Scand J Statist 11: 193–209, 1984

A Marked Point Process Approach to Censored Failure Data with Complicated Covariates

ELJA ARJAS and PENTTI HAARA
University of Oulu

ABSTRACT. Complicated failure time data which can involve, e.g., random covariates, censored observations and multiple failures, is here considered as a sample path of a marked point process (MPP). Our main task is to derive likelihood expressions for parametric statistical models under such general circumstances. To do this, and motivated by concrete examples, we split each marked point into two characteristic parts, called innovation and non-innovation, and then characterize this representation in terms of the statistical model. Technically the paper is based on the martingale approach to point processes.

Key words: failure time, likelihood, covariate, censoring, innovation, point process, martingale

1. Introduction

In the simplest setting, failure time data is thought of as a random sample from an unknown distribution. This simple model applies only rarely in practice, however: for example, often some of the observations are censored, the failures can be due to many different causes, etc.

A natural possibility to model such complicated data is to base the model on a random process describing the time evolution actually observed. Point processes (also called counting processes) provide a particularly convenient framework for this. Their use has made possible the application of strong results from modern probability theory, notably martingale methods. The works of Aalen (1978), Gill (1980), Andersen, Borgan, Gill & Keiding (1982), Johansen (1983) and Jacobsen (1982), to mention a few, are examples of the stochastic process approach, skillfully built into statistical theory.

Our goal here is twofold. Firstly, we want to stress the immense flexibility of a particular class of stochastic process models in failure time analysis, viz. the marked point processes. It appears that this class of models is capable of accommodating almost all conceivably occurring variations of the failure problem. We do not suggest any new techniques for statistical inference in this paper. In fact, all we do is that we derive likelihood expressions under very general assumptions. Therefore, this paper should be seen more as a contribution to the concepts and thinking than to the statistical practice.

Secondly, while deriving the likelihood expressions, it becomes apparent that information contained in the "marked points" can be usefully divided into two categories. Representative to the first category, *the innovations*, would be the actual failures as they, given the past as a condition, instantaneously provide information about the model parameters. Censored observations, on the other hand, would typically be classified into the second category of non-innovative points. As we shall see, to this latter category there corresponds a factor in the likelihood function which does not depend on the parameter, and which therefore can be deleted when inferring about the parameter. This phenomenon is well known from simple censoring models, and it is also related to the role of exogenous variables in econometric models (see Hendry & Richard (1983)).

The plan of this paper is as follows. In Section 2 we introduce our basic model, the marked point process (MPP), and explain the particular structure of the chosen mark space. We also provide motivation to the definitions and terminology, and consider examples. In Section 3, having explained the reasoning in a special case in intuitive terms, we state our condition for non-innovation. It is most conveniently expressed in terms of the compensator measure associated with the basic marked point process. The relationship between the condition and *non-informative censoring* is also discussed, and three particular mechanisms for censoring are considered as examples. In Section 4 we derive general expressions for the likelihood ratio and likelihood function, basing the presentation on results of Jacod (1975). We also work out a number of special cases. In the final section we consider the practical specification of the statistical model governing the observed MPP, also discussing questions related to interpretation.

2. The structure of the mark space

Motivated primarily by biomedical applications we now introduce our model, also adopting the corresponding terminology. Let $0<T_1<T_2<\dots$ be the successive time epochs at which observations are made and let the marks $X_1, X_2, \dots$ specify the qualities of these findings. Thus the data can be viewed as a sample path of a *marked point process* (abbr. MPP) $(T,X)=(T_n,X_n)_{n\geq 1}$.

Perhaps the simplest example of this arises if the data consists of the lifetimes of a set of n individuals. If the individuals enter the follow-up study at the same time instant, or if they are independent, we can measure their life lengths from a common origin. In terms of an MPP, $T_1<T_2<\dots$ are then the corresponding order statistics and X_j labels the individual whose lifetime equals T_j. (In the case of ties, X_j is chosen to be set-valued.)

A slight generalization of this is the common situation where some of the observations are censored before death. Also, it might be necessary to register the entry time into the study for each individual in order to investigate a time dependent factor. Both cases are easily included in the MPP model as is seen below.

Further, it is often desirable to link the failure times to a set of explanatory variables called *covariates*. Typically a medical researcher might be interested in the effect a treatment, or a set of prognostic factors, has on the remaining life length of the patient. For a thorough discussion about covariates the reader is referred to the book of Kalbfleisch & Prentice (1980). What concerns us here is that the covariate information, random or not, can be "embedded" into the sample path of an MPP if only the mark space E is made large enough.

Given this background of commonly occurring situations we suggest that the collected information, expressed in terms of marked points, can be usefully divided into two characteristic categories. Some of the marked points are called *innovations* and others *non-innovations*. (We may also decide that "a part of the mark" X_j observed at T_j forms an innovation.)

Before giving exact definitions, or considering examples in detail, we try to motivate such terminology. Consider first the censoring of an individual. Given that an individual has lived up to time t, the fact that he is then censored at t should in general not provide new information in the statistical inference problem. In this sense we would call his censoring at t non-innovative. If, however, he died at t (again given that he had lived up to time t), this would be usually considered as an innovation. A further example of information a researcher would generally judge to form a non-innovative point is the initial status of a patient at the entry time. Another would be a medical treatment given to a patient during the follow-up, given that the factors leading to the decision to undertake such treatment were determined—apart from possible random factors ancillary in the considered inference problem—by events registered previously in the sample path of the MPP.

We now attempt to formalize the above ideas, beginning with the structure of the mark space.

We assume that the elements of the mark space E are of the form $x=(x', x'')$, where x' is the *innovative part* and x'' the *non-innovative part* of x. A purely innovative mark is denoted by $(x', \emptyset'')$ and a purely non-innovative mark by $(\emptyset', x'')$, while $(\emptyset', \emptyset'')$ is excluded from the mark space. For mathematical tractability, the collection E' of all innovative parts is assumed to be countable while the collection E'' of all non-innovative parts can be more general (see Section 3).

If $T_n(\omega)=\infty$ for some $n\geq 1$, we follow the usual convention defining $X_n(\omega)=\Delta$, where $\Delta \notin E$ is a fictitious mark.

Example 2.1. If we need only consider the number of dead and censored individuals at each time point, as is often the case in a right-censored sample of n independent identically distributed lifetimes, we can take $E'=\{\emptyset', 1, 2, ..., n\}$ and $E''=\{\emptyset'', 1, 2, ..., n\}$.

Example 2.2. Let J denote the pool of individuals who, in the course of the study, may be under observation, and assume that J is either $\{1, 2, ..., n\}$ (with n finite) or $\{1, 2, ...\}$. In addition to death (denote by "d"), entry into the study ("e") and departure through censoring ("c"), as time proceeds, various covariate values can be registered about each individual, e.g., in the form of a medical record. As E' we take the set of sequences $x'=(x'_1, x'_2, ...)$, where x'_j may be "d" (j dies), "$\emptyset$" (no observation about j), or belong to a countable set Z' of "innovative covariate values". Since E' was assumed to be countable, we require that $x'_j\neq\emptyset$ for only finitely many j. Similarly, as E'' we take the sequences $x''=(x''_1, x''_2, ...)$, where $x''_j\in\{e, c, \emptyset\}\cup Z''$ ($j\in J$) and Z'' denotes the set of "non-innovative covariate values". In particular we have $\emptyset'=\emptyset''=(\emptyset, \emptyset, ...)$. If the study involves external covariates (see Kalbfleisch & Prentice (1980), p. 123) it is reasonable to regard them non-innovative. We can include them in the model as the first coordinate x''_0 of the sequence x''. Note that then it is possible to have points even when there are no individuals at risk. The covariate values may be names, real numbers, vectors or functions. Essential is that random effects occur only at points $T_0\equiv 0<T_1<T_2<\dots$.

To illustrate the flexibility of the rather complex mark space in Ex. 2.2., denote by $\pi_j=(\pi'_j, \pi''_j)$ the natural projection defined by $\pi'_j(x)=x'_j$ and $\pi''_j(x)=x''_j$ for $x=((x'_1, x'_2, ...), (x''_1, x''_2, ...))\in E$. Consider the stopping times

$$T^{z',j}=\inf\{T_n; \pi'_j(X_n)=z'\}, \quad z'\in\{d\}\cup Z',$$

$$T^{z'',j}=\inf\{T_n; \pi''_j(X_n)=z''\}, \quad z''\in\{e, c\}\cup Z''.$$

Then individual $j\in J$ enters the study at $T^{e,j}$, is under observation during the time interval $(T^{e,j}, T^{c,j}\wedge T^{d,j})$ and leaves forever through either death or censoring. The set

$$R(t)=\{j\in J;\ T^{e,j}<t\leq T^{c,j}\wedge T^{d,j}\} \tag{2.1}$$

is commonly called the *risk set* at time t.

3. The condition for non-innovation

We shall consider a parametric statistical model, i.e., a family $\{\mathbf{P}^\theta; \theta\in\Theta\}$ of equivalent probability measures on the measurable space $(\Omega, \mathcal{F})$ on which the variables $(T_n, X_n)_{n\geq 1}$ are defined.

What assumptions concerning such a model would justify our calling the observations in

E'' non-innovative? We consider this question first in the simple case where both E' and E'' are denumerable and the random variables T_n, $n \geq 1$, are integer-valued.

Denote $\mathbf{N}_0 = \{0, 1, 2, \ldots\}$ and define the "state-process" $(X(t))_{t \in \mathbf{N}_0} = (X'(t), X''(t))_{t \in \mathbf{N}_0}$ by

$$(X'(t), X''(t)) = \begin{cases} (X_i', X_i'') & \text{if } t = T_i \text{ for some } i \geq 1, \\ (\varnothing', \varnothing'') & \text{otherwise.} \end{cases} \tag{3.1}$$

Then there is a one-to-one correspondence between the sample paths of $(X(t))_{t \in \mathbf{N}_0}$ and $(T_i, X_i)_{i \geq 1}$. (Recall that $(\varnothing', \varnothing'')$ was not an element of the mark space E.) If $(t_i, (x_i', x_i''))_{1 \leq i \leq n}$ is a fixed observed sample path (stopped at t_n) from the marked point process and $(x(t))_{0 \leq t \leq t_n} = (x'(t), x''(t))_{0 \leq t \leq t_n}$ is the corresponding path of $(X(t))_{t \in \mathbf{N}_0}$, then the likelihood function has the obvious expression

$$\begin{aligned} L(\theta) &= \prod_{t=1}^{t_n} \mathbf{P}^\theta(X(t) = x(t) \mid X(s) = x(s),\ 0 \leq s \leq t-1) \\ &= \prod_{t=1}^{t_n} \mathbf{P}^\theta(X'(t) = x'(t) \mid X(s) = x(s),\ 0 \leq s \leq t-1) \\ &\quad \times \mathbf{P}^\theta(X''(t) = x''(t) \mid X(s) = x(s),\ 0 \leq s \leq t-1;\ X'(t) = x'(t)). \end{aligned} \tag{3.2}$$

Therefore, by assuming that

(A_1) "The conditional distributions

$$\mathbf{P}^\theta(X''(t) \in \cdot \mid X(s),\ 0 \leq s \leq t-1, \quad X'(t)),\ t \geq 0,$$

do not depend on θ"

holds, the likelihood function becomes proportional to the product

$$\prod_{t=1}^{t_n} \mathbf{P}^\theta(X'(t) = x'(t) \mid X(s) = x(s),\ 0 \leq s \leq t-1) \tag{3.3}$$

and so depends only on the conditional probabilities of the "innovative parts" $x'(t)$, $0 \leq t \leq t_n$.

The product (3.3) does not generalize into continuous time in a trivial way. However, the following provides an alternative formulation to (3.3) for which the generalization is more straightforward. Denote, for $t \in \mathbf{N}_0$,

$$\begin{aligned} p_t^\theta(x', x'') &= \mathbf{P}^\theta(X(t) = (x', x'') \mid X(s),\ 0 \leq s \leq t-1),\ (x', x'') \in E, \\ p_t^\theta(x', E'') &= \mathbf{P}^\theta(X'(t) = x' \mid X(s),\ 0 \leq s \leq t-1) \\ &= \sum_{x'' \in E''} p_t^\theta(x', x''), \quad x' \neq \varnothing', \\ r_t^\theta &= \mathbf{P}^\theta(X'(t) \neq \varnothing' \mid X(s), \quad 0 \leq s \leq t-1) \\ &= \sum_{x' \neq \varnothing'} p_t^\theta(x', E''), \quad \text{and} \\ q_t^\theta &= \mathbf{P}^\theta(X(t) \in E \mid X(s), \quad 0 \leq s \leq t-1) = \sum_{(x', x'') \in E} p_t^\theta(x', x''). \end{aligned} \tag{3.4}$$

Then, by an elementary calculation on conditional probabilities, we find that

$$\mathbf{P}^\theta(X''(t)=x''|X(s),\ 0\leqslant s\leqslant t-1,\ X'(t))=\begin{cases}\dfrac{p_t^\theta(X'(t),x'')}{p_t^\theta(X'(t),E'')}, & \text{if } X'(t)\neq\varnothing',\\[2ex] \dfrac{p_t^\theta(\varnothing',x'')}{1-r_t^\theta}, & \text{if } X'(t)=\varnothing',\ x''\neq\varnothing'',\\[2ex] \dfrac{1-q_t^\theta}{1-r_t^\theta}, & \text{if } X'(t)=\varnothing',\ x''=\varnothing'',\end{cases} \tag{3.5}$$

where we note that

$$\frac{1-q_t^\theta}{1-r_t^\theta}+\sum_{x''\neq\varnothing''}\frac{p_t^\theta(\varnothing',x'')}{1-r_t^\theta}=1.$$

But then assumption (A_1) can be rephrased as

(A_2) "The expressions

$$\frac{p_t^\theta(x',x'')}{p_t^\theta(x',E'')}\quad (x'\neq\varnothing',\ x''\in E'',\ t\geqslant 0)\quad\text{and}$$

$$\frac{p_t^\theta(\varnothing',x'')}{1-r_t^\theta}\quad (x''\neq\varnothing'',\ t\geqslant 0)$$ do not depend on θ".

Furthermore, under (A_2) the likelihood expression

$$\prod_{\substack{t=1\\ x'(t)\neq\varnothing'}}^{t_n} p_t^\theta(x'(t),E'')\prod_{\substack{s=1\\ x'(s)=\varnothing'}}^{t_n}(1-r_s^\theta), \tag{3.6}$$

considered on the event $\{X(t)=x(t),\ 0\leqslant t\leqslant t_n\}$, is equal to (3.3).

Let us now extend the above considerations to continuous time and a general mark space.

We assume that the mark space E has the structure given in Section 2 and make the technical assumption that E'' is a Polish space. We equip E' with the σ-field $\mathscr{E}'$ of all subsets, E'' with the Borel σ-field $\mathscr{E}''$, and E with the corresponding product-σ-field $\mathscr{E}$.

Let $(\mathscr{F}_t^0)_{t\geqslant 0}$ be the (right-continuous) internal history of the process (T,X), and let $\mathscr{F}_0$ be some sub-σ-field of $\mathscr{F}$, generated by a random element S_0 and completed, for every $\theta\in\Theta$, by $\mathbf{P}^\theta$-null-sets of $\mathscr{F}$. We then set

$$\mathscr{F}_t=\mathscr{F}_0\vee\mathscr{F}_t^0,\quad t\geqslant 0, \tag{3.7}$$

and assume this structure of the history throughout the rest of the paper. We denote by ν^θ the $(\mathbf{P}^\theta,\mathscr{F}_t)$-compensator of (T,X) (see e.g. Jacod (1975) or Davis (1976) for the definition).

The following representation result is essentially well known.

Lemma 3.1. (i) *The compensator (w.r.t.* $\mathbf{P}^\theta$*) of* (T,X) *has a version given by*

$$\nu^\theta(\omega;dt,dx)=d\Lambda_t^\theta(\omega)\,\varphi_t^\theta(\omega;dx). \tag{3.8}$$

Further, $(\Lambda_t^\theta)_{t\geqslant 0}$ is predictable and $\varphi_t^\theta(\omega;dx)$ is a transition probability from $(\Omega\times\mathbf{R}_+,\mathscr{P})$ into $(E,\mathscr{E})$ ($\mathscr{P}$ being the predictable σ-field);

(ii) The transition probability $\varphi_t^\theta(dx)$ satisfies, for all $1 \leq n < \infty$,

$$\varphi^\theta_{T_n}(A) = \mathbf{P}^\theta(X_n \in A | S_0, T_1, X_1, \ldots, T_{n-1}, X_{n-1}, T_n) \tag{3.9}$$

on $\{T_n < \infty\}$.

It follows from (3.9) that $d\Lambda_t^\theta$ has the role of q_t^θ in (3.4), while $\nu^\theta(dt, dx) = \nu^\theta(dt, dx', dx'')$ corresponds to $p_t^\theta(x', x'')$. From now on much use is made of the fact that E' is countable. We use $\nu^\theta(dt, x', dx'')$ as shorthand for $\nu^\theta(dt, \{x'\}, dx'')$. Define, for $t>0$ and $\theta \in \Theta$,

$$\varrho_t^\theta(\omega) = \nu^\theta(\omega; \{t\}, E' \setminus \{\emptyset'\}, E''),$$

$$\psi_t^{\theta, x'}(\omega; dx'') = \frac{\varphi_t^\theta(\omega; x', dx'')}{\varphi_t^\theta(\omega; x', E'')}, \quad x' \neq \emptyset', \ x'' \in E'',$$

$$\eta^\theta(\omega; dt, dx'') = \frac{\nu^\theta(\omega; dt, \emptyset', dx'')}{1 - \varrho_t^\theta(\omega)}, \quad x'' \neq \emptyset''. \tag{3.10}$$

(Note that (i) $\varrho_t^\theta(\omega) = 0$ unless the sample path $(\Lambda_s^\theta(\omega))_{s \geq 0}$ has a jump at t, (ii) $\psi_t^{\theta, x'}(\omega; \cdot)$ is a probability on $(E'', \mathscr{E}'')$, and (iii) $\eta^\theta(\omega; \cdot)$ is a measure on $(\mathbf{R}_+ \times E'', \mathscr{R}_+ \otimes \mathscr{E}'')$ such that $\eta^\theta(\omega; \cdot)$ vanishes if $d\Lambda_t^\theta(\omega) = 0$.) Then $\psi_t^{\theta, x'}(dx'')$ $(x' \neq \emptyset')$ can be understood as the conditional $\mathbf{P}^\theta$-probability of obtaining at time t a marked point whose non-innovative part is in dx'', given $\mathscr{F}_{t-}$ and the non-trivial innovative part x' of the marked point at t. Similarly, $\eta^\theta(dt, dx'')$ can be viewed as the $\mathbf{P}^\theta$-probability of obtaining in dt a marked point whose non-innovative part is in dx'', given $\mathscr{F}_{t-}$ and a trivial innovation at t. Note that innovation at t can be trivial either because there is a marked point of the form $(\emptyset', x'')$ which occurs with probability $d\Lambda_t^\theta(\omega)\, \varphi_t^\theta(\emptyset', dx'')$ or because there is no marked point at all at t, which happens with probability $1 - d\Lambda_t^\theta(\omega)$. Thus the probability of obtaining $(\emptyset', dx'')$ given a trivial innovative component at t becomes

$$d\Lambda_t^\theta(\omega)\, \varphi_t^\theta(\emptyset', dx'') / \{(1 - d\Lambda_t^\theta(\omega)) + d\Lambda_t^\theta(\omega)\, \varphi_t^\theta(\emptyset', E'' \setminus \{\emptyset''\})$$
$$= \nu^\theta(\omega, dt, \emptyset', dx'') / (1 - \varrho_t^\theta(\omega)) = \eta^\theta(\omega, dt, dx'').$$

We are now ready to state our main condition.

Condition A. For each $\omega \in \Omega$ the measures $\psi^{\theta, x'}(\omega; \cdot)$, $x' \neq \emptyset'$, and $\eta^\theta(\omega; \cdot)$ do not depend on θ.

It is immediate from the above explanation that, for every t, Condition A expresses a *conditional sufficiency* principle: Given $\mathscr{F}_{t-}$, the innovative part of the finding at t, trivial or not, is sufficient for θ.

Obviously, also the innovative parts form an MPP (with mark space $E' \setminus \{\emptyset'\}$), embedded in the original MPP (T, X): Define $(U, \hat{X}) = (U_n, \hat{X}_n)_{n \geq 1}$ by

$$\varkappa_0' = 0, \quad \varkappa_n' = \inf\{i > \varkappa_{n-1}'; X_i' \neq \emptyset'\}, \quad U_n = T_{\varkappa_n'}, \quad \hat{X}_n = X'_{\varkappa_n'} \tag{3.11}$$

We call the process $(U, \hat{X})$ *the innovation process*. For later use we denote

$$N_t^{x'} = \sum_{n=1}^{\infty} 1_{\{U_n \leq t, \hat{X}_n = x'\}}, \quad x' \neq \emptyset', \ t \geq 0,$$

$$d\Lambda_t^{\theta, x'} = \nu^\theta(dt, x', E''), \quad x' \in E', \ t \geq 0, \tag{3.12}$$

and note that $\{\Lambda_t^{\theta,x'};\, x'\neq\emptyset',\, t\geq 0\}$ is the $(\mathbf{P}^\theta, \mathscr{F}_t)$-compensator of the counting process $\{N_t^{x'};\, x'\neq\emptyset',\, t\geq 0\}$.

The following warning about the correct interpretation of Condition A is necessary: Condition A does *not* imply, when considering a sample path of (T, X) as the data set, say $\{(t_i, x_i', x_i'');\ 1\leq i\leq n\}$, that the reduced data $\{(t_i, x_i');\ 1\leq i\leq n,\ x_i'\neq\emptyset'\}$ corresponding to the innovation process $(U, \hat{X})$ would be sufficient for θ. This is because the conditioning, being on $(\mathscr{F}_t)$, may involve the full generating point process.

In the particular case of Example 2.1, Condition A can be seen as an exact mathematical formulation of the *non-informative censoring* discussed in Kalbfleisch & Prentice (1980, p. 121). However, for reasons that are obvious from the above warning we have preferred the term *non-innovative*. (Note the different meaning of the term non-informative censoring in Williams & Lagakos (1977)).

To conclude this section we consider some well known censoring mechanisms and the role of Condition A in this context.

Example 3.1. ("Progressive censorship (of Type II)", Gill (1980), Ex. 3.1.5). Suppose we start with n individuals alive. At the time of the first failure a random selection of r_1 individuals out of the $n-1$ still alive is removed from the test, all $\binom{n-1}{r_1}$ selections being equally likely. When the next recorded failure occurs, we withdraw r_2 individuals out of the $n-r_1-2$ still on test, etc., until a total of s failures have been observed, with r_k individuals being withdrawn at the kth stage $k=1, \dots, s$, $\Sigma_{k=1}^{s}(r_k+1)=n$.

Here we can choose, for $1\leq i\leq s$,

$T_i =$ the ith failure time,

$X_i' = \{j\}$ if j is the individual failing at T_i,

$X_i'' =$ the index set of the individuals censored at T_i.

Then, if $N(t)=\Sigma_i 1_{\{T_i\leq t\}}$ is the number of failures up to time t, the risk set $R(t)$ at time t is of size $|R(t)|=n-\Sigma_{i\leq N(t-)}(r_i+1)$. Moreover, it is easy to see that $\psi_t^{\theta, j}(x'')=\binom{|R(t)|-1}{r_{N(t-)+1}}^{-1}$ for every subset $x''\subset R(t)$ of size $r_{N(t-)+1}$ and index $j\in R(t)\setminus x''$. This is sufficient for Condition A to hold since the innovative parts in the marked points are always non-trivial (censoring takes place only at failure times), and therefore $\emptyset'$ (and consequently η^θ) can be dropped from the model.

In this example one could even have the selection probabilities depend on previous selections and failures and on the index numbers of the individuals, who need not have identical life-time distributions either. In this way one can extend the example to a k-sample setup without equal censoring in the k samples.

Example 3.2. ("Renewal testing", Gill (1980), Ex. 3.1.6). Here we have n components simultaneously on test. When a component fails before a fixed time u, it is immediately replaced by a new component. At time u all n components on test are censored. At the end of the test a random number N of failures have been observed (multiple failures are not ruled out) and there are exactly n censored observations.

We index by k, $1\leq k\leq n$, the used n "test sites", and set on $\{1\leq i\leq N\}$:

$T_i =$ the ith failure time,

$X_i' =$ the index set of test sites of components failing at T_i,

$X_i'' = \emptyset$ (the empty set),

while

$$T_{N+1} = u,$$
$$X'_{N+1} = \emptyset \text{ (the empty set)},$$
$$X''_{N+1} = \{1, 2, ..., n\}.$$

Thus all points before time u are purely innovative and for them Condition A is trivial. But the final point at time u is purely non-innovative with deterministic mark $(\emptyset, \{1, ..., n\})$, and so Condition A holds also there with $\eta^\theta(dt, dx'')$ being the unit mass at $(u, \{1, 2, ..., n\})$.

Example 3.3. Here we continue to assume that only censorings and deaths are registered but now postulate that all individuals behave independently (including their censoring). Then it is obvious that censoring is non-innovative, i.e., Condition A holds if it holds separately for each individual with respect to its own history. Considering the jth individual, let (T'_j, δ_j) be the registered observation: T'_j is the failure time or the censoring time, whichever occurs first, and $\delta_j=1$ or 0 depending on whether j dies or is censored at T'_j. (We assume that no individual is censored at its own failure time.) Denoting $G^\theta_j(dt, k)=\mathbf{P}^\theta(T'_j\in dt, \delta_j=k)$, $k=0, 1$, $G^\theta_j(dt)=\mathbf{P}^\theta(T'_j\in dt)$ $(=G^\theta_j(dt, 1)+G^\theta_j(dt, 0))$, assuming absolute continuity and writing $g^\theta_j(\cdot, k)$ for the density corresponding to $G^\theta_j(\cdot, k)$, $k=0, 1$, we define the hazard rates for censoring and death by

$$r^\theta_j(t, k) = \frac{g^\theta_j(t, k)}{1-G^\theta_j(t)}, \quad k=0, 1. \tag{3.13}$$

It is well known that $r^\theta_j(t, 1)\, 1_{\{T'_j\geq t\}}$ and $r^\theta_j(t, 0)\, 1_{\{T'_j\geq t\}}$ are then the intensities for respectively "j dies" and "j is censored". Clearly, Condition A holds if $r^\theta_j(t, 0)$ does not depend on θ.

Note that the *random censorship* model (see e.g. Efron (1967), Gill (1980), Ex. 3.1.3, or Lagakos (1979)), requiring that T'_j is the minimum of two independent random variables, is a special case of Example 3.3.

4. The likelihood ratio process

Our discussion in Section 3 already indicated that non-innovative points give rise to a factor in the likelihood expression which does not depend on the parameter θ. Consequently this factor can be ignored when inferring about θ. We now give a rigorous proof of this fact. We start from the likelihood *ratio* (corresponding to two parameter values) and show that under Condition A a factor arising from non-innovative marked points cancels out. From this result we can easily derive an expression for the likelihood function (up to a proportionality factor).

Unfortunately the exact mathematical derivation of the likelihood expression remains rather involved, in spite of our attempts to find convenient notations. We emphasize, however, that the final result (4.15) is the exact continuous time analogue of the elementary formula (3.6).

We fix $\theta, \tilde\theta\in\Theta$, $\theta\neq\tilde\theta$, and drop "θ" from the upper-indices. Thus we write $\mathbf{P}$ and $\tilde{\mathbf{P}}$ instead of $\mathbf{P}^\theta$ and $\mathbf{P}^{\tilde\theta}$ etc.

First we sketch some necessary results from Jacod (1975). Recall that $(E, \mathscr{E})$ is Polish and the history $(\mathscr{F}_t)_{t\geq 0}$ is of the form (A 1) in Jacod (1975).

Proceeding as in the proof of Proposition (2.3) of Jacod (1975), one can choose a version of ν satisfying (3.8) and

$$\Delta\Lambda_t(\omega) \leq 1 \tag{4.1}$$

for all $t \geq 0$ and $\omega \in \Omega$ ($\Delta\Lambda_t(\omega)$ denotes the jump at t of the mapping $t \mapsto \Lambda_t(\omega)$). Of course the same holds for the $\tilde{\mathbf{P}}$-compensator $\tilde{\nu}$. Now $\tilde{\mathbf{P}} \ll \mathbf{P}$ by assumption and thus according to Jacod (1975), Theorem (4.1), there exists a finite $\mathcal{P}\otimes\mathcal{E}$-measurable mapping $(\omega, t, x) \rightarrow Y(\omega, t, x)$ such that

$$\tilde{\nu}(\omega; dt, dx) = Y(\omega, t, x)\, \nu(\omega; dt, dx) \tag{4.2}$$

for $\mathbf{P}$-almost all $\omega \in \Omega$. Further Y satisfies identically

$$\hat{Y}_t \stackrel{\text{def.}}{=} \int_E Y(t, x)\, \nu(\{t\} \times dx) \leq 1 \quad \text{and} \quad \hat{Y}_t = 1, \quad \text{whenever } \Delta\Lambda_t = 1. \tag{4.3}$$

We see that for $\mathbf{P}$-almost all $\omega \in \Omega$ we have $\tilde{\nu}(\omega; \cdot) \ll \nu(\omega; \cdot)$ and $\tilde{\Lambda}(\omega; \cdot) \ll \Lambda(\omega; \cdot)$, the respective Radon-Nikodym derivatives being $Y(\omega, \cdot)$ and

$$t \rightarrow \int_E Y(\omega, t, x)\, \varphi_t(\omega; dx). \tag{4.4}$$

Note that (4.4) defines a predictable process.

Let L denote the Radon-Nikodym derivative of $\tilde{\mathbf{P}}$ with respect to $\mathbf{P}$. Then $L_t = \mathbf{E}(L|\mathcal{F}_t)$ gives the corresponding Radon-Nikodym derivative on σ-field $\mathcal{F}_t$, $0 \leq t \leq \infty$ ($\mathcal{F}_\infty = V_{t=0}^{\infty} \mathcal{F}_t$). We see from Jacod (1975), Theorem (5.1) (or Bremaud & Jacod (1977), Theorem 1, p. 388) that L_t has the version

$$L_t = L_0 \left(\prod_{T_n \leq t} Y(T_n, X_n) \right) \left(\prod_{\substack{s \leq t \\ s \notin (T_n)}} D_s \right) \exp\left\{ \int_0^t \int_E (1 - Y(s, x))\, \nu^c(ds, dx) \right\}, \tag{4.5}$$

where

$$D_s = 1 + \frac{\Delta\Lambda_s - \hat{Y}_s}{1 - \Delta\Lambda_s}, \tag{4.6}$$

and

$$\nu^c(dt, dx) = 1_{\{\Delta\Lambda_t = 0\}}\, \nu(dt, dx) \tag{4.7}$$

is the "continuous part" of ν. Note that $\hat{Y}_t = \Delta\tilde{\Lambda}_t$ (by (4.2)) and thus $D_t = (1 - \Delta\tilde{\Lambda}_t)/(1 - \Delta\Lambda_t)$ ($=1$, if $\Delta\Lambda_t = \Delta\tilde{\Lambda}_t = 1$) giving the more illuminating version of (4.5)

$$L_t = L_0 \left(\prod_{T_n \leq t} \frac{d\tilde{\nu}}{d\nu}(T_n, X_n) \right) \left(\prod_{\substack{s \leq t \\ s \notin (T_n)}} \frac{1 - \Delta\tilde{\Lambda}_s}{1 - \Delta\Lambda_s} \right) \frac{\exp(-\tilde{\Lambda}_t^c)}{\exp(-\Lambda_t^c)}. \tag{4.5'}$$

Furthermore, formula (15) of Jacod (1975) gives us the representation

$$L_t = L_0 + \int_0^t \int_E L_{s-}(D_s - Y(s, x))\, (\nu(ds, dx) - N(ds, dx)), \quad t > 0, \tag{4.8}$$

where N is the counting measure on $(\mathbf{R}_+ \times E, \mathcal{R}_+ \otimes \mathcal{E})$ corresponding to (T, X).

Our aim is now to calculate (4.5) explicitly for the failure model. When doing so we find how a proportionality factor, arising from Condition A, cancels out.

We use the notation Y^* for our candidate for Y in (4.2). Denoting by C the complement of $\{\omega|\tilde{\nu}(\omega;\cdot)\ll\nu(\omega;\cdot)\}$, we define for $\omega\in\Omega\setminus C$

$$Y^*(\omega,t,x',x'')=\begin{cases}\dfrac{d\tilde{\Lambda}(\omega)}{d\Lambda(\omega)}(t)\dfrac{\tilde{\varphi}_t(\omega;x',E'')}{\varphi_t(\omega;x',E'')}, & \text{if } x'\neq\varnothing',\\ \dfrac{1-\tilde{\varrho}_t(\omega)}{1-\varrho_t(\omega)}, & \text{if } x'=\varnothing',\ \varrho_t(\omega)<1,\\ 1, & \text{if } x'=\varnothing',\ \varrho_t(\omega)=1,\end{cases} \tag{4.9}$$

where $d\tilde{\Lambda}(\omega)/d\Lambda(\omega)$ is defined according to (4.4), and for $\omega\in C$ let $Y^*(\omega,\cdot,\cdot)\equiv 1$.

In order to use this Y^* in the construction of the likelihood ratio (4.5) we must check (4.2) and (4.3). Now (4.2) holds, because on $\Omega\setminus C$, for $x'\neq\varnothing'$

$$\begin{aligned}\nu(dt,x',dx'')\,Y^*(t,x',x'')&=d\Lambda_t\,\varphi_t(x',dx'')\frac{d\tilde{\Lambda}}{d\Lambda}(t)\frac{\tilde{\varphi}_t(x',E'')}{\varphi_t(x',E'')}\\ &=d\Lambda_t\frac{d\tilde{\Lambda}}{d\Lambda}(t)\,\tilde{\varphi}_t(x',dx'')\\ &=\tilde{\nu}(dt,x',dx'')\end{aligned} \tag{4.10 a}$$

by Condition A, and similarly for $x'=\varnothing'$

$$\begin{aligned}\nu(dt,\varnothing',dx'')\,Y^*(t,\varnothing',x'')&=\nu(dt,\varnothing',dx'')\frac{1-\tilde{\varrho}_t}{1-\varrho_t}\\ &=\tilde{\nu}(dt,\varnothing',dx'').\end{aligned} \tag{4.10 b}$$

(The first equality in (4.10 b) follows from (4.9) and the fact that $\varrho_t=1$ implies $\nu(\{t\},\varnothing',E'')=0$.) But then also (4.3) holds since it holds trivially on C, and on $\Omega\setminus C$ conditions (4.10) imply that

$$\hat{Y}^*_t=\tilde{\nu}(\{t\},E)=\Delta\tilde{\Lambda}_t. \tag{4.11}$$

From now on we can write Y instead of Y^*. To begin with the computation of (4.5), note first that corresponding to (4.6) we have on $\Omega\setminus C$

$$\begin{aligned}D_t&=\frac{1-\Delta\tilde{\Lambda}_t}{1-\Delta\Lambda_t}\quad\text{(by (4.11))}\\ &=\frac{1-\tilde{\varrho}_t-\Delta\tilde{\Lambda}_t^{\varnothing'}}{1-\varrho_t-\Delta\Lambda_t^{\varnothing'}}\quad\text{(for notation see (3.12))}\\ &=\frac{(1-\tilde{\varrho}_t)\left(1-\dfrac{\Delta\tilde{\Lambda}^{\varnothing'}}{1-\tilde{\varrho}_t}\right)}{(1-\varrho_t)\left(1-\dfrac{\Delta\Lambda^{\varnothing'}}{1-\varrho_t}\right)}\\ &=\frac{1-\tilde{\varrho}_t}{1-\varrho_t}\quad\text{(by Condition A),}\end{aligned} \tag{4.12}$$

if $0<\Delta\Lambda_t<1$, and $D_t=1$ otherwise (by absolute continuity when $\Delta\Lambda_t=0$, and by (4.3) and the convention $0/0=0$ of Jacod (1975) when $\Delta\Lambda_t=1$).

Next, considering an arbitrary $\omega\in\Omega$ and defining Y^{**} by

$$Y^{**}(\omega; t, x', x'') = \begin{cases} \dfrac{d\tilde{\nu}^c(\omega; \cdot)}{d\nu^c(\omega; \cdot)}, & \text{if } x' \neq \varnothing', \\ 1, & \text{if } x' = \varnothing', \end{cases}$$

we find that $Y(\omega; \cdot) = Y^{**}(\omega; \cdot)\ \nu^c(\omega; \cdot)$–a.e. Therefore

$$\int_0^t \int_E (1 - Y(s,x))\, \nu^c(ds, dx) = \nu^c((0,t], E' \setminus \{\varnothing'\}, E'') - \tilde{\nu}^c((0,t], E' \setminus \{\varnothing'\}, E'')$$

$$= \sum_{x' \neq \varnothing'} (\Lambda_t^{c,x'} - \tilde{\Lambda}_t^{c,x'}). \tag{4.13}$$

But then (4.5) together with (4.9), (4.12) and (4.13) implies (4.14) in the following basic result, where we use the notation of (3.10), (3.11) and (3.12).

Theorem 4.1. *Under Condition A, if the probabilities* **P** *and* $\tilde{\mathbf{P}}$ *agree on* $\mathcal{F}_0$, *the martingale* $L_t = \mathbf{E}(d\tilde{\mathbf{P}}/d\mathbf{P} | \mathcal{F}_t)$, $t \geq 0$, *has the representation*

$$L_t = \left[\prod_{U_n \leq t} \frac{d\tilde{\Lambda}^{\hat{X}_n}}{d\Lambda^{\hat{X}_n}}(U_n) \right] \left[\prod_{\substack{s \leq t \\ s \notin (U_n)}} \frac{1 - \tilde{\varrho}_s}{1 - \varrho_s} \right] \exp\left\{ \sum_{x' \neq \varnothing'} (\Lambda_t^{c,x'} - \tilde{\Lambda}_t^{c,x'}) \right\}. \tag{4.14}$$

Note that if $\Delta\Lambda_s = 1$ and $T_n \neq s$ for all $n \geq 1$, then there is a factor $D_s = 1$ in (4.5) but a factor $(1 - \tilde{\varrho}_s)/(1 - \varrho_s)$ in (4.14). However, this occurs only with probability zero.

Let us then derive an expression for the likelihood function. Consider (4.14) for a fixed value $\theta = \theta_0$ in the denominator and write $\mathbf{P}^\theta$ (instead of $\tilde{\mathbf{P}}$) in the numerator. Then, for fixed $\omega \in \Omega$, $t \geq 0$, and with θ varying, (4.14) implies that the likelihood function based on the sample

$$\tilde{y}_t = \tilde{Y}_t(\omega) = (S_0(\omega), (T_i(\omega), X_i(\omega))_{1 \leq i \leq N_t(\omega)}),$$

where $N_t(\omega)$ denotes the number of points on the interval $[0, t]$, is proportional to the expression

$$H_t^\theta = \left[\prod_{U_n \leq t} d\Lambda_{U_n}^{\theta, \hat{x}_n} \right] \left[\prod_{\substack{s \leq t \\ s \notin (U_n)}} (1 - \varrho_s^\theta) \right] \exp\left\{ \sum_{x' \neq \varnothing'} \Lambda_t^{\theta, c, x'} \right\}. \tag{4.15}$$

Here $d\Lambda_s^{\theta, x'}$ should be read as the density of $\Lambda^{\theta, x'}$ with respect to any (for all θ) dominating measure $\mu^{x'}$. We note that H_t^θ depends on ω only through $(\tilde{Y}_t)$, and by (4.5) it is, for fixed $\omega \in \{H_t^{\theta_0} > 0\}$ and as a function of θ, proportional to $L_t^\theta = d\mathbf{P}^\theta / d\mathbf{P}^{\theta_0}$ (on $\mathcal{F}_t$). This conclusion is also valid if $\mathbf{P}^{\theta_0}$ is replaced by any dominating measure μ on $\mathcal{F}_t$ (the proportionality factor gets multiplied by $(d\mathbf{P}^{\theta_0}/d\mu)(\tilde{y}_t)$). Thus under Condition A we may view $\theta \to H_t^\theta(\tilde{y}_t)$ as the *likelihood function* based on the data $\tilde{y}_t$.

Remark 1. If Condition A is not satisfied, expression (4.15) can be considered as a *partial likelihood.* Under the usual kind of regularity conditions it leads to the familiar asymptotic properties of likelihood based inferences (see Cox (1975), Kalbfleisch & Prentice (1980), pp. 127–132).

Remark 2. There is an alternative formulation to the likelihood formula (4.15) which is based on the notion of a product integral (see Dollard & Friedman (1979), and also

Kalbfleisch & Prentice (1980), p. 9): If μ is a measure on the real line with all point masses less than one, the product integral $\mathscr{P}_0^t(1-\mu(ds))$ is defined according to "the exponential formula" as

$$\mathscr{P}_0^t(1-\mu(ds)) = \exp\left\{-\int_0^t \mu^c(ds)\right\}\prod_{s\leq t}(1-\mu(\{s\})),$$

where μ^c is the continuous part of μ. Recall from (3.12) that the $(\mathbf{P}^\theta, \mathscr{F}_t)$-compensator of the innovation process $(U, \hat{X})$ is $d\Lambda_t^{\theta,x'} = \nu^\theta(dt, x', E'')$. Also denote $d\hat{\Lambda}_t^\theta = \nu^\theta(dt, E' \setminus \{\varnothing'\}, E'')$ and $\hat{N}_t = \Sigma_{x' \neq \varnothing'} N_t^{x'}$, $t \geq 0$. Then we find easily that (4.15) can be written as

$$H_t^\theta = \left[\prod_{s\leq t}\prod_{x'\neq\varnothing'}(d\Lambda_s^{\theta,x'})^{dN_s^{x'}}\right][\mathscr{P}_0^t(1-d\hat{\Lambda}_s^\theta)^{1-d\hat{N}_s}], \tag{4.15a}$$

where the product $\Pi_{s\leq t}$ has only countably many non-trivial terms.

We conclude this section by again studying examples.

Example 2.2 (continuation). We assume Condition A to hold, and additionally that only failures are innovative, i.e., $Z' = \varnothing$. Then each innovation can be identified with some finite subset I of J, and we can conveniently replace "x'" by "I". We denote by $|I|$ the cardinality of I.

We first consider the case when the distributions of the lifetimes are *continuous*. This is achieved by making the additional assumptions

$$\varphi_s^\theta(I, E'') = 0, \quad \text{when } |I| \geq 2, \tag{4.16a}$$

$$\Delta\Lambda_s^\theta \varphi_s^\theta(I, E'') = 0, \quad \text{when } I \neq \varnothing, \tag{4.16b}$$

for all $\theta \in \Theta$, $s \geq 0$. We also assume that for each $\theta \in \Theta$ the continuous part of Λ^θ has an intensity. In this case, we get from (4.15) that the likelihood for θ at t is proportional to the expression

$$\left[\prod_{U_n\leq t}\lambda_{U_n}^{\theta,\hat{X}_n}\right]\exp\left\{-\sum_{\substack{I\subset J\\0<|I|<\infty}}\int_0^t \lambda_s^{\theta,I}ds\right\}$$

$$= \left[\prod_{U_n\leq t}\lambda_{U_n}^{\theta,\hat{X}_n}\right]\exp\left\{-\sum_{j\in J}\int_0^t \lambda_s^{\theta,\{j\}}ds\right\}. \quad \text{(by 4.16a)} \tag{4.17}$$

Here, obviously, the innovation process $(U, \hat{X})$ gives the failure times and $\hat{X}_n$ labels the individual failing at U_n (by (4.16a) the $\mathbf{P}^\theta$-probability of simultaneous failures is zero).

If the continuous part of Λ^θ vanishes for each $\theta \in \Theta$, we have the important special case of a purely *discrete* process. Then (4.15) reduces to

$$\left[\prod_{U_n\leq t}\Delta\Lambda_{U_n}^{\theta,\hat{X}_n}\right]\left[\prod_{\substack{s\leq t\\ s\notin(U_n)}}\left(1-\sum_{0<|I|<\infty}\Delta\Lambda_s^{\theta,I}\right)\right] \tag{4.18}$$

Here simultaneous failures are usually possible with positive probability and thus the finite subsets of J can come up as failure patterns.

5. Partially specified models and comments on interpretation

In practice it is often impractical to specify explicitly the full parametric statistical model $\{\mathbf{P}^\theta; \theta \in \Theta\}$ governing an MPP. As we have seen above, however, the likelihood (4.15) is expressed solely in terms of the compensator $\{\Lambda_t^{\theta, x'}; x \in E', t \geq 0\}$ associated with *innovations*. Indeed, it is this economy aspect which makes the classification of the marked points into innovations and non-innovations so useful: As long as Condition A is satisfied, only the innovation hazards need explicit parametric specification. We call such an approach to setting up a statistical model *a partial specification*. (Note, however, that Condition A can be only expressed in terms of the full model. Thus in a sense one has to investigate the full model in order to be able to concentrate on the partial model.) The purpose of this last section is to consider the partial specification in some detail, and also discuss the interpretation of the consequent model.

The most convenient starting point is Example 3.3.

Example 3.3 (*continuation*). Because of the assumed independence between individuals, the likelihood expression involves only functions $\{r_j^\theta(t, 1); j \in J, t \geq 0\}$, $r_j^\theta(t, 1)$ being the death hazard individual j faces, if alive and uncensored, at time t. Thus, under the non-innovation condition on censoring (according to which the functions $\{r_j^\theta(t, 0); j \in J, t \geq 0\}$ must not depend on θ) it suffices to specify the death hazards $r_j^\theta(t, 1)$. Usually one then assumes that the life length of individual j, say ξ_j, is distributed according to some parametric distribution F_j^θ, called here a *reference distribution*. Supposing that F_j^θ has a density f_j^θ, one gets the corresponding hazard rate h_j^θ from $h_j^\theta(t) = f_j^\theta(t)[1 - F_j^\theta(t)]^{-1}$ and defines $r_j^\theta(t, 1)$ by

$$r_j^\theta(t, 1) = h_j^\theta(t), \quad t \geq 0 \tag{5.1}$$

It is then easily seen by substitution into (4.17) that the likelihood expression corresponding to the data set $\{(T_j', \delta_j); j \in J\}$ gets the familiar form

$$\prod_j (f_j^\theta(T_j'))^{\delta_j} (1 - F_j^\theta(T_j'))^{1-\delta_j}.$$

Condition (5.1) was shown by Kalbfleisch & MacKay (1979) to be equivalent to *the constant sum condition* of Williams & Lagakos (1977). In Kalbfleisch & Prentice (1980) it is called *independent censoring* (see Assumption 2 on p. 120). Compare also with Assumption 3.1.1 in Gill (1980).

The equality (5.1) expresses a very natural relationship between the statistical models for $\{(T_j', \delta_j'); j \in J\}$ and for $\{\xi_j; j \in J\}$: In the former, $r_j^\theta(t, 1)$ is the hazard rate of an *uncensored* individual, while the latter model does not contain censoring at all. Thus we can take (5.1) to mean that censoring does not introduce a bias into the model for $\{(T_j', \delta_j); j \in J\}$ by altering the death hazards.

Remark 1. Often censoring is unavoidable in practical experimentation and then a companion experiment in which "the censoring mechanism has been inactivated" remains purely fictitious. However, we believe that in most cases such a *reference experiment*, and the application of the consequent model as in (5.1), make good sense. In the above case, for example, it is reasonable to think that it is really the values of the variables ξ_j an investigator would like to observe in order to study the statistical model based on the distributions F_j^θ. In a sense, the life lengths ξ_j "exist in reality", but their measurement is sometimes prevented by the censoring.

Remark 2. An alternative mathematical formulation, also leading to (5.1), is the well known *random censoring model*: There one postulates the existence of a censoring time for individual j, say ζ_j, taken to be independent from ξ_j, and then defines (T_j', δ_j) by $T_j'=\xi_j \wedge \zeta_j$ and $\delta_j = 1_{\{\xi_j \leq \zeta_j\}}$. A serious drawback of this model is, however, that in most experiments the random variable ζ_j has no real meaning if death is observed. Then also the independence becomes an artefact of the model. (This is essentially the usual critique directed towards the *competing risks* models.)

In general the individuals under study cannot be considered separately because of lacking independence between them. (As we have seen in Examples 3.1 and 3.2, even if the failure mechanisms are independent the observations (T_j', δ_j) can be dependent because of censoring.) Also conditioning on common random covariates can bring in dependence. This difficulty can be avoided by splitting the time axis into random intervals of the form $(T_{n-1}, T_n]$ and then considering separately each such interval. The following example illustrates the partial model specification in such a case.

Example 5.1. We consider Ex. 2.2 in the situation where (4.16a–b) holds. Since only deaths are innovative, the partial specification requires merely that the death rates $\lambda_t^{\theta, \{j\}}$ be defined. Choosing predictable versions for such rates amounts to writing

$$\lambda_t^{\theta, \{j\}}(\omega) = \sum_{n \geq 1} \lambda_j^{\theta, n}(t, \omega)\, 1_{\{T_{n-1}(\omega) < t \leq T_n(\omega)\}}, \tag{5.2}$$

where each $\lambda_j^{\theta, n}$ is $\mathcal{F}_{T_{n-1}} \otimes \mathcal{R}_+$-measurable (see e.g. Bremaud (1981)). Also recall that for $T_{n-1} < t \leq T_n$ the hazard rate $\lambda_t^{\theta, \{j\}}$ can be expressed in terms of conditional $\mathbf{P}^\theta$-distributions as

$$\begin{aligned} \lambda_t^{\theta, \{j\}}\, dt &= \mathbf{P}^\theta(T_n \in dt,\ \pi_j'(X_n) = d \mid \mathcal{F}_{T_{n-1}};\ T_n \geq t) \\ &\underset{\text{def.}}{=} \frac{\mathbf{P}^\theta(T_n \in dt,\ \pi_j'(X_n) = d \mid \mathcal{F}_{T_{n-1}})}{P^\theta(T_n \geq t \mid \mathcal{F}_{T_{n-1}})} \end{aligned}$$

(cf. the interpretation of $r_j^\theta(t, 1)$ above).

In order to determine the functions $\lambda_j^{\theta, n}$ on the right hand side of (5.2) we fix the risk set and the values of the covariates (up to functional changes in time) by conditioning on $\mathcal{F}_{T_{n-1}}$, and then consider a (possibly only hypothetical) reference experiment performed under those fixed conditions, starting at T_{n-1} and terminating at the next failure. The partial specification of the model on $(T_{n-1}, T_n]$ then consists of choosing a probability distribution for this failure, letting $\lambda_j^{\theta, n}$ be the corresponding hazard rate for individual j, and finally asking the equality $\lambda_t^{\theta, \{j\}} = \lambda_j^{\theta, n}$ to hold for $t \in (T_{n-1}, T_n]$.

For instance, in studying covariate influence on failures one often models the death hazard for a person at risk to depend on ω only through a p-dimensional vector of covariate values, $p \geq 1$. The well known proportional hazards model of Cox (1972) is an example of this. Allowing for time-dependent stochastic covariates, one then uses the representation

$$\lambda_j^{\theta, n}(t, \omega) = \lambda_j^\theta(t, Z_j^n(\omega, t))\, 1_{\{j \in R(\omega, t)\}}, \quad n \geq 1,$$

in (5.2), where the mapping $t \to Z_j^n(\omega, t) \in \mathbf{R}^p$ is determined by the covariate history of $j \in J$ up to the point $T_{n-1}(\omega)$. In the particular case of finite J and fixed covariates, i.e., if

$Z_j^n(\omega, t)=z_j$ does not depend on ω, t and n, one can derive the likelihood expression (5.2) of Kalbfleisch & Prentice (1980) by substitution into (4.17) at $t=\infty$.

Let us then consider the partial specification of the general MPP model. Corresponding to the partition

$$(0, t] = \sum_{n\leq N(t)} (T_{n-1}, T_n]+(T_{N(t)}, t]$$

we factorize the likelihood (4.15) into a product form

$$H_t^\theta = \left[\prod_{n\leq N(t)} H^\theta_{(T_{n-1}, T_n]}\right]\cdot H^\theta_{(T_{N(t)}, t]}, \tag{5.3}$$

where (recall that $\varrho_s^\theta=\Sigma_{x'\neq\varnothing'}\Delta\Lambda_s^{\theta,x'}$)

$$H^\theta_{(T_{n-1}, T_n]} = \begin{cases} \left[\prod_{s\in(T_{n-1}, T_n]}\left(1-\sum_{x'\neq\varnothing'}\Delta\Lambda_s^{\theta,x'}\right)\right]\exp\left\{-\sum_{x'\neq\varnothing'}\left[\Lambda_{T_n}^{\theta,c,x'}-\Lambda_{T_{n-1}}^{\theta,c,x'}\right]\right\} & \text{if } X'_n=\varnothing' \\ \left[\prod_{s\in(T_{n-1}, T_n)}\left(1-\sum_{x'\neq\varnothing'}\Delta\Lambda_s^{\theta,x'}\right)\right]\exp\left\{-\sum_{x'\neq\varnothing'}\left[\Lambda_{T_n}^{\theta,c,x'}-\Lambda_{T_{n-1}}^{\theta,c,x'}\right]\right\}\cdot d\Lambda_{T_n}^{\theta,X'_n} & \text{if } X'_n\neq\varnothing' \end{cases} \tag{5.4}$$

and

$$H^\theta_{(T_{N(t)}, t]} = \begin{cases} \left[\prod_{s\in(T_{N(t)}, t]}\left(1-\sum_{x'\neq\varnothing'}\Delta\Lambda_s^{\theta,x'}\right)\right]\exp\left\{-\sum_{x'\neq\varnothing'}\left[\Lambda_t^{\theta,c,x'}-\Lambda_{T_{N(t)}}^{\theta,c,x'}\right]\right\} & \text{if } T_{N(t)}<t \\ 1 & \text{otherwise} \end{cases} \tag{5.5}$$

For each $n\geq 1$, the term $H^\theta_{(T_{n-1}, T_n]}$ depends on an $\mathcal{F}_{T_{n-1}}$-conditional distribution and the innovation X'_n at T_n (which may be trivial). In more detail, we have (a.s.) on $\{T_{n-1}<t\leq T_n\}$

$$\begin{aligned} d\Lambda_t^{\theta,x'} &= \mathbf{P}^\theta(T_n\in dt,\ X'_n=x'\mid\mathcal{F}_{T_{n-1}};\ T_n\geq t) \\ &\underset{\text{def.}}{=} \frac{\mathbf{P}^\theta(T_n\in dt,\ X'_n=x'\mid\mathcal{F}_{T_{n-1}})}{\mathbf{P}^\theta(T_n\geq t\mid\mathcal{F}_{T_{n-1}})},\quad x'\neq\varnothing'. \end{aligned} \tag{5.6}$$

(Conversely, the conditional distribution can be determined from the compensator by using the exponential inversion formula (cf. (5.7) & (5.9) below).)

It is clear that $d\Lambda_t^{\theta,x'}$ has here the same role as $r_j^\theta(t, 1)\,dt$ in Ex. 3.3 and $\lambda_t^{\theta,\{j\}}\,dt$ in Ex. 5.1.

To introduce the corresponding reference distributions, let, for $n\geq 1$ and $\theta\in\Theta$, $Q^{\theta,n}(dt,x')$ be a given transition probability from Ω to $(0,\infty]\times E'$ such that $Q^{\theta,n}(B,x')$ is $\mathcal{F}_{T_{n-1}}$-measurable for any fixed Borel set B and $x'\in E'$. Let $Q^{\theta,n}$ also satisfy $Q^{\theta,n}((0,T_{n-1}],E')=0$. In setting up the model we give to $Q^{\theta,n}$ the role of the conditional distribution, given $\mathcal{F}_{T_{n-1}}$, of the first innovation in the post-T_{n-1}-process under the (possibly only fictitious) circumstances where the generation of non-innovative points has been discontinued at T_{n-1}. Define the (multivariate) hazard function $\{\Gamma_t^{\theta,n,x'};\ x'\in E',\ t>T_{n-1}\}$ corresponding to $Q^{\theta,n}$ in the usual way by

$$d\Gamma_t^{\theta,n,x'} = \frac{Q^{\theta,n}(dt,x')}{Q^{\theta,n}([t,\infty],E')}. \tag{5.7}$$

Then $d\Gamma_t^{\theta,n,x'}$ corresponds to $\lambda_j^{\theta,n}(t)\,dt$ in Example 5.1.

The partial specification of the general MPP model is accomplished by requiring that $\Lambda_t^{\theta,x'}$ satisfy

$$d\Lambda_t^{\theta,x'} = d\Gamma_t^{\theta,\mathrm{n},x'} \quad \text{on } \{T_{n-1}<\mathrm{t}\leqslant T_n\}. \tag{5.8}$$

By substituting (5.8) into (5.4) and (5.5), and using the exponential inversion formula

$$Q^{\theta,n}(dt,x') = \left[\prod_{s<t}\left(1-\sum_{x'\neq\varnothing}\Delta\Gamma_s^{\theta,n,x'}\right)\right]\cdot\exp\left\{-\sum_{x'\neq\varnothing'}\Gamma_t^{\theta,n,c,x'}\right\}d\Gamma_t^{\theta,n,x'}, \tag{5.9}$$

we find that

$$H^{\theta}_{(T_{n-1},T_n]} = \begin{cases} Q^{\theta,n}(dT_n,X'_n), & \text{if } X'_n\neq\varnothing, \\ Q^{\theta,n}((T_n,\infty],E'), & \text{if } X'_n=\varnothing, \end{cases}$$

and

$$H^{\theta}_{(T_{N(t)},t]} = Q^{\theta,N(t)}((t,\infty],E'), \quad t>T_{N(t)}.$$

Hence, by (5.3), the partial specification in terms of the distributions $Q^{\theta,n}$ leads to the likelihood expression

$$H_t^{\theta} = \left[\prod_{\substack{T_n\leqslant t\\ X'_n\neq\varnothing}} Q^{\theta,n}(dT_n,X'_n)\right]\left[\prod_{\substack{T_n\leqslant t\\ X'_n=\varnothing'}} Q^{\theta,n}((T_n,\infty],E')\right] Q^{\theta,N(t)}((t,\infty],E'). \tag{5.10}$$

Acknowledgements

We are grateful to Søren Johansen and Richard Gill for their many helpful comments on earlier versions of this paper.

References

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

Aalen, O. (1982). Discussion of paper by P. K. Andersen, Ø. Borgan, R. Gill and N. Keiding. *Internat. Statist. Rev.* **50**, 244–246.

Andersen, P. K., Borgan, Ø., Gill, R. D. & Keiding, N. (1982). Linear nonparametric tests for comparison of counting processes, with applications to censored survival data (with discussion). *Internat. Statist. Rev.* **50**, 219–258.

Arjas, E. (1981). The failure and hazard processes in multivariate reliability systems. *Math. Oper. Res.* **6**, 551–562.

Bremaud, P. (1981). *Point processes and queues, martingale dynamics*. Springer, New York.

Bremaud, P. & Jacod, J. (1977). Processus ponctuels et martingales: Résultats récents sur la modélisation et le filtrage. *Adv. in Appl. Probab.* **9**, 362–416.

Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34**, 187–200.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.

Davis, M. H. A. (1976). The representation of martingale of jump processes. *SIAM J. Control Optim.* **14**, 623–638.
Dollard, J. D. & Friedman, C. N. (1979). *Product integration*. Addison-Wesley, London.
Efron, B. (1967). The two sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Probability* **4**, pp. 831–853. Univ. of California Press.
Gill, R. D. (1980). *Censoring and stochastic integrals*. M. C. Tracts, Mathematical Centre, Amsterdam.
Hendry, D. F. & Richard, J. F. (1983). The econometric analysis of econometric time series. *Internat. Statist. Rev.* **51**, 111–165.
Jacobsen, M. (1983). Statistical analysis of counting processes. Lecture Notes in Statist. **12**. Springer-Verlag, New York.
Jacod, J. (1975). Multivariate point processes; Predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrsch. Verw. Gebiete* **31**, 235–253.
Johansen, S. (1983). An extension of Cox's regression model. *Internat. Statist. Rev.* **51**, 165–174.
Kalbfleisch, J. D. & MacKay, R. J. (1979). On constant-sum models for censored survival data. *Biometrica* **66**, 87–90.
Kalbfleisch, J. D. & Prentice, R. L. (1980). *The statistical analysis of failure time data*. Wiley, New York.
Lagakos, S. (1979). General right censoring and its impact on the analysis of survival data. *Biometrics* **35**, 139–156.
Liptser, R. & Shiryayev, A. (1978). *Statistics of random processes*. Vol. II, Applications. Springer, Heidelberg.
Williams, J. S. & Lagakos, S. (1977). Models for censored survival analysis: constant-sum and variable sum models. *Biometrica* **64**, 215–224.

Received November 1983, in final form October 1984

Elja Arjas
University of Oulu
Dept. of Applied Math. and Stat.
SF-90570 Oulu 57
Finland

14–848194