



Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

P.K. Andersen, Ø. Borgan, R.D. Gill, N. Keiding

Censoring, truncation and filtering in statistical
models based on counting processes

Department of Mathematical Statistics Report MS-R8803 March

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

Censoring, Truncation and Filtering in Statistical Models Based on Counting Processes

Per Kragh Andersen
*Statistical Research Unit,
University of Copenhagen,
Blegdamsvej 3, 2200 Copenhagen N,
Denmark*

Ørnulf Borgan
*Institute of Mathematics,
University of Oslo,
P.O. Box 1053 - Blindern, 0316 Oslo 3,
Norway*

Richard D. Gill
*Centre for Mathematics and Computer Science,
Kruislaan 413,
1098 SJ Amsterdam,
The Netherlands*

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

A survey is given of the formulation of statistical models for life history data based on counting processes. Examples include survival data with (possibly time-dependent) covariates and continuous time Markov processes. First, complete life history data are discussed and next, explicit modelling of mechanisms causing incomplete information like (right) censoring, filtering, and left truncation is superimposed onto the model for the complete data. The concepts of independent censoring and noninformative censoring are defined, and inference from such models based on (partial) likelihoods is discussed.

Note: To appear in *Contemporary Mathematics* (AMS) volume on *Inference from Stochastic Processes*.

Keywords & Phrases: statistical inference for stochastic processes, partial likelihood, informative censoring, independent censoring.

1980 Mathematics Subject Classification: 62M99, 62P10.

1. INTRODUCTION

The first systematic use of the terms censoring and truncation was due to HALD (1949; 1952, p. 144), who credited J.E. Kerrich for suggesting the term censoring for use in statistics. *Truncation* is 'sampling an incomplete population' - we would nowadays perhaps prefer 'sampling from a conditional distribution'. *Censoring* occurs 'when we are able to sample the complete population but the individual values of observations below (or above) a given value are not specified'. Obviously Hald's definition of censoring immediately extends to more general types of incomplete observation, such as grouped data.

Report MS-R8803
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The purpose of this paper is to survey the mathematical theory of censoring - incomplete sampling from a complete distribution - and of truncation - complete sampling from an incomplete distribution - not for statistical models in general, but for longitudinal data that may be described by *counting processes*. The simplest example of such a model is the classical life-testing situation, where the life times of n individuals are modelled as independent identically distributed nonnegative random variables and where the nature of the observational situation implies that some individuals are *lost to follow-up* or have *delayed entry*.

A main point of this paper is to demonstrate that loss to follow-up and delayed entry are both special cases of a particular kind of incomplete observation that is very natural in the stochastic process context: what we shall here term the *Aalen filter* (AALEN, 1978). For this reason *right censoring* and *left truncation* may be conveniently handled here, whereas other kinds of censoring (such as left censoring) or truncation (such as right truncation) may not. Indeed, it might be more methodologically consistent to speak about *left and right filtering* rather than about delayed entry and right censoring in the stochastic process situation.

Nonparametric estimation of a distribution function under left truncation and right censoring was surveyed by KAPLAN & MEIER (1958) (although several later authors have overlooked the part about left truncation). The subsequently published literature has been much richer for right censoring than for left truncation, important contributors being KALBFLEISCH & PRENTICE (1980, Chapter 5), ARJAS & HAARA (1984) and JACOBSEN (1986). There are two approaches: one is the *random censorship model* where censoring is treated (in the model) equivalently with life-length. This leads to a neat and easily comprehensible mathematical theory within the general area of competing risk models. However, the assumptions needed to complete the ambitious task of having an explicit stochastic model not only for the life-length but also for the censoring are often too restrictive, which has led to an interest in the *second approach*, of *partial models*, identifying minimal conditions for the censoring mechanism that allow correct inference for the distribution of the life-lengths, this being the distribution 'of interest'.

Two concepts are important in the partial modelling of right censoring, both given early formulations by KALBFLEISCH & PRENTICE (1980, Chapter 5). There is the *statistical* concept of *noninformative censoring*: the censoring mechanism should be ancillary in some sense, not contributing information about the unknown parameter. And there is the *probabilistic* concept of *independent censoring*, very heuristically stating that the extra randomness, and the reduced information, caused by the censoring mechanism, should be 'orthogonal' to the (conceptual) situation without censoring.

We shall present here an account of the important counting process framework of ARJAS & HAARA (1984) for Kalbfleisch & Prentice's concept of noninformative censoring, based upon a particular class of marked point processes. And we shall continue the analysis of independent censoring by these authors as well as by JACOBSEN (1986).

For left truncated survival data, one may similarly choose the easy way of embedding everything in one model, treating the truncation time symmetrically to the life time. A comprehensive exposition with a complete asymptotic theory was given by WOODROOFE (1985), who was motivated by applications to astronomy and did not connect to life testing. WANG, JEWELL & TSAI (1986) put Woodrooffe's results into the survival analysis framework, while KEIDING & GILL (1987) demonstrated how the (exact as well as asymptotic) results of this *random truncation model* may be obtained as corollaries of the existing statistical theory for counting processes.

An alternative theory of partial modelling of left truncation is not available in the existing literature. We provide in this paper some introductory remarks in this direction for the counting process framework but we regard it unlikely that this theory will be as rich as that of right censoring.

Other kinds of censoring and truncation are less easily tractable by genuine stochastic process methods except for some tricks in very special situations, and we provide some explanation for this towards the end of the paper.

The structure of the paper is that Section 2 recalls the multiplicative intensity model for counting processes and basic examples of completely observed processes. Section 3 introduces right censoring

and various models for right censoring mechanisms, and the concepts of independent censoring and non-informative censoring are discussed. Section 4 contains discussion of *left truncation*. In Section 5 the concept of *filtering* is introduced and more *general patterns of censoring* are discussed, also in connection with left truncation. Section 6 outlines the situation when only a *discrete skeleton* is observed. Often the statistician is not interested in specifying a model for the whole system under consideration including the distribution of covariates and the censoring mechanism (or he or she is simply unable to do this), and therefore in Section 7 possibilities of analysing *partially specified* models are discussed.

2. INTENSITY MODELS FOR COUNTING PROCESSES AND EXAMPLES OF UNCENSORED MODELS

Several expositions of the basics of statistical models for counting processes already exist (AALEN, 1978, ANDERSEN et al., 1982, JACOBSEN, 1982, ANDERSEN & BORGAN, 1985) so we may be brief in defining our framework.

Consider a measurable space (Ω, \mathcal{F}) with a right-continuous *filtration* $(\mathcal{F}_t)_{t \in \mathcal{T}}$, where $\mathcal{T} = [0, \tau]$ or $[0, \infty)$ for a given time instant τ , $0 < \tau \leq \infty$. We write $\mathcal{T} = [0, \tau]$ and define $\mathcal{F}_\tau = \bigvee_{t \in \mathcal{T}} \mathcal{F}_t$ if $\tau \notin \mathcal{T}$. It is assumed that, for each member $P_{\theta\phi}$ of a family

$$\mathcal{P} = \{P_{\theta\phi} : (\theta, \phi) \in \Theta \times \Phi\}$$

of mutually equivalent probability measures on \mathcal{F}_τ , \mathcal{F}_0 (and hence each \mathcal{F}_t) is *complete* in the sense of containing all subsets of null sets of \mathcal{F}_τ , although we shall not explicitly include null sets in \mathcal{F}_0 in the examples.

On $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathcal{P})$ we consider a *multivariate counting process* $\mathbf{N} = (\mathbf{N}(t), t \in \mathcal{T}) = (N_1(t), \dots, N_k(t), t \in \mathcal{T})$ adapted to (\mathcal{F}_t) . That is, each component $N_h(t)$, $t \in \mathcal{T}$ is a stochastic process with sample functions right continuous non-decreasing step functions, 0 at time 0, and with jumps of unit size. Moreover, it is assumed that with probability one, no two components jump simultaneously and that $N_h(t) < \infty$ for $t \in \mathcal{T}$.

A counting process $\mathbf{N}(t)$ has *compensator* $\Lambda(t)$ such that $\mathbf{N}(t) - \Lambda(t)$ is a local martingale, and $\Lambda(t)$ is predictable and has paths of locally bounded variation. Under regularity conditions (see e.g. AALEN, 1978, Section 3.2), $\Lambda(t)$ will be absolutely continuous

$$\Lambda_h(t) = \int_0^t \lambda_h(s) ds, \quad h = 1, \dots, k$$

where $\lambda_h(t)$ is predictable and has the property

$$\lambda_h(t+) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} P\{N_h(t + \Delta t) - N_h(t) = 1 | \mathcal{F}_t\};$$

$\lambda_h(t)$ is denoted the *intensity process* for $N_h(t)$.

We shall often construct compensators for multivariate counting processes from compensators of its components, combined with an independence assumption, cf. JACOBSEN (1982, pp. 72-73) for the case of *canonical* counting processes. This *product construction* is given in the Appendix.

The family \mathcal{P} is doubly indexed by (θ, ϕ) , where θ is the parameter of interest parametrizing the transition intensities for the events under study and ϕ a nuisance parameter typically parametrizing the distribution of censoring and covariates. In some cases there are no nuisance parameters ϕ . It is assumed that the $P_{\theta\phi}$ -compensator for \mathbf{N} with respect to (\mathcal{F}_t) is $\Lambda^\theta(t) = \Lambda(t, \theta)$ and hence that it does not depend on ϕ . The sets Θ and Φ may be subsets of either finite dimensional Euclidean spaces, corresponding to parametric models, or more general function spaces, corresponding to nonparametric models.

The statistical model \mathcal{P} corresponding to a given filtration (\mathcal{F}_t) is called a *multiplicative intensity model* (AALEN, 1975, 1978) if its intensity process admits a decomposition

$$\lambda_h(t, \theta) = \alpha_h(t, \theta) Y_h(t), \quad h = 1, \dots, k, \quad t \in \mathcal{T}$$

where $Y_h(t)$ is predictable and does not depend on θ , while α is deterministic.

In a moment, we shall see how the counting process N and the filtration (\mathcal{F}_t) may be constructed in several examples when there is no censoring. We will either have $\mathcal{F}_t = \mathcal{N}_t = \sigma\{N(u), 0 \leq u \leq t\}$ (the *self-exciting filtration*) or $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{N}_t$ with \mathcal{F}_0 generated by a random variable X_0 which can be thought of as being realised at time 0. Thus the observations available to the researcher at time t consist of $(N(u), 0 \leq u \leq t)$ and (when relevant) X_0 . We shall sometimes use the convention that the observations at time t are the σ -algebra \mathcal{F}_t hereby meaning that at time t it can be determined whether or not any event $A \in \mathcal{F}_t$ has occurred. Let $P_{\theta, \phi; t}$ denote the restriction of $P_{\theta, \phi}$ to \mathcal{F}_t , for any $t \in \overline{\mathcal{T}}$. Whenever (\mathcal{F}_t) has the structure $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{N}_t$, a key result by JACOD (1975) expresses the likelihood function $dP_{\theta, \phi; t}$ or Radon-Nikodym derivative $dP_{\theta, \phi; t} / dP_{\theta_0, \phi_0; t}$ in terms of the intensity of N and the likelihood function (or R.-N. derivative) for the data X_0 at time 0. Using product-integral notation (see JOHANSEN, 1987, GILL & JOHANSEN, 1987) the result can be summarized as

$$dP_{\theta, \phi; t} = dP_{\theta, \phi; 0} \cdot \prod_{s \leq t} \{(1 - \lambda(s; \theta, \phi) ds)^{1 - dN(s)} \prod_{h=1}^k (\lambda_h(s; \theta, \phi) ds)^{dN_h(s)}\} \quad (2.1)$$

where $N = \sum_h N_h$ and $\lambda = \sum_h \lambda_h$ and (for later use) λ may depend on ϕ as well as on θ . Formula (2.1) yields the Radon-Nikodym derivative by forming the quotient of left and right hand sides with the same expressions for $(\theta, \phi) = (\theta_0, \phi_0)$, say. The same result holds for non (absolutely) continuous $\Lambda^{\theta, \phi}$, simply replacing λds and $\lambda_h ds$ by $d\Lambda$ and $d\Lambda_h$ respectively. Note that by the interpretation of the intensity as conditional probability (density), the term in braces in (2.1) can be interpreted as $P_{\theta, \phi}(dN(s) | \mathcal{F}_s^-)$.

EXAMPLE 2.1. *A single non-negative random variable.* Let X be a non-negative random variable on a space (Ω, \mathcal{F}) with absolutely continuous distribution function F , survival function $S = 1 - F$, density f , and hazard rate function $\alpha = f/S$. We assume that the distribution of X depends on a (finite or infinite dimensional) parameter θ and write $\alpha^\theta(t)$ or $\alpha(t, \theta)$ for the hazard function, $F(t, \theta)$ for the distribution function etc. Furthermore we assume that τ , the upper endpoint of the support of F^θ does not depend on θ and we let $\mathcal{T} = [0, \tau)$. Then

$$\int_0^t \alpha(u, \theta) du = -\log(1 - F(t, \theta)) < \infty$$

for all $t \in \mathcal{T}$, though $\int_0^\tau \alpha(u, \theta) du = \infty$.

Define the stochastic process

$$N(t) = I(X \leq t). \quad (2.2)$$

Then N is a univariate counting process counting +1 only at X . We let

$$Y(t) = I(X \geq t) = 1 - N(t-). \quad (2.3)$$

It is then a direct consequence of an important representation of JACOD (1975), expressing the intensity of a counting process N (with respect to a filtration of the type $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{N}_t$) in terms of the conditional distributions of the time and type of each jump given all preceding ones and given \mathcal{F}_0 , that N has compensator

$$\Lambda(t, \theta) = \int_0^t \alpha(u, \theta) Y(u) du$$

with respect to the self-exciting filtration $(\mathcal{F}_t) = (\mathcal{N}_t)$ and the probability P_θ corresponding to the distribution F^θ of X . It further follows from Jacod's result (2.1) that the likelihood

$$L(\theta) = \prod_{t \in \mathcal{T}} \{(1 - \alpha(t, \theta) Y(t) dt)^{1 - dN(t)} (\alpha(t, \theta) Y(t))^{dN(t)}\}$$

$$\begin{aligned}
&= \prod_{t < X} (1 - \alpha(t, \theta) dt) \alpha(X, \theta) \\
&= \exp\left(-\int_0^X \alpha(t, \theta) dt\right) \alpha(X, \theta) = S(X, \theta) \alpha(X, \theta) = f(X, \theta),
\end{aligned}$$

the density function $f(\cdot, \theta)$ evaluated at X . \square

EXAMPLE 2.2. Uncensored survival data. Let X_1, \dots, X_n be independent non-negative random variables, X_i having hazard function $\alpha_i(t, \theta)$. Assume that their distributions have common support not depending on θ , and let τ be the upper endpoint for this support. As in Example 2.1 we let $\mathcal{T} = [0, \tau)$, and define for each $i = 1, \dots, n$ stochastic processes $N_i(t)$ and $Y_i(t)$ by (2.2) and (2.3).

Identifying i with an 'individual' and X_i with the 'survival time' or 'failure time' of that individual, then N_i counts 1 only at the time X_i when individual i dies and $Y_i(t) = 1$ if individual i is still 'alive' or 'at risk' just before time t .

Obviously $\mathbf{N} = (N_1, \dots, N_n)$ is a multivariate counting process with respect to the self-exciting filtration (\mathcal{N}_t) generated by \mathbf{N} . The compensator with respect to (\mathcal{N}_t) and P_θ , the joint distribution of the X_i 's, may be derived directly from Jacod's representation as in the previous example, or alternatively from the *product construction* given in the Appendix.

It follows that N_i has intensity process $(\alpha_i(t, \theta) Y_i(t))$ and compensator

$$\Lambda_i(t, \theta) = \int_0^t \alpha_i(u, \theta) Y_i(u) du.$$

The likelihood for \mathbf{N} is

$$L(\theta) = \prod_{t \in \mathcal{T}} \{(1 - \sum \alpha_i(t, \theta) Y_i(t) dt)^{1 - dN.(t)} \prod_{i=1}^n (\alpha_i(t, \theta) Y_i(t))^{dN_i(t)}\},$$

where

$$N.(t) = \sum_{i=1}^n N_i(t).$$

Similarly to Example 2.1 the likelihood

$$\begin{aligned}
L(\theta) &= \exp\left(-\sum_{i=1}^n \int_0^{X_i} \alpha_i(u, \theta) du\right) \prod_{i=1}^n \alpha_i(X_i, \theta) \\
&= \prod_{i=1}^n S_i(X_i, \theta) \alpha_i(X_i, \theta)
\end{aligned}$$

reduces to the product of the density functions $f_i(\cdot, \theta)$ evaluated at X_i .

When X_1, \dots, X_n are independent and identically distributed (*i.i.d.*) with hazard function $\alpha(\cdot, \theta)$ the likelihood reduces to

$$L(\theta) = \prod_{t \in \mathcal{T}} \{(1 - \alpha(t, \theta) Y.(t) dt)^{1 - dN.(t)} \prod_{i=1}^n (\alpha(t, \theta) Y_i(t))^{dN_i(t)}\}$$

where

$$Y.(t) = \sum_{i=1}^n Y_i(t) = n - N.(t-).$$

Thus in this case $L(\theta)$ is equal to

$$\prod_{t \in \mathcal{T}} (1 - \alpha(t, \theta) [n - N.(t-)] dt)^{1 - dN.(t)} \alpha(t, \theta)^{dN.(t)}$$

showing that the *aggregated* process N . is *sufficient* for θ corresponding to the fact that the ordered observations $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ are sufficient. We may also say that the self-exciting filtration after aggregation is sufficient for θ . Note that in the i.i.d. case N . is a univariate counting process with intensity process

$$\lambda(t, \theta) = \alpha(t, \theta)Y(t)$$

with respect to (\mathcal{G}_t) and P_θ . Thus the counting process N . obtained by *aggregation* of the *individual* counting processes N_i each following a multiplicative intensity model is a univariate counting process with the same intensity structure. The intensity process for the aggregated counting process N . is a product of an *individual intensity* $\alpha^\theta(t)$ and a process $Y(t)$ which can be interpreted as the *number of individuals at risk* for failing just before time t . \square

EXAMPLE 2.3. A model for relative mortality. Let, as in Example 2.2, X_1, \dots, X_n be independent non-negative random variables and assume that the distribution of X_i is absolutely continuous with hazard rate function $\mu_i(t)\alpha_0(t, \theta)$. Here $\mu_i(\cdot)$ is assumed to be a *known* hazard rate function, e.g. a population based quantity known from vital statistics, and $\alpha_0(\cdot, \theta)$ is an unknown (time- or age-dependent) *relative mortality* common to all i . Now $\mathbf{N} = (N_1, \dots, N_n)$ is a multivariate counting process, N_i having intensity process given by

$$\lambda_i(t, \theta) = \alpha_0(t, \theta)\mu_i(t)I(X_i \geq t), \quad i = 1, \dots, n$$

with respect to (\mathcal{G}_t) and P_θ . In this case the likelihood is proportional to

$$\prod_{t \in \mathfrak{G}} (1 - \alpha_0(t, \theta)Y^\mu(t)dt)^{1 - dN(t)} \alpha_0(t, \theta)^{dN(t)},$$

where

$$Y^\mu(t) = \sum_{i=1}^n \mu_i(t)I(X_i \geq t).$$

Thus the pair (N, Y^μ) is sufficient for θ , whereas the self-exciting filtration after aggregation is not. By aggregation, a univariate counting process $N = N_1 + \dots + N_n$ is obtained with intensity process $\lambda(t, \theta) = \alpha_0(t, \theta)Y^\mu(t)$. Thus again the intensity process for the aggregated counting process N . has a multiplicative form but in this case $Y^\mu(t)$ is no longer simply the number at risk for failing at t . \square

EXAMPLE 2.4. A finite state Markov process. Let $(X(t), t \in \mathfrak{G})$, $\mathfrak{G} = [0, \tau]$, be a Markov process with finite state space S and right continuous sample paths and suppose that the initial distribution, i.e. the distribution of $X(0) = J_0$, say, depends on parameters ϕ (and possibly on θ too). We let T_ν be the ν 'th jump time of X and J_ν the state reached at T_ν . Then the Markov process X is equivalent to J_0 and the marked point process $(\mathbf{T}, \mathbf{J}) = \{(T_\nu, J_\nu); \nu = 1, 2, \dots\}$ in the sense that observation of $X(u), 0 \leq u \leq t$ gives the same data as observing J_0 and (\mathbf{T}, \mathbf{J}) on $[0, t]$.

We shall assume the existence of integrable *transition intensities* $\alpha_{hj}^\theta(t) = \alpha_{hj}(t, \theta)$ from state h to state j , $h \neq j$ (some $\alpha_{hj}^\theta(\cdot)$ may be zero for all values of θ). According to JACOBSEN (1972) $\alpha_{hj}^\theta(t)$ can then be decomposed into

$$\alpha_{hj}^\theta(t) = \mu_h^\theta(t)\pi_{hj}^\theta(t),$$

where μ_h^θ equals $\sum_{j \in S} \alpha_{hj}^\theta$ and satisfies

$$P_{\theta\phi}(T_{r+1} > t | J_0, (T_\nu, J_\nu), \nu = 1, \dots, r, J_r = h) = \exp\left\{-\int_T^t \mu_h^\theta(s) ds\right\} \quad (2.4a)$$

and π_{hj}^θ equals $\alpha_{hj}^\theta/\mu_h^\theta$ and satisfies

$$\pi_{hj}^\theta(t) = P_{\theta\phi}(J_{r+1} = j | J_0, (T_\nu, J_\nu), \nu = 1, \dots, r, J_r = h, T_{r+1} = t). \quad (2.4b)$$

Thus μ_h^θ is the force of transition out of the state h and when $\mu_h^\theta(t) = 0$ for all t (and θ) we say that the state h is *absorbing*. (Transition intensities into an absorbing state need only be *locally* integrable on $[0, \tau)$.)

Let $N_{hj}(t)$ be the number of direct transitions for X from h to j , $h \neq j$ in $[0, t]$. Then also $\mathbf{N} = (N_{hj}(\cdot), h \neq j)$ and J_0 are equivalent to X in the above mentioned sense and we define (\mathcal{O}_t) to be the self-exciting filtration for \mathbf{N} and let $\mathcal{F}_t = \mathcal{O}_t \vee \mathcal{F}_0$ with \mathcal{F}_0 generated by J_0 . The $P_{\theta\phi}$ -intensity process for the multivariate counting process \mathbf{N} with respect to (\mathcal{F}_t) can now again be found from the representation of JACOD (1975) and (2.4) since

$$P_{\theta\phi}(J_{r+1} = j | J_0, (T_\nu, J_\nu), \nu = 1, \dots, r; T_{r+1}) = \pi_{J,j}(T_{r+1}).$$

This shows that N_{hj} has $P_{\theta\phi}$ -intensity process

$$\begin{aligned} \lambda_{hj}^{\theta\phi}(t) &= \lambda_{hj}^\theta(t) = \mu_h^\theta(t) \pi_{hj}^\theta(t) Y_h(t) \\ &= \alpha_{hj}^\theta(t) Y_h(t), \end{aligned}$$

where $Y_h(t) = I(X(t-) = h)$ is the indicator for X being in the state h just before time t (see also JACOBSEN, 1982, p.120, and GILL & JOHANSEN, 1987). So the intensity only depends on θ and again we have a multiplicative intensity structure.

Next, assume that given $J_{i0}, i = 1, \dots, n$, independent copies $X_1(\cdot), \dots, X_n(\cdot)$ of $X(\cdot)$ are constructed with $X_i(0) = J_{i0}$; let $\mathbf{J}_0 = (J_{10}, \dots, J_{n0})$ and define a multivariate counting process $\mathbf{N} = (N_{hji}, i = 1, \dots, n; h \neq j)$ from $(X_1(\cdot), \dots, X_n(\cdot))$ as above. Then, by the conditional independence of the $X_i(\cdot)$'s and by the product construction given in the Appendix it is seen that N_{hji} has $P_{\theta\phi}$ -intensity process

$$\lambda_{hji}^{\theta\phi}(t) = \lambda_{hji}^\theta(t) = \alpha_{hji}^\theta(t) Y_{hi}(t),$$

where $Y_{hi}(t) = I(X_i(t-) = h)$ is the indicator for $X_i(\cdot)$ being in state h just before time t .

So, the multivariate counting process $\mathbf{N} = (N_{hji}, i = 1, \dots, n; h, j \in S, h \neq j)$ has a multiplicative intensity with respect to (\mathcal{F}_t) which only depends on θ . By (2.1) the likelihood takes the form

$$L(\theta, \phi) = L_0(\theta, \phi) L_\tau(\theta)$$

where

$$L_0(\theta, \phi) = P_{\theta\phi}(\mathbf{J}_0)$$

and

$$L_\tau(\theta) = \prod_{t \in \mathcal{G}} \left\{ \left(1 - \sum_i \sum_{h \neq j} \alpha_{hji}^\theta(t) Y_{hi}(t) dt \right)^{1 - \sum_i \sum_{h \neq j} dN_{hji}(t)} \prod_{i, h \neq j} (\alpha_{hji}^\theta(t) Y_{hi}(t))^{dN_{hji}(t)} \right\}.$$

Then $L_\tau(\theta)$ equals

$$\prod_{t \in \mathcal{G}} \left\{ \left(1 - \sum_{h \neq j} \alpha_{hj}^\theta(t) Y_{h\cdot}(t) dt \right)^{1 - \sum_{h \neq j} dN_{hj}(t)} \prod_{h \neq j} \alpha_{hj}^\theta(t)^{dN_{hj}(t)} \right\}$$

where

$$N_{hj\cdot}(t) = \sum_{i=1}^n N_{hji}(t)$$

and

$$Y_{h\cdot}(t) = \sum_{i=1}^n Y_{hi}(t),$$

the latter being a function of \mathbf{J}_0 and $(N_{hji}, h \neq j)$. Thus if the distribution of \mathbf{J}_0 only depends on ϕ , then for each fixed $\phi \in \Phi$, $L_\tau(\theta)$ is the full likelihood for θ , otherwise $L_\tau(\theta)$ is only a *partial* likelihood (COX, 1975; KALBFLEISCH & PRENTICE, 1980, Ch.5). At any rate, $L_\tau(\theta)$ is the full *conditional*

likelihood given the initial states \mathbf{J}_0 and $(\mathbf{J}_0, N_{hj}; h \neq j)$ is *sufficient* for θ , the second part being a multivariate counting process with intensity process

$$\lambda_{hj}^{\theta}(t) = \alpha_{hj}^{\theta}(t) Y_h(t).$$

Again aggregation leads to a counting process following a multiplicative intensity model where the first factor is an intensity on the individual level and the second is a process indicating the number at risk just before time t for experiencing events of given types. \square

EXAMPLE 2.5. Competing risks. As a very special case of Example 2.4 one may consider two states 0 ('alive') and 1 ('dead') and assume $\alpha_{10}(t, \theta) \equiv 0$ (that is, state 1 is absorbing) and the initial distribution degenerate at 0, yielding exactly the independent identically distributed uncensored survival times of Example 2.2, with hazard function equal to the transition intensity $\alpha_{01}(t, \theta)$.

A more general special case of the Markov process example is the *competing risks* model, obtained by considering one transient state 0 ('alive') and absorbing states $h = 1, \dots, k$ (so that $\alpha_{hj}(t, \theta) \equiv 0$ for $h = 1, \dots, k$ and all j). State h corresponds to 'dead by cause h '. The initial distribution is degenerate at 0 and the transition intensities $\alpha_{0h}(t, \theta)$, $h = 1, \dots, k$, are termed 'cause specific hazard functions'.

It is easily seen that the competing risks model is equivalent to considering independent random variables X_{i1}, \dots, X_{ik} , $i = 1, \dots, n$, with hazard functions $\alpha_{01}(t, \theta), \dots, \alpha_{0k}(t, \theta)$ and the multivariate counting process

$$\mathbf{N}(t) = (N_1(t), \dots, N_k(t))$$

with

$$N_h(t) = \sum_{i=1}^n I(\min_l X_{il} = X_{ih} \leq t).$$

In reliability theory, $\min_l X_{il}$ is interpreted as the life time for a series system of k independent components with life times X_{i1}, \dots, X_{ik} . It has been debated extensively how useful this reliability interpretation is in biomedical contexts. In particular, even if the competing risks model may be generated from a set of independent 'latent' (or 'underlying') life times, these are often hypothetical. References to a discussion of the interpretability and testability of the latent life time model include Cox (1959), TSIATIS (1975), KALBFLEISCH & PRENTICE (1980, Chapter 7), COX & OAKES (1984, Chapter 9). \square

In Examples 2.1 — 2.4 the *individual* counting processes N_{hi} satisfied the *multiplicative intensity model*

$$\lambda_{hi}^{\theta}(t) = \alpha_{hi}^{\theta}(t) Y_{hi}(t), \quad h = 1, \dots, k, \quad i = 1, \dots, n.$$

Here $\alpha_{hi}^{\theta}(t)$ was an *individual* force of transition, relative hazard, or hazard of *type* h and $Y_{hi}(t)$ a *predictable* process which was *observable* in the sense that it did not depend on the parameter θ . The process $Y_{hi}(t)$ contained information on whether or not individual i was at risk for experiencing an event of type h at time t . For instance (with a slight abuse of notation) h could correspond to a transition from one state to another in a Markov process.

When $\alpha_{hi}^{\theta} = \alpha_h^{\theta}$ for all i , the *aggregated counting process* N_h satisfied the multiplicative intensity model

$$\lambda_h^{\theta}(t) = \alpha_h^{\theta}(t) Y_h(t), \quad h = 1, \dots, k$$

with the predictable and observable process $Y_h(t)$ giving the size of the *risk set* for that type of event just before time t (in Examples 2.2 and 2.4 but not in Example 2.3, $Y_h(t)$ was simply the *number* at risk for a type h transition just before time t).

The next example presents a model for completely observed life history data for which aggregation does not lead to Aalen's multiplicative intensity model.

EXAMPLE 2.6. Relative risk regression models with time-independent covariates. Let (X_i, Z_i) be random

variables with X_i non-negative and each Z_i p -dimensional. We shall assume that X_1, \dots, X_n are conditionally independent given $\mathbf{Z} = (Z_1, \dots, Z_n)$, that the marginal distribution of \mathbf{Z} depends on parameters ϕ (and possibly θ too) and that the conditional distribution of X_i given $\mathbf{Z} = \mathbf{z} = (z_1, \dots, z_n)$ has hazard function $\alpha_i(t, \theta)$ as in Example 2.2. We shall consider models of the form

$$\alpha_i(t, \theta) = \alpha_0(t, \gamma)r(\beta^T z_i) \quad (2.5)$$

with $\theta = (\gamma, \beta)$, $\beta \in \mathbb{R}^p$ and the *relative risk* function $r(\cdot)$ being non-negative. The main example is the semi-parametric *Cox regression model* (COX, 1972) where r is the exponential function and γ is infinite dimensional, but parametric models with $\gamma \in \mathbb{R}^q$ may also be considered. In any case Z_1, \dots, Z_n are *covariates* upon which we want to condition and ϕ is a nuisance parameter. In some cases it is reasonable to assume $(X_1, Z_1), \dots, (X_n, Z_n)$ to be i.i.d. but when some of the X_i 's correspond to life times of individuals from the same family or community the above assumption of conditional independence of the X_i 's given \mathbf{Z} is more realistic.

From X_i , $i = 1, \dots, n$ we define stochastic processes $N_i(\cdot)$ and $Y_i(\cdot)$ from (2.2) and (2.3) and we let (\mathcal{N}_t) be the filtration generated by the multivariate counting process $\mathbf{N} = (N_1, \dots, N_n)$. Furthermore we let \mathcal{F}_0 be generated by \mathbf{Z} , define $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{N}_t$ and let $P_{\theta\phi}$ be the probability measure corresponding to the distribution of (X_i, Z_i) , $i = 1, \dots, n$. The $(P_{\theta\phi}, (\mathcal{F}_t))$ -compensator for \mathbf{N} can now be found directly from Jacod's representation or alternatively from the conditional independence - version of the product construction.

This shows that the compensator for N_i only depends on θ and that it equals

$$\Lambda_i(t, \theta) = \int_0^t Y_i(u) \alpha_i(u, \theta) du.$$

The likelihood takes the form

$$L(\theta, \phi) = L_0(\theta, \phi) L_\tau(\theta)$$

as in Example 2.4, where

$$L_0(\theta, \phi) = P_{\theta\phi}(\mathbf{Z})$$

and

$$L_\tau(\theta) = \prod_{i \in \mathcal{I}} \left\{ (1 - \sum_i \alpha_i(t, \theta) Y_i(t) dt)^{1 - dN_i(t)} \prod_{i=1}^n (\alpha_i(t, \theta) Y_i(t))^{dN_i(t)} \right\}.$$

If the distribution of \mathbf{Z} does not depend on θ then for each fixed $\phi \in \Phi$ $L_\tau(\theta)$ is the full likelihood for θ ; otherwise it is a partial likelihood and the full conditional likelihood given \mathbf{Z} . No sufficiency reduction of $(N_i(\cdot), Z_i)$, $i = 1, \dots, n$ is possible when β is unknown. It is easily seen that $L_\tau(\theta)$ is equal to

$$\prod_{i=1}^n \alpha_0(X_i, \gamma) r(\beta^T z_i) \exp \left\{ -r(\beta^T z_i) \int_0^{X_i} \alpha_0(u, \gamma) du \right\}$$

the density for the conditional distribution of $\mathbf{X} = (X_1, \dots, X_n)$ given $\mathbf{Z} = \mathbf{z}$, evaluated at \mathbf{X} .

Combining the present example with Example 2.3 a regression model for the *relative* mortality is obtained (ANDERSEN et al., 1985). Here the hazard rate function for X_i given $\mathbf{Z} = \mathbf{z}$ is

$$\alpha_i(t, \theta) = \alpha_0(t, \gamma) \mu_i(t) r(\beta^T z_i)$$

where $\theta = (\gamma, \beta)$ and $\mu_i(\cdot)$ is *known*. Also Example 2.4 can be combined with the present one into regression models for the transition intensities in a Markov process of the form

$$\alpha_{hji}(t, \theta) = \alpha_{hj0}(t, \gamma) r(\beta^T z_{hi}).$$

Here *type specific covariates* Z_{hi} may be defined from the vector of basic covariates Z_i for individual i reflecting the fact that some of these basic covariates may affect the different transition intensities

differently (e.g. ANDERSEN & BORGAN, 1985). \square

In the previous examples we have re-formulated well known models for complete life history data in terms of multivariate counting processes. We shall conclude this section by two models which are conveniently formulated directly as counting processes.

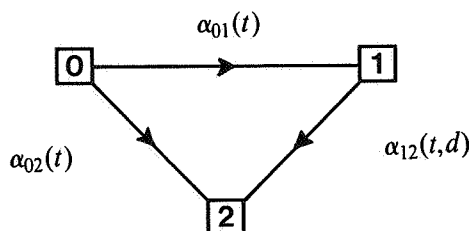
EXAMPLE 2.7. *A model for matings of Drosophila flies* (unpublished data of F.B. CHRISTIANSEN, 1969, cf. AALEN, 1978). In each mating experiment 30 female virgin flies and 40 male virgin flies were inserted into an observation arena, the 'pornoscope', consisting of a circular plastic bowl 1 cm high, with diameter 17 cm covered by a transparent lid. The flies were observed continuously and times of initiation and termination of matings were observed. Each fly mates at most once. In particular the number of ongoing matings is known for each time. The observation times are unpaired in the sense that it is unknown which termination times correspond to which initiation times.

We let $N(t)$ be the number of matings initiated in the interval $[0, t]$ and $F(t)$ and $M(t)$ the number of female and male flies respectively not yet having initiated a mating just before time t . Thus $F(t) = f_0 - N(t-)$ and $M(t) = m_0 - N(t-)$ where $f_0 = F(0)$ and $m_0 = M(0)$ are the number of female and male flies respectively in the pornoscope. Let $\mathcal{G}_t = \sigma(N(u), 0 \leq u \leq t)$ be generated by $N(\cdot)$ on an interval \mathcal{T} . Then a model for a univariate counting process $N(\cdot)$ can be set up by assuming that for a given locally integrable function $\alpha^\theta(t)$ parametrized by some θ the $(P_\theta, (\mathcal{G}_t))$ -intensity process for $N(t)$ is

$$\lambda^\theta(t) = \alpha^\theta(t)F(t)M(t).$$

It can be shown from Jacod's representation that a counting process with this intensity process exists and is unique on (\mathcal{G}_t) . Thus we have another example of Aalen's multiplicative intensity model. The interpretation of $\alpha^\theta(t)$ is that of an individual mating intensity since it is the intensity of $N(t)$ when $M(t) = F(t) = 1$. \square

EXAMPLE 2.8. *An illness-death process with duration dependence.* Let states 0, 1 and 2 denote healthy, diseased and dead and define the counting process of transitions



between these states by $N(t) = (N_{01}(t), N_{02}(t), N_{12}(t))$ where $N_{0h}(t)$, $h = 1, 2$, has intensity process $\alpha_{0h}(t)Y_0(t)$ with $Y_0(t) = 1 - N_{01}(t-) - N_{02}(t-)$, while $N_{12}(t)$ has intensity process $\alpha_{12}(t, t-T)Y_1(t)$ with $Y_1(t) = N_{01}(t-) - N_{12}(t-)$ and $T = \inf\{t : N_{01}(t) = 1\}$. Thus $Y_h(t)$ indicates that the individual is in state h at time $t-$, while T is the time of transition from 0 to 1 (if this transition ever occurs). It is seen that the intensity $\alpha_{12}(t, d)$ of dying while diseased depends on both *time* t and *duration* d . This is *not* a multiplicative intensity model, since $\alpha_{12}(t, t-T)Y_1(t)$ cannot generally be written as a product of a deterministic function and a stochastic process independent of the parameter, except in the particular case when $\alpha_{12}(t, d)$ only depends on t , and the process corresponds to a *Markov illness-death process*, cf. Example 2.4 above. (When $\alpha_{12}(t, d)$ only depends on d , one has a special case of a *semi-Markov* or *Markov renewal process*). \square

3. RIGHT CENSORING

3.1. Introduction

In this section we consider multivariate counting processes composed of n individual processes (each of which may be multivariate) and we shall see how the most common form of incomplete observation, *right censoring*, may be 'superimposed' onto such a model.

Right censoring will often introduce extra random variation in which case we first have to *enlarge* the filtrations compared to the (\mathcal{F}_t) considered in the uncensored examples in Section 2, where $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{N}_t$ with $\mathcal{F}_0 = \sigma(X_0)$. So, now a multivariate counting process

$$N = (N_{hi}; i = 1, \dots, n; h = 1, \dots, k)$$

defined on some space (Ω, \mathcal{F}) is considered with compensator Λ^θ with respect to some filtration $(\mathcal{G}_t) \supseteq (\mathcal{F}_t)$ and a family of probability measures

$$\mathcal{P} = \{P_{\theta\phi}, \theta \in \Theta, \phi \in \Phi\}.$$

Typically \mathcal{P} now also describes the censoring distribution. However, we do assume that *the compensator is the same as if there had been no censoring, i.e. the same as with respect to the original ('small') filtration, (\mathcal{F}_t)* . We consider this to be the most appropriate mathematical formulation of the notion of *independent censoring*, to which we shall return in Subsection 3.3. We shall indicate explicitly in the examples below how this may be achieved. As before, i indexes *individuals* and h types of events that the individuals may experience; for example $h = 1, \dots, k$ may indicate the different causes of death in a competing risks model (cf. Example 2.5).

Right censoring of N is the situation where observation of $N_{hi}(\cdot)$, $h = 1, \dots, k$ is ceased after some (possibly random) time U_i , i.e. N_{hi} is only observed on the random set $E_i = \{t \leq U_i\} \subseteq \mathcal{T}$ or equivalently when the process

$$C_i(t) = I(t \in E_i) = I(t \leq U_i) \quad (3.1)$$

is unity. Thus, right censoring is imposed onto N by *individual right censoring processes* $C_1(\cdot), \dots, C_n(\cdot)$; AALEN (1975, 1978), AALEN & JOHANSEN (1978).

We shall assume that the censoring process $C = (C_i, i = 1, \dots, n)$ is *predictable* with respect to (\mathcal{G}_t) . Since each $C_i(\cdot)$ is left continuous this is the case if C is *adapted*, i.e. if the U_i 's are *stopping times* with respect to (\mathcal{G}_t) . The interpretation is that *censoring may depend only on the past and not on future events*. In the concrete examples of censoring to be discussed in the following, the censoring process will typically depend only on i and not on h . It is easily seen, however, that the calculations will go through virtually unchanged with $C_i(\cdot)$ replaced by $C_{hi}(\cdot)$. Thus different censoring mechanisms for the different types h of transitions can be handled within the framework in which we are working, the important thing being that the censoring processes are *predictable*.

After (possibly) having enlarged the filtrations from (\mathcal{F}_t) to (\mathcal{G}_t) to include any additional random variation in the right censoring times we shall now *reduce* the filtrations again by specifying which data are available to the researcher at any time t after censoring. Thus we do *not* in general assume \mathcal{G}_t to represent the data at time t . First of all, the observable part of N or the *right censored counting process* $N^c = (N_{hi}^c)$ is given by

$$N_{hi}^c(t) = \int_0^t C_i(s) dN_{hi}(s). \quad (3.2)$$

Since

$$N_{hi}(t) = \Lambda_{hi}(t) + M_{hi}(t)$$

where M_{hi} is a local square integrable martingale we have that

$$N_{hi}^c(t) = \int_0^t C_i(s) d\Lambda_{hi}(s, \theta) + \int_0^t C_i(s) dM_{hi}(s)$$

$$= \Lambda_{hi}^c(t, \theta) + M_{hi}^c(t).$$

Here the latter term is again a local square integrable martingale by the predictability of $C_i(\cdot)$. Thus N_{hi}^c has $(P_{\theta\phi}, (\mathcal{G}_t))$ -compensator

$$\Lambda_{hi}^c(t, \theta) = \int_0^t C_i(s) d\Lambda_{hi}(s, \theta).$$

For the special case where N satisfies Aalen's *multiplicative intensity model* $\lambda_{hi}(t, \theta) = \alpha_{hi}(t, \theta) Y_{hi}(t)$ with respect to (\mathcal{G}_t) it is seen that also N^c satisfies the multiplicative intensity model with respect to (\mathcal{G}_t) with intensity process given by

$$\lambda_{hi}^c(t, \theta) = \alpha_{hi}(t, \theta) Y_{hi}^c(t)$$

where

$$Y_{hi}^c(t) = C_i(t) Y_{hi}(t).$$

That is, the observable counting process has the same 'individual intensity' α_{hi}^θ as the uncensored process but the random part $Y_{hi}^c(t)$ of the intensity process must in most examples be interpreted as the (predictable) indicator process for individual i being *observed to be at risk* for experiencing a type h event just before time t .

Next, we turn to a discussion of the available data at time t in addition to the right censored counting process $(N^c(u); 0 \leq u \leq t)$. As in Section 2 we assume that X_0 is observed (when relevant). We do *not* assume the whole censoring process C to be observed and it is then a question for each individual i of whether or not the value of U_i is observed. The situation is most easily thought of by introducing the concept of an *absorption time* by which we shall mean a (possibly random) time $\tau_i \in \bar{\mathcal{T}}$ with the property that all $\lambda_{hi}^\theta(t) \equiv 0$, $h = 1, \dots, k$ for $t \geq \tau_i$. In the case of uncensored survival data (Examples 2.2, 2.3) we have $\tau_i = X_i$. In an uncensored Markov process (Example 2.4) we have $\tau_i = \inf_{t \in \mathcal{G}} \{t : X_i(t) \in A\}$ where $A \subseteq S$ is the subset of absorbing states.

The idea is that typically, when the time of absorption for individual i precedes U_i , then $(N_{hi}(t), h = 1, \dots, k; t \in \mathcal{G})$ is observed since in this case $N_{hi}(t) \equiv N_{hi}^c(t)$, but U_i itself is usually *not observed*. If there is no absorption of individual i before time U_i then $(N_{hi}^c(t), h = 1, \dots, k, t \leq U_i)$ is observed *together with the value of U_i* .

In the case when N satisfies Aalen's multiplicative intensity model the observations at time t can thus be specified as

$$(X_0, (N^c(u); Y^c(u)); 0 \leq u \leq t)$$

where $N^c = (N_{hi}^c; h = 1, \dots, k, i = 1, \dots, n)$ and $Y^c = (Y_{hi}^c; h = 1, \dots, k, i = 1, \dots, n)$. Equivalently the observed data at time t can be specified as the σ -algebra

$$\mathcal{F}_t^c = \sigma(X_0, (N^c(u), Y^c(u)); 0 \leq u \leq t).$$

Then it follows that the observed counting process N^c also satisfies a multiplicative intensity model with respect to the filtration (\mathcal{F}_t^c) generated by the observed family of σ -algebras since this intensity process is given by

$$E_{\theta\phi}(\lambda_{hi}^c(t, \theta) | \mathcal{F}_t^c) = \alpha_{hi}(t, \theta) Y_{hi}^c(t),$$

by the innovation theorem (see e.g. AALEN, 1978, Theorem 3.4). Here we used the fact that $Y_{hi}^c(t)$ (by definition of \mathcal{F}_t^c) is adapted to \mathcal{F}_t^c . So, in this respect censoring by a predictable process $C(\cdot)$ preserves the multiplicative intensity model.

In general, the compensator $\Lambda_{hi}^c(\cdot, \theta)$ may depend on θ and on the past observations in a more complicated way than specified by Aalen's model. We shall assume that the observations up to time t enable the researcher to calculate $\Lambda_{hi}^c(t, \theta)$ for any given value of the unknown parameter θ . We may

therefore specify the observations available at time t as

$$(X_0, (N^c(u), 0 \leq u \leq t); \mathcal{L}^c(t))$$

where

$$\mathcal{L}^c(t) = (\Lambda^c(u, \theta); \theta \in \Theta, 0 \leq u \leq t) \quad (3.3)$$

is the family of $P_{\theta\phi}$ -compensators for N^c with respect to (\mathcal{G}_t) . Alternatively the data at time t may be given as the σ -algebra

$$\mathcal{F}_t^c = \sigma(X_0, (N^c(u), 0 \leq u \leq t); \mathcal{L}^c(t)). \quad (3.4)$$

Then (3.3) is also the family of $(P_{\theta\phi}, (\mathcal{F}_t^c))$ -compensators for the right censored counting process N^c .

In non-degenerate cases the two definitions of (\mathcal{F}_t^c) for the multiplicative model coincide, e.g. when $\alpha_{hi}(\cdot, \theta)$ is positive on \mathfrak{F} .

3.2. Examples of models for right censoring mechanisms

We shall first consider the special case where a multivariate counting process N is defined from independent survival times X_1, \dots, X_n (Examples 2.2, 2.3, 2.6). Here, observation of $(N_i^c, Y_i^c; i = 1, \dots, n)$ amounts to observing $(\tilde{X}_i, D_i; i = 1, \dots, n)$ where $\tilde{X}_i = X_i \wedge U_i$; and $D_i = I(\tilde{X}_i = X_i)$. Thus for each individual an observation time \tilde{X}_i and information on whether or not X_i is a failure time or a censoring time is available. If $D_i = 1$ then the censoring time U_i is usually not observed; see, however, Example 3.3 below.

We shall now see how some commonly used models for right censoring fit into this set-up. In Examples 2.2 and 2.3 we had a multivariate counting process with components defined by $N_i(t) = I(X_i \leq t)$ and we studied the compensator with respect to the self-exciting filtration (\mathcal{G}_t) . We first consider two examples of censoring processes $C_i(t) = I(t \leq U_i)$ predictable with respect to this filtration, i.e. examples where the original filtration (\mathcal{G}_t) need not be enlarged to include the censoring.

EXAMPLE 3.1. Survival data and simple type I censorship. Here the observation of each individual is ceased at a common, deterministic time u_0 so $C_i(t) = I(t \leq u_0)$ is non-random and trivially predictable with respect to any filtration. This censoring scheme is most common in industrial life testing where n identical items are put on test simultaneously and observed on a fixed interval $[0, u_0]$. \square

EXAMPLE 3.2. Survival data and simple type II censorship. In this case the experiment is terminated at the time of the r 'th failure, $r \leq n$, i.e. $U_i = X_{(r)}, i = 1, \dots, n$. Then $C_i = I(t \leq X_{(r)})$ is predictable with respect to (\mathcal{G}_t) , $X_{(r)}$ being a stopping time with respect to this filtration. Type II censorship is rarely applied in medical contexts but it is more common in industrial life testing experiments. Notice that in this example the censored observation times $\tilde{X}_1, \dots, \tilde{X}_n$ are dependent. \square

When N has $(P_{\theta\phi}, (\mathcal{G}_t))$ -compensator Λ^θ then by Jacod's representation also the $P_{\theta\phi}$ -compensator with respect to any enlarged filtration (\mathcal{G}_t) given by $\mathcal{G}_t = \mathcal{G}_0 \vee \mathcal{G}_t$ with \mathcal{G}_0 generated by a random variable X_0 independent of (X_1, \dots, X_n) is Λ^θ . (Here ϕ may be parameters for the distribution of X_0). This is used in the next two examples.

EXAMPLE 3.3. Survival data and progressive type I censorship. In clinical trials patients often enter the study consecutively while the study is closed at a particular date. When the interest focuses on the life time from entry (which might be the case if patients are randomised to some treatment at entry) the maximal time under study for patient i will be the time U_i from entry to the closing date. If the survival times from entry are independent of the entry times X_0 , say, and if we let \mathcal{G}_0 be generated by the entry times then the censoring process C with components $C_i(t) = I(t \leq U_i)$ is predictable with

respect to (\mathcal{G}_t) . Another way of stating this is to say that given \mathcal{G}_0 , the censoring times $U_1 = u_1, \dots, U_n = u_n$ are *deterministic* and thus we have the generalisation of Example 3.1 known as progressive type I censorship.

In this example ϕ may parametrize the arrival time process. Also in this example *all* the censoring times will be observable since it will be known when a patient would have left the study if he or she had not died before the closing date. This is of course a consequence of the assumption that *all* censoring is caused by patients being alive at the closing date, cf. the remark at the end of this subsection. \square

EXAMPLE 3.4. Survival data and random censorship. A generalisation of Examples 3.1 and 3.3 is the *general random censorship model* where $\mathbf{U} = (U_1, \dots, U_n)$ is independent of $\mathbf{X} = (X_1, \dots, X_n)$ but where \mathbf{U} may have an arbitrary distribution.

The *classical* or *simple* random censorship model in which U_1, \dots, U_n are assumed to be i.i.d. is the mathematically most tractable model for the censoring mechanism and it underlies the majority of papers on the analysis of survival data. In any case, as long as \mathbf{U} is independent of \mathbf{X} we can introduce $\mathcal{G}_0 = \sigma(\mathbf{U})$ and the censoring processes are then adapted to $\mathcal{G}_t = \mathcal{G}_0 \vee \mathcal{A}_t$ as required. However, except for the progressive type I censorship model in the previous example it is usually intuitively very unnatural to model the censoring times as being realised at time 0, and this would also be contrary to the interpretation that at time t the data can be summarized as the σ -algebra \mathcal{F}_t .

Therefore, we shall now discuss the random censorship model from another, mathematically equivalent, point of view. The approach that we shall adopt is the marked point process approach of ARJAS & HAARA (1984). This may at first glance look unnecessarily complicated for the study of the random censorship model. But we take this rather simple case as an introduction to the marked point process approach that we shall use in much more generality in the next subsection.

The basic tool of Arjas & Haara is a *marked point process* on the time interval \mathcal{T} specified by time epochs $0 < T_1 < T_2 < \dots$ and marks $x_i \in E$, $i = 1, 2, \dots$. Here the *mark space* E is written as $E = E' \times E''$ where E' contains the so-called *innovative* marks x' and E'' the *non-innovative* marks x'' . The interpretation is that the innovative marks signal the occurrence of the events of interest (e.g. failures) and hence contain information on the parameter θ , whereas the non-innovative marks signal censoring events. Either of the components of (x', x'') may be the *empty mark* \emptyset' or \emptyset'' respectively but $\emptyset = (\emptyset', \emptyset'') \notin E$. In fact, by $E' \times E''$ we mean precisely $E' \times E'' \setminus \{\emptyset\}$.

We first consider the univariate case and let X and U be non-negative random variables which are mutually independent and assume that X has hazard rate function $\alpha(\cdot, \theta)$ and that the distribution of U depends on parameters ϕ . (The distribution of U need not be absolutely continuous). We can represent the complete observation of X and U by a marked point process with $E' = \{\emptyset', d\}$ and $E'' = \{\emptyset'', c\}$ in the following way: there is an event at time t with mark x if and only if

$$\begin{array}{lll} \text{either} & t = X \text{ and } t \neq U \text{ in which case} & x = (d, \emptyset'') \\ \text{or} & t = U \text{ and } t \neq X \text{ in which case} & x = (\emptyset', c) \\ \text{or} & t = U = X & \text{in which case} & x = (d, c). \end{array}$$

We now let (\mathcal{G}_t) be the filtration generated by this marked point process. Then the process $N(t) = I(X \leq t)$ counting the number of *innovative marks* in $[0, t]$ has the same compensator

$$\Lambda(t, \theta) = \int_0^t \alpha(u, \theta) I(X \geq u) du$$

with respect to (\mathcal{G}_t) as it has in the model without censoring, i.e. with respect to $(\mathcal{F}_t) = (\mathcal{A}_t)$. This is (by the independence of X and U) an immediate consequence of Jacod's representation.

That the observation of 'the failure time' X may be prevented by right censoring at U now corresponds to the situation where observation is terminated at the *stopping time* $X \wedge U$ and that the mark $x = (x', x'')$ is observed at that time, i.e. whether $X \wedge U$ is a failure time or a censoring time (or

both). Thus the censoring process $C(t) = I(U \geq t)$ is (\mathcal{G}_t) -predictable and it follows that the censored process N^c has compensator

$$\Lambda^c(t, \theta) = \int_0^t \alpha(u, \theta) I(X \wedge U \geq u) du$$

with respect to (\mathcal{G}_t) and hence also with respect to the filtration (\mathcal{F}_t) generated by the *censored marked point process* corresponding to the observation of $X \wedge U$ and the mark x at that time.

In the case $n > 1$ with X_1, \dots, X_n i.i.d. and (U_1, \dots, U_n) independent of X_1, \dots, X_n one can go through the same arguments. From the 'large' marked point process given by all the X 's and all the U 's, we can first define a multivariate counting process N counting the *innovative* marks corresponding to the X 's. We can then calculate the $P_{\theta\phi}$ -compensator $\Lambda(\cdot, \theta)$ with respect to the entire history (\mathcal{G}_t) of the large process and notice that it coincides with the original (\mathcal{F}_t) -compensator in the model without censoring. Next, we can define a censored process N^c counting only the observed innovative marks and we can define observed censoring times. Finally, we can calculate the $P_{\theta\phi}$ -compensator $\Lambda^c(\cdot, \theta)$ of N^c with respect to the filtration (\mathcal{F}_t) generated by the process (say, N^*) consisting of observed failures and censoring times and notice that the multiplicative intensity structure is preserved. If U_1, \dots, U_n are mutually independent this result can alternatively be derived from the univariate case and the product construction. Thus the general random censorship model fits into our framework without modelling censoring times as being realised at time 0. \square

EXAMPLE 3.5. Censoring by competing risks. If interest in a study of survival data focuses on deaths from one specific cause then one may wish to consider deaths due to other causes as right censorings. As seen in Example 2.5 the competing risks model is a simple random censorship model as just discussed. However, as mentioned in that example the existence of the independent latent failure times is debatable. Hence we shall now demonstrate how the situation can be modelled using the marked point process approach introduced in Example 3.4 to the competing risks model.

Let $\alpha_{01}(t, \theta)$ denote the hazard function for the cause of interest and let $\alpha_{0j}(t, \theta, \phi)$, $j = 2, \dots, k$ be the other cause specific hazards. Consider a single individual, i , define $N_i = (N_{0ji}, j = 1, \dots, k)$ as in Example 2.4 and consider the self-exciting filtration (\mathcal{N}_t) . We can then identify N_i with a marked point process with mark space $E = E' \times E''$ as in Example 3.4, a non-empty innovative mark $j = 1$ and non-empty non-innovative marks $j = 2, \dots, k$. Then Example 2.5 shows that the $(P_{\theta\phi}, (\mathcal{N}_t))$ -compensator for the component $N_{01i}(t)$ counting the number of innovative marks in $[0, t]$ is

$$\Lambda_{1i}(t, \theta) = \int_0^t \alpha_{01}(u, \theta) Y_{0i}(u) du.$$

Here $Y_{0i}(t) = 1 - N_{0i}(t-)$ indicates whether individual i is alive (i.e. in state 0) at time $t-$. We can now define the right censoring process $C_i(t) = I(t \leq U_i)$ where $U_i \leq \tau$ is the (\mathcal{N}_t) -stopping time:

$$U_i = \inf_{t \in \mathcal{D}} \{t : N_{0ji}(t) = 1; j = 2, \dots, k\}.$$

Observation of the *censored marked point process* now corresponds to observing $\tilde{X}_i = X_i \wedge U_i$ where

$$X_i = \inf_{t \in \mathcal{D}} \{t : N_{01i}(t) = 1\},$$

and the mark (\emptyset', c) if $\tilde{X}_i = U_i$ or the mark (d, \emptyset'') if $\tilde{X}_i = X_i$. The component $N_{01i}^c(t) = N_{01i}(t)$ counting the number of innovative marks in $[0, t]$ in the censored marked point process then has $(P_{\theta\phi}, (\mathcal{N}_t))$ -compensator $\Lambda_{1i}(\cdot, \theta)$ which is adapted also to the filtration (\mathcal{F}_t) generated by the censored marked point process (since $Y_{0i}(t) = I(\tilde{X}_i \geq t)$). Thus $\Lambda_{1i}(\cdot, \theta)$ is also the $(P_{\theta\phi}, (\mathcal{F}_t))$ -compensator for N_{01i}^c and this means that we can make inference on θ (and hence of the cause specific hazard of interest, α_{01}^θ) in the presence of the competing risks by considering deaths from other causes as censorings. \square

EXAMPLE 3.6. *The illness-death process.* In the illness-death model defined in Example 2.8, consider a separate study of $N_{01}(t)$, the healthy-diseased transition. This furnishes an example of (random) censoring by a competing risk: the transition $0 \rightarrow 2$ (death, while healthy) removes the individual from being at risk for the transition $0 \rightarrow 1$ (becoming ill). \square

In Examples 3.1 to 3.6 we have demonstrated how some right censoring mechanisms only depending on the previous history of \mathbf{N} or on outside random variation preserve the multiplicative intensity model in the case of survival data. Next, we shall consider right censoring in the Markov process of Example 2.4.

EXAMPLE 3.7. *Right censoring in a Markov process.* It is easily seen going through Examples 3.1 to 3.4 that the same arguments will apply starting with the uncensored Markov process model of Example 2.4. Thus for example 'censoring at the r 'th transition from state h to state j ' would be an admissible censoring scheme in this setting, being generated by an (\mathcal{F}_t) -stopping time.

It should also be mentioned that in a Markov process model censoring may depend on the initial states generating \mathcal{F}_0 . For example one could have random censorship with different distributions according to the state in which the individuals were at time 0. \square

The last remark in Example 3.7 leads to another general class of censoring mechanisms relevant for regression models such as that of Example 2.6.

EXAMPLE 3.8. *Censoring depending on covariates.* In the relative risk regression models (Example 2.6) the filtration considered was of the form $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{G}_t$ with \mathcal{F}_0 generated by time-independent covariates Z_1, \dots, Z_n . This means that the previously mentioned models for censoring mechanisms (Examples 3.1-3.6) can be combined with censoring depending on the covariates generating \mathcal{F}_0 . Thus in a simple two sample case there may be different censoring distributions in the two samples. Also, a possible censoring scheme in a survival study with time since entry as the basic time scale and age at entry and sex included as covariates would be every year to censor e.g. the oldest woman still alive. Recalling that the censoring process has to be adapted to an extended filtration (\mathcal{G}_t) , it is crucial that the extension generated by the covariates upon which censoring depends does not change the compensator of \mathbf{N} . \square

Let us finally mention that all the models for right censoring mechanisms discussed in this subsection may be *combined* to more general models for predictable processes \mathbf{C} . Examples include censoring in a clinical trial with staggered entry (Example 3.3). Here censoring may, in addition to being caused by patients surviving until the closing date be a consequence of patients dying from causes unrelated to the one being studied (as in Example 3.5).

Censoring of type I or type II (Examples 3.1 and 3.2) are also relevant for the porthoscope model in Example 2.7. Thus one might here choose to terminate the experiment at a fixed time u_0 or at the time of initiation of the r 'th mating. In the actual experiment the censoring that was used was as follows. Let $X_{(1)}$ denote the time of initiation of the first mating. Then the experiment was terminated at the first time after $X_{(1)} + 45$ minutes where no matings were going on — no later, however, than at time $X_{(1)} + 60$ minutes. It is seen that this censoring satisfies our requirements provided that the times of termination of matings are included in the filtration (\mathcal{G}_t) and do not thereby alter the mating intensities.

3.3. Independent and non-informative right censoring

We now return to the general set-up considered in Subsection 3.1 with the purpose of discussing the concepts of *independent* right censoring and *non-informative* right censoring. So, we let $\mathbf{N} = (N_{hi}; i = 1, \dots, n; h = 1, \dots, k)$ be a multivariate counting process on some space (Ω, \mathfrak{F}) and we let $\mathfrak{F}_t = \mathfrak{F}_0 \vee \mathcal{N}_t$ with \mathfrak{F}_0 generated by a random variable \mathbf{X}_0 realised at time 0. Furthermore $\mathbf{C} = (C_1, \dots, C_n)$ is a right censoring process, \mathbf{C} is predictable with respect to a given, possibly larger filtration $(\mathfrak{G}_t) \supseteq (\mathfrak{F}_t)$ on the same space. Suppose further that a family of probabilities $\mathcal{P} = \{P_{\theta\phi}; (\theta, \phi) \in \Theta \times \Phi\}$ with all $P_{\theta\phi}$'s mutually equivalent is given and that the compensator of \mathbf{N} with respect to $P_{\theta\phi}$ and (\mathfrak{G}_t) , $\Lambda^\theta = (\Lambda_{hi}^\theta; i = 1, \dots, n; h = 1, \dots, k)$ is identical to the (\mathfrak{F}_t) -compensator for \mathbf{N} . In particular the compensator does not depend on ϕ . We shall denote right censoring generated by a process \mathbf{C} with these properties *independent right censoring*. Thus θ is the parameter of interest whereas \mathbf{C} depends on the nuisance parameter $\phi \in \Phi$ and it may depend on θ as well.

The (\mathfrak{F}_t) -likelihood based on observation of \mathbf{N} and \mathbf{X}_0 , whose distribution may also depend on ϕ and maybe on θ too, is by Jacod's formula (2.1)

$$\begin{aligned} L(\theta, \phi) &= L_0(\theta, \phi) \prod_{t \in \mathfrak{T}} \{(1 - d\Lambda^\theta_{..}(t))^{1 - dN_{..}(t)} \prod_{i=1}^n \prod_{h=1}^k d\Lambda_{hi}^\theta(t)^{dN_w(t)}\} \\ &= L_0(\theta, \phi) L_\tau(\theta). \end{aligned} \quad (3.5)$$

Here $N_{..} = \sum_{i=1}^n \sum_{h=1}^k N_{hi}$, and $\Lambda^\theta_{..}$ is the $P_{\theta\phi}$ -compensator of $N_{..}$ with respect to (\mathfrak{F}_t) . If \mathbf{X}_0 does not depend on θ then $L_\tau(\theta)$ is the full likelihood for θ based on observation of \mathbf{N} , otherwise $L_\tau(\theta)$ is a partial likelihood and the full conditional likelihood given \mathfrak{F}_0 . Consider now the right censored counting process

$$\mathbf{N}^c = (N_{hi}^c, i = 1, \dots, n; h = 1, \dots, k),$$

where

$$N_{hi}^c(t) = \int_0^t C_i(s) dN_{hi}(s).$$

Then for all $\phi \in \Phi$ the $(P_{\theta\phi}, (\mathfrak{G}_t))$ -compensator of N_{hi}^c is

$$\Lambda_{hi}^c(t, \theta) = \int_0^t C_i(s) d\Lambda_{hi}(s, \theta).$$

As explained in Subsection 3.1 we assume that the observations available at time t include \mathbf{X}_0 and $(\mathbf{N}^c(u), 0 \leq u \leq t)$ together with right censoring times $U_i \leq t$ for individuals for which there is no time of absorption before U_i . Under non-degeneracy conditions (e.g. $\alpha_{hi}(t, \theta) > 0$ for $t \in \mathfrak{T}$ for the multiplicative intensity model), the observations may be formalised as $(\mathbf{X}_0, \mathbf{N}^c, \mathcal{L}(t))$ with $\mathcal{L}(t)$ given by (3.3). In what follows we shall consider these observations as a *marked point process*

$$\mathbf{N}^* = (\mathbf{N}^c, \mathbf{N}'')$$

with *innovative* marks at jump times for \mathbf{N}^c and *non-innovative* marks at observed right censoring times as in Examples 3.4 and 3.5 and at $t=0$. There may be simultaneous marks. Thus \mathbf{N}'' is the process recording the observed right censoring times with marks and which also has a mark at $t=0$ recording \mathbf{X}_0 . These two ways of describing the data are equivalent in the sense that for any $t \in \mathfrak{T}$

$$\sigma(\mathbf{X}_0, (\mathbf{N}^c(u), 0 \leq u \leq t); \mathcal{L}(t)) = \sigma(\mathbf{N}^*(u), 0 \leq u \leq t)$$

both being equal to \mathfrak{F}_t^* (defined in (3.4)).

We shall now write down parallel to (3.5) the $P_{\theta\phi}$ -likelihood $L^*(\theta, \phi)$ for \mathbf{N}^* with respect to (\mathfrak{F}_t^*) and rewrite it along the lines of ARJAS & HAARA (1984). We shall keep the calculations at an informal and

intuitive level and refer the reader to ARJAS & HAARA (1984) for the full mathematical details.

An element x in the mark space $E = E' \times E''$ of N^* can be written as a pair $x = (x', x'')$. Here at time t , $x' = (i, h)$ if N_{hi}^c jumps at t while $x' = \emptyset'$ if only censorings occur at t , and x'' indicates the subset of $\{1, \dots, n\}$ of individuals censored at t where we may have $x'' = \emptyset''$. Thus

$$N^* = (N_x^*; x = (x', x''))$$

and

$$N_{hi}^c = \sum_{x''} N_{(x', x'')}^* \quad \text{with } x' = (i, h).$$

We can now write, starting from Jacod's formula (2.1) applied to N^* ,

$$\begin{aligned} L_\tau^*(\theta, \phi) &= L_0(\theta, \phi) \prod_t P_{\theta\phi} (dN^*(t) | \mathcal{F}_t^-) \\ &= \prod_t P_{\theta\phi} (dN^c(t) | \mathcal{F}_t^-) \cdot L_0(\theta, \phi) \prod_t P_{\theta\phi} (dN^*(t) | dN^c(t), \mathcal{F}_t^-) \\ &= L_\tau^c(\theta) L_\tau''(\theta, \phi), \end{aligned} \quad (3.6)$$

where the contribution from X_0 has been absorbed in the second factor. Here the first factor equals

$$L_\tau^c(\theta) = \prod_t \{ (1 - d\Lambda^c(\cdot, \theta))^{1 - dN^c(\cdot, t)} \prod_{i,h} d\Lambda_{hi}^c(t, \theta)^{dN_{hi}^c(t)} \}, \quad (3.7)$$

where $N^c = \sum_{h,i} N_{hi}^c$ and $\Lambda^c(\cdot, \theta)$ is the $(P_{\theta\phi}, (\mathcal{F}_t^-))$ -compensator for N^c , which does not depend on ϕ . This *partial* likelihood function has the same form as the (partial) likelihood $L_\tau(\theta)$ in (3.5) based on the uncensored process N . (It should be noticed that whether or not a given U_i is observed does not alter $L_\tau(\theta)$, cf. the discussion in Section 3.1). Thus *independent right censoring mechanisms preserve the form of the (partial) likelihood*.

The fact that the form of the partial likelihood is preserved after independent right censoring has the consequence that its *martingale properties* stay the same. For instance, the 'score-process' $(\partial/\partial\theta \log L_\tau^c(\theta))$ is a $(P_{\theta\phi}, (\mathcal{F}_t^-))$ -martingale just as $(\partial/\partial\theta \log L_\tau(\theta))$ is a $(P_{\theta\phi}, (\mathcal{F}_t^-))$ -martingale in the model without censoring. This means that large sample statistical inference for independently censored data based on the partial likelihood will be much the same as that for uncensored data based on the full likelihood, since this martingale structure plays such a central role in asymptotic theory. Note also that if N satisfies Aalen's multiplicative intensity model then $Y_{hi}(t)$ in (3.5) is simply replaced by $Y_{hi}^c(t)$ in (3.7).

In the next example we shall relate our definition of independent right censoring to other suggestions in the literature in the special case of i.i.d. survival times.

EXAMPLE 3.9. Independent censoring of i.i.d. survival times. WILLIAMS & LAGAKOS (1977) considered right censoring of i.i.d. survival times X_1, \dots, X_n with hazard function $\alpha^\theta(t)$. They showed that if the model for the censored data satisfies a certain 'constant sum' condition then the likelihood for θ is proportional to (3.7). KALBFLEISCH & MACKAY (1979) showed that the constant sum condition is equivalent to another condition which is a *consequence* of our definition of independent censoring, namely that the failure intensity at time t for an individual i at risk at that time (i.e. $\tilde{X}_i \geq t$) is $\alpha^\theta(t)$. Formulated in our notation this condition simply states that the $(P_{\theta\phi}, (\mathcal{F}_t^-))$ -compensator for N_i^c is $(\int_0^t \alpha^\theta(u) Y_i^c(u) du)$. This condition was verbally formulated by COX (1975) and further discussed by KALBFLEISCH & PRENTICE (1980, p.120). It was given a precise mathematical formulation, not restricted to the (absolutely) continuous case, by GILL (1980a, Theorem 3.1.1). Thus our requirement, being a condition on the larger filtration $(\mathcal{G}_t) \supseteq (\mathcal{F}_t^-)$, is a stronger requirement for independent censoring than those considered by these various authors. However, as seen in Section 3.2 it does cover all the interesting models for right censoring in the case of survival data and furthermore it can be generalized to other models based on counting processes. \square

In the definition of independent right censoring it is required that the censoring process C is predictable. So examples of *dependent* right censoring patterns include cases where C is not adapted. This is for instance the case if C depends on covariates which are *not* included in the model. Another example is the following.

EXAMPLE 3.10. Testing with replacement. (GILL, 1980b, 1981). Suppose that objects (e.g. light bulbs) are life tested one at a time and at each failure time replaced by a new one. If observation is terminated at a fixed calendar time the last object put on test will typically still be working and hence contribute with a censored life time. Thus censoring of the last object (at life time t , say) depends on the life times of previous objects which may well exceed t . A similar situation arises in the clinical trial Example 3.3 if observation had not been terminated at a fixed calendar time but instead at the r 'th observed failure $X_{(r)}$ (as in type II censorship). With this stopping rule patients with entry times later than that of the patient with failure time equal to $X_{(r)}$ may still be alive and thus censored after a time under study less than $X_{(r)}$. For a further discussion, the reader is referred to SELLEKE & SIEGMUND (1983), SLUD (1984) and ARJAS (1985). \square

For independent censoring mechanisms an important question is now whether $L_r^c(\theta)$ is the full likelihood for θ for each fixed $\phi \in \Phi$ based on observation of (N^c, \mathcal{E}^c) . This is of course the case when for each fixed $\phi \in \Phi$ the second factor $L_r''(\theta, \phi)$ in $L_r^*(\theta, \phi)$ does not depend on θ . In this case we term the independent right censoring mechanism C *non-informative* for the parameter θ . This precise definition (but without consideration of a nuisance parameter ϕ) is due to ARJAS & HAARA (1984) who made the discussion by KALBFLEISCH & PRENTICE (1980, p.126) rigorous. In fact, Arjas & Haara termed it *non-innovative* censoring and considered a more general situation with other kinds of censoring and with time-dependent covariates. Thus their discussion included the concept of non-innovative covariates. We shall return to these situations in Sections 5 and 7.

To find conditions for C to be non-informative for θ we now look closely at $L_r''(\theta, \phi)$ denoting by $\Lambda_x^*(t, \theta, \phi)$ the $(P_{\theta\phi}, \mathcal{F}_t^c)$ -compensator for $N_x^*(t)$, where $x = (x', x'')$ and $x' = (i, h)$ or $x' = \emptyset'$. Furthermore we let $N^* = \sum_{x \in E} N_x^*$ be the total aggregated point process. There are three possible kinds of contributions to $L_r''(\theta, \phi)$ at time t corresponding to

- a) no marks at all
- b) an empty innovative mark \emptyset' and a non-empty non-innovative mark $x'' \in E'' \setminus \{\emptyset''\}$
- c) an innovative mark $(i, h) = x' \neq \emptyset'$ and a non-innovative mark $x'' \in E''$, (possibly empty).

In the first case we get the factor

$$P_{\theta\phi}(dN^*(t) = 0 | dN^c(t) = 0, \mathcal{F}_t^-) = \frac{1 - d\Lambda^*(t, \theta, \phi)}{1 - d\Lambda^c(t, \theta)}, \quad (3.8)$$

in the second case we get

$$P_{\theta\phi}(dN_{(\emptyset', x'')}^*(t) = 1 | dN^c(t) = 0, \mathcal{F}_t^-) = \frac{d\Lambda_{(\emptyset', x'')}^*(t, \theta, \phi)}{1 - d\Lambda^c(t, \theta)}, \quad (3.9)$$

while in the third case we get

$$P_{\theta\phi}(dN_{(x', x'')}^*(t) = 1 | dN_{hi}^c(t) = 1, \mathcal{F}_t^-) = \frac{d\Lambda_{(x', x'')}^*(t, \theta, \phi)}{d\Lambda_{hi}^c(t, \theta)}. \quad (3.10)$$

The (partial) likelihood for the censoring thus becomes

$$L_r''(\theta, \phi) = L_0(\theta, \phi) \prod_t \left\{ \frac{1 - d\Lambda^*(t, \theta, \phi)}{1 - d\Lambda^c(t, \theta)} \right\}^{1 - dN^*(t)} \quad (3.11)$$

$$\prod_{x'' \neq \emptyset''} \left[\frac{d\Lambda_{(\emptyset', x'')}^*(t, \theta, \phi)}{1 - d\Lambda_{\cdot}^c(t, \theta)} \right]^{dN_{(\emptyset', x'')}^*(t)} \prod_{x'=(i,h)} \prod_{x''} \left[\frac{d\Lambda_x^*(t, \theta, \phi)}{d\Lambda_{hi}^c(t, \theta)} \right]^{dN_x^*(t)}$$

Collecting factors in (3.7) and (3.11) one can recover, after some calculation,

$$L_{\tau}^c(\theta)L_{\tau}''(\theta, \phi) = L_0(\theta, \phi) \prod_t (1 - d\Lambda_{\cdot}^*(t, \theta, \phi))^{1 - dN_{\cdot}^*(t)} \prod_{x \in E} d\Lambda_x^*(t, \theta, \phi)^{dN_x^*(t)}$$

as in (3.6). These heuristic calculations, for which the probabilistic interpretation is very clear, are of course only formal algebra. However, each calculation can be rigorized when we take the product over t and the ratio over two probability measures, since then each factor becomes a term in a product integral or a Radon-Nikodym derivative for which the corresponding algebra is valid.

In order that $L_{\tau}''(\theta, \phi)$ should be independent of θ then to start with $L_0(\theta, \phi)$ should not depend on θ , and furthermore

$$\frac{d\Lambda_{(\emptyset', x'')}^*(t, \theta, \phi)}{1 - d\Lambda_{\cdot}^c(t, \theta)} \quad \text{for all } x'' \in E'' \quad (3.12a)$$

(corresponding to (3.9)) and

$$\frac{d\Lambda_{(x', x'')}^*(t, \theta, \phi)}{d\Lambda_{hi}^c(t, \theta)} \quad \text{for all } (i, h) = x' \text{ and all } x'' \in E'' \quad (3.12b)$$

(corresponding to (3.10)) should be independent of θ for all $\phi \in \Phi$. From (3.12) and the relation

$$\sum_{x'' \neq \emptyset''} \frac{d\Lambda_{(\emptyset', x'')}^*(t, \theta, \phi)}{1 - d\Lambda_{\cdot}^c(t, \theta)} + \frac{1 - d\Lambda_{\cdot}^*(t, \theta, \phi)}{1 - d\Lambda_{\cdot}^c(t, \theta)} = 1$$

it follows that then also (3.8) corresponding to the first factor in (3.11) is independent of θ . Condition (3.12) states that if for each fixed $\phi \in \Phi$ the conditional distribution of $dN^*(t)$ given \mathcal{F}_-^c and $dN^c(t)$ does not depend on θ then C is non-informative for θ . In other words, if for each fixed $\phi \in \Phi$ and each $t \in \mathcal{T}$ the conditional intensity of certain individuals being censored at t given the past up till just before t and given a possible failure at t does not depend on θ then the censoring is non-informative for θ .

In the special case of Aalen's multiplicative intensity model

$$\lambda_{hi}^c(t, \theta) = \alpha_{hi}(t, \theta) Y_{hi}^c(t, \theta)$$

where α_{hi} does not depend on i , i.e. $\alpha_{hi} = \alpha_h$, it is easily seen using (3.5) - (3.7) that if the censoring is non-informative for θ then $(N_{h\cdot}^c, Y_{h\cdot}^c; h = 1, \dots, k)$ is sufficient for θ .

Here

$$Y_{h\cdot}^c(t) = \sum_{i=1}^n Y_{hi}^c(t)$$

can often be interpreted as the total number of individuals observed to be at risk for experiencing a type h event just before time t . Thus, again, a process satisfying the multiplicative intensity model is obtained by aggregation. However, in contrast to Examples 2.2 and 2.4 with uncensored data where the aggregated counting processes $(N_{h\cdot}, h = 1, \dots, k)$ (and J_0) were themselves sufficient, $Y_{h\cdot}$ being a function of $N_{h\cdot}$ (and J_0) it is now the pairs $(N_{h\cdot}^c, Y_{h\cdot}^c, h = 1, \dots, k)$ which are sufficient under non-informative right censoring.

In Example 3.3 we have non-informative censoring provided that the entry time process does not depend on θ and in Example 3.4 if the censoring distribution does not depend on θ . In Example 3.8 we have non-informative censoring when there is random censorship with censoring distribution depending on covariates and if this distribution does not depend on θ .

Informative right censoring may occur if censoring is due to competing causes of deaths with cause specific intensities depending on θ (cf. Example 3.5). One such example is the Koziol-Green model

(KOZIOL & GREEN, 1976) where failure and censoring intensities are *proportional*. The more simple statistical procedures based on $L_i^c(\theta)$ will lose some information if we have informative censoring and in some such examples more efficient methods may be applied. We have seen, however, that most sensible models for *right* censoring mechanisms were non-informative. This is in contrast to other kinds of censoring, including *left* censoring, to which we return in Section 5.

3.4. Identifiability of independent right censoring mechanisms

So far we have strived to put conditions on the censoring pattern to make the resulting observable processes tractable. The opposite wish is to see how much of the underlying structure is uniquely given, if the observable processes are tractable. We conclude this section by indicating some recent results by JACOBSEN (1986), who studied i.i.d. survival data with hazard function $\alpha(t)$ (the restriction to identical distributions being made for convenience only).

Jacobsen's concept of independent censoring is also more restrictive than the one considered in Example 3.9 and differs slightly from ours. He studied the marked point process N^* with non-empty innovative marks at observed failure times (i.e. at jump times for N^c) and considered the joint distribution of N^* and *all* the failure times $X = (X_1, \dots, X_n)$. He then showed first that if for all t the conditional distribution of $N^*(t)$ given (X_1, \dots, X_n) only depends on the X 's through what we actually observe about them at time $t-$, i.e. that

$$X_i \geq t \text{ if } X_i \wedge U_i \geq t \quad (3.13)$$

$$X_i = x_i \text{ if } X_i = x_i < t \text{ and } U_i \geq x_i$$

$$X_i > u_i \text{ if } U_i = u_i < t \text{ and } X_i \geq u_i$$

then N_i^c has (\mathcal{F}_i^c) -compensator

$$\Lambda_i^c(t) = \int_0^t \alpha(u) Y_i^c(u) du \quad (3.14)$$

with respect to the probability measure corresponding to the joint distribution of N^* and X . Secondly, he showed that given an (\mathcal{F}_i^c) -compensator Λ^* for N^* satisfying (3.14) there exists one and only one joint distribution of N^* and (X_1, \dots, X_n) satisfying (3.13) such that X_1, \dots, X_n are i.i.d. with hazard $\alpha(\cdot)$ and N^* has compensator Λ^* .

This distribution can be *simulated* in the way described in the following Example 3.11 concerning a randomized version of what is known as *progressive type II censorship*.

EXAMPLE 3.11. Survival data and randomized progressive type II censorship. Suppose that n identical items are put on test simultaneously, as in Example 3.9, and let X_1, \dots, X_n be the i.i.d. life times that would have been observed had there been no censoring. Next, generate n potential (possibly mutually dependent) right censoring times $U_1^{(1)}, \dots, U_n^{(1)}$ independent of the X_i 's and find $X_{(1)}$, the smallest X_i with $X_i \leq U_i^{(1)}$. Items j for which $U_j^{(1)} < X_{(1)}$ and $U_j^{(1)} < X_j$ are removed at the time points $U_j^{(1)}$. At time $X_{(1)}$ new potential right censoring times $U_j^{(2)} > X_{(1)}$ are generated for each item j still on test. The joint distribution of the $U_j^{(2)}$ may depend on $X_{(1)}$ and the censoring times for those items actually removed in $[0, X_{(1)})$ (and on the labels of these items) but not on the life times X_j of the items still on test. Next, $X_{(2)}$, the smallest X_i with $X_i \leq U_i^{(2)}$ is found and items j still on test and for which $U_j^{(2)} < X_{(2)}$ and $U_j^{(2)} < X_j$ are removed at $U_j^{(2)}$. At $X_{(2)}$ the censoring times for items still on test are once more updated and allowed to depend also on $X_{(2)}$ and censorings in $[X_{(1)}, X_{(2)})$ and so on. At every point in time the decision to censor an item 'still working' may depend arbitrarily on the past observations but not on the future. \square

Jacobsen illustrated the concepts by the following 'counterexample'.

EXAMPLE 3.12. Let X_1 and X_2 be i.i.d. with hazard function $\alpha(t)$ and let the censoring variables U_1 and U_2 be given by assuming that U_1 is independent of X_1 and X_2 and exponentially (1) distributed, and by defining

$$U_2 = \begin{cases} \infty & \text{if } X_1 \leq U_1 \\ X_1 + a & \text{if } U_1 < X_1 \end{cases}$$

One may check that this strange censoring pattern satisfies Jacobsen's as well as our definition of independent censoring. Of course it is not *defined* by a simulation experiment as described in Example 3.11 above, because the censoring time U_2 depends on the value of X_1 exactly when this is *not* observed. Jacobsen's result then tells us that there exists one and only one simulation experiment resulting in exactly the same observable process as in this example. The reason one feels unhappy about this example is that, though the censoring is *independent*, it is also *informative*. Thus the problem with it is *statistical*, not *probabilistic*. \square

In our view the simulation experiment represents the canonical form of what one should understand by well-behaved right censoring. It would be desirable to obtain a more abstract formulation (perhaps as a suitable stopping time condition on the U_1, \dots, U_n) so that the concept may be defined for general counting processes.

4. LEFT TRUNCATION

The most common kind of incomplete information on life history data, *right censoring* was discussed in the previous section.

To exemplify a different kind of incomplete observation consider a study of survival among insulin dependent diabetics in Fyn county (GREEN et al., 1981, GREEN & HOUGAARD, 1985). Out of the about 450,000 inhabitants in Fyn county, Denmark, it was ascertained from prescriptions in the National Health Service files that $n = 1499$ suffered from insulin dependent diabetes mellitus on 1 July 1973. They were all followed until 1 July 1980 with the purpose of assessing the age-specific mortality of diabetics. Since a diabetic was only included in the sample conditionally on being alive on 1 July 1973, the relevant distribution to consider for the survival times X_i , $i = 1, \dots, n$ is the conditional distribution of X given $X > V$ where the entry time, V is the time since birth at 1 July 1973. The survival data are then said to be *left truncated*.

In this section we shall consider in more generality counting process based models for left truncated life history data. The set-up is analogous to that studied in Section 3. We consider a single individual, i , at a time and we drop the subscript, i . We let

$$N = (N_h, h = 1, \dots, k)$$

be a basic, untruncated multivariate counting process on a space (Ω, \mathcal{F}) with $P_{\theta\phi}$ -compensator Λ^θ and intensity process λ^θ with respect to a filtration (\mathcal{F}_t) of the form $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{N}_t$, cf. Section 3.1. We assume the existence of a larger filtration $(\mathcal{G}_t) \supseteq (\mathcal{F}_t)$ such that the $(P_{\theta\phi}, (\mathcal{G}_t))$ -compensator of N is also Λ^θ ; this is intended to carry the possible extra random variation involved in the truncation time. Furthermore, we let V be a (\mathcal{G}_t) -stopping time and consider an event $A \in \mathcal{G}_V$. The process N started at V is defined as

$${}_V N(t) = N(t) - N(t \wedge V). \quad (4.1)$$

We want to study the process N , starting from the time V , given that the event A (prior to V) has actually occurred. We call the process ${}_V N$, under this conditional distribution, a left truncated process. The proposition below (proved in the Appendix) shows that left truncation of N by the event A (before V) preserves the intensity of N after time V . For ease of presentation we suppose $P_{\theta\phi}(A) > 0$;

similar results can be obtained for any event $A \in \mathcal{G}_V$ using the technical apparatus of proper regular conditional probabilities and Blackwell spaces; see JACOBSEN (1982, Exercise 8, p. 51, and Appendix 1).

PROPOSITION 4.1. *The left truncated counting process ${}_V N$ has intensity process*

$${}_V \lambda^\theta(t) = \lambda^\theta(t) I(t > V) \quad (4.2)$$

with respect to the filtration $({}_V \mathcal{G}_t)$ given by

$${}_V \mathcal{G}_t = \mathcal{G}_t \vee \mathcal{G}_V$$

and the conditional probability $P_{\theta\phi}^A$ given by

$$P_{\theta\phi}^A(F) = P_{\theta\phi}(F \cap A) / P_{\theta\phi}(A), \quad F \in \mathcal{F}$$

Suppose (just as in the right censoring case) that as well as observing ${}_V N$, the available data also allows us to write down ${}_V \lambda^\theta$ for any $\theta \in \Theta$. If \mathcal{G}_t has the special form

$$\mathcal{G}_t = \mathcal{G}_0 \vee \mathcal{O}_t$$

then

$${}_V \mathcal{G}_t = \mathcal{G}_V \vee \sigma\{{}_V N(u); 0 \leq u \leq t\}$$

and the $(P_{\theta\phi}^A, ({}_V \mathcal{G}_t))$ conditional likelihood for ${}_V N$ given \mathcal{G}_V , also the partial likelihood for ${}_V N$, is

$${}_V L(\theta) = \prod_{t > V} \{(1 - {}_V \lambda^\theta(t) dt)^{1 - d_{{}_V N}(t)} \prod_{h=1}^k ({}_V \lambda_h^\theta(t) dt)^{d_{{}_V N_h}(t)}\}. \quad (4.3)$$

For instance, if N satisfies Aalen's multiplicative intensity model

$$\lambda_h^\theta(t) = \alpha_h^\theta(t) Y_h(t), \quad t \in \mathcal{T}, \quad h = 1, \dots, k$$

with respect to $P_{\theta\phi}$ and (\mathcal{F}_t) then the left truncated process ${}_V N$ satisfies the multiplicative intensity model

$${}_V \lambda_h^\theta(t) = \alpha_h^\theta(t) {}_V Y_h(t), \quad t > V$$

with respect to $P_{\theta\phi}^A$ and $({}_V \mathcal{G}_t)$, where

$${}_V Y_h(t) = Y_h(t) I(t > V).$$

In this case the data needed at time t , $t > V$ is $({}_V N(u), {}_V Y(u); V < u \leq t)$.

A general discussion of how the truncated data becomes available as time proceeds in parallel with the situation for censored data, cf. (3.3) and (3.4), is possible, but we do not give it here. In the most trivial example $V = v_0$ is deterministic and we have $\mathcal{G}_t = \mathcal{F}_t$. More interesting is the following example.

EXAMPLE 4.1. *Random left truncation of a survival time.* Let the random variable $X > 0$ have hazard function α_X^θ . Define $N(t) = N_X(t) = I(X \leq t)$ and let (\mathcal{O}_t) be the filtration generated by N_X . Assume that $V > 0$ is independent of X with distribution depending on parameters ϕ . Define the bivariate counting process (N_X, N_V) with components $N_X(t)$ and $N_V(t) = I(V \leq t)$ and let (\mathcal{G}_t) be the filtration it generates. The (\mathcal{G}_t) -intensity process $\lambda_X^{\theta\phi}$ for N_X with respect to the joint distribution $P_{\theta\phi}$ of X and V is (as previously, cf. Example 3.4)

$$\lambda_X^{\theta\phi}(t) = \lambda_X^\theta(t) = \alpha_X^\theta(t) I(t \leq X)$$

which is the same as the intensity process with respect to (\mathcal{O}_t) . If the event

$$A = \{X > V\}$$

has positive probability then the intensity process for the left truncated process ${}_vN_X = N_X(t) - N_X(t \wedge V)$ with respect to the conditional joint distribution $P_{\theta\phi}^A$ of X and V given A is

$${}_v\lambda_X^{\theta\phi}(t) = {}_v\lambda_X^\theta(t) = \alpha_X^\theta(t)I(V < t \leq X)$$

as shown above. We see that random left truncation preserves the multiplicative structure of the intensity process and in particular, ${}_v\lambda_X$ still only depends on θ . The conditional (given V) or partial likelihood for ${}_vN_X$ with respect to the conditional distribution given $X > V$ is by (4.3)

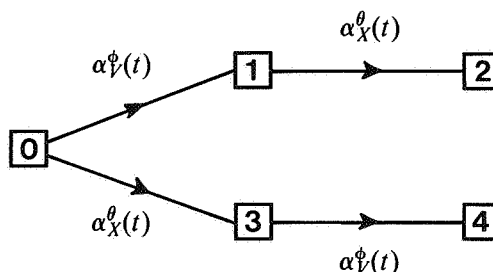
$${}_vL(\theta) = \prod_{t > V} (1 - {}_v\lambda_X^\theta(t)dt)^{1 - d_{{}_vN_X}(t)} {}_v\lambda_X^\theta(t)^{d_{{}_vN_X}(t)} = S_X^\theta(X)\alpha_X^\theta(X)/S_X^\theta(V).$$

It is seen that this is the conditional density of X given $X > v$, evaluated at (X, V) . □

When the distribution of X depends on covariates Z (Examples 2.6 and 3.8) the basic filtration is given by $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{G}_t$ with $\mathcal{F}_0 = \sigma(Z)$. In order to write down the intensity process ${}_v\lambda^\theta$ for the left truncated process ${}_vN$ we must include not only V but also Z in the ‘observed filtration’.

In Examples 2.5, 3.4 and 3.5 the relationship between the random censorship model for survival data and a certain Markov process, the competing risks model, was studied. The next example establishes the relationship between the random truncation model and a certain Markov process.

EXAMPLE 4.2. *The random truncation model for a survival time viewed as a Markov process* (KEIDING & GILL, 1987). In the model of Example 4.1, assume for convenience the distribution of V absolutely continuous with hazard $\alpha_V^\phi(t)$. Define the Markov process $U(t)$ by $U(0)=0$ and transition intensities as specified in the diagram. The random variables



X and V correspond to the times of transition from 1 to 2 or 0 to 3, and from 0 to 1 or 3 to 4, respectively. Observation of (V, X) given $V < X$ is equivalent to observing $U(t)$ in the conditional distribution given $U(\tau) = 2$ (ultimate absorption in 2). The counting process

$${}_vN_X(t) = N_X(t) - N_X(t \wedge V)$$

is identical to that counting transitions from state 1 to state 2:

$${}_vN_X(t) = N_{12}(t),$$

and it is a standard result for Markov processes (see HOEM (1969) for an explicit formulation) that the intensity $1 \rightarrow 2$ in the conditional Markov process is

$$\alpha_X^\theta(t) \frac{P(U(\tau)=2 | U(t)=2)}{P(U(\tau)=2 | U(t)=1)} = \alpha_X^\theta(t).$$

It now follows from Example 2.4, used for the conditional Markov process given $\{X > V\}$, that $N_{12}(t)$ has intensity process

$$\alpha_X^\theta(t)Y_1(t) = \alpha_X^\theta(t)I\{V < t \leq X\}$$

with respect to the conditional distribution given $V < X$ and the filtration given by

$$\mathcal{F}_t = \sigma(I\{V \leq u, V < X\}, I\{V < X \leq u\}, 0 \leq u \leq t).$$

The likelihood with respect to θ consequently has the form (4.3) which may thus in this case be obtained without recourse to the proof given there. \square

EXAMPLE 4.3. Left truncation of a Markov process. Let the Markov process $X = (X(t), t \in \mathcal{T})$, the counting process N and the filtration (\mathcal{F}_t) be defined as in Example 2.4. Let $V > 0$ be independent of X with distribution depending on parameters ϕ and define $N_V(t) = I(V \leq t)$. Then N has the same compensator both with respect to (\mathcal{F}_t) and with respect to the filtration (\mathcal{G}_t) defined by

$$\mathcal{G}_t = \mathcal{F}_t \vee \sigma(N_V(u), 0 \leq u \leq t)$$

and we can apply Proposition 4.1 to any event $A \in \mathcal{G}_V = \sigma\{V, J_0; N(t \wedge V), t \geq 0\}$ with $P_{\theta\phi}(A) > 0$. Examples include events A indicating that a certain component N_h had at least or at most some specified number of jumps before V .

As an alternative to such partly *external* truncation (V independent of X) one might consider *internal* truncation, where V is an (\mathcal{F}_t) -stopping time. The obvious examples are the first arrival time to a particular state (cf. the illness-death process to be further discussed below) or the first (or p 'th) time that the process (or one of its components) jumps. \square

EXAMPLE 4.4. The illness-death process. Consider the illness-death process earlier discussed in Examples 2.8 and 3.6. A separate study of the transition $1 \rightarrow 2$ (death when diseased) may be performed by studying the counting process N_{12} which has intensity process $\alpha_{12}(t, t-T)Y_1(t)$ where $Y_1(t) = I(\text{the individual is in state 1 at time } t-)$, $T =$ entry time into state 1; define also $X =$ time of death. Consider left truncation by the event $A = \{T \leq V < X < \tau\}$ at some random time V .

Internal truncation is obtained by choosing for instance $V = T$, that is, follow the individual from the time of disease occurrence conditioning on it occurring: the truncated counting process would be

$${}_V N_{12}(t) = N_{12}(t) - N_{12}(t \wedge V) = N_{12}(t)$$

with intensity process

$$\alpha_{12}(t, t-V)Y_1(t)I(t > V) = \alpha_{12}(t, t-T)Y_1(t).$$

In this case, except for the conditioning, left truncation is equivalent to observation of the original process N_{12} .

External truncation is exemplified by choosing V independent of the illness-death process. On the conditioning event A , the truncated counting process $N_{12}(t) - N_{12}(t \wedge V)$ has intensity process

$$\alpha_{12}(t, t-T)Y_1(t)I(t > V) = \alpha_{12}(t, t-T)I(V < t \leq X).$$

It is seen that this is only observable (for given α_{12}) if we not only follow the diseased individual from time $V (> T)$ until death at time X , but also actually know the time T of disease occurrence. The latter condition (technically: that a nontrivial part of \mathcal{G}_V is needed) is not always fulfilled and forms the basic problem in the modelling of latency times for epidemics such as AIDS.

Finally, in the particular case where $\alpha_{12}(t, d)$ does not depend on d (the original process is a Markov illness-death process), it is easily seen that both types of truncation (that is, following diseased individuals only, either from disease occurrence or from some random time, until death) are equivalent to studying a left truncated random variable with hazard given by the death intensity of the diseased. \square

So far we have only considered a (possibly multivariate) counting process corresponding to a single individual. We shall now briefly study left truncation of the processes corresponding to several independent individuals simultaneously. Let $N_i = (N_{hi}, h = 1, \dots, k)$, $i = 1, \dots, n$ be independent and

define $\mathbf{N} = (N_i, i = 1, \dots, n)$. Assume that \mathbf{N} has the same P_{θ_ϕ} -compensator with respect to (\mathcal{G}_t) and (\mathcal{F}_t) , where $(\mathcal{G}_t) \supseteq (\mathcal{F}_t)$ and $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{N}_t$.

If V is a (\mathcal{G}_t) -stopping time then we may apply the general result to the process

$${}_V\mathbf{N}(t) = \mathbf{N}(t) - \mathbf{N}(t \wedge V)$$

corresponding to truncation of all individuals at the same time. Alternatively we can let $(V_1, N_1), \dots, (V_n, N_n)$ be mutually independent. In this case

$${}_V N_i(t) = N_i(t) - N_i(t \wedge V_i), \quad i = 1, \dots, n$$

are independent and the theorem may be applied to each individual separately after which the relevant intensity process for the multivariate counting process

$$({}_V N_i, i = 1, \dots, n)$$

can be found using the product construction (Appendix).

More general cases with V_1, \dots, V_n being dependent seem to be more difficult to handle and the theory of left truncation as a whole seems to be less rich than the theory of right censoring.

Very frequently in practice there will be both left truncation and right censoring. Here we briefly indicate how the methods from this and the previous section can be combined. If in the construction of our model truncation precedes censoring then conditionally on an event before a stopping time V_i the individual process $N_{hi}(\cdot)$ started at V_i is observed on a set of the form $(V_i, U_i]$ with $V_i \geq 0$ and $U_i \geq V_i$. If censoring precedes truncation then the right censored process N_{hi}^c started at V_i is only observed conditionally on an event before V_i .

In either case, one needs a specification of the conditional joint distribution of V_i and U_i given $U_i \geq V_i$ for $i = 1, \dots, n$, possibly via a specification of their joint unconditional distribution.

5. GENERAL CENSORSHIP, FILTERING AND TRUNCATION

In Section 3 we studied the case where the observation of the individual counting processes $\mathbf{N}_i = (N_{1i}, \dots, N_{ki})$ was right censored, i.e. the component i was observed not on \mathcal{T} but only on a set of the form $E_i = [0, U_i]$. This is the most important example of incomplete observation but there may be other observational plans of interest where observation of \mathbf{N}_i is restricted to a subset $E_i \subseteq \mathcal{T}$.

Left censoring corresponds to a set $E_i = (V_i, \tau]$, $V_i \geq 0$; as an example we may recall the problem of recording the time of descent of baboons from the trees (WAGNER & ALTMANN, 1973). Troops of baboons in the Amboseli Reserve, Kenya, sleep in the trees and descend for foraging at some time of the day. Observers often arrive later in the day (say, at time V_i) than this descent and for such days they can only ascertain that descent took place *before* V_i , so that the descent times are left censored.

When defined in terms of random variables, left censoring is of course a concept symmetric to right censoring, and indeed WARE & DEMETS (1976) solved the baboon estimation problem by reversing time and using standard methods for right censored data. This trick however violates the basic role of the filtration in our framework, and as we shall see presently, in more complicated models left censoring presents special problems because of this.

Notice the difference between left censoring where \mathbf{N}_i is only observed on $(V_i, \tau]$ and left truncation (discussed in Section 4) where N_i , started at a stopping time V_i , is only observed on $(V_i, \tau]$ conditionally on an event prior to V_i . We shall return to a comparison of these two concepts later on in this section.

Censoring on intervals and *combination of left and right censoring* correspond to observing \mathbf{N}_i on a set of the form

$$E_i = \bigcup_{j=1}^r (V_{ji}, U_{ji}] \quad (5.1)$$

where

$$0 \leq V_{1i} \leq U_{1i} \leq \dots \leq V_{ri} \leq U_{ri} \leq \tau.$$

Thus, censoring corresponds to observing N on (possibly random) subintervals of \mathcal{T} only. Closely related to censoring is the concept of *filtering* where dN is observed on certain subintervals of \mathcal{T} only. To exemplify the difference between these two concepts we can consider a simple two-state Markov process model for a reversible disease (states *healthy* and *diseased*, transitions possible both ways, mortality disregarded). We can think of the case where occurrences of a recurrent disease are being studied in an individual on two intervals $[0, U_{1i}]$ and $(V_{2i}, U_{2i}]$ where $U_{1i} < V_{2i}$. If only new information is being collected during the two intervals we are observing the disease process *via a filter* and the number of disease occurrences in $(U_{1i}, V_{2i}]$ will not be known. If, however, at time V_{2i} this number can be observed (via hospital records, interviews or whatever) the observation of the disease process is *censored*. So, for a set E_i of the form (5.1) more information is available after a censored observation of the process than after observation via a filter, and only in the right censoring case $E_i = [0, U_i]$ do the two concepts coincide. As we shall see presently, however, there may be cases where one deliberately throws away some pieces of information about the censored process and analyses it as if it had been observed via a filter.

We shall now extend the method for handling right censoring in Section 3 to the more general plans of observation of N_i considered above. Corresponding to the set E_i we define a *censoring* or *filtering* process C by

$$C_i(t) = I(t \in E_i)$$

and the *filtered* counting process by

$$N_{hi}^c(t) = \int_0^t C_i(u) dN_{hi}(u).$$

We shall assume the existence of a filtration $(\mathcal{G}_t) \supseteq (\mathcal{F}_t)$ such that $N = (N_i, i = 1, \dots, n)$ has the same $P_{\theta\phi}$ -compensator Λ^θ with respect to both. We also assume that the set E_i is such that C_i is (\mathcal{G}_t) -predictable, i.e. that the U_j 's and V_j 's are (\mathcal{G}_t) -stopping times. Then the $P_{\theta\phi}$ -compensator for N_{hi}^c with respect to (\mathcal{G}_t) is

$$\Lambda_{hi}^c(t, \theta) = \int_0^t C_i(u) d\Lambda_{hi}(u, \theta).$$

As in the cases of right censoring we assume that the available data include X_0 and N^c . For a set E_i of the form (5.1) time points V_{ji}, U_{ji} before the time τ_i of absorption are also observed. If τ_i is not observed i.e. when τ_i belongs to some interval $(U_{ji}, V_{j+1,i}]$ then we may also observe the smallest V_{ji} such that $V_{ji} \geq \tau_i$. This is for instance the case when we have a left censored survival time X_i and observe V_i when $V_i > X_i$, see Example 5.2 below. In the case of censoring we also observe the values of N_{hi} at the observed *entry times* V_{ji} . In the case of filtering these values are not observed. We do, however, assume that the data at time t enable us to calculate $\Lambda^c(t, \theta)$ for any given value of θ . Viewing these observed data as a marked point process $N^* = (N^c, N'')$ we can calculate the likelihood $L_r^*(\theta, \phi)$ corresponding to the filtration (\mathcal{F}_t) generated by N^* . Letting jumps in N^c carry the innovative marks of N^* the corresponding partial likelihood

$$L_r^c(\theta) = \prod_{t \in \mathcal{T}} (1 - d\Lambda^c(t, \theta))^{1 - dN^c(t)} \prod_{h,i} d\Lambda_{hi}^c(t, \theta)^{dN_{hi}^c(t)}$$

is identical to (3.7) and thus it has the same form as (3.5). We therefore term the predictable process C an *independent filter*. We shall also denote it an *Aalen filter*.

When N_{hi} satisfies Aalen's multiplicative intensity model

$$\lambda_{hi}(t, \theta) = \alpha_{hi}^\theta(t) Y_{hi}(t)$$

with respect to (\mathcal{F}_t) it follows that $N_{hi}^c(t)$ follows the multiplicative intensity model

$$\lambda_{hi}^c(t, \theta) = \alpha_{hi}^c(t) Y_{hi}^c(t)$$

with respect to (\mathcal{F}_t^c) . Here, $Y_{hi}^c(t) = Y_{hi}(t)C_i(t)$.

EXAMPLE 5.1. *Mau's concept of partitioned counting processes* (MAU, 1985). Assume $k=1$, let, for $i = 1, \dots, n$ and some m , $0 \leq S_{i0} \leq S_{i1} \leq \dots \leq S_{im} \leq S_{i,m+1} = \tau$ be (\mathcal{G}_t) -stopping times and define stochastic processes

$$C_i^{(j)}(t) = I(S_{i,j-1} < t \leq S_{ij}), \quad i = 1, \dots, n; \quad j = 1, \dots, m+1.$$

Then to each component N_i there exists an $(m+1)$ -variate *partitioned counting process* given by

$$N_{ij}(t) = \begin{cases} 0, & 0 \leq t \leq S_{ij} \\ N_i(t) - N_i(S_{ij}), & S_{ij} < t \leq S_{i,j+1} \\ N_i(S_{i,j+1}) - N_i(S_{ij}), & S_{i,j+1} < t \leq \tau, \end{cases}$$

$i = 1, \dots, k$, $j = 1, \dots, m+1$. Obviously $(N_{ij}(t))$ counts the events in the random interval $(S_{ij}, S_{i,j+1}]$. MAU (1985) noted that the partitioned counting process still satisfies the multiplicative intensity model, now with intensity processes

$$\alpha_{ij}^c(t) Y_{ij}(t) = \alpha_i^c(t) C_i^{(j)}(t) Y_i(t).$$

In our terminology $\mathbf{C} = (C_i^{(j)})$ is an Aalen filter, so that the analysis of what happens to the counting process in particular random intervals may be performed using the powerful tools of the multiplicative intensity model. MAU (1987) showed how this allows monitoring of clinical trials, e.g. by separately analysing the information from several calendar time intervals in a trial with staggered entry, cf. also KEIDING, BAYER & WATT-BOOLSEN (1987). \square

When inference is based on $L_r^c(\theta)$ alone using the Aalen filter, the non-innovative marks of \mathbf{N}^* are disregarded. These marks will at time t contain information on certain individuals either leaving the risk set or entering the risk set at that time and they may also be defined to carry information on occurrences of earlier events the exact times of which are not observed. In that case the marks will typically carry information on θ , the parameter of interest, and we then term the censoring mechanism (or the filter) \mathbf{C} *informative* for θ . If \mathbf{C} is non-informative for θ , $L_r^c(\theta)$ is the full likelihood (or at least the full conditional likelihood given \mathbf{X}_0) and no information is lost by basing the statistical inference on it. In the example mentioned above concerning a disease process observed on the set $E_i = [0, U_{1i}] \cup (V_{2i}, U_{2i}]$ the mark at V_{2i} may thus contain information of disease occurrences in $(U_{1i}, V_{2i}]$ if this piece of information is available. In this case some information is lost by only considering the process counting the number of disease occurrences filtered via the process \mathbf{C} with components $C_i = I(t \in E_i)$. Thus \mathbf{C} is informative (for the parameters of the disease intensity) and it would be more efficient to base inference on the entire likelihood $L_r^*(\theta, \phi)$ than on the partial likelihood $L_r^c(\theta)$. On the other hand, the entire likelihood may depend on the nuisance parameter ϕ which is often inconvenient. In fact, one may not even be prepared to write down a full statistical model for \mathbf{N}, \mathbf{C} .

To return to a comparison of *left truncation*, *left censoring* and *left filtering* note first the technical difference that the latter two keep the original sample space and probability measure whereas left truncation is a conditional procedure, restricted to a subset of the sample space and the corresponding conditional probability. Some further aspects are best considered in the simplest possible example.

EXAMPLE 5.2. *Random left censoring of a non-negative random variable*. We consider once more the set-up from Examples 3.4 and 4.1: X and V are independent non-negative random variables, X has hazard function $\alpha_X(\cdot, \theta)$ and V has distribution function $F_V(\cdot, \phi)$ (and hazard function $\alpha_V(\cdot, \phi)$ if it exists). Furthermore complete observation of X and V is considered as observation of a marked point process as described in those examples. Right censoring at V corresponds to only being able to

observe $X \wedge V$ and the mark at that time, i.e. the counting process $N_X(t) = I(X \leq t)$ is only observed on the random interval $E = [0, V]$. Similarly left censoring corresponds to the case where $N_X(t)$ is only observed on a set $E = (V, \tau]$ and we shall assume that V is always observed. Thus we observe the *filtered* counting process

$$N^c(t) = \int_0^t C(u) dN(u) = (N(t) - N(V))I(t > V),$$

its \mathcal{G}_t -compensator (except for the value of the unknown θ)

$$\Lambda^c(t, \theta) = \int_0^t C(u) d\Lambda(u, \theta) = \int_0^t \alpha(u, \theta) Y^c(u) du,$$

(where $Y^c(t) = C(t)Y(t) = I(V < t \leq X)$), and the value $N(V) = I(X \leq V)$. That is, if $X > V$ then we observe *both* V and X , and if $X \leq V$ then we observe V and know that $X \leq V$. The partial likelihood $L_r^c(\theta)$ is in this case, according to (3.7)

$$\begin{aligned} L_r^c(\theta) &= \prod_{V < t < X} (1 - \alpha_X(t, \theta) dt) \alpha_X(X, \theta)^{I(X > V)} \\ &= \left[\frac{S_X(X, \theta)}{S_X(V, \theta)} \alpha_X(X, \theta) \right]^{I(X > V)}. \end{aligned} \quad (5.2)$$

The second factor of the full likelihood (3.6) is

$$L_r''(\theta, \phi) = S_V(V, \phi) \alpha_V(V, \phi) S_X(V, \theta)^{I(V < X)} F_X(V, \theta)^{I(V \geq X)},$$

and it *does* depend on θ . So, obviously observation of $N(V)$ gives us some information on θ meaning that $C(\cdot)$ is informative for θ and that inference based on $L_r^c(\theta)$ only will not be fully efficient.

The fact that V is always observed is in contrast to the case of right censoring (see, however, Example 3.2). A more direct parallel definition of left censoring would be to assume that $X \vee V$ were observed together with the mark at that time. That situation gives a different likelihood but it can be handled in a similar way. \square

The way in which left censored data are analysed using the Aalen filter is by treating the counting process as being observed with delayed entry or as being *left filtered*. The partial likelihood (5.2) for the left filtered process N^c is identical to (4.3) in the sense that the data used in the two situations are the same and that the parameter θ enters into the two likelihoods in the same way. Formally, however, the likelihood (4.3) is with respect to a conditional distribution $P_{\theta\phi}^A$ whereas (5.2) is with respect to the original probability measure $P_{\theta\phi}$. These results show that *left truncated* counting processes can be correctly analysed as counting processes observed with *delayed entry*. If individuals $i = 1, \dots, n$ are observed then individual i is included in the relevant risk set from the time V_i . Analysing left censored data as data with delayed entry is, however, not fully efficient.

For survival data the basic difference between a left censored and a left truncated survival time X_i is that in the latter case individual i is only included in the sample conditionally on its survival time exceeding the entry time V_i whereas in the former case individual i is always included in the sample but observation of the exact failure time may be prevented for some reason. So for n independent left truncated observations the partial likelihood will be a product of a *fixed number* (n) of factors of the form (5.2) whereas for n independent left censored observations the partial likelihood is a product of a *random number* ($\leq n$) of factors of this form.

Also the situation with both left truncation and right censoring can be handled using the Aalen filter. In this case we can define the filtering process by

$$C_i(t) = I(V_i < t \leq U_i)$$

and base the inference on θ on the (partial) likelihood $L_r^c(\theta)$ with the form (3.7). In many realistic

models for the distribution of N , V and U the filtering process $C = (C_1, \dots, C_n)$ will be non-informative about θ and this analysis will be efficient. It should, however, be emphasized once more that for other incomplete plans of observation a more efficient analysis can be carried out using the entire likelihood $L_r^*(\theta, \phi)$.

Right truncation and more general types of truncation may be defined similarly to left truncation, but none of these are conveniently dealt with in the present framework. As for left versus right censoring, the explanation is that the time direction given by the filtration destroys the symmetry between left and right, except for some simple cases where one may study right truncation by reversal of time. See KEIDING (1986) for applications to the disease intensity ('incidence') in the illness-death model under special epidemiological sampling plans.

The literature on general censoring and truncation is rather unsophisticated as regards concepts of independence and non-informativity of the censoring (and truncation) patterns, in effect assuming these to be deterministic. However aspects of the iterative methods (versions of the *EM* algorithm) necessary to study the full likelihood and primarily developed by TURNBULL (1974, 1976) and DEMPSTER, LAIRD & RUBIN (1977, Section 4.2) are instructive in the general modelling framework of this paper. While censoring is readily interpreted as being an example of incompletely observed data, it is at first sight more surprising that truncation may also be interpreted in this way. The idea is to consider among the unobserved data also the number of individuals who were never observed, because their values are outside the relevant truncation set. TURNBULL (1976) termed these the 'ghosts' and DEMPSTER, LAIRD & RUBIN (1977) gave a comprehensive discussion.

Later authors have primarily been concerned with the (difficult) task of proving asymptotic properties of estimators derived this way. An interesting modelling contribution was made by SAMUELSEN (1988), who suggested a stochastic process model for double censoring, generalizing the competing risk framework for random right censoring (Example 3.5) and the Markov process model for random left truncation (Example 4.2).

6. INTERMITTENT OBSERVATION OF A COUNTING PROCESS: MODELS FOR GROUPED DATA

An extreme example of incomplete observation formally covered by the concept of interval censoring is observation of a *discrete skeleton* of the process, that is, the h 'th component N_h is observed at times $0 = \tau_0^{(h)} < \tau_1^{(h)} < \dots < \tau_r^{(h)} \leq \tau$. Because one will then (with probability 1) never observe the exact time of a transition, the Aalen filter will reduce observation to nothing: the partial (filtered) likelihood $L_r^c(\theta) = 1$. This is for instance the case in connection with various kinds of grouped data from a Markov process as exemplified in this section.

Let the Markov process X with state space S , the counting process N and the filtration (\mathcal{F}_t) be defined as in Example 2.4. The statistical model is given by assuming some transitions impossible and the rest of the transition intensities (specified by the set $R \subseteq \{(h, j) : h, j \in S, h \neq j\}$) arbitrarily varying. Intermittent observation of the counting process $N = (N_{hj}, (h, j) \in R)$ is observation of $(N_{hj}(\tau_1^{(hj)}), \dots, N_{hj}(\tau_r^{(hj)}))$, where the τ 's are assumed to be deterministic times unless otherwise specified. Note that $N_{hj}(\tau_{i+1}^{(hj)}) - N_{hj}(\tau_i^{(hj)})$ counts the number of transitions $h \rightarrow j$ in the time interval $(\tau_i^{(hj)}, \tau_{i+1}^{(hj)}]$, and therefore intermittent observation of the transition counts corresponds to *grouped observation* of the transition times. As indicated above, filtering removes all information, in the sense that (with probability one) we have $N^c = 0$ and $L_r^c(\theta) = 1$. So, in this case the partial likelihood contains no information on θ . In connection with other kinds of grouped data even less information may be available. Sometimes, only the state occupied at τ_0, \dots, τ_r is observed, i.e. $(X(\tau_j), j = 0, 1, \dots, r)$ or equivalently $(Y(\tau_j +), j = 0, 1, \dots, r)$. Also in this case $L_r^c(\theta) = 1$ and one has to consider the full likelihood. KALBFLEISCH & LAWLESS (1985) studied maximum likelihood estimation in the model with constant transition intensities based on *panel data*, i.e. observation of independent Markov processes $X_i(\cdot)$, $i = 1, \dots, n$ at time points $\tau_0, \tau_1, \dots, \tau_r$.

There may even be cases where the individual panel data are not available but only the number of individuals in each state. In this case the *aggregated* data are $(Y_h(\tau_j); j = 0, 1, \dots, r, h = 1, \dots, k)$.

EXAMPLE 6.1. *The illness-death process.* This example was discussed earlier in Examples 2.8, 3.6 and 4.4. While the time of death can usually be observed exactly, it is often difficult to assess exactly when disease occurred.

One group of problems of this kind is *long-term animal carcinogenicity trials* where it is usually assumed that it can always be assessed after death whether disease had occurred or not. Often supplementary data are obtained by *serial sacrifice*, that is, animals are killed at pre-specified times and it is assessed whether or not they already had the tumour. Most of the literature on designs of such trials (see e.g. BORGAN et al, 1984) studies deterministic observational plans as opposed to plans determined adaptively by the development of the process. An exception is MAU (1986) who formulated an explicitly random 'associated design process'. The likelihood function usually becomes complicated and sometimes only certain functionals of the process are identifiable. Recent reviews well in tone with the approach taken here are by MCKNIGHT & CROWLEY (1984), MCKNIGHT (1985) and DEWANJI & KALBFLEISCH (1986); see also the monograph by GART et al. (1986).

A somewhat different application of the simple illness-death process is to non-reversible complications of chronic diseases such as diabetes (ANDERSEN, 1988) or cancer. Here patients are examined at visits to the hospital and the determination of whether a transition $0 \rightarrow 1$ (onset of disease complication) has happened may only be performed at those times. It is here very important (though often overlooked in practice) to know whether the visits to the hospital are planned independently of the underlying disease process (as would be true for deterministic observational plans) or whether they may be triggered by the disease. Motivated by these problems GRÜGER (1986) developed an interesting theory of *noninformative observational plans* for counting processes much along the lines of the present paper. \square

7. PARTIAL MODEL SPECIFICATION. TIME-DEPENDENT COVARIATES

The essence of the concept of independent censoring or filtering is that under such a scheme of observation it is possible to write down a partial likelihood for θ , the parameter of interest, which has the same form as the likelihood for the full data and which does not depend on the nuisance parameter ϕ . Thus, the partial likelihood can be computed without actually specifying a model for the censoring mechanism, in fact as if censoring had been at fixed given times. Another example of a partially specified model is the Cox regression model (Example 2.6) where inference could be performed conditionally on the covariates and without specifying a model for their distribution.

In these examples the covariates were *time-independent*, i.e. they were fixed given \mathcal{F}_0 , but in several examples it is also of interest to study for instance a death intensity conditionally on covariates which change in time. Some such time-dependent covariates may be deterministic or at least fixed given \mathcal{F}_0 ; in the example mentioned at the beginning of Section 4 concerning survival among insulin dependent diabetics in Fyn county time was taken to be the age of the patients whereas the age at diagnosis was included as a time-independent covariate. The death intensity may also depend on the time-dependent covariate 'disease duration' which can be computed for each age t knowing the age at diagnosis. Thus the stochastic process $Z_i(t)$ = 'disease duration for patient i at age t ' is adapted to the filtration generated by the data.

In such a case, where the intensity depends on what KALBFLEISCH & PRENTICE (1980, p.123) termed a *defined time-dependent covariate*, the (partial) likelihood stays the same and inference based on the likelihood can be performed as if the covariate paths had been fixed in advance.

In other examples there may be time-dependent covariates which are *truly random* in the sense that the processes $Z_i(\cdot)$, $i = 1, \dots, n$ are *not* automatically adapted to the filtration under consideration. KALBFLEISCH & PRENTICE (1980, Section 5.3) distinguished between *ancillary covariates* and *internal covariates* (much corresponding to exogenous and endogenous variables in econometrics, cf. HENDRY & RICHARD, 1983). An ancillary covariate could for instance be the level of air pollution in a study of the occurrence of asthma attacks, while an important class of internal covariates are 'disease complications' developing in a fashion unpredictable from the history of the process itself.

In order to include such covariates in the model we must extend the filtration. One way of doing that is to consider the whole system of uncensored observations as developing according to a (very large) marked point process N_Z recording, with innovative marks, failures (and other transitions or events of interest) and, with non-innovative marks, changes of covariate values. To consider everything as a point process does pose some restrictions on the types of covariates considered in that (random) changes of covariate values have to be generated by an underlying process changing at discrete (possibly random) points in time and not continuously. So, if a continuously observed time-dependent covariate, which is not adapted, is to be included in the model then its path has to be discretised in some way; for instance by defining its changes of values to happen at discrete points in time or at least to let its path vary deterministically except at a discrete set of points.

As in the previous sections one may superimpose censoring or filtering onto the marked point process N_Z via a process C which is predictable with respect to a filtration (\mathcal{G}_t) larger than that generated by N_Z . In this way a *censored* or *filtered* marked point process, say N_Z^* , is obtained, and we may write

$$N_Z^* = (N^c, N_Z'')$$

Here N^c , as before, counts the *observed* transitions of interest and carries the innovative marks, whereas N_Z'' carries the non-innovative marks including information on individuals entering or leaving the risk sets and information on observed changes in covariate values, ARJAS & HAARA (1984). We assume that observation of N_Z^* enables us to calculate for each value of θ the $(P_{\theta, \phi}, (\mathcal{G}_t))$ -compensator for N^c . We can then calculate the full likelihood for N_Z^* with respect to the filtration (\mathcal{F}_t) generated by itself and factorise it into a partial likelihood $L_r^c(\theta)$ not depending on the nuisance parameter ϕ and a second factor $L_r''(\theta, \phi)$ which may or may not depend on θ . This means that inference on θ can be based on $L_r^c(\theta)$ only and it can be made without specifying the model for the censoring mechanism and the covariate processes. However, as before, a more efficient inference on θ may be obtained from the full likelihood if the second factor does in fact depend on θ , i.e. if censoring or covariates are informative.

The partially specified model specifying only the (\mathcal{F}_t) -compensator Λ_θ^c for N^c has, however, got some limitations due to the fact that only a small part of a big system is modelled. If one wants to make predictions then this is not directly possible if the model for N^c includes time-dependent covariates whose development in time is *not* modelled (ANDERSEN, 1986). So, if prediction making is an important issue of a study one has to either disregard time-dependent covariates or to model them. The latter possibility corresponds to labelling the marks for changes in these covariates innovative and to include the parameters for them in θ rather than in ϕ . It should be emphasized that the labelling of marks as innovative or non-innovative is up to the statistician and it depends on the purposes of the study.

An example of this problem was seen previously in that one may sometimes be interested in studying several cause specific hazard functions in a competing risks model (Example 2.5) and sometimes only deaths due to one cause are of interest while deaths due to other causes are treated as censorings and the corresponding cause specific hazard functions as nuisance parameters (Example 3.6). Another problem with censored or filtered observation of time-dependent covariates is that values of these covariates may not be observed, which may prevent one from computing even the partial likelihood $L_r^c(\theta)$. As discussed in Example 4.4 this might be the case in the illness-death model with duration dependence introduced in Example 2.8. Suppose that the $1 \rightarrow 2$ transition intensity is modelled as $\alpha_{12}(t, t-T) = \alpha_0(t) \exp(\beta(t-T))$ using the time-dependent covariate $z(t) = t-T =$ 'sojourn time in state 1 at time t ', and suppose that at the entry time $V > T$, the value of T is unknown. Then the value of $z(t)$ is unobservable.

8. APPENDIX

8.1. The compensator of a product of independent counting processes

The product construction, without loss of generality for $k=2$ components, proceeds as follows. From $(\Omega^{(i)}, \mathcal{F}^{(i)}, (\mathcal{F}_t^{(i)}, t \in \mathcal{T}), P^i), i = 1, 2$, we define

$$\Omega = \Omega^{(1)} \times \Omega^{(2)}, \mathcal{F} = \mathcal{F}^{(1)} \otimes \mathcal{F}^{(2)},$$

$$\mathcal{F}_t = \mathcal{F}_t^{(1)} \otimes \mathcal{F}_t^{(2)}, P = P^{(1)} \otimes P^{(2)};$$

thus P is the product probability measure on \mathcal{F} . If the counting processes $N^{(i)}(t)$ have compensators $\Lambda^{(i)}(t)$, with respect to P^i and $(\mathcal{F}_t^{(i)}), i = 1, 2$, then it is easy to check that the $\Lambda_h^{(i)}(t)$ are also predictable viewed as defined on Ω , with respect to P and (\mathcal{F}_t) , and that the $N_h^{(i)}(t) - \Lambda_h^{(i)}(t)$ are (local) martingales. However \mathcal{F}_t is not necessarily right-continuous. A sufficient condition for this is that $\mathcal{F}_t^{(i)} = \mathcal{F}_0^{(i)} \vee \sigma\{N^{(i)}(s): s \leq t\}$ in which case

$$\mathcal{F}_t = \mathcal{F}_0^{(1)} \otimes \mathcal{F}_0^{(2)} \vee \sigma\{N^{(1)}(s), N^{(2)}(s): s \leq t\}$$

which is obviously right-continuous. The filtrations also have to be completed, which creates no further problems.

Instead of this combination of independent component process we will occasionally need to combine *conditionally independent* components. The situation now is that there is one probability space (Ω, \mathcal{F}, P) on which $N(t) = (N^{(1)}(t), N^{(2)}(t))$ is defined; we consider two filtrations $(\mathcal{F}_t^{(1)})$ and $(\mathcal{F}_t^{(2)})$ and assume that they are *conditionally independent* given some σ -algebra $\mathcal{C} \subseteq \mathcal{F}$, that is, if $A \in \mathcal{F}_t^{(1)}, B \in \mathcal{F}_t^{(2)}, C \in \mathcal{C}, P(C) > 0$, then

$$P(A \cap B | C) = P(A | C)P(B | C).$$

(Often $\mathcal{F}_t^{(i)} = \sigma\{N^{(i)}(s): s \leq t\}$). Define

$$\mathcal{F}_t = \mathcal{C} \vee \mathcal{F}_t^{(1)} \vee \mathcal{F}_t^{(2)}.$$

One may then check directly that any $(\mathcal{C} \vee \mathcal{F}_t^{(1)})$ -martingale is also an (\mathcal{F}_t) -martingale, which is the key step in verifying that the (\mathcal{F}_t) -compensator of $N(t)$ may be obtained by combining the $(\mathcal{C} \vee \mathcal{F}_t^{(i)})$ -compensators of $N^{(i)}(t)$.

8.2. Proof of Proposition 4.1.

The basic point is to show that if M is a $(P, (\mathcal{G}_t))$ -martingale and $A \in \mathcal{G}_V$ has $P(A) > 0$, then

$${}_V M(t) = M(t) - M(t \wedge V)$$

is a $(P^A, ({}_V \mathcal{G}_t))$ -martingale. This may be seen as follows (M. JACOBSEN, personal communication, September 1986).

Clearly ${}_V M$ is adapted to $({}_V \mathcal{G}_t)$. We have to verify that

$$E[I(A \cap F)\{{}_V M(u) - {}_V M(t)\}] = 0 \tag{*}$$

for all $F \in {}_V \mathcal{G}_t, t < u$. Since $\{V > t\} \in {}_V \mathcal{G}_t$, it suffices to consider the two special cases $F \subseteq \{V > t\}$ and $F \subseteq \{V \leq t\}$.

If $F \subseteq \{V \leq t\}$, we have

$$A \cap F = (A \cap \{V \leq t\}) \cap (F \cap \{V \leq t\}) \in \mathcal{G}_t$$

(definition of \mathcal{G}_V resp. of ${}_V \mathcal{G}_t$) and since on $\{V \leq t\}$ we have

$${}_V M(u) - {}_V M(t) = M(u) - M(t),$$

(*) follows from the martingale property of M :

$$E[I(A \cap F)\{M(u) - M(t)\}] = 0 \text{ for } A \cap F \in \mathcal{G}_t.$$

If $F \subseteq \{V > t\}$, define $B = \{V \leq u\}$. The left hand side of (*) may be written as

$$E[I(A \cap B \cap F)\{M(u) - M(u \wedge V)\}].$$

Now since $F \in \mathcal{G}_{V \vee V}$ and $F \subseteq \{V > t\}$, $F \in \mathcal{G}_V$, and hence $A \cap B \cap F \in \mathcal{G}_V$. Furthermore $A \cap F \in \mathcal{G}_V$ implies

$$A \cap B \cap F = (A \cap F) \cap \{V \leq u\} \in \mathcal{G}_u,$$

hence $A \cap B \cap F \in \mathcal{G}_{u \wedge V} = \mathcal{G}_u \cap \mathcal{G}_V$ and (*) follows by optional stopping. \square

REFERENCES

- O.O. AALEN, (1975). *Statistical inference for a family of counting processes*. Ph.D. Dissertation, Department of Statistics, University of California, Berkeley.
- O.O. AALEN, (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.*, **6**, pp. 701-726.
- O.O. AALEN & S. JOHANSEN, (1978). An empirical transition matrix for non-homogeneous Markov chains based on censored observations. *Scand. J. Statist.*, **5**, pp. 141-150.
- P.K. ANDERSEN, (1986). Time-dependent covariates and Markov processes. In: *Modern Statistical Methods in Chronic Disease Epidemiology* (S.H. Moolgavkar & R.L. Prentice, eds.) Wiley, pp. 82-103.
- P.K. ANDERSEN, (1988). Multi-state models in survival analysis: A study of nephropathy and mortality in diabetes. *Statist. in Medicine*, (to appear).
- P.K. ANDERSEN, K. BORCH-JOHNSEN, T. DECKERT, A. GREEN, P. HOUGAARD, N. KEIDING & S. KREINER, (1985). A Cox regression model for the relative mortality and its application to diabetes mellitus survival data. *Biometrics*, **41**, pp. 921-932.
- P.K. ANDERSEN & Ø. BORGAN, (1985). Counting process models for life history data: A review (with discussion). *Scand. J. Statist.*, **12**, pp. 97-158.
- P.K. ANDERSEN, Ø. BORGAN, R.D. GILL & N. KEIDING, (1982). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data (with discussion). *Int. Statist. Rev.*, **50**, pp. 219-258. Correction **52**, p. 225.
- E. ARJAS, (1985). Discussion of P.K. Andersen & Ø. Borgan: Counting process models for life history data: A review. *Scand. J. Statist.*, **12**, pp. 150-153.
- E. ARJAS & P. HAARA, (1984). A marked point process approach to censored failure data with complicated covariates. *Scand. J. Statist.*, **11**, pp. 193-209.
- Ø. BORGAN, K. LIESTØL & P. EBBESEN, (1984). Efficiencies of experimental designs for an illness-death model. *Biometrics*, **40**, pp. 627-638.
- D.R. COX, (1959). The analysis of exponentially distributed lifetimes with two types of failure. *J.R. Statist. Soc (B)*, **21**, pp. 411-21.
- D.R. COX, (1972). Regression models and life tables (with discussion). *J.R. Statist. Soc (B)*, **34**, pp. 187-220.
- D.R. COX, (1975). Partial likelihood. *Biometrika*, **62**, pp. 269-76.
- D.R. COX & D. OAKES, (1984). *Analysis of survival data*. Chapman & Hall, London.
- A.P. DEMPSTER, N.M. LAIRD & D.B. RUBIN, (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Roy. Statist. Soc. (B)*, **39**, pp. 1-38.
- A. DEWANJI & J.D. KALBFLEISCH, (1986). Nonparametric methods for survival/sacrifice experiments. *Biometrics*, **42**, pp. 325-341.
- J.J. GART, D. KREWSKI, P.N. LEE, R.E. TARONE & J. WAHRENDORF, (1986). *The design and analysis of long-term animal experiments*. Lyon: International Agency for Research on Cancer.
- R.D. GILL, (1980a). *Censoring and stochastic integrals*. Math. Centre Tracts, **124**, Mathematical Centre, Amsterdam.
- R.D. GILL, (1980b). Nonparametric estimation based on censored observations of a Markov renewal

- process. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, **53**, pp. 97-116.
- R.D. GILL, (1981). Testing with replacement and the product-limit estimator. *Ann. Statist.*, **9**, pp. 853-860.
- R.D. GILL & S. JOHANSEN, (1987). *Product-integrals and counting process*. Report MS-R8709, Centre for Mathematics and Computer Science, Amsterdam.
- A. GREEN, M. HAUGE, N.V. HOLM & L.L. RASCH, (1981). Epidemiological studies of diabetes mellitus in Denmark: II. A prevalence study based on insulin prescriptions. *Diabetologia*, **20**, pp. 468-470.
- A. GREEN & P. HOUGAARD, (1984). Epidemiological studies of diabetes mellitus in Denmark: 5. Mortality and causes of death among insulin-treated diabetic patients. *Diabetologia*, **26**, pp. 190-194.
- J. GRÜGER, (1986). *Nichtparametrische Analyse sporadisch beobachtbarer Krankheitsverlaufsdaten*. Dissertation, Universität Dortmund.
- A. HALD, (1949). Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point. *Skand. Aktuarietidskr.*, **32**, pp. 119-134.
- A. HALD, (1952). *Statistical theory with engineering applications*. Wiley, New York.
- D.F. HENDRY & J.F. RICHARD, (1983). The econometric analysis of econometric time series. *Int. Statist. Rev.*, **51**, pp. 111-165.
- J.M. HOEM, (1969). Purged and partial Markov chains. *Skand. Aktuarietidskr.*, **52**, pp. 147-155.
- M. JACOBSEN, (1972). A characterization of minimal Markov jump processes. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, **23**, pp. 32-46.
- M. JACOBSEN, (1982). *Statistical analysis of counting processes*. Lecture Notes in Statistics, **12**, Springer-Verlag, Berlin.
- M. JACOBSEN, (1986). Right censoring and the Kaplan-Meier and Nelson-Aalen estimators. Preprint no. 6, Institute of Mathematical Statistics, University of Copenhagen (to appear in *Ann. Statist.*).
- J. JACOD, (1975). Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, **31**, pp. 235-253.
- S. JOHANSEN, (1987). Product integrals and Markov processes. *CWI Newsletter*, **12**, pp. 3-13; originally appeared (1977) as: Preprint no. 3, Institute of Mathematical Statistics, Univ. of Copenhagen.
- J.D. KALBFLEISCH & J.F. LAWLESS, (1985). The analysis of panel data under a Markov assumption. *J. Amer. Stat. Assoc.*, **80**, pp. 863-871.
- J.D. KALBFLEISCH & R.J. MACKAY, (1979). On constant-sum models for censored survival data. *Biometrika*, **66**, pp. 87-90.
- J.D. KALBFLEISCH & R.L. PRENTICE, (1980). *The statistical analysis of failure time data*. Wiley, New York.
- E.L. KAPLAN & P. MEIER, (1958). Non-parametric estimation from incomplete observations. *J. Amer. Statist. Assoc.*, **53**, pp. 457-481.
- N. KEIDING, (1986). *The epidemiological information in a cross-sectional sample: A statistical perspective*. Research Report 86/3, Statistical Research Unit, University of Copenhagen.
- N. KEIDING, T. BAYER & S. WATT-BOOLSEN, (1987). Confirmatory analysis of survival data using left truncation of the life times of primary survivors. *Statist. in Medicine* (to appear).
- N. KEIDING & R.D. GILL, (1987). *Random truncation models and Markov processes*. Research Report 87/3, Statistical Research Unit, University of Copenhagen. Also issued as Report MS-R8702, Centre for Mathematics and Computer Science, Amsterdam.
- J.A. KOZIOL & S.B. GREEN, (1976). A Cramer-von Mises statistic for randomly censored data, *Biometrika*, **63**, pp. 465-474.
- J. MAU, (1985). Statistical modelling via partitioned counting processes. *J. Statist. Plann. Inf.*, **12**, pp. 171-176.
- J. MAU, (1986). Nonparametric estimation of the integrated intensity of an unobservable transition in a Markov illness-death process. *Stoch. Proc. Appl.*, **21**, pp. 275-289.
- J. MAU, (1987). *Monitoring of therapeutical studies via partitioned counting processes*. Research Report no. 1/87, SFB 175 Implantology, University of Tübingen.
- B. MCKNIGHT, (1985). Discussion of session on statistical tests for carcinogenic effects. In:

- Proceedings of the symposium on long-term animal carcinogenicity studies: A statistical perspective.* Amer. Statist. Assoc. Washington D.C. pp. 107-111.
- B. MCKNIGHT & J. CROWLEY, (1984). Tests for differences in tumor incidence based on animal carcinogenesis experiments. *J. Amer. Statist. Assoc.*, **79**, pp. 639-648.
- S.O. SAMUELSEN, (1988). Nonparametric estimation of the cumulative intensity from doubly censored data. Asymptotic theory. *Scand. J. Statist.*, (to appear).
- T. SELKE & D. SIEGMUND, (1983). Sequential analysis of the proportional hazards model. *Biometrika* **70**, pp. 315-326.
- E.V. SLUD, (1984). Sequential linear rank tests for two-sample censored survival data. *Ann. Statist.*, **12**, pp. 551-571.
- A. TSIATIS, (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Natl. Acad. Sci.*, **72**, pp. 20-22.
- B.W. TURNBULL, (1974). Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.*, **69**, pp. 169-173.
- B.W. TURNBULL, (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *J. Roy. Statist. Soc., (B)*, **38**, pp. 290-295.
- S.S. WAGNER & S.A. ALTMANN, (1973). What time do the baboons come down from the trees? (An estimation problem). *Biometrics*, **29**, pp. 623-635.
- M.-C. WANG, N.P. JEWELL & W.-Y. TSAI, (1986). Asymptotic properties of the product limit estimate under random truncation. *Ann. Statist.*, **14**, pp. 1597-1605.
- J.H. WARE & D.L. DEMETS, (1976). Reanalysis of some baboon descent data. *Biometrics*, **32**, pp. 459-463.
- J.S. WILLIAMS & S.W. LAGAKOS, (1977). Models for censored survival analysis: Constant sum and variable sum models. *Biometrika*, **64**, pp. 215-224.
- M. WOODROOFE, (1985). Estimating a distribution function with truncated data. *Ann. Statist.*, **13**, pp. 163-177. Correction Note *ibid.* **15**, pp. 883, (1987).