The Annals of Statistics
1989, Vol. 17, No. 3, 1133–1156

RIGHT CENSORING AND MARTINGALE METHODS FOR FAILURE TIME DATA

BY MARTIN JACOBSEN

University of Copenhagen

Statistical models are considered for (partial) observation of independent, identically distributed failure times, subjected to censoring from the right. A minimal class of censoring patterns is determined under the assumption that the failure intensities are as they would have been without censoring. This class is then used to discuss the statistical models for right censored survival data, where, e.g., the Kaplan–Meier estimator exploits all available information about the failure time distribution. The counting process description of survival data is used throughout.

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 were 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 ([4], 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 [4] in particular used this framework for his study of censoring patterns, and it is also the backbone of the present paper.

Received November 1986; revised May 1988.
AMS 1980 *subject classifications*. Primary 62G05; secondary 62M09, 60G55.
Key words and phrases. Survival analysis, counting process, intensity process, multiplicative intensity model, martingale estimator.

Following [8, 4], the basic assumptions made here are (i) that the failure times are independent and identically distributed and (ii) that the distribution of observed failures and censorings is such that the intensities for individual failures have Aalen's multiplicative structure. These conditions of course only partially specify a model for the observations: What is lacking is precisely a description of the interplay between failures and censorings. The main results below now characterize a minimal class of censoring patterns, which are compatible with (i) and (ii). Formally, the class is described using conditions on the joint distribution of all failure times (observed or not) *and* observed failures and censorings. The class is then shown to be minimal in the sense that there is a one-to-one correspondence between these joint distributions and the class of all (marginal) distributions of the observations alone that agree with (i) and (ii). The class appears to contain all censoring systems studied in the literature.

In Section 2 we introduce the counting process setup used throughout the paper and also formalize the problems to be solved in the following sections. Section 3 contains the definition of the minimal class of censoring patterns and a number of distributional results, valid for every member of the class. Based on the concept of informative versus noninformative censorings (see [8], page 121), the results are used in Section 4 to characterize the statistical models for which the Nelson–Aalen estimator exploits all information about the failure time distribution available from the observations.

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 $X_1, \dots, X_n$ be i.i.d. and let $U_1, \dots, U_n$ be the censoring times with all X_i, U_i strictly positive. Then the observations consist of the pairs

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

where $T_i = X_i \wedge U_i$ and $\delta_i = 1_{(X_i \le U_i)}$ 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 $X_1, \dots, X_n$ and censoring times $U_1, \dots, U_n$ are strictly positive, possibly infinite, random variables such that $U_i = \infty$ whenever $U_i \ge X_i$, and such 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.,

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

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 $s \le t$ we write

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

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

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

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 $t^+ = \inf\{t: \int_0^t \mu = \infty\} < \infty$, much of what is said below will be valid only on the time interval $[0, t^+)$.] We shall however not assume that $\int_0^\infty \mu = \infty$, so the X_i are allowed to take the value $+\infty$. Apart from 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 more of them may coincide or coincide 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 allow 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

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

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

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

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

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

The simplest situation where it is reasonable to use the estimators (2.3) and (2.4) is the model for *random censorship* where the X_i are i.i.d. with hazard μ and the U_i are mutually independent and independent of $(X_1, \dots, X_n)$. [With (D) in force, the description is different, but still very simple.]

On the other hand, it is easy to construct formal censoring patterns, where it is absurd to use the estimators (2.3) and (2.4): If $U_i < X_i$ for all i, (2.3) and (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) and (2.4). Clearly, the censorings in an explicit manner anticipate future failures and this is precisely what must not happen if (2.3) and (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 [4], which is a main reference for counting processes and censoring and the main reference for this paper.

Consider a collection of failure times $X_1, \dots, X_n$ and censoring times $U_1, \dots, U_n$ satisfying (D). In particular it may be assumed that all finite X_i are distinct.

To observe $(T_i, \delta_i)_{1 \le i \le n}$ is equivalent to observing the occurrence in time of a sequence of events. The possible events consist in either a failure simultaneously with a number of censorings or in the occurrence 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)

$$\begin{aligned} E = &\{(i, A): 1 \le i \le n,\ A \subseteq \{1, \dots, n\} \setminus \{i\}\} \\ &\cup \{(c, B): \varnothing \ne B \subseteq \{1, \dots, n\}\} \end{aligned}$$

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." For each $y \in E$ we define $K(y)$, the set of individuals involved in y, by

$$K(y) = \begin{cases} \{i\} \cup A & \text{if } y = (i, A), \\ B & \text{if } y = (c, B). \end{cases} \tag{2.5}$$

In finite time a random number ($\le n$) and a random selection of events are observed. Let τ_k be the time of occurrence of the kth event and let Y_k be the name of the event. If precisely m events, $0 \le m \le n$, are observed on $(0, \infty)$, we have $0 < \tau_1 < \cdots < \tau_m < \infty$. Define $\tau_{m+1} = \cdots = \tau_n = \infty$ and leave $Y_{m+1}, \dots, Y_n$ unspecified.

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 and with $\mathscr{F}_t$ the σ-algebra generated by $(N(s))_{s \le t}$, we may write in an essentially unique fashion

$$N^y = M^y + \Lambda^y,$$

where M^y is a right continuous $\mathscr{F}_t$-martingale for each y and Λ^y is predictable, right continuous and increasing, $\Lambda^y(0) = 0$; cf. Jacod [6] or, for the special case of absolutely continuous compensators, Jacobsen ([5], Section 2.2). [Some proofs below will use the following definition of predictable processes: A process $(Z_t)_{t \geq 0}$ is predictable if it is measurable and for every $t \geq 0$, Z_t is $\mathscr{F}_{t-}$-measurable, where $\mathscr{F}_{0-} = \mathscr{F}_0$ and $\mathscr{F}_{t-} = \sigma(N(s))_{0 \leq s < t}$. Otherwise, any standard definition will do.]

Occasionally it will be convenient to view each Λ^y as a positive random measure on $[0, \infty)$, identifying for each t the measure of the interval $[0, t]$ with the process value $\Lambda^y(t)$.

The distribution of N is uniquely determined by the compensator $\Lambda = (\Lambda^y)_{y \in E}$ as may be seen from the following expression (cf. [5], Proposition 2.5.13): Introduce the total intensity

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

and consider the infinitesimal event that on $(0, t]$ precisely m jumps occur at times in $dt_1, \dots, dt_m$, the kth jump occurring in component y_k. The probability of this event is

$$\begin{aligned} &P(\tau_1 \in dt_1, Y_1 = y_1, \dots, \tau_m \in dt_m, Y_m = y_m, \tau_{m+1} > t) \\ &\qquad = \prod_{\substack{0 < s \leq t \\ s \neq t_k}} (1 - \bar{\Lambda}(ds, w)) \prod_{k=1}^{m} \Lambda^{y_k}(dt_k, w), \end{aligned} \tag{2.6}$$

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

NOTE. Adding up infinitesimal probabilities like (2.6) shows in particular that for any $F \in \mathscr{F}_t$, $P(F)$ is determined by the behaviour of Λ on F.

As a multivariate counting process, N has a special structure. More specifically, each component has at most one jump and if jumps in components y, y' are observed, then $K(y)$ and $K(y')$ are disjoint. Defining W as the space of all sample paths w for N with these properties and defining $N_t(w) = w(t)$ and $\mathscr{F}$, $\mathscr{F}_t$, $\mathscr{F}_{t-}$ as σ-algebras of subsets of W, we introduce

2.7. DEFINITION. A (canonical) *failure-censoring* process is a probability on $(W, \mathscr{F})$.

In the sequel, failure-censoring process is abbreviated FC-process.

We shall need some more notation:

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

registers the failure of i, the sum extending over all $A \subseteq \{1, \dots, n\} \setminus \{i\}$, and

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

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

$$\Lambda^i = \sum_A \Lambda^{(i,A)}, \qquad \tilde{\Lambda} = \sum_i \Lambda^i.$$

With the standard convention $\inf \varnothing = \infty$, introduce the *observed lifetimes*

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

and note that because of convention (D), the censoring times U_i themselves may be given a similar description. In particular all X_i^* and U_i are stopping times with respect to the filtration $(\mathscr{F}_t)$.

At each time point t, the collection $\{1, \dots, n\}$ of individuals splits into three disjoint sets, the *risk set* $R(t-)$, the *censoring set* $C(t-)$ and the observed *failure set* $D(t-)$, where

$$R(t-) = \{i\colon U_i \geq t,\ X_i^* \geq t\},$$
$$C(t-) = \{i\colon U_i < t\},$$
$$D(t-) = \{i\colon X_i^* < t\}.$$

Each of these random sets is $\mathscr{F}_{t-}$-measurable, for instance $R(t-)$ is the set of individuals at risk immediately before t. For $t = \infty$ we obtain the set of individuals always at risk $R(\infty)$, the set of censored individuals $C(\infty)$ and the set $D(\infty)$ of individuals observed to fail. Occasionally we shall use

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

and the analogues $C(t+)$ and $D(t+)$.

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

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.

2.8. FACT. Let $\Lambda = (\Lambda^y)_{y \in E}$ be a collection of processes $\Lambda^y\colon [0, \infty) \to [0, \infty)$ defined on W and consider the following conditions. For all $y \in E$:

(i) Λ^y is right continuous and increasing with $\Lambda^y(0) = 0$.
(ii) $\Lambda^y(t)$ is $\mathscr{F}_{t-}$-measurable for all t.
(iii) $\Delta\overline{\Lambda}(t) \leq 1$ 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) reflects the special jump structure of FC-processes. In (iii) above as elsewhere, Δ is the notation for the size of a jump of a process or function.

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 $X = (X_1, \ldots, X_n)$ is the vector of all failure times (observed or unobserved).

The basic assumption we shall make about the model for the observation N, is that for $i = 1, \ldots, n$,

$$\Lambda^i(dt) = \mu(t) I_i(t)\, dt, \tag{2.9}$$

where

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

i.e., we assume that the model for the observed failures is a *multiplicative intensity model* as introduced by Aalen [1].

This condition, which we shall refer to as the *martingale condition*, was introduced by Gill [4], but is also the same as (5.5) of Kalbfleisch and Prentice [8]. As a consequence of the condition, if P has compensator Λ satisfying (2.9), then 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 martingale, where $\hat{\beta}$ is defined by (2.3) and

$$I(s) = 1_{(R(s-) \neq \varnothing)}.$$

With (2.9) and the preceding remarks in mind, our main purpose is to discuss the following three problems:

1. 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 (2.9)?
2. (The embedding problem.) Supposing the distribution of N to satisfy (2.9) 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?
3. What must be the structure of statistical models for the distribution of N, satisfying (2.9) for all μ, in order that no essential information about μ is lost when using the Nelson–Aalen estimator?

When answering problems 1 and 2, it is enough to fix an arbitrary μ at a time, while 3 involves the structure of the complete model obtained when μ varies.

Gill [4] gave several examples of censoring patterns, where (2.9) is satisfied, including random censorship and progressive type 2 censorship. Also, for the case of i.i.d. pairs (X_i, U_i) studied by Williams and Lagakos [12], it is easily checked that the constant sum condition from [12] as reformulated by Kalbfleisch and MacKay [7] is precisely (2.9). (See Example 3.31 below for further remarks on this case.)

The solutions we shall give of the three problems are presented in detail in the following two sections. But already at this stage we find it important to indicate what the solutions look like and what is achieved by obtaining them.

First a condition is introduced on the structure of the conditional distribution of the observations given *all* the failure times. In a precise dynamical manner this condition states that "present censorings depend only on past observations, but not on unobserved failures." It is then shown that with i.i.d. failure times and this type of conditional distribution, the martingale condition holds.

Thus a collection of censoring mechanisms has been delimited, which fits with the assumptions from problem 1. It is next shown that this collection is minimal in the sense that not only does the embedding problem always have a solution, but the solution is unique within the above mentioned collection. A number of distributional results about this minimal class of censoring patterns are next developed, which because of the answers to the embedding question, one can always assume to be true, as long as one knows the failure times to be i.i.d., (2.9) to hold and is otherwise interested only in the *marginal* distribution of the observations.

For statistical inference, we consider a model which is a given family of marginal distributions for the observations satisfying (2.9) with μ unknown. Assuming also that the failure times are i.i.d. we need only consider the minimal class of censoring systems from above, and are thereby able to write down the total likelihood for all observations, on a given interval of time, and thus it becomes possible to discuss the structure of those models where only the observed failures contribute to the inference about μ.

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 $x = (x_1, \dots, x_n)$ with $0 < x_i \leq \infty$ and such that no two finite x_i are equal. Also write $X_i(x) = x_i$, define $\mathscr{H} = \sigma(X_i)_{1 \leq i \leq n}$ and let $\mathscr{H}_t = \sigma(X_i 1_{(X_i \leq t)})_{1 \leq i \leq n}$ be the σ-algebra measuring all failure times less than or equal to t, whether observed or not. Note that $(X_i > t) \in \mathscr{H}_t$.

A realization of all failure times and the observations is a point $\omega = (x, w) \in L \times W$ such that x and w are *compatible*, i.e., ω belongs to Ω defined as the space of pairs $(x, w) \in L \times W$ 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 $i \in R(\infty, w)$.

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

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

We shall say that $x \in L$ and $w \in W$ are *t-compatible* 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 $X_i(x,w) = x_i$. A set $F \in \mathscr{F}$ may be viewed as the set $F = \{(x,w) \in \Omega\colon w \in F\} \in \mathscr{G}$; a set $H \in \mathscr{H}$ as the set $H = \{(x,w) \in \Omega\colon x \in H\}$. Then also $\mathscr{H}_t$, $\mathscr{F}_t$ may be considered sub-σ-algebras of $\mathscr{G}$ and then $\mathscr{G}_t = \mathscr{H}_t \vee \mathscr{F}_t$, the smallest σ-algebra containing both $\mathscr{H}_t$ and $\mathscr{F}_t$.

We shall construct probabilities on Ω by letting the X_i be i.i.d. μ (the distribution of X is $Q = Q_\mu$). 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 $(W, \mathscr{F})$ with $P_x(W_x) = 1$ and such that $x \to P_x(F)$ is measurable for all $F \in \mathscr{F}$. Then [see (2.11)]

$$\mathbb{P} = \int P_x Q(dx) \tag{2.10}$$

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. If $f \geq 0$ defined on Ω is measurable, then by (2.10),

$$\int f\, d\mathbb{P} = \int Q(dx) \int f(x,w) P_x(dw). \tag{2.11}$$

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 2.8 plus some extra conditions that we now list.

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

$$\Delta\Lambda_x^{(c,B)}(x_i) = 0 \quad \text{on } (i \in R(x_i -)) \tag{2.12}$$

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

$$\Delta\Lambda_x^{(i,A)}(x_i) = \Delta\Lambda_x^{(i,A)}(x_i) 1_{(\{i\} \cup A \subseteq R(x_i -))}. \tag{2.13}$$

Finally,

$$\Delta\Lambda_x^i(x_i) = 1 \quad \text{on } (i \in R(x_i -)) \tag{2.14}$$

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 (2.12)–(2.14) on all of W; it is enough that these conditions hold on W_x.

2.15. LEMMA. *Let $x \in L$ and let $\Lambda_x = (\Lambda_x^y)_{y \in E}$ be the intensity for some FC-process P_x. If* (2.12)–(2.14) *hold for the restriction of Λ_x to W_x, then* $P_x(W_x) = 1$.

For the proof one builds the P_x-process step by step: the time of the first jump, the type of the first jump, the time of the second jump, etc. One then checks at each step that almost all paths for the part of the process constructed so far belong to W_x, and that only knowledge of Λ_x on W_x is required for the construction. We omit the details.

3. Main results—problems 1 and 2. We shall begin by quoting Gill's [4] answer to question 1. 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 $\mathscr{F}_t$ the X_i for $i \in R(t+)$ are i.i.d. on $(t, \infty]$ with hazard μ.

3.1. **Theorem** (Gill [4], Theorem 3.1). *If the joint distribution of* (X, N) *is such that* (G) *is satisfied, then the martingale condition* (2.9) *holds.*

If (G) is true, taking $t = 0$ shows all X_i to be i.i.d. with hazard μ. Also given $\mathscr{F}_t$ 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 $\mathscr{F}_t$, 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 of 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):

(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 that $x_i > t$ 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 (2.14) it is essential that in (C) an x t-compatible with w is required to satisfy $x_i > t$ for $i \in R(t-, w)$ rather than $x_i \geq t$.

Because each $\Lambda_x(t, w)$ is right continuous in t, one verifies easily that (C) is equivalent to the following, seemingly stronger condition (C′) and implies another condition (C″).

(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 $i \in R(t-, w)$.

(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 $j \in R(t-, w)$, $j \neq i$.

In the sequel, when applying (C), we allow either of (C), (C′) or (C″).

Note. 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 $Q(X \in A) = 0$. 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 $\mathscr{H}_t = \sigma(X_i 1_{(X_i \le t)})_{1 \le i \le n}$ and introduce $\mathscr{H}^t = \sigma(X_i 1_{(X_i > t)})_{1 \le i \le n}$, the σ-algebra spanned by those $X_i > t$. Now consider the following condition:

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

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

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

on the set $(\{i\colon X_i > t\} = S) \in \mathscr{H}_t$: By the conditional independence, $\mathbb{P}(H|\mathscr{G}_t) = \mathbb{P}(H|\mathscr{H}_t)$ for $H \in \mathscr{H}^t$ 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

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

we can give yet another version of (M). It is standard terminology to call a random time τ a *randomized stopping time* for the filtration $(\mathscr{H}_t)$, if for any t, $\mathbb{P}(\tau > t|\mathscr{H}) = \mathbb{P}(\tau > t|\mathscr{H}_t)$. 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 $\mathscr{H}$ (i.e., X) leaves only the U_i as random and hence (3.3) holds for all $F \in \mathscr{F}_t$ iff it holds for all $F \in \mathscr{U}_t := \sigma(U_i 1_{(U_i \le t)})_{1 \le i \le n}$. Thus (M) is equivalent to the statement that $(U_1, \dots, U_n)$ is a multivariate randomized stopping time if by this (nonstandard) statement, we mean that (3.3) holds for all t and all $F \in \mathscr{U}_t$. Finally, note that since $\mathbb{P}(\cdot\,|\mathscr{H}) = P_x$ on $(X = x)$, (3.3) is a condition on the P_x.

In a competing risks setting, conditions of a nature similar to (G), (C) and (M) have been proposed by Arjas [2].

3.4. **Proposition.** (C) $\Rightarrow$ (M) $\Rightarrow$ (G) *and neither implication can be reversed.*

Proof. 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′) and (C″) that for all such w, $\Lambda_x(s, w) = \Lambda_{x'}(s, w)$ for $s \le t$. 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 $\mathscr{F}_t$. But conditioning on $\mathscr{H}_t$ amounts precisely to specifying for each i the value of X_i if $X_i \le t$ and the event $(X_i > t)$ otherwise. Thus we have shown that (3.3) and hence (M) holds.

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

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

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

That the implications cannot be reversed will be shown in Example 3.33 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 $\mathscr{F}_{t-}$.*

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 $\mathscr{F}_s$. Since $\mathscr{F}_{t-}$ is the σ-algebra spanned by $(\mathscr{F}_s)_{s<t}$, the lemma follows. □

Our first 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 (2.9) holds. Of course this fact alone follows directly from Gill's theorem and Proposition 3.4, but with the restrictive condition (C) much more can be said.

In the sequel, if $x \in L$, $t > 0$, we write $x|i,t$ for the vector $(x_1,\dots,x_{i-1},t,x_{i+1},\dots,x_n)$. 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*

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

$$\Lambda^{(c,B)} = \Lambda^{(c,B)}_X \tag{3.8}$$

for $i = 1,\dots,n$, $A \subseteq \{1,\dots,n\}\setminus\{i\}$, $\varnothing \neq B \subseteq \{1,\dots,n\}$. 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 $(P, \mathscr{F}_t)$-martingale and* (2.9) *holds,*

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

REMARKS. In the proof we show that each M^y is a $(\mathbb{P}, \mathscr{G}_t)$-martingale.

Equation (3.7) is best understood recalling that the measure $\Lambda^{(i,A)}_x$ is concentrated at x_i. Also note that because of (2.13), the factor $I_i(t)$ may be omitted from (3.7).

PROOF OF THEOREM 3.6. We begin by showing that Λ, which formally depends on both x and w, is determined by w alone, i.e., that whenever

$(x, w) \in \Omega$ and $(x', w) \in \Omega$, 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

$$\Delta\Lambda_{x|i,t}^{(i,A)}(t, w)I_i(t, w) = \Delta\Lambda_{x'|i,t}^{(i,A)}(t, w)I_i(t, w),$$
$$\Lambda_x^{(c,B)}(t, w) = \Lambda_{x'}^{(c,B)}(t, w)$$

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, e.g., by right continuity.

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 $w \sim_{t-} w'$, i.e., $w(s) = w'(s)$ for all $s \in [0, t)$, then $\Lambda^{(i,A)}(t, w) = \Lambda^{(i,A)}(t, w')$. But

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

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 time points s. Thus $\Lambda^{(i,A)}$ is predictable.

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

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

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 $w \sim_{t-} w'$ and choose arbitrary $x, x' \in L$ such that $(x, w) \in \Omega$ and $(x', w') \in \Omega$, 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 $\tilde{\Lambda}^{(c,B)}(t, w) = \tilde{\Lambda}^{(c,B)}(t, w')$. 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 (2.12), $\Delta\Lambda_x^{(c,B)}(t, w) = 0$, 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 (2.14). 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 $(\mathbb{P}, \mathscr{G}_t)$-martingale, which certainly renders M^y a $(P, \mathscr{F}_t)$ – martingale. Recalling that $\mathscr{G} = \Omega \cap (\mathscr{H}_t \otimes \mathscr{F}_t)$ 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 $y \in E$, $s \le t$, $G = \Omega \cap (H \times F)$ with $H \in \mathscr{H}_s$, $F \in \mathscr{F}_s$. But because $M_x^y = N^y - \Lambda_x^y$ is a $(P_x, \mathscr{F}_t)$-martingale, using (2.11) we see that

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

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,

$$\mathbb{P}\left(\Delta\Lambda_{X|i, X_i+\delta}^{(c,B)}(X_i);\ X_i < \infty\right) = 0. \tag{3.14}$$

For every $x \in L$, the function

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

is $\mathscr{F}_{x_i-}$-measurable. Hence by Lemma 3.17 below the integral from (3.14) equals

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

By (C), $\Delta\Lambda_{X|i, t+\delta}^{(c,B)}(t) = \Delta\Lambda_X^{(c,B)}(t)$ for $t < X_i$. Since the measure $\Lambda_X^{(c,B)}$ 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

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

is $\mathscr{F}_{x_i-}$-measurable. So by Lemma 3.17,

$$\begin{aligned}&\mathbb{P}\left(\Delta\Lambda_X^{(i,A)}(X_i); G, s < X_i \le t\right) \\ &\quad = \mathbb{P}\left(\int_0^{X_i} \mu(u)\,\Delta\Lambda_{X|i,u}^{(i,A)}(u)1_G(X|i, u;\ N)1_{(s,t]}(u)\,du\right).\end{aligned}$$

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 $H \in \mathscr{H}_s$, it is seen that this indicator is constant in u, and hence the Lebesgue integral may be written

$$\int_s^{t \wedge X_i} \mu(u)\,\Delta\Lambda_{X|i,u}^{(i,A)}(u)\,du\,1_{(X_i > s, G)}. \tag{3.15}$$

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

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

and we have shown that

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

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.

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

$$\mathbb{P}(f_i;\ X_i < \infty) = \mathbb{P}\left(\int_0^{X_i}\mu(t)f_i(X|i, t;\ N)\,dt\right). \tag{3.18}$$

Note. Since $f_i(x, w)$ depends on w only through the behaviour of w on $[0, x_i)$, $f_i(x, w)$ is well defined whenever x and w are x_i-compatible. Thus, if $(x, w) \in \Omega$ and $t \leq x_i$, since $x|i, t$ and w are t-compatible, $f_i(x|i, t;\ w)$ is defined for the all but finitely many $t \leq x_i$ for which $x|i, t \in L$ and hence the integral in (3.18) is well defined.

Proof of Lemma 3.17. By (2.11),

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

Because $f(x, \cdot)$ is $\mathscr{F}_{x_i-}$-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 and using that Q is a product measure 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 (2.11) 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 3.19 should also be compared to (G) and (3.2).

3.19. 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 μ on (U_i, ∞) and for $i \in R(t+)$, X_i has hazard μ on (t, ∞).*

(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, ∞).*

PROOF. Since (b) is a consequence of (a) for $t \to \infty$, we only consider (a) and argue that for $F \in \mathcal{F}_t$ of the form

$$F = \big(C(t+) = C,\ R(t+) = R,\ U_i \in du_i,\ i \in C,\ X_i \in ds_i,\ i \in D\big),$$

where $t > 0$, C, R are disjoint subsets of $\{1, \dots, n\}$, $D = \{1, \dots, n\} \setminus C \cup R$ and u_i for $i \in C$, s_i for $i \in D$ belong to $(0, t]$, it holds that

$$\mathbb{P}(X_i > x_i,\ i \in C \cup R,\ F) = \prod_C G(x_i | u_i) \prod_R G(x_i | t) \mathbb{P}(F) \tag{3.20}$$

for all x_i with $x_i > u_i$, $i \in C$ and $x_i > t$, $i \in D$.

On F, $D(t+) = D$, and using (2.11) the probability above becomes

$$\begin{aligned} \prod_D F(ds_i) \int_{s_i > x_i,\ i \in C \cup R} & P_s(U_i \in du_i,\ i \in C, \\ & R(t+) = R,\ D(t+) = D) \prod_{C \cup R} F(ds_i), \end{aligned} \tag{3.21}$$

where $s = (s_1, \dots, s_n)$. The integrand is the P_s-probability of an event F' in $\mathcal{F}_t$; hence by the note following (2.6), it is determined from the behaviour of Λ_s on F'. Using (C), it is then clear that the probability does not change if s is replaced by any $z = (z_1, \dots, z_n)$ such that $z_i = s_i$, $i \in D$, $z_i > u_i$, $i \in C$ and $z_i > t$, $i \in R$. Thus the integrand can be replaced by

$$\prod_C G^{-1}(u_i) \prod_R G^{-1}(t) \int_{\substack{z_i > u_i,\ i \in C \\ z_i > t,\ i \in R}} P_z(F') \prod_{C \cup R} F(dz_i)$$

not depending on s and it is an easy matter to reduce (3.21) to

$$\prod_C G(x_i | u_i) \prod_R G(x_i | t) \prod_D F(ds_i) \int_{\substack{z_i > u_i,\ i \in C \\ z_i > t,\ i \in R}} P_z(F') \prod_{C \cup R} F(dz_i),$$

which by (2.11) is precisely the right-hand side of (3.20). □

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

3.22. THEOREM. *Let Λ be the intensity for a FC-process P and suppose that for $i = 1, \dots, n$, $t > 0$,*

$$\Lambda^i(dt) = \mu(t) I_i(t)\, dt, \tag{3.23}$$

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

PROOF. Proposition 3.19 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 (3.23) we shall solve (3.7) and (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 (3.23) holds, solving (3.7) and (3.8) for the Λ_x suggests that

$$\Lambda_x^{(c,B)} = \Lambda^{(c,B)} \tag{3.24}$$

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

$$\Delta\Lambda_x^{(i,A)}(x_i) = \frac{d\Lambda^{(i,A)}}{d\Lambda^i}(x_i) I_i(x_i), \tag{3.25}$$

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

The Λ_x must satisfy (2.12)–(2.14). Defining $\Lambda_x^{(c,B)}$ by (3.24) gives a problem with (2.12), while of course (3.25) 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

$$P\big(\Delta\Lambda^{(c,B)}(X_i^*);\ X_i^* < \infty\big) = 0. \tag{3.26}$$

But it is even true that $\Delta\overline{\Lambda}(X_i^*) = 0$ P-a.s. on $(X_i^* < \infty)$: If finite, X_i^* is a jump time τ_n for N and the jump Y_n is of the form (i, A). For every n,

$$P\big(\Delta\overline{\Lambda}(\tau_n);\ Y_n = (i,A),\ \tau_n < \infty\big) = P\big(\Delta\Lambda^{(i,A)}(\tau_n);\ \tau_n < \infty\big)$$

as is seen using, e.g., the construction of multivariate counting processes presented in [5], Chapter 2. Since by assumption, Λ^i and, a fortiori, $\Lambda^{(i,A)}$ is absolutely continuous, (3.26) follows.

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

$$\Lambda_x^{(c,B)} = \tilde{\Lambda}^{(c,B)}. \tag{3.27}$$

To pick out the correct value of the derivative in (3.25) 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 defining for any $x \in L$,

$$\begin{aligned} &\Delta\Lambda_x^{(i,A)}(x_i, w) \\ &\quad = \liminf_{k\to\infty} \frac{\Lambda^{(i,A)}(x_i, w) - \Lambda^{(i,A)}(x_i - (1/k), w)}{\Lambda^i(x_i, w) - \Lambda^i(x_i - (1/k), w)} I_i(x_i, w) \end{aligned} \tag{3.28}$$

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

$$\Delta\Lambda_x^{(i,\varnothing)}(x_i, w) = \Big(1 - \sum_{A\neq\varnothing} \Delta\Lambda_x^{(i,A)}(x_i, w)\Big) I_i(x_i, w), \tag{3.29}$$

it is clear that with w, i fixed, for $\Lambda^i(\cdot, w)$-almost all x_i, (3.25) holds for all A.

Adding the requirement that $\Lambda_x^{(i,A)}$ be concentrated at x_i, (3.27)–(3.29) provides an explicit definition of Λ_x on all of W. While obviously each $\Lambda_x^{(i,A)}$ is predictable on all of W, $\Lambda_x^{(c,B)}$ is predictable only when restricted to W_x. But on W_x, certainly Λ_x satisfies (i)–(ii) of Fact 2.8 and (2.12)–(2.14). Invoking Lemma 2.15, this ensures that Λ_x is the intensity for a uniquely determined 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, $x_i \le t$ 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 (3.28) and (3.29)], $\Lambda_x^{(i,A)}(t, w) = \Lambda_{x'}^{(i,A)}(t, w)$ 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 (3.27), and that the $\Lambda^{(i,A)}$ from (3.7) equals the $\Lambda^{(i,A)}$ for the P we started off with. Here we give the proof if $A \neq \varnothing$, the case $A = \varnothing$ being an easy consequence. But by the remarks above, for any $w \in W$, $x \in L$ and almost all t, $\Delta\Lambda_{x|i,t}^{(i,A)}(t, w)$ is a derivative with respect to Λ^i and, using (3.23), then

$$\int_0^s \mu(t)\, \Delta\Lambda_{X|i,t}^{(i,A)}(t) I_i(t)\, dt = \int_0^s \frac{d\Lambda^{(i,A)}}{d\Lambda^i}(t) \Lambda^i(dt) = \Lambda^{(i,A)}(s). \qquad \square$$

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 (2.9).

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 $(U_1^{(1)}, \dots, U_n^{(1)})$ of possible censoring times, $0 < U_i^{(1)} \le \infty$, stochastically independent of X and with an arbitrary distribution. Find the smallest X_i, X_{i_1} say, such that $X_{i_1} \le U_{i_1}^{(1)}$. 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, $X_{i_1} \le U_i^{(2)} \le \infty$, using an arbitrary distribution depending on i_1, X_{i_1}, C_1 and $(U_j)_{j \in C_1}$, 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 $X_{i_2} \le U_{i_2}^{(2)}$. 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 proceeds. The kth vector of possible censoring times $(U_i^{(k)})_{i \in R_{k-1}}$ are chosen from an arbitrary distribution depending on $i_1, \dots, i_{k-1}$, $X_{i_1}, \dots, X_{i_{k-1}}$, $C_1, \dots, C_{k-1}$ and $(U_j)_{j \in C_1 \cup \cdots \cup C_{k-1}}$, i.e., everything observed on $[0, X_{i_{k-1}})$ and i_{k-1}, $X_{i_{k-1}}$, but independent of $(X_i)_{i \in R_{k-1}}$.

3.30. EXAMPLE. Consider the scheme for progressive type 2 censorship, where at the time of the kth 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 kth observed failure is $|R_k| = n - k - (r_1 + \cdots + r_k)$.

Choosing X_i i.i.d. μ, the conditional intensities Λ_x are specified by $\Lambda_x^{(c, B)} = 0$ 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 $\{1, \dots, n\} \setminus \{i\}$ of cardinality $r_{\tilde{N}(x_i-)+1}$ and equal to 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 $U_i^{(1)} = \infty$, 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.

3.31. EXAMPLE. Suppose first that $n = 1$ and consider a FC-process satisfying the martingale condition (2.9). We only have the two intensities

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

for the failure, respectively, censoring of the individual 1. The failure intensity is prescribed by (2.9) and in order to be predictable, the censoring intensity must necessarily be of the form above with ν some hazard measure on $(0,\infty)$. The description of the simulation above, shows that the distribution of the FC-process may be obtained by choosing X_1 and U_1' independent, X_1 with hazard μ, U_1' with hazard ν and observing the failure time X_1 if $X_1 \leq U_1'$ and the censoring time U_1' if $U_1' < X_1$. Equivalently, the observed failure time, respectively, censoring time, is

$$X_1^* = \begin{cases} X_1 & \text{if } X_1 \leq U_1', \\ \infty & \text{if } X_1 > U_1', \end{cases} \qquad U_1 = \begin{cases} U_1' & \text{if } X_1 > U_1', \\ \infty & \text{if } X_1 \leq U_1'. \end{cases} \tag{3.32}$$

Now consider the case of independent pairs (X_i, U_i'), discussed, e.g., by Williams and Lagakos [12], Kalbfleisch and MacKay [7], with observed failures X_i^* and censorings U_i defined in analogy with (3.32) for each i. (One may allow the hazard for U_i' to depend on i.)

By the independence, (2.9) holds iff it holds separately for each i and, as noted in Section 2, is then equivalent to the constant sum condition from [12] as reformulated in [7]. Thus it emerges, that as long as only the distribution of observed failures and censorings is of interest, and Lagakos' constant sum condition holds, it may always be assumed that in addition to the independence of the (X_i, U_i'), X_i and U_i' are independent for each i.

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

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

$$\Lambda_x^{(i,\varnothing)}(dt) = \varepsilon_{x_i}(dt) 1_{(i \in R(t-))}, \qquad i = 1,2, \tag{3.34}$$

$$\Lambda_x^{(c,1)}(dt) = dt\, 1_{(1 \in R(t-),\, x_1 > t)}, \tag{3.35}$$

$$\Lambda_x^{(c,2)}(dt) = \varepsilon_{x_1 + a}(dt) 1_{(N^{(c,1)}(t-) = 1,\, 2 \in R(t-),\, x_2 > t)}. \tag{3.36}$$

Here (3.34) comes from (2.14), while (3.35) shows that individual 1 is censored at an exponential time, independent of everything else, provided this time is less than x_1. Finally (3.36) 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 (3.34)–(3.36) imply 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 (3.34)–(3.36) imply (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 (3.34)–(3.36) imply (G) is completed by observing that (G′) $\Rightarrow$ (G). This may be argued along the lines of the proof of Proposition 3.19. We omit the details.

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

$$\begin{aligned} \Lambda^{(i,\varnothing)}(dt) &= \mu(t) I_i(t)\,dt, \qquad i = 1,2, \\ \Lambda^{(c,1)}(dt) &= I_1(t)\,dt, \end{aligned} \tag{3.37}$$

$$\Lambda^{(c,2)}(dt) = \begin{cases} \mu(t-a) I_2(t) 1_{(U_1 < t-a)}\,dt \\ \quad \text{if } a > 0, \\ \mu(t-a) \dfrac{G(t-a)}{G(t-a) + G(U_1) - G(U_1 - a)} I_2(t) 1_{(U_1 < t)}\,dt, \\ \quad \text{if } a < 0. \end{cases} \tag{3.38}$$

By Theorem 3.22 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 (3.34), (3.35) and replacing (3.36) by letting $\Lambda_x^{(c,2)}$ be given by the right-hand side of (3.38); cf. (3.8).

In conclusion it should be mentioned that the results of this section, with some obvious changes, remain valid if only the X_i are independent, but not necessarily identically distributed.

4. Statistical models—problem 3. We shall discuss models for right censored survival data, where each member of the model is a FC-process, derived from i.i.d. failure times with an unknown hazard μ and satisfying the martingale condition (2.9).

Formally the model is a family of distributions for N and evidently, the model is only partially specified by allowing an arbitrary μ in the description above. Indeed, as shown in Kalbfleisch and Prentice ([8], Section 5.2), the likelihood for observed failures and censorings splits into a product of a factor where μ appears in a natural manner [(4.1) below] and a factor due to the censorings [see [8], (5.6)].

In [8] the censoring is called informative if this extra factor depends on μ, noninformative otherwise. With this in mind, the point to be made here is the following: By the results of Section 3, each distribution for N derives in a unique manner from a joint distribution of X and N which satisfies condition (C) and makes the X_i i.i.d. In particular the model for N can be specified by considering an arbitrary hazard μ and considering for each μ and failure time vector x, some family of conditional intensities Λ_x, all satisfying (C). Expressing the likelihood in terms of μ and the Λ_x enables us to characterize the structure of those models where only the observed failures contribute information about μ. We shall refer to the models built in this manner as FC(C)-models.

Suppose N is observed on $[0, t]$ and following [8], consider the condition that the censorings on $[0, t]$ are noninformative, i.e., that the likelihood $L(t)$ has the form

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

apart from factors not depending on μ (see also, e.g., Lagakos [10]).

The structure (4.1) is also the natural one to assume for the Nelson–Aalen and Kaplan–Meier estimators to exploit all information about μ.

Our characterization of FC(C)-models with noninformative censoring involves in particular the concept of noninnovation as introduced by Arjas and Haara [see [3], condition (A)].

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

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

$$\exp\left(-\int_0^t \mu(s)|R(s-)|\,ds\right) \prod_{i:\ X_i^* \le t} \mu(X_i^*). \tag{4.3}$$

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

(iii) *All censorings are noninnovative in the sense of Arjas and Haara.*

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

PROOF OF THEOREM 4.2 The likelihood function for observation of N on $[0, t]$ is [cf. (2.6)]

$$\prod_{\substack{0 < s \le t \\ s \ne \tau_k}} \left(1 - \overline{\Lambda}(ds)\right) \prod_{k=1}^{\overline{N}(t)} \Lambda^{Y_k}(d\tau_k), \tag{4.4}$$

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

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

with the first term absolutely continuous. Thus the product integral in (4.4) becomes

$$\prod_{\substack{0<s\le t\\ s\neq\tau_k}} (1-\overline{\Lambda}(ds)) = \exp\left(-\int_0^t \mu(s)|R(s-)|\,ds\right) \times \prod_{\substack{0<s\le t\\ s\neq\tau_k}}\left(1-\sum_B \Lambda_x^{(c,B)}(ds)\right). \tag{4.5}$$

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

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

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

$$\mu(X_i^*)\,\Delta\Lambda_X^{(i,A)}{}_{|i,X_i^*}(X_i^*)\,dX_i^* \tag{4.7}$$

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 (ii) holds. Thus (ii) and (i) are equivalent.

In [3] the type set (mark space) E is split into two parts E' and E'' of innovative, respectively, noninnovative marks. In our case

$$E' = \{(i,A): 1\le i\le n,\ A\subseteq\{1,\dots,n\}\setminus\{i\}\},$$
$$E'' = \{(c,B): \varnothing\neq B\subseteq\{1,\dots,n\}\}.$$

Recognizing that because the intensities $\Lambda^{(i,A)}$ are absolutely continuous, the ρ_t defined in ([3], page 198) vanishes, it is easy to see that condition (A) of [3] for the censorings to be noninnovative 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),

$$\frac{d\Lambda^{(i,A)}}{d\Lambda^i}(s) = \Delta\Lambda_{X|i,s}^{(i,A)}(s)I_i(s),$$
$$\Lambda^{(c,B)} = \Lambda_X^{(c,B)}$$

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

For FC(C)-models not satisfying the conditions of the theorem, even though the Nelson–Aalen estimator is still a martingale estimator, one can do better by using also the information about μ provided by the observed censorings. The classical example is the Koziol–Green model obtained from i.i.d. censoring times also independent of X and with a hazard proportional to μ. Another example is provided by considering (3.37) and (3.38) with arbitrary μ and a given or unknown.

Acknowledgments. It is a pleasure to thank Per Kragh Andersen, Ørnulf Borgan, Richard Gill and Niels Keiding for their comments and good advice. Also, the comments and suggestions made by the referees are greatly appreciated.

REFERENCES

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

[2] ARJAS, E. (1982). On the consistency of Kaplan–Meier and Nelson plots in the case of multiple causes of failure. Report 2/82, Dept. Mathematics, Univ. Oulu.

[3] ARJAS, E. and HAARA, P. (1984). A marked point process approach to censored failure data with complicated covariates. *Scand. J. Statist.* **11** 193–209.

[4] GILL, R. D. (1980). *Censoring and Stochastic Integrals. Math. Centre Tracts* **124**. Mathematisch Centrum, Amsterdam.

[5] JACOBSEN, M. (1982). *Statistical Analysis of Counting Processes. Lecture Notes in Statist.* **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.

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN
5, UNIVERSITETSPARKEN
DK-2100 COPENHAGEN Ø
DENMARK