ISSN 1399-2538 MPS-LN 3 www.maphysto.dk

Marked Point Processes and Piecewise Deterministic Processes

Martin Jacobsen
Department of Theoretical Statistics, University of Copenhagen,
MaPhySto[1]

[1]Centre for Mathematical Physics and Stochastics, funded by a grant from the Danish National Research Foundation.

Contents

Preface

A first version of these lecture notes was prepared for a graduate course given at the University of Copenhagen 1995–96. Visits to the University of Aarhus at the Centre for Mathematical Physics and Stochastics during the autumn of 1998 and at the Stochastic Centre, Chalmers University of Technology and the University of Gothenburg in early 1999, made it possible to prepare the version presented here.

In some respects the notes represent a modern version of the lecture notes Jacobsen [10] and deal with the time dynamics of point processes. The focus is on canonical marked point processes (MPP's) (equivalently, MPP's on filtered probablity spaces with the filtration that generated by the process itself), and their use in the study of piecewise deterministic processes (PDP's), notably piecewise deterministic Markov processes (PDMP's). This approach conforms with the author's attitude that all understanding of MPP's stems from an understanding of the canonical setup! Nevertheless, MPP's adapted to more general filtrations are also discussed, and some of the main structural differences between the canonical and other setups are pointed out.

The MPP part of the notes (Chapters 1 to 5) constructs MPP's from the regular conditional distributions generating jump times, respectively marks, given the past history. Compensators and compensating measures are then defined, it is shown (in the canonical case) that the compensating measure determines the distribution of the MPP, and the basic martingales characterizing the compensators are derived. A martingale representation theorem is also given, together with a form of Itô's formula for MPP's. The final general result presents a simple necessary and sufficient condition for local absolute continuity between the distributions of two different MPP's.

Among the examples discussed are Markov chains (homogeneous and non-homogeneous) viewed as MPP's. Chapter 6 shows how to generalize this to PDMP's, where the basic theory is developed from scratch and the Markov property established through a key lemma already used heavily for the MPP theory. In a certain sense, a PDP is nothing but a process adapted to the filtration generated by a MPP. For the process to be Markov (homogeneous or non-homogeneous) a special structure is of course required, as discussed in Chapter 6, but once that structure is present, martingale properties, Itô's formula, likelihood processes etc. for PDMP's, are immediately available from the MPP theory.

Most of the results presented in these notes are certainly well known. Perhaps though, some of the proofs and the approach used to develop parts of the theory is different from what has been seen elsewhere.

The notes are intended for courses at graduate to post-doc level. They are still in a somewhat preliminary form, e.g. with some proofs only sketched, and doubtless contain a number of typographical and other errors.

The short bibliography at the end of the notes mentions only some of the most important references on counting processes, marked point processes and piecewise deterministic Markov processes. Most of the items are monographs including several only parts of which are directly relevant for the subject matter of the notes. For detailed bibliographies, see Last and Brandt [16] (MPP's) and Davis [8] (PDMP's).

I am very grateful for the hospitality and fine working conditions I enjoyed while visiting MaPhySto in Aarhus and the Stochastic Centre in Gothenburg. Special thanks are due to Jacob Krabbe Pedersen from the Department of Theoretical Statistics at the University of Aarhus, who voluntarily undertook the huge task of converting the handwritten manuscript of the original Copenhagen notes into LaTeX.

Copenhagen, July 1999
Martin Jacobsen

Chapter 1

Simple and marked point processes

1.1 The definition of SPP's and MPP's

Let $(\Omega, \mathcal{F}, P)$ be a probability space: Ω a non-empty set, the sample space, $\mathcal{F}$ a $\sigma-$algebra of subsets of Ω and P a probability measure on $\mathcal{F}$.

Definition 1.1.1 *A* simple point process *(SPP for short) is a sequence* $\mathcal{T} = (T_n)_{n\geq 1}$ *of* $\overline{\mathbb{R}}_+ -$*valued random variables defined on* $(\Omega, \mathcal{F}, P)$ *such that*

(i) $P(T_1 \leq T_2 \leq \cdots) = 1,$

(ii) $P(T_n < T_{n+1}, T_n < \infty) = P(T_n < \infty) \quad (n \geq 1),$

(iii) $P\left(\lim_{n\to\infty} T_n = \infty\right) = 1.$

Notation. The intervals referring to the time axis are denoted as follows: $\mathbb{R}_0 = [0, \infty[\,,\ \mathbb{R}_+ =]0, \infty[\,,\ \overline{\mathbb{R}}_0 = [0, \infty]\,,\ \overline{\mathbb{R}}_+ =]0, \infty]$. The corresponding Borel $\sigma-$algebras are written $\mathcal{B}_0$, $\mathcal{B}_+$, $\overline{\mathcal{B}}_0$, $\overline{\mathcal{B}}_+$.

Thus, a SPP is an almost surely increasing sequence of strictly positive, possibly infinite random variables, strictly increasing as long as they are finite and with almost sure limit ∞. The interpretation is that the T_n mark the occurrence in time of some event, the n'th occurrence happening at time T_n if $T_n < \infty$ and less than n events occurring altogether (on the time axis $\mathbb{R}_0$) if $T_n = \infty$. By the definition, no event can happen at time 0, nevertheless we shall mostly use $\mathbb{R}_0$ (rather than $\mathbb{R}_+$) as time axis.

The condition (iii) in Definition 1.1.1 is important. It is equivalent to the statement that only finitely many events can occur in any finite time interval. The more general class of SPP's obtained by retaining only (i) and (ii) from Definition 1.1.1 is the class of simple point processes with *explosion*. It will be denoted $\mathrm{SPP}_{\mathrm{ex}}$ and will be discussed further in Section 2.1 below.

Introducing the sequence space

$$K = \left\{ (t_n)_{n \geq 1} \in \overline{\mathbb{R}}_+^{\mathbb{N}} : t_1 \leq t_2 \leq \cdots \uparrow \infty,\ t_n < t_{n+1} \text{ if } t_n < \infty \right\}$$

together with the $\sigma-$algebra $\mathcal{K}$ generated by the coordinate projections $T_n^\circ (t_1, t_2, \ldots) = t_n \quad (n \geq 1)$, we may view the SPP $\mathcal{T}$ as a $(K, \mathcal{K})-$valued random variable, defined $P-$a.s. The *distribution of* $\mathcal{T}$ is the probability $\mathcal{T}(P)$ on $(K, \mathcal{K})$ obtained by transformation,

$$\mathcal{T}(P)(B) = P\{\omega : \mathcal{T}(\omega) \in B\} \quad (B \in \mathcal{K}).$$

Similarly, introducing

$$\overline{K} = \left\{ (t_n)_{n \geq 1} \in \overline{\mathbb{R}}_+^{\mathbb{N}} : t_1 \leq t_2 \leq \cdots,\ t_n < t_{n+1} \text{ if } t_n < \infty \right\}$$

with $\overline{\mathcal{K}}$ the $\sigma-$algebra generated by the coordinate projections on $\overline{K}$, the distribution of a $\mathrm{SPP}_{\mathrm{ex}}$, $\overline{\mathcal{T}}$, is the transformed probability $\overline{\mathcal{T}}(P)$ on $\left(\overline{K}, \overline{\mathcal{K}}\right)$. Note that $K = \left\{ (t_n) \in \overline{K} : \lim_{n \to \infty} t_n = \infty \right\}$.

Now suppose also given a measurable space $(E, \mathcal{E})$, the *mark space*. Adjoin to E the *irrelevant mark* ∇, write $\overline{E} = E \cup \{\nabla\}$ and let $\overline{\mathcal{E}} = \sigma(\mathcal{E}, \{\nabla\})$ denote the $\sigma-$algebra of subsets of $\overline{E}$ generated by the measurable subsets of E and the singleton $\{\nabla\}$.

Definition 1.1.2 *A* marked point process *(MPP for short) with mark space E, is a double sequence $(\mathcal{T}, \mathcal{Y}) = \left((T_n)_{n \geq 1}, (Y_n)_{n \geq 1}\right)$ of $\overline{\mathbb{R}}_+-$valued random variables T_n and $\overline{E}-$valued random variables Y_n defined on $(\Omega, \mathcal{F}, P)$ such that $\mathcal{T} = (T_n)$ is a SPP and*

(i) $P(Y_n \in E, T_n < \infty) = P(T_n < \infty)$,

(ii) $P(Y_n = \nabla, T_n = \infty) = P(T_n = \infty)$.

Thus, as in Definition 1.1.1 we have a sequence of time points marking the occurrence of events, but now these events may be of different types, with the type (or name or label) of the n'th event denoted by the n'th mark, Y_n. Note that the irrelevant mark appears only for events that never occur. It was introduced only in order to have Y_n always defined for all n.

A MPP $(\mathcal{T}, \mathcal{Y})$ may be viewed as a $(K_E, \mathcal{K}_E)$ −valued random variable, where

$$K_E = \left\{ ((t_n), (y_n)) \in \overline{\mathbb{R}}_+^{\mathbb{N}} \times \overline{E}^{\mathbb{N}} : (t_n) \in K,\ y_n \in E \text{ iff } t_n < \infty \right\}$$

with $\mathcal{K}_E$ the σ−algebra of subsets of K_E generated by the coordinate projections $T_n^\circ((t_k), (y_k)) = t_n$, $Y_n^\circ((t_k), (y_k)) = y_n$. The *distribution of* $(\mathcal{T}, \mathcal{Y})$ is then the transformed probability $(\mathcal{T}, \mathcal{Y})(P)$ on $(K_E, \mathcal{K}_E)$.

MPP's with *explosion* are introduced in the obvious manner as $\left(\overline{K}_E, \overline{\mathcal{K}}_E\right)$− valued random variables, where

$$\overline{K}_E = \left\{ ((t_n), (y_n)) \in \overline{\mathbb{R}}_+^{\mathbb{N}} \times \overline{E}^{\mathbb{N}} : (t_n) \in \overline{K},\ y_n \in E \text{ iff } t_n < \infty \right\}$$

with $\overline{\mathcal{K}}_E$ the σ−algebra generated by the projections on $\overline{K}_E$. The distribution of a MPP_{ex} $\left(\overline{\mathcal{T}}, \mathcal{Y}\right)$ is of course the probability $\left(\overline{\mathcal{T}}, \mathcal{Y}\right)(P)$ on $\left(\overline{K}_E, \overline{\mathcal{K}}_E\right)$.

1.2 Counting processes and counting measures

Let $\mathcal{T} = (T_n)_{n \geq 1}$ be a SPP and define the *counting process* (CP) associated with $\mathcal{T}$ as $N = (N_t)_{t \geq 0}$, where

$$N_t = \sum_{n=1}^{\infty} 1_{(T_n \leq t)}. \tag{1.1}$$

Thus N_t counts the number of events in the time interval $[0, t]$ with $N_0 \equiv 0$. Clearly each N_t is a $\mathbb{N}_0$−valued random variable. Also, for P−a.a. ω the sample path $t \to N_t(\omega)$ belongs to the space W of *counting process paths,*

$$W = \left\{ w \in \mathbb{N}_0^{\mathbb{R}_0} : \begin{array}{l} w(0) = 0,\ w \text{ is right-continuous, increasing,} \\ \Delta w(t) = 0 \text{ or } 1 \text{ for all } t \end{array} \right\}.$$

Notation. $\mathbb{N}_0$ denotes the non-negative integers $\mathbb{N} \cup \{0\}$. We also write $\overline{\mathbb{N}}_0 = \mathbb{N} \cup \{\infty\}$, $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$.

If $t \to f(t)$ is a *cadlag*-function (right-continuous with left limits) such as any $w \in W$, Δf is the function of discontinuities for f, $\Delta f(t) = f(t) - f(t-)$.

Note that it is the assumption that the T_n be strictly increasing as long as they are finite (Definition 1.1.1 (ii)), that ensures that $t \to N_t$ increases only in jumps of size 1.

On W define the *canonical counting process* $N^\circ = (N_t^\circ)_{t \geq 0}$ by $N_t^\circ(w) = w(t)$, so in fact $N^\circ : W \to W$ is just the identity, and let $\mathcal{H} = \sigma(N_t^\circ)_{t \geq 0}$ be the smallest $\sigma-$algebra of subsets of W such that all N_t° are measurable. Then we may view N defined by (1.1) as a $(W, \mathcal{H})-$valued random variable with distribution $Q = N(P)$ the probability on $(W, \mathcal{H})$ obtained by transformation of P.

Note that $\mathcal{T}$ is easily recovered from N since a.s.

$$\begin{aligned} T_n &= \inf\{t \geq 0 : N_t = n\}, \\ (T_n \leq t) &= (N_t \geq n), \end{aligned} \tag{1.2}$$

where here as elsewhere we define $\inf \emptyset = \infty$. Thus we have shown that *any SPP may be identified with its associated CP.*

The discussion above could also be carried out for $\overline{\mathcal{T}}$ a $\mathrm{SPP}_{\mathrm{ex}}$, only now $N_t = \infty$ can occur with probability > 0, and W should be replaced by the space $\overline{W}$ of paths $\overline{w} \in \overline{\mathbb{N}}_0^{\mathbb{R}_0}$ that otherwise satisfy the conditions already imposed on $w \in W$.

We have seen that a SPP can be identified with a counting process. In a similar fashion, a MPP $(\mathcal{T}, \mathcal{Y}) = ((T_n), (Y_n))$ can be identified with a *random counting measure* (RCM) μ, viz.

$$\mu = \sum_{n \in \mathbb{N}: T_n < \infty} \varepsilon_{(T_n, Y_n)}. \tag{1.3}$$

Here $\varepsilon_{(T_n,Y_n)}(\omega) = \varepsilon_{(T_n(\omega),Y_n(\omega))}$ is the measure on the product space $(\mathbb{R}_0 \times E, \mathcal{B}_0 \otimes \mathcal{E})$ attaching unit mass to the point $(T_n(\omega), Y_n(\omega))$ and identically 0 elsewhere. Thus, for $P-$a.a. ω, $\mu(\omega)$ is a *discrete counting measure* on $(\mathbb{R}_0 \times E, \mathcal{B}_0 \otimes \mathcal{E})$, i.e. $\mu(\omega)$ is a positive $\sigma-$finite measure such that

$$\begin{aligned} &\mu(\omega, C) \in \overline{\mathbb{N}}_0 \quad (C \in \mathcal{B}_0 \otimes \mathcal{E}), \\ &\mu(\omega, \{0\} \times E) = 0, \\ &\mu(\omega, \{t\} \times E) \leq 1 \quad (t \geq 0), \\ &\mu(\omega, [0,t] \times E) < \infty \quad (t \geq 0). \end{aligned}$$

The identity

$$\mu(C) = \sum_{n=1}^{\infty} 1_C(T_n, Y_n) \quad (C \in \mathcal{B}_0 \otimes \mathcal{E})$$

shows that for all measurable C, $\mu(C)$ is a $\overline{\mathbb{N}}_0-$valued random variable.

We shall denote by $\mathcal{M}$ the space of discrete counting measures on $(\mathbb{R}_0 \times E, \mathcal{B}_0 \otimes \mathcal{E})$. Elements in $\mathcal{M}$ are denoted m and we write μ° for the identity map on $\mathcal{M}$. For $C \in \mathcal{B}_0 \otimes \mathcal{E}$, $\mu^\circ(C)$ denotes the function $m \to m(C)$ from $\mathcal{M}$ to $\overline{\mathbb{N}}_0$.

On $\mathcal{M}$ we use the $\sigma-$algebra $\mathcal{H} = \sigma(\mu^\circ(C))_{C \in \mathcal{B}_0 \otimes \mathcal{E}}$, the smallest $\sigma-$algebra such that all $\mu^\circ(C)$ are measurable. Thus, with μ the RCM above,

μ becomes an a.s. defined $(\mathcal{M},\mathcal{H})$ −valued random variable. Its distribution is the probability $Q=\mu(P)$ on $(\mathcal{M},\mathcal{H})$.

It is also possible to describe $(\mathcal{T},\mathcal{Y})$ through a family of counting processes: for $A\in\mathcal{E}$ define $N(A)=(N_t(A))_{t\geq 0}$ by

$$N_t(A)=\sum_{n=1}^{\infty} 1_{(T_n\leq t, Y_n\in A)}$$

so $N_t(A)$ counts the number of events on $[0,t]$ matching a mark belonging to the set A. Note that

$$N_t(A)=\mu([0,t]\times A).$$

The total number of events on $[0,t]$ is denoted $\overline{N}_t$,

$$\overline{N}_t=N_t(E)=\sum_{n=1}^{\infty} 1_{(T_n\leq t)},$$

(cf. (1.1)) and $\overline{N}$ is the counting process $\left(\overline{N}_t\right)_{t\geq 0}$.

It is easily verified that $\mathcal{H}=\sigma\left(N_t(A)\right)_{t\geq 0, A\in\mathcal{E}}$.

We shall now show how the T_n and Y_n may be recovered in a measurable fashion from μ (or rather, the counting processes $N(A)$). Clearly

$$\begin{aligned} T_n &= \inf\left\{t\geq 0:\overline{N}_t=n\right\}, \\ (T_n\leq t) &= \left(\overline{N}_t\geq n\right), \end{aligned} \tag{1.4}$$

cf. (1.2). It is more tricky to find the Y_n, however using the right-continuity of counting processes one finds that

$$(Y_n\in A)=\bigcup_{K_0=1}^{\infty}\bigcap_{K=K_0}^{\infty}\bigcup_{k=1}^{\infty}\left(\overline{N}_{(k-1)/2^K}=n-1, N_{k/2^K}(A)-N_{(k-1)/2^K}(A)=1\right),$$

showing that if μ is a random variable, so is Y_n. Note however that without some further structure on $(E,\mathcal{E})$ it may not be possible to compute $Y_n(\omega)$ from $\mu(\omega)$. It is possible if e.g. either all singletons $\{y\}\in\mathcal{E}$ or more generally if $\mathcal{E}$ separates points in the sense that for all $y\neq y'\in E$ there exists $A\in\mathcal{E}$ such that $y\in A$, $y'\in A^c$. With this kind of structure imposed on $(E,\mathcal{E})$, we have seen that *any MPP may be identified with its associated RCM.*

If E (and $\mathcal{E}$) is uncountable, it is much more convenient to identify $(\mathcal{T},\mathcal{Y})$ with the RCM μ rather dthan the collection $(N(A))_{A\in\mathcal{E}}$. If however E is at most countably infinite with $\mathcal{E}$ comprising all subsets of E, it is enough to keep track of just $N^y:=N^{\{y\}}$ for all $y\in E$ since $N(A)=\sum_{y\in A}N^y$.

It is of course possible to identify $\text{MPP}_{\text{ex}}^{\prime}$s with suitable exploding RCM's where $\overline{N}_t=\infty$ is possible. We shall not go into the details.

Chapter 2

Construction of SPP's and MPP's

2.1 Creating SPP's

We shall construct SPP's by constructing probabilities on the sequence space $(K, \mathcal{K})$, see Section 1.1. We shall call a probability on $(K, \mathcal{K})$ a *canonical SPP*. Through the bimeasurable bijection $\varphi : (K, \mathcal{K}) \to (W, \mathcal{H})$ given by (cf. (1.1), (1.2))

$$N_t^\circ (\varphi (t_1, t_2, \ldots)) = \sum_{n=1}^{\infty} 1_{[0,t_n]}(t) \quad ((t_1, t_2, \ldots) \in K),$$
$$T_n^\circ \left(\varphi^{-1}(w)\right) = \inf \{t \geq 0 : N_t^\circ(w) = n\} \quad (w \in W), \tag{2.1}$$

we at the same time obtain a construction of *canonical counting processes*, i.e. probabilities on $(W, \mathcal{H})$.

The reader is reminded that if $(D_i, \mathcal{D}_i)$ for $i = 1, 2$ are measurable spaces, a *Markov kernel* or *transition probability* from $(D_1, \mathcal{D}_1)$ to $(D_2, \mathcal{D}_2)$ is a map $p : D_1 \times \mathcal{D}_2 \to [0, 1]$ such that

(i) $x_1 \to p_{x_1}(A_2)$ is $\mathcal{D}_1$ – measurable for all $A_2 \in \mathcal{D}_2$,
(ii) $A_2 \to p_{x_1}(A_2)$ is a probability on $(D_2, \mathcal{D}_2)$ for all $x_1 \in D_1$.

Markov kernels in particular serve as regular conditional probabilities.

The idea underlying the construction of the distribution of a SPP $\mathcal{T} = (T_n)$ is to start by specifying the marginal distribution of T_1 and then, successively for each $n \in \mathbb{N}$, the conditional distribution of T_{n+1} given $(T_1, \ldots, T_n)$. More precisely, let

$$K^{(n)} = \{(t_1, \ldots, t_n) : 0 < t_1 \leq \cdots \leq t_n \leq \infty,\ t_k < t_{k+1} \text{ if } t_k < \infty\}$$

be the space of $n-$sequences that can appear as the first n coordinates of an element in K, equipped with the $\sigma-$algebra $\mathcal{K}^{(n)}$ spanned by the coordinate projections (equivalently, the trace on $K^{(n)}$ of the $n-$dimensional Borel $\sigma-$algebra).

Assume given a probability $P^{(0)}$ on $\overline{\mathbb{R}}_+$ and also for every $n \in \mathbb{N}$, a Markov kernel $P^{(n)}$ from $\left(K^{(n)}, \mathcal{K}^{(n)}\right)$ to $\left(\overline{\mathbb{R}}_+, \overline{\mathcal{B}}_+\right)$.

Notation. Write z_n for a typical point $(t_1, \ldots, t_n) \in K^{(n)}$ and $\overline{T}_n^\circ$ for the projection $\overline{T}_n^\circ((t_k)) = t_n$ on $\overline{K}$.

Theorem 2.1.1 *(a) For every choice of the probability $P^{(0)}$ and the Markov kernels $P^{(n)}$ for $n \geq 1$ satisfying*

$$\begin{array}{ll} P_{z_n}^{(n)}(]t_n, \infty]) = 1 & \text{if } t_n < \infty \\ P_{z_n}^{(n)}(\{\infty\}) = 1 & \text{if } t_n = \infty, \end{array} \tag{2.2}$$

there is a unique probability $\overline{R}$ on the sequence space $\left(\overline{K}, \overline{\mathcal{K}}\right)$ allowing explosions, such that $\overline{T}_1^\circ(\overline{R}) = P^{(0)}$ and for every $n \geq 1$, $z_n \in K^{(n)}$, the probability $P_{z_n}^{(n)}(\cdot)$ is a regular conditional distribution of $\overline{T}_{n+1}^\circ$ given $\left(\overline{T}_1^\circ, \ldots, \overline{T}_n^\circ\right) = z_n$.

(b) $\overline{R}$ defines a canonical SPP R, i.e. $\overline{R}(K) = 1$ with R the restriction to K of $\overline{R}$, if and only if

$$\overline{R}\left(\lim_{n\to\infty} \overline{T}_n^\circ = \infty\right) = 1. \tag{2.3}$$

We shall not give the proof here. The theorem follows easily from the Kolmogorov consistency theorem by showing consistency of the finite-dimensional distributions, cf. the expression (2.8) below, which refers to the MPP case.

Remark 2.1.1 *It should be clear that (2.2) ensures that the sequence $\left(\overline{T}_n^\circ\right)$ is increasing a.s., strictly increasing as long as $\overline{T}_n^\circ$ is finite. And it is also clear that (2.3) is exactly the condition ensuring that no explosion occurs.*

Note that several choices of Markov kernels may lead to the same R: subject only to the measurability conditions, $P_{z_n}^{(n)}$ may be changed arbitrarily for $z_n \in B^{(n)} \in \mathcal{K}^{(n)}$ provided $R\left((T_1^\circ, \ldots T_n^\circ) \in B^{(n)}\right) = 0$.

While arbitrary choices of $P^{(n)}$ for $n \geq 0$ lead to possibly exploding SPP's, there is no simple characterization of the $P^{(n)}$ that result in genuine SPP's. Indeed, it may be extremely difficult to decide whether a canonical SPP_{ex} is a true SPP or not. This stability problem *will be discussed on various occasions later in these notes.*

Notation. We shall often describe the $P^{(n)}_{z_n}$ through their *survivor functions* $\overline{P}_{z_n}$,

$$\overline{P}_{z_n}(t) = P^{(n)}_{z_n}\left(]t, \infty]\right).$$

By (2.2), $\overline{P}^{(n)}_{z_n}(t) = 1$ for $t \leq t_n$ if $t_n < \infty$, so it suffices to give $\overline{P}^{(n)}_{z_n}(t)$ for $t \geq t_n$, imposing the value 1 for $t = t_n < \infty$.

Example 2.1.1 *The* dead process *is the canonical SPP with no jumps in finite time: $R\left(T^\circ_1 = \infty\right) = 1$. It is completely specified by the requirement $P^{(0)} = \varepsilon_\infty$ while the choice of $P^{(n)}$ for $n \geq 1$ is immaterial, cf. the remarks following Theorem 2.1.1.*

Example 2.1.2 *The canonical* Poisson process *(SPP version) is the probability R that makes the waiting times $V^\circ_n := T^\circ_n - T^\circ_{n-1}$ for $n \geq 1$ (with $T^\circ_0 \equiv 0$) independent and identically distributed (iid), exponential with some rate $\lambda > 0$. Thus*

$$\overline{P}^{(0)}(t) = e^{-\lambda t} \quad (t \geq 0),$$

$$\overline{P}^{(n)}_{z_n}(t) = e^{-\lambda(t-t_n)} \quad (t \geq t_n).$$

The corresponding probability $Q = \phi(R)$ on $(W, \mathcal{H})$ makes N° into a homogeneous Poisson process with parameter λ: for $s < t$, $N^\circ_t - N^\circ_s$ is independent of $\left(N^\circ_u\right)_{u \leq s}$ and

$$Q\left(N^\circ_t - N^\circ_s = n\right) = \frac{\left(\lambda\left(t-s\right)\right)^n}{n!} e^{-\lambda(t-s)} \quad \left(n \in \mathbb{N}_0\right).$$

This familiar fact will appear as a consequence of Examples 3.8.1 and 3.8.3 below.

The dead process may be viewed as the Poisson process with parameter $\lambda = 0$.

Example 2.1.3 *A canonical* renewal process *is a canonical SPP such that the waiting times V°_n are iid. If $\overline{G}$ is the survivor function for the waiting time distribution,*

$$\overline{P}^{(0)}(t) = \overline{G}(t) \quad (t \geq 0),$$

$$\overline{P}^{(n)}_{z_n}(t) = \overline{G}\left(t - t_n\right) \quad (t \geq t_n).$$

The Poisson process is in particular a renewal process. That renewal processes do not explode is of course a consequence of the simple fact that if $U_n > 0$ for $n \geq 1$ are iid random variables, then $\sum U_n = \infty$ a.s.

Example 2.1.4 *Suppose the waiting times V_n° are independent, V_n° exponential at rate $\lambda_{n-1} \geq 0$. Thus*

$$\overline{P}^{(0)}(t) = e^{-\lambda_0 t} \quad (t \geq 0),$$

$$\overline{P}_{z_n}^{(n)}(t) = e^{-\lambda_n (t-t_n)} \quad (t \geq t_n).$$

If $\lambda_n = 0$ for some $n \geq 0$ and n_0 is the smallest such n, precisely n_0 jumps occur and $T_{n_0+1}^\circ = \infty$ a.s.

With this example explosion may be possible and either happens with probability 0 (and we have a SPP) or with probability 1 (and we have a genuine SPP_{ex}). Stability (no explosion) occurs iff $\sum_{n\geq 1} EV_n^\circ = \infty$, i.e. iff $\sum_{n\geq 0} \lambda_n^{-1} = \infty$ (with $1/0 = \infty$).

The canonical counting process corresponding to a SPP or SPP_{ex} as described here, is a continuous time, homogeneous Markov chain, moving from state 0 to state 1, from 1 to 2 etc. The process $N^\circ + 1$ is what is commonly called a birth process.

(For a general construction of time-homogeneous Markov chains, see Example 2.2.3 below).

2.2 Creating MPP's

The construction of MPP's $(\mathcal{T}, \mathcal{Y})$ with mark space $(E, \mathcal{E})$ consists in the construction of *canonical MPP's*, i.e. probabilities R on the sequence space $(K(E), \mathcal{K}(E))$, cf. Section 1.1. If $\mathcal{E}$ separates points, (see p. 11), the bimeasurable bijection $\varphi : (K(E), \mathcal{K}(E)) \to (\mathcal{M}, \mathcal{H})$ given by

$$\varphi(t_1, t_2, \ldots; y_1, y_2, \ldots) = \sum_{n: t_n < \infty} \varepsilon_{(t_n, y_n)}$$

and

$$T_n^\circ\left(\varphi^{-1} m\right) = \inf\{t : m([0,t]) = n\} \quad (m \in \mathcal{M}, n \in \mathbb{N}), \tag{2.4}$$

$$\left(Y_n^\circ \circ \varphi^{-1} \in A\right) = \tag{2.5}$$
$$\bigcup_{K_0=1}^{\infty} \bigcap_{K=K_0}^{\infty} \bigcup_{k=1}^{\infty} \left(\overline{N}_{(k-1)/2^K}^\circ = n-1, N_{k/2^K}^\circ(A) - N_{(k-1)/2^K}^\circ(A) = 1\right)$$

for $n \in \mathbb{N}$, $A \in \mathcal{E}$, where $N_t^\circ(\mu^\circ, A) = \mu^\circ([0,t] \times A)$, $\overline{N}_t^\circ = N_t^\circ(E)$, provides a construction of a *canonical random counting measure* by transformation, yielding the probability $\varphi(R)$ on $(\mathcal{M}, \mathcal{H})$.

The idea behind the construction of a MPP is to start with the marginal distribution of the first jump time T_1 and then successively specify the conditional distribution of T_{n+1} given $(T_1, \ldots, T_n; Y_1, \ldots, Y_n)$ and of Y_{n+1} given $(T_1, \ldots, T_n, T_{n+1}; Y_1, \ldots, Y_n)$. Formally, let

$$K^{(n)}(E) = \{(t_1, \ldots, t_n; y_1, \ldots, y_n) : \begin{array}{l} 0 < t_1 \leq \cdots \leq t_n \leq \infty, \ t_k < t_{k+1} \\ \text{and } y_k \in E \text{ if } t_k < \infty\} \end{array}$$

denote the space of finite sequences of n timepoints and n marks that form the beginning of sequences in $K(E)$, equipped with the $\sigma-$algebra $\mathcal{K}^{(n)}(E)$ spanned by the coordinate projections (equivalently the trace on $K^{(n)}(E)$ of the product $\sigma-$algebra $\mathcal{B}^n \otimes \mathcal{E}^n$). Similarly, for $n \in \mathbb{N}_0$ define

$$J^{(n)}(E) = \{(t_1, \ldots, t_n, t; y_1, \ldots, y_n) : \begin{array}{l} (t_1, \ldots, t_n; y_1, \ldots, y_n) \in K^{(n)}(E), \\ t_n \leq t \text{ with } t_n < t \text{ if } t_n < \infty\} \end{array}$$

equipped with the obvious $\sigma-$algebra $\mathcal{J}^{(n)}(E)$.

Assume given a probability $P^{(0)}$ on $\overline{\mathbb{R}}_+$ and, for every $n \in \mathbb{N}$, a Markov kernel $P^{(n)}$ from $\left(K^{(n)}(E), \mathcal{K}^{(n)}(E)\right)$ to $\left(\overline{\mathbb{R}}_+, \overline{\mathcal{B}}_+\right)$, as well as, for every $n \in \mathbb{N}_0$, a Markov kernel $\pi^{(n)}$ from $\left(J^{(n)}(E), \mathcal{J}^{(n)}(E)\right)$ to $\left(\overline{E}, \overline{\mathcal{E}}\right)$.

Notation. Write z_n for a typical point $(t_1, \ldots, t_n; y_1, \ldots, y_n) \in K^{(n)}(E)$ and (z_n, t) or z_n, t for a typical point $(t_1, \ldots, t_n, t; y_1, \ldots, y_n) \in J^{(n)}(E)$. Also we write $\overline{T}^\circ_n((t_k, y_k)) = t_n$ and $\overline{Y}^\circ_n((t_k, y_k)) = y_n$ for coordinate projections on $\overline{K}(E)$.

Theorem 2.2.1 *(a) For every choice of the probability $P^{(0)}$ and the Markov kernels $P^{(n)}$ for $n \geq 1$, $\pi^{(n)}$ for $n \geq 0$ satisfying*

$$\begin{array}{ll} P^{(n)}_{z_n}(]t_n, \infty]) = 1 & \text{if } t_n < \infty \\ P^{(n)}_{z_n}(\{\infty\}) = 1 & \text{if } t_n = \infty, \\ \pi^{(n)}_{z_n,t}(E) = 1 & \text{if } t < \infty, \\ \pi^{(n)}_{z_n,t}(\{\nabla\}) = 1 & \text{if } t = \infty, \end{array} \tag{2.6}$$

there is a unique probability $\overline{R}$ on the sequence space $\left(\overline{K}(E), \overline{\mathcal{K}}(E)\right)$ allowing explosions, such that $\overline{T}^\circ_1(Q) = P^{(0)}$ and for every $n \geq 1$, $z_n \in K^{(n)}(E)$, the probability $P^{(n)}_{z_n}(\cdot)$ is a regular conditional distribution of $\overline{T}^\circ_{n+1}$ given $\left(\overline{T}^\circ_1, \ldots, \overline{T}^\circ_n; \overline{Y}^\circ_1, \ldots, \overline{Y}^\circ_n\right) = z_n$, and for every $n \geq 0$, $(z_n, t) \in J^{(n)}(E)$, the probability $\pi^{(n)}_{z_n,t}(\cdot)$ is a regular conditional distribution of $\overline{Y}^\circ_{n+1}$ given $\left(\overline{T}^\circ_1, \ldots, \overline{T}^\circ_n, \overline{T}^\circ_{n+1}; \overline{Y}^\circ_1, \ldots, \overline{Y}^\circ_n;\right) = (z_n, t)$.

(b) $\overline{R}$ defines a canonical MPP R, i.e. $\overline{R}(K(E)) = 1$ with R the restriction to $K(E)$ of $\overline{R}$, if and only if

$$\overline{R}\left(\lim_{n\to\infty} \overline{T}_n^\circ = \infty\right) = 1. \tag{2.7}$$

Similar remarks apply to this result as those given after Theorem 2.1.1. The proof is based on the Kolmogorov consistency theorem working with finite-dimensional distributions, which have the following appearance: for $n \in \mathbb{N}$, $C^{(n)} \in K^{(n)}(E)$,

$$R\left(Z_n \in C^{(n)}\right) = \tag{2.8}$$
$$\int_{\mathbb{R}_0} P^{(0)}(dt_1) \int_E \pi_{t_1}^{(0)}(dy_1) \cdots \int_{\mathbb{R}_0} P_{z_{n-1}}^{(n-1)}(dt_n) \int_E \pi_{z_{n-1},t_{n-1}}^{(n-1)}(dy_n)\ 1_{C^{(n)}}(z_n).$$

Notation. As in the SPP case we shall write $\overline{P}_{z_n}^{(n)}(t) = P_{z_n}^{(n)}(]t,\infty])$, c.f. p.15.

Example 2.2.1 *Suppose $E = \{1,\ldots,r\}$ is finite with $r \geq 2$, and let $\mathcal{E}$ be the $\sigma-$algebra of all subsets of E. Define*

$$\overline{P}^{(0)}(t) = e^{-\lambda t} \quad (t \geq 0), \qquad \overline{P}_{z_n}^{(n)}(t) = e^{-\lambda(t-t_n)} \quad (t \geq t_n),$$

$$\pi_{z_n,t}^{(n)}(\{y\}) = p_y \quad (y \in E)$$

for some $\lambda > 0$ and some probability function $(p_y)_{y\in E}$ (i.e. $p_y \geq 0$, $\sum_E p_y = 1$). In other words, the waiting times between jumps are iid exponential, the Y_n° are iid (p_y) and the sequences (T_n°) and (Y_n°) are independent. The resulting probability R on $(K(E),\mathcal{K}(E))$ is the canonical Poisson process with mark space E and parameter vector $(\lambda_y)_{y\in E}$ *where $\lambda_y = \lambda p_y$. If $Q = \phi(R)$ is the corresponding canonical RCM one finds, that under Q, the counting process $N^{\circ,y} := N^\circ(\{y\})$ is Poisson λ_y, and that the $N^{\circ,y}$ are mutually stochastically independent. (For a proof, see Example 3.8.3). Note that by Example 2.1.2, $\overline{N}^\circ$ is Poisson λ.*

Example 2.2.2 *Suppose $(X_n)_{n\in\mathbb{N}_0}$ is a stochastic process in* discrete *time with state space $(G,\mathcal{G})$ (each X_n is a $\mathcal{G}-$measurable, $G-$valued random variable). The distribution of (X_n)* conditionally *on $X_0 = x_0$ for an arbitrary $x_0 \in G$, may be viewed as the MPP with mark space $(G,\mathcal{G})$ generated by the Markov kernels $P_{z_n|x_0}^{(n)} = \varepsilon_{n+1}$ and*

$$\pi_{z_n,t|x_0}^{(n)}(A) = P\left(X_{n+1} \in A \,|X_0 = x_0, (X_1,\ldots,X_n) = z_n\right).$$

It is one of the main purposes of these notes to construct *piecewise deterministic processes* (PDP's) from MPP's and to use MPP theory to discuss the properties of the PDP's. We now briefly outline how the connection arises.

Suppose $X = (X_t)_{t\geq 0}$ is a stochastic process defined on some probability space with a state space $(G, \mathcal{G})$ which is a topological space equipped with its Borel $\sigma-$algebra. Assume further that X is right-continuous and piecewise continuous with only finitely many discontinuities (jumps) on any finite time interval. Then a MPP is easily constructed from X by letting T_n be the time of the n'th jump, and defining $Y_n = X_{T_n}$ (if $T_n < \infty$) to be the state reached by X at the time of the n'th jump. In general it is of course not possible to reconstruct X from the MPP, but if further structure is imposed the reconstruction can be done in a natural way, i.e. such that knowledge about the MPP on $[0, t]$ yields $(X_s)_{0\leq s\leq t}$ for any t: suppose that the initial value $x_0 = X_0$ of X is non-random and that for every $n \in \mathbb{N}_0$ there is a suitably measurable function $f^{(n)}_{z_n|x_0}(t)$ of z_n and t and the initial state x_0 such that

$$X_t = f^{(\overline{N}_t)}_{Z_t|x_0}(t) \tag{2.9}$$

where $Z_t := \left(T_1, \ldots, T_{\overline{N}_T}; Y_1, \ldots, Y_{\overline{N}_T}\right)$. Thus, until the time of the first jump, $X_t = f^{(0)}_{|x_0}(t)$ is a deterministic function of x_0 and t, between the first and second jump, X_t is a function of x_0, t and T_1, Y_1 etc. The functions $f^{(n)}$ provide algorithms for computing X between jumps, based on the past history of the process. Note that the fact that $Y_n = X_{T_n}$ on $(T_n < \infty)$ translates into the boundary condition

$$f^{(n)}_{z_n|x_0}(t_n) = y_n.$$

We shall more formally call a process X of the form (2.9) *piecewise deterministic*. It is a *piecewise continuous* process if all $t \to f^{(n)}_{z_n}(t)$ are continuous on $[t_n, \infty[$, a *step process* (or *piecewise constant* process) if $f^{(n)}_{z_n}(t) = y_n$ and a *piecewise linear* process with slope α if $f^{(n)}_{z_n}(t) = y_n + \alpha(t - t_n)$ for some constant α not depending on n.

Example 2.2.3 *We shall outline the MPP description of time-homogeneous Mar-kov chains in continuous time. That what follows is a construction of such chains is well known but will follow in any case from Theorem 6.1.1. Let $X = (X_t)_{t\geq 0}$ be a homogeneous Markov chain with an at most countably infinite state space E with $\mathcal{E}$ the $\sigma-$algebra of all subsets. In accordance with traditional notation we write i, j, i_k for elements of E rather than e.g. y, y_k. Assume that $X_0 \equiv i_0 \in E$ is non-random (alternatively, look at X*

conditionally on X_0) and that X has only finitely many jumps in finite time intervals. The distribution of the chain is completely specified by the initial state i_0 and the transition probabilities (not depending on i_0)

$$p_{ij}(t) = P(X_{s+t} = j \,|X_s = i) = P(X_{s+t} = j \,|(X_u)_{u \leq s},\ X_s = i)$$

for any s, t, i, j and also by i_0 and the transition intensities

$$q_{ij} := \lim_{t \to 0} \frac{1}{t}(p_{ij}(t) - \delta_{ij}),$$

where $\delta_{ij} = 1$ if $i = j$ and $= 0$ otherwise. The transition probabilities form a semigroup, $P(s+t) = P(s)P(t)$ for $s, t \geq 0$ with $P(0) = I$, where $P(t) = (p_{ij}(t))_{i,j \in E}$ is the transition matrix for time intervals of length t and $I = (\delta_{ij})$ is the $E \times E$ identity matrix. The q_{ij} satisfy $\lambda_i := -q_{ii} \geq 0$, $q_{ij} \geq 0$ for $i \neq j$ and $\sum_j q_{ij} = 0$ for all i.

X is a step process as described above, i.e. X is given by (2.9) with, for $z_n = (t_1, \ldots, t_n; i_1, \ldots, i_n)$,

$$f^{(n)}_{z_n|i_0}(t) = i_n,$$

letting T_n be the time of the n'th jump of X and $Y_n = X_{T_n}$ on $(T_n < \infty)$. Then the distribution of the MPP $((T_n), (Y_n))$ is determined by the Markov kernels $P^{(n)}_{\cdot|i_0}$, $\pi^{(n)}_{\cdot|i_0}$ given by

$$\overline{P}^{(0)}_{|i_0}(t) = e^{-\lambda_{i_0} t}, \qquad \overline{P}^{(n)}_{z_n|i_0}(t) = e^{-\lambda_{i_n}(t - t_n)} \quad (t \geq t_n), \tag{2.10}$$

$$\pi^{(n)}_{z_n,t|i_0}(A) = \sum_{j \in A \setminus i_n} \frac{q_{i_n j}}{\lambda_{i_n}}, \tag{2.11}$$

corresponding to the well known fact that for minimal jump chains with stationary transitions, conditionally on the past the waiting time to the next jump is exponential at rate λ_{i_n}, where i_n is the present state, while the jump itself is governed by the intensities $q_{i_n j}$ in the manner described.

Note that if $\lambda_i = 0$, the state i is absorbing *as is seen from (2.10): once the chain enters state i, it remains there forever. In particular the definition of $\pi^{(n)}_{z_n,t|i_0}$ if $\lambda_{i_n} = 0$ (in which case (2.11) does not make sense) is therefore immaterial.*

Chapter 3

Compensators and martingales

3.1 Hazard measures

Let $\mathbb{P}$ be a probability on $\left(\overline{\mathbb{R}}_+, \overline{\mathcal{B}}_+\right)$ with survivor function $\overline{\mathbb{P}}$. Thus $\overline{\mathbb{P}}(t) = \mathbb{P}\left(]t, \infty]\right)$ and $\overline{\mathbb{P}}(t-) = \mathbb{P}\left([t, \infty]\right)$. Also write $\Delta\mathbb{P}(t) := \mathbb{P}\left(\{t\}\right)$ with in particular $\Delta\mathbb{P}(\infty) := \mathbb{P}\left(\{\infty\}\right)$.

Definition 3.1.1 *The* hazard measure *for* $\mathbb{P}$ *is the positive measure* ν *on* $(\mathbb{R}_+, \mathcal{B}_+)$ *with* $\nu \ll \mathbb{P}$ *and*

$$\frac{d\nu}{d\mathbb{P}}(t) = \begin{cases} \frac{1}{\overline{\mathbb{P}}(t-)} & \textit{if } \overline{\mathbb{P}}(t-) > 0, \\ 0 & \textit{otherwise.} \end{cases} \tag{3.1}$$

Formally, the Radon-Nikodym derivative is with respect to the restriction to $\mathbb{R}_+$ of $\mathbb{P}$ rather than $\mathbb{P}$ itself. By the definition of ν,

$$\nu(B) = \int_B \frac{1}{\overline{\mathbb{P}}(s-)}\, \mathbb{P}\left(ds\right)$$

for *all* $B \in \mathcal{B}_+$ as follows from the observation that

$$\mathbb{P}\left(\left\{t > 0 : \overline{\mathbb{P}}(t-) = 0\right\}\right) = 0, \tag{3.2}$$

i.e. the definition of $d\nu/d\mathbb{P}(t)$ when $\overline{\mathbb{P}}(t-) = 0$ is immaterial. (With $t^\dagger$ as defined in (3.3) below,the set appearing in (3.2) is $= \emptyset$ if $t^\dagger = \infty$, $= \left[t^\dagger, \infty\right[$ if $t^\dagger < \infty$ and $\Delta\mathbb{P}\left(t^\dagger\right) = 0$, $= \left]t^\dagger, \infty\right[$ if $t^\dagger < \infty$ and $\Delta\mathbb{P}\left(t^\dagger\right) > 0$).

The reader is reminded about the standard more informal definition of hazard measure: if U is a $\overline{\mathbb{R}}_+-$valued random variable, then the hazard measure for the distribution of U is given by

$$\nu\left([t, t + dt[\right) = P\left(t \leq U < t + dt\, | U \geq t\right),$$

i.e. if U is the timepoint at which some event occurs, ν measures the risk of that event happening now or in the immediate future given that it has not occurred yet.

If $\mathbb{P}$ has density f with respect to Lebesgue measure, ν has density u where $u(t) = f(t)/\overline{\mathbb{P}}(t)$ if $\overline{\mathbb{P}}(t) > 0$. The function u is then called the *hazard function* for $\mathbb{P}$.

Define

$$t^\dagger := \inf\left\{t > 0 : \overline{\mathbb{P}}(t) = 0\right\}, \tag{3.3}$$

the *termination point* for $\mathbb{P}$. Note that because $\overline{\mathbb{P}}(0) = 1$, $t^\dagger > 0$. Further, $t^\dagger = \infty$ iff $\overline{\mathbb{P}}(t) > 0$ for all $t \in \mathbb{R}_+$ and if $t^\dagger < \infty$ always $\overline{\mathbb{P}}(t^\dagger) = 0$ and in addition either $\Delta\mathbb{P}\left(t^\dagger\right) > 0$, or $\Delta\mathbb{P}\left(t^\dagger\right) = 0$ and $\overline{\mathbb{P}}(t^\dagger - \varepsilon) > 0$ for all $0 < \varepsilon < t^\dagger$. Also $\nu\left(\left]t^\dagger, \infty\right[\right) = 0$ because of our definition (3.1) – but that need of course not hold with other definitions of ν beyond $t^\dagger$.

The basic properties of hazard measures are summarized in the next result where $\Delta\nu(t) := \nu\left(\{t\}\right)$.

Theorem 3.1.1 *Let ν be the hazard measure for some probability $\mathbb{P}$ on $\left(\overline{\mathbb{R}}_+, \overline{\mathcal{B}}_+\right)$. Then*

(i) ν is locally finite at 0: $\nu\left(]0, t]\right) < \infty$ for $t > 0$ sufficiently small;

(ii) $\nu\left(]0, t]\right) < \infty$ if $t < t^\dagger$;

(iii) $\Delta\nu(t) \le 1$ for all $t \in \mathbb{R}_0$, $\Delta\nu(t) < 1$ for all $t < t^\dagger$;

(iv) $\Delta\nu\left(t^\dagger\right) = 1$ iff $t^\dagger < \infty$ and $\Delta\mathbb{P}\left(t^\dagger\right) > 0$;

(v) if $\Delta\mathbb{P}\left(t^\dagger\right) > 0$, then $\nu\left(\left]0, t^\dagger\right]\right) < \infty$ if $t^\dagger < \infty$ and $\nu\left(\mathbb{R}_+\right) < \infty$ if $t^\dagger = \infty$;

(vi) if $\Delta\mathbb{P}\left(t^\dagger\right) = 0$, then $\nu\left(\left]0, t^\dagger\right[\right) = \infty$ whether $t^\dagger$ is finite or not.

If conversely ν is a positive, possibly infinite measure on $(\mathbb{R}_+, \mathcal{B}_+)$, locally finite at 0 as in (i) above with $\Delta\nu\,(t) \le 1$ for all $t \in \mathbb{R}_+$, then ν is the hazard measure for a uniquely determined probability $\mathbb{P}$ on $\left(\overline{\mathbb{R}}_+, \overline{\mathcal{B}}_+\right)$ in the following sense: the termination point $t^\dagger$ for $\mathbb{P}$ is

$$t^\dagger = \inf\left\{t > 0 : \Delta\nu\,(t) = 1 \text{ or } \nu\left(]0, t]\right) = \infty\right\} \tag{3.4}$$

and the survivor function $\overline{\mathbb{P}}$ for $\mathbb{P}$ is given by the product integral

$$\overline{\mathbb{P}}(t) = \begin{cases} \prod_{0<s\le t}\left(1 - \nu\,(ds)\right) & \text{if } t < t^\dagger, \\ 0 & \text{if } t \ge t^\dagger. \end{cases}$$

Note. The product integral is described in detail below, see (3.6).

Proof. If $\overline{\mathbb{P}}(t-) > 0$ then $\nu(]0,t]) \le \int_{]0,t]} \frac{1}{\overline{\mathbb{P}}(t-)} \mathbb{P}(ds) < \infty$, proving (i), (ii) and (v). (iii) and (iv) follow from $\Delta\nu(t) = \Delta\mathbb{P}(t)/\overline{\mathbb{P}}(t-)$ if $\overline{\mathbb{P}}(t-) > 0$. To prove (vi), note that $\mathbb{P}\left([t,t^\dagger[\right) \downarrow 0$ as $t \uparrow t^\dagger$ and then define $t_0 = 0$ and recursively choose t_k such that $\mathbb{P}\left([t_k,t^\dagger[\right) \le \frac{1}{2}\mathbb{P}\left([t_{k-1},t^\dagger[\right)$. Then $t_k \uparrow t^\dagger$ and hence

$$\begin{aligned}\nu\left(]0,t^\dagger[\right) &= \sum_{k=1}^\infty \nu\left([t_{k-1},t_k[\right)\\ &= \sum_{k=1}^\infty \int_{[t_{k-1},t_k[} \frac{1}{\overline{\mathbb{P}}(s-)} \mathbb{P}(ds)\\ &\ge \sum_{k=1}^\infty \frac{\mathbb{P}\left([t_{k-1},t_k[\right)}{\mathbb{P}\left([t_{k-1},t^\dagger[\right)}\end{aligned}$$

where the assumption $\Delta\mathbb{P}(t^\dagger) = 0$ has been used in the last step. But the series in the last line diverges since each term is $\ge \frac{1}{2}$.

For the proof of the last part of the proposition we focus on two cases, noting that the definition of $t^\dagger$ is certainly the only possible in view of (iv) and (vi): suppose first that $\mathbb{P}$ has a continuous density f on $]0,t^\dagger[$. Then $\nu(dt) = u(t)\,dt$ where $u(t) = f(t)/\overline{\mathbb{P}}(t) = -\left(\log\overline{\mathbb{P}}(t)\right)'$ forcing $\overline{\mathbb{P}}(t) = \exp\left(-\int_0^t u(s)\,ds\right) = \exp\left(-\nu(]0,t])\right)$ for $t < t^\dagger$. Suppose next that $\mathbb{P}$ is discrete in the special sense that there is a finite or infinite strictly increasing sequence $0 < t_1 < t_2 < \cdots < \infty$ such that all $\Delta\mathbb{P}(t_k) > 0$ and $\sum \Delta\mathbb{P}(t_k) + \Delta\mathbb{P}(\infty) = 1$. Then $\Delta\nu(t_k) = \Delta\mathbb{P}(t_k)/\overline{\mathbb{P}}(t_k-) = 1 - \overline{\mathbb{P}}(t_k)/\overline{\mathbb{P}}(t_{k-1})$ and consequently $\overline{\mathbb{P}}(t_k) = \prod_{j=1}^k (1 - \Delta\nu(t_j))$ resulting in $\overline{\mathbb{P}}(t) = \prod_{0<s\le t}(1-\Delta\nu(s))$. ■

The following immediate consequence of Theorem 3.1.1 will be used later:

Corollary 3.1.2 *If U is a $\overline{\mathbb{R}}_+$−valued random variable with hazard measure ν, then*

$$P\left(\nu(]0,U] \cap \mathbb{R}_+) < \infty\right) = 1. \tag{3.5}$$

The last part of the proof of Proposition 3.1.1 gives two instances of how the product integral

$$\overline{\mathbb{P}}(t) = \prod_{0<s\le t} (1 - \nu(ds)) \tag{3.6}$$

should be interpreted. It may be shown more generally that if $\mathbb{P}$ is *continuous* ($\Delta\mathbb{P}(t) = 0$ for all $t \in \mathbb{R}_+$), then

$$\overline{\mathbb{P}}(t) = \exp\left(-\nu(]0,t])\right),$$

and if $\mathbb{P}$ is *discrete* (there is an at most countably infinite subset B_d of $\mathbb{R}_+$ such that $\mathbb{P}(\mathbb{R}_+ \backslash B_d) = 0$), then

$$\overline{\mathbb{P}}(t) = \prod_{0<s\leq t} (1 - \Delta\nu(s))$$

with only the factors $1-\Delta\nu(s)$ for $s \in B_d$ contributing to the continuous time product. For *general* $\mathbb{P}$, suppose ν is a given hazard measure and compute $t^\dagger$ by (3.4). To recover $\overline{\mathbb{P}}$ it suffices to find $\overline{\mathbb{P}}(t)$ for $t < t^\dagger$. Consider ν restricted to $\left]0, t^\dagger\right[$ and split ν into its continuous and discrete parts, $\nu = \nu^c + \nu^d$, where for Borel sets $B \subset \left]0, t^\dagger\right[$, $\nu^d(B) = \sum_{t \in B \cap B_d} \Delta\nu(t)$ writing $B_d = \left\{t \in \left]0, t^\dagger\right[: \Delta\nu(t) > 0\right\}$. Then the claim is that

$$\overline{\mathbb{P}}(t) = \exp\left(-\nu^c\left(]0, t]\right)\right) \prod_{0<s\leq t} \left(1 - \Delta\nu^d(s)\right). \tag{3.7}$$

The idea behind the proof is essentially to identify ν^c and ν^d with hazard measures, one of which must have termination point $t^\dagger$. One then exploits that if U^c, U^d are independent random variables with hazard measures ν^c, ν^d (so that since U^c has a continuous distribution and U^d a discrete distribution,

$$P\left(U^c > t\right) = \exp\left(-\nu^c\left(]0, t]\right)\right), \quad P\left(U^d > t\right) = \prod_{0<s\leq t} \left(1 - \Delta\nu^d(s)\right)),$$

then $U := \min\left(U^c, U^d\right)$, which trivially has survivor function $\overline{\mathbb{P}}$ as in (3.7), has hazard measure $\nu = \nu^c + \nu^d$ by Proposition 3.1.3 (ii) below.

Example 3.1.1 *(i) $\nu \equiv 0$ is the hazard measure for the probability ε_∞ with point mass 1 at ∞.*

(ii) $\nu = \lambda\ell$ where $\lambda > 0$ and ℓ denotes Lebesgue measure, is the hazard measure for the exponential distribution with rate λ, i.e. $\overline{\mathbb{P}}(t) = e^{-\lambda t}$.

(iii) $\nu = \sum_{n=1}^\infty p\,\varepsilon_n$ where $0 < p < 1$, is the hazard measure for the geometric distribution on $\mathbb{N}$ given by $\Delta\mathbb{P}(n) = p(1-p)^{n-1}$ for $n \in \mathbb{N}$.

Hazard measures have some further nice properties, that we quote without proofs.

Proposition 3.1.3 *(i) If $\mathbb{P}$ has hazard measure ν and $t_0 > 0$ is such that $\overline{\mathbb{P}}(t_0) > 0$, then the hazard measure for the conditional probability $\mathbb{P}(\cdot\, |]t_0, \infty])$ is the restriction to $]t_0, \infty[$ of ν.*

(ii) If U_1, U_2 are $\overline{\mathbb{R}}_+-$valued and independent random variables with distributions $\mathbb{P}_1, \mathbb{P}_2$ and hazard measures ν_1, ν_2 respectively, then provided $\Delta\mathbb{P}_1(t)\Delta\mathbb{P}_2(t) = 0$ for all $t > 0$, the distribution of $U := \min(U_1, U_2)$ has hazard measure $\nu = \nu_1 + \nu_2$.

The assumption in (ii) that $\mathbb{P}_1, \mathbb{P}_2$ do not have atoms in common on $\mathbb{R}_+$ is essential for the assertion to be valid.

For the proof of Proposition 3.1.3 and other proofs omitted here, one may use the techniques discussed in Appendix A.

3.2 Adapted and predictable processes

We shall in most of this chapter be dealing with canonical counting processes and canonical random counting measures, i.e. probabilities on the spaces $(W, \mathcal{H})$ and $(\mathcal{M}, \mathcal{H})$ respectively, cf. Sections 2.1 and 2.2. Recall that any such probability may be viewed as a SPP, respectively a MPP through the following definition of the sequence of jump times (τ_n) on W,

$$\tau_n = \inf\{t > 0 : N_t^\circ = n\} \quad (n \in \mathbb{N}),$$

see (2.1), and the following definition of the sequence of jump times (τ_n) and marks (η_n) on $\mathcal{M}$,

$$\tau_n = \inf\left\{t > 0 : \overline{N}_t^\circ = n\right\} \quad (n \in \mathbb{N}),$$

$$(\eta_n \in A) = \bigcup_{K_0=1}^{\infty} \bigcap_{K=K_0}^{\infty} \bigcup_{k=1}^{\infty} \left(\overline{N}_{(k-1)/2^K}^\circ = n-1, N_{k/2^K}^\circ(A) - N_{(k-1)/2^K}^\circ(A) = 1\right) \tag{3.8}$$

for $n \in \mathbb{N}$, $A \in \mathcal{E}$, see (2.4), (2.5). In order to be able to identify the exact value of η_n from (3.8), we assume from now on that $\mathcal{E}$ separates point, cf. p. 11. Otherwise the notation used throughout is

$$N^\circ(w) = w, \quad N_t^\circ(w) = w(t) \qquad (w \in W, t \geq 0),$$

and

$$\mu^\circ(m) = m, \quad \mu^\circ(C, m) = m(C)$$

for $m \in \mathcal{M}$, $C \in \mathcal{B}_0 \otimes \mathcal{E}$, while $\overline{N}_t^\circ : \mathcal{M} \to \mathbb{N}_0$ and $\overline{N}_t^\circ(A) : \mathcal{M} \to \mathbb{N}_0$ (for $A \in \mathcal{E}$) are defined by

$$\overline{N}_t^\circ(m) = \sum_{n=1}^{\infty} 1_{(\tau_n \leq t)}(m)$$

$$N_t^\circ(A,m) = \sum_{n=1}^{\infty} 1_{(\tau_n \le t, \eta_n \in A)}(m) = \mu^\circ([0,t]\times A, m)$$

for $m \in \mathcal{M}$, and where we write e.g. $N_t^\circ(A,m)$ rather than $N_t^\circ(A)(m)$. Finally we define

$$\xi_n = \begin{cases} (\tau_1,\ldots,\tau_n) & \text{on } W, \\ (\tau_1,\ldots,\tau_n;\eta_1,\ldots,\eta_n) & \text{on } \mathcal{M} \end{cases} \tag{3.9}$$

as well as

$$\xi_t = \begin{cases} \xi_{N_t^\circ}, \\ \xi_{\overline{N}_t^\circ}, \end{cases} \qquad \xi_{t-} = \begin{cases} \xi_{N_{t-}^\circ}, \\ \xi_{\overline{N}_{t-}^\circ}, \end{cases} \tag{3.10}$$

with the convention that on $(N_t^\circ = 0)$, resp. $\left(\overline{N}_t^\circ = 0\right)$, $\xi_t \equiv 0$ (the important thing is that ξ_0 should be something non-informative and if viewed as a random variable, should generate the trivial $\sigma-$algebra, $\sigma(\xi_0) = \{\emptyset, W\}$ or $\{\emptyset,\mathcal{M}\}$). Note that ξ_t summarizes the jump times and marks occurring in $[0,t]$, ξ_{t-} those occurring in $[0,t[$.

The fundamental *filtration on* $(W,\mathcal{H})$ is $(\mathcal{H}_t)_{t\ge 0}$ where

$$\mathcal{H}_t = \sigma\left((N_s^\circ)_{0\le s\le t}\right)$$

is the smallest $\sigma-$algebra such that all N_s° for $s \le t$ are measurable. Similarly the fundamental *filtration on* $(\mathcal{M},\mathcal{H})$ is $(\mathcal{H}_t)_{t\ge 0}$ where

$$\mathcal{H}_t = \sigma\left((N_s^\circ(A))_{0\le s\le t, A\in\mathcal{E}}\right).$$

For $t = 0$ we get the trivial $\sigma-$algebra,

$$\mathcal{H}_0 = \begin{cases} \{\emptyset, W\} \\ \{\emptyset, \mathcal{M}\}. \end{cases}$$

Note that if $\mathcal{E}' \subset \mathcal{E}$ generates $\mathcal{E}$ (the $\sigma-$algebra $\sigma(\mathcal{E}')$ generated by the sets in $\mathcal{E}'$ is $\mathcal{E}$ itself) with $E \in \mathcal{E}'$ and $\mathcal{E}'$ is closed under the formation of finite intersections, $A_1 \cap A_2 \in \mathcal{E}'$ if $A_1, A_2 \in \mathcal{E}'$, then $\mathcal{H}_t = \sigma\left((N_s^\circ(A))_{0\le s\le t, A\in\mathcal{E}'}\right)$.

A *process* $X = (X_t)_{t\ge 0}$ with state space $(G,\mathcal{G})$ defined on W or $\mathcal{M}$ is a family of $G-$valued random varaibles. The process is *measurable* if the mapping $(t,w) \overset{\Phi}{\to} X_t(w)$, respectively $(t,m) \overset{\Phi}{\to} X_t(m)$, is measurable from $(\mathbb{R}_0 \times W, \mathcal{B}_0 \otimes \mathcal{H})$, respectively $(\mathbb{R}_0 \times \mathcal{M}, \mathcal{B}_0 \otimes \mathcal{H})$, to $(G,\mathcal{G})$. X is *adapted* if it is measurable and each X_t is $\mathcal{H}_t-$measurable; X is *predictable* (or *previsible*) if X_0 is constant (i.e. $\mathcal{H}_0-$measurable) and if Φ restricted to

$(\mathbb{R}_+ \times W, \mathcal{B}_+ \otimes \mathcal{H})$, or $(\mathbb{R}_+ \times \mathcal{M}, \mathcal{B}_+ \otimes \mathcal{H})$, is measurable with respect to the *predictable* $\sigma-$*algebra* $\mathcal{P}$. Here $\mathcal{P}$ is generated by the subsets

$$]t, \infty[\times H \quad (t \geq 0, H \in \mathcal{H}_t)$$

of $\mathbb{R}_+ \times W$ or $\mathbb{R}_+ \times \mathcal{M}$. Recall that any $\mathbb{R}^d-$valued process $X = (X_t)_{t \geq 0}$ which is *left-continuous* and adapted is predictable. (See Appendix B for a more general discussion of the concepts from process theory introduced here).

The following useful result characterizes $\mathcal{H}_t$ and adapted and predictable processes and also shows that the filtration $(\mathcal{H}_t)$ is right-continuous. Recall the definitions (3.9), (3.10) of ξ_n, ξ_t and ξ_{t-}.

Proposition 3.2.1 *(a) Consider the space* $(W, \mathcal{H})$.

(i) A set $H \subset W$ *belongs to* $\mathcal{H}_t$ *iff for every* $n \in \mathbb{N}_0$ *there exists* $B_n \in \mathcal{B}_+^n \cap]0, t]^n$ *such that*

$$H \cap (N_t^\circ = n) = (\xi_n \in B_n, \tau_{n+1} > t).$$

(ii) For every $t \geq 0$, $\mathcal{H}_{t+} = \mathcal{H}_t$ *where* $\mathcal{H}_{t+} := \bigcap_{\varepsilon > 0} \mathcal{H}_{t+\varepsilon}$.

(iii) A real-valued process $X = (X_t)_{t \geq 0}$ *is adapted iff for every* $n \in \mathbb{N}_0$ *there exists a measurable function* $(z_n, t) \to f_{z_n}^{(n)}(t)$ *from* $\mathbb{R}_+^n \times \mathbb{R}_0$ *to* $\mathbb{R}$ *such that identically on* W *and for all* $t \in \mathbb{R}_0$,

$$X_t = f_{\xi_t}^{(N_t^\circ)}(t).$$

(iv) A real-valued process $X = (X_t)_{t \geq 0}$ *is predictable iff for every* $n \in \mathbb{N}_0$ *there exists a measurable function* $(z_n, t) \to f_{z_n}^{(n)}(t)$ *from* $\mathbb{R}_+^n \times \mathbb{R}_0$ *to* $\mathbb{R}$ *such that identically on* W *and for all* $t \in \mathbb{R}_0$

$$X_t = f_{\xi_{t-}}^{\left(N_{t-}^\circ\right)}(t). \tag{3.11}$$

(b) Consider the space $(\mathcal{M}, \mathcal{H})$.

(i) A set $H \subset \mathcal{M}$ *belongs to* $\mathcal{H}_t$ *iff for every* $n \in \mathbb{N}_0$ *there exists* $C_n \in (\mathcal{B}_+^n \otimes \mathcal{E}^n) \cap (]0, t]^n \times E^n)$ *such that*

$$H \cap \left(\overline{N}_t^\circ = n\right) = (\xi_n \in C_n, \tau_{n+1} > t). \tag{3.12}$$

(ii) For every $t \geq 0$, $\mathcal{H}_{t+} = \mathcal{H}_t$ *where* $\mathcal{H}_{t+} := \bigcap_{\varepsilon > 0} \mathcal{H}_{t+\varepsilon}$.

(iii) *A real-valued process* $X = (X_t)_{t\geq 0}$ *is adapted iff for every* $n \in \mathbb{N}_0$ *there exists a measurable function* $(z_n, t) \to f_{z_n}^{(n)}(t)$ *from* $(\mathbb{R}_+^n \times E^n) \times \mathbb{R}_0$ *to* $\mathbb{R}$ *such that identically on* $\mathcal{M}$ *and for all* $t \in \mathbb{R}_0$

$$X_t = f_{\xi_t}^{(\overline{N}_t^\circ)}(t). \tag{3.13}$$

(iv) *A real-valued process* $X = (X_t)_{t\geq 0}$ *is predictable iff for every* $n \in \mathbb{N}_0$ *there exists a measurable function* $(z_n, t) \to f_{z_n}^{(n)}(t)$ *from* $(\mathbb{R}_+^n \times E^n) \times \mathbb{R}_0$ *to* $\mathbb{R}$ *such that identically on* $\mathcal{M}$ *and for all* $t \in \mathbb{R}_0$

$$X_t = f_{\xi_{t-}}^{(\overline{N}_{t-}^\circ)}(t). \tag{3.14}$$

Remark 3.2.1 *The description of adapted processes merely states (apart from measurability properties) that a process is adapted iff its value at* t *can be computed from the number of jumps on* $[0,t]$ *and the timepoints and marks for these jumps. In particular, an adapted* $\mathbb{R}-$*valued process on e.g.* $(\mathcal{M}, \mathcal{H})$ *is piecewise deterministic, cf. (2.9). For a process to be predictable, to find its value at* t *it suffices to know the number of jumps just before* t*, their location and the marks.*

Example 3.2.1 *On* $(W, \mathcal{H})$ *, the counting process* N° *is adapted but* not *predictable. Similarly, on* $(\mathcal{M}, \mathcal{H})$ *the counting processes* $N^\circ(A)$ *are adapted for all* $A \in \mathcal{E}$*, but not predictable except for* $A = \emptyset$*. To see the latter, fix* $A \neq \emptyset$*,* $t > 0$ *and just note that there is* $m \in \mathcal{M}$ *with* $N_t^\circ(m, A) = 0$ *and a different* $\widetilde{m}$ *with* $N_{t-}^\circ(\widetilde{m}, A) = 0$*,* $N_t^\circ(\widetilde{m}, A) = 1$ *(the first jump for* $\widetilde{m}$ *occurs at time* t*, resulting in a mark in* A*); but were* $N^\circ(A)$ *predictable, by Proposition 3.2.1,* $N_t^\circ(m, A) = N_t^\circ(\widetilde{m}, A)$*.*

As an example of the representation (3.13), note that on $(\mathcal{M}, \mathcal{H})$ *,* $N_t^\circ(A)$ *has the representation (3.13) with*

$$f_{z_n}^{(n)}(t) = \sum_{k=1}^{n} 1_A(y_k)$$

where, as usual, $z_n = (t_1, \ldots, t_n; y_1, \ldots, y_n)$ *.*

Proof. *(Proposition 3.2.1).* We just prove (b).

(i). To show that all $H \in \mathcal{H}_t$ have the representation (3.12), it suffices to show that (3.12) holds for the members of the $\mathcal{H}_t-$generating class $((N_s^\circ(A) = l))_{l\in\mathbb{N}_0, s\leq t, A\in\mathcal{E}}$. But

$$\left(N_s^\circ(A) = l,\ \overline{N}_t^\circ = n\right) = \left(\sum_{k=1}^{n} 1_{(\tau_k \leq s, \eta_k \in A)} = l,\ \tau_n \leq t < \tau_{n+1}\right)$$

so (3.12) holds with

$$C_n = \left\{ z_n : \sum_{k=1}^{n} 1_{]0,s]\times A}(t_k, y_k) = l,\ t_n \le t \right\}.$$

Suppose conversely that $H \subset \mathcal{M}$ satisfies (3.12) for all n. Since $H = \bigcup_{n=0}^{\infty} H \cap \left(\overline{N}_t^\circ = n\right)$, $H \in \mathcal{H}_t$ follows if we show $H \cap \left(\overline{N}_t^\circ = n\right) \in \mathcal{H}_t$. But for that, by (3.12), it suffices to show that

$$(\xi_n \in C_n, \tau_{n+1} > t) \in \mathcal{H}_t$$

for C_n of the form

$$C_n = \{z_n : t_k \le s, y_k \in A\} \tag{3.15}$$

for $1 \le k \le n$, $s \le t$, $A \in \mathcal{E}$, since these sets generate $\left(\mathcal{B}_+^n \otimes \mathcal{E}^n\right) \cap \left(]0,t]^n \times E^n\right)$. But with C_n given by (3.15), use a variation of (3.8) to verify that

$$\begin{aligned} &(\xi_n \in C_n, \tau_{n+1} > t) = (\tau_k \le s, \tau_{n+1} > t) \cap \\ &\bigcup_{K_0=1}^{\infty} \bigcap_{K=K_0}^{\infty} \bigcup_{j=1}^{2^K} \left(\overline{N}_{(j-1)s/2^K}^\circ = k-1,\ N_{js/2^K}^\circ(A) - N_{(j-1)s/2^K}^\circ(A) = 1\right) \end{aligned}$$

which clearly is a set in $\mathcal{H}_t$.

(ii). We must show that $\mathcal{H}_{t+} \subset \mathcal{H}_t$. So suppose that $H \in \mathcal{H}_{t+}$, i.e. that $H \in \mathcal{H}_{t+\frac{1}{k}}$ for all $k \in \mathbb{N}$. By (i), for every n, k there is $C_{n,k} \subset \left]0, t+\frac{1}{k}\right]^n \times E^n$ measurable such that

$$H \cap \left(\overline{N}_{t+\frac{1}{k}}^\circ = n\right) = \left(\xi_n \in C_{n,k}, \tau_{n+1} > t + \tfrac{1}{k}\right). \tag{3.16}$$

Now consider the set of $m \in \mathcal{M}$ belonging to the set (3.16) for k sufficiently large. Because $t \to \overline{N}_t^\circ$ is right-continuous and piecewise constant and because $m \in \left(\tau_{n+1} > t + \frac{1}{k}\right)$ for k sufficiently large iff $m \in (\tau_{n+1} > t)$, this leads to the identity

$$H \cap \left(\overline{N}_t^\circ = n\right) = \left(\xi_n \in \bigcup_{K_0=1}^{\infty} \bigcap_{k=K_0}^{\infty} C_{n,k}\right) \cap (\tau_{n+1} > t).$$

But since $C_{n,k}$ is a measurable subset of $\left]0, t+\frac{1}{k}\right]^n \times E^n$, $\bigcup_{K_0=1}^{\infty} \bigcap_{k=K_0}^{\infty} C_{n,k}$ is a measurable subset of $]0,t]^n \times E^n$, hence (3.12) holds and by (i), $H \in \mathcal{H}_t$.

(iii). Let X be a $\mathbb{R}-$valued, adapted process. Writing $X = X^+ - X^-$, the difference between the positive and negative part of X, it is seen that to establish (3.13) it is enough to consider $X \geq 0$. Further, writing

$$X_t\left(m\right) = \lim_{K\to\infty} \sum_{k=0}^{\infty} \frac{k}{2^K} 1_{\left(\frac{k}{2^K} \leq X < \frac{k+1}{2^K}\right)}\left(t, m\right)$$

it is clear that it suffices to consider adapted X of the form

$$X_t\left(m\right) = 1_D\left(t, m\right) \tag{3.17}$$

for some $D \in \mathcal{B}_0 \otimes \mathcal{H}$, i.e. D has the property that

$$D_t := \{m : (t, m) \in D\} \in \mathcal{H}_t$$

for all t. But by (i), for every $n \in \mathbb{N}_0$, $t \geq 0$,

$$D_t \cap \left(\overline{N}_t^\circ = n\right) = \left(\xi_n \in C_{n,t}, \tau_{n+1} > t\right)$$

for some $C_{n,t} \in \mathcal{B}_+^n \otimes \mathcal{E}^n$ and it follows that (3.13) holds with

$$f_{z_n}^{(n)}(t) = 1_{C_{n,t}}\left(z_n\right).$$

It remains to show that $f^{(n)}$ is a measurable function of $\left(z_n, t\right)$. But since X is measurable, so is

$$(t, m) \to X_t(m) 1_{\left(\overline{N}_t^\circ = n\right)}\left(m\right) = f_{\xi_n(m)}^{(n)}\left(t\right)$$

and the assertion follows easily.

For the converse, suppose that X is given by (3.13) with all $(z_n, t) \to f_{z_n}^{(n)}(t)$ measurable. It is immediately checked that X is then measurable, and it remains to see that X_t is $\mathcal{H}_t-$measurable. For this, the standard approximation techniques imply that it suffices to consider $f^{(n)}$ of the form

$$f_{z_n}^{(n)}(t) = 1_{C_n}(z_n) 1_{[s_n,\infty[}(t)$$

for some $C_n \in \mathcal{B}_+^n \otimes \mathcal{E}^n$, $s_n \geq 0$. But then, for every n,

$$X_t 1_{\left(\overline{N}_t^\circ = n\right)} = 1_{(\xi_n \in C_n)} 1_{(\tau_n \leq t < \tau_{n+1})} 1_{[s_n,\infty[}(t)$$

which is $\mathcal{H}_t-$measurable by (i).

(iv). Arguing as in the proof of (iii), to prove that any predictable X has the form (3.14), it suffices to consider

$$X_t(m) = 1_D\left(t, m\right)$$

for $D \in \mathcal{P}$. Since the desired representation holds trivially for $D = \mathbb{R}_+ \times \mathcal{M}$, it here suffices to consider D of the form

$$D =]s, \infty[\times H \quad (s > 0, H \in \mathcal{H}_s), \tag{3.18}$$

since these sets form a class closed under the formation of finite intersections that generate $\mathcal{P}$.

Since $H \in \mathcal{H}_s$, (3.12) holds and consequently, if D is given by (3.18), for $t \geq 0$,

$$\begin{aligned} X_t &= 1_{]s,\infty[}(t) 1_H \\ &= 1_{]s,\infty[}(t) \sum_{n=0}^{\infty} 1_{\left(\overline{N}^{\circ}_{t-}=n\right)} \sum_{k=0}^{n} 1_{(\xi_k \in C_k, \tau_{k+1} > s)} \end{aligned}$$

using that when $t > s$, $\overline{N}^{\circ}_{t-} \geq \overline{N}^{\circ}_{s}$. Thus (3.13) holds with

$$f^{(n)}_{z_n}(s) = 1_{]s,\infty[}(t) \left(\sum_{k=0}^{n-1} 1_{C_k}(z_k)\, 1_{]s,\infty[}(t_{k+1}) + 1_{C_n}(z_n) \right).$$

Finally, let X be given by (3.14) and let us show that X is predictable. Clearly X_0 is constant on $\mathcal{M}$, hence $\mathcal{H}_0-$measurable. To show that X is $\mathcal{P}-$measurable on $\mathbb{R}_+ \times \mathcal{M}$ it suffices to consider $f^{(n)}$ of the form

$$f^{(n)}_{z_n}(t) = 1_{C_n}(z_n)\, 1_{]s_n,\infty[}(t)$$

for some $C_n \in \mathcal{B}^n_+ \otimes \mathcal{E}^n$, $s_n \geq 0$. We need

$$D_n := \left\{ (t, m) : t > 0,\ X_t(m) = 1,\ \overline{N}^{\circ}_{t-}(m) = n \right\} \in \mathcal{P}$$

and find, using $s_{n,j,K} := s_n + j2^{-K}$ to approximate the value of $t > s$, that

$$D_n = \bigcup_{K_0=1}^{\infty} \bigcap_{K=K_0}^{\infty} \bigcup_{j=0}^{\infty}]s_{n,j,K}, s_{n,j+1,K}] \times \left(\xi_n \in B_n, \overline{N}^{\circ}_{s_{n,j,K}} = n \right)$$

which is in $\mathcal{P}$ because $\left(\xi_n \in B_n, \overline{N}^{\circ}_{s_{n,j,K}} = n \right) \in \mathcal{H}_{s_{n,j,K}}$ and the interval $]s_{n,j,K}, s_{n,j+1,K}]$ is open to the left. ■

Proposition 3.2.1 has the following intuitively obvious and very useful consequence: conditioning on $\mathcal{H}_t$ is the same as conditioning on the number of jumps on $[0, t]$, their location in time and their associated marks. More formally we have

Corollary 3.2.2 *If Q is a probability on $(\mathcal{M}, \mathcal{H})$ (or $(W, \mathcal{H})$) and U is a $\mathbb{R}-$valued random variable with $E|U| < \infty$, then*

$$E(U|\mathcal{H}_t) = \sum_{n=0}^{\infty} 1_{(\overline{N}_t^\circ = n)} E(U|\xi_n, \tau_{n+1} > t). \tag{3.19}$$

Note. $E(U|\xi_n, \tau_{n+1} > t)$ is the conditional expectation of U given the random variables ξ_n and $1_{(\tau_{n+1}>t)}$ considered on the set $(\tau_{n+1} > t)$ only.

Proof. By its definition $E(U|\xi_n, \tau_{n+1} > t)$ is a measurable function of ξ_n and $1_{(\tau_{n+1}>t)}$, evaluated on the set $(\tau_{n+1} > t)$ only where the indicator is 1, cf. (3.20) below. Thus we may write

$$1_{(\overline{N}_t^\circ = n)} E(U|\xi_n, \tau_{n+1} > t) = 1_{(\overline{N}_t^\circ = n)} \varphi_n(\xi_n)$$

and (3.13) shows that the n'th term in (3.19) is $\mathcal{H}_t-$measurable for all n. Next, let $H \in \mathcal{H}_t$ and use (3.12) and the definition of $E(U|\xi_n, \tau_{n+1} > t)$ to obtain

$$\begin{aligned}
&\int_H \sum_{n=0}^{\infty} 1_{(\overline{N}_t^\circ = n)} E(U|\xi_n, \tau_{n+1} > t)\, dP \\
=\ & \sum_{n=0}^{\infty} \int_{(\xi_n \in C_n, \tau_{n+1} > t)} E(U|\xi_n, \tau_{n+1} > t)\, dP \\
=\ & \sum_{n=0}^{\infty} \int_{(\xi_n \in C_n, \tau_{n+1} > t)} U\, dP \\
=\ & \int_H U\, dP.
\end{aligned}$$

■

Remark 3.2.2 *The usefulness of the result comes from the construction of SPP's and MPP's which makes it natural to work with the conditional expectations on the right of (3.19). Note that on $(\tau_{n+1} > t)$,*

$$E(U|\xi_n, \tau_{n+1} > t) = \frac{E\left(U 1_{(\tau_{n+1}>t)} |\xi_n\right)}{Q(\tau_{n+1} > t|\xi_n)} = \frac{1}{\overline{P}_{\xi_n}^{(n)}(t)} E\left(U 1_{(\tau_{n+1}>t)} |\xi_n\right). \tag{3.20}$$

3.3 Compensators and compensating measures

Let Q be a probability on $(W, \mathcal{H})$ determined by the sequence $\left(P^{(n)}\right)$ of Markov kernels, see Theorem 2.1.1. Write $\nu_{z_n}^{(n)}$ for the hazard measure for $P_{z_n}^{(n)}$, cf. Section 3.1 so that

$$\frac{d\nu_{z_n}^{(n)}}{dP_{z_n}^{(n)}} = \frac{1}{\overline{P}_{z_n}^{(n)}(t-)}$$

when $\overline{P}_{z_n}^{(n)}(t-) > 0$.

Definition 3.3.1 *The* compensator *for Q is the process $\Lambda^\circ = (\Lambda_t^\circ)_{t \geq 0}$ on $(W, \mathcal{H})$ given by*

$$\Lambda_t^\circ = \sum_{n=0}^{N_t^\circ} \nu_{\xi_n}^{(n)} \left( ]\tau_n, \tau_{n+1} \wedge t] \right). \tag{3.21}$$

Note that only for the last term, $n = N_t^\circ$, is $\tau_{n+1} \wedge t = t$.

Clearly, for all $w \in W$, $t \to \Lambda_t^\circ(w)$ is ≥ 0, 0 at time 0, increasing and right-continuous. The ambiguity in the choice of the Markov kernels (see p.14) translates into the following ambiguity about the compensator: if $\left(\widetilde{P}^{(n)}\right)$ is another sequence of Markov kernels generating Q with resulting compensator $\widetilde{\Lambda}^\circ$, then Λ° and $\widetilde{\Lambda}^\circ$ are $Q-$indistinguishable, i.e.

$$Q \bigcap_{t \geq 0} \left(\widetilde{\Lambda}_t^\circ = \Lambda_t^\circ \right) = 1. \tag{3.22}$$

As it stands Λ_t° can take the value ∞. However, by (3.5) it follows that

$$Q \bigcap_{t \geq 0} (\Lambda_t^\circ < \infty) = 1.$$

Another important property of the compensator, see Theorem 3.1.1 (iii), is that

$$Q \bigcap_{t \geq 0} (\Delta\Lambda_t^\circ \leq 1) = 1. \tag{3.23}$$

The definition of what corresponds to the compensator for a probability Q on $(\mathcal{M}, \mathcal{H})$ is more involved.

Let $\left(P^{(n)}, \pi^{(n)}\right)$ be the sequences of Markov kernels generating Q. We start by defining the *total compensator* as the process $\overline{\Lambda}^\circ = \left(\overline{\Lambda}_t^\circ\right)_{t \geq 0}$ on $(\mathcal{M}, \mathcal{H})$ given by

$$\overline{\Lambda}_t^\circ = \sum_{n=0}^{\overline{N}_t^\circ} \nu_{\xi_n}^{(n)} \left( ]\tau_n, \tau_{n+1} \wedge t] \right) \tag{3.24}$$

with $\nu_{z_n}^{(n)}$ the hazard measure for $P_{z_n}^{(n)}$. This definition mimics (3.21) and $\overline{\Lambda}^{\circ}$ has the same properties as Λ° in the CP-case as listed above. In particular, for $Q-$a.a. m, the right-continuous function $t \to \overline{\Lambda}_t^{\circ}(m)$ defines a positive measure $\overline{\Lambda}^{\circ}(dt, m)$ on $\mathbb{R}_0$ with $\overline{\Lambda}^{\circ}(\{0\}, m) = \overline{\Lambda}_0^{\circ}(m) = 0$, $\overline{\Lambda}^{\circ}([0,t], m) = \overline{\Lambda}_t^{\circ}(m) < \infty$ for all t and $\Delta\overline{\Lambda}_t^{\circ}(m) = \overline{\Lambda}^{\circ}(\{t\}, m) \leq 1$ for all t.

For $A \in \mathcal{E}$ we define the $Q-$compensator for the CP $N^{\circ}(A)$ as the process $\Lambda^{\circ}(A)$ given by

$$\Lambda_t^{\circ}(A) = \int_{]0,t]} \pi_{\xi_{s-},s}^{(\overline{N}_{s-}^{\circ})}(A)\, \overline{\Lambda}^{\circ}(ds). \tag{3.25}$$

Definition 3.3.2 *The* compensating measure *for Q is the random, non-negative, $Q-$a.s. $\sigma-$finite measure L° on $\mathbb{R}_0 \times E$ given by*

$$L^{\circ}(C) = \int_{\mathbb{R}_0} \int_E 1_C(s,y)\, \pi_{\xi_{s-},s}^{(\overline{N}_{s-}^{\circ})}(dy)\, \overline{\Lambda}^{\circ}(ds) \quad (C \in \mathcal{B}_0 \otimes \mathcal{E}).$$

That L° is a random measure as described in the definition means of course that for $Q-$a.a. m, $C \to L^{\circ}(C, m)$ is a positive, $\sigma-$finite measure on $\mathcal{B}_0 \otimes \mathcal{E}$.

Note that

$$\Lambda_t^{\circ}(A) = L^{\circ}([0,t] \times A), \quad \overline{\Lambda}_t^{\circ} = L^{\circ}([0,t] \times E).$$

An essential property of compensators is presented in

Proposition 3.3.1 *(a) The compensator Λ° for a probability Q on $(W, \mathcal{H})$ is predictable.*

(b) The compensators $\Lambda^{\circ}(A)$ for the counting processes $N^{\circ}(A)$ under a probability Q on $(\mathcal{M}, \mathcal{H})$ are predictable for all $A \in \mathcal{E}$.

Proof. We prove (a), which is good enough to pinpoint the critical part of the argument. Keeping Proposition 3.2.1 (aiv) in mind, on $\left(N_{t-}^{\circ} = n\right)$ where either $N_t^{\circ} = n$ or $N_t^{\circ} = n+1$,

$$\Lambda_t^{\circ} = \sum_{k=0}^{n-1} \nu_{\xi_k}^{(k)}(]\tau_k, \tau_{k+1}]) + \nu_{\xi_n}^{(n)}(]\tau_n, t]) \tag{3.26}$$

if $N_t^{\circ} = n$, and

$$\Lambda_t^{\circ} = \sum_{k=0}^{n} \nu_{\xi_k}^{(k)}(]\tau_k, \tau_{k+1}]) + \nu_{\xi_{n+1}}^{(n+1)}(]\tau_{n+1}, t])$$

if $N_t^\circ = n+1$. But in this latter case, $\tau_{n+1} = t$ and it is seen that the expression (3.26) still holds and thus also the representation (3.11). ■

It is critically important that it is possible to reconstruct the Markov kernels generating a probability Q from the compensators: consider first the CP case. Then using (3.21) one finds that for any n, $z_n = (t_1, \ldots, t_n)$,

$$\nu_{z_n}^{(n)}(]t_n, t]) = \Lambda_t^\circ(w) - \Lambda_t^\circ(w)$$

for any $w \in \left(N_{t-}^\circ = n, \xi_n = z_n\right)$. Similarly, in the MPP case, cf. (3.24), for any n, $z_n = (t_1, \ldots, t_n; y_1, \ldots, y_n)$,

$$\nu_{z_n}^{(n)}(]t_n, t]) = \overline{\Lambda}_t^\circ(m) - \overline{\Lambda}_t^\circ(m) \tag{3.27}$$

for any $m \in \left(\overline{N}_{t-}^\circ = n, \xi_n = z_n\right)$. To extract the kernels $\pi_{z_n,t}^{(n)}$ is more elaborate and based on the fact, obvious from (3.25), that the measure $\Lambda^\circ(dt, A)$ on $\mathbb{R}_0$ determined from the right-continuous process $t \to \Lambda_t^\circ(A)$ is absolutely continuous with respect to the measure $\overline{\Lambda}^\circ(dt)$ with Radon-Nikodym derivative

$$\frac{d\Lambda^\circ(A)}{d\overline{\Lambda}^\circ}(t) = \pi_{\xi_{t-},t}^{(\overline{N}_{t-}^\circ)}(A).$$

Thus

$$\pi_{z_n,t}^{(n)}(A) = \frac{d\Lambda^\circ(A)}{d\overline{\Lambda}^\circ}(t, m) \tag{3.28}$$

for any $m \in \left(\overline{N}_{t-}^\circ = n, \xi_n = z_n\right)$. The only problem here is that the Radon-Nikodym derivative is determined for $\overline{\Lambda}^\circ(dt, m)$ –a.a t only, with an exceptionel set depending on A and m, so care is needed to obtain e.g. that $\pi_{z_n,t}^{(n)}$ is always a probability.

Even though the Markov kernels are obtainable from the compensators, it is just conceivable that two different Q's might have the same compensators. That this is not the case follows from the next result which informally stated shows that *compensators characterize probabilities on* $(W, \mathcal{H})$ *or* $(\mathcal{M}, \mathcal{H})$.

Theorem 3.3.2 *(a) Suppose Λ° is the compensator for some probability Q on $(W, \mathcal{H})$. Then that Q is uniquely determined.*

(b) Suppose L° is the compensating measure for some probability Q on $(\mathcal{M}, \mathcal{H})$. Then that Q is uniquely determined.

Proof. We just consider (b). Suppose that $Q \neq \widetilde{Q}$ are two probabilities on $(\mathcal{M}, \mathcal{H})$ with compensating measures $L^\circ, \widetilde{L}^\circ$. Since $Q \neq \widetilde{Q}$ there is a smallest

$n \in \mathbb{N}$ such that the $Q-$distribution of ξ_n is different from the $\widetilde{Q}-$distribution of ξ_n. Define

$$C_{n-1} = \{z_{n-1} : \begin{array}{l}\text{the conditional } Q - \text{distribution of } (\tau_n, \eta_n) \text{ given} \\ \xi_{n-1} = z_{n-1} \text{ is different from the corresponding} \\ \text{conditional } \widetilde{Q} - \text{distribution}\}.\end{array}$$

By the definition of n, $Q\left(\xi_{n-1} \in C_{n-1}\right) = \widetilde{Q}\left(\xi_{n-1} \in C_{n-1}\right) > 0$, while by the definition of C_{n-1}, the compensating measures $L^\circ, \widetilde{L}^\circ$ for $Q, \widetilde{Q}$ satisfy that $L^\circ(m) \neq \widetilde{L}^\circ(m)$ for all m such that $\xi_{n-1}(m) \in C_{n-1}$. Thus L° is not $\widetilde{Q}-$indistinguishable from $\widetilde{L}^\circ$ and hence, cf. (3.22), L° cannot be a compensating measure for $\widetilde{Q}$. ■

Example 3.3.1 *Suppose Q on W makes N° Poisson λ, see Example 2.1.2. Then the hazard measure for $P^{(n)}_{z_n}$ is the restriction to $]t_n, \infty[$ of λ times Lebesgue measure. Thus*

$$\Lambda^\circ_t = \lambda t,$$

in particular the compensator is deterministic.

Example 3.3.2 *Suppose Q on W makes N° a renewal process with waiting time distribution with hazard measure ν, see Example 2.1.3. Then*

$$\Lambda^\circ_t = \sum_{n=0}^{N^\circ_t} \nu\left(]0, \tau_{n+1} \wedge t - \tau_n]\right).$$

If in particular the waiting time distribution is absolutely continuous with hazard function u, then

$$\Lambda^\circ_t = \int_0^t u\left(s - \tau_{N^\circ_s}\right)\, ds. \tag{3.29}$$

Example 3.3.3 *Suppose E is finite and let Q be the canonical RCM with the Markov kernels described in Example 2.2.1, i.e. the Poisson process with mark space E and parameter vector $(\lambda_y)_{y \in E}$. Then*

$$\Lambda^\circ_t(A) = t \sum_{y \in A} \lambda_y$$

and

$$L^\circ = \ell \otimes \kappa$$

where ℓ is Lebesgue measure and κ is the measure on E given by $\kappa(A) = \sum_A \lambda_y$.

We shall conclude this section with a key lemma that will prove immensely useful in the future. For $s \geq 0$, define the *shift* θ_s mapping W, respectively $\mathcal{M}$, into itself, by

$$(\theta_s N^\circ)_t = \begin{cases} N_t^\circ - N_s^\circ & \text{if } t \geq s \\ 0 & \text{if } t < s, \end{cases} \qquad \theta_s \mu^\circ = \mu^\circ\left(\cdot \cap (]s, \infty[\times E)\right).$$

Thus θ_s only contains the points from the original process that belong strictly after time s. Writing $(\tau_{n,s})_{n\geq 1}$, $\left(\eta_{n,s}\right)_{n\geq 1}$ for the sequence of jump times and marks determining θ_s, we have for instance in the MPP case that

$$\tau_{n,s} = \tau_{k+n}, \quad \eta_{n,s} = \eta_{k+n}$$

on $\left(\overline{N}_s^\circ = k\right)$.

Similarly, for $k_0 \in \mathbb{N}$ define $\vartheta_{k_0} = \theta_{\tau_{k_0}}$ as the map $\vartheta_{k_0} : (\tau_{k_0} < \infty) \to W$ or $\mathcal{M}$ given by

$$(\vartheta_{k_0} N^\circ)_t = \begin{cases} N_t^\circ - N_{\tau_{k_0}}^\circ & \text{if } t \geq \tau_{k_0} \\ 0 & \text{if } t < \tau_{k_0}, \end{cases} \qquad \vartheta_{k_0} \mu^\circ = \mu^\circ\left(\cdot \cap (]\tau_{k_0}, \infty[\times E)\right).$$

Lemma 3.3.3 *(a) Let Q be a probability on $(W, \mathcal{H})$ with compensator Λ°, generated by the Markov kernels $P^{(n)}$.*

(i) The conditional distribution of $\theta_s N^\circ$ given $N_s^\circ = k$, $\xi_k = z_k$ for an arbitrary $k \in \mathbb{N}_0$, $z_k = (t_1, \ldots, t_k) \in K^{(k)}$ with $t_k \leq s$ is the probability $\widetilde{Q} = \widetilde{Q}_{|k,z_k}$ on $(W, \mathcal{H})$ generated by the Markov kernels $\widetilde{P}^{(n)}_{|k,z_k}$ given by

$$\widetilde{P}^{(0)}_{|k,z_k} = P^{(k)}_{z_k}\left(\cdot\, |]s, \infty]\right)$$

$$\widetilde{P}^{(n)}_{\widetilde{z}_n|k,z_k} = P^{(k+n)}_{\mathrm{join}(z_k,\widetilde{z}_k)} \quad (n \geq 1),$$

where for $\widetilde{z}_n = \left(\widetilde{t}_1, \ldots, \widetilde{t}_n\right) \in K^{(n)}$ with $\widetilde{t}_1 > s$,

$$\mathrm{join}\left(z_k, \widetilde{z}_n\right) = \left(t_1, \ldots, t_k, \widetilde{t}_1, \ldots, \widetilde{t}_n\right).$$

(ii) The conditional distribution of $\vartheta_{k_0} N^\circ$ given $\xi_{k_0} = z_{k_0}$ for an arbitrary $z_{k_0} = (t_1, \ldots, t_{k_0}) \in K^{(k_0)}$ with $t_{k_0} < \infty$ is the probability $\widetilde{Q} = \widetilde{Q}_{\left|k_0, z_{k_0}\right.}$ on $(W, \mathcal{H})$ generated by the Markov kernels $\widetilde{P}^{(n)}_{\left|k_0,z_{k_0}\right.}$ given by

$$\widetilde{P}^{(0)}_{\left|k_0,z_{k_0}\right.} = P^{(k_0)}_{z_{k_0}}$$

$$\widetilde{P}^{(n)}_{\widetilde{z}_n\left|k_0,z_{k_0}\right.} = P^{(k_0+n)}_{\mathrm{join}\left(z_{k_0},\widetilde{z}_{k_0}\right)} \quad (n \geq 1),$$

where for $\widetilde{z}_n = \left(\widetilde{t}_1, \ldots, \widetilde{t}_n\right) \in K^{(n)}$ *with* $\widetilde{t}_1 > t_{k_0}$,

$$\operatorname{join}\left(z_{k_0}, \widetilde{z}_n\right) = \left(t_1, \ldots, t_{k_0}, \widetilde{t}_1, \ldots, \widetilde{t}_n\right).$$

(b) Let Q *be a probability on* $(\mathcal{M}, \mathcal{H})$, *generated by the Markov kernels* $\left(P^{(n)}\right)$, $\left(\pi^{(n)}\right)$.

(i) The conditional distribution of $\theta_s \mu^\circ$ *given* $\overline{N}_s = k$, $\xi_k = z_k$ *for an arbitrary* $k \in \mathbb{N}_0$, $z_k = (t_1, \ldots, t_k; y_1, \ldots, y_k) \in K^{(k)}(E)$ *with* $t_k \leq s$ *is the probability* $\widetilde{Q} = \widetilde{Q}_{|k,z_k}$ *on* $(\mathcal{M}, \mathcal{H})$ *generated by the Markov kernels* $\tilde{P}^{(n)}_{|k,z_k}$, $\tilde{\pi}^{(n)}_{|k,z_k}$ *given by*

$$\widetilde{P}^{(0)}_{|k,z_k} = P^{(k)}_{z_k}\left(\cdot \,|]s,\infty]\right),$$

$$\widetilde{P}^{(n)}_{\widetilde{z}_n|k,z_k} = P^{(k+n)}_{\operatorname{join}(z_k,\widetilde{z}_n)} \quad (n \geq 1),$$

$$\widetilde{\pi}^{(n)}_{\widetilde{z}_n,t|k,z_k} = \pi^{(k+n)}_{\operatorname{join}(z_k,\widetilde{z}_n),t} \quad \left(n \geq 0, t > \widetilde{t}_n\right)$$

where for $\widetilde{z}_n = \left(\widetilde{t}_1, \ldots, \widetilde{t}_n; \widetilde{y}_1, \ldots, \widetilde{y}_n\right) \in K^{(n)}(E)$ *with* $\widetilde{t}_1 > s$,

$$\operatorname{join}\left(z_k, \widetilde{z}_n\right) = \left(t_1, \ldots, t_k, \widetilde{t}_1, \ldots, \widetilde{t}_n; y_1, \ldots, y_k, \widetilde{y}_1, \ldots \widetilde{y}_n\right).$$

(ii) The conditional distribution of $\vartheta_{k_0} \mu^\circ$ *given* $\xi_{k_0} = z_{k_0}$ *for an arbitrary* $z_{k_0} = (t_1, \ldots, t_{k_0}; y_1, \ldots, y_{k_0}) \in K^{(k_0)}(E)$ *with* $t_{k_0} < \infty$ *is the probability* $\widetilde{Q} = \widetilde{Q}_{\left|k_0, z_{k_0}\right.}$ *on* $(\mathcal{M}, \mathcal{H})$ *generated by the Markov kernels* $\tilde{P}^{(n)}_{\left|k_0,z_{k_0}\right.}$, $\tilde{\pi}^{(n)}_{\left|k_0,z_{k_0}\right.}$ *given by*

$$\widetilde{P}^{(0)}_{\left|k_0,z_{k_0}\right.} = P^{(k_0)}_{z_{k_0}},$$

$$\widetilde{P}^{(n)}_{\widetilde{z}_n\left|k_0,z_{k_0}\right.} = P^{(k_0+n)}_{\operatorname{join}\left(z_{k_0},\widetilde{z}_n\right)} \quad (n \geq 1),$$

$$\widetilde{\pi}^{(n)}_{\widetilde{z}_n,t\left|k_0,z_{k_0}\right.} = \pi^{(k_0+n)}_{\operatorname{join}\left(z_{k_0},\widetilde{z}_n\right),t} \quad \left(n \geq 0, t > \widetilde{t}_n\right)$$

where for $\widetilde{z}_n = \left(\widetilde{t}_1, \ldots, \widetilde{t}_n; \widetilde{y}_1, \ldots, \widetilde{y}_n\right) \in K^{(n)}(E)$ *with* $\widetilde{t}_1 > t_{k_0}$,

$$\operatorname{join}\left(z_{k_0}, \widetilde{z}_n\right) = \left(t_1, \ldots, t_{k_0}, \widetilde{t}_1, \ldots, \widetilde{t}_n; y_1, \ldots, y_{k_0}, \widetilde{y}_1, \ldots \widetilde{y}_n\right).$$

Note. By Corollary 3.2.2, the conditional probabilities described in (ai) and (bi) simply determine the conditional distribution of θ_s given $\mathcal{H}_s$.

Proof. We only outline the proof. The expressions for $\widetilde{P}^{(0)}_{|k,z_k}$ in (ai), (bi) follow immediately from Corollary 3.2.2 and (3.20). The remaining assertions are consequences of the fact that conditining first on e.g. the first k jump times and marks, and then, within that conditioning, also on the times and marks for the next n jumps, is the same as conditioning from the start on the $k+n$ first jump times and marks. ■

The formulation we have given of Lemma 3.3.3 uses the Markov kernels to desribe the conditional distributions of θ_s and ϑ_{k_0}. Alternatively they may be described using compensators and compensating measures and here, using Proposition 3.1.3 (i), it is seen e.g. that if Q is a probability on $(\mathcal{M}, \mathcal{H})$ with compensating measure L°, then the conditional distribution of θ_s given $\overline{N}_s^\circ = k$, $\xi_k = z_k$ has a compensating measure $L^\circ_{|k,z_k}$ which is the restriction of L° to $[s, \infty[$ in the following sense: for any $m_0 \in \mathcal{M}$ such that $\overline{N}_s(m_0) = k$, $\xi_k(m_0) = z_k$,

$$L^\circ_{|k,z_k}(m, C) = L^\circ(\mathrm{cross}(m_0, m), C \cap (]s, \infty[\times E))$$

for arbitrary $m \in \mathcal{M}$, $C \in \mathcal{B}_0 \otimes \mathcal{E}$, where $\mathrm{cross}(m_0, m) \in \mathcal{M}$ is obtained by using m_0 on $[0, s] \times E$, m on $]s, \infty[\times E$, i.e.

$$\mathrm{cross}(m_0, m) = m_0(\cdot \cap [0, s] \times E) + m(\cdot \cap]s, \infty[\times E).$$

We have in this section only discussed compensators for CP's and RCM's that do not explode (see p.9). The definitions carry over verbatim to processes with explosion (adding e.g. the requirement $L^\circ([\tau_\infty, \infty[\times E) \equiv 0$ with $\tau_\infty = \inf\left\{t : \overline{N}_t^\circ = \infty\right\}$ the time of explosion). One may then show that in fact $\tau_\infty = \inf\left\{t : \overline{\Lambda}_t^\circ = \infty\right\}$ a.s. which yields a (useless) criterion for deciding whether explosions are possible or not.

A more useful criterion for explosion in terms of compensators is the following: let Q be a probability on $(\mathcal{M}, \mathcal{H})$ (no explosions) and let $\widetilde{Q}$ be a probability on $\left(\overline{\mathcal{M}}, \overline{\mathcal{H}}\right)$, the space of exploding discrete counting measures $\overline{m}$ (i.e. $\overline{N}_t^\circ(\overline{m}) = \overline{m}([0, t] \times E) = \infty$ is possible). Let Q have total compensator $\overline{\Lambda}^\circ$, $\widetilde{Q}$ have total compensator $\widetilde{\overline{\Lambda}}^\circ$. Finally, for $m, \widetilde{m} \in \mathcal{M}$ write

$$\widetilde{m} \preceq m$$

if for all t,

$$\overline{N}_t^\circ(\widetilde{m}) \le \overline{N}_t^\circ(m).$$

Proposition 3.3.4 *Assume that $\overline{\Lambda}^\circ$ and $\widetilde{\overline{\Lambda}}^\circ$ are continuous. A sufficient condition for $\widetilde{Q}$ to be non-exploding is that*

$$\widetilde{\overline{\Lambda}}_t^\circ(\widetilde{m}) - \widetilde{\overline{\Lambda}}_s^\circ(\widetilde{m}) \le \overline{\Lambda}_t^\circ(m) - \overline{\Lambda}_s^\circ(m) \tag{3.30}$$

for all $s \le t$ and all $\widetilde{m} \preceq m$ with $\overline{N}_t^\circ(\widetilde{m}) = \overline{N}_s^\circ(\widetilde{m}) = \overline{N}_t^\circ(m) = \overline{N}_s^\circ(m)$.

Proof. An outline of the proof: let $\widetilde{P}^{(n)}_{\widetilde{z}_n}$ and $P^{(n)}_{z_n}$ denote the Markov kernels generating the jump times for $\widetilde{Q}$ and Q respectively. (3.30) implies that for any n and any $\widetilde{z}_n = (\widetilde{t}_1, \ldots, \widetilde{t}_n; \widetilde{y}_1, \ldots, \widetilde{y}_n)$, $z_n = (t_1, \ldots, t_n; y_1, \ldots, y_n)$ with $\widetilde{t}_k \le t_k$ for $k = 1, \ldots, n$, the distribution $\widetilde{P}^{(n)}_{\widetilde{z}_n}$ is stochastically larger than $P^{(n)}_{z_n}$, i.e.

$$\overline{\widetilde{P}}^{(n)}_{\widetilde{z}_n}(t) \ge \overline{P}^{(n)}_{z_n}(t) \tag{3.31}$$

for all $t \ge t_n$. (For this one needs the continuity of the total compensators: because of this assumption, in terms of hazard measures (3.31) reads

$$\exp\left(-\widetilde{\nu}^{(n)}_{\widetilde{z}_n}\left(]\widetilde{t}_n, t]\right)\right) \ge \exp\left(-\nu^{(n)}_{\widetilde{z}_n}\left(]t_n, t]\right)\right)$$

which follows easily from (3.30)). But by a coupling construction it is then possible to define on some probability space, random sequences $\left(\widetilde{T}_n, \widetilde{Y}_n\right)$ and (T_n, Y_n) of two sets of jump times and marks such that $\widetilde{\mu} := \sum_{n:\widetilde{T}_n<\infty} \varepsilon_{\left(\widetilde{T}_n, \widetilde{Y}_n\right)}$ has distribution $\widetilde{Q}$, $\mu := \sum_{n:T_n<\infty} \varepsilon_{(T_n,Y_n)}$ has distribution Q, *and* for every n, $\widetilde{T}_n \ge T_n$. Since by assumption μ does not explode, neither does $\widetilde{\mu}$. ■

A quite useful consequence of this result is presented in Corollary 3.4.4 in the next section.

3.4 Intensity processes

We shall in this section discuss the case where compensators can be represented as ordinary Lebesgue integrals.

Definition 3.4.1 *(a) Let Q be a probability on $(W, \mathcal{H})$ with compensator Λ°. A predictable process $\lambda^\circ \ge 0$ is an* intensity process *for Q if $Q-$a.s.*

$$\Lambda^\circ_t = \int_0^t \lambda^\circ_s \, ds \quad (t \ge 0).$$

(b) Let Q be a probability on $(\mathcal{M}, \mathcal{H})$ with compensating measure L°.

(i) Let $A \in \mathcal{E}$. A predictable process $\lambda^\circ(A) \ge 0$ is an intensity process *for the counting process $N^\circ(A)$ under Q if $Q-$a.s.*

$$\Lambda^\circ_t(A) = \int_0^t \lambda^\circ_s(A) \, ds \quad (t \ge 0).$$

(ii) *If for arbitrary* $A \in \mathcal{E}$ *there is an intensity process* $\lambda^\circ(A)$ *for* $N^\circ(A)$ *under* Q *such that* $Q-$*a.s.* $A \to \lambda_t^\circ(A)$ *is a positive measure on* $(E, \mathcal{E})$ *for all* t*, then the collection* $(\lambda^\circ(A))_{A\in\mathcal{E}}$ *is an* intensity measure *for* Q.

(iii) *If* κ *is a positive,* $\sigma-$*finite measure on* $(E, \mathcal{E})$ *and if* $\lambda^\circ = \left(\lambda^{\circ y}\right)_{y\in E}$ *is a collection of predictable processes* $\lambda^{\circ y} \geq 0$ *such that* $Q-$*a.s.*

$$\Lambda_t^\circ(A) = \int_0^t \int_A \lambda_s^{\circ y} \, \kappa\,(dy)\; ds \quad (t \geq 0, A \in \mathcal{E})\,,$$

then λ° *is a* $\kappa-$intensity process *for* Q.

Note that we have demanded that all intensities be predictable. It is perfectly possible to work with intensities that are adapted but not predictable and still give predictable (since adapted and continuous) compensators by integration, cf. Proposition 3.4.2. However, in some contexts (see Proposition 3.4.3 and Theorem 4.0.2 below) it is essential to have the intensities predictable.

The requirement in (biii) stipulates that the compensating measure L° for Q should have a density with respect to $\ell\otimes\kappa$: $L^\circ\,(dt, dy) = \lambda_t^y \, dt\, \kappa\,(dy)\,.$

Of course if λ° is a $\kappa-$intensity for Q, then $(\lambda^\circ(A))_{A\in\mathcal{E}}$ is an intensity measure for Q, where

$$\lambda_t^\circ(A) = \int_A \lambda_t^{\circ y} \, \kappa\,(dy)\,,$$

and each $\lambda^\circ(A)$ defined this way is an intensity process for $N^\circ(A)$.

The following result gives sufficient conditions for existence of intensities. Recall the definition p.22 of hazard functions.

Proposition 3.4.1 *(a) Let* Q *be a probability on* $(W, \mathcal{H})$ *determined by the Markov kernels* $P^{(n)}$*. Assume that* $Q-$*a.s. for every* n*,* $P_{\xi_n}^{(n)}$ *is absolutely continuous with respect to Lebesgue measure with hazard function* $u_{\xi_n}^{(n)}$*. Then*

$$\lambda_t^\circ = u_{\xi_{t-}}^{\left(N_{t-}^\circ\right)}(t) \tag{3.32}$$

is an intensity process for Q.

(b) Let Q *be a probability on* $(\mathcal{M}, \mathcal{H})$ *determined by the Markov kernels* $P^{(n)}$ *and* $\pi^{(n)}$.

(i) *Assume that Q–a.s. for every n, $P^{(n)}_{\xi_n}$ is absolutely continuous with respect to Lebesgue measure with hazard function $u^{(n)}_{\xi_n}$. Then $(\lambda^\circ(A))_{A\in\mathcal{E}}$ is an intensity measure for Q, where*

$$\lambda^\circ_t(A) = u^{(\overline{N}^\circ_{t-})}_{\xi_{t-}}(t)\pi^{(\overline{N}^\circ_{t-})}_{\xi_{t-},t}(A). \tag{3.33}$$

(ii) *If in addition to the assumption in (i), it also holds that there is a positive, σ–finite measure κ on $(E,\mathcal{E})$ such that Q–a.s. for every n and Lebesgue-a.a. t, $\pi^{(n)}_{\xi_n,t}$ is absolutely continuous with respect to κ with density $p^{(n)}_{\xi_n,t}$, then $\left(\lambda^{\circ y}\right)_{y\in E}$ is a κ–intensity for Q, where*

$$\lambda^{\circ y}_t = u^{(\overline{N}^\circ_{t-})}_{\xi_{t-}}(t)p^{(\overline{N}^\circ_{t-})}_{\xi_{t-},t}(y). \tag{3.34}$$

Proof. The predictability of the intensity processes is ensured by the left-limits $\overline{N}^\circ_{t-}$, ξ_{t-} appearing in the expressions. Everything else follows from the definitions of the compensators and the fact that

$$\nu^{(n)}_{\xi_n}(dt) = u^{(n)}_{\xi_n}(t)\,dt.$$

■

Example 3.4.1 *The canonical Poisson counting process with parameter λ has intensity process*

$$\lambda^\circ_t \equiv \lambda,$$

cf. Example 3.3.1.

Example 3.4.2 *For the renewal process (Example 3.3.2), if the waiting time distribution has hazard function u,*

$$\lambda^\circ_t = u\left(t - \tau_{N^\circ_{t-}}\right)$$

is an intensity process. Note that the compensator $\Lambda^\circ_t = \int_0^t \lambda^\circ_s\,ds$ is the same as that given by (3.29), but that the integrand in (3.29) is not predictable.

Example 3.4.3 *Consider the finite-dimensional Poisson process from Example 3.3.3. Then the deterministic quantities*

$$\lambda^{\circ y}_t \equiv \lambda_y$$

define a κ_0–intensity process, where κ_0 is counting measure on E.

The following description is perhaps the one most often associated with the concept of (non-predictable) intensity processes.

Proposition 3.4.2 *(a) Let Q be a probability on $(W, \mathcal{H})$ and assume that Q has an intensity process λ° given by (3.32) such that $Q-$a.s. all limits from the right, $\lambda^\circ_{t+} = \lim_{h \downarrow 0, h>0} \lambda^\circ_{t+h}$ exist. Then for all t, $Q-$a.s.*

$$\lambda^\circ_{t+} = \lim_{h \downarrow 0, h>0} \frac{1}{h} Q\left(N^\circ_{t+h} - N^\circ_t \geq 1 \,|\mathcal{H}_t\right).$$

(b) Let Q be a probability on $(\mathcal{M}, \mathcal{H})$, let $A \in \mathcal{E}$ and assume that $N^\circ(A)$ has intensity process $\lambda^\circ(A)$ given by (3.33) such that $Q-$a.s. all limits from the right, $\lambda^\circ_{t+}(A) = \lim_{h \downarrow 0, h>0} \lambda^\circ_{t+h}(A)$ exist. Then for all t, $Q-$a.s.

$$\lambda^\circ_{t+}(A) = \lim_{h \downarrow 0, h>0} \frac{1}{h} Q\left(\overline{N}^\circ_{t+h} - \overline{N}^\circ_t \geq 1, \eta_{1,t} \in A \,|\mathcal{H}_t\right).$$

Proof. Recall from p.37 that $\eta_{1,t}$ is the mark for the first jump on $]t, \infty]$. The proof is now based on Lemma 3.3.3 and follows by explicit calculations that we show in the MPP case:

$$\begin{aligned}
& \frac{1}{h} Q\left(\overline{N}^\circ_{t+h} - \overline{N}^\circ_t \geq 1, \eta_{1,t} \in A \,|\mathcal{H}_t\right) \\
= \quad & \frac{1}{h} \int_t^{t+h} u^{\left(\overline{N}^\circ_t\right)}_{\xi_t}(s) \pi^{\left(\overline{N}^\circ_t\right)}_{\xi_t, s}(A)\, ds \\
\to \quad & \lambda^\circ_{t+}(A).
\end{aligned}$$

■

Remark 3.4.1 *Proposition 3.4.2 shows in what sense it is possible to build a model (SPP or MPP) from intensities of the form that specifies (approximately) the probability of a certain event happening in the near future, conditionally on the entire past: as we have seen, it is sometimes possible to interpret these kind of intensities as genuine intensity processes, in which case we have a SPP or MPP. There is of course always the usual problem with explosions, but other than that the importance of the result lies in the fact that from an intuitive understanding of the phenomenon one wants to describe, one can often argue the form of the intensities given by limits as in Proposition 3.4.2. That this may lead to a complete specification of the model is certainly a non-trivial observation!*

Warning. In the literature one may often find alternative expressions for e.g. the counting process intensities λ°_{t+} such as

$$\lim_{h\downarrow 0}\frac{1}{h}Q\left(N^\circ_{t+h}-N^\circ_t=1\,|\mathcal{H}_t\right),\quad \lim_{h\downarrow 0}\frac{1}{h}E\left(N^\circ_{t+h}-N^\circ_t\,|\mathcal{H}_t\right).$$

Although often valid, these expressions are not valid in general (for the second version, not even if all $EN^\circ_s<\infty$). Something like

$$\lim_{h\downarrow 0}\frac{1}{h}Q\left(N^\circ_{t+h}-N^\circ_t\geq 2\,|\mathcal{H}_t\right)=0,$$

is required, but this may fail (although the examples may appear artificial).

The right limit intensities in Proposition 3.4.2 are typically not predictable. They must *never* be used in expressions such as (4.7) below. Also the next result will not in general hold for right limit intensities.

Proposition 3.4.3 *(a) Let Q be a probability on $(W,\mathcal{H})$ with intensity process given by (3.32). Then for all $n\geq 1$,*

$$Q\left(\lambda^\circ_{\tau_n}>0,\tau_n<\infty\right)=Q\left(\tau_n<\infty\right).$$

(b) Let Q be a probability on $(\mathcal{M},\mathcal{H})$ with $\kappa-$intensity process given by (3.34). Then for all $n\geq 1$,

$$Q\left(\lambda^{\circ\eta_n}_{\tau_n}>0,\tau_n<\infty\right)=Q\left(\tau_n<\infty\right).$$

Proof. (b). By explicit calculation

$$\begin{aligned}
&Q\left(\lambda^{\circ\eta_n}_{\tau_n}=0,\tau_n<\infty\right)\\
=\;&Q\left(u^{(n-1)}_{\xi_{n-1}}(\tau_n)\,p^{(n-1)}_{\xi_{n-1},\tau_n}(\eta_n)=0,\tau_n<\infty\right)\\
=\;&E1_{(\tau_{n-1}<\infty)}\int_{]\tau_{n-1},\infty[}P^{(n-1)}_{\xi_{n-1}}(dt)\int_E\pi^{(n-1)}_{\xi_{n-1},t}(dy)\;1_C(t,y)
\end{aligned}$$

where

$$C=\left\{(t,y):u^{(n-1)}_{\xi_{n-1}}(t)\,p^{(n-1)}_{\xi_{n-1},t}(y)=0\right\}.$$

Since

$$P^{(n-1)}_{\xi_{n-1}}(dt)=u^{(n-1)}_{\xi_{n-1}}(t)\overline{P}^{(n-1)}_{\xi_{n-1}}(t)\;dt,\quad \pi^{(n-1)}_{\xi_{n-1},t}(dy)=p^{(n-1)}_{\xi_{n-1},t}(y)\;\kappa(dy)$$

by assumption, it is clear that

$$Q\left(\lambda_{\tau_n}^{\circ\eta_n} = 0, \tau_n < \infty\right) = 0.$$

■

Let $Q, \widetilde{Q}$ be probabilities on $(\mathcal{M}, \mathcal{H})$ and $\left(\overline{\mathcal{M}}, \overline{\mathcal{H}}\right)$ respectively (i.e. $\widetilde{Q}$ allows explosions) such that the Markov kernels $P_{z_n}^{(n)}, \widetilde{P}_{\widetilde{z}_n}^{(n)}$ generating the jump times are absolutely continuous with hazard functions $u_{z_n}^{(n)}, \widetilde{u}_{\widetilde{z}_n}^{(n)}$. Let also $\overline{\lambda}^{\circ}, \overline{\overline{\lambda}}^{\circ}$ denote the total intensity processes,

$$\overline{\lambda}_t^{\circ} = u_{\xi_{t-}}^{\left(\overline{N}_{t-}^{\circ}\right)}, \quad \overline{\overline{\lambda}}_t^{\circ} = \widetilde{u}_{\xi_{t-}}^{\left(\overline{N}_{t-}^{\circ}\right)}.$$

Corollary 3.4.4 *A sufficient condition for $\widetilde{Q}$ to be non-exploding is that*

$$\overline{\overline{\lambda}}_t^{\circ}(\widetilde{m}) \leq \overline{\lambda}_t^{\circ}(m)$$

for all $t \geq 0$ and all $\widetilde{m}, m \in \mathcal{M}$ with $\widetilde{m} \preceq m$, $\overline{N}_t^{\circ}(\widetilde{m}) = \overline{N}_t^{\circ}(m)$.

The proof is obvious from Proposition 3.3.4. See p.39 for the meaning of the 'inequality' $\widetilde{m} \preceq m$.

Example 3.4.4 *If $\widetilde{Q}$ is such that there exists constants $a_n \geq 0$ with $\sum 1/a_n = \infty$ and $\overline{\overline{\lambda}}_t^{\circ} \leq a_{\overline{N}_t^{\circ}}$ everywhere on $\mathcal{M}$ for every t, then $\widetilde{Q}$ is non-exploding, cf. p.16.*

3.5 The basic martingales

In this section we shall characterize compensators and compensating measures through certain martingale properties. The main results provide Doob-Meyer decompositions of the counting process N° on $(W, \mathcal{H})$ and the counting processes $N^{\circ}(A)$ on $(\mathcal{M}, \mathcal{H})$: these counting processes are $\mathcal{H}_t-$adapted and increasing, hence they are trivially local submartingales and representable as a local martingale plus an increasing, predictable process, 0 at time 0. As we shall see, the increasing, predictable process is simply the compensator.

The fact that the increasing, predictable process in the Doob-Meyer decomposition is unique (when assumed to equal 0 at time 0) in our setup amounts to Proposition 3.5.1 below.

Recall that a $\overline{\mathbb{R}}_0-$valued map defined on $(W, \mathcal{H})$ or $(\mathcal{M}, \mathcal{H})$ is a *stopping time* provided $(\tau < t) \in \mathcal{H}_t$ for all t, (equivalently, since the filtration is right continuous, $(\tau \leq t) \in \mathcal{H}_t$ for all t). The *pre-τ $\sigma-$algebra* $\mathcal{H}_\tau$ is the collection $\{H \in \mathcal{H} : H \cap (\tau < t) \in \mathcal{H}_t \text{ for all } t\} = \{H \in \mathcal{H} : H \cap (\tau \leq t) \in \mathcal{H}_t \text{ for all } t\}$ of measurable sets. In particular each τ_n is a stopping time, and as may easily be verified, $\mathcal{H}_{\tau_n}$ is the $\sigma-$algebra generated by ξ_n.

Let Q be a probability on $(W, \mathcal{H})$ or $(\mathcal{M}, \mathcal{H})$. A *local $Q-$martingale* is a real-valued, $\mathcal{H}_t-$adapted process $M = (M_t)_{t \geq 0}$, such that each $M^{\rho_n} := \left(M_{\rho_n \wedge t}\right)_{t \geq 0}$ is a true martingale – still with respect to the filtration $(\mathcal{H}_t)$ – for some increasing sequence (ρ_n) of stopping times with $\rho_n \uparrow \infty$ $Q-$a.s. The sequence (ρ_n) is called a *reducing sequence* and we write that M is a *local $Q-$martingale* (ρ_n). By far the most important reducing sequence is the sequence (τ_n) of jump times. (See Appendix B for a general discussion of stopping times, local martingales and the optional sampling theorem.)

Proposition 3.5.1 *Suppose M is a right-continuous, $\mathcal{H}_t-$predictable local Q – martingale on $(W, \mathcal{H})$ or $(\mathcal{M}, \mathcal{H})$. Then M is constant,*

$$Q \bigcap_{t \geq 0} (M_t = M_0) = 1.$$

Proof. Suppose first that M is a right-continuous, predictable martingale. By optional sampling, which applies only because M is right-continuous, for any t,

$$EM_{\tau_1 \wedge t} = EM_0 = M_0$$

since M_0 is a constant random variable, being $\mathcal{H}_0-$measurable. But since M is predictable, by Proposition 3.2.1 there is a Borel function f such that $M_{\tau_1 \wedge t} = f(\tau_1 \wedge t)$ and thus

$$f(0) = f(t)\overline{P}^{(0)}(t) + \int_{]0,t]} f(s)\, P^{(0)}(ds).$$

From this follows by differentiation with respect to $P^{(0)}$ (see Appendix A), that f is constant and so, M is constant on $[0, \tau_1]$. Again by optional sampling, $E\left(M_{\tau_{n+1} \wedge t} \,|\mathcal{H}_{\tau_n \wedge t}\right) = M_{\tau_n \wedge t}$. But the conditional expectation on the left is

$$\begin{aligned} & 1_{(\tau_n \leq t)} E\left(M_{\tau_{n+1} \wedge t} \,|\mathcal{H}_{\tau_n}\right) + 1_{(\tau_n > t)} E\left(M_{\tau_{n+1} \wedge t} \,|\mathcal{H}_t\right) \\ = \; & 1_{(\tau_n \leq t)} E\left(M_{\tau_{n+1} \wedge t} \,|\xi_n\right) + 1_{(\tau_n > t)} M_t \end{aligned}$$

and so, by the same argument using the Markov kernel $P_{\xi_n}^{(n)}$ instead of $P^{(0)}$, one finds that M is constant on $]\tau_n, \tau_{n+1}]$ and equal to M_{τ_n}. It follows immediately that $M \equiv M_0$.

Let now M be a right-continuous and predictable local $Q-$martingale with reducing sequence (ρ_n). We claim that for each n the martingale M^{ρ_n} is in fact predictable and from this the assertion follows from what has been proved already: for each n we have $M^{\rho_n} \equiv M_0$ a.s. and since $\rho_n \uparrow \infty$ a.s. also $M \equiv M_0$ a.s.

It remains to show that if X is a right-continuous and predictable $\mathbb{R}-$valued process and ρ is a stopping time, then X^ρ is predictable. But

$$X_t^\rho = X_t 1_{(\rho \geq t)} + X_\rho 1_{(\rho < t)}$$

and here the first term on the right defines a predictable process since X and the left-continuous indicator process $\left(1_{(\rho \geq t)}\right)_{t \geq 0}$ are predictable. The second term is predictable since the process $\left(X_\rho 1_{(\rho < t)}\right)_{t \geq 0}$ is left-continuous and for all t,

$$X_\rho 1_{(\rho < t)} = \lim_{K \to \infty} \sum_{k=1}^{\infty} X_{\frac{k}{2^K} \wedge t} 1_{\left(\frac{k-1}{2^K} \wedge t \leq \rho < \frac{k}{2^K} \wedge t\right)}$$

is $\mathcal{H}_t-$measurable. ■

Remark 3.5.1 *The result is peculiar to the point process setup: Brownian motion is the most famous example of a continuous martingale which is not constant! The assumption that M be right-continuous is also important: it is easy to find cadlag (in particular right-continuous) martingales M that are not constant, and such that $Q(M_t = M_{t-}) = 1$ for all t. But then (M_{t-}) is a left-continuous, hence predictable martingale, which is non-constant.*

Theorem 3.5.2 *(a) Let Q be a probability on $(W, \mathcal{H})$ with compensator Λ°. Then $M^\circ := N^\circ - \Lambda^\circ$ is a local $Q-$martingale (τ_n) and Λ° is, up to $Q-$indistinguishability, the unique right-continuous $\mathcal{H}_t-$predictable process, 0 at time 0, such that M° is a $Q-$local martingale. A sufficient condition for M° to be a $Q-$martingale is that $EN_t^\circ < \infty$ for all t.*

(b) Let Q be a probability on $(\mathcal{M}, \mathcal{H})$ with compensating measure L° and compensators $\Lambda^\circ(A)$, $\Lambda_t^\circ(A) = L^\circ([0,t] \times A)$. Then, for any $A \in \mathcal{E}$, $M^\circ(A) := N^\circ(A) - \Lambda^\circ(A)$ is a local $Q-$martingale (τ_n) and $\Lambda^\circ(A)$ is, up to $Q-$indistinguishability, the unique right-continuous $\mathcal{H}_t-$predictable process, 0 at time 0, such that $M^\circ(A)$ is a $Q-$local martingale. A sufficient condition for $M^\circ(A)$ to be a $Q-$martingale is that $EN_t^\circ(A) < \infty$ for all t.

Proof. The proof relies on a technique that will be used also on several occasions in the sequel, and is therefore here presented in detail. We consider the more difficult case (b) only and will start by showing that $M^{\circ}(A)$ is a $Q-$martingale if $EN_t^{\circ}(A) < \infty$ for all t.

The idea is to argue that for this it suffices to prove that

$$EN^{\circ}_{\tau_1 \wedge t}(A) = E\Lambda^{\circ}_{\tau_1 \wedge t}(A), \tag{3.35}$$

for *all* probabilities Q on $(\mathcal{M}, \mathcal{H})$, where of course $\Lambda^{\circ}(A)$ is the compensator for the Q considered, and then verify (3.35) by explicit calculation. (Note that since $0 \le N^{\circ}_{\tau_1 \wedge t}(A) \le 1$ the expectation on the left is trivially finite for all Q, in particular it follows from (3.35) that $\Lambda^{\circ}_{\tau_1 \wedge t}(A)$ is $Q-$integrable for all Q and all t and A).

We claim first that from Lemma 3.3.3 (b) and (3.35) it follows that for all $n \in \mathbb{N}_0$ and all $t \ge 0$,

$$E\left(N^{\circ}_{\tau_{n+1} \wedge t}(A) - N^{\circ}_{\tau_n \wedge t}(A) \,|\xi_n\right) = E\left(\Lambda^{\circ}_{\tau_{n+1} \wedge t}(A) - \Lambda^{\circ}_{\tau_n \wedge t}(A) \,|\xi_n\right). \tag{3.36}$$

This identity is obvious on $(\tau_n > t)$ and on $(\tau_n \le t)$ is just (3.35) applied to the conditional distribution of the shifted process $\vartheta_n \mu^{\circ}$, see Lemma 3.3.3.

It is an immediate consquence of (3.35) and (3.36) that for all Q, A, and all n and t,

$$EN^{\circ}_{\tau_n \wedge t}(A) = E\Lambda^{\circ}_{\tau_n \wedge t}(A)$$

with both expectations finite since $N^{\circ}_{\tau_n \wedge t}(A) \le n$. Let $n \uparrow \infty$ and use monotone convergence to deduce that for all Q, A, and all t,

$$EN^{\circ}_t(A) = E\Lambda^{\circ}_t(A), \tag{3.37}$$

whether the expectatons are finite or not. Finally, assuming that $EN_t^{\circ}(A) < \infty$ for all t it follows first for $s < t$ that $E\left(N_t^{\circ}(A) \,|\mathcal{H}_s\right) < \infty$ $Q-$a.s. and then from Lemma 3.3.3 (b) and (3.37) applied to the conditional distribution of $\theta_s \mu^{\circ}$ given $\mathcal{H}_s$, that

$$E\left(N^{\circ}_t(A) - N^{\circ}_s(A) \,|\mathcal{H}_s\right) = E\left(\Lambda^{\circ}_t(A) - \Lambda^{\circ}_s(A) \,|\mathcal{H}_s\right)$$

which by rearrangement of the terms, that are all finite, results in the desired martingale property

$$E\left(M^{\circ}_t(A) \,|\mathcal{H}_s\right) = M^{\circ}_s(A).$$

It remains to establish (3.35). But

$$\begin{aligned} EN^{\circ}_{\tau_1 \wedge t}(A) &= Q\left(\tau_1 \le t, \eta_1 \in A\right) \\ &= \int_{]0,t]} \pi_s^{(0)}(A)\, P^{(0)}(ds), \end{aligned}$$

while

$$\begin{aligned} E\Lambda^\circ_{\tau_1\wedge t}(A) &= E\int_{]0,\tau_1\wedge t]} \pi^{(0)}_s(A)\,\nu^{(0)}(ds) \\ &= \overline{P}^{(0)}(t)\int_{]0,t]} \pi^{(0)}_s(A)\,\nu^{(0)}(ds) \\ &\quad + \int_{]0,t]}\left(\int_{]0,u]} \pi^{(0)}_s(A)\,\nu^{(0)}(ds)\right) P^{(0)}(du) \end{aligned}$$

and (3.35) follows by partial integration, cf. Appendix A.

That $\Lambda^\circ(A)$ is the only predictable process, 0 at time 0, such that $N^\circ(A)-\Lambda^\circ(A)$ is a local $Q-$martingale follows immediately from Proposition 3.5.1: if also $N^\circ(A)-\widetilde{\Lambda}^\circ$ is a local martingale, where $\widetilde{\Lambda}^\circ$ is predictable and 0 at time 0, then $\Lambda^\circ(A)-\widetilde{\Lambda}^\circ$ is a predictable local martingale, 0 at time 0, hence identically equal to 0.

The remaining assertion of the theorem, that $M^\circ(A)$ is always a local $Q-$martingale (τ_n), is easy to verify: the distribution Q_n of the stopped RCM $\mu^{\circ\tau_n} := \mu^\circ(\cdot\cap[0,\tau^\circ_n]\times E)$ obviously has compensating measure $L^{\circ\tau_n} = L^\circ(\cdot\cap[0,\tau^\circ_n]\times E)$, (the Markov kernels $P^{n,(k)}$, $\pi^{n,(k)}$ generating Q_n are those of Q for $k\le n$, while $P^{n,(n+1)}=\varepsilon_\infty$), and since $E_nN^\circ_t(A)\le n<\infty$, by what has been proved above, $M^{\circ\tau_n}(A)$, which is Q_n-indistinguishable from $M^\circ(A)$, is a Q_n-martingale for all n, equivalently $M^\circ(A)$ is a local $Q-$martingale (τ_n). ■

Some of the other important martingales arising directly from the compensators are presented in the next result. Note that the result does not hold in the form presented here without the assumption about continuity of Λ° and $\overline{\Lambda}^\circ$.

Proposition 3.5.3 *(a) Let Q be a probability on $(W,\mathcal{H})$ with continuous compensator Λ°. Then $M^{\circ^2}-\Lambda^\circ$ is a local $Q-$martingale (τ_n), which is a $Q-$martingale if $EN^\circ_t<\infty$ for all t.*

(b) Let Q be a probability on $(\mathcal{M},\mathcal{H})$ with compensators $\Lambda^\circ(A)$ and continuous total compensator $\overline{\Lambda}^\circ$.

(i) For every $A\in\mathcal{E}$, $M^{\circ^2}(A)-\Lambda^\circ(A)$ is a local $Q-$martingale (τ_n), which is a $Q-$martingale if $EN^\circ_t(A)<\infty$ for all t.

(ii) For every $A,A'\in\mathcal{E}$ with $A\cap A'=\emptyset$, $M^\circ(A)M^\circ(A')$ is a local $Q-$martingale (τ_n) which is a $Q-$martingale if $EN^\circ_t(A)<\infty$, $EN^\circ_t(A')<\infty$ for all t.

Proof. Note for (b) that all $\Lambda^\circ(A)$ are continuous when $\overline{\Lambda}^\circ$ is. Otherwise the technique from the proof of Theorem 3.5.2 is used, i.e. for the two parts of the Proposition it is argued that it suffices to show that for all Q, A, A' and t,

$$EM^{\circ 2}_{\tau_1\wedge t} = E\Lambda^\circ_{\tau_1\wedge t},$$

$$EM^{\circ 2}_{\tau_1\wedge t}(A) = E\Lambda^\circ_{\tau_1\wedge t}(A), \quad EM^\circ_{\tau_1\wedge t}(A)M^\circ_{\tau_1\wedge t}(A') = 0$$

respectively. These identities are then verified directly through straightforward calculations leading to some not so straightforward partial integrations.

It is important to point out one step in the argument: to show e.g. in (b) that $M^{\circ 2}(A) - \Lambda^\circ(A)$ is a martingale if all $EN^\circ_t(A) < \infty$ one must deduce that

$$EM^{\circ 2}_t(A) = E\Lambda^\circ_t(A) \tag{3.38}$$

for all t when knowing that

$$EM^{\circ 2}_{\tau_n\wedge t}(A) = E\Lambda^\circ_{\tau_n\wedge t}(A) \tag{3.39}$$

for all t and n. (Since we know from Theorem 3.5.2 that $E\Lambda^\circ_t(A) = EN^\circ_t(A)$ it follows in particular from (3.38) that $EM^{\circ 2}_t(A) < \infty$ as is certainly required for the martingale assertion). To deduce (3.38) from (3.39), use Fatou's lemma and monotone convergence to obtain

$$\begin{aligned} EM^{\circ 2}_t(A) &= E \liminf_{n\to\infty} M^{\circ 2}_{\tau_n\wedge t}(A) \\ &\le \liminf_{n\to\infty} EM^{\circ 2}_{\tau_n\wedge t}(A) \\ &= \lim_{n\to\infty} E\Lambda^\circ_{\tau_n\wedge t}(A) \\ &= E\Lambda^\circ_t(A) \\ &< \infty. \end{aligned} \tag{3.40}$$

But with $M^\circ(A)$ a martingale and $M^{\circ 2}_t(A)$ integrable for all t, $M^{\circ 2}(A)$ is a submartingale and in particular, by optional sampling,

$$EM^{\circ 2}_t(A) \ge EM^{\circ 2}_{\tau_n\wedge t}(A) = E\Lambda^\circ_{\tau_n\wedge t}$$

for all n, t. Letting $n \uparrow \infty$ gives $EM^{\circ 2}_t(A) \ge E\Lambda^\circ_t$ which combined with the inequality (3.40) yields (3.38).

Note, as is relevant for the proof of (bii), that the argument above implies that all $M^\circ_t(A)M^\circ_t(A')$ are integrable if $EN^\circ_t(A)$ and $EN^\circ_t(A')$ are both finite for all t. ∎

Remark 3.5.2 *For (bi) it suffices to assume that $\Lambda^\circ(A)$ is continuous, and for (bii) that $\Delta(\Lambda^\circ(A)\Lambda^\circ(A')) \equiv 0$. Without these assumptions it is still possible to find in (bi) a predictable, increasing process $\widetilde{\Lambda}^\circ$ such that $M^{\circ^2}(A) - \widetilde{\Lambda}^\circ$ is a local martingale, and in (bii) a predictable process Υ, 0 at time 0, such that $M^\circ(A)M^\circ(A') - \Upsilon$ is a local martingale, see Section 3.8 below.*

3.6 Stochastic integrals and martingales

Let Q be a probability on $(\mathcal{M}, \mathcal{H})$ with compensating measure L° (for the results in this section it is natural to consider canonical CP's as a special case of canonical RCM's). We shall first discuss stochastic integrals with respect to μ° and L° and then use them to arrive at the martingale representation theorem, Theorem 3.6.1 below.

The integrands are functions of $m \in \mathcal{M}$, $t \geq 0$ and $y \in E$. A typical integrand is denoted S where $(m,t,y) \to S_t^y(m)$ is assumed to be $\mathbb{R}-$valued and measurable (with respect to the Borel $\sigma-$algebra on $\mathbb{R}$ and the product $\sigma-$algebra $\mathcal{H} \otimes \mathcal{B}_0 \otimes \mathcal{E}$ on $\mathcal{M} \times \mathbb{R}_0 \times E$). Often we shall think of S as a family $(S^y)_{y\in E}$ of processes $S^y = (S_t^y)_{t\geq 0}$ and then refer to S as a *flow* of processes. Particularly important are *predictable flows* which are flows with each S^y predictable.

The *stochastic integral*

$$N^\circ(S) = (N_t^\circ(S))_{t\geq 0}, \quad N_t^\circ(S) := \int_{]0,t]\times E} S_s^y \, \mu^\circ(ds, dy)$$

is always well defined as a $\mathbb{R}-$valued process, and the stochastic intgral is just a finite sum,

$$N_t^\circ(S) = \sum_{\substack{n=1 \\ \tau_n < \infty}}^{\infty} S_{\tau_n}^{\eta_n}.$$

If each S^y is adapted, also $N^\circ(S)$ is adapted.

The stochastic integral

$$\Lambda^\circ(S) = (\Lambda_t^\circ(S))_{t\geq 0}, \quad \Lambda_t^\circ(S)(m) := \int_{]0,t]\times E} S_s^y(m) \; L^\circ(m, ds, dy)$$

is always well defined if $S \geq 0$ (or $S \leq 0$) as a $\overline{\mathbb{R}}_0-$valued process (respectively a process with values in $[-\infty, 0]$) with the integral an ordinary Lebesgue-Stieltjes integral for each m. If $S \geq 0$, in order for $\Lambda^\circ(S)$ to be $Q-$a.s. finite, i.e.

$$Q \bigcap_{t\geq 0} (\Lambda_t^\circ(S) < \infty) = 1,$$

it suffices that

$$Q\bigcap_{t\geq 0}\left(\sup_{s\leq t, y\in E} S_s^y < \infty\right) = 1.$$

For arbitrary S, write $S = S^+ - S^-$ where $S^+ = S\vee 0$, $S^- = -S\wedge 0$ and define $\Lambda^\circ(S) = \Lambda^\circ(S^+) - \Lambda^\circ(S^-)$ whenever $\Lambda^\circ(S^+)$ or $\Lambda^\circ(S^-)$ is $Q-$a.s. finite. In particular $\Lambda^\circ(S)$ is well defined with

$$Q\bigcap_{t\geq 0}\left(|\Lambda_t^\circ(S)| < \infty\right) = 1$$

provided

$$Q\bigcap_{t\geq 0}\left(\sup_{s\leq t, y\in E} |S_s^y| < \infty\right) = 1.$$

If $\Lambda^\circ(S)$ is well defined, it is an adapted process if each S^y is adapted and a predictable process if S is a predictable flow.

Note. If discussing counting processes there is of course no need for predictable flows: S is just a predictable process.

Suppose $\Lambda^\circ(S)$ is well defined and then define the process $M^\circ(S)$ by

$$M_t^\circ(S) = N_t^\circ(S) - \Lambda_t^\circ(S). \tag{3.41}$$

Theorem 3.6.1 *Let Q be a probability on $(W, \mathcal{H})$ or $(\mathcal{M}, \mathcal{H})$.*

(i) Suppose M is a right-continuous local $Q-$martingale. Then there exists a predictable flow $S = (S^y)$ such that

$$M_t = M_0 + M_t^\circ(S). \tag{3.42}$$

(ii) If $S \geq 0$ is a predictable flow, then

(1) $M^\circ(S)$ given by (3.41) is a local $Q-$martingale (τ_n) if for all $n \geq 1$, $t \geq 0$,

$$EN_{\tau_n\wedge t}^\circ(S) < \infty;$$

(2) $M^\circ(S)$ given by (3.41) is a $Q-$martingale if for all $t \geq 0$,

$$EN_t^\circ(S) < \infty.$$

(iii) If S is a predictable flow, then

(1) $M^\circ(S)$ given by (3.41) is a local $Q-$martingale (τ_n) if for all $n \geq 1$, $t \geq 0$,

$$EN^\circ_{\tau_n \wedge t}(|S|) < \infty;$$

(2) $M^\circ(S)$ given by (3.41) is a $Q-$martingale if for all $t \geq 0$,

$$EN^\circ_t(|S|) < \infty.$$

Proof. We outline the main parts of the proof and note first that (iii) follows trivially from (ii). We start with the proof of

(ii). By the technique introduced in the proof of Theorem 3.5.2, it suffices to prove that

$$E\int_{]0,\tau_1\wedge t]\times E} S^y_s\, \mu^\circ(ds,dy) = E\int_{]0,\tau_1\wedge t]\times E} S^y_s\, L^\circ(ds,dy) \tag{3.43}$$

for all t and all Q, with L° the compensating for Q. By Proposition 3.2.1 (biv) there is a function $f(s,y)$, jointly measurable in s and y, such that $S^y_s = f(s,y)$ on $\left(\overline{N}^\circ_{s-} = 0\right)$. Thus (3.43) reduces to

$$E1_{(\tau_1 \leq t)} f(\tau_1,\eta_1) = E\int_{]0,\tau_1\wedge t]} \nu^{(0)}(ds) \int_E \pi^{(0)}_s(dy)\ f(s,y)$$

or

$$\begin{aligned}
&\int_{]0,t]} P^{(0)}(ds) \int_E \pi^{(0)}_s(dy)\ f(s,y)\\
= \ &\overline{P}^{(0)}(t)\int_{]0,t]} \nu^{(0)}(ds) \int_E \pi^{(0)}_s(dy)\ f(s,y)\\
&+\int_{]0,t]} P^{(0)}(du)\int_{]0,u]} \nu^{(0)}(ds) \int_E \pi^{(0)}_s(dy)\ f(s,y).
\end{aligned}$$

This you verify directly by partial integration or differentiation with respect to $P^{(0)}$ (see Appendix A).

An alternative way of proving (ii) is to start with flows S of the form

$$S^y_s = 1_{H_0} 1_{]s_0,\infty[}(s)\, 1_{A_0}(y) \tag{3.44}$$

where $s_0 \geq 0$, $H_0 \in \mathcal{H}_{s_0}$, $A_0 \in \mathcal{E}$, and then extend to all $S \geq 0$ by standard arguments. For S of the form (3.44) the (local) martingale property of $M^\circ(S)$ follows from that of $M^\circ(A_0)$.

(i). Suppose just that M is a right-continuous true $Q-$martingale. Because M is adapted we can write

$$M_t = f^{(n)}_{\xi_n}(t) \quad \text{on } \left(\overline{N}^\circ_t = n\right) \tag{3.45}$$

for all n, t, cf. Proposition 3.2.1 (biii). By optional sampling

$$E\left(M_{\tau_{n+1}\wedge t} - M_{\tau_n} \left| \xi_n \right.\right) = 0 \quad \text{on } \left(\tau_n \leq t\right),$$

an identity which using (3.45) we may write

$$\begin{aligned}&\overline{P}^{(n)}_{\xi_n}(t)\left(f^{(n)}_{\xi_n}(t) - f^{(n)}_{\xi_n}(\tau_n)\right)\\ &+\int_{]\tau_n,t]} P^{(n)}_{\xi_n}(ds)\int_E \pi^{(n)}_{\xi_n,s}(dy)\left(f^{(n+1)}_{\mathrm{join}(\xi_n,(s,y))}(s) - f^{(n)}_{\xi_n}(\tau_n)\right) = 0\end{aligned} \tag{3.46}$$

on $(\tau_n \leq t)$, where $\mathrm{join}(\xi_n, (s,y)) = (\tau_1, \ldots, \tau_n, s; \eta_1, \ldots, \eta_n, y)$.

We want to find S such that (3.42) holds. Since each S^y is predictable we may write

$$S^y_s = g^{(n)}_{\xi_n}(s,y) \quad \text{on } \left(\overline{N}^\circ_{s-} = n\right)$$

so that on $(\tau_n < t)$ we have that

$$\begin{aligned}&M^\circ_{\tau_{n+1}\wedge t}(S) - M^\circ_{\tau_n}(S)\\ =&\begin{cases} -\int_{]\tau_n,t]}\nu^{(n)}_{\xi_n}(ds)\int_E \pi^{(n)}_{\xi_n,s}(dy)\, g^{(n)}_{\xi_n}(s,y)\\ g^{(n)}_{\xi_n}\left(\tau_{n+1},\eta_{n+1}\right) - \int_{]\tau_n,\tau_{n+1}]}\nu^{(n)}_{\xi_n}(ds)\int_E \pi^{(n)}_{\xi_n,s}(dy)\, g^{(n)}_{\xi_n}(s,y)\end{cases}\end{aligned}$$

with the top expression valid if $t < \tau_{n+1}$, that on the bottom if $t \geq \tau_{n+1}$. But also on $(\tau_n < t)$,

$$M_{\tau_{n+1}\wedge t} - M_{\tau_n} = \begin{cases} f^{(n)}_{\xi_n}(t) - f^{(n)}_{\xi_n}(\tau_n) & \text{if } t < \tau_{n+1},\\ f^{(n+1)}_{\xi_{n+1}}(\tau_{n+1}) - f^{(n)}_{\xi_n}(\tau_n) & \text{if } t \geq \tau_{n+1},\end{cases}$$

and it is seen that $M^\circ(S) \equiv M$ if for all n, t, y on $(\tau_n < t)$ it holds that

$$\begin{aligned}-\int_{]\tau_n,t]}\nu^{(n)}_{\xi_n}(ds)\int_E \pi^{(n)}_{\xi_n,s}(dy)\, g^{(n)}_{\xi_n}(s,y) &= f^{(n)}_{\xi_n}(t) - f^{(n)}_{\xi_n}(\tau_n),\\ g^{(n)}_{\xi_n}(t,y) - \int_{]\tau_n,t]}\nu^{(n)}_{\xi_n}(ds)\int_E \pi^{(n)}_{\xi_n,s}(dy)\, g^{(n)}_{\xi_n}(s,y) &= f^{(n+1)}_{\mathrm{join}(\xi_n,(t,y))}(t) - f^{(n)}_{\xi_n}(\tau_n)\end{aligned}$$

or that

$$-\int_{]\tau_n,t]}\nu^{(n)}_{\xi_n}(ds)\int_E \pi^{(n)}_{\xi_n,s}(dy)\, g^{(n)}_{\xi_n}(s,y) = f^{(n)}_{\xi_n}(t) - f^{(n)}_{\xi_n}(\tau_n) \tag{3.47}$$

and

$$g^{(n)}_{\xi_n}(t,y) = f^{(n+1)}_{\mathrm{join}(\xi_n,(t,y))}(t) - f^{(n)}_{\xi_n}(t), \tag{3.48}$$

where equation (3.48) defines $g^{(n)}_{\xi_n}$. But from (3.46) it follows that $t \to f(t) := f^{(n)}_{\xi_n}(t)$ is differentiable with respect to $P := P^{(n)}_{\xi_n}$ and by the differentiation rule

$$D_P(F_1F_2)(t) = (D_PF_1)(t)F_2(t) + F_1(t-)(D_PF_2)(t),$$

(see Appendix A), (3.46) implies that

$$-\left(f\left(t\right)-f\left(\tau_n\right)\right)+\overline{P}\left(t-\right)D_P f(t)=-\int_E \pi^{(n)}_{\xi_n,t}\left(dy\right)\,\left(f^{(n+1)}_{\mathrm{join}(\xi_n,(t,y))}\left(t\right)-f\left(\tau_n\right)\right),$$

or equivalently, using (3.48)

$$\overline{P}\left(t-\right)D_P f(t)=-\int_E \pi^{(n)}_{\xi_n,t}\left(dy\right)\,g^{(n)}_{\xi_n}\left(t,y\right). \tag{3.49}$$

From (3.46) we see in particular that $f(t)$ is right-continuous for $t\geq\tau_n$, and since both sides of (3.47) vanish as $t\downarrow\tau_n$, with the left hand side obviously differentiable with resect to P, to prove (3.47) it suffices to show that

$$D_P f(t)=D_P\left(-\int_{]\tau_n,t]}\nu^{(n)}_{\xi_n}\left(ds\right)\,\int_E \pi^{(n)}_{\xi_n,s}\left(dy\right)\,g^{(n)}_{\xi_n}\left(s,y\right)\right).$$

But recalling the definition (3.1) of hazard measures, it is seen that this is precisely (3.49). ■

Remark 3.6.1 *It is often important to be able to show that a local martingale is a martingale. The conditions in Theorem 3.6.1 (iib) and (iiib) are sufficient for this but far from necessary.*

We shall conclude this section by quoting some important identities involving the socalled *quadratic characteristics* and *cross characteristics* for (local) martingales.

Suppose M is a $Q-$martingale with $EM_t^2<\infty$ for all t. Then M^2 is a submartingale and by the Doob-Meyer decomposition theorem,

$$M^2=\text{local martingale}+A$$

where A is predictable, cadlag, increasing, 0 at time 0. A is in general process theory called the quadratic characteristic for M and is denoted $\langle M\rangle$ (not to be confused with the quadratic variation process $[M]$).

More generally, if M_1,M_2 are two martingales with second moments, the cross characteristic between M_1,M_2 is the process

$$\langle M_1,M_2\rangle:=\tfrac{1}{4}\left(\langle M_1+M_2\rangle-\langle M_1-M_2\rangle\right). \tag{3.50}$$

For us, with Theorem 3.6.1 available, we need only find the quadratic characteristics for the stochastic integrals $M^\circ(S)$. So let Q be a probability on $(\mathcal{M},\mathcal{H})$and let $S,\widetilde{S}$ be predictable flows such that the stochastic integrals $\Lambda^\circ\left(S^2\right),\Lambda^\circ\left(\widetilde{S}^2\right)$ are well defined, and *define*

$$\begin{aligned}\left\langle M^\circ(S)\right\rangle &= \Lambda^\circ\left(S^2\right),\\ \left\langle M^\circ(S),M^\circ(\widetilde{S})\right\rangle &= \Lambda^\circ\left(S\widetilde{S}\right).\end{aligned}$$

(Note that these definitions conform with (3.50)).

Proposition 3.6.2 *Assume that $\overline{\Lambda}^\circ$ is continuous. Assume also that for all n and t,*

$$\sup_{s\le t,y\in E,m\in\mathcal{M}}\left|S^{*y}_{\tau_n\wedge s}(m)\right|<\infty \tag{3.51}$$

where $S^*=S$ or $=\widetilde{S}$. Then

$$M^\circ\left(S^2\right)-\left\langle M^\circ(S)\right\rangle,\quad M^\circ(S)M^\circ\left(\widetilde{S}\right)-\left\langle M^\circ\left(S\right),M^\circ\left(\widetilde{S}\right)\right\rangle \tag{3.52}$$

are local $Q-$martingales $\left(\tau_n\right)$.

If instead of (3.51)

$$E\overline{N}^\circ_t<\infty,\qquad\sup_{s\le t,y\in E,m\in\mathcal{M}}\left|S^{*y}_s(m)\right|<\infty$$

holds for all n and t, then the local martingales in (3.52) are $Q-$martingales.

For the proof, a critical step is to verify by explicit calculation that for all Q,

$$EM^{\circ^2}_{\tau_1\wedge t}(S)=E\Lambda^\circ_{\tau_1\wedge t}\left(S^2\right),\quad EM^\circ_{\tau_1\wedge t}(S)M^\circ_{\tau_1\wedge t}(\widetilde{S})=E\Lambda^\circ_{\tau_1\wedge t}\left(S\widetilde{S}\right).$$

Note that Proposition 3.5.3 corresponds to the special case $S^y_t=1_A\left(y\right)$, $\widetilde{S}^y_t=1_{A'}\left(y\right)$.

The integrabilitry conditions imposed above are not the most general available.

3.7 Compensators and filtrations

We have so far exclusively discussed compensators and compensating measures for canonical CP's and RCM's, i.e. probabilities on $(W,\mathcal{H})$ and $(\mathcal{M},\mathcal{H})$ respectively. Both concepts make perfect sense for CP's and RCM's defined on arbitrary filtered spaces, but they are defined through martingale properties rather than using quantities directly related to the distribution of the process. Working on canonical spaces it is the probability Q that decides the structure of the compensators, hence the terminology '$Q-$compensator' used earlier. On general spaces it is the filtration that matters, hence we shall write '$\mathcal{F}_t-$compensator' below.

Suppose that μ is a RCM defined on $(\Omega, \mathcal{F}, P)$ and let $(\mathcal{F}_t^\mu)_{t\geq 0}$ be the filtration generated by μ, $\mathcal{F}_t^\mu = \sigma\left(N_s(A)\right)_{0\leq s<t, A\in\mathcal{E}}$, where as usual $N_s(A) = \mu\left([0,s]\times A\right)$. In particular the filtration $(\mathcal{F}_t^\mu)$ is right-continuous. Also, let $Q = \mu(P)$ be the distribution of μ and let L° be the compensating measure for Q.

The initial important point to make is that all results about Q and L° carry over to results about μ, the filtration $(\mathcal{F}_t^\mu)$ and the positive random measure $L := L^\circ \circ \mu$. Thus e.g. $\Lambda(A) = (\Lambda_t(A))_{t\geq 0}$ is $\mathcal{F}_t^\mu-$predictable for all $A \in \mathcal{E}$, where $\Lambda_t(A) = L\left([0,t]\times A\right)$, and also

(i) for all $A \in \mathcal{E}$, $M(A) := N(A) - \Lambda(A)$ is a $\mathcal{F}_t^\mu-$local martingale (T_n) (where of course $T_n = \inf\left\{t : \overline{N}_t = n\right\}$);

(ii) up to $P-$indistinguishability, $\Lambda(A)$ is the unique $\mathcal{F}_t^\mu-$predictable process $\widetilde{\Lambda}$, 0 at time 0, such that $M(A)$ is a $\mathcal{F}_t^\mu-$local martingale.

Note also that any $\mathcal{F}_t^\mu-$predictable, right-continuous $\mathcal{F}_t^\mu-$local martingale is constant.

We shall now discuss CP's and RCM's defined on filtered spaces $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. It will always be assumed that *the filtration* $(\mathcal{F}_t)$ *is right-continuous.*

Let N be an adapted counting process defined on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. In particular $\mathcal{F}_t^N \subset \mathcal{F}_t$ for all t, where $\left(\mathcal{F}_t^N\right)$ is the filtration generated by N, $\mathcal{F}_t^N = \sigma\left(N_s\right)_{0\leq s\leq t}$. We shall call Λ the $\mathcal{F}_t-$*compensator for* N if $\Lambda_0 = 0$ $P-$a.s., Λ is increasing, right-continuous, $\mathcal{F}_t-$predictable and satisfies that $M := N - \Lambda$ is a $\mathcal{F}_t-$local martingale. This compensator exists and is unique by the Doob-Meyer decomposition theorem. Note that by the preceding discussion the $\mathcal{F}_t^N-$compensator for N, as just defined for arbitrary filtrations, is $\Lambda^\circ \circ N$, where Λ° is the compensator for the distribution $Q = N(P)$ of N.

Similarly, if μ is an adapted random counting measure on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, (so in particular $\mathcal{F}_t^\mu \subset \mathcal{F}_t$ for all t) the $\mathcal{F}_t-$*compensating measure for* μ is the positive random measure L with $L\left(\{0\}\times E\right) = 0$ $P-$a.s. such that for all $A \in \mathcal{E}$, $\Lambda(A)$, where $\Lambda_t(A) = L\left([0,t]\times A\right)$, defines a right-continuous $\mathcal{F}_t^\mu-$predictable process, necessarily increasing, such that $M(A) := N(A) - \Lambda(A)$ is a $\mathcal{F}_t-$local martingale. Thus, as a special case of this definition, the $\mathcal{F}_t^\mu-$compensating measure is $L^\circ \circ \mu$ with L° the compensating measure for $Q = \mu(P)$.

Note. If L is a positive random measure such that all $\Lambda(A)$ are $\mathcal{F}_t-$predictable we shall say that L *is* $\mathcal{F}_t-$*predictable.*

In general compensators and compensating measures depend of course on the filtration. Furthermore, and this is an important point, while e.g. we know that the $\mathcal{F}_t^\mu-$compensating measure for a RCM μ determines the distribution Q of μ, *it is not in general true that the* $\mathcal{F}_t-$*compensating measure for* μ *determines* Q (and of course, much less will it determine P). This also applies to canonical processes: if Q is a probability on $(\mathcal{M},\mathcal{H})$ with compensating measure L°, the $\mathcal{H}_t-$compensator for the counting process $N^\circ(A)$ is $\Lambda^\circ(A)$, but $\Lambda^\circ(A)$ does not determine the distribution of $N^\circ(A)$ – *marginals of compensating measures do not determine the distribution of the corresponding marginals of the RCM.*

To elaborate further on this point, let Q be a probability on $(\mathcal{M},\mathcal{H})$ with compensating measure L°. The $\mathcal{H}_t-$compensator for the counting process $\overline{N}^\circ$ is $\overline{\Lambda}^\circ$,

$$\overline{\Lambda}_t^\circ = \sum_{n=0}^{\overline{N}_t^\circ} \nu_{\xi_n}^{(n)}\left(\left]\tau_n, \tau_{n+1}\wedge t\right]\right)$$

while the $\mathcal{H}_t^{\overline{N}^\circ}-$compensator for $\overline{N}^\circ$ is

$$\Lambda_t^{\overline{N}^\circ} = \sum_{n=0}^{\overline{N}_t^\circ} \overline{\nu}_{\overline{\xi}_n}^{(n)}\left(\left]\tau_n, \tau_{n+1}\wedge t\right]\right)$$

where $\overline{\xi}_n = (\tau_1,\ldots,\tau_n)$ and $\overline{\nu}_{\overline{\xi}_n}^{(n)}$ is the hazard measure for the conditional $Q-$distribution of τ_{n+1} given $\overline{\xi}_n$, a conditional distribution typically different from $P_{\xi_n}^{(n)}$. And typically it is impossible to obtain $\Lambda^{\overline{N}^\circ}$ (and the distribution of $\overline{N}^\circ$) from knowledge of $\overline{\Lambda}^\circ$ alone – to achieve this complete knowledge of L° may be required.

In one important fairly obvious case one can say that a $\mathcal{F}_t-$compensating measure determines the distribution of a RCM μ.

Proposition 3.7.1 *Let* μ *be an adapted RCM on* $(\Omega,\mathcal{F},\mathcal{F}_t,P)$ *with* $\mathcal{F}_t-$*compensating measure* L. *If* L *is* $\mathcal{F}_t^\mu-$*predictable, then* $L = L^\circ \circ \mu$ *up to* $P-$*indistinguishability, where* L° *is the compensating measure for the distribution* $Q = \mu(P)$ *of* μ.

Proof. By definition, for all A, $M(A) = N(A) - \Lambda(A)$ is a $\mathcal{F}_t-$local martingale. If we can show that in fact $M(A)$ is a $\mathcal{F}_t-$local martingale (T_n), since each T_n is a $\mathcal{F}_t^\mu-$stopping time, we have that $M(A)$ is a $\mathcal{F}_t^\mu-$local martingale and the assertion follows from the discussion at the beginning of this section.

Thus, let $\left(\widetilde{T}_n\right)$ be a reducing sequence for $M(A)$ (so each $\widetilde{T}_n$ is a $\mathcal{F}_t-$stopping time) and note first that since each T_k is a $\mathcal{F}_t-$stopping time, for every n and k, $M^{\widetilde{T}_n \wedge T_k}$ is a $\mathcal{F}_t-$martingale. In particular, since $E\left|M_{\widetilde{T}_n \wedge T_k \wedge t}(A)\right|$ $= E\left|N_{\widetilde{T}_n \wedge T_k \wedge t}(A) - \Lambda_{\widetilde{T}_n \wedge T_k \wedge t}(A)\right| < \infty$ for all t and $0 \leq N_{\widetilde{T}_n \wedge T_k \wedge t}(A) \leq k$ we see that $E\Lambda_{\widetilde{T}_n \wedge T_k \wedge t}(A) < \infty$ for all t and thus, for $s < t$, $F \in \mathcal{F}_s$,

$$\int_F \left(N_{\widetilde{T}_n \wedge T_k \wedge t}(A) - N_{\widetilde{T}_n \wedge T_k \wedge s}(A)\right) dP = \int_F \left(\Lambda_{\widetilde{T}_n \wedge T_k \wedge t}(A) - \Lambda_{\widetilde{T}_n \wedge T_k \wedge s}(A)\right) dP \tag{3.53}$$

as is seen writing down the martingale property for $M^{\widetilde{T}_n \wedge T_k}(A)$ and rearranging the terms. Let now $n \uparrow \infty$ and use monotone convergence to obtain

$$\int_F \left(N_{T_k \wedge t}(A) - N_{T_k \wedge s}(A)\right) dP = \int_F \left(\Lambda_{T_k \wedge t}(A) - \Lambda_{T_k \wedge s}(A)\right) dP. \tag{3.54}$$

Since $N_{T_k \wedge t}(A) \leq k$, this equation for $s = 0$, $F = \Omega$ shows that all $\Lambda_{T_k \wedge t}(A)$ are integrable. It is therefore safe to rearrange the terms in (3.54) and it is then clear that $M^{T_k}(A)$ is a $\mathcal{F}_t-$martingale for all k. ■

Example 3.7.1 *Let N be a counting process on $(\Omega, \mathcal{F}, P)$ and let $U \geq 0$ be a $\mathcal{F}-$measurable random variable. N is a* Cox process *if conditionally on $U = \lambda$, N is homogeneous Poisson (λ). The distribution of N is thus a mixture of Poisson process distributions and by explicit calculation of the Markov kernels $P_{z_n}^{(n)}$ generating the distribution of N, one finds that the $\mathcal{F}_t^N-$compensator is $\Lambda_t^N = \int_0^t \lambda_s^N \, ds$ with $\mathcal{F}_t^N-$predictable intensity process*

$$\lambda_t^N = \frac{\int_{\mathbb{R}_0} \mathbb{P}(d\lambda)\; e^{-\lambda t} \lambda^{N_{t-}+1}}{\int_{\mathbb{R}_0} \mathbb{P}(d\lambda)\; e^{-\lambda t} \lambda^{N_{t-}}}$$

with $\mathbb{P}$ the distribution of the U. This unlovely expression for Λ^N of course serves to describe the unlovely distribution of N. By contrast, defining $\mathcal{F}_t = \sigma\left(U, \mathcal{F}_t^N\right)$ (in particular U is $\mathcal{F}_0-$measurable) the $\mathcal{F}_t-$compensator Λ for N is

$$\Lambda_t = Ut$$

which describes the conditional distribution of N given U only, and does not contain any information whatever about the distribution of U.

In Section 5.5 below we shall discuss some of the basic models in survival analysis. These are partially specified in the sense that only parts of the $\mathcal{F}_t^\mu-$compensating measure for a RCM μ is given. As will be seen, this applies in particular to models for right-censored survival data, where hardly ever the compensators for censorings are described.

Even though there is not a general mechanism for determining compensators using Markov kernels as we did in the canonical case, it is possible to give a prescription for certain filtrations: let μ be a RCM and let $0 = a_0 < a_1 < a_2 < \cdots$ be given timepoints with $a_k \uparrow \infty$ and $a_k = \infty$ allowed. Let $\mathcal{A}_k$ for $k \geq 1$ be given $\sigma-$algebras, increasing with k, and consider the filtration $(\mathcal{F}_t)$ given by

$$\mathcal{F}_t = \sigma\left(\mathcal{A}_k, \mathcal{F}_t^{\mu}\right) \quad \left(t \in [a_k, a_{k+1}[\,,\ k \in \mathbb{N}_0\right).$$

Then the restriction to any interval $[a_k, a_{k+1}[$ of the $\mathcal{F}_t-$compensating measure for μ is found as the restriction of the $\mathcal{F}_t^{\mu}-$compensating measure with respect to the conditional probability $P\left(\cdot\,|\mathcal{F}_{a_k}\right)$.

3.8 Itô's formula for MPP's

Let Q be a probability on $(\mathcal{M}, \mathcal{H})$ and let X be an adapted $\mathbb{R}-$valued process. Itô's formula shows that X can be decomposed as a sum of a predictable process and a local martingale. Uniqueness of the decomposition is achieved when the initial values of the terms are fixed.

By Proposition 3.2.1,

$$X_t = f_{\xi_t}^{\left(\overline{N}_t^{\circ}\right)}(t). \tag{3.55}$$

Recall that X is piecewise continuous if all $f_{z_n}^{(n)}(t)$ are continuous functions of $t \geq t_n$, and define X^c, the *continuous part* of X as

$$X_t^c = X_t - \sum_{0<s\leq t} \Delta X_s.$$

Note that X can only have discontinuities at the timepoints τ_n.

Theorem 3.8.1 (Itô's formula). *Suppose that $\overline{\Lambda}^{\circ}$ is continuous and let X be an adapted $\mathbb{R}-$valued process which is piecewise continuous. Then*

$$X_t = X_0 + U_t + M_t^{\circ}(S) \quad (t \geq 0), \tag{3.56}$$

where $S = (S^y)$ is the predictable flow

$$S_t^y = f_{\mathrm{join}\left(\xi_{t-},(t,y)\right)}^{\left(\overline{N}_{t-}^{\circ}+1\right)} - f_{\xi_{t-}}^{\left(\overline{N}_{t-}^{\circ}\right)}(t),$$

and where U is continuous and predictable. Subject to $U_0 = 0$, $M_0^{\circ}(S) = 0$ a.s., the processes U and the local $Q-$martingale $M^{\circ}(S)$ are unique up to $Q-$indistinguishability.

Proof. By (3.55), X is cadlag and the process ΔX of jumps is well defined. Now identify ΔX and $\Delta M^\circ(S)$: since $\overline{\Lambda}^\circ$ is continuous,

$$\Delta M_t^\circ(S) = S_t^{\eta_{\overline{N}_t^\circ}} \Delta \overline{N}_t^\circ \tag{3.57}$$

while

$$\begin{aligned}
\Delta X_t &= \Delta X_t \Delta \overline{N}_t^\circ \\
&= \left(f_{\xi_t}^{(\overline{N}_t^\circ)}(t) - f_{\xi_{t-}}^{(\overline{N}_{t-}^\circ)}(t) \right) \Delta \overline{N}_t^\circ \\
&= \left(f_{\mathrm{join}\left(\xi_{t-},\left(t, Y_{\overline{N}_t^\circ}\right)\right)}^{(\overline{N}_{t-}^\circ+1)} - f_{\xi_{t-}}^{(\overline{N}_{t-}^\circ)}(t) \right) \Delta \overline{N}_t^\circ .
\end{aligned}$$

Thus (3.56) holds with

$$S_t^y = f_{\mathrm{join}\left(\xi_{t-},(t,y)\right)}^{(\overline{N}_{t-}^\circ+1)} - f_{\xi_{t-}}^{(\overline{N}_{t-}^\circ)}(t),$$

which is a predictable flow, and

$$U_t = X_t - X_0 - M_t^\circ(S) = X_t^c + \Lambda_t^\circ(S) \tag{3.58}$$

which is continuous and adapted, hence predictable.

The uniqueness of the representation (3.56) is immediate from Proposition 3.5.1. ■

Remark 3.8.1 *There is a general decomposition of* any *adapted,* $\mathbb{R}-$*valued process, but in general* U *will not be continuous, perhaps not even cadlag.*

The proof of the theorem yields the decomposition in explicit form, but this is typically too unwieldy to use in practice. Instead, with X *cadlag and piecewise continuous as required in the theorem, one finds* S *by directly identifying* $\Delta X_t = S_t^{\eta_{\overline{N}_t^\circ}} \Delta \overline{N}_t^\circ$, *cf. (3.57) and then using (3.58) to find* U.

Example 3.8.1 *Let* Q *be the canonical Poisson process on* $(W, \mathcal{H})$ *with parameter* $\lambda > 0$. *Fix* $n_0 \in \mathbb{N}$ *and define*

$$X_t = 1_{(N_t^\circ = n_0)}.$$

Since

$$\Delta X_t = \left(1_{\left(N_{t-}^\circ = n_0 - 1\right)} - 1_{\left(N_{t-}^\circ = n_0\right)}\right) \Delta N_t^\circ$$

we obtain

$$X_t = U_t + M_t^\circ(S)$$

with

$$S_t = 1_{\left(N^\circ_{t-}=n_0-1\right)} - 1_{\left(N^\circ_{t-}=n_0\right)}$$

and, since X *is a step process with* $X_0 \equiv 0$ *so that* $X^c \equiv 0$,

$$\begin{aligned} U_t &= \int_0^t \left(1_{\left(N^\circ_{s-}=n_0-1\right)} - 1_{\left(N^\circ_{s-}=n_0\right)}\right) \lambda\, ds \\ &= \int_0^t \left(1_{(N^\circ_s=n_0-1)} - 1_{(N^\circ_s=n_0)}\right) \lambda\, ds. \end{aligned}$$

Since $|S| \leq 1$, *by Theorem 3.6.1 (iii2)* $M^\circ(S)$ *is a* $Q-$*martingale and thus*

$$p_{n_0}(t) := EX_t = EU_t = \int_0^t \left(p_{n_0-1}(s) - p_{n_0}(s)\right) \lambda\, ds, \tag{3.59}$$

a formula valid for $n_0 \geq 1$, $t \geq 0$. *But we know from Example 2.1.2 that* $p_0(t) = Q\left(\tau_1 > t\right) = e^{-\lambda t}$, *and since (3.59) shows that*

$$p'_{n_0}(t) = \lambda\left(p_{n_0-1}(t) - p_{n_0}(t)\right), \quad p_{n_0}(0) = 0$$

for $n_0 \geq 1$, *by induction or otherwise the well known formula*

$$p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

follows.

Example 3.8.2 *We can use Itô's formula to establish that if* Q *is a probability on* $(\mathcal{M}, \mathcal{H})$ *and* $\overline{\Lambda}^\circ$ *is continuous, then* $X = M^{\circ^2}(A) - \Lambda^\circ(A)$ *is a local* $Q-$*martingale* (τ_n) *(Proposition 3.5.3): first, it is easily checked that*

$$\Delta X_t = \left(2M^\circ_{t-}(A) + 1\right) 1_{\left(\eta_{\overline{N}^\circ_t} \in A\right)} \Delta \overline{N}^\circ_t$$

and so (3.56) holds with

$$S^y_t = \left(2M^\circ_{t-}(A) + 1\right) 1_A(y). \tag{3.60}$$

To show that X *is a local martingale we must first show that* $U \equiv 0$, *where, cf. (3.58),*

$$U_t = X^c_t + \int_0^t \int_E S^y_s\, \pi^{\left(\overline{N}^\circ_{s-}\right)}_{\xi_{s-}}(dy)\, \overline{\Lambda}^\circ(ds).$$

But if $\overline{\Lambda}^\circ(ds) = \overline{\lambda}^\circ_s\, ds$, *by differentiation between jumps in the defining expression for* X,

$$\dot{X}^c_t := \tfrac{d}{dt} X^c_t = 2M^\circ_t(A)\left(-\lambda^\circ_t(A)\right) - \lambda^\circ_t(A)$$

where $\Lambda^\circ(A)\,(ds) = \pi_{\xi_s}^{(\overline{N}_s^\circ)}(A)\,ds$. *Using (3.60) it now follows that the continuous process* U *satisfies* $\dot{U}_t = 0$, *hence* $U \equiv 0$.

Using Theorem 3.6.1 (iii1) it is an easy matter to show that X *is a local martingale:*

$$N^\circ_{\tau_n \wedge t}(|S|) \le \int_{]0,\tau_n\wedge t]} \left(2\left(\overline{N}^\circ_{\tau_n\wedge t} + \overline{\Lambda}^\circ_{\tau_n\wedge t}\right) + 1\right) \overline{N}^\circ(ds)$$
$$\le \left(2\left(n + \overline{\Lambda}^\circ_{\tau_n\wedge t}\right)+1\right) n$$

and since $E\overline{\Lambda}^\circ_{\tau_n\wedge t} = E\overline{N}^\circ_{\tau_n\wedge t} \le n < \infty$ *it follows that* $EN^\circ_{\tau_n\wedge t}(|S|) < \infty$. *But the condition in Theorem 3.6.1 (iii2) is too weak to give that* X *is a* $Q-$*martingale when* $E\overline{N}^\circ_t < \infty$ *for all* t, *as was shown in Proposition 3.5.3.*

By similar reasoning one may show that $M^\circ(A)M^\circ(A')$ *is a local martingale when* $A \cap A' = \emptyset$ *(Proposition 3.5.3).*

Example 3.8.3 *Let* Q *be the probability on* $(\mathcal{M}, \mathcal{H})$ *determined by*

$$\overline{P}^{(0)}(t) = e^{-\lambda t}, \qquad \overline{P}^{(n)}_{z_n}(t) = e^{-\lambda(t-t_n)} \quad (n \ge 1, t \ge t_n),$$
$$\pi^{(n)}_{z_n,t}(A) = \kappa(A) \quad (n \ge 0, A \in \mathcal{E})$$

where $\lambda > 0$ *and* κ *is a probability on* $(E, \mathcal{E})$, *cf. Example 3.3.3 above. Thus, under* Q, *the waiting times* $(\tau_n - \tau_{n-1})_{n\ge1}$ *are i.i.d. exponential* λ, *the marks* $(\eta_n)_{n\ge1}$ *are i.i.d with distribution* κ, *and the sequences* (τ_n) *and* (η_n) *are independent.*

We shall show that for any $r \in \mathbb{N}$, *and any* $A_1, \ldots, A_r \in \mathcal{E}$ *mutually disjoint, under* Q *the counting processes* $(N^\circ(A_j))_{1\le j\le r}$ *are independent, homogeneous Poisson processes with intensities* $\lambda_j := \lambda\kappa(A_j)$ *for* $j = 1, \ldots, r$.

We shall show this by showing that for any $s < t$, *the increments* $(N^\circ_t(A_j) - N^\circ_s(A_j))_{1\le j\le r}$ *are independent of* $\mathcal{H}_s$ *and also mutually independent with* $N^\circ_t(A_j) - N^\circ_s(A_j)$ *following a Poisson distribution with parameter* $\lambda_j(t-s)$. *And this in turn will be shown by showing that for all* $(u_1, \ldots, u_r) \in \mathbb{R}^r$,

$$M_t := \exp\left(i\sum_{j=1}^r u_j N^\circ_t(A_j) - t\sum_{j=1}^r \lambda_j\left(e^{iu_j}-1\right)\right)$$

is a $\mathbb{C}-$*valued* $Q-$*martingale: the martingale property* $E(M_t\,|\mathcal{H}_s) = M_s$ *is easily rewritten as*

$$E\left(\exp\left(i\sum_{j=1}^r \left(N^\circ_t(A_j) - N^\circ_s(A_j)\right)\right)|\mathcal{H}_s\right)$$
$$= \exp\left((t-s)\sum_{j=1}^r \lambda_j\left(e^{iu_j}-1\right)\right),$$

which is precisely to say, (when $(u_1, \ldots, u_r)$ *varies) that the joint characteristic function of* $(N_t^\circ(A_j) - N_s^\circ(A_j))_{1 \le j \le r}$ *given* $\mathcal{H}_s$ *is that of* r *inedependent Poisson random variables with parameters* $(\lambda_j (t-s))_{1 \le j \le r}$ *.*

We use Itô's formula to show that M *is a martingale. Because*

$$\Delta M_t = M_{t-} \left(\sum_{j=1}^{r} 1_{A_j} \left(\eta_{\overline{N}_t^\circ} \right) \left(e^{iu_j} - 1 \right) \right) \Delta \overline{N}_t^\circ,$$

we have the representation

$$M_t = 1 + U_t + M_t^\circ(S)$$

with U *continuous,* $U_0 = 0$ *and*

$$S_t^y = M_{t-} \sum_{j=1}^{r} 1_{A_j}(y) \left(e^{iu_j} - 1 \right).$$

We now show that $U \equiv 0$ *by verifying that between jumps* $\dot{U}_t := \frac{d}{dt} U_t = 0$*. But clearly, since* $N^\circ(S)$ *is constant between jumps,*

$$\dot{U}_t = \dot{M}_t + \dot{\Lambda}_t^\circ(S).$$

By computation, if $t \in \left]\tau_{n-1}, \tau_n\right[$ *for some* n*, writing* $c = \sum_{j=1}^r \lambda_j \left(e^{iu_j} - 1 \right)$*,*

$$\dot{M}_t = -cM_t$$

while

$$\begin{aligned} \dot{\Lambda}_t^\circ(S) &= \frac{d}{dt} \int_0^t \int_E S_s^y \, \kappa(dy) \, \lambda \, ds \\ &= \lambda \int_E S_t^y \, \kappa(dy) \\ &= \lambda M_t \sum_{j=1}^{r} \kappa(A_j) \left(e^{iu_j} - 1 \right) \\ &= cM_t \end{aligned}$$

(using that between jumps, $M_{t-} = M_t$*).*

Thus $U \equiv 0$ *and* $M_t = 1 + M_t^\circ(S)$*. It remains to verify that* M *is a* $Q-$*martingale, and this follows from Theorem 3.6.1 (iii2) if we show that* $EN_t^\circ(|S|) < \infty$ *for all* t*. But*

$$|S_t^y| = e^{-ct},$$

so $EN_t^\circ(|S|) = e^{-ct} E\overline{N}_t^\circ = e^{-ct} \lambda t < \infty$*.*

In the version of Itô's formula given above, Theorem 3.8.1, it was assumed in particular that the total compensator $\overline{\Lambda}^{\circ}$ should be continuous. We shall by two examples show how a martingale decomposition may be obtained when this assumption is not fulfilled.

Example 3.8.4 *Let Q be a probability on $(W, \mathcal{H})$ with compensator Λ° that need not be continuous. If Λ° is continuous, we know from Proposition 3.5.3 that $M^{\circ^2} - \Lambda^\circ$ is a local $Q-$martingale, but it was also noted that if Λ° is not continuous, this is no longer true. We are thus, for general Λ°, looking for $\widetilde{\Lambda}$ right-continuous and predictable, 0 at time 0, such that $M^{\circ^2} - \widetilde{\Lambda}$ is a local martingale, i.e. we also need S predictable such that*

$$M^{\circ^2} - \widetilde{\Lambda} = N^\circ(S) - \Lambda^\circ(S). \tag{3.61}$$

In particular $X := M^{\circ^2} - N^\circ(S)$ must be predictable, and this fact is used to identify S. (It is no longer as in the proof of Theorem 3.8.1 and the preceding examples, a matter of simply identifying the jumps of M°^2} occurring when N° jumps, with those of $N^\circ(S)$. Note that the two processes N° and Λ° may share discontinuities, but that it is also possible that one of them is continuous, the other discontinuous at a certain point).

If $\Delta N_t^\circ = 0$ we have

$$X_t = \left(M_{t-}^\circ - \Delta\Lambda_t^\circ\right)^2 - N_{t-}^\circ(S)$$

while if $\Delta N_t^\circ = 1$,

$$X_t = \left(M_{t-}^\circ + 1 - \Delta\Lambda_t^\circ\right)^2 - N_{t-}^\circ(S) - S_t.$$

If X is to be predictable the two expressions must be the same, i.e. we must have

$$S_t = 1 + 2\left(M_{t-}^\circ - \Delta\Lambda_t^\circ\right).$$

With S determined, $\widetilde{\Lambda}$ is of course found from (3.61), but a simpler expression is available by showing that $\widetilde{\Lambda}$ is differentiable with respect to Λ° and finding $D_{\Lambda^\circ}\widetilde{\Lambda}$. (For a discussion of differentiability and the results used below, see Appendix A).

We have $D_{\Lambda^\circ}\Lambda^\circ(S) = S$. To find $D_{\Lambda^\circ}X$, fix t and for a given K find $k \in \mathbb{N}$ such that $t_K = \frac{k-1}{2^K} < t \leq \frac{k}{2^K} = \widetilde{t}_K$. The task is to compute (for $\Lambda^\circ-$a.a. t),

$$\lim_{K\to\infty} \frac{X\left(\widetilde{t}_K\right) - X\left(t_K\right)}{\Lambda^\circ\left(\widetilde{t}_K\right) - \Lambda^\circ\left(t_K\right)}$$

and there are two cases, (i) $\Delta N_t^\circ = 0$*, (ii)* $\Delta N_t^\circ = 1$ *and* $\Delta \Lambda_t^\circ > 0$*. (Because the limit is only wanted for* $\Lambda^\circ -$*a.a.* t *we may ignore the case* $\Delta N_t^\circ = 1$ *and* $\Delta \Lambda_t^\circ = 0$*). In case (i) one finds that the limit is*

$$\lim_{K\to\infty} \frac{\left(M^\circ_{\widetilde{t}_K} + M^\circ_{t_K}\right)\left(M^\circ_{\widetilde{t}_K} - M^\circ_{t_K}\right)}{\Lambda^\circ\left(\widetilde{t}_K\right) - \Lambda^\circ\left(t_K\right)} = -\left(2M^\circ_{t-} - \Delta\Lambda^\circ_t\right)$$

since for K *sufficiently large,* $M^\circ_{\widetilde{t}_K} = M^\circ_{t_K} - \left(\Lambda^\circ\left(\widetilde{t}_K\right) - \Lambda^\circ\left(t_K\right)\right)$*. A similar argument in case (ii) results in the same limit and verifying the other requirements for differentiability, one ends up with*

$$D_{\Lambda^\circ} X_t = -\left(2M^\circ_{t-} - \Delta\Lambda^\circ_t\right).$$

Thus

$$\begin{aligned} D_{\Lambda^\circ}\widetilde{\Lambda}_t &= D_{\Lambda^\circ} X_t + S_t \\ &= 1 - \Delta\Lambda^\circ_t \end{aligned}$$

and, using the results from Appendix A,

$$\widetilde{\Lambda}_t = \int_{]0,t]} D_{\Lambda^\circ}\widetilde{\Lambda}_s \, d\Lambda^\circ_s = \Lambda^\circ_t - \sum_{0<s\le t} \left(\Delta\Lambda^\circ_s\right)^2.$$

Example 3.8.5 *Let* Q *be a probability on* $(\mathcal{M}, \mathcal{H})$ *with compensating measure* L°*. Let* $A, A' \in \mathcal{E}$ *and look for a decomposition of* $M^\circ(A)M^\circ(A')$*, i.e. we want* $\widetilde{\Lambda}$ *right-continuous and predictable,* 0 *at time* 0*, and a predictable flow* (S^y) *such that*

$$M^\circ_t(A)M^\circ_t(A') = \widetilde{\Lambda}_t + N^\circ(S) - \Lambda^\circ(S).$$

From Proposition 3.5.3 we know that if $\overline{\Lambda}^\circ$ *is continuous,* $\widetilde{\Lambda} = \Lambda^\circ(A)$ *if* $A = A'$ *and* $\widetilde{\Lambda} \equiv 0$ *if* $A \cap A' = \emptyset$*. Here we do not assume that* $\overline{\Lambda}^\circ$ *is continuous.*

We find S *by using that* $X := M^\circ(A)M^\circ(A') - N^\circ(S)$ *is predictable. If* $\Delta\overline{N}^\circ_t = 0$*,*

$$X_t = \left(M^\circ_{t-}(A) - \Delta\Lambda^\circ_t(A)\right)\left(M^\circ_{t-}(A') - \Delta\Lambda^\circ_t(A')\right) - N^\circ_{t-}(S),$$

while if $\Delta\overline{N}^\circ_t = 1$*,*

$$\begin{aligned} X_t = {} & \left(M^\circ_{t-}(A) + 1_A(\eta_t) - \Delta\Lambda^\circ_t(A)\right)\left(M^\circ_{t-}(A') + 1_{A'}(\eta_t) - \Delta\Lambda^\circ_t(A')\right) \\ & - N^\circ_{t-}(S) - S^{\eta_t}_t \end{aligned}$$

where we write $\eta_t = \eta_{\overline{N}_t^\circ}$. *Predictability of* X *forces the two expressions to be identical, hence*

$$S_t^y = 1_{A\cap A'}(y) + 1_A(y)\left(M_{t-}^\circ(A') - \Delta\Lambda_t^\circ(A')\right) + 1_{A'}(y)\left(M_{t-}^\circ(A) - \Delta\Lambda_t^\circ(A)\right). \tag{3.62}$$

As in the previous Example 3.8.4, we identify $\widetilde{\Lambda}$ *through its derivative* $D_{\overline{\Lambda}^\circ}\widetilde{\Lambda} = D_{\overline{\Lambda}^\circ}X + D_{\overline{\Lambda}^\circ}\Lambda^\circ(S)$. *Here*

$$D_{\overline{\Lambda}^\circ}\Lambda_t^\circ(S) = \int_E S_t^y\,\pi_t\,(dy)\,, \tag{3.63}$$

where π_t *is short for* $\pi_{\xi_{t-},t}^{(\overline{N}_{t-}^\circ)}$, *and by computations along the same lines as those in Example 3.8.4, treating the cases* $\Delta\overline{N}_t^\circ = 0$ *and* $\Delta\overline{N}_t^\circ = 1$, $\Delta\overline{\Lambda}_t^\circ > 0$ *separately one finds*

$$D_{\overline{\Lambda}^\circ}X_t = -M_{t-}^\circ(A)\pi_t(A') - M_{t-}^\circ(A')\pi_t(A) + \Delta\overline{\Lambda}_t^\circ\pi_t(A)\pi_t(A').$$

Thus, recalling (3.62), (3.63)

$$D_{\overline{\Lambda}^\circ}\widetilde{\Lambda}_t = D_{\overline{\Lambda}^\circ}X_t + D_{\overline{\Lambda}^\circ}L_t^\circ(S) = \pi_t\,(A\cap A') - \Delta\overline{\Lambda}_t^\circ\pi_t(A)\pi_t(A')$$

so finally,

$$\widetilde{\Lambda}_t = \int_{]0,t]} D_{\overline{\Lambda}^\circ}\widetilde{\Lambda}_s\,d\overline{\Lambda}_s^\circ = \Lambda_t^\circ\,(A\cap A') - \sum_{0<s\le t}\Delta\Lambda_s^\circ(A)\Delta\Lambda_s^\circ(A').$$

Chapter 4

Likelihood processes

In this short chapter we derive the likelihood function corresponding to observing a CP or RCM completely on a finite time interval $[0,t]$. In statistical terms, one would suppose given a family of distributions for the point process, choose a reference measure from the family and define the likelihood function as the Radon-Nikodym derivative between the distribution of the process observed on $[0,t]$ under an arbitrary measure from the family and under the reference measure. The essence is therefore to be able to find the relevant Radon-Nikodym derivatives between two different distributions of the process on $[0,t]$.

Let $Q, \widetilde{Q}$ be two probability measures on $(W, \mathcal{H})$ or $(\mathcal{M}, \mathcal{H})$, and let for $t \geq 0$, $Q_t, \widetilde{Q}_t$ denote the restrictions of $Q, \widetilde{Q}$ to $\mathcal{H}_t$.

Definition 4.0.1 $\widetilde{Q}$ *is* locally absolutely continuous *with respect to* Q *if* $\widetilde{Q}_t \ll Q_t$ *for all* $t \in \mathbb{R}_0$.

If $\widetilde{Q}$ is locally absolutely continuous with respect to Q, we write $\widetilde{Q}_t \ll_{\text{loc}} Q_t$ and define the *likelihood process* $\mathfrak{L} = (\mathfrak{L}_t)_{t \geq 0}$ by

$$\mathfrak{L}_t := \frac{d\widetilde{Q}_t}{dQ_t}. \tag{4.1}$$

Since $\mathcal{H}_0$ is the trivial $\sigma-$algebra, $\mathfrak{L}_0 \equiv 1$. Otherwise each $\mathfrak{L}_t$ is $\mathcal{H}_t-$measurable, ≥ 0 $Q-$a.s. and

$$\widetilde{Q}(H) = \widetilde{Q}_t(H) = \int_H \mathfrak{L}_t \, dQ_t = \int_H \mathfrak{L}_t \, dQ \quad (t \in \mathbb{R}_0, H \in \mathcal{H}_t). \tag{4.2}$$

With E_Q denoting expectation with respect to Q, it follows in particular that

$$E_Q \mathfrak{L}_t = 1 \quad (t \in \mathbb{R}_0).$$

If $s < t$ and $H \in \mathcal{H}_s \subset \mathcal{H}_t$, it follows from (4.2) that

$$\int_H \mathfrak{L}_s \, dQ = \int_H \mathfrak{L}_t \, dQ,$$

in other words, $\mathfrak{L}$ *is a* $Q-$*martingale* which, since the filtration $(\mathcal{H}_t)$ is right-continuous, has a cadlag version. In the sequel we shall always assume $\mathfrak{L}$ to be cadlag.

Since $\mathfrak{L} \geq 0$, $\mathfrak{L}_\infty := \lim_{t\to\infty} \mathfrak{L}_t$ exists $Q-$a.s. and by Fatou's lemma $E_Q \mathfrak{L}_\infty \leq 1$. The reader is reminded that the Lebesgue decomposition of $\widetilde{Q}$ with respect to Q (on all of $\mathcal{H}$) is

$$\widetilde{Q} = \mathfrak{L}_\infty \cdot Q + \zeta$$

where $\mathfrak{L}_\infty \cdot Q$ is the bounded measure $(\mathfrak{L}_\infty \cdot Q)(H) = \int_H \mathfrak{L}_\infty \, dQ$ on $\mathcal{H}$, and where ζ is a bounded positive measure on $\mathcal{H}$ with $Q \perp \zeta$ (i.e. there is a set $H_0 \in \mathcal{H}$ such that $Q(H_0) = 1$, $\zeta(H_0) = 0$). In particular, $\widetilde{Q} \perp Q$ iff $Q(\mathfrak{L}_\infty = 0) = 1$, and $\widetilde{Q} \ll Q$ iff $E_Q \mathfrak{L}_\infty = 1$, this latter condition also being equivalent to the condition that $\mathfrak{L}$ be uniformly integrable with respect to Q.

Before stating the main result, we need one more concept: if $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are probabilities on $\overline{\mathbb{R}}_+$, we write $\widetilde{\mathbb{P}} \ll_{\text{loc}} \mathbb{P}$ if

(i) $\widetilde{\mathbb{P}}_{\mathbb{R}_+} \ll \mathbb{P}_{\mathbb{R}_+}$, the subscript denoting restriction to $\mathbb{R}_+$ (from $\overline{\mathbb{R}}_+$);

(ii) $\mathbb{P}(]t,\infty]) > 0$ whenever $\widetilde{\mathbb{P}}(]t,\infty]) > 0$.

Note that if $t^\dagger, \tilde{t}^\dagger$ are the termination points for $\mathbb{P}, \widetilde{\mathbb{P}}$ (see p.22) the last condition is equivalent to the condition $\tilde{t}^\dagger \leq t^\dagger$. Note also that $\widetilde{\mathbb{P}} \ll_{\text{loc}} \mathbb{P}$ if $\widetilde{\mathbb{P}} \ll \mathbb{P}$ (on all of $\overline{\mathbb{R}}_+$) while it is possible to have $\widetilde{\mathbb{P}}_{\mathbb{R}_+} \ll \mathbb{P}_{\mathbb{R}_+}$ without $\widetilde{\mathbb{P}} \ll_{\text{loc}} \mathbb{P}$: if $\widetilde{\mathbb{P}} = \varepsilon_\infty$, $\widetilde{\mathbb{P}}_{\mathbb{R}_+} \ll \mathbb{P}_{\mathbb{R}_+}$ for any $\mathbb{P}$ since $\widetilde{\mathbb{P}}_{\mathbb{R}_+}$ is the null measure, but $\widetilde{\mathbb{P}} \ll_{\text{loc}} \mathbb{P}$ iff $\mathbb{P}(]t,\infty]) > 0$ for all $t \in \mathbb{R}_0$.

If $\widetilde{\mathbb{P}} \ll_{\text{loc}} \mathbb{P}$, we write

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}(t) = \frac{d\widetilde{\mathbb{P}}_{\mathbb{R}_+}}{d\mathbb{P}_{\mathbb{R}_+}}(t) \quad (t \in \mathbb{R}_+).$$

We shall as usual denote by $P^{(n)}$, $\pi^{(n)}$ the Markov kernels generating Q and of course write $\widetilde{P}^{(n)}$, $\widetilde{\pi}^{(n)}$ for those generating $\widetilde{Q}$. $\overline{P}^{(n)}_{z_n}$ and $\overline{\widetilde{P}}^{(n)}_{z_n}$ are the survivor functions for $P^{(n)}_{z_n}$ and $\widetilde{P}^{(n)}_{z_n}$.

Theorem 4.0.2 *(a) Let $Q, \widetilde{Q}$ be probabilities on $(W, \mathcal{H})$. In order that $\widetilde{Q} \ll_{\text{loc}} Q$ it is sufficient that $\widetilde{P}^{(0)} \ll_{\text{loc}} P^{(0)}$ and that for every $n \in \mathbb{N}$*

there exists an exceptional set $B_n^{null} \in \mathcal{B}_+^n$ *with* $Q\left(\xi_n \in B_n^{null}\right) = 0$ *such that* $\widetilde{P}_{z_n}^{(n)} \ll_{\text{loc}} P_{z_n}^{(n)}$ *for all* $z_n = (t_1, \ldots, t_n) \notin B_n^{null}$ *with* $0 < t_1 < \cdots < t_n$.

If this condition for $\widetilde{Q} \ll_{\text{loc}} Q$ *is satisfied, the cadlag* $Q-$*martingale* $\mathfrak{L}$ *is up to* $Q-$*indistinguishability given by*

$$\mathfrak{L}_t = \left(\prod_{n=1}^{N_t^\circ} \frac{d\widetilde{P}_{\xi_{n-1}}^{(n-1)}}{dP_{\xi_{n-1}}^{(n-1)}}(\tau_n)\right) \frac{\overline{\widetilde{P}}_{\xi_t}^{(N_t^\circ)}(t)}{\overline{P}_{\xi_t}^{(N_t^\circ)}(t)} \qquad (t \in \mathbb{R}_0).$$

(b) Let $Q, \widetilde{Q}$ *be probabilities on* $(\mathcal{M}, \mathcal{H})$. *In order that* $\widetilde{Q} \ll_{\text{loc}} Q$ *it is sufficient that* $\widetilde{P}^{(0)} \ll_{\text{loc}} P^{(0)}$ *and (i) for every* $n \in \mathbb{N}$ *there exists an exceptional set* $C_n^{null} \in \mathcal{B}_+^n \otimes \mathcal{E}^n$ *with* $Q\left(\xi_n \in C_n^{null}\right) = 0$ *such that* $\widetilde{P}_{z_n}^{(n)} \ll_{\text{loc}} P_{z_n}^{(n)}$ *for all* $z_n = (t_1, \ldots, t_n; y_1, \ldots, y_n) \notin C_n^{null}$ *with* $0 < t_1 < \cdots < t_n$, *and (ii) for every* $n \in \mathbb{N}_0$ *there exists an exceptional set* $D_n^{null} \in \mathcal{B}_+^{n+1} \otimes \mathcal{E}^n$ *with* $Q\left((\xi_n, \tau_{n+1}) \in D_n^{null}\right) = 0$ *such that* $\widetilde{\pi}_{z_n,t}^{(n)} \ll \pi_{z_n,t}^{(n)}$ *for all* $(z_n, t) = (t_1, \ldots, t_n, t; y_1, \ldots, y_n) \notin D_n^{null}$ *with* $0 < t_1 < \cdots < t_n < t$.

If this condition for $\widetilde{Q} \ll_{\text{loc}} Q$ *is satisfied, the cadlag* $Q-$*martingale* $\mathfrak{L}$ *is up to* $Q-$*indistinguishability given by*

$$\mathfrak{L}_t = \left(\prod_{n=1}^{\overline{N}_t^\circ} \frac{d\widetilde{P}_{\xi_{n-1}}^{(n-1)}}{dP_{\xi_{n-1}}^{(n-1)}}(\tau_n)\, \frac{d\widetilde{\pi}_{\xi_{n-1},\tau_n}^{(n-1)}}{d\pi_{\xi_{n-1},\tau_n}^{(n-1)}}(\eta_n)\right) \frac{\overline{\widetilde{P}}_{\xi_t}^{(\overline{N}_t^\circ)}(t)}{\overline{P}_{\xi_t}^{(\overline{N}_t^\circ)}(t)} \qquad (t \in \mathbb{R}_0).$$

Proof. We just give the proof of (b). For $n \in \mathbb{N}$, let $\widetilde{R}_n, R_n$ denote the distribution of ξ_n under $\widetilde{Q}$ and Q respectively, restricted to $(\tau_n < \infty)$ so that e.g. $\widetilde{R}_n$ is a subprobability on $\mathbb{R}_+^n$ with

$$\widetilde{R}_n(C_n) = \widetilde{Q}\left(\xi_n \in C_n, \tau_n < \infty\right) \quad \left(C_n \in \mathcal{B}_+^n \otimes \mathcal{E}^n\right).$$

Because $\left(\widetilde{P}_{z_n}^{(n)}\right)_{\mathbb{R}_+} \ll \left(P_{z_n}^{(n)}\right)_{\mathbb{R}_+}$ and $\widetilde{\pi}_{z_n,t}^{(n)} \ll \pi_{z_n,t}^{(n)}$ (for almost all z_n and (z_n, t)) it follows that $\widetilde{R}_n \ll R_n$ and

$$\frac{d\widetilde{R}_n}{dR_n}(z_n) = \prod_{k=1}^{n} \left(\frac{d\widetilde{P}_{z_{k-1}}^{(k-1)}}{dP_{k-1}^{(k-1)}}(t_k) \frac{d\widetilde{\pi}_{z_{k-1},t_k}^{(k-1)}}{d\pi_{z_{k-1},t_k}^{(k-1)}}(y_k)\right) \tag{4.3}$$

for R_n-a.a. $z_n \in \mathbb{R}_+^n \times E^n$.

We want to show that for $t \in \mathbb{R}_+$, $H \in \mathcal{H}_t$,

$$\widetilde{Q}(H) = \int_H \mathfrak{L}_t \, dQ$$

and recalling the representation of sets in $\mathcal{H}_t$, Proposition 3.2.1 (bi), and writing $H = \bigcup_{n=0}^{\infty} H \cap \left(\overline{N}_t^{\circ} = n\right)$, this follows if we show

$$\widetilde{Q}\left(\xi_n \in C_n, \tau_{n+1} > t\right) = \int_{(\xi_n \in C_n, \tau_{n+1} > t)} \mathfrak{L}_t \, dQ$$

where $H \cap \left(\overline{N}_t^{\circ} = n\right) = \left(\xi_n \in C_n, \tau_{n+1} > t\right)$. But

$$\begin{aligned}
\widetilde{Q}\left(\xi_n \in C_n, \tau_{n+1} > t\right) &= \int_{(\xi_n \in C_n)} \overline{\widetilde{P}}^{(n)}_{\xi_n}(t) \, d\widetilde{Q} \\
&= \int_{C_n} \overline{\widetilde{P}}^{(n)}_{z_n}(t) \, \widetilde{R}_n(dz_n) \\
&= \int_{C_n} \overline{\widetilde{P}}^{(n)}_{z_n}(t) \frac{d\widetilde{R}_n}{dR_n}(z_n) \, R_n(dz_n) \\
&= \int_{(\xi_n \in C_n)} \overline{\widetilde{P}}^{(n)}_{\xi_n}(t) \frac{d\widetilde{R}_n}{dR_n}\left(\xi_n\right) \, dQ.
\end{aligned}$$

Define $H_n = \left(\overline{\widetilde{P}}^{(n)}_{\xi_n}(t) > 0\right)$. Then the last integral above is the same as the integral over the set $(\xi_n \in C_n) \cap H_n$ and since by the assumption $\widetilde{P}^{(n)}_{\xi_n} \ll_{\text{loc}} P^{(n)}_{\xi_n}$ we have $\overline{P}^{(n)}_{\xi_n}(t) > 0$ on H_n, the domain of integration may be replaced by $\left(\xi_n \in C_n, \overline{P}^{(n)}_{\xi_n}(t) > 0\right) \cap H_n$, and the integral may be written

$$\int_{\left(\xi_n \in C_n, \tau_{n+1} > t, \overline{P}^{(n)}_{\xi_n}(t) > 0\right) \cap H_n} \frac{\overline{\widetilde{P}}^{(n)}_{\xi_n}(t)}{\overline{P}^{(n)}_{\xi_n}(t)} \frac{d\widetilde{R}_n}{dR_n}\left(\xi_n\right) \, dQ$$

as is seen by conditioning on ξ_n in this last integral. But by the definition of H_n and the appearance of the factor $\overline{\widetilde{P}}^{(n)}_{\xi_n}(t)$ in the integrand, this is the same as the integral over $\left(\xi_n \in C_n, \tau_{n+1} > t, \overline{P}^{(n)}_{\xi_n}(t) > 0\right)$, and finally, because

$$Q\left(\tau_{n+1} > t, \overline{P}^{(n)}_{\xi_n}(t) = 0\right) = E_Q 1_{\left(\overline{P}^{(n)}_{\xi_n}(t) = 0\right)} \overline{P}^{(n)}_{\xi_n}(t) = 0,$$

we may as well integrate over $(\xi_n \in C_n, \tau_{n+1} > t) = H \cap \left(\overline{N}_t^\circ = n\right)$ and we have arrived at the identity

$$\widetilde{Q}\left(H \cap \left(\overline{N}_t^\circ = n\right)\right) = \int_{H\cap\left(\overline{N}_t^\circ = n\right)} \frac{\overline{\widetilde{P}}_{\xi_n}^{(n)}(t)}{\overline{P}_{\xi_n}^{(n)}(t)} \frac{d\widetilde{R}_n}{dR_n}(\xi_n)\, dQ.$$

Using (4.3), the assertion of the theorem follows immediately. ∎

Remark 4.0.2 *It may be shown that the sufficient conditions for $\widetilde{Q} \ll_{\mathrm{loc}} Q$ as given in (a), (b) are in fact also necessary. The perhaps most peculiar condition, viz. that $\widetilde{P}_{z_n}^{(n)} \ll_{\mathrm{loc}} P_{z_n}^{(n)}$ rather than just $\left(\widetilde{P}_{z_n}^{(n)}\right)_{\mathbb{R}_+} \ll \left(P_{z_n}^{(n)}\right)_{\mathbb{R}_+}$ was certainly used in the proof, and that it is necessary may be seen from the following CP example: suppose $\widetilde{Q}$ is the distribution of the dead process while under Q, τ_1 is bounded by 1 say. Then $\widetilde{P}_{\mathbb{R}_+}^{(0)} \ll P_{\mathbb{R}_+}^{(0)}$ trivially since $\widetilde{P}_{\mathbb{R}_+}^{(0)}$ is the null measure, $\widetilde{P}^{(0)} \ll_{\mathrm{loc}} P^{(0)}$ does not hold and for $t \geq 1$ it does not hold either that $\widetilde{Q}_t \ll Q_t$ since $Q_t(\tau_1 \leq t) = 1$, $\widetilde{Q}_t(\tau_1 \leq t) = 0$.*

It is in general possible to express the formula for $\mathfrak{L}_t$ in terms of compensators or compensating measures. We shall see how this may be done in some special cases.

Consider first the CP case and assume that the compensators $\widetilde{\Lambda}^\circ, \Lambda^\circ$ under $\widetilde{Q}, Q$ have predictable intensities,

$$\widetilde{\Lambda}_t^\circ = \int_0^t \widetilde{\lambda}_s^\circ\, ds, \quad \Lambda_t^\circ = \int_0^t \lambda_s^\circ\, ds$$

with

$$\widetilde{\lambda}_s^\circ = \widetilde{u}_{\xi_{s-}}^{\left(N_{s-}^\circ\right)}(s), \quad \lambda_s^\circ = u_{\xi_{s-}}^{\left(N_{s-}^\circ\right)}(s)$$

where $\widetilde{u}_{z_n}^{(n)}$ $(u_{z_n}^{(n)})$ is the hazard function for $\widetilde{P}_{z_n}^{(n)}$ $(P_{z_n}^{(n)})$, cf. Proposition 3.4.1 (a). If $\widetilde{Q} \ll_{\mathrm{loc}} Q$ we have

$$\frac{d\widetilde{P}_{\xi_{k-1}}^{(k-1)}}{dP_{\xi_{k-1}}^{(k-1)}}(\tau_k) = \frac{\widetilde{u}_{\xi_{k-1}}^{(k-1)}(\tau_k) \exp\left(-\int_{\tau_{k-1}}^{\tau_k} \widetilde{u}_{\xi_{k-1}}^{(k-1)}(s)\, ds\right)}{u_{\xi_{k-1}}^{(k-1)}(\tau_k) \exp\left(-\int_{\tau_{k-1}}^{\tau_k} u_{\xi_{k-1}}^{(k-1)}(s)\, ds\right)}, \tag{4.4}$$

$$\frac{\overline{\widetilde{P}}_{\xi_t}^{(N_t^\circ)}(t)}{\overline{P}_{\xi_t}^{(N_t^\circ)}(t)} = \exp\left(-\int_{\tau_{N_t^\circ}}^t \left(\widetilde{u}_{\xi_{s-}}^{\left(N_{s-}^\circ\right)}(s) - u_{\xi_{s-}}^{\left(N_{s-}^\circ\right)}(s)\right) ds\right) \tag{4.5}$$

and it follows that

$$\mathfrak{L}_t = \exp\left(-\widetilde{\Lambda}_t^\circ + \Lambda_t^\circ\right) \prod_{n=1}^{N_t^\circ} \frac{\widetilde{\lambda}_{\tau_n}^\circ}{\lambda_{\tau_n}^\circ}. \tag{4.6}$$

Similarly, if $\widetilde{Q}, Q$ have compensating measures $\widetilde{L}^\circ, L^\circ$ with $\kappa-$intensities (Proposition 3.4.1 (bii), the same κ for $\widetilde{Q}$ and Q),

$$\widetilde{\lambda}_t^{\circ y} = \widetilde{u}_{\xi_{t-}}^{\left(\overline{N}_{t-}^\circ\right)}(t)\, \widetilde{p}_{\xi_{t-},t}^{\left(\overline{N}_{t-}^\circ\right)}(y), \quad \widetilde{\lambda}_t^{\circ y} = u_{\xi_{t-}}^{\left(\overline{N}_{t-}^\circ\right)}(t)\, p_{\xi_{t-},t}^{\left(\overline{N}_{t-}^\circ\right)}(y)$$

we still have analogues of (4.4), (4.5) and in addition

$$\frac{d\widetilde{\pi}_{\xi_{k-1},\tau_k}^{(k-1)}}{d\pi_{\xi_{k-1},\tau_k}^{(k-1)}}(\eta_k) = \frac{\widetilde{p}_{\xi_{k-1},\tau_k}^{(k-1)}(\eta_k)}{p_{\xi_{k-1},\tau_k}^{(k-1)}(\eta_k)}$$

and it follows that

$$\mathfrak{L}_t = \exp\left(-\overline{\widetilde{\Lambda}}_t^\circ + \overline{\Lambda}_t^\circ\right) \prod_{n=1}^{\overline{N}_t^\circ} \frac{\widetilde{\lambda}_{\tau_n}^{\circ \eta_n}}{\lambda_{\tau_n}^{\circ \eta_n}}. \tag{4.7}$$

The derivation of (4.6) and (4.7) shows that it is the *predictable* intensities that must enter into the expressions: it is the preceding intensity that fires the next jump. Using the right-continuous intensities of Proposition 3.4.2 would drastically change the expressions and yield non-sensical results.

We quote a particular case of (4.7), important for statistical applications. Here $\widetilde{Q}$ is the important measure, while the particular Q displayed serves only as a convenient reference. Recall Examples 3.3.3 and 3.8.3.

Corollary 4.0.3 *On $(\mathcal{M}, \mathcal{H})$ let Q be the Poisson process with compensating measure $L^\circ = \ell \otimes \kappa$, where κ is a bounded positive measure on $(E, \mathcal{E})$ with $\overline{\lambda} := \kappa(E) > 0$. If $\widetilde{Q}$ has predictable $\kappa-$intensity $\left(\widetilde{\lambda}_t^{\circ y}\right)$, then $\widetilde{Q} \ll_{\mathrm{loc}} Q$ and*

$$\mathfrak{L}_t = \exp\left(-\overline{\widetilde{\Lambda}}_t^\circ + \overline{\lambda} t\right) \prod_{n=1}^{\overline{N}_t^\circ} \widetilde{\lambda}_{\tau_n}^{\circ \eta_n}.$$

Proof. With Q Poisson, all $P_{z_n}^{(n)}$ have Lebesgue densities that are > 0 on $]t_n, \infty[$ and have termination point ∞. By assumption also all $\widetilde{P}_{z_n}^{(n)}$ have Lebesgue densities and hence $\widetilde{P}_{z_n}^{(n)} \ll_{\mathrm{loc}} P_{z_n}^{(n)}$. Because Q has $\kappa-$intensity process $\lambda_t^{\circ y} \equiv 1$ for all t, y, the assumptions about $\widetilde{Q}$ imply also that $\widetilde{\pi}_{z_n,t}^{(n)} \ll$

$\pi^{(n)}_{z_n,t}$ always. Thus $\widetilde{Q} \ll_{\text{loc}} Q$ by Theorem 4.0.2 and the expression for $\mathfrak{L}_t$ emerges as a special case of (4.7). ∎

The conditions for local absolute continuity in Theorem 4.0.2 were expressed in terms of the Markov kernels generating Q and $\widetilde{Q}$. It is natural to ask for conditions expressed in terms of the compensating measures instead and such conditions exist in the literature, but there does not appear to be an easy translation between the two sets of conditions.

Example 4.0.6 *Let Q be an arbitrary probability on $(W, \mathcal{H})$ and let $\widetilde{Q}$ be the canonical dead process. Then $\widetilde{Q} \ll_{\text{loc}} Q$ if (and only if) $\overline{P}^{(0)}(t) > 0$ for all t ($P^{(0)}$ has termination point ∞), and in that case*

$$\mathfrak{L}_t = \frac{1}{\overline{P}^{(0)}(t)} 1_{(N_t^\circ = 0)}. \tag{4.8}$$

Example 4.0.7 *If $Q, \widetilde{Q}$ are probabilities on $(W, \mathcal{H})$, Q is Poisson (λ) where $\lambda > 0$, $\widetilde{Q}$ is Poisson $\left(\widetilde{\lambda}\right)$, then $\widetilde{Q} \ll_{\text{loc}} Q$ and*

$$\mathfrak{L}_t = e^{-\left(\widetilde{\lambda} - \lambda\right)t} \left(\frac{\widetilde{\lambda}}{\lambda}\right)^{N_t^\circ}.$$

Note that on $\mathcal{H}$, $\widetilde{Q} \perp Q$ if $\widetilde{\lambda} \neq \lambda$ since by the strong law of large numbers $\widetilde{Q}\left(\lim_{t\to\infty} \frac{1}{t} N_t^\circ = \widetilde{\lambda}\right) = 1$ and $Q\left(\lim_{t\to\infty} \frac{1}{t} N_t^\circ = \lambda\right) = 1$.

Example 4.0.8 *Recall the description in Example 2.2.3 of time-homogeneous Markov chains on an at most countably infinite state space E. On $(\mathcal{M}, \mathcal{H})$ the chain X° is defined by*

$$X_t^\circ = \eta_{\overline{N}_t^\circ}$$

with $X_0^\circ \equiv i_0$, a given state in E. Suppose that under Q, X° is Markov with transition intensities (q_{ij}) and that under $\widetilde{Q}$, X° is Markov with transition intensities $(\widetilde{q}_{ij})$. Then $\widetilde{Q} \ll_{\text{loc}} Q$ if whenever $\widetilde{q}_{ij} > 0$ for $i \neq j$ also $q_{ij} > 0$, and in that case

$$\mathfrak{L}_t = \exp\left(-\int_0^t \left(\widetilde{\lambda}_{X_s^\circ} - \lambda_{X_s^\circ}\right) ds\right) \prod_{(i,j): i \neq j} \left(\frac{\widetilde{q}_{ij}}{q_{ij}}\right)^{N_t^{\circ ij}}$$

where $\lambda_i = -q_{ii}$, $\widetilde{\lambda}_i = -\widetilde{q}_{ii}$ and

$$N_t^{\circ ij} = \sum_{0 < s \leq t} 1_{\left(X_{s-}^\circ = i, X_s^\circ = j\right)} = \int_{]0,t]} 1_{\left(X_{s-}^\circ = i\right)} dN_s^{\circ j}$$

for $i \neq j$ is the number of jumps from i to j on $[0,t]$, writing $N^{\circ j} = N^{\circ}(\{j\})$ for the number of jumps with mark j, i.e. the number of jumps X° makes into the state j.

If $\widetilde{Q} \ll_{\text{loc}} Q$, $\mathfrak{L}$ is a $Q-$martingale, hence has a representation

$$\mathfrak{L}_t = 1 + M^{\circ}(S),$$

cf. Theorem 3.6.1. If $\mathfrak{L}$ is of the form (4.7), identifying the jumps of $\mathfrak{L}$ one finds that the predictable flow S is given by

$$S_t^y = \mathfrak{L}_{t-}\left(\frac{\widetilde{\lambda}_t^y}{\lambda_t^y} - 1\right).$$

Suppose $Q, \widetilde{Q}$ are as in Example 4.0.6. If, say, $P^{(0)}$ has termination point 1, the expression (4.8) for $\mathfrak{L}_t$ still makes sense $Q-$a.s. for all t (since $Q\left(N_t^{\circ} = 0\right) = 0$ for $t \geq 1$), but $\widetilde{Q} \perp Q$ and $\mathfrak{L}$ is certainly not a $Q-$martingale, not even a $Q-$local martingale (since $Q \bigcap_{t \geq 1}(\mathfrak{L}_t = 0) = 1$).

It is also possible to obtain expressions of the form (4.7) for instance, that are $Q-$local martingales but not $Q-$martingales: if $\widetilde{Q}$ corresponds to a MPP_{ex} (a MPP with explosion possible, see p.9) but the conditions on the Markov kernels from Theorem 4.0.2 are satisfied, then $\mathfrak{L}$ is a $Q-$local martingale (τ_n) (because at the time of the n'th jump explosion has not yet occurred, and the conditions for local absolute continuity between the processes stopped at τ_n are obviously satisfied), but if $\widetilde{Q}(\tau_\infty \leq t) > 0$ (with $\tau_\infty = \lim_{n\to\infty} \tau_n$ the time of explosion), then $E_Q \mathfrak{L}_t < 1$ so $\mathfrak{L}$ is not a $Q-$martingale.

In principle this observation may be used to test for explosions: suppose $\mathfrak{L}$ is given by (4.7) but that one does not know whether $\widetilde{Q}$ may explode or not. Then the condition $E_Q \mathfrak{L}_t = 1$ for all t forces the $Q-$local martingale $\mathfrak{L}$ to be a true martingale and $\widetilde{Q}$ cannot explode in finite time.

Chapter 5

Examples of models

5.1 Independent point processes

Let $r \in \mathbb{N}$ with $r \geq 2$ and consider r given MPP's, each viewed as a RCM μ^i with mark space $(E^i, \mathcal{E}^i)$, $1 \leq i \leq r$, all defined on the same probability space $(\Omega, \mathcal{F}, P)$. In particular μ^i may correspond to a SPP, in which case E^i is a one point set and μ^i may be thought of as a counting process.

Let $(T^i_n)_{n \geq 1}$ denote the sequence of jump times for μ^i and $(Y^i_n)_{n \geq 1}$ the sequence of marks, so $Y^i_n \in \overline{E}^i = E^i \cup \{\nabla\}$, where ∇ is the irrelevant mark, common to all μ^i.

Now define $E = \{(i, y^i) : 1 \leq i \leq r,\ y^i \in E^i\}$ and let $\mathcal{E}$ be the $\sigma-$algebra of subsets of E generated by all sets of the form $\{i\} \times A^i$ where $1 \leq i \leq r,\ A^i \in \mathcal{E}_i$. Next put

$$\mu = \sum_{i=1}^{r} \sum_{\substack{n=1 \\ T^i_n < \infty}}^{\infty} \varepsilon_{(T^i_n, (i, Y^i_n))}. \tag{5.1}$$

Clearly μ is a random, $\overline{\mathbb{N}}_0-$valued measure such that

$$\mu([0,t] \times A) < \infty \qquad (t \in \mathbb{R}_0\ ,\ A \in \mathcal{E}). \tag{5.2}$$

However, for μ to be a RCM (with mark space E), we need (p.10) $\mu(\{t\} \times E) \leq 1$ for all t, at least $P-$almost surely, and for this to hold it is necessary to assume that no two finite jump times for different μ^i can agree, i.e.

$$P(\bigcup_{\substack{i,j=1 \\ i \neq j}}^{r} \bigcup_{k,n=1}^{\infty} \{T^i_k = T^j_n < \infty\}) = 0. \tag{5.3}$$

If (5.3) holds, μ is a RCM, and the corresponding MPP has mark space E and is the *aggregate* of all the μ^i : μ consists of all the points determined from the μ^i, $1 \leq i \leq r$.

Remark 5.1.1 *If all $E^i = E^\circ$ and (5.3) is satisfied one can also define the* superposition *of the μ^i as the MPP $\widetilde{\mu}$ with mark space E° given by*

$$\widetilde{\mu} = \sum_{i=1}^{r} \sum_{\substack{n=1 \\ T_n^i < \infty}}^{\infty} \varepsilon_{(T_n^i, Y_n^i)}. \tag{5.4}$$

The difference with the corresponding aggregate μ given by (5.1) is of course that μ keeps track of which of the original μ^i a mark came from while $\widetilde{\mu}$ does not.

Remark 5.1.2 *Even if (5.3) does not hold it is possible to define an aggregate of the μ^i using a larger mark space than E above, viz. the new mark space should consist of all finite subsets of E, a mark designating that all μ^i singled out in this subset jump simultaneously.*

From now on assume that (5.3) holds, and even, after discarding a $P-$null set, that

$$T_k^i(\omega) \neq T_n^j(\omega) \tag{5.5}$$

for $(i,k) \neq (j,n)$ and *all* $\omega \in \Omega$.

Note that with this assumption, if

$$N_t^i(A^i) = \mu^i([0,t] \times A^i) \quad (1 \leq i \leq r,\ A^i \in \mathcal{E}^i), \tag{5.6}$$

then $N_t(A) = \mu([0,t] \times A)$ for $A \in \mathcal{E}$ is given by

$$N_t(A) = \sum_{i=1}^{r} N_t^i(\widetilde{A}^i), \tag{5.7}$$

where $\widetilde{A}^i = \{y^i \in E^i : (i, y^i) \in A\}$.

For $1 \leq i \leq r$, let L^i denote the $\mathcal{F}_t^{\mu^i}-$compensating measure for μ^i and let

$$\Lambda_t^i(A^i) = L^i([0,t] \times A^i) \quad (t \in \mathbb{R}_0\ ,\ A^i \in \mathcal{E}^i) \tag{5.8}$$

denote the corresponding compensator for the counting process $N^i(A^i)$.

Theorem 5.1.1 *(a) Suppose the RCM's $\mu^1, \ldots, \mu^r$ are stochastically independent and satisfy (5.5). Then the aggregate μ of the μ^i has $\mathcal{F}_t^\mu-$ compensating measure L determined by*

$$L([0,t] \times A) = \Lambda_t(A) \quad (t \in \mathbb{R},\ A \in \mathcal{E}) \tag{5.9}$$

where

$$\Lambda_t(A) = \sum_{i=1}^{r} \Lambda_t^i\left(\widetilde{A}^i\right). \tag{5.10}$$

(b) *Suppose* $\mu^1, \dots, \mu^r$ *satisfy (5.5) and that the aggregate* μ *of the* μ^i *has* $\mathcal{F}_t^\mu-$*compensating measure determined by (5.9) and (5.10). Then* $\mu^1, \dots, \mu^r$ *are stochastically independent.*

Proof. (a). By Theorem 3.5.2 and the discussion in Section 3.7, we must show that $\Lambda(A)$ given by (5.10) is $\mathcal{F}_t^\mu-$predictable (which is obvious since the i'th term in the sum on the right is $\mathcal{F}_t^{\mu^i}-$predictable and $\mathcal{F}_t^{\mu^i} \subset \mathcal{F}_t^\mu$) and that $M(A) := N(A) - \Lambda(A)$ is a local $\mathcal{F}_t^\mu-$martingale.

To prove this last assertion, assume first that $E\overline{N}_t < \infty$ for all t. Then by Theorem 3.5.2, for all i, $M^i\left(\widetilde{A}^i\right) := N^i\left(\widetilde{A}^i\right) - \Lambda^i\left(\widetilde{A}^i\right)$ is a $\mathcal{F}_t^{\mu^i}-$martingale and further, the $M^i\left(\widetilde{A}^i\right)$ are independent since the μ^i are. We want to show that for $s < t$, $F \in \mathcal{F}_s^\mu$,

$$\int_F M_t(A)\, dP = \int_F M_s(A)\, dP. \tag{5.11}$$

Now $M(A) = \sum_{i=1}^r M^i\left(\widetilde{A}^i\right)$, so if $F = \bigcap_{i=1}^r F^i$ where $F^i \in \mathcal{F}_s^{\mu^i}$, using the independence and the martingale property of each $M^i\left(\widetilde{A}^i\right)$, it follows that

$$\begin{aligned}
\int_F M_t(A)\, dP &= \sum_{i=1}^r \int_F M_t^i\left(\widetilde{A}^i\right)\, dP \\
&= \sum_{i=1}^r P\left(\bigcap_{j:j\neq i} F^j\right) \int_{F^i} M_t^i\left(\widetilde{A}^i\right)\, dP \\
&= \sum_{i=1}^r P\left(\bigcap_{j:j\neq i} F^j\right) \int_{F^i} M_s^i\left(\widetilde{A}^i\right)\, dP \\
&= \int_F M_s(A)\, dP.
\end{aligned}$$

This proves (5.11) for $F \in \mathcal{F}_s^\mu$ of the form $F = \bigcap_i F^i$ with $F^i \in \mathcal{F}_s^{\mu^i}$. But since $\mathcal{F}_s^\mu = \sigma\left(\mathcal{F}_s^{\mu^i}\right)_{1\le i\le r}$, the collection of sets $F = \bigcap_i F^i$ generates $\mathcal{F}_s^\mu$ and as it is closed under the formation of finite intersections, (5.11) holds for all $F \in \mathcal{F}_s^\mu$.

Suppose now that the assumption $E\overline{N}_t < \infty$ for all t does not hold. For any $n \in \mathbb{N}$, $1 \le i \le r$ the stopped process $M^{T_n^i}(\widetilde{A}^i)$ is a $\mathcal{F}_t^{\mu^i}-$martingale, see Theorem 3.5.2. But the processes $M^{T_n^i}(\widetilde{A}^i)$ are independent, and copying the argument above one finds that

$$\sum_{i=1}^{r} M^{T_n^i}\left(\widetilde{A}^i\right)$$

is a $\mathcal{F}_t^\mu-$martingale for each n. With $T_n = \inf\left\{t \geq 0 : \overline{N}_t = n\right\}$, for all i, $T_n \leq T_n^i$, so by optional sampling, $M^{T_n}(A) = \left(\sum_{i=1}^r M^{T_n^i}\left(\widetilde{A}^i\right)\right)^{T_n}$ is a $\mathcal{F}_t^\mu-$martingale, i.e. $M(A)$ is a $\mathcal{F}_t^\mu-$local martingale.

(b). Here it is assumed that the compensator $\Lambda(A)$ is the sum of the compensators $\Lambda^i\left(\widetilde{A}^i\right)$ as in (5.10). From (a) we know this to be true if the μ^i are independent, and from Theorem 3.3.2 it now follows that if (5.10) holds they are indeed independent. ■

Remark 5.1.3 *More informally the theorem could be stated as follows: if the μ^i satisfy (5.5) (or just (5.3)), they are independent iff the compensating measure of the aggregate μ is the sum of the compensating measures for the μ^i. The precise meaning of this phrase is provided by (5.9) and (5.10).*

Remark 5.1.4 *Note that if the μ^i are independent, (5.3) is satisfied if the restriction to $\mathbb{R}_+$ of the distribution of each T_n^i is continuous.*

Example 5.1.1 *Suppose the μ^i are independent, and that each μ^i has a $\mathcal{F}_t^{\mu^i}$ − predictable intensity process $\left(\lambda^i\left(A^i\right)\right)_{A^i\in\mathcal{E}^i}$, so $\Lambda_t^i\left(A^i\right) = \int_0^t \lambda_s^i\left(A^i\right)\, ds$. Then (5.3) is satisfied and the aggregate μ has an intensity process $(\lambda(A))_{A\in\mathcal{E}}$ given by*

$$\lambda_t(A) = \sum_{i=!}^{r} \lambda_t^i\left(\widetilde{A}^i\right).$$

5.2 Homogeneous Poisson random measures and processes with stationary, independent increments

Let μ be RCM with mark space E, defined on some probability space $(\Omega, \mathcal{F}, P)$. As usual for $A \in \mathcal{E},\ t \in \mathbb{R}_0$,

$$N_t(A) = \mu([0,t] \times A).$$

Definition 5.2.1 *μ is a* homogeneous Poisson random measure *if there exists a positive bounded measure κ on $(E, \mathcal{E})$ such that* *(i) for every $A \in \mathcal{E}$, $N(A)$ is a homogeneous Poisson process with parameter $\kappa(A)$,*

(ii) for every $k \geq 2$, $A_1, \ldots, A_k \in \mathcal{E}$ mutually disjoint, the counting processes $N(A_1), \ldots, N(A_k)$ are independent.

It would suffice to assume that κ be a $\mathbb{R}_0-$valued function on $\mathcal{E}$ since (i) implies that κ is then a measure: by monotone convergence, if $A_1, A_2, \ldots \in \mathcal{E}$ are disjoint and $A = \bigcup_j A_j$, then

$$\begin{aligned}
\kappa(A) &= EN_1(A) \\
&= E\mu\left([0,1] \times A\right) \\
&= E\sum_j \mu\left([0,1] \times A_j\right) \\
&= \sum_j E\mu\left([0,1] \times A_j\right) \\
&= \sum_j \kappa(A_j).
\end{aligned}$$

κ is called the *intensity measure* for μ.

As we shall now see, homogeneous Poisson measures are precisely the RCM's discussed in Example 3.8.3, (see also Example 3.3.3).

Proposition 5.2.1 *A RCM μ is a homogeneous Poisson random measure if and only if the compensating measure L for μ (with respect to the filtration $(\mathcal{F}_t^\mu)$) is non-random and of the form*

$$L = \ell \otimes \kappa, \tag{5.12}$$

where ℓ denotes Lebesgue measure on $\mathbb{R}_0$ and κ is a positive bounded measure on $(E, \mathcal{E})$. If L has this form, κ is the intensity measure for μ.

Proof. Suppose μ is homogeneous Poisson, and let κ denote the intensity measure. In particular $\overline{N}$ is homogeneous Poisson $\kappa(E)$ so $E\overline{N}_t = t\kappa(E) < \infty$. Therefore, by Theorem 3.5.2, to show that $L = \ell \otimes \kappa$, it suffices to show that for every $A \in \mathcal{E}$, $M(A)$, where $M_t(A) = N_t(A) - t\kappa(A)$, is a $\mathcal{F}_t^\mu-$martingale. By Lemma 5.2.2 below however, for $s < t$, $N_t(A) - N_s(A)$ is independent of $\mathcal{F}_s^\mu$, and so

$$E(N_t(A) - N_s(A)|\mathcal{F}_s^\mu) = E(N_t(A) - N_s(A)) = (t-s)\kappa(A).$$

The martingale property follows immediately.

Now suppose conversely that μ is a MPP with $\mathcal{F}_t^\mu$−compensating measure L given by (5.12). That (i) and (ii) from Definition 5.2.1 hold was shown in Example 3.8.3. ■

To complete the proof of Proposition 5.2.1 we still need

Lemma 5.2.2 *If μ is homogeneous Poisson, then for $s < t,\ A \in \mathcal{E},\ N_t(A) - N_s(A)$ is independent of $\mathcal{F}_s^\mu$.*

Proof. Since $\mathcal{F}_s^\mu = \sigma(N_u(\widetilde{A})|0 \le u \le s,\ \widetilde{A} \in \mathcal{E})$, it suffices to show that for any $k \in \mathbb{N}$, $u_1, \ldots, u_k \le s,\ A_1, \ldots, A_k \in \mathcal{E},\ A \in \mathcal{E},\ (N_{u_1}(A_1), \ldots, N_{u_k}(A_k))$ is independent of $N_t(A) - N_s(A)$. We do this for $k = 1$, writing $u_1 = u,\ A_1 = \widetilde{A}$. Then

$$\begin{aligned} N_t(A) - N_s(A) &= (N_t(A \cap \widetilde{A}) - N_s(A \cap \widetilde{A})) + (N_t(A \cap \widetilde{A}^c) - N_s(A \cap \widetilde{A}^c)), \\ N_u(\widetilde{A}) &= N_u(A \cap \widetilde{A}) + N_u(A \cap \widetilde{A}^c). \end{aligned}$$

Because μ is homogeneous Poisson, it is seen that each of the two terms in the expression for $N_u(\widetilde{A})$ are independent of each of the two terms in the expression for $N_t(A) - N_s(A)$. ■

Having characterized homogeneous Poisson random measures, we shall now see how they may be used to describe certain Lévy processes, i.e. processes with stationary independent increments.

Let $X = (X_t)_{t \ge 0}$ be a $\mathbb{R}^d$−valued process, defined on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. Assume that X is a step process, i.e. X has only finitely many jumps on finite intervals, is piecewise constant between jumps and right-continuous. Also assume that $X_0 \equiv x_0$ for some $x_0 \in \mathbb{R}^d$, and that X is adapted.

Definition 5.2.2 *A $\mathbb{R}^d$−valued step process X has* stationary, independent increments *with respect to the filtration $(\mathcal{F}_t)$ if it is adapted and for all $0 \le s < t$,*

(i) $X_t - X_s$ is independent of $\mathcal{F}_s$,

(ii) the distribution of $X_t - X_s$ depends on s, t only through the difference $t - s$.

A process with stationary, independent increments is also called a *Lévy process.*

Since X is adapted, if for $s < t,\ X_t - X_s$ is independent of $\mathcal{F}_s,\ X_t - X_s$ is also independent of $\mathcal{F}_s^X$. Note that $X_t - X_s$ is independent of $\mathcal{F}_s^X$ iff for all $n \in \mathbb{N},\ 0 \le s_1 < \cdots < s_n \le s,\ X_t - X_s$ is independent of $(X_{s_1}, \ldots, X_{s_n})$.

Lemma 5.2.3 *Suppose X has stationary, independent increments with respect to $(\mathcal{F}_t)$. Then, for every $s \geq 0$ the process $\widetilde{X} = (\widetilde{X}_u)_{u \geq 0}$ defined by*

$$\widetilde{X}_u = X_{u+s} - X_s$$

is independent of $\mathcal{F}_s$, and has the same distribution as the process $(X_t - x_0)_{t \geq 0}$, where x_0 is the initial value of X.

Proof. The lemma follows if we show that for $n \in \mathbb{N}$, $0 \leq u_1 < \cdots < u_n$, $(\widetilde{X}_{u_1}, \ldots, \widetilde{X}_{u_n})$ is independent of $\mathcal{F}_s$ and with the same distribution as $(X_{u_1} - x_0, \ldots, X_{u_n} - x_0)$. This in turn follows if it is shown that $(\widetilde{X}_{u_1}, \widetilde{X}_{u_2} - \widetilde{X}_{u_1}, \ldots, \widetilde{X}_{u_n} - \widetilde{X}_{u_{n-1}})$ is independent of $\mathcal{F}_s$ and has the same distribution as $(X_{u_1} - x_0, X_{u_2} - X_{u_1}, \ldots, X_{u_n} - X_{u_{n-1}})$ (since $(X_{u_j} - x_0) - (X_{u_{j-1}} - x_0) = X_{u_j} - X_{u_{j-1}}$). But $\widetilde{X}_{u_n} - \widetilde{X}_{u_{n-1}} = X_{u_n+s} - X_{u_{n-1}+s}$ is independent of $\mathcal{F}_{u_{n-1}+s}$ and therefore independent of $\mathcal{F}_s$ and $(\widetilde{X}_{u_1}, \widetilde{X}_{u_2} - \widetilde{X}_{u_1}, \ldots, \widetilde{X}_{u_{n-1}} - \widetilde{X}_{u_{n-2}})$. Also, $\widetilde{X}_{u_n} - \widetilde{X}_{u_{n-1}}$ has the same distribution as $X_{u_n} - X_{u_{n-1}}$. Repeating this argument, one finds that $\mathcal{F}_s$, $\widetilde{X}_{u_1}, \widetilde{X}_{u_2} - \widetilde{X}_{u_1}, \ldots, \widetilde{X}_{u_n} - \widetilde{X}_{u_{n-1}}$ are independent, and for $j \geq 2$, $\widetilde{X}_{u_j} - \widetilde{X}_{u_{j-1}}$ has the same distribution as $X_{u_j} - X_{u_{j-1}}$. Since also $\widetilde{X}_{u_1}$ has the same distribution as $X_{u_1} - X_0 = X_u - x_0$, the desired conclusion follows. ■

Let now X be an arbitrary $\mathbb{R}^d-$valued step process. We identify X with the MPP $(\mathcal{T}, \mathcal{Y}) = (T_n, Y_n)_{n \geq 1}$ with mark space $(\mathbb{R}^d_{\backslash 0}, \mathcal{B}^d_{\backslash 0})$, $(\left(\mathbb{R}^d, \mathcal{B}^d\right)$ with 0 removed) where T_n is the time of the n'th jump of X, and $Y_n = \triangle X_{T_n}$ is the size of that jump (so here we use a different set of marks from that normally used for describing piecewise deterministic processes, see p.19). Write μ for the RCM defined by $(\mathcal{T}, \mathcal{Y})$ or X, and note that with $N_t(B) = \mu([0,t] \times B)$,

$$N_t(B) = \sum_{0 < s \leq t} 1_{(\triangle X_s \in B)} \quad (t \in \mathbb{R}_0 \ , \ B \in \mathcal{B}^d_{\backslash 0}), \tag{5.13}$$

while X because it is a step process is determined by μ through the equations

$$X_t = x_0 + \int_{[0,t] \times \mathbb{R}^d_{\backslash 0}} y \, \mu(ds, dy) \quad (t \in \mathbb{R}_0). \tag{5.14}$$

In particular we see that $\mathcal{F}^X_t = \mathcal{F}^\mu_t$.

Proposition 5.2.4 *The step process X with $X_0 \equiv x_0$ has stationary independent increments with respect to $(\mathcal{F}^X_t)$ if and only if the RCM μ determining X is a homogeneous Poisson random measure.*

Remark 5.2.1 *It follows from the proposition and Example 3.8.3 that X has stationary independent increments iff X is a* compound Poisson process*, i.e.*

$$X_t = x_0 + \sum_{n=0}^{N_t} U_n$$

where N is a homogeneous Poisson counting process and the U_n are i.i.d. $\mathbb{R}^d_{\backslash 0}$−valued random variables, independent of N.

Proof. Suppose first that X has stationary independent increments. For $B \in \mathcal{B}^d_{\backslash 0}$, $s < t$, by (5.13)

$$\begin{aligned} N_t(B) - N_s(B) &= \sum_{u:s<u\le t} 1_{(\triangle X_u \in B)} \\ &= \sum_{v:0<v\le t-s} 1_{\left(\triangle \widetilde{X}_v \in B\right)} \end{aligned}$$

with $\widetilde{X}_v = X_{s+v} - X_s$. From Lemma 5.2.3 it follows that $N_t(B) - N_s(B)$ is independent of $\mathcal{F}^X_s$ with a distribution depending on s and t through $t - s$ only. Assuming for the moment that $E\overline{N}_u < \infty$ for all u, we see that for any B

$$\begin{aligned} E(N_t(B) - N_s(B)|\mathcal{F}_s) &= E(N_t(B) - N_s(B)) \\ &= \gamma_B(t-s) \end{aligned} \tag{5.15}$$

where $\gamma_B(u) = EN_u(B)$. But for $u, v \geq 0$,

$$\gamma_B(u+v) = EN_{u+v}(B) = E(N_u(B) + (N_{u+v}(B) - N_u(B))) = \gamma_B(u) + \gamma_B(v),$$

and since γ_B is non-decreasing, this implies that $\gamma_B(u) = \lambda_B u$ for some $\lambda_B \geq 0$. (5.15) now shows $M(B)$ to be an $(\mathcal{F}_t)$−martingale, where $M_t(B) = N_t(B) - \lambda_B t$, and thus $N(B)$ is Poisson (λ_B).

If $B_1, \ldots, B_k \in \mathcal{B}^d_{\backslash 0}$ are disjoint, the $(\mathcal{F}^X_t)$−compensators for $N(B_1), \ldots,$ $N(B_k)$ are $(\lambda_j t)_{t\geq 0}$, where $\lambda_j = \lambda_{B_j}$. Since by the definition of the $N(B_j)$ no two of them can jump simultaneously, (5.5) holds and from Theorem 5.1.1 it follows that the $N(B_j)$ are independent, and we have shown that μ is homogeneous Poisson, assuming that $E\overline{N}_t < \infty$ for all t.

To get rid of this assumption, we show directly that $\overline{N}$ is a Poisson process. Let for $t \geq 0$, $T_{1,t} = \inf\{u > t | X_u \neq X_t\}$ denote the time of the first jump of X after time t. Because

$$(T_{1,t} - t > v) = \bigcap_{0<u\le v} (X_{u+t} - X_t = 0),$$

Lemma 5.2.3 shows that this event is independent of $\mathcal{F}_t^X$ and has the same probability as the event

$$\bigcap_{0<u\leq v} (X_u - X_0 = 0) = (T_1 > v).$$

But then, if $\psi(v) = P(T_1 > v)$,

$$\psi(u+v) = P(T_1 > u+v) = P(T_1 > u\ ,\ T_{1,u} > u+v) = \psi(u)\psi(v)$$

and therefore

$$P(T_1 > v) = e^{-\overline{\lambda} v}$$

for some $\overline{\lambda} \geq 0$. If $\overline{\lambda} = 0$, $\overline{N}$ is the dead process (and $X_t \equiv x_0$, for all t) and $E\overline{N}_t = 0 < \infty$. If $\overline{\lambda} > 0$, we get for $u > t$,

$$P(T_{1,t} > u | \mathcal{F}_t^X) = e^{-\overline{\lambda}(u-t)} = P(T_{1,t} > u | \mathcal{F}_t^{\overline{N}}).$$

But if $P^{(n)}_{t_1 \ldots t_n}$ denotes the conditional jump time distributions for $\overline{N}$, then by Lemma 3.3.3 (bi),

$$P(T_{1,t} > u | \mathcal{F}_t^{\overline{N}}) = \frac{P^{(\overline{N}_t)}_{Z_t}(]u,\infty])}{P^{(\overline{N}_t)}_{Z_t}(]t,\infty])},$$

where $Z_t = \left(T_1, \ldots, T_{\overline{N}_t}\right)$, and this will equal $e^{-\overline{\lambda}(u-t)}$ precisely when

$$P^{(\overline{N}_t)}_{T_1 \cdots T_{\overline{N}_t}}(]t,\infty]) = e^{-\overline{\lambda}(t - T_{\overline{N}_t})}.$$

Using this on $(\overline{N}_t = n)$ for all n determines the $P^{(n)}_{t_1 \cdots t_n}$ and shows that $\overline{N}$ is Poisson $\overline{\lambda}$, in particular $E\overline{N}_t = \overline{\lambda} t < \infty$, and the preceding argument that μ is homogeneous Poisson applies.

For the converse, suppose that the RCM μ determining the step process X is homogeneous Poisson. By (5.14), for $s < t$,

$$X_t - X_s = \int_{]s,t] \times \mathbb{R}^d_{\backslash 0}} y\, \mu(du, dy)$$

which we must show is independent of $\mathcal{F}_s^X = \mathcal{F}_s^\mu$ with a distribution depending on s, t through $t - s$ only. But this follows immediately from the observation that if μ is homogeneous Poisson, then for any s the restriction of μ to $]s,\infty[\, \times E$ is independent of $\mathcal{F}_s^\mu$ with the same distribution as $\widetilde{\mu}_s := \sum_{n: T_n < \infty} \varepsilon_{(T_n + s, Y_n)}$. ■

Let X be a $\mathbb{R}-$valued step process with stationary, independent increments and finitely many jumps on finite intervals so that Proposition 5.2.4 applies. If ν_t is the distribution of $X_{s+t} - X_s$ (for any s) clearly

$$\nu_s * \nu_t = \nu_{s+t} \quad (s, t \in \mathbb{R}_0)$$

with $\nu_t \to \varepsilon_0$ weakly as $t \downarrow 0$, i.e. $(\nu_t)_{t \geq 0}$ is a weakly continuous convolution semigroup of probability measures on $\overline{\mathbb{R}}$ and in particular each ν_t is infinitely divisible. As such the characteristic function ϕ_t of ν_t is given by (a special form of) the famous Lévy-Khinchine formula. It is however possible to arrive at this directly using Proposition 5.2.4: Let for $K \in \mathbb{N}$, $k \in \mathbb{Z}$, $I_{K,k}$ denote the interval $\left]\frac{k-1}{2^K}, \frac{k}{2^K}\right]$ if $k \neq 0$ and the interval $\left]-\frac{1}{2^K}, 0\right[$ if $k = 0$. From (5.14) it follows that

$$X_t - X_0 = \lim_{K \to \infty} \sum_{k \in \mathbb{Z}} \frac{k}{2^K} N_t(I_{K,k})$$

and since μ is homogeneous Poisson, because the $(I_{K,k})_{k \in \mathbb{Z}}$ are mutually disjoint we find that for $u \in \mathbb{R}$

$$\begin{aligned} \phi_t(u) &= E \exp\left(iu\left(X_t - X_0\right)\right) \\ &= \lim_{K \to \infty} E \exp\left(iu \sum_{k \in \mathbb{Z}} \frac{k}{2^K} N_t(I_{K,k})\right) \\ &= \lim_{K \to \infty} \prod_{k \in \mathbb{Z}} E \exp\left(iu \frac{k}{2^K} N_t(I_{K,k})\right), \end{aligned}$$

since the random variables $(N_t(I_{n,k})_{k \in \mathbb{Z}^p}$ are independent. Since by Proposition 5.2.1 the compensating measure for μ is of the form $\ell \otimes \kappa$ with κ a positive bounded measure on $\mathbb{R}_{\backslash 0}$ and the process $N(I_{n,k})$ is Poisson $\kappa(I_{n,k})$ so that

$$E \exp\left(ivN_t(I_{n,k})\right) = \exp\left(t\kappa(I_{n,k})(e^{iv} - 1)\right),$$

we deduce that

$$\begin{aligned} \phi_t(u) &= \lim_{K \to \infty} \prod_{k \in \mathbb{Z}} \exp\left(t\kappa(I_{K,k})(e^{iuk2^{-K}} - 1)\right) \\ &= \exp\left(\lim_{K \to \infty} t \sum_{k \in \mathbb{Z}} \kappa(I_{K,k})(e^{iuk2^{-K}} - 1)\right) \\ &= \exp\left(\lim_{K \to \infty} t \int_{\mathbb{R}_{\backslash 0}} f_K(x)\kappa(dx)\right) \end{aligned}$$

where

$$\begin{aligned} f_K(x) &= \sum_{k\in\mathbb{Z}} \left(e^{iuk2^{-K}} - 1\right) 1_{I_{K,k}}(x) \\ &\underset{K\to\infty}{\to} e^{iux} - 1. \end{aligned}$$

Using dominated convergence, we finally arrive at

$$\phi_t(u) = \exp\left(t \int_{\mathbb{R}_{\backslash 0}} \left(e^{iux} - 1\right) \kappa(dx)\right).$$

Note that if X is $\mathbb{R}-$valued with stationary independent increments with finitely many jumps on finite intervals, but is no longer a step process, then

$$\phi_t(u) = \exp\left(iat + t \int_{\mathbb{R}_{\backslash 0}} \left(e^{iux} - 1\right) \kappa(dx)\right) \tag{5.16}$$

for some $a \in \mathbb{R}$ corresponding to considering $X_t = \widetilde{X}_t + at$ with $\widetilde{X}$ a step process with stationary independent increments as studied above. Finally, if κ is an unbounded positive measure, in particular $\kappa\left(\mathbb{R}_{\backslash 0}\right) = \infty$, and satisfies that

$$\kappa\left(\mathbb{R}\backslash\left[-h,h\right]\right) < \infty, \quad \int_{[-h,h]\backslash 0} \frac{1}{y}\,\kappa\left(dy\right) < \infty \quad (h > 0),$$

then (5.16) is the characteristic function of $X_t - X_0$ where X has stationary independent increments and $X_t - X_0 - at = \sum_{0<s\le t} \Delta X_s$ is the sum of jumps for X, where the series converges absolutely, $\sum_{0<s\le t} |\Delta X_s| < \infty$ a.s. With κ unbounded, the set of jump times is countably infinite and dense in $\mathbb{R}_+$. κ is called the *Lévy measure* for X. The most general form of the Lévy-Khinchine formula for $\mathbb{R}-$ valued processes incorporates of course a Gaussian component (corresponding to adding an independent scaled Brownian motion to the X corresponding to (5.16)) and also allows for sums of jumps that need not converge absolutely.

We only considered $\mathbb{R}^d-$valued processes above, but it should be fairly clear that Proposition 5.2.4 generalizes in a straightforward manner to $V-$ valued processes with stationary independent increments, where V is a vector space.

5.3 Deterministic compensators, Poisson measures

From the point of view of describing the distribution of a CP or RCM through its (canonical) compensator Λ° or compensating measure L°, the simplest case is that where Λ° or L° is deterministic (non-random). The examples we have seen so far are the homogeneous Poisson process (Examples 2.1.2 and 3.3.1), where

$$\Lambda_t^\circ = \lambda t$$

for some constant $\lambda \geq 0$, and the homogeneous Poisson measures of Section 5.2, where

$$L^\circ = \ell \otimes \kappa$$

with ℓ Lebesgue measure on $\mathbb{R}_0$ and κ a positive bounded measure on E.

We shall now discuss the processes determined by more general forms of deterministic compensators or compensating measures.

Let $D : \mathbb{R}_0 \to \mathbb{R}_0$ be an increasing, right-continuous function with $D(0) = 0$.

Definition 5.3.1 *A $\mathbb{N}_0-$valued step process X, defined on some probability space, is a* non-homogeneous Poisson process with rate function *D provided $X_0 = 0$ a.s. and X has independent increments such that for $s \leq t$, $X_t - X_s$ follows a Poisson distribution with parameter $D(t) - D(s)$.*

Note. The definition requires that for all $n \geq 1$, $0 = t_0 < t_1 < \cdots < t_n$ the random variable $X_{t_n} - X_{t_{n-1}}$ be independent of $\left(X_{t_1}, \ldots, X_{t_{n-1}}\right)$ and Poisson $\left(D(t_n) - D(t_{n-1})\right)$. As argued in Lemma 5.2.3 it then follows that for $s \leq t$, $X_t - X_s$ is independent of $\mathcal{F}_s^X$.

Homogeneous Poisson processes are special cases of non-homogeneous processes, so the terminology 'non-homogeneous' should really be understood as 'possibly non-homogeneous'!

Let Q_1 denote the canonical homogeneous Poisson process with parameter 1, i.e. the compensator for the probability Q_1 on $(W, \mathcal{H})$ is $\Lambda_{1,t}^\circ = t$. With D right-continuous and increasing as above with $D(0) = 0$, define the process N^* on $(W, \mathcal{H})$ by

$$N_u^* = N_{D(u)}^\circ \quad (u \geq 0).$$

Note that N^* is increasing and right-continuous.

Proposition 5.3.1 *(i) The process N^* is a non-homogeneous Poisson process with rate function D.*

(ii) The process N^* is a counting process if and only if D is continuous.

(iii) If D is continuous, the distribution Q^* of N^* has deterministic compensator $\Lambda^{\circ *}$ given by

$$\Lambda_t^{\circ *} = D(t).$$

Proof. (i). That N^* has independent increments follows directly because under Q_1, N° has independent increments. And because N° is Poisson (1), for $u < v$, $N^*_v - N^*_u$ is Poisson $(D(v) - D(u))$.

(ii). We have $\Delta N^*_u = N^\circ_{D(u)} - N^\circ_{D(u-)-}$. If D is continuous, $\Delta N^*_u = \Delta N^\circ_{D(u)} \leq 1$ everywhere on W for all u, and N^* is a counting process. If D has a discontinuity, $\Delta D(u_0) = D(u_0) - D(u_0-) > 0$, at $u_0 \in \mathbb{R}_+$, then

$$\Delta N^*_{u_0} = \lim_{h \downarrow 0} \left(N^\circ_{D(u_0)} - N^\circ_{D(u_0-h)} \right)$$

as a limit of Poisson random variables with parameters $D(u_0) - D(u_0 - h)$, is itself Poisson with parameter $\Delta D(u_0)$. In particular $Q_1 \left(\Delta N^*_{u_0} \geq 2 \right) > 0$ and N^* is not a counting process.

(iii). The martingale properties of N° immediately show that

$$M^*_u := N^*_u - D(u)$$

is a $\mathcal{F}^{N^*}_u = \mathcal{H}_{D(u)}-$martingale (the equality between filtrations holds only because D is continuous). Since D is deterministic, as a process it is $\mathcal{F}^{N^*}_u-$ predictable, hence D is the $\mathcal{F}^{N^*}_u-$compensator for N^*, cf. Theorem 3.5.2 (a) and Section 3.7. ■

The possible candidates for deterministic compensators of counting processes are all functions D of the form above, satisfying in addition that, cf. (3.23),

$$\Delta D(u) \leq 1 \quad (u \geq 0)\,. \tag{5.17}$$

The proposition takes care of the case where D is continuous. If D has discontinuities, but of course satisfies (5.17), things are more difficult and the description of the process with compensator D not so nice. To treat this case, identify N^* with the MPP $(T^*_n, Y^*_n)_{n \geq 1}$ where T^*_n is the time of the n'th jump of N^* and $Y^*_n = \Delta N^*_{T^*_n}$ is the size of that jump. For $y \in \mathbb{N}$, introduce the counting process N^{*y}, where

$$N^{*y}_u = \sum_{0 < r \leq u} 1_{(\Delta N^*_r = y)} \tag{5.18}$$

is the number of jumps for N^* of size y on $[0, u]$. Going back to the definition of N^* in terms of N°, argue that for $u < v$,

$$\left(N^{*y}_v - N^{*y}_u\right)_{y \geq 1} \text{ is independent of } \mathcal{H}_{D(u)}. \tag{5.19}$$

Since

$$N^*_u = \sum_{y=1}^{\infty} y N^{*y}_u, \tag{5.20}$$

in particular $N_u^* \geq N_u^{*y}$ for all y, and so $EN_u^{*y} \leq EN_u^* < \infty$. From this and (5.19) it immediately follows that each M^{*y}, where

$$M_u^{*y} = N_u^{*y} - EN_u^{*y},$$

is a $\mathcal{H}_{D(u)}-$martingale. Since $\mathcal{F}_u^{N^*} \subset \mathcal{H}_{D(u)}$ (with strict inclusion if D has discontinuities on $[0,u]$), we have now shown that the RCM $\mu^* = \sum_{n:T_n^*<\infty} \varepsilon_{(T_n^*,Y_n^*)}$ has *deterministic compensating measure* $\mathfrak{D}$ on $\mathbb{R}_0 \times \mathbb{N}$, where

$$D^{*y}(u) := \mathfrak{D}\left([0,u] \times \{y\}\right) = EN_u^{*y}.$$

It remains to find EN_u^{*y}, which is easily done from (5.18) if $y \geq 2$ since then the only terms appearing in the sum are for $r-$values with $\Delta D(r) > 0$, in which case ΔN_u^* is Poisson $(\Delta D(u))$, cf. Proposition 5.3.1 and its proof. Thus

$$EN_u^{*y} = \sum_{0<r\leq u} \frac{1}{y!} \left(\Delta D(r)\right)^y e^{-\Delta D(r)} \quad (y \geq 2).$$

For $y = 1$, use (5.18), (5.20) to obtain

$$\begin{aligned} EN_u^{*1} &= EN_u^* - \sum_{y=2}^{\infty} yEN_u^{*y} \\ &= \sum_{0<r\leq u} \Delta D(r) e^{-\Delta D(r)} + D^c(u), \end{aligned}$$

where $D^c(u) := D(u) - \sum_{r\leq u} \Delta D(r)$ is the continuous part of D.

Defining $\overline{N}^*$ by

$$\overline{N}_u^* = \sum_{y=1}^{\infty} N_u^{*y},$$

the total number of jumps for N^*, from (5.19) it follows that $\overline{M}_u^* = \overline{N}_u^* - E\overline{N}_u^*$ is a $\mathcal{H}_{D(u)}-$martingale, and since

$$E\overline{N}_u^* = D^c(u) + \sum_{0<r\leq u} \left(1 - e^{-\Delta D(r)}\right)$$

we have shown

Proposition 5.3.2 *If $\widetilde{D} : \mathbb{R}_0 \to \mathbb{R}$ is right-continuous and increasing with $\widetilde{D}(u) = 0$ and $\Delta\widetilde{D}(u) < 1$ for all u, then the deterministic function $\widetilde{D}$ is the compensator of the counting process with independent increments $\overline{N}^*$, where*

$\overline{N}^*_u$ is the total number of jumps on $[0,u]$ for the non-homogeneous Poisson process process N^* with rate function D given by

$$D^c = \widetilde{D}^c, \quad \Delta D(u) = \log\left(1 - \Delta\widetilde{D}(u)\right),$$

$D^c, \widetilde{D}^c$ denoting the continuous part of D and $\widetilde{D}$ respectively.

This result still excludes the possibility $\Delta\widetilde{D}(u) = 1$. But if the counting process N has deterministic compensator $\widetilde{D}$ and $\Delta\widetilde{D}(u_0) = 1$ for at least one u_0, then $\Delta N_{u_0} = 1$ a.s., i.e. a jump at u_0 is forced. Proposition 5.3.2 is easily adjusted to take this exciting option into account!

Returning to the non-homogeneous Poisson process N^* with rate function D discussed above, since we have found the $\mathcal{F}^{N^*}_u$–compensator for all N^{*y}, we are able to read off the Markov kernels $P^{*(n)}_{z_n}$, $\pi^{*(n)}_{z_n,u}$ generating the jump times T^*_n (which are also the jump times of $\overline{N}^*$) and the jump sizes Y^*_n for N^*. The result is (using (3.27), (3.28) and Section 3.1), writing $z_n = (u_1, \ldots, u_n; y_1, \ldots, y_n)$ where only z_n such that $y_k \geq 2$ implies $\Delta D(u_k) > 0$ are relevant,

$$\overline{P}^{*(n)}_{z_n}(u) = \exp\left(-\left(D(u) - D(u_n)\right)\right)$$

for $u \geq u_n$,

$$\pi^{*(n)}_{z_n,u}(\{y\}) = \begin{cases} 1 & \text{if } \Delta D(u) = 0,\ y = 1 \\ \dfrac{(\Delta D(u))^y}{y!}\dfrac{e^{-\Delta D(u)}}{1 - e^{-\Delta D(u)}} & \text{if } \Delta D(u) > 0,\ y \in \mathbb{N} \end{cases}$$

for $u > u_n$, listing only the π–values > 0. Note that if $\Delta D(u) > 0$, $\pi^{*(n)}_{z_n,u}$ is the distribution of a Poisson $\Delta D(u)$ random variable conditioned to be ≥ 1.

We turn now to the discussion of RCM's with deterministic compensating measures, so let μ be a RCM with mark space $(E, \mathcal{E})$ and non-random $\mathcal{F}^\mu_t$–compensating measure $\mathfrak{D}$ – retaining the traditional 't' as notation for timepoints rather than the 'u' used above.

Defining $D(t,A) = \mathfrak{D}\left([0,t] \times A\right)$, since $ED(t,A) = D(t,A) < \infty$, $M(A) = N(A) - D(\cdot, A)$ is a martingale and Theorem 3.5.2 (b) and the results above yield the distribution of $N(A)$. If in particular $D(\cdot, A)$ is continuous, $N(A)$ is non-homogeneous Poisson with rate function $D(\cdot, A)$.

To study the joint distribution of several $N(A)$, assume for convenience and simplicity that $\mathfrak{D}$ is continuous, $\mathfrak{D}\left(\{t\} \times E\right) = 0$ for all t. Next, let for $r \geq 2$, $A_1, \ldots, A_r \in \mathcal{E}$ be mutually disjoint, let $(u_1, \ldots, u_r) \in \mathbb{R}^r$, define

$$\xi(t) = E\exp\left(i\sum_{j=1}^r u_j N_t\left(A_j\right)\right)$$

(which is continuous because $\mathfrak{D}$ is), and then show, proceeding exactly as in Example 3.8.3, that M is a $\mathbb{C}-$valued martingale, where

$$M_t = \frac{1}{\xi(t)} \exp\left(i \sum_{j=1}^{r} u_j N_t(A_j)\right).$$

The conclusion is

Proposition 5.3.3 *If μ has a compensating measure which is deterministic and continuous, then for any $r \geq 2$ and any $A_1, \ldots, A_r \in \mathcal{E}$ mutually disjoint, the counting processes $N(A_1), \ldots, N(A_r)$ are independent and $N(A_j)$ is a non-homogeneous Poisson process with rate function $D(\cdot, A_j)$.*

We have deliberately ignored any mention of filtrations in this result. The point is of course that if μ is $\mathcal{F}_t-$adapted for some filtration $(\mathcal{F}_t)$ and has deterministic $\mathcal{F}_t-$compensating measure $\mathfrak{D}$ (not necessarily continuous), then $\mathfrak{D}$ is automatically $\mathcal{F}_t-$predictable and hence characterizes the distribution of μ.

If again we assume that $\mathfrak{D}$ is continuous, the Markov kernels generating the jump times for μ are given by (use (3.27))

$$\overline{P}^{(n)}_{z_n}(t) = \exp\left(-\left(\overline{D}(t) - \overline{D}(t_n)\right)\right) \quad (t \geq t_n), \tag{5.21}$$

where of course $\overline{D}(t) = \mathfrak{D}([0,t] \times E)$. The kernels generating the jumps are given by

$$\pi^{(n)}_{z_n,t}(A) = \frac{dD(\cdot, A)}{d\overline{D}}(t) \quad (t > t_n, A \in \mathcal{E}),$$

cf. (3.28), in particular the conditional distribution of Y_{n+1} given $(T_1, \ldots, T_n, T_{n+1}; Y_1, \ldots, Y_n)$ depends on T_{n+1} only. Combining this with (5.21) we therefore see that if $\mathfrak{D} = \rho \otimes \kappa$ is the product of a continuous, positive and $\sigma-$finite mesure on $\mathbb{R}_0$ and a positive, bounded measure on E, then the Y_n are iid with distribution $\kappa/\kappa(E)$ and independent of the sequence (T_k).

The final remark we shall make is that if μ has deterministic, continuous compensator $\mathfrak{D}$, then it follows from Proposition 5.3.3 that μ viewed as an ordinary *simple* point process on $\mathbb{R}_0 \times E$ is a *Poisson process with intensity measure* $\mathfrak{D}$, i.e. for any $C \in \mathcal{B}_0 \otimes \mathcal{E}$, $\mu(C)$ is Poisson $\mathfrak{D}(C)$ and for any $r \geq 2$, $C_1, \ldots, C_r \in \mathcal{B}_0 \otimes \mathcal{E}$ mutually disjoint the random variables $\mu(C_1), \ldots, \mu(C_r)$ are independent. (That μ is simple just means that a.s. all atoms of μ have mass 1, $P \bigcap_{t,y} (\mu(\{(t,y)\}) = 0 \text{ or } 1) = 1$).

5.4 Non-homogeneous Markov chains

Let X be a $(G, \mathcal{G})$ −valued step process with $X_0 \equiv x_0$, defined on $(\Omega, \mathcal{F}, P)$. Let $E = G$ and identify X with the RCM μ with mark space G determined by the MPP $(T_n, Y_n)_{n \geq 1}$, where T_n is the time of the n'th jump for X and $Y_n = X_{T_n}$ is the state recahed by that jump when $T_n < \infty$. Thus, if $Y_0 \equiv x_0$,

$$X_t = Y_{\overline{N}_t}.$$

If $Q^\circ = \mu(P)$ is the distribution of μ, we want to discuss the structure of the compensating measure L° for Q° when X is a (in general non-homogeneous) Markov chain. More precisely, we shall find sufficient conditions on the Markov kernels $P^{(n)}$, $\pi^{(n)}$ determining Q° which ensure that X is Markov with respect to the filtration $\left(\mathcal{F}_t^X\right)$.

First some remarks on the definition and basic properties of general Markov processes.

Let $(\mathcal{F}_t)_{t \geq 0}$ be a filtration and let $X^* = (X_t^*)_{t \geq 0}$ be an arbitrary measurable and adapted process defined on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ with values in $(G, \mathcal{G})$. We shall assume that one-point sets in G are measurable: $\{x\} \in \mathcal{G}$ for $x \in G$.

Definition 5.4.1 *The process X^* is a* Markov process *with respect to the filtration $(\mathcal{F}_t)$ if for every $s \leq t$ there exists a Markov kernel $p_{st}(\cdot, \cdot)$ from G to $\mathcal{G}$ such that*

$$P\left(X_t^* \in C \,|\mathcal{F}_s\right) = p_{st}\left(X_s^*, C\right) \quad (C \in \mathcal{G}). \tag{5.22}$$

X^* is a time-homogeneous Markov process *if in addition one may choose the p_{st} to depend on (s, t) through the difference $t - s$ only.*

The Markov kernel p_{st} is called the *transition probability* from time s to time t. Note that (5.22) does not determine $p_{st}(x, \cdot)$ uniquely for all $x \in G$, but that it does hold that if p_{st}, $\widetilde{p}_{st}$ are both transition probabilities, then $P(X_s^* \in C_{st}) = 1$, where $C_{st} = \{x \in G | p_{st}(x, \cdot) = \widetilde{p}_{st}(x, \cdot)\}$. Note also that one may always take $p_{ss}(x, \cdot) = \varepsilon_x$.

In the time-homogeneous case we write p_t for any of the transition probabilities $p_{s,s+t}$ with $s, t \geq 0$.

A time-homogeneous Markov process is also called a Markov process with *stationary transition probabilities.*

It is customary to call a Markov process, which is a step process, a *Markov chain.*

Example 5.4.1 *A $\mathbb{R}^d-$valued process X with independent increments (i.e. such that $X_t - X_s$ is independent of $\mathcal{F}_s$ for $s \le t$, cf. Definition 5.2.2) with respect to $(\mathcal{F}_t)$, is a Markov process with transition probabilities*

$$p_{st}(x, B) = \nu_{st}(B - x) \quad (s \le t\ ,\ B \in \mathcal{B}^d), \tag{5.23}$$

where ν_{st} is the distribution of $X_t - X_s$ and $B - x = \{x' - x | x' \in B\}$. If in addition X has stationary increments, X becomes a time-homogeneous Markov process with transition probabilities

$$p_t(x, B) = \nu_t(B - x), \tag{5.24}$$

where ν_t is the distribution of any increment $X_{s+t} - X_s$.

Suppose X^* is Markov with transition probabilities (p_{st}), or, in the homogeneous case, (p_t). We say that the transition probabilities satisfy the *Chapman-Kolmogorov equations* if for all $s \le t \le u,\ x \in G,\ C \in \mathcal{G}$,

$$p_{su}(x, C) = \int_G p_{st}(x, dx')\, p_{tu}(x', C). \tag{5.25}$$

or, in the homogeneous case, if for all $s, t \ge 0,\ x \in G,\ C \in \mathcal{G}$,

$$p_{s+t}(x, C) = \int_G p_s(x, dx')\, p_t(x', C). \tag{5.26}$$

It is essential to note that e.g. (5.25) holds almost surely in the following sense for any Markov process with transition probabilities (p_{st}): for $s \le t \le u,\ C \in \mathcal{G}$,

$$\begin{aligned} p_{su}(X_s^*, C) &= P(X_u^* \in C | \mathcal{F}_s) \\ &= E(P(X_u^* \in C | \mathcal{F}_t) | \mathcal{F}_s) \\ &= E(p_{tu}(X_t^*, C) | \mathcal{F}_s) \\ &= \int_G p_{st}(X_s^*, dx')\, p_{tu}(x', C), \end{aligned}$$

where the equalities for each given s, t, u, hold $P-$almost surely. Thus, given $s \le t \le u,\ C \in \mathcal{G}$, (5.25) holds for $X_s^*(P)-$almost all x.

We need to mention two more important facts about general Markov processes.

Let $\nu_0 = X_0^*(P)$ denote the distribution of X_0^*. Then the finite-dimensional distributions for the Markov process X^* are uniquely determined by ν_0 and the transition probabilities p_{st}. This follows by an induction argument, using that

$$\begin{aligned} P\left(X_t^* \in C\right) &= E\left(P\left(X_t^* \in C \,|Z_0\right)\right) \\ &= \int_G \nu_0(dx')\, p_{0t}(x', C) \end{aligned}$$

and that for $n \geq 2,\ 0 \leq t_1 < \cdots < t_n,\ C_1, \ldots, C_n \in \mathcal{G}$,

$$\begin{aligned} & P\left(X_{t_1}^* \in C_1, \ldots, X_{t_n}^* \in C_n\right) \\ = \quad & E\left(1_{\left(X_{t_1}^* \in C_1, \ldots, X_{t_{n-1}}^* \in C_{n-1}\right)} P\left(X_{t_n}^* \in C_n \,|\mathcal{F}_{t_{n-1}}\right)\right) \\ = \quad & E\left(1_{\left(X_{t_1}^* \in C_1, \ldots, X_{t_{n-1}}^* \in C_{n-1}\right)} p_{t_{n-1}t_n}(X_{t_{n-1}}^*, C_n)\right). \end{aligned}$$

The second fact we need is a generalization of the Markov property (5.22): for $t \geq 0$, let $\mathcal{F}^{t,X^*} = \sigma(X_u^* | u \geq t)$. Then for any $F \in \mathcal{F}^{t,X^*}$,

$$P(F|\mathcal{F}_t) = P(F|X_t^*), \tag{5.27}$$

a fact which informally may be phrased as follows: the future depends on the past only through the present.

(5.27) is proved considering $F \in \mathcal{F}^{t,X^*}$ of the form

$$F = \left(X_{t_1}^* \in C_1, \ldots, X_{t_n}^* \in C_n\right)\},$$

where $n \in \mathbb{N},\ t \leq t_1 < \cdots < t_n,\ C_1, \ldots, C_n \in \mathcal{G}$, and then proceeding by induction: the case $n = 1$ is just (5.22), and the induction step from $n-1$ to n is obtained using that

$$P(F|\mathcal{F}_t) = E(P(F|\mathcal{F}_{t_{n-1}})|\mathcal{F}_t).$$

A standard extension argument finally gives (5.27) for all $F \in \mathcal{F}^{t,X^*}$.

We now go back to the setup from the beginning of this section with X a step process and μ the RCM describing X.

The basic ingredient needed to make X a Markov chain (with respect to the filtration $(\mathcal{F}_t^X)$) is a collection of time dependent transition intensities, i.e. a function $q : \mathbb{R}_0 \times G \times \mathcal{G} \to \mathbb{R}_0$ such that

(i) $(t, x) \to q_t(x, C)$ is $\mathcal{B}_0 \otimes \mathcal{G}-$measurable for any $C \in \mathcal{G}$,

(ii) $C \to q_t(x, C)$ is a positive bounded measure on $(G, \mathcal{G})$ for any $t \in \mathbb{R}_0,\ x \in G$,

(iii) $q_t(x, \{x\}) = 0$ for all $t \in \mathbb{R}_0,\ x \in G$.

Here (iii) ensures that at each finite T_n, X performs a genuine jump, $X_{T_n} \neq X_{T_n-}$ a.s. which means in particular that ignoring a null set the filtrations agree, $\left(\mathcal{F}_t^X\right) = \left(\mathcal{F}_t^\mu\right)$.

Introduce $\overline{q}_t(x) = q_t(x, G)$. The next result in particular establishes formally the RCM construction of time-homogeneous Markov chains from Example 2.2.3.

Theorem 5.4.1 *A sufficient condition for the step process X to be Markov with respect to the filtration $(\mathcal{F}_t^X)$, is that there exist time dependent transition intensities $(q_t(x, C))$ such that the Markov kernels $P^{(n)}$, $\pi^{(n)}$ determining the distribution Q° of the RCM μ are given by*

$$\overline{P}^{(0)}(t) = \exp\left(-\int_0^t \overline{q}_s(x_0)\,ds\right), \quad \pi_t^{(0)}(C) = \frac{q_t(x_0, C)}{\overline{q}_t(x_0)},$$

and for $n \in \mathbb{N},\ t_1 < \cdots < t_n \le t,\ y_1, \ldots, y_n \in G$,

$$\overline{P}_{z_n}^{(n)}(t) = \exp\left(-\int_{t_n}^t \overline{q}_s(y_n)\,ds\right), \quad \pi_{z_n,t}^{(n)}(C) = \frac{q_t(y_n, C)}{\overline{q}_t(y_n)}$$

with $t > t_n$ in the last identity.

If q_t does not depend on t, then X is a time-homogeneous Markov chain.

Proof. The proof relies on Lemma 3.3.3. Let $t > 0$ and consider the conditional distribution of $\theta_s X := (X_u)_{u \ge s}$ given $\mathcal{F}_s^X$. We want to show that it depends on the past $\mathcal{F}_s^X$ through X_s only. Conditioning on $\mathcal{F}_s^X$ amounts to conditioning on $\overline{N}_s = k,\ T_1 = t_t, \ldots, T_k = t_k,\ Y_1 = y_1, \ldots, Y_k = y_k$ for some $k \in \mathbb{N}$, $0 < t_1 < \cdots < t_k \le s,\ y_1, \ldots, y_k \in G$. Since also $\theta_s X$ is determined by $\theta_s \mu$, the restriction of μ to $]s, \infty[\, \times E$, cf. p.37, and since $X_s = y_k$ on the set of conditioning, the desired Markov property for X follows, using the lemma, if we show that

$$\overline{P}_{z_k}^{(k)}(t)/P_{z_k}^{(k)}(s)$$

for $t \ge s$,

$$\pi_{z_k,t}^{(k)}$$

for $t > s$ and

$$P_{\mathrm{join}(z_k,\widetilde{z}_n)}^{(k+n)}, \quad \pi_{\mathrm{join}(z_k,\widetilde{z}_n),t}^{(k+n)}$$

for $n \in \mathbb{N},\ \widetilde{z}_n = \left(\overline{t}_1, \ldots, \overline{t}_n; \overline{y}_1, \ldots, \overline{y}_n\right)$ with $t < \widetilde{t}_1 < \cdots < \widetilde{t}_n < t,\ \widetilde{y}_1, \ldots, \widetilde{y}_n \in G$, all of them depend on $z_k = (t_1, \ldots, t_k; y_1, \ldots, y_k)$ only through y_k. But this is immediate using the explicit expressions for the $P^{(n)}$, $\pi^{(n)}$ from the statement of the theorem.

The time-homogeneous case follows by noting that when q_t does not depend on t, the conditional distribution of $\widetilde{\mu}_s$ given $\mathcal{F}_s^\mu$ depends on s and X_s through X_s only. Here

$$\widetilde{\mu}_s = \sum_{n:s<T_n<\infty} \varepsilon_{(T_n - s, Y_n)}$$

is $\theta_s \mu$ translated backwards in time to start at time 0. ■

With the setup used in Theorem 5.4.1, the RCM μ determining the Markov chain X, has a compensating measure L, which has a predictable intensity process $\lambda = (\lambda(C))_{C \in \mathcal{G}}$ (recall that $L([0,t] \times C) = \int_0^t \lambda_s(C) ds$) given by

$$\lambda_t(C) = q_t(X_{t-}, C),$$

as follows from Proposition 3.4.1. Note that

$$\lambda_t(C) = q_t(X_{t-}, C \backslash \{X_{t-}\})$$

and that if $t \to \lambda_t(C)$ has limits from the right, it follows from Proposition 3.4.2 (b), using (5.27) that

$$\begin{aligned} \lambda_{t+}(C) &= q_{t+}(X_t, C \backslash \{X_t\}) \\ &= \lim_{h \downarrow 0} \frac{1}{h} P(T_{t,1} \leq t + h,\ X_{t,1} \in C \backslash \{X_t\} | X_t), \end{aligned}$$

where $X_{t,1}$ is the state reached by X at the time of the first jump strictly after t and $T_{t,1}$ is the time of that jump. Typically the right hand side equals $\lim \frac{1}{h} p_{t,t+h} (X_t, C \backslash \{X_t\})$ and the identity may be written

$$q_{t+}(X_t, C) = \lim_{h \downarrow 0} \frac{1}{h} \left(p_{t,t+h}(X_t, C) - \varepsilon_{X_t}(C) \right),$$

the expression usually associated with the concept of transition intensities. (The diagonal intensities $q_t(x, x)$ in Markov chain theory are defined as $\lim \frac{1}{h} (p_{t,t+h}(x, \{x\}) - 1) = -\overline{q}_t(x)$).

It is perfectly possible to have Markov chains without transition intensities as described above. The proof of Theorem 5.4.1 carries over to the case where the Markov kernels $P^{(n)}$, $\pi^{(n)}$ have the form

$$\begin{aligned} &\overline{P}^{(0)}(t) = \overline{F}_{x_0}(t), && \pi_t^{(0)}(C) = r_t(x_0, C), \\ &\overline{P}_{z_n}^{(n)}(t) = \tfrac{\overline{F}_{y_n}(t)}{\overline{F}_{y_n}(t_n)}, && \pi_{z_n,t}^{(n)}(C) = r_t(y_n, C). \end{aligned}$$

where for each $x \in G$, F_x is the distribution function for a probability on $\overline{\mathbb{R}}_+$ with $\overline{F}_x = 1 - F_x$ the corresponding survivor function, and where for each t, r_t is a transition probability on G, such that $r_t(x, \{x\}) = 0$ for all x (corresponding to $q_t(x_0, C)/\overline{q}_t(x_0)$ from Theorem 5.4.1).

For the expression for $P^{(n)}$ to make sense, it is natural to assume that $\overline{F}_x(t) > 0$ for all $t \in \mathbb{R}_0$. Even this restriction can be omitted by using families $\left(\overline{F}_{x|s}\right)_{x \in G, s \geq 0}$ of survivor functions, with each $\overline{F}_{x|s}$ the survivor function for a probability on $]s, \infty]$, consistent in the sense that if $s < t$ and $\overline{F}_{x|s}(t) > 0$, then

$$\frac{1}{\overline{F}_{x|s}(t)} \overline{F}_{x|s} = \overline{F}_{x|t} \quad \text{on }]t, \infty] .$$

One then defines

$$\overline{P}^{(n)}_{z_n}(t) = \overline{F}_{y_n|t_n}(t).$$

5.5 The basic models from survival analysis

Let $X_1, \ldots, X_r$ be $\overline{\mathbb{R}}_+$-valued random variables, to be thought of as the failure times of r different items.

We shall assume that the X_i are independent and that the distribution of X_i has a Lebesgue density corresponding to the hazard function u_i,

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

Define the counting process N^i by

$$N_t^i = 1_{(X_i \leq t)}.$$

Then N^i has at most one jump and has $\mathcal{F}_t^{N^i}$-compensator $\Lambda_t^i = \int_0^t \lambda_s^i ds$, where

$$\begin{aligned} \lambda_s^i &= u_i(s) 1_{(X_i \geq s)} \\ &= u_i(s) 1_{\left(N_{s-}^i = 0\right)}. \end{aligned}$$

The aggregate μ of $(N^1, \ldots, N^r)$ is the RCM with mark space $E = \{1, \ldots, r\}$ such that

$$\mu([0,t] \times A) = N_t(A) = \sum_{i \in A} N_t^i \quad (A \subset E).$$

(Since the X_i have continuous distributions and are independent, (5.3) is satisfied).

By Theorem 5.1.1 (a), the $\mathcal{F}_t^\mu$−compensating measure for μ is given by the compensators $\Lambda(A) = \sum_{i\in A} \Lambda^i$, with Λ^i as above.

It is of particular interest to consider the case where the X_i are identically distributed, $u_i = u$ for $1 \le i \le r$. With $\overline{N} = \sum_{i=1}^r N^i$ (the counting process recording the total number of failures observed at a given time), the $\mathcal{F}_t^\mu$−compensator for $\overline{N}$ is $\overline{\Lambda} = \sum_{i=1}^r \Lambda^i = \int_0^\cdot \overline{\lambda}_s\, ds$, where

$$\overline{\lambda}_s = u(s)R_{s-}$$

with R_{s-} the number of *items at risk* just before time s,

$$R_{s-} = \sum_{i=1}^r 1_{\left(N_{s-}^i = 0\right)} = \sum_{i=1}^r (1 - N_{s-}^i) = r - \overline{N}_{s-}.$$

Notice in particular that $\overline{M} := \overline{N} - \overline{\Lambda}$ is a $\mathcal{F}_t^\mu$−martingale. Since however $\overline{\lambda}_s$ depends on s and $\overline{N}_{s-}$ alone, $\overline{\Lambda}$ is also $(\mathcal{F}_t^{\overline{N}})$−predictable and the $(\mathcal{F}_t^{\overline{N}})$−compensator for $\overline{N}$. It follows therefore, that if Q° is the distribution (on $(W, \mathcal{H})$) of $\overline{N}$, the compensator Λ° for Q° is $\Lambda_t^\circ = \int_0^t \lambda_s^\circ ds$, where

$$\lambda_s^\circ = u(s)(r - N_{s-}^\circ)^+ \tag{5.28}$$

with the $^+$−sign added to give an intensity process defined on all of W, which is everywhere ≥ 0.

From the discussion in Section 5.4 it follows that the fact that $\overline{\lambda}_s$ is a function of s and $\overline{N}_{s-}$ only, also implies that $\overline{N}$ is a non-homogeneous Markov chain, with state space $\{0, 1, \ldots, r\}$, initial state $\overline{N}_0 \equiv 0$. The transition probabilities are

$$p_{st}(k, n) = P(\overline{N}_t = n | \overline{N}_s = k),$$

non-zero only if $0 \le k \le n \le r,\ 0 \le s \le t$, and may be computed explicitly: let

$$\overline{F}(t) = \exp\left(-\int_0^t u(s)\, ds\right)$$

be the survivor function for the distribution of the X_i, $F = 1 - \overline{F}$ the distribution function. Then for $k \le n,\ s \le t$,

$$\begin{aligned}
p_{st}(k, n) &= P\left(X_i \in]0, t] \text{ for } n \text{ values of } i \,|X_i \in]0, s] \text{ for } k \text{ values of } i\right) \\
&= \frac{\binom{r}{k,n,r-n}(F(s))^k(F(t) - F(s))^{n-k}(\overline{F}(t))^{r-n}}{\binom{r}{k}(F(s))^k(\overline{F}(s))^{r-k}} \\
&= \binom{r-k}{r-n}\left(1 - \exp\left(-\int_s^t u\right)\right)^{n-k} \exp\left(-(r-n)\int_s^t u\right).
\end{aligned}$$

Thus, conditionally on $\overline{N}_s = k,\ \overline{N}_t - k$ follows a binomial distribution $(r - k,\ 1 - \exp\left(-\int_s^t u\right))$.

The transition probabilities are also determined by the time dependent transition intensities

$$q_t(k,n) = \lim_{h \downarrow 0} \frac{1}{h}(p_{t,t+h}(k,n) - \delta_{kn}),$$

(at least when e.g. u is continuous, so the limits exists), where the only non-zero intensities are $q_t(k,k+1) = -q_t(k,k)$, and where

$$q_t(k,k+1) = (r-k)u(t)$$

as is seen by either using the expressions for the transition probabilities directly or referring to Proposition 3.4.2 (a), (5.28) and the fact that

$$Q^\circ(N^\circ_{t+h} - N^\circ_t \geq 1 | \mathcal{H}_t) = 1 - p_{t,t+h}(N^\circ_t, N^\circ_t).$$

What we have discussed so far is the most elementary of all models in survival analysis. An important generalization arises by considering models for *right-censored survival data.* It is still assumed that the X_i are independent with hazard functions u_i, but due to right-censoring, not all X_i are observed. Formally, apart from $X_1, \ldots, X_r$, we are also given the censoring variables $V_1, \ldots, V_r$, which are $\overline{\mathbb{R}}_+-$valued random variables. What is observed are the pairs (S_i, δ_i) of random variables where

$$S_i = X_i \wedge V_i, \quad \delta_i = 1_{(X_i \leq V_i)}.$$

If $\delta_i = 1,\ X_i$ is observed, while if $\delta_i = 0$, the censoring time V_i is observed, and all that is known about the unobserved failure time X_i is that it exceeds (is strictly larger than) V_i.

For each i, introduce a counting process N^i with at most one jump,

$$N^i_t = 1_{(X_i \leq t \wedge V_i)}.$$

Thus N^i has one jump precisely when $X_i < \infty$ and X_i is observed, and no jump if either $X_i = \infty$ or X_i is not observed. If there is a jump, it occurs at time X_i.

So far we have said nothing about the joint distribution of the V_i, either on their own or jointly with the X_i. Now let $\mathcal{F}_t$ be the $\sigma-$algebra generated by the failures and censorings observed to occur in $[0,t]$. Formally,

$$\mathcal{F}_t = \sigma(S_i 1_{(S_i \leq t)},\ \delta_i 1_{(S_i \leq t)})_{1 \leq i \leq r}.$$

(Note that $(S_i > t) \in \mathcal{F}_t$: since $S_i > 0$, $S_i 1_{(S_i \le t)} = 0$ iff $S_i > t$). What is usually assumed about a model for right-censored survival data, is then that for every i, M^i is a $\mathcal{F}_t-$martingale, where

$$M_t^i = N_t^i - \int_0^t \lambda_s^i ds, \qquad \lambda_s^i = u^i(s) I_s^i, \tag{5.29}$$

that $(M^i)^2 - \lambda^i$ is a martingale, and finally that $M^i M^j$ is a martingale for $i \neq j$. Here I^i in (5.29) is the $\mathcal{F}_t-$predictable indicator

$$I_s^i = 1_{(S_i \ge s)},$$

which is one when item i is still at *risk at time* s, i.e. just before s, i had as yet neither been censored nor observed to fail.

There are several comments to be made on these martingale assumptions. The first is that even though it is always assumed that the X_i are independent with hazard functions u_i, requiring that M^i be a martingale is not enough to specify the joint distribution of $(X_1, \ldots, X_r; V_1, \ldots, V_r)$. To see why this is so, think of the observations (S_i, δ_i) as a MPP $(\mathcal{T}, \mathcal{Y})$, where the T_n when finite are the distinct finite values of the S_i ordered according to size, and where Y_n, if $T_n < \infty$, lists either the item i observed to fail at time T_n (if any), and/or those items j observed to be censored at T_n. (Without any assumptions on the structure of the censoring pattern, it is perfectly possible for the censorings of several items to occur simultaneously, and even to coincide with the observed failure time of some other item). As mark space we could use

$$E := \{(i, A) : 1 \le i \le r,\ A \subset \{1, \ldots, r\} \backslash \{i\}\} \cup \{(c, A) : \emptyset \neq A \subset \{1, \ldots, r\}\},$$

where the mark (i, A) means 'i observed to fail, all $j \in A$ censored' and the pure censoring mark (c, A) means 'no failures observed, all $j \in A$ censored'. With this setup the filtration generated by $(\mathcal{T}, \mathcal{Y})$ is precisely $(\mathcal{F}_t)$ and we know that to specify the distribution of the MPP, one must write down the $\mathcal{F}_t-$compensators for all the counting processes N^y, $N_t^y = \sum_{n=1}^{\infty} 1_{(T_n \le t, Y_n = y)}$ for arbitrary $y \in E$. All that is done in (5.29) is to give the compensator for

$$N^i = \sum_{A:(i,A) \in E} N^{(i,A)},$$

so not even the distribution of the observations (S_i, δ_i) is determined by (5.29), much less of course the joint distribution of all (X_i, V_i). The message is that with (5.29) and the assumption that the X_i are independent with hazards u_i, the model for censored survival data is only *partially specified.*

Note that the assumptions about the martingale structure of $(M^i)^2$ and M^iM^j presented after (5.29) are direct consequences of (5.29) and Proposition 3.5.3 (b), so are not new conditions.

In the model without censoring discussed initially we have $V_1 \equiv V_2 \equiv \cdots V_r \equiv \infty$, and then $I_s^i = 1_{(X_i \geq s)}$ and as we saw earlier, (5.29) is in fact satisfied.

We shall give some other examples of censoring models, where (5.29) holds.

Suppose first that for $1 \leq i \leq r$, $V_i \equiv v_i$, where $v_i \in]0, \infty]$ is a given constant. Observing all (S_i, δ_i) then amounts to observing the independent counting processes N^i,

$$N_t^i = 1_{(X_i \leq t \wedge v_i)}.$$

With T_1^i the first (and only) jump time for N^i,

$$P(T_1^i > t) = \overline{F}_i(t \wedge v_i),$$

$\overline{F}_i$ denoting the survivor function $\overline{F}_i(t) = \exp(-\int_0^t u_i)$ for the distribution of X_i. Thus the distribution of T_1^i has hazard function

$$u^{i,(0)}(t) = \begin{cases} u_i(t) & (t \leq v_i) \\ 0 & (t > v_i) \end{cases}$$

and the compensator for N^i has $\mathcal{F}_t$–predictable intensity process

$$\lambda_t^i = u_i(t) 1_{[0,v_i]}(t) 1_{\left(N_{t-}^i = 0\right)} = u_i(t) I_t^i,$$

i.e. (5.29) holds.

As a second example, suppose that for each i, X_i and V_i are independent, that the different pairs (X_i, V_i) are independent and that V_i for all i has a distribution with Lebesgue density g_i. Identify any given observable pair (S_i, δ_i) with a MPP described by the two counting processes

$$N_t^i = 1_{(X_i \leq t \wedge V_i)}, \quad N_t^{c,i} = 1_{(V_i \leq t, V_i < X_i)},$$

(which makes sense since $P(X_i = V_i < \infty) = 0$). Then N^i, $N^{c,i}$ combined have at most one jump in all, occurring at time S_i. Hence, to find the compensating measure (with respect to $(\mathcal{F}_t^{N^i, N^{c,i}})$), we need only find the hazard function $u^{i,(0)}$ for the distribution of S_i, and the conditional jump distribution

$$\rho_i(t) := \pi_t^{i,(0)}(\{i\}) = P(X_i \leq V_i | S_i = t).$$

Clearly, if u_i^c is the hazard function for g_i,

$$u^{i,(0)}(t) = u_i(t) + u_i^c(t), \tag{5.30}$$

while ρ_i is determined by the condition

$$\int_B \rho_i(t)\, P(S_i \in dt) = P(X_i \le V_i,\ S_i \in B) \quad (B \in \mathcal{B}_0).$$

But (f_i denoting the density for X_i, $\overline{G}_i$ the survivor function for V_i)

$$\begin{aligned} P(X_i \le V_i,\ S_i \in B) &= P\left(X_i \le V_i,\ X_i \in B\right)) \\ &= \int_B \overline{G}_i(x_i) f_i(x_i)\, dx_i, \end{aligned}$$

and since $P(S_i > t) = \overline{F}_i(t)\overline{G}_i(t)$, S_i has density $\overline{F}_i g_i + f_i\overline{G}_i$ and so

$$\int_B \rho_i(t)\, P(S_i \in dt) = \int_B \rho_i(t) \left(\overline{F}_i g_i + f_i \overline{G}_i\right)\, dt.$$

Consequently

$$\rho_i = \frac{f_i\overline{G}_i}{\overline{F}_i g_i + f_i\overline{G}_i} = \frac{u_i}{u_i + u_i^c}$$

(since $f_i = u_i\overline{F}_i$, $g_i = u_i^c\overline{G}_i$), and comparing with (5.30) it is seen that the intensity process λ^i for the $\mathcal{F}_t^{N^i,N^{c,i}}-$compensator for N^i is given by (5.29). Because of the independence of the pairs (X_i, V_i), using Theorem 5.1.1 (a) it follows that (5.29) holds (with $(\mathcal{F}_t)$ the filtration generated by $(N^i, N^{c,i})_{1\le i\le r}$).

It is in fact true that (5.29) holds under much more general independence assumptions: it is enough that $X_1, \dots, X_r$ are independent with hazards u_i, and that the random vector $V = (V_1, \dots, V_r)$ is independent of $(X_1, \dots, X_r)$. To see this, consider first the conditional distribution of the counting processes N^i given $V = v$, $v = (v_1, \dots, v_r)$. Since V is independent of the X_i this is the same as the distribution of the N^i, assuming the censoring times V_i to be identically equal to the given constants v_i, a situation where we have already seen that (5.29) holds, and have then found the compensators for the N^i. A small amount of work shows that this amounts to saying that the N^i (with V random and independent of the X_i) have compensators

$$\int_0^{\cdot} \widetilde{\lambda}_s^i\, ds, \quad \widetilde{\lambda}_s^i = \mu_i(s) I_i(s)$$

with respect to the filtration $(\widetilde{\mathcal{F}}_t)$, where $\widetilde{\mathcal{F}}_t = \sigma(V, N_s^i)_{0\le s\le t, 1\le i\le r}$. But $\mathcal{F}_t \subset \widetilde{\mathcal{F}}_t$ and since the compensators are $\mathcal{F}_t-$predictable, we deduce that (5.29) is satisfied.

A different example of a model for right-censoring for which (5.29) holds: let $X_1, \ldots, X_r$ be independent with hazard function u_i for X_i. Define $V_1 = \cdots = V_r = X_{(m)}$ with $X_{(1)} \le X_{(2)} \le \cdots \le X_{(r)}$ the order statistics. (Since the X_i have continuous distributions, the finite X_i-values are distinct). Clearly, there is a strong dependence between the X_i and the V_i. Nevertheless it is easy to see that (5.29) holds: start with $\widetilde{N}_t^i = 1_{(X_i \le t)}$, and let $\widetilde{\mu} = \sum_{n:\widetilde{T}_n<\infty} \varepsilon_{(\widetilde{T}_n, \widetilde{Y}_n)}$ denote the RCM with mark space $\{1, \ldots, r\}$ determined from the $\widetilde{N}^i$. The observations $(S_i, \delta_i)_{1 \le i \le r}$ arising from censoring at $X_{(m)}$ are identified with the counting processes

$$N_t^i = 1_{\left(X_i \le t, X_i \le X_{(m)}\right)},$$

corresponding to a RCM $\mu = \sum_{n:T_n<\infty} \varepsilon_{(T_n, Y_n)}$ with mark space $\{1, \ldots, r\}$. Clearly $(T_k, Y_k) = (\widetilde{T}_k, \widetilde{Y}_k)$ for $k \le m$, $T_{m+1} = T_{m+1} = \cdots = \infty$. So the Markov kernels determining μ agree with those of $\widetilde{\mu}$ up to and including the time of the m'th jump. We know the compensators for $\widetilde{\mu}$ and deduce the desired structure for those of μ.

In conclusion of this discussion of censoring models we note that there are lots of other censoring problems for which (5.29) holds (always assuming that the X_i are independent with hazards u_i), without V and the X_i being independent.

A different question, that can be given an affirmative answer, is the following: suppose only that the joint distribution of the (S_i, δ_i) satisfies the martingale property (5.29) for all i, but do not otherwise suppose anything about the distribution of the X_i. Is this assumption always compatible with the requirement that the X_i be independent with hazard functions u_i?

One of the most versatile of all models used in survival analysis is the *Cox regression model*. Here the intensity for failure of an item is allowed to depend on an observable process of *covariates* which contains information about different characteristics for that item.

The model is typically only partially specified, describing the intensity for failure, but not the distribution of the covariates. The model allows for censoring.

Formally, suppose given a filtered space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ and let for $1 \le i \le r$, ζ^i be an $\mathbb{R}^p-$valued $\mathcal{F}_t-$predictable process (the covariate process for i represented as a column vector), let $V_i \in]0, \infty]$ be the censoring time, X_i the failure time for item i, and assume that the counting process N^i,

$$N_t^i = 1_{(X_i \le t \wedge V_i)}$$

is $\mathcal{F}_t-$adapted, and that the $\{0,1\}-$valued process I^i,

$$I_t^i = 1_{(t \leq X_i \wedge V_i)}$$

is $\mathcal{F}_t-$predictable.

The fundamental assumption in the Cox model is then that for every i the $\mathcal{F}_t-$compensator for N^i has $\mathcal{F}_t-$predictable intensity λ^i,

$$\lambda_t^i = u(t) e^{\beta^T \zeta_t^i} I_t^i, \tag{5.31}$$

where u is the hazard function for some distribution on $\overline{\mathbb{R}}_+$ and $\beta^T = (\beta_1, \ldots, \beta_p)$ is a row vector of regression parameters. The 'model' arises by allowing the *baseline hazard* u and the $\beta-$parameters to vary.

Note that if $\beta = 0$ we get a model for right-censoring as discussed above (with all $u_i = u$).

In the case of no censoring (all $V_i \equiv \infty$) and each $\zeta_t^i = \varsigma^i(t)$ a given, non-random function of t, take $(\mathcal{F}_t)$ to be the filtration generated by $(N^i)_{1 \leq i \leq r}$. It is then clear that $X_1, \ldots, X_r$ are independent such that for $1 \leq i \leq r$, the distribution of X_i has hazard function

$$u_i(t) = u(t) e^{\beta^T \varsigma^i(t)}.$$

In general with random covariates no such expression is valid for the failure time hazards and the best interpretation of the Cox model is through expressions like (cf. Proposition 3.4.2),

$$u(t+) e^{\beta^T \zeta_{t+}^i} = \lim_{h \downarrow 0} \frac{1}{h} P(t < X_i \leq t + h | \mathcal{F}_t) \quad \text{on the set } \ (X_i \wedge V_i > t)$$

describing the imminent risk of failure.

With only the intensities (5.31) given of course hardly anything is said about the distribution of the censoring times and covariates. With all the ζ^i e.g. piecewise deterministic processes determined from certain RCM's, a full specification of the model could be given involving not only the failure intensities (5.31) but also censoring intensities (possibly jointly or jointly with a failure) and intensities describing the jumps of the ζ^i. Such specifications are never given in statistical practice which makes model based prediction of mean survival times and other such relevant quantities virtually impossible in the Cox regression model!

Chapter 6

Piecewise deterministic Markov processes

6.1 Definition and construction

Markov chains are Markov processes that are piecewise constant with only finitely many jumps on finite time intervals. They were treated in Example 2.2.3 (time-homogeneous chains) and in greater generality and detail in Section 5.4 and are special cases of the class of Markov processes we shall now discuss, which in turn are special cases of the piecewise deterministic processes introduced p.19. For some of the basic properties of Markov processes, see Section 5.4.

So let $X = (X_t)_{t \geq 0}$ be a process with state space $(G, \mathcal{G})$, where G is a topological space such that $\{x\} \in \mathcal{G}$ for all $x \in G$, assume that $X_0 \equiv x_0$ for some given $x_0 \in G$, that X has only finitely many discontinuities on finite time intervals, and that with T_n the time of the n'th jump (discontinuity), Y_n the state reached by the jump,

$$X_t = f^{(\overline{N}_t)}_{Z_t|x_0}(t) \tag{6.1}$$

with $\overline{N}_t$ the total number of jumps on $[0, t]$, $Z_t = \left(T_1, \ldots, T_{\overline{N}_t}; Y_1, \ldots, Y_{\overline{N}_t}\right)$ and with each $t \to f^{(n)}_{z_n|x_0}(t)$ a *continuous* function assumed to satisfy that $f^{(n)}_{z_n|x_0}(t)$ is jointly measurable in the arguments x_0, $z_n = (t_1, \ldots, t_n; y_1, \ldots, y_n)$ and t. Finally, the $f^{(n)}$ satisfy the boundary conditions

$$f^{(0)}_{|x_0}(0) = x_0, \qquad f^{(n)}_{z_n|x_0}(t_n) = y_n.$$

The process X, piecewise continuous by assumption, is identified with the MPP (T_n, Y_n) with mark space $(G, \mathcal{G})$, and with the corresponding RCM

μ. It is utterly boring but perfectly possible to show that (ignoring a null set if necessary) the random variables (T_n, Y_n) are $\mathcal{F}^X-$measurable ($\mathcal{F}^X = \sigma(X_t)_{t\geq 0}$) and that the filtrations $(\mathcal{F}_t^X)$ and $(\mathcal{F}_t^\mu)$ agree, at least if (6.2) below holds. (For instance, that T_1 is $\mathcal{F}^X-$measurable follows from

$$(T_1 > t) = \bigcap_{q\in\mathbb{Q}_0, q\leq t} \left(X_q = f_{x_0}^{(0)}(q)\right)).$$

Note that by (6.1) and Proposition 3.2.1 (bii), it is clear that X is $\mathcal{F}_t^\mu-$adapted.

The distribution of X is determined by that of μ and its Markov kernels $P_{z_n|x_0}^{(n)}$ and $\pi_{z_n,t|x_0}^{(n)}$ (depending on x_0), through (6.1). However, the identification between X and μ and the filtrations $\left(\mathcal{F}_t^X\right)$ and $(\mathcal{F}_t^\mu)$ works only if, with the $f_{z_n|x_0}^{(n)}(t)$ continuous functions of t as above, the T_n are true jump times for X: on $(T_n < \infty)$ it must hold a.s. that $X_{T_n} \neq X_{T_n-}$. In terms of the Markov kernels generating the distribution of (T_n, Y_n) this means that

$$\pi_{z_n,t|x_0}^{(n)}\left(\left\{f_{z_n|x_0}^{(n)}(t)\right\}\right) = 0 \tag{6.2}$$

for almost all values (z_n, t) of (Z_n, T_{n+1}). (While always (6.1) determines X from μ, to go the other way one needs some mechanism for identifying the T_n – finding the $Y_n = X_{T_n}$ is then immediate. Here we focus on the T_n being the jump times for X, but in general one might just add to X the information about when the T_n occur, for instance by introducing states (x, mark) to signify the occurrence of a jump for μ. In that case one need not assume the $f_{z_n|x_0}^{(n)}$ to be continuous and X is nothing but a $\mathcal{F}_t^\mu-$adapted process, cf. Proposition 3.2.1. Also, as marks Y_n one need not use X_{T_n} – other possibilities are $Y_n = (X_{T_n-}, X_{T_n})$ or, in the case of $\mathbb{R}^d-$valued processes, $Y_n = \Delta X_{T_n}$ as was done in Section 5.2. We shall not discuss these generalizations further).

The distribution of X being described by the distribution of μ, the problem we shall now discuss is that of finding out what structure must be imposed on the Markov kernels $P_{z_n|x_0}^{(n)}$, $\pi_{z_n,t|x_0}^{(n)}$ *and* the functions $f_{z_n|x_0}^{(n)}$, in order for X to be a Markov process, and, in particular, a time-homogeneous Markov process. We shall furthermore impose the constraint that the transition probabilities of the Markov processes do not depend on the initial state x_0.

For $s \geq 0$, consider the conditional distribution of $\theta_s X = (X_t)_{t\geq s}$ given $\mathcal{F}_s^X$. By (5.27) X is Markov iff this depends on the past $\mathcal{F}_s^X$ through X_s only. Conditioning on $\mathcal{F}_s^X$ amounts to conditioning on $\overline{N}_s = k,\ Z_k = z_k = (t_1, \ldots, t_k; y_1, \ldots, y_k)$ for some $k \in \mathbb{N}_0,\ 0 < t_1 < \cdots < t_k \leq t,\ y_1, \ldots, y_k \in G$ (cf. Corollary 3.2.2). On the set $(\overline{N}_s = k, Z_k = z_k)$, by (6.1),

$$X_s = f_{z_k|x_0}^{(k)}(s), \tag{6.3}$$

and until the time $T_{s,1}$ of the first jump after s, X follows the deterministic function

$$t \to f^{(k)}_{z_k|x_0}(t) \qquad (t \geq s).$$

Copying the proof of Theorem 5.4.1 and referring to Lemma 3.3.3, it now follows that for X to be Markov with transitions that do not depend on x_0, it is sufficient that the following six quantities (for arbitrary $k, x_0, t_1, \ldots, t_k$, $y_1, \ldots, y_k$) depend on these $2k+2$ variables through X_s as given by (6.3) only:

$$\begin{aligned} & f^{(k)}_{z_k|x_0}(t) \qquad (t \geq s), \\ & \overline{P}^{(k)}_{z_k|x_0}(t)/\overline{P}^{(k)}_{z_k|x_0}(s) \qquad (t \geq s) \\ & \pi^{(k)}_{z_k,t|x_0} \qquad (t \geq s) \end{aligned} \tag{6.4}$$

and, for $n \in \mathbb{N}$, $s < \widetilde{t}_1 < \cdots < \widetilde{t}_n$, $\widetilde{y}_1, \ldots, \widetilde{y}_n \in G$ with $\widetilde{z}_n = (\widetilde{t}_1, \ldots, \widetilde{t}_n; \widetilde{y}_1, \ldots, \widetilde{y}_n)$

$$\begin{aligned} & f^{(k+n)}_{\text{join}(z_k,\widetilde{z}_n)|x_0} \\ & P^{(k+n)}_{\text{join}(z_k,\widetilde{z}_n)|x_0} \\ & \pi^{(k+n)}_{\text{join}(z_k,\widetilde{z}_n),t|x_0} \qquad \left(t > \widetilde{t}_n\right). \end{aligned}$$

Starting with (6.4), the requirement that this quantity depends on X_s only amounts to requiring that for some function ϕ_{st},

$$f^{(k)}_{z_k|x_0}(t) = \phi_{st}(f^{(k)}_{z_k|x_0}(s)). \tag{6.5}$$

Taking $s = t_k$ and recalling the boundary condition $f^{(k)}_{z_k|x_0}(t_k) = y_k$ gives

$$f^{(k)}_{z_k|x_0}(t) = \phi_{t_k,t}(y_k).$$

Inserting this general expression for $f^{(k)}$ in (6.5), and changing the notation a little gives

$$\phi_{su}(y) = \phi_{tu}(\phi_{st}(y)) \qquad (0 \leq s \leq t \leq u,\ y \in G), \tag{6.6}$$

which together with the boundary conditions

$$\phi_{tt}(y) = y \qquad (t \in \mathbb{R}_0) \tag{6.7}$$

are the basic functional equations describing the deterministic behavior of a piecewise deterministic Markov process: (6.1) becomes

$$X_t = \phi_{T_{\overline{N}_t},t}(Y_{\overline{N}_t}) \tag{6.8}$$

with $T_{\overline{N}_t} = 0$, $Y_{\overline{N}_t} = Y_0 \equiv x_0$ on $(\overline{N}_t = 0)$. More compactly (6.6) and (6.7) may be written

$$\phi_{su} = \phi_{tu} \circ \phi_{st} \quad (0 \leq s \leq t \leq u), \qquad \phi_{tt} = \mathrm{id} \quad (t \geq 0),$$

id denoting the identity on G. In particular the ϕ_{st} form a two-parameter semigroup under composition.

Having determined the structure of the piecewise deterministic part of the process, it is now an easy matter to prove the following result, where $D := \{(s,t,y) \in \mathbb{R}_0^2 \times G : s \leq t\}$ and $(t,y) \to q_t(y)$ is a measurable function from $\mathbb{R}_0 \times G$ to $\mathbb{R}_0$, while for every $t \in \mathbb{R}_0$, p_t is a transition probability (Markov kernel) on G such that $(t,y) \to p_t(y,C)$ is measurable for all C and such that for all t, y

$$p_t\left(y, \{y\}\right) = 0, \tag{6.9}$$

as is essential to ensure that X has a discontinuity at each finite T_n, cf. (6.2).

Theorem 6.1.1 *(a) Suppose $\phi : D \to G$ is a measurable function which satisfies (6.6), (6.7) and is such that $t \to \phi_{st}(y)$ is continuous on $[s, \infty[$ for all $s \in \mathbb{R}_0$, $y \in G$. Then the piecewise deterministic process X given by $X_0 = Y_0 \equiv x_0$ and*

$$X_t = \phi_{T_{\overline{N}_t},t}(Y_{\overline{N}_t})$$

is a Markov process with transition probabilities that do not depend on x_0, provided the Markov kernels $P^{(n)}_{z_n|x_0}$, $\pi^{(n)}_{z_n,t|x_0}$ determining the distribution of the RCM μ recording the jump times for X and the states reached by X at the time of each jump, are of the form

$$\overline{P}^{(0)}_{|x_0}(t) = \exp\left(-\int_0^t q_s(\phi_{0,s}(x_0))\,ds\right), \qquad (t \in \mathbb{R}_0)$$

$$\pi^{(0)}_{t|x_0}(C) = p_t(\phi_{0,t}(x_0), C) \qquad (t \in \mathbb{R}_0,\ C \in \mathcal{G})$$

and for $n \in \mathbb{N}$, $t_1 < \cdots < t_n \leq t$, $y_1, \ldots, y_n \in G$

$$\overline{P}^{(n)}_{z_n|x_0}(t) = \exp\left(-\int_{t_n}^t q_s(\phi_{t_n,s}(y_n))\,ds\right),$$

$$\pi^{(n)}_{z_n,t|x_0}(C) = p_t(\phi_{t_n,t}(y_n), C) \qquad (C \in \mathcal{G}),$$

with $t > t_n$ in the last identity.

(b) The piecewise deterministic Markov process X determined by $\phi_{tu}(y)$, $q_t(y)$, $p_t(y,C)$ is time-homogeneous with transition probabilities that do not depend on x_0, if there exists a measurable function $\widetilde{\phi} : \mathbb{R}_0 \times$

$G \to G$ with $\widetilde{\phi}_0(\cdot) = \mathrm{id}$ and $t \to \widetilde{\phi}_t(y)$ continuous on $\mathbb{R}_0$ for all y, a measurable function $\widetilde{q} : G \to \mathbb{R}_0$ and a transition probability $\widetilde{p}$ on G with $\widetilde{p}(y, \{y\}) = 0$, such that for all $s \le t,\ y \in G$

$$\phi_{st}(y) = \widetilde{\phi}_{t-s}(y), \quad q_t(y) = \widetilde{q}(y), \quad p_t(y, C) = \widetilde{p}(y, C).$$

Notation. In the sequel we always write ϕ, q, p rather than $\widetilde{\phi}, \widetilde{q}, \widetilde{p}$ in the time-homogeneous case.

Remark 6.1.1 *$q_t(y)$ is the intensity for a jump to occur at time t if the process at that time is in state y. $p_t(y, C)$ is interpreted as the conditional probability that a jump leads to a state in C, given that the jump occurs at time t from state y.*

Note that in the time-homogeneous case (6.6), (6.7) becomes

$$\phi_{s+t}(y) = \phi_s(\phi_t(y)) \quad (s, t \ge 0,\ y \in G), \qquad \phi_0(y) = y \quad (y \in G) \tag{6.10}$$

or, in compact form

$$\phi_{s+t} = \phi_s \circ \phi_t, \quad \phi_0 = \mathrm{id}\,.$$

Thus the ϕ_t form a one-parameter commutative semigroup under composition.

Proof. *(Theorem 6.1.1).* For part (a) we must show that (6.4) and the three quantities after that depend on k, x_0, z_k through $X_s = \phi_{t_k,s}(y_k)$ only. For (6.4) this follows from (6.5) and (6.6), cf. the argument on p.109 leading to (6.6). Next

$$\begin{aligned}
\frac{\overline{P}^{(k)}_{z_k|x_0}(t)}{\overline{P}^{(k)}_{z_k|x_0}(s)} &= \exp\left(-\int_s^t q_u(\phi_{t_k,u}(y_k))\,du\right) \\
&= \exp\left(-\int_s^t q_u(\phi_{su}(X_s))\,du\right), \\
\pi^{(k)}_{z_k,t|x_0} &= p_t(\phi_{t_k,t}(y_k)) \\
&= p_t(\phi_{st}(X_s)),
\end{aligned}$$

again using (6.6). As for the last three quantities, $f^{(k+n)}_{\mathrm{join}(z_k,\widetilde{z}_n)|x_0}$, $P^{(k+n)}_{\mathrm{join}(z_k,\widetilde{z}_n)|x_0}$, $\pi^{(k+n)}_{\mathrm{join}(z_k,\widetilde{z}_n),t|x_0}$, they depend on $\widetilde{t}_n$, $\widetilde{y}_n$ and (in the case of π) t only, in particular they do not depend on either of k, x_0 or z_k. This completes the proof of (a).

The proof of part (a) relies on an identification of the conditional distribution of $\theta_s X$ given $\mathcal{F}^X_s$. Here $s \in \mathbb{R}_0$ is given but arbitrary, so when saying that the conditional distribution depends on X_s only, we really mean that it depends on X_s *and* the constant s. To obtain a time-homogeneous X, this dependence on s must be eliminated and thus, to prove (b), one must show that when ϕ, q, p are of the form given in part (b), (6.4) and the four quantities after (6.4), which we already know to depend on s and $X_s = \phi_{t_k,s}(y_k) = \widetilde{\phi}_{s-t_k}(y_k)$ only, either do not depend on s or, when evaluating a certain quantity at a time-point $t \geq s$ depends on t through $t-s$ alone. But this is immediate since for instance

$$\begin{aligned} \frac{\overline{P}^{(k)}_{z_k}(t)}{\overline{P}^{(k)}_{z_k}(s)} &= \exp\left(-\int_s^t \widetilde{q}(\widetilde{\phi}_{u-s}(X_s))\,du\right) \\ &= \exp\left(-\int_0^{t-s} \widetilde{q}(\widetilde{\phi}_u(X_t))\,du\right) \end{aligned}$$

etc. (See also the last part of the proof of Theorem 5.4.1 for a more formal argument). ■

We shall discuss further the structure of PDMP's and begin with a discussion of the solutions to the homogeneous semigroup equation (6.10).

It is immediately checked that a general form of solutions are obtained by considering a continuous bijection $S : G \to \widetilde{G}$, where $\widetilde{G}$ is a topological vector space, and defining

$$\phi_t(y) = S^{-1}(S(y) + tv_0) \tag{6.11}$$

for some $v_0 \in \widetilde{G}$. In particular, if $G = \mathbb{R}^d$ one may take $\widetilde{G} = G$ with S a homeomorphism on $\mathbb{R}^d$.

Assuming that $G = \mathbb{R}$, it is also possible to obtain partial differential equations for solutions of (6.10). By assumption $t \to \phi_t(y)$ is continuous, in particular $\lim_{s\downarrow 0} \phi_s(y) = \phi_0(y) = y$. Suppose now that ϕ is differentiable at $t = 0$,

$$\lim_{s\downarrow 0} \frac{1}{s}(\phi_s(y) - y) = a(y)$$

exists as a limit in $\mathbb{R}$. Then for $t \in \mathbb{R}_0$, using (6.10)

$$\begin{aligned} D_t\phi_t(y) &= \lim_{s\downarrow 0} \frac{1}{s}(\phi_{s+t}(y) - \phi_t(y)) \\ &= \lim_{s\downarrow 0} \frac{1}{s}(\phi_s(y) - y)\frac{\phi_t(\phi_s(y)) - \phi_t(y)}{\phi_s(y) - y} \\ &= a(y)D_y\phi_t(y) \end{aligned}$$

so, assuming that the partial derivatives exist we arrive at the first order linear partial differential equation

$$D_t\phi_t(y) = a(y)D_y\phi_t(y) \tag{6.12}$$

with the boundary condition $\phi_0(y) = y$.

Becuse the ϕ_t commute under composition, a differential equation different from (6.12) is also available, viz.

$$D_t\phi_t(y) = \lim_{s\downarrow 0} \frac{1}{s}\left(\phi_s\left(\phi_t(y)\right) - \phi_t(y)\right)$$

resulting in the non-linear differential equation

$$D_t\phi_t(y) = a(\phi_t(y)). \tag{6.13}$$

Examples of solutions corresponding to different choices of a are essentially (apart from the first example and for the others, apart from possible problems with domains of definition) of the form (6.11) with S satisfying $v_0/S' = a$ (where now $v_0 \in \mathbb{R}$). Some examples where K is a constant:

(i) If $a \equiv 0$, $\phi_t(y) = y$, corresponding to the step process case.

(ii) If $a(y) = K$, then $\phi_t(y) = y + Kt$, yielding piecewise linear processes.

(iii) If $a(y) = Ky$, then $\phi_t(y) = ye^{Kt}$.

(iv) If $y > 0$ only and $a(y) = \frac{K}{y}$, then $\phi_t(y) = \sqrt{y^2 + 2Kt}$, a solution that cannot be used on all of $\mathbb{R}_0$ if $K < 0$. (However, it may still be possible to extend to a complete solution of (6.10) by forcing a jump when the $\phi_t(y)$ reaches the critical value 0).

(v) More generally, if $y > 0$ and $a(y) = Ky^\beta$, where $\beta \neq 1$, then $\phi_t(y) = \left(y^{1-\beta} + K\left(1-\beta\right)t\right)^{1/(1-\beta)}$.

(vi) If $a(y) = Ke^{-y}$, then $\phi_t(y) = \log(e^y + Kt)$.

A fairly general form of solutions to the functional equation (6.6) in the non-homogeneous case is obtained by recalling the standard space-time device: if X is non-homogeneous Markov, then (t, X_t) is time-homogeneous and piecewise deterministic with, trivially, the time component increasing linearly over time with slope 1. With $\widetilde{S}$ a homeomorphism from the state space $\mathbb{R}_0 \times G$ for (t, X_t) to $\mathbb{R}_0 \times \widetilde{G}$, with $\widetilde{G}$ a topological vector space, this makes it natural to look for $\widetilde{S}$ of the form $\widetilde{S}(t, y) = (t, S_t(y))$ with the deterministic part of (t, X_t) given by (6.11), using $\widetilde{S}$ instead of S, (s, y) instead of y and $(1, v_0)$ instead of v_0. The end result is that with $(S_t)_{t \geq 0}$ a family of homeomorphisms from G to $\widetilde{G}$, the functions $(\phi_{st})_{0 \leq s \leq t}$ given by

$$\phi_{st}(y) = S_t^{-1}\left(S_s(y) + (t - s)\, v_0\right)$$

for some $v_0 \in \widetilde{G}$ satisfy the non-homogeneous equation (6.6) and the boundary condition (6.7).

We next discuss intensity processes. Since all $P^{(n)}$ have hazard functions

$$u_{z_n|x_0}(t) = q_t(\phi_{t_n,t}(y_n)),$$

as a consequence the MPP $(T_n, Y_n)_{n \geq 1}$ has a $\mathcal{F}_t^{\mu}-$predictable intensity process λ, so that $\Lambda_t(C) = \int_0^t \lambda_s(C)ds$. Using Proposition 3.4.1 and the expressions for the $u^{(n)}$, $\pi^{(n)}$, it follows using (6.8) that

$$\lambda_t(C) = q_t(X_{t-})p_t(X_{t-}, C), \tag{6.14}$$

and in the time-homogeneous case,

$$\lambda_t(C) = q(X_{t-})p(X_{t-}, C). \tag{6.15}$$

The Markov property of X is reflected in the fact that $\lambda_t(C)$, which is $\mathcal{F}_t^X-$predictable, depends on the past $(X_s)_{0 \leq s < t}$ through X_{t-} only.

It is quite easy to generalize Theorem 6.1.1 to the case, where the $P^{(n)}$ do not have densities: imitating the proof one only has to verify that (6.4) and the quantities listed after that, depend on the past through X_s only. The result is

Theorem 6.1.2 *(a) Suppose $\phi : D \to G$ is a measurable function which satisfies (6.6), (6.7) and is such that $t \to \phi_{st}(y)$ is continuous on $[s, \infty[$ for all $s \in \mathbb{R}_0$, $y \in G$. The piecewise deterministic process X given by $X_0 \equiv x_0$ and*

$$X_t = \phi_{T_{\overline{N}_t},t}(Y_{\overline{N}_t})$$

is a Markov process with transitions that do not depend on x_0, provided the $\pi^{(n)}$ are as in Theorem 6.1.1, and if furthermore the $P^{(n)}$ are of the form

$$\overline{P}^{(n)}_{z_n|x_0}(t) = \overline{Q}_{t_n,y_n}(t) \qquad (t \geq t_n),$$

where for each s, y, Q_{sy} is a probability on $]s, \infty]$ with survivor function $\overline{Q}_{sy}$, and the Q_{sy} satisfy that for $s \leq t \leq u$, $y \in G$,

$$\overline{Q}_{sy}(u) = \overline{Q}_{sy}(t)\overline{Q}_{t,\phi_{st}(y)}(u).$$

(b) The Markov process X from (a) is time-homogeneous if $\phi_{st} = \phi_{t-s}$, $p_t(y, C) = p(y, C)$ and if in addition $\overline{Q}_{sy}(t)$ depends on s and t through $t - s$ only.

Remark 6.1.2 *Theorem 6.1.1 is the special case of Theorem 6.1.2 corresponding to*

$$\overline{Q}_{sy}(t) = \exp\left(-\int_s^t q_u(\phi_{su}(y))\, du\right).$$

6.2 Examples of PDMP's

6.2.1 Renewal processes

A SPP $\mathcal{T} = (T_n)_{n\geq 1}$ is a (0−delayed) renewal process (cf. Example 2.1.3) if the waiting times $V_n = T_n - T_{n-1}$ (with $T_0 \equiv 0$) are independent and identically distributed. Defining the *backward recurrence time process* X by

$$X_t = t - T_{N_t},$$

(the time since the most recent renewal, T_n denoting the renewal times), we claim that X is time-homogeneous Markov (with respect to the filtration $(\mathcal{F}_t^X)$).

To verify this, we use Theorem 6.1.1 and assume that the distribution of V_n has hazard function u,

$$P(V_n > v) = \exp\left(-\int_0^v u(s)ds\right).$$

(X will be time-homogeneous Markov even without this assumption, which is made to make the example fit into the framework of Theorem 6.1.1. For the more general result, use Theorem 6.1.2).

As state space we may use $G = \mathbb{R}_0$ or $G = [0, t^\dagger[$, with $t^\dagger$ the termination point for the distribution of the V_n. X is then identified with the MPP (T_n, Y_n), where the T_n are as above, and $Y_n = X_{T_n}$ is seen to be always 0 on $(T_n < \infty)$. Hence, when determining the distribution of (T_n, Y_n) we need only worry about $P_{z_n}^{(n)}$, where $y_1 = y_2 = \cdots = y_n = 0$ and similarly for the $\pi^{(n)}$. By calculation it is clear that (cf. Example 2.1.3)

$$\overline{P}^{(n)}_{t_1\cdots t_n, 0\cdots 0}(t) = \exp\left(-\int_0^{t-t_n} u(s)\, ds\right)$$

and since $\pi^{(n)}_{t_1\cdots t_n t 0\cdots 0} = \varepsilon_0$, and the deterministic behavior of X is described by $X_t = \phi_{t-T_{\overline{N}_t}}(Y_{\overline{N}_t})$ where $\phi_t(y) = y + t$ satisfies (6.10), it follows from Theorem 6.1.1 that X is time-homogeneous Markov with

$$\phi_t(y) = y + t, \quad q(y) = u(y), \quad p(y, \cdot) = \varepsilon_0.$$

Consider again the renewal process $\mathcal{T} = (T_n)$ but assume now that $P(V_n < \infty) = 1$. The *forward recurrence time process* $\widetilde{X}$ is defined by

$$\widetilde{X}_t = T_{\overline{N}_t+1} - t,$$

the time until the next renewal. The state space for $\widetilde{X}$ is again $\mathbb{R}_0$ or $\left[0, t^\dagger\right[$. We have $\widetilde{X}_0 = T_1$, and to fix the value, $\widetilde{X}_0 \equiv \widetilde{x}_0$, we must condition on $T_1 = V_1 = \widetilde{x}_0$, i.e. all distributions must be evaluated under the conditional probability $P(\cdot | T_1 = \widetilde{x}_0)$.

The MPP describing $\widetilde{X}$ is (T_n, Y_n) with T_n as always the time of the n'th renewal and

$$Y_n = \widetilde{X}_{T_n} = T_{n+1} - T_n = V_{n+1}.$$

The deterministic part of $\widetilde{X}$ is given by $\widetilde{X}_t = \phi_{t-s}(Y_{\overline{N}_t})$, where $\phi_t(y) = y - t$. Since also

$$P^{(0)} = \varepsilon_{\tilde{x}_0}, \qquad P^{(n)}_{z_n} = \varepsilon_{t_n + y_n} \quad (n \geq 1),$$

$$\pi^{(n)}_{z_n}(]y, \infty[) = P(V_1 > y)$$

it follows quite easily from Theorem 6.1.2, that $\widetilde{X}$ is time-homogeneous Markov with respect to $(\mathcal{F}^{\widetilde{X}}_t)$. Note that since $T_{n+1} = T_n + Y_n$, one need only consider $P^{(n)}_{z_n}$ when $t_1 = \widetilde{x}_0$, $t_{k+1} = t_k + y_k$ and $\pi^{(n)}_{z_n, t}$ when in addition $t = t_n + y_n$.

Note that the filtrations $(\mathcal{F}^X_t)$ and $(\mathcal{F}^{\widetilde{X}}_t)$ for the backward and forward processes are quite different: $T_{\overline{N}_t+1}$ is $\mathcal{F}^{\widetilde{X}}_t-$ but not $\mathcal{F}^X_t-$measurable (except in the case where the the V_n are degenerate, $V_n \equiv v_0$ for some $v_0 > 0$).

The process $\mathcal{T}$ discussed above is a *0-delayed renewal process* since all waiting times V_n for $n \geq 1$ are iid, corresponding to $X_0 \equiv 0$. But as a Markov process, the backward recurrence time process X may be started from any $x_0 \in [0, t^\dagger[$ in which case, by Theorem 6.1.1 (b), with $q(y) = u(y)$ and $\phi_t(y) = y + t$, we find

$$\overline{P}^{(0)}(t) = \exp\left(-\int_0^t u(x_0 + s) ds\right),$$

while the remaining $P^{(n)}$ and all $\pi^{(n)}$ are as before. In the case where $x_0 > 0$, $\mathcal{T} = (T_n)$ is an example of a *delayed renewal process*, with the waiting times V_n independent for $n \geq 1$ but only $(V_n)_{n\geq 2}$ identically distributed. Note that if U has the same distribution as the V_n for $n \geq 2$, the distribution of V_1 is that of $U - x_0$ given $U > x_0$.

6.2.2 Processes derived from homogeneous Poisson measures

Let N be a time-homogeneous Poisson process with parameter $\lambda > 0$, and let $(U_n)_{n\geq 1}$ be an iid sequence of $\mathbb{R}^d$−valued random variables such that $P(U_n = 0) = 0$. Let $\mathbb{P}$ denote the distribution of the U_n. (This setup was used to describe a compound Poisson process in Remark 5.2.1).

Let $\phi : D \to G = \mathbb{R}^d$ satisfy the conditions of Theorem 6.1.1 (a) and consider the piecewise deterministic process $X = (X_t)_{t\geq 0}$ given by $X_0 \equiv x_0$,

$$X_t = \phi_{T_{\overline{N}_t},t}(Y_{\overline{N}_t})$$

where as usual $Y_n = X_{T_n}$, while the U_n determine the jump sizes for X,

$$\triangle X_{T_n} = U_n.$$

Identifying X with the MPP (T_n, Y_n), we find

$$\overline{P}^{(n)}_{z_n}(t) = e^{-\lambda(t-t_n)}$$

and

$$\begin{aligned}\pi^{(n)}_{z_n,t}(C) &= P(\phi_{t_n,t}(y_n) + U_{n+1} \in C)\\ &= \mathbb{P}(C - \phi_{t_n,t}(y_n))\end{aligned}$$

It follows from Theorem 6.1.1 that X is a Markov process, which is time homogeneous if $\phi_{tu}(y)$ is of the form $\phi_{u-t}(y)$. For the compound Poisson process (a step process: $\phi_t(y) = y$) and the piecewise linear process with $\phi_t(y) = y + \alpha t$ we obtain processes with stationary, independent increments, cf. Section 5.2.

6.2.3 A PDMP that solves a SDE

Let N be a homogeneous Poisson process with parameter $\lambda > 0$ and consider the stochastic differential equation (SDE)

$$dX_t = a(X_t)\,dt + \sigma(X_{t-})\,dN_t, \quad X_0 \equiv x_0 \tag{6.16}$$

for a $\mathbb{R}-$valued cadlag process X, where a and σ are given functions with $\sigma(x) \neq 0$ for all x. If a solution X exists it satisfies

$$X_t = X_0 + \int_0^t a(X_s)\,ds + \int_{]0,t]} \sigma(X_{s-})\,dN_s$$

and is necessarily cadlag. To find the solution note that between jumps, writing $\dot{X}_t = \frac{d}{dt} X_t$,

$$\dot{X}_t = a(X_t),$$

which should be compared with (6.13), while because $\sigma(x)$ is never 0, X jumps precisely when N does and

$$\Delta X_t = \sigma(X_{t-})\Delta N_t.$$

Suppose now that for every x, the non-linear differential equation

$$f'(t) = a(f(t)), \quad f(0) = x$$

has a unique solution $f(t) = \phi_t(x)$. Fixing $s \geq 0$ and looking at the functions $t \to \phi_{t+s}(x)$, $t \to \phi_t(\phi_s(x))$, it is immediately checked that both solve the differential equation $f'(t) = a(f(t))$, $f(0) = \phi_s(x)$, hence, as expected, $\phi_{t+s} = \phi_t \circ \phi_s$, $\phi_0 = \mathrm{id}$, i.e. (6.10) holds.

It is now clear that (6.16) has a unique solution given by

$$X_t = \phi_{t-T_{N_t}}(Y_{N_t}), \quad \Delta X_{T_{n+1}} = \sigma\left(\phi_{t-T_n}(Y_n)\right)$$

where the T_n are the jump times for N (and X) and $Y_n = X_{T_n}$ with $Y_0 \equiv x_0$. (Note that it is used in an essential manner that it is the term $\sigma(X_{t-})\,dN_t$ rather than $\sigma(X_t)\,dN_t$ that appears in (6.16)). It is now an easy matter to see that X is a time-homogeneous PDMP: the Markov kernels for the (T_n, Y_n) are

$$\overline{P}^{(n)}_{z_n}(t) = e^{-\lambda(t-t_n)} \quad (t \geq t_n),$$
$$\pi^{(n)}_{z_n,t} = \varepsilon_{y_n^*} \quad (t > t_n),$$

where $y_n^* = \phi_{t-t_n}(y_n) + \sigma\left(\phi_{t-t_n}(y_n)\right)$, and hence X satisfies the conditions of Theorem 6.1.1 (b) with $\widetilde{\phi}_t = \phi_t$, $\widetilde{q}(y) = \lambda$ for all y, and

$$\widetilde{p}(y, \cdot) = \varepsilon_{y+\sigma(y)}.$$

The reasoning above works particularly well because N is Poisson. With other choices for N one might still find that X is Markov, but it is clear that the hazard functions for the conditional jump time distributions of N must then in a critical way depend on the ϕ_t. We shall not here discuss further the possibility of obtaining homogeneous PDMP's with non-Poisson jump times as solutions to SDE's.

6.2.4 An example from the theory of branching processes

Suppose one wants to set up a model describing the evolution of a one-sex population, where each individual (mother) can give birth to a new individual with a birth rate depending on the age of the mother, and where each individual may die according to an age dependent death rate. How should one go about this? We of course shall do it by defining a suitable MPP and associating with that a process which turns out to be a homogeneous PDMP.

Let $\beta : \mathbb{R}_0 \to \mathbb{R}_0$, $\delta : \mathbb{R}_0 \to \mathbb{R}_0$ be given functions, to be understood as the age-dependent birth rate and death rate respectively. We suppose given an initial population $X_0 \equiv x_0$ consisting of $k_0 \geq 1$ individuals, labelled by their ages $a_1^{(0)}, \ldots, a_{k_0}^{(0)}$ (all ≥ 0). There are different possibilities for the choice of the state space G. Here we shall simply let G consist of all finite subsets of $\mathbb{R}_0$, allowing for repetitions of elements, where the state $x = \{a_1, \ldots, a_k\}$ for $k \in \mathbb{N}_0$, $a_1, \ldots, a_k \geq 0$ is interpreted as a population consisting of k individuals of ages $a_1, \ldots, a_k$. Thus $X_0 \equiv x_0 = \left\{a_1^{(0)}, \ldots, a_{k_0}^{(0)}\right\}$. In general individuals may have the same age, denoted by a repeated element: a population consisting of 5 individuals, one pair of twins and one set of triplets, would be denoted $\{a, a, b, b, b\}$. An extinct population (0 individuals) is denoted $\emptyset$. If $x = \{a_1, \ldots, a_k\} \in G$, we write $|x|$ for the population size k and $x + t = \{a_1 + t, \ldots, a_k + t\}$ (Note: the choice of G is quite crude and ignores information concerning the life histories of single individuals. It is not a problem to define a population process on a state space where one keeps track of this type of information – the details are left to the reader).

In the model we shall define (which can easily be generalized to allow e.g. for multiple births), all births and deaths occur at separate time points (so one only sees individuals of the same age if they are present at time 0). In particular, the population size can only increase or decrease in jumps of size 1. If the process is observed in state $x = \{a_1, \ldots, a_k\}$ and a jump occurs, the state reached by the jump is either $\{a_1, \cdots, a_k\} \cup 0$ corresponding to a birth or $\{a_1, \ldots, a_k\} \backslash a_i$ for some $1 \leq i \leq k$ corresponding to a death (where of course, if there were several individuals of age a_i in x, $\{a_1, \ldots, a_k\} \backslash a_i$ denotes the population where precisely one of those of age a_i is removed, the others retained).

It is now a simple matter to set up the model. The process is piecewise deterministic,

$$X_t = \phi_{t - T_{\overline{N}_t}} \left(Y_{\overline{N}_t}\right)$$

with T_n the time of the n'th jump, Y_n the state reached by that jump, and the deterministic behaviour given by

$$\phi_t(x) = x + t$$

if $|x| \geq 1$ and $\phi_t(\emptyset) = \emptyset$. The Markov kernels determining the distribution of the MPP (T_n, Y_n) are given by, for $t \geq t_n$, with only the atoms for $\pi^{(n)}$ listed,

$$\begin{aligned}
\overline{P}^{(n)}_{z_n|x_0}(t) &= \exp\left(-\int_0^{t-t_n} [\beta+\delta]_{y_n}(s)\,ds\right),\\
\pi_{z_n,t|x_0}\left((y_n + t - t_n) \cup 0\right) &= \frac{\sum_{i=1}^k \beta(a_i + t - t_n)}{[\beta+\delta]_{y_n}(t-t_n)}\\
\pi_{z_n,t|x_0}\left((y_n + t - t_n) \backslash a_{i_0} + t - t_n\right) &= \frac{\delta(a_{i_0} + t - t_n)}{[\beta+\delta]_{y_n}(t-t_n)} \quad (1 \leq i_0 \leq k).
\end{aligned}$$

Here

$$\pi_{z_n,t|x_0}\left((y_n + t - t_n) \backslash a_{i_0} + t - t_n\right) = \frac{\delta(a_{i_0} + t - t_n)}{[\beta+\delta]_{y_n}(t-t_n)} \quad (1 \leq i_0 \leq k).$$

$$[\beta+\delta]_{y_n}(s) := \sum_{i=1}^k \left(\beta(a_i+s) + \delta(a_i+s)\right)$$

and $z_n = (t_1, \ldots, t_n; y_1, \ldots, y_n)$ with $y_n = \{a_1, \ldots, a_k\}$, using of course $t_0 = 0$, $y_0 = x_0$ if $n = 0$. Naturally, empty sums equal 0, hence it follows in particular that once the population becomes extinct, it remains extinct forever.

It may be shown that the process defined above does not explode. Clearly the ϕ_t satsify (6.10) and it is then immediately verified from Theorem 6.1.1 (b) that X is time-homogeneous Markov with

$$\begin{aligned}
q(x) &= \sum_{i=1}^k \left(\beta(a_i) + \delta(a_i)\right),\\
p(x, x \cup 0) &= \frac{\sum_{i=1}^k \beta(a_i)}{\sum_{i=1}^k \left(\beta(a_i) + \delta(a_i)\right)},\\
p(x, x \backslash a_{i_0}) &= \frac{\delta(a_{i_0})}{\sum_{i=1}^k \left(\beta(a_i) + \delta(a_i)\right)} \quad (1 \leq i_0 \leq k),
\end{aligned}$$

where $x = \{a_1, \ldots, a_k\}$, and where only the atoms for $p(x, \cdot)$ have been specified. The corresponding predictable intensities, $\lambda_t(C)$ for special C, are given by

$$\begin{aligned}\overline{\lambda}_t &= \sum_{a\in X_{t-}} \left(\beta\left(a\right)+\delta\left(a\right)\right), \\ \lambda_t\left(X_{t-}\cup 0\right) &= \sum_{a\in X_{t-}} \beta(a), \\ \lambda\left(X_{t-}\backslash a_0\right) &= \delta\left(a_0\right) \quad \left(a_0\in X_{t-}\right),\end{aligned}$$

justifying the interpretation of β, δ as age-dependent birth and death intensities.

A particularly simple case of the model above is obtained when both β and δ are constants. In that case the intensities above depend on X_{t-} through the population size $|X_{t-}|$ only and it follows that the process $|X|$ is a time-homogeneous Markov chain with state space $\mathbb{N}_0$, the familiar linear birth- and death process with transition intensities

$$q_{i,i+1} = i\beta, \quad q_{i,i-1} = i\delta.$$

The process we have constructed here has the *branching property* characteristic of branching processes: with the initial state $x_0 = \left\{a_1^0,\ldots,a_{k_0}^{(0)}\right\}$ as above, define k_0 independent processes $X^{(i)}$ with age-dependent birth- and death intensities β and δ respectively such that $X_0^{(i)} = \left\{a_i^{(0)}\right\}$. Then $\widetilde{X} := \bigcup_{i=1}^{k_0} X^{(i)}$ has the same distribution as X. (A proof of this may be given using Theorem 5.1.1 to identify the relevant compensating measure for $\widetilde{X}$).

A quantity that is easily calculated and critical for the ultimate behaviour of the population is γ, the expected number of children born to an individual throughout her lifetime. To find γ, consider an initial population $\{0\}$ consisting of one newborn individual ι. Let N^b be the counting process that registers the times at which ι gives birth. Let also ζ denote the time at which ι dies and define $N_t^d = 1_{(\zeta\le t)}$. then $\left(N^b,N^d\right)$ has $\mathcal{F}^{\left(N^b,N^d\right)}-$compensator $\left(\Lambda^b,\Lambda^d\right)$ where

$$\Lambda_t^b = \int_0^t 1_{\left(N_{s-}^d=0\right)}\beta(s)\,ds, \quad \Lambda_t^d = \int_0^t 1_{\left(N_{s-}^d=0\right)}\delta(s)\,ds.$$

Since Λ^d is $\mathcal{F}^{N^d}-$predictable we recognize that

$$P\left(\zeta>t\right) = \exp\left(-\int_0^t \delta\left(s\right)\,ds\right)$$

and it then follows that

$$EN_t^b = E\Lambda_t^b = \int_0^t \exp\left(-\int_0^s \delta(u)\,du\right)\beta(s)\,ds,$$

and defining $N_\infty^b = \lim_{t\to\infty} N_t^b$ that

$$\gamma = EN_\infty^b = \int_0^\infty \exp\left(-\int_0^s \delta(u)\,du\right)\beta(s)\,ds.$$

One would expect the population to become extinct almost surely (no matter what the value of X_0) if $\gamma < 1$, while if $\gamma > 1$ for some initial populations (e.g. those containing at least one newborn) the population will grow to ∞ over time with probability > 0. Note that for the linear birth- and death process $\gamma \gtreqless 1$ according as $\beta \gtreqless \delta$.

6.3 Itô's formula for homogeneous PDMP's

In this section we shall discuss in more detail the structure of $\mathbb{R}-$valued piecewise deterministic Markov processes, that are time-homogeneous.

The processes will be of the type described in Theorem 6.1.1 (b). Thus, if $X = (X)_{t\geq 0}$ denotes the process with now $(\mathbb{R}, \mathcal{B})$ as state space, X is completely specified by its initial value $X_0 \equiv x_0$ and the MPP $(T_n, Y_n)_{n\geq 1}$, where T_n is the time of the n'th discontinuity (n'th jump) for X, and $Y_n = X_{T_n} \in \mathbb{R}$ is the state reached by the n'th jump, cf. (6.8),

$$X_t = \phi_{t-T_{\overline{N}_t}}(Y_{\overline{N}_t})\ ,$$

with ϕ describing the deterministic behavior of X between jumps so that $t \to \phi_t(y)$ is continuous and satisfies the semigroup equation $\phi_{s+t} = \phi_s \circ \phi_t$ $(s, t \geq 0)$ with the initial condition $\phi_0 = \mathrm{id}$, cf. (6.10). Recall also that the distribution of $(T_n, Y_n)_{n\geq 1}$ is determined by

$$\begin{aligned}
\overline{P}^{(n)}_{z_n|x_0}(t) &= \exp\left(-\int_0^t q(\phi_{t-s}(y_n))\,ds\right), \\
\pi^{(n)}_{z_n,t|x_0}(C) &= p(\phi_{t-t_n}(y_n), C)
\end{aligned}$$

with $q \geq 0$ and such that $t \to q(\phi_t(y))$ is a hazard function every y, and with p a Markov kernel on G such that, cf. (6.9),

$$p(y, \{y\}) = 0 \quad (y \in \mathbb{R}). \tag{6.17}$$

If μ is the RCM determined by (T_n, Y_n), we see that μ has $\mathcal{F}_t^\mu = \mathcal{F}_t^X-$ compensating measure L given by

$$L([0,t] \times C) = \Lambda_t(C) = \int_0^t \lambda_s(C)\,ds$$

with the predictable intensity $\lambda(C)$ determined as in (6.15), in other words

$$L\,(ds,dy) = q\,(X_{s-})\;ds\,p\,(X_{s-},dy)\,. \tag{6.18}$$

Finally, we shall assume that $t \to \phi_t(y)$ is continuously differentiable, and writing

$$a(y) = D_t\phi_t(y)|_{t=0}$$

we then have the differential equation (6.13),

$$D_t\phi_t(y) = a\,(\phi_t(y))\,.$$

Suppose now that $f : \mathbb{R}_0 \times \mathbb{R} \to \mathbb{R}$ is a continuous function, and consider the process $(f(t,X_t))_{t\geq 0}$. We may then use Itô's formula for MPP's, Section 3.8, to obtain a decomposition of $f(\cdot,X)$ into a local martingale and a predictable process. To formulate the result, introduce the operator $\mathcal{A}$ acting on the space $\mathcal{D}\,(\mathcal{A})$ of functions f that are continuously differentiable in t and y and satisfy that $\int_{\mathbb{R}} p\,(y,d\widetilde{y})\;|f\,(t,\widetilde{y}) - f\,(t,y)| < \infty$ for all y, and with $\mathcal{A}f$ given by

$$\mathcal{A}f\,(t,y) = D_tf(t,y) + a(y)D_yf(t,y) + q(y)\int_{\mathbb{R}} p(y,d\widetilde{y})\,(f(t,\widetilde{y}) - f(t,y)).$$

Theorem 6.3.1 *(a) For $f \in \mathcal{D}\,(\mathcal{A})$ the process $f(\cdot,X)$ may be written*

$$f(t,X_t) = f(0,x_0) + \int_{]0,t]\times\mathbb{R}} S_s^y\,M(ds,dy) + U_t, \tag{6.19}$$

where M is the martingale measure $\mu - L$, $(S_s^y)_{s\geq 0, y\in\mathbb{R}}$ is the $\mathcal{F}_t^X-$predictable flow given by

$$S_t^y = f(t,y) - f(t,X_{t-}) \tag{6.20}$$

and U is continuous and $\mathcal{F}_t^X-$predictable with

$$U_t = \int_0^t \mathcal{A}f\,(s,X_s)\;ds. \tag{6.21}$$

The decomposition is unique up to indistinguishability.

(b) If $E\overline{N}_t < \infty$ for all t and if $f \in \mathcal{D}\,(\mathcal{A})$ is such that

$$\mathcal{A}f \equiv 0 \tag{6.22}$$

and the function $(s,y,\widetilde{y}) \to f\,(s,\widetilde{y}) - f\,(s,y)$ is bounded on $[0,t] \times \mathbb{R}^2$ for all t, then the process $f\,(\cdot,X)$ is a $\mathcal{F}_t^X-$martingale.

Proof. (a). As in Section 3.8 we identify S by identifying the jumps in (6.19). Since $f \in \mathcal{D}(\mathcal{A})$ and X is continuous between jumps, $f(\cdot, X)$ jumps only when μ does and with the requirement that U in (6.19) be continuous it follows that

$$\Delta f(t, X_t)\, \Delta \overline{N}_t = S_t^{X_t} \Delta \overline{N}_t$$

which certainly holds if S is given by (6.20).

Having found S, we define U by solving (6.19) for U_t. It then follows that between jumps U is differentiable in t and that $\dot{U}_t := \frac{d}{dt} U_t$ is given by

$$\dot{U}_t = \tfrac{d}{dt} f(t, X_t) + \tfrac{d}{dt} \int_{]0,t] \times \mathbb{R}} S_s^y \, L(ds, dy).$$

But

$$\tfrac{d}{dt} f(t, X_t) = D_t f(t, X_t) + D_y f(t, X_t)\, \dot{X}_t$$

and by (6.13), between jumps

$$\dot{X}_t = \tfrac{d}{dt} \phi_{t - T_{\overline{N}_t}} \left(Y_{\overline{N}_t} \right) = a\left(\phi_{t - T_{\overline{N}_t}} \left(Y_{\overline{N}_t} \right) \right) = a(X_t).$$

Since also, by (6.18)

$$\begin{aligned} \tfrac{d}{dt} \int_{]0,t] \times \mathbb{R}} S_s^y \, L(ds, dy) &= \tfrac{d}{dt} \int_0^t ds\, q(X_{s-}) \int_{\mathbb{R}} p(X_{s-}, dy)\, S_s^y \\ &= q(X_t) \int_{\mathbb{R}} p(X_t, dy)\, S_t^y, \end{aligned}$$

using (6.20) it follows that between jumps

$$\dot{U}_t = \mathcal{A} f(t, X_t)$$

and we have shown that (6.19) holds with S, U given by (6.20), (6.21).

The uniqueness of the decomposition follows from Proposition 3.5.1.

(b). This is immediate from (a) and Theorem 3.6.1 (iii2). ■

Suppose that $h : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and bounded. Then $f \in \mathcal{D}(\mathcal{A})$ where $f(t, y) = h(y)$ and $\mathcal{A} f(t, y) = Ah(y)$ where

$$Ah(y) = a(y) h'(y) + q(y) \int_{\mathbb{R}} p(y, d\widetilde{y}) \left(h(\widetilde{y}) - h(y) \right). \tag{6.23}$$

The operator A, acting on a suitable domain of functions h, is the *infinitesimal generator* for the time-homogeneous Markov process X. (The operator $\mathcal{A}$ is the generator for the *time-space process* $(t, X_t)_{t \geq 0}$). Note that if

$E\overline{N}_t < \infty$ for all t, since h is bounded the stochastic integral $M(S)$ (where now $S_t^y = h(y) - h(X_{t-})$) is a martingale, and hence, taking expectations in (6.19),

$$Eh(X_t) = h(x_0) + \int_0^t EAh(X_s)\,ds,$$

from which follows the familiar formula for the generator,

$$Ah(x_0) = \lim_{t\to 0} \frac{1}{t}\left(E^{x_0}h(X_t) - h(x_0)\right), \tag{6.24}$$

valid for arbitrary x_0, and where we have written E^{x_0} instead of E to emphasize that $X_0 \equiv x_0$.

6.3.1 An example involving ruin probabilities

In general there is not much hope of solving an equation like (6.22) explicitly. However, martingales of the form $f(\cdot, X)$ are nice to deal with when available, as we shall now see when discussing some classical *ruin problems.*

Let N be a one-dimensional, homogeneous Poisson process with parameter $\lambda > 0$ and let $(U_n)_{n\geq 1}$ be a sequence of iid $\mathbb{R}-$valued random variables with $P(U_n = 0) = 0$, independent of N. Finally, let $x_0 > 0$, let $\alpha \in \mathbb{R}$ and define

$$X_t = x_0 + \alpha t + \sum_{n=1}^{N_t} U_n,$$

i.e. X is a compound Poisson process (Remark 5.2.1) with a linear drift $t \to \alpha t$ added on. In particular X has stationary independent increments, cf. Section 5.2, and is a time-homogeneous, piecewise deterministic Markov process with

$$\phi_t(y) = y + \alpha t, \quad q(y) = \lambda$$

for all t, y, and

$$p(y,\cdot) = \text{the distribution of } y + U_1.$$

We define the *time to ruin* as

$$\tau_{\text{ruin}} = \inf\{t : X_t \leq 0\}.$$

The problem is then to find the *ruin probability* $p_{\text{ruin}} = P(\tau_{\text{ruin}} < \infty)$ and, if possible, the distribution of τ_{ruin}.

We shall focus on two different setups: (i) the *simple ruin problem* corresponding to the case where $P(U_1 > 0) = 1$, $\alpha < 0$, where X decreases linearly between the strictly positive jumps, (ii) the *difficult ruin problem* where $P(U_1 < 0) = 1,\ \alpha > 0$ with X increasing linearly between the strictly negative jumps.

It should be clear that for Problem (i),

$$X_{\tau_{\mathrm{ruin}}} = 0 \quad \text{on} \quad (\tau_{\mathrm{ruin}} < \infty), \tag{6.25}$$

and it is this property that makes (i) simple, while for Problem (ii) it may well happen that

$$X_{\tau_{\mathrm{ruin}}} < 0,$$

and it is this possibility of *undershoot*, which makes (ii) difficult.

It is easy to find that $p_{\mathrm{ruin}} = 1$ in some special cases: in both Problem (i) and (ii), $\xi = EU_1$ is well defined, but may be infinite. Now

$$X_{T_n} = x_0 + \alpha T_n + \sum_{k=1}^{n} U_k$$

and so, by the strong law of large numbers

$$\frac{1}{n} X_{T_n} \stackrel{\text{a.s.}}{\to} \frac{\alpha}{\lambda} + \xi. \tag{6.26}$$

Consequently, for both Problem (i) and (ii)

$$p_{\mathrm{ruin}} = 1 \quad \text{if} \quad \frac{\alpha}{\lambda} + \xi < 0. \tag{6.27}$$

(For (i) this is possible only if $\xi < \infty$. For (ii) (6.27) is satisfied if in particular $\xi = -\infty$).

Let $\mathbb{P}$ denote the distribution of the U_n and introduce

$$\psi(\theta) = Ee^{-\theta U_1} = \int e^{-\theta u}\, \mathbb{P}(du).$$

In the case of Problem (i), this is finite if $\theta \geq 0$, and ψ is the Laplace transform for U_1. For Problem (ii), ψ is finite if $\theta \leq 0$. (It is of course possible that $\psi(\theta) < \infty$ for other values of θ than those just mentioned in either case (i) or (ii)). From now on, when discussing Problem (i), assume $\theta \geq 0$ and when treating Problem (ii), assume $\theta \leq 0$.

By a simple calculation

$$\begin{aligned} E\exp(-\theta(X_t - x_0)) &= \sum_{n=1}^{\infty} P(\overline{N}_t = n) e^{-\theta\alpha t} E\exp\left(-\theta \textstyle\sum_1^n U_k\right) \\ &= \sum_{n=1}^{\infty} \frac{(\lambda t)^n}{n!} e^{\lambda t} e^{-\theta\alpha t} \psi^n(\theta) \\ &= \exp(t\rho(\theta)) \end{aligned}$$

where

$$\rho(\theta) = \lambda\psi(\theta) - \lambda - \theta\alpha. \tag{6.28}$$

Because X has stationary, independent increments, also

$$E\exp(-\theta(X_{s+t} - X_s))|\mathcal{F}_s^X) = \exp(t\rho(\theta))$$

for any $s \geq 0$, so defining

$$V_t(\theta) = \exp(-\theta(X_t - x_0) - t\rho(\theta)),$$

the following result follows immediately:

Proposition 6.3.2 *For Problem (i), $(V_t(\theta), \mathcal{F}_t^X)$ is a martingale for each $\theta \geq 0$. For Problem (ii), $(V_t(\theta), \mathcal{F}_t^X)$ is a martingale for each $\theta \leq 0$.*

We proved this without using (6.22), but it is quite instructive to verify that (6.22) holds: we have

$$f(t, y) = \exp(-\theta(y - x_0) - t\rho(\theta))$$

and $a(y) = \alpha$ for all y, so

$$D_t f(t, y) + a(y) D_y f(t, y) = f(t, y)(-\rho(\theta) - \alpha\theta)$$

while

$$\begin{aligned} q(y) \int_{\mathbb{R}} p(y, d\tilde{y})\,(f(t, \tilde{y}) - f(t, y)) &= \lambda \left(\int f(t, y + u)\,\mathbb{P}(du) - f(t, y) \right) \\ &= f(t, y)(\lambda\psi(\theta) - \lambda). \end{aligned}$$

Using (6.28) we see that (6.22) does indeed hold.

Problem (i)

First note, that as a Laplace transform, ψ is strictly convex on $\mathbb{R}_0$, and that ψ is differentiable on $\mathbb{R}_+$ with

$$\psi'(\theta) = -EU_1 e^{-\theta U_1}.$$

Let $\theta \downarrow 0$ and use monotone convergence to obtain $\psi'(0) := \lim_{\theta \downarrow 0} \psi'(\theta)$,

$$\psi'(0) = -\xi,$$

also if $\xi = \infty$.

Since ψ is strictly convex, so is ρ, and

$$\rho(0) = 0, \quad \rho'(0) = -\lambda\xi - \alpha.$$

Defining $\theta_0 = \sup\{\theta \geq 0 : \rho(\theta) = 0\}$, it follows, that $\theta_0 = 0$ iff $\alpha/\lambda + \xi \leq 0$ (iff $\rho'(0) \geq 0$), cf. (6.27), and that $\theta_0 > 0$ iff $\alpha/\lambda + \xi > 0$, (iff $\rho'(0) < 0$).

Since $V_0(\theta) \equiv 1$, the martingales $V(\theta)$ for $\theta \geq 0$ have constant expectation 1, so applying optional sampling to the bounded stopping times $t \wedge \tau_{\text{ruin}}$ for $t \geq 0$ we obtain

$$1 = EV_{t\wedge\tau_{\text{ruin}}}(\theta) = E1_{\left(\tau_{\text{ruin}}\leq t\right)} e^{\theta x_0 - \tau_{\text{ruin}}\rho(\theta)} + E1_{(\tau_{\text{ruin}}>t)} e^{-\theta(X_t - x_0) - t\rho(\theta)}$$

using (6.25) on the way. Now take $\theta > \theta_0$ and consider the last term. Since $t < \tau_{\text{ruin}},\ X_t > 0$ so the integrand is dominated by the constant $e^{\theta x_0}$ (because also $\rho(\theta) > 0$). Since $e^{-t\rho(\theta)} \to 0$ as $t \to \infty$, the last term vanishes as $t \to \infty$. On the first, use monotone convergence to obtain

$$1 = E1_{(\tau_{\text{ruin}}<\infty)} e^{\theta x_0 - \tau_{\text{ruin}}\rho(\theta)} \qquad (\theta > \theta_0).$$

By monotone convergence

$$E1_{(\tau_{\text{ruin}}<\infty)} e^{-\tau_{\text{ruin}}\rho(\theta)} \uparrow p_{\text{ruin}} \quad \text{as} \quad \theta \downarrow \theta_0$$

and consequently

$$p_{\text{ruin}} = e^{-\theta_0 x_0}.$$

Furthermore, again since $\rho(\theta) > 0$ for $\theta > \theta_0$,

$$Ee^{-\rho(\theta)\tau_{\text{ruin}}} = E1_{(\tau_{\text{ruin}}<\infty)} e^{-\rho(\theta)\tau_{\text{ruin}}} = e^{-\theta x_0}$$

and we have shown the following result:

Proposition 6.3.3 *(a) For Problem (i), the ruin probability is*

$$p_{\text{ruin}} = e^{-\theta_0 x_0}.$$

where $\theta_0 = \sup\{\theta \geq 0 : \rho(\theta) = 0\}$.

(b) For $\vartheta > 0$, *the Laplace transform of the (possibly infinite) random variable* τ_{ruin} *is given by*

$$Ee^{-\vartheta\tau_{\text{ruin}}} = e^{-\rho^{-1}(\vartheta)x_0}, \tag{6.29}$$

where $\rho^{-1} : \mathbb{R}_0 \to [\theta_0, \infty[$ *is the strictly increasing inverse of the function* ρ *restricted to the interval* $[\theta_0, \infty[$.

Note that (6.29) is valid only for $\vartheta > 0$, but that by monotone convergence

$$p_{\text{ruin}} = e^{-\theta_0 x_0} = \lim_{\vartheta\downarrow 0, \vartheta>0} e^{-\rho^{-1}(\vartheta)x_0} = \lim_{\vartheta\downarrow 0, \vartheta>0} Ee^{-\vartheta\tau_{\text{ruin}}}.$$

Problem (ii)

For the solution of Problem (i), (6.25) was used in an essential manner. With the possibility of undershoot occurring in Problem (ii), (6.25) no longer holds. The basic idea now is to replace X by the process $\widetilde{X}$ given by

$$\widetilde{X}_t = \begin{cases} X_t & \text{if } t < \tau_{\text{ruin}} \\ 0 & \text{if } t \geq \tau_{\text{ruin}}, \end{cases}$$

i.e. at the time of ruin, the non-positive value of X is replaced by 0 and kept there forever. Clearly

$$\widetilde{X}_{\tau_{\text{ruin}}} = 0 \quad \text{on} \quad (\tau_{\text{ruin}} < \infty),$$

so (6.25) holds for $\widetilde{X}$. However, going from X to $\widetilde{X}$ destroys the independent increments property, so it is not so easy to find simple functions of $\widetilde{X}$ that are martingales.

We certainly have that $\widetilde{X}$ is a piecewise deterministic process with state space $\mathbb{R}_0$. If $\widetilde{T}_n$ is the time of the n'th jump for $\widetilde{X}$ and $\widetilde{Y}_n = \widetilde{X}_{\widetilde{T}_n} \geq 0$ the state reached by that jump, it is immediately seen that the Markov kernels $\widetilde{P}^{(n)}$, $\widetilde{\pi}^{(n)}$ determining the distribution of the MPP $(\widetilde{T}_n, \widetilde{Y}_n)$ are given as follows:

$$\overline{\widetilde{P}}^{(n)}_{z_n|x_0}(t) = \begin{cases} \exp(-\lambda(t-t_n)) & \text{if } y_n > 0 \\ 1 & \text{if } y_n = 0 \end{cases}$$

for $t \geq t_n$, while for $t > t_n,\ y_n > 0$,

$$\widetilde{\pi}^{(n)}_{z_n,t|x_0} = \text{the distribution of } (y_n + \alpha(t-t_n) + U_1) \vee 0,$$

or more formally

$$\widetilde{\pi}^{(n)}_{z_n,t|x_0}(]a, y_n + \alpha(t-t_n)[) = \mathbb{P}(]a - y_n - \alpha(t-t_n), 0[)$$

if $0 < a \leq y_n + \alpha(t-t_n)$ and

$$\widetilde{\pi}^{(n)}_{z_n,t|x_0}(\{0\}) = \mathbb{P}(]-\infty, -y_n - \alpha(t-t_n)]).$$

(Recall that $\mathbb{P}$ is the distribution of U_1 and that $U_1 < 0$).

The expression for $\widetilde{P}^{(n)}$ and $\widetilde{\pi}^{(n)}$ show that $\widetilde{X}$ is a piecewise deterministic timehomogeneous Markov process with

$$\begin{aligned} \widetilde{\phi}_t(y) &= \begin{cases} y + \alpha t & \text{if } y > 0, \\ 0 & \text{if } y = 0, \end{cases} \\ \widetilde{q}(y) &= \begin{cases} \lambda & \text{if } y > 0, \\ 0 & \text{if } y = 0, \end{cases} \\ \widetilde{p}(y, \cdot) &= \text{the distribution of } (y + U_1) \vee 0. \end{aligned}$$

From now on assume that $-U_1$ follows an exponential distribution,

$$P(U_1 \le u) = e^{\beta u} \qquad (u \le 0),$$

where $\beta > 0$. Thus $\xi = -\frac{1}{\beta}$, so from (6.27),

$$p_{\text{ruin}} = 1 \quad \text{if} \quad \alpha\beta < \lambda.$$

We therefore also assume that $\alpha\beta \ge \lambda$, and are now looking for a bounded function $g : \mathbb{R}_0 \to \mathbb{R}$ such that $g(\widetilde{X})$ is a $\mathcal{F}_t^{\widetilde{X}}-$martingale.

Lemma 6.3.4 *With*

$$g(y) = \left(\kappa + e^{\theta^* y}\right) 1_{\mathbb{R}_+}(y) \qquad (y \in \mathbb{R}_0)\,,$$

where $\kappa = -\frac{\beta}{\beta+\theta}$, $\theta^* = \frac{\lambda}{\alpha} - \beta$, *the process* $g\left(\widetilde{X}\right)$ *is a* $\mathcal{F}_t^{\widetilde{X}}-$*martingale.*

Note that $g(0) = 0$ and that $\theta^* \le 0$ so g is bounded on $\mathbb{R}_0$.

Proof. Imitating the proof of Theorem 6.3.1 we want to show that

$$g\left(\widetilde{X}_t\right) = g\left(x_0\right) + \int_{]0,t]\times\mathbb{R}_0} \widetilde{S}_s^y \,\widetilde{M}\left(ds, dy\right)$$

with $\widetilde{M} = \widetilde{\mu} - \widetilde{L}$, writing $\widetilde{\mu}$ for the RCM describing the jump times $\widetilde{T}_n$ and marks $\widetilde{Y}_n$ for $\widetilde{X}$ and writing $\widetilde{L}$ for the $\mathcal{F}_t^{\widetilde{\mu}} = \mathcal{F}_t^{\widetilde{X}}-$compensating measure for $\widetilde{\mu}$. Identifying jumps gives

$$\widetilde{S}_t^y = g(y) - g\left(\widetilde{X}_{t-}\right)$$

and by differentiation with respect to t between jumps and with $t < \tau_{\text{ruin}}$ (so $\widetilde{X}_t > 0$), we find

$$\frac{d}{dt} g\left(\widetilde{X}_t\right) = \alpha g'\left(\widetilde{X}_t\right), \tag{6.30}$$

$$-\frac{d}{dt}\int_{]0,t]\times\mathbb{R}_0} \widetilde{S}_s^y \, L(ds, dy) = -\frac{d}{dt}\int_0^t ds\, \widetilde{q}\left(\widetilde{X}_{s-}\right) \int_{\mathbb{R}_0} \widetilde{p}\left(\widetilde{X}_{s-}, dy\right)\, \widetilde{S}_s^y. \tag{6.31}$$

That these two derivatives are the same amounts to showing that for $y > 0$,

$$\begin{aligned}
\alpha\theta^* e^{\theta^* y} &= -\lambda\left(Eg\left(\left(y + U_1\right) \vee 0\right) - g(y)\right) \\
&= -\lambda\left(E1_{(U_1 > -y)}\left(\kappa + e^{\theta^*(y+U_1)}\right) - \kappa - e^{\theta^* y}\right) \\
&= \lambda\left(\kappa + \frac{\beta}{\beta + \theta^*}\right) e^{\beta y} + \lambda \frac{\theta^*}{\beta + \theta^*} e^{\theta^* y},
\end{aligned}$$

i.e. we need that

$$\kappa = -\frac{\beta}{\beta + \theta^*}, \quad \alpha = \frac{\lambda}{\beta + \theta^*},$$

which are precisely the conditions on κ and θ^* given in the statement of the lemma.

For $t > \tau_{\text{ruin}}$ there are no jumps for $\widetilde{X}$, $\widetilde{X}_t = 0$, and also $\widetilde{L}$ vanishes (since $\widetilde{q} = 0$). Hence the time derivatives (6.30), (6.31) agree on all of $\mathbb{R}_0$ between jumps, and the proof is complete. ■

Proposition 6.3.5 *When the $-U_n$ are exponential, $P(U_1 < u) = e^{\beta u}$ for $u \leq 0$, and $\alpha\beta \geq \lambda$, the ruin probability is given by*

$$p_{\text{ruin}} = e^{\left(\frac{\lambda}{\alpha} - \beta x_0\right)} \frac{\lambda}{\alpha\beta}. \tag{6.32}$$

Proof. The martingale $g\left(\widetilde{X}\right)$ is bounded, hence $\widetilde{g}_\infty := \lim_{t\to\infty} g\left(\widetilde{X}_t\right)$ exists a.s. and $E\widetilde{g}_\infty = Eg\left(\widetilde{X}_0\right) = g(x_0)$. Since $g(0) = 0$, on $(\tau_{\text{ruin}} < \infty)$ we have $\widetilde{g}_\infty = 0$. On $(\tau_{\text{ruin}} = \infty)$, $\widetilde{g}_\infty = \lim \left(\kappa + e^{\theta^* \widetilde{X}_t}\right)$ and if $\alpha\beta > \lambda$, since $\theta^* < 0$ and $\widetilde{X}_t = X_t \to \infty$ by the strong law of large numbers, see (6.26), we have $\widetilde{g}_\infty = \kappa$. Thus $g(x_0) = \kappa\left(1 - p_{\text{ruin}}\right)$ if $\alpha\beta > \lambda$, yielding the desired result. If $\alpha\beta = \lambda$ we have $\theta^* = 0$, $\kappa = -1$ and learn nothing from the argument above. However, using e.g. a coupling argument one shows that p_{ruin} is a decreasing function of β and it then follows that $p_{\text{ruin}} = 1$ if $\alpha\beta = \lambda$. (For the coupling, take $\beta_1 < \beta_2$ and define on the same probability space two processes X_1, X_2 with the same initial value, same λ and same α and the same jump times, but such that the jump sizes $U_{1,n}$ for X_1 are all larger than those, $U_{2,n}$, for X_2 – viz. $U_{1,n} = \frac{\beta_2}{\beta_1} U_{2,n}$. Then for all t, $X_{1,t} \leq X_{2,t}$ and the probability of ruin for X_1 exceeds that for X_2). ■

6.4 Likelihood functions for PDMP's

Using the results of Chapter 4 together with the construction from Section 6.1, it is an easy matter to derive likelihood processes for observation of PDMP's.

Suppose given a RCM $\mu = \sum_{n:T_n<\infty} \varepsilon_{(T_n,Y_n)}$ on some measurable space $(\Omega, \mathcal{F})$, and let $P, \widetilde{P}$ be two probability measures on $(\Omega, \mathcal{F})$. Write $P_t, \widetilde{P}_t$ for the restrictions of $P, \widetilde{P}$ to $\mathcal{F}_t^\mu$ and write $\widetilde{P} \ll_{\text{loc}} P$ if $\widetilde{P}_t \ll P_t$ for all $t \in \mathbb{R}_0$. The likelihood process $\mathfrak{L}^\mu = \left(\mathfrak{L}_t^\mu\right)_{t\geq 0}$ for observing μ is then given by the results from Chapter 4, more specifically if $Q, \widetilde{Q}$ are the distributions of μ under $P, \widetilde{P}$ respectively, then $\widetilde{P} \ll_{\text{loc}} P$ iff $\widetilde{Q} \ll_{\text{loc}} Q$ and

$$\mathfrak{L}_t^{\mu} = \mathfrak{L}_t \circ \mu,$$

where $\mathfrak{L}_t = \frac{d\widetilde{Q}_t}{dQ_t}$ is the likelihood process from Chapter 4.

Now, suppose further that $X = (X_t)_{t \geq 0}$ is a $(G, \mathcal{G})$ −valued process defined by $X_0 \equiv x_0$ and

$$X_t = \phi_{T_{\overline{N}_t}, t}\left(Y_{\overline{N}_t}\right),$$

where each $t \to \phi_{st}(y)$ is continuous and the ϕ_{st} satisfy the semigroup property (6.6) and the boundary condition (6.7), cf. Theorem 6.1.2. Finally assume that the Markov kernels generating P are as in Theorem 6.1.2, while those generating $\widetilde{P}$ have a similar structure,

$$\begin{aligned}
\overline{\widetilde{P}}^{(n)}_{z_n|x_0}(t) &= \exp\left(-\int_{t_n}^{t} \widetilde{q}_s\left(\phi_{t_n,s}(y_n)\right) ds\right) \\
\widetilde{\pi}^{(n)}_{z_n,t|x_0}(C) &= \widetilde{p}_t\left(\phi_{t_n,t}(y_n), C\right),
\end{aligned}$$

where $\widetilde{q}_s \geq 0$ and each $\widetilde{p}_t$ is a Markov kernel on G such that $\widetilde{p}_t(y, \{y\}) = 0$ for all t, y, cf. (6.9).

Thus X is a PDMP under both P and $\widetilde{P}$ and from Theorem 4.0.2 and (6.14) we immediately obtain the following result:

Theorem 6.4.1 *A sufficient condition for $\widetilde{P} \ll_{\text{loc}} P$ is that $q_t(y) > 0$ for all t, y, that $\int_s^t q_u(\phi_{su}(y))\, du < \infty$ for all $s < t$, all y, and that $\widetilde{p}_t(y, \cdot) \ll p_t(y, \cdot)$ for all t, y.*

In that case a $\left(\mathcal{F}_t^X\right)$ −adapted version of the likelihood process $\mathfrak{L}^{\mu}$ is $\mathfrak{L}^X = \left(\mathfrak{L}_t^X\right)_{t \geq 0}$ given by

$$\mathfrak{L}_t^X =$$

$$\exp\left(-\int_0^t \left(\widetilde{q}_s(X_s) - q(X_s)\right) ds\right) \prod_{n=1}^{\overline{N}_t^X} \frac{\widetilde{q}_{T_n^X}\left(X_{T_n^X-}\right)}{q_{T_n^X}\left(X_{T_n^X-}\right)} \frac{d\widetilde{p}_{T_n^X}\left(X_{T_n^X-}, \cdot\right)}{dp_{T_n^X}\left(X_{T_n^X-}, \cdot\right)} \left(X_{T_n^X}\right).$$

Notation. Of course $\overline{N}_t^X$ is the total number of jumps for X on $[0, t]$ and T_n^X is the time of the n'th jump for X. Note that $\overline{N}^X$ is P−indistinguishable from $\overline{N}$ and that for all n, $T_n^X = T_n$ P−a.s.

We have not given Theorem 6.4.1 in its most general form, which would amount to a direct translation of Theorem 4.0.2. The result gives (in special cases) the likelihood function for observing a PDMP completely on an interval

$[0,t]$. Formally, that likelihood is the Radon-Nikodym derivative between the distributions on finite intervals of two different PDMP's, and here it is important to emphasize that for the likelihood to make sense at all (i.e. local absolute continuity to hold), the piecewise deterministic behaviour of the two processes must be the same because (some of) the ϕ_{st} can be read off from the sample path. Thus, as was done above, while the q_t, p_t can change into $\widetilde{q}_t, \widetilde{p}_t$, the ϕ_{st} must remain the unchanged when switching from one process to the other.

6.5 Differential equations for transitions

Let X be a time-homogeneous PDMP with state space G as in Theorem 6.1.2 (b) with $X_0 \equiv x_0$ for an arbitrary x_0, and let $h : G \to \mathbb{R}$ be bounded and measurable. Consider the function $\mathfrak{P}_t h : G \to \mathbb{R}$ given by

$$\mathfrak{P}_t h\left(x_0\right) = E^{x_0} h\left(X_t\right),$$

i.e. $\mathfrak{P}_t$ is the *transition operator* for X for time intervals of length t, related to the generator A through the formal expression

$$A = \lim_{t \to 0} \frac{1}{t}\left(\mathfrak{P}_t - \mathrm{id}\right),$$

cf. (6.24). The transition operators form a semigroup, $\mathfrak{P}_{s+t} = \mathfrak{P}_s \mathfrak{P}_t$, as is (essentially) seen from the Chapman-Kolmogorov equations (5.26) in the time-homogeneous case.

Writing x rather than x_0 for the arbitrary initial state, we shall first quote the backward integral equations for computing $\mathfrak{P}_t h(x)$.

For $n \in \mathbb{N}_0$, define

$$\mathfrak{P}_t^{(n)} h(x) = E^x \left(h\left(X_t\right) 1_{\left(\overline{N}_t = n\right)}\right).$$

Then clearly

$$\mathfrak{P}_t^{(0)} h(x) = h\left(\phi_t\left(x\right)\right) \exp\left(-\int_0^t q\left(\phi_s\left(x\right)\right)\, ds\right) \tag{6.33}$$

and recursively, conditioning on the time of the first jump and using Lemma 3.3.3 (bii),

$$\mathfrak{P}_t^{(n+1)} h(x) = \int_0^t ds\, f_q\left(s, x\right) \int_G p\left(\phi_s\left(x\right), dy\right)\, \mathfrak{P}_{t-s}^{(n)} h(y),$$

where

$$f_q(s,x) = q\left(\phi_s\left(x\right)\right)\exp\left(-\int_0^s q\left(\phi_u\left(x\right)\right)\,du\right)$$

is the density for the distribution of the first jump time T_1. Summing on n finally gives

$$\begin{aligned}\mathfrak{P}_t h(x) &= h\left(\phi_t\left(x\right)\right)\exp\left(-\int_0^t q\left(\phi_s\left(x\right)\right)\,ds\right)\\ &\quad +\int_0^t ds\, f_q\left(s,x\right)\int_G p\left(\phi_s\left(x\right),dy\right)\,\mathfrak{P}_{t-s}h(x).\end{aligned}$$

In the Markov chain case ($\phi_t(x) = x$) this after differentiation with respect to t yields the famous backward Feller-Kolmogorov differential equations. No simple analogue of these are available for general PDMP's, although for h smooth enough,

$$\frac{d}{dt}\mathfrak{P}_t h(x) = A\left(\mathfrak{P}_t h\right)(x),$$

(see (6.23) for the definition of the generator A).

We shall also quote the forward equations although they are more difficult to derive and less transparent. For $n \geq 0$, conditioning on Z_{n+1} we first find

$$\mathfrak{P}_t^{(n+1)}h(x) =$$

$$E^x 1_{(T_{n+1}\leq t)} h\left(\phi_{t-T_{n+1}}\left(Y_{n+1}\right)\right)\exp\left(-\int_{T_{n+1}}^{t} q\left(\phi_{u-T_{n+1}}\left(Y_{n+1}\right)\right)\,du\right).$$

Next, condition on Z_n to obtain

$$\begin{aligned}= E^x 1_{(T_n\leq t)}\int_{T_n}^t ds\, f_q(s-T_n,Y_n) &\int_E p\left(\phi_{s-T_n}\left(Y_n\right),dy\right)\\ &h\left(\phi_{t-s}(y)\right)\exp\left(-\int_s^t q\left(\phi_{u-s}(y)\right)\,du\right).\end{aligned}$$

Use the device $1_{(T_n\leq t)}\int_{T_n}^t ds = \int_0^t ds\, 1_{(T_n\leq s)}$ to take the integral with respect to x outside the expectation, and then use

$$f_q\left(s-T_n,Y_n\right) = q\left(\phi_{s-T_n}\left(Y_n\right)\right) P^x\left(T_{n+1}>s\,|Z_n\right)$$

on $(T_n \leq s)$ to obtain, cf. (6.33),

$$\begin{aligned}&= \int_0^t ds\, E^x 1_{\left(\overline{N}_s=n\right)} q\left(\phi_{s-T_n}\left(Y_n\right)\right)\int_E p\left(\phi_{s-T_n}\left(Y_n\right),dy\right)\,\mathfrak{P}_{t-s}^{(0)}h(y)\\ &= \int_0^t ds\, E^x 1_{\left(\overline{N}_s=n\right)} q\left(X_s\right)\int_E p\left(X_s,dy\right)\,\mathfrak{P}_{t-s}^{(0)}h(y).\end{aligned}$$

Thus, changing s to $t-s$,

$$\mathfrak{P}_t^{(n+1)} h(x) = \int_0^t ds\, \mathfrak{P}_{t-s}^{(n)} \left(q\left(\cdot, \mathfrak{P}_s^{(0)} h \right)\right)(x),$$

where

$$q\left(\cdot, \widetilde{h} \right)(\widetilde{x}) := q(\widetilde{x}) \int_E p(\widetilde{x}, dy)\, \widetilde{h}(y).$$

Summing on n finally gives

$$\begin{aligned} \mathfrak{P}_t h(x) &= h(\phi_t(x)) \exp\left(-\int_0^t q(\phi_s(x))\, ds \right) \\ &\quad + \int_0^t ds\, \mathfrak{P}_{t-s} \left(q\left(\cdot, \mathfrak{P}_s^{(0)} h \right)\right)(x). \end{aligned}$$

For Markov chains on a state space which is at most countable, after differentiation this gives the forward Feller-Kolmogorov differential equations for the transition probabilities.

Appendix A

Differentiation of cadlag functions

Let A be a positive measure on $(\mathbb{R}_0, \mathcal{B}_0)$ such that $A(t) := A([0,t]) < \infty$ for all t and $A(0) = 0$. Let f be a Borel function on $\mathbb{R}_0$ such that $\int_{]0,t]} |f| \; dA < \infty$ for all t, and define

$$F(t) = \int_{]0,t]} f(s)\, A(ds). \tag{A.1}$$

Clearly F is cadlag (right-continuous with left limits), $F(0) = 0$ and F inherits the following properties of A: (i) if for some $s < t$, $A(s) = A(t)$ also $F(s) = F(t)$ (and F is constant on $]s,t]$); (ii) if $\Delta F(t) \neq 0$ also $\Delta A(t) > 0$.

It is natural to say that F *is absolutely continuous with respect to* A *with Radon-Nikodym derivative* f. For us it is of particular interest that the derivative may be computed in a certain way.

For any function g on $\mathbb{R}_0$ define for $k, K \in \mathbb{N}$, $g_{k,K} = g\left(\frac{k}{2^K}\right)$ and g_K as the function

$$g_K = \sum_{k=1}^{\infty} 1_{\left]\frac{k-1}{2^K}, \frac{k}{2^K}\right]} \frac{g_{k,K} - g_{k-1,K}}{A_{k,K} - A_{k-1,K}}$$

with the convention $\frac{x}{0} = 0$ (where in practice we shall only have to worry about $\frac{0}{0}$).

Proposition A.0.1 *If F is given by (A.1), then $\lim_{K\to\infty} F_K(t) = f(t)$ for $A-$a.a. t and $\lim_{K\to\infty} \int_{]0,t]} |F_K - f| \; dA = 0$ for all t.*

Proof. Suppose that A is a probability measure and that $\int_{\mathbb{R}_0} |f| \; dA < \infty$ and let $\mathcal{G}_K = \sigma\left(\left]\frac{k-1}{2^K}, \frac{k}{2^K}\right]\right)_{k\geq 1}$. Then $E\left(f \,|\mathcal{G}_K\right) = F_K$, hence $\left(F_K\right)_{K\geq 1}$ is a uniformly integrable martingale on $(\mathbb{R}_0, \mathcal{B}_0, A)$ converging $A-$a.e. (and in

$L^1(A)$) to $E(f\,|\mathcal{G}_\infty)$ where $\mathcal{G}_\infty$ is the smallest $\sigma-$algebra containing all the $\mathcal{G}_K$, i.e. $\mathcal{G}_\infty = \mathcal{B}_0$ so that $E(f\,|\mathcal{G}_\infty) = f$.

The argument obviously applies also if A is a bounded measure with $\int |f|\, dA < \infty$, and for general A and f with $\int_{[0,t]} |f|\, dA < \infty$ for all t, by e.g. considering the restriction of A to $[0, t_0]$ for an arbitrarily large t_0, it is seen that the $A-$a.e. convergence and the L^1-convergence on $[0, t_0]$ remains valid. ■

Let now $F : \mathbb{R}_0 \to \mathbb{R}$ be a cadlag function, but not apriori of the form (A.1). We say that F *is (pointwise) differentiable with respect to* A if

(i) whenever $A(s) = A(t)$ for some $s < t$, also $F(s) = F(t)$;

(ii) whenever $\Delta F(t) \neq 0$ also $\Delta A(t) > 0$;

(iii) $\lim_{K\to\infty} F_K(t) = f(t)$ exists for $A-$a.a. t.

We then call f *the derivative of* F *with respect to* A, and write $f = D_A F$.

Proposition A.0.2 *If the cadlag function F is pointwise differentiable with respect to A with $D_A F = f$, then a sufficient condition for*

$$F(t) = F(0) + \int_{]0,t]} f(s)\, A(ds) \quad (t \in \mathbb{R}_0) \tag{A.2}$$

to hold, is that f be bounded on finite intervals and that

$$|F(t) - F(s)| \le \int_{]s,t]} |f|\, dA \tag{A.3}$$

for all $s < t$.

Note. Clearly (A.3) is necessary for (A.2) to hold.

Proof. Given t, write $t_K = \left[2^K t\right]/2^K$ (where $[x]$ is the integer part of x), $\widetilde{t}_K = t_K + \frac{1}{2^K}$. Using (A.3) we find for $s \le t$,

$$|F_K(s)| \le \sum_{k=1}^{\infty} 1_{\left]\frac{k-1}{2^K},\frac{k}{2^K}\right]}(s)\, \frac{\int_{\left]\frac{k-1}{2^K},\frac{k}{2^K}\right]} |f|\, dA}{A_{k,K} - A_{k-1,K}} \le \sup_{[0,t]} |f|$$

and hence, by dominated convergence,

$$\int_{]0,t]} f\, dA = \lim_{K\to\infty} \int_{]0,t]} F_K\, dA.$$

But

$$\begin{aligned}
\int_{]0,t]} F_K \, dA &= \sum_{k=1}^{\infty} \left(F_{k,K} - F_{k,K-1}\right) \frac{A\left(\frac{k}{2^K} \wedge t\right) - A\left(\frac{k-1}{2^K} \wedge t\right)}{A_{k,K} - A_{k-1,K}} \\
&= \sum_{k=1}^{\left[2^K t\right]} \left(F_{k,K} - F_{k-1,K}\right) + \left(F\left(\widetilde{t}_K\right) - F\left(t_K\right)\right) \frac{A(t) - A\left(t_K\right)}{A\left(\widetilde{t}_K\right) - A\left(t_K\right)} \\
&= F\left(\widetilde{t}_K\right) - F(0) - \left(F\left(\widetilde{t}_K\right) - F\left(t_K\right)\right) \frac{A\left(\widetilde{t}_K\right) - A(t)}{A\left(\widetilde{t}_K\right) - A\left(t_K\right)}
\end{aligned}$$

where (i) above has been used for the second equality. Since F is right-continuous, $F\left(\widetilde{t}_K\right) \to F(t)$ as $K \to \infty$. If $\Delta A(t) = 0$, by (ii) F is continuous at t and so also $F\left(t_K\right) \to F(t)$ and since the ratio involving the $A-$increments is bounded by 1, we have convergence to $F(t) - F(0)$ of the entire expression. If $\Delta A(t) > 0$, let c be an upper bound for $|F|$ on $[0, t+1]$ and use

$$\left| \left(F\left(\widetilde{t}_K\right) - F\left(t_K\right)\right) \frac{A\left(\widetilde{t}_K\right) - A(t)}{A\left(\widetilde{t}_K\right) - A\left(t_K\right)} \right| \leq 2c \frac{A\left(\widetilde{t}_K\right) - A(t)}{\Delta A(t)} \to 0$$

to again obtain convergence to $F(t) - F(0)$. ■

The following useful differentiation rule is easily proved: if F_1, F_2 are differentiable with respect to A, so is the product $F_1 F_2$ and

$$D_A\left(F_1 F_2\right)(t) = \left(D_A F_1\right)(t) F_2(t) + F_1(t-)\left(D_A F_2\right)(t).$$

Note that this expression is not symmetric in the indices $1, 2$. Switching between 1 and 2 gives an alternative expression for the same derivative. In practice one expression may prove more useful than the other. If the conditions from Proposition A.0.2 are satisfied for $\left(F_i, D_A F_i\right)$, $i = 1, 2$, one obtains the partial integration formula

$$\left(F_1 F_2\right)(t) = \left(F_1 F_2\right)(0) + \int_{]0,t]} \left(\left(D_A F_1\right)(s) F_2(s) + F_1(s-)\left(D_A F_2\right)(s)\right) A(ds).$$

Appendix B

Filtrations, processes, martingales.

We shall quickly go through some of the basics from the general theory of stochastic processes. All results below are quoted without proofs.

A *filtered probability space* is a quadruple $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, where $(\Omega, \mathcal{F}, P)$ is a probability space and $(\mathcal{F}_t)_{t\geq 0}$, the *filtration*, is an increasing family of sub $\sigma-$algebras of $\mathcal{F}$, $\mathcal{F}_s \subset \mathcal{F}_t$ if $s \leq t$. (Notation: $\subset$ means 'contained in or equal to').

A probability space is *complete* if any subset of a $P-$null set is measurable: if $F_0 \in \mathcal{F}$, $P(F_0) = 0$, then $F \in \mathcal{F}$ for any $F \subset F_0$.

A probability space $(\Omega, \mathcal{F}, P)$ may always be completed: define $\mathcal{N} = \{F \subset \Omega : \exists F_0 \in \mathcal{F} \text{ with } P(F_0) = 0 \text{ such that } F \subset F_0\}$, and let $\overline{\mathcal{F}}$ be the smallest $\sigma-$algebra containing $\mathcal{F}$ and $\mathcal{N}$. Then $\overline{\mathcal{F}} = \{F \cup N : F \in \mathcal{F},\ N \in \mathcal{N}\}$ and P extends uniquely to a probability $\overline{P}$ on $(\Omega, \overline{\mathcal{F}})$ using the definition $\overline{P}(\overline{F}) = P(F)$ for any $\overline{F} \in \overline{\mathcal{F}}$ and any representation $\overline{F} = F \cup N$ of F with $\overline{F} \in \mathcal{F}$, $N \in \mathcal{N}$. The probability space $(\Omega, \overline{\mathcal{F}}, \overline{P})$ is complete and is called the *completion* of $(\Omega, \mathcal{F}, P)$.

A filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ satisfies the *usual conditions* if $(\Omega, \mathcal{F}, P)$ is complete, if $\mathcal{N} \subset \mathcal{F}_0$ where now $\mathcal{N} = \{N \in \mathcal{F} : P(N) = 0\}$, and if the filtration is right-continuous,

$$\mathcal{F}_t = \mathcal{F}_{t+} \quad (t \geq 0),$$

where $\mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s$.

Much of the literature on process theory presents results and definitions, assuming that the usual conditions are satisfied. We shall not make this assumption here, although we have in mind applications where the filtration is automatically right-continuous, but where we do not complete the $\sigma-$algebras.

Let $(G, \mathcal{G})$ be a measurable space. A *stochastic process* (in continuous time) with state space $(G, \mathcal{G})$, defined on $(\Omega, \mathcal{F}, P)$, is a family $X = (X_t)_{t \geq 0}$ of random variables $X_t : (\Omega, \mathcal{F}) \to (G, \mathcal{G})$. A stochastic process X is *measurable*, if the map

$$(t, \omega) \to X_t(\omega)$$

from $(\mathbb{R}_0 \times \Omega, \mathcal{B}_0 \otimes \mathcal{F})$ to $(G, \mathcal{G})$ is measurable.

The filtration generated by the process X is the family $(\mathcal{F}_t^X)_{t \geq 0}$ of $\sigma-$algebras, where $\mathcal{F}_t^X = \sigma(X_s)_{0 \leq s \leq t}$.

A process X with a state space which is a measurable subspace of $(\mathbb{R}^d, \mathbb{R}^d)$ is *right-continuous* if $t \to X_t(\omega)$ is right-continuous for $P-$almost all ω. Similarly, X is *left-continuous, cadlag, increasing, continuous* if for $P-$almost all ω, $t \to X_t(\omega)$ is respectively left-continuous, cadlag (right-continuous with left limits), increasing (in each of the d coordinates), continuous.

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a filtered probability space. A process X defined on $(\Omega, \mathcal{F}, P)$ is *adapted* if it is measurable and each $X_t : (\Omega, \mathcal{F}) \to (G, \mathcal{G})$ is $\mathcal{F}_t-$measurable. X is *predictable* (or *previsible*) if it is measurable, X_0 is $\mathcal{F}_0-$measurable and the map $(t, \omega) \to X_t(\omega)$ from $(\mathbb{R}_+ \times \Omega, \mathcal{B}_+ \otimes \mathcal{F})$ to $(G, \mathcal{G})$ is $\mathcal{P}-$measurable, where $\mathcal{P}$, the $\sigma-$algebra of *predictable sets*, is the sub $\sigma-$algebra of $\mathcal{B}_+ \otimes \mathcal{F}$ generated by the subsets of the form

$$]s, \infty[\times F \quad (s \in \mathbb{R}_0, F \in \mathcal{F}_s).$$

If X, X' are two processes on $(\Omega, \mathcal{F}, P)$ with state space $(G, \mathcal{G})$, they are *versions* of each other if for all t, $P(X_t = X'_t) = 1$. They are *indistinguishable* if $F_0 := \bigcap_{t \geq 0} (X_t = X'_t) \in \mathcal{F}$ and $P(F_0) = 1$.

Proposition B.0.3 *Let X be a process on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ with state space $(G, \mathcal{G}) \subset (\mathbb{R}^d, \mathcal{B}^d)$.*

(i) *If X is right-continuous or left-continuous, X is indistinguishable from a measurable process.*

(ii) *If X is right-continuous and each X_t is $\mathcal{F}_t-$measurable, X is indistinguishable from an adapted process.*

(iii) *If X is left-continuous and X_t is $\mathcal{F}_t-$measurable for all t, X is indistinguishable from a predictable process.*

Notation used above: $(G, \mathcal{G}) \subset (\mathbb{R}^d, \mathcal{B}^d)$ means that $(G, \mathcal{G})$ is a measurable subspace of $(\mathbb{R}^d, \mathcal{B}^d)$.

We shall now proceed to define martingales and submartingales in continuous time.

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a filtered probability space and let X be a real valued process (state space $(G, \mathcal{G}) \subset (\mathbb{R}, \mathcal{B})$).

Definition B.0.1 *X is a* martingale (submartingale) *if for all t, $E|X_t| < \infty$, X_t is $\mathcal{F}_t-$measurable and*

$$E(X_t|\mathcal{F}_s) = X_s \quad (0 \leq s \leq t)$$
$$(E(X_t|\mathcal{F}_s) \geq X_s \quad (0 \leq s \leq t)).$$

X is a supermartingale *if $-X$ is a submartingale.*

The definition depends in a crucial manner on the underlying filtration. We shall therefore include the filtration in the notation and e.g. write that $(X_t, \mathcal{F}_t)$ is a martingale.

The next result describes transformations that turn (sub)martingales into submartingales.

Proposition B.0.4 *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be convex.*

(a) If $(M_t, \mathcal{F}_t)$ is a martingale and $E|\varphi(M_t)| < \infty$ for all t, then $(\varphi(M_t), \mathcal{F}_t)$ is a submartingale.

(b) If $(X_t, \mathcal{F}_t)$ is a submartingale, ϕ is increasing (and convex) and $E|\varphi(X_t)| < \infty$ for all t, then $(\varphi(X_t), \mathcal{F}_t)$ is a submartingale.

Notation. If f is a $\mathbb{R}-$valued function, write f^+, f^- for the positive and negative part of f : $f^+ = f \vee 0$, $f^- = -(f \wedge 0)$.

Proposition B.0.5 *Let $(X_t, \mathcal{F}_t)$ be a submartingale.*

(a) If $t > 0$ and $D \subset [0, t]$ is at most countable, then for every $x > 0$

$$P\left(\sup_{s \in D} X_s > x\right) \leq \frac{1}{x} EX_t^+$$
$$P\left(\inf_{s \in D} X_s < -x\right) \leq \frac{1}{x}\left(EX_t^+ - EX_0\right)$$

(b) If in addition X is right-continuous or left-continuous, for all $t \geq 0$, $x > 0$

$$P(\sup_{s \leq t} X_s > x) \leq \frac{1}{x} EX_t^+$$
$$P(\inf_{s \leq t} X_s < -x) \leq \frac{1}{x}(E_t^+ - EX_0).$$

Let $D \subset \mathbb{R}_0$, and let $f : D \to \mathbb{R}$ be a function. For $a < b \in \mathbb{R}$, the number of *upcrossings* from a to b for f on D is defined as

$$\beta_D(f;a,b) = \sup\{n \in \mathbb{N}_0 : \quad \exists t_1 < t_2 < \cdots < t_{2n} \in D \\ \text{with } f(t_{2k-1}) < a < b < f(t_{2k}),\ 1 \le k \le n\}.$$

The following analytic fact is a basic tool for establishing the main theorem on continuous time martingales.

Lemma B.0.6 *Let* $f : \mathbb{Q}_0 \to \mathbb{R}$.

(a) *The following two conditions are equivalent:*

(i) *the limits*

$$f(t+) := \lim_{s \downarrow t,\ s>t} f(s), \quad f(t-) := \lim_{s \uparrow t,\ s<t} f(s)$$

exists as limits in $\overline{\mathbb{R}}$, *simultaneously for all* $t \in \mathbb{R}_0$ *in the case of* $f(t+)$ *and for all* $t \in \mathbb{R}_+$ *in the case of* $f(t-)$.

(ii) $\beta_{\mathbb{Q}_o \cap [0,t]}(f;a,b) < \infty \qquad (t \in \mathbb{R}_0,\ a < b \in \mathbb{R})$.

(b) *If (i), (ii) are satisfied, then the function* $t \to f(t+)$ *from* $\mathbb{R}_0$ *to* $\overline{\mathbb{R}}$ *is cadlag.*

In order to show that there are finitely many upcrossings, one uses

Lemma B.0.7 *Let* $(X_t, \mathcal{F}_t)$ *be a submartingale, let* $t \in \mathbb{R}_+$ *and let* $D \subset [0,t]$ *be at most countable. For all* $a < b \in \mathbb{R}$, $\beta_D(X;a,b)$ *is then a* $\mathcal{F}_t-$*measurable random variable and*

$$E\beta_D(X;a,b) \le \frac{1}{b-a} E(X_t - a)^+.$$

We are now ready to formulate the main theorem for martingales and submartingales in continuous time.

Theorem B.0.8 (a) *Let* $(M_t, \mathcal{F}_t)$ *be a martingale.*

(i) *For* $P-$*almost all* ω *the limits*

$$M_{t+}(\omega) := \lim_{s \downarrow\downarrow t,\ s \in \mathbb{Q}_0} M_s(\omega), \quad M_{t-}(\omega) := \lim_{s \uparrow\uparrow t,\ s \in \mathbb{Q}_0} M_s(\omega)$$

exists as limits in $\mathbb{R}$, *simultaneously for all* $t \ge 0$ *in the case of* $M_{t+}(\omega)$ *and for all* $t > 0$ *in the case of* $M_{t-}(\omega)$. *Moreover, for every* t,

$$E|M_{t+}| < \infty, \quad E|M_{t-}| < \infty.$$

(ii) *For all* $t \geq 0$, $E(M_{t+}|\mathcal{F}_t) = M_t$ *a.s., and for all* $0 \leq s < t$, $E(M_t|\mathcal{F}_{s+}) = M_{s+}$ *a.s. Moreover, for a given* $t \geq 0$, $M_{t+} = M_t$ *holds a.s. if one of the two following conditions are satisfied:*

$$\begin{array}{ll} (*) & \mathcal{F}_{t+} = \mathcal{F}_t, \\ (**) & M \text{ is right-continuous in probability at } t. \end{array} \tag{B.1}$$

(iii) *The process* $M_+ := (M_{t+})_{t\geq 0}$ *may be chosen in such a way that all* M_{t+} *are* $\mathcal{F}_{t+}-$*measurable, in which case* $(M_{t+}, \mathcal{F}_{t+})$ *is a cadlag martingale. Furthermore, if one of the conditions in (B.1) is satisfied for every* $t \geq 0$, *then* M_+ *is a version of* M.

(iv) *If* $\sup_{t\geq 0} EM_t^+ < \infty$ *or if* $\sup_{t\geq 0} EM_t^- < \infty$, *then*

$$M_\infty := \lim_{t\to\infty,\ t\in\mathbb{Q}_0} M_t = \lim_{t\to\infty} M_{t+}$$

exists a.s. and $E|M_\infty| < \infty$.

(b) *Let* $X = (X_t, \mathcal{F}_t)$ *be a submartingale.*

(i) *For* $P-$*almost all* ω *the limits*

$$X_{t+}(\omega) := \lim_{s\downarrow\downarrow t,\ s\in\mathbb{Q}_0} X_s(\omega), \quad X_{t-}(\omega) := \lim_{s\uparrow\uparrow t,\ s\in\mathbb{Q}_0} X_s(\omega)$$

exists as limits in $\mathbb{R}$, *simultaneously for all* $t \geq 0$ *in the case of* $X_{t+}(\omega)$ *and for all* $t > 0$ *in the case of* $X_{t-}(\omega)$. *Moreover, for every* t,

$$E|X_{t+}| < \infty, \quad E|X_{t-}| < \infty.$$

(ii) *For all* $t \geq 0$, $E(X_{t+}|\mathcal{F}_t) \geq X_t$ *a.s., and for all* $0 \leq s < t$, $E(X_t|\mathcal{F}_{s+}) \geq X_{s+}$ *a.s. Moreover, for a given* $t \geq 0$, $X_{t+} \geq X_t$ *holds a.s. if (B.2∗) is satisfied and* $X_{t+} = X_t$ *holds a.s. if (B.2∗∗) is satisfied:*

$$\begin{array}{ll} (*) & \mathcal{F}_{t+} = \mathcal{F}_t, \\ (**) & X \text{ is right-continuous in probability at } t. \end{array} \tag{B.2}$$

(iii) *The process* $X_+ := (X_{t+})_{t\geq 0}$ *may be chosen in such a way that all* X_{t+} *are* $\mathcal{F}_{t+}-$*measurable, in which case* $(X_{t+}, \mathcal{F}_{t+})$ *is a cadlag submartingale. Furthermore, if (B.2∗∗) is satisfied for every* $t \geq 0$, *then* X_+ *is a version of* X.

(iv) If $\sup_{t\geq 0} EX_t^+ < \infty$, *then*

$$X_\infty := \lim_{t\to\infty,\ t\in\mathbb{Q}_0} X_t = \lim_{t\to\infty} X_{t+}$$

exists a.s. and $E|X_\infty| < \infty$.

Note. A process $V = (V_t)_{t\geq 0}$ is *right-continuous in probability at* t, if for every sequence (t_n) with $t_n \geq t$, $\lim_{n\to\infty} t_n = t$ it holds that $V_{t_n} \to V_t$ in probability.

We next proceed with a brief discussion of stopping times and the optional sampling theorem.

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ be a filtered probability space. A map $\tau : \Omega \to \overline{\mathbb{R}}_0$ is a *stopping time* if

$$(\tau < t) \in \mathcal{F}_t \quad (t \in \mathbb{R}_+).$$

Let τ be a stopping time and define

$$\mathcal{F}_\tau := \{F \in \mathcal{F} : F \cap (\tau < t) \in \mathcal{F}_t \text{ for all } t \in \mathbb{R}_+\},$$

which is a $\sigma-$algebra. Note that if $\tau \equiv t_0$ for some $t_0 \in \mathbb{R}_0$, then $\mathcal{F}_\tau = \mathcal{F}_{t_0+}$. Also, if $\sigma \leq \tau$ are stopping times, then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.

If the filtration is right-continuous, $\mathcal{F}_{t+} = \mathcal{F}_t$ for all t, $\tau : \Omega \to \overline{\mathbb{R}}_0$ is a stopping time iff $(\tau \leq t) \in \mathcal{F}_t$ for all $t \in \mathbb{R}_0$ and $\mathcal{F}_\tau = \{F \in \mathcal{F} : F \cap (\tau \leq t) \in \mathcal{F}_t$ for all $t \in \mathbb{R}_0\}$. On general filtered spaces such τ are called *strict stopping times* and are special cases of stopping times.

Now let X be a $\mathbb{R}-$valued process defined on $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ and let τ be a stopping time. Define X_τ by

$$X_\tau(\omega) = X_{\tau(\omega)}(\omega)$$

if $\tau(\omega) < \infty$. This defines X_τ almost surely if $P(\tau < \infty) = 1$. If $P(\tau = \infty) > 0$ define X_τ (as an a.s. surely defined random variable) only in the case where $X_\infty = \lim_{t\to\infty} X_t$ exists a.s. and then put

$$X_\tau = X_\infty \quad \text{on} \quad (\tau = \infty).$$

Lemma B.0.9 *Assume that* $(\mathcal{F}_t)$ *is right-continuous and let* X *be a* $\mathbb{R}-$*valued process which is right-continuous and adapted, and let* τ *be a stopping time such that* X_τ *is defined almost surely. Then* X_τ *is a.s. equal to a* $\mathcal{F}_\tau-$*meas-urable random variable.*

The lemma justifies that one may always assume that X_τ (when defined) is $\mathcal{F}_\tau$−measurable if X is right-continuous and adapted.

For the statement of the next result, recall that a family $(U_i)_{i\in I}$ of real-valued random variables is *uniformly integrable* if (i) $\sup_{i\in I} E|U_i| < \infty$ and (ii) $\lim_{x\to\infty} \sup_{i\in I} \int_{(|U_i|>x)} |U_i|\, dP = 0$. In particular, (U_i) is uniformly integrable if (a) there exists a random variable $U \in L^1$ such that $P(|U_i| \le |U|) = 1$ for all i, (b) there exists $p > 1$ such that (U_i) is bounded in L^p: $\sup_{i\in I} E|U_i|^p < \infty$.

Theorem B.0.10 (Optional sampling). *Assume that the filtration $(\mathcal{F}_t)$ is right-continuous.*

(a) Let $(M_t, \mathcal{F}_t)$ be a right-continuous martingale and let $\sigma \le \tau$ be stopping times.

If either of the following two conditions (i), (ii) are satisfied, then $E|M_\sigma| < \infty$, $E|M_\tau| < \infty$ and

$$E(M_\tau|\mathcal{F}_\sigma) = M_\sigma; \tag{B.3}$$

(i) τ is bounded.

(ii) $(M_t)_{t\ge0}$ is uniformly integrable.

(b) Let $(X_t, \mathcal{F}_t)$ be a right-continuous submartingale and let $\sigma \le \tau$ be stopping times.

If either of the following two conditions (i), (ii) are satisfied, then $E|X_\sigma| < \infty$, $E|X_\tau| < \infty$ and

$$E(X_\tau|\mathcal{F}_\sigma) \ge X_\sigma; \tag{B.4}$$

(i) τ is bounded.

(ii) $(X_t^+)_{t\ge0}$ is uniformly integrable.

Note that if (aii) (or (bii)) holds, then $\lim_{t\to\infty} M_t$ ($\lim_{t\to\infty} X_t$) exists a.s. and M_τ (respectively X_τ) is well defined for any stopping time τ. Thus, with (aii) satisfied (B.3) holds for all pairs $\sigma \le \tau$ of stopping times, while if (bii) is satisfied (B.4) holds for all pairs $\sigma \le \tau$.

We finally need to discuss local martingales. We still assume that the filtration is right-continuous.

First note that if M is a cadlag $\mathcal{F}_t-$martingale and τ is a stopping time, then M^τ is also a $\mathcal{F}_t-$martingale, where *M^τ is M stopped at τ*,

$$M_t^\tau := M_{\tau \wedge t}.$$

(Note that by optional sampling, for $s < t$

$$E\left(M_t^\tau \left| \mathcal{F}_{\tau\wedge s}\right.\right) = M_s^\tau.$$

To obtain the stronger result

$$E\left(M_t^\tau \left| \mathcal{F}_s\right.\right) = M_s^\tau,$$

one shows that if $F \in \mathcal{F}_s$, then $F \cap (\tau > s) \in \mathcal{F}_{\tau\wedge s}$ and therefore

$$\begin{aligned}
\int_F M_t^\tau \, dP &= \int_{F\cap(\tau>s)} M_t^\tau \, dP + \int_{F\cap(\tau\le s)} M_\tau \, dP \\
&= \int_{F\cap(\tau>s)} M_s^\tau \, dP + \int_{F\cap(\tau\le s)} M_s^\tau \, dP \\
&= \int_F M_s^\tau \, dP.)
\end{aligned}$$

Definition B.0.2 *An adapted, $\mathbb{R}-$valued cadlag process M is a* local $\mathcal{F}_t-$martingale *if there exists a sequence $(\tau_n)_{n\geq 1}$ of stopping times, increasing to ∞ a.s., such that for every n, M^{τ_n} is a $\mathcal{F}_t-$martingale.*

That (τ_n) increases to ∞ a.s. means that for all n, $\tau_n \leq \tau_{n+1}$ a.s. and that $\lim_{n\to\infty} \tau_n = \infty$ a.s. The sequence (τ_n) is called a *reducing sequence* for the local martingale M, and we write that M is a local $\mathcal{F}_t-$martingale (τ_n).

Clearly any martingale is also a local martingale (use $\tau_n \equiv \infty$ for all n). If (τ_n) is a reducing sequence and (ρ_n) is a sequence of stopping times increasing to ∞ a.s., since $(M^{\tau_n})^{\rho_n} = M^{\tau_n \wedge \rho_n}$ it follows immediately that $(\tau_n \wedge \rho_n)$ is also a reducing sequence.

It is often important to be able to show that a local martingale is a true martingale. This may be very difficult, but a useful criterion is the following:

Proposition B.0.11 *Let M be a local $\mathcal{F}_t-$martingale. For M to be a $\mathcal{F}_t-$martingale it is sufficient that for all t,*

$$E \sup_{s:s\leq t} |M_s| < \infty. \tag{B.5}$$

Warning. A local martingale M need not be a martingale even though $E\,|M_t| < \infty$ for all t. There are even examples of local martingales M that are not martingales although the exponential moments $E\exp(\theta\,|M_t|)$ are $< \infty$ for *all* $t, \theta \geq 0$. Thus moment conditions on the individual M_t are not sufficient to argue that a local martingale is a true martingale – some kind of uniformity as in (B.5) is required.

Bibliography

[1] Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1992). *Statistical Models based on Counting Processes.* Springer, New York.

[2] Asmussen, S. (1987). *Applied Probability and Queues.* Wiley, Chichester.

[3] Baccelli, F. and Brémaud, P. (1994). *Elements of Queueing Theory.* Springer, New York.

[4] Boel, R., Varaiya, P. and Wong, E. (1975). Martingales on Jump Processes, Part I: Representation Results, Part II: Applications. *SIAM J. Control* **13,** 999–1061.

[5] Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics.* Springer, Berlin.

[6] Daley, D.J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes.* Springer, New York.

[7] Davis, M.H.A. (1984). Piecewise-deterministic Markov Processes: a General Class of Non-diffusion Stochastic Models (with discussion). *J. Royal Statist. Soc. (B)* **46,** 353–388.

[8] Davis, M.H.A.(1993). *Markov Models and Optimization.* Chapman & Hall, London.

[9] Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.

[10] Jacobsen, M. *Statistical Analysis of Counting Processes.* Lecture Notes in Statistics 12, Springer, New York.

[11] M. Jacobsen (1989). Right Censoring and Martingale Methods for Failure Time Data. *Ann. Statist.* **17,** 1133–1156.

[12] Jacod, J. (1975). Multivariate Point Processes: Predictable Projection, Radon-Nikodym Derivatives, Representation of Martingales. *Z. Wahrsch. verw. Gebiete* **31,** 235-253.

[13] Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes.* Springer, Berlin.

[14] Kabanov, Y., Liptser, R.S. and Shiryaev, A.N. (1976). Criteria of Absolute Continuity of Measures corresponding to Multivariate Point Processes. In *Proc. 3rd Japan-USSR Symposium on Probability.* Lecture Notes in Mathematics 550, Springer, Berlin.

[15] Kallenberg, O. (1983). *Random Measures.* Akademie-Verlag, Berlin and Academic Press, London.

[16] Last, G. and Brandt, A. (1995). *Marked Point Processes on the Real Line. The Dynamic Approach.* Springer, New York.

[17] Lindvall, T. (1988). *Lectures on the Coupling Method.* Wiley, New York.

[18] Liptser, R.S. and Shiryaev, A.N. (1978). *Statistics of Random Processes, Vol. II Applications.* Springer, Berlin.

[19] Neveu, J. (1977). Processus ponctuels. *École d'Eté de Probabilités de Saint-Flour VI,* Lecture Notes in Mathematics 598, 249–445, Springer, Berlin.

[20] Sigman, K. (1995). *Stationary Marked Point Processes. An Intuitive Approach.* Chapman & Hall, New York.