1 .$$

Using this, and the fact that given which X_i correspond to $X_{(1)}, \ldots, X_{(n_0)}$ and given $X_{(n_0)} = x$, the remaining X_i are i.i.d. with hazard μ on $[x, \infty)$, the reader can readily verify that θ is determined F_μ-almost everywhere by

$$\theta(t) = \begin{cases} G_\mu(t) , & t < x_p^- \\ (1 - \frac{\pi}{1-p}) G_\mu(t) , & t > x_p^+ . \end{cases}$$

□

Acknowledgement. It is a pleasure to thank Per Kragh Andersen, Ørnulf Borgan, Richard Gill and Niels Keiding for their comments and good advice. Special thanks are due to Richard Gill for telling me about "convergence, boundedly in probability".

References.

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

[2] Arjas, E. and Haara, P. (1984). A marked point process approach to censored failure data with complicated covariates.
Scand. J. Statist. 11, 193 - 209.

[3] Gill, R. D. (1980). Censoring and stochastic integrals.
Mathematical Centre Tracts 124,
Mathematisch Centrum, Amsterdam.

[4] Gill, R. D. (1984). A note on two papers in central limit theory.
Proc. 44th Session ISI, Madrid; Bull. Inst. Internat. Statist. 50, Vol. L, Book 3, 239-243.

[5] Jacobsen, M. (1982). Statistical analysis of counting processes.
Lecture Notes in Statistics 12 , Springer, New York.

[6] Jacod, J. (1975). Multivariate point processes: predictable projection, Radon - Nikodym derivatives, representation of martingales.
Z. Wahrsch. Verw. Gebiete, 31, 235 - 253.

[7] Kalbfleisch, J. D. and MacKay, R. J. (1979). On constant - sum models for censored survival data.
Biometrika 66, 87 - 90.

[8] Kalbfleisch, J. D. and Prentice, R. L. (1980). The statistical analysis of failure time data.
Wiley, New York.

[9] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations.
J. Amer. Statist. Assoc. 53, 457 - 481.

[10] Lagakos, S. W. (1979). General right censoring and its impact on the analysis of survival data.
Biometrics 35, 139 - 156.

[11] Nelson, W. (1969). Hazard plotting for incomplete failure data. J. Qual. Tech. 1, 27 - 52.

[12] Williams, J. S. and Lagakos, S. (1977). Models for censored survival analysis: constant - sum and variable - sum models. Biometrika 64, 215 - 224.

PREPRINTS 1985

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

No. 1 Cohn, Harry: Almost Sure Convergence for Stochastically Monotone Temporally Homogeneous Markov Processes and Applications.

No. 2 Rootzén, Holger: Maxima and Exceedances of Stationary Markov Chains.

No. 3 Asmussen, Søren und Johansen, Helle: Über eine Stetigkeitsfrage betreffend das Bedienungssystem GI/GI/s.

No. 4 Asmussen, Søren and Thorisson, Hermann: A Markov Chain Approach to Periodic Queues.

No. 5 Hald, Anders: Galileo as Statistician.

No. 6 Johansen, Søren: The Mathematical Structure of Error Correction Models.

No. 7 Johansen, Søren and Johnstone, Iain: Some Uses of Spherical Geometry in Simultaneous Inference and Data Analysis.

PREPRINTS 1986

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

No. 1 Jespersen, N.C.B.: On the Structure of Simple Transformation Models.

No. 2 Dalgaard, Peter and Johansen, Søren: The Asymptotic Properties of the Cornish-Bowden-Eisenthal Median Estimator.

No. 3 Jespersen, N.C.B.: Dichotomizing a Continuous Covariate in the Cox Regression Model.

No. 4 Asmussen, Søren: On Ruin Problems and Queues of Markov-Modulated M/G/1 Type.

No. 5 Asmussen, Søren: The Heavy Traffic Limit of a Class of Markovian Queueing Models.

No. 6 Jacobsen, Martin: Right Censoring and the Kaplan - Meier and Nelson - Aalen Estimators.