Scand J Statist 19: 111–132, 1992

Observation Scheme and Likelihood

ELJA ARJAS and PENTTI HAARA

University of Oulu

ABSTRACT. In the statistical analysis of longitudinal data, it is useful to distinguish between different categories of information that data points may represent. In simple terms, some points may represent the actual development of interest, whereas others can be viewed to be consequences of the observation scheme or protocol followed. In addition, data may contain indirect information about variables which are of interest but not directly observable. The purpose of this paper is to provide an exact mathematical framework in which these general questions can be discussed, and give conditions under which the relevant information to the statistical problem, in the form of a likelihood expression, can be extracted from the data. The general form of the likelihood is given in the case where there are several individuals who, apart from the control of the protocol, behave independently. As an example we consider animal carcinogenicity experiments involving occult tumors and serial sacrificing.

Key words: marked point process, filtering, event history analysis, likelihood, censoring, survival/sacrifice experiment

1. Introduction

Longitudinal data are conveniently viewed as a sample path of a marked point process (abbr. MPP). In principle, this interpretation offers enormous flexibility in the modeling because the way the marked points are chosen, in order to arrive at an realistic and scientifically well motivated description, depends entirely upon the investigator. There is also a well developed dynamical theory describing the time evolution of such processes. Only mild mathematical regularity properties need to be satisfied.

This is not the entire truth, however. Observed longitudinal data may come from many different sources, and their roles for statistical inference may be different. In particular, while most marked points describe the development of actual interest, others may arise directly from the observation scheme, or come from some other external source. Therefore they will not be of direct inferential interest given the objectives of the study. Right censored survival times are a classical example: A model describing the censoring mechanism will typically have the same roles as nuisance parameters have in parametric models; they themselves are not of interest, but cannot be avoided in practice when a model is specified. Another example is provided by animal carcinogenicity experiments in which the actual variables of interest are the times from exposure to tumor. Often the presence of a tumor can be established only in an autopsy, and therefore serial sacrificing of the experimental animals may be necessary. Sacrificing, or more generally, periodic inspection, is an example of an observation scheme which generates both data of actual interest and undesirable but in practice unavoidable random or optional variables.

The questions are then: Under what circumstances can the information of actual inferential interest be extracted from the complete data set? If this is possible, how can it be done? We are primarily interested in likelihood-based inference and therefore these problems concern the general structure of the likelihood expression arising from an MPP. These questions led to the formation of a general condition concerning "non-innovative marks" in Arjas & Haara (1984). This, to our knowledge, was the first mathematically precise way to define the concept of non-informative censoring. The aim of this paper is to further extend this analysis into general algorithmic study plans, perhaps also involving variables which are not directly observable. The results will enable one to separate from the likelihood expression a factor arising directly from the protocol, giving to the remaining expression a mathematically simple form. In particular, we show that, under a natural independence condition regarding the natural evolution of different individuals in a follow-up, the likelihood becomes a product of "individual contributions". The proofs use in an essential way filtering ideas for MPP's. Our derivations can be viewed as an illustration of the techniques presented in section 3.2 of Arjas (1989) and, more generally, an Arjas, Haara & Norros (1991) (abbr. AHN).

The plan of this paper is as follows. In section 2 we introduce the marked point process model, distinguishing between the "full process" which describes the whole history of the experiment including latent events, and the observed process. In section 3 we use the filter formula of AHN to deduce the fundamental conditional independence properties between individuals. In section 4 we then show how the general likelihood expression of Jacod (1975) reduces to a simple product over the individuals. Finally, in section 5, a general survival/sacrifice experiment is considered as an illustration.

2. The marked point process model

We start by describing the statistical model in terms of two marked point processes: *the full process* and *the observed process*. The full process describes the entire development of the studied individuals, including underlying events such as an unobserved defect in a monitored system or a tumor onset in an experimental animal, as well as events which are primarily related to the observation scheme, such as scheduled inspections or censorings. In order to make a distinction between these different types of occurrences we shall talk about *state events* and *protocol events*, dividing the former further into *directly observable* and *latent* events.

The framework is similar to that of AHN and of section 3.2 in Arjas (1989). However, here the marks are given more structure. First, in both MPP's we distinguish between a population and an individual level. A population mark is of the form $\mathbf{x} = (x_j)_{j \in J}$, where x_j is the "individual mark" concerning individual j. Here $J := \{1, \ldots, n\}$ is the population (or sample) of individuals under study. Second, individual marks in the full process and in the observed process are of the form $x_j = (x_j', x_j'')$ and $(\hat{x}_j, x_j'')$ respectively. Here x_j' describes an event in the actual state evolution of individual j, x_j'' describes an event imposed on individual j by the observation scheme or protocol, and $\hat{x}_j$ describes what is observed directly about the current state of individual j. The mark $\hat{x}_j$ may also contain information about some previous latent development concerning j, such as a failed component detected in a periodic inspection. Thus the full process includes both the actual state evolution of the studied population and the study protocol, the latter being copied without change from the full to the observed process.

We distinguish between population and individual marks by using boldface letters for the former. This applies also to the notation for mark spaces and mark-valued random elements.

For stochastic processes $X := (X_t; t \geqslant 0)$ and $Y := (Y_t; t \geqslant 0)$, whose sample paths are increasing or of bounded variation on finite intervals, we often use the shorter "measure notation" $dX_t = dY_t$, $t \geqslant 0$, to mean that X and Y have the same sample paths outside a null set. For unexplained notation see AHN, and for unexplained terminology on the general theory of stochastic processes see the two volumes of Dellacherie & Meyer (1978, 1982).

2.1. *The full process*

We first define the MPP that describes the entire development of all n individuals we want to model. Let E' be the assortment of marks x' describing the state events a single individual may experience during the follow-up. We assume that E' is finite and split it into two parts by writing

$$E' = E'_0 + E'_1. \tag{2.1}$$

The sets E'_0 and E'_1 are now given the following interpretations:

—E'_0 is the set of (latent) events in the state development that may be observed only afterwards, and then often only partially, in connection with some observable event concerning the same individual (e.g. the onset of a tumor whose presence can only be determined in an autopsy);
—E'_1 is the set of directly observable events in the state development (e.g. death from a disease).

In addition we have the finite collection E'' of individual marks x'', called *protocol marks*, corresponding to the events that either the protocol may inflict on an individual or that are of secondary interest (such as censoring due to the manifestation of a competing risk). Such events are obviously always observable. For example, a mark indicating the right censoring of an individual would be an element in E''.

Typically only a subset of the individuals in J experience something at any given marked point. This brings up the need for some extra notation. First we complete the individual mark spaces by adding the empty mark "$\emptyset$". We indicate this by an upper bar, e.g. $\bar{E}' := E' \cup \{\emptyset\}$. We shall also use the symbol $\boldsymbol{\emptyset}_n$ to mean the n-tuple whose every element is $\emptyset$. More generally, we abbreviate $\boldsymbol{\emptyset}_{njx'} := (\emptyset, \ldots, \emptyset, x', \emptyset, \ldots, \emptyset)$, where x' appears in the j-th coordinate. Finally, we denote by $\mathbf{E}'' := \{\mathbf{x}'' | \mathbf{x}'' := (x''_1, \ldots, x''_n) \in E''^n, \mathbf{x}'' \neq \boldsymbol{\emptyset}_n\}$ the set of population level protocol marks and add the upper bar to the notation if we want the mark $\boldsymbol{\emptyset}_n$ to be included in the set.

We are now ready to give the formal definition of the full process.

Definition 2.1
The full process is the canonical marked point process

$$(\mathbf{T}, \mathbf{X}) := (T_i, \mathbf{X}_i)_{i \geqslant 1}$$

with mark space

$$\mathbf{E} := \{\mathbf{x} = (x'_1, \ldots, x'_n, \mathbf{x}'') \,|\mathbf{x} \in \bar{E}'^n \times \bar{\mathbf{E}}'', \mathbf{x} \neq \boldsymbol{\emptyset}_{2n}\},$$

defined on the canonical path space of $\mathbf{E}$-marked points

$$\Omega := \{(t_i, \mathbf{x}_i)_{i \geqslant 1} | 0 < t_1 \leqslant t_2 \leqslant \ldots \uparrow +\infty; t_i < \infty \Rightarrow t_{i+1} > t_i \& \mathbf{x}_i \in \mathbf{E};\ t_i = \infty \Rightarrow \mathbf{x}_i = \Delta\}.$$

(Here $\Delta \notin \mathbf{E}$ is a fictitious mark.)

The process $(\mathbf{T}, \mathbf{X})$ is thus the coordinate process of Ω: $T_i(\omega) = t_i$ and $\mathbf{X}_i(\omega) = \mathbf{x}_i$, $i \geqslant 1$, when $\omega := (t_i, \mathbf{x}_i)_{i \geqslant 1}$. The mark space $\mathbf{E}$ is finite. The auxiliary mark $\emptyset$ has the obvious interpretation that nothing happens in that coordinate: $X'_{ij} = \emptyset$ (resp. $X''_{ij} = \emptyset$) if j experiences no state (resp. protocol) event at T_i.

An equivalent representation for the full process is given by the counting processes

$$N_t(\mathbf{x}) := \sum_{i \geqslant 1} 1_{\{T_i \leqslant t, \mathbf{X}_i = \mathbf{x}\}}, \quad t \geqslant 0, \quad \mathbf{x} \in \mathbf{E}.$$

The internal history $\mathbb{F} := (\mathscr{F}_t; t \geqslant 0)$ of the full process is defined by

$$\mathscr{F}_t := \sigma(\{N_s(\mathbf{x}) | 0 \leqslant s \leqslant t, \mathbf{x} \in \mathbf{E}\}), \quad t \geqslant 0.$$

We shall also need the individual state histories $\mathbb{F}'_j := (\mathscr{F}'_{jt}; t \geqslant 0)$, $j \in J$, defined by the counting processes

$$N_{jt}(x') := \sum_{\mathbf{x}:\, x_j = x'} N_t(\mathbf{x}), \tag{2.2}$$

via

$$\mathscr{F}'_{jt} := \sigma(\{N_{js}(x') \mid 0 \leqslant s \leqslant t, x' \in E'\}).$$

2.2. The observed process

Let $\hat{E}$ be the countable set of possible observations that can be made about a single individual. A mark $\hat{x} \in \hat{E}$ may include a directly observable state mark $x' \in E'_1$ and/or contain some informtion about the latent events that the individual in question experienced earlier. The presence of a tumor, which could be established in an autopsy, would be in this latter category. We now assume the following two conditions:

(C1) For each observed individual mark $(\hat{x}, x'')$ there is in the full process a unique individual mark $(p(\hat{x}), x'')$ that can generate $(\hat{x}, x'')$. The function p has its values in $\bar{E}'_1$, and $p(\emptyset) = \emptyset$.

Condition C1 is only an expression of the idea that the marks in E'_1 and E'' are directly observable (and those in E'_0 are not), and $p(\hat{x})$ is the directly observable part in $\hat{x}$. Obviously $p^{-1}(x') = \emptyset$ for $x' \in E'_0$. Which particular mark $\hat{x} \in p^{-1}(x')$ is then actually observed when (x', x'') occurs depends on the history of the individual through an indicator link function:

(C2) For each $j \in J$ and $(\hat{x}, x'') \in \bar{\hat{E}} \times \bar{E}''$ there exists a 0–1-valued process $(I_{jt}(\hat{x}, x''); t \geqslant 0)$ such that

(i) $(I_{jt}(\hat{x}, x''); t \geqslant 0)$ is $\mathbb{F}'_j$-predictable;
(ii) If $x' \in \bar{E}'_1$ then

$$\sum_{\hat{x} \in p^{-1}(x')} I_{jt}(\hat{x}, x'') = 1, \quad t \geqslant 0,$$

for each (j, x'');
(iii) $I_{jt}(\emptyset, \emptyset) = 1$, $t \geqslant 0$, for each j.

Thus $I_{jt}(\hat{x}, x'') = 1$ if the evolution of individual j is "at risk at time t of producing observation $\hat{x}$ about j". Condition C2(iii) implies that a retrospective finding about the latent events of a particular individual is possible only in connection with a directly observable state event or a protocol event for the same individual. Now we are ready to give the construction of the observed process (cf. Arjas, 1989: 3.2.1.(2)):

Definition 2.2

The observed process is the marked point process $(\hat{\mathbf{T}}, \hat{\mathbf{X}}) := (\hat{T}_i, \hat{\mathbf{X}}_i)_{i \geqslant 1}$ with mark space

$$\hat{\mathbf{E}} := \{\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_n, \mathbf{x}'') \mid \hat{\mathbf{x}} \in \bar{\hat{E}}^n \times \bar{\mathbf{E}}'', \hat{\mathbf{x}} \neq \boldsymbol{\emptyset}_{2n}\},$$

derived from $(\mathbf{T}, \mathbf{X})$ via the counting processes

$$\hat{N}_t(\hat{\mathbf{x}}) := \int_0^t \left[\prod_{j \in J} I_{js}(\hat{x}_j, x''_j)\right] dN_s(\mathbf{p}(\hat{\mathbf{x}}), \mathbf{x}''), \; t \geqslant 0, \hat{\mathbf{x}} \in \hat{\mathbf{E}}. \tag{2.3}$$

Here we denote $\mathbf{p}(\hat{\mathbf{x}}) := (p(\hat{x}_j))_{j \in J}$.

Let

$$\hat{\mathbb{F}} := (\hat{\mathscr{F}}_t, t \geqslant 0)$$

be the internal history of $(\hat{\mathbf{T}}, \hat{\mathbf{X}})$. The following summarizes the intuitive content of definition 2.2: The transformation from $(\mathbf{T}, \mathbf{X})$ to $(\hat{\mathbf{T}}, \hat{\mathbf{X}})$ ignores the latent marks in E'_0, copies the directly observable marks in E'_1 and E'' without change, and supplements the former with possible retrospective findings about the latent marks. Let

$$\hat{N}_{jt}(\hat{x}, x'') := \sum \hat{N}_t(\hat{\mathbf{x}}),\ t \geqslant 0, \tag{2.4}$$

where the summation is over those $\hat{\mathbf{x}} \in \hat{\mathbf{E}}$ for which $(\hat{x}_j, x''_j) = (\hat{x}, x'')$. Then the counting processes $(\hat{N}_{jt}(\hat{x}, x''); t \geqslant 0)$, $(\hat{x}, x'') \neq \emptyset_2$, define the observed process about individual j. We let $\hat{\mathbb{F}}_j := (\hat{\mathscr{F}}_{jt}; t \geqslant 0)$ be the corresponding internal history. Note that the definition in (2.3) is coordinate-wise: the observed process about j is completely determined by the development in the full process which concerns individual j.

2.3. *The probability model*

We shall define our probability model in terms of the compensators of the counting process $N(\mathbf{x})$, $\mathbf{x} \in \mathbf{E}$. This involves two main ingredients. One is that for each individual $j \in J$ there are intensity processes $a_j(x')$, $x' \in E'$, which form a simple non-parametric model of the "natural development" of individual j. As a collective, these intensities define an absolutely continuous probability model for a reference experiment where the individuals behave independently of each other and where there are no protocol events (such as censoring or sacrifices, cf. Arjas & Haara (1984)). The latent events are allowed to influence the subsequent development of individual j, however.

The second ingredient of the probability model are the processes $B(\mathbf{x})$, $\mathbf{x} \in \mathbf{E}$, that model the observation scheme. They are allowed to depend on the entire past observed history, and the individuals need not be treated independently in the scheme. For example, the protocol could require type II censorship, or it could state that every time there is a directly observable failure in the studied population the remaining individuals are checked for the presence of some latent feature. Furthermore, the corresponding probability distributions may have a discrete component. This leads us to the following structure.

The model assumptions

Let the probability $\mathbb{P}$ on the measurable space $(\Omega, \mathscr{F}_\infty)$ be such that the $(\mathbb{P}, \mathbb{F})$-compensator of each counting process $(N_t(\mathbf{x}); t \geqslant 0)$, with $\mathbf{x} \in \mathbf{E}$, is of the form

$$d\Lambda_t(\mathbf{x}) := \left[\prod_{j \in J} a_{jt}(x'_j) \right] dB_t(\mathbf{x}). \tag{2.5}$$

More specifically, we assume that $\Sigma_{\mathbf{x} \in \mathbf{E}} \Lambda_t(\mathbf{x}) < \infty$ for all $t \geqslant 0$, and that the following conditions are satisfied.

(A) For each $x' \in \bar{E}'$ the process $a_j(x')$ is $\mathbb{F}'_j$-adapted, has sample paths that are non-negative-valued, uniformly bounded on finite intervals, and left-continuous with right-hand limits. In addition we assume

$$a_{jt}(\emptyset) = 1, \quad t \geqslant 0. \tag{2.6}$$

(B) For each $\mathbf{x} \in \mathbf{E}$ the process $B(\mathbf{x})$ is $\hat{\mathbb{F}}$-predictable and has sample paths which start from zero, are non-negative non-decreasing and bounded on finite intervals, and do not have a singular continuous part. We denote by $b(\mathbf{x})$ the intensity process of the absolutely continuous part of $B(\mathbf{x})$. The process $b(\mathbf{x})$ is assumed to be $\hat{\mathbb{F}}$-predictable and bounded on finite intervals. We further assume

(i) if $x'_j \neq \emptyset$ for some j, then the jump part of $B(\mathbf{x})$ vanishes;
(ii) if $x'_j \neq \emptyset$ for more than one j, then $B(\mathbf{x})$ vanishes;
(iii) for any $(i, j) \in J^2$ such that $x'_i \in E'_0$ and $x''_j \neq \emptyset$, the process $B(\mathbf{x})$ vanishes;
(iv) for each $x' \in E'$ the process $\delta_j(x')$ is 0–1-valued and $\hat{\mathbb{F}}_j$-predictable, where

$$\delta_{jt}(x') := \sum b_t(\mathbf{x})$$

and the summation is over those $\mathbf{x} \in \mathbf{E}$ for which $x'_j = x'$. We complete all σ-algebras on Ω with the $\mathscr{F}_\infty$-null sets.

Remarks. 1. Once the processes $a_j(x')$, $(j, x') \in J \times E'$, and $B(\mathbf{x})$, $x \in \mathbf{E}$, are specified, they determine uniquely a probability $\mathbb{P}$ on $(\Omega, \mathscr{F}_\infty)$ in such a way that (2.5) with A and B holds. This probability is such that the set of the exploding sample paths is a null set.

2. Conditions B(i) and B(ii) imply that the individual state events describing the natural development have absolutely continuous risks and they cannot occur simultaneously (with positive probability). This, together with the requirement that each $a_j(x')$ is $\mathbb{F}'_j$-adapted, constitutes the fundamental independence assumption leading to the product distribution form in theorem 3.1.

3. Condition B(iii) excludes the possibility that protocol events could occur simultaneously with a latent state event.

4. Of course, protocol events can be attached to directly observable state events. By condition B(iv), formula (2.5) corresponds to 4.3.(3) in Arjas (1989) when $\mathbf{x}$ is such that $x'_k \neq \emptyset$ for some k. In the terminology of Arjas (1989) we can then view x'_k as a pre-mark, whereas $b_t(\mathbf{x})$ is the conditional probability of obtaining the protocol events $(x''_j)_{j \in J}$ given that x'_k is realized for individual k.

5. We have $\delta_{jt}(x') = 0$ whenever the observed pre-t history implies that individual j cannot experience x' at t. In other words, the set $R_t(x') := \{j \in J \mid \delta_{jt}(x') = 1\}$ can be interpreted as the observed risk set for state event x'. This interpretation is valid also for the latent events $x' \in E'_0$. (Just consider individuals which are known to have died.) On the other hand, the appearance of a latent event, say z' at time s, in the state history of j can have the effect that the $\mathbb{F}$-risk of j experiencing x' at $t \geqslant s$ becomes zero, i.e. $a_{jt}(x') = 0$. However, j may stay in $R_t(x')$ (and the observation based risk for j experiencing x' positive) until the occurrence of z' is observed.

6. The cartesian product mark space $\mathbf{E}$ is overly large in the sense that it includes many mark combinations whose appearance is restricted by the above model assumptions to a null subset of the path space (see e.g. remark 2 above). However, it fits well to the product structure of the independent "natural development processes" of the individuals. It also allows for the possible generalization where there may be ties between individual state events but where a product form is assumed about the corresponding compensator increments, thus retaining the independence between the individual state histories.

3. The conditional distributions of the individual state histories given the observed history

In order to define the conditional distributions of the histories we introduce some additional concepts and notation. The mark space $\mathbf{E}$ defines the set $\mathbb{H}$ of history sets H, which are finite subsets of $\mathbb{R}_+ \times \mathbf{E}$ such that $(t, \mathbf{x}_1) \in H$ and $(t, \mathbf{x}_2) \in H$ imply $\mathbf{x}_1 = \mathbf{x}_2$ (cf. Norros, 1986). We equip $\mathbb{H}$ with a topology which makes it a Polish space, and with the corresponding Borel sets $\mathscr{H}$ (cf. AHN). Outside the $\mathbb{P}$-null set of exploding sample paths the $\mathbb{H}$-valued pre-t

histories of $(\mathbf{T}, \mathbf{X})$ correspond to the history process $(H_t; t \geqslant 0)$, defined by

$$H_t(\omega) := \{(T_i(\omega), \mathbf{X}_i(\omega)) \mid T_i(\omega) \leqslant t\}, \tag{3.1a}$$

and its left continuous version $(H_{t-}; t \geqslant 0)$, defined by

$$H_{t-}(\omega) := \{(T_i(\omega), \mathbf{X}_i(\omega)) \mid T_i(\omega) < t\}. \tag{3.1b}$$

For each $\mathbb{F}$-predictable process Z we can use the representation

$$Z_t(\omega) := Z(t; H_{t-}(\omega)), \quad t \geqslant 0, \quad \omega \in \Omega, \tag{3.2}$$

where on the right-hand side Z appears as a jointly measurable function of (t, H) (cf. AHN: 2.3). We continue to use the same symbol for a process and a corresponding function of (t, H), but the position of the time-parameter indicates which one of these two is meant.

The concepts and relationships above are valid for any MPP with a countable mark space. In the sequel we need the analogous concepts for mark space E', which we accordingly indicate by adding a prime (′) to the upper right corner of the corresponding symbols. More precisely, for each $j \in J$ we need the individual state history process $(H'_{jt}; t \geqslant 0)$ defined in a similar manner as in (3.1) by the E'-marked point process corresponding to the counting processes (2.2). Note that the σ-fields $\mathscr{F}_t$, $\mathscr{F}_{t-}$, $\mathscr{F}'_{jt}$ and $\mathscr{F}'_{jt-}$ are generated by the random elements H_t, H_{t-}, H'_{jt} and H'_{jt-}, respectively.

For given $j \in J$ and $H \in \mathbb{H}$ we define

$$h'_j(H) := \{(t, x'_j) \mid (t, \mathbf{x}) \in H, x'_j \neq \emptyset\}.$$

Then as a direct consequence of the definitions we get

Lemma 3.1

For each $j \in J$, the mapping h'_j from $(\mathbb{H}, \mathscr{H})$ to $(\mathbb{H}', \mathscr{H}')$ is measurable and for each $t \geqslant 0$ it satisfies the relationships $H'_{jt} = h'_j(H_t)$, and $H'_{jt-} = h'_j(H_{t-})$.

Thus h'_j is a projection mapping that extracts from the full history the state history of individual j.

We now state the main result of this section.

Theorem 3.1

Suppose that (2.5), (A), (B), (C1) *and* (C2) *hold. Then for each $t > 0$:*

(*i*) *The individual state histories H'_{jt-}, $j \in J$, are conditionally independent given the observed history $\hat{\mathscr{F}}_{t-}$; and*

(*ii*) *For each $j \in J$ the state history H'_{jt-} of individual j and the observed history $\hat{\mathscr{F}}_{t-}$ are conditionally independent given the observed history $\hat{\mathscr{F}}_{jt-}$ of individual j.*

According to theorem 3.1, in estimating the underlying history from the observations it suffices to consider each individual separately. This makes the necessary computations much simpler.

The proof of theorem 3.1 is based on the fitering theory for derived MPP's of AHN and is split into four lemmas. As an intermediate step we consider the MPP $(\tilde{\mathbf{T}}, \mathbf{Y}) := (\tilde{T}_i, \mathbf{Y}_i)_{i \geqslant 1}$ with mark space

$$\tilde{\mathbf{E}} := \{\mathbf{y} = (\hat{y}_1, y'_2, \ldots, y'_n, \mathbf{y}'') \mid \mathbf{y} \in \bar{\hat{E}} \times \bar{E}'^{n-1} \times \bar{\mathbf{E}}'', \mathbf{y} \neq \boldsymbol{\emptyset}_{2n}\},$$

defined by the counting processes

$$d\tilde{N}_t(\mathbf{y}) := I_{1t}(\hat{y}_1, y''_1)\, dN_t(p(\hat{y}_1), y'_2, \ldots, y'_n, \mathbf{y}''), \quad t \geqslant 0, \quad \mathbf{y} \in \tilde{\mathbf{E}}. \tag{3.3}$$

In the sequel we also abbreviate $\mathbf{y} = (\hat{y}_1, \mathbf{z})$ with $\mathbf{z} := (y'_2, \ldots, y'_n, \mathbf{y}'')$. Let $\mathbb{G} := (\mathcal{G}_t; t \geq 0)$ be the internal history of $(\tilde{\mathbf{T}}, \mathbf{Y})$. The history $\mathbb{G}$ reveals all events in the full process except for the latent events concerning individual 1. Individual 1 is chosen here for the sake of notational convenience only.

Consider the conditional probabilities $(\tilde{\pi}_t; t \geq 0)$ on $(\mathbb{H}, \mathcal{H})$ and $(\pi_t; t \geq 0)$ on $(\mathbb{H}', \mathcal{H}')$ defined by

$$\tilde{\pi}_t(B) := \mathbb{P}(H_t \in B \mid \mathcal{G}_t), \quad B \in \mathcal{H},$$

$$\pi_t(A) := \mathbb{P}(H'_{1t} \in A \mid \mathcal{G}_t), \quad A \in \mathcal{H}'.$$

It is a straightforward consequence of our model assumptions and (3.3) that the system of the two MPP's $(\mathbf{T}, \mathbf{X})$ and $(\tilde{\mathbf{T}}, \mathbf{Y})$ satisfies the conditions (i)–(iii) of AHN. Therefore there exists a regular version of $(\tilde{\pi}_t; t \geq 0)$ for which lemma 1 of AHN holds. This and the relationship

$$\pi_t(A) = \tilde{\pi}_t(h_1'^{-1}(A)), \quad A \in \mathcal{H}', \quad t \geq 0, \tag{3.4}$$

imply that we can choose a regular version of $(\pi_t; t \geq 0)$ satisfying:

(a) π_t is a regular conditional distribution (r.c.d.) of H'_{1t} given $\mathcal{G}_t$;
(b) π_t, $t \geq 0$, is a cadlag process on the space $\mathcal{P}(\mathcal{H}')$ of probability measures on $\mathbb{H}'$ in the sense that, outside a common null set, $\pi_t(A)$, $t \geq 0$, has cadlag sample paths for each $A \in \mathcal{H}'$;
(c) for each $t > 0$, the left-hand limit π_{t-} is a r.c.d. of H'_{1t-} given $\mathcal{G}_{t-}$.

The application of lemma 3.1 to the representation (3.2) for a bounded or non-negative $\mathbb{F}'_1$-predictable process Z gives $Z_t(\omega) = Z_t(t, h'_1(H_{t-}(\omega)))$, which implies by (3.4) and (c) above that

$$\int_{\mathbb{H}} \tilde{\pi}_{t-}(dH) Z(t; h'_1(H)) = \int_{\mathbb{H}'} \pi_{t-}(dH') Z(t; H') = \mathbb{E}(Z_t \mid \mathcal{G}_{t-}). \tag{3.5}$$

By conditions A and C2(i) the processes $a_1(x')$, $x' \in E'$, and $I_1(\hat{x}, x'')$, $(\hat{x}, x'') \in \bar{\hat{E}} \times \bar{E}''$, are $\mathbb{F}'_1$-predictable. Therefore there exist $\mathcal{R}_+ \times \mathcal{H}'$-measurable mappings $(t, H') \to a_1(t, x'; H')$ and $(t, H') \to I_1(t, \hat{x}, x''; H')$ such that $a_{1t}(x') = a_1(t, x'; H'_{1t-})$ and $I_{1t}(\hat{x}, x'') = I_1(t, \hat{x}, x''; H'_{1t-})$. Following AHN we shall use the notation $\mu[f]$ for the integrals $\int f\, d\mu$ when the integration is over $\mathbb{H}$ or $\mathbb{H}'$.

Lemma 3.2

The $(\mathbb{P}, \mathbb{G})$-compensator $\tilde{\Lambda}(\mathbf{y})$ of the counting process $\tilde{N}(\mathbf{y})$ is given by

$$\tilde{\Lambda}_t(\mathbf{y}) = \int_0^t \pi_{s-}[\tilde{a}_1(s, \hat{y}_1, y''_1)] \left\{ \prod_{j \neq 1} a_{js}(y'_j) \right\} dB_s(p(\hat{y}_1), \mathbf{z}), \quad t \geq 0, \tag{3.6}$$

where $\tilde{a}_1(t, \hat{x}, x''; H') := I_1(t, \hat{x}, x''; H') a_1(t, p(\hat{x}); H')$ for $(\hat{x}, x'') \in \bar{\hat{E}} \times \bar{E}''$.

Proof. From (3.3) and (2.5) we get for the $(\mathbb{P}, \mathbb{F})$-compensator $\Lambda(\mathbf{y})$ of $\tilde{N}(\mathbf{y})$ the representation

$$d\Lambda_t(\mathbf{y}) = I_{1t}(\hat{y}_1, y''_1)\, d\Lambda_t(p(\hat{y}_1), \mathbf{z}) = \tilde{a}_{1t}(\hat{y}_1, y''_1) \left\{ \prod_{j \neq 1} a_{jt}(y'_j) \right\} dB_t(p(\hat{y}), \mathbf{z}).$$

Using 3.2.2.(4) of Arjas (1989) and the $\mathbb{G}$-predictability of the processes $a_j(x')$, $x' \in E'$, $j \neq 1$, and $B(\mathbf{x})$, $\mathbf{x} \in \mathbf{E}$, we get from above

$$d\tilde{\Lambda}_t(\mathbf{y}) = \mathbb{E}(\tilde{a}_{1t}(\hat{y}_1, y_1'') \mid \mathscr{G}_{t-}) \left\{ \prod_{j \neq 1} a_{jt}(y_j') \right\} dB_t(p(\hat{y}), \mathbf{z}).$$

Now (3.5) implies (3.6). □

Remark. Our method of proof is rather informal. See the remark following the proof of proposition 4.1.

Next we give a filter formula which is a straightforward application of the more general formula in theorem 1 of AHN to our model. However, the derivation requires a tedious calculation. Therefore the proofs of the next three lemmas are in the appendix.

Following AHN we abbreviate $g(t, x', A; H') := 1_A(H' \cup \{(t, x')\})$ ($:= 1_A(H')$ if $x' = \emptyset$) and $d(t, x', A; H') := g(t, x', A; H') - 1_A(H')$. Thus $d(t, x', A; H') = 1$, -1 or 0 according to whether the (individual state) history H' moves into, out of, or stays inside/outside the set $A \in \mathscr{H}'$, respectively, when a new event $x' \in E'$ is included in it at time t. Recall from (2.4) the definition of the counting process $\hat{N}_1(\hat{x}, x'')$. We also adopt the convention $\frac{0}{0} := 0$.

Lemma 3.3

Outside a $\mathbb{P}$-null set, common to all $A \in \mathscr{H}'$, one has

$$\begin{aligned}\pi_t(A) = 1_A(\emptyset) &+ \sum_{x' \in E_0'} \int_0^t \pi_{s-}[a_1(s, x')\, d(s, x', A)]\, \delta_{1s}(x')\, ds \\ &+ \sum_{x' \in E_1'} \int_0^t \{\pi_{s-}[a_1(s, x')]\pi_{s-}(A) - \pi_{s-}[a_1(s, x') 1_A]\}\, \delta_{1s}(x')\, ds \\ &+ \sum_{(\hat{x}, x'') \neq \emptyset_2} \int_0^t C_{1s}(\hat{x}, x'')\, d\hat{N}_{1s}(\hat{x}, x''), \quad t \geqslant 0,\end{aligned} \tag{3.7a}$$

where the innovation gain is given by

$$C_{1t}(\hat{x}, x'') := \frac{\pi_{t-}[\tilde{a}_1(t, \hat{x}, x'') g(t, p(\hat{x}), A)]}{\pi_{t-}[\tilde{a}_1(t, \hat{x}, x'')]} - \pi_{t-}(A). \tag{3.7b}$$

Note that (3.7) depends on the protocol processes $B(\mathbf{x})$, $\mathbf{x} \in \mathbf{E}$, only through the risk indicator processes $\delta_1(x')$, $x' \in E'$, of individual 1. This is essentially a consequence of the $\hat{\mathbb{F}}$-predictability of the processes $B(\mathbf{x})$, $\mathbf{x} \in \mathbf{E}$, and the assumption B(iv). It explains why the assumed independence between individual state histories is retained after conditioning on the observed history.

In lemmas 3.4 and 3.5 below we show that $(\pi_t; t \geqslant 0)$ depends on ω only through what is observed about individual 1. Recall that $\hat{\mathscr{F}}_{1t}$ is the σ-field generated by the pre-t observed history of individual 1, i.e. $\hat{\mathscr{F}}_{1t} = \sigma(\hat{H}_{1t})$ where

$$\hat{H}_{1t}(\omega) := \{(s, \hat{x}, x'') \mid \Delta\hat{N}_{1s}(\omega, \hat{x}, x'') = 1, s \leqslant t, (\hat{x}, x'') \neq \emptyset_2\}.$$

The proofs of these lemmas are based on (3.7) and the equivalence class characterization of measurability (cf. Jacobsen, 1982, p. 7).

Lemma 3.4

Let $t \in [0, \infty)$, and let $\omega_1 \neq \omega_2$ be such that:

(*i*) $\hat{H}_{1s}(\omega_1) = \hat{H}_{1s}(\omega_2)$ *for* $s \leqslant t$;

(*ii*) *the conditional distributions* $\pi_t(\omega_1; \cdot)$ *and* $\pi_t(\omega_2; \cdot)$ *agree on* $(\mathbb{H}', \mathscr{H}')$.

Then the conditional distributions $\pi_u(\omega_1, \cdot)$ *and* $\pi_u(\omega_2; \cdot)$ *agree on* $(\mathbb{H}', \mathscr{H}')$ *for all* $u \in [t, V(\omega_1) \wedge V(\omega_2))$, *where*

$$V(\omega) := \inf\{\tilde{T}_i(\omega) \mid \tilde{T}_i(\omega) > t\}, \quad \omega \in \Omega.$$

Lemma 3.5
For each $A \in \mathscr{H}'$ *the process* $(\pi_t(A); t \geq 0)$ *is* $\hat{\mathbb{F}}_1$*-adapted.*

The proof theorem 3.1. Let $t \in (0, \infty)$ and let $g_j: \mathbb{H}' \to \mathbb{R}$, $j \in J$, be bounded $\mathscr{H}'$-measurable functions. Denote by $\mathscr{G}_{jt-}$ the σ-field which is generated by the pre-t-evolution of the full process except for possible latent events concerning individual j. (Thus $\mathscr{G}_{1t-} = \mathscr{G}_{t-}$ in lemmas 3.2–3.5, and the other $\mathscr{G}_{jt-}$'s are defined analogously.) By the regularity properties (a)–(c) of $(\pi_s; s \geq 0)$ and lemma 3.5, we see that $(\pi_{s-}(A); s > 0)$ is $\hat{\mathbb{F}}_1$-predictable for each $A \in \mathscr{H}'$. This means that the state history H'_{1t-} of individual 1 and the history $\mathscr{G}_{1t-}$ are conditionally independent given the observed history $\hat{\mathscr{F}}_{1t-}$ of individual 1. By symmetry between individuals the analogous result holds also for $j \neq 1$. Therefore

$$\mathbb{E}(g_j(H'_{jt-}) \mid \mathscr{G}_{jt-}) = \mathbb{E}(g_j(H'_{jt-}) \mid \hat{\mathscr{F}}_{jt-}), \quad j \in J. \tag{3.8}$$

Note that since $\hat{\mathscr{F}}_{t-} \subset \mathscr{G}_{jt-}$, $j \in J$, we have in fact already proven part (ii) of theorem 3.1.

For each $k \in J$ we have

$$\begin{aligned}
\mathbb{E}\left(\prod_{j=k}^{n} g_j(H'_{jt-}) \mid \hat{\mathscr{F}}_{t-}\right) &= \mathbb{E}\left(\mathbb{E}\left(\prod_{j=k}^{n} g_j(H'_{jt-}) \mid \mathscr{G}_{kt-}\right) \mid \hat{\mathscr{F}}_{t-}\right) \\
&= \mathbb{E}\left(\prod_{j=k+1}^{n} g_j(H'_{jt-}) \mathbb{E}(g_k(H'_{kt-}) \mid \mathscr{G}_{kt-}) \mid \hat{\mathscr{F}}_{t-}\right) \\
&= \mathbb{E}\left(\prod_{j=k+1}^{n} g_j(H'_{jt-}) \mathbb{E}(g_k(H'_{kt-}) \mid \hat{\mathscr{F}}_{kt} -) \mid \hat{\mathscr{F}}_{t-}\right) \\
&= \mathbb{E}(g_k(H'_{kt-}) \mid \hat{\mathscr{F}}_{kt-}) \mathbb{E}\left(\prod_{j=k+1}^{n} g_j(H'_{jt-}) \mid \hat{\mathscr{F}}_{t-}\right)
\end{aligned} \tag{3.9}$$

By a simple interation of (3.9) we get the required conditional independence property:

$$\mathbb{E}\left(\prod_{j=1}^{n} g_j(H'_{jt-}) \mid \hat{\mathscr{F}}_{t-}\right) = \prod_{j=1}^{n} \mathbb{E}(g_j(H'_{jt-}) \mid \hat{\mathscr{F}}_{jt-}).$$

□

4. Likelihood of the observation $(\hat{\mathbf{T}}, \hat{\mathbf{X}})$

In this section we derive the likelihood expression corresponding to the observed process $(\hat{\mathbf{T}}, \hat{\mathbf{X}})$. We consider the intensity processes $a_j(x')$, $j \in J$, $x' \in E'$, to involve an unknown quantity θ of interest which we want to estimate. On the other hand each process $B(\mathbf{x})$, $\mathbf{x} \in \mathbf{E}$, is considered as a known function determined progressively by the previous observations. It turns out that under the assumptions of theorem 1 the likelihood expression arising from the observed process factors into the product of two terms: One factor is the product over individuals $j \in J$ of terms such that the j-th term is determined solely by the observations about individual j. The other factor does not involve the intensity processes $a_j(x')$ at all and appears therefore in likelihood-based inference only as a constant proportionality factor. Thus a familiar product-form likelihood is attainable under quite complicated observation schemes.

We start from the general likelihood expression of Jacod (1975):

$$L_t = \prod_{\hat{T}_i \leqslant t} d\hat{\Lambda}_{\hat{T}_i}(\hat{\mathbf{X}}_i) \prod_{\substack{s \leqslant t \\ s \notin (\hat{T}_k)}} (1 - \Delta\bar{\hat{\Lambda}}_s) \exp\{-\bar{\hat{\Lambda}}_t^c\}, \quad 0 < t < \infty, \tag{4.1}$$

where $(\hat{\Lambda}_t(\hat{\mathbf{x}}); t \geqslant 0)$ is the $(\mathbb{P}, \hat{\mathbb{F}})$-compensator of the counting process $(\hat{N}_t(\hat{\mathbf{x}}); t \geqslant 0)$ (see e.g. Arjas, 1989, section 4.2.).

Remark. In order to define (4.1) more formally, we should assume that the probability $\mathbb{P}$ belongs to a family $\mathscr{P} := \{\mathbb{P}_\theta \mid \theta \in \Theta\}$ of probability measures that are mutually equivalent on $(\Omega, \mathscr{F}_t)$. Each $\mathbb{P}_\theta$ is assumed to satisfy our model assumptions, with the process $a_j(x')$ depending but the processes $B(\mathbf{x})$ not depending on θ. Here the set Θ may be a subset of either a finite dimensional Euclidean space (parametric model) or a more general function space (non-parametric model). Theorem 5.1 of Jacod (1975) gives an expression for the Radon–Nikodym derivatives $(d\mathbb{P}_\theta / d\mathbb{P}_{\theta'})$ on $(\Omega, \hat{\mathscr{F}}_t)$. We recover (4.1) as the numerator of this expression when we consider θ varying and θ' fixed (see e.g. Andersen *et al.*, 1988, ch. 2, for more details). However, we continue to dispense with "θ" in the notation and use (4.1) without any further reference to the family $\mathscr{P}$.

Clearly (4.1) is determined by the observations $\{(\hat{T}_i, \hat{\mathbf{X}}_i) \mid \hat{T}_i \leqslant t\}$ and the $(\mathbb{P}, \hat{\mathbb{F}})$-compensators of the counting processes $\hat{N}(\hat{\mathbf{x}})$, $\hat{\mathbf{x}} \in \hat{\mathbf{E}}$. To develop (4.1) further we need an expression for these compensators which is analogous to (3.6).

From our model assumptions and (2.3) it follows that the system of the two MPP's $(\mathbf{T}, \mathbf{X})$ and $(\hat{\mathbf{T}}, \hat{\mathbf{X}})$ satisfies the conditions (i)–(iii) of AHN. As in (3.4) we can choose for each $j \in J$ a regular version $(\hat{\pi}_{jt}; t \geqslant 0)$ of the conditional distributions

$$\hat{\pi}_{jt}(A) := \mathbb{P}(H'_{jt} \in A \mid \hat{\mathscr{F}}_{jt}), \quad A \in \mathscr{H}', \quad t \geqslant 0,$$

such that (a)–(c) of p. 118 are satisfied with π_t, H'_{1t}, $\mathscr{G}_t$ replaced by $\hat{\pi}_{jt}$, H'_{jt}, $\hat{\mathscr{F}}_{jt}$, respectively. In particular, recalling (3.2) and (3.5), one has for any $\mathbb{F}'_j$-predictable non-negative or bounded process Z

$$\hat{\pi}_{jt-}[Z(t)] = \mathbb{E}(Z_t \mid \hat{\mathscr{F}}_{jt-}). \tag{4.2}$$

Proposition 4.1

The $(\mathbb{P}, \mathbb{F})$-compensator $\hat{\Lambda}(\hat{\mathbf{x}})$ of the counting process $\hat{N}(\hat{\mathbf{x}})$ is given by

$$\hat{\Lambda}_t(\hat{\mathbf{x}}) = \int_0^t \prod_{j \in J} \hat{\pi}_{js-}[\tilde{a}_j(s, \hat{x}_j, x''_j)]\, dB_s(\mathbf{p}(\hat{\mathbf{x}}), \mathbf{x}''), \tag{4.3a}$$

where

$$\tilde{a}_j(t, \hat{x}, x''; H') := I_j(t, \hat{x}, x''; H') a_j(t, p(\hat{x}); H'). \tag{4.3b}$$

Proof. From (2.3) and (2.5) we get for the $(\mathbb{P}, \mathbb{F})$-compensator $\Lambda(\hat{\mathbf{x}})$ of $\hat{N}(\hat{\mathbf{x}})$ the representation

$$d\Lambda_t(\hat{\mathbf{x}}) = \prod_{j \in J} I_{jt}(\hat{x}_j, x''_j)\, d\Lambda_t(\mathbf{p}(\hat{\mathbf{x}}), \mathbf{x}'') = \prod_{j \in J} \tilde{a}_{jt}(\hat{x}_j, x''_j)\, dB_t(\mathbf{p}(\hat{\mathbf{x}}), \mathbf{x}'').$$

Using 3.2.2.(4) of Arjas (1989) and the $\hat{\mathbb{F}}$-predictability of the process $B(\mathbf{p}(\hat{\mathbf{x}}), \mathbf{x}'')$ we get from this

$$d\hat{\Lambda}_t(\hat{\mathbf{x}}) = \mathbb{E}\Big(\prod_{j \in J} \tilde{a}_{jt}(\hat{x}_j, x''_j) \mid \hat{\mathscr{F}}_{t-}\Big)\, dB_t(\mathbf{p}(\hat{\mathbf{x}}), \mathbf{x}'').$$

Now theorem 3.1 and (4.2) imply (4.3). □

Remark. The formal proof of proposition 4.1 is a standard check of the definition of the compensator, using the fact that $\int_0^t \prod_{j \in J} \hat{\pi}_{js-}[\tilde{a}_j(s, \hat{x}_j, x_j'')]\, dB_s(\mathbf{p}(\hat{\mathbf{x}}), \mathbf{x}'')$ is the dual $(\mathbb{P}, \hat{\mathbb{F}})$-predictable projection of the increasing process $\int_0^t \prod_{j \in J} \tilde{a}_{js}(\hat{x}_j, x_j'')\, dB_s(\mathbf{p}(\hat{\mathbf{x}}), x'')$ by our assumptions and lemma 2 of AHN (see also Brémaud & Jacod, 1977, ex. 4, p. 372).

Theorem 4.1

Under the assumptions of theorem 3.1 *and when the processes* $B(\mathbf{x}), \mathbf{x} \in \mathbf{E}$, *are considered as known functions of the data, the likelihood* (4.1) *is proportional* (*in* θ) *to the product*

$$L_t = \prod_{j \in J} L_{jt}, \tag{4.4a}$$

where

$$L_{jt} := \prod_{\hat{T}_i \leqslant t} \hat{\pi}_{j,\hat{T}_i-}[\tilde{a}_j(\hat{T}_i, \hat{X}_{ij}, X_{ij}'')] \exp\left\{-\sum_{x' \in E_1'} \int_0^t \hat{\pi}_{js-}[a_j(s, x')]\, \delta_{js}(x')\, ds\right\}. \tag{4.4b}$$

Proof. Let $S_0 := \{\hat{\mathbf{x}} \in \hat{\mathbf{E}} \mid p(\hat{x}_j) = \emptyset\, \forall j\}$ be the set of the observed marks that do not include any directly observable state events. On the other hand, from (4.3b), (2.6), C2(ii) and C2(iii) it follows that

$$\sum_{\hat{x} \in p^{-1}(\emptyset)} \tilde{a}_j(t, \hat{x}, x''; H') = 1 \tag{4.5}$$

for all j, t, x'' and H'. From (4.3a) and (4.5) it follows that

$$\begin{aligned}\sum_{\hat{\mathbf{x}} \in S_0} d\hat{\Lambda}_t(\hat{\mathbf{x}}) &= \sum_{\mathbf{x}'' \in \mathbf{E}''} \sum_{\hat{x}_1 \in p^{-1}(\emptyset)} \cdots \sum_{\hat{x}_n \in p^{-1}(\emptyset)} d\hat{\Lambda}_t(\hat{\mathbf{x}}) \\ &= \sum_{\mathbf{x}'' \in \mathbf{E}''} \prod_{j \in J} \hat{\pi}_{jt-}\left[\sum_{\hat{x} \in p^{-1}(\emptyset)} \tilde{a}_j(t, \hat{x}, x_j'')\right] dB_t(\boldsymbol{\emptyset}_n, \mathbf{x}'') \\ &= \sum_{\mathbf{x}'' \in \mathbf{E}''} dB_t(\boldsymbol{\emptyset}_n, \mathbf{x}'').\end{aligned} \tag{4.6}$$

Next we consider the corresponding sum over $\hat{\mathbf{x}} \notin S_0$. Let $S_j(x') := \{\hat{\mathbf{x}} \in \hat{\mathrm{E}} \mid p(\hat{x}_j) = x'$, and $p(\hat{x}_i) = \emptyset\, \forall i \neq j\}$ be the set of observed population marks that include the directly observable state mark $x' \in E_1'$ for individual j and no directly observable state marks for the other individuals. By condition B(ii) we have

$$\sum_{\hat{\mathbf{x}} \notin S_0} d\hat{\Lambda}_t(\hat{\mathbf{x}}) = \sum_{j \in J} \sum_{x' \in E_1'} \sum_{\hat{\mathbf{x}} \in S_j(x')} d\hat{\Lambda}_t(\hat{\mathbf{x}}).$$

Using (4.3a) and (4.5) again as in (4.6), the iterated sum above reduces to

$$\begin{aligned}\sum_{\hat{\mathbf{x}} \notin S_0} d\hat{\Lambda}_t(\hat{\mathbf{x}}) &= \sum_{j \in J} \sum_{x' \in E_1'} \sum_{\mathbf{x}'' \in \mathbf{E}''} \hat{\pi}_{jt-}\left[a_j(t, x') \sum_{\hat{x} \in p^{-1}(x')} I_j(t, \hat{x}, x_j'')\right] dB_t(\boldsymbol{\emptyset}_{njx'}, \mathbf{x}'') \\ &= \sum_{j \in J} \sum_{x' \in E_1'} \hat{\pi}_{jt-}[a_j(t, x')]\, \delta_{jt}(x')\, dt,\end{aligned} \tag{4.7}$$

where also conditions B(i) and B(iv) have been used.

From (4.3a) and B(i) it follows that $\Delta\hat{\Lambda}_t(\hat{\mathbf{x}}) > 0$ only if $\hat{\mathbf{x}} \in S_0$. By (4.6) we can thus drop the product of the terms $1 - \Delta\bar{\hat{\Lambda}}_s$ from the likelihood expression (4.1). Further, (4.6) and (4.7) imply that the factor $\exp\{-\bar{\hat{\Lambda}}_t^c\}$ is proportional to

$$\exp\left\{-\sum_{\hat{\mathbf{x}} \notin S_0} \hat{\Lambda}_t(\hat{\mathbf{x}})\right\} = \prod_{j \in J} \exp\left\{-\sum_{x' \in E_1'} \int_0^t \hat{\pi}_{js-}[a_j(s, x')]\, \delta_{js}(x')\, ds\right\}.$$

Since $\max(b_{\hat{T}_i}(\hat{\mathbf{X}}_i), \Delta B_{\hat{T}_i}(\hat{\mathbf{X}}_i)) > 0$, $i \geqslant 1$, $\mathbb{P}$-a.s., the factor $\prod_{\hat{T}_i \leqslant t} d\hat{\Lambda}_{\hat{T}_i}(\hat{\mathbf{X}}_i)$ is proportional to

$\Pi_{\hat{T}_i \leqslant t} \Pi_{j \in J} \hat{\pi}_{j, \hat{T}_i -}[\tilde{a}_j(\hat{T}_i, \hat{X}_{ij}, X''_{ij})]$ by (4.3a). The claim (4.4) follows then by regrouping the terms. □

5. An example: a model for occult tumors

As an application of our framework we consider the well-known problem concerning occult tumors in animal carcinogenicity experiments. Some animals are exposed to a carcinogenic agent, and they may or may not develop a tumor during the follow-up. The variables of primary interest in the experiment are the times to tumor onset and these are not directly observable. If an animal dies during the follow-up it is usually assumed possible to determine whether it had a tumor, and if it had, whether death followed from the tumor or from some other cause. Since it is only possible to determine the presence of an occult tumor in an autopsy, the experiment protocol typically involves serial sacrificing: Animals are killed, according to some scheme, in order to perform an autopsy. The lifetimes can also be censored from the right. Another example, concerning periodic inspections, is given in a companion paper Arjas & Haara (1992).

The full process $(\mathbf{T}, \mathbf{X})$ and the triple $(\Omega, \mathbb{F}, \mathscr{F}_\infty)$ are defined by the choice $E'_0 := \{T\}$, $E'_1 := \{DT, DO\}$ and $E'' := \{S, C\}$, where the individual events have the following interpretations (cf. McKnight & Crowley, 1984):

T—onset of tumor;
DT—death from tumor;
DO—death from some other cause than tumor;
S—sacrificing;
C—censoring.

For the observed process we choose the set of individual marks to be $\hat{E} := \{DONF, DOTF, DTTF, NF, TF\}$ giving them the following interpretations:

$DONF$—death from some other cause than tumor, no tumor found;
$DOTF$—death from some other cause than tumor, tumor found;
$DTTF$—death from tumor, tumor found (of course);
NF—no tumor found;
TF—tumor found

Here DO and DT are the directly observable state marks, and TF and NF are the possible retrospective findings about the latent event "tumor onset". Accordingly, we define the function $p: \bar{\hat{E}} \to \bar{E}'_1$ by

$$p(\hat{x}) := \begin{cases} DT & \text{if } \hat{x} = DTTF, \\ DO & \text{if } \hat{x} \in \{DOTF, DONF\}, \\ \emptyset & \text{if } \hat{x} \in \{TF, NF, \emptyset\}. \end{cases} \tag{5.1}$$

The definition of the observed process $(\hat{\mathbf{T}}, \hat{\mathbf{X}})$ is completed by specifying the indicator processes $I_j(\hat{x}, x'')$. For each $j \in J$ let

$$U_j := \inf\{t > 0 \mid N_{jt}(T) = 1\}$$

be the tumor onset time for individual j. Define for each $j \in J$ and $t > 0$:

$$I_{jt}(\hat{x}, x'') := \begin{cases} 1_{\{U_j < t\}} & \text{if } \hat{x} = DOTF \text{ or } (\hat{x}, x'') = (TF, S), \\ 1_{\{U_j \geqslant t\}} & \text{if } \hat{x} = DONF \text{ or } (\hat{x}, x'') = (NF, S), \\ 0 & \text{if } \hat{x} \in \{TF, NF\} \text{ and } x'' \in \{\emptyset, C\}, \text{ or } (\hat{x}, x'') = (\emptyset, S), \\ 1 & \text{otherwise.} \end{cases} \tag{5.2}$$

It is easy to check that conditions C2 are then satisfied. Note in particular that in order to satisfy C2(ii) we define $I_{jt}(\omega, DTTF, x'') = 1$ for all $\omega \in \Omega$ and $x'' \in \bar{E}''$, and let the probability model below assign zero-probability to the set of those sample paths ω where DT appears before T for the same individual.

In order to define the probability model for the tumor experiment we need to specify, up to an unknown parameter θ, for each individual $j \in J$ the following non-negative functions of time t:

$h_1^{j,T}(t)$—the hazard rate of tumor onset;
$h_1^{j,DO}(t)$—the hazard rate of death when no tumor is present;
$h_2^{j,DO}(t|u)$—the hazard rate of death from some other cause than tumor when tumor is present and tumor onset was at time u;
$h_2^{j,DT}(t|u)$—the hazard rate of death from tumor when tumor is present and tumor onset was at time u.

The last two functions need to be specified only for $t \geqslant u$, $u \geqslant 0$. These four risk functions are sufficient to define a probability model for an experiment where the individuals behave independently of each other, where death (DO or DT) removes an individual from the study, and where there is no censoring or sacrificing.

Next we define the corresponding intensity processes. Let

$$V_j := \inf\{t > 0 | N_{jt}(DO) + N_{jt}(DT) > 0\}$$

be the time of "natural" death of individual j. For each $j \in J$ we define the processes $a_j(x')$, $x' \in \bar{E}'$, by

$$a_{jt}(x') := \begin{cases} h_1^{j,T}(t) 1_{\{t \leqslant \min(U_j, V_j)\}} & \text{if } x' = T, \\ h_1^{j,DO}(t) 1_{\{t \leqslant \min(U_j, V_j)\}} + h_2^{j,DO}(t|U_j) 1_{\{U_j < t \leqslant V_j\}} & \text{if } x' = DO, \\ h_2^{j,DT}(t|U_j) 1_{\{U_j < t \leqslant V_j\}} & \text{if } x' = DT, \\ 1 & \text{if } x' = \emptyset. \end{cases} \tag{5.3}$$

Obviously these processes satisfy condition A.

As usual, we want that also C and S remove the individual from the study. Therefore we define

$$\tau_j := \inf\{t > 0 | \exists \mathbf{x} \in \mathbf{E}: N_t(\mathbf{x}) = 1, x'_j \in \{DO, DT\} \text{ or } x''_j \in \{S, C\}\},$$

$$R_t := \{j \in J | \tau_j \geqslant t\}.$$

We call the set R_t *risk set* at time t. Define, for any $\mathbf{x} \in \mathbf{E}$, the set $J(\mathbf{x})$ and the $\mathbb{F}$-stopping time $\tau_{\mathbf{x}}$ by

$$J(\mathbf{x}) := \{j \in J | (x'_j, x''_j) \neq \emptyset_2\},$$

$$\tau_{\mathbf{x}} := \min\{\tau_j | j \in J(\mathbf{x})\}.$$

Our model for the tumor experiment is the probability $\mathbb{P}$ on $(\Omega, \mathscr{F}_\infty)$ defined by the compensators $\Lambda(\mathbf{x})$, $\mathbf{x} \in \mathbf{E}$, of the form (2.5) such that:

—the processes $a_j(x')$, $j \in J$, $x' \in \bar{E}'$, are as in (5.3);

—the processes $\bar{B}(\mathbf{x})$, $\mathbf{x} \in \mathbf{E}$, satisfy conditions B and the conditions

$$B_t(\mathbf{x}) = B_{t \wedge \tau_{\mathbf{x}}}(\mathbf{x}), \quad t \geqslant 0, \quad \mathbf{x} \in \mathbf{E}; \tag{5.4a}$$

$$\delta_{jt}(x') = 1_{\{\tau_j \geqslant t\}}, \quad t \geqslant 0, \quad x' \in E', \quad j \in J. \tag{5.4b}$$

Here the processes $B(\mathbf{x})$, $\mathbf{x} \in \mathbf{E}$, describe the censoring and sacrificing mechanisms, and $a_j(x')$, $x' \in E'$, are the $(\mathbb{P}, \mathbb{F})$-intensities for j to experience tumor onset or natural death.

Example. Consider the serial sacrifice scheme of McKnight & Crowley (1984). They have fixed sacrificing times $0 < t_1 < t_2 < \ldots,$ and at times t_i each individual $j \in R_{t_i}$ at risk is sacrificed with probability π_{ij} independently of the others. No censoring is assumed. Then the protocol is specified by the intensity processes

$$b_t(\mathbf{x}) := \begin{cases} 1_{\{j \in R_t\}} & \text{if } \mathbf{x} = \boldsymbol{\emptyset}_{2n,j,x'}, x' \in E', j \in J, \\ 0 & \text{for the other } \mathbf{x} \in \mathbf{E}, \end{cases}$$

for $t > 0$ and by the jumps

$$\Delta B_t(\mathbf{x}) := \begin{cases} 1_{\{J(\mathbf{x}) \subset R_t\}} \prod_{\substack{j \in R_t \\ x_j'' = S}} \pi_{ij} \prod_{\substack{j \in R_t \\ x_j'' = \emptyset}} (1 - \pi_{ij}) & \text{if } x_j' = \emptyset \text{ and } x_j'' \in \{S, \emptyset\}\ \forall j \in J, \\ 0 & \text{for the other } \mathbf{x} \in \mathbf{E}, \end{cases}$$

for $t \in \{t_1, t_2, \ldots, \}$, the jumps being identically zero for the other t.

The condition (5.4a) restricts the whole mass of $\mathbb{P}$ to that part of the sample space Ω where each individual experiences at most one of the events DO, DT, S, or C. DO, C and S are possibly, and DT always preceded by T. Then with probability one $(\mathbf{T}, \mathbf{X})$ has at most $2n$ points inside $\mathbb{R}_+ \times \mathbf{E}$. This implies that with probability one each individual j experiences at most one event in the observed process, which for j happens at the time τ_j. Comparing this with remark 5 in section 2.3 note that under (5.4a) we have $R_s(x') \subset R_s$ for $x' \in \{T, DO, DT\}$, and (5.4b) strengthens the inclusion into equality. Therefore the intensity for individual j to experience $x' \in E'$ is $a_j(x')$ until τ_j, and zero thereafter. Obviously it is not a restriction in the present application to assume that τ_j is observed for each $j \in J$. We can therefore consider $t = \infty$ in (4.4). We then have in (4.4b):

$$\prod_i \hat{\pi}_{j, \hat{T}_i -}[\tilde{a}_j(\hat{T}_i, \hat{X}_{ij}, X_{ij}'')] = \hat{\pi}_{j\tau_j -}[\tilde{a}_j(\tau_j, \hat{X}_{\eta_j, j}, X_{\eta_j, j}'')], \tag{5.5a}$$

where $\eta_j := i$ if $\hat{T}_i = \tau_j$, and in the exponential part

$$\sum_{x' \in E_1'} \int_0^t \hat{\pi}_{js-}[a_j(s, x')]\, \delta_{js}(x')\, ds = \int_0^{\tau_j} (\hat{\pi}_{js-}[a_j(s, DO)] + \hat{\pi}_{js-}[a_j(s, DT)])\, ds. \tag{5.5b}$$

in the exponential part of (4.4b).

In order to get an explicit form for the likelihood, expressed in terms of the hazard functions for tumor onset and natural death, we reason as follows. By (5.2) and (5.3) one only needs to consider the conditional distribution of $U_j 1\{U_j < t\}$, given $\hat{\mathscr{F}}_{jt-}$, where the latter is restricted to the set $\{j \in R_t\}$. For this it suffices to consider the conditional distribution of the tumor onset time U_j, given that j has not died and has not been sacrificed or censored before t. From our model assumptions and (5.3) it follows that the hazards for tumor onset and natural death of j are not affected by the knowledge that j has not yet been sacrified or censored (cf. Arjas & Haara 1984, sect. 5). Therefore the required conditional distribution for U_j corresponds to conditioning on the event "j has not experienced a natural death before t", and is thus the same as in the experiment without censoring and sacrifices. Denote for $0 < u \leqslant t < \infty$:

$$\bar{H}_1^j(u) := \int_0^u (h_1^{j,T}(v) + h_1^{j,DO}(v))\, dv,$$

$$\bar{H}_2^j(t \mid u) := \int_u^t (h_2^{j,DO}(v \mid u) + h_2^{j,DT}(v \mid u))\, dv,$$

$$m^j(u, t) := \exp\{-\bar{H}_2^j(t \mid u)\} h_1^{j,T}(u) \exp\{-\bar{H}_1^j(u)\},$$

$$S^j(t) := \exp\{-\bar{H}_1^j(t)\} + \int_0^t m^j(v, t)\, dv,$$

$$f^j(u, t) := S^j(t)^{-1} m^j(u, t),$$

$$P^j(t) := \int_0^t f^j(u, t)\, du.$$

The differential $f^j(u, t)\, du$ can be interpreted as the conditional probability of the event "j had tumor onset at du" given that "j is alive at t", and $P^j(t)$ is the corresponding prevalence at time t. Thus on the set $\{\omega | j \in R_t(\omega)\}$ we have

$$\mathbb{P}(U_j \leqslant u | \hat{\mathscr{F}}_{jt-}) = \int_0^u f^j(v, t)\, dv, \quad u \leqslant t,$$

$$\mathbb{P}(U_j > t | \hat{\mathscr{F}}_{jt-}) = 1 - P^j(t). \tag{5.6}$$

By (4.2) and (4.3b) we have

$$\hat{\pi}_{jt-}[\tilde{a}_j(t, \hat{x}, x'')] = \mathbb{E}(I_{jt}(\hat{x}, x'') a_{jt}(p(\hat{x})) | \hat{\mathscr{F}}_{jt-}).$$

By substituting the definitions (5.1)–(5.3) and the results (5.6) into this we get

$$\hat{\pi}_{jt-}[\tilde{a}_j(t, \hat{x}, x'')] = \begin{cases} \int_0^t h_2^{j, DT}(t|u) f^j(u, t)\, du & \text{if } \hat{x} = DTTF, \\ \int_0^t h_2^{j, DO}(t|u) f^j(u, t)\, du & \text{if } \hat{x} = DOTF, \\ h_1^{j, DO}(t)(1 - P^j(t)) & \text{if } \hat{x} = DONF, \\ P^j(t) & \text{if } (\hat{x}, x'') = (TF, S), \\ 1 - P^j(t) & \text{if } (\hat{x}, x'') = (NF, S), \\ 1 & \text{if } (\hat{x}, x'') = (C, \emptyset). \end{cases} \tag{5.7}$$

From (5.7) the likelihood factors (5.5a) corresponding to the marked points $(\tau_j, \hat{X}_{\eta_j, j}, X''_{\eta_j, j})$, $j \in J$, can be written down. Note that in (5.7) we have ignored the pairs $(\hat{x}, x'') \in \{TF, NF\} \times \{\emptyset, C\} \cup \{(\emptyset, S), (\emptyset, \emptyset)\}$, since, by (5.2) and the definition of τ_j, with probability one these marks cannot appear as values of $(\hat{X}_{\eta_j, j}, X''_{\eta_j, j})$. In addition, according to (4.4b), (5.5b), (5.3) and (5.6) each individiual $j \in J$ contributes to the likelihood the factor $\exp\{-A_j\}$, where

$$A_j := \int_0^{\tau_j} \left[\int_0^s (h_2^{j, DO}(s|u) + h_2^{j, DT}(s|u)) f^j(u, s)\, du + h_1^{j, DO}(s)(1 - P^j(s)) \right] ds. \tag{5.8}$$

Summarizing the results, we get the following proposition.

Proposition 5.1

In the considered survival/sacrifice experiment, each individual j contributes to the likelihood expression the factor $\exp\{-A_j\}$ *arising from the survival up to τ_j, with A_j given in* (5.8), *and a factor arising as the contribution of the finding $(\hat{X}_{\eta_j, j}, X''_{\eta_j, j})$ at the removal time τ_j. The latter is obtained from* (5.7) *by the substitution* $(t, \hat{x}, x'') = (\tau_j, \hat{X}_{\eta_j, j}, X''_{\eta_j, j})$.

As stated before, this expression is essentially identical to the likelihood derived by other authors under much more restrictive assumptions.

6. Discussion

Finally we consider briefly the implications which the above, admittedly rather tedious, analysis has to statistical inference. Theorem 4.1 shows that for very general study designs, under the stated regularity conditions, the likelihood expression has the "usual" product form over the set of individuals as long as the protocol events, such as censorings or interventions made by the investigator, do not depend on the model parameters of interest or on events in the future. The latter requirement seems to be satisfied completely generally in practice, but the former, corresponding to the notion of non-informative censoring, may be a real concern in data analysis.

As a consequence, under these conditions all inferential methods satisfying the likelihood principle can be used "as if the data had come from independent observations". In particular, this concerns Bayesian inference, ML-estimation, likelihood ratio rates, and score tests. In asymptotic theory, however, some concern should be shown to the interpretation, in terms of the considered study protocol, of the limits where the sample size and/or the observation time tend to infinity. It should also be emphasized that the above conclusion does not generalize from likelihood-based inference to other forms of statistical inference. Thus, for example, in the case of simple renewal process observations, right censoring of the last observation at a fixed time corresponding to a fixed interval of observation can destroy the logic leading to well-known martingale estimators (Gill, 1980; Arjas, 1985).

Finally we should mention the recent related paper by Gill (1992), in which is shown that a marginal partial likelihood, obtained by integrating out some component of a data vector, still possesses some of the most important properties of partial likelihoods, providing that the "missing term", left out by the original partial likelihood, does not depend on the integration variable. The model and methods used by Gill are rather different from ours, and it is possible that they could be used as an alternative route to the results presented here.

Acknowledgement

We are grateful to Ørnulf Borgan, Richard Gill and Niels Keiding for their insightful comments.

References

Andersen, P. K., Borgan, Ø, Gill, R. D. & Keiding, N. (1988). Censoring, truncation and filtering in statistical models based on counting processes. In *Statistical inference from stochastic processes* (ed. N. U. Prabhu). Contemporary Mathematics, Vol. 80, pp. 19–60. American Mathematical Society.

Arjas, E. (1985). Contribution to the discussion on the paper by P. K. Andersen & Ø. Borgan. *Scand. J. Statist.* **12,** 150–153.

Arjas, E. (1989). Survival models and martingale dynamics (with discussion). *Scand. J. Statist.* **16,** 177–225.

Arjas, E. & Haara, P. (1984). A marked point process approach to censored failure data with complicated covariates. *Scand. J. Statist.* **11,** 193–209.

Arjas, E. & Haara, P. (1992). Periodic inspections in a longitudinal study: viewing occult tumors through a filter. In *Survival analysis: state of the art* (eds J. P. Klein & P. K. Goel). Kluwer, Dordrecht. (To appear.)

Arjas, E., Haara, P. & Norros, I. (1992). Filtering the histories of a partially observed marked point process. *Stochastic Process. Appl.* (to appear).

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

Dellacherie, C. & Meyer, P.-A. (1978). *Probabilities and potential A.* North-Holland, Amsterdam.

Dellacherie, C. & Meyer, P.-A. (1982). *Probabilities and potential B.* North-Holland, Amsterdam.

Gill, R. D. (1980). *Censoring and stochastic integrals.* M. C. Tracts, Mathematical Centre, Amsterdam.

Gill, R. D. (1992). Marginal partial likelihood. *Scand. J. Statist.* **19**, 133–137.
Jacobsen, M. (1982). Statistical analysis of counting processes. *Lecture Notes in Statistics* **12**. Springer, Heidelberg.
Jacod, J. (1975). Multivariate point processes: predictable projection, Radon–Nikodym derivatives, representation of martingales. *Z. Wahrsch. verw. Gebiete* **31**, 235–253.
McKnight, B. & Crowley, J. (1984). Tests for differences in tumor incidence based on animal carcinogenesis experiments. *J. Amer. Statist. Assoc.* **79**, 639–647.
Norros, I. (1986). A compensator representation of multivariate life length distributions, with applications. *Scand. J. Statist.* **13**, 99–112.

Received November 1989, in final form June 1991

Elja Arjas, Department of Applied Mathematics and Statistics, University of Oulu, SF-90570, Oulu, Finland.

Appendix

The proof of lemma 3.3

The filter formula (3.1)–(3.3) of AHN applies directly to the family $(\tilde{\pi}_t; t \geqslant 0)$ with $d\hat{N}_t(y)$ and $d\hat{\Lambda}_t(y)$ of AHN replaced respectively by $d\tilde{N}_t(\mathbf{y})$ and $d\tilde{\Lambda}_t(\mathbf{y})$. The claim is that this formula simplifies to (3.7) when it is applied to the process $(\tilde{\pi}_t(D); t \geqslant 0)$ with $D = (h'_1)^{-1}(A)$ and $A \in \mathscr{H}'$ (see (3.4)).

For each $\mathbf{x} \in \mathbf{E}$ we have

$$1_D(H \cup \{(t, \mathbf{x})\}) - 1_D(H) = 1_A(h'_1(H \cup \{(t, \mathbf{x})\})) - 1_A(h'_1(H)) = d(t, x'_1, A; h'_1(H)),$$

which is zero for $x'_1 = \emptyset$ by definition. Since for each $x' \in E'$

$$\sum_{\mathbf{x}:\, x'_1 = x'} d\Lambda_t(\mathbf{x}) = a_{1t}(x') \sum_{\mathbf{x}'' \in \bar{E}''^n} b_t(\emptyset_{n1x'}, \mathbf{x}'')\, dt \quad \text{(by (2.5), B(i), B(ii))}$$

$$= a_{1t}(x')\, \delta_{1t}(x'), \qquad \text{(by B(iv))}$$

we get

$$\sum_{\mathbf{x} \in \mathbf{E}} \tilde{\pi}_{t-}[(1_D(\cdot \cup \{(t, \mathbf{x})\}) - 1_D)\Lambda(dt, \mathbf{x})] = \sum_{\mathbf{x} \in \mathbf{E}} \tilde{\pi}_{t-}[d(t, x'_1, A; h'_1(\cdot))\Lambda(dt, \mathbf{x})]$$

$$= \sum_{x' \in E'} \tilde{\pi}_{t-}[d(t, x', A; h'_1(\cdot)) \sum_{\mathbf{x}:\, x'_1 = x'} \Lambda(dt, \mathbf{x})]$$

$$= \sum_{x' \in E'} \pi_{t-}[d(t, x', A) a_1(t, x')]\, \delta_{1t}(x'), \tag{A.1}$$

where the last identity follows from the model assumptions, (3.4) and (3.5).

In order to evaluate the martingale part of the filter formula we must calculate the innovation gains $\hat{Z}_t(\mathbf{y}) - \hat{Z}_t(\emptyset)$, for $\mathbf{y} \in \tilde{\mathbf{E}}$, as defined in (3.2) and (3.3) of AHN. By (2.5), (2.6) and B(i) we have

$$\Delta\Lambda(t, \mathbf{x}; H) = \Delta B(t, \mathbf{x}; H), \quad \mathbf{x} \in \mathbf{E} \tag{A.2}$$

and (A.2) holds in the form "0 = 0" whenever $x'_j \neq \emptyset$ for some j. Now (A.2) clearly implies

$$p(t, \emptyset; H) := 1 - \sum_{\mathbf{x} \in \mathbf{E}} \Delta\Lambda(t, \mathbf{x}; H) = 1 - \sum_{\mathbf{x} \in \mathbf{E}} \Delta B(t, \mathbf{x}; H). \tag{A.3}$$

By (3.6) and B(i), $\Delta\tilde{\Lambda}_t(\mathbf{y}) > 0$ holds only if $\mathbf{y} \in \tilde{\mathbf{E}}$ is such that $p(\hat{y}_1) = \emptyset$, $y'_2 = \ldots = y'_n = \emptyset$

and $y''_j \neq \emptyset$ for at least one $j \in J$. Therefore we get

$$\begin{aligned}\hat{p}_t(\emptyset) &:= 1 - \sum_{\mathbf{y} \in \tilde{\mathbf{E}}} \Delta\tilde{\Lambda}_t(\mathbf{y}) \\ &= 1 - \sum_{\mathbf{x}'' \in \mathbf{E}''} \sum_{\hat{x} \in p^{-1}(\emptyset)} \Delta\tilde{\Lambda}_t(\hat{x}, \boldsymbol{\emptyset}_{n-1}, \mathbf{x}'') \\ &= 1 - \sum_{\mathbf{x}'' \in \mathbf{E}''} \pi_{t-}\left[\sum_{\hat{x} \in p^{-1}(\emptyset)} I_1(t, \hat{x}, x''_1)\right] \Delta B_t(\boldsymbol{\emptyset}_n, \mathbf{x}'') \quad \text{(by (3.6) and (2.6))} \\ &= 1 - \sum_{\mathbf{x} \in \mathbf{E}} \Delta B_t(\mathbf{x}). \quad \text{(by C2(ii) and B(i))} \end{aligned} \tag{A.4}$$

From (2.1) of AHN and (3.3) it follows that the functions $c(t, \mathbf{x}, \mathbf{y}; H)$ and $\bar{c}(t, \mathbf{y}; H)$, $\mathbf{x} \in \mathbf{E}$, $\mathbf{y} = (\hat{y}, \mathbf{z}) \in \tilde{\mathbf{E}}$, appearing in the filter formula of AHN are in our case

$$c(t, \mathbf{x}, \mathbf{y}; H) = \begin{cases} I_1(t, \hat{y}_1, x''_1; h'_1(H)) & \text{if } \mathbf{z} = (x'_2, \dots, x'_n, x''_1, \dots, x''_n) \text{ and } p(\hat{y}_1) = x'_1, \\ 0 & \text{otherwise,} \end{cases} \tag{A.5a}$$

and, consequently,

$$\bar{c}(t, \mathbf{x}; H) := \sum_{\mathbf{y} \in \tilde{\mathbf{E}}} c(t, \mathbf{x}, \mathbf{y}; H) = \sum_{\hat{x} \in p^{-1}(x'_1)} I_1(t, \hat{x}, x''_1; h'_1(H)). \tag{A.5b}$$

By (A.5b) and C2(ii) we have $1 - \bar{c}(t, \mathbf{x}; H) = 0$ whenever $x'_1 \in \bar{E}'_1$. On the other hand B(i) implies that $\Delta\Lambda(t, \mathbf{x}) > 0$ only if $x'_1 = \emptyset \in \bar{E}'_1$. Therefore (3.3) of AHN simplifies in the present case to

$$\hat{Z}_t(\emptyset) = \frac{0 + \tilde{\pi}_{t-}[p(t, \emptyset) 1_D]}{\hat{p}_t(\emptyset)} = \tilde{\pi}_{t-}(D) = \pi_{t-}(A) \tag{A.6a}$$

by (A.3) and (A.4) also when $\hat{p}_t(\emptyset) > 0$.

Next we fix $\mathbf{y} = (\hat{y}_1, \mathbf{z}) \in \tilde{\mathbf{E}}$ and consider what comes out when (A.5a) and (2.5) are substituted into (3.2) of AHN. We note first that by (A.5a) each of the four sums appearing on the rhs of (3.2) of AHN have at most one nonzero term, corresponding to the mark $\mathbf{x} = (p(\hat{y}_1), \mathbf{z})$. Thus we have by (3.5)

$$\begin{aligned}&\tilde{\pi}_{t-}\Bigg[I_1(t, \hat{y}_1, y''_1; h'_1(\cdot)) a_1(t, p(\hat{y}_1); h'_1(\cdot)) \\ &\quad \times \left\{\prod_{j \neq 1} a_j(t, y'_j)\right\} b(t, p(\hat{y}_1), \mathbf{z}) 1_D(\cdot \cup \{(t, p(\hat{y}_1), \mathbf{z})\})\Bigg] \\ &= \pi_{t-}[\tilde{a}_1(t, \hat{y}_1, y''_1) g(t, p(\hat{y}_1), A)] \left\{\prod_{j \neq 1} a_{jt}(y'_j)\right\} b_t(p(\hat{y}_1), \mathbf{z})\end{aligned}$$

in the numerator of $\hat{Z}_t(\mathbf{y})$, when $\hat{p}_t(\emptyset) = 1$. The denominator and the case $\hat{p}_t(\emptyset) < 1$ are treated similarly. In particular, the factor

$$\chi_t(\mathbf{y}) := \begin{cases} \prod_{j \neq 1} a_{jt}(y'_j) b_t(p(\hat{y}_1), \mathbf{z}) & \text{if } \hat{p}_t(\emptyset) = 1, \\ \prod_{j \neq 1} a_{jt}(y'_j) \Delta B_t(p(\hat{y}_1), \mathbf{z}) & \text{if } \hat{p}_t(\emptyset) < 1, \end{cases}$$

cancels out. This is also the case when $\chi_t(\mathbf{y})$ is zero, since outside a $\mathbb{P}$-null set

$$1_{\{\chi_t(\mathbf{y}) = 0\}} (d\tilde{N}_t(\mathbf{y}) - d\tilde{\Lambda}_t(\mathbf{y})), \quad t \geqslant 0,$$

is the zero-measure on $\mathbb{R}_+$. This leaves us with

$$\hat{Z}_t(\mathbf{y}) = \frac{\pi_{t-}[\tilde{a}_1(t, \hat{y}_1, y''_1)g(t, p(\hat{y}_1), A)]}{\pi_{t-}[\tilde{a}_1(t, \hat{y}_1, y''_1)]}. \tag{A.6b}$$

From (A.6a, b) it follows that the innovation gain corresponding to the observation $\mathbf{y} \in \tilde{\mathbf{E}}$ at time t is $C_{1t}(\hat{y}_1, y''_1)$, where $C_{1t}(\hat{x}, x'')$ is as in (3.7b). On the other hand, $C_{1t}(\emptyset, \emptyset) = 0$ by C2(iii), (2.6) and the definition of $\tilde{a}_1(t, \hat{x}, x'')$ in (3.6). Therefore we get

$$\sum_{\mathbf{y} \in \tilde{\mathbf{E}}} (\hat{Z}_t(\mathbf{y}) - \hat{Z}_t(\emptyset))\, d\tilde{N}_t(\mathbf{y}) = \sum_{(\hat{x}, x'') \neq \emptyset_2} C_{1t}(\hat{x}, x'') \sum_{\substack{\mathbf{y} \in \tilde{\mathbf{E}}: \\ (\hat{y}_1, y''_1) = (\hat{x}, x'')}} d\tilde{N}_t(\mathbf{y})$$

$$= \sum_{(\hat{x}, x'') \neq \emptyset_2} C_{1t}(\hat{x}, x'')\, d\hat{N}_{1t}(\hat{x}, x''). \tag{A.7}$$

Assume now that $p(\hat{y}_1) = \emptyset$. Then $\tilde{a}_1(t, \hat{y}_1, y''_1) = I_1(t, \hat{y}_1, y''_1)$ by (2.6) and in (A.6b) we have

$$\hat{Z}_t(\mathbf{y}) = \frac{\pi_{t-}[I_1(t, \hat{y}_1, y''_1)1_A]}{\pi_{t-}[I_1(t, \hat{y}_1, y''_1)]}.$$

Since $d\tilde{\Lambda}_t(\mathbf{y}) = \pi_{t-}[I_1(t, \hat{y}_1, y''_1)]\, \{\Pi_{j \neq 1} a_{jt}(y'_j)\}\, dB_t(\emptyset, \mathbf{z})$ by (3.6), we get

$$\sum_{\mathbf{y}: p(\hat{y}_1) = \emptyset} C_{1t}(\hat{y}_1, y''_1)\, d\tilde{\Lambda}_t(\mathbf{y}) = \sum_{\mathbf{y}: p(\hat{y}_1) = \emptyset} \{\pi_{t-}[I_1(t, \hat{y}_1, y''_1)1_A]$$

$$- \pi_{t-}(A)\pi_{t-}[I_1(t, \hat{y}_1, y''_1)]\} \left\{ \prod_{j \neq 1} a_{jt}(y'_j) \right\} dB_t(\emptyset, \mathbf{z})$$

$$= \sum_{\mathbf{z}: (\emptyset, \mathbf{z}) \in \mathbf{E}} \left\{ \pi_{t-}\left[\sum_{\hat{x} \in p^{-1}(\emptyset)} I_1(t, \hat{x}, y''_1)1_A \right] - \pi_{t-}(A) \right.$$

$$\left. \times\, \pi_{t-}\left[\sum_{\hat{x} \in p^{-1}(\emptyset)} I_1(t, \hat{x}, y''_1) \right] \right\} \left\{ \sum_{j \neq 1} a_{jt}(y'_j) \right\} dB_t(\emptyset, \mathbf{z})$$

$$= 0 \tag{A.8a}$$

by C2(ii). If $p(\hat{y}_1) = x' \in E'_1$, then a similar calculation gives

$$\sum_{\mathbf{y}: p(\hat{y}_1) = x'} C_{1t}(\hat{y}_1, y''_1)\, d\tilde{\Lambda}_t(\mathbf{y}) = \sum_{\mathbf{z}: (x', \mathbf{z}) \in \mathbf{E}} \left\{ \pi_{t-}\left[\sum_{\hat{x} \in p^{-1}(x')} I_1(t, \hat{x}, y''_1)a_1(t, x')g(t, x', A) \right] \right.$$

$$\left. - \pi_{t-}(A)\pi_{t-}\left[\sum_{\hat{x} \in p^{-1}(x')} I_1(t, \hat{x}, y''_1)a_1(t, x') \right] \right\}$$

$$\times \left\{ \prod_{j \neq 1} a_{jt}(y'_j) \right\} dB_t(x', \mathbf{z})$$

$$= \{\pi_{t-}[a_1(t, x')g(t, x', A)] - \pi_{t-}[a_1(t, x')]\pi_{t-}(A)\}\, \delta_{1t}(x'), \tag{A.8b}$$

where the last equality is implied by C2(ii), (2.6), B(ii) and B(iv).

From (A.1) and (A.8) it follows that

$$\sum_{\mathbf{x} \in \mathbf{E}} \tilde{\pi}_{t-}[(1_D(\cdot \cup \{(t, \mathbf{x})\}) - 1_D)\Lambda(dt, \mathbf{x})] - \sum_{\mathbf{y} \in \tilde{\mathbf{E}}} (\hat{Z}_t(\mathbf{y}) - \hat{Z}_t(\emptyset))\, d\tilde{\Lambda}_t(\mathbf{y})$$

$$= \sum_{x' \in E'} \pi_{t-}[a_1(t, x')\, d(t, x', A)]\, \delta_{1t}(x') - \sum_{x' \in E'_1} \{\pi_{t-}[a_1(t, x')g(t, x', A)]$$

$$- \pi_{t-}[a_1(t, x')]\pi_{t-}(A)\}\, \delta_{1t}(x')$$

$$= \sum_{x' \in E'_0} \pi_{t-}[a_1(t, x')\, d(t, x', A)]\, \delta_{1t}(x')$$

$$+ \sum_{x' \in E'_1} \{\pi_{t-}[a_1(t, x')]\pi_{t-}(A) - \pi_{t-}[a_1(t, x')1_A]\}\, \delta_{1t}(x'), \tag{A.9}$$

since $E' = E'_0 + E'_1$ and $d(t, x', A) = g(t, x', A) - 1_A$. Now theorem 1 of AHN, (A.7) and (A.9) imply lemma 3.3. □

The proof of lemma 3.4

Abbreviate $\nu_{iu} := \pi_u(\omega_i; \cdot)$, $i = 1, 2$, and $v := \min(V(\omega_1), V(\omega_2))$. From (3.7a) it follows that

$$\nu_{iu}(A) = \nu_{it}(A) + \int_t^u \sum_{x' \in E} G_{x'}(\nu_{is-}[f_k(s, x', A)]; 1 \leqslant k \leqslant 4)\, \delta_{1s}(\omega_i, x')\, ds, \tag{A.10}$$

for $t \leqslant u < v$, $A \in \mathscr{H}'$, $i = 1, 2$. Here

$$G_{x'}(c_1, c_2, c_3, c_4) := \begin{cases} c_1 & \text{if } x' \in E'_0, \\ c_2 c_3 - c_4 & \text{if } x' \in E'_1, \end{cases}$$

and

$$f_k(s, x', A; H') := \begin{cases} a_1(s, x'; H')\, d(s, x', A; H') & \text{if } k = 1, \\ a_1(s, x'; H') & \text{if } k = 2, \\ 1_A(H') & \text{if } k = 3, \\ a_1(s, x'; H') 1_A(H') & \text{if } k = 4. \end{cases}$$

For each $x' \in E'$ the function $G_{x'}: \mathbb{R}^4 \to \mathbb{R}$ satisfies the Lipschitz condition on any compact subset of $\mathbb{R}^4$. By condition A each function $f_k(s, x', A; H')$, $1 \leqslant k \leqslant 4$, is $\mathscr{H}'$-measurable in H', and is bounded by a finite constant common to all $t \leqslant s < v$, $x' \in E'$, $A \in \mathscr{H}'$ and $H' \in \mathbb{H}'$.

Since $\delta_{1s}(\omega_1, x') = \delta_{1s}(\omega_2, x')$, $t \leqslant s < v$, $x' \in E'$, by assumption (i) of the lemma and condition B(iv), we see that ν_{1u} and ν_{2u}, $t \leqslant u < v$, are solutions of the same integral equation (A.10). Since $\nu_{1t} = \nu_{2t}$ by assumption (ii) and E' is finite we get by the same arguments as in (3.12) of AHN the evaluation

$$|\nu_{1u}(A) - \nu_{2u}(A)| \leqslant K \int_t^u \|\nu_{1s-} - \nu_{2s-}\|\, ds,$$

for $A \in H'$, $t < u < v$. The rest of the proof is identical to that of theorem 2 in AHN. □

The proof of lemma 3.5

Let $A \in \mathscr{H}'$ and $t > 0$. Since $\hat{\mathscr{F}}_{1t} = \sigma(\hat{H}_{1t})$ it suffices to prove the implication

$$\hat{H}_{1t}(\omega_1) = \hat{H}_{1t}(\omega_2) \Rightarrow \pi_t(\omega_1; A) = \pi_t(\omega_2; A). \tag{A.11}$$

Take ω_1 and ω_2 such that lhs of (A.11) holds. Denote again $\nu_{is} := \pi_s(\omega_i; \cdot)$, $0 \leqslant s \leqslant t$, $i = 1, 2$. Let $0 < t_1 < t_2 < \ldots < t_k < t$ be the ordering of the point set

$$\left\{ 0 < s < t \,\middle|\, \sum_{y \in \mathbb{E}} (\Delta\tilde{N}_s(\omega_1; y) + \Delta\tilde{N}_s(\omega_2; y)) > 0 \right\}.$$

By the integrability assumption following (2.5) it is not a restriction to assume that ω_1 and ω_2 are such that k is finite. Denote $t_0 := 0$ and $t_{k+1} := t$.

By lemma 3.4 the measure-valued functions ν_1 and ν_2 coincide on $[t_{h-1}, t_h)$ provided they coincide at t_{h-1}, $h = 1, 2, \ldots, k + 1$. Since $\nu_{10} = \varepsilon_\emptyset = \nu_{20}$, where $\varepsilon_\emptyset$ is the unit mass at $\emptyset \in \mathbb{H}'$, it suffices to show that

$$\nu_{1t_h-} = \nu_{2t_h-} \Rightarrow \nu_{1t_h} = \nu_{2t_h}, \tag{A.12}$$

where $h = 1, 2, \ldots, k$. From (3.7a) we get

$$\nu_{it_h}(A) = \nu_{it_h-}(A) + \sum_{(\hat{x}, x'') \neq \emptyset_2} C_{1t_h}(\omega_i, \hat{x}, x'')\Delta\hat{N}_{1t_h}(\omega_i, \hat{x}, x''), \quad A \in \mathscr{H}', \quad i = 1, 2. \quad \text{(A.13)}$$

Now for each $(\hat{x}, x'')$, the innovation gains $C_{1t_h}(\hat{x}, x'')$ depend on ω only through π_{t_h-} (by (3.7b)) and also $\Delta\hat{N}_{1t_h}(\omega_1, \hat{x}, x'') = \Delta\hat{N}_{1t_h}(\omega_2, \hat{x}, x'')$ by the choice of ω_1 and ω_2. Therefore (A.13) implies (A.12).

We have thus proved (A.11) in the seemingly stronger form

$$\hat{H}_{1t}(\omega_1) = \hat{H}_{1t}(\omega_2) \Rightarrow \pi_s(\omega_1; \cdot\,) = \pi_s(\omega_2; \cdot\,), \quad 0 \leqslant s \leqslant t.$$

□