Stochastic Processes and their Applications 40 (1992) 225–250
North-Holland

Filtering the histories of a partially observed marked point process

Elja Arjas and Pentti Haara

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

Ikka Norros

Technical Research Centre of Finland, Telecommunications Laboratory, SF-02150 Espoo, Finland

Received 23 November 1989
Revised 14 January 1991

We consider a situation in which the evolution of an 'underlying' marked point process is of interest, but where this process is not directly observable. Instead, we assume that another marked point process, which is fully determined by the underlying process, can be observed. The problem is then the estimation, at any given time t, of the underlying development so far, given the corresponding observations. The solution, in the sense of a conditional distribution of the underlying pre-t history, is shown to satisfy a recursive filter formula. Sufficient conditions for the uniqueness of the solution are given. Two non-trivial examples are considered in detail.

marked point process * compensator * filtering * history set * disruption problem * alternating renewal process

1. Introduction

Marked point processes (MPP's) often provide a very convenient framework for the statistical modelling of evolutionary behaviour in time. Examples of their use include survival studies in clinical research, duration analysis in demography and econometrics, and the analysis of reliability in engineering, to mention a few. For a recent survey on survival models, and a list of references, see Arjas (1989).

An MPP is defined as a time ordered sequence of marked points $(T, X) = (T_i, X_i)_{i\geq 1}$, with the finite epochs satisfying $(0<)\ T_1 < T_2 < \cdots$ and the marks X_i taking values in a conveniently defined set E. In applications T_i are typically the time epochs when 'something of interest occurs', and X_i is then a description of that corresponding event. In survival studies, for example, the event at T_i can be the death or the censoring of an individual or the change of the model covariates to a new level. The probability law of an MPP is often most conveniently specified in terms of conditional mark-specific intensities (more generally, compensators), where the conditioning at time t corresponds to the pre-t evolution of the MPP itself.

One reason why MPP's are such a natural tool for modelling particularly evolutionary (transient) behaviour of individuals is that there is no need to assume that the process has a Markov, Markov renewal, or some other special 'state' structure. In

0304-4149/92/$05.00

the above mentioned applications the considered individuals typically go through various stages of development and do not return to 'states' which were visited earlier.

The purpose of this paper is to consider situations in which an MPP $(T_i, X_i)_{i \geq 1}$ is not fully observable. Instead, we assume that another MPP, say $(\hat{T}, \hat{X}) = (\hat{T}_i, \hat{X}_i)_{i \geq 1}$, which is completely determined by the former, is observed. The question of interest to us is then: Suppose that by the time t we have observed the pre-t evolution of $(\hat{T}, \hat{X})$, i.e., the marked points $(\hat{T}_i, \hat{X}_i)$ such that $\hat{T}_i \leq t$. What can then be said about the corresponding pre-t evolution of the underlying process (T, X)?

The complete answer is obviously expressed in terms of the conditional distribution of the underlying pre-t history, given the observations. This could, in principle, be determined directly from the law of (T, X) by Bayes' rule. However, the necessary integration over the sample paths (which are consistent with the observations) will usually not be feasible in practice. It is also of more interest to us here that such probabilities can be updated recursively in continuous time, as t increases, where the updating depends directly on what is observed to happen in $(\hat{T}, \hat{X})$. In this way our approach is a dynamic one and falls within the general framework of partially observed processes and filtering as considered, in the context of point processes, e.g. by Galchuck and Rozovsky (1971), Segall, Davis and Kailath (1975), Hadjiev (1978), Brémaud and Jacod (1977), Brémaud (1981) and Kliemann, Koch and Marchetti (1990). Our approach differs from that of the authors above in the sense that due to the MPP structure of the state process the pertinent semimartingales do not have continuous martingale parts. This allows for the *pathwise* solution of the filtering problem presented in this paper.

The plan of this paper is as follows. In Section 2 we introduce the necessary mathematical framework, stressing the explicit use of MPP histories, and prove some fundamental properties which are needed later. Section 3 contains the main result of this paper, the filter formula, and its proof. Section 4 illustrates the use of the filter formula, providing a detailed analysis and explicit solutions of two non-trivial examples.

2. Mathematical preliminaries

Let E be a nonempty countable set. The considered *underlying* process is a marked point process (MPP)

$$(T, X) = (T_i, X_i)_{i \geq 1}$$

with mark space E, defined on some abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The counting processes $N(x)$, $x \in E$, and the internal history $\mathbb{F} := (\mathcal{F}_t)_{t \geq 0}$ of (T, X) are defined on Ω by

$$N_t(x) := \sum_{i \geq 1}^{\infty} 1_{\{T_i \leq t, X_i = x\}}, \quad t \geq 0, \ x \in E,$$

$$\mathcal{F}_t := \sigma\{N_s(x) \mid s \leq t, x \in E\}, \quad t \geq 0.$$

We also denote $\bar{N} := \sum_{x \in E} N(x)$. We often omit the time-parameter from the notation for stochastic processes. We shall use the upper bar as a generic notation for summation of processes over a mark space. For unexplained terminology on the general theory of stochastic processes see the two volumes of Dellacherie and Meyer (1978, 1982).

As mentioned in the introduction, we view the underlying process (T, X) as one which cannot be fully observed. Instead, we assume that there is another MPP $(\hat{T}, \hat{X})$ which is completely determined by (T, X) and which can be observed. The mathematical definition is given most conveniently in terms of the corresponding counting processes: Let $\hat{E}$ be a nonempty countable set and let $c(x, y)$, $x \in E$, $y \in \hat{E}$, be a family of $\{0, 1\}$-valued $\mathbb{F}$-predictable processes such that for each $x \in E$,

$$\bar{c}(x) := \sum_{y \in \hat{E}} c(x, y) \leq 1.$$

Then the counting processes

$$\hat{N}_t(y) := \sum_{x \in E} \int_0^t c_s(x, y)\, \mathrm{d}N_s(x), \quad y \in \hat{E}, \ t \geq 0, \tag{2.1}$$

define an MPP with mark space $\hat{E}$, derived from (T, X) in the sense of definition 3.2.1.(2) in Arjas (1989). Let $\mathbb{G} := (\mathcal{G}_t)_{t \geq 0}$ be the internal history of this MPP. We call $\mathbb{F}$ *the underlying history* and $\mathbb{G}$ *the observed history*. Clearly $\mathcal{F}_t \supset \mathcal{G}_t$ holds for all t.

For a stochastic process B with almost surely (abbr. a.s.) càdlàg sample paths we adopt the usual notation B_-, $\Delta B := B - B_-$, $B^{\mathrm{d}} := \sum \Delta B$, and $B^{\mathrm{c}} := B - B^{\mathrm{d}}$, respectively for the left-continuous version, jumps, discrete part, and continuous part of B.

We complete all σ-algebras on Ω with the $\mathcal{F}$-null sets. As internal histories of MPP's, $\mathbb{F}$ and $\mathbb{G}$ then satisfy the 'usual conditions'. Let $\hat{\mathcal{O}}$ (resp. $\hat{\mathcal{P}}$) be the $\mathbb{G}$-optional (resp. $\mathbb{G}$-predictable) σ-field on $\mathbb{R}_+ \times \Omega$. If $\mathcal{E}$ is a σ-field on some set, we denote by $b(\mathcal{E})$ (resp. $\mathcal{E}_+$) the class of bounded (resp. nonnegative) real-valued $\mathcal{E}$-measurable functions on that set, with the exception that $\mathcal{R}_+$ denotes the Borel-σ-field on $\mathbb{R}_+$. For a stochastic process $B \in b(\mathcal{R}_+ \otimes \mathcal{F}) \cup (\mathcal{R}_+ \otimes \mathcal{F})_+$ we denote by $\mathbb{E}(B \mid \hat{\mathcal{O}})$ (resp. $\mathbb{E}(B \mid \hat{\mathcal{P}})$) the $\mathbb{G}$-optional (resp. $\mathbb{G}$-predictable) projection of B. We also denote by $\Lambda(x)$ the $(\mathbb{P}, \mathbb{F})$-compensator of the counting process $N(x)$, and by $\hat{B}$ the dual $(\mathbb{P}, \mathbb{G})$-predictable projection of an $\mathbb{F}$-adapted locally integrable increasing process B (with a slight abuse of notation, since $\hat{N}(y)$ is not generally $\mathbb{G}$-predictable). In particular,

$$\hat{\Lambda}(y) := \widehat{\hat{N}(y)}$$

is the $(\mathbb{P}, \mathbb{G})$-compensator of the counting process $\hat{N}(y)$.

For the later development we need the following *model assumptions*: The probability $\mathbb{P}$ on $(\Omega, \mathcal{F})$ is such that

(i) the sample paths of $\bar{N}$ are a.s. finite-valued;

(ii) $\Lambda^{c}(x)$ possesses an intensity $\lambda(x)$, $x \in E$;

(iii) $(\widehat{\Lambda^{d}(x)})^{c} \equiv 0, \quad x \in E.$

By (i) the compensators $\Lambda(x)$, $x \in E$, have versions that satisfy $\bar{\Lambda}_t < \infty$, $t \in \mathbb{R}_+$. In fact, it is not a further restriction to assume that *all* sample paths of $\bar{N}$ are non-explosive. Thus we don't worry about explosive sample paths in the sequel.

The pre-t sample path of an MPP is obviously a collection of marked points (T_i, X_i) such that $T_i \leq t$. With this in mind, and following Norros (1986), we call a finite subset H of $\mathbb{R}_+ \times E$ *a history set* if it is such that $(t, x) \in H$ and $(t, x') \in H$ imply $x = x'$. We denote by $\mathbb{H}$ the space of all history sets.

We can embed $\mathbb{H}$ into the topological sum

$$\Sigma := \{0\} + \sum_{n=1}^{\infty} ((0, \infty) \times E)^n$$

by the open injection

$$j(H) := \begin{cases} 0 & \text{if } H = \emptyset, \\ ((t_1, x_1), \ldots, (t_n, x_n)) & \text{if } H = \{(t_1, x_1), \ldots, (t_n, x_n)\} \\ & \qquad \text{and } t_1 < \cdots < t_n, \ n \in \mathbb{N}. \end{cases}$$

It is an exercise in elementary topology to show that $\mathbb{H}$ equipped with the topology

$$\{j^{-1}(U) \mid U \text{ is open in } \Sigma\}$$

is locally compact Hausdorff with a countable base. Then well-known results (see e.g. Revuz, 1975, p. 6) imply that the space $\mathbb{H}$ (with this topology) is Polish and the corresponding σ-field $\mathscr{H}$ of Borel sets is countably generated.

The $\mathbb{H}$-valued pre-t histories of (T, X) correspond to the *history process* $\boldsymbol{H} := (H_t)_{t \geq 0}$, defined by

$$H_t(\omega) := \{(T_i(\omega), X_i(\omega)) \mid T_i(\omega) \leq t\}, \tag{2.2a}$$

and its left-continuous version $\boldsymbol{H}_- := (H_{t-})_{t>0}$, defined by

$$H_{t-}(\omega) := \{(T_i(\omega), X_i(\omega)) \mid T_i(\omega) < t\}. \tag{2.2b}$$

Note that, up to the completion by null sets, $\mathscr{F}_t$ is generated by H_t. From condition (i) and (A2, T34) in Brémaud (1981) it follows that for each $\mathbb{F}$-predictable process Y there exists a non-random $\mathscr{R}_+ \otimes \mathscr{H}$-measurable function $(t, H) \mapsto Y'(t; H)$ such that the process

$$Y'_t(\omega) := Y'(t; H_{t-}(\omega)), \quad t \geq 0, \ \omega \in \Omega, \tag{2.3}$$

is indistinguishable from Y. From now on we drop the prime and use the position of the time-variable to indicate the difference between an $\mathbb{F}$-predictable process and the corresponding function of (t, H).

The basic filtering problem outlined in Section 1 now is to calculate the conditional distribution of the underlying pre-t history H_t given the observations $\mathscr{G}_t$. For this purpose, denote

$$\hat{\pi}_t(A) := \mathbb{P}(H_t \in A \mid \mathscr{G}_t), \tag{2.4}$$

where $A \in \mathscr{H}$ and $t \in \mathbb{R}_+$. We also denote

$$\mu[f] := \int_{\mathbb{H}} \mu(\mathrm{d}H) f(H),$$

when μ is a measure on $(\mathbb{H}, \mathscr{H})$ and the integral exists. The following lemma shows that $\hat{\pi}$ admits a regular version both as a function of (ω, A) with t fixed and as a function of (t, ω) with A fixed.

Lemma 1. *There exist transition probabilities $\hat{\pi}_t$, $t \geq 0$, (resp. $\hat{\pi}_{t-}$, $t > 0$) from $(\Omega, \mathscr{G}_t)$ (resp. from $(\Omega, \mathscr{G}_{t-})$) to $(\mathbb{H}, \mathscr{H})$ such that*

(a) *for any finite $\mathbb{G}$-stopping time T (resp. for any finite $\mathbb{G}$-predictable stopping time T), $\hat{\pi}_T$ (resp. $\hat{\pi}_{T-}$) is a regular conditional distribution (abbr. r.c.d.) of H_T given $\mathscr{G}_T$ (resp. H_{T-} given $\mathscr{G}_{T-}$);*

(b) *for each $f \in b(\mathscr{H})$ the process $\hat{\pi}_t[f]$, $t \geq 0$, (resp. $\hat{\pi}_{t-}[f]$, $t > 0$) has a.s. càdlàg (resp. càglàd) sample paths;*

(c) *for any $f \in b(\mathscr{H})$, denote by $\hat{\pi}[f]_{t-}$ the left continuous version of the process $\hat{\pi}_t[f]$, that is,*

$$\hat{\pi}[f]_{t-} = \lim_{s \uparrow\uparrow t} \hat{\pi}_s[f], \quad t > 0;$$

then the processes $\hat{\pi}[f]_{t-}$ and $\hat{\pi}_{t-}[f]$ are indistinguishable;

(d) *for each $f \in b(\mathscr{R}_+ \otimes \mathscr{H})$ the process $\hat{\pi}_{t-}[f(t)]$, $t > 0$, is $\mathbb{G}$-predictable, and*

$$\hat{\pi}_{T-}[f(T)] = \mathbb{E}(f_T \mid \mathbb{G}_{T-}) \tag{2.5}$$

for any finite $\mathbb{G}$-predictable stopping time T.

Proof. (a) In order to define $\hat{\pi}_t$ and $\hat{\pi}_{t-}$ we use the argument on p. 260 of Yor (1977). Up to indistinguishability, the mapping $f \mapsto \mathbb{E}(f \circ \boldsymbol{H} \mid \hat{\mathscr{O}})$ from $b(\mathscr{H})$ to $b(\hat{\mathscr{O}})$ is well-defined, linear, positive and continuous in the sense of monotone increasing limits. Then, by Proposition 4.1 of Getoor (1975), there exists a bounded kernel $\hat{\pi}$ from $(\mathbb{R}_+ \times \Omega, \hat{\mathscr{O}})$ to $(\mathbb{H}, \mathscr{H})$ such that, given $f \in b(\mathscr{H})$,

$$\int_{\mathbb{H}} \hat{\pi}((t, \omega); \mathrm{d}H) f(H) = \mathbb{E}(f \circ \boldsymbol{H} \mid \hat{\mathscr{O}})_t(\omega), \quad t \geq 0, \tag{2.6a}$$

for almost all $\omega \in \Omega$. Since $\mathbb{E}(1 \circ \boldsymbol{H} \mid \hat{\mathscr{O}}) = 1$ a.s. and the history $\mathbb{G}$ has been completed w.r.t. $\mathbb{P}$, we can modify the kernel $\hat{\pi}$ on a null set to obtain a transition probability. Then, for each $t \geq 0$, we define

$$\hat{\pi}_t(\omega; A) := \hat{\pi}((t, \omega); A), \quad \omega \in \Omega, \ A \in \mathscr{H}.$$

The transition probabilities $\hat{\pi}_{t-}$, $t>0$, are found by the same arguments. We apply Getoor's result to the mapping $f \mapsto \mathbb{E}(f \circ \boldsymbol{H}_{-} \mid \hat{\mathscr{P}})$ to obtain a transition probability $\hat{\pi}_{-}$ from $(\mathbb{R}_+ \times \Omega, \hat{\mathscr{P}})$ to $(\mathbb{H}, \mathscr{H})$ such that, given $f \in b(\mathscr{H})$,

$$\int_{\mathbb{H}} \hat{\pi}_{-}((t,\omega); \mathrm{d}H) f(H) = \mathbb{E}(f \circ \boldsymbol{H}_{-} \mid \hat{\mathscr{P}})_t(\omega), \quad t>0, \tag{2.6b}$$

for almost all $\omega \in \Omega$, and define

$$\hat{\pi}_{t-}(\omega; A) := \hat{\pi}_{-}((t,\omega); A), \quad t>0,\ \omega \in \Omega,\ A \in \mathscr{H}.$$

The assertion that the measure-valued random variable $\hat{\pi}_T$ (resp. $\hat{\pi}_{T-}$) is a r.c.d. follows from the well known fact that for any $f \in b(\mathscr{H})$ the random variable $\mathbb{E}(f \circ \boldsymbol{H} \mid \hat{\mathscr{O}})_T$ is a version of $\mathbb{E}(f(H_T) \mid \mathscr{G}_T)$ (resp. $\mathbb{E}(f \circ \boldsymbol{H}_{-} \mid \hat{\mathscr{P}})_T$ is a version of $\mathbb{E}(f(H_{T-}) \mid \mathscr{G}_{T-})$). This finishes the proof of (a).

Property (b) is a direct consequence of (2.6) and (VI, T47) in Dellacherie and Meyer (1982), since $f \circ \boldsymbol{H}$ (resp. $f \circ \boldsymbol{H}_{-}$) is a piecewise constant càdlàg (resp. càglàd) process.

In order to show (c), note that both processes are predictable (because they are left-continuous). Let T be any finite predictable stopping time, and let S_n be an increasing sequence of stopping times such that $S_n < T$ and $S_n \uparrow T$.

Denote $\|f\|_\infty := \sup\{|f(H)| \mid H \in \mathbb{H}\}$. Let $G \in \mathscr{G}_{T-}$ and let $\varepsilon>0$. Since $\mathscr{G}_{S_n} \uparrow \mathscr{G}_{T-}$, there exists $G_\varepsilon \in \mathscr{G}_{T-}$ such that $4\|f\|_\infty P(G \Delta G_\varepsilon) < \varepsilon$ and $G_\varepsilon \in \mathscr{G}_{S_n}$ for large enough $n \in \mathbb{N}$. By the dominated convergence theorem we have for large enough n,

$$\begin{aligned}
&\left| \int_G \hat{\pi}[f]_{T-}\, \mathrm{d}\mathbb{P} - \int_G f(H_{T-})\, \mathrm{d}\mathbb{P} \right| \\
&\quad \leqslant \left| \int_G \hat{\pi}[f]_{T-}\, \mathrm{d}\mathbb{P} - \int_{G_\varepsilon} \hat{\pi}_{S_n}[f]\, \mathrm{d}\mathbb{P} \right| + \left| \int_{G_\varepsilon} f(H_{S_n})\, \mathrm{d}\mathbb{P} - \int_G f(H_{T-})\, \mathrm{d}\mathbb{P} \right| \\
&\quad \leqslant \int_{G \cap G_\varepsilon} |\hat{\pi}[f]_{T-} - \hat{\pi}_{S_n}[f]|\, \mathrm{d}\mathbb{P} + \int_{G \cap G_\varepsilon} |f(H_{S_n}) - f(H_{T-})|\, \mathrm{d}\mathbb{P} \\
&\qquad + 4\|f\|_\infty \mathbb{P}(G \Delta G_\varepsilon) \\
&\quad < 3\varepsilon.
\end{aligned}$$

This implies that $\hat{\pi}[f]_{T-}$ gives a version of $\mathbb{E}(f(H_{T-}) \mid \mathscr{G}_{T-})$. We have shown that the two predictable processes coincide at every finite predictable stopping time almost surely. Thus they are indistinguishable.

In (d) we first consider the $\mathbb{G}$-predictability. Since $\hat{\pi}_{-}(A)$ is a.s. left-continuous and $\mathbb{G}$-adapted, it is $\mathbb{G}$-predictable. Thus the claim holds for

$$f(t; H) = 1_B(t) 1_A(H),$$

where $B \in \mathscr{R}_+$ and $A \in \mathscr{H}$. Since the mapping

$$f \to \hat{\pi}_{t-}[f(t)], \quad t \geqslant 0,$$

from $b(\mathscr{R}_+ \otimes \mathscr{H})$ to the set of stochastic processes on Ω is linear and continuous in the sense of monotone increasing convergence, the predictability for general f follows by a monotone class argument. Finally, from (a) it follows that

$$\hat{\pi}_{T-}[1_B(T)1_A] = \mathbb{E}(1_B(T)1_A(H_{T-}) \mid \mathscr{G}_{T-}), \quad B \in \mathscr{R}_+, \ A \in \mathscr{H},$$

since $1_B(T)$ is $\mathscr{G}_{T-}$-measurable. Again the claim (2.5) follows by a monotone class argument. □

Remarks. (a) Lemma 1(d) extends to $(\mathscr{R}_+ \otimes \mathscr{H})_+$. To deal with this case it suffices to apply the above result to $f \wedge n$, $n \in \mathbb{N}$, and let n tend to infinity. Of course, without extra integrability conditions $\hat{\pi}_{t-}[f(t)]$ may then take the value $+\infty$ with positive probability.

(b) In the proof sketch on p. 261 of Yor (1977) the transition probabilities $\hat{\pi}_{t-}$, $t > 0$, are obtained as left-hand limits in the sense of weak convergence. However, the argument which is given is not convincing, because the space of real-valued bounded continuous mappings of a Polish noncompact space is *not* separable in the topology of uniform convergence. Therefore we have chosen a different approach here.

Recall that by (2.3) there exist $\mathscr{R}_+ \otimes \mathscr{H}$-measurable $\mathbb{R}_+$-valued functions $(t, H) \mapsto \Lambda(t, x; H)$ and $(t, H) \mapsto \lambda(t, x; H)$, $x \in E$, such that $\Lambda_t(\omega, x) = \Lambda_t(t, x; H_{t-}(\omega))$ and $\lambda_t(\omega, x) = \lambda(t, x; H_{t-}(\omega))$. In particular, by condition (ii),

$$\Lambda(t, x; H_{t-}(\omega)) = \int_0^t \lambda(s, x; H_{s-}(\omega))\, \mathrm{d}s + \sum_{s \le t} \Delta\Lambda(s, x; H_{s-}(\omega)), \quad (t, \omega) \in \mathbb{R}_+ \times \Omega,$$

where $\Delta\Lambda(s, x; H_{s-}(\omega)) := \Lambda(s, x; H_{s-}(\omega)) - \Lambda(s-, x; H_{s-}(\omega))$.

We are now in a position to calculate dual $\mathbb{G}$-predictable projections explicitly in terms of the transition probabilities $\hat{\pi}_{t-}$, $t > 0$. This will be necessary in proving our main results in Section 3.

Lemma 2. *Let $f: \mathbb{R}_+ \times \mathbb{H} \to \mathbb{R}$ be $\mathscr{R}_+ \otimes \mathscr{H}$-measurable and bounded. The dual $\mathbb{G}$-predictable projection of the process $\int_0^t f_s \, \mathrm{d}\Lambda_s(x)$, $t \geq 0$, is given by*

$$\int_0^t \hat{\pi}_{s-}[f(s)\Lambda(\mathrm{d}s, x)] := \int_0^t \hat{\pi}_{s-}[f(s)\lambda(s, x)]\, \mathrm{d}s + \sum_{s \le t} \hat{\pi}_{s-}[f(s)\Delta\Lambda(s, x)], \quad t \geq 0. \tag{2.7}$$

Proof. From condition (i) in our model assumptions it follows that the integrals on the r.h.s. of (2.7) exist for each $f \in b(\mathscr{R}_+ \otimes \mathscr{H})$. We consider the case $f \geq 0$ only. Now

$$\int_0^t f_s \, \mathrm{d}\Lambda_s(x) = \int_0^t f_s \lambda_s(x)\, \mathrm{d}s + \sum_{s \le t} f_s \Delta\Lambda_s(x)$$

by (ii), and from Lemma 1(d) it follows that $\int_0^t \hat{\pi}_{s-}[f(s)\lambda(s,x)]\,\mathrm{d}s$, $t \geq 0$, is the dual-$\mathbb{G}$-predictable projection of $\int_0^t f_s \lambda_s(x)\,\mathrm{d}s$, $t \geq 0$.

Denote by $G(x)$ the dual $\mathbb{G}$-predictable projection of $\sum_{s \leq t} f_s \Delta \Lambda_s(x)$, $t \geq 0$, and by $D(x)$ the process $1_{\{\Delta \Lambda(x) > 0\}}$. Let $B \geq 0$ be $\mathbb{G}$-predictable. Since

$$\mathbb{E}\left(\int_0^\infty B_s(1-D_s(x))\,\mathrm{d}\Lambda_s^{\mathrm{d}}(x)\right) = \mathbb{E}\left(\sum_{s>0} B_s(1-D_s(x))\,\widehat{\Delta\Lambda_s(x)}\right) = 0$$

by (iii) and the $\mathbb{G}$-predictability of $D(x)$, we also have

$$\mathbb{E}\left(\int_0^\infty B_s(1-D_s(x)) f_s\,\mathrm{d}\Lambda_s^{\mathrm{d}}(x)\right) = 0.$$

Thus

$$\mathbb{E}\left(\int_0^\infty B_s D_s(x)\,\mathrm{d}G_s(x)\right) = \mathbb{E}\left(\int_0^\infty B_s D_s(x) f_s\,\mathrm{d}\Lambda_s^{\mathrm{d}}(x)\right)$$
$$= \mathbb{E}\left(\int_0^\infty B_s f_s\,\mathrm{d}\Lambda_s^{\mathrm{d}}(x)\right) = E\left(\int_0^\infty B_s\,\mathrm{d}G_s(x)\right).$$

In particular, the continuous part of $G(x)$ is identically zero. Finally

$$\Delta G_t(x) = \mathbb{E}(f_s \Delta \Lambda_s(x) \mid \mathcal{G}_{s-}) = \hat{\pi}_{s-}[f(s)\Delta\Lambda(s,x)]$$

by Dellacherie and Meyer (1982, VI, 81.2) and Lemma 1(d). □

From (2.1) it follows that $\hat{\Lambda}(y)$ is the dual $\mathbb{G}$-predictable projection of the process $\sum_{x \in E} \int_0^t c_s(x,y)\,\mathrm{d}\Lambda_s(x)$, $t \geq 0$. By Lemma 2 we have the representation

$$\mathrm{d}\hat{\Lambda}_t(y) = \sum_x \hat{\pi}_{t-}[c(t,x,y)\Lambda(\mathrm{d}t,x)]$$
$$= \sum_x \hat{\pi}_{t-}[c(t,x,y)\lambda(t,x)]\,\mathrm{d}t + \sum_x \hat{\pi}_{t-}[c(t,x,y)\Delta\Lambda(t,x)] \tag{2.8}$$

for the $(\mathbb{P},\mathbb{G})$-compensator of $\hat{N}(y)$.

3. The filter formula

We are now ready to state the main result of this paper. For this denote

$$\bar{c}(t,x;H) := \sum_y c(t,x,y;H),$$

$$p(t,\emptyset;H) := 1 - \sum_x \Delta\Lambda(t,x;H),$$

$$\hat{p}_t(\emptyset) := 1 - \sum_y \Delta\hat{\Lambda}_t(y) = 1 - \sum_y \hat{\pi}_{t-}\left[\sum_x c(t,x,y)\Delta\Lambda(t,x)\right],$$

$$g(t,x,A;H) := 1_A(H \cup \{(t,x)\}),\quad g_t(x,A) := g(t,x,A;H_{t-}).$$

Note that $g_t(x,A) = 1_A(H_t)$ if $\Delta N_t(x) = 1$. We also adopt the convention $0/0 := 0$.

The reader should recall here from Section 2 the definition (2.1), the notation $\Lambda(x)$ (resp. $\hat{\Lambda}(y)$) for the compensator of the counting process $N(x)$ (resp. $\hat{N}(y)$) w.r.t. the state history $\mathbb{F}$ (resp. the observed history $\mathbb{G}$), the representation (2.3) of an $\mathbb{F}$-predictable process such as $\Lambda(x)$ in terms of the history process (2.2b), and the abbreviation (2.7).

Theorem 1. *Let $A \in \mathscr{H}$ be arbitrary. Then, under the model assumptions* (i)–(iii), *the regular version of the conditional distribution* (2.4), *given in Lemma* 1, *satisfies the recursive equation*

$$\hat{\pi}_t(A) = 1_A(\emptyset) + \int_0^t \sum_x \hat{\pi}_{s-}[(g(s,x,A) - 1_A)\Lambda(\mathrm{d}s, x)]$$
$$+ \sum_y \int_0^t (\hat{Z}_s(y) - \hat{Z}_s(\emptyset))(\mathrm{d}\hat{N}_s(y) - \mathrm{d}\hat{\Lambda}_s(y)), \quad t \geq 0, \tag{3.1}$$

where

$$\hat{Z}_t(y) := \begin{cases} \dfrac{\sum_x \hat{\pi}_{t-}[c(t,x,y)\lambda(t,x)g(t,x,A)]}{\sum_x \hat{\pi}_{t-}[c(t,x,y)\lambda(t,x)]} & \text{if } \hat{p}_t(\emptyset) = 1, \\[2ex] \dfrac{\sum_x \hat{\pi}_{t-}[c(t,x,y)\Delta\Lambda(t,x)g(t,x,A)]}{\sum_x \hat{\pi}_{t-}[c(t,x,y)\Delta\Lambda(t,x)]} & \text{if } \hat{p}_t(\emptyset) < 1, \end{cases} \tag{3.2}$$

and

$$\hat{Z}_t(\emptyset)$$
$$:= \begin{cases} \dfrac{\sum_x \hat{\pi}_{t-}[(1-\bar{c}(t,x))\Delta\Lambda(t,x)g(t,x,A)] + \hat{\pi}_{t-}[p(t,\emptyset)1_A]}{\hat{p}_t(\emptyset)} & \text{if } \hat{p}_t(\emptyset) > 0, \\[2ex] \hat{\pi}_{t-}(A) & \text{if } \hat{p}_t(\emptyset) = 0. \end{cases} \tag{3.3}$$

Formula (3.1) *is valid outside a $\mathbb{P}$-null set, common to all $A \in \mathscr{H}$. (Note that the expression for $\hat{Z}_t(\emptyset)$ collapses into $\hat{\pi}_{t-}(A)$ also when $\hat{p}_t(\emptyset) = 1$.)*

Proof. We fix $A \in \mathscr{H}$.

The $\mathbb{F}$-semimartingale $1_A \circ \boldsymbol{H}$ has the canonical decomposition

$$1_A \circ \boldsymbol{H} = 1_A(\emptyset) + B + M, \tag{3.4}$$

where

$$B_t := \sum_x \int_0^t (g_s(x,A) - 1_A(H_{s-}))\,\mathrm{d}\Lambda_s(x),$$

and

$$M_t := \sum_x \int_0^t (g_s(x,A) - 1_A(H_{s-}))(\mathrm{d}N_s(x) - \mathrm{d}\Lambda_s(x)).$$

By Theorem 1 of Hadjiev (1978), applied to (3.4),

$$\hat{\pi}_t(A) = 1_A(\emptyset) + \hat{B}_t + \sum_{y \in \hat{E}} \int_0^t K_s(y)(\mathrm{d}\hat{N}_s(y) - \mathrm{d}\hat{\Lambda}_s(y)), \quad t \geqslant 0. \tag{3.5}$$

outside a $\mathbb{P}$-null set, where $\hat{B} :=$ the dual $\mathbb{G}$-predictable projection of B,

$$K_t(y) := \begin{cases} V_t(y) + \dfrac{\sum_y V_t(y)\Delta\hat{\Lambda}_t(y)}{\hat{p}_t(\emptyset)} & \text{if } \hat{p}_t(\emptyset) > 0, \\ V_t(y) & \text{if } \hat{p}_t(\emptyset) = 0, \end{cases} \tag{3.6}$$

and

$$V_t(y) := \hat{Z}_t(y) - \hat{\pi}_{t-}(A) - \Delta\hat{B}_t. \tag{3.7}$$

In (3.7) $\hat{Z}(y)$ is a $\mathbb{G}$-predictable process such that

$$\mathbb{E}\left(\int_0^\infty D_t 1_A(H_t)\, \mathrm{d}\hat{N}_t(y)\right) = \mathbb{E}\left(\int_0^\infty D_t \hat{Z}_t(y)\, \mathrm{d}\hat{N}_t(y)\right) \tag{3.8a}$$

for all nonnegative $\mathbb{G}$-predictable D, and

$$\sum_y \hat{Z}_t(y)\Delta\hat{\Lambda}_t(y) = \hat{\pi}_{t-}(A) + \Delta\hat{B}_t \quad \text{if } \hat{p}_t(\emptyset) = 0. \tag{3.8b}$$

By Lemma 2 we have

$$\hat{B}_t = \int_0^t \sum_x \hat{\pi}_{s-}[(g(s, x, A) - 1_A)\Lambda(\mathrm{d}s, x)].$$

Therefore we must prove that (3.2) and (3.3) imply (3.8), and

$$K_t(y) = \hat{Z}_t(y) - \hat{Z}_t(\emptyset). \tag{3.9}$$

Note first that the processes $\hat{Z}(y)$, $y \in \hat{E}$, and $\hat{Z}(\emptyset)$, are $\mathbb{G}$-predictable by Lemma 1(d). By (2.8) and (3.2), for $y \in \hat{E}$,

$$\int_0^t \hat{Z}_s(y)\, \mathrm{d}\hat{\Lambda}_s(y) = \int_0^t \sum_x \hat{\pi}_{s-}[c(s, x, y)g(s, x, A)\Lambda(\mathrm{d}s, x)].$$

Then for any $\mathbb{G}$-predictable process $D \geqslant 0$,

$$\begin{aligned}
&\mathbb{E}\left(\int_0^\infty D_t \hat{Z}_t(y)\, \mathrm{d}\hat{N}_t(y)\right) \\
&= \mathbb{E}\left(\int_0^\infty D_t \sum_x \hat{\pi}_{t-}[c(t, x, y)g(t, x, A)\Lambda(\mathrm{d}t, x)]\right) \\
&= \mathbb{E}\left(\int_0^\infty D_t \sum_x g_t(x, A)c_t(x, y)\, \mathrm{d}\Lambda_t(x)\right) \quad \text{(by Lemma 2)} \\
&= \mathbb{E}\left(\int_0^\infty D_t \sum_x g_t(x, A)c_t(x, y)\, \mathrm{d}N_t(x)\right) \\
&= \mathbb{E}\left(\int_0^\infty D_t 1_A(H_t)\, \mathrm{d}\hat{N}_t(y)\right) \quad (\text{since } \Delta N_t(x) = 1 \Rightarrow H_{t-} \cup \{(t, x)\} = H_t).
\end{aligned}$$

Thus (3.2) implies (3.8a). Clearly

$$\hat{p}_t(\emptyset)=0 \;\Rightarrow\; \hat{\pi}_{t-}\left[\sum_x \bar{c}(t,x)\Delta\Lambda(t,x)\right]=1.$$

Since $\bar{c}(t,x)$ is $\{0,1\}$-valued and $0\leqslant\sum_x \Delta\Lambda(t,x)\leqslant 1$ $\hat{\pi}_{t-}$-a.e., we have, whenever $\hat{p}_t(\emptyset)=0$, that $\sum_x \bar{c}(t,x)\Delta\Lambda(t,x)=1$ $\hat{\pi}_{t-}$-a.e., and moreover $\hat{\pi}_{t-}[\Delta\Lambda(t,x)]>0$ implies $\bar{c}(t,x)=1$ $\hat{\pi}_{t-}$ a.e. Thus $\hat{p}_t(\emptyset)=0$ implies

$$\begin{aligned}\sum_y \hat{Z}_t(y)\Delta\hat{\Lambda}_t(y) &= \sum_y\sum_x \hat{\pi}_{t-}[c(t,x,y)\Delta\Lambda(t,x)g(t,x,A)]\\ &= \sum_x \hat{\pi}_{t-}[\bar{c}(t,x)\Delta\Lambda(t,x)(g(t,x,A)-1_A)]\\ &\quad+\hat{\pi}_{t-}\left[\sum_x \bar{c}(t,x)\Delta\Lambda(t,x)1_A\right]\\ &= \Delta\hat{B}_t+\hat{\pi}_{t-}(A).\end{aligned}$$

Thus (3.2) implies (3.8b).

Next, if $\hat{p}_t(\emptyset)>0$, we have

$$\begin{aligned}&\hat{Z}_t(y)-K_t(y)\\ &\quad=\hat{\pi}_{t-}(A)+\Delta\hat{B}_t-\frac{\sum_y V_t(y)\Delta\hat{\Lambda}_t(y)}{\hat{p}_t(\emptyset)}\\ &\quad=\hat{\pi}_{t-}(A)+\Delta\hat{B}_t-\frac{\sum_x \hat{\pi}_{t-}[\bar{c}(t,x)\Delta\Lambda(t,x)g(t,x,A)]-(\hat{\pi}_{t-}(A)+\Delta\hat{B}_t)(1-\hat{p}_t(\emptyset))}{\hat{p}_t(\emptyset)}\\ &\quad=\frac{-\sum_x \hat{\pi}_{t-}[\bar{c}(t,x)\Delta\Lambda(t,x)g(t,x,A)]+\hat{\pi}_{t-}(A)+\sum_x \hat{\pi}_{t-}[(g(t,x,A)-1_A)\Delta\Lambda(t,x)]}{\hat{p}_t(\emptyset)}\\ &\quad=\frac{\sum_x \hat{\pi}_{t-}[(1-\bar{c}(t,x))\Delta\Lambda(t,x)g(t,x,A)]+\hat{\pi}_{t-}[1_A-1_A\sum_x \Delta\Lambda(t,x)]}{\hat{p}_t(\emptyset)}\\ &\quad=\hat{Z}_t(\emptyset).\end{aligned}$$

Finally, for the case $\hat{p}_t(\emptyset)=0$, we have from (3.6) and (3.7)

$$\hat{Z}_t(y)-K_t(y)=\hat{\pi}_{t-}(A)+\Delta\hat{B}_t=\hat{Z}_t(\emptyset)+\Delta\hat{B}_t.$$

But we can delete $\Delta\hat{B}_t$ since $\hat{p}_t(\emptyset)=0$ implies

$$\sum_{y\in\hat{E}}(\Delta\hat{N}_t(y)-\Delta\hat{\Lambda}_t(y))=0$$

$\mathbb{P}$-a.s, and thus (3.5) remains unchanged.

In order to show that one can choose a common null set to all $A\in\mathscr{H}$, note that, a priori as a function of $A\in\mathscr{H}$, the right-hand side of (3.1) is countably additive for each $t\in\mathbb{R}_+$ and $\omega\in\Omega$. On the other hand, $\mathscr{H}$ is generated by a countable field. The claim follows from the uniqueness of the extension of measures. □

Remark. The filter equation (36) of Brémaud and Jacod (1977) (abbr. BJ) is also general enough to apply in the present situation. Unfortunately, their expression for the innovation gain K appears to be incorrect.

First, the + and − signs in (39) of BJ should be interchanged. We think that this is due to some sign errors in an earlier proof (Brémaud, 1975).

A second, and from our point of view more serious observation is that even after the signs have been corrected, (38) and (39) of BJ do not give the correct innovation gain if the predictable part A of the semimartingale X is allowed to have jumps. (In Brémaud (1975) the process A was assumed to have absolutely continuous sample paths, and this fact is needed in step c) of Brémaud's proof.)

Here is a counterexample. Consider the time interval [0, 2]. Let the state history $(\mathscr{F}_t)$ be generated by the process

$$Z_1 1_{\{1<t<2\}} + Z_2 1_{\{t=2\}},$$

where Z_1 and Z_2 are {0, 1}-valued, and distributed according to the transition mechanism in Figure 1.

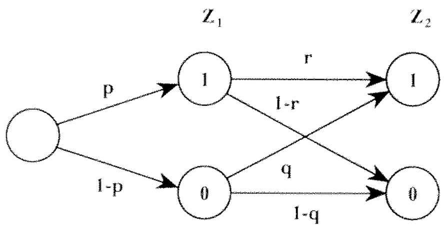

Fig. 1.

Adopting the notation of BJ, let the semimartingale X and the observed counting process $\tilde{Y}$ be identically zero on [0, 2), and define

$$X_2 := 1_{\{Z_2=0\}}, \qquad \tilde{Y}_2 := 1_{\{Z_1=Z_2\}}.$$

Then by elementary calculations one sees that on the set $\{\tilde{Y}_2 = 1\}$,

$$E(X_2 \mid \mathscr{Y}_2) = P(Z_2 = 0 \mid Z_1 = Z_2) = ((1-p)(1-q)+pr)^{-1}(1-p)(1-q),$$

but (38)-(40) of BJ gives, after correcting the signs in (39), the expression

$$((1-p)(1-q)+pr)^{-1}((1-p)(1-q)(1+pq)-p(1-p)r(1-r)).$$

In applications we can sometimes use the filter formula for calculating explicitly the conditional probabilities of certain events of interest in the state history, given the observed history. More precisely: We construct, by 'intelligent guessing', transition probabilities $\tilde{\pi}_t$ from $(\Omega, \mathscr{G}_t)$ to $(\mathbb{H}, \mathscr{A})$, $t \geq 0$, and verify that $\tilde{\pi}_t$, $t \geq 0$, satisfy (3.1)-(3.3) for $A \in \mathscr{A}$, where the sub-σ-field $\mathscr{A} \subset \mathscr{H}$ represents 'the events of interest' in the state history. Finally, we conclude that the conditional probabilities $\mathbb{P}(H_t \in A \mid \mathscr{G}_t)$ are indeed given by $\tilde{\pi}_t(A)$, $A \in \mathscr{A}$, $t \geq 0$.

In order to fulfill the last step we need conditions that guarantee the pathwise uniqueness of the solution $\hat{\pi}$ of the filter formula, as a $P(\mathscr{A})$-valued stochastic process. Here $P(\mathscr{A})$ denotes the set of probability measures on $(\mathbb{H}, \mathscr{A})$. The extra condition, sufficient for uniqueness, says that the filter formula reduces to a form whose sample paths behave smoothly: The jump points of $\hat{\pi}(A)$ do not accumulate, and between the jump points a Lipschitz condition is valid. The possibility to consider a sub-σ-field $\mathscr{A} \subset \mathscr{H}$ is important for two reasons: Often it is not practical to construct the conditional probabilities of all $A \in \mathscr{H}$, and the verification of the filter equations for all $A \in \mathscr{H}$ may be both difficult and unnecessary.

Theorem 2. *Assume* (i)–(iii). *Let $\mathscr{A} \subset \mathscr{H}$ be a σ-field such that for each $A \in \mathscr{A}$ the formula* (3.1) *admits a representation*

$$\hat{\pi}_t(A) = 1_A(\emptyset) + \int_0^t G(\hat{\pi}_{s-}[f_j(s, A)];\ 1 \leq j \leq n)\,\mathrm{d}s$$

$$+ \sum_{y \in \hat{E}} \int_0^t G_y(\hat{\pi}_{s-}[f_j(s, y, A)];\ 1 \leq j \leq n_y)\,\mathrm{d}\hat{N}_s(y), \quad t \geq 0, \tag{3.10}$$

where

(a) $n, n_y \in \mathbb{N}$;

(b) *the function $G: \mathbb{R}^n \to \mathbb{R}$ satisfies the Lipschitz condition on each compact subset of $\mathbb{R}^n$*;

(c) *for each $1 \leq j \leq n$ the functions $s \mapsto f_j(s, A; H)$, $H \in \mathbb{H}$, $A \in \mathscr{A}$, are uniformly bounded on finite intervals*; *and*

(d) *all functions $H \mapsto f_j(s, A; H)$ and $H \mapsto f_j(s, y, A; H)$ are $\mathscr{A}$-measurable.*

Let $\check{\pi}$ be a $P(\mathscr{A})$-valued stochastic process with càdlàg sample paths in the following sense: *$\check{\pi}(A)$ has càdlàg sample paths for each $A \in \mathscr{A}$ and the set functions $\check{\pi}_{t-}(\omega; \cdot)$, $\omega \in \Omega$, $t > 0$, are also probability measures.*

Then, if $\check{\pi}$ satisfies the filter equations (3.10) *for each $A \in \mathscr{A}$, $\check{\pi}$ is indistinguishable from $\hat{\pi}$* (*as $P(\mathscr{A})$-valued processes*).

Proof. Theorem 1 implies the existence of a $\mathbb{P}$-null set Ω^* such that $\hat{\pi}(A)$ has càdlàg sample paths for each $A \in \mathscr{A}$ on $\Omega \backslash \Omega^*$.

In the sequel it is convenient to fix an $\omega \in \Omega \backslash \Omega^*$ and consider $P(\mathscr{A})$- and $\mathbb{R}$-valued functions of $\mathbb{R}_+$, instead of the corresponding processes. However, we continue to omit ω from our notation. Let $t_k := \hat{T}_k(\omega)$ be the kth observed point, $k \in \mathbb{N}$. By (i) the sequence $t_0 := 0 < t_1 < t_2 < \cdots$ does not have a finite accumulation point.

From (3.10) it is straightforward to see that the possible jump points of $\hat{\pi}$ (resp. $\check{\pi}$) are $t_1, t_2, \ldots$, and the size of the jump at t_k depends on $\hat{\pi}$ (resp. $\check{\pi}$) only through $\hat{\pi}_{t_k-}$ (resp. $\check{\pi}_{t_k-}$). Thus it suffices to show that $\hat{\pi}$ and $\check{\pi}$ agree on the interval (t_{k-1}, t_k) provided they agree at t_{k-1}.

On the interval (t_{k-1}, t_k) the filter formula (3.10) can be written as

$$\hat{\pi}_t(A) = \hat{\pi}_{t_{k-1}}(A) + \int_{t_{k-1}}^t G(\hat{\pi}_{s-}[f_j(s, A)];\ 1 \leq j \leq n)\,\mathrm{d}s, \quad t_{k-1} < t < t_k, \tag{3.11}$$

and similarly for $\check{\pi}$.

Denote by $\|\nu\|$ the total variation norm of a bounded signed measure ν on the space $(\mathbb{H}, \mathscr{A})$. If $\nu(\mathbb{H})=0$, then

$$\|\nu\| = 2 \sup_{A\in\mathscr{A}} |\nu(A)|.$$

Obviously $\hat{\pi}_0 = \varepsilon_\emptyset = \check{\pi}_0$. We consider $\|\hat{\pi}_t - \check{\pi}_t\|$ for $t\in(0, t_1)$. The other intervals are treated in a similar manner. From (3.11) we have

$$\begin{aligned}
&|\hat{\pi}_t(A) - \check{\pi}_t(A)| \\
&\quad\leqslant \int_0^t |G(\hat{\pi}_{s-}[f_j(s, A)];\ 1\leqslant j\leqslant n) - G(\check{\pi}_{s-}[f_j(s, A)];\ 1\leqslant j\leqslant n)|\,\mathrm{d}s \\
&\quad\leqslant \int_0^t K' \max_{1\leqslant j\leqslant n} |\hat{\pi}_{s-}[f_j(s, A)] - \check{\pi}_{s-}[f_j(s, A)]|\,\mathrm{d}s \\
&\quad\leqslant K\int_0^t \|\hat{\pi}_{s-} - \check{\pi}_{s-}\|\,\mathrm{d}s, \quad A\in\mathscr{A},\ 0<t<t_1,
\end{aligned} \tag{3.12}$$

where K' and K are finite constants whose existence is implied by the assumptions following (3.10). Obviously this implies

$$\|\hat{\pi}_t - \check{\pi}_t\| < 2K\int_0^t \|\hat{\pi}_{s-} - \check{\pi}_{s-}\|\,\mathrm{d}s, \quad 0<t<t_1. \tag{3.13a}$$

On the other hand, considering $t\uparrow u<t_1$ in (3.12) one has

$$|\hat{\pi}_{u-}(A) - \check{\pi}_{u-}(A)| < K\int_0^u \|\hat{\pi}_{s-} - \check{\pi}_{s-}\|\,\mathrm{d}s$$

for $A\in\mathscr{A}$, $0<u<t_1$. Thus

$$\|\hat{\pi}_{u-} - \check{\pi}_{u-}\| < 2K\int_0^u \|\hat{\pi}_{s-} - \check{\pi}_{s-}\|\,\mathrm{d}s < 4Ku, \quad 0<u<t_1, \tag{3.13b}$$

which implies

$$\begin{aligned}
\|\hat{\pi}_{u-} - \check{\pi}_{u-}\| &< 2K\int_0^u 4Ks\,\mathrm{d}s = 2(2Ku)^2/2! \\
&\ \ \vdots \\
&< 2K\int_0^u 2(2Ks)^n/n!\,\mathrm{d}s = 2(2Ku)^{n+1}/(n+1)!, \quad n\in\mathbb{N}.
\end{aligned}$$

Necessarily $\|\hat{\pi}_{u-} - \check{\pi}_{u-}\| = 0$, $0<u<t_1$, and by (3.13a), $\|\hat{\pi}_t - \check{\pi}_t\| = 0$, $0<t<t_1$, as well. □

For absolutely continuous models we have the following simpler uniqueness condition.

Corollary 1. *Assume that* (i) *holds, $\bar{\Lambda}$ is absolutely continuous and the sample paths of $\bar{\lambda}$ are uniformly bounded on finite intervals. Then the filter equations* (3.1)–(3.3) *determine $\hat{\pi}$ uniquely in the sense of Theorem* 2 *provided that the sub-σ-field $\mathscr{A}$ is rich enough to make the functions $H \mapsto \lambda(s, x; H)$ and $H \mapsto c(s, x, y; H)$, $\mathscr{A}$-measurable.*

Proof. Note that conditions (ii) and (iii) are trivially true. Thus the solution $\hat{\pi}$ exists. For uniqueness we must find a representation (3.10) for the filter formula. It is a straightforward calculation to see that under the absolute continuity of $\bar{\Lambda}$ the filter formula (3.1)–(3.3) transforms into

$$\hat{\pi}_t(A) = 1_A(\emptyset) + \int_0^t \left\{ \hat{\pi}_{s-}\left[\sum_x ((1 - \bar{c}(s, x))g(s, x, A) - 1_A)\lambda(s, x) \right] + \hat{\pi}_{s-}(A)\hat{\pi}_{s-}\left[\sum_x \bar{c}(s, x)\lambda(s, x) \right] \right\} \mathrm{d}s$$
$$+ \sum_y \int_0^t \left\{ \frac{\hat{\pi}_{s-}[\sum_x c(s, x, y)\lambda(s, x)g(s, x, A)]}{\hat{\pi}_{s-}[\sum_x c(s, x, y)\lambda(s, x)]} - \hat{\pi}_{s-}(A) \right\} \mathrm{d}\hat{N}_s(y) \tag{3.14}$$

for all $A \in \mathscr{H}$. This is of the form (3.10) with $n_y = n = 3$, $G(\pi_1, \pi_2, \pi_3) = \pi_1 + \pi_2\pi_3$ and $G_y(\pi_1, \pi_2, \pi_3) = \pi_2^{-1}\pi_1 - \pi_3$ since we can choose the functions $(s, H) \mapsto \lambda(s, x; H)$, $x \in E$, so that $s \mapsto \lambda(s, x; H)$, $x \in E$, $H \in \mathbb{H}$, are uniformly bounded on finite intervals. □

Remarks. (1) Corollary 1 is sufficient for the examples of Section 4. A nontrivial example where the representation (3.10) becomes necessary is the model in Arjas and Haara (1991). There the compensator $\bar{\Lambda}$ can have discontinuities (even in a dense subset of $\mathbb{R}_+$).

(2) A straightforward generalization of (3.10) is obtained by replacing $\int G(\cdot)\,\mathrm{d}s$ by $\sum_{i \in I} \int G_i(\cdot)\,\mathrm{d}Q_s(i)$, where I is a finite set, each mapping $(s, \omega) \to G_i(s, \omega)$ is of the same form as $(s, \omega) \to G(s, \omega)$ in (3.10), each process

$$Q_t(i) = \int_0^t q_s(i)\,\mathrm{d}s + \sum_{s \leq t} \Delta Q_s(i), \quad t \geq 0,$$

is $\mathbb{G}$-predictable and such that the intensity $q_s(i)$ is uniformly bounded over $\omega \in \Omega$ on finite intervals, and, for all $\omega \in \Omega$, the set $\{t > 0 \mid \Delta Q_t(\omega, i) > 0\}$ has no accumulation points. The proof of Theorem 2 goes through virtually unchanged since the jump points of $Q_t(\omega, i)$, $i \in I$, can be absorbed into the increasing sequence t_k, $k \in \mathbb{N}$.

(3) The assumptions (b) and (c) are crucial for obtaining (3.12). We are not aware of ways to relax them.

4. Applications of the filter formula

In this section we consider two problems where the filter formula (3.1) provides an explicit solution. A third and much more elaborate example involving periodic inspections, also allowing for censored observations, is considered in Arjas and Haara (1991). In the definitions and calculations that follow we use the following simplifying notation:

$$\delta(t) := \begin{cases} 1 & \text{if } t<\infty, \\ 0 & \text{if } t=\infty, \end{cases} \qquad \delta_u(t) := \begin{cases} 1 & \text{if } t\le u, \\ 0 & \text{if } t>u, \quad u\in\mathbb{R}_+, \end{cases}$$

$$g(t,x,A;H) := 1_A(H\cup\{(t,x)\}),$$

$$d(t,x,A;H) := g(t,x,A;H) - 1_A(H), \quad t\in\mathbb{R}_+,\ x\in E,\ A\in\mathscr{H},\ H\in\mathbb{H}.$$

4.1. The disruption problem

Let $\sigma>0$ be a random time epoch, and let $0<a<b<\infty$. Suppose that we can observe a counting process $\hat{N}=(\hat{N}_t)$ which, for given σ, is Poisson(a) for times $t\le\sigma$ and Poisson(b) thereafter. The change point σ, however, is not observed, and the task is to determine, given the sample path of $\hat{N}$ up to the observation time t, the conditional distribution of $\sigma\wedge t$.

This is an elaboration of a well-known problem which, using filtering techniques, has been considered at least by Galchuk and Rozovsky (1971), Segall, Davis and Kailath (1975), Brémaud (1981) (whose calculations contain some minor flaws and give an incorrect result) and by Koch (1986). Our calculations extend these results in two ways: First, we consider the (sub)distribution $P(\sigma\in \mathrm{d}s\,|\,\mathscr{G}_t)$, $s\le t$, and not only the 'prevalence' probabilities $P(\sigma\le t\,|\,\mathscr{G}_t)$. Second, we give this distribution in a closed form, instead of only deriving a filter equation which these probabilities satisfy.

In order to express $\hat{N}$ as a process *derived* from an underlying MPP we must define the latter in such a way that it counts both at σ and at the points of $\hat{N}$. For this purpose, let $E=\{0,1\}$ and suppose that the mark-specific $(\mathbb{P},\mathbb{F})$-intensities of $N(0)$ and $N(1)$ are given respectively by

$$\lambda_t(0)=a+(b-a)N_{t-}(1), \quad \lambda_t(1)=h(t)(1-N_{t-}(1)), \qquad t\ge 0, \tag{4.1}$$

where $t\mapsto h(t)$ is a non-random hazard rate for the change point σ. We assume that h is bounded on finite intervals. The observed process $\hat{N}$ is then obviously the same as $N(0)$, and (2.1) is satisfied if we let $\hat{E}=\{0\}$ and

$$c_t(x,y) := \begin{cases} 1 & \text{if } (x,y)=(0,0), \\ 0 & \text{otherwise.} \end{cases}$$

Expressed explicitly as functions of a history set we have

$$\begin{aligned} &c(t,x,y;H) := 1_{\{(0,0)\}}(x,y), \qquad \tau(H) := \inf\{t\,|\,(t,1)\in H\}, \\ &\lambda(t,0;H) := a+(b-a)\delta(\tau(H)), \qquad \lambda(t,1;H) := h(t)(1-\delta(\tau(H))). \end{aligned} \tag{4.2}$$

The mapping τ defines on $(\Omega, \mathscr{F}, \mathbb{P})$ the process $\tau \circ H_t$, $t \geq 0$, which has the property that $\tau \circ H_t = \sigma$ on $\{\sigma \leq t\}$ and $\tau \circ H_t = \infty$ otherwise. The $\mathbb{G}$-intensity of $\hat{N}$ is given by

$$\hat{\lambda}_t := \hat{\lambda}_t(0) = \int_{\mathbb{H}} \lambda(t, 0; H) \hat{\pi}_{t-}(\mathrm{d}H) = a + (b-a)\hat{\pi}_{t-}[\delta(\tau)]. \tag{4.3}$$

In order to construct the above mentioned closed form solution we need some further notation. Let $\mathscr{A} := \sigma(\tau)$; clearly $\mathscr{A} \subset \mathscr{H}$. Denote by f the density

$$f(t) := h(t) \exp\left\{ -\int_0^t h(s)\,\mathrm{d}s \right\}, \quad t \in \mathbb{R}_+, \tag{4.4}$$

corresponding to the given hazard rate h, and let $L(u, t)$ be the Poisson likelihood corresponding to $\sigma = u$,

$$L(u, t) := a^{\hat{N}_u} b^{\hat{N}_t - \hat{N}_u} \mathrm{e}^{-au} \mathrm{e}^{-b(t-u)}, \quad u \in [0, t], \ t \in \mathbb{R}_+. \tag{4.5}$$

$L(u, t)$ is nonnegative and càdlàg in u, and $\mathscr{G}_t$-measurable for each $u \in [0, t]$. Considering Borel sets B, the formula

$$F_t(B) := C_t^{-1} \left[\int_B \delta_t(u) L(u, t) f(u)\,\mathrm{d}u + L(t, t) \int_t^\infty f(u)\,\mathrm{d}u \cdot \varepsilon_\infty(B) \right], \tag{4.6}$$

where ε_∞ is the unit mass at ∞ and

$$C_t := \int_0^t L(u, t) f(u)\,\mathrm{d}u + L(t, t) \int_t^\infty f(u)\,\mathrm{d}u, \tag{4.7}$$

defines a transition probability F_t from $(\Omega, \mathscr{G}_t)$ to $(\bar{\mathbb{R}}_+, \bar{\mathscr{R}}_+)$. Further, since $\tau : \mathbb{H} \to (0, \infty]$ is surjective, the formula

$$\check{\pi}_t(\tau^{-1}(B)) := F_t(B), \quad B \in \bar{\mathscr{R}}_+, \tag{4.8}$$

defines a transition probability $\check{\pi}_t$ from $(\Omega, \mathscr{G}_t)$ to $(\mathbb{H}, \mathscr{A})$ such that F_t is the $\check{\pi}_t$-distribution of τ.

The above definitions imply the existence of $\check{\pi}_{t-}$ and F_{t-} as left hand limits in t. Here F_{t-} is given by (4.6) and (4.7) when $L(u, t)(u < t)$ and $L(t, t)$ are replaced respectively by

$$L(u, t-) = b^{-\Delta \hat{N}_t} L(u, t), \quad u < t, \tag{4.9a}$$

and

$$L(t, t-) = a^{-\Delta \hat{N}_t} L(t, t). \tag{4.9b}$$

The desired conditional distribution is now given by the following proposition.

Proposition 1. $\mathbb{P}(\tau \circ H_t \in B \,|\, \mathcal{G}_t) = F_t(B)$, $B \in \bar{\mathcal{R}}_+$.

Proof. The general filter formula (3.1)–(3.3) obtains now the (differential) form

$$\begin{aligned}
\mathrm{d}\hat{\pi}_t(A) &= \sum_{x=0}^{1} \hat{\pi}_{t-}[\lambda(t,x)d(t,x,A)]\,\mathrm{d}t \\
&\quad +\{\bar{\lambda}_{t-}^{-1}\hat{\pi}_{t-}[\lambda(t,0)g(t,0,A)] - \hat{\pi}_{t-}(A)\}(\mathrm{d}\hat{N}_t - \bar{\lambda}_t\,\mathrm{d}t) \\
&= \{\hat{\pi}_{t-}[\lambda(t,0)]\hat{\pi}_{t-}(A) - \hat{\pi}_{t-}[\lambda(t,0)1_A] + \hat{\pi}_{t-}[\lambda(t,1)d(t,1,A)]\}\,\mathrm{d}t \\
&\quad +\{\bar{\lambda}_{t-}^{-1}\hat{\pi}_{t-}[\lambda(t,0)g(t,0,A)] - \hat{\pi}_{t-}(A)\}\Delta\hat{N}_t \\
&= \{(b-a)(\hat{\pi}_{t-}[\delta(\tau)]\hat{\pi}_{t-}(A) - \hat{\pi}_{t-}[\delta(\tau)1_A]) \\
&\qquad +h(t)\hat{\pi}_{t-}[(1-\delta(\tau))d(t,1,A)]\}\,\mathrm{d}t \\
&\quad +\{\hat{\lambda}_{t-}^{-1}\hat{\pi}_{t-}[\lambda(t,0)g(t,0,A)] - \hat{\pi}_{t-}(A)\}\Delta\hat{N}_t. \qquad (4.10)
\end{aligned}$$

By Corollary 1 it is sufficient to show that for given $t>0$ the family $(\check{\pi}_s)_{s\geq 0}$ satisfies (4.10) for $A = B := \{\tau \leq u\}$, $u \in \mathbb{R}_+$. We have two cases:

(a) $\Delta\hat{N}_t = 1$: Since $g(t,0,B;H) = 1_B(H)$, $H \in \mathbb{H}$, $t \in \mathbb{R}_+$, the calculation

$$\begin{aligned}
&\frac{\check{\pi}_{t-}[\lambda(t,0)1_B]}{\check{\pi}_{t-}[\lambda(t,0)]} \\
&\quad = \frac{b\check{\pi}_{t-}(B)}{a+(b-a)\check{\pi}_{t-}[\delta(\tau)]} \\
&\quad = \frac{b\int_0^{u\wedge t} L(v,t-)f(v)\,\mathrm{d}v}{aL(t,t-)\int_t^{\infty} f(v)\,\mathrm{d}v + b\int_0^t L(v,t-)f(v)\,\mathrm{d}v} \quad \text{(by (4.6)–(4.8))} \\
&\quad = \frac{\int_0^{u\wedge t} L(v,t)f(v)\,\mathrm{d}v}{L(t,t)\int_t^{\infty} f(v)\,\mathrm{d}v + \int_0^t L(v,t)f(v)\,\mathrm{d}v} \quad \text{(by (4.9) and } \Delta\hat{N}_t = 1) \\
&\quad = \check{\pi}_t(B) \quad \text{(by (4.6)–(4.8))}
\end{aligned}$$

implies $\{\check{\lambda}_{t-}^{-1}\check{\pi}_{t-}[\lambda(t,0)g(t,0,B)] - \check{\pi}_{t-}(B)\}\Delta\hat{N}_t = \Delta\check{\pi}_t(B)$, i.e. (4.10) gives correctly the jumps of $\check{\pi}(B)$.

(b) $\Delta\hat{N}_t = 0$: By (4.10) it is sufficient to prove that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\check{\pi}_t(B) &= (b-a)(\check{\pi}_{t-}[\delta(\tau)]\check{\pi}_{t-}(B) - \check{\pi}_{t-}[\delta(\tau)1_B]) \\
&\quad +h(t)\check{\pi}_{t-}[(1-\delta(\tau))d(t,1,B)]. \qquad (4.11)
\end{aligned}$$

Denote $F_t(u) := F_t((0,u])$ and note that $F_t(u) = F_t(u\wedge t)$ for $u \in \mathbb{R}_+$. Having assumed that $\Delta\hat{N}_t = 0$, the function $s \mapsto \hat{N}_s$ is constant in a neighbourhood of t. Thus we can differentiate in (4.5), obtaining

$$\frac{\mathrm{d}}{\mathrm{d}t}L(u,t) = \begin{cases} -bL(u,t) & \text{if } u<t, \\ -aL(t,t) & \text{if } u=t. \end{cases}$$

This implies that

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^u L(v,t)f(v)\,\mathrm{d}v=-b\int_0^u L(v,t)f(v)\,\mathrm{d}v \quad \text{if } u<t,$$

and,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}&\int_0^t L(v,t)f(v)\,\mathrm{d}v\\ &=\left[\frac{\mathrm{d}}{\mathrm{d}s_1}\int_0^{s_1} L(v,s_2)f(v)\,\mathrm{d}v+\int_0^{s_1}\frac{\mathrm{d}}{\mathrm{d}s_2}L(v,s_2)f(v)\,\mathrm{d}v\right]_{s_1=t,s_2=t}\\ &=L(t,t)f(t)-b\int_0^t L(v,t)f(v)\,\mathrm{d}v.\end{aligned}$$

Consequently from (4.7),

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}C_t&=L(t,t)f(t)-b\int_0^t L(v,t)f(v)\,\mathrm{d}v-aL(t,t)\int_t^\infty f(v)\,\mathrm{d}v-L(t,t)f(t)\\ &=(a-b)\int_0^t L(v,t)f(v)\,\mathrm{d}v-aC_t,\end{aligned}$$

hence

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}F_t(u)&=\frac{-b\int_0^u L(v,t)f(v)\,\mathrm{d}v\cdot C_t-(\mathrm{d}/\mathrm{d}t)\,C_t\cdot\int_0^u L(v,t)f(v)\,\mathrm{d}v}{C_t^2}\\ &=(-b)\frac{\int_0^u L(v,t)f(v)\,\mathrm{d}v}{C_t}\\ &\quad-\left[(a-b)\frac{\int_0^t L(v,t)f(v)\,\mathrm{d}v}{C_t}-a\right]\frac{\int_0^u L(v,t)f(v)\,\mathrm{d}v}{C_t}\\ &=(b-a)(F_t(t)F_t(u)-F_t(u)) \quad \text{if } u<t.\end{aligned}$$

Similarly, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}F_t(t)=(b-a)(F_t(t)^2-F_t(t))+C_t^{-1}L(t,t)f(t).$$

Thus we have shown that

$$\frac{\mathrm{d}}{\mathrm{d}t}F_t(u)=(b-a)(F_t(t)F_t(u)-F_t(u))+C_t^{-1}L(t,t)f(t)\delta_u(t) \tag{4.12}$$

for $(u,t)\in\mathbb{R}_+^2$.

We obviously have $\check{\pi}_{t-} = \check{\pi}_t$ on the set $\{\Delta\hat{N}_t = 0\}$. Further

$$(1-\delta(\tau(H_{t-})))d(t, 1, B; H_{t-}) = (1-\delta(\tau(H_{t-})))\delta_u(t).$$

Thus we have, when $\Delta\hat{N}_t = 0$,

$$\begin{aligned}
&\text{r.h.s. of (4.11)}\\
&= (b-a)(\check{\pi}_t[\delta(\tau)]\check{\pi}_t(B) - \check{\pi}_t[\delta(\tau)1_B]) + h(t)\check{\pi}_t[(1-\delta(\tau))]\delta_u(t)\\
&= (b-a)(F_t(t)F_t(u) - F_t(u)) + h(t)C_t^{-1}L(t,t)\\
&\quad \times\left(\int_t^\infty f(s)\,\mathrm{d}s\right)\delta_u(t) \quad \text{(by (4.8))}\\
&= \text{r.h.s. of (4.12)} \quad \left(\text{by } h(t)\int_t^\infty f(s)\,\mathrm{d}s = f(t)\right).
\end{aligned}$$

But then (4.11) follows since the left hand sides of (4.11) and (4.12) are obviously equal. □

Remark. Consider $A = \{\tau < \infty\}$. Then $1_A = \delta(\tau)$ and, in particular, $1_A(H_t) = \delta(\tau(H_t)) = 1_{\{\sigma \le t\}}$. Thus the substitution of $A = \{\tau < \infty\}$ into (4.10) recovers the filter formula for the 'prevalence' probabilities $\hat{\pi}_t[\delta(\tau)] = \mathbb{P}(\sigma \le t \mid \mathcal{G}_t)$, $t \ge 0$, considered by the earlier authors:

$$\mathrm{d}\hat{\pi}_t[\delta(\tau)] = (h(t) - (b-a)\hat{\pi}_{t-}[\delta(\tau)])(1-\hat{\pi}_{t-}[\delta(\tau)])\,\mathrm{d}t + \frac{(b-a)\hat{\pi}_{t-}[\delta(\tau)](1-\hat{\pi}_{t-}[\delta(\tau)])}{a+(b-a)\hat{\pi}_{t-}[\delta(\tau)]}\,\mathrm{d}\hat{N}_t$$

(cf. Brémaud, 1981, IV.1.35, which contains a sign error; or Segall, Davis and Kailath, 1975, 4.12, where the distributional assumptions about σ are somewhat different).

4.2. Poisson process modulated by an alternating renewal process

Our second illustration is in some ways similar to the first, but it is more complicated: We consider an alternating renewal process (ARP) which modulates randomly the rate of a Poisson process. Only the latter is observed, and the task is to estimate which of the two states the underlying ARP is in, and what is the corresponding backward sojourn time in that state.

Our starting point is again the formulation of this problem in the MPP and filter framework. Suppose that the sojourns in the two states of the ARP are governed by absolutely continuous distributions F_1 and F_2. Let the corresponding hazard rates be h_1 and h_2, satisfying $F_i(t) = 1 - \exp\{-\int_0^t h_i(s)\,\mathrm{d}s\}$, $i = 1, 2$. Again we assume that h_1 and h_2 are bounded on finite intervals. The corresponding Poisson intensities (constants) are denoted by a_1 and a_2. We then define the underlying MPP (T, X) to have three possible marks: $E = \{0, 1, 2\}$, with 0 corresponding to a 'Poisson' point, and i corresponding to a 'state i' in the ARP, i.e., an initiation of a sojourn with distribution F_i, $i = 1, 2$.

Consequently, the definitions of $\hat{N}$, $\hat{E}$ and $c(x, y)$ are the same as in Section 4.1.

The statistical model is specified most conveniently in terms of the conditional intensities of (T, X) with respect to its internal history. Making explicit use of the process histories we set

$$U_1(H) := \inf\{t > 0 \mid (t, 1) \in H\},$$

$$U_i(H) := \inf\{t > U_{i-1}(H) \mid (t, \langle i\rangle) \in H\},$$

where

$$\langle i\rangle := \begin{cases} 1 & \text{if } i \text{ is odd,} \\ 2 & \text{otherwise,} \end{cases}$$

and

$$N(H) := \text{card}\{i \geq 1 \mid U_i(H) < \infty\}, \qquad Z(H) := \langle N(H)\rangle,$$

$$U(H) := \begin{cases} U_N(H) & \text{if } N(H) > 0, \\ 0 & \text{if } N(H) = 0. \end{cases}$$

The pair $(U(H), Z(H))$ can be interpreted as giving the last renewal time in H and the state of the ARP after H. In particular, the ARP is assumed to start from state 2.

The $\mathbb{F}$-intensities of the underlying process are then specified by the functions

$$\lambda(t, 0; H) := \sum_{i=1}^{2} a_i 1_{\{Z=i\}}(H),$$

$$\lambda(t, i; H) := h_{\langle i+1\rangle}(t - U(H)) 1_{\{Z=\langle i+1\rangle\}}(H), \quad i \in \{1, 2\},$$

by letting

$$\lambda_s(i) = \lambda(s, i; H_{s-}), \quad i = 0, 1, 2.$$

Clearly

$$\bar{\lambda}_t = \hat{\pi}_{t-}[\lambda(t, 0)] = \sum_{i=1}^{2} a_i \hat{\pi}_{t-}(Z = i).$$

The construction of the conditional distribution of (U, Z) given $\mathcal{G}_t$ is as follows: Let $\mathcal{A} := \sigma(\{U, Z\})$; clearly then $\mathcal{A} \subset \mathcal{H}$. We define the following random variables:

$$\beta_i(u, v) := (1 - F_i(v - u)) a_i^{\hat{N}_v - \hat{N}_u} e^{-a_i(v-u)}, \quad v \geq u \geq 0, \ i \in \{1, 2\},$$

$$L_n(u_1, \ldots, u_n; t) := \left[\prod_{j=1}^{n} h_{\langle j-1\rangle}(u_j - u_{j-1}) \beta_{\langle j-1\rangle}(u_{j-1}, u_j)\right] \beta_{\langle n\rangle}(u_n, t), \tag{4.13}$$

$n \in \mathbb{N}$, $0 =: u_0 \leq u_1 \leq \cdots \leq u_n \leq t$. Note that $L_n(u_1, \ldots, u_n; t)$ is $\mathcal{G}_t$-measurable, and

$$L_n(u_1, \ldots, u_{n-1}, t; t) = h_{\langle n-1\rangle}(t - u_{n-1}) L_{n-1}(u_1, \ldots, u_{n-1}; t). \tag{4.14}$$

We denote by ℓ_n the Lebesgue measure on $\mathbb{R}^n$, by $\mathbb{N}_2$ (resp. $\mathbb{N}_1$) the set of even (resp. odd) positive integers, and define the sets

$$A(n, t) := \{(u_1, \ldots, u_n) \in \mathbb{R}^n \mid 0 < u_1 < \cdots < u_n < t\}, \quad n \in \mathbb{N}, \ t \in (0, \infty).$$

For each $u \in \mathbb{R}_+$ and $t \in \mathbb{R}_+$ define

$$K_0(u; t) := \beta_2(0, t), \tag{4.15a}$$

$$K_n(u; t) := \int_{A(n, u\wedge t)} L_n(\cdot\,; t)\, \mathrm{d}l_n, \quad n \in \mathbb{N}, \tag{4.15b}$$

$$C_t := \sum_{n\geq 0} K_n(\infty; t) = \sum_{n\geq 0} K_n(t; t), \tag{4.15c}$$

$$G_t(u, x) := \begin{cases} C_t^{-1} \sum_{n\in\mathbb{N}_1} K_n(u; t) & \text{if } x = 1, \\ C_t^{-1}(K_0(u; t) + \sum_{n\in\mathbb{N}_2} K_n(u; t)) & \text{if } x = 2. \end{cases} \tag{4.15d}$$

Note that for fixed $t > 0$,

$$0 \leq L_n(u_1, \ldots, u_n; t) \leq \left[\prod_{j=1}^{n} f_{(j-1)}(u_j - u_{j-1})\right](a_1 \vee a_2)^{\hat{N}_t}$$

$$\leq \left[\prod_{j=1}^{n} (f_1 + f_2)(u_j - u_{j-1})\right](a_1 \vee a_2)^{\hat{N}_t},$$

hence

$$0 < C_t \leq (a_1 \vee a_2)^{\hat{N}_t}(R_1(t) + R_2(t)),$$

where $R_i(t)$ denotes the renewal function at t corresponding to the density f_i.

We conclude that for each $(t, \omega) \in \mathbb{R}_+ \times \Omega$ the series in (4.15c) converges, and (4.15d) defines a transition probability from $(\Omega, \mathcal{G}_t)$ to $((0, \infty) \times \{1\}) + (\mathbb{R}_+ \times \{2\})$. Note that for each (ω, t) the support of $G_t(\omega; \cdot, 1) + G_t(\omega; \cdot, 2)$ is a subset of $[0, t]$. Since $K_0(\cdot\,; t)$, and consequently $G_t(\cdot\,; 2)$, have a positive mass at 0, the symbol $\int_0^t$ is in the sequel taken to mean an integral over the closed interval $[0, t]$.

Since $(U, Z): \mathbb{H} \to ((0, \infty) \times \{1\}) + (\mathbb{R}_+ \times \{2\})$ is surjective, the formula

$$\check{\pi}_t(U \leq u, Z = x) = G_t(u, x) \tag{4.16}$$

defines a transition probability from $(\Omega, \mathcal{G}_t)$ to $(\mathbb{H}, \mathcal{A})$ such that G_t gives the $\check{\pi}_t$-distribution of the pair (U, Z). Since

$$\beta_i(u, t-) = a_i^{-\Delta\hat{N}_t} \beta_i(u, t) \quad \text{if } u < t, \tag{4.17a}$$

and accordingly,

$$L_n(u_1, \ldots, u_n; t-) = a_{(n)}^{-\Delta\hat{N}_t} L_n(u_1, \ldots, u_n; t) \quad \text{if } u < t, \tag{4.17b}$$

we see that the pathwise left-hand limits $K_n(u; t-)$, C_{t-}, $G_{t-}(u, x)$ and $\check{\pi}_{t-}$ exist. Furthermore, $\check{\pi}_{t-}$ is a transition probability from $(\Omega, \mathcal{G}_{t-})$ to $(\mathbb{H}, \mathcal{A})$, such that G_{t-} is the $\check{\pi}_{t-}$-distribution of the pair (U, Z). We denote $U_t := U \circ H_t$, the last renewal epoch $\leq t$, and $Z_t := Z \circ H_t$, the state of the ARP after this renewal epoch. Since the renewal time distributions are absolutely continuous, we have $\mathbb{P}$-a.s.

$$U_t = U_{t-} := U \circ H_{t-},$$
$$Z_t = \langle N_t(1) + N_t(2) \rangle = \langle N_{t-}(1) + N_{t-}(2) \rangle.$$

The following proposition now gives the wanted conditional distribution in an explicit form.

Proposition 2. $\mathbb{P}(U_t \leq u, Z_t = x \mid \mathcal{G}_t) = G_t(u, x)$, $u \in \mathbb{R}_+$, $x \in \{1, 2\}$.

Proof. The filter formula (3.1)–(3.3) now takes the form

$$\begin{aligned} \mathrm{d}\hat{\pi}_t(A) &= \sum_{x=0}^{2} \hat{\pi}_{t-}[\lambda(t,x)d(t,x,A)]\,\mathrm{d}t \\ &\quad + \{\hat{\lambda}_t^{-1}\hat{\pi}_{t-}[\lambda(t,0)g(t,0,A)] - \hat{\pi}_{t-}(A)\}(\mathrm{d}\hat{N}_t - \hat{\lambda}_t\,\mathrm{d}t) \\ &= \Bigg\{ \hat{\pi}_{t-}[\lambda(t,0)]\hat{\pi}_{t-}(A) - \hat{\pi}_{t-}[\lambda(t,0)1_A] \\ &\quad + \sum_{x=1}^{2} \hat{\pi}_{t-}[\lambda(t,x)d(t,x,A)] \Bigg\}\,\mathrm{d}t \\ &\quad + \left\{ \frac{\hat{\pi}_{t-}[\lambda(t,0)g(t,0,A)]}{\hat{\pi}_{t-}[\lambda(t,0)]} - \hat{\pi}_{t-}(A) \right\} \Delta\hat{N}_t. \end{aligned} \tag{4.18}$$

Again by Corollary 1 it suffices to check that for given $t > 0$ the constructed family $(\breve{\pi}_s)_{s>0}$ satisfies (4.18) when $A = B \cap D$, where $B := \{U \leq u\}$, $u \in \mathbb{R}_+$, and $D := \{Z = i\}$, $i \in \{1, 2\}$. We do the calculations only in the case $D = \{Z = 2\}$ since the case $D = \{Z = 1\}$ is similar.

(a) $\Delta\hat{N}_t = 1$: Since

$$g(t, 0, B \cap D; H) = 1_{B \cap D}(H), \quad H \in \mathbb{H}, \quad t \in \mathbb{R}_+,$$

it suffices to calculate

$$\begin{aligned} \frac{\breve{\pi}_{t-}[\lambda(t,0)1_{B\cap D}]}{\breve{\pi}_{t-}[\lambda(t,0)]} &= \frac{a_2\breve{\pi}_{t-}[\{Z=2\}\cap B\}}{\sum_{i=1}^{2} a_i\breve{\pi}_{t-}(Z=i)} \\ &= \left(\sum_{i=1}^{2} a_i G_{t-}(\infty, i)\right)^{-1} a_2 G_{t-}(u,2) \quad \text{(by (4.16))} \\ &= \left(\sum_{i=1}^{2} G_t(\infty, i)\right)^{-1} G_t(u,2) \quad \text{(by (4.15) and (4.17))} \\ &= G_t(u,2) = \breve{\pi}_t(B \cap D) \quad \text{(by (4.16))}. \end{aligned}$$

(b) $\Delta\hat{N}_t = 0$: Since $s \mapsto \hat{N}_s$ is constant in a neighbourhood of t, the following differentials exist:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\beta_i(u,t) &= -f_i(t-u)a_i^{\hat{N}_t - \hat{N}_u}\mathrm{e}^{-a_i(t-u)} - a_i(1 - F_i(t-u))a_i^{\hat{N}_t - \hat{N}_u}\mathrm{e}^{-a_i(t-u)} \\ &= -(h_i(t-u) + a_i)\beta_i(u,t), \end{aligned} \tag{4.19a}$$

and for $n\in\mathbb{N}$,

$$\frac{\mathrm{d}}{\mathrm{d}t}L_n(u_1,\ldots,u_n;t)=-(h_{\langle n\rangle}(t-u_n)+a_{\langle n\rangle})L_n(u_1,\ldots,u_n;t), \tag{4.19b}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}K_n(u;t)&=\frac{\mathrm{d}}{\mathrm{d}t}\int_0^{u\wedge t}\left[\int_{A(n-1,u_n)}L_n(\cdot,u_n;t)\,\mathrm{d}\ell_{n-1}\right]\mathrm{d}u_n\\
&=\delta_u(t)\int_{A(n-1,t)}L_n(\cdot,t;t)\,\mathrm{d}\ell_{n-1}\\
&\quad+\int_0^{u\wedge t}\left[\int_{A(n-1,u_n)}\frac{\mathrm{d}}{\mathrm{d}t}L_n(\cdot,u_n;t)\,\mathrm{d}l_{n-1}\right]\mathrm{d}u_n\\
&=\delta_u(t)\int_{A(n-1,t)}h_{\langle n-1\rangle}(t-u_{n-1})\\
&\qquad\qquad\times L_{n-1}(u_1,\ldots,u_{n-1};t)\,\mathrm{d}u_1,\ldots,\mathrm{d}u_{n-1}\\
&\quad-\int_{A(n,u\wedge t)}(h_{\langle n\rangle}(t-u_n)+a_{\langle n\rangle})L_n(u_1,\ldots,u_n;t)\,\mathrm{d}u_1,\ldots,\mathrm{d}u_n\\
&\quad\text{(by (4.14) and (4.19b))}\\
&=\delta_u(t)\int_0^t h_{\langle n-1\rangle}(t-s)K_{n-1}(\mathrm{d}s;t)\\
&\quad-\int_0^u h_{\langle n\rangle}(t-s)K_n(\mathrm{d}s;t)-a_{\langle n\rangle}K_n(u;t)\\
&\quad\text{(by (4.15a,b)).}\end{aligned} \tag{4.19c}$$

The formulas (4.19) imply

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}C_t&=\sum_{n\geq 0}\frac{\mathrm{d}}{\mathrm{d}t}K_n(t;t)\\
&=-(h_2(t)+a_2)\beta_2(0,t)\\
&\quad+\sum_{n\geq 1}\left\{\int_0^t h_{\langle n-1\rangle}(t-s)K_{n-1}(\mathrm{d}s;t)\right.\\
&\qquad\left.-\int_0^t h_{\langle n\rangle}(t-s)K_n(\mathrm{d}s;t)-a_{\langle n\rangle}K_n(\infty;t)\right\}\\
&=-a_2\beta_2(0,t)-\lim_{n\to\infty}\int_0^t h_{\langle n\rangle}(t-s)K_n(\mathrm{d}s;t)-\sum_{n=1}^{\infty}a_{\langle n\rangle}K_n(\infty;t)\\
&\qquad\left(\text{since }\int_0^t h_{\langle 2\rangle}(t-s)K_0(\mathrm{d}s;t)=h_2(t)\beta_2(0,t)\right)\\
&=-a_2\beta_2(0,t)-a_1\sum_{n\in\mathbb{N}_1}K_n(\infty;t)-a_2\sum_{n\in\mathbb{N}_2}K_n(\infty;t)\\
&=-C_t(a_1G_t(\infty,1)+a_2G_t(\infty,2))\quad\text{(by (4.15)),}\end{aligned}$$

and, similarly,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(K_0(u;t)+\sum_{n\in\mathbb{N}_2}K_n(u;t)\right)$$

$$=-(h_2(t)+a_2)\beta_2(0,t)$$

$$+\sum_{n\in\mathbb{N}_2}\left\{\delta_u(t)\int_0^u h_{\langle n-1\rangle}(t-s)K_{n-1}(\mathrm{d}s;t)\right.$$

$$\left.-\int_0^u h_{\langle n\rangle}(t-s)K_n(\mathrm{d}s;t)-a_{\langle n\rangle}K_n(u;t)\right\}$$

$$=C_t\left[\delta_u(t)\int_0^u h_1(t-s)G_t(\mathrm{d}s,1)-\int_0^u h_2(t-s)G_t(\mathrm{d}s,2)-a_2G_t(u,2)\right].$$

Thus we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\check{\pi}_t(B\cap D)=\frac{\mathrm{d}}{\mathrm{d}t}G_t(u,2)$$

$$=C_t^{-2}\left[\left(\frac{\mathrm{d}}{\mathrm{d}t}\sum_{n\in\mathbb{N}_2\cup\{0\}}K_n(u;t)\right)C_t\right.$$

$$\left.-\frac{\mathrm{d}}{\mathrm{d}t}C_t\left(\sum_{n\in\mathbb{N}_2\cup\{0\}}K_n(u;t)\right)\right]$$

$$=\delta_u(t)\int_0^u h_1(t-s)G_t(\mathrm{d}s,1)-\int_0^u h_2(t-s)G_t(\mathrm{d}s,2)$$

$$-a_2G_t(u,2)+(a_1G_t(\infty,1)+a_2G_t(\infty,2))G_t(u,2)$$

$$=\delta_u(t)\check{\pi}_t[h_1(t-U)1_{\{Z=1\}}]-\check{\pi}_t[h_2(t-U)1_{B\cap D}]$$

$$+\left(\sum_{i=1}^2 a_i\check{\pi}_t(Z=i)\right)\check{\pi}_t(B\cap D)-a_2\check{\pi}_t(B\cap D)$$

(note that $\check{\pi}_t(B)=1$ if $\delta_u(t)=1$)

$$=\sum_{i=1}^2\check{\pi}_{t-}[\lambda(t,i)d(t,i,B\cap D)]$$

$$+\check{\pi}_{t-}[\lambda(t,0)]\check{\pi}_{t-}(B\cap D)-\check{\pi}_{t-}[\lambda(t,0)1_{B\cap D}]. \tag{4.20}$$

For the last equality in (4.20) use the following facts:

(i) $\check{\pi}_{t-}=\check{\pi}_t$ by (4.18) and the assumption $\Delta\hat{N}_t=0$;

(ii) for $H\in H_{t-}(\Omega)$ we have

$$\lambda(t,1;H)d(t,1,B\cap D;H)=-h_2(t-U(H))1_{B\cap D}(H)$$

since $H\cup\{(t,1)\}\notin D$;

(iii) for $H\in H_{t-}(\Omega)$ we have

$$\lambda(t,2;H)d(t,2,B\cap D;H)=\delta_u(t)h_1(t-U(H))1_{\{Z=1\}}(H)$$

since $Z(H)=2$ implies $d(t,2,B\cap D;H)=0$ and $Z(H)=1$ implies that the equivalence $H\cup\{(t,2)\}\in B\cap D\Leftrightarrow t\leq u$ holds. □

References

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

E. Arjas and P. Haara, Observation scheme and likelihood (1991), submitted for publication.

P. Brémaud, La méthode des semi-martingales en filtrage lorsque l'observation est un processus ponctuel marqué, in: Sem. Probab. IX, Lecture Notes in Math. No. 511 (Springer, Berlin, 1975) pp. 1–18.

P. Brémaud, Point Processes and Queues (Springer, New York, 1981).

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

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

C. Dellachierie and P.-A. Meyer, Probabilities and Potential B (North-Holland, Amsterdam, 1982).

L.I. Galchuk and B.L. Rozovsky, The disorder problem for a Poisson process, Theory Probab. Appl. 16 (1971) 712–716.

R.K. Getoor, On the construction of kernels, in: Sem. Probab. IX, Lecture Notes in Math. No. 511 (Springer, Berlin, 1975) pp. 443–463.

D.I. Hadjiev, On the filtering of semimartingales in case of observations of point processes, Theory Probab. Appl. 23 (1978) 169–178.

W.H. Kliemann, G. Koch and F. Marchetti, On the unnormalized solution of the filtering problem with counting process observations, IEEE Trans. Inform. Theory. (1990).

G. Koch, A dynamical approach to reliability theory, in: Proc. Int. School of Phys. 'Enrico Fermi' CXIV (North-Holland, Amsterdam, 1986) pp. 215–240.

I. Norros, A compensator representation of multivariate life length distributions, with applications, Scand. J. Statist. 13 (1986) 99–112.

D. Revuz, Markov Chains (North-Holland, Amsterdam, 1975).

A. Segall, M.H.A. Davis and T. Kailath, Nonlinear filtering with counting observations, IEEE Trans. Inform. Theory. IT-21 (1975) 143–149.

M. Yor, Sur les theories du filtrage et de la prediction, in: Sem. Probab. XI, Lecture Notes in Math. No. 581 (Springer, Berlin, 1977) pp. 257–297.