Probability and Its Applications

Martin Jacobsen

Point Process Theory and Applications

Marked Point and Piecewise Deterministic Processes

Birkhäuser

Probability and its Applications

Series Editors

Charles Newman
Sidney Resnick

Martin Jacobsen

Point Process Theory and Applications

Marked Point and Piecewise Deterministic Processes

Birkhäuser
Boston • Basel • Berlin

Martin Jacobsen
University of Copenhagen
Institute of Mathematical Sciences
Department of Applied Mathematics and Statistics
5 Universitetsparken
DK-2100 Copenhagen Ø
Denmark

Mathematics Subject Classification (2000): 60G07, 60G44, 60G55, 60G57, 60H05, 60J25, 60J35, 60J75, 60J80, 60K25, 62N01, 91B28, 91B30 (primary); 60G30, 60G40, 60G51, 60J57, 60K15 (secondary)

Library of Congress Control Number: 2005934409

ISBN-10 0-8176-4215-3 eISBN 0-8176-4463-6
ISBN-13 978-0-8176-4215-0

Printed on acid-free paper.

Birkhäuser

Printed in the United States of America. (TXQ/EB)

9 8 7 6 5 4 3 2 1

www.birkhauser.com

Contents

Preface

The book aims at presenting a detailed and mathematically rigorous exposition of the theory and applications of a class of point processes and piecewise deterministic processes. The framework is sufficiently general to unify the treatment of several classes of stochastic phenomena: point processes, Markov chains and other Markov processes in continuous time, semi-Markov processes, queueing and storage models, and likelihood processes. There are applications to finance, insurance and risk, population models, survival analysis, and congestion models. A major aim has been to show the versatility of piecewise deterministic Markov processes for applications and to show how they may also become useful in areas where thus far they have not been much in evidence.

Originally the plan was to develop a graduate text on marked point processes indexed by time which would focus on probabilistic structure and be essentially self-contained. However, it soon became apparent that the discussion should naturally include a traditional class of continuous time stochastic processes constructed from certain marked point processes. This class consists of 'piecewise deterministic processes'; that is, processes with finitely many jumps on finite time intervals which, roughly speaking, develop deterministically between the random jump times. The exposition starts with the point process theory and then uses this to treat the piecewise deterministic processes.

Throughout the focus is constructive, emphasizing canonical versions, which often means that a process is discussed relative to its own filtration, rather than with respect to a general and larger filtration to which the process may be adapted. Many of the main results are proved within this canonical setting, which makes it easier to develop the proofs. But of course these main results then appear only as special cases of well-established results from 'the general theory of processes'; even so, we believe that by treating the canonical setup directly, additional insight into the structure of the processes is gained.

Among the piecewise deterministic processes those that are Markov are especially important, and they are also the ones treated most thoroughly in this book. The pioneering work here was done by Mark Davis and many of his results duly reappear here—but again, basing everything on marked point process theory leads to a somewhat different approach and to a very general construction of not only time-homogeneous piecewise deterministic Markov processes (the ones considered by Davis), but also of those that are non-homogeneous.

The text is designed for advanced topics courses or self-study by graduate students who are at least in the fourth year of a European style degree program or at least in the second year of an American style Ph.D program. The text will also be useful to researchers specializing in the use of probabilistic models for point processes and piecewise deterministic processes. A course can easily be fashioned from selected parts of the book, and we suggest Chapters 2, 3, 4 and Sections 7.1–Chapter 7. This material should be supplemented by discussion of some of the models and applications treated in Part II.

The reader who wishes to master all details of the text will need a background in measure-theoretic probability theory. Readers with a narrower foundation will also benefit from reading the book. Short introductions to each chapter, apart from pointing to material that is considered essential, also list parts of the text (entire sections, technical proofs, etc) that may be omitted.

Acknowledgements

An initial version of this text started as lecture notes for a graduate course at the University of Copenhagen in 1995–1996 and went through successive revisions while lecturing at Copenhagen, the University of Aarhus (1998) and at Chalmers University of Technology and the University of Gothenburg (1999). Grateful acknowledgement is made to the University of Aarhus' Centre for Mathematical Physics and Stochastics (MaPhySto, a network funded by the Danish National Research Foundation), with special thanks to Ole Barndorff-Nielsen, the then director of the Centre. Many thanks also to Holger Rootzén for arranging my stay at the Stochastic Centre at Chalmers University of Technology and the University of Gothenburg in 1999 and for the support received from the Centre.

I am also much indebted to colleagues and students in Copenhagen and elsewhere for stimulation and feedback which contributed much to the improvement of the text. I especially thank Jacob Krabbe Pedersen, now at the University of Southern Denmark, who voluntarily undertook the huge task of converting the handwritten manuscript of the original Copenhagen notes into LaTeX. My gratitude to him, Bo Markussen and Anders Tolver Jensen (both from Copenhagen) for their many insightful comments, questions and suggestions. I am also very grateful to Kasper Larsen, presently at Carnegie Mellon, for reading, commenting on and improving the section on finance.

Sid Resnick from Cornell University first encouraged me to write a book based on the MaPhySto lecture notes. I would like to thank him and also the anonymous reviewers who offered many helpful suggestions. Finally, many thanks are due to the staff at Birkhäuser Boston, Ann Kostant in particular, for their help and patience during the slow progress of this project.

Martin Jacobsen Copenhagen, May 2005

Point Process Theory and Applications

Part I

Theory

1

Introduction

1.1 Overview

While point processes in general describe the random scattering of points in a certain space, the point processes considered in this book are special in the sense that the specific points are associated with the times at which certain random events occur. The resulting stochastic phenomena developing over time then form the main object of study. Just recording the time occurrences of the events results in a simple point process (SPP), a marked point process (MPP) is obtained if at each time occurrence the type or mark of the event is also registered. Apart from discussing the theory of such processes, one of the main points of the book is to show how certain 'ordinary' stochastic processes may be viewed as MPPs, e.g., by considering the time points at which a given stochastic process jumps, with the type of event or marks associated with each occurrence some characteristic of the jump. The result is a piecewise deterministic process (PDP), formally defined as a process adapted to the filtration generated by the MPP of timepoints of jumps and their characteristics.

In the first chapters, Chapters 2 and 3, the definition and construction of simple and marked point processes are given together with the equivalent description of, respectively, an SPP as a counting process (CP) and an MPP as a random counting measure (RCM). It is part of the definition that the processes should be stable, i.e., only finitely many events are allowed in finite intervals of time, but point processes with explosion, where infinitely many time points accumulate in finite time, arise naturally and they are therefore discussed as well, albeit somewhat summarily, at various places in the text.

The long Chapter 4 contains the basic theory of CPs and RCMs, including the characterization of their distributions through compensators, compensating measures and martingale properties. In Chapter 5 likelihood processes (processes of Radon–Nikodym derivatives) are treated. Chapter 6 treats stochastic dependence between finitely many RCMs as well as RCMs with independent increments.

One of the most important chapters, Chapter 7, develops the theory of piecewise deterministic Markov processes (PDMPs) using the theory of RCMs from the earlier chapters. In particular it investigates the description of general PDPs as MPPs using jump times and characteristics of jumps to describe the times when events occur and their marks.

The theoretical Part I of the book contains a number of examples and exercises, but the main examples are found in Part II, where it is shown how different aspects of the theory may be used to discuss models of relevance for topics in statistics and applied probability: survival analysis, branching processes, ruin probabilities, soccer (football), mathematical finance and queueing theory.

Part III contains two appendices: one on a frequently used differentiation technique and one on some of the fundamentals from the general theory of stochastic processes with an introduction to martingales in continuous time.

Each chapter starts with a short introduction in which some of the main references for the topics covered in the chapter are listed; the reader is alerted as to what material from the chapter is required reading and what is not. The bibliography with the bibliographical notes are found at the end of the book together with a notation glossary and subject index.

The emphasis on the time dynamics of the point processes under consideration already prevalent in the basic construction of SPPs and MPPs from Chapter 3 is *not* suited for the discussion of some important topics in the general theory of point processes such as palm probabilities, stationarity (although Chapter 7 does contain a section on stationarity of homogeneous PDMPs) and thinning of point processes. Also, as mentioned above, point processes where infinitely many points may accumulate within small bounded sets are excluded, with the exception of exploding SPPs and MPPs that allow for exactly one accumulation point.

Some notation: $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$. 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 σ-algebras are written $\mathcal{B}_0$, $\mathcal{B}_+$, $\overline{\mathcal{B}}_0$, $\overline{\mathcal{B}}_+$. In higher dimensions, $\mathcal{B}^d$ denotes the d-dimensional Borel σ-algebra of subsets of $\mathbb{R}^d$.

An abstract probability space is denoted $(\Omega, \mathcal{F}, \mathbb{P})$. The space of marks for a marked point process is written $(E, \mathcal{E})$, and the state space for a Markov process or other stochastic process is denoted by $(G, \mathcal{G})$. Probabilities on these and other concrete spaces are usually denoted by roman letters, e.g., P, while the italics Q are reserved for probabilities on the canonical space of counting processes or the canonical space of random counting measures. The symbol P arises exclusively in connection with the basic Markov kernels $P^{(n)}$ describing the conditional distribution of the $(n+1)$-th jump time given the previous jump times and marks.

If $(D_i, \mathcal{D}_i)$ for $i = 1, 2$ are measurable spaces and P_1 is a probability on $(D_1, \mathcal{D}_1)$ and $\psi : (D_1, \mathcal{D}_1) \to (D_2, \mathcal{D}_2)$ is measurable, we write $\psi(\mathrm{P}_1)$ for the probability on $(D_2, \mathcal{D}_2)$ obtained by transformation of P_1,

$$\psi(\mathrm{P}_1)(A_2) = \mathrm{P}_1\left(\psi^{-1}(A_2)\right) \quad (A_2 \in \mathcal{D}_2).$$

The phrases ‘almost surely’, ‘almost all’, ‘almost everywhere’, are abbreviated a.s., a.a., and a.e., respectively.

Expectations are written using the same font as the matching probability: $\mathbb{E}$ (for $\mathbb{P}$), E (for P), E (for Q). If it is necessary to emphasize which probability the expectation refers to, this is done using an index, $\mathbb{E}_{\mathbb{P}}$, E_{P}, E_Q.

If $(D, \mathcal{D})$ is a measurable space, ε_x for $x \in D$ denotes the probability measure degenerate at x,

$$\varepsilon_x(A) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Integrals are written in the standard form, e.g.,

$$\int f(x)\,\mathrm{P}(dx) \quad \text{or} \quad \int f\,d\mathrm{P}$$

but also quite often as

$$\int \mathrm{P}(dx)\, f(x);$$

in particular this is used in connection with multiple integrals. Often we write

$$\mathbb{E}[U; F] \quad \text{instead of} \quad \int_F U\,d\mathbb{P}.$$

$\mathbb{R}$-valued functions that are defined for example on $\mathbb{R}_0$ and are right-continuous and have limits from the left play an important role in the book. They are referred to by the French acronym *cadlag* (*c*ontinus *à d*roite avec les *l*imites *à g*auche) rather than the English *corlol* (*c*ontinuous *o*n the *r*ight with *l*imits *o*n the *l*eft).

Differentiation is denoted by the symbol D: $D_t f(t) = \frac{d}{dt} f(t)$.

Normally the random variables forming a stochastic process indexed by time t are denoted as X_t, while a given function (non-random) of time is written as $f(t)$.

If $a \in \mathbb{R}$ we write a^+, a^- for the positive and negative part of a: $a^+ = a \vee 0$, $a^- = -(a \wedge 0)$. The same notation is used for $\mathbb{R}$-valued processes and functions.

1.2 Conditional expectations and probabilities

Throughout the book conditional expectations and conditional probabilities are used in an absolutely essential manner. A short review of the basic definitions and most relevant results follow below.

Suppose X is an $\mathbb{R}$-valued random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{F}'$ be a sub σ-algebra of $\mathcal{F}$. If $\mathbb{E}|X| < \infty$, there exists a $\mathbb{P}$-integrable and $\mathcal{F}'$-measurable random variable $\mathbb{E}\left[X \,\middle|\mathcal{F}'\right]$ defined on Ω, the *conditional expectation of X given $\mathcal{F}'$*, such that

$$\int_{F'} \mathbb{E}\left[X \,\middle|\mathcal{F}'\right] d\mathbb{P} = \int_{F'} X\,d\mathbb{P}$$

for all $F' \in \mathcal{F}'$. The conditional expectation is uniquely determined $\mathbb{P}$-a.s.

If $\mathcal{F}'$ is the σ-algebra $\sigma\left(X'\right)$ generated by a random variable X', we write $\mathbb{E}\left[X \,\middle|\, X'\right]$ instead of $\mathbb{E}\left[X \,\middle|\, \mathcal{F}'\right]$. If X' is e.g., $\mathbb{R}$-valued, any $\mathbb{R}$-valued and $\sigma\left(X'\right)$-measurable random variable U has the form $g\left(X'\right)$ for some measurable $g : \mathbb{R} \to \mathbb{R}$. In particular $\mathbb{E}\left[X \,\middle|\, X'\right] = g\left(X'\right)$ and we then write

$$\mathbb{E}\left[X \,\middle|\, X' = x'\right] \quad \text{for} \quad g\left(x'\right).$$

If X is an indicator, $X = 1_F$ for some $F \in \mathcal{F}'$, $\mathbb{E}\left[X \,\middle|\, \mathcal{F}'\right]$ is the *conditional probability of F given $\mathcal{F}'$* and is denoted $\mathbb{P}\left(F \,\middle|\, \mathcal{F}'\right)$.

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 $\mathrm{p} : D_1 \times \mathcal{D}_2 \to [0, 1]$ such that

(i) $x_1 \mapsto \mathrm{p}(x_1, A_2)$ is $\mathcal{D}_1$-measurable for all $A_2 \in \mathcal{D}_2$,
(ii) $A_2 \mapsto \mathrm{p}\left(x_1, A_2\right)$ is a probability on $(D_2, \mathcal{D}_2)$ for all $x_1 \in D_1$.

Markov kernels in particular serve as *regular conditional distributions*: if X_1, X_2 are random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $(D_1, \mathcal{D}_1)$ and $(D_2, \mathcal{D}_2)$, respectively, then a regular conditional distribution of X_2, given $X_1 = x_1$ for all $x_1 \in D_1$, is a Markov kernel p from D_1 to D_2 such that

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2) = \int_{A_1} \mathrm{P}_{X_1}(dx_1)\, \mathrm{p}(x_1, A_2) \quad (A_1 \in \mathcal{D}_1, A_2 \in \mathcal{D}_2),$$

where $\mathrm{P}_{X_1} = X_1(\mathbb{P})$ is the distribution of X_1. Such regular conditional distributions need not exist (see below for conditions for existence), but if they do they may be used for finding conditional expectations: if $\psi : (D_2, \mathcal{D}_2) \to (\mathbb{R}, \mathcal{B})$ is measurable and $\mathbb{E}\,|\psi(X_2)| < \infty$, then

$$\mathbb{E}[\psi(X_2)\,|X_1 = x_1] = \int_{D_2} \psi(x_2)\, \mathrm{p}(x_1, dx_2)$$

for P_{X_1}-a.a. $x_1 \in D_1$ and

$$\mathbb{E}[\psi(X_2)\,|X_1] = \int_{D_2} \psi(x_2)\, \mathrm{p}(X_1, dx_2)$$

$\mathbb{P}$-a.s.

More generally, if X is a $(D, \mathcal{D})$-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{F}'$ is a sub σ-algebra of $\mathcal{F}$, a *regular conditional distribution of X given $\mathcal{F}'$* is a Markov kernel p' from $\left(\Omega, \mathcal{F}'\right)$ to $(D, \mathcal{D})$ such that

$$\mathbb{P}\left((X \in A) \cap F'\right) = \int_{F'} \mathbb{P}(d\omega)\, \mathrm{p}'(\omega, A)$$

for all $A \in \mathcal{D}$, $F' \in \mathcal{F}'$. Then, if $\psi : (D, \mathcal{D}) \to (\mathbb{R}, \mathcal{B})$ is measurable and $\mathbb{E}\,|\psi(X)| < \infty$,

$$\mathbb{E}\left[\psi\left(X\right)\middle|\mathcal{F}'\right](\omega)=\int_{D}\psi\left(x\right)\,\mathrm{p}'\left(\omega,dx\right)$$

for $\mathbb{P}$-a.a. ω. If, in particular, $X_1 and X_2$ are as above and p is a regular conditional distribution of X_2 given $X_1 = x_1$ for all x_1, then $\mathrm{p}'(\omega, A_2) = \mathrm{p}(X_1(\omega), A_2)$ defines a regular conditional distribution of X_2 given X_1 (given $\mathcal{F}' = \sigma(X_1)$).

A regular conditional distribution of X given $\mathcal{F}'$ always exists if the space $(D, \mathcal{D})$ is a *Borel space*, i.e., there is a bijection $\iota : D \to B_0$, a Borel subset of $\mathbb{R}$, such that ι and ι^{-1} are measurable when the σ-algebra $\mathcal{D}$ is used on D and the trace $\mathcal{B} \cap B_0$ of the Borel σ-algebra is used on B_0. In particular, $\mathcal{D}$ is then countably generated and all singletons $\{x\}$ for $x \in D$ are measurable. Recall that $\mathbb{R}^d$ (with the d-dimensional Borel σ-algebra $\mathcal{B}^d$) and the sequence space $\mathbb{R}^{\mathbb{N}}$ (equipped with the σ-algebra $\mathcal{B}^{\infty}$ generated by the projection maps $(x_n)_{n\in\mathbb{N}} \mapsto x_{n_0}$ for $n_0 \in \mathbb{N}$) are Borel spaces. Also a host of function spaces are Borel spaces, e.g., the space of functions $\widetilde{w} : \mathbb{R}_0 \to \mathbb{R}^d$ that are cadlag (right-continuous with left limits) equipped with the σ-algebra generated by the projection maps $(\widetilde{w}(t))_{t\in\mathbb{R}_0} \mapsto \widetilde{w}(t_0)$ for $t_0 \in \mathbb{R}_0$.

2

Simple and Marked Point Processes

This chapter contains the definitions of simple and marked point processes (SPPs and MPPs, respectively) recording the random occurrences over time of random events and shows how to identify the point processes with counting processes (CPs) and random counting measures (RCMs). The canonical spaces K, and K_E of sequences of timepoints of events and their marks are introduced together with the space W of counting process paths and the space $\mathcal{M}$ of discrete counting measures, counting timepoints and marks. It is demonstrated how SPPs and MPPs may be viewed as random variables with values in the sequence spaces K and K_E respectively, while CPs are W-valued and RCMs are $\mathcal{M}$-valued random variables. The definitions and notation given in the chapter are fundamental for everything that follows.

Reference. The definitions given here are standard, see e.g., Last and Brandt [81], Sections 1.2 and 1.3.

2.1 The definition of SPPs and MPPs

In general, point processes are probabilistic models for the random scatterings of points on some space (often a subset of $\mathbb{R}^d$). In this book we shall exclusively be concerned with point processes describing the occurrence over time of random events in which the occurrences are revealed one-by-one as time evolves. If only the occurrences of events are recorded, the space is always the time axis $\mathbb{R}_0 = [0, \infty[$, and we shall determine the random points in $\mathbb{R}_0$ by referring to the time of the 1st occurrence, 2nd occurrence, and so on. If, in addition, there are different types or marks of events, then the state space has the form $\mathbb{R}_0 \times E$ where E is the mark space. For the nth occurrence, together with the time at which it occurs, there is also the nth mark. Thus, for the processes we consider, time is an essential characteristic.

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

Definition 2.1.1 A *simple point process* (SPP for short) is a sequence $\mathcal{T} = (T_n)_{n \geq 1}$ of $\overline{\mathbb{R}}_0$-valued random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that

(i) $\mathbb{P}(0 < T_1 \le T_2 \le \cdots) = 1$,
(ii) $\mathbb{P}(T_n < T_{n+1}, T_n < \infty) = \mathbb{P}(T_n < \infty) \quad (n \ge 1)$,
(iii) $\mathbb{P}(\lim_{n\to\infty} T_n = \infty) = 1$.

Thus, an 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 of T_n is that, if finite, it is the timepoint at which the nth recording of an event takes place with 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 the time axis.

The condition (iii) in Definition 2.1.1 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 SPPs obtained by retaining only (i) and (ii) from Definition 2.1.1 is the class of simple point processes with *explosion*. It will be denoted SPP_{ex} and together with its MPP counterpart is discussed below and at various points in the text.

Introducing the sequence space

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

together with the σ-algebra $\mathcal{K}$ generated by the coordinate projections

$$T_n^\circ (t_1, t_2, \dots) = t_n \quad (n \ge 1),$$

we may view the SPP $\mathcal{T}$ as a $(\mathrm{K}, \mathcal{K})$-valued random variable, defined $\mathbb{P}$-a.s. The *distribution of* $\mathcal{T}$ is the probability $\mathcal{T}(\mathbb{P})$ on $(\mathrm{K}, \mathcal{K})$ obtained by transformation,

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

Similarly, introducing

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

with $\overline{\mathcal{K}}$ as the σ-algebra generated by the coordinate projections on $\overline{\mathrm{K}}$ (as those on K, also denoted T_n°), an SPP_{ex} may be viewed as a random variable $\overline{\mathcal{T}} = (T_n)$ with values in $\left(\overline{\mathrm{K}}, \overline{\mathcal{K}}\right)$, and the distribution of the SPP_{ex} is the transformed probability $\overline{\mathcal{T}}(\mathbb{P})$. Note that $\mathrm{K} = \left\{ (t_n) \in \overline{\mathrm{K}} : \lim_{n\to\infty} t_n = \infty \right\}$.

Now suppose we are also given a measurable space $(E, \mathcal{E})$ called the *mark space*. Adjoin to E the *irrelevant mark* ∇, to be used for describing the mark of an event that never occurs, write $\overline{E} = E \cup \{\nabla\}$ and let $\overline{\mathcal{E}} = \sigma(\mathcal{E}, \{\nabla\})$ denote the σ-algebra of subsets of $\overline{E}$ generated by the measurable subsets of E and the singleton $\{\nabla\}$.

Definition 2.1.2 A *marked point process* (MPP for short) with mark space E is a double sequence $(\mathcal{T}, \mathcal{Y}) = \left((T_n)_{n\ge 1}, (Y_n)_{n\ge 1}\right)$ of $\overline{\mathbb{R}}_+$-valued random variables T_n and $\overline{E}$-valued random variables Y_n defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathcal{T} = (T_n)$ is an SPP and

(i) $\mathbb{P}(Y_n \in E, T_n < \infty) = \mathbb{P}(T_n < \infty) \quad (n \ge 1)$,

(ii) $\mathbb{P}(Y_n = \nabla, T_n = \infty) = \mathbb{P}(T_n = \infty) \quad (n \geq 1)$.

Thus, as in Definition 2.1.1 we have a sequence of timepoints marking the occurrence of events, but now these events may be of different types, with the type (or name or label or mark) of the nth event denoted by Y_n. Note that the use of the irrelevant mark permits the definition of Y_n for all n, even if the nth event never occurs, i.e., when $T_n = \infty$.

An MPP $(\mathcal{T}, \mathcal{Y})$ may be viewed as a $(\mathrm{K}_E, \mathcal{K}_E)$-valued random variable, where

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

with $\mathcal{K}_E$ the σ-algebra of subsets of K_E generated by the coordinate projections

$$T_n^\circ((t_k), (y_k)) = t_n, \quad Y_n^\circ((t_k), (y_k)) = y_n.$$

The *distribution of* $(\mathcal{T}, \mathcal{Y})$ is then the transformed probability $(\mathcal{T}, \mathcal{Y})(\mathbb{P})$ on $(\mathrm{K}_E, \mathcal{K}_E)$.

MPPs with *explosion*, $\mathrm{MPP}_{\mathrm{ex}}$, are introduced in the obvious manner as $\left(\overline{\mathrm{K}}_E, \overline{\mathcal{K}}_E\right)$-valued random variables $\left(\overline{\mathcal{T}}, \overline{\mathcal{Y}}\right)$, where

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

with $\overline{\mathcal{K}}_E$ the σ-algebra generated by the projections on $\overline{\mathrm{K}}_E$. The distribution of an $\mathrm{MPP}_{\mathrm{ex}}$ is of course the probability $\left(\overline{\mathcal{T}}, \overline{\mathcal{Y}}\right)(\mathbb{P})$ on $\left(\overline{\mathrm{K}}_E, \overline{\mathcal{K}}_E\right)$.

It was only assumed above that $(E, \mathcal{E})$ be a measurable space, but from now on we shall impose more structure and assume that $(E, \mathcal{E})$ be sufficiently 'nice': in the remainder of this book we shall simply require the following:

Assumption *The mark space* $(E, \mathcal{E})$ *is a Borel space.*

Borel spaces were introduced on p. 7. There might be other choices for 'nice' spaces as well, but Borel spaces have at least three properties that will prove to be useful:

(i) Singletons are measurable, $\{y\} \in \mathcal{E}$ for all $y \in E$, so that in particular $\mathcal{E}$ *separates points,* i.e., for any $y \neq y' \in E$ there exists $A \in \mathcal{E}$ such that $y \in A$ and $y' \in A^c$.

(ii) The σ-algebra $\mathcal{E}$ is countably generated.

(iii) If Y is an $(E, \mathcal{E})$-valued random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{F}' \subset \mathcal{F}$ is a sub-σ-algebra, there always exists a regular conditional distribution of Y given $\mathcal{F}'$.

2.2 Counting processes and counting measures

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

$$N_t = \sum_{n=1}^{\infty} 1_{(T_n \le t)}. \tag{2.1}$$

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

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

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

If $t \mapsto f(t)$ is a *cadlag* function, Δf is the function $t \mapsto f(t) = f(t) - f(t-)$.

Note that the assumption that the T_n be strictly increasing as long as they are finite (Definition 2.1.1 (ii)) ensures that $t \mapsto N_t$ increases only in jumps of size 1.

Define on W the *canonical counting process* $N^\circ = \left(N_t^\circ\right)_{t \ge 0}$ by $N_t^\circ(w) = w(t)$, (so $N^\circ : W \to W$ is just the identity), and let

$$\mathcal{H} = \sigma\left(N_t^\circ\right)_{t \ge 0} \tag{2.2}$$

the smallest σ-algebra of subsets of W such that all N_t° are measurable. Then we may view N defined by (2.1) as a $(W, \mathcal{H})$-valued random variable with *distribution* $Q = N(\mathbb{P})$, the probability on $(W, \mathcal{H})$ obtained by transformation of $\mathbb{P}$.

The sequence $\mathcal{T}$ is easily recovered from N since $\mathbb{P}$-a.s. (consider $\omega \in \Omega$ such that $\mathcal{T}(\omega) \in \mathrm{K}$ only)

$$\begin{gathered} T_n = \inf\{t \ge 0 : N_t = n\}, \\ (T_n \le t) = (N_t \ge n), \end{gathered} \tag{2.3}$$

where here as elsewhere we define $\inf \emptyset = \infty$. Thus we have shown that *any SPP may be identified with its associated CP.* Later it will be the probability Q rather than $\mathcal{T}(\mathbb{P})$ that will serve as the basic description of the distribution of an SPP.

The discussion above could also be carried out for $\overline{\mathcal{T}}$ as an $\mathrm{SPP}_{\mathrm{ex}}$, only now $N_t = \infty$ can occur with probability > 0, and W may be replaced by the space $\overline{W}$ of exploding counting process paths, $\mathrm{CP}_{\mathrm{ex}}$, $\overline{w} \in \overline{\mathbb{N}}_0^{\mathbb{R}_0}$, which otherwise satisfy the conditions already imposed on $w \in W$. On $\overline{W}$ we use the σ-algebra $\overline{\mathcal{H}}$ defined analogously to $\mathcal{H}$ in (2.2).

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

$$\mu = \sum_{n \in \mathbb{N}: T_n < \infty} \varepsilon_{(T_n, Y_n)}. \tag{2.4}$$

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, i.e.,

$$\varepsilon_{(T_n,Y_n)}(C,\omega) = 1_C(T_n(\omega), Y_n(\omega)) \quad (C \in \mathcal{B}_0 \otimes \mathcal{E}).$$

Thus, for $\mathbb{P}$-a.a. ω, $\mu(\omega)$ is a *discrete counting measure* on $(\mathbb{R}_0 \times E, \mathcal{B}_0 \otimes \mathcal{E})$, and satisfies in particular that $\mu(\omega)$ be a positive σ-finite measure m such that

$$\begin{aligned}
&m(C) \in \overline{\mathbb{N}}_0 \quad (C \in \mathcal{B}_0 \otimes \mathcal{E}),\\
&m(\{0\} \times E) = 0,\\
&m(\{t\} \times E) \le 1 \quad (t \ge 0),\\
&m([0,t] \times E) < \infty \quad (t \ge 0).
\end{aligned} \tag{2.5}$$

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 an $\overline{\mathbb{N}}_0$-valued random variable.

We denote by $\mathcal{M}$ *the space of discrete counting measures on* $(\mathbb{R}_0 \times E, \mathcal{B}_0 \otimes \mathcal{E})$, i.e., measues m of the form

$$m = \sum_{n:t_n<\infty} \varepsilon_{(t_n,y_n)} \tag{2.6}$$

for some sequence $((t_n), (y_n)) \in \mathrm{K}_E$. 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 $m \mapsto m(C)$ from $\mathcal{M}$ to $\overline{\mathbb{N}}_0$.

On $\mathcal{M}$ we use the σ-algebra

$$\mathcal{H} = \sigma\left(\mu^\circ(C)\right)_{C \in \mathcal{B}_0 \otimes \mathcal{E}}, \tag{2.7}$$

the smallest σ-algebra such that all $\mu^\circ(C)$ are measurable. Thus, with μ the RCM from (2.4) above, μ becomes an a.s. defined $(\mathcal{M}, \mathcal{H})$-valued random variable. Its *distribution* is the probability $Q = \mu(\mathbb{P})$ on $(\mathcal{M}, \mathcal{H})$.

Remark 2.2.1 With the mark space $(E, \mathcal{E})$ a Borel space, any measure m satisfying the four conditions (2.5) will be a discrete counting measure: defining $t_n = \inf\{t : m([0,t] \times E) = n\}$ it is not difficult to verify that $(t_n) \in \mathrm{K}$, so the problem is to extract the sequence of marks (y_n) from m and then to verify that m is given by (2.6). For the recipe for obtaining (y_n) from m, see (2.9) and (3.7) below.

Instead of using an RCM, it is also possible to describe $(\mathcal{T}, \mathcal{Y})$ through a family of counting processes; for $A \in \mathcal{E}$ define $N(A) = (N_t(A))_{t \ge 0}$ by

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

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

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

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

On the space $\mathcal{M}$, the counting processes corresponding to $N(A)$ and $\overline{N}$ are denoted $N^\circ(A)$ and $\overline{N}^\circ$ and defined by

$$N_t^\circ(A) = \mu^\circ\left([0,t] \times A\right), \quad \overline{N}_t^\circ = \mu^\circ\left([0,t] \times E\right).$$

Exercise 2.2.1 *Show that* $\mathcal{H} = \sigma\left(N_t^\circ(A)\right)_{t\geq 0, A\in E}$.

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, $\mathbb{P}$-almost surely,

$$\begin{array}{c} T_n = \inf\left\{t \geq 0 : \overline{N}_t = n\right\}, \\ (T_n \leq t) = \left(\overline{N}_t \geq n\right); \end{array} \tag{2.8}$$

cf. (2.3). It is trickier to find the Y_n; however it holds for any $A \in \mathcal{E}$ that $\mathbb{P}$-a.s.

$$(Y_n \in A) = \bigcup_{K'=1}^{\infty} \bigcap_{K=K'}^{\infty} \bigcup_{k=1}^{\infty} \left(\overline{N}_{(k-1)/2^K} = n-1, N_{k/2^K}(A) - N_{(k-1)/2^K}(A) = 1\right), \tag{2.9}$$

showing that if μ is a random variable, then so is Y_n.

The identity (2.9) is important, so we shall give a detailed argument based on a device to be used also in the sequel, which consists in subdividing the time axis $\mathbb{R}_+$ into intervals of length 2^{-K} for an arbitrary $K \in \mathbb{N}$,

$$\mathbb{R}_+ = \bigcup_{k=1}^{\infty} \left]\frac{k-1}{2^K}, \frac{k}{2^K}\right].$$

In order to prove (2.9) we shall simply show that any $\omega \in \Omega$ such that $(\mathcal{T}(\omega), \mathcal{Y}(\omega)) \in \mathrm{K}_E$ (which by Definition 2.1.2 amounts to $\mathbb{P}$-a.a. ω) belongs to the set $(Y_n \in A)$ on the left of (2.9) iff it belongs to the set on the right.

With ω as above, suppose first that $Y_n(\omega) \in A$ (with $A \in \mathcal{E}$), so that in particular $T_n(\omega) < \infty$, and determine for every $K \in \mathbb{N}$ the unique $k = k_K(\omega) \in \mathbb{N}$ such that $(k-1)/2^K < T_n(\omega) \leq k/2^K$. Because the sequence $(T_{n'}(\omega))_{n'\geq 1}$ belongs to K, we have that $T_{n-1}(\omega) < T_n(\omega) < T_{n+1}(\omega)$, and therefore also for K sufficiently large that $T_{n-1}(\omega) \leq (k-1)/2^K < T_n(\omega) \leq k/2^K < T_{n+1}(\omega)$ which is precisely to say that ω belongs to the set on the right of (2.9). If conversely ω belongs to that set, by assumption it holds for K sufficiently large that there exists $k = k_K(\omega)$ such that the interval $I_{k,K} = \left](k-1)/2^K, k/2^K\right]$ contains at least one of the jump times $T_{n'}(\omega)$

with $n' \geq n$ and precisely one with $Y_{n'}(\omega) \in A$, while also $T_n(\omega) > (k-1)/2^K$. In particular $I_{k,K}$ is that interval in the Kth partitioning of $\mathbb{R}_+$ that contains $T_n(\omega)$; hence k is uniquely determined and as argued above, $T_n(\omega)$, for K sufficiently large, will be the only $T_{n'}(\omega)$ inside $I_{k,K}$; it follows that necessarily $Y_n(\omega) \in A$.

Note that even though (2.9) always holds, without assuming that $(E, \mathcal{E})$ is a Borel space (as we are assuming throughout, cf. p. 11) it may not be possible to compute $Y_n(\omega)$ from $\mu(\omega)$: here it is at least required that $\mathcal{E}$ separate points as described in (i), p. 11). But when $(E, \mathcal{E})$ is a Borel space, we have now 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 than with 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 MPP_{ex}s with suitable exploding RCMs where $\overline{N}_t = \infty$ is possible. Such RCMs are random variables with values in a space of discrete counting measures m, which are not required to be σ-finite, i.e., it is allowed that $m([0,t] \times E) = \infty$ for $t \in \mathbb{R}_+$. We denote the space of these RCM_{ex}'s by $\left(\overline{\mathcal{M}}, \overline{\mathcal{H}}\right)$ with the σ-algebra $\overline{\mathcal{H}}$ defined in the obvious way, in complete analogy with the definition (2.7) of $\mathcal{H}$ above.

3

Construction of SPPs and MPPs

This chapter contains the fundamental construction of canonical point processes (i.e., probabilities on the sequence spaces K and K_E), canonical counting processes (probabilities on the space W) and canonical random counting measures (probabilities on the space $\mathcal{M}$). The construction is performed using successive regular conditional distributions. The chapter also has a section on how to view certain types of continuous time stochastic processes as MPPs, an approach examined in detail in Chapters 6 and 7. Finally, a number of basic examples that will reappear at various points in the text are presented.

Reference. The construction of canonical point processes given here is the same as that used in Jacobsen [53].

3.1 Creating SPPs

The probabilistic properties of an SPP $\mathcal{T} = (T_n)$ are captured by the distribution of $\mathcal{T}$. As we saw in Chapter 2, this distribution is the probability $\mathcal{T}(\mathbb{P})$ on the sequence space $(\mathrm{K}, \mathcal{K})$. In general we shall call a probability measure R on $(\mathrm{K}, \mathcal{K})$ a *canonical SPP;* and these are the 'processes' (i.e., probability measures) we shall now proceed to construct. Through the bimeasurable bijection (i.e., a map which is one-to-one and onto with the map and its inverse both measurable) $\varphi : (\mathrm{K}, \mathcal{K}) \to (W, \mathcal{H})$ given by (cf. (2.1), (2.3))

$$N_t^\circ\left(\varphi\left((t_n)\right)\right) = \sum_{n=1}^{\infty} 1_{[0,t]}(t_n) \quad ((t_n) \in \mathrm{K}, t \geq 0),$$

$$T_n^\circ\left(\varphi^{-1}(w)\right) = \inf\{t \geq 0 : w(t) = n\} \quad (w \in W), \tag{3.1}$$

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

The idea underlying the construction of the distribution of an 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 $Z_n := (T_1, \ldots, T_n)$. More precisely, let

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

be the space of n-sequences that can appear as the first n-coordinates of an element in K, equipped with the σ-algebra $\mathcal{K}^{(n)}$ spanned by the coordinate projections T_k° for $1 \leq k \leq n$, (equivalently, the trace on $\mathrm{K}^{(n)}$ of the n-dimensional Borel σ-algebra).

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

Notation. Write z_n for a typical point $z_n = (t_1, \ldots, t_n) \in \mathrm{K}^{(n)}$ and Z_n° for the mapping from $\left(\overline{\mathrm{K}}, \overline{\mathcal{K}}\right)$ to $\left(\mathrm{K}^{(n)}, \mathcal{K}^{(n)}\right)$ given by $Z_n^\circ((t_k)) = (t_1, \ldots, t_n)$, equivalently $Z_n^\circ = \left(T_1^\circ, \ldots, T_n^\circ\right)$. For an SPP $\mathcal{T} = (T_n)$ defined on an arbitrary probability space, we write $Z_n = (T_1, \ldots, T_n)$.

Theorem 3.1.1 (a) *For every choice of the probability $P^{(0)}$ and the Markov kernels $P^{(n)}$ for $n \geq 1$ satisfying*

$$\begin{aligned} P_{z_n}^{(n)}(]t_n, \infty]) &= 1 \qquad \left(z_n \in \mathrm{K}^{(n)}, t_n < \infty\right) \\ P_{z_n}^{(n)}(\{\infty\}) &= 1 \qquad \left(z_n \in \mathrm{K}^{(n)}, t_n = \infty\right), \end{aligned} \tag{3.2}$$

there is a unique probability $\overline{\mathrm{R}}$ on the sequence space $\left(\overline{\mathrm{K}}, \overline{\mathcal{K}}\right)$ allowing explosions, such that $T_1^\circ(\overline{\mathrm{R}}) = P^{(0)}$ and for every $n \geq 1$, the probability $P_{z_n}^{(n)}(\cdot)$ is a regular conditional distribution of T_{n+1}° given $Z_n^\circ = z_n$ for all z_n.

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

$$\overline{\mathrm{R}}\left(\lim_{n\to\infty} T_n^\circ = \infty\right) = 1. \tag{3.3}$$

We shall not give the proof here. The theorem follows easily from the Kolmogorov consistency theorem by showing consistency of the finite-dimensional distributions (distributions of Z_n° for all $n \in \mathbb{N}$), which have the following structure: for $B^{(n)} \in \mathcal{K}^{(n)}$,

$$\overline{\mathrm{R}}\left(Z_n^\circ \in B^{(n)}\right) = \int_{\overline{\mathbb{R}}_+} P^{(0)}(dt_1) \int_{\overline{\mathbb{R}}_+} P_{z_1}^{(1)}(dt_2) \cdots \int_{\overline{\mathbb{R}}_+} P_{z_{n-1}}^{(n-1)}(dt_n)\ 1_{B^{(n)}}(z_n). \tag{3.4}$$

Remark 3.1.1 It should be clear that (3.2) ensures that the sequence $\left(T_n^\circ\right)$ is increasing a.s., strictly increasing as long as T_n° is finite. And it is also clear that (3.3) is exactly the condition ensuring that no explosion occurs. In the sequel the Markov kernels $P_{z_n}^{(n)}$ will only be specified for $z_n = (t_1, \ldots, t_n) \in \mathrm{K}^{(n)}$ with $t_n < \infty$; due to the second condition in (3.2) this is of course sufficient.

Note that several choices of Markov kernels may lead to the same $\overline{\mathrm{R}}$ or 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 $\overline{\mathrm{R}}\left(Z_n^\circ \in B^{(n)}\right) = 0$.

While arbitrary choices of $P^{(n)}$ for $n \geq 0$ lead to possibly exploding SPPs, there is no simple characterization of the $P^{(n)}$ that results in genuine SPPs. Indeed, it may be extremely difficult to decide whether or not a canonical $\mathrm{SPP}_{\mathrm{ex}}$ is a true SPP. This *stability problem* will be discussed on various occasions later in this book, see p. 61 for example.

If $\mathcal{T} = (T_n)$ is an SPP defined on an arbitrary probability space, then the regular conditional distribution of T_{n+1} given $Z_n = (T_1, \ldots, T_n) = z_n$ for all z_n always exist (cf. p. 7), and the corresponding Markov kernels for n arbitrary satisfy (3.2).

Notation. For $z_n \in \mathrm{K}^{(n)}$ with $t_n < \infty$, we shall often describe the $P_{z_n}^{(n)}$ through their *survivor functions* $\overline{P}_{z_n}^{(n)}$,

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

By (3.2), $\overline{P}_{z_n}^{(n)}(t) = 1$ for $t \leq t_n$ if $t_n < \infty$, so it suffices to give $\overline{P}_{z_n}^{(n)}(t)$ for $t \geq t_n$, imposing the value 1 for $t = t_n < \infty$.

Example 3.1.1 The *dead process* is the canonical SPP with no jumps in finite time: $\mathrm{R}\left(T_1^\circ = \infty\right) = 1$. It is completely specified by the requirement that $P^{(0)} = \varepsilon_\infty$, while the choice of $P^{(n)}$ for $n \geq 1$ (subject to (3.2)) is irrelevant; cf. Remark 3.1.1.

Example 3.1.2 The canonical *homogeneous Poisson process* (SPP version) is the probability R on $(\mathrm{K}, \mathcal{K})$ that makes the waiting times $V_n^\circ := T_n^\circ - T_{n-1}^\circ$ for $n \geq 1$ (with $T_0^\circ \equiv 0$) independent and identically distributed (iid), exponential with some rate $\lambda > 0$. Thus

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

$$\overline{P}_{z_n}^{(n)}(t) = e^{-\lambda(t - t_n)} \quad (n \geq 1, t \geq t_n).$$

The corresponding probability $Q = \varphi(\mathrm{R})$ on $(W, \mathcal{H})$ makes N° into a homogeneous Poisson process (in the traditional sense) with parameter λ, meaning that N° has stationary, independent increments that follow Poisson distributions, i.e., for $s < t$, $N_t^\circ - N_s^\circ$ is independent of $\left(N_u^\circ\right)_{u \leq s}$ and

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

with the Poisson distribution depending on s and t through $t-s$ only. This familiar fact, which is not so easy to prove from the properties of R, will appear as a consequence of Example 4.7.1 and Proposition 4.7.2 below. At this stage we shall be satisfied with just determining the distribution of N_t° directly: for $t > 0$ trivially

$$Q\left(N_t^\circ = 0\right) = \mathrm{R}\left(T_1^\circ > t\right) = e^{-\lambda t},$$

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

$$\begin{aligned} Q\left(N_t^\circ = n\right) &= \mathrm{R}\left(T_n^\circ \le t < T_{n+1}^\circ\right) \\ &= \mathrm{E}_\mathrm{R}\left[\mathrm{R}\left(T_{n+1}^\circ > t \,\middle|\, Z_n^\circ\right); T_n^\circ \le t\right] \\ &= \mathrm{E}_\mathrm{R}\left[\overline{P}_{Z_n^\circ}^{(n)}(t)\,; T_n^\circ \le t\right] \\ &= \mathrm{E}_\mathrm{R}\left[e^{-\lambda(t-T_n^\circ)}; T_n^\circ \le t\right]. \end{aligned}$$

But under R, T_n° follows a Γ-distribution with density

$$\frac{\lambda^n}{(n-1)!} s^{n-1} e^{-\lambda s}$$

for $s > 0$, and therefore

$$Q\left(N_t^\circ = n\right) = \int_0^t e^{-\lambda(t-s)} \frac{\lambda^n}{(n-1)!} s^{n-1} e^{-\lambda s}\, ds = \frac{(\lambda t)^n}{n!} e^{-\lambda t},$$

as desired.

The dead process may be viewed as the Poisson process with parameter $\lambda = 0$. That Poisson processes are stable is argued in Example 3.1.3.

Exercise 3.1.1 Try to give a direct proof that if R is Poisson (λ), then under $Q = \varphi(\mathrm{R})$, N° has stationary and independent Poisson distributed increments as asserted in Example 3.1.2.

Example 3.1.3 A canonical *renewal process* is a canonical SPP such that the waiting times V_n° are iid and finite. If $\overline{G}$ is the survivor function for the waiting time distribution, then

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

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

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

The renewal processes defined here are 0-*delayed*. It is customary in the renewal process literature to consider the first event (renewal) of a 0-delayed renewal process to always occur at time 0, but in the SPP theory that occurrence is of course ignored.

Example 3.1.4 Suppose the waiting times V_n° are independent, V_n° exponential at rate $\lambda_{n-1} \ge 0$ for $n \ge 1$. Thus

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

$$\overline{P}^{(n)}_{z_n}(t) = e^{-\lambda_n(t-t_n)} \quad (n \geq 1, t \geq t_n).$$

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

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

The canonical counting process corresponding to an 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 3.3.1 below).

Exercise 3.1.2 In Example 3.1.4, show that explosion occurs with probability 1 if $\sum \lambda_n^{-1} < \infty$ (easy), and that the process is stable if $\sum \lambda_n^{-1} = \infty$ (compute and analyze $\mathrm{E_R} \exp\left(- \lim_{n\to\infty} T^{\circ}_n\right)$).

3.2 Creating MPPs

In complete analogy with the construction of SPPs, the construction of MPPs $(\mathcal{T}, \mathcal{Y})$ with a mark space $(E, \mathcal{E})$ that is a Borel space space (cf. p. 11), consists in the construction of *canonical MPPs*, i.e., probabilities R on the sequence space $(\mathrm{K}_E, \mathcal{K}_E)$; cf. Section 2.1. Furthermore, since $(E, \mathcal{E})$ is a Borel space there is a bimeasurable bijection $\varphi : (\mathrm{K}_E, \mathcal{K}_E) \to (\mathcal{M}, \mathcal{H})$, the space of discrete counting measures defined on p. 13, given by

$$\varphi\left((t_n), (y_n)\right) = \sum_{n:t_n<\infty} \varepsilon_{(t_n, y_n)} \tag{3.5}$$

with the inverse φ^{-1} determined by (cf. (2.8) and (2.9)),

$$T^{\circ}_n \circ \varphi^{-1} = \inf\left\{t : \overline{N}^{\circ}_t = n\right\} \quad (n \in \mathbb{N}), \tag{3.6}$$

$$\begin{aligned}&\left(Y^{\circ}_n \circ \varphi^{-1} \in A\right) \\ &= \bigcup_{K'=1}^{\infty} \bigcap_{K=K'}^{\infty} \bigcup_{k=1}^{\infty} \left(\overline{N}^{\circ}_{(k-1)/2^K} = n-1, N^{\circ}_{k/2^K}(A) - N^{\circ}_{(k-1)/2^K}(A) = 1\right)\end{aligned} \tag{3.7}$$

for $n \in \mathbb{N}$, $A \in \mathcal{E}$, (where as in Section 2.1, $N^{\circ}_t(A) = \mu^{\circ}([0, t] \times A)$, $\overline{N}^{\circ}_t = N^{\circ}_t(E)$). Using φ one then obtains a construction of a *canonical random counting measure* by transformation, viz. the probability $\varphi(\mathrm{R})$ on $(\mathcal{M}, \mathcal{H})$.

Note. There are some subtleties involved in using (3.7) to actually define the Y°_n: not only is it (obviously) needed that $\mathcal{E}$ separates points, but it is certainly also helpful that $\mathcal{E}$ is countably generated, and the assumption that $(E, \mathcal{E})$ is a Borel space therefore comes in handy. Suppose for simplicity that $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$. Taking $A = \mathbb{R}$ in (3.7) one verifies that the set given by the expression on the bottom line equals $\left(T_n^\circ \circ \varphi^{-1} < \infty\right)$ as it should. Now consider a nested sequence of partitions $\left(\mathfrak{I}_{K_0}\right)_{K_0 \geq 1}$ of $\mathbb{R}$ into intervals of length 2^{-K_0}, and such that each interval from $\mathfrak{I}_{K_0+1}$ is contained in one from $\mathfrak{I}_{K_0}$. Consider an arbitrary $m \in \mathcal{M}$ such that $\left(T_n^\circ \circ \varphi^{-1}\right)(m) < \infty$. The properties of RCMs listed in (2.5) imply that for each K_0 there is precisely one interval $I \in \mathfrak{I}_{K_0}$ such that with $A = I$ the set in the second line of (3.7) contains m. With I_{K_0} as this interval, it also holds that $I_{K_0+1} \subset I_{K_0}$ and therefore $\bigcap_{K_0} I_{K_0} = \{y\}$ for some $y \in \mathbb{R}$ and of course $\left(Y_n^\circ \circ \varphi^{-1}\right)(m) = y$.

The idea behind the construction of an 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

$$\mathrm{K}_E^{(n)} = \Big\{ (t_1, \ldots, t_n; y_1, \ldots, y_n) \in \overline{\mathbb{R}}_+^n \times \overline{E}^n : \begin{array}{l} (t_1, \ldots, t_n) \in \mathrm{K}^{(n)} \\ \text{and } y_k \in E \text{ iff } t_k < \infty \Big\} \end{array}$$

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

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

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

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

Notation. Write z_n for a typical point $z_n = (t_1, \ldots, t_n; y_1, \ldots, y_n) \in \mathrm{K}_E^{(n)}$ and (z_n, t) or z_n, t for a typical point $(z_n, t) = (t_1, \ldots, t_n, t; y_1, \ldots, y_n) \in \mathrm{J}_E^{(n)}$. Also we write Z_n° for the mapping from $\overline{\mathrm{K}}_E$ to $\mathrm{K}_E^{(n)}$ given by $Z_n^\circ\left((t_k), (y_k)\right) = (t_1, \ldots, t_n; y_1, \ldots, y_n)$, equivalently $Z_n^\circ = (T_1^\circ, \ldots, T_n^\circ; Y_1^\circ, \ldots, Y_n^\circ)$. For an MPP $(\mathcal{T}, \mathcal{Y}) = ((T_n), (Y_n))$ defined on an arbitrary probability space, we write $Z_n = (T_1, \ldots, T_n; Y_1, \ldots, Y_n)$.

Theorem 3.2.1 (a) *Given a probability $P^{(0)}$ and Markov kernels $P^{(n)}$ for $n \geq 1$ and $\pi^{(n)}$ for $n \geq 0$ satisfying*

$$\begin{aligned} &P_{z_n}^{(n)}\left(]t_n,\infty]\right) = 1 && \textit{if } t_n < \infty \\ &P_{z_n}^{(n)}\left(\{\infty\}\right) = 1 && \textit{if } t_n = \infty, \\ &\pi_{z_n,t}^{(n)}\left(E\right) = 1 && \textit{if } t < \infty, \\ &\pi_{z_n,t}^{(n)}\left(\{\nabla\}\right) = 1 && \textit{if } t = \infty, \end{aligned} \tag{3.8}$$

there is a unique probability $\overline{\mathrm{R}}$ *on the sequence space* $\left(\overline{\mathrm{K}}_E, \overline{\mathcal{K}}_E\right)$ *allowing explosions, such that* $T_1^\circ(\overline{\mathrm{R}}) = P^{(0)}$ *and for every* $n \geq 1$, $z_n \in \mathrm{K}_E^{(n)}$, *the probability* $P_{z_n}^{(n)}(\cdot)$ *is a regular conditional distribution of* T_{n+1}° *given* $Z_n^\circ = z_n$ *for all* z_n, *and for every* $n \geq 0$, $(z_n, t) \in \mathrm{J}_E^{(n)}$, *the probability* $\pi_{z_n,t}^{(n)}(\cdot)$ *is a regular conditional distribution of* Y_{n+1}° *given* $\left(Z_n^\circ, T_{n+1}^\circ\right) = (z_n, t)$ *for all* (z_n, t).

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

$$\overline{\mathrm{R}}\left(\lim_{n\to\infty} T_n^\circ = \infty\right) = 1. \tag{3.9}$$

In the sequel, when describing the Markov kernels, it obviously suffices to consider $P_{z_n}^{(n)}$ and $\pi_{z_n}^{(n)}$ for $z_n \in \mathrm{K}_E^{(n)}$ with $t_n < \infty$ only.

Similar remarks apply to this result as those given after Theorem 3.1.1. The proof can be based on the Kolmogorov consistency theorem if, as we are assuming, the mark space $(E, \mathcal{E})$ is a Borel space, but the theorem holds in fact with $(E, \mathcal{E})$ as an arbitrary measurable space and is then a consequence of the result known as the Ionescu–Tulcea theorem. The finite-dimensional distributions have the following appearance (cf. (3.4)): for $n \in \mathbb{N}$, $C^{(n)} \in \mathcal{K}_E^{(n)}$,

$$\begin{aligned} &\overline{\mathrm{R}}\left(Z_n^\circ \in C^{(n)}\right) \\ &= \int_{\overline{\mathbb{R}}_+ \times \overline{E}} P^{(0)}\left(dt_1\right) \pi_{t_1}^{(0)}\left(dy_1\right) \cdots \int_{\overline{\mathbb{R}}_+ \times \overline{E}} P_{z_{n-1}}^{(n-1)}\left(dt_n\right) \pi_{z_{n-1},t_n}^{(n-1)}\left(dy_n\right) 1_{C^{(n)}}\left(z_n\right). \end{aligned} \tag{3.10}$$

It should be noted that if $(\mathcal{T}, \mathcal{Y})$ is an MPP defined on some probability space and the mark space $(E, \mathcal{E})$ is a Borel space (cf. p.7), the regular conditional distributions of Y_{n+1} given $Z_n = (T_1, \ldots, T_n; Y_1, \ldots Y_n) = z_n$ and $T_{n+1} = t$ for all (z_n, t) exist. The resulting Markov kernels together with those describing the regular conditional distributions of T_{n+1} given $Z_n = z_n$ for all z_n (that always do exist) will satisfy (3.8).

Notation. As in the SPP case we shall write $\overline{P}_{z_n}^{(n)}(t) = P_{z_n}^{(n)}\left(]t, \infty]\right)$; cf. p. 19.

Example 3.2.1 Let $(E, \mathcal{E})$ be an arbitrary mark space, let $\lambda > 0$ and let κ be a probability measure on $(E, \mathcal{E})$. Write $\rho = \lambda\kappa$, a bounded positive measure on E. The canonical *homogeneous Poisson process with mark space* E *and intensity measure* ρ is the MPP determined by the Markov kernels

$$\overline{P}^{(0)}(t) = e^{-\lambda t} \quad (t \geq 0), \qquad \overline{P}^{(n)}_{z_n}(t) = e^{-\lambda(t-t_n)} \quad (n \geq 1, t \geq t_n),$$

$$\pi^{(n)}_{z_n,t} = \kappa \quad (n \geq 0, t > t_n).$$

It follows that the waiting times $V_n^\circ = T_n^\circ - T_{n-1}^\circ$ are iid exponential with rate λ; in particular the canonical process R is stable and the SPP $\left(T_n^\circ\right)$ of jump times is a Poisson process (λ); cf. Example 3.1.2. Furthermore, it is also seen that under R the Y_n° are iid with distribution κ and that the sequence $\left(Y_n^\circ\right)$ is stochastically independent of the sequence $\left(T_n^\circ\right)$.

Let $Q = \varphi(\mathrm{R})$ be the corresponding canonical RCM, the *homogeneous Poisson random measure with intensity measure* ρ. Then the following is true: under Q, for each $A \in \mathcal{E}$, the counting process $N^\circ(A)$ is a homogeneous Poisson process $(\rho(A))$, and for $A_1, \ldots, A_r \in \mathcal{E}$ mutually disjoint, the counting processes $N^\circ(A_1), \ldots, N^\circ(A_r)$ are stochastically independent. These (non-trivial) properties of Q will be established in Proposition 4.7.2. But similar to what was done in Example 3.1.2 we can at least find the distribution of $N_t^\circ(A)$ for $t > 0$ and $A \in \mathcal{E}$ directly: from the earlier example we know that $\overline{N}_t^\circ$ follows a Poisson distribution with parameter λt, and using this and the independence under R between the sequences $\left(T_n^\circ\right)$ and $\left(Y_n^\circ\right)$, it follows that for $n \in \mathbb{N}_0$,

$$\begin{aligned}
Q\left(N_t^\circ(A) = n\right) &= \sum_{n'=n}^{\infty} Q\left(N_t^\circ(A) = n, \overline{N}_t^\circ = n'\right) \\
&= \sum_{n'=n}^{\infty} \mathrm{R}\left(\sum_{k=1}^{n'} 1_{(Y_k^\circ \in A)} = n,\, T_{n'}^\circ \leq t < T_{n'+1}^\circ\right) \\
&= \sum_{n'=n}^{\infty} \mathrm{R}\left(\sum_{k=1}^{n'} 1_{(Y_k^\circ \in A)} = n\right) \mathrm{R}\left(T_{n'}^\circ \leq t < T_{n'+1}^\circ\right) \\
&= \sum_{n'=n}^{\infty} \binom{n'}{n} (\kappa(A))^n (1-\kappa(A))^{n'-n} \frac{(\lambda t)^{n'}}{n'!} e^{-\lambda t} \\
&= \frac{1}{n!} (t\lambda\kappa(A))^n e^{-\lambda t} \sum_{n'=n}^{\infty} \frac{1}{(n'-n)!} (\lambda t (1-\kappa(A)))^{n'-n} \\
&= \frac{1}{n!} (t\lambda\kappa(A))^n e^{-t\lambda\kappa(A)},
\end{aligned}$$

as desired.

Exercise 3.2.1 If you have solved Exercise 3.1.1, now try to establish the properties of the homogeneous Poisson measure asserted in Example 3.2.1! (This is not supposed to be easy).

Example 3.2.2 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 $(G, \mathcal{G})$-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 $(G, \mathcal{G})$, given by $T_n \equiv n$ and $Y_n = X_n$ and generated by the probability $P^{(0)}_{|x_0} = \varepsilon_1$ and the Markov kernels $P^{(n)}_{z_n|x_0} = \varepsilon_{n+1}$ and

$$\pi^{(n)}_{z_n,t|x_0}(A) = \mathbb{P}\left(X_{n+1} \in A \,|X_0 = x_0, (X_1, \ldots, X_n) = (y_1, \ldots, y_n)\right) \tag{3.11}$$

where only $z_n = (t_1, \ldots, t_n; y_1, \ldots, y_n)$ and t with $t_k = k$ and $t = n+1$ are relevant.

If in particular (X_n) is a Markov chain in discrete time, (3.11) simplifies to

$$\pi^{(n)}_{z_n,t|x_0}(A) = p_n(y_n, A)$$

(with $y_0 = x_0$ if $n = 0$), where p_n is the transition probability of the chain from time n to time $n + 1$. The chain is homogeneous if all the transition probabilities p_n for $n \geq 0$ are the same.

Note that the identification of (X_n) with the MPP $((T_n), (Y_n))$ is possible only after assuming the value of X_0 to be known.

3.3 From MPPs to PDPs

One of the main purposes of this book is to construct *piecewise deterministic processes* (PDPs) from MPPs and to use MPP theory to discuss the properties of the PDPs. 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 σ-algebra. Assume further that X is right-continuous and piecewise continuous with only finitely many discontinuities (jumps) on any finite time interval. Then an MPP is easily constructed from X by letting T_n be the time of the nth jump, and defining $Y_n = X_{T_n}$ (if $T_n < \infty$) to be the state reached by X at the time of the nth 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 measurable G-valued function $f^{(n)}_{z_n|x_0}(t)$ of z_n (with $t_n < \infty$) and $t \geq t_n$ and the initial state x_0 such that

$$X_t = f^{(\overline{N}_t)}_{Z_{\langle t\rangle}|x_0}(t) \tag{3.12}$$

where

$$Z_{\langle t\rangle} := \left(T_1, \ldots, T_{\overline{N}_t}; Y_1, \ldots, Y_{\overline{N}_t}\right).$$

Thus, until the time of the first jump (where the MPP contains no information about X), $X_t = f^{(0)}_{|x_0}(t)$ is a given, 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

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

for all z_n with $t_n < \infty$, in particular

$$f^{(0)}_{|x_0}(0) = x_0.$$

We shall more formally call a process X of the form (3.12) *piecewise deterministic*. It is a *step process* (or *piecewise constant* process) if $f^{(n)}_{z_n}(t) = y_n$ and, if $G = \mathbb{R}$, 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.

Notation. In the sequel, whenever $\overline{N}_t$ appears as an upper or lower index, we shall write $\langle t \rangle$ as short for $\overline{N}_t$ to simplify the notation. Always, $\langle t \rangle$ will refer to the total number of jumps on $[0, t]$ whether we are dealing with an MPP or RCM or just an SPP or a CP, and whether the process in question is defined on an abstract probability space or on one of the canonical spaces. Thus, on $(W, \mathcal{H})$, $\langle t \rangle$ will be short for N°_t and on $(\mathcal{M}, \mathcal{H})$, $\langle t \rangle$ will be the notation used for $\overline{N}^\circ_t$.

With this notational convention, (3.12) becomes when also preferring the notation $\langle t \rangle$ for $(\langle t \rangle)$ in the upper index,

$$X_t = f^{\langle t \rangle}_{Z_{\langle t \rangle}|x_0}(t) \tag{3.13}$$

where

$$Z_{\langle t \rangle} := \left(T_1, \ldots, T_{\langle t \rangle}; Y_1, \ldots, Y_{\langle t \rangle}\right).$$

In analogy with $\langle t \rangle$, we shall also use the notation $\langle t- \rangle$ for e.g., the left limit $\overline{N}_{t-} = \lim_{s \uparrow t, s < t} \overline{N}_s$.

The identification of X with an MPP need not refer exclusively to jump times, but could be based on other selected time points and marks. Between these selected time points, the computation rule (3.12) should still apply. We then call X *piecewise continuous* if all $t \mapsto f^{(n)}_{z_n}(t)$ are continuous on $[t_n, \infty[$. For the general piecewise deterministic processes, and even for all step processes, it is not required that G be a topological space.

Above we used as marks $Y_n = X_{T_n}$, the states X arives at at the selected time points. Other choices are possible, e.g., if $G = \mathbb{R}$ (or just some topological vector space) and X is cadlag (right-continuous with left limits) it is sometimes relevant to use $Y_n = \Delta X_{T_n} = X_{T_n} - X_{T_n-}$, the size of the jump X makes at time T_n.

Example 3.3.1 We shall outline the MPP description of *time-homogeneous Markov chains* in continuous time. That what follows is a construction of such chains is standard (although a rigorous proof is not always given); see Example 4.3.6 below for a detailed argument.

Let G be the at most countably infinite state space with $\mathcal{G}$ the σ-algebra of all subsets. G will be the mark space, but in agreement with standard Markov chain notation we write i, i_n for elements in G rather than y, y_n. Next, let $\lambda_i \geq 0$ for $i \in G$ and let $0 \leq \pi_{ij} \leq 1$ for $i \neq j \in G$ be parameters such that $\sum_{j:j\neq i} \pi_{ij} = 1$ for all $i;$ consider also an MPP$_{ex}$ $(\mathcal{T}, \mathcal{Y}) = ((T_n), (Y_n))$ defined on some measurable space $(\Omega, \mathcal{F})$ equipped with a Markov family of probabilities $\left(\mathbb{P}_{|i_0}\right)_{i_0 \in G}$ (e.g., use $(\Omega, \mathcal{F}) = \left(\overline{\mathrm{K}}_G, \overline{\mathcal{K}}_G\right)$ and define $\mathbb{P}_{|i_0}$ using the Markov kernels (3.14) and (3.15)) such that the $\mathbb{P}_{|i_0}$-distribution of $(\mathcal{T}, \mathcal{Y})$ for an arbitrary $i_0 \in G$ is generated by the Markov kernels

$$\overline{P}^{(0)}_{|i_0}(t) = e^{-\lambda_{i_0} t}, \qquad \overline{P}^{(n)}_{z_n|i_0}(t) = e^{-\lambda_{i_n}(t-t_n)} \quad (t \geq t_n), \tag{3.14}$$

$$\pi^{(n)}_{z_n,t|i_0}(A) = \sum_{j \in A \setminus i_n} \pi_{i_n j}, \tag{3.15}$$

where of course $z_n = (t_1, \ldots, t_n; i_1, \ldots, i_n)$. Assuming that the process is stable (which may be non-trivial to verify in concrete cases), define for a given i_0 the step process

$$X_t = X_{t|i_0} = Y_{\langle t \rangle}$$

with initial state $X_0 \equiv Y_0 \equiv i_0$ corresponding to taking $f^{(n)}_{z_n|i_0}(t) = i_n$. Thus the waiting times between jumps are exponential given the past determined by the last state reached by the chain; this last state also governs the destination of the jumps. It now holds that there are *transition probabilities* $p_{ij}(t) := \mathbb{P}_{|i}\left(X_{t|i} = j\right)$ not depending on i_0 such that

$$p_{ij}(t) = \mathbb{P}_{|i_0}(X_{s+t} = j \,|X_s = i) = \mathbb{P}_{|i_0}\left(X_{s+t} = j \,\middle|(X_u)_{0 \leq u \leq s}, X_s = i\right) \tag{3.16}$$

for all $i, j \in G$ and all $s, t \geq 0$. Here $\sum_j p_{ij}(t) = 1$ for all i and t (because the MPP is assumed to be stable). Furthermore the transition probabilities satisfy the *Chapman–Kolmogorov equations,*

$$p_{ij}(s+t) = \sum_{k \in G} p_{ik}(s) p_{kj}(t)$$

for all $i, j \in G$, $s, t \geq 0$ with $p_{ij}(0) = \delta_{ij}$, and the *transition intensities*

$$q_{ij} = \lim_{t \to 0} \tfrac{1}{t}\left(p_{ij}(t) - \delta_{ij}\right)$$

exist for all i, j, and are given by $q_{ii} = -\lambda_i$, while for $i \neq j$, $q_{ij} = \lambda_i \pi_{ij}$, in particular $q_{ii} \leq 0$ and $\sum_j q_{ij} = 0$ for all i.

Note that if $\lambda_i = 0$, the state i is *absorbing,* as is seen from (3.14): once the chain enters state i it remains there forever and consequently the π_{ij} for $j \neq i$ are irrelevant for the description of the chain.

If the MPP is not stable, defining $X_{t|i_0} = \nabla$ for $t \geq \lim_{n\to\infty} T_n$, with ∇ an additional state not in G, the conclusions above about the properties of X are still valid. The resulting Markov chain is the so-called *minimal jump chain* determined by the transition intensities q_{ij}, with substochastic transition probabilities on G, i.e., $\sum_{j\in G} p_{ij}(t) < 1$ is possible.

For a proof of the properties of X given above and a stronger form of the Markov property (3.16), see Example 4.3.6 below and also, for Markov chains on general state spaces; see Section 7.2. At this stage, just note that one way of showing (3.16) is to show that

$$\mathbb{P}_{|i_0}\left(X_{t_1} = i_1, \ldots, X_{t_n} = i_n\right) = \mathbb{P}_{|i_0}\left(X_{t_1} = i_1, \ldots, X_{t_{n-1}} = i_{n-1}\right) p_{i_{n-1}i_n}(t_n - t_{n-1})$$

for all $n \geq 1$ and all $0 = t_0 < t_1 < \cdots < t_n$, $i_0, i_1, \ldots, i_n \in G$.

Example 3.3.2 Let G be a finite set, and $\mathcal{G}$ the σ-algebra of all subsets. A step process $X = \left(X_{t|i_0}\right)$is a *semi-Markov process* or *Markov renewal process* in continuous time with initial state $i_0 \in G$, provided it is determined from an MPP $(\mathcal{T}, \mathcal{Y}) = ((T_n), (Y_n))$ by $X_{t|i_0} = Y_{\langle t\rangle}$ (just as in Example 3.3.1 and still using $Y_0 \equiv i_0$), but where now the sequence $(V_n, Y_n)_{n\geq 1}$, where $V_n = T_n - T_{n-1}$, is a discrete time-homogeneous Markov chain with transition probabilities of the form

$$\begin{aligned} &\mathbb{P}_{|i_0}(Y_{n+1} = j, V_{n+1} \in B | Y_n = i, V_n) \\ &\quad = \mathbb{P}_{|i_0}(Y_{n+1} = j, V_{n+1} \in B | (V_k, Y_k)_{1\leq k\leq n-1}, Y_n = i, V_n) \qquad (3.17) \\ &\quad = \mathrm{P}_{ij}(B) \end{aligned}$$

for $n \geq 0$, $i, j \in G$ and $B \in \mathcal{B}_+$. Here each P_{ij} is a non-negative finite measure (of total mass ≤ 1) on $\mathbb{R}_+$ such that all $\mathrm{P}_{ii} \equiv 0$ and $\sum_{j\neq i} \mathrm{P}_{ij}$ is a probability measure for all $i \in G$, i.e., $\sum_{j\neq i} \pi_{ij} = 1$ where $\pi_{ij} = \mathrm{P}_{ij}(\mathbb{R}_+)$.

From (3.17) one finds that

$$\mathbb{P}_{|i_0}(Y_{n+1} = j | Y_n = i, V_n) = \pi_{ij},$$

and since this does not depend on V_n it follows that $(Y_n)_{n\geq 0}$ is a homogeneous Markov chain with initial state i_0 and transition probabilities $\left(\pi_{ij}\right)$. It is then also clear from (3.17) that the Markov kernels generating the jump times T_n for $(\mathcal{T}, \mathcal{Y})$ are determined by

$$\overline{P}^{(0)}_{|i_0}(t) = \sum_{j\neq i_0} \overline{\mathrm{P}}_{i_0 j}(t), \qquad \overline{P}^{(n)}_{z_n|i_0}(t) = \sum_{j\neq i_n} \overline{\mathrm{P}}_{i_n j}(t - t_n) \quad (n \geq 1, t \geq t_n),$$

with $z_n = (t_1, \ldots, t_n; i_1, \ldots, i_n)$, writing $\overline{\mathrm{P}}_{ij}(u) = \mathrm{P}_{ij}(]u, \infty[)$. To find the Markov kernels generating the marks Y_n is more complicated, but using (3.17) to find the conditional distribution of Y_{n+1} given (Z_n, T_{n+1}) one eventually arrives at

$$\pi^{(n)}_{z_n,t|i_0}(\{j\}) = \frac{d\mathrm{P}_{i_n j}}{d\sum_{j'\neq i_n}\mathrm{P}_{i_n j'}}(t-t_n) \qquad (t > t_n,\, j \neq i_n)\,.$$

Here the Radon–Nikodym derivatives are determined by

$$\int_{]u,\infty[} \frac{d\mathrm{P}_{ij}}{d\sum_{j'\neq i}\mathrm{P}_{ij'}}(v) \sum_{j'\neq i}\mathrm{P}_{ij'}(dv) = \overline{\mathrm{P}}_{ij}(u)$$

for any $u \geq 0,\ i \neq j$. If in particular there are probability measures P_i on $\mathbb{R}_+$ and constants $c_{ij'}$ such that $\mathrm{P}_{ij'} = c_{ij'}\mathrm{P}_i$ for $j' \neq i$, then necessarily $c_{ij'} = \pi_{ij'}$ and it is immediately verified that

$$\pi^{(n)}_{z_n,t|i_0}(\{j\}) = \pi_{i_n j}.$$

If $\overline{\mathrm{P}}_{ij}(u) = \pi_{ij}e^{-\lambda_i u}$ for all $i \neq j$ for some $\lambda_i \geq 0$, then X becomes a homogeneous Markov chain as in Example 3.3.1. If all $\mathrm{P}_{ij} = \pi_{ij}\mathrm{P}$ for $i \neq j$ with P a probability on $\mathbb{R}_+$, then the SPP $(T_n)_{n\geq 1}$ is a 0-delayed renewal process, as in Example 3.1.3, with iid waiting times V_n that have distribution P.

The final example we shall present, like the two preceding examples, gives an MPP construction of a certain type of stochastic process in continuous time but it is more advanced in two respects: the obtained process is not a PDP, but it is more complicated, and the mark space required is much larger than the ones encountered so far. The example is included mainly to illustrate that it is relevant to allow marks that are functions rather than just points in $\mathbb{R}^d$ or some countable set; but since this is outside the main scope of this book, the example is not reused or referred to later.

Example 3.3.3 We shall construct a simple example of a so-called *branching diffusion* where particles or individuals are born and die and during their lifetime, moving randomly through space (in this case $\mathbb{R}$) following the sample path of a Brownian motion. To this end, let G be the space of all continuous functions $w : \mathbb{R}_0 \to \mathbb{R}$ and define $\mathcal{G}$ as the smallest σ-algebra of subsets of G that makes all the projection maps $w \overset{\pi_t^\circ}{\mapsto} w(t)$ measurable. Furthermore, let $\mathrm{P}_{\mathrm{B}|x}$ be the probability on $(G,\mathcal{G})$ such that under $\mathrm{P}_{\mathrm{B}|x}$ the process $\left(\pi_t^\circ\right)_{t\geq 0}$ is a standard *Brownian motion* on $\mathbb{R}$, starting from x: $\mathrm{P}_{\mathrm{B}|x}(X_0 = x) = 1$; and for all $n \in \mathbb{N}$, $0 = t_0 < t_1 < \cdots < t_n$, the increments $\left(\pi^\circ_{t_k} - \pi^\circ_{t_{k-1}}\right)_{1\leq k\leq n}$ are independent with $\pi_t^\circ - \pi_s^\circ$ for $s < t$ Gaussian with mean 0 and variance $t - s$.

For $r \geq 1$, let $\{G\}^r$ denote the set of all sets $\{w_1, \ldots, w_r\}$ consisting of r elements of G. As a mark space we shall use

$$E = \{0\} \cup \bigcup_{r=1}^{\infty}\{G\}^r$$

with the understanding that if the mark 0 occurs, the branching diffusion has become extinct, while if the mark $\{w_1, \ldots, w_r\} \in \{G\}^r$ results, then the branching diffusion has reached a population size r and the locations in $\mathbb{R}$ of the r individuals at any (!) future point in time are specified by the collection $\{w_1, \ldots, w_r\}$ of Brownian paths (for technical reasons to be used though only until the time of the next jump for the MPP).

In order to describe the MPP we finally need a birth rate $\beta > 0$ at which new individuals are born, and a death rate $\delta > 0$ at which individuals die.

Write $y = (r, \{w_1, \ldots, w_r\})$ for the elements of E, (just $y = 0$ if $r = 0$) and let $y_0 = \left(r_0, \left\{w_1, \ldots, w_{r_0}\right\}\right)$ be a given but arbitrarily chosen initial state with $r_0 \geq 1$. Then define the Markov kernels $P^{(n)}_{z_n|y_0}$ generating the jump times (T_n) by

$$\overline{P}^{(0)}_{|y_0}(t) = e^{-r_0(\beta+\delta)t}, \qquad \overline{P}^{(n)}_{z_n|y_0}(t) = e^{-r_n(\beta+\delta)(t-t_n)} \quad (n \geq 1, t \geq t_n),$$

writing $z_n = (t_1, \ldots, t_n; y_1, \ldots, y_n)$ with $y_n = \left(r_n, \left\{w_1, \ldots, w_{r_n}\right\}\right)$.

Note that if $r_n = 0$, $P^{(n)}_{z_n|y_0} = \varepsilon_\infty$, then the population has become extinct and the $(n+1)$th jump never occurs. So in order to define the kernel $\pi^{(n)}_{z_n,t|y_0}$ we need only consider the case $r_n \geq 1$ and in that case we describe $\pi^{(n)}_{z_n,t|y_0}$ as follows: just before the time t of the $(n+1)$th jump the r_n individuals in the population are located at the positions $w_1(t - t_n), \ldots, w_{r_n}(t - t_n)$. Now, with probability

$$\frac{\beta}{r_n(\beta+\delta)}$$

choose any one of these positions, $w_j(t - t_n)$, and then generate Y_{n+1} as the collection of $r_n + 1$ Brownian paths chosen independently according to the distributions $\mathrm{P}_{\mathrm{B}|w_i(t-t_n)}$ for $1 \leq i \leq r_n$ *and* an extra $\mathrm{P}_{\mathrm{B}|w_j(t-t_n)}$.

Similarly, with probability

$$\frac{\delta}{r_n(\beta+\delta)}$$

choose any of the positions $w_j(t - t_n)$ and let Y_{n+1} consist of the collection of $r_n - 1$ Brownian paths chosen independently according to the distributions $\mathrm{P}_{\mathrm{B}|w_i(t-t_n)}$ for $1 \leq i \leq r_n$ with $i \neq j$.

Thus, at each finite jump time T_n, a new individual is born with probability $\beta/(\beta+\delta)$ and starts moving from the same location as its parent, with all individuals following independent Brownian motions, by construction made continuous across the jump times T_n. Similarly, with probability $\delta/(\beta+\delta)$, one of the existing individuals dies and is removed from the population, the remaining individuals still moving as independent Brownian motions.

Having constructed the MPP $((T_n), (Y_n))$ it is a simple matter to define the branching diffusion $\rho = (\rho_t)_{t\geq 0}$ itself: ρ_t is the random variable

$$\rho_t = Y_{\langle t\rangle, t}$$

where

$$Y_{\langle t\rangle,t}(\omega) = \left\{w_1(t - T_n(\omega)), \ldots, w_{r_n}(t - T_n(\omega))\right\}$$

when $\overline{N}_t(\omega) = n$ and $Y_n(\omega) = \left(r_n, \left\{w_1, \ldots, w_{r_n}\right\}\right)$. Here the definition of $Y_{\langle t\rangle,t}(\omega)$ as a collection of locations should allow for repetitions: it will happen once in a while that two of the different Brownian motions meet, and also, if at time $T_n(\omega)$ a birth occurs, two of the locations describing $Y_{\langle t\rangle,T_n(\omega)}(\omega)$ are certainly the same.

In the construction above new Brownian motions are generated at each jump time T_n for all individuals surviving immediately after T_n. It is not necessary to do this. One might just as well have continued using the Brownian paths already in place for the individuals in the population immediately before T_n and then generate just one new Brownian motion for the newborn individual if a birth takes place at T_n. By the Markov property of Brownian motion this alternative branching diffusion will have the same distribution as the one we have constructed — even though the resulting MPP is certainly different from the one defined above. The construction with new Brownian motions for all individuals at each jump time was chosen because it appears simpler to describe.

We have defined ρ as a process with a state space which is the set of all finite subsets of $\mathbb{R}$, when allowing for repetitions. An equivalent definition allows one to view a branching diffusion as a process $\widetilde{\rho} = (\widetilde{\rho}_t)_{t\geq 0}$ with values in the space of discrete counting measures on $\mathbb{R}$, viz. by letting $\widetilde{\rho}_t(B)$ be the number of elements of $Y_{\langle t\rangle,t}$ (when counting repetitions of course) that belong to the Borel set B.

A final comment on the branching diffusion is that the MPP involved is in fact stable. It is easily seen from the description of the Markov kernels when referring back to Example 3.3.1 that the process $(\widetilde{\rho}_t(\mathbb{R}))_{t\geq 0}$ is a homogeneous Markov chain on $\mathbb{N}_0$ with transition intensities q_{ij} given by $q_{ii} = -i(\beta + d)$, while for $i \neq j$ the only non-zero intensities are $q_{i,i+1} = i\beta$ and $q_{i,i-1} = i\delta$ for $i \geq 1$. Intuitively this implies that $\widetilde{\rho}(\mathbb{R})$ jumps more slowly than the birth process from Example 3.1.4 with $\lambda_i = i(\beta + \delta)$, and since by Exercise 3.1.2 that birth process is stable, so is $\widetilde{\rho}(\mathbb{R})$. A formal proof may be given using Corollary 4.4.4.

4

Compensators and Martingales

This chapter contains the basic theory of probability measures on the space W of counting process paths and the space $\mathcal{M}$ of discrete counting measures. Compensators and compensating measures are defined using hazard measures, and by exploring the structure of adapted and predictable processes on W and $\mathcal{M}$ (a structure that is discussed in detail), it is shown how the various forms of compensators characterize probabilities on canonical spaces. Also, these probabilities are described by the structure of basic martingales. Stochastic integrals (all elementary) are discussed and the martingale representation theorem is established. It is shown (Itô's formula) how processes adapted to the filtration generated by an RCM may be decomposed into a predictable process and a local martingale. Finally, there is a discussion of compensators and compensating measures for counting processes and random counting measures defined on arbitrary filtered probability spaces.

Much of the material presented in this chapter is essential for what follows and complete proofs are given for the main results. Some of these proofs are quite long and technical and rely on techniques familiar only to readers well acquainted with measure theory and integration. It should be reasonably safe to omit reading the proofs of e.g., the following results: Theorem 4.1.1 (the proof of the last assertion), Proposition 4.2.1, Theorem 4.3.2, Proposition 4.3.5, Proposition 4.5.1 (although it is useful to understand why (4.65) implies that f is constant) and Theorem 4.6.1.

References. Hazard measures (Section 4.1) are treated in Jacod [60] and Appendix A5 in Last and Brandt [81]. Other than these, everything in this chapter apart from Section 4.8 refers to canonical SPPs, CPs, MPPs and RCMs. The corresponding theory for processes on general filtered spaces was initiated by Jacod [60]; a nice survey is given in Chapter 2 of Last and Brandt [81].

4.1 Hazard measures

Let P be a probability on $\left(\overline{\mathbb{R}}_+, \overline{\mathcal{B}}_+\right)$ with survivor function $\overline{\mathrm{P}}$. Thus $\overline{\mathrm{P}}(t) = \mathrm{P}\left(]t, \infty]\right)$ with $\overline{\mathrm{P}}(0) = 1$ and $\overline{\mathrm{P}}(t-) = \mathrm{P}\left([t, \infty]\right)$. Also write $\Delta \mathrm{P}(t) := \mathrm{P}\left(\{t\}\right)$ with $\Delta \mathrm{P}(\infty) := \mathrm{P}\left(\{\infty\}\right)$.

Define

$$t^\dagger := \inf\left\{t > 0 : \overline{\mathrm{P}}(t) = 0\right\}, \tag{4.1}$$

the *termination point* for P. Note that $t^\dagger > 0$ because $\overline{\mathrm{P}}(0) = 1$ and $\overline{\mathrm{P}}$ is right-continuous. Further, $t^\dagger = \infty$ iff $\overline{\mathrm{P}}(t) > 0$ for all $t \in \mathbb{R}_+$ and if $t^\dagger < \infty$, then $\overline{\mathrm{P}}(t^\dagger) = 0$ always; in addition either $\Delta\mathrm{P}\left(t^\dagger\right) > 0$ and $\mathrm{P}\left(t^\dagger -\right) > 0$, or $\Delta\mathrm{P}\left(t^\dagger\right) = 0$ and $\overline{\mathrm{P}}(t^\dagger - \epsilon) > 0$ for all $0 < \epsilon < t^\dagger$.

Definition 4.1.1 The *hazard measure* for P is the $\overline{\mathbb{R}}_0$-valued measure ν on $(\mathbb{R}_+, \mathcal{B}_+)$ with $\nu \ll \mathrm{P}$ and, for $t \in \mathbb{R}_+$,

$$\frac{d\nu}{d\mathrm{P}}(t) = \begin{cases} 1/\overline{\mathrm{P}}(t-) & \text{if } \overline{\mathrm{P}}(t-) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

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

$$\nu(B) = \int_B \frac{1}{\overline{\mathrm{P}}(s-)}\,\mathrm{P}(ds) \tag{4.3}$$

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

$$\mathrm{P}\left(\left\{t \in \mathbb{R}_+ : \overline{\mathrm{P}}(t-) = 0\right\}\right) = 0, \tag{4.4}$$

i.e., the definition of $d\nu/d\mathrm{P}(t)$ when $\overline{\mathrm{P}}(t-) = 0$ is irrelevant. (For us $\nu\left(\left]t^\dagger, \infty\right[\right) = 0$ because of (4.2) — but that need of course not hold with other definitions of ν beyond $t^\dagger$). To argue that (4.4) holds, just note that the set appearing in (4.4) is $= \emptyset$ if $t^\dagger = \infty$, $= \left[t^\dagger, \infty\right[$ if $t^\dagger < \infty$ and $\Delta\mathrm{P}\left(t^\dagger\right) = 0$, $= \left]t^\dagger, \infty\right[$ if $t^\dagger < \infty$ and $\Delta\mathrm{P}\left(t^\dagger\right) > 0$).

It will be clear from Theorem 4.1.1 below that hazard measures are not in general σ-finite; it is possible that $\nu([s,t]) = \infty$ for some $0 < s < t < \infty$.

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

$$\nu\left([t, t + dt[\right) = \mathbb{P}\left(t \le U < t + dt \,|U \ge t\right),$$

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 yet occurred.

There are two particularly important structures for P, where the hazard measure ν has a nice form, and where it is easy to recover P from ν.

Suppose first that P has a density f with respect to Lebesgue measure. Then ν has density u where $u(t) = f(t)/\overline{\mathrm{P}}(t)$ if $t < t^\dagger$, and the function u is called the *hazard function* of P. If in addition f, and therefore also u, is continuous on $\left]0, t^\dagger\right[$, u is found by differentiation,

$$u(t) = -D_t \log \overline{\mathrm{P}}(t)$$

for $t < t^\dagger$, and since $\overline{\mathrm{P}}(0) = 1$ it also follows that

$$\overline{\mathrm{P}}(t) = \exp\left(-\int_0^t u(s)\,ds\right). \tag{4.5}$$

For the second structure, suppose that there is a finite or infinite sequence (t_k) in $\mathbb{R}_+$ with $0 < t_1 < t_2 < \cdots$ such that P is concentrated on the t_k and, possibly, ∞ with all t_k atoms for P: $\mathrm{P}(\mathbb{R}_+ \setminus \{(t_k)\}) = 0$ and $\Delta\mathrm{P}(t_k) > 0$ for all k. Then ν is concentrated on $\{(t_k)\}$. Writing $\Delta\nu(t) := \nu(\{t\})$, we have that

$$\Delta\nu\,(t_k) = \frac{\Delta\mathrm{P}(t_k)}{\overline{\mathrm{P}}(t_k-)} = 1 - \frac{\overline{\mathrm{P}}(t_k)}{\overline{\mathrm{P}}(t_{k-1})}$$

(with $t_0 = 0$ if $k = 1$) so that

$$\overline{\mathrm{P}}(t_k) = \prod_{j=1}^{k}\left(1 - \Delta\nu(t_j)\right). \tag{4.6}$$

The basic properties of hazard measures are summarized in the next result.

Theorem 4.1.1 *Let ν be the hazard measure for some probability* P *on* $\left(\overline{\mathbb{R}}_+, \overline{\mathcal{B}}_+\right)$. *Then*

(i) $\nu\,(]0, t]) < \infty$ *if $t < t^\dagger$;*
(ii) $\Delta\nu(t) \le 1$ *for all $t \in \mathbb{R}_+$,* $\Delta\nu(t) < 1$ *for all $t < t^\dagger$;*
(iii) *if $t^\dagger < \infty$, then $\Delta\nu\left(t^\dagger\right) = 1$ iff $\Delta\mathrm{P}\left(t^\dagger\right) > 0$;*
(iv) *if $\Delta\mathrm{P}\left(t^\dagger\right) > 0$, then $\nu\left(]0, t^\dagger]\right) < \infty$ if $t^\dagger < \infty$ and $\nu\,(\mathbb{R}_+) < \infty$ if $t^\dagger = \infty$;*
(v) *if $\Delta\mathrm{P}\left(t^\dagger\right) = 0$, then $\nu\left(]0, t^\dagger[\right) = \infty$ whether or not $t^\dagger$ is finite.*

If, conversely, ν is a $\overline{\mathbb{R}}_0$-valued measure on $(\mathbb{R}_+, \mathcal{B}_+)$, and locally finite at 0 *so that $\nu\,(]0, t]) < \infty$ for $t > 0$ sufficiently small, and with $\Delta\nu\,(t) \le 1$ for all $t \in \mathbb{R}_+$, then ν is the hazard measure for a uniquely determined probability* P *on* $\left(\overline{\mathbb{R}}_+, \overline{\mathcal{B}}_+\right)$ *in the following sense: the termination point $t^\dagger$ for* P *is*

$$t^\dagger = \inf\{t > 0 : \nu\,(]0, t]) = \infty \text{ or } \Delta\nu\,(t) = 1\}, \tag{4.7}$$

the survivor function $\overline{\mathrm{P}}$ *for* P *is given by the* product integral

$$\overline{\mathrm{P}}(t) = \begin{cases} \prod_{0<s\le t}(1 - \nu\,(ds)) & \text{if } t < t^\dagger, \\ 0 & \text{if } t \ge t^\dagger \end{cases} \tag{4.8}$$

and the hazard measure for P *agrees with ν on* $\left\{t \in \mathbb{R}_+ : \overline{\mathrm{P}}(t-) > 0\right\}$.

Note. The product integral (4.8) reduces to (4.5) and (4.6) in the special cases where either $\nu\,(dt) = u(t)\,dt$ or where ν is concentrated on $\{(t_k)\}$; cf. the two special structures for P used to arrive at (4.5) and (4.6). To define the product integral for general

ν, define $t^\dagger$ as in (4.7) and restrict ν to $\left]0, t^\dagger\right]$ if $t^\dagger < \infty$, $\nu\left(\left]0, t^\dagger\right[\right) < \infty$ and (necessarily) $\Delta\nu\left(t^\dagger\right) = 1$, and restrict ν to $\left]0, t^\dagger\right[$ otherwise. The restriction is denoted by ν as well.

Now split the restricted ν into its discrete part ν^d and its continuous part ν^c, i.e., $\nu = \nu^c + \nu^d$, where

$$\nu^d(B) = \nu\left(B \cap A_\nu\right) \quad (B \in \mathcal{B}_+)$$

with $A_\nu = \{t \in \mathbb{R}_+ : \Delta\nu(t) > 0\}$, at most the countably infinite set of atoms for ν. (The definition of $t^\dagger$ implies that $\nu\left(]0,t]\right) < \infty$ for any $t < t^\dagger$, hence on each $]0,t]$ with $t < t^\dagger$, ν can have at most countably many atoms). Then define for $t < t^\dagger$

$$\prod_{0<s\le t}\left(1 - \nu(ds)\right) = \exp\left(-\nu^c\left(]0,t]\right)\right)\prod_{0<s\le t}\left(1 - \Delta\nu^d(s)\right). \tag{4.9}$$

(The continuous product extends of course only over s such that $s \le t$ with $s \in A_\nu$).

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

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

where the assumption $\Delta\mathrm{P}(t^\dagger) = 0$ 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 theorem, let $t^\dagger$ be given by (4.7), restrict ν to $\left]0, t^\dagger\right]$ or $\left]0, t^\dagger\right[$ as described in the note above, and let ν now denote this restriction.

Consider the function

$$\overline{\mathrm{P}}(t) = \exp\left(-\nu^c\left(]0,t]\right)\right)\prod_{0<s\le t}\left(1 - \Delta\nu^d(s)\right) \tag{4.10}$$

defined by the right-hand side of (4.9). The basic structure of the product integral that we shall explore is that (obviously) for $0 \le s < t < t^\dagger$,

$$\overline{\mathrm{P}}(t) = \overline{\mathrm{P}}(s)\exp\left(-\nu^c\left(]s,t]\right)\right)\prod_{s<u\le t}\left(1 - \Delta\nu^d(u)\right). \tag{4.11}$$

By the elementary inequality $1 - x \le e^{-x}$ for $x \ge 0$ it follows that for $0 \le s < t < t^\dagger$,

$$\prod_{s<u\le t}\left(1-\Delta\nu^d(u)\right)\le \exp\left(-\nu^d\,(]s,t])\right) \tag{4.12}$$

implying that

$$\overline{\mathrm{P}}(t)\le \overline{\mathrm{P}}(s)\exp\left(-\nu\,(]s,t])\right). \tag{4.13}$$

Also, for fixed $t_0 < t^\dagger$, since by the definition of $t^\dagger$, $\Delta\nu^d(t) < 1$ for $t \le t_0$, and since on $]0, t_0]$ there can be at most finitely many t with $\Delta\nu(t) > \epsilon$ for any given $\epsilon > 0$, it follows that

$$\sup_{0<t\le t_0}\Delta\nu^d(t)\le c_0<1;$$

determining $\gamma_0 > 1$ to be so large that $1 - x \ge e^{-\gamma_0 x}$ for $0 \le x \le c_0$, it follows that

$$\prod_{s<u\le t}\left(1-\Delta\nu^d(u)\right)\ge \exp\left(-\gamma_0\nu^d\,(]s,t])\right) \tag{4.14}$$

and so

$$\begin{aligned}\overline{\mathrm{P}}(t)&\ge \overline{\mathrm{P}}(s)\exp\left(-\nu^c\,(]s,t])-\gamma_0\nu^d\,(]s,t])\right)\\&\ge \overline{\mathrm{P}}(s)\exp\left(-\gamma_0\nu\,(]s,t])\right).\end{aligned} \tag{4.15}$$

Letting $t \downarrow s$ with $t > s$ in (4.13) and (4.15) it follows that $\overline{\mathrm{P}}$ is right-continuous on $]0, t_0[$ for any $t_0 < t^\dagger$, and so, trivially, $\overline{\mathrm{P}}$ is right-continuous on $\mathbb{R}_+$. Since $\overline{\mathrm{P}}(0) = 1$ and $\overline{\mathrm{P}}$ is decreasing, $\overline{\mathrm{P}}$ is the survivor function for a (uniquely determined) probability P on $\overline{\mathbb{R}}_+$.

We now proceed to show that P has termination point $t^\dagger(\mathrm{P}) = t^\dagger$, and that P has hazard measure ν. Since by the definition (4.10) of $\overline{\mathrm{P}}$, $\overline{\mathrm{P}}(t^\dagger) = 0$ if $t^\dagger < \infty$, we see that $t^\dagger(\mathrm{P}) \le t^\dagger$. On the other hand, for $t_0 < t^\dagger$, (4.15) gives

$$\overline{\mathrm{P}}(t_0)\ge \exp\left(-\gamma_0\nu\,(]0,t_0])\right)>0$$

so that $t^\dagger(\mathrm{P}) \ge t_0$, forcing $t^\dagger(\mathrm{P}) \ge t^\dagger$, and we conclude that $t^\dagger(\mathrm{P}) = t^\dagger$.

To show that P has hazard measure ν, we use the differentiation technique from Appendix A, and start by showing that on $\left]0, t^\dagger\right[$, $\nu(t) := \nu\,(]0, t])$ is pointwise differentiable with derivative $1/\overline{\mathrm{P}}(t-)$ with respect to the distribution function $F_{\mathrm{P}} = 1 - \overline{\mathrm{P}}$ for P. This means that

(i) whenever $F_{\mathrm{P}}(s) = F_{\mathrm{P}}(t)$ for some $s < t < t^\dagger$, then also $\nu(s) = \nu(t)$;
(ii) whenever $\Delta\nu(t) > 0$ for some $t < t^\dagger$, then also $\Delta\mathrm{P}(t) > 0$;
(iii) $\lim_{K\to\infty} F_K(t) = 1/\overline{\mathrm{P}}(t-)$ for P-a.a $t \in \left]0, t^\dagger\right[$,

where, in (iii), writing $I_{k,K} = \left]\frac{k-1}{2^K}, \frac{k}{2^K}\right]$,

$$F_K(t)=\sum_{k=1}^{\infty}1_{I_{k,K}}(t)\frac{\nu\left(I_{k,K}\right)}{\mathrm{P}\left(I_{k,K}\right)} \tag{4.16}$$

with $0/0 = 0$ if the denominator (and by (i), also the numerator) in the kth term vanishes.

(i) follows directly from (4.11). Letting $s \uparrow t$ with $s < t$ in (4.11) gives $\overline{\mathrm{P}}(t) = \overline{\mathrm{P}}(t-)(1 - \Delta\nu(t))$ so that

$$\Delta \mathrm{P}(t) = \overline{\mathrm{P}}(t-)\Delta\nu(t) \tag{4.17}$$

and (ii) follows. Finally, if $t < t^\dagger$, letting $]a_K, b_K]$ denote the interval $I_{k,K}$ containing t, we have

$$\begin{aligned} F_K(t) &= \frac{\nu\left(]a_K, b_K]\right)}{\mathrm{P}_K\left(]a_K, b_K]\right)} \\ &= \frac{\nu_K\left(]a_K, b_K]\right)}{\overline{\mathrm{P}}(a_K)\left(1 - \exp\left(-\nu^c\left(]a_K, b_K]\right)\right)\prod_{a_K < u \le b_K}\left(1 - \Delta\nu^d(u)\right)\right)}. \end{aligned}$$

Since $]a_K, b_K] \downarrow \{t\}$ and $\overline{\mathrm{P}}(a_K) \to \overline{\mathrm{P}}(t-)$, it is clear from (4.17) that if $\Delta\nu(t) > 0$,

$$\lim_{K\to\infty} F_K(t) = \frac{\Delta\nu(t)}{\Delta\mathrm{P}(t)} = \frac{1}{\overline{\mathrm{P}}(t-)}.$$

If $\Delta\nu(t) = 0$,, writing $\nu_K^d = \nu^d\left(]a_K, b_K]\right)$, we have inequalities from (4.12) and (4.14) of the form

$$e^{-\gamma\nu_K^d} \le \prod_{a_K < u \le b_K}(1 - \Delta\nu(u)) \le e^{-\nu_K^d} \tag{4.18}$$

for $\gamma > 1$, chosen so that $1 - x \ge e^{-\gamma x}$ if $0 \le x \le c_K$, where $c_K = \sup\{\Delta\nu(u) : a_K < u \le b_K\}$. But since $\Delta\nu(t) = 0$, $\nu_K := \nu\left(]a_K, b_K]\right) \to 0$ and consequently also $c_K \to 0$ as $K \to \infty$. It is clear that for any $\gamma > 1$, (4.18) holds for K sufficiently large. Therefore

$$\frac{\nu_K}{\overline{\mathrm{P}}(a_K)\left(1 - e^{-\gamma\nu_K}\right)} \le F_K(t) \le \frac{\nu_K}{\overline{\mathrm{P}}(a_K)\left(1 - e^{-\nu_K}\right)}$$

for $\gamma > 1$, K sufficiently large, and since the expression on the right converges to $1/\overline{\mathrm{P}}(t-)$, and on the left to $1/(\gamma\overline{\mathrm{P}}(t-))$ as $K \to \infty$, the desired conclusion $\lim F_K(t) = 1/\overline{\mathrm{P}}(t-)$ follows.

To show that P has hazard measure ν it remains to show that ν is obtained by integrating the derivative,

$$\nu(t) = \int_{]0,t]} \frac{1}{\overline{\mathrm{P}}(s-)}\,\mathrm{P}(ds) \tag{4.19}$$

for $t < t^\dagger$. (This must also hold for $t = t^\dagger$ if $t^\dagger < \infty$ and $\Delta\mathrm{P}\left(t^\dagger\right) > 0$, but that is automatic if (4.19) is valid for $t < t^\dagger$; let $t \uparrow t^\dagger$ to obtain

$$\nu\left(\left]0, t^\dagger\right[\right) = \nu\left(t^\dagger-\right) \le \frac{1}{\overline{\mathrm{P}}\left(t^\dagger-\right)} < \infty$$

and so, by the definition of $t^\dagger$, $\Delta\nu\left(t^\dagger\right) = 1$, and $\Delta\nu\left(t^\dagger\right) = \Delta\mathrm{P}\left(t^\dagger\right)/\overline{\mathrm{P}}\left(t^\dagger-\right)$ follows since $t^\dagger$ is the termination point for P).

By Proposition A.0.2, (4.19) will hold if for every $t_0 < t^\dagger$ there is a constant c such that for $0 \le s < t \le t_0$,

$$\nu(t) - \nu(s) \le c\left(\mathrm{F_P}(t) - \mathrm{F_P}(s)\right).$$

Since by (4.11) and (4.13)

$$\begin{aligned}\mathrm{P}\left(]s,t]\right) &\ge \overline{\mathrm{P}}(s)\left(1-\exp\left(-\nu\left(]s,t]\right)\right)\right)\\ &\ge \overline{\mathrm{P}}(t_0)b_0\nu\left(]s,t]\right)\end{aligned}$$

for $b_0 > 0$ chosen so small that $1 - e^{-x} \ge b_0 x$ for $0 \le x \le \nu\left(]0,t_0]\right)$, we may use $c = 1/\left(\overline{\mathrm{P}}(t_0)b_0\right)$.

The final assertion of the theorem is that there is only one probability that has a given hazard measure ν as its hazard measure, i.e., the probability with survivor function given by (4.8). So assume that $\mathrm{P}_1 \ne \mathrm{P}_2$ are two probabilities with hazard measure ν. Certainly, both probabilities have termination point $t^\dagger$ given by (4.7) and defining $t_0 = \inf\left\{t : \overline{\mathrm{P}}_1(t) \ne \overline{\mathrm{P}}_2(t)\right\}$ we must have $t_0 < t^\dagger$. For $t \in \left[t_0, t^\dagger\right[$ consider

$$d(t) := \left|\mathrm{P}_1\left([t_0,t[\right) - \mathrm{P}_2\left([t_0,t[\right)\right| = \left|\int_{[t_0,t[}\left(\overline{\mathrm{P}}_1(s-) - \overline{\mathrm{P}}_2(s-)\right)\,\nu(ds)\right|.$$

Here by the definition of t_0,

$$\overline{\mathrm{P}}_1(s-) - \overline{\mathrm{P}}_2(s-) = \mathrm{P}_2\left([t_0,s[\right) - \mathrm{P}_1\left([t_0,s[\right),$$

so it follows that

$$d(t) \le \int_{[t_0,t[} d(s)\,\nu(ds). \tag{4.20}$$

Choose any $t_1 \in \left]t_0, t^\dagger\right[$ and put

$$K = \sup\left\{d(s) : t_0 < s < t_1\right\}.$$

By (4.20), for $t \in [t_0, t_1[$,

$$d(t) \le K\nu\left([t_0,t[\right) \le K\nu\left([t_0,t_1[\right)$$

and hence

$$K \le K\nu\left([t_0,t_1[\right).$$

But for $t_1 \downarrow t_0$, $\nu\left([t_0,t[\right) \downarrow \nu\left(\{t_0\}\right) < 1$, by choosing $t_1 > t_0$ sufficiently close to t_0 it follows that $K = 0$, and therefore also that $d(t) = 0$ for $t > t_0$ sufficiently close to t_0; thus, for these t, $\overline{\mathrm{P}}_1(t) = \overline{\mathrm{P}}_2(t)$, in contradiction with the definition of t_0. □

The following immediate consequence of Theorem 4.1.1 will be used later:

Corollary 4.1.2 *If U is a $\overline{\mathbb{R}}_+$-valued random variable with hazard measure ν, then*

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

Example 4.1.1 (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., $\overline{\mathrm{P}}(t) = e^{-\lambda t}$.

(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 $\Delta\mathrm{P}\,(n) = p\,(1-p)^{n-1}$ for $n \in \mathbb{N}$.

Hazard measures have some further nice properties that we quote, leaving the proofs as an exercise.

Proposition 4.1.3 (i) *If* P *has hazard measure ν and $t_0 > 0$ is such that $\overline{\mathrm{P}}(t_0) > 0$, then the hazard measure for the conditional probability* $\mathrm{P}\left(\cdot\,|\,]t_0,\infty]\right)$ *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 $\mathrm{P}_1, \mathrm{P}_2$ and hazard measures ν_1, ν_2 respectively, then provided $\Delta\mathrm{P}_1(t)\Delta\mathrm{P}_2(t) = 0$ 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 P_1 and P_2 do not have atoms in common on $\mathbb{R}_+$ is essential for the assertion to be valid, as is made clear from the following.

Exercise 4.1.1 Show part (i) of Proposition 4.1.3. Also show the following sharpening of part (ii): if U_1, U_2 are $\overline{\mathbb{R}}_+$-valued and independent random variables with distributions $\mathrm{P}_1, \mathrm{P}_2$ and hazard measures ν_1, ν_2 respectively, then $U = \min(U_1, U_2)$ has hazard measure ν given by

$$\nu\left(]0,t]\right) = \nu_1\left(]0,t]\right) + \nu_2\left(]0,t]\right) - \sum_{0<s\le t} \Delta\nu_1(s)\Delta\nu_2(s).$$

(*Hint:* show and use that

$$\nu\left(]0,t]\right) = \mathbb{E}\left[\frac{1}{\overline{\mathrm{P}}_1(U-)\overline{\mathrm{P}}_2(U-)}; U \le t\right]).$$

For later use we give the following result:

Lemma 4.1.4 *There exists a universal constant $c < 1$ such that for any probability* P *on $\overline{\mathbb{R}}_+$ with hazard measure ν, it holds that*

$$\int_{]0,\infty[} \mathrm{P}(ds)\, e^{-\nu(s)} \le c,$$

where $\nu(s) := \nu\left(]0,s]\right)$.

Proof. As motivation, suppose first that P has Lebesgue density f. Then the integral above equals

$$\int_0^\infty f(s)\overline{\mathrm{P}}(s)\,ds = \lim_{t\to\infty}\left(-\tfrac{1}{2}\left(\overline{\mathrm{P}}(t)\right)^2 + \tfrac{1}{2}\right) \le \tfrac{1}{2}.$$

For the general case, introduce

$$I(t) := \int_{]0,t]} \mathrm{P}(ds)\, e^{-\nu(s)},$$

and note that by partial integration

$$I(t) = 1 - \overline{\mathrm{P}}(t)e^{-\nu(t)} + \int_{]0,t]} \mathrm{P}(ds)\,\overline{\mathrm{P}}(s-)D_{\mathrm{P}}e^{-\nu(s)}, \tag{4.21}$$

where the differentiation D_{P} is in the sense described in Appendix A, and where the formula itself is most easily verified by differentiation with respect to P — it is always valid for $t < t^\dagger$ and extends to $t = t^\dagger$ in case $t^\dagger < \infty$ and $\Delta\mathrm{P}(t^\dagger) > 0$.

Now, by direct computation if $\overline{\mathrm{P}}(s-) > 0$,

$$D_{\mathrm{P}}e^{-\nu(s)} = e^{-\nu(s-)}\frac{1}{\overline{\mathrm{P}}(s-)}\frac{\exp\left(-\Delta\mathrm{P}(s)/\overline{\mathrm{P}}(s-)\right) - 1}{\Delta\mathrm{P}(s)/\overline{\mathrm{P}}(s-)}$$

using the obvious definition $\left(e^{-x} - 1\right)/x = -1$ if $x = 0$, and so

$$D_{\mathrm{P}}e^{-\nu(s)} \le e^{-\nu(s-)}\frac{1}{\overline{\mathrm{P}}(s-)}\gamma \tag{4.22}$$

where

$$\gamma := \sup_{0\le x\le 1}\left(e^{-x} - 1\right)/x$$

is < 0 and ≥ -1. But since also $\nu(s-) \le \nu(s)$, we may replace $\nu(s-)$ by $\nu(s)$ in (4.22) and from (4.21) we deduce that

$$(1-\gamma)\,I(t) \le 1 - \overline{\mathrm{P}}(t)e^{-\nu(t)} \le 1$$

and the assertion of the proposition follows with $c = 1/(1-\gamma)$. $\square$

4.2 Adapted and predictable processes

In most of this chapter we shall deal 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 3.1 and 3.2. Recall that any such probability may be viewed as an SPP, respectively an MPP through the following definition of the sequence of jump times (τ_n) on W:

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

see (3.1), and the following definition of the sequence of jump times (τ_n) and marks (η_n) on $\mathcal{M}$,

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

$$(\eta_n \in A) = \bigcup_{K'=1}^{\infty} \bigcap_{K=K'}^{\infty} \bigcup_{k=1}^{\infty} \left(\overline{N}_{(k-1)/2^K}^\circ = n-1, N_{k/2^K}^\circ(A) - N_{(k-1)/2^K}^\circ(A) = 1\right) \tag{4.23}$$

for $n \in \mathbb{N}$, $A \in \mathcal{E}$, see (3.6), (3.7). Because we have assumed that $(E, \mathcal{E})$ is a Borel space (see p. 11) we are able to identify the exact value of η_n from (4.23) and may therefore write

$$\mu^\circ = \sum_{n:\tau_n<\infty} \varepsilon_{(\tau_n,\eta_n)},$$

see (2.4). Other 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 $N_t^\circ(A) : \mathcal{M} \to \mathbb{N}_0$ (for $A \in \mathcal{E}$) are defined by

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

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

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

$$\xi_n = \begin{cases} (\tau_1, \ldots, \tau_n) & \text{on } W, \\ (\tau_1, \ldots, \tau_n; \eta_1, \ldots, \eta_n) & \text{on } \mathcal{M} \end{cases} \tag{4.24}$$

as well as (cf. p. 26)

$$\xi_{\langle t\rangle} = \begin{cases} \xi_{N_t^\circ}, \\ \xi_{\overline{N}_t^\circ}, \end{cases} \quad \xi_{\langle t-\rangle} = \begin{cases} \xi_{N_{t-}^\circ}, \\ \xi_{\overline{N}_{t-}^\circ}, \end{cases} \tag{4.25}$$

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

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

$$\mathcal{H}_t = \sigma\left(\left(N_s^\circ\right)_{0\leq s\leq t}\right) \tag{4.26}$$

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

$$\mathcal{H}_t = \sigma\left(\left(N_s^\circ(A)\right)_{0\leq s\leq t, A\in\mathcal{E}}\right). \tag{4.27}$$

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

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

Considering the space $(\mathcal{M}, \mathcal{H})$ it should be noted that if $\mathcal{E}' \subset \mathcal{E}$ is a collection of sets in $\mathcal{E}$ that generates $\mathcal{E}$ (the σ-algebra $\sigma\left(\mathcal{E}'\right)$ generated by the sets in $\mathcal{E}'$ is $\mathcal{E}$ itself) such that $E \in \mathcal{E}'$ and $\mathcal{E}'$ is closed under the formation of finite intersections, i.e., $A_1 \cap A_2 \in \mathcal{E}'$ if $A_1, A_2 \in \mathcal{E}'$, then $\mathcal{H}_t = \sigma\left(\left(N_s^\circ(A)\right)_{0\leq s\leq t, A\in\mathcal{E}'}\right)$.

Remark 4.2.1 The filtrations $(\mathcal{H}_t)$ are defined entirely from the structure of the two spaces W and $\mathcal{M}$. If, in addition, a probability Q is given on the relevant space, it is customary in general process theory as part of the so-called usual conditions to enlarge each $\mathcal{H}_t$ by including all subsets ($\mathcal{H}$-measurable or not) of Q-null sets from $\mathcal{H}$. In this book, however, we always use the $\mathcal{H}$ as defined originally and the $\mathcal{H}_t$ as in (4.26) and (4.27) — *no completed σ-algebras are ever considered.* Keeping to these (restrictive) definitions is essential for results such as in Proposition 4.2.1 below to be valid! (For more on the usual conditions and completions, see Appendix B).

A *process* $X = (X_t)_{t\geq 0}$ with state space $(G, \mathcal{G})$ defined on W or $\mathcal{M}$ is a family of G-valued random variables; each X_t is a measurable map from $(W, \mathcal{H})$ or $(\mathcal{M}, \mathcal{H})$ to $(G, \mathcal{G})$. The process is *measurable* if the mapping $(t, w) \overset{\mathbf{X}}{\mapsto} X_t(w)$, respectively $(t, m) \overset{\mathbf{X}}{\mapsto} X_t(m)$, 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 $\mathcal{H}_0$-measurable (i.e., if $\mathcal{G}$ separates points, $X_0 \equiv x_0$ for some $x_0 \in G$) and if $\mathbf{X}$ 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 σ- algebra $\mathcal{P}$*. Here $\mathcal{P}$ is generated by the subsets

$$]t, \infty[\times H \quad (t \geq 0, H \in \mathcal{H}_t) \tag{4.28}$$

of $\mathbb{R}_+ \times W$ or $\mathbb{R}_+ \times \mathcal{M}$.

If $(G, \mathcal{G}) = \left(\mathbb{R}^d, \mathcal{B}^d\right)$, the process X is *right-continuous* if all *sample paths* are right-continuous, i.e., in the case of $(\mathcal{M}, \mathcal{H})$, the map $t \mapsto X_t(m)$ is right-continuous on $\mathbb{R}_0$ for all $m \in \mathcal{M}$; similarly X is *left-continuous, continuous, cadlag, increasing* if for all m, $t \mapsto X_t(m)$ is respectively left-continuous on $\mathbb{R}_+$, continuous, right-continuous on $\mathbb{R}_0$ with left limits on $\mathbb{R}_+$, increasing in each of the d coordinates.

Appendix B contains a more general discussion of the concepts from process theory introduced here.

Example 4.2.1 Let X be an $\mathbb{R}^d$-valued process defined on $(\mathcal{M}, \mathcal{H})$ (or $(W, \mathcal{H})$). Then X is measurable if it is either right-continuous or left-continuous, e.g., if X is right-continuous this follows from the representation

$$X_t(m) = \lim_{K \to \infty} \sum_{k=1}^{\infty} 1_{[(k-1)2^{-K}, k2^{-K}[}(t) X_{k2^{-K}}(m)$$

where each term in the sum is $\mathcal{B}_0 \otimes \mathcal{H}$-measurable since it is a product of a measurable function of t and a measurable function of m.

The concept of predictability is crucial and therefore deserves some comments. Note that from (4.28) it follows immediately that all sets of the form $]s, t] \times H$ or $]s, t[\times H$ for $0 \le s < t$ and $H \in \mathcal{H}_s$ belong to $\mathcal{P}$; therefore also a process X defined, for example, on $(\mathcal{M}, \mathcal{H})$ of the form

$$X_t(m) = 1_{]s_0, t_0]}(t)\, U(m) \quad \text{or} \quad 1_{]s_0, t_0[}(t)\, U(m)$$

for some $0 \le s_0 < t_0 \le \infty$ and some $\mathbb{R}$-valued and $\mathcal{H}_{s_0}$-measurable random variable U is predictable. It is critically important in (4.28) that the interval $]t, \infty[$ is open to the left!

Intuitively a process X should be understood to be predictable, if for every $t > 0$ it is possible to compute the value of X_t from everything that has been observed strictly before time t, i.e., observations referring to the time interval $[0, t[$. The prime example is

Example 4.2.2 An $\mathbb{R}$-valued process X on e.g., $(\mathcal{M}, \mathcal{H})$, that is left-continuous with each X_t $\mathcal{H}_t$-measurable (i.e., X is left-continuous and adapted), is predictable: just use the representation

$$X_t(m) = \lim_{K \to \infty} \sum_{k=1}^{\infty} 1_{](k-1)2^{-K}, k2^{K}]}(t) X_{(k-1)2^{-K}}(m)$$

for $t > 0$, $m \in \mathcal{M}$.

The following very useful result characterizes the σ-algebras $\mathcal{H}_t$ (as defined by (4.26), (4.27)) as well as adapted and predictable processes, and also shows that the filtration $(\mathcal{H}_t)$ is right-continuous. (For this the definitions (4.26) and (4.27) of the $\mathcal{H}_t$ are essential, cf. Remark 4.2.1). Recall the definitions (4.24), (4.25) of ξ_n, $\xi_{\langle t \rangle}$ and $\xi_{\langle t- \rangle}$.

Proposition 4.2.1 (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*

$$H \cap \left(N_t^\circ = n\right) = (\xi_n \in B_n, \tau_{n+1} > t). \tag{4.29}$$

(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* $(z_n, t) \mapsto f_{z_n}^{(n)}(t)$ *from* $\mathbb{R}_+^n \times \mathbb{R}_0$ *to* $\mathbb{R}$ *such that identically on* W *and for all* $t \in \mathbb{R}_0$,

$$X_t = f_{\xi_{\langle t \rangle}}^{\langle t \rangle}(t).$$

(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* $(z_n, t) \mapsto f_{z_n}^{(n)}(t)$ *from* $\mathbb{R}_+^n \times \mathbb{R}_0$ *to* $\mathbb{R}$ *such that identically on* W *and for all* $t \in \mathbb{R}_0$

$$X_t = f_{\xi_{\langle t- \rangle}}^{\langle t- \rangle}(t). \tag{4.30}$$

(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*

$$H \cap \left(\overline{N}_t^\circ = n\right) = (\xi_n \in C_n, \tau_{n+1} > t). \tag{4.31}$$

(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* $t \in \mathbb{R}_0$

$$X_t = f_{\xi_{\langle t \rangle}}^{\langle t \rangle}(t). \tag{4.32}$$

(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* $t \in \mathbb{R}_0$

$$X_t = f_{\xi_{\langle t- \rangle}}^{\langle t- \rangle}(t). \tag{4.33}$$

Note. Because $\mathcal{H}_0$ is the trivial σ-algebra, the meaning of (4.29) and (4.31) for $n = 0$ is that $H \in \mathcal{H}_0$ iff $H \cap \left(N_t^\circ = 0\right)$, resp. $H \cap \left(\overline{N}_t^\circ = 0\right)$, equals either the empty set or the set $(\tau_1 > t)$. Also, for $n = 0$ the function $f_{z_n}^{(n)}(t)$ is a function $f^{(0)}(t)$ of t only.

Remark 4.2.2 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. (3.12)). For a process to be predictable, to find its value at t it suffices to know the number of jumps on $[0, t[$, their location in time and the marks.

Example 4.2.3 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 an $m \in \mathcal{M}$ with $N_t^\circ(m, A) = 0$ and a different m' with $N_{t-}^\circ\left(m', A\right) = 0$, $N_t^\circ\left(m', A\right) = 1$ (the first jump for m' occurs at time t, resulting in a mark in A), but were $N^\circ(A)$ predictable, by Proposition 4.2.1, $N_t^\circ(m, A) = N_t^\circ(m', A)$.

As an example of the representation (4.32), note that on $(\mathcal{M}, \mathcal{H})$, $N_t^\circ(A)$ has this representation 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 of Proposition 4.2.1. We just prove (b).

(i). Note first that since $\xi_n \in \mathrm{K}^{(n)}(E)$ always (see p. 22 for the definition of $\mathrm{K}^{(n)}(E)$), in (4.31) it is only the part of C_n inside $\mathrm{K}^{(n)}(E)$ that is relevant: C_n may be replaced by $C_n \cap \mathrm{K}^{(n)}(E)$.

It is easy to see that the class of sets $H \in \mathcal{H}_t$ that have the representation (4.31) is a σ-algebra (e.g., if (4.31) holds for H, it also holds for H^c, replacing C_n by $\left(]0, t]^n \times E^n\right) \backslash C_n$). It therefore suffices to show that (4.31) holds for the members of the $\mathcal{H}_t$-generating class $\left(\left(N_s^\circ(A) = l\right)\right)_{l \in \mathbb{N}_0, s \leq t, A \in \mathcal{E}}$. 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 (4.31) holds with

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

where here and below it is notationally convenient to allow z_n – the vector $(t_1, \ldots, t_n;$ $y_1, \ldots, y_n)$ – to vary freely in all of $]0, t]^n \times E^n$ rather than the part of $\mathrm{K}^{(n)}(E)$ with $t_n \leq t$. Suppose conversely that $H \subset \mathcal{M}$ satisfies (4.31) 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, since the class of sets C_n for which $(\xi_n \in C_n, \tau_{n+1} > t) \in \mathcal{H}_t$ forms a σ-algebra of measurable subsets of $]0, t]^n \times E^n$, it suffices to consider C_n of the form

$$C_n = \{z_n : t_k \le s, y_k \in A\} \tag{4.34}$$

for some k and s with $1 \le k \le n$, $s \le t$, $A \in \mathcal{E}$, 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 (4.34), use a variation of (4.23) to verify that

$$\begin{aligned}(\xi_n \in C_n, \tau_{n+1} > t) &= (\tau_k \le s, \eta_k \in A, \tau_{n+1} > t) \\ &= \left(\overline{N}_s^\circ \ge k, \overline{N}_t^\circ < n+1\right) \cap H_n\end{aligned}$$

where

$$H_n = \bigcup_{K'=1}^\infty \bigcap_{K=K'}^\infty \bigcup_{j=1}^{2^K} \left(\overline{N}_{(j-1)s/2^K}^\circ = k-1,\ N_{js/2^K}^\circ(A) - N_{(j-1)s/2^K}^\circ(A) = 1\right),$$

and it is clear that $(\xi_n \in C_n, \tau_{n+1} > t) \in \mathcal{H}_t$.

(ii). We must show that $\mathcal{H}_{t+} \subset \mathcal{H}_t$. So suppose that $H \in \mathcal{H}_{t+}$, i.e., that $H \in \mathcal{H}_{t+k^{-1}}$ 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

$$H \cap \left(\overline{N}_{t+\frac{1}{k}}^\circ = n\right) = \left(\xi_n \in C_{n,k}, \tau_{n+1} > t + \tfrac{1}{k}\right). \tag{4.35}$$

Now consider the set of $m \in \mathcal{M}$ belonging to the set (4.35) for k sufficiently large. Because $t \mapsto \overline{N}_t^\circ$ is right-continuous and piecewise constant and because $m \in \left(\tau_{n+1} > t + k^{-1}\right)$ for k sufficiently large iff $m \in (\tau_{n+1} > t)$, this leads to the identity

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

But since $C_{n,k}$ is a measurable subset of $\left]0, t+k^{-1}\right]^n \times E^n$, $\bigcup_{K=1}^\infty \bigcap_{k=K}^\infty C_{n,k}$ is a measurable subset of $]0, t]^n \times E^n$, hence (4.31) holds and by (i), $H \in \mathcal{H}_t$.

(iii). Let X be an $\mathbb{R}$-valued, adapted process. Writing $X = X^+ - X^-$, the difference between the positive and negative part of X, it is seen that to establish (4.32) it is enough to consider $X \ge 0$. Further, writing

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

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

$$X_t(m) = 1_D(t, m) \tag{4.36}$$

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

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

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

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

for some measurable $C_{n,t} \subset]0,t]^n \times E^n$, and it follows that (4.32) holds with

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

It remains to show that this $f^{(n)}$, which is measurable in z_n for each t, is a measurable function of (z_n, t) . But since $X = 1_D$ is measurable, so is

$$(t,m) \overset{\rho}{\mapsto} X_t(m) 1_{\left(\overline{N}_t^\circ = n\right)}(m) = f_{\xi_n(m)}^{(n)}(t)\, 1_{(\tau_{n+1} > t)}(m),$$

and hence, so is $\rho \circ (\mathrm{id}, \varphi)$ where $(\mathrm{id}, \varphi) : \mathbb{R}_+ \times K(E) \to \mathbb{R}_+ \times \mathcal{M}$ is given by $(\mathrm{id}, \varphi)(t, z_\infty) = (t, \varphi(z_\infty))$ for $z_\infty = (t_1, t_2, \ldots ; y_1, y_2, \ldots) \in K(E)$. (For the definition of φ, see (3.5); the σ-algebras on the two product spaces are the obvious product σ-algebras). Since, with obvious notation,

$$\rho \circ (\mathrm{id}, \varphi)(t, z_\infty) = f_{z_n}^{(n)}(t) 1_{]t,\infty]}(t_{n+1})$$

the assertion follows easily.

For the converse, suppose that X is given by (4.32) with all $(z_n, t) \mapsto f_{z_n}^{(n)}(t)$ measurable. It is immediately checked that X is then measurable, and it remains to see that X_t is $\mathcal{H}_t$-measurable. But for $B \in \mathcal{B}$, $n \in \mathbb{N}_0$,

$$\begin{aligned}\left(X_t \in B, \overline{N}_t^\circ = n\right) &= \left(f_{\xi_n}^{(n)}(t) \in B, \tau_n \le t < \tau_{n+1}\right) \\ &= (\xi_n \in C_n, \tau_{n+1} > t)\end{aligned}$$

where

$$C_n = \left\{z_n \in]0,t]^n \times E^n : f_{z_n}^{(n)}(t) \in B\right\}.$$

Hence by (i), $\left(X_t \in B, \overline{N}_t^\circ = n\right) \in \mathcal{H}_t$ and it follows that X_t is $\mathcal{H}_t$-measurable.

(iv). Arguing as in the beginning of the proof of (iii), to prove that any predictable X has the form (4.33), by standard extension arguments it suffices to consider

$$X_t(m) = 1_D(t,m)$$

for $D \in \mathcal{P}$. But it is clear that the class of $D \in \mathcal{P}$ such that $X = 1_D$ has the form (4.33) is closed under the operations $\backslash$ and $\uparrow$: if $D_1 \supset D_2$ belong to the class, so does $D_1 \backslash D_2$, and if $D_1 \subset D_2 \subset \cdots$ belong to the class, so does $\bigcup_n D_n$. Since the desired representation holds trivially for $D = \mathbb{R}_+ \times \mathcal{M}$, it therefore suffices to consider D of the form

$$D =]s, \infty[\times H \quad (s > 0, H \in \mathcal{H}_s), \tag{4.37}$$

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

Since $H \in \mathcal{H}_s$, (4.31) holds,

$$H \cap \left(\overline{N}_s^\circ = k\right) = (\xi_k \in C_k, \tau_{k+1} > s)$$

for some measurable $C_k \subset]0, s]^k \times E^k$, and consequently, if D is given by (4.37), for $t \geq 0$,

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

using that when $t > s$, then $\overline{N}_{t-}^\circ \geq \overline{N}_s^\circ$. Thus (4.32) holds with

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

Finally, let X be given by (4.33) and let us show that X is predictable. Clearly X_0 is $\mathcal{H}_0$-measurable. To show that X is $\mathcal{P}$-measurable on $\mathbb{R}_+ \times \mathcal{M}$, it suffices to consider $f^{(n)}$ of the form

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

for some $C_n \in \mathcal{B}_+^n \otimes \mathcal{E}^n$, $s_n \geq 0$; in particular X only takes the values 0 and 1. We need

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

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

$$D_n = \bigcup_{K'=1}^{\infty} \bigcap_{K=K'}^{\infty} \bigcup_{j=0}^{\infty} \left]s_{n,j,K}, s_{n,j+1,K}\right] \times \left(\xi_n \in C_{n,j,K}, \tau_{n+1} > s_{n,j,K}\right)$$

where

$$C_{n,j,K} = C_n \cap \left(]0, s_{n,j,K}]^n \times E^n\right).$$

But this shows that $D_n \in \mathcal{P}$ because by (i) $\left(\xi_n \in C_{n,j,K}, \tau_{n+1} > s_{n,j,K}\right) \in \mathcal{H}_{s_{n,j,K}}$ and the interval $\left]s_{n,j,K}, s_{n,j+1,K}\right]$ is open to the left. □

Proposition 4.2.1 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 4.2.2 *If Q is a probability on $(\mathcal{M}, \mathcal{H})$ (or $(W, \mathcal{H})$) and U is an $\mathbb{R}$-valued random variable with $E|U| < \infty$, then*

$$E[U|\mathcal{H}_t] = \sum_{n=0}^{\infty} 1_{\left(\overline{N}_t^\circ = n\right)} E\left[U|\xi_n, \tau_{n+1} > t\right]. \tag{4.38}$$

Note. $E = E_Q$ denotes Q-expectation. $E\left[U\,|\xi_n, \tau_{n+1} > t\right]$ is the conditional expectation of U given the random variables ξ_n and $1_{(\tau_{n+1}>t)}$ considered only on the set $(\tau_{n+1} > t)$.

Proof. By definition $E\left[U\,|\xi_n, \tau_{n+1} > t\right]$ 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 (as displayed in (4.39) below). Thus we may write for some measurable function ρ_n that

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

and (4.32) shows that the sum on the right of (4.38) is $\mathcal{H}_t$-measurable. Next, let $H \in \mathcal{H}_t$ and use (4.31) and the definition of $E\left[U\,|\xi_n, \tau_{n+1} > t\right]$ to obtain

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

(That it is permitted to interchange the order of summation and integration in the second line follows by monotone convergence if $U \geq 0$. In particular, therefore, for $U \geq 0$ with $EU < \infty$, the random variable on the right of (4.38) is Q-integrable. For arbitrary U with $E\,|U| < \infty$ we then have

$$E\left|\sum_{n=0}^{\infty} 1_{\left(\overline{N}_t^{\circ}=n\right)} E\left[U\,|\xi_n, \tau_{n+1} > t\right]\right| \leq E\sum_{n=0}^{\infty} 1_{\left(\overline{N}_t^{\circ}=n\right)} E\left[|U|\,|\xi_n, \tau_{n+1} > t\right] < \infty$$

and dominated convergence can be used to argue for the interchange). □

Remark 4.2.3 The usefulness of Corollary 4.2.2 comes from the construction of SPPs and MPPs which makes it natural to work with the conditional expectations on the right of (4.38). Note that on $(\tau_{n+1} > t)$,

$$E\left[U\,|\xi_n, \tau_{n+1} > t\right] = \frac{E\left[U 1_{(\tau_{n+1}>t)}\,|\xi_n\right]}{Q\left(\tau_{n+1} > t\,|\xi_n\right)} = \frac{1}{\overline{P}^{(n)}_{\xi_n}(t)} E\left[U 1_{(\tau_{n+1}>t)}\,|\xi_n\right]. \tag{4.39}$$

4.3 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 3.1.1. If $t_n < \infty$, write $\nu^{(n)}_{z_n}$ for the hazard measure for $P^{(n)}_{z_n}$, cf. Section 4.1, so that

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

when $\overline{P}_{z_n}^{(n)}(t-) > 0$. Since $P_{z_n}^{(n)}$ is concentrated on $]t_n, \infty]$, so is $\nu_{z_n}^{(n)}$, wherefore $\nu_{z_n}^{(n)}(]0, t_n]) = 0$.

Definition 4.3.1 The *compensator* for Q is the non-negative process $\Lambda^\circ = \left(\Lambda_t^\circ\right)_{t\geq 0}$ on $(W, \mathcal{H})$ given by

$$\Lambda_t^\circ = \sum_{n=0}^{N_t^\circ} \nu_{\xi_n}^{(n)}\left(\left]\tau_n, \tau_{n+1} \wedge t\right]\right). \tag{4.40}$$

The compensator combines the hazard measures $\nu_{\xi_n}^{(n)}$, following $\nu^{(0)}$ on the interval $]0, \tau_1]$, $\nu_{\xi_1}^{(1)}$ on $]\tau_1, \tau_2]$ etc. It is critical that the half-open intervals, where each hazard measure operates, are closed to the right, open to the left.

Note that in (4.40) only for the last term, $n = N_t^\circ$, is $\tau_{n+1} \wedge t = t$.

Clearly, for all $w \in W$, $t \mapsto \Lambda_t^\circ(w)$ is ≥ 0, 0 at time 0, increasing and right-continuous. The ambiguity in the choice of the Markov kernels (see p.19) 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.,

$$Q \bigcap_{t\geq 0} \left(\widetilde{\Lambda}_t^\circ = \Lambda_t^\circ\right) = 1. \tag{4.41}$$

(For a proof, use that $\widetilde{P}^{(0)} = P^{(0)}$ and that for $n \geq 1$, we have $\widetilde{P}_{z_n}^{(n)} = P_{z_n}^{(n)}$ except for $z_n \in B_n$, where $Q(\xi_n \in B_n) = 0$. Thus, outside the Q-null set $\bigcup_n (\xi_n \in B_n)$ we have that $\widetilde{\Lambda}_t^\circ = \Lambda_t^\circ$ for all t).

As it stands Λ_t° can take the value ∞. However, by Corollary 4.1.2 it follows that

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

Another important property of the compensator, see Theorem 4.1.1 (iii), is that

$$Q \bigcap_{t\geq 0} \left(\Delta\Lambda_t^\circ \leq 1\right) = 1. \tag{4.42}$$

With these properties of Λ° in place, Λ° may be identified with a random positive and σ-finite measure on $(\mathbb{R}_0, \mathcal{B}_0)$ that we also denote by Λ°: $\Lambda^\circ([0, t]) = \Lambda_t^\circ$. In particular for Q-a.a $w \in W$ it holds that $\Lambda^\circ(\{0\})(w) = 0$, that $\Lambda^\circ([0, t])(w) < \infty$ for all $t \in \mathbb{R}_0$ and that $\Lambda^\circ(\{t\})(w) \leq 1$ for all $t \in \mathbb{R}_0$.

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

Let $\left(P^{(n)}\right)$, $\left(\pi^{(n)}\right)$ be the sequences of Markov kernels generating Q, and let $\nu_{z_n}^{(n)}$ denote the hazard measure for $P_{z_n}^{(n)}$.

Definition 4.3.2 (i) The *total compensator* for Q is the process $\overline{\Lambda}^{\circ} = \left(\overline{\Lambda}_t^{\circ}\right)_{t\geq 0}$ on $(\mathcal{M}, \mathcal{H})$ given by

$$\overline{\Lambda}_t^{\circ} = \sum_{n=0}^{\overline{N}_t^{\circ}} \nu_{\xi_n}^{(n)} \left(\left]\tau_n, \tau_{n+1} \wedge t\right]\right). \tag{4.43}$$

(ii) The *Q-compensator for the counting process* $N^{\circ}(A)$ is the process $\Lambda^{\circ}(A) = \left(\Lambda_t^{\circ}(A)\right)_{t\geq 0}$ on $(\mathcal{M}, \mathcal{H})$ given by

$$\Lambda_t^{\circ}(A) = \int_{]0,t]} \pi_{\xi_{\langle s-\rangle},s}^{\langle s-\rangle}(A)\, \overline{\Lambda}^{\circ}(ds). \tag{4.44}$$

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

$$L^{\circ}(C) = \int_{\mathbb{R}_0} \int_E 1_C(s, y)\, \pi_{\xi_{\langle s-\rangle},s}^{\langle s-\rangle}(dy)\, \overline{\Lambda}^{\circ}(ds) \quad (C \in \mathcal{B}_0 \otimes \mathcal{E}). \tag{4.45}$$

Part (i) of the definition mimics (4.40) and $\overline{\Lambda}^{\circ}$ and all $\Lambda^{\circ}(A)$ have the same properties as Λ° in the CP-case as listed above. In particular, for Q-a.a m, the right-continuous functions $t \mapsto \overline{\Lambda}^{\circ}(m)$ and $t \mapsto \Lambda_t^{\circ}(A, m)$ for arbitrary $A \in \mathcal{E}$ define positive measures $\Lambda^*(dt, m)$, all of which satisfy that $\Lambda^*(\{0\}, m) = 0$, $\Lambda^*([0, t], m) < \infty$ for all t and $\Lambda^*(\{t\}, m) \leq 1$ for all t.

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

Note that

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

Also note that if $\left(\widetilde{P}^{(n)}\right)$ and $\left(\widetilde{\pi}^{(n)}\right)$ are a second set of sequences of Markov kernels that generate Q, and that if $\widetilde{L}^{\circ}$ is the corresponding compensating measure, then

$$Q \bigcap_{C \in \mathcal{B}_0 \otimes \mathcal{E}} \left(\widetilde{L}^{\circ}(C) = L^{\circ}(C)\right) = 1;$$

cf. the argument leading to (4.41).

An essential property of compensators is presented in

Proposition 4.3.1 (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 4.2.1 (aiv) in mind, on $\left(N_{t-}^{\circ} = n\right)$ where either $N_t^{\circ} = n$ or $N_t^{\circ} = n + 1$,

$$\Lambda_t^{\circ} = \sum_{k=0}^{n-1} \nu_{\xi_k}^{(k)} \left(\left]\tau_k, \tau_{k+1}\right]\right) + \nu_{\xi_n}^{(n)} \left(\left]\tau_n, t\right]\right) \tag{4.46}$$

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

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

if $N_t^{\circ} = n + 1$. But in this latter case, $\tau_{n+1} = t$ and it is seen that the expression (4.46) still holds and thus also the representation (4.30). □

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 (4.40) one finds that for any n, $z_n = (t_1, \ldots, t_n) \in K^{(n)}$,

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

for any $w \in \left(N_{t-}^{\circ} = n, \xi_n = z_n\right)$. Similarly, in the MPP case, cf. (4.43), for any n, $z_n = (t_1, \ldots, t_n; y_1, \ldots, y_n)$,

$$\nu_{z_n}^{(n)} \left(\left]t_n, t\right]\right) = \overline{\Lambda}_t^{\circ}(m) - \overline{\Lambda}_{t_n}^{\circ}(m) \tag{4.47}$$

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 (4.44), that the measure $\Lambda^{\circ}(dt, A)$ on $\mathbb{R}_0$ determined from the right-continuous process $t \mapsto \Lambda_t^{\circ}(A)$ is absolutely continuous with respect to the measure $\overline{\Lambda}^{\circ}(dt)$ with the Radon–Nikodym derivative

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

Thus

$$\pi_{z_n,t}^{(n)}(A) = \frac{d\Lambda^{\circ}(m, A)}{d\overline{\Lambda}^{\circ}(m)}(t) \tag{4.48}$$

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 only for $\overline{\Lambda}^{\circ}(m, dt)$-a.a. t, with an exceptionel set depending on A and m, so care is needed to obtain e.g., that $A \mapsto \pi_{z_n,t}^{(n)}(A)$ is always a probability on $(E, \mathcal{E})$.

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 4.3.2 (a) *Suppose that 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° and $\widetilde{L}^\circ$ determined by (4.45) using a specific but arbitrary choice of Markov kernels $P^{(n)}, \pi^{(n)}$ and $\widetilde{P}^{(n)}, \widetilde{\pi}^{(n)}$ that generate Q and $\widetilde{Q}$ respectively. We shall show that no matter how the kernels are chosen, it holds that

$$\widetilde{Q}\left(\widetilde{L}^\circ \neq L^\circ\right) > 0, \tag{4.49}$$

so that L° is not $\widetilde{Q}$-indistinguishable from $\widetilde{L}^\circ$, and hence cannot be a compensating measure for $\widetilde{Q}$, cf. (4.41).

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

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

and note that for any $z_{n-1} = (t_1, \ldots, t_{n-1}; y_1, \ldots, y_{n-1}) \in C_{n-1}$, necessarily then $t_{n-1} < \infty$ and $y_{n-1} \in E$. Also, by the definition of n,

$$Q\left(\xi_{n-1} \in C_{n-1}\right) = \widetilde{Q}\left(\xi_{n-1} \in C_{n-1}\right) > 0. \tag{4.50}$$

For $m \in (\xi_{n-1} \in C_{n-1})$, consider the conditional $\widetilde{Q}$-distribution of (τ_n, η_n) given ξ_{n-1} evaluated at m,

$$\widetilde{\rho}_m(dt, dy) = \widetilde{P}^{(n-1)}_{z_{n-1}}(dt)\, \widetilde{\pi}^{(n-1)}_{z_{n-1},t}(dy),$$

where $z_{n-1} = \xi_{n-1}(m)$, together with the corresponding Q-conditional distribution $\rho_m \neq \widetilde{\rho}_m$. Introduce $\sigma_{n-1}(m)$ as the smallest timepoint $\geq \tau_{n-1}(m)$ such that for $\tau_{n-1}(m) \leq t < \sigma_{n-1}(m)$, $\widetilde{\rho}_m \equiv \rho_m$ on all measurable subsets of $\left[\tau_{n-1}(m), t\right] \times E$, while for $t > \sigma_{n-1}(m)$, $\widetilde{\rho}_m$ and ρ_m are not identical on the subsets of $\left[\tau_{n-1}(m), t\right] \times E$. Then necessarily $\sigma_{n-1}(m) < \infty$, and since

$$\widetilde{\Lambda}^\circ_t(m, A) - \widetilde{\Lambda}^\circ_{\tau_{n-1}(m)}(m, A) = \int_{]\tau_{n-1}(m), t]} \widetilde{\nu}^{(n-1)}_{z_{n-1}}(ds)\, \widetilde{\pi}^{(n-1)}_{z_{n-1},s}(A)$$

if $\tau_n(m) \geq t$ with a similar formula valid for $\Lambda^\circ(A)$, we deduce that $\widetilde{L}^\circ \neq L^\circ$ if either

(i) $\tau_n(m) > \sigma_{n-1}(m)$, or
(ii) $\tau_n(m) = \sigma_{n-1}(m) = t_0$, and

$$\Delta\widetilde{\nu}^{(n-1)}_{z_{n-1}}(t_0)\, \widetilde{\pi}^{(n-1)}_{z_{n-1},t_0} \neq \Delta\nu^{(n-1)}_{z_{n-1}}(t_0)\, \pi^{(n-1)}_{z_{n-1},t_0}. \tag{4.51}$$

We now claim that

$$\widetilde{Q}\left(\widetilde{L}^{\circ} \neq L^{\circ} \,|\xi_{n-1}\right)(m) > 0 \tag{4.52}$$

for all $m \in (\xi_{n-1} \in C_{n-1})$, which will certainly imply (4.49) since then, using (4.50),

$$\widetilde{Q}\left(\widetilde{L}^{\circ} \neq L^{\circ}\right) \geq \widetilde{E}\left[\widetilde{Q}\left(\widetilde{L}^{\circ} \neq L^{\circ} \,|\xi_{n-1}\right); \xi_{n-1} \in C_{n-1}\right] > 0.$$

To establish (4.52), write $\widetilde{t}_{n-1}^{\dagger}(m)$ for the termination point of $\widetilde{P}_{z_{n-1}}^{(n-1)}$. If $\widetilde{t}_{n-1}^{\dagger}(m) > \sigma_{n-1}(m)$, (4.52) follows directly since then, by (i) above and the definition of $\widetilde{t}_{n-1}^{\dagger}(m)$,

$$\widetilde{Q}\left(\widetilde{L}^{\circ} \neq L^{\circ} \,|\xi_{n-1}\right)(m) \geq \widetilde{Q}\left(\tau_n > \sigma_{n-1}(m) \,|\xi_{n-1}\right)(m) > 0.$$

If $\widetilde{t}_{n-1}^{\dagger}(m) = \sigma_{n-1}(m) = t_0$, since by the definition of $\sigma_{n-1}(m)$, $\widetilde{\rho}_m$ and ρ_m are identical strictly before t_0, t_0 must be an atom for $\widetilde{P}_{z_{n-1}}^{(n-1)}$, and since by assumption $\widetilde{\rho}_m$ has no mass strictly after t_0, also (4.51) must hold. Thus for m with $\widetilde{t}_{n-1}^{\dagger}(m) = \sigma_{n-1}(m)$,

$$\widetilde{Q}\left(\widetilde{L}^{\circ} \neq L^{\circ} \,|\xi_{n-1}\right)(m) \geq \widetilde{Q}\left(\tau_n = \sigma_{n-1}(m) \,|\xi_{n-1}\right)(m) > 0,$$

as desired. □

Note. In the proof it was tacitly assumed that the set $\left(\widetilde{L}^{\circ} \neq L^{\circ}\right) \in \mathcal{H}$ and that the random time σ_{n-1} is measurable. Both facts follow when using that $\widetilde{L}^{\circ} \equiv L^{0}$ iff $\widetilde{\Lambda}_t^{\circ}(A) \equiv \Lambda_t^{\circ}(A)$ for all rational $t \geq 0$ and all $A \in \mathcal{E}_0$, with $\mathcal{E}_0$ a countable collection of sets from $\mathcal{E}$ such that $\mathcal{E}_0$ generates $\mathcal{E}$ and $\mathcal{E}_0$ is closed under the formation of finite intersections.

Example 4.3.1 Suppose Q on W makes N° Poisson λ; see Example 3.1.2. Then the hazard measure for $P_{z_n}^{(n)}$ is the restriction to $]t_n, \infty[$ of λ times Lebesgue measure. Thus

$$\Lambda_t^{\circ} = \lambda t;$$

in particular the compensator is deterministic.

Example 4.3.2 Suppose Q on W makes N° a renewal process with waiting time distribution with hazard measure ν; see Example 3.1.3. Then

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

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

$$\Lambda_t^{\circ} = \int_0^t u\left(s - \tau_{\langle s \rangle}\right) ds. \tag{4.53}$$

Example 4.3.3 Consider the homogeneous Poisson random measure with intensity measure $\rho = \lambda\kappa$ from Example 3.2.1. Then

$$\overline{\Lambda}_t^\circ = \lambda t, \quad \Lambda_t^\circ(A) = t\rho(A)$$

and

$$L^\circ = \ell \otimes \rho,$$

where ℓ is Lebesgue measure on $\mathbb{R}_0$. In particular L° is a deterministic product measure.

Example 4.3.4 Consider the MPP construction from Example 3.3.1 of time-homogeneous Markov chains on an at most countably infinite state space G. Thus, assuming that stability holds, Q is the probability on $(\mathcal{M}, \mathcal{H})$ generated by the Markov kernels (3.14) and (3.15). Consequently, with i_0 as the initial state for the chain, which itself is the step process $X^\circ = \left(X_t^\circ\right)$ given by $X_t^\circ = \eta_{\langle t\rangle}$, where $\eta_0 \equiv i_0$,

$$\overline{\Lambda}_t^\circ = \sum_{n=0}^{\overline{N}_t^\circ} \lambda_{\eta_n}\left((\tau_{n+1} \wedge t) - \tau_n\right) = \int_0^t \lambda_{X_{s-}^\circ}\, ds;$$

writing $\Lambda^{\circ i} = \Lambda^\circ(\{i\})$ for $i \in G$ so that $\Lambda^\circ(A) = \sum_{i \in A} \Lambda^{\circ i}$, *we have*

$$\Lambda_t^{\circ i} = \int_0^t 1_{(X_{s-}^\circ \neq i)} \lambda_{X_{s-}^\circ} \pi_{X_{s-}^\circ i}\, ds = \int_0^t 1_{(X_{s-}^\circ \neq i)} q_{X_{s-}^\circ i}\, ds.$$

We come now to 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

$$(\theta_s N^\circ)_t = \begin{cases} N_t^\circ - N_s^\circ & \text{if } t \geq s \\ 0 & \text{if } t < s, \end{cases} \qquad \theta_s \mu^\circ = \mu^\circ\left(\cdot \cap (]s, \infty[\times E)\right). \tag{4.54}$$

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

$$\tau_{n,s} = \tau_n \circ \theta_s = \tau_{k+n}, \quad \eta_{n,s} = \eta_n \circ \theta_s = \eta_{k+n}$$

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

$$\theta_s \mu^\circ = \sum_{n: s < \tau_n < \infty} \varepsilon_{(\tau_n, \eta_n)}.$$

For later use, we also introduce the *translated shift* $\theta_s^* : \mathcal{M} \to \mathcal{M}$ (with the obvious analogue on W) by

$$\theta_s^* \mu^\circ = \sum_{n: s < \tau_n < \infty} \varepsilon_{(\tau_n - s, \eta_n)}, \tag{4.55}$$

equivalently, for $C \in \mathcal{B}_0 \otimes \mathcal{E}$,

$$\theta_s^* \mu^\circ(C) = \theta_s \mu^\circ (C + s)$$

where $C + s := \{(t + s, y) : (t, y) \in C\}$. (Note that for $C \subset \mathbb{R}_+ \times E$, $\theta_s^* \mu^\circ(C) = \mu^\circ (C + s)$ since then $C + s \subset]s, \infty[\times E$).

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

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

Lemma 4.3.3 (Key lemma.) (a) *Let Q be a probability on $(W, \mathcal{H})$ with compensator Λ°, generated by the Markov kernels $\left(P^{(n)}\right)_{n \geq 0}$.*

(i) *For any $s \in \mathbb{R}_0$, the conditional distribution of $\theta_s N^\circ$ given $N_s^\circ = k$, $\xi_k = z_k$ for an arbitrary $k \in \mathbb{N}_0$, $z_k = (t_1, \ldots, t_k) \in \mathrm{K}^{(k)}$ with $t_k \leq s$ is the probability $\widetilde{Q} = \widetilde{Q}_{|k, z_k}$ on $(W, \mathcal{H})$ generated by the Markov kernels $\left(\widetilde{P}_{|k,z_k}^{(n)}\right)_{n \geq 0}$ from $\mathrm{K}^{(n)}$ to $\overline{\mathbb{R}}_+$ given by*

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

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

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

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

(ii) *For any $k_0 \in \mathbb{N}$, the conditional distribution of $\vartheta_{k_0} N^\circ$ given $\xi_{k_0} = z_{k_0}$ for an arbitrary $z_{k_0} = \left(t_1, \ldots, t_{k_0}\right) \in \mathrm{K}^{(k_0)}$ 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 $\left(\widetilde{P}_{\left|k_0,z_{k_0}\right.}^{(n)}\right)_{n \geq 0}$ from $\mathrm{K}^{(n)}$ to $\overline{\mathbb{R}}_+$ given by*

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

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

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

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

(b) *Let Q be a probability on $(\mathcal{M}, \mathcal{H})$, generated by the Markov kernels $\left(P^{(n)}\right)_{n\geq 0}$ and $\left(\pi^{(n)}\right)_{n\geq 0}$.*

(i) *For any $s \in \mathbb{R}_0$, the conditional distribution of $\theta_s\mu^\circ$ given $\overline{N}_s = k$, $\xi_k = z_k$ for an arbitrary $k \in \mathbb{N}_0$, $z_k = (t_1, \ldots, t_k; y_1, \ldots, y_k) \in \mathrm{K}_E^{(k)}$ 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 $\left(\widetilde{P}^{(n)}_{|k,z_k}\right)_{n\geq 0}$ from $\mathrm{K}_E^{(n)}$ to $\overline{\mathbb{R}}_+$ and $\left(\widetilde{\pi}^{(n)}_{|k,z_k}\right)_{n\geq 0}$ from $\mathrm{J}_E^{(n)}$ to $\overline{E}$ given by*

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

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

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

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

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

(ii) *For any $k_0 \in \mathbb{N}$, the conditional distribution of $\vartheta_{k_0}\mu^\circ$ given $\xi_{k_0} = z_{k_0}$ for an arbitrary $z_{k_0} = \left(t_1, \ldots, t_{k_0}; y_1, \ldots, y_{k_0}\right) \in \mathrm{K}_E^{(k_0)}$ with $t_{k_0} < \infty$ is the probability $\widetilde{Q} = \widetilde{Q}_{\big|k_0,z_{k_0}}$ on $(\mathcal{M}, \mathcal{H})$ generated by the Markov kernels $\left(\widetilde{P}^{(n)}_{\big|k_0,z_{k_0}}\right)_{n\geq 0}$ from $\mathrm{K}_E^{(n)}$ to $\overline{\mathbb{R}}_+$ and $\left(\widetilde{\pi}^{(n)}_{\big|k_0,z_{k_0}}\right)_{n\geq 0}$ from $\mathrm{J}_E^{(n)}$ to $\overline{E}$ given by*

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

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

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

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

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

Note. By Corollary 4.2.2, the conditional probabilities described in (ai) and (bi) simply determine the conditional distribution of θ_s given $\mathcal{H}_s$ on the set $\left(N_s^\circ = k, \xi_k = z_k\right)$, respectively $\left(\overline{N}_s^\circ = k, \xi_k = z_k\right)$. Thus, for example, $Q\left(\theta_s \mu^\circ \in \cdot \,|\mathcal{H}_s\right)$ is the probability on $(\mathcal{M}, \mathcal{H})$ generated by the Markov kernels $\widetilde{P}^{(n)}_{|\langle s\rangle, \xi_{\langle s\rangle}}$ from $\mathrm{K}_E^{(n)}$ to $\overline{\mathbb{R}}_+$ and $\widetilde{\pi}^{(n)}_{|\langle s\rangle, \xi_{\langle s\rangle}}$ from $\mathrm{J}_E^{(n)}$ to $\overline{E}$.

Proof. We only outline the proof. The expressions for $\widetilde{P}^{(0)}_{|k,z_k}$ in (ai), (bi) follow immediately from Corollary 4.2.2 and (4.39). 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 conditioning 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: if U, U' are random variables with values in two arbitrary measurable spaces, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and if $\mathbb{P}_{|U=u}$ denotes a regular conditional probability of $\mathbb{P}$ given $U = u$ for all u, then for each u one may consider a regular conditional probability of $\mathbb{P}_{|U=u}$ given $U' = u'$ for all u'. The result of this repeated conditioning, $\left(\mathbb{P}_{|U=u}\right)_{|U'=u'}$, is simply a regular conditional probability of $\mathbb{P}$ given $\left(U, U'\right) = \left(u, u'\right)$. $\square$

Example 4.3.5 As a first example on how the lemma may be used, consider the canonical homogeneous Poisson random measure Q (Examples 3.2.1 and 4.3.3), i.e., Q has compensating measure $\lambda\ell \otimes \kappa$ and the waiting times $\tau_n - \tau_{n-1}$ are iid exponential at rate λ, while the η_n are iid with distribution κ and are independent of the τ_n. But then the conditional distribution of $\theta_s \mu^\circ$ given $\overline{N}_s^\circ = k, \xi_k = z_k$ has

$$\overline{\widetilde{P}}^{(0)}_{|k,z_k}(t) = \frac{e^{-\lambda(t-t_k)}}{e^{-\lambda(s-t_k)}} = e^{-\lambda(t-s)}$$

for $t \geq s$, and after the first jump, $\theta_s \mu^\circ$ behaves exactly as μ°. Shifting back to time 0, it is seen that Q has the lack of memory property

$$Q\left(\theta_s^* \in \cdot \,|\mathcal{H}_s\right) = Q, \tag{4.56}$$

i.e., $\theta_s^* \mu^\circ$ is independent of $\mathcal{H}_s$ and has distribution Q.

Exercise 4.3.1 Show that if Q is a probability on $(\mathcal{H}, \mathcal{M})$ such that the lack of memory property (4.56) holds for all s, then Q is a homogeneous Poisson random measure.

Example 4.3.6 We shall illustrate the usefulness of the lemma by establishing the Markov property in its usual form for the homogeneous Markov chains on at most countably infinite state spaces, as they were constructed in Example 3.3.1; see also Example 4.3.4. We assume that stability holds.

For a given initial state $i_0 \in G$, we let $Q_{|i_0}$ denote the distribution of the RCM determined by the MPP from Example 3.3.1, and recall that on $\mathcal{M}$, the Markov chain is given by $X_t^\circ = \eta_{\langle t\rangle}$. By Lemma 4.3.3, given $\mathcal{H}_s$ on the set $\left(\overline{N}_s^\circ = k\right)$, the conditional distribution of the first jump time of θ_s has survivor function

$$\overline{\widetilde{P}}^{(0)}_{|k,\xi_k}(t) = e^{-\lambda_{\eta_k}(t-s)} \tag{4.57}$$

for $t \geq s$, while

$$\widetilde{\pi}^{(0)}_{t|k,\xi_k}(\{i\}) = \pi_{\eta_k,i} 1_{(\eta_k \neq i)}$$

determines the conditional distribution given $\mathcal{H}_s$ and $\tau_{1,s}$ of the state reached by θ_s through its first jump. Both quantities depend on the past ξ_k through $\eta_k = X^\circ_s$ only; since it is clear from the lemma that for $n \geq 1$ all $\overline{\widetilde{P}}^{(n)}_{\widetilde{z}_n|k,\xi_k}$ and $\widetilde{\pi}^{(n)}_{\widetilde{z}_n,t|k,\xi_k}$ depend neither on ξ_k nor on s (but only on $\widetilde{i}_n$, the last state listed in $\widetilde{z}_n$), we have shown the Markov property,

$$Q_{|i_0}(\theta_s \in \cdot \,|\mathcal{H}_s) = Q_{|i_0}\left(\theta_s \in \cdot \,\middle|X^\circ_s\right),$$

i.e., given the past the entire future depends only on the present. Because of the time homogeneity expressed in (4.57) by the fact that the dependence on s and t is through $t-s$ only, even more is true: with θ^*_s the translated shift given by (4.55), it is clear from the above that

$$Q_{|i_0}\left(\theta^*_s \in \cdot \,|\mathcal{H}_s\right) = Q_{|X^\circ_s}.$$

This is the *homogeneous Markov property* which implies in particular that under all $Q_{|i_0}$, the chain X° has the same transition probabilities.

The formulation we have given of Lemma 4.3.3 uses the Markov kernels to describe the conditional distributions of θ_s and ϑ_{k_0}. Alternatively they may be described using compensators and compensating measures. Here, using Proposition 4.1.3 (i), it is seen that if Q is a probability on $(\mathcal{M}, \mathcal{H})$ with compensating measure L°, then the conditional distribution of θ_s given $\overline{N}^\circ_s = k$, $\xi_k = z_k$ 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 $\overline{N}_s(m_0) = k$, $\xi_k(m_0) = z_k$,

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

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

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

In particular, with $\Lambda^\circ_{t|k,z_k}(A) = L^\circ_{t|k,z_k}([0,t] \times A)$,

$$\Lambda^\circ_{t|k,z_k}(m, A) = \begin{cases} \Lambda^\circ_t(\widetilde{m}, A) - \Lambda^\circ_s(\widetilde{m}, A) & \text{if } t \geq s, \\ 0 & \text{if } t < s. \end{cases} \tag{4.58}$$

For a probability Q on $(W, \mathcal{H})$, the conditional distribution of θ_s given $N^\circ_s = k$, $\xi_k = z_k$ has a compensator $\Lambda^\circ_{|k,z_k}$ given by

$$\Lambda^{\circ}_{t|k,z_k}(w) = \begin{cases} \Lambda^{\circ}_t(\widetilde{w}) - \Lambda^{\circ}_s(\widetilde{w}) & \text{if } t \geq s, \\ 0 & \text{if } t < s, \end{cases}$$

where for any $w_0 \in \left(N^{\circ}_s = k, \xi_k = z_k\right)$, $\widetilde{w} = \text{cross}\,(w_0, w)$ is given by $\widetilde{w}(u) = w_0(u)$ for $u \leq s$ and $= w_0(s) + (w(u) - w(s))$ for $u \geq s$.

We have in this section only discussed compensators for CPs and RCMs that do not explode (see p. 11). The definitions carry over verbatim to processes with explosion with Λ° in the CP case and L° in the RCM case now defined on the spaces allowing explosions, $\left(\overline{W}, \overline{\mathcal{H}}\right)$ and $\left(\overline{\mathcal{M}}, \overline{\mathcal{H}}\right)$ respectively, adding e.g., the requirement that $L^{\circ}\left([\tau_\infty, \infty[\times E\right) \equiv 0$ with $\tau_\infty = \inf\left\{t : \overline{N}^{\circ}_t = \infty\right\}$ the time of explosion. (For the definitions of $\left(\overline{W}, \overline{\mathcal{H}}\right)$ and $\left(\overline{\mathcal{M}}, \overline{\mathcal{H}}\right)$, see p. 12 and p. 15). One may then give the following characterization of τ_∞, which is of theoretical interest, but not very useful in practice.

Proposition 4.3.4 *Let Q be a probability on $\left(\overline{\mathcal{M}}, \overline{\mathcal{H}}\right)$. With $\overline{\Lambda}^{\circ}$ the total compensator for Q, it then holds Q-a.s. that*

$$\tau_\infty = \inf\left\{t : \overline{\Lambda}^{\circ}_t = \infty\right\}.$$

Proof. Define $\tau'_\infty = \inf\left\{t : \overline{\Lambda}^{\circ}_t = \infty\right\}$. We still have from Corollary 4.1.2 that for all n, $\overline{\Lambda}^{\circ}_{\tau_n} < \infty$ Q-a.s. and hence $\tau'_\infty \geq \tau_\infty$ a.s. For the converse, consider

$$\alpha_n := E\left(e^{-\overline{\Lambda}^{\circ}_{\tau_n}}; \tau_n < \infty\right).$$

Conditioning on ξ_n we get for $n \geq 0$,

$$\alpha_{n+1} = E\left[e^{-\overline{\Lambda}^{\circ}_{\tau_n}} \int_{]\tau_n,\infty[} P^{(n)}_{\xi_n}(ds)\, \exp\left(-\nu^{(n)}_{\xi_n}(]\tau_n, s])\right); \tau_n < \infty\right].$$

By Lemma 4.1.4 there is a constant $c < 1$ such that the inner integral is always $\leq c$. Thus $\alpha_{n+1} \leq c\alpha_n$ implying that $\alpha_n \leq c^n \to 0$ as $n \to \infty$. But then

$$E\left(e^{-\overline{\Lambda}^{\circ}_{\tau_\infty}}; \tau_\infty < \infty\right) \leq E\left(\lim_{n\to\infty} e^{-\overline{\Lambda}^{\circ}_{\tau_n}} 1_{(\tau_n < \infty)}\right) = \lim_{n\to\infty} \alpha_n = 0$$

by monotone convergence, and thus $\overline{\Lambda}^{\circ}_{\tau_\infty} = \infty$ a.s. on $(\tau_\infty < \infty)$ forcing $\tau'_\infty \leq \tau_\infty$ a.s. □

Exercise 4.3.2 Let Q be a probability on $\left(\overline{\mathcal{M}}, \overline{\mathcal{H}}\right)$. Show that $\overline{\Lambda}^{\circ}_\infty = \infty$ Q-a.s. on $\bigcap_n (\tau_n < \infty)$. (Hint: find $\lim_{n\to\infty} e^{-\overline{\Lambda}^{\circ}_{\tau_n}} 1_{(\tau_n<\infty)}$).

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

$$m \preceq_t w$$

if for all $t' \leq t$, m is delayed relative to w,

$$\overline{N}^{\circ}_{t'}(m) \leq N^{\circ}_{t'}(w)\,.$$

Proposition 4.3.5 *Assume that Λ° and $\overline{\widetilde{\Lambda}}^{\circ}$ are continuous. A sufficient condition for $\widetilde{Q}$ to be stable (non-exploding) is that there exists a set $H \in \overline{H}$ with $\widetilde{Q}(H) = 1$ such that*

$$\overline{\widetilde{\Lambda}}^{\circ}_t(m) - \overline{\widetilde{\Lambda}}^{\circ}_{s'}(m) \leq \overline{\Lambda}^{\circ}_t(w) - \overline{\Lambda}^{\circ}_s(w) \tag{4.59}$$

for all $s \leq s' \leq t$ and all $m \in H$, $w \in W$ with $m \preceq_t w$ and $\overline{N}^{\circ}_t(m) = \overline{N}^{\circ}_{s'}(m) = N^{\circ}_t(w) = N^{\circ}_s(w)$.

Proof. An outline of the proof: the idea is to construct on the same probability space, sequences $\mathcal{T} = (T_n)$ and $\left(\widetilde{\mathcal{T}}, \widetilde{\mathcal{Y}}\right) = \left(\left(\widetilde{T}_n\right), \left(\widetilde{Y}_n\right)\right)$ such that the CP determined by $\mathcal{T}$ has distribution Q, the RCM$_{\text{ex}}$ determined by $\left(\widetilde{\mathcal{T}}, \widetilde{\mathcal{Y}}\right)$ has distribution $\widetilde{Q}$ *and* such that for all n, $\widetilde{T}_n \geq T_n$. Since Q is stable this *coupling* will force $\widetilde{Q}$ to also be stable. The construction of the coupling is itself based on a simpler and standard coupling. If U and V are real-valued random variables with survivor functions

$$\overline{F}_U(x) = \mathbb{P}(U > x)\,, \quad \overline{F}_V(x) = \mathbb{P}(V > x)\,,$$

and U is *stochastically larger* than V,

$$\overline{F}_U(x) \geq \overline{F}_V(x)$$

for all x, then it is possible to define U' and V' on the same probability space such that U' has the same distribution as U, V' has the same distribution as V, *and* $U' \geq V'$.

Now let $\widetilde{P}^{(n)}_{\widetilde{z}_n}$, $\widetilde{\pi}^{(n)}_{\widetilde{z}_n,t}$ denote the Markov kernels generating the jump times and marks for $\widetilde{Q}$ and let $P^{(n)}_{z_n}$ denote the Markov kernels generating the jump times for Q. Because $\overline{\widetilde{\Lambda}}^{\circ}$ and Λ° are continuous we have (cf. (4.8) and (4.9))

$$\overline{\widetilde{P}}^{(n)}_{\widetilde{z}_n}(t) = \exp\left(-\widetilde{\nu}^{(n)}_{\widetilde{z}_n}\left(\left]\widetilde{t}_n, t\right]\right)\right), \quad \overline{P}^{(n)}_{z_n}(t) = \exp\left(-\nu^{(n)}_{z_n}\left(\left]t_n, t\right]\right)\right) \tag{4.60}$$

where $\widetilde{z}_n = (\widetilde{t}_1, \ldots, \widetilde{t}_n; \widetilde{y}_1, \ldots, \widetilde{y}_n)$ and $z_n = (t_1, \ldots, t_n)$. But since

$$\widetilde{\nu}^{(n)}_{\widetilde{z}_n}\left(\left]\widetilde{t}_n, t\right]\right) = \overline{\widetilde{\Lambda}}^{\circ}_t(m) - \overline{\widetilde{\Lambda}}^{\circ}_{\widetilde{t}_n}(m)$$

for any $m \in H$ such that $\overline{N}^{\circ}_{\widetilde{t}_n}(m) = \overline{N}^{\circ}_{t-}(m) = n$ and $\xi_n(m) = \widetilde{z}_n$, and since similarly,

$$\nu^{(n)}_{z_n}\left(\left]t_n, t\right]\right) = \Lambda^{\circ}_t(w) - \Lambda^{\circ}_{t_n}(w)$$

for any $w \in W$ such that $N^\circ_{t_n}(w) = N^\circ_{t-}(w) = n$ and $\xi_n(w) = z_n$, the assumption (4.59) translates into the inequality

$$\widetilde{\nu}^{(n)}_{\widetilde{z}_n}\left(\left]\widetilde{t}_n, t\right]\right) \leq \nu^{(n)}_{z_n}\left(\left]t_n, t\right]\right)$$

for all $n \geq 0$, all $\widetilde{z}_n$ and z_n such that $\widetilde{t}_k \geq t_k$ for $1 \leq k \leq n$, and all $t \geq \widetilde{t}_n$, i.e., because of (4.60),

$$\overline{\widetilde{P}}^{(n)}_{\widetilde{z}_n}(t) \geq \overline{P}^{(n)}_{z_n}(t). \tag{4.61}$$

Using this for $n = 0$, apply the simple coupling described above to obtain $T_1 \leq \widetilde{T}_1$. Then generate $\widetilde{Y}_1$ using the Markov kernel $\widetilde{\pi}^{(0)}_{\widetilde{T}_1}$. Since $T_1 \leq \widetilde{T}_1$, use (4.61) for $n = 1$ with $\widetilde{z}_1 = \left(\widetilde{T}_1, \widetilde{Y}_1\right)$ and $z_1 = T_1$ to generate $T_2 \leq \widetilde{T}_2$ and $\widetilde{\pi}^{(1)}_{(\widetilde{T}_1, \widetilde{Y}_1), \widetilde{T}_2}$ to generate $\widetilde{Y}_2$. To continue, it is clear the desired coupling construction of $\mathcal{T}$ and $(\mathcal{T}, \mathcal{Y})$ results. □

A simpler and more useful consequence of this result is presented in Corollary 4.4.4 in the next section.

4.4 Intensity processes

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

Definition 4.4.1 (a) Let Q be a probability on $(W, \mathcal{H})$ with compensator Λ°. A predictable process $\lambda^\circ \geq 0$ is an *intensity process* for Q if Q-a.s.

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

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

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

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

(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. $A \mapsto \lambda^\circ_t(A)$ 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, σ-finite measure on $(E, \mathcal{E})$ and if $\lambda^\circ = (\lambda^{\circ y})_{y \in E}$ is a collection of predictable processes $\lambda^{\circ y} \geq 0$ such that $(m, t, y) \mapsto \lambda^{\circ y}_t(m)$ is measurable and such that Q-a.s.

$$\Lambda^\circ_t(A) = \int_0^t \int_A \lambda^{\circ y}_s \, \kappa(dy) \, ds \quad (t \geq 0, A \in \mathcal{E})$$

or, equivalently,

$$L^\circ(C) = \int_{\mathbb{R}_0} \int_E 1_C(s, y)\, \lambda_s^{\circ y}\, \kappa(dy)\, ds \quad (C \in \mathcal{B}_0 \otimes \mathcal{E}),$$

then λ° is a κ*-intensity process* for Q.

The family $(\lambda^{\circ y})$ of processes in (biii) is an example of a *predictable field* that will appear in a more general form later, see p. 78.

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 4.4.2. However, in some contexts (see Proposition 4.4.3 and Theorem 5.1.1 below) it is essential to have the intensities predictable.

The requirement in (biii) stipulates that Q-a.s. the compensating measure L° for Q should have a density with respect to the product measure $\ell \otimes \kappa$:

$$L^\circ(dt, dy) = \lambda_t^y\, dt\, \kappa(dy).$$

Of course if λ° is a κ-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.34 of hazard functions.

Proposition 4.4.1 (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*

$$\lambda_t^\circ = u_{\xi_{\langle t- \rangle}}^{\langle t- \rangle}(t) \tag{4.62}$$

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, $P_{\xi_n}^{(n)}$ is absolutely continuous with respect to Lebesgue measure with hazard function $u_{\xi_n}^{(n)}$. Then $(\lambda^\circ(A))_{A \in \mathcal{E}}$ is an intensity measure for Q, where*

$$\lambda_t^\circ(A) = u_{\xi_{\langle t- \rangle}}^{\langle t- \rangle}(t)\, \pi_{\xi_{\langle t- \rangle}, t}^{\langle t- \rangle}(A). \tag{4.63}$$

(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 a density $p^{(n)}_{\xi_n,t}$ such that $(z_n, t, y) \mapsto p^{(n)}_{z_n,t}(y)$ is measurable, then $(\lambda^{\circ y})_{y\in E}$ is a κ-intensity for Q, where

$$\lambda^{\circ y}_t = u^{\langle t-\rangle}_{\xi_{\langle t-\rangle}}(t)\, p^{\langle t-\rangle}_{\xi_{\langle t-\rangle},t}(y). \tag{4.64}$$

Proof. The predictability of the intensity processes is ensured by the left-limits $\overline{N}^{\circ}_{t-}$, $\xi_{\langle t-\rangle}$ appearing in the expressions. Everything else follows from the definitions of the compensators; see (4.43) and (4.44), and the fact that

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

Example 4.4.1 The canonical homogeneous Poisson counting process with parameter λ has intensity process

$$\lambda^{\circ}_t \equiv \lambda;$$

cf. Example 4.3.1.

Example 4.4.2 For the renewal process (Example 4.3.2), if the waiting time distribution has hazard function u,

$$\lambda^{\circ}_t = u\left(t - \tau_{\langle t-\rangle}\right)$$

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 (4.53), but that the integrand in (4.53) is not predictable.

Example 4.4.3 Consider the homogeneous Poisson random measure from Examples 3.2.1 and 4.3.3 with intensity measure $\rho = \lambda\kappa$, where $\lambda > 0$ and κ is the probability on E that determines the distribution of the marks. Then

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

for all t and y defines a κ-intensity process.

Example 4.4.4 Consider the MPP construction of homogeneous Markov chains from Example 3.3.1. Letting κ denote counting measure on the state space G, it is clear from Example 4.3.4 that the RCM describing the chain has κ-intensity process

$$\lambda^{\circ i}_t = \lambda_{X^{\circ}_{t-}} \pi_{X^{\circ}_{t-},i} 1_{(X^{\circ}_{t-}\neq i)} = q_{X^{\circ}_{t-},i} 1_{(X^{\circ}_{t-}\neq i)}$$

for all $t \geq 0$, $i \in G$.

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

Proposition 4.4.2 (a) *Let Q be a probability on $(W, \mathcal{H})$ and assume that Q has an intensity process λ° given by (4.62) 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 (4.63) 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. 56 that $\eta_{1,t}$ is the mark for the first jump on $]t, \infty]$. The proof is now based on Lemma 4.3.3 and follows by explicit calculations that we show in the MPP case:

$$\begin{aligned}
&\frac{1}{h} Q\left(\overline{N}^\circ_{t+h} - \overline{N}^\circ_t \geq 1, \eta_{1,t} \in A \,|\mathcal{H}_t\right) \\
&= \frac{1}{h} Q\left(\tau_{1,t} \leq t+h, \eta_{1,t} \in A \,|\mathcal{H}_t\right) \\
&= \frac{1}{h} \int_t^{t+h} u^{\langle t \rangle}_{\xi_{\langle t \rangle}}(s) \exp\left(-\int_t^s u^{\langle t \rangle}_{\xi_{\langle t \rangle}}(v)\, dv\right) \pi^{\langle t \rangle}_{\xi_{\langle t \rangle}, s}(A)\, ds \\
&\to \lambda^\circ_{t+}(A),
\end{aligned}$$

as $h \downarrow 0$. □

Remark 4.4.1 Proposition 4.4.2 shows in what sense it is possible to build a model (CP or RCM) 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 CP or RCM. 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 4.4.2. 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 4.4.2 are typically not predictable. They must *never* be used in expressions such as (5.7) below. Also the next result will not in general hold for right limit intensities.

Proposition 4.4.3 (a) *Let Q be a probability on $(W, \mathcal{H})$ with the intensity process given by (4.62). 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 κ-intensity process given by (4.64). 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

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

where C is a random set of timepoints and marks, determined by ξ_{n-1}:

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

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

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

Exercise 4.4.1 In the same vein as Proposition 4.4.3, show that for any probability Q on $(\mathcal{M}, \mathcal{H})$ it holds for all $n \geq 1$ that

$$Q\left(\overline{\Lambda}^\circ_{\tau_{n+1}} = \overline{\Lambda}^\circ_{\tau_n}; \tau_{n+1} < \infty\right) = 0$$

and more generally that for any $A \in \mathcal{E}$

$$Q\left(\Lambda^\circ_{\tau_{n+1}}(A) = \Lambda^\circ_{\tau_n}(A); \tau_{n+1} < \infty, \eta_{n+1} \in A\right) = 0.$$

Let Q be a probability on $(W, \mathcal{H})$ (hence stable), and $\widetilde{Q}$ a probability on $\left(\overline{\mathcal{M}}, \overline{\mathcal{H}}\right)$ as in Proposition 4.3.5. Assume that the Markov kernels $P^{(n)}_{z_n}$, $\widetilde{P}^{(n)}_{\widetilde{z}_n}$ generating the jump times for Q and $\widetilde{Q}$ respectively are absolutely continuous with hazard functions $u^{(n)}_{z_n}$, $\widetilde{u}^{(n)}_{\widetilde{z}_n}$. Let also λ°, $\overline{\overline{\lambda}}^\circ$ denote the intensity processes,

$$\lambda^\circ_t = u^{\langle t-\rangle}_{\xi_{\langle t-\rangle}}(t), \quad \overline{\overline{\lambda}}^\circ_t = \widetilde{u}^{\langle t-\rangle}_{\xi_{\langle t-\rangle}}(t)$$

(with $\xi_{\langle t-\rangle}$ as defined on W in the first case and as defined on $\mathcal{M}$ in the second).

Corollary 4.4.4 *A sufficient condition for $\widetilde{Q}$ to be stable is that there exists $H \in \overline{\mathcal{H}}$ with $\widetilde{Q}(H) = 1$ such that*

$$\overline{\overline{\lambda}}^\circ_t(m) \le \lambda^\circ_t(w)$$

for all $t \ge 0$ and all $m \in H$, $w \in W$ with $m \preceq_t w$ and $\overline{N}^\circ_{t-}(m) = N^\circ_{t-}(w)$.

The proof is obvious (exercise) from Proposition 4.3.5. See p. 62 for the meaning of the 'inequality' $m \preceq_t w$.

Example 4.4.5 If $\widetilde{Q}$ is such that there exists constants $a_n \ge 0$ with $\sum 1/a_n = \infty$ and $\overline{\overline{\lambda}}^\circ_t \le a_{\langle t-\rangle}$ everywhere on $\mathcal{M}$ for every t, then $\widetilde{Q}$ is non-exploding; cf. Example 3.1.4.

Exercise 4.4.2 Suppose $a_n \ge 0$ with $\sum 1/a_n = \infty$ as in the example above and let $\alpha : \mathbb{R}_0 \to \mathbb{R}_0$ be a function such that $A(t) := \int_0^t \alpha(s)\, ds < \infty$ for all t.

(i) Show that the probability Q on $\left(\overline{W}, \overline{\mathcal{H}}\right)$ with intensity process $\lambda^\circ_t = \alpha(t) a_{\langle t-\rangle}$ is stable. (Hint: one possibility is to consider the stable Q' on $(W, \mathcal{H})$ with intensity process $\lambda'^\circ_t = a_{\langle t-\rangle}$ and then show that the under Q' the counting process $(N_u)_{u\ge 0}$ given by $N_u = N^\circ_{A(u)}$ has distribution Q).

(ii) Let $\widetilde{Q}$ be a probability on $\left(\overline{\mathcal{M}}, \overline{\mathcal{H}}\right)$ with a total intensity process $\overline{\overline{\lambda}}^\circ_t$ satisfying that there is a set $H \in \overline{\mathcal{H}}$ with $\widetilde{Q}(H) = 1$ such that

$$\overline{\overline{\lambda}}^\circ_t(m) \le \alpha(t) a_{\langle t-\rangle(m)}$$

for all t and all $m \in H$. Show that $\widetilde{Q}$ is stable.

4.5 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 4.5.1 below.

It is important for what follows that we know the filtration to be right-continuous, $\mathcal{H}_t = \mathcal{H}_{t+}$; see Proposition 4.2.1 (aii) and (bii).

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, $(\tau \le t) \in \mathcal{H}_t$ for all t). The *pre-τ σ-algebra* $\mathcal{H}_\tau$ is the collection

$$\{H \in \mathcal{H} : H \cap (\tau < t) \in \mathcal{H}_t \text{ for all } t\} = \{H \in \mathcal{H} : H \cap (\tau \le t) \in \mathcal{H}_t \text{ for all } t\}$$

of measurable sets. In particular each τ_n is a stopping time and $\mathcal{H}_{\tau_n}$ is the σ-algebra generated by ξ_n:

Exercise 4.5.1 Show that $\mathcal{H}_{\tau_n} = \sigma(\xi_n)$, the σ-algebra generated by ξ_n. (Consider the space $(\mathcal{M}, \mathcal{H})$ and use Proposition 4.2.1 (b): if $H \in \sigma(\xi_n)$, H has the form $(\xi_n \in C_n)$ for some set C_n, and from the proposition it follows readily that $H \in \mathcal{H}_{\tau_n}$. The converse is more difficult, but if $H \in \mathcal{H}_{\tau_n}$, it is at least easy to see that $H \cap (\tau_n = \infty) \in \sigma(\xi_n)$ simply because on $(\tau_n = \infty)$, ξ_n determines μ°. Then show that $H \cap (\tau_n < \infty) \in \sigma(\xi_n)$ by e.g., showing and using that

$$H \cap (\tau_n < \infty) = \bigcup_{K'=1}^{\infty} \bigcap_{K=K'}^{\infty} \bigcup_{k=1}^{\infty} H \cap \left(\tfrac{k-1}{2^K} < \tau_n \le \tfrac{k}{2^K}, \overline{N}^\circ_{k/2^K} = n \right)$$

together with Proposition 4.2.1 (b)).

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 $M = (M_t)_{t \ge 0}$, such that each $M^{\rho_n} := \left(M_{\rho_n \wedge t}\right)_{t \ge 0}$ 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, martingales and local martingales and the optional sampling theorem.)

Proposition 4.5.1 *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,*

$$Q \bigcap_{t \ge 0} (M_t = c) = 1,$$

where c is the constant value of the $\mathcal{H}_0$-measurable random variable M_0.

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,

$$EM_{\tau_1 \wedge t} = EM_0 = c.$$

But since M is predictable and $\overline{N}^{\circ}_{(\tau_1 \wedge t)-} = 0$, by Proposition 4.2.1 there is a Borel function f such that $M_{\tau_1 \wedge t} = f(\tau_1 \wedge t)$ and thus $c = f(0)$ and

$$f(0) = f(t)\overline{P}^{(0)}(t) + \int_{]0,t]} f(s)\, P^{(0)}(ds), \tag{4.65}$$

where $\int_{]0,t]} |f|\, dP^{(0)} < \infty$ because $E\left|M_{\tau_1 \wedge t}\right| < \infty$. Of course the identity (4.65) is required only for t such that $\overline{P}^{(0)}(t) > 0$, i.e., for $t \in \left[0, t^\dagger\right[$ if $\overline{P}^{(0)}\left(t^\dagger -\right) = 0$ or $t^\dagger = \infty$, and for $t \in \left[0, t^\dagger\right]$ if $\overline{P}^{(0)}\left(t^\dagger -\right) > 0$ and $t^\dagger < \infty$ with $t^\dagger$ the termination point for $P^{(0)}$.

If you are lazy, just assume that $P^{(0)}$ has a continuous density, deduce from (4.65) first that f is continuous, then that f is differentiable and differentiate to see that $f' \equiv 0$ so f is constant. For the general case, (4.65) gives that f is cadlag and finite on the domain of interest, and that $f\overline{P}^{(0)}$ is differentiable with respect to $P^{(0)}$ with $D_{P^{(0)}} f\overline{P}^{(0)} = -f$ (see Appendix A for a discussion of this form of differentiability). But it is verified directly that

$$D_{P^{(0)}} \frac{1}{\overline{P}^{(0)}}(t) = \frac{1}{\overline{P}^{(0)}(t-)\overline{P}^{(0)}(t)},$$

hence by the differentiation rule for products, (A.4) from Appendix A, it follows that f is differentiable with respect to $P^{(0)}$ with derivative

$$\begin{aligned}
D_{P^{(0)}} f(t) &= D_{P^{(0)}} \left(f\overline{P}^{(0)}/\overline{P}^{(0)}\right)(t) \\
&= f\overline{P}^{(0)}(t) D_{P^{(0)}} \left(1/\overline{P}^{(0)}\right)(t) + D_{P^{(0)}} \left(f\overline{P}^{(0)}\right)(t)/\overline{P}^{(0)}(t-) \\
&= 0.
\end{aligned} \tag{4.66}$$

That this implies that f is constant follows from Proposition A.0.2 in Appendix A: we show that f is constant on $[0, t_0[$ for a fixed $t_0 < t^\dagger$ by showing that there is a constant c' such that

$$|f(t) - f(s)| \le c' P^{(0)}(]s,t]) \tag{4.67}$$

for $0 \le s < t \le t_0$. But (4.65) gives

$$|f(t) - f(s)| = \left| \left(\frac{1}{\overline{P}^{(0)}(t)} - \frac{1}{\overline{P}^{(0)}(s)} \right) f\overline{P}^{(0)}(s) - \frac{1}{\overline{P}^{(0)}(t)} \int_{]s,t]} f\, dP^{(0)} \right|,$$

so (4.67) follows with

$$c' = \frac{1}{\left(\overline{P}^{(0)}(t_0)\right)^2} \sup_{[0,t_0]} \left| f\overline{P}^{(0)} \right| + \frac{1}{\overline{P}^{(0)}(t_0)} \sup_{[0,t_0]} |f|,$$

which is finite because f cadlag on $\left[0, t^\dagger\right[$ implies that f is bounded on $[0, t_0[$.

Thus it has been shown that f is constant on $\left[0, t^{\dagger}\right[$. We need to extend this to $\left[0, t^{\dagger}\right]$ if $t^{\dagger} < \infty$ and $\overline{P}^{(0)}\left(t^{\dagger}-\right) > 0$, but then $f\left(t^{\dagger}\right) = f\left(t^{\dagger}-\right)$ is immediate from (4.66).

To proceed beyond τ_1, by optional sampling for $n \geq 1$, $E\left[M_{\tau_{n+1}\wedge t} \left| \mathcal{H}_{\tau_n \wedge t}\right.\right] = M_{\tau_n \wedge t}$. But since $(\tau_n < t) \in \mathcal{H}_{\tau_n \wedge t}$ and, as is easily argued (see Exercise 4.5.2 below), $\mathcal{H}_{\tau_n \wedge t} \cap (\tau_n < t) = \mathcal{H}_{\tau_n} \cap (\tau_n < t)$, this implies

$$1_{(\tau_n < t)} E\left[M_{\tau_{n+1}\wedge t} \left| \mathcal{H}_{\tau_n}\right.\right] = 1_{(\tau_n < t)} E M_{\tau_n}, \tag{4.68}$$

which since $M_{\tau_n} = f^{(n-1)}_{\xi_{n-1}}(\tau_n)$ and $M_{\tau_{n+1}\wedge t} = f^{(n)}_{\xi_n}(\tau_{n+1} \wedge t)$ when $\tau_n < t$, see Proposition 4.2.1 (aiv) and (biv), reduces to the identity

$$1_{(\tau_n < t)} f^{(n-1)}_{\xi_{n-1}}(\tau_n) = 1_{(\tau_n < t)} \left[f^{(n)}_{\xi_n}(t) \overline{P}^{(n)}_{\xi_n}(t) + \int_{]\tau_n, t]} f^{(n)}_{\xi_n}(s)\, P^{(n)}_{\xi_n}(ds) \right];$$

using the argument above, this implies that $s \mapsto f^{(n)}_{\xi_n}(s)$ is constant on $\left]\tau_n, \tau_{n+1}\right]$ and equal to $f^{(n-1)}_{\xi_{n-1}}(\tau_n) = M_{\tau_n}$. Thus, by induction M is constant on each closed interval $\left[\tau_n, \tau_{n+1}\right]$ and $M \equiv M_0$ follows.

Now let 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 that M is constant 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

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

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. □

Exercise 4.5.2 In the proof of (4.68) it was used that

$$\mathcal{H}_{\tau_n \wedge t} \cap (\tau_n < t) = \mathcal{H}_{\tau_n} \cap (\tau_n < t).$$

Show that this is true. (Hint: the inclusion $\subset$ is obvious. For the converse, use that since $\mathcal{H}_{\tau_n} = \sigma(\xi_n)$, any set $H \in \mathcal{H}_{\tau_n}$ is of the form $(\xi_n \in C_n)$ for some measurable $C_n \subset \mathrm{K}^{(n)}_E$).

Remark 4.5.1 The result is peculiar to the point process setup: Brownian motion is the most famous example of a continuous, hence predictable and right-continuous, martingale which is not constant! The assumption that M be right-continuous is also vitally 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 — a simple example is provided by the basic Poisson martingale from Example 4.5.1 below. But then obviously (M_{t-}) is a left-continuous, hence predictable martingale, which is non-constant.

Theorem 4.5.2 (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 V°,* 0 *at time* 0, *such that $M = N^\circ - V^\circ$ 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 $\Lambda^\circ(A)$, $\Lambda_t^\circ(A) = L^\circ\left([0,t] \times A\right)$. 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 V°,* 0 *at time* 0, *such that $M = N^\circ(A) - V^\circ$ 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

$$EN_{\tau_1 \wedge t}^\circ(A) = E\Lambda_{\tau_1 \wedge t}^\circ(A), \tag{4.69}$$

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 (4.69) by explicit calculation. (Note that since $0 \le N_{\tau_1 \wedge t}^\circ(A) \le 1$ the expectation on the left is trivially finite for all Q; in particular it follows from (4.69) that $\Lambda_{\tau_1 \wedge t}^\circ(A)$ is Q-integrable for all Q and all t and A).

We claim first that from Lemma 4.3.3 (b) and (4.69) it follows that for all $n \in \mathbb{N}_0$ and all $t \ge 0$,

$$E\left[N_{\tau_{n+1} \wedge t}^\circ(A) - N_{\tau_n \wedge t}^\circ(A) \,|\xi_n\right] = E\left[\Lambda_{\tau_{n+1} \wedge t}^\circ(A) - \Lambda_{\tau_n \wedge t}^\circ(A) \,|\xi_n\right]. \tag{4.70}$$

This identity is obvious on $(\tau_n > t)$ and on $(\tau_n \le t)$ is just (4.69) applied to the conditional distribution of the shifted process $\vartheta_n \mu^\circ$; see Lemma 4.3.3 and (4.58) (with the fixed time point s there replaced by τ_n).

It is an immediate consquence of (4.69) and (4.70) that for all Q, A, and all n and t,

$$EN_{\tau_n \wedge t}^\circ(A) = E\Lambda_{\tau_n \wedge t}^\circ(A)$$

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

$$EN_t^\circ(A) = E\Lambda_t^\circ(A), \tag{4.71}$$

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 4.3.3 (b) and (4.71) applied to the conditional distribution of $\theta_s\mu^\circ$ given $\mathcal{H}_s$ that

$$E\left[N_t^\circ(A) - N_s^\circ(A)\,|\mathcal{H}_s\right] = E\left[\Lambda_t^\circ(A) - \Lambda_s^\circ(A)\,|\mathcal{H}_s\right],$$

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

$$E\left[M_t^\circ(A)\,|\mathcal{H}_s\right] = M_s^\circ(A).$$

It remains to establish (4.69). But

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

while

$$\begin{aligned} E\Lambda_{\tau_1\wedge t}^\circ(A) &= E\int_{]0,\tau_1\wedge t]} \pi_s^{(0)}(A)\, \nu^{(0)}\,(ds) \\ &= \overline{P}^{(0)}(t)\int_{]0,t]} \pi_s^{(0)}(A)\, \nu^{(0)}\,(ds) \\ &\quad + \int_{]0,t]} \left(\int_{]0,u]} \pi_s^{(0)}(A)\, \nu^{(0)}\,(ds)\right) P^{(0)}(du) \end{aligned} \tag{4.72}$$

and (4.69) follows by differentiation with respect to $P^{(0)}$: using (A.4) and Proposition A.0.1, the last expression in (4.72) is easily seen to have derivative $\pi_s^{(0)}(A)$, which is evidently the derivative of $EN_{\tau_1\wedge t}^\circ(A)$. It is also easy to satisfy the conditions of Proposition A.0.2 (with g constant), hence $EN_{\tau_1\wedge t}^\circ(A) - E\Lambda_{\tau_1\wedge t}^\circ(A)$ is constant as a function of t. Since both expectations are 0 for $t = 0$, (4.69) has been proved.

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 4.5.1: if also $N^\circ(A) - V^\circ$ is a local martingale, where V° is predictable, right-continuous and 0 at time 0, then $\Lambda^\circ(A) - V^\circ$ is a predictable, right-continuous 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 $\mu^{\circ\tau_n} := \mu^\circ(\cdot \cap \left[0, \tau_n^\circ\right] \times E)$ obviously has compensating measure $L^{\circ\tau_n} = L^\circ(\cdot \cap \left[0, \tau_n^\circ\right] \times E)$, (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 $E_n N_t^\circ(A) \le n < \infty$, 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). $\square$

Example 4.5.1 Let Q be the Poisson λ probability on $(W, \mathcal{H})$, where $\lambda > 0$. Then $M^\circ = \left(N_t^\circ - \lambda t\right)_{t \geq 0}$ is a Q-martingale, the *basic Poisson martingale.*

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 $\overline{\Lambda}^\circ$. (For a generalization, see Examples 4.7.3 and 4.7.4 below).

Proposition 4.5.3 (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 $EN_t^\circ < \infty$ 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 $EN_t^\circ(A) < \infty$ for all t. It always holds that $EM^{\circ 2}_{\tau_n \wedge t}(A) < \infty$ for all n and t, while for a given t, $EM_t^{\circ 2}(A) < \infty$ if $EN_t^\circ(A) < \infty$*

(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 $EN_t^\circ(A)$, $EN_t^\circ(A') < \infty$ for all t. It always holds that $E\left|M^\circ_{\tau_n \wedge t}(A)\, M^\circ_{\tau_n \wedge t}\left(A'\right)\right| < \infty$ for all n and t, while for a given t, $E\left|M_t^\circ(A)\, M_t^\circ\left(A'\right)\right| < \infty$ if $EN_t^\circ(A)$, $EN_t^\circ(A') < \infty$.*

Proof. We shall just prove (b), and for this assume that $EN_t^\circ(A) < \infty$ in the proof of (i) and that $EN_t^\circ(A) < \infty$, $EN_t^\circ(A') < \infty$ in the proof of (ii): to obtain in general the local martingales with reducing sequence (τ_n), just consider the distribution Q_n of Q stopped at τ_n as was done at the end of the proof of Theorem 4.5.2.

For the proof of (b), note first that all $\Lambda^\circ(A)$ are continuous when $\overline{\Lambda}^\circ$ is. Otherwise the technique from the proof of Theorem 4.5.2 is used, i.e., for the two parts of the Proposition it is argued that it suffices to show that for all Q, all A, all A' disjoint from A and all t,

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

$$EM^{\circ 2}_{\tau_1 \wedge t}(A) = E\Lambda^\circ_{\tau_1 \wedge t}(A), \quad EM^\circ_{\tau_1 \wedge t}(A)M^\circ_{\tau_1 \wedge t}(A') = 0, \tag{4.73}$$

respectively. These identities are then verified directly (see p. 76 below) through straightforward calculations leading to some not so straightforward partial integrations.

We now discuss the proof of (b) in some detail. The argument requires repeated use of the fact that if M is a Q-martingale and $\sigma \leq \sigma'$ are bounded stopping times, then if $M_\sigma, M_{\sigma'} \in L^2(Q)$,

$$E\left[(M_{\sigma'} - M_\sigma)^2 \,|\mathcal{H}_\sigma\right] = E\left[M^2_{\sigma'} - M^2_\sigma \,|\mathcal{H}_\sigma\right] \tag{4.74}$$

as follows writing

$$(M_{\sigma'} - M_\sigma)^2 = M^2_{\sigma'} - M^2_\sigma - 2M_\sigma\left(M_{\sigma'} - M_\sigma\right)$$

and using optional sampling. As a consequence of (4.74) we obtain the orthogonality relation

$$E\,(M_{\sigma'} - M_{\sigma})^2 = EM_{\sigma'}^2 - EM_{\sigma}^2. \tag{4.75}$$

As a first step in the proof, use (4.73) on $\vartheta_n \mu^\circ$ given ξ_n to obtain,

$$E\left[\left(M^\circ_{\tau_{n+1}\wedge t}(A) - M^\circ_{\tau_n\wedge t}(A)\right)^2 |\xi_n\right] = E\left[\Lambda^\circ_{\tau_{n+1}\wedge t}(A) - \Lambda^\circ_{\tau_n\wedge t}(A)\,|\xi_n\right],$$

$$E\left[\left(M^\circ_{\tau_{n+1}\wedge t}(A) - M^\circ_{\tau_n\wedge t}(A)\right)\left(M^\circ_{\tau_{n+1}\wedge t}(A') - M^\circ_{\tau_n\wedge t}(A')\right)|\xi_n\right] = 0,$$

and take expectations, remembering that $E\Lambda^\circ_{\tau_n\wedge t}(A) = EN^\circ_{\tau_n\wedge t}(A) \le n$, to arrive at

$$E\left(M^\circ_{\tau_{n+1}\wedge t}(A) - M^\circ_{\tau_n\wedge t}(A)\right)^2 = E\left(\Lambda^\circ_{\tau_{n+1}\wedge t}(A) - \Lambda^\circ_{\tau_n\wedge t}(A)\right) < \infty, \tag{4.76}$$

$$E\left(M^\circ_{\tau_{n+1}\wedge t}(A) - M^\circ_{\tau_n\wedge t}(A)\right)\left(M^\circ_{\tau_{n+1}\wedge t}(A') - M^\circ_{\tau_n\wedge t}(A')\right) = 0$$

with the integrand in the last expression integrable by (4.76) and the Cauchy–Schwarz inequality. (4.76) also implies that all $M^\circ_{\tau_n\wedge t}(A) \in L^2(Q)$, hence by (4.75),

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

and, by analogous reasoning,

$$EM^\circ_{\tau_n\wedge t}(A) M^\circ_{\tau_n\wedge t}(A') = 0. \tag{4.77}$$

Now, use Fatou's lemma and monotone convergence to obtain (for any Q such that $EN_t^\circ(A) < \infty$ for all t)

$$\begin{aligned} EM_t^{\circ 2}(A) &= E \liminf_{n\to\infty} M^{\circ 2}_{\tau_n\wedge t}(A) \\ &\le \liminf_{n\to\infty} EM^{\circ 2}_{\tau_n\wedge t}(A) \\ &= \lim_{n\to\infty} E\Lambda^\circ_{\tau_n\wedge t}(A) \\ &= E\Lambda_t^\circ(A) \\ &= EN_t^\circ(A) \\ &< \infty. \end{aligned} \tag{4.78}$$

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

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

for all n, t. Letting $n \uparrow \infty$ gives $EM_t^{\circ 2}(A) \geq E\Lambda_t^\circ$ which combined with the opposite inequality resulting from (4.78) yields

$$EM_t^{\circ 2}(A) = E\Lambda_t^\circ(A). \tag{4.79}$$

Applying this to $\theta_s\mu^\circ$ given $\mathcal{H}_s$ together with (4.74) finally shows that $M^{\circ 2}(A) - \Lambda^\circ(A)$ is a Q-martingale whenever $EN_t^\circ(A) < \infty$ for all t.

To show similarly that $M^\circ(A)M^\circ(A')$ is a Q-martingale if $EN_t^\circ(A) < \infty$, $EN_t^\circ(A') < \infty$ for all t, the critical step is to deduce from (4.77) that

$$EM_t^\circ(A)M_t^\circ(A') = 0.$$

The product is certainly Q-integrable since $M_t^\circ(A), M_t^\circ(A') \in L^2(Q)$, so the problem is to verify that

$$0 = \lim_{n\to\infty} EM_{\tau_n\wedge t}^\circ(A)M_{\tau_n\wedge t}^\circ(A') = EM_t^\circ(A)M_t^\circ(A').$$

As argued above, by (4.75) e.g.,

$$E\left(M_t^\circ(A) - M_{\tau_n\wedge t}^\circ(A)\right)^2 = EM_t^{\circ 2}(A) - EM_{\tau_n\wedge t}^{\circ 2}(A) \to 0,$$

and similarly for A'. Thus $M_{\tau_n\wedge t}^\circ(A) \to M_t^\circ(A)$ and $M_{\tau_n\wedge t}^\circ(A') \to M_t^\circ(A')$ in $L^2(Q)$ and consequently

$$\begin{aligned}
&|E\left(M_t^\circ(A)M_t^\circ(A') - M_{\tau_n\wedge t}^\circ(A)M_{\tau_n\wedge t}^\circ(A')\right)| \\
&= |E\left(M_t^\circ(A) - M_{\tau_n\wedge t}^\circ(A)\right)\left(M_t^\circ(A') - M_{\tau_n\wedge t}^\circ(A)'\right)| \\
&\leq \left[E\left(M_t^\circ(A) - M_{\tau_n\wedge t}^\circ(A)\right)^2 E\left(M_t^\circ(A') - M_{\tau_n\wedge t}^\circ(A)'\right)^2\right]^{\frac{1}{2}} \to 0,
\end{aligned}$$

as desired.

The proof of the proposition is completed by the verification of (4.73). For this, introduce

$$I_u(A) = \int_{]0,u]} \pi_s^{(0)}(A)\,\nu^{(0)}(ds), \quad I_u(A') = \int_{]0,u]} \pi_s^{(0)}(A')\,\nu^{(0)}(ds),$$

the (non-random) values of $\Lambda_u^\circ(A)$ and $\Lambda_u^\circ(A')$ on $(\tau_1 \geq u)$. Then note first that

$$\begin{aligned}
EM_{\tau_1\wedge t}^{\circ^2}(A) = {} & \overline{P}^{(0)}(t)\,(I_t(A))^2 \\
& + \int_{]0,t]} \pi_u^{(0)}(A)\,(1 - I_u(A))^2\; P^{(0)}(du) \\
& + \int_{]0,t]} \left(1 - \pi_u^{(0)}(A)\right)(I_u(A))^2\; P^{(0)}(du)
\end{aligned}$$

which by differentiation is seen to equal (4.72) — the proper approach is to use the method from Appendix A, but if things are smooth enough as functions of t, ordinary differentiation will do. Similarly, the second claim in (4.73) follows by differentiation in the expression

$$\begin{aligned} EM^\circ_{\tau_1\wedge t}(A)M^\circ_{\tau_1\wedge t}(A') &= \overline{P}^{(0)}(t)I_t(A) \\ &+ \int_{]0,t]} \pi_u^{(0)}(A)\,(1-I_u(A))\left(-I_u(A')\right)\,P^{(0)}\,(du) \\ &+ \int_{]0,t]} \pi_u^{(0)}(A')\,(-I_u(A))\left(1-I_u(A')\right)\,P^{(0)}\,(du) \\ &+ \int_{]0,t]} \left(1-\pi_u^{(0)}(A\cup A')\right)(-I_u(A))\left(-I_u(A')\right)\,P^{(0)}\,(du)\,. \end{aligned}$$

The assumption $A \cap A' = \emptyset$ is used here in an essential manner: the first mark is either inside A or inside A' or outside $A \cup A'$ with the three possibilities mutually exclusive. □

Example 4.5.2 Suppose Q is the canonical homogeneous Poisson RCM with compensating measure $L^\circ = \ell \otimes \rho$, where $\rho = \lambda\kappa$ with $\lambda > 0$ the total intensity for a jump and κ the distribution of the iid marks; cf. Example 4.3.3. From Theorem 4.5.2 it follows that $M_t^\circ(A) = N_t^\circ(A) - t\rho(A)$ defines a Q-martingale, while Proposition 4.5.3 gives that also $M_t^{\circ 2}(A) - t\rho(A)$ and $M_t^\circ(A)\,M_t^\circ\left(A'\right)$ when $A \cap A' = \emptyset$ define Q-martingales. These facts certainly agree with the assertion that each counting process $N^\circ(A)$ is homogeneous Poisson with intensity $\rho(A)$ such that the processes $N^\circ(A)$ and $N^\circ(A')$ are independent when A and A' are disjoint. But it must be remembered that this Poisson process structure is not apparent from the original definition in Example 3.2.1! The assertion will be firmly established in Proposition 4.7.2 below.

Remark 4.5.2 For (bi) it suffices to assume that $\Lambda^\circ(A)$ is continuous, and for (bii) that $\Lambda^\circ(A)$ and $\Lambda^\circ(A')$ are continuous. Without these continuity 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 Examples 4.7.3 and 4.7.4 below.

Exercise 4.5.3 Let Q be a probability on $(\mathcal{M}, \mathcal{H})$ such that $\overline{\Lambda}^\circ$ is continuous. Show that for any $A, A' \in \mathcal{E}$, the process $M^\circ(A)M^\circ(A') - \Lambda^\circ\left(A \cap A'\right)$ is a local Q-martingale (τ_n) always, and a true martingale if $EN_t^\circ(A) < \infty$ and $EN_t^\circ(A') < \infty$ for all t.

4.6 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 CPs as a special case of canonical RCMs). We shall first discuss stochastic integrals with respect to μ° and L° and then use them to arrive at the martingale representation theorem, Theorem 4.6.1 below.

The integrands will be functions of $m \in \mathcal{M}$, $t \geq 0$ and $y \in E$. A typical integrand is denoted S where $(m, t, y) \mapsto S_t^y(m)$ is assumed to be $\mathbb{R}$-valued and measurable (with respect to the Borel σ-algebra on $\mathbb{R}$ and the product σ-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 $S^y = \left(S_t^y\right)_{t \geq 0}$ and then refer to S as a field of processes. Particularly important are *predictable fields* which are fields with each S^y predictable.

The *stochastic integral*

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

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

$$N_t^\circ(S) = \sum_{n: \tau_n \leq t} S_{\tau_n}^{\eta_n} = \sum_{n=1}^{\overline{N}_t^\circ} S_{\tau_n}^{\eta_n}.$$

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

The stochastic integral $\Lambda^\circ(S) = \left(\Lambda_t^\circ(S)\right)_{t \geq 0}$,

$$\Lambda_t^\circ(S) := \int_{]0,t] \times E} S_s^y \, L^\circ(ds, dy) = \int_{]0,t]} \int_E S_s^y \, \pi_{\xi_{\langle s- \rangle}, s}^{\langle s- \rangle}(dy) \, \overline{\Lambda}^\circ(ds)$$

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

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

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

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

provided

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

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 field.

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

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

$$M_t^\circ(S) = N_t^\circ(S) - \Lambda_t^\circ(S). \tag{4.80}$$

This may also be understood as

$$M_t^\circ(S) = \int_{]0,t]} S_s^y \, M^\circ(ds, dy)$$

with $M^\circ := \mu^\circ - L^\circ$ being the *fundamental martingale measure* for Q. However, M° should be used with care: formally it is a random signed measure on $\mathbb{R}_0 \times E$ but need not be defined on all sets in $\mathcal{B}_0 \otimes \mathcal{E}$ since e.g., $\mu^\circ(\mathbb{R}_0 \times E) = L^\circ(\mathbb{R}_0 \times E) = \infty$ Q-a.s. is perfectly possible. The restriction of M° to $[0, t] \times E$ for any $t \in \mathbb{R}_0$ is of course well defined and bounded Q-a.s.

Theorem 4.6.1 (The martingale representation theorem) *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 field $S = (S^y)$ such that*

$$M_t = M_0 + M_t^\circ(S). \tag{4.81}$$

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

(1) *for all $n \geq 1, t \geq 0$,*

$$EN_{\tau_n \wedge t}^\circ(S) = E\Lambda_{\tau_n \wedge t}^\circ(S) \leq \infty,$$

and $M^\circ(S)$ given by (4.80) 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) *for all $t \geq 0$,*

$$EN_t^\circ(S) = E\Lambda_t^\circ(S) \leq \infty,$$

and $M^\circ(S)$ given by (4.80) is a Q-martingale if for all $t \geq 0$,

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

(iii) *If S is a predictable field, then*

(1) *$M^\circ(S)$ given by (4.80) 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 (4.80) is a Q-martingale if for all $t \geq 0$,*

$$EN^\circ_t(|S|) < \infty. \tag{4.82}$$

Note. (4.82) is satisfied in particular if S (as a function of m, s and y) is uniformly bounded on $[0, t]$ for all t *and* $E\overline{N}^\circ_t < \infty$ for all t.

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 4.5.2, it suffices to prove that

$$E \int_{]0,\tau_1 \wedge t] \times E} S^y_s \, \mu^\circ(ds, dy) = E \int_{]0,\tau_1 \wedge t] \times E} S^y_s \, L^\circ(ds, dy) \tag{4.83}$$

for all t and all Q, with L° the compensating for Q. By Proposition 4.2.1 (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 (4.83) reduces to

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

or

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

This is verified directly by partial integration or differentiation with respect to $P^{(0)}$ (see Appendix A), or by ordinary differentiation if things are smooth and you are lazy.

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

$$S^y_s = 1_{H_0} 1_{]s_0,\infty[}(s)\, 1_{A_0}(y) \tag{4.84}$$

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 (4.84) 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

$$M_t = f^{(n)}_{\xi_n}(t) \quad \text{on} \quad \left(\overline{N}^\circ_t = n\right) \tag{4.85}$$

for all n, t, cf. Proposition 4.2.1 (biii). By optional sampling

$$E\left[M_{\tau_{n+1}\wedge t} - M_{\tau_n} \left|\mathcal{H}_{\tau_n}\right.\right] = E\left[M_{\tau_{n+1}\wedge t} - M_{\tau_n} \left|\xi_n\right.\right] = 0 \quad \text{on } (\tau_n \le t),$$

an identity which using (4.85) we may write

$$\begin{aligned}&\overline{P}^{(n)}_{\xi_n}(t)\left(f^{(n)}_{\xi_n}(t) - f^{(n)}_{\xi_n}(\tau_n)\right)\\&+\int_{]\tau_n,t]} P^{(n)}_{\xi_n}(ds)\int_E \pi^{(n)}_{\xi_n,s}(dy)\left(f^{(n+1)}_{\text{join}\,(\xi_n,(s,y))}(s) - f^{(n)}_{\xi_n}(\tau_n)\right) = 0\end{aligned} \tag{4.86}$$

on $(\tau_n \le t)$, where join $(\xi_n,(s,y)) = (\tau_1,\dots,\tau_n,s;\eta_1,\dots,\eta_n,y)$.

We want to find S such that (4.81) holds. Since each S^y is predictable we may write

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

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

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

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

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

It is seen that $M^{\circ}(S) \equiv M$ if for all n, t, y on $(\tau_n \le t)$ it holds that

$$-\int_{]\tau_n,t]} \nu^{(n)}_{\xi_n}(ds)\int_E \pi^{(n)}_{\xi_n,s}(dy)\, g^{(n)}_{\xi_n}(s,y) = f^{(n)}_{\xi_n}(t) - f^{(n)}_{\xi_n}(\tau_n), \tag{4.87}$$

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

or that (4.87) holds and

$$g^{(n)}_{\xi_n}(t,y) = f^{(n+1)}_{\text{join}\,(\xi_n,(t,y))}(t) - f^{(n)}_{\xi_n}(t), \tag{4.88}$$

where equation (4.88) defines $g^{(n)}_{\xi_n}$.

It remains to see that $g^{(n)}_{\xi_n}$ satisfies (4.87). But from (4.86) it follows that $t \mapsto f(t) := f^{(n)}_{\xi_n}(t)$ 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, (A.4)), (4.86) implies that

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

or equivalently, using (4.88)

$$\overline{P}\,(t-)\,D_P f(t) = -\int_E \pi^{(n)}_{\xi_n,t}\,(dy)\;g^{(n)}_{\xi_n}\,(t,y)\,. \tag{4.89}$$

Since both sides of (4.87) vanish as $t \downarrow \tau_n$, with the left-hand side obviously differentiable with respect to P, to prove (4.87) it suffices to show that

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

But recalling the definition (4.2) of hazard measures, it is seen that this is precisely (4.89). □

Remark 4.6.1 It is often important to be able to show that a local martingale is a martingale. The conditions in Theorem 4.6.1 (ii2) and (iii2) are sufficient for this but far from necessary.

Remark 4.6.2 Suppose M is a right-continuous local Q-martingale. It is important to note that if the total compensator $\overline{\Lambda}^\circ$ is continuous, then (4.81) immediately shows that M is *piecewise continuous,* i.e., continuous on each interval $\left[\tau_n, \tau_{n+1}\right[$.

Example 4.6.1 A particularly simple example of the representation from Theorem 4.6.1 is obtained by considering a sequence $(X_n)_{n\in\mathbb{N}}$ of independent $(G,\mathcal{G})$-valued random variables and then viewing the sequence as an MPP $((T_n)\,,(Y_n))$ with mark space $(G,\mathcal{G})$ by taking $T_n \equiv n$, $Y_n \equiv X_n$ as was done in Example 3.2.2. The distribution Q of this MPP is determined by the Markov kernels (cf. (3.11))

$$P^{(n)}_{z_n} = \varepsilon_{n+1}, \qquad \pi^{(n)}_{z_n,n+1}\,(A) = \mathbb{P}\,(X_{n+1} \in A)$$

where only z_n of the form $(1,\ldots,n;\, y_1,\ldots,y_n)$ have to be considered.

Since $\nu^{(n)}_{z_n} = \varepsilon_{n+1}$ the compensators for Q are given by

$$\overline{\Lambda}^\circ_t \equiv [t]\,, \qquad \Lambda^\circ_t\,(A) = \sum_{n=1}^{[t]} \mathbb{P}\,(X_n \in A) \quad (A \in \mathcal{G})\,,$$

$[t]$ denoting the integer part of t. With $(S^y)_{y\in G}$ a predictable field,

$$S^y_t = f^{\langle t-\rangle}_{\xi_{\langle t-\rangle}}\,(t,y)\,,$$

it therefore follows that the stochastic integral $M^\circ(S)$ becomes

$$M_t^\circ(S) = \sum_{n=1}^{[t]} \left\{ f_{\xi_{n-1}}^{(n-1)}(n, \eta_n) - \int_G f_{\xi_{n-1}}^{(n-1)}(n, y)\, \mathbb{P}(X_n \in dy) \right\}.$$

By Theorem 4.6.1 all local Q-martingales have this form and a sufficient condition for $M^\circ(S)$ to be a martingale is that for every $n \geq 1$, the function

$$(y_1, \ldots, y_n) \mapsto f_{z_{n-1}}^{(n-1)}(n, y_n)$$

is bounded when $z_{n-1} = (1, \ldots, n-1; y_1, \ldots, y_{n-1})$. Of course the martingale property can be verified directly from the independence of the X_n: it amounts to the simple fact that if $g : G^n \to \mathbb{R}$ is bounded and measurable, then

$$\mathbb{E}\left[g(X_1 \ldots, X_n) \,|X_1 \ldots, X_{n-1}\right] = \int_G g(X_1 \ldots, X_{n-1}, y)\, \mathbb{P}(X_n \in dy).$$

We shall conclude this section by quoting some important identities involving the so-called *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 $[M]$, see p.84 below).

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

$$\langle M, M' \rangle := \tfrac{1}{4}\left(\langle M + M' \rangle - \langle M - M' \rangle\right). \tag{4.90}$$

For us, with Theorem 4.6.1 available, we need only find the quadratic characteristics for the stochastic integrals $M^\circ(S)$. So let Q be a probability on $(\mathcal{M}, \mathcal{H})$ with $\overline{\Lambda}^\circ$ *continuous* and let $S, \widetilde{S}$ be predictable fields such that the stochastic integrals $\Lambda_t^\circ(S^2), \Lambda_t^\circ(\widetilde{S}^2)$ are all finite, and *define*

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

(Note that these definitions conform with (4.90)).

Proposition 4.6.2 *Assume that $\overline{\Lambda}^\circ$ is continuous. Assume also that for all n and t,*

$$EN_{\tau_n \wedge t}^\circ\left(S^{*2}\right) < \infty \tag{4.91}$$

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

$$\left(M^\circ(S)\right)^2 - \langle M^\circ(S)\rangle, \quad M^\circ(S)M^\circ\left(\widetilde{S}\right) - \left\langle M^\circ(S), M^\circ\left(\widetilde{S}\right)\right\rangle \tag{4.92}$$

are local Q*-martingales* (τ_n).

If instead of (4.91)

$$EN_t^\circ\left(S^{*2}\right) < \infty \tag{4.93}$$

holds for all t, *then the local martingales in (4.92) are* Q*-martingales.*

The condition (4.91) is satisfied in particular if for all n and t,

$$\sup_{s \le t, y \in E, m \in \mathcal{M}} \left|S^{*y}_{\tau_n \wedge s}(m)\right| < \infty, \tag{4.94}$$

while for (4.93) it suffices that for all t,

$$E\overline{N}_t^\circ < \infty, \qquad \sup_{s \le t, y \in E, m \in \mathcal{M}} \left|S_s^{*y}(m)\right| < \infty.$$

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

$$EM^{\circ 2}_{\tau_1 \wedge t}(S) = E\Lambda^\circ_{\tau_1 \wedge t}\left(S^2\right), \quad EM^\circ_{\tau_1 \wedge t}(S)M^\circ_{\tau_1 \wedge t}(\widetilde{S}) = E\Lambda^\circ_{\tau_1 \wedge t}\left(S\widetilde{S}\right). \tag{4.95}$$

Note that Proposition 4.5.3 corresponds to the special case $S_t^y = 1_A(y)$, $\widetilde{S}_t^y = 1_{A'}(y)$ when $A \cap A' = \emptyset$.

The integrability conditions given in Proposition 4.6.2 are sufficient for the stated martingale properties, but far from necessary.

Exercise 4.6.1 Show (4.95) by direct calculation.

Apart from quadratic characteristics, in the general theory of processes one also operates with the *quadratic variation* process for local martingales M, denoted $[M]$. For continuous M, $[M] \equiv \langle M\rangle$, but for processes with jumps the two processes differ: assume for simplicity that $\overline{\Lambda}^\circ$ is continuous so that as noted in Remark 4.6.2 above, M is continuous on each interval $\left[\tau_n, \tau_{n+1}\right[$, and then define

$$[M]_t := \sum_{0 < s \le t} (\Delta M_s)^2 = \sum_{n=1}^{\overline{N}_t^\circ} \left(\Delta M_{\tau_n}\right)^2.$$

If $M_t = M_0 + M_t^\circ(S)$ as in (4.81), it follows that

$$[M]_t = \sum_{n=1}^{\overline{N}_t^\circ} \left(S^{\eta_n}_{\tau_n}\right)^2 = N_t^\circ\left(S^2\right).$$

Combining this with Theorem 4.5.2 we see that under suitable boundedness conditions on S, if $\overline{\Lambda}^\circ$ is continuous then $[M] - \langle M\rangle$ is a local Q-martingale.

4.7 Itô's formula for MPPs

Let Q be a probability on $(\mathcal{M}, \mathcal{H})$ (or $(W, \mathcal{H})$) and let X be an adapted, piecewise continuous $\mathbb{R}$-valued process. It holds in great generality that X can be written as the sum of a predictable process and a stochastic integral $M^\circ(S)$ with S a predictable field. Uniqueness of this decomposition is achieved when $M^\circ(S)$ is a local Q-martingale; cf. Theorem 4.6.1. The version of Itô's formula in Theorem 4.7.1 below gives the explicit form of the decomposition under mild smoothness conditions. Examples 4.7.3 and 4.7.4 illustrate how the decomposition may be achieved when some of these conditions do not hold.

By Proposition 4.2.1,

$$X_t = f^{\langle t \rangle}_{\xi_{\langle t \rangle}}(t). \tag{4.96}$$

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

$$X^c_t := X_t - \sum_{0<s\leq t} \Delta X_s = X_t - \sum_{n=1}^{\overline{N}^\circ_t} \Delta X_{\tau_n},$$

since X can only have discontinuities at the timepoints τ_n. For the formulation of the next result, recall also that

$$\text{join}\left(\xi_{\langle t- \rangle}, (t, y)\right) = \left(\tau_1, \ldots, \tau_{\langle t- \rangle}, t; \eta_1, \ldots, \eta_{\langle t- \rangle}, y\right).$$

Theorem 4.7.1 (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 provided*

$$\int_{[0,t]\times E} \left|S^y_s\right| \, L^\circ(ds, dy) < \infty$$

Q-a.s. for all t, it holds that

$$X_t = X_0 + U_t + M^\circ_t(S) \quad (t \geq 0), \tag{4.97}$$

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

$$S^y_t = f^{\langle t- \rangle+1}_{\text{join}\left(\xi_{\langle t- \rangle},(t,y)\right)}(t) - f^{\langle t- \rangle}_{\xi_{\langle t- \rangle}}(t), \tag{4.98}$$

and where U defined by

$$U_t = X_t - X_0 - M^\circ_t(S)$$

is continuous and predictable. Furthermore, $U_0 \equiv 0$, $M^\circ_0(S) \equiv 0$ and provided the process $M^\circ(S)$ is a local Q-martingale, this decomposition of X into the sum of its initial value, a predictable process 0 at time 0 and a local Q-martingale is unique up to Q-indistinguishability.

Note. S may also be written $\widetilde{X}_t^y - X_{t-}$, where $\widetilde{X}_t^y$ is the value of X_t obtained by retaining the behaviour of μ° on $[0, t[$ and pretending that a jump with mark y occurs at time t.

Proof. By (4.96), 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,

$$\Delta M_t^\circ(S) = S_t^{\eta_{\langle t \rangle}} \Delta \overline{N}_t^\circ \tag{4.99}$$

while

$$\begin{aligned}\Delta X_t &= \Delta X_t \Delta \overline{N}_t^\circ \\ &= \left(f_{\xi_{\langle t \rangle}}^{\langle t \rangle}(t) - f_{\xi_{\langle t- \rangle}}^{\langle t- \rangle}(t) \right) \Delta \overline{N}_t^\circ \\ &= \left(f_{\text{join}\,(\xi_{\langle t- \rangle}, (t, \eta_{\langle t \rangle}))}^{\langle t- \rangle + 1} - f_{\xi_{\langle t- \rangle}}^{\langle t- \rangle}(t) \right) \Delta \overline{N}_t^\circ.\end{aligned}$$

Thus (4.97) holds with

$$S_t^y = f_{\text{join}\,(\xi_{\langle t- \rangle}, (t, y))}^{\langle t- \rangle + 1} - f_{\xi_{\langle t- \rangle}}^{\langle t- \rangle}(t),$$

which defines a predictable field, and

$$U_t = X_t - X_0 - M_t^\circ(S) = X_t^c - X_0 + \Lambda_t^\circ(S) \tag{4.100}$$

which defines a process that is continuous and adapted, hence predictable.

The uniqueness of the representation (4.97) when $M^\circ(S)$ is a local martingale is immediate from Proposition 4.5.1. □

Remark 4.7.1 There are more general decompositions of adapted processes (see Examples 4.7.3 and 4.7.4), but in general U will not be continuous.

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 $\Delta X_t = S_t^{\eta_{\langle t \rangle}} \Delta \overline{N}_t^\circ$; cf. (4.99), and then use (4.100) to find U.

Example 4.7.1 Let Q be the canonical Poisson process on $(W, \mathcal{H})$ with parameter $\lambda > 0$, (Example 3.1.2). Fix $n_0 \in \mathbb{N}_0$ and define

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

Since

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

we obtain

$$X_t = X_0 + U_t + M_t^\circ(S)$$

with

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

and, since X is a step process with $X_0 \equiv 0$ so that $X^c \equiv 0$, by (4.100),

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

Since $|S| \leq 1$, we find $EN^\circ_t(|S|) \leq EN^\circ_t = E\Lambda^\circ_t = \lambda t < \infty$ so that by Theorem 4.6.1 (iii2) $M^\circ(S)$ is a Q-martingale and thus

$$p_{n_0}(t) := EX_t = 1_{\{0\}}(n_0) + EU_t = 1_{\{0\}}(n_0) + \int_0^t \left(p_{n_0-1}(s) - p_{n_0}(s)\right) \lambda\, ds, \tag{4.101}$$

a formula valid for $n_0 \geq 0$, $t \geq 0$. For $n_0 = 0$ this gives $p'_0(t) = -\lambda p_0(t)$ with $p_0(0) = 1$, i.e., $p_0(t) = e^{-\lambda t}$ (which since $p_0(t) = Q(\tau_1 > t)$ is obvious anyway), and since (4.101) 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 it follows that under Q, N°_t has a Poisson distribution with mean λt,

$$p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (t \geq 0, n \in \mathbb{N}_0).$$

Exercise 4.7.1 Let E be at most countably infinite and let $Q_{|i}$ denote the probability on $(\mathcal{M}, \mathcal{H})$ that makes X° given by $X^\circ_t = \eta_{\langle t\rangle}$ a homogeneous Markov chain with intensity matrix (q_{ij}) and initial state $X^\circ_0 \equiv i$, see Examples 3.3.1 and 4.3.6; in particular the q_{ij} are such that all $Q_{|i}$ exist as stable canonical RCMs. Recall that $\lambda_j = -q_{jj}$ is the rate for the exponential waiting time in j, and that $\pi_{jk} = q_{jk}/\lambda_j$ for $j \neq k$ is the probability that a jump from j leads to k.

Let $j \in \mathbb{E}$, $n \in \mathbb{N}_0$. Use Itô's formula on the step process

$$1_{\left(X^\circ_t=j,\overline{N}^\circ_t=n\right)} \quad (t \geq 0)$$

to show that

$$\begin{aligned} 1_{\left(X^\circ_t=j,\overline{N}^\circ_t=n\right)} &= \delta_{ij}\delta_{0n} \\ &\quad + \int_0^t \lambda_{X^\circ_s}\left(1_{\left(X^\circ_s\neq j,\overline{N}^\circ_s=n-1\right)}\pi_{X^\circ_s,j} - 1_{\left(X^\circ_s=j,\overline{N}^\circ_s=n\right)}\right) ds + M^\circ_t(S), \end{aligned}$$

where S is the predictable field

$$S_t^y = 1_{y=j} 1_{\left(\overline{N}_{t-}^\circ = n-1\right)} - 1_{\left(X_{t-}^\circ = j, \overline{N}_{t-}^\circ = n\right)}.$$

Show next using Theorem 4.6.1(iii) that $M^\circ(S)$ is a $Q_{|i}$ -martingale, and defining

$$p_{ij}^{(n)}(t) = Q_{|i}\left(X_t^\circ = j, \overline{N}_t^\circ = n\right),$$

the *n-step transition probability from i to j*, deduce the formula

$$p_{ij}^{(n)}(t) = \delta_{ij}\delta_{0n} + \int_0^t \left(\sum_{k:k\neq j} p_{ik}^{(n-1)}(s) q_{kj} + p_{ij}^{(n)}(s) q_{jj} \right) ds. \tag{4.102}$$

(Note that for $n = 0$ this gives

$$p_{ij}^{(0)}(t) = \delta_{ij} + q_{jj} \int_0^t p_{ij}^{(0)}(s)\, ds,$$

i.e., $p_{ij}^{(0)}(t) = \delta_{ij} e^{q_{jj} t}$ in agreement with the obvious fact

$$p_{ij}^{(0)}(t) = Q_{|i}\left(X_t^\circ = j, \tau_1 > t\right) = \delta_{ij} Q_{|i}\,(\tau_1 > t)).$$

From (4.102), deduce that the transition probabilities $p_{ij}(t) = Q_{|i}\left(X_t^\circ = j\right)$ satisfy the integral equations

$$p_{ij}(t) = \delta_{ij} + \int_0^t \sum_k p_{ik}(s) q_{kj}\, ds,$$

which by formal differentiation with respect to t yield the *forward Feller–Kolmogorov differential equations*,

$$D_t p_{ij}(t) = \sum_k p_{ik}(t) q_{kj}.$$

The differentiation is trivially justified if e.g., $\sup_{k:k\neq j} q_{kj} < \infty$ for all j, in particular if E is finite, but the differential equations are in fact always valid.

The reference state i singled out above is unimportant: the n-step transition probabilities are the same for all initial states, i.e., for all i', i, j and n and all $s, t \geq 0$ it holds that

$$p_{ij}^{(n)}(t) = Q_{|i'}\left(X_{s+t}^\circ = j, \overline{N}_{s+t}^\circ - \overline{N}_s^\circ = n \,|\mathcal{H}_s\right)$$

on the set $\left(X_s^\circ = i\right)$, cf. Example 4.3.6.

Using Lemma 4.3.3, conditioning on ξ_1, show that for $n \geq 1$,

$$p_{ij}^{(n)}(t) = \int_0^t \sum_{k:k\neq i} q_{ik} e^{-\lambda_i s} p_{kj}^{(n-1)}(t-s)\, ds.$$

From this, deduce that

$$p_{ij}(t) = \delta_{ij} e^{-\lambda_i t} + \int_0^t \sum_{k:k \neq i} q_{ik} e^{-\lambda_i (t-s)} p_{kj}(s)\, ds,$$

and then derive (without any conditions on the transition intensities) the *backward Feller–Kolmogorov differential equations*,

$$D_t p_{ij}(t) = \sum_k q_{ik} p_{kj}(t).$$

Example 4.7.2 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 4.5.3): first, it is easily checked that

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

and so (4.97) holds with

$$S^y_t = \left(2M^\circ_{t-}(A) + 1\right) 1_A(y). \tag{4.103}$$

To show that X is a local martingale we must first show that $U \equiv 0$, where, cf. (4.100),

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

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

$$D_t X^c_t = 2M^\circ_t(A)\left(-\lambda^\circ_t(A)\right) - \lambda^\circ_t(A)$$

where $\Lambda^\circ(A)\,(ds) = \lambda^\circ_s(A)\, ds = \overline{\lambda}^\circ_s \pi^{\langle s \rangle}_{\xi_{\langle s \rangle}}(A)\, ds$. Using (4.103) it now follows that the continuous process U satisfies $D_t U_t = 0$, hence $U \equiv 0$.

Using Theorem 4.6.1 (iii1) it is an easy matter to show that X is a local martingale:

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

and since $E\overline{\Lambda}^\circ_{\tau_n \wedge t} = E\overline{N}^\circ_{\tau_n \wedge t} \leq n < \infty$ it follows that $EN^\circ_{\tau_n \wedge t}(|S|) < \infty$. But the condition in Theorem 4.6.1 (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 4.5.3.

By similar reasoning one may show that $M^\circ(A)M^\circ(A')$ is a local martingale when $A \cap A' = \emptyset$ (Proposition 4.5.3).

A particularly important application of Itô's formula yields the basic Poisson point process structure of homogeneous Poisson random measures. Let Q be such a measure

(Example 3.2.1) so that the compensating measure is $L^\circ = \lambda\ell \otimes \kappa$ and the Markov kernels are given by

$$\overline{P}^{(0)}(t) = e^{-\lambda t}, \qquad \overline{P}^{(n)}_{z_n}(t) = e^{-\lambda(t-t_n)} \quad (n \geq 1, t \geq t_n),$$

$$\pi^{(n)}_{z_n,t}(A) = \kappa(A) \quad (n \geq 0, A \in \mathcal{E})$$

where $\lambda > 0$ and κ is a probability on $(E, \mathcal{E})$.

Proposition 4.7.2 *Let Q be the homogeneous Poisson random measure described above. Then under Q, for any $r \in \mathbb{N}$, and any $A_1, \ldots, A_r \in \mathcal{E}$ mutually disjoint, the counting processes $\left(N^\circ(A_j)\right)_{1 \leq j \leq r}$ are independent, homogeneous Poisson processes in the traditional sense with intensities $\lambda_j := \lambda\kappa(A_j)$ for $j = 1, \ldots, r$.*

Remark 4.7.2 When saying that each $N^\circ(A_j)$ is homogeneous Poisson in the traditional sense, we mean that $N^\circ(A_j)$ has stationary and independent increments that follow Poisson distributions: we shall show that for any $s < t$, the increments $(N^\circ_t(A_j) - N^\circ_s(A_j))_{1 \leq j \leq r}$ 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 $\lambda_j(t-s)$.

Proof. The key step in the proof consists in showing that for r and $A_1, \ldots, A_r$ given with the A_j mutually disjoint, it holds for all $(u_1, \ldots, u_r) \in \mathbb{R}^r$ that

$$M_t := \exp\left(i \sum_{j=1}^r u_j N^\circ_t(A_j) - t \sum_{j=1}^r \lambda_j \left(e^{iu_j} - 1\right)\right) \tag{4.104}$$

is a $\mathbb{C}$-valued Q-martingale. Once this is established, the martingale property $E(M_t \,|\mathcal{H}_s) = M_s$ gives that for $s \leq t$,

$$E\left[\exp\left(i \sum_{j=1}^r u_j \left(N^\circ_t(A_j) - N^\circ_s(A_j)\right)\right) |\mathcal{H}_s\right] = \exp\left((t-s) \sum_{j=1}^r \lambda_j \left(e^{iu_j} - 1\right)\right), \tag{4.105}$$

which is precisely to say (when $(u_1, \ldots, u_r)$ varies) that the joint characteristic function of $\left(N^\circ_t(A_j) - N^\circ_s(A_j)\right)_{1 \leq j \leq r}$ given $\mathcal{H}_s$ is that of r independent Poisson random variables with parameters $\left(\lambda_j(t-s)\right)_{1 \leq j \leq r}$. (Strictly speaking, the identity (4.105) only holds Q-a.s. for any given $\left(u_j\right)$. But then it also holds Q-a.s. simultaneously for all vectors $\left(u_j\right)$ with the $u_j \in \mathbb{Q}$, and since there is a regular conditional distribution of $\left(N^\circ_t(A_j) - N^\circ_s(A_j)\right)_{1 \leq j \leq r}$ given $\mathcal{H}_s$ and therefore also a (necessarily continuous) conditional characteristic function, the latter has indeed been identified by (4.105), at least Q-a.s.). That the $N^\circ\left(A_j\right)$ are independent homogeneous Poisson counting processes with intensities λ_j now follows directly.

We complete the proof by using Itô's formula to show that M given by (4.104) is a martingale. Because

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

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 $D_t U_t = 0$. But clearly, since $N^{\circ}(S)$ is constant between jumps,

$$D_t U_t = D_t M_t + D_t \Lambda_t^{\circ}(S).$$

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

$$D_t M_t = -cM_t$$

while

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

(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 4.6.1 (iii2) if we show that $EN_t^{\circ}\left(|S|\right) < \infty$ for all t. But there are constants c_1, c_2 such that

$$\left|S_t^y\right| \le e^{c_1 t} c_2,$$

so $EN_t^{\circ}\left(|S|\right) \le e^{c_1 t} c_2 E\overline{N}_t^{\circ} = e^{c_1 t} c_2 \lambda t < \infty$, see also the note after Theorem 4.6.1. □

Exercise 4.7.2 With Q the homogeneous Poisson measure from Proposition 4.7.2, let $A_1, \ldots, A_r \in \mathcal{E}$ be mutually disjoint and write

$$\tau_n\left(A_j\right) = \inf\left\{t : N_t^\circ\left(A_j\right) = n\right\}$$

for the time of the nth jump for $N^\circ\left(A_j\right)$. Show that the r sequences $\left(\tau_n\left(A_j\right)\right)_{n\geq 1}$ for $j = 1, \ldots, r$ are stochastically independent and that for each j, the waiting times $\tau_n\left(A_j\right) - \tau_{n-1}\left(A_j\right)$ are iid exponential at rate λ_j. (Please remember that we now know that a sequence of iid exponential waiting times may be identified with a counting process that has stationary independent and Poisson distributed increments)!

In the version of Itô's formula given above, Theorem 4.7.1, 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 4.7.3 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 4.5.3 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

$$M^{\circ 2} - \widetilde{\Lambda} = N^\circ(S) - \Lambda^\circ(S). \tag{4.106}$$

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 4.7.1 and the preceding examples, a matter of simply identifying the jumps of $M^{\circ 2}$ 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 either 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 (4.106), but a simpler expression is available by showing that $\widetilde{\Lambda}$ is differentiable with respect to Λ° and finding $D_{\Lambda^\circ}\widetilde{\Lambda} = D_{\Lambda^\circ}(X + \Lambda^\circ(S))$ according to the methods from 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 $t_K = \frac{k-1}{2^K} < t \leq \frac{k}{2^K} = t'_K$. The task is to compute (for Λ°-a.a. t),

$$\lim_{K\to\infty} \frac{X_{t'_K} - X_{t_K}}{\Lambda_{t'_K}^\circ - \Lambda_{t_K}^\circ}$$

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 Λ°-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

$$\lim_{K\to\infty} \frac{\left(M_{t_K'}^\circ + M_{t_K}^\circ\right)\left(M_{t_K'}^\circ - M_{t_K}^\circ\right)}{\Lambda_{t_k'}^\circ - \Lambda_{t_K}^\circ} = -\left(2M_{t-}^\circ - \Delta\Lambda_t^\circ\right)$$

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

$$D_{\Lambda^\circ} X_t = -\left(2M_{t-}^\circ - \Delta\Lambda_t^\circ\right).$$

Thus

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

and

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

Example 4.7.4 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 field (S^y) such that

$$M_t^\circ(A)M_t^\circ(A') = \widetilde{\Lambda}_t + N_t^\circ(S) - \Lambda_t^\circ(S).$$

From Proposition 4.5.3 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 and A and A' are arbitrary.

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

$$X_t = \left(M_{t-}^\circ(A) - \Delta\Lambda_t^\circ(A)\right)\left(M_{t-}^\circ(A') - \Delta\Lambda_t^\circ(A')\right) - N_{t-}^\circ(S),$$

while if $\Delta\overline{N}_t^\circ = 1$,

$$\begin{aligned} X_t = \left(M_{t-}^\circ(A) + 1_A\left(\eta_{\langle t\rangle}\right) - \Delta\Lambda_t^\circ(A)\right)\left(M_{t-}^\circ(A') + 1_{A'}\left(\eta_{\langle t\rangle}\right) - \Delta\Lambda_t^\circ(A')\right) \\ - N_{t-}^\circ(S) - S_t^{\eta_{\langle t\rangle}}. \end{aligned}$$

Predictability of X forces the two expressions to be identical, hence

$$S_t^y = 1_{A\cap A'}(y) + 1_A(y)\left(M_{t-}^\circ(A') - \Delta\Lambda_t^\circ(A')\right) + 1_{A'}(y)\left(M_{t-}^\circ(A) - \Delta\Lambda_t^\circ(A)\right). \tag{4.107}$$

As in the previous Example 4.7.3, we identify $\widetilde{\Lambda}$ through its derivative $D_{\overline{\Lambda}^\circ}\widetilde{\Lambda} = D_{\overline{\Lambda}^\circ}X + D_{\overline{\Lambda}^\circ}L^\circ(S)$. Here

$$D_{\overline{\Lambda}^\circ}L_t^\circ(S) = \int_E S_t^y\, \pi_t\,(dy)\,, \tag{4.108}$$

where π_t is short for $\pi^{\langle t-\rangle}_{\xi_{\langle t-\rangle},t}$, and by computations along the same lines as those in Example 4.7.3, 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 (4.107), (4.108)

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

so finally,

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

Exercise 4.7.3 Let Q be a probability on $(\mathcal{M},\mathcal{H})$ with compensating measure L°. Let S, S' be predictable fields such that $\Lambda_t^\circ(|S|)$ and $\Lambda_t^\circ(\left|S'\right|) < \infty$ for all t, and find a decomposition of $M^\circ(S)M^\circ(S')$, i.e., show that there is a predictable field $\left(\widetilde{S}\right)$ such that

$$M_t^\circ(S)M_t^\circ(S') = \widetilde{\Lambda}_t + M_t^\circ(\widetilde{S}),$$

where

$$\widetilde{\Lambda}_t = \Lambda_t^\circ\left(SS'\right) - \sum_{0<s\le t}\Delta\Lambda_s^\circ\,(S)\,\Delta\Lambda_s^\circ\left(S'\right).$$

Show in particular that $\widetilde{\Lambda}$ is increasing if $S \equiv S'$.

The result is the easy to guess at generalization of the decomposition in Example 4.7.4.

4.8 Compensators and filtrations

We have so far exclusively discussed compensators and compensating measures for canonical CPs and RCMs, i.e., probabilities on $(W,\mathcal{H})$ and $(\mathcal{M},\mathcal{H})$ respectively. Both concepts make perfect sense for CPs and RCMs 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.

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space, i.e., $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $(\mathcal{F}_t)_{t\geq 0}$ is an increasing family of sub σ-algebras of $\mathcal{F}$. Since the discussion below focuses on right-continuous martingales, we may and shall assume that the filtration is right-continuous, $\mathcal{F}_{t+} = \mathcal{F}_t$ for all $t \geq 0$, cf. Corollary B.0.9.

Let $X = (X_t)_{t\geq 0}$ be a stochastic process with state space $(G, \mathcal{G})$, defined on $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$. Recall that X is measurable, if the map $(t, \omega) \mapsto X_t(\omega)$ from $(\mathbb{R}_0 \times \Omega, \mathcal{B}_0 \otimes \mathcal{F})$ to $(G, \mathcal{G})$ is measurable and that X is adapted to the filtration $(\mathcal{F}_t)$ if it is measurable and each $X_t : (\Omega, \mathcal{F}) \to (G, \mathcal{G})$ is $\mathcal{F}_t$-measurable. Also recall (cf. p.302) that X is predictable if X_0 is $\mathcal{F}_0$-measurable and the map $(t, \omega) \mapsto X_t(\omega)$ from $(\mathbb{R}_+ \times \Omega, \mathcal{B}_+ \otimes \mathcal{F})$ to $(G, \mathcal{G})$ is $\mathcal{P}$-measurable, where $\mathcal{P}$ is the σ-algebra of predictable sets, i.e., the sub σ-algebra of $\mathcal{B}_+ \otimes \mathcal{F}$ generated by the subsets of the form

$$]s, \infty[\times F \quad (s \in \mathbb{R}_0, F \in \mathcal{F}_s).$$

We denote by $(\mathcal{F}_t^X)_{t\geq 0}$ the filtration generated by X, $\mathcal{F}_t^X = \sigma(X_s)_{0\leq s\leq t}$. (For more on concepts and results from the general theory of processes used here and below, see Appendix B).

If N is a counting process, i.e., a $(W, \mathcal{H})$-valued random variable defined on $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$, it is always assumed without further comment that N is $\mathcal{F}_t$-adapted. We write $\mathcal{T} = (T_n)_{n\geq 1} = (\tau_n \circ N)_{n\geq 1}$ for the SPP determined by N, with T_n the time of the nth jump for N, cf. Section 2.2, and $\left(\mathcal{F}_t^N\right)_{t\geq 0}$ for the filtration generated by N, $\mathcal{F}_t^N = \sigma\,(N_s)_{0\leq s\leq t}$ (so that $\mathcal{F}_t^N \subset \mathcal{F}_t$ since N is adapted). Recalling the definition of $\mathcal{H}_t$, it is clear that

$$\mathcal{F}_t^N = N^{-1}(\mathcal{H}_t). \tag{4.109}$$

Similarly, if μ is a random counting measure, i.e., a $(\mathcal{M}, \mathcal{H})$-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, for $A \in \mathcal{E}$ we define the counting process $N(A)$ by $N_t(A) = \mu([0, t] \times A)$ and always assume that μ is $\mathcal{F}_t$-adapted, i.e., that $N_t(A)$ is $\mathcal{F}_t$-measurable for all t and all A. We write $(\mathcal{T}, \mathcal{Y}) = ((T_n), (Y_n))_{n\geq 1} = ((\tau_n \circ \mu), (\eta_n \circ \mu))_{n\geq 1}$ for the MPP determined by μ, and $\left(\mathcal{F}_t^\mu\right)_{t\geq 0}$ for the filtration generated by μ, $\mathcal{F}_t^\mu = \sigma\,(N_s(A))_{0\leq s\leq t, A\in\mathcal{E}}$ (so that $\mathcal{F}_t^\mu \subset \mathcal{F}_t$ for all t). In analogy with (4.109) we have

$$\mathcal{F}_t^\mu = \mu^{-1}(\mathcal{H}_t). \tag{4.110}$$

Note that if N is a CP, then each T_n is an $\mathcal{F}_t$-stopping time. Similarly, if μ is an $\mathcal{F}_t$-adapted RCM, then each T_n is an $\mathcal{F}_t$-stopping time and Y_n is $\mathcal{F}_{T_n}$-measurable.

Suppose now e.g., that μ is an RCM. Write $Q = \mu(\mathbb{P})$ for the distribution of μ and L° for 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 $\left(\mathcal{F}_t^\mu\right)$ and the positive random measure $L^\mu := L^\circ \circ \mu$. Thus e.g., $\Lambda^\mu(A) = \left(\Lambda_t^\mu(A)\right)_{t\geq 0}$ is $\mathcal{F}_t^\mu$-predictable for all $A \in \mathcal{E}$, where $\Lambda_t^\mu(A) = L^\mu([0, t] \times A) = \Lambda_t^\circ(A) \circ \mu$, and also

(i) for all $A \in \mathcal{E}$, $M(A) := N(A) - \Lambda^{\mu}(A)$ is an $\mathcal{F}_t^{\mu}$-local martingale (T_n);
(ii) up to P-indistinguishability, $\Lambda^{\mu}(A)$ is the unique right-continuous and $\mathcal{F}_t^{\mu}$-predictable process $\widetilde{\Lambda}$, 0 at time 0, such that $N(A) - \widetilde{\Lambda}$ is an $\mathcal{F}_t^{\mu}$-local martingale.

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

The proof of (i) above involves nothing more than transformation of integrals: for any n and any $s \le t$, $F \in \mathcal{F}_s^{\mu}$, it must be shown that

$$\int_F M_{T_n \wedge t}(A)\, d\mathbb{P} = \int_F M_{T_n \wedge s}(A)\, d\mathbb{P}.$$

But $M_{T_n \wedge u}(A) = M^{\circ}_{\tau_n \wedge u}(A) \circ \mu$ for any u, and since by (4.110) $F = \mu^{-1}(H)$ for some $H \in \mathcal{H}_s$ this identity is equivalent to the basic local martingale property

$$\int_H M^{\circ}_{\tau_n \wedge t}(A)\, dQ = \int_H M^{\circ}_{\tau_n \wedge s}(A)\, dQ.$$

We shall now discuss compensators and compensating measures for CPs and RCMs defined on general filtered probability spaces.

Let N be an adapted counting process defined on $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$. We shall call Λ the *$\mathcal{F}_t$-compensator for N* if $\Lambda_0 = 0$ $\mathbb{P}$-a.s., Λ is increasing, right-continuous, $\mathcal{F}_t$-predictable and satisfies that $M := N - \Lambda$ is an $\mathcal{F}_t$-local martingale. This compensator exists and is unique by the Doob–Meyer decomposition theorem, see Theorem B.0.10. 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 $Q = N(\mathbb{P})$ of N.

Similarly, if μ is an adapted random counting measure on $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$, the *$\mathcal{F}_t$-compensating measure for μ* is the positive random measure L with $L(\{0\} \times E) = 0$ $\mathbb{P}$-a.s. such that for all $A \in \mathcal{E}$, $\Lambda(A)$, where $\Lambda_t(A) = L([0,t] \times A)$, defines a right-continuous $\mathcal{F}_t^{\mu}$-predictable process, necessarily increasing, such that $M(A) := N(A) - \Lambda(A)$ is an $\mathcal{F}_t$-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 $Q = \mu(\mathbb{P})$. (About the existence of L: by the Doob–Meyer decomposition theorem each compensator $\Lambda(A)$ exists and is unique up to indistinguishability. Some care is needed to fit the compensators together to arrive at the compensating measure, but with $(E, \mathcal{E})$ a Borel space, L always exists).

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 an 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 $\mathbb{P}$). 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

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

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

Theorem 4.8.1 *Let μ be an RCM on $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ with $\mathcal{F}_t$-compensating measure L. If L is $\mathcal{F}_t^\mu$-predictable, then $L = L^\circ \circ \mu$ up to $\mathbb{P}$-indistinguishability, where L° is the compensating measure for the distribution $Q = \mu(\mathbb{P})$ of μ.*

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

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

$$\int_F \left(N_{\widetilde{T}_n \wedge T_k \wedge t}(A) - N_{\widetilde{T}_n \wedge T_k \wedge s}(A) \right) d\mathbb{P} = \int_F \left(\Lambda_{\widetilde{T}_n \wedge T_k \wedge t}(A) - \Lambda_{\widetilde{T}_n \wedge T_k \wedge s}(A) \right) d\mathbb{P}, \tag{4.111}$$

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

$$\int_F \left(N_{T_k\wedge t}(A) - N_{T_k\wedge s}(A)\right) d\mathbb{P} = \int_F \left(\Lambda_{T_k\wedge t}(A) - \Lambda_{T_k\wedge s}(A)\right) d\mathbb{P}. \qquad (4.112)$$

Since $N_{T_k\wedge t}(A) \le k$, 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 (4.112) and it is then clear that $M^{T_k}(A)$ is an $\mathcal{F}_t$-martingale for all k. □

Using Theorem 4.8.1 and Example 3.2.1 we immediately get

Corollary 4.8.2 *Let μ be an RCM with a deterministic (non-random) $\mathcal{F}_t$-compensating measure $L = \lambda\ell \otimes \kappa$, where $\lambda > 0$ is a constant and κ is a probability on $(E, \mathcal{E})$. Then μ is a homogeneous Poisson random measure with intensity measure $\rho = \lambda\kappa$.*

Example 4.8.1 *Deterministic time change.* Let Q be a probability on $(\mathcal{M}, \mathcal{H})$ (or $(W, \mathcal{H})$) and let $\Phi : \mathbb{R}_0 \to \mathbb{R}_0$ be a continuous and increasing (not necessarily strictly increasing) function with $\Phi(0) = 0$. Defining $\overline{N} = \left(\overline{N}_u\right)_{u\ge 0}$ by $\overline{N}_u = \overline{N}^\circ_{\Phi(u)}$, it follows from the continuity of Φ that $\overline{N}$ is a counting process. With $T_n := \inf\left\{u : \overline{N}_u = n\right\}$, by looking closely one finds that $T_n = \inf\{u : \Phi(u) = \tau_n\}$, in particular $\Phi(T_n) = \tau_n$ whenever $T_n < \infty$. Letting $Y_n := \eta_n$, consider the RCM $\mu := \sum_{T_n<\infty} \varepsilon_{(T_n, Y_n)}$, the RCM obtained from μ° through the deterministic time change Φ.

For any $A \in \mathcal{E}$, it is clear that $N_u(A) := \mu([0, u] \times A) = N^\circ_{\Phi(u)}(A)$, hence it follows that $\mathcal{F}^\mu_u \subset \mathcal{H}_{\Phi(u)}$ for all u. We now claim that the $\mathcal{F}^\mu_u$-compensator for $N(A)$ is the process $\Lambda(A) := \left(\Lambda^\circ_{\Phi(u)}(A)\right)_{u\ge 0}$: we have

$$\begin{aligned}\Lambda^\circ_{\Phi(u)}(A) &= \sum_{n=0}^{\overline{N}^\circ_{\Phi(u)}} \int_{]\tau_k, \tau_{k+1}\wedge\Phi(u)]} \nu^{(k)}_{\xi_k}(ds)\, \pi^{(k)}_{\xi_k, s}(A) \\ &= \sum_{n=0}^{\overline{N}_u} \int_{]\Phi(T_k), \Phi(T_{k+1})\wedge\Phi(u)]} \nu^{(k)}_{Z^*_k}(ds)\, \pi^{(k)}_{Z^*_k, s}(A)\end{aligned}$$

where $Z^*_k = (\Phi(T_1), \ldots, \Phi(T_k); Y_1, \ldots, Y_k)$, which is enough to show that $\Lambda(A)$ is $\mathcal{F}^\mu_u$-predictable. It remains to argue that $M(A) := N(A) - \Lambda(A)$ is an $\mathcal{F}^\mu_u$-local martingale, but for $u < v$ and $H \in \mathcal{F}^\mu_u \subset \mathcal{H}_{\Phi(u)}$,

$$\begin{aligned}\int_H M_{v\wedge T_n}(A)\, dQ &= \int_H M^\circ_{\Phi(v)\wedge\Phi(T_n)}(A)\, dQ \\ &= \int_H M^\circ_{\Phi(v)\wedge\tau_n}(A)\, dQ,\end{aligned}$$

and since $\left(M^\circ_t(A)\right)_{t\ge 0}$ is a Q-local martingale (τ_n), v can be replaced by u and the assertion follows.

We have thus determined the $\mathcal{F}^\mu_t$-compensating measure for μ and hence, according to Theorem 4.8.1, determined the distribution $\widetilde{Q}$ of the time changed RCM μ. It is seen in particular that if $\widetilde{P}^{(n)}_{\widetilde{z}_n}$ and $\widetilde{\pi}^{(n)}_{\widetilde{z}_n,u}$ denote the Markov kernels generating $\widetilde{Q}$, where $\widetilde{z}_n = (u_1, \ldots, u_n; y_1, \ldots, y_n)$, then $\widetilde{P}^{(n)}_{\widetilde{z}_n}$ has hazard measure

$$\widetilde{\nu}^{(n)}_{\widetilde{z}_n}\left(\left]u_n, u\right]\right) = \nu^{(n)}_{z^*_n}\left(\left]\Phi(u_n), \Phi(u)\right]\right),$$

writing $z^*_n = (\Phi(u_1), \ldots, \Phi(u_n); y_1, \ldots, y_n)$, while

$$\widetilde{\pi}^{(n)}_{\widetilde{z}_n,u}(A) = \pi^{(n)}_{z^*_n,\Phi(u)}(A).$$

Example 4.8.2 As a special case of the preceding example, consider the probability Q on $(W, \mathcal{H})$ that makes N° a homogeneous Markov chain with intensity process $\lambda^\circ_s = a_{\langle s-\rangle}$ with the a_n constants ≥ 0 such that $\sum a_n^{-1} = \infty$, see Example 3.1.4. Consider the time change $\Phi(u) = \int_0^u \alpha(s)\, ds$ where $\alpha \geq 0$ is measurable with all the integrals finite and define $N_u = N^\circ_{\Phi(u)}$. By the previous example, the $\mathcal{F}^N_u$-compensator for N is

$$\begin{aligned}
\Lambda^\circ_{\Phi(u)} &= \int_0^{\Phi(u)} a_{\langle s-\rangle}\, ds \\
&= \sum_{n=0}^{N^\circ_{\Phi(u)}} a_n \left(\tau_{n+1} \wedge \Phi(u) - \tau_n\right) \\
&= \sum_{n=0}^{N_u} a_n \left(\Phi\left(T_{n+1} \wedge u\right) - \Phi\left(T_n\right)\right) \\
&= \int_0^u \alpha(v) a_{N_{v-}}\, dv.
\end{aligned}$$

It follows that the distribution $\widetilde{Q} = N(Q)$ of N has intensity process $\widetilde{\lambda}^\circ_u = \alpha(u) a_{\langle u-\rangle}$.

Example 4.8.3 Let N be a counting process on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $U \geq 0$ be an $\mathcal{F}$-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 $P^{(n)}_{z_n}$ generating the distribution of N, one finds that the $\mathcal{F}^N_t$-compensator is $\Lambda^N_t = \int_0^t \lambda^N_s\, ds$ with $\mathcal{F}^N_t$-predictable intensity process

$$\lambda^N_t = \frac{\int_{\mathbb{R}_0} \mathrm{P}(d\lambda)\, e^{-\lambda t} \lambda^{N_{t-}+1}}{\int_{\mathbb{R}_0} \mathrm{P}(d\lambda)\, e^{-\lambda t} \lambda^{N_{t-}}} \tag{4.113}$$

with P the distribution of U. This not-so-lovely expression for Λ^N of course serves to describe the not-so-lovely distribution of N. By contrast, defining $\mathcal{F}_t = \sigma\left(U, \mathcal{F}^N_t\right)$ (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 at all about the distribution of U.

The Cox process is a simple example of a doubly stochastic Poisson process, see Example 6.2.1.

Exercise 4.8.1 In Example 4.8.3, show that λ^N is given by the expression (4.113). (Hint: somewhat informally, for all $n \in \mathbb{N}_0$ and $0 < t_1 < \cdots < t_n \leq t$, calculate the numerator and denominator in

$$\mathbb{P}\left(T_{n+1} > t \,|T_1 = t_1, \ldots, T_n = t_n\right) = \frac{\mathbb{P}\left(T_1 \in dt_1, \ldots, T_n \in dt_n, T_{n+1} > t\right)}{\mathbb{P}\left(T_1 \in dt_1, \ldots, T_n \in dt_n\right)}$$

by conditioning on U).

Even though there is no 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 an 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 σ-algebras, increasing with k, and consider the filtration $(\mathcal{F}_t)$ given by

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

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

We shall conclude this section with a discussion of how compensators in general change with filtrations.

Consider a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ and suppose $\lambda = (\lambda_t)_{t \geq 0}$ is an $\mathcal{F}_t$-adapted and non-negative process. Define $\Lambda_t = \int_0^t \lambda_s \, ds$ and assume that $\mathbb{E}\Lambda_t < \infty$ for all t. Let $(\mathcal{G}_t)$ be a sub-filtration of $(\mathcal{F}_t)$, i.e., a filtration such that $\mathcal{G}_t \subset \mathcal{F}_t$ for all t. Now consider for each t a version $\widetilde{\lambda}_t$ of $\mathbb{E}\left[\lambda_t \,|\mathcal{G}_t\right]$ and assume that the conditional expectations combine to define a measurable, hence $\mathcal{G}_t$-adapted, process $\widetilde{\lambda} \geq 0$. Finally, define $\widetilde{\Lambda}_t = \int_0^t \widetilde{\lambda}_s \, ds$.

Lemma 4.8.3 *For all* $0 \leq s \leq t$, $G \in \mathcal{G}_s$, *it holds that*

$$\int_G (\Lambda_t - \Lambda_s) \, d\mathbb{P} = \int_G \left(\widetilde{\Lambda}_t - \widetilde{\Lambda}_s\right) d\mathbb{P}. \tag{4.114}$$

Proof. By calculation,

$$\begin{aligned}
\int_G \left(\widetilde{\Lambda}_t - \widetilde{\Lambda}_s\right) d\mathbb{P} &= \int_G d\mathbb{P} \int_s^t du \, \widetilde{\lambda}_u \\
&= \int_s^t du \int_G d\mathbb{P} \, \widetilde{\lambda}_u \\
&= \int_s^t du \int_G d\mathbb{P} \, \lambda_u \\
&= \int_G (\Lambda_t - \Lambda_s) \, d\mathbb{P}.
\end{aligned}$$

□

Based on this lemma it is easy to prove

Proposition 4.8.4 *Let U be an $\mathcal{F}_t$-adapted process with $\mathbb{E}\,|U_t| < \infty$ for all t. Suppose that $M := U - \Lambda$ is an $\mathcal{F}_t$-adapted martingale, where $\Lambda_t = \int_0^t \lambda_s\, ds$ with $\lambda \geq 0$ $\mathcal{F}_t$-adapted and $\mathbb{E}\Lambda_t < \infty$ for all t. Let $(\mathcal{G}_t)$ be a sub-filtration of $(\mathcal{F}_t)$ and assume that there are $\mathcal{G}_t$-adapted processes $\widetilde{U}$ and $\widetilde{\lambda}$ such that $\widetilde{U}_t = \mathbb{E}\,[U_t\,|\mathcal{G}_t\,]$ and $\widetilde{\lambda}_t = \mathbb{E}\,[\lambda_t\,|\mathcal{G}_t\,]$ a.s. for every t.*

Then the process $\widetilde{M} := \widetilde{U} - \widetilde{\Lambda}$ is a $\mathcal{G}_t$-martingale, where $\widetilde{\Lambda}_t = \int_0^t \widetilde{\lambda}_s\, ds$.

Proof. For $0 \leq s \leq t$, $G \in \mathcal{G}_s$,

$$\begin{aligned}\int_G \widetilde{M}_t\, d\mathbb{P} &= \int_G \left(\widetilde{U}_t - \widetilde{\Lambda}_t\right)\, d\mathbb{P} \\ &= \int_G (U_t - \Lambda_t)\, d\mathbb{P} + \int_G \left(\Lambda_t - \widetilde{\Lambda}_t\right)\, d\mathbb{P}.\end{aligned}$$

Here one can replace t by s in the first term because M is an $\mathcal{F}_t$-martingale, and in the second because of (4.114). □

Now consider a $\mathcal{G}_t$-adapted counting process N with $\mathbb{E}N_t < \infty$ for all t, and let Λ be the $\mathcal{F}_t$-compensator for N, assumed to be of the form $\Lambda_t = \int_0^t \lambda_s\, ds$, as in Proposition 4.8.4. Also define $\lambda, \widetilde{\lambda}, \widetilde{\Lambda}$ as it was done there. Since $\mathbb{E}\,[N_t\,|\mathcal{G}_t\,] = N_t$ we immediately get

Corollary 4.8.5 *With $(\mathcal{G}_t)$ a sub-filtration of $(\mathcal{F}_t)$, if N is $\mathcal{G}_t$-adapted with $\mathcal{F}_t$-compensator Λ, and $N - \Lambda$ is a true $\mathcal{F}_t$-martingale, then the $\mathcal{G}_t$-compensator for N is $\widetilde{\Lambda}$ and $N - \widetilde{\Lambda}$ is a true $\mathcal{G}_t$-martingale.*

One of course calls λ the *$\mathcal{F}_t$-intensity process* for N, and $\widetilde{\lambda}$ is then the $\mathcal{G}_t$-intensity. The relationship

$$\widetilde{\lambda}_t = \mathbb{E}\,[\lambda_t\,|\mathcal{G}_t\,]$$

between the two intensities is known in the literature as the *innovation theorem.*

We conclude with another interesting consequence of Lemma 4.8.3. Here we say that a well behaved (e.g., cadlag) $\mathbb{R}$-valued process U is of *finite variation on finite intervals* if for all t,

$$\sup \sum_{k=1}^{n} \left|U_{t_k} - U_{t_{k-1}}\right| < \infty \quad \text{a.s.},$$

where the sup extends over all $n \in \mathbb{N}$ and all subdivisions $0 = t_0 < t_1 < \cdots < t_n = t$ of $[0, t]$. The condition simply stipulates that almost all sample paths of the process U must be of finite variation on any interval $[0, t]$.

Corollary 4.8.6 *Let $\Lambda, \widetilde{\Lambda}$ be as in Lemma 4.8.3. Then the process defined by $\mathbb{E}\,[\Lambda_t\,|\mathcal{G}_t\,] - \widetilde{\Lambda}_t$ for all t, is a $\mathcal{G}_t$-martingale. In particular, if a.s. $t \mapsto \mathbb{E}\,[\Lambda_t\,|\mathcal{G}_t\,]$ is continuous and of finite variation on finite intervals, then a.s. $\mathbb{E}\,[\Lambda_t\,|\mathcal{G}_t\,] = \widetilde{\Lambda}_t$ simultaneously for all t.*

Proof. The first assertion follows directly from (4.114), rearranging the terms and using e.g., that $\int_G \Lambda_t \, d\mathbb{P} = \int_G \mathbb{E}[\Lambda_t \,|\mathcal{G}_t] \, d\mathbb{P}$ when $G \in \mathcal{G}_s$. For the second assertion one uses the well-known result from process theory which states that any continuous martingale of finite variation on finite intervals is constant. Since the martingale we consider here has value 0 at time 0, the constant is 0. $\square$

It should be noted that even though Λ is continuous, $t \mapsto \mathbb{E}[\Lambda_t \,|\mathcal{G}_t]$ need not be: take $\mathcal{G}_t = \{\Omega, \emptyset\}$ for $t < 1$, $= \mathcal{F}_t$ for $t \geq 1$ to get easy examples. Quite possibly also, even though Λ is increasing, $t \mapsto \mathbb{E}[\Lambda_t \,|\mathcal{G}_t]$ may not have finite variation.

5

Likelihood Processes

In this chapter the structure of the likelihood process (process of Radon–Nikodym derivatives) is derived when considering two probability measures on the canonical spaces W and $\mathcal{M}$, assuming that one of the probabilities is locally absolutely continuous with respect to the other. Also, it is shown how to change measure using martingales or just local martingales.

It is important to understand the contents of Theorem 5.1.1, but not necessary to read the proof in detail. Also, Section 5.2 may be omitted on a first reading.

References. For likelihood processes for MPPs and RCMs, see Boel et al [12] and Jacod [60]. For general likelihood processes, see e.g., Jacod and Shiryaev [65], Section III.5.

5.1 The structure of the likelihood

In this section 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 $[0, t]$.

Recall that if $\widetilde{\mathbb{P}}, \mathbb{P}$ are probabilities on a measurable space $(\Omega, \mathcal{F})$, $\widetilde{\mathbb{P}}$ is *absolutely continuous* with respect to $\mathbb{P}$ (written $\widetilde{\mathbb{P}} \ll \mathbb{P}$) if $\widetilde{\mathbb{P}}(F) = 0$ whenever $\mathbb{P}(F) = 0$, and that in that case

$$\widetilde{\mathbb{P}}(F) = \int_F \frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}}\, d\mathbb{P} \quad (F \in \mathcal{F})$$

with $d\widetilde{\mathbb{P}}/d\mathbb{P}$ the *Radon–Nikodym* derivative of $\widetilde{\mathbb{P}}$ with respect to $\mathbb{P}$. Recall also that $\widetilde{\mathbb{P}}$ and $\mathbb{P}$ are *singular* (written $\widetilde{\mathbb{P}} \perp \mathbb{P}$) if there exists $F \in \mathcal{F}$ with $\mathbb{P}(F) = 1$ and $\widetilde{\mathbb{P}}(F) = 0$.

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 5.1.1 $\tilde{Q}$ *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 $\widetilde{Q} \ll_{\text{loc}} Q$ and define the *likelihood process* $\mathcal{L} = (\mathcal{L}_t)_{t\geq 0}$ by

$$\mathcal{L}_t := \frac{d\widetilde{Q}_t}{dQ_t}. \tag{5.1}$$

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

$$\widetilde{Q}(H) = \widetilde{Q}_t(H) = \int_H \mathcal{L}_t \, dQ_t = \int_H \mathcal{L}_t \, dQ \quad (t \in \mathbb{R}_0, H \in \mathcal{H}_t). \tag{5.2}$$

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

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

If $s < t$ and $H \in \mathcal{H}_s \subset \mathcal{H}_t$, it follows from (5.2) that

$$\int_H \mathcal{L}_s \, dQ = \int_H \mathcal{L}_t \, dQ;$$

in other words, $\mathcal{L}$ *is a* Q*-martingale* which, since the filtration $(\mathcal{H}_t)$ is right-continuous, has a cadlag version, (Theorem B.0.8 in Appendix B). In the sequel we shall always assume $\mathcal{L}$ to be cadlag.

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

$$\widetilde{Q} = \mathcal{L}_\infty \cdot Q + Q^\perp$$

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

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

(i) $\widetilde{\mathrm{P}}_{\mathbb{R}_+} \ll \mathrm{P}_{\mathbb{R}_+}$, the subscript denoting restriction to $\mathbb{R}_+$ (from $\overline{\mathbb{R}}_+$);
(ii) $\mathrm{P}(]t, \infty]) > 0$ whenever $\widetilde{\mathrm{P}}(]t, \infty]) > 0$.

Note that if $t^\dagger$, $\widetilde{t}^\dagger$ are the termination points for P, $\widetilde{\mathrm{P}}$ (see p.34) the last condition is equivalent to the condition $\widetilde{t}^\dagger \leq t^\dagger$. Note also that $\widetilde{\mathrm{P}} \ll_{\text{loc}} \mathrm{P}$ if $\widetilde{\mathrm{P}} \ll \mathrm{P}$ (on all of $\overline{\mathbb{R}}_+$) while it is possible to have $\widetilde{\mathrm{P}}_{\mathbb{R}_+} \ll \mathrm{P}_{\mathbb{R}_+}$ without $\widetilde{\mathrm{P}} \ll_{\text{loc}} \mathrm{P}$: if $\widetilde{\mathrm{P}} = \varepsilon_\infty$, $\widetilde{\mathrm{P}}_{\mathbb{R}_+} \ll \mathrm{P}_{\mathbb{R}_+}$

for any P since $\widetilde{\mathrm{P}}_{\mathbb{R}_+}$ is the null measure, but $\widetilde{\mathrm{P}} \ll_{\mathrm{loc}} \mathrm{P}$ iff $\mathrm{P}\left(]t,\infty]\right) > 0$ for all $t \in \mathbb{R}_0$. The same example also shows that it is possible to have $\widetilde{\mathrm{P}} \ll_{\mathrm{loc}} \mathrm{P}$ without $\widetilde{\mathrm{P}} \ll \mathrm{P}$: $\varepsilon_\infty \ll \mathrm{P}$ iff $\mathrm{P}(\{\infty\}) > 0$, a condition not implied by the condition for $\varepsilon_\infty \ll_{\mathrm{loc}} \mathrm{P}$.

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

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

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 5.1.1 is formulated in as general a version as possible, allowing for certain exceptional sets denoted by the superscript $^{\mathrm{null}}$. A simpler version of the theorem is obtained by ignoring these exceptional sets, i.e., assuming that they are empty.

Theorem 5.1.1 (a) *Let $Q, \widetilde{Q}$ be probabilities on $(W, \mathcal{H})$. In order that $\widetilde{Q} \ll_{\mathrm{loc}} Q$ it is sufficient that $\widetilde{P}^{(0)} \ll_{\mathrm{loc}} P^{(0)}$ and that for every $n \in \mathbb{N}$ there exists an exceptional set $B_n^{\mathrm{null}} \in \mathcal{B}_+^n$ with $Q\left(\xi_n \in B_n^{\mathrm{null}}\right) = 0$ such that $\widetilde{P}^{(n)}_{z_n} \ll_{\mathrm{loc}} P^{(n)}_{z_n}$ for all $z_n = (t_1, \ldots, t_n) \notin B_n^{\mathrm{null}}$ with $0 < t_1 < \cdots < t_n < \infty$.*

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

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

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

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

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

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

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

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

$$\frac{d\widetilde{\mathrm{R}}_n}{d\mathrm{R}_n}(z_n) = \prod_{k=1}^{n} \left(\frac{d\widetilde{P}^{(k-1)}_{z_{k-1}}}{dP^{(k-1)}_{z_{k-1}}}(t_k) \frac{d\widetilde{\pi}^{(k-1)}_{z_{k-1},t_k}}{d\pi^{(k-1)}_{z_{k-1},t_k}}(y_k) \right) \tag{5.3}$$

for R_n-a.a. $z_n \in \mathbb{R}^n_+ \times E^n$. (A little care is needed here: (5.3) is automatic if $\widetilde{P}^{(n)}_{z_n,\mathbb{R}_+} \ll P^{(n)}_{z_n,\mathbb{R}_+}$ and $\widetilde{\pi}^{(n)}_{z_n,t} \ll \pi^{(n)}_{z_n,t}$ for *all* z_n and t. With exceptional sets appearing, one may proceed by induction: by assumption $\widetilde{P}^{(0)}_{\mathbb{R}_+} \ll P^{(0)}_{\mathbb{R}_+}$, hence if $\mathrm{R}_1(C_1) = 0$, where $C_1 \in \mathcal{B}_+ \otimes \mathcal{E}$,

$$\begin{aligned}
\widetilde{\mathrm{R}}_1(C_1) &= \int_{\mathbb{R}_+} \widetilde{P}^{(0)}(dt) \int_E \widetilde{\pi}^{(0)}_t(dy)\, 1_{C_1}(t, y) \\
&= \int_{\mathbb{R}_+} P^{(0)}(dt) \int_E \widetilde{\pi}^{(0)}_t(dy)\, 1_{C_1}(t, y) \frac{d\widetilde{P}^{(0)}}{dP^{(0)}}(t) \\
&= \int_{\mathbb{R}_+} P^{(0)}(dt) \int_E \pi^{(0)}_t(dy)\, 1_{C_1}(t, y) \frac{d\widetilde{P}^{(0)}}{dP^{(0)}}(t) \frac{d\widetilde{\pi}^{(0)}_t}{d\pi^{(0)}_t}(y) \\
&= 0,
\end{aligned}$$

where for the second but last equality it is used exactly that for $P^{(0)}$-a.a t we have $\widetilde{\pi}^{(0)}_t \ll \pi^{(0)}_t$. Thus $\widetilde{\mathrm{R}}_1 \ll \mathrm{R}_1$ and similar reasoning yields that if $\widetilde{\mathrm{R}}_n \ll \mathrm{R}_n$, also $\widetilde{\mathrm{R}}_{n+1} \ll \mathrm{R}_{n+1}$).

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

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

Recalling the representation of sets in $\mathcal{H}_t$, Proposition 4.2.1 (bi), and writing $H = \bigcup_{n=0}^{\infty} H \cap \left(\overline{N}^\circ_t = n\right)$, this follows if we show that

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

where $H \cap \left(\overline{N}^\circ_t = n\right) = (\xi_n \in C_n, \tau_{n+1} > t)$ with $C_n \subset]0, t]^n \times E^n$. But

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

Define $H_n = (\overline{\widetilde{P}}{}^{(n)}_{\xi_n}(t) > 0)$. Then the last integral above is the same as the integral over the set $(\xi_n \in C_n) \cap H_n$; 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, then 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

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

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 $\left(\xi_n \in C_n, \tau_{n+1} > t, \overline{P}^{(n)}_{\xi_n}(t) > 0\right)$. Finally, because

$$Q\left(\tau_{n+1} > t, \overline{P}^{(n)}_{\xi_n}(t) = 0\right) = E_Q\left[\overline{P}^{(n)}_{\xi_n}(t); \overline{P}^{(n)}_{\xi_n}(t) = 0\right] = 0,$$

we may as well integrate over $(\xi_n \in C_n, \tau_{n+1} > t) = H \cap \left(\overline{N}^{\circ}_t = n\right)$, and we have arrived at the identity

$$\widetilde{Q}\left(H \cap \left(\overline{N}^{\circ}_t = n\right)\right) = \int_{H \cap \left(\overline{N}^{\circ}_t = n\right)} \frac{\overline{\widetilde{P}}{}^{(n)}_{\xi_n}(t)}{\overline{P}^{(n)}_{\xi_n}(t)} \frac{d\widetilde{\mathrm{R}}_n}{d\mathrm{R}_n}(\xi_n)\, dQ.$$

Using (5.3), the assertion of the theorem follows immediately. □

Remark 5.1.1 It may be shown that the sufficient conditions for $\widetilde{Q} \ll_{\text{loc}} Q$ as given in (a), (b) are in fact also necessary. The perhaps most peculiar condition, viz. that $\widetilde{P}^{(n)}_{z_n} \ll_{\text{loc}} P^{(n)}_{z_n}$ rather than just $\widetilde{P}^{(n)}_{z_n,\mathbb{R}_+} \ll P^{(n)}_{z_n,\mathbb{R}_+}$ 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, say, 1. Then $\widetilde{P}^{(0)}_{\mathbb{R}_+} \ll P^{(0)}_{\mathbb{R}_+}$ trivially since $\widetilde{P}^{(0)}_{\mathbb{R}_+}$ is the null measure, $\widetilde{P}^{(0)} \ll_{\text{loc}} P^{(0)}$ does not hold, and for $t \geq 1$ it does not hold either that $\widetilde{Q}_t \ll Q_t$ since $Q_t(\tau_1 \leq t) = 1$, $\widetilde{Q}_t(\tau_1 \leq t) = 0$, i.e., $\widetilde{Q}_t \perp Q_t$ for $t \geq 1$. (See also Example 5.1.1 below).

In general it is possible to express the formula for $\mathcal{L}_t$ 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 $\widetilde{\Lambda}^{\circ}$, Λ° under $\widetilde{Q}$, and Q have predictable intensities,

$$\widetilde{\Lambda}^{\circ}_t = \int_0^t \widetilde{\lambda}^{\circ}_s\, ds, \quad \Lambda^{\circ}_t = \int_0^t \lambda^{\circ}_s\, ds$$

with

$$\widetilde{\lambda}_s^\circ = \widetilde{u}_{\xi_{\langle s-\rangle}}^{\langle s-\rangle}(s), \quad \lambda_s^\circ = u_{\xi_{\langle s-\rangle}}^{\langle s-\rangle}(s)$$

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 4.4.1 (a). If $\widetilde{Q} \ll_{\text{loc}} Q$ (e.g., if all $u_{z_n}^{(n)} > 0$ with $\int_{t_n}^t u_{z_n}^{(n)}(s)\, ds < \infty$ for $t > t_n$) we have

$$\frac{d\widetilde{P}_{\xi_{k-1}}^{(k-1)}}{dP_{\xi_{k-1}}^{(k-1)}}(\tau_k) = \frac{\widetilde{u}_{\xi_{k-1}}^{(k-1)}(\tau_k)\exp\left(-\int_{\tau_{k-1}}^{\tau_k}\widetilde{u}_{\xi_{k-1}}^{(k-1)}(s)\,ds\right)}{u_{\xi_{k-1}}^{(k-1)}(\tau_k)\exp\left(-\int_{\tau_{k-1}}^{\tau_k}u_{\xi_{k-1}}^{(k-1)}(s)\,ds\right)}, \tag{5.4}$$

$$\frac{\overline{\widetilde{P}}_{\xi_{\langle t\rangle}}^{\langle t\rangle}(t)}{\overline{P}_{\xi_{\langle t\rangle}}^{\langle t\rangle}(t)} = \exp\left(-\int_{\tau_{\langle t\rangle}}^t \left(\widetilde{u}_{\xi_{\langle s-\rangle}}^{\langle s-\rangle}(s) - u_{\xi_{\langle s-\rangle}}^{\langle s-\rangle}(s)\right) ds\right). \tag{5.5}$$

It follows that

$$\mathcal{L}_t = \exp\left(-\widetilde{\Lambda}_t^\circ + \Lambda_t^\circ\right)\prod_{n=1}^{N_t^\circ}\frac{\widetilde{\lambda}_{\tau_n}^\circ}{\lambda_{\tau_n}^\circ}. \tag{5.6}$$

Similarly, if $\widetilde{Q}$, Q have compensating measures $\widetilde{L}^\circ$, L° with κ-intensities (Proposition 4.4.1 (bii), the same κ for $\widetilde{Q}$ and Q),

$$\widetilde{\lambda}_t^{\circ y} = \widetilde{u}_{\xi_{\langle t-\rangle}}^{\langle t-\rangle}(t)\,\widetilde{p}_{\xi_{\langle t-\rangle},t}^{\langle t-\rangle}(y), \quad \lambda_t^{\circ y} = u_{\xi_{\langle t-\rangle}}^{\langle t-\rangle}(t)\,p_{\xi_{\langle t-\rangle},t}^{\langle t-\rangle}(y)$$

we still have the analogues of (5.4) and (5.5), and in addition

$$\frac{d\widetilde{\pi}_{\xi_{k-1},\tau_k}^{(k-1)}}{d\pi_{\xi_{k-1},\tau_k}^{(k-1)}}(\eta_k) = \frac{\widetilde{p}_{\xi_{k-1},\tau_k}^{(k-1)}(\eta_k)}{p_{\xi_{k-1},\tau_k}^{(k-1)}(\eta_k)}.$$

It follows that

$$\mathcal{L}_t = \exp\left(-\overline{\widetilde{\Lambda}}_t^\circ + \overline{\Lambda}_t^\circ\right)\prod_{n=1}^{\overline{N}_t^\circ}\frac{\widetilde{\lambda}_{\tau_n}^{\circ\eta_n}}{\lambda_{\tau_n}^{\circ\eta_n}}. \tag{5.7}$$

The derivation of (5.6) and (5.7) 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 4.4.2 would drastically change the expressions and yield non-sensical results!

We quote a particular case of (5.7), important for statistical applications. Here $\widetilde{Q}$ is the important measure, while the particular Q displayed serves only as a convenient reference. Recall Example 3.2.1.

Corollary 5.1.2 *On* $(\mathcal{M},\mathcal{H})$, *let* Q *be the homogeneous Poisson random measure with intensity measure* $\rho = \lambda\kappa$, *where* $\lambda > 0$ *is a constant and* κ *is a probability measure on* $(E,\mathcal{E})$. *If* $\widetilde{Q}$ *has predictable* κ*-intensity* $\left(\widetilde{\lambda}_t^{\circ y}\right)$, *then* $\widetilde{Q} \ll_{\text{loc}} Q$ *and*

$$\mathcal{L}_t = \exp\left(-\overline{\widetilde{\Lambda}}_t^\circ + \lambda t\right)\prod_{n=1}^{\overline{N}_t^\circ}\frac{\widetilde{\lambda}_{\tau_n}^{\circ\eta_n}}{\lambda}.$$

Proof. With Q Poisson, all $P^{(n)}_{z_n}$ have Lebesgue densities that are > 0 on $]t_n, \infty[$ and have termination point ∞. By assumption also all $\widetilde{P}^{(n)}_{z_n}$ have Lebesgue densities and hence $\widetilde{P}^{(n)}_{z_n} \ll_{\text{loc}} P^{(n)}_{z_n}$. Because Q has κ-intensity process $\lambda^{\circ y}_t \equiv \lambda$ for all t, y, the assumptions about $\widetilde{Q}$ imply also that $\widetilde{\pi}^{(n)}_{z_n,t} \ll \pi^{(n)}_{z_n,t}$ always. Thus $\widetilde{Q} \ll_{\text{loc}} Q$ by Theorem 5.1.1 and the expression for $\mathcal{L}_t$ emerges as a special case of (5.7). □

Example 5.1.1 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 this case

$$\mathcal{L}_t = \frac{1}{\overline{P}^{(0)}(t)} 1_{(N^\circ_t = 0)}. \tag{5.8}$$

Example 5.1.2 *If Q, and $\widetilde{Q}$ are probabilities on $(W, \mathcal{H})$, and Q is homogeneous Poisson (λ) where $\lambda > 0$, $\widetilde{Q}$ is homogeneous Poisson $(\widetilde{\lambda})$, then $\widetilde{Q} \ll_{\text{loc}} Q$ and*

$$\mathcal{L}_t = e^{-(\widetilde{\lambda}-\lambda)t} \left(\frac{\widetilde{\lambda}}{\lambda}\right)^{N^\circ_t}.$$

Note that on $\mathcal{H}$, $\widetilde{Q} \perp Q$ if $\widetilde{\lambda} \neq \lambda$ since by the strong law of large numbers $\widetilde{Q}\left(\lim_{t\to\infty} \frac{1}{t} N^\circ_t = \widetilde{\lambda}\right) = 1$ and $Q\left(\lim_{t\to\infty} \frac{1}{t} N^\circ_t = \lambda\right) = 1$.

Example 5.1.3 Recall the description in Examples 3.3.1 and 4.3.6 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

$$X^\circ_t = \eta_{\langle t \rangle}$$

with $X^\circ_0 \equiv i_0$, 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 $q_{ij} > 0$. In this case

$$\mathcal{L}_t = \exp\left(-\int_0^t \left(\widetilde{\lambda}_{X^\circ_s} - \lambda_{X^\circ_s}\right) ds\right) \prod_{(i,j):i\neq j} \left(\frac{\widetilde{q}_{ij}}{q_{ij}}\right)^{N^{\circ ij}_t}$$

where $\lambda_i = -q_{ii}$, $\widetilde{\lambda}_i = -\widetilde{q}_{ii}$ and

$$N^{\circ ij}_t = \sum_{0<s\leq t} 1_{(X^\circ_{s-}=i, X^\circ_s=j)} = \int_{]0,t]} 1_{(X^\circ_{s-}=i)}\, dN^{\circ j}_s$$

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.

Exercise 5.1.1 Consider the statistical model for observation of a counting process N, specifying that N is homogeneous Poisson with unknown parameter λ. Suppose N is observed completely on the interval $[0, t]$. Find the maximum-likelihood estimator for λ.

Exercise 5.1.2 Let G be a finite state space and consider the statistical model for observing a homogeneous Markov chain X with state space G, given the initial state $X_0 \equiv i_0$ and unknown transition intensities q_{ij} for $i \neq j \in G$. Suppose that X is observed completely on the interval $[0, t]$. Show that the maximum-likelihood estimator for q_{ij} is the occurrence-exposure rate

$$\widehat{q}_{ij} = \frac{N_t^{ij}}{\int_0^t 1_{(X_s=i)}\, ds}$$

for $i \neq j$, where N_t^{ij} is the number of jumps from i to j made by X on $[0, t]$.

If $\widetilde{Q} \ll_{\text{loc}} Q$, $\mathcal{L}$ is a Q-martingale, hence it has a representation

$$\mathcal{L}_t = 1 + M_t^\circ(S);$$

cf. Theorem 4.6.1. If $\mathcal{L}$ is of the form (5.7), identifying the jumps of $\mathcal{L}$ one finds that the predictable field S is given by

$$S_t^y = \mathcal{L}_{t-}\left(\frac{\widetilde{\lambda}_t^y}{\lambda_t^y} - 1\right).$$

Suppose Q and $\widetilde{Q}$ are as in Example 5.1.1. If, say, $P^{(0)}$ has termination point 1, with the convention $\frac{1}{0} \cdot 0 = 0$, the expression (5.8) for $\mathcal{L}_t$ still makes sense Q-a.s. for all t: with Q-probability one, $\overline{P}^{(0)}(\tau_1-) > 0$, and at the timepoint $\tau_1 \leq 1$, $\mathcal{L}$ jumps from the value $1/\overline{P}^{(0)}(\tau_1-)$ to 0, where it remains forever since $1_{(N_t^\circ=0)} = 0$ for $t \geq \tau_1$. Thus $(\mathcal{L}_t)_{t<1}$ is a Q-martingale on the time interval $[0, 1[$ but for $t \geq 1$ we have $\widetilde{Q}_t \perp Q_t$ and $\mathcal{L}$ is certainly not a Q-martingale since for $t \geq 1$, $E_Q\mathcal{L}_t = 0 \neq E_Q\mathcal{L}_0 = 1$. If $\Delta P^{(0)}(1) > 0$, then $\mathcal{L}$ is uniformly bounded (by $1/\overline{P}^{(0)}(1-)$) and is therefore not even a local martingale. Indeed, suppose $\mathcal{L}$ is a local Q-martingale with reducing sequence (ρ_n). Then $E_Q\mathcal{L}_{\rho_n \wedge t} = 1$ for all n and t, but for $t = 1$,

$$E_Q\mathcal{L}_{\rho_n \wedge 1} = E_Q\left[\mathcal{L}_{\rho_n}; \rho_n < 1\right] \leq \frac{1}{\overline{P}^{(0)}(1-)} Q(\rho_n < 1)$$

which $\to 0$ as $n \to \infty$ and we have a contradiction.

There are also expressions of the analytical form (5.7) that are Q-local martingales but not Q-martingales. They typically arise if $\widetilde{Q}$ corresponds to an exploding RCM; see Theorem 5.2.1(ii) below.

5.2 Constructing RCMs from martingales

In the previous section we saw that if Q, and $\widetilde{Q}$ are probabilities on $(\mathcal{M}, \mathcal{H})$ such that $\widetilde{Q} \ll_{\text{loc}} Q$, then the likelihood process $\mathcal{L}$ is a non-negative Q-martingale with $E_Q\mathcal{L}_t = 1$ for all t, in particular $\mathcal{L}_0 \equiv 1$. In this section we start instead just with the probability Q and a non-negative cadlag Q-martingale M with $M_0 \equiv 1$ — so that also $EM_t = E_Q M_t = 1$ for all t — and we wish to construct a probability $\widetilde{Q}$ such that for all t, (5.12) below holds. As Theorem 5.2.1 will show this is in fact possible, and it is even possible to construct a possibly exploding $\widetilde{Q}$ if M is just a Q-local martingale with (τ_n) as a reducing sequence.

Suppose that $M \geq 0$ is a cadlag Q-martingale with $M_0 \equiv 1$. For each t, define a probability $\widetilde{Q}_t$ on $\mathcal{H}_t$ by

$$\widetilde{Q}_t(H) = \int_H M_t \, dQ \quad (H \in \mathcal{H}_t), \tag{5.9}$$

as in Theorem 5.2.1 below, see (5.12). The martingale property of M ensures precisely that for all $s \leq t$, $\widetilde{Q}_t = \widetilde{Q}_s$ on $\mathcal{H}_s$. Hence (5.9) may be used to define a set function $\widetilde{Q}$ on the algebra $\bigcup_{t \geq 0} \mathcal{H}_t$ of subsets of $\mathcal{M}$ with restriction $\widetilde{Q}_t$ to $\mathcal{H}_t$ for every t. The problem is then to extend this set function to all of $\mathcal{H}$ so that it is σ-additive and therefore a probability. For this it is enough to show that $\widetilde{Q}$ is σ-additive on the algebra, but doing this directly does not appear easy, so instead we use a consistency argument by presenting the candidates for the $\widetilde{Q}$-distribution of ξ_n for every n. This approach has the advantage that it becomes possible to handle also the local martingales described above, where the end result is a probability $\widetilde{Q}$ on the space $\overline{\mathcal{M}}$ of possibly exploding RCMs that may of course but need not be concentrated on $\mathcal{M}$; i.e., starting with a probability on $\mathcal{M}$ it is possible to construct a probability on the larger space $\overline{\mathcal{M}}$ in a non-trivial manner.

Consider again a martingale M as above. With M in particular non-negative, $M_\infty = \lim_{t \to \infty} M_t$ exists Q-a.s. by the martingale convergence theorem. By Fatou's lemma, $EM_\infty \leq 1$, and if $EM_\infty = 1$, the construction of $\widetilde{Q}$ is trivial: define

$$\widetilde{Q}(H) = \int_H M_\infty \, dQ \quad (H \in \mathcal{H}). \tag{5.10}$$

If, by contrast $EM_\infty = 0$, as is entirely possible, this identity is of course useless and the $\widetilde{Q}$ to be obtained from Theorem 5.2.1 will then in fact be singular with respect to Q, $\widetilde{Q} \perp Q$. (Note that if $EM_\infty = 1$ it follows from (5.10) that necessarily $E[M_\infty | \mathcal{H}_t] = M_t$ for all t so that (5.9) holds: for any $H \in \mathcal{H}_t$, by Fatou's lemma

$$\int_H M_\infty \, dQ \leq \liminf_{u \to \infty} \int_H M_u \, dQ = \int_H M_t \, dQ$$

so that $E[M_\infty | \mathcal{H}_t] \leq M_t$. But then $Q(H_0) = 0$, where $H_0 = (E[M_\infty | \mathcal{H}_t] < M_t)$, since

$$\begin{aligned} 1 &= EM_\infty \\ &= \int_{H_0^c} + \int_{H_0} E[M_\infty | \mathcal{H}_t] \, dQ \\ &= EM_t + \int_{H_0} (E[M_\infty | \mathcal{H}_t] - M_t) \, dQ \end{aligned}$$

so the integral is 0).

If M is a Q-local martingale with some reducing sequence (ρ_n), non-negative with $M_0 \equiv 1$, then M is a Q-supermartingale: by Fatou's lemma,

$$EM_t = E \lim_{n\to\infty} M_t^{\rho_n} \leq \lim\inf_{n\to\infty} EM_t^{\rho_n} = 1$$

so that M_t is Q-integrable, and then by the same reasoning applied to conditional expectations (understood as integrals with respect to a regular conditional probability of Q given $\mathcal{H}_s$),

$$E\left[M_t \left|\mathcal{H}_s\right.\right] \leq \lim\inf_{n\to\infty} E\left[M_t^{\rho_n} \left|\mathcal{H}_s\right.\right] = \lim_{n\to\infty} M_s^{\rho_n} = M_s. \tag{5.11}$$

Again, $M_\infty = \lim_{t\to\infty} M_t$ exists Q-a.s. (because $M \geq 0$ is a supermartingale) with now $EM_\infty \leq \inf_{t\geq 0} EM_t = \lim_{t\to\infty} EM_t$.

If $M \geq 0$ is a cadlag Q-local martingale with $M_0 \equiv 1$, then M is a true martingale iff $EM_t = 1$ for all t: one way is trivial so suppose that all $EM_t = 1$ and consider for given $0 \leq s < t$ the difference

$$0 = EM_s - EM_t = E\left[M_s - E\left[M_t \left|\mathcal{H}_s\right.\right]\right]$$

and conclude from (5.11) that $M_s = E\left[M_t \left|\mathcal{H}_s\right.\right]$ Q-a.s. .

We shall recapitulate the notation to be used: on the space $\left(\overline{\mathcal{M}}, \overline{\mathcal{H}}\right)$ of RCMs allowing explosions, the time of explosion is $\tau_\infty = \lim_{n\to\infty} \tau_n \leq \infty$ – we retain the notations $\overline{N}^\circ, \tau_n, \eta_n, \xi_n$ also when working with $\overline{\mathcal{M}}$ instead of $\mathcal{M}$ (of course on $\overline{\mathcal{M}}$, $\overline{N}_t^\circ = \infty$ is possible).

We shall also use the sequence spaces $\mathrm{K}_E, \mathrm{K}_E^{(n)}$ and $\overline{\mathrm{K}}_E$ introduced in Section 2.1 and here we will use the notation $Z_k^\circ = \left(T_1^\circ, \ldots, T_k^\circ; Y_1^\circ, \ldots, Y_k^\circ\right)$ on all three spaces to denote the coordinate projections taking an infinite or finite sequence $(t_1, t_2, \ldots; y_1, y_2, \ldots)$ into $z_k = (t_1, \ldots, t_k; y_1, \ldots, y_k)$. (On $\mathrm{K}_E^{(n)}$, Z_k° is defined only for $k \leq n$). The special notation

$$\left(\underline{\infty}, \underline{\nabla}\right)_n = (\infty, \ldots, \infty; \nabla, \ldots, \nabla)$$

is used for the sequence of n ∞'s and n ∇'s.

Finally, we shall use the notation

$$\zeta = \inf\{n : \tau_n = \infty\}$$

for the number of the last jump plus one (finite iff the RCM has only finitely many jumps altogether) on the spaces $\overline{\mathcal{M}}$ or $\mathcal{M}$ and, correspondingly,

$$S^\circ = \inf\left\{k : T_k^\circ = \infty\right\}$$

on the spaces $\mathrm{K}_E, \mathrm{K}_E^{(n)}$ and $\overline{\mathrm{K}}_E$. (On $\mathrm{K}_E^{(n)}$, S° will be used only on the subset $\bigcup_{k=1}^n \left(T_k^\circ = \infty\right)$).

Theorem 5.2.1 (i) *Let Q be a probability on $(\mathcal{M}, \mathcal{H})$ and let $M = (M_t)_{t\geq 0}$ be a non-negative cadlag Q-martingale with $M_0 \equiv 1$. Then there exists a uniquely defined probability $\widetilde{Q}$ on $(\mathcal{M}, \mathcal{H})$ such that for all $t \geq 0$,*

$$\widetilde{Q}(H) = \int_H M_t \, dQ \quad (H \in \mathcal{H}_t). \tag{5.12}$$

For any $n \in \mathbb{N}$, the $\widetilde{Q}$-distribution of ξ_n is determined by

$$\widetilde{Q}(\xi_n \in C_n) = \widetilde{Q}(\xi_n \in C_n, \zeta > n) + \sum_{k=1}^{n} \widetilde{Q}(\xi_n \in C_n, \zeta = k) \tag{5.13}$$

for C_n an arbitrary measurable subset of $\overline{\mathbb{R}}_+^n \times \overline{E}^n$, where

$$\widetilde{Q}(\xi_n \in C_n, \zeta > n) = \lim_{t\to\infty} \int_{(\xi_n \in C_n, \tau_n \leq t)} M_t \, dQ, \tag{5.14}$$

$$\widetilde{Q}(\xi_n \in C_n, \zeta = k) = \lim_{t\to\infty} \int_{\left(\xi_{k-1} \in C_{n,k-1}, \overline{N}_t^\circ = k-1\right)} M_t \, dQ, \tag{5.15}$$

and in (5.15), $C_{n,k-1}$ for $1 \leq k \leq n$ is the set

$$C_{n,k-1} = \left\{ z_{k-1} \in \mathbb{R}^{k-1} \times E^{k-1} : \text{join}\left(z_{k-1}, \left(\underline{\infty}, \underline{\nabla}\right)_{n-k+1}\right) \in C_n \right\}. \tag{5.16}$$

(ii) *Let Q be a probability on $(\mathcal{M}, \mathcal{H})$ and let $M = (M_t)_{t\geq 0}$ be a non-negative cadlag Q-local martingale with reducing sequence (τ_n) and $M_0 \equiv 1$. Then there exists a uniquely defined probability $\widetilde{Q}$ on the space $\left(\overline{\mathcal{M}}, \overline{\mathcal{H}}\right)$ of RCMs allowing explosions such that for all $t \geq 0$,*

$$\widetilde{Q}(H \cap (\tau_\infty > t)) = \int_{H \cap \mathcal{M}} M_t \, dQ \quad \left(H \in \overline{\mathcal{H}}_t\right). \tag{5.17}$$

The probability $\widetilde{Q}$ is concentrated on $\mathcal{M}$, the space of stable RCMs, if and only if M is a Q-martingale.
For any $n \in \mathbb{N}$, the $\widetilde{Q}$-distribution of ξ_n is determined by

$$\widetilde{Q}(\xi_n \in C_n) = \widetilde{Q}(\xi_n \in C_n, \zeta > n) + \sum_{k=1}^{n} \widetilde{Q}(\xi_n \in C_n, \zeta = k) \tag{5.18}$$

for C_n an arbitrary measurable subset of $\overline{\mathbb{R}}_+^n \times \overline{E}^n$, where

$$\widetilde{Q}(\xi_n \in C_n, \zeta > n) = \lim_{t\to\infty} \int_{(\xi_n \in C_n, \tau_n \leq t)} M_t^{\tau_n} \, dQ, \tag{5.19}$$

$$\widetilde{Q}(\xi_n \in C_n, \zeta = k) = \lim_{t\to\infty} \int_{\left(\xi_{k-1} \in C_{n,k-1}, \overline{N}_t^\circ = k-1\right)} M_t^{\tau_n} \, dQ, \tag{5.20}$$

and $C_{n,k-1}$ for $1 \leq k \leq n$ is defined as in (5.16).

Note. Note that while C_n in (5.13) and (5.18) is a subset of $\overline{\mathbb{R}}_+^n \times \overline{E}^n$ allowing for infinite timepoints, $C_{n,k-1}$, as given by (5.16), involves only finite timepoints. It is clear from (5.14), (5.15), (5.19) and (5.20) that throughout C_n may be replaced by $C_n \cap \overline{K}_E$. (5.14) and (5.19) involve only the part of C_n comprising finite timepoints with the part involving infinite timepoints appearing in (5.15) and (5.20). Finally note that for $k = 1$, the set $C_{n,0}$ given by (5.16) should be understood as 'everything' if $\left(\underline{\infty}, \underline{\nabla}\right)_n \in C_n$ and as 'nothing' otherwise, so that e.g.,

$$\widetilde{Q}\left(\xi_n \in C_n, \zeta = 1\right) = \lim_{t\to\infty} \int_{\left(\overline{N}_t^{\circ}=0\right)} M_t \, dQ$$

if $\left(\underline{\infty}, \underline{\nabla}\right)_n \in C_n$ while if $\left(\underline{\infty}, \underline{\nabla}\right)_n \notin C_n$, the probability $= 0$.

Proof. (i) The basic idea is to construct $\widetilde{Q}$ by using (5.13), (5.14) and (5.15) to define the $\widetilde{Q}$-distribution of ξ_n for any n, show that the results form a consistent family of finite-dimensional distributions and then use Kolmogorov's consistency theorem to obtain a uniquely defined probability R on the product space $\overline{\mathbb{R}}_+^{\mathbb{N}} \times \overline{E}^{\mathbb{N}}$ with the correct finite-dimensional distributions. It is obvious that R is concentrated on the set $\overline{K}(E)$ of possible sequences of jump times and marks allowing explosion; hence by transformation from R we obtain a probability $\widetilde{Q} = \varphi(\mathrm{R})$ on $\left(\overline{\mathcal{M}}, \overline{\mathcal{H}}\right)$ with φ the basic bimeasurable bijection between $\overline{K}(E)$ and $\overline{\mathcal{M}}$ from Section 3.2. We then finally show that $\widetilde{Q}$ is concentrated on the space $\mathcal{M}$ of stable RCMs and satisfies (5.12), an equation that as noted above in the introduction to this section determines $\widetilde{Q}$ on the algebra $\bigcup_{t\geq 0} \mathcal{H}_t$ of subsets of $\mathcal{M}$, which is enough to see that $\widetilde{Q}$ is uniquely determined when extending to the full σ-algebra $\mathcal{H}$.

Let M be a non-negative Q-martingale with $M_0 \equiv 1$. For n given, define a set function $\mathrm{R}^{(n)}$ on $\overline{\mathcal{B}}_+^n \otimes \overline{\mathcal{E}}^n$ using the recipes (5.14) and (5.15), i.e.,

$$\mathrm{R}^{(n)}\left(Z_n^{\circ} \in C_n\right) = \mathrm{R}^{(n)}\left(Z_n^{\circ} \in C_n, T_n^{\circ} < \infty\right) + \sum_{k=1}^{n} \mathrm{R}^{(n)}\left(Z_n^{\circ} \in C_n, S^{\circ} = k\right) \quad (5.21)$$

with

$$\mathrm{R}^{(n)}\left(Z_n^{\circ} \in C_n, T_n^{\circ} < \infty\right) = \lim_{t\to\infty} \int_{(\xi_n \in C_n, \tau_n \leq t)} M_t \, dQ, \quad (5.22)$$

$$\mathrm{R}^{(n)}\left(Z_n^{\circ} \in C_n, S^{\circ} = k\right) = \lim_{t\to\infty} \int_{\left(\xi_{k-1} \in C_{n,k-1}, \overline{N}_t^{\circ} = k-1\right)} M_t \, dQ. \quad (5.23)$$

Here we first note that the limit in (5.22) exists because the integrals increase with t: for $t \leq u$,

$$\begin{aligned}\int_{(\xi_n \in C_n, \tau_n \leq t)} M_t \, dQ &= \int_{(\xi_n \in C_n, \tau_n \leq t)} M_u \, dQ \\ &\leq \int_{(\xi_n \in C_n, \tau_n \leq u)} M_u \, dQ\end{aligned}$$

using that $(\xi_n \in C_n, \tau_n \le t) \in \mathcal{H}_t$ and that M is a martingale. Since $\left(\overline{N}_t^\circ = k-1\right) = (\tau_{k-1} \le t) \setminus (\tau_k \le t)$, the integral in (5.23) may be written as the difference between two integrals, both of the form considered in (5.22), hence the limit in (5.23) also exists.

Because the limit in (5.22) is increasing it follows by monotone convergence that $\mathrm{R}^{(n)}\left(Z_n^\circ \in C_n, T_n^\circ < \infty\right)$ is σ-additive as a function of C_n. Furthermore, if $C_{j,n}$ for $j \in \mathbb{N}$ are mutually disjoint sets in $\overline{\mathcal{B}}^n \otimes \overline{\mathcal{E}}^n$, for each k, the sets $C_{j,n,k-1}$ are mutually disjoint sets in $\mathcal{B}^{k-1} \otimes \mathcal{E}^{k-1}$. Writing the integral in (5.23) as the difference between two integrals as was done above, it follows that also $\mathrm{R}^{(n)}\left(Z_n^\circ \in C_n, S^\circ = k\right)$ is σ-additive in C_n for each k, $1 \le k \le n$.

Thus $\mathrm{R}^{(n)}$ defined by (5.21) is σ-additive on $\overline{\mathcal{B}}^n \otimes \overline{\mathcal{E}}^n$. Of course $\mathrm{R}^{(n)}(\emptyset) = 0$, $\mathrm{R}^{(n)} \ge 0$, hence $\mathrm{R}^{(n)}$ is a probability if the total mass is 1. But since $(\tau_n \le t) = \left(\overline{N}_t^\circ \ge n\right)$, using (5.22) and (5.23) with $C_n = \overline{\mathbb{R}}_+^n \times \overline{E}^n$ gives

$$\begin{aligned}
\mathrm{R}^{(n)}\left(\overline{\mathbb{R}}_+^n \times \overline{E}^n\right) &= \lim_{t\to\infty}\left(\int_{\left(\overline{N}_t^\circ \ge n\right)} + \sum_{k=1}^n \int_{\left(\overline{N}_t^\circ = k-1\right)} M_t \, dQ\right) \\
&= \lim_{t\to\infty} EM_t \\
&= 1.
\end{aligned}$$

The next step is to show consistency of the family $\left(\mathrm{R}^{(n)}\right)_{n\ge1}$ of probabilities, i.e., it must be shown that

$$\mathrm{R}^{(n+1)}\left(Z_{n+1}^\circ \in C_{n+1}\right) = \mathrm{R}^{(n)}\left(Z_n^\circ \in C_n\right) \tag{5.24}$$

for all n and all $C_n \in \overline{\mathcal{B}}^n \otimes \overline{\mathcal{E}}^n$ with $C_{n+1} \in \overline{\mathcal{B}}^{n+1} \otimes \overline{\mathcal{E}}^{n+1}$ the set of all z_{n+1} such that $Z_n^\circ(z_{n+1}) = z_n \in C_n$. But using (5.21), (5.22) and (5.23) with n replaced by $n+1$ and C_n replaced by C_{n+1}, since clearly $C_{n+1,k-1} = C_{n,k-1}$ for $1 \le k \le n$ and $C_{n+1,n} = C_n$, it follows that

$$\begin{aligned}
\mathrm{R}^{(n+1)}\left(Z_{n+1}^\circ \in C_{n+1}\right) = \lim_{t\to\infty}\Bigg(&\int_{(\xi_n \in C_n, \tau_{n+1} \le t)} + \sum_{k=1}^n \int_{\left(\xi_{k-1} \in C_{n,k-1}, \overline{N}_t^\circ = k-1\right)} M_t \, dQ \\
&+ \int_{\left(\xi_n \in C_n, \overline{N}_t^\circ = n\right)} M_t \, dQ\Bigg).
\end{aligned}$$

The limit of the sum over k is recognized from (5.23) as $\mathrm{R}^{(n)}\left(Z_n^\circ \in C_n, S^\circ \le n\right)$ and since

$$\begin{aligned}
&\lim_{t\to\infty}\left(\int_{(\xi_n \in C_n, \tau_{n+1} \le t)} + \int_{\left(\xi_n \in C_n, \overline{N}_t^\circ = n\right)} M_t \, dQ\right) \\
&= \lim_{t\to\infty} \int_{(\xi_n \in C_n, \tau_n \le t)} M_t \, dQ \\
&= \mathrm{R}^{(n)}\left(Z_n^\circ \in C_n, T_n^\circ < \infty\right)
\end{aligned}$$

by (5.22), (5.24) follows.

By Kolmogorov's consistency theorem there is therefore a unique probability R on $\overline{\mathbb{R}}_+^{\mathbb{N}} \times \overline{E}^{\mathbb{N}}$ such that $\mathrm{R}\left(Z_n^\circ \in C_n\right) = \mathrm{R}^{(n)}\left(Z_n^\circ \in C_n\right)$ for all n and all $C_n \in \overline{\mathcal{B}}^n \otimes \overline{\mathcal{E}}^n$. By the construction through (5.21), (5.22) and (5.23) it is clear that $\mathrm{R}^{(n)}$ is concentrated on $\mathrm{K}_E^{(n)}$; hence R is concentrated on $\overline{\mathrm{K}}_E$ and $\widetilde{Q} = \varphi(\mathrm{R})$ with φ given by (3.5) (used on all of $\overline{\mathrm{K}}_E$) is well defined as a probability on $\left(\overline{\mathcal{M}}, \overline{\mathcal{H}}\right)$, the space of RCMs allowing explosion.

It remains to show that $\widetilde{Q}$ is concentrated on $\mathcal{M} \subset \overline{\mathcal{M}}$ and satisfies (5.12). Let $H \subset \mathcal{M}$ be such that $H \in \mathcal{H}_t$. For every n, by (4.31),

$$
\begin{aligned}
H \cap \left(\overline{N}_t^\circ = n\right) &= (\xi_n \in C_n, \tau_{n+1} > t) \qquad (5.25)\\
&= (\xi_n \in C_n) \setminus (\xi_n \in C_n, \tau_{n+1} \le t)\\
&= (\xi_n \in C_n, \zeta > n) \setminus (\xi_n \in C_n, \tau_{n+1} \le t, \zeta > n+1)
\end{aligned}
$$

for some measurable $C_n \subset]0,t]^n \times E^n$. But then by (5.14),

$$
\widetilde{Q}\left(H \cap \left(\overline{N}_t^\circ = n\right)\right) = \lim_{u\to\infty}\left(\int_{(\xi_n \in C_n, \tau_n \le u)} - \int_{(\xi_n \in C_n, \tau_{n+1} \le t, \tau_{n+1} \le u)} M_u \, dQ\right)
$$

which, since $C_n \subset]0,t]^n \times E^n$ so that the sets of integration are in $\mathcal{H}_t$, and because M is a martingale, reduces to

$$
\begin{aligned}
\lim_{u\to\infty}\left(\int_{(\xi_n \in C_n)} - \int_{(\xi_n \in C_n, \tau_{n+1} \le t)} M_u \, dQ\right) &= \int_{(\xi_n \in C_n, \tau_{n+1} > t)} M_t \, dQ\\
&= \int_{\left(H \cap \left(\overline{N}_t^\circ = n\right)\right)} M_t \, dQ.
\end{aligned}
$$

Thus, summing over finite n and using that $H \subset \mathcal{M}$, we get

$$
\widetilde{Q}(H) = \sum_{n=0}^{\infty} \widetilde{Q}\left(H \cap \left(\overline{N}_t^\circ = n\right)\right) = \int_H M_t \, dQ,
$$

proving (5.12). Taking $H = \mathcal{M}$ finally shows that $\widetilde{Q}(\mathcal{M}) = EM_t = 1$.

(ii) Let M be a Q-local martingale (τ_n), non-negative with $M_0 \equiv 1$. Applying (i) to the stopped martingale M^{τ_n}, we obtain for each n a probability $\widetilde{Q}^{\tau_n}$ on $(\mathcal{M}, \mathcal{H})$ such that for $t \ge 0$,

$$
\widetilde{Q}^{\tau_n}(H) = \int_H M_t^{\tau_n} \, dQ \quad (H \in \mathcal{H}_t). \qquad (5.26)
$$

Here $\widetilde{Q}^{\tau_n} = \varphi\left(\mathrm{R}^{\tau_n}\right)$ where R^{τ_n} is the uniquely determined probability on $\overline{\mathbb{R}}_+^{\mathbb{N}} \times \overline{E}^{\mathbb{N}}$ satisfying that

$$
\mathrm{R}^{\tau_n}\left(Z_k^\circ \in C_k\right) = \lim_{t\to\infty}\left(\int_{(\xi_k \in C_k, \tau_k \le t)} + \sum_{j=1}^{k} \int_{\left(\xi_{j-1} \in C_{k,j-1}, \overline{N}_t^\circ = j-1\right)} M_t^{\tau_n} \, dQ\right) \qquad (5.27)
$$

for all k, all $C_k \in \overline{\mathcal{B}}_+^k \otimes \overline{\mathcal{E}}^k$. But all sets of integration belong to $\mathcal{H}_{\tau_k \wedge t}$; hence by optional sampling applied to the bounded stopping times $\tau_k \wedge t$ and $\tau_n \wedge t$, one may replace the integrand $M_t^{\tau_n}$ by $M_t^{\tau_k}$ provided $n \geq k$. For each $k \in \mathbb{N}$ we can therefore define a probability $\mathrm{R}^{(k)}$ on $\overline{\mathbb{R}}_+^k \times \overline{E}^k$ by

$$\mathrm{R}^{(k)}\left(Z_k^\circ \in C_k\right) = \mathrm{R}^{\tau_n}\left(Z_k^\circ \in C_k\right)$$

for an arbitrary $n \geq k$. It is then clear that the $\mathrm{R}^{(k)}$ are consistent: with $C_k \in \overline{\mathcal{B}}_+^k \otimes \overline{\mathcal{E}}^k$, define as in the proof of (i), $C_{k+1} = \{z_{k+1} : z_k \in C_k\}$ and note that for $n \geq k+1$,

$$\mathrm{R}^{(k+1)}\left(Z_{k+1}^\circ \in C_{k+1}\right) = \mathrm{R}^{\tau_n}\left(Z_{k+1}^\circ \in C_{k+1}\right) = \mathrm{R}^{\tau_n}\left(Z_k^\circ \in C_k\right) = \mathrm{R}^{(k)}\left(Z_k^\circ \in C_k\right).$$

By the consistency theorem there is a unique probability R on $\overline{\mathbb{R}}_+^{\mathbb{N}} \times \overline{E}^{\mathbb{N}}$, by its construction concentrated on $\overline{\mathrm{K}}(E)$, such that for all k and all $C_k \in \overline{\mathcal{B}}_+^k \otimes \overline{\mathcal{E}}^k$, $\mathrm{R}\left(Z_k^\circ \in C_k\right) = \mathrm{R}^{(k)}\left(Z_k^\circ \in C_k\right)$. We of course now define $\widetilde{Q} = \varphi(\mathrm{R})$ as a probability on $\left(\overline{\mathcal{M}}, \overline{\mathcal{H}}\right)$ and note that e.g., for any $C_k \in \overline{\mathcal{B}}_+^k \otimes \overline{\mathcal{E}}^k$, writing $C_k' = C_k \cap \left(\mathbb{R}_+^k \times E^k\right)$ and using (5.27) and the fact that $C_{k,j-1}' = \emptyset$ for $1 \leq j \leq k$, it follows that

$$\begin{aligned}
\widetilde{Q}(\xi_k \in C_k, \zeta > k) &= \mathrm{R}\left(Z_k^\circ \in C_k, T_k^\circ < \infty\right) \\
&= \mathrm{R}^{(k)}\left(Z_k^\circ \in C_k'\right) \\
&= \mathrm{R}^{\tau_n}\left(Z_k^\circ \in C_k'\right) \\
&= \lim_{t \to \infty} \int_{(\xi_k \in C_k, \tau_k \leq t)} M_t^{\tau_n}\, dQ \qquad (5.28)
\end{aligned}$$

for any $n \geq k$. Taking $n = k$ yields (5.19), and (5.20) is then shown similarly, using that for $j \leq k$,

$$\begin{aligned}
\widetilde{Q}(\xi_k \in C_k, \zeta = j) &= \widetilde{Q}\left(\xi_{j-1} \in C_{k,j-1}, \zeta = j\right) \\
&= \widetilde{Q}\left(\xi_{j-1} \in C_{k,j-1}, \zeta > j-1\right) - \widetilde{Q}\left(\xi_{j-1} \in C_{k,j-1}, \zeta > j\right)
\end{aligned}$$

and evaluating both terms using (5.28).

Finally, for $H \subset \overline{\mathcal{M}}$ with $H \in \overline{\mathcal{H}}_t$, we may still represent $H \cap \left(\overline{N}_t^\circ = n\right)$ as in (5.25) with $C_n \subset\,]0,t]^n \times E^n$, and then get

$$\begin{aligned}
\widetilde{Q}\left(H \cap \left(\overline{N}_t^\circ = n\right)\right) &= \mathrm{R}\left(Z_n^\circ \in C_n\right) - \mathrm{R}\left(Z_n^\circ \in C_n, T_{n+1}^\circ \leq t\right) \\
&= \mathrm{R}^{\tau_{n+1}}\left(Z_n^\circ \in C_n\right) - \mathrm{R}^{\tau_{n+1}}\left(Z_n^\circ \in C_n, T_{n+1}^\circ \leq t\right).
\end{aligned}$$

Here, by (5.27) and optional sampling,

$$\begin{aligned}
\mathrm{R}^{\tau_{n+1}}\left(Z_n^\circ \in C_n, T_{n+1}^\circ \leq t\right) &= \lim_{u \to \infty} \int_{(\xi_n \in C_n, \tau_{n+1} \leq t, \tau_{n+1} \leq u)} M_u^{\tau_{n+1}}\, dQ \\
&= \int_{(\xi_n \in C_n, \tau_{n+1} \leq t)} M_t^{\tau_{n+1}}\, dQ,
\end{aligned}$$

and similarly (because $\tau_n \le t$ on $(\xi_n \in C_n)$),

$$\mathrm{R}^{\tau_{n+1}}\left(Z_n^\circ \in C_n\right) = \int_{(\xi_n \in C_n)} M_t^{\tau_{n+1}}\, dQ.$$

Thus

$$\begin{aligned}\widetilde{Q}\left(H \cap \left(\overline{N}_t^\circ = n\right)\right) &= \int_{(\xi_n \in C_n)} - \int_{(\xi_n \in C_n, \tau_{n+1} \le t)} M_t^{\tau_{n+1}}\, dQ \\ &= \int_{(\xi_n \in C_n, \tau_{n+1} > t)} M_{\tau_{n+1} \wedge t}\, dQ \\ &= \int_{\left(H \cap \left(\overline{N}_t^\circ = n\right)\right)} M_t\, dQ\,.\end{aligned}$$

Summing on $n \in \mathbb{N}$ we finally get, using that Q viewed as a probability on $\overline{\mathcal{M}}$ is concentrated on $\mathcal{M} \subset \left(\overline{N}_t^\circ < \infty\right)$ for all t,

$$\begin{aligned}\widetilde{Q}\left(H \cap (\tau_\infty > t)\right) &= \widetilde{Q}\left(H \cap \left(\overline{N}_t^\circ < \infty\right)\right) \\ &= \int_{H \cap \left(\overline{N}_t^\circ < \infty\right)} M_t\, dQ \\ &= \int_{H \cap \mathcal{M}} M_t\, dQ,\end{aligned}$$

which proves (5.17). Taking $H = \overline{\mathcal{M}}$ gives

$$\widetilde{Q}\left(\tau_\infty > t\right) = EM_t$$

and therefore

$$\widetilde{Q}\left(\mathcal{M}\right) = \widetilde{Q}\left(\tau_\infty = \infty\right) = \lim_{t \to \infty} \widetilde{Q}\left(\tau_\infty > t\right) = \lim_{t \to \infty} EM_t$$

which equals 1 iff M is a true martingale: all $EM_t = 1$ (cf. p. 112). □

Remark 5.2.1 In (5.17) only subsets H of $\overline{\mathcal{M}}$ such that $H \in \overline{\mathcal{H}}_t$ are allowed. It is tempting, but *wrong* — and certainly gives a non-sensical result — to use the formula with $H = \mathcal{M}$: as a subset of $\overline{\mathcal{M}}$, $\mathcal{M}$ does not belong to $\overline{\mathcal{H}}_t$!

6

Independence

In the first part of this chapter it is shown how stochastic independence between finitely many marked point processes may be characterized in terms of the structure of the compensating measures. The last part of the chapter is devoted to the study of CPs and RCMs with independent increments (stationary or not), which are characterized as those having deterministic compensating measures, and of other stochastic processes with independent increments, with, in particular, a discussion of compound Poisson processes and more general Lévy processes.

The discussion of general Lévy processes at the end of the chapter is intended for information only and is not required reading. One may also omit reading the technical parts of the proof of Theorem 6.2.1(i) and the proof of Proposition 6.2.3(i).

References. Two main references on Lévy processes are the books by Bertoin [8] and Sato [108].

6.1 Independent point processes

Let $r \in \mathbb{N}$ with $r \geq 2$ and consider r given MPPs, each viewed as an RCM μ^i with mark space $(E^i, \mathcal{E}^i)$, $1 \leq i \leq r$, all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In particular μ^i may correspond to an SPP, in which case E^i is a one-point set and μ^i may be thought of as a counting process.

Let $(T_n^i)_{n\geq 1}$ denote the sequence of jump times for μ^i and $(Y_n^i)_{n\geq 1}$ the sequence of marks, so $Y_n^i \in \overline{E}^i = E^i \cup \{\nabla\}$, where ∇ is the irrelevant mark, common to all μ^i.

Now define $E = \{(i, y^i) : 1 \leq i \leq r,\ y^i \in E^i\}$, the disjoint union of the E^i, and let $\mathcal{E}$ be the σ-algebra of subsets of E generated by all sets of the form $\{i\} \times A^i$ where $1 \leq i \leq r,\ A^i \in \mathcal{E}^i$. Thus any $A \in \mathcal{E}$ has the form

$$A = \bigcup_{i=1}^{r} \{i\} \times A^i$$

where $A^i \in \mathcal{E}^i$ for each i (A is the disjoint union of the A^i), and then

$$A^i = \left\{ y^i \in E^i : \left(i, y^i\right) \in A \right\}.$$

Next put

$$\mu = \sum_{i=1}^{r} \sum_{\substack{n=1 \\ T_n^i < \infty}}^{\infty} \varepsilon_{(T_n^i, (i, Y_n^i))}. \tag{6.1}$$

Clearly μ is a random, $\overline{\mathbb{N}}_0$-valued measure such that

$$\mu([0,t] \times A) < \infty \quad (t \in \mathbb{R}_0, A \in \mathcal{E}). \tag{6.2}$$

However, for μ to be an RCM (with mark space E), we need (p.12) $\mu(\{t\} \times E) \leq 1$ for all t, at least $\mathbb{P}$-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.,

$$\mathbb{P} \bigcup_{\substack{i,j=1 \\ i \neq j}}^{r} \bigcup_{k,n=1}^{\infty} \left(T_k^i = T_n^j\right) < \infty = 0. \tag{6.3}$$

If (6.3) holds, μ is an 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 6.1.1 If all $E^i = E'$ and (6.3) is satisfied one can also define the *superposition* of the μ^i as the RCM μ' with mark space E' given by

$$\mu' = \sum_{i=1}^{r} \mu^i = \sum_{i=1}^{r} \sum_{\substack{n=1 \\ T_n^i < \infty}}^{\infty} \varepsilon_{(T_n^i, Y_n^i)}. \tag{6.4}$$

The difference from the corresponding aggregate μ given by (6.1) is of course that μ keeps track of which of the original μ^i a mark came from, while μ' does not.

Remark 6.1.2 Even if (6.3) 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 marks should be objects of the form

$$\left((i_1, \ldots, i_{r'}) ; \left(y^{i_1}, \ldots, y^{i_{r'}}\right)\right)$$

for $r' \geq 1$, $1 \leq i_1 < \cdots < i_{r'} \leq r$ and $y^{i_j} \in E^{i_j}$ for $1 \leq j \leq r'$, this mark designating that precisely the μ^i singled out in the set $D = \{i_1, \ldots, i_{r'}\}$ jump simultaneously with marks y^i for $i \in D$.

From now on assume that (6.3) holds, and even, after discarding a $\mathbb{P}$-null set, that for *all* $\omega \in \Omega$

$$T_k^i(\omega) \neq T_n^j(\omega) \tag{6.5}$$

for $(i,k) \neq (j,n)$ whenever $T_k^i(\omega)$ or $T_n^j(\omega)$ is finite.

Note that with this assumption, if

$$N_t^i(A^i) = \mu^i([0,t] \times A^i) \quad (1 \le i \le r,\ A^i \in \mathcal{E}^i), \tag{6.6}$$

then $N_t(A) = \mu([0,t] \times A)$ for $A \in \mathcal{E}$ is given by

$$N_t(A) = \sum_{i=1}^{r} N_t^i(A^i), \tag{6.7}$$

with $A^i := \{y^i \in E^i : (i, y^i) \in A\}$ as above.

It is immediately verified that for all t,

$$\mathcal{F}_t^{\mu} = \sigma\left(\mathcal{F}_t^{\mu^i}\right)_{1 \le i \le r}. \tag{6.8}$$

For $1 \le i \le r$, let L^i denote the $\mathcal{F}_t^{\mu^i}$-compensating measure for μ^i and let

$$\Lambda_t^i(A^i) = L^i([0,t] \times A^i) \quad (t \in \mathbb{R}_0,\ A^i \in \mathcal{E}^i) \tag{6.9}$$

denote the corresponding $\mathcal{F}_t^{\mu^i}$-compensator for the counting process $N^i(A^i)$.

Theorem 6.1.1 *Let $\mu^1, \ldots, \mu^r$ be RCMs satisfying (6.5) and for each i and $A^i \in \mathcal{E}^i$, let $\Lambda_t^i(A^i)$ be given by (6.9).*

(i) *If $\mu^1, \ldots, \mu^r$ are stochastically independent, then the $\mathcal{F}_t^{\mu}$-compensating measure L for the aggregate μ of $\mu^1, \ldots, \mu^r$ is given by*

$$\Lambda_t(A) = \sum_{i=1}^{r} \Lambda_t^i\left(A^i\right) \quad (t \in \mathbb{R}_0, A \in \mathcal{E}), \tag{6.10}$$

with $A^i = \left\{y^i : (i, y^i) \in A\right\}$, where

$$\Lambda_t(A) = L\left([0,t] \times A\right). \tag{6.11}$$

(ii) *If conversely (6.10) holds and in addition (6.3) holds for $\widetilde{\mu}^1, \ldots, \widetilde{\mu}^r$ when the $\widetilde{\mu}^i$ are stochastically independent with each $\widetilde{\mu}^i$ having the same marginal distribution as μ^i, then $\mu^1, \ldots, \mu^r$ are independent.*

Proof. Assume first that the μ^i are independent. By Theorem 4.5.2 and the discussion in Section 4.8, we must show that $\Lambda(A)$ given by (6.10) is $\mathcal{F}_t^{\mu}$-predictable (which is obvious since the ith 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, suppose first that $\mathbb{E}\overline{N}_t < \infty$ for all t. Then by Theorem 4.5.2, for all i, $M^i\left(A^i\right) := N^i\left(A^i\right) - \Lambda^i\left(A^i\right)$ is an $\mathcal{F}_t^{\mu^i}$-martingale and further, the $M^i\left(A^i\right)$ are independent since the μ^i are. We want to show that for $s < t$, $F \in \mathcal{F}_s^{\mu}$,

$$\int_F M_t(A)\,d\mathbb{P} = \int_F M_s(A)\,d\mathbb{P}. \tag{6.12}$$

Now $M(A) = \sum_{i=1}^r M^i\left(A^i\right)$, 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 $M^i\left(A^i\right)$, it follows that

$$\begin{aligned}
\int_F M_t(A)\,d\mathbb{P} &= \sum_{i=1}^r \int_F M_t^i\left(A^i\right)\,d\mathbb{P} \\
&= \sum_{i=1}^r \mathbb{P}\left(\bigcap_{j:j\neq i} F^j\right)\int_{F^i} M_t^i\left(A^i\right)\,d\mathbb{P} \\
&= \sum_{i=1}^r \mathbb{P}\left(\bigcap_{j:j\neq i} F^j\right)\int_{F^i} M_s^i\left(A^i\right)\,d\mathbb{P} \\
&= \int_F M_s(A)\,d\mathbb{P}.
\end{aligned}$$

This proves (6.12) 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 the class of sets $F \in \mathcal{F}_s^\mu$ for which (6.12) holds is closed under the formation of set differences and increasing limits of sets. Since the collection of sets $F = \bigcap_i F^i$ generates $\mathcal{F}_s^\mu$, cf. (6.8), contains Ω and is closed under the formation of finite intersections, (6.12) holds for all $F \in \mathcal{F}_s^\mu$.

Suppose now that the assumption $\mathbb{E}\overline{N}_t < \infty$ for all t does not hold. For any $n \in \mathbb{N}$, $1 \le i \le r$ the stopped process $M^{T_n^i}(A^i)$ is an $\mathcal{F}_t^{\mu^i}$-martingale; see Theorem 4.5.2. But the processes $M^{T_n^i}(A^i)$ are independent, and repeating the argument above one finds that

$$\sum_{i=1}^r M^{T_n^i}\left(A^i\right)$$

is an $\mathcal{F}_t^\mu$-martingale for each n. With $T_n = \inf\left\{t \ge 0 : \overline{N}_t = n\right\}$, for all i, $T_n \le T_n^i$, so by optional sampling (see p. 308 for general remarks concerning stopped martingales), $M^{T_n}(A) = \left(\sum_{i=1}^r M^{T_n^i}\left(A^i\right)\right)^{T_n}$ is an $\mathcal{F}_t^\mu$-martingale, i.e., $M(A)$ is an $\mathcal{F}_t^\mu$-local martingale.

Now suppose conversely that the compensator $\Lambda(A)$ is the sum of the compensators $\Lambda^i\left(A^i\right)$ as in (6.10). From the first part of the proof and the assumption that independent $\widetilde{\mu}^i$ with each $\widetilde{\mu}^i$ having the same law as μ^i satisfy (6.3), we know (6.10) to be true if the μ^i are independent, and from Theorem 4.3.2 and the fact that the distribution of μ determines the joint distribution of the μ^i (simply because the μ^i are functions of μ), it now follows that if (6.10) holds, the μ^i are indeed independent. □

Remark 6.1.3 More informally the theorem could be stated as follows: if the μ^i and the independent copies $\widetilde{\mu}^i$ both satisfy (6.5) (or just (6.3)), the μ^i 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 (6.11) and (6.10).

Remark 6.1.4 Note that if the μ^i are independent, (6.3) is satisfied if the restriction to $\mathbb{R}_+$ of the distribution of each T_n^i is continuous, but not if e.g., there exists a t_0 and $i \neq i'$ such that both μ^i and $\mu^{i'}$ have probability > 0 of performing a jump at time t_0.

Example 6.1.1 Suppose the μ^i are independent, and each μ^i has an $\mathcal{F}_t^{\mu^i}$-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 (6.3) is satisfied and the aggregate μ has an $\mathcal{F}_t^{\mu}$-predictable intensity process $(\lambda(A))_{A \in \mathcal{E}}$ given by

$$\lambda_t(A) = \sum_{i=1}^{r} \lambda_t^i\left(A^i\right)$$

with $A^i = \left\{y^i : \left(i, y^i\right) \in A\right\}$.

Example 6.1.2 Suppose the μ^i are independent with each μ^i a homogeneous Poisson random measure. Then (6.3) is satisfied and the aggregate μ is also a homogeneous Poisson random measure: if the $\mathcal{F}_t^{\mu^i}$-compensating measure for μ^i is the deterministic product measure $\ell \otimes \rho^i$ (cf. Example 4.3.3), then μ has $\mathcal{F}_t^{\mu}$-compensating measure $\ell \otimes \rho$, where $\rho(A) = \sum_i \rho^i\left(A^i\right)$ with A^i determined from A as above. If all $E^i = E'$, the superposition $\mu' = \sum_i \mu^i$ is also homogeneous Poisson with $\mathcal{F}_t^{\mu'}$-compensating measure $\ell \otimes \rho'$, where $\rho' = \sum_i \rho^i$.

6.2 Independent increments, Lévy processes

Let μ be an RCM with mark space E, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We use the standard notation T_n for the time of the nth jump, Y_n for the nth mark, and write $Z_n = (T_1, \ldots, T_n; Y_1 \ldots Y_n)$ and $Z_{\langle t \rangle} = Z_{\overline{N}_t}$. Recall the definitions (4.54) and (4.55) of the shift θ_s and the translated shift θ_s^*, respectively.

Definition 6.2.1 (i) μ has *independent increments* if for every $s \in \mathbb{R}_0$, $\theta_s \mu$ is independent of $\mathcal{F}_s^{\mu}$.

(ii) μ has *stationary independent increments* if it has independent increments and the distribution of $\theta_s^* \mu$ is the same for all $s \in \mathbb{R}_0$.

The definition applies in particular to counting processes. Thus, a counting process has independent increments if for every s, $\theta_s N$ is independent of $\mathcal{F}_s^N$.

Note. Note that the definition refers to the filtration $\left(\mathcal{F}_t^{\mu}\right)$ only, and thus, for an RCM μ to have independent increments is a property of the distribution of μ; cf. Theorem 6.2.1 below. One may easily define what it means for μ to have independent increments with respect to an arbitrary filtration to which μ is adapted, and then in particular μ has independent increments in the sense of Definition 6.2.1. Later, when discussing e.g., $\mathbb{R}$-valued processes with independent increments, we shall consider general filtrations.

It is easy to show (cf. Exercise 6.2.1 below) that μ has independent increments iff for all $s < t$ and all $A_1, \dots, A_r \in \mathcal{E}$ it holds that $\mathcal{N}_{st} = (N_t(A_i) - N_s(A_i))_{1 \le i \le r}$ is independent of $\mathcal{F}_s^\mu$. μ has stationary independent increments iff this condition holds and in addition the distribution of the vector $\mathcal{N}_{st}$ of increments depends on s and t through $t - s$ only.

Exercise 6.2.1 Show that μ has independent increments iff for all $s < t$ it holds that for all $n', r \in \mathbb{N}$, all $s_1 \le \cdots \le s_{n'} \le s$, all $A'_1, \dots, A'_{n'} \in \mathcal{E}$ and all $A_1, \dots, A_r \in \mathcal{E}$ the vector $\mathcal{N}_{st} = (N_t(A_i) - N_s(A_i))_{1 \le i \le r}$ is independent of $\left(N_{s_j}\left(A'_j\right)\right)_{1 \le j \le n'}$.

(*Hint*: the problem is to show that if this condition on $\mathcal{N}_{st}$ and the $N_{s_j}\left(A'_j\right)$ holds, then μ has independent increments. So assuming the condition to hold, first use the definition of $\mathcal{F}_s^\mu$ to show that $\mathcal{N}_{st}$ is independent of $\mathcal{F}_s^\mu$. Then argue that for any choice of $t_1 < \cdots < t_{n''}$ with $t_1 > s = t_0$ the collection $\left(\mathcal{N}_{t_{k-1}t_k}\right)_{1 \le k \le n''}$ is independent of $\mathcal{F}_s^\mu$ (with each $\mathcal{N}_{t_{k-1}t_k}$ defined similarly to $\mathcal{N}_{st}$, but with different choices for r and the A_i for each $\mathcal{N}_{t_{k-1}t_k}$) and deduce from this that $\theta_s\mu$ is independent of $\mathcal{F}_s^\mu$).

Show also that μ has stationary independent increments iff it has independent increments and for all choices of $s < t$ and $A_1, \dots, A_r \in \mathcal{E}$, the distribution of $\mathcal{N}_{st}$ depends on s and t through $t - s$ only.

The next result characterizes RCMs with independent increments as those having deterministic compensators. We shall denote deterministic compensating measures by Γ and note that Γ is a σ-finite positive measure on $\mathbb{R}_0 \times E$ such that $\Gamma(\{0\} \times E) = 0$, $\Gamma([0, t] \times E) < \infty$ for all $t \in \mathbb{R}_0$ and $\Gamma(\{t\} \times E) \le 1$ for all t. Below we use the notation

$$\overline{\gamma}(t) = \nu^*(]0, t]) = \Gamma([0, t] \times E), \quad \gamma(t; A) = \Gamma([0, t] \times A). \tag{6.13}$$

Especially important is the decomposition

$$\gamma(t; A) = \int_{]0,t]} \pi^*(s; A)\, \nu^*(ds), \tag{6.14}$$

with π^* a Markov kernel from $\mathbb{R}_0$ to E. (The notation $\nu^*(ds)$ is preferred to $\overline{\gamma}(ds)$ to indicate that ν^* indeed has something to do with a hazard measure, see Remark 6.2.2 below).

Theorem 6.2.1 (i) *μ has independent increments if and only if the $\mathcal{F}_t^\mu$-compensating measure Γ for μ is deterministic. If μ has independent increments, it holds in particular that $\mathbb{E}\overline{N}_t < \infty$ for all t and that $\gamma(t; A) = \mathbb{E}N_t(A)$ for all t and all $A \in \mathcal{E}$ where $\gamma(t; A)$ is as in (6.13).*

(ii) *μ has stationary independent increments if and only if μ is a homogeneous Poisson random measure.*

Note. By a deterministic compensating measure we mean a non-random measure, i.e., the measure $\Gamma(\cdot, \omega)$ on $\mathbb{R}_0 \times E$ does not depend on $\omega \in \Omega$. Among the homogeneous Poisson measures in (ii) we include the dead measure $\mu \equiv 0$, which has the 0-measure as compensating measure.

Proof. (i) Suppose that μ has independent increments. By Proposition 6.2.3(i) below it then follows that $\mathbb{E}\overline{N}_t < \infty$ for all t, and so $\gamma(t; A) = \mathbb{E}N_t(A) < \infty$ for all t and all $A \in \mathcal{E}$. But then, for $s < t$,

$$\mathbb{E}\left[N_t(A) - N_s(A) \,\middle|\mathcal{F}_s^{\mu}\right] = \mathbb{E}\left[N_t(A) - N_s(A)\right] = \gamma(t; A) - \gamma(s; A) \qquad (6.15)$$

and it follows directly that $M(A) = (M_t(A))_{t\geq 0}$ given by

$$M_t(A) = N_t(A) - \gamma(t; A)$$

is an $\mathcal{F}_t^{\mu}$-martingale, and therefore, by Theorem 4.5.2, that Γ is given by

$$\Gamma\left([0, t] \times A\right) = \gamma(t; A),$$

in particular Γ is deterministic. (One must of course argue that $A \mapsto \gamma(t; A)$ is a measure, but that is obvious from the definition $\gamma(t; A) = \mathbb{E}N_t(A)$ and monotone convergence).

If conversely μ has deterministic $\mathcal{F}_t^{\mu}$-compensating measure Γ, since always $\overline{\Lambda}_t < \infty$ for all t a.s., defining $\overline{\gamma}(t) := \mathbb{E}\overline{\Lambda}_t$ we have $\overline{\gamma}(t) = \overline{\Lambda}_t < \infty$ for all t and consequently, for every t, $\mathbb{E}\overline{N}_t = \overline{\gamma}(t) < \infty$. The function $t \mapsto \overline{\gamma}(t)$ is right-continuous and increasing with $\overline{\gamma}(0) = 0$, hence may be identified with a positive measure ν^* on $\mathbb{R}_+$ through the relation $\nu^*\left(]0, t]\right) = \overline{\gamma}(t)$. But then there is also a Markov kernel π^* from $\mathbb{R}_0$ to E such that (6.14) holds for all t, all $A \in \mathcal{E}$, and from this one reads off the Markov kernels $P^{(n)}$ and $\pi^{(n)}$ generating the distribution of μ, using the recipe from (4.47) and (4.48). By the basic properties of compensating measures, $\Delta\nu^*(t) \leq 1$ always and the set $\mathrm{A}^* = \{t \in \mathbb{R}_0 : \Delta\nu^*(t) = 1\}$ is at most countably infinite with, since $\nu^*\left(]0, t]\right) < \infty$ for all t, only finitely many elements inside any interval $]0, t]$. Thus we may label the elements of the set A^* in increasing order, $0 < a_1 < a_2 < \cdots$ and it is then first seen that $P^{(0)}$ has hazard measure ν^* with a possibly finite termination point $t^\dagger$ (in which case necessarily $t^\dagger = a_1$). Next we deduce that $P_{z_n}^{(n)}$ has a hazard measure $\nu_{z_n}^{(n)}$ which is the restriction to $]t_n, \infty[$ of ν^*, i.e., if $t_n < \infty$,

$$P_{z_n}^{(n)} = P^{*j_n}\left(\cdot\,|]t_n, \infty]\right)$$

(see Proposition 4.1.3), where j_n is chosen so that $a_{j_n} \leq t_n < a_{j_n+1}$, and P^{*j} is the probability on $\left]a_j, \infty\right]$ with a hazard measure that is the restriction of ν^* to $\left]a_j, \infty\right[$, in particular $P^{*0} = P^{(0)}$. From (4.48) and (6.14) one finds that $\pi_{z_n,t}^{(n)} = \pi^*\left(t; \cdot\right)$ and so, by Lemma 4.3.3, for any s, the conditional distribution of θ_s given $\overline{N}_s = k$, $Z_k = z_k$ is determined by the Markov kernels

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

where $a_{j(s)} \leq s < a_{j(s)+1}$, and for $n \geq 1$

$$\widetilde{P}_{\widetilde{z}_n|k,z_k}^{(n)} = P^{*\widetilde{j}_n}\left(\cdot\,\middle|\left]\widetilde{t}_n, \infty\right]\right),$$

where $a_{\tilde{j}_n} \leq \widetilde{t}_n < a_{\tilde{j}_n+1}$, and finally for $n \geq 0$

$$\widetilde{\pi}^{(n)}_{\widetilde{z}_n,t|k,z_k} = \pi^* (t; \cdot) .$$

None of these kernels depend on k or z_k, and we have therefore shown that μ has independent increments. (The arguments above simplify if you assume that $\Delta\nu^* (t) < 1$ for all t: then ν^* is the hazard measure for a probability P^* on $\overline{\mathbb{R}}_+$ with termination point ∞, and all the fuss about j_n and $j(s)$ can be ignored).

(ii) If μ has stationary independent increments, with $\gamma(t; A) = \mathbb{E}N_t(A)$ as in the proof of (i), the stationarity implies that for $s \leq t$, $A \in \mathcal{E}$,

$$\gamma(t; A) - \gamma(s; A) = \widetilde{\rho}\,(t - s; A)$$

for some function $\widetilde{\rho}$. Since $\gamma(0; A) = 0$, taking $s = 0$ gives $\widetilde{\rho}(t; A) = \gamma(t; A)$, and the linearity property

$$\widetilde{\rho}(u + v; A) = \widetilde{\rho}(u; A) + \widetilde{\rho}(v; A) \quad (u, v \geq 0)$$

with $\widetilde{\rho}(0; A) = 0$ follows. Hence, since $t \mapsto \widetilde{\rho}(t; A)$ is increasing and right-continuous,

$$\widetilde{\rho}(t; A) = t\rho(A)$$

for some constant $\rho(A)$, with $A \mapsto \rho(A)$ a measure (cf. the argument in (i) above). Thus $M_t(A) = N_t(A) - t\rho(A)$ defines an $\mathcal{F}^\mu_t$-martingale, therefore μ has compensating measure $\ell \otimes \rho$ and consequently μ is a homogeneous Poisson random measure; see Example 4.3.3.

If conversely μ is a homogeneous Poisson random measure, it was shown in Example 4.3.5 that μ has stationary independent increments; see also Proposition 4.7.2.

□

Remark 6.2.1 Theorem 6.2.1 applies in particular to a counting process N: (i) N has independent increments iff the $\mathcal{F}^N_t$-compensator γ for N is deterministic. If N has independent increments, then $\mathbb{E}N_t = \gamma\,(t) < \infty$ for all t. N has stationary independent increments iff N is a homogeneous Poisson process (including the dead process corresponding to the Poisson parameter being equal to 0).

Remark 6.2.2 Suppose μ has independent increments, write $\nu^* \,(]0, t]) = \overline{\gamma}(t) = \mathbb{E}\overline{N}_t$ as in the proof of Theorem 6.2.1, and let $0 = a_0 < a_1 < a_2 < \cdots$ denote the atoms for ν^* (if any) such that $\Delta\nu^*(t) = 1$. As shown in the proof of Theorem 6.2.1, the Markov kernels generating the distribution of μ are then given by

$$P^{(n)}_{z_n} = P^{*j_n}\,(\cdot\, |]t_n, \infty])\,, \qquad \pi^{(n)}_{z_n,t} = \pi^*\,(t; \cdot)\,,$$

where j_n is determined so that $a_{j_n} \leq t_n < a_{j_n+1}$, P^{*j} is the probability on $\left]a_j, \infty\right]$ with hazard measure ν^* restricted to $\left]a_j, \infty\right[$, and π^* is given by the decomposition (6.14).

Exercise 6.2.2 Suppose μ has independent increments and let Γ be the deterministic $\mathcal{F}_t^\mu$-compensating measure for μ. Show that

$$\mathbb{E}\int_{\mathbb{R}_0\times E} f(t,y)\ \mu(dt,dy) = \int_{\mathbb{R}_0\times E} f(t,y)\ \Gamma(dt,dy)$$

for all measurable functions $f : \mathbb{R}_0 \times E \to \mathbb{R}_0$.

In the proof of Theorem 6.2.1 it was used in a crucial fashion that if μ has independent increments, then $\mathbb{E}\overline{N}_t < \infty$ for all t. This fact is part of Propostion 6.2.3 below, but before coming to that we give two important definitions.

Definition 6.2.2 (i) An $\mathbb{R}^d$-valued process X on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ has *independent increments with respect to* $(\mathcal{F}_t)$ if it is adapted, cadlag and for every $s < t$, $X_t - X_s$ is independent of $\mathcal{F}_s$.

(ii) An $\mathbb{R}^d$-valued process X on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ has *stationary, independent increments with respect to* $(\mathcal{F}_t)$ if it has independent increments and for every $s < t$, the distribution of $X_t - X_s$ depends on s, t through $t - s$ only.

A process with stationary independent increments is also called a *Lévy process.*

Below, when just writing that a process X has independent increments, we always mean that it has independent increments with respect to the filtration $\left(\mathcal{F}_t^X\right)$.

Remark 6.2.3 If X has independent increments with respect to $(\mathcal{F}_t)$, it also has independent increments with respect to $\left(\mathcal{F}_t^X\right)$. Note also, as will be used frequently below, if X has independent increments with respect to $(\mathcal{F}_t)$, then for every s, the process $(X_{s+u} - X_s)_{u\geq 0}$ is independent of $\mathcal{F}_s$ (cf. Exercise 6.2.1). If in addition the increments are stationary, then the distribution of this future increment process does not depend on s.

Definition 6.2.3 (i) An $\mathbb{N}_0$-valued process X on $(\Omega, \mathcal{F}, \mathbb{P})$ is a *Poisson process* if $X_0 = 0$ a.s. and provided X has independent increments with respect to $\left(\mathcal{F}_s^X\right)$ such that for every $s < t$, $X_t - X_s$ follows a Poisson distribution.

(ii) A counting process N on $(\Omega, \mathcal{F}, \mathbb{P})$ is a *curtailed Poisson process* if N has the same distribution as the process given by

$$\sum_{0<s\leq t} 1_{(\Delta X_s \geq 1)} + w^*(t) \quad (t \in \mathbb{R}_0), \tag{6.16}$$

where X is a Poisson process, $w^* \in W$ is a given counting process path, and X and w^* are compatible in the sense that for all t^* with $\Delta w^*(t^*) = 1$ it holds that

$$\mathbb{P}(\Delta X_{t^*} \geq 1) = 0. \tag{6.17}$$

Note. In this definition and everywhere below, the class of Poisson distributions includes the distribution degenerate at 0 (Poisson parameter 0).

Note. Recall (Example 3.1.2) that a homogeneous Poisson process is a counting process N such that $N_t - N_s$ for $s < t$ is independent of $\mathcal{F}_s^N$ and Poisson distributed with parameter $\lambda(t-s)$. Among all Poisson processes X, the homogeneous ones are characterized by having increments that are also stationary: for $s < t$, the distribution of $X_t - X_s$ depends on s, t through $t - s$ only; cf. Theorem 6.2.1(ii), Proposition 6.2.2 and Proposition 6.2.3.

Remark 6.2.4 If X is non-homogeneous Poisson, we have in particular that $\gamma_X(t) := \mathbb{E}X_t < \infty$ for all t. The distribution of X is completely determined by the *mean function* γ_X, in particular for $s < t$ the increment $X_t - X_s$ is Poisson distributed with parameter $\gamma_X(t) - \gamma_X(s)$. Note also that X is necessarily increasing, by definition cadlag and in fact a step process.

The sum in (6.16) counts the number of jumps for X on $[0, t]$. The contribution from w^* occurs at given time points $0 < a_1 < a_2 < \cdots$ (possibly none at all, finitely many or infinitely many, but only finitely many a_j in any finite time interval). Thus (6.17) is equivalent to the assertion that X and w^* never jump at the same time,

$$\mathbb{P} \bigcup_{t \geq 0, \Delta w^*(t)=1} (\Delta X_t \neq 0) = \mathbb{P} \bigcup_j \left(\Delta X_{a_j} \geq 1\right) = 0.$$

The term w^* in (6.16) describes the time points a_j at which N is certain to jump. These deterministic jump times cannot be accounted for by the Poisson process X: it is a consequence of Proposition 6.2.2 below that if X is Poisson, then $\mathbb{P}(\Delta X_t = 0) > 0$ for all t.

If the Poisson process X is itself a counting process (e.g., if X is homogeneous Poisson), X is also curtailed Poisson: (6.16) holds with $w^* \equiv 0$ and the sum of jumps for X equals X itself.

It is easy to see that if X is non-homogeneous Poisson, then the mean function γ_X is increasing and right-continuous with $\gamma_X(0) = 0$: but any such function is a mean function.

Proposition 6.2.2 *Let $\gamma : \mathbb{R}_0 \to \mathbb{R}_0$ be increasing and right-continuous with $\gamma(0) = 0$.*

(i) *There exists a Poisson process X with mean function γ.*
(ii) *X is a counting process if and only if γ is continuous.*
(iii) *If for a given $t > 0$, $\Delta\gamma(t) > 0$, then ΔX_t follows a Poisson distribution with parameter $\Delta\gamma(t)$.*
(iv) *If for a given $t > 0$, $\Delta\gamma(t) = 0$, then $\mathbb{P}(\Delta X_t = 0) = 1$.*

Proof. Let N be a homogeneous Poisson process with rate 1 (mean function t). Defining

$$X_t = N_{\gamma(t)} \tag{6.18}$$

for $t \geq 0$, it is clear that $X_0 \equiv 0$ and that X is $\mathbb{N}_0$-valued and cadlag. And for $s < t$, $X_t - X_s = N_{\gamma(t)} - N_{\gamma(s)}$ is independent of $\mathcal{F}^N_{\gamma(s)} \supset \mathcal{F}^X_s$ and follows a Poisson distribution with parameter $\gamma(t) - \gamma(s)$.

For a given $t > 0$,

$$\Delta X_t = \begin{cases} 0 & \text{if } \gamma(t-\epsilon) = \gamma(t) \text{ for some } 0 < \epsilon < t, \\ N_{\gamma(t)} - N_{\gamma(t-)-} & \text{if } \gamma(t-\epsilon) < \gamma(t) \text{ for all } 0 < \epsilon < t. \end{cases}$$

It follows that if γ is continuous at t, then $\Delta X_t \le \Delta N_{\gamma(t)} \le 1$, in particular X is a counting process if γ is everywhere continuous. Next, for any t, since $\Delta X_t = \lim_{h\downarrow 0, h>0} \left(N_{\gamma(t)} - N_{\gamma(t-h)}\right)$ it follows that ΔX_t is the limit of Poisson variables with parameters $\gamma(t) - \gamma(t-h)$, hence itself Poisson with parameter $\lim(\gamma(t) - \gamma(t-h)) = \Delta\gamma(t)$. In particular, if $\Delta\gamma(t) = 0$, then $\Delta X_t = 0$ a.s., while if $\Delta\gamma(t) > 0$, then ΔX_t is non-degenerate Poisson so that $\mathbb{P}(\Delta X_t \ge 2) > 0$ and X is not a counting process. $\square$

A further discussion of non-homogeneous and curtailed Poisson processes is contained in the next result and its proof.

Proposition 6.2.3 *Suppose N is a counting process with independent increments. Then*

(i) $\gamma(t) := \mathbb{E}N_t < \infty$ *for all t and γ is the deterministic $\mathcal{F}_t^N$-compensator for N.*
(ii) *If γ is continuous, N is Poisson with mean function γ, in particular for $s < t$, $N_t - N_s$ is Poisson distributed with parameter $\gamma(t) - \gamma(s)$.*
(iii) *If γ has discontinuities, N is a curtailed Poisson process such that in the representation (6.16), w^* jumps precisely at the times a where* $\Delta\gamma(a) = 1$, *and the Poisson process X has mean function*

$$\gamma_X(t) = \gamma^c(t) - \sum_{0<s\le t, \Delta\gamma(s)<1} \log(1 - \Delta\gamma(s)), \tag{6.19}$$

γ^c denoting the continuous part of γ, $\gamma^c(t) = \gamma(t) - \sum_{0<s\le t} \Delta\gamma(s)$.

Proof. (i) Let $P^{(n)}$ denote the Markov kernels for the distribution of N, and let $t^\dagger$ be the termination point for $P^{(0)}$. For $s < t < t^\dagger$, by the independent increment assumption

$$\begin{aligned} \mathbb{P}(N_t - N_s = 0) &= \mathbb{P}(N_t - N_s = 0 \mid N_s = 0) \\ &= \mathbb{P}(T_1 > t \mid T_1 > s) \\ &= \frac{\overline{P}^{(0)}(t)}{\overline{P}^{(0)}(s)}. \end{aligned}$$

But then also, conditionally on $N_s = k$, $Z_k = z_k \in \mathrm{K}^{(k)}$ with $t_k \le s$, by Lemma 4.3.3

$$\begin{aligned} \frac{\overline{P}^{(0)}(t)}{\overline{P}^{(0)}(s)} &= \mathbb{P}(N_t - N_s = 0 \mid N_s = k, Z_k = z_k) \\ &= \frac{\overline{P}_{z_k}^{(k)}(t)}{\overline{P}_{z_k}^{(k)}(s)}. \end{aligned}$$

Since $\overline{P}^{(k)}_{z_k}(t_k) = 1$ this forces

$$\overline{P}^{(k)}_{z_k}(t) = \frac{\overline{P}^{(0)}(t)}{\overline{P}^{(0)}(t_k)} \tag{6.20}$$

for all $t < t^\dagger$, all k and almost all possible values of z_k of Z_k with $t_k \leq t$.

Let $\nu^{(0)}$ be the hazard measure for $P^{(0)}$. We claim that if $t^\dagger < \infty$, then

$$\nu^{(0)}\left(\left]0, t^\dagger\right[\right) < \infty$$

(implying if $t^\dagger < \infty$ that $\Delta P^{(0)}\left(t^\dagger\right) > 0$ and $\Delta\nu^{(0)}\left(t^\dagger\right) = 1$, see Theorem 4.1.1(iii)): suppose that $t^\dagger < \infty$ and $\nu^{(0)}\left(\left]0, t^\dagger\right[\right) = \infty$, in particular $\Delta P^{(0)}\left(t^\dagger\right) = 0$. Using (6.20) it is clear that if $T_k < t^\dagger$, so is T_{k+1}, and since $\mathbb{P}\left(T_1 < t^\dagger\right) = 1$ it follows that N explodes on $\left[0, t^\dagger\right[$ which is a contradiction.

Next, suppose first that $t^\dagger = \infty$ and fix $t \in \mathbb{R}_0$. We then have $\overline{P}^{(0)}(t) = p > 0$ and find from (6.20) that

$$\begin{aligned}
\mathbb{P}\left(T_{k+1} \leq t\right) &= \mathbb{E}\left[1 - \overline{P}^{(k)}_{Z_k}(t)\,;\, T_k \leq t\right] \\
&= \mathbb{E}\left[1 - \frac{\overline{P}^{(0)}(t)}{\overline{P}^{(0)}(T_k)}\,;\, T_k \leq t\right] \\
&\leq (1-p)\,\mathbb{P}\left(T_k \leq t\right).
\end{aligned}$$

Thus $\mathbb{P}\left(N_t \geq k\right) = \mathbb{P}\left(T_k \leq t\right) \leq (1-p)^k$ and $\mathbb{E}N_t = \sum_{k=1}^{\infty} \mathbb{P}\left(N_t \geq k\right) < \infty$ follows.

If $t^\dagger < \infty$, since $\nu^{(0)}\left(\left]0, t^\dagger\right]\right) < \infty$ and $\Delta P^{(0)}\left(t^\dagger\right) > 0$, $\Delta\nu^{(0)}\left(t^\dagger\right) = 1$ as argued above, by similar reasoning

$$\mathbb{P}\left(T_{k+1} < t\right) \leq \left(1 - \Delta P^{(0)}\left(t^\dagger\right)\right)\mathbb{P}\left(T_k < t\right)$$

and since $\mathbb{P}\left(T_k < t^\dagger\right) = \mathbb{P}\left(N_{t^\dagger-} \geq k\right)$ it follows that $\mathbb{E}N_{t^\dagger-} < \infty$, and hence also (a jump at $t^\dagger$ is bound to happen), $\mathbb{E}N_{t^\dagger} = \mathbb{E}N_{t^\dagger-} + 1 < \infty$. Proceeding beyond $t^\dagger$, the independent increments property now yields a probability P^* on $\left]t^\dagger, \infty\right]$ (determined from any $P^{(n)}_{z_n}$ with $t_n = t^\dagger$) with termination point t^* such that (6.20) holds with $\overline{P}^{(0)}$ replaced by $\overline{\mathrm{P}}^*$ whenever $t^\dagger \leq t_k < t^*$. Arguing as above one finds if $t^* = \infty$ that $\mathbb{E}N_t - \mathbb{E}N_{t^\dagger} < \infty$ for $t > t^\dagger$ so that $\mathbb{E}N_t < \infty$ for all t, or if $t^* < \infty$ that $\mathbb{E}N_{t^*} - \mathbb{E}N_{t^\dagger} < \infty$ so that $\mathbb{E}N_{t^*} < \infty$, with a jump certain to happen at t^*. Continuing beyond t^* if necessary, it is clear that always $\mathbb{E}N_t < \infty$ for all t: the worst that can happen is that one gets an infinite sequence of successive finite termination points, but at each of them a jump is certain; hence since N is a counting process, in any interval $[0, t]$ there can be at most finitely many such points, and between any two of them that follow each other, the expected number of jumps for N is finite.

This completes the proof that $\gamma(t) = \mathbb{E}N_t < \infty$ for all t and therefore also the proof of Theorem 6.2.1. From that result it follows in particular that γ is the deterministic $\mathcal{F}_t^N$-compensator for N.

(ii) Follows immediately from Proposition 6.2.2(ii).

(iii) Let X be Poisson with mean function γ_X given by (6.19) and define

$$N_t = \sum_{0<s\le t} 1_{(\Delta X_s \ge 1)} + w^*(t),$$

cf. (6.16). By Proposition 6.2.2(iv) and the definition of γ_X, for any t such that $\Delta w^*(t) = 1$, we have $\Delta X_t = 0$ so that X does not jump at any of the at most countably infinite such t. Thus the compatibility requirement of Definition 6.2.3(ii) is fulfilled, and N is a counting process. With w^* deterministic, since X has independent increments with respect to $\left(\mathcal{F}_t^X\right)$, it follows (see Remark 6.2.3) that so has N. Also $\mathcal{F}_t^N \subset \mathcal{F}_t^X$, so N has independent increments with respect to $\left(\mathcal{F}_t^N\right)$ and by Theorem 6.2.1(i) it therefore remains only to show that $\mathbb{E}N_t = \gamma(t)$ for all t. But

$$\mathbb{E}\sum_{0<s\le t} 1_{(\Delta X_s \ge 1)} = \mathbb{E}\sum_{0<s\le t} 1_{(\Delta X_s = 1)} + \sum_{x=2}^{\infty}\mathbb{E}\sum_{0<s\le t} 1_{(\Delta X_s = x)}$$

and here, in the last sum for $x \ge 2$ only $s \in A^* = \{a : \Delta\gamma(a) > 0\}$ contribute (cf. Proposition 6.2.2(iii) and (iv)), hence

$$\begin{aligned}\mathbb{E}\sum_{0<s\le t} 1_{(\Delta X_s = x)} &= \sum_{a\in A^*, a\le t} \mathbb{P}(\Delta X_a = x)\\ &= \sum_{a\in A^*, a\le t} \frac{(\Delta\gamma_X(a))^x}{x!}e^{-\Delta\gamma_X(a)}\end{aligned}$$

for $x \ge 2$. Finally we may use

$$\begin{aligned}\gamma_X(t) &= \mathbb{E}X_t\\ &= \mathbb{E}\sum_{0<s\le t}\Delta X_s\\ &= \sum_{x=1}^{\infty}\mathbb{E}\sum_{0<s\le t} x1_{(\Delta X_s = x)}\\ &= \mathbb{E}\sum_{0<s\le t} 1_{(\Delta X_s = 1)} + \sum_{x=2}^{\infty} x\mathbb{E}\sum_{0<s\le t} 1_{(\Delta X_s = x)}\\ &= \mathbb{E}\sum_{0<s\le t} 1_{(\Delta X_s = 1)} + \sum_{x=2}^{\infty} x\left(\sum_{a\in A^*, a\le t}\frac{(\Delta\gamma_X(a))^x}{x!}e^{-\Delta\gamma_X(a)}\right)\end{aligned}$$

to compute $\mathbb{E}\sum_{0<s\le t} 1_{(\Delta X_s = 1)}$. The end result is

$$\begin{aligned}\mathbb{E}N_t &= \gamma_X(t) + \sum_{x=2}^{\infty}(1-x)\sum_{a\in A^*, a\le t}\frac{(\Delta\gamma_X(a))^x}{x!}e^{-\Delta\gamma_X(a)} + w^*(t)\\ &= \gamma_X(t) + \sum_{a\in A^*, a\le t}\left(1 - \Delta\gamma_X(a) - e^{-\Delta\gamma_X(a)}\right) + w^*(t)\end{aligned}$$

$$
\begin{aligned}
&= \gamma_X^c(t) + \sum_{0<s\le t} \left(1 - e^{-\Delta\gamma_X(s)}\right) + w^*(t) \\
&= \gamma^c(t) + \sum_{0<s\le t,\Delta\gamma(s)<1} \Delta\gamma(s) + \sum_{0<s\le t,\Delta\gamma(s)=1} \Delta\gamma(s) \\
&= \gamma(t)
\end{aligned}
$$

as desired. □

Example 6.2.1 A counting process $N = (N_t)_{t\ge 0}$ is a *doubly stochastic Poisson process* if conditionally on $(U_t)_{t\ge 0}$, where (U_t) is some non-negative measurable process, it holds that N is Poisson with (conditional) mean function

$$\gamma_{|U}(t) = \int_0^t U_s\, ds.$$

Since $\gamma_{|U}(t)$ is continuous, Proposition 6.2.2 ensures that N is indeed a counting process.

If $(X_t)_{t\ge 0}$ is a homogeneous Markov chain on a finite state space and $U_t = f(X_t)$ for some $f \ge 0$, N is called a *Markov modulated Poisson process* .

The Cox process from Example 4.8.3 is a doubly stochastic Poisson process with $U_t \equiv U_0$ for all t.

The next result gives some important properties of RCMs with independent increments, known to be true by Theorem 6.2.1(ii) and Example 4.7.2 if the increments are also stationary.

Corollary 6.2.4 *Let μ be an RCM with independent increments and let Γ denote the deterministic $\mathcal{F}_t^\mu$-compensating measure for μ.*

(i) *The counting process $N(A)$ has independent increments for any $A \in \mathcal{E}$ with deterministic $\mathcal{F}_t^{N(A)}$-compensator $\gamma(t; A) = \Gamma([0,t] \times A)$. In particular, if $\gamma(\cdot; A)$ is continuous, $N(A)$ is a Poisson process with mean function $\gamma(\cdot; A)$, while if $\gamma(\cdot; A)$ has discontinuities, $N(A)$ is curtailed Poisson as described in Proposition 6.2.3(iii).*

(ii) *For any $r \ge 2$ and any $A_1, \dots, A_r \in \mathcal{E}$ pairwise disjoint it holds that the counting processes $N(A_1), \dots, N(A_r)$ are stochastically independent provided no two of the counting processes $\widetilde{N}(A_1), \dots, \widetilde{N}(A_r)$ jump simultaneously as required by (6.3), where the $\widetilde{N}(A_i)$ are defined to be independent with each $\widetilde{N}(A_i)$ having the same distribution as $N(A_i)$.*

Proof. Because μ has independent increments, it follows directly that so has any $N(A)$. Thus the $\mathcal{F}_t^{N(A)}$-compensator for $N(A)$ is $\gamma(\cdot; A)$, and (i) follows from Proposition 6.2.3(ii) and (iii). For the proof of (ii), note that since the A_j are disjoint, no two of the counting processes $N(A_j)$ jump simultaneously, hence the aggregate of the r counting processes is well defined as an RCM $\widetilde{\mu}$ with mark space

$\{1, \ldots, r\}$. Because μ has independent increments, so has $\widetilde{\mu}$ and the deterministic $\mathcal{F}_t^{\widetilde{\mu}}$-compensating measure $\widetilde{\Gamma}$ for $\widetilde{\mu}$ is determined by Γ in the obvious manner, i.e., $\widetilde{\Gamma}([0,t] \times \{j\}) = \gamma(t; A_j)$. Thus (6.10) holds and the independence of the $N(A_j)$ follows from Theorem 6.1.1 when using the assumption about the $\widetilde{N}(A_i)$. $\square$

Remark 6.2.5 The assumption about the independent counting processes $\widetilde{N}(A_i)$ made in Corollary 6.2.4(ii) is essential: suppose that there exists t^* and A, A' disjoint such that $\Delta\gamma(t^*; A) > 0$ and $\Delta\gamma\left(t^*; A'\right) > 0$. Then $N(A)$ and $N\left(A'\right)$ are not independent since $\Delta N_{t^*}(A)\,\Delta N_{t^*}\left(A'\right) \equiv 0$ but

$$\mathbb{P}(\Delta N_{t^*}(A) = 1)\,\mathbb{P}\left(\Delta N_{t^*}\left(A'\right) = 1\right) = \Delta\gamma\left(t^*; A\right)\Delta\gamma\left(t^*; A'\right) > 0.$$

(See Exercise 6.2.5(i) below for the formula for $\mathbb{P}(\Delta N_t = 1)$ used here when N is an arbitrary counting process with independent increments).

Exercise 6.2.3 Let X be a Poisson process with mean function γ_X. Define the RCM μ with mark space $\mathbb{N}$ by

$$N_t^x := \mu([0,t] \times \{x\}) = \sum_{0<s\le t} 1_{(\Delta X_s = x)}$$

for $t \ge 0$, $x \in \mathbb{N}$. In particular, since $X_0 = 0$,

$$X_t = \sum_{x=1}^{\infty} x N_t^x$$

so that μ determines X.

(i) For any s, show that $\theta_s\mu$ is determined by the future increments process $(X_t - X_s)_{t\ge s}$, and use this to show that μ has indpendent increments.
(ii) Find the deterministic $\mathcal{F}_t^{\mu}$-compensating measure for μ by showing that

$$\mathbb{E}N_t^x = \begin{cases} \gamma_X^c(t) + \sum_{0<s\le t} \Delta\gamma_X(s) e^{-\Delta\gamma_X(s)} & \text{if } x = 1, \\ \sum_{0<s\le t} \frac{1}{x!}\left(\Delta\gamma_X(s)\right)^x e^{-\Delta\gamma_X(s)} & \text{if } x \ge 2. \end{cases}$$

(iii) Show that the Markov kernels generating the distribution of μ are determined by

$$\overline{P}_{z_n}^{(n)}(t) = \exp\left(-\left(\gamma_X(t) - \gamma_X(t_n)\right)\right) \quad (t \ge t_n)$$

whether γ_X is continuous or not, and

$$\pi_{z_n,t}^{(n)}(\{x\}) = \frac{\left(\Delta\gamma_X(t)\right)^x e^{-\Delta\gamma_X(t)}}{x!\left(1 - e^{-\Delta\gamma_X(t)}\right)} \quad (x \ge 1).$$

(The π-value is defined by continuity if $\Delta\gamma_X(t) = 0$, and if this is the case it therefore equals 1 for $x = 1$ and equals 0 for $x \ge 2$).

(iv) Use Corollary 6.2.4 and Remark 6.2.5 to show that the counting processes $(N^x)_{x\in\mathbb{N}}$ are stochastically independent if γ_X is continuous, but not if γ_X has discontinuities.

Exercise 6.2.4 Let μ be an RCM with independent increments and assume that the total compensator, $\mathbb{E}\overline{N}_t$, is continuous as a function of t. Show that for any $C \in \mathcal{B}_0 \otimes \mathcal{E}$ it holds that $\mu(C)$ has a Poisson distribution $\Gamma(C)$, where Γ denotes the deterministic compensating measure for μ. Also show that if $r \geq 2$ and $C_1, \ldots, C_r \in \mathcal{B}_0 \otimes \mathcal{E}$ are pairwise disjoint, then $\mu(C_1), \ldots, \mu(C_r)$ are independent. (Hint: the assertions are true by Corollary 6.2.4 if the C_j (and C) are of the form $\left]s_j, t_j\right] \times A_j$. Now consider e.g., the class $\mathcal{C}$ of $C \in \mathcal{B}_0 \otimes \mathcal{E}$ such that $\mu(C)$ has a Poisson distribution, and show that $\mathcal{C}$ is a monotone class containing the algebra of sets that are finite disjoint unions of sets of the form $\left]s_j, t_j\right] \times A_j$).

The properties listed here for μ are those used for the definition of Poisson point processes on arbitrary spaces (in our case the space $\mathbb{R}_0 \times E$).

We shall now use the results obtained for RCMs with independent increment to derive some basic properties of processes with independent increments.

To begin with, let X denote an $\mathbb{R}^d$-valued process step process with independent increments with respect to the filtration $(\mathcal{F}_t)$; cf. Definition 6.2.2. As a step process X has only has finitely many jumps on finite intervals and we write T_n for the time of the nth jump and $Y_n = \Delta X_{T_n}$ for the size of the nth jump. Thus $((T_n), (Y_n))$ is an MPP with mark space $\mathbb{R}^d_{\backslash 0}$ (the notation used for $\mathbb{R}^d \backslash 0$, equipped with the matching Borel σ-algebra which we denote $\mathcal{B}^d_{\backslash 0}$), and we call the corresponding RCM μ, the RCM *associated with the jump sizes for the step process* X. Because X is a step process,

$$\begin{aligned} X_t &= X_0 + \sum_{0<s\leq t} \Delta X_s \\ &= X_0 + \sum_{n: T_n \leq t} Y_n \\ &= X_0 + \int_{]0,t]\times\mathbb{R}^d_{\backslash 0}} y\, \mu(ds, dy). \end{aligned} \tag{6.21}$$

Theorem 6.2.5 *Let X be an $\mathbb{R}^d$-valued step process and let μ be the RCM associated with the jump sizes for X.*

(i) *X has independent increments with respect to $\left(\mathcal{F}^X_t\right)$ if and only if μ is independent of X_0 and has independent increments.*
(ii) *X has stationary independent increments (is a Lévy process) with respect to $\left(\mathcal{F}^X_t\right)$ if and only if μ is independent of X_0 and is a homogeneous Poisson random measure.*

Proof. Note first that (6.21) shows that $\mathcal{F}^\mu_t = \mathcal{F}^{X-X_0}_t$ for all t. If X has independent increments the process $X - X_0$ is independent of X_0, hence μ is independent of X_0. Further, for any s the process $\widetilde{X} := (X_{s+u} - X_s)_{u\geq 0}$ is independent of $\mathcal{F}^X_s \supset \mathcal{F}^\mu_s$ and since

$$X_t - X_s = \sum_{n: s < T_n \le t} Y_n$$
$$= \int_{]s,t] \times \mathbb{R}^d_{\backslash 0}} y \, (\theta_s \mu) (du, dy), \tag{6.22}$$

it is seen that the jump times and marks for $\theta_s \mu$ are determined by $\widetilde{X}$ and it follows that μ has independent increments. If in addition X has stationary increments, the distribution of $\widetilde{X}$ will not depend on s, implying that μ has stationary increments and is therefore homogeneous Poisson by Theorem 6.2.1(ii).

If conversely μ is independent of X_0 and has independent increments, from (6.22) it follows that $X_t - X_s$ is independent of $\mathcal{F}_s^{\mu} = \mathcal{F}_s^{X - X_0}$ and of X_0, hence $X_t - X_s$ is independent of $\sigma\left(\mathcal{F}_s^{X-X_0}, X_0\right) = \mathcal{F}_s^X$ and X has independent increments with respect to $\left(\mathcal{F}_t^X\right)$. If in addition μ has stationary increments,

$$X_t - X_s = \int_{]0,t-s] \times \mathbb{R}^d_{\backslash 0}} y \, \left(\theta_s^* \mu\right) (du, dy)$$

so the distribution of $X_t - X_s$ depends on s, t through $t - s$ only. □

A Lévy process X which is a step process is also called a *compound Poisson process*. By Theorem 6.2.5(ii) and the definition of homogeneous Poisson random measures it follows that the jump sizes Y_n are independent and identically distributed and independent of the jump times T_n, which are jump times of a homogeneous Poisson process. One often represents a compound Poisson process as

$$X_t = X_0 + \sum_{n=1}^{N_t} U_n$$

where X_0, N and the sequence (U_n) are independent with N a homogeneous counting Poisson process and the U_n iid random variables with values in $\mathbb{R}^d_{\backslash 0}$.

Proposition 6.2.6 *Let X be an $\mathbb{R}^d$-valued step process with independent increments with respect to $\left(\mathcal{F}_t^X\right)$.*

(i) *If X is a Lévy process (compound Poisson process), then the characteristic function for $X_t - X_0$ is given by*

$$\mathbb{E} e^{i \langle u, X_t - X_0 \rangle} = \exp\left(-t \int_{\mathbb{R}^d_{\backslash 0}} \left(1 - e^{i \langle u, y \rangle}\right) \rho(dy)\right) \quad (u \in \mathbb{R}^d), \tag{6.23}$$

where ρ is the intensity measure for μ, the homogeneous Poisson random measure associated with the jump sizes for X.

(ii) *In general, if the RCM μ associated with the jump sizes for X has a deterministic compensating measure Γ, which is continuous, then for $s < t$,*

$$\mathbb{E}e^{i\langle u, X_t - X_s\rangle} = \exp\left(-\int_{]s,t]} \nu^*(dv) \int_{\mathbb{R}^d_{\setminus 0}} \pi^*(v; dy) \left(1 - e^{i\langle u, y\rangle}\right)\right) \quad (u \in \mathbb{R}^d),$$

where ν^* is the total continuous compensator and π^* is the Markov kernel generating the jump sizes as in (6.14).

Note. $\langle \cdot, \cdot \rangle$ is the usual inner product on $\mathbb{R}^d$, $\langle u, y\rangle = \sum_{j=1}^d u_j y_j$.

Proof. We focus on a detailed proof of (i). So let X be a compound Poisson process, and let ρ be the intensity measure for μ. For every $K \in \mathbb{N}$ and every $k = (k_1, \ldots, k_d) \in \mathbb{Z}^d$, define the d-dimensional interval

$$I_{k,K} = \left(\prod_{j=1}^d \left]\frac{k_j - 1}{2^K}, \frac{k_j}{2^K}\right]\right) \setminus 0$$

(omitting 0 is relevant only if $k_1 = \cdots = k_d = 1$), and verify that for any given $u \in \mathbb{R}^d$ (see (6.21)),

$$\begin{aligned}\langle u, X_t - X_0\rangle &= \int_{]0,t]\times\mathbb{R}^d_{\setminus 0}} \langle u, y\rangle\, \mu(ds, dy)\\ &= \lim_{K\to\infty} \sum_{k\in\mathbb{Z}^d} \left\langle u, \tfrac{k}{2^K}\right\rangle N_t\left(I_{k,K}\right)\end{aligned}$$

whence, using dominated convergence and the homogeneous Poisson structure of μ,

$$\begin{aligned}\mathbb{E}e^{i\langle u, X_t - X_0\rangle} &= \lim_{K\to\infty} \prod_{k\in\mathbb{Z}^d} \exp\left(-t\rho\left(I_{k,K}\right)\left(1 - e^{i\left\langle u, k2^{-K}\right\rangle}\right)\right)\\ &= \exp\left(-t\int_{\mathbb{R}^d_{\setminus 0}} \left(1 - e^{i\langle u, y\rangle}\right) \rho(dy)\right),\end{aligned}$$

(recall that if U has a Poisson distribution with parameter λ, then $\mathbb{E}e^{iU} = \exp\left(-\lambda\left(1 - e^{iu}\right)\right)$). The proof of (ii) is completely similar, only it should be remembered that since ν^* is continuous, by Theorem 6.2.5 and Corollary 6.2.4, $N_t\left(I_{k,K}\right)$ is Poisson with parameter $\int_{]0,t]} \pi_s^*\left(I_{k,K}\right)\, \nu^*(ds)$. □

The equation (6.23) is a special case of the famous Lévy–Khinchine formula for the characteristic function of an infinitely divisible distribution on $\mathbb{R}^d$. For $d = 1$ the equation reads

$$\mathbb{E}e^{iu(X_t - X_0)} = \exp\left(-t\int_{\mathbb{R}_{\setminus 0}} \left(1 - e^{iuy}\right) \rho(dy)\right) \quad (u \in \mathbb{R}), \tag{6.24}$$

and one recognizes the infinitesimal contribution

$$\exp\left(-t\left(1-e^{iuy}\right)\,\rho(dy)\right)$$

as the characteristic function of yU, where U is Poisson with parameter $t\rho\,(dy)$. For a more thorough discussion of infinite divisibility and processes with independent increments, see p.139 below.

Exercise 6.2.5 The formulas for characteristic functions given so far assumes that compensators are continuous. This exercise treats examples with discontinuous compensators.

(i) Let N be a counting process with independent increments, i.e., a curtailed Poisson process, and let γ denote the deterministic compensator for N. Show that N_t has characteristic function

$$\mathbb{E}e^{iuN_t} = e^{-\gamma^c(t)\left(1-e^{iu}\right)} \prod_{0<s\le t}\left(\Delta\gamma\,(s)\,e^{iu} + 1 - \Delta\gamma\,(s)\right).$$

(*Hint:* one possibility is to use that N is curtailed Poisson. Another is to argue directly that for any t, $\mathbb{E}\Delta N_t = \Delta\gamma\,(t)$, which since ΔN_t only takes the values 0 or 1, implies that $\mathbb{P}\,(\Delta N_t = 1) = \Delta\gamma\,(t)$).

(ii) Let X be an $\mathbb{R}^d$-valued step process with independent increments as in Proposition 6.2.6. Show that for $s < t$,

$$\mathbb{E}e^{i\langle u, X_t - X_s\rangle} = \exp\left(-\int_{]s,t]} \nu^{*c}(dv) \int_{\mathbb{R}^d_{\backslash 0}} \pi^*\,(v; dy)\,\left(1 - e^{i\langle u,y\rangle}\right)\right)$$
$$\times \prod_{s<v\le t}\left(\Delta\overline{\gamma}\,(v) \int_{\mathbb{R}^d_{\backslash 0}} \pi^*\,(v; dy)\; e^{i\langle u,y\rangle} + 1 - \Delta\overline{\gamma}\,(v)\right)$$

where ν^*, π^* are determined from the deterministic compensating measure Γ for μ as in (6.14), ν^{*c} is the continuous part of ν^* and $\overline{\gamma}\,(v) = \nu^*\,(]0, v])$.

Of all independent increment processes with jumps, the step processes are the simplest. Yet the structure of the jumps described in Theorem 6.2.5 pertains in great generality as we shall now see.

Let X be $\mathbb{R}^d$-valued with independent increments with respect to a filtration $(\mathcal{F}_t)$. In particular X is cadlag and adapted, $\mathcal{F}_t \supset \mathcal{F}_t^X$ for all t, and for every s, the process $\widetilde{X} = (X_{s+t} - X_s)_{t\ge 0}$ is independent of $\mathcal{F}_s$. Fix $\epsilon > 0$ and consider the MPP $(\mathcal{T}_\epsilon, \mathcal{Y}_\epsilon) = \left(\left(T_{n,\epsilon}\right), \left(Y_{n,\epsilon}\right)\right)$ comprising only the jump times and jump sizes for X such that $\|\Delta X_t\| > \epsilon$ (with $\|\cdot\|$ any standard norm on $\mathbb{R}^d$). The assumption that X be cadlag guarantees that there are only finitely many such jumps on any finite time interval, so $(\mathcal{T}_\epsilon, \mathcal{Y}_\epsilon)$ is a bona fide MPP with mark space $E_\epsilon = B_\epsilon^c$, where B_ϵ is the closed ball $\{x : \|x\| \le \epsilon\}$ in $\mathbb{R}^d$. Letting μ_ϵ denote the corresponding RCM and viewing μ_ϵ as a random measure on all of $\mathbb{R}_0 \times \mathbb{R}^d_{\backslash 0}$, it is clear that for $0 < \epsilon' < \epsilon$, μ_ϵ is the restriction to $\mathbb{R}_0 \times B_\epsilon^c$ of $\mu_{\epsilon'}$.

Let $s \in \mathbb{R}_0$. Since all jumps for $\theta_s \mu_\epsilon$ are determined by $\widetilde{X}$, it follows that $\theta_s \mu_\epsilon$ is independent of $\mathcal{F}_s \supset \mathcal{F}_s^X \supset \mathcal{F}_s^{\mu_\epsilon}$ and thus μ^ϵ has independent increments and μ^ϵ has a deterministic $\mathcal{F}_t^{\mu_\epsilon}$-compensating measure Γ_ϵ (with Γ_ϵ the restriction to $\mathbb{R}_0 \times B_\epsilon^c$ of $\Gamma_{\epsilon'}$ if $\epsilon' < \epsilon$). From Theorem 6.2.1 and Corollary 6.2.4 we now read off

Theorem 6.2.7 *Let X be a cadlag process with independent increments with respect to $(\mathcal{F}_t)$, and let for $\epsilon > 0$, μ_ϵ denote the RCM of jump sizes for X with $\|\Delta X_t\| > \epsilon$ as described above.*

(i) *μ_ϵ has independent increments.*
(ii) *If X is a Lévy process, then each μ_ϵ is a homogeneous Poisson random measure with some bounded intensity measure ρ_ϵ concentrated on B_ϵ^c, and where for $0 < \epsilon' < \epsilon$, ρ_ϵ is the restriction to B_ϵ^c of $\Gamma_{\epsilon'}$.*

The collection $(\mu_\epsilon)_{\epsilon>0}$ of RCMs describes all jump times and jump sizes for X, for instance through the object

$$\widetilde{\mu} := \mu_1\left(\cdot \cap \left(\mathbb{R}_0 \times B_1^c\right)\right) + \sum_{k=2}^{\infty} \mu_{1/k}\left(\cdot \cap \left(\mathbb{R}_0 \times B_{k-1} \backslash B_k\right)\right).$$

It may of course happen that X has only finitely many jumps on finite intervals, in which case $\widetilde{\mu}$ is an RCM and there is no need to first consider the μ_ϵ. But it may certainly also happen that the jumps for X are (with probability one) dense in $\mathbb{R}_0$, cf. Examples 6.2.2 and 6.2.3 below, and in that case the approximation through the μ_ϵ, discarding the dense set of jumps $\leq \epsilon$, is essential for the MPP description of the jump structure.

Consider again a step process X with independent increments with respect to $(\mathcal{F}_t)$, and let $g : \mathbb{R}_0 \to \mathbb{R}^d$ be a given continuous function. Then the process $X' := (X_t + g(t))_{t \geq 0}$ also has independent increments with respect to $(\mathcal{F}_t)$ in the sense of Definition 6.2.2 and X' and X share the same RCM μ of jump times and jump sizes. If X is a Lévy process, it is easy to see that X' is also a Lévy process iff $g(t) = t v_0$ for some $v_0 \in \mathbb{R}^d_{\backslash 0}$. (Of course, for any g it holds that $X'_t - X'_s$ is independent of $\mathcal{F}_s$ whenever $s < t$, and if g is cadlag, X' is cadlag and the requirements of Definition 6.2.2 are satisfied. But if g has discontinuities, the jump structure for X' includes the deterministic jumps for g and need not (because the jumps for g may e.g., be dense in $\mathbb{R}_0$) be characterized by an RCM). Note that X' has the property that the process $X' - X'_0$ is $\mathcal{F}_t^\mu$-adapted. It may be shown that if X has independent increments with respect to $(\mathcal{F}_t)$ and has only finitely many jumps on finite intervals with μ the RCM of jump times and jump sizes for X, then if the process $X - X_0$ is $\mathcal{F}_t^\mu$-adapted (so $X_t - X_0$ has the form $f_{Z_{\langle t \rangle}}^{\langle t \rangle}(t)$, see Proposition 4.2.1) necessarily $X_t = X_t^{\text{step}} + g(t)$ for some continuous g, where

$$X_t^{\text{step}} = X_0 + \sum_{0 < s \leq t} \Delta X_s$$

is the step process determined by the jumps for X.

We conclude by recording some of the standard facts about general processes X with independent increments. Let for $0 \le s \le t$, α_{st} denote the distribution of $X_t - X_s$. If $0 \le s \le t \le u$, from

$$X_u - X_s = (X_t - X_s) + (X_u - X_t)$$

and the fact that X is right-continuous, it follows that the probability measures α_{st} form a two-parameter, weakly right-continuous convolution semigroup,

$$\alpha_{su} = \alpha_{st} * \alpha_{tu} \quad (0 \le s \le t \le u), \tag{6.25}$$

$$\alpha_{su} \xrightarrow{w} \alpha_{st} \quad (0 \le s \le t \le u, u \downarrow t)$$

with all $\alpha_{ss} = \varepsilon_0$. If X is a Lévy process, $\alpha_{st} = \alpha_{t-s}$ depends on $t - s$ only, and the α_t form a one-parameter, weakly continuous convolution semigroup,

$$\alpha_{s+t} = \alpha_s * \alpha_t \quad (s, t \ge 0) \tag{6.26}$$

$$\alpha_s \xrightarrow{w} \alpha_t \quad (s, t \ge 0, s \to t)$$

with $\alpha_0 = \varepsilon_0$.

For Lévy processes X on $\mathbb{R}$, the classical *Lévy–Khinchine formula* describes the possible structure for the characteristic functions of the α_t,

$$\int_{\mathbb{R}} e^{iux}\, \alpha_t(dx) = \exp\left(-\frac{\sigma^2}{2}tu + i\beta tu - t\int_{\mathbb{R}_{\backslash 0}} \left(1 - e^{iuy} + iuy 1_{[-1,1]}(y)\right)\rho(dy)\right), \tag{6.27}$$

for $u \in \mathbb{R}$, where $\sigma^2 \ge 0$, $\beta \in \mathbb{R}$ and ρ is the *Lévy measure,* i.e., a positive measure on $\mathbb{R}_{\backslash 0}$ such that for all $\epsilon > 0$

$$\rho\left([-\epsilon, \epsilon]^c\right) < \infty,$$
$$\int_{[-\epsilon,\epsilon]\backslash 0} y^2\, \rho(dy) < \infty.$$

In (6.27) the contribution $-\frac{\sigma^2}{2}tu$ to the exponential corresponds to a standard Brownian motion multiplied by σ, the term $i\beta tu$ corresponds to the deterministic linear drift $t \mapsto \beta t$, and the integral describes a compensated sum of jumps. Note that if $\int_{[-\epsilon,\epsilon]\backslash 0} |y|\, \rho(dy) < \infty$, the last term in the integral may be incorporated into the drift term, leaving the contribution

$$\exp\left(-t\int_{\mathbb{R}_{\backslash 0}} \left(1 - e^{iuy}\right)\rho(dy)\right)$$

which is the same as the expression from (6.24), only now $\rho\left(\mathbb{R}_{\setminus 0}\right) = \infty$ is possible. This establishes the fact that for any Lévy process X on $\mathbb{R}$, if the characteristic function for the increments $X_{s+t} - X_s$ is given by (6.27), then for any $\epsilon > 0$, the intensity measure ρ_ϵ for the homogeneous Poisson random measure μ_ϵ of jumps $> \epsilon$ in absolute value of size is the restriction to $[-\epsilon, \epsilon]^c$ of the Lévy measure ρ.

It is important to be able to extract the Lévy measure from the distributions α_t: it can be shown that for all bounded and continuous functions $f : \mathbb{R}_{\setminus 0} \to 0$ with support bounded away from 0 (i.e., there exists $\epsilon > 0$ such that $f \equiv 0$ on $B_\epsilon \setminus 0$), then

$$\int_{\mathbb{R}_{\setminus 0}} f(y)\,\rho(dy) = \lim_{t \to 0} \tfrac{1}{t} \int_{\mathbb{R}} f(x)\,\alpha_t(dx), \tag{6.28}$$

(for an incomplete argument, see Exercise 6.2.6 below).

Example 6.2.2 Let α_t be the Γ-distribution with scale parameter $\beta > 0$ and shape parameter λt, where $\lambda > 0$, i.e.,

$$\alpha_t(dx) = \frac{1}{\beta^{\lambda t}\Gamma(\lambda t)} x^{\lambda t - 1} e^{-\frac{1}{\beta}x}\, dx$$

for $x > 0$. The α_t form a weakly continuous convolution semigroup and determine a Lévy process X, the Γ*-process,* which is obviously increasing. Since

$$\frac{1}{t\beta^{\lambda t}\Gamma(\lambda t)} x^{\lambda t-1} e^{-\frac{1}{\beta}x} = \frac{\lambda}{\beta^{\lambda t}\Gamma(\lambda t + 1)} x^{\lambda t - 1} e^{-\frac{1}{\beta}x} \to \frac{\lambda}{x} e^{-\frac{1}{\beta}x}$$

as $t \to 0$, it follows from (6.28) that X has Lévy measure

$$\rho(dy) = \frac{\lambda}{y} e^{-\frac{1}{\beta}y}\, dy$$

for $x > 0$. The fact that $\rho(\mathbb{R}_+) = \infty$ implies that (with probability one) the Γ-process has jumps that are dense on $\mathbb{R}_0$. It may be shown that X is built entirely from its jumps, $X_t = X_0 + \sum_{0<s\le t} \Delta X_s$, e.g., by identifying the characteristic function for the Γ-distribution of $X_t - X_0$ with that of $\sum_{0<s\le t} \Delta X_s = \lim_{\epsilon \downarrow 0} \int_{]0,t]\times]\epsilon,\infty[} y\, \mu_\epsilon(ds, dy)$ which is easily found from (6.24).

Example 6.2.3 Let α_t be the Cauchy distribution with scale parameter βt where $\beta > 0$, i.e.,

$$\alpha_t(dx) = \frac{1}{\pi\beta t\left(1 + \frac{1}{\beta^2 t^2}x^2\right)}\, dx$$

for $x \in \mathbb{R}$.

The α_t form a weakly continuous convolution semigroup and determine a Lévy process X, the *Cauchy process.* Since

$$\frac{1}{t\pi\beta t\left(1 + \frac{1}{\beta^2 t^2}x^2\right)} \to \frac{\beta}{\pi x^2}$$

as $t \to 0$ we conclude that the Cauchy process has Lévy measure

$$\rho\,(dy) = \frac{\beta}{\pi y^2}\,dy.$$

As in the previous example, ρ is unbounded, $\rho\left(\mathbb{R}_{\backslash 0}\right) = \infty$, and the jumps are dense on $\mathbb{R}_0$. Note that $\int_{[-\epsilon,\epsilon]\backslash 0} |y|\;\rho\,(dy) = \infty$ for any $\epsilon > 0$ implying that in the Lévy–Khinchine representation (6.27) of α_t it is necessary to include the term $iuy1_{[-1,1]}\,(y)$ in the integral. Probabilistically this means that the sum of the absolute values of the small jumps diverges, $\lim_{\epsilon\downarrow 0}\int_{]0,t]\times]\epsilon,1]} |y|\;\mu_\epsilon\,(ds, dy) = \infty$ a.s.

Exercise 6.2.6 Let X be a cadlag, $\mathbb{R}$-valued process, and let $f : \mathbb{R}_{\backslash 0} \to \mathbb{R}$ be bounded and continuous with support $\subset [-\epsilon_0, \epsilon_0]^c$ for some $\epsilon_0 > 0$. Let $0 < \epsilon < \epsilon_0$ and let μ_ϵ denote the RCM of jump times corresponding to jump sizes $> \epsilon$ in absolute value with $\left(\left(T_{n,\epsilon}\right), \left(Y_{n,\epsilon}\right)\right)$ the corresponding MPP. Show that

$$\begin{aligned}\lim_{K\to\infty}\sum_{k=1}^{2^K} f\left(X_{k/2^K} - X_{(k-1)/2^K}\right) &= \int_{]0,1]\times[-\epsilon,\epsilon]^c} f(y)\,\mu_\epsilon\,(ds, dy) \qquad (6.29)\\ &= \sum_{n:T_{n,\epsilon}\le 1} f\left(Y_{n,\epsilon}\right).\end{aligned}$$

(*Hints:* argue this for the random variables evaluated at an arbitrary $\omega \in \Omega$. Show that for K sufficiently large, each $T_{n,\epsilon}(\omega)$ is placed in its own interval $\left](k-1)/2^K, k/2^K\right]$, and that the contributions to the sum in (6.29) for these k-values converge to the expression on the right. Next, use the cadlag property of X to show that with k' denoting an arbitrary of the remaining k-values, it holds that, for K sufficiently large, $\max_{k'}\left|X_{k'/2^K}\,(\omega) - X_{(k'-1)/2^K}\,(\omega)\right| < \epsilon_0$).

If X is a Lévy process, taking expectations in (6.29) one finds, *provided the operations of taking limit and expectation can be interchanged*, that

$$\lim_{K\to\infty} 2^K\mathbb{E}f\left(X_{1/2^K} - X_0\right) = \mathbb{E}\int_{]0,1]\times[-\epsilon,\epsilon]^c} f(y)\,\mu_\epsilon\,(ds, dy).$$

Show that this identity is the same as

$$\lim_{K\to\infty} 2^K\int f(x)\,\alpha_{1/2^K}(dx) = \int f(y)\,\rho_\epsilon\,(dy),$$

which at least may indicate the truth of (6.28).

A final comment on cadlag processes with independent increments: we assumed above that the processes had values in $\mathbb{R}^d$. It should be fairly clear, however, that a result like Theorem 6.2.7 remains valued if the processes take their values in, say, an infinite-dimensional Banach space.

7

Piecewise Deterministic Markov Processes

This chapter contains the basic theory for piecewise deterministic Markov processes, whether homogeneous or not, based exclusively on the theory of marked point processes from the previous chapters and presented through the device of viewing a PDMP as a process adapted to the filtration generated by an RCM. The strong Markov property is established, various versions of Itô's formula for PDMPs are given, the socalled full infinitesimal generator for a homogeneous PDMP is discussed, invariant measures are treated, and the chapter concludes with a section on likelihood processes for PDMPs.

At a first reading one may omit Sections 7.5, 7.7, 7.8 and the last part of Section 7.9 (dealing with multiplicative functionals).

References. The standard reference for homogeneous PDMPs is the book [30] by Mark Davis, but the little known earlier dissertation by Wobst [122] should also be mentioned.

7.1 Markov processes

In this chapter we discuss Markov processes — homogeneous or not — that can be described completely in terms of an initial state and an MPP, i.e., processes that are Markovian and piecewise deterministic, see p.25. First some definitions and remarks concerning Markov processes in general.

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 $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ with values in a state space $(G, \mathcal{G})$ which, since it will serve as a mark space for an RCM, we shall assume to be a Borel space (see p. 11).

Definition 7.1.1 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 $p_{st}(\cdot, \cdot)$ on G such that

$$\mathbb{P}(X_t \in C \mid \mathcal{F}_s) = p_{st}(X_s, C) \quad (C \in \mathcal{G}). \tag{7.1}$$

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 (7.1) does not determine $p_{st}(x, \cdot)$ uniquely for all $x \in G$, but because $\mathcal{G}$ is countably generated it does hold that if $p_{st},\ p'_{st}$ are both transition probabilities, then $\mathbb{P}(X_s \in C_{st}) = 1$, where $C_{st} = \{x \in G | p_{st}(x, \cdot) = p'_{st}(x, \cdot)\}$. 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*.

Example 7.1.1 An $\mathbb{R}^d$-valued process X with independent increments with respect to $(\mathcal{F}_t)$, is a Markov process with transition probabilities

$$p_{st}(x, B) = \alpha_{st}(B - x) \quad (s \leq t\ ,\ B \in \mathcal{B}^d), \tag{7.2}$$

where α_{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

$$p_t(x, B) = \alpha_t(B - x), \tag{7.3}$$

where α_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 $s \leq t \leq u,\ x \in G,\ C \in \mathcal{G}$,

$$p_{su}(x, C) = \int_G p_{st}(x, dy)\, p_{tu}(y, C). \tag{7.4}$$

or, in the homogeneous case, if for all $s, t \geq 0,\ x \in G,\ C \in \mathcal{G}$,

$$p_{s+t}(x, C) = \int_G p_s(x, dy)\, p_t(y, C). \tag{7.5}$$

It is essential to note that e.g., (7.4) holds almost surely in the following sense: for $s \leq t \leq u,\ C \in \mathcal{G}$,

$$\begin{aligned} p_{su}(X_s, C) &= \mathbb{P}(X_u \in C | \mathcal{F}_s) \\ &= \mathbb{E}(\mathbb{P}(X_u \in C | \mathcal{F}_t) | \mathcal{F}_s) \\ &= \mathbb{E}(p_{tu}(X_t, C) | \mathcal{F}_s) \\ &= \int_G p_{st}(X_s, dy)\, p_{tu}(y, C), \end{aligned}$$

$\mathbb{P}$-a.s. Thus, given $s \leq t \leq u,\ C \in \mathcal{G}$, (7.4) holds for $X_s(\mathbb{P})$-almost all x.

Example 7.1.2 If X has independent increments as in Example 7.1.1, then (7.4) or, if X has stationary increments, (7.5) holds and is equivalent to the convolution property (6.25), resp. (6.26).

The *transition operators* for the Markov process are operators on the space of bounded measurable functions $f : G \to \mathbb{R}$, defined by

$$\mathrm{P}_{st} f(x) = \int_G p_{st}(x, dy)\ f(y)$$

for $0 \le s \le t$, and in the homogeneous case,

$$\mathrm{P}_t f(x) = \int_G p_t(x, dy)\ f(y)$$

for $t \ge 0$. The Chapman–Kolmogorov equations translate into the semigroup property,

$$\mathrm{P}_{su} = \mathrm{P}_{st}\mathrm{P}_{tu}, \quad \mathrm{P}_{tt} = \mathrm{id}$$

for $0 \le s \le t \le u$ with id the identity operator, and in the homogeneous case

$$\mathrm{P}_{s+t} = \mathrm{P}_s\mathrm{P}_t, \quad \mathrm{P}_0 = \mathrm{id}$$

for $s, t \ge 0$.

We mention two important facts about general Markov processes.

Let $\nu_0 = X_0(\mathbb{P})$ denote the distribution of X_0, the *initial distribution* for X. 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

$$\begin{aligned}\mathbb{P}(X_t \in C) &= \mathbb{E}\left[\mathbb{P}(X_t \in C \mid X_0)\right] \\ &= \int_G \nu_0(dx)\, p_{0t}(x, C)\end{aligned}$$

and that for $n \ge 2,\ 0 \le t_1 < \cdots < t_n,\ C_1, \ldots, C_n \in \mathcal{G}$,

$$\begin{aligned}&\mathbb{P}\left(X_{t_1} \in C_1, \ldots, X_{t_n} \in C_n\right) \\ &\quad = \mathbb{E}\left[\mathbb{P}\left(X_{t_n} \in C_n \middle| \mathcal{F}_{t_{n-1}}\right); X_{t_1} \in C_1, \ldots, X_{t_{n-1}} \in C_{n-1}\right] \\ &\quad = \mathbb{E}\left[p_{t_{n-1}t_n}(X_{t_{n-1}}, C_n); X_{t_1} \in C_1, \ldots, X_{t_{n-1}} \in C_{n-1}\right].\end{aligned}$$

The next fact we need is a generalization of the Markov property (7.1): for $t \ge 0$, let $\mathcal{F}^{t,X} = \sigma(X_u)_{u \ge t}$. Then for any bounded and $\mathcal{F}^{t,X}$-measurable $\mathbb{R}$-valued random variable U,

$$\mathbb{E}\left[U \mid \mathcal{F}_t\right] = \mathbb{E}\left[U \mid X_t\right], \tag{7.6}$$

a fact which informally may be phrased as follows: the future depends on the past only through the present.

(7.6) is proved, considering random variables U of the form

$$U = \prod_{k=1}^{n} f_k\left(X_{t_k}\right) \tag{7.7}$$

where $n \in \mathbb{N},\ t \le t_1 < \cdots < t_n$, and each $f_k : G \to \mathbb{R}$ is bounded and measurable, and then proceeding by induction: the case $n = 1$ is an obvious extension of (7.1) from conditional probabilities to conditional extensions,

$$\mathbb{E}\left[f\left(X_u\right)|\mathcal{F}_t\right] = \mathrm{P}_{tu} f\left(X_t\right)$$

for $t \le u$, f bounded and measurable, and the induction step from $n-1$ to n is obtained using that with U as in (7.7),

$$\begin{aligned}\mathbb{E}\left[U|\mathcal{F}_t\right] &= \mathbb{E}\left[\mathbb{E}\left[U|\mathcal{F}_{t_{n-1}}\right]|\mathcal{F}_t\right]\\ &= \mathbb{E}\left[\left(\prod_{k=1}^{n-2} f_k\left(X_{t_k}\right)\right) f\left(X_{t_{n-1}}\right)|\mathcal{F}_t\right]\end{aligned}$$

where $f = f_{n-1} \cdot \mathrm{P}_{t_{n-1}t_n} f_n$. A standard extension argument finally gives (7.6) for all U that are bounded and $\mathcal{F}^{t,X}$-measurable.

7.2 Markov chains

Let μ be an RCM with mark space $E = G$, and let $\left(\left(T_n\right), \left(Y_n\right)\right)$ denote the corresponding MPP. Fixing an arbitrary initial state x_0, define the G-valued step process X by

$$X_t = Y_{\langle t \rangle}, \tag{7.8}$$

where $Y_0 \equiv x_0$. Provided $Y_n \ne Y_{n-1}$ whenever $T_n < \infty$, T_n will be the time of the nth jump for X with $Y_n = X_{T_n}$ the state reached by that jump. This is the situation treated below, although formally this identification of jumps for X with those of μ is not required. Note that always $\mathcal{F}_t^X \subset \mathcal{F}_t^\mu$ for all t.

If $Q = Q_{|x_0} = \mu(\mathbb{P})$ is the distribution of μ, the aim is now to find sufficient conditions on the Markov kernels $P^{(n)}$, $\pi^{(n)}$ determining Q which ensure that X is Markov with respect to the filtration $\left(\mathcal{F}_t^\mu\right)$, and such that the transition probabilities do not depend on x_0. We then refer to X as a *Markov chain with (general) state space* G.

The basic ingredient needed to achieve this is a collection of time-dependent *transition intensities*, i.e., quantities $q_t(x, C)$ for $t \in \mathbb{R}_0$, $x \in G$, $C \in \mathcal{G}$ of the form

$$q_t(x, C) = q_t(x) r_t(x, C)$$

where

(i) $q_t(x) = q_t\left(x, G\right) \ge 0$ is the *total intensity* for a jump from x at time t, satisfying that $(t, x) \mapsto q_t(x)$ is $\mathcal{B}_0 \otimes \mathcal{G}$-measurable and locally integrable from the right, i.e., for all t, x, $\int_t^{t+h} q_s(x)\, ds < \infty$ for $h = h(t, x) > 0$ sufficiently small;

(ii) for each t, r_t is a Markov kernel on G such that $(t, x) \mapsto r_t(x, C)$ is $\mathcal{B}_0 \otimes \mathcal{G}$-measurable for every $C \in \mathcal{G}$;

(iii) $r_t(x, \{x\}) = 0$ for all $t \in \mathbb{R}_0,\ x \in G$.

Here (iii) will ensure 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, $\left(\mathcal{F}_t^X\right) = \left(\mathcal{F}_t^\mu\right)$.

The next result as a particular case contains the RCM construction of time-homogeneous Markov chains (on a finite or countably infinite state space) from Example 3.3.1.

Theorem 7.2.1 *A sufficient condition for the step process X from (7.8) to be Markov with respect to the filtration $(\mathcal{F}_t^\mu)$ with transition probabilities that do not depend on x_0, is that there exist time-dependent transition intensities $q_t(x, C) = q_t(x) r_t(x, C)$ with q_t and r_t satisfying (i)–(iii) such that the Markov kernels $P^{(n)}$, $\pi^{(n)}$ determining the distribution $Q = Q_{|x_0}$ of the RCM μ define a stable RCM for all $x_0 \in G$ and are given by*

$$\overline{P}^{(0)}_{|x_0}(t) = \exp\left(-\int_0^t q_s(x_0)\, ds\right), \quad \pi^{(0)}_{t|x_0}(C) = r_t(x_0, C); \tag{7.9}$$

and for $n \in \mathbb{N}$, $t_1 < \cdots < t_n \leq t$, $y_1, \ldots, y_n \in G$,

$$\overline{P}^{(n)}_{z_n|x_0}(t) = \exp\left(-\int_{t_n}^t q_s(y_n)\, ds\right), \quad \pi^{(n)}_{z_n,t|x_0}(C) = r_t(y_n, C) \tag{7.10}$$

with $t > t_n$ in the last identity.

Notation. For a stochastic process X, determined from a sequence of jump times and jumps with a given arbitrary initial state x_0, we shall write $P^{(n)}_{z_n|x_0}$ and $\pi^{(n)}_{z_n,t|x_0}$ for the Markov kernels generating the jump times and jumps in order to exhibit the dependence on x_0. Note that the special structure of (7.9) and (7.10) implies that only for $n = 0$ do the kernels actually depend on x_0.

Theorem 7.2.2 *If $q_t(x)$ and $r_t(x, C)$ do not depend on t, then X is a time-homogeneous Markov chain.*

In the homogeneous case one obtains a generalization of Example 3.3.1 with exponential holding times

$$\overline{P}^{(n)}_{z_n|x_0}(t) = \exp\left(-(t - t_n)\, q(y_n)\right)$$

and jump probabilities

$$\pi^{(n)}_{z_n,t|x_0}(C) = r(y_n, C).$$

Note that the right integrability of the total intensity from (i) above is required in order that (7.9) and (7.10) define survivor functions for probabilities on $]0, \infty]$ and $]t_n, \infty]$ respectively. It is not required, however, that $\int_t^{t+h} q_s(x)\, ds < \infty$ for all $h > 0$, i.e., (7.9) and (7.10) may specify jump time distributions with finite termination points.

Proof. The proof relies on Lemma 4.3.3. Let $t > 0$ and consider the conditional distribution of $\widetilde{X} = (X_u)_{u \geq s}$ given $\mathcal{F}_s^{\mu}$. We want to show that it depends on the past $\mathcal{F}_s^{\mu}$ through X_s only. Conditioning on $\mathcal{F}_s^{\mu}$ 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 $k \in \mathbb{N}$, $0 < t_1 < \cdots < t_k \leq s,\ y_1, \ldots, y_k \in G$. Since also $\widetilde{X}$ is determined by $\theta_s \mu$, the restriction of μ to $]s, \infty[\times E$, (see p. 56 for the definition of θ_s), 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

$$\overline{P}^{(k)}_{z_k|x_0}(t) / \overline{P}^{(k)}_{z_k|x_0}(s)$$

for $t \geq s$,

$$\pi^{(k)}_{z_k,t|x_0}$$

for $t > s$ and

$$P^{(k+n)}_{\mathrm{join}(z_k,\widetilde{z}_n)|x_0}, \qquad \pi^{(k+n)}_{\mathrm{join}(z_k,\widetilde{z}_n),t|x_0}$$

for $n \in \mathbb{N}$, $\widetilde{z}_n = \left(\widetilde{t}_1, \ldots, \widetilde{t}_n; \widetilde{y}_1, \ldots, \widetilde{y}_n\right)$ with $t < \widetilde{t}_1 < \cdots < \widetilde{t}_n < t,\ \widetilde{y}_1, \ldots, \widetilde{y}_n \in G$, all of them depend on $z_k = (t_1, \ldots, t_k; y_1, \ldots, y_k)$ and x_0 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 $q_t(x)$ and $r_t(x, C)$ do not depend on t, the conditional distribution of $\theta_s^* \mu$ given $\mathcal{F}_s^{\mu}$ depends on X_s only and not on s. Here $\theta_s^* \mu$ is $\theta_s \mu$ translated backwards in time to start at time 0; cf. (4.55). □

In order for Theorem 7.2.1 to apply, one must of course first verify that the RCM μ determined by the Markov kernels (7.9) and (7.10) is stable. A useable sufficient condition for this is contained in Exercise 4.4.2.

By defining the transition probabilities appropriately, one may obtain that the Chapman–Kolmogorov equations (7.4) ((7.5) in the homogeneous case) are satisfied: consider first the general (non-homogeneous) case, and for $s \geq 0, x \in G$, let $Q^{s,x}$ denote the probability on $(\mathcal{M}, \mathcal{H})$ (tacitly assuming that explosions do not occur) determined by the Markov kernels

$$\overline{P}^{(0)}_{|s,x}(t) = \begin{cases} 1 & (t < s) \\ \exp\left(-\int_s^t q_u(x)\,du\right) & (t \geq s) \end{cases}$$

$$\overline{P}^{(n)}_{z_n|s,x}(t) = \exp\left(-\int_{t_n}^t q_u(x)\,du\right), \quad (n \geq 1, t \geq t_n)$$

$$\pi^{(n)}_{z_n,t|s,x} = r_t(y_n, \cdot), \quad (n \geq 0, t > t_n).$$

Then simply define

$$p_{st}(x, C) = Q^{s,x}\left(X_t^{\circ} \in C\right) \tag{7.11}$$

where $X_t^\circ = \eta_{\langle t \rangle}$ is the canonical chain on $\mathcal{M}$ with initial state $X_0^\circ \equiv \eta_0 \equiv x$. (The proof that the Chapman–Kolmogorov equations are indeed satisfied is left as an exercise). Note that in terms of the probabilities $Q^{s,x}$, the generalized Markov property (7.6) for the Markov chain X may be written

$$\mathbb{P}\left(\theta_t \mu \in H \left| \mathcal{F}_t^\mu \right.\right) = Q^{t,X_t}(H) \tag{7.12}$$

for all t and all $H \in \mathcal{H}$. (Since $(X_u)_{u \geq t}$ is determined by X_t and $\theta_t \mu$, this in particular gives $\mathbb{P}\left(F \left| \mathcal{F}_t^\mu \right.\right) = \mathbb{P}\left(F \left| X_t \right.\right)$ for all $F \in \mathcal{F}^{t,X}$, cf. (7.6)).

In the homogeneous case, just define Q^x for every x, using the Markov kernels from (7.9) and (7.10) when neither $q_t(x')$ nor $r_t(x', C)$ depends on t and with x_0 there replaced by x. The Markov property (7.12) now becomes

$$\mathbb{P}\left(\theta_t^* \mu \in H \left| \mathcal{F}_t^\mu \right.\right) = Q^{X_t}(H). \tag{7.13}$$

With the setup used in Theorem 7.2.1, the RCM μ determining the Markov chain X has a compensating measure L, which has an $\mathcal{F}_t^\mu$-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

$$\lambda_t(C) = q_t(X_{t-}, C), \tag{7.14}$$

as follows from Proposition 4.4.1. Note that

$$\lambda_t(C) = q_t(X_{t-}, C \backslash X_{t-})$$

and that if $t \mapsto \lambda_t(C)$ has limits from the right, it follows from Proposition 4.4.2 (b), using (7.6) that

$$\begin{aligned} \lambda_{t+}(C) &= q_{t+}(X_t, C \backslash X_t) \\ &= \lim_{h \downarrow 0} \frac{1}{h} \mathbb{P}(T_{t,1} \leq t + h,\ X_{t,1} \in C \backslash X_t | X_t), \end{aligned}$$

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. If the conditional probability that at least two jumps occur in $]t, t+h]$ is of the order $o(h)$, then the limit equals $\lim \frac{1}{h} p_{t,t+h}\left(X_t, C \backslash X_t\right)$ 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 $q_t\left(x, x\right)$ in the theory of Markov chains on finite or countably infinite state spaces are defined as $\lim \frac{1}{h}\left(p_{t,t+h}\left(x, \{x\}\right) - 1\right) = -q_t(x)$).

Consider again a Markov chain with transition intensities $q_t\left(x, C\right)$ as constructed in Theorem 7.2.1. With the transition probabilities on the form (7.11), it is possible to write down the *backward and forward integral and differential equations* (Feller–Kolmogorov equations): to obtain the backward equations, let $s < t$, $C \in \mathcal{G}$ and use

Lemma 4.3.3(bii) with $k_0 = 1$ (condition on the time and value of the first jump) on the probability $Q^{s,x}$ to obtain

$$p_{st}(x,C) = 1_C(x)\exp\left(-\int_s^t q_u(x)\,du\right) + \int_s^t du \int_{G\backslash x} q_u(x,dy)\exp\left(-\int_s^u q_v(x)\,dv\right) p_{ut}(y,C).$$

This shows $p_{st}(x,C)$ to be continuous in s and if $u \mapsto q_u(x)$ is continuous, it is then an easy matter to see that $p_{st}(x,C)$ is differentiable in s with

$$D_s p_{st}(x,C) = q_s(x)p_{st}(x,C) - \int_{G\backslash x} q_s(x,dy)\; p_{st}(y,C),$$

reducing to (and always true)

$$D_t p_t(x,C) = \int_{G\backslash x} q(x,dy)\,p_t(y,C) - q(x)\,p_t(x,C) \tag{7.15}$$

in the homogeneous case where $p_{st} = p_{t-s}$.

The forward systems are more delicate and additional assumptions are needed for the derivation. Again, fix s, let $C \in \mathcal{G}$ and now apply Itô's formula (Section 4.7) to the process $1_{(X_t^\circ \in C)}$ for $t \geq s$ under the probability $Q^{s,x}$ on $(\mathcal{M},\mathcal{H})$ (or for $t \geq 0$ since we have that $X_t^\circ \equiv x$ for $t \leq s$ $Q^{s,x}$-a.s.), i.e., we are looking for a decomposition

$$1_{(X_t^\circ \in C)} = 1_C(x) + U_t + M_t^\circ(S) \tag{7.16}$$

for $t \geq s$ with U predictable and continuous, $U_s \equiv M_s^\circ(S) \equiv 0$ and S a predictable field. The indicator process on the left jumps only when μ° does and

$$\Delta 1_{(X_t^\circ \in C)} = \Delta\overline{N}_t^\circ\left(1_{(X_{t-}^\circ \notin C, X_t^\circ \in C)} - 1_{(X_{t-}^\circ \in C, X_t^\circ \notin C)}\right).$$

Identifying these jumps with those of the stochastic integral on the right of (7.16), $\Delta M_t^\circ(S) = (\Delta\overline{N}_t^\circ)S_t^{X_t^\circ}$ (recall that $\eta_{\langle t\rangle} = X_t^\circ$), yields

$$\begin{aligned} S_t^y &= 1_C(y)1_{(X_{t-}^\circ \notin C)} - 1_{C^c}(y)1_{(X_{t-}^\circ \in C)} \\ &= 1_C(y) - 1_{(X_{t-}^\circ \in C)}. \end{aligned}$$

Since the indicator process is piecewise constant, it only remains to choose U so that the process on the right of (7.16) is also a step process, i.e., $U_t = L_t^\circ(S)$. But

$$L^\circ(dt,dy) = \overline{\lambda}_t^\circ\,dt\,\pi_{\xi_{\langle t-\rangle},t}^{\langle t-\rangle}(dy) = q_t\left(X_{t-}^\circ\right)dt\,r_t\left(X_{t-}^\circ,dy\right)$$

and hence

$$\begin{aligned} U_t &= \int_s^t du \int_G q_u\left(X_{u-}^\circ,dy\right)\left(1_C(y) - 1_{(X_{u-}^\circ \in C)}\right) \\ &= \int_s^t du\left(q_u\left(X_u^\circ,C\right) - 1_{(X_u^\circ \in C)}q_u\left(X_u^\circ\right)\right). \end{aligned}$$

Because S is uniformly bounded, it follows from Theorem 4.6.1(iii) that $M^{\circ}(S)$ is a local $Q^{s,x}$-martingale (τ_n), which is a true martingale if $E^{s,x}\overline{N}_t^{\circ} = \mathbb{E}\left[\overline{N}_t \,|X_s = x\right] < \infty$. Assuming that this condition holds, (7.16) shows that $E^{s,x}|U_t| < \infty$ and taking expectations in (7.16) then gives

$$p_{st}(x, C) = 1_C(x) + \int_s^t du \int_G p_{su}(x, dy)\ (q_u(y, C\backslash y) - 1_C(y)\, q_u(y)), \quad (7.17)$$

provided it is allowed to interchange $E^{s,x}$ and $\int_s^t$, e.g. if

$$\int_s^t du \int_G p_{su}(x, dy)\ q_u(y) < \infty.$$

A formal differentiation after t in (7.17) finally yields

$$D_t p_{st}(x, C) = \int_G p_{st}(x, dy)\,(q_t(y, C\backslash y) - 1_C(y)\, q_t(y)),$$

in the homogeneous case reducing to

$$D_t p_t(x, C) = \int_G p_t(x, dy)\ (q(y, C\backslash y) - 1_C(y)\, q(y)). \quad (7.18)$$

When X is a homogeneous Markov chain on an at most countably infinite state space, the backward and forward Feller–Kolmogorov equations were presented in Exercise 4.7.1. For a different derivation, see Example 7.7.2 below.

It is perfectly possible to have Markov chains (necessarily non-homogeneous) that do not have transition intensities and we shall conclude this section with a discussion of how they are constructed. The proof of Theorem 7.2.1 carries over to the case where the Markov kernels $P^{(n)}$, $\pi^{(n)}$ have the form

$$\begin{aligned} \overline{P}^{(0)}_{|x_0}(t) &= \overline{F}_{x_0}(t), & \pi^{(0)}_{t|x_0}(C) &= r_t(x_0, C), \\ \overline{P}^{(n)}_{z_n|x_0}(t) &= \frac{\overline{F}_{y_n}(t)}{\overline{F}_{y_n}(t_n)}, & \pi^{(n)}_{z_n,t|x_0}(C) &= r_t(y_n, C), \end{aligned}$$

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 as in Theorem 7.2.1. The construction works in particular for distribution functions F_x with atoms.

For the expression for $P^{(n)}$ to make sense, it is natural to assume that $\overline{F}_x(t) > 0$ for all x and all $t \in \mathbb{R}_0$. In the transition intensity case this would mean that $\int_0^t q_s(x)\, ds < \infty$ for all $t \in \mathbb{R}_0$, a condition that is not needed as was noted above. In general the condition $\overline{F}_x(t) > 0$ may be dispensed with 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

$$\frac{1}{\overline{F}_{x|s}(t)}\overline{F}_{x|s} \equiv \overline{F}_{x|t} \quad \text{on }]t, \infty].$$

One then defines

$$\overline{P}^{(n)}_{z_n|x_0}(t) = \begin{cases} \overline{F}_{y_n|t_n}(t) & \text{if } t < t^\dagger_{z_n}, \\ 0 & \text{if } t \geq t^\dagger_{z_n} \end{cases}$$

where $t^\dagger_{z_n}$ the termination point for the distribution with distribution function $F_{y_n|t_n}$.

7.3 Construction and basic properties of PDMPs

We shall now discuss in detail the structure of piecewise deterministic Markov processes (PDMPs), where the Markov chains treated in the previous section appear as a simple special case. In turn the PDMPs are special cases of the piecewise deterministic processes introduced on p. 25.

Just as for the Markov chains, we shall denote the state space for the PDMPs by $(G, \mathcal{G})$, but apart from assuming as before that $\mathcal{G}$ is countably generated with all singletons $\{x\} \in \mathcal{G}$, we shall now also assume that G is a topological space with $\mathcal{G}$ the Borel σ-algebra (the σ-algebra generated by the open sets).

As was the case for Markov chains, the starting point is an RCM μ with mark space $E = G$ with $((T_n), (Y_n))$ denoting the corresponding MPP. Also fix an arbitrary initial state $x_0 \in G$, and as in (3.12) define

$$X_t = f^{\langle t\rangle}_{Z_{\langle t\rangle}|x_0}(t) \tag{7.19}$$

with as usual $\overline{N}_t$ the total number of jumps for μ on $[0, t]$,

$$Z_{\langle t\rangle} = (T_1, \ldots, T_{\langle t\rangle}; Y_1, \ldots, Y_{\langle t\rangle}),$$

and with each $t \to f^{(n)}_{z_n|x_0}(t)$ a *continuous* function on $[t_n, \infty[$ 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 $f^{(n)}$ are assumed to satisfy the boundary conditions

$$f^{(0)}_{|x_0}(0) = x_0, \qquad f^{(n)}_{z_n|x_0}(t_n) = y_n$$

so that if $T_n < \infty$,

$$X_{T_n} = Y_n,$$

i.e., Y_n is the state reached by X at the time of the nth jump for μ.

Because of the continuity of the $f^{(n)}$, the process X is piecewise continuous with jumps possible only at the time points T_n. Also $\mathcal{F}^X_t \subset \mathcal{F}^\mu_t$ for all t. Markov chains are obtained by taking $f^{(n)}_{z_n|x_0}(t) = y_n$ (with $y_0 = x_0$).

The distribution of X is determined by that of μ, i.e., through the Markov kernels $P^{(n)}_{z_n|x_0}$ and $\pi^{(n)}_{z_n,t|x_0}$ (depending on x_0). To ensure that the jump times for μ agree precisely with those of X, i.e., that $X_{T_n} \neq X_{T_n-}$ whenever $T_n < \infty$ we shall assume that

$$\pi^{(n)}_{z_n,t|x_0}\left(\left\{f^{(n)}_{z_n|x_0}(t)\right\}\right) = 0 \tag{7.20}$$

for almost all values (z_n, t) of (Z_n, T_{n+1}). (As was the case for Markov chains, (7.20) is not required for the results below, nor is the continuity of the $f^{(n)}$. For convenience we shall however maintain both assumptions).

The main problem to be discussed is that of finding out what structure must be imposed on the Markov kernels $P^{(n)}_{z_n|x_0}$, $\pi^{(n)}_{z_n,t|x_0}$ *and* the functions $f^{(n)}_{z_n|x_0}$, 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 $(X_u)_{u\geq s}$ given $\mathcal{F}^{\mu}_s$. By (7.6) X is Markov with respect to $\left(\mathcal{F}^{\mu}_t\right)$ iff this depends on the past $\mathcal{F}^{\mu}_s$ through X_s only. As always, conditioning on $\mathcal{F}^{\mu}_s$ 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 4.2.2). On the set $(\overline{N}_s = k, Z_k = z_k)$, by (7.19),

$$X_s = f^{(k)}_{z_k|x_0}(s), \tag{7.21}$$

and until the time $T_{s,1}$ of the first jump after s, X follows the deterministic function

$$t \mapsto f^{(k)}_{z_k|x_0}(t) \quad (t \geq s).$$

Copying the proof of Theorem 7.2.1 and referring to Lemma 4.3.3, 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 (7.21) only:

$$\begin{aligned} &f^{(k)}_{z_k|x_0}(t) \quad (t \geq s) \\ &\overline{P}^{(k)}_{z_k|x_0}(t)/\overline{P}^{(k)}_{z_k|x_0}(s) \quad (t \geq s) \\ &\pi^{(k)}_{z_k,t|x_0} \quad (t \geq s) \end{aligned} \tag{7.22}$$

and, for $n \in \mathbb{N}$, writing $\widetilde{z}_n = (\widetilde{t}_1, \ldots, \widetilde{t}_n; \widetilde{y}_1, \ldots, \widetilde{y}_n)$ with $s < \widetilde{t}_1 < \cdots < \widetilde{t}_n$, $\widetilde{y}_1, \ldots, \widetilde{y}_n \in G$,

$$\begin{aligned} &f^{(k+n)}_{\text{join}(z_k,\widetilde{z}_n)|x_0} \\ &P^{(k+n)}_{\text{join}(z_k,\widetilde{z}_n)|x_0} \\ &\pi^{(k+n)}_{\text{join}(z_k,\widetilde{z}_n),t|x_0} \end{aligned} \tag{7.23}$$

with $t > \widetilde{t}_n$ in the last expression. We start by exploring the first of the three quantities in (7.22). The requirement that this quantity depends on X_s only amounts to requiring that for some function ϕ_{st},

$$f^{(k)}_{z_k|x_0}(t) = \phi_{st}(f^{(k)}_{z_k|x_0}(s)). \tag{7.24}$$

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 (7.24), and changing the notation gives

$$\phi_{su}(y) = \phi_{tu}(\phi_{st}(y)) \quad (0 \le s \le t \le u,\ y \in G), \tag{7.25}$$

which together with the boundary conditions

$$\phi_{tt}(y) = y \quad (t \in \mathbb{R}_0) \tag{7.26}$$

are the basic functional equations describing the deterministic behavior of a piecewise deterministic Markov process: (7.19) becomes

$$X_t = \phi_{T_{\langle t\rangle},t}(Y_{\langle t\rangle}) \tag{7.27}$$

with $T_{\langle t\rangle} = 0,\ Y_{\langle t\rangle} = Y_0 \equiv x_0$ on $(\overline{N}_t = 0)$. More compactly (7.25) and (7.26) may be written

$$\phi_{su} = \phi_{tu} \circ \phi_{st} \quad (0 \le s \le t \le u), \qquad \phi_{tt} = \mathrm{id} \quad (t \ge 0),$$

id denoting the identity on G. In particular the ϕ_{st} form a two-parameter semigroup under composition. The interpretation of $\phi_{st}(y)$ is that it denotes the state of the process at time t if at time s it was in state y, and there were no jumps on the time interval $]s, t]$.

The time-homogeneous case, where ϕ_{st} depends on s, t through $t - s$ only, is of particular interest. Writing $\phi_t = \phi_{s,s+t}$ for any s, (7.25) and (7.26) become

$$\phi_{s+t}(y) = \phi_s(\phi_t(y)) \quad (s, t \ge 0,\ y \in G), \qquad \phi_0(y) = y \quad (y \in G) \tag{7.28}$$

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. The interpretation of $\phi_t(y)$ is that it denotes the state of the process at the end of a time interval of length t, provided that at the start of the interval the process was in state y and there were no jumps during the interval.

Having determined the structure of the piecewise deterministic part of the process, it is now an easy matter to prove the following result, where $D := \{(s, t, y) \in \mathbb{R}_0^2 \times G : s \le t\}$ and

(i) $q_t(x) = q_t(x, G) \ge 0$ is the total intensity for a jump from x at time t, satisfying that $(t, x) \mapsto q_t(x)$ is $\mathcal{B}_0 \otimes \mathcal{G}$-measurable and locally integrable from the right in the sense that for all t, x, $\int_t^{t+h} q_s(\phi_{st}(x))\, ds < \infty$ for $h = h(t, x) > 0$ sufficiently small;

(ii) for each t, r_t is a Markov kernel on G such that $(t, x) \mapsto r_t(x, C)$ is $\mathcal{B}_0 \otimes \mathcal{G}$-measurable for every $C \in \mathcal{G}$; $r_t(x, C)$ 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 x.

(iii) $r_t(x, \{x\}) = 0$ for all $t \in \mathbb{R}_0,\ x \in G$.

The condition

$$r_t\,(x, \{x\}) = 0, \tag{7.29}$$

from (iii) is essential to ensure that X has a discontinuity at each finite T_n; cf. (7.20).

Theorem 7.3.1 (a) *Suppose $\phi : D \to G$ is a measurable function which satisfies (7.25), (7.26) and is such that $t \mapsto \phi_{st}(y)$ is continuous on $[s, \infty[$ for all $s \in \mathbb{R}_0,\ y \in G$ and suppose that q_t and r_t satisfy (i)–(iii) above. Then the piecewise deterministic process X given by $X_0 = Y_0 \equiv x_0$ and*

$$X_t = \phi_{T_{\langle t\rangle},t}(Y_{\langle t\rangle})$$

(with $T_0 \equiv 0$) is a piecewise continuous 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 $Q = Q_{|x_0}$ of the RCM μ recording the jump times for X and the states reached by X at the time of each jump, define a stable RCM for all $x_0 \in G$ and are of the form

$$\overline{P}^{(0)}_{|x_0}(t) = \exp\left(-\int_0^t q_s(\phi_{0s}(x_0))\,ds\right), \quad (t \in \mathbb{R}_0)$$

$$\pi^{(0)}_{t|x_0}(C) = r_t(\phi_{0t}(x_0), C) \quad (t \in \mathbb{R}_0, C \in \mathcal{G})$$

and for $n \in \mathbb{N},\ t_1 < \cdots < t_n \le t,\ y_1, \ldots, y_n \in G$

$$\overline{P}^{(n)}_{z_n|x_0}(t) = \exp\left(-\int_{t_n}^t q_s(\phi_{t_n s}(y_n))\,ds\right), \tag{7.30}$$

$$\pi^{(n)}_{z_n,t|x_0}(C) = r_t(\phi_{t_n t}(y_n), C)(C \in \mathcal{G}), \tag{7.31}$$

with $t > t_n$ in the last identity.

(b) *The piecewise deterministic Markov process X determined by $\phi_{st}(y),\ q_t(y),\ r_t(y, C)$ 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 $\widetilde{r}$ on G with $\widetilde{r}\,(y, \{y\}) = 0$, such that for all $s \le t,\ y \in G$*

$$\phi_{st}(y) = \widetilde{\phi}_{t-s}(y), \quad q_t(y) = \widetilde{q}(y), \quad r_t(y, C) = \widetilde{r}(y, C).$$

The Markov kernels (7.30) and (7.31) then take the form

$$\overline{P}^{(n)}_{z_n|x_0}(t) = \exp\left(-\int_0^{t-t_n} q(\widetilde{\phi}_s(y_n))\,ds\right),$$

$$\pi^{(n)}_{z_n,t|x_0}(C) = r(\widetilde{\phi}_{t-t_n}(y_n), C) \quad (C \in \mathcal{G}),$$

using $t_0 = 0$ *and* $y_0 = x_0$.

Notation. In the sequel we always write ϕ, q, r rather than $\widetilde{\phi}, \widetilde{q}, \widetilde{r}$ in the time-homogeneous case, with ϕ then satisfying (7.28).

Remark 7.3.1 Note that for $n \geq 1$, $P^{(n)}_{z_n|x_0}$ and $\pi^{(n)}_{z_n,t|x_0}$ do not depend on the initial state x_0.

Proof. For part (a) we must show that the three quantities in (7.22) and (7.23) depend on k, x_0, z_k through $X_s = \phi_{t_k s}(y_k)$ only. For the first entry in (7.22) this follows from (7.24) and (7.25); cf. the argument on p. 154 leading to (7.25). Next

$$\frac{\overline{P}^{(k)}_{z_k|x_0}(t)}{\overline{P}^{(k)}_{z_k|x_0}(s)} = \exp\left(-\int_s^t q_u(\phi_{t_k u}(y_k))\,du\right)$$
$$= \exp\left(-\int_s^t q_u(\phi_{su}(X_s))\,du\right),$$
$$\pi^{(k)}_{z_k,t|x_0} = r_t(\phi_{t_k t}(y_k))$$
$$= r_t(\phi_{st}(X_s)),$$

again using (7.25). As for the three quantities from (7.23), 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, z_k or s. This completes the proof of (a).

To prove (b), one must show that when ϕ, q, p are of the form given in part (b), everything in (7.22) and (7.23), which we already know 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

$$\frac{\overline{P}^{(k)}_{z_k|x_0}(t)}{\overline{P}^{(k)}_{z_k|x_0}(s)} = \exp\left(-\int_s^t \widetilde{q}(\widetilde{\phi}_{u-s}(X_s))\,du\right)$$
$$= \exp\left(-\int_0^{t-s} \widetilde{q}(\widetilde{\phi}_u(X_s))\,du\right).$$

(See also the last part of the proof of Theorem 7.2.1 for a more detailed argument). □

With the PDMP X constructed as in Theorem 7.3.1(a), it follows that the $\mathcal{F}^{\mu}_t$-compensating measure L for the RCM μ determining the jump times and jumps for X is given by

$$L(dt, dy) = q_t(X_{t-})\, dt\, r_t(X_{t-}, dy). \tag{7.32}$$

Equivalently, for any $C \in \mathcal{G}$, the compensator $\Lambda_t(C) = L([0, t] \times C)$ has an $\mathcal{F}_t^{\mu}$-predictable intensity process given by

$$\lambda_t(C) = q_t(X_{t-}) r_t(X_{t-}, C), \tag{7.33}$$

and in the time-homogeneous case,

$$\lambda_t(C) = q(X_{t-}) r(X_{t-}, C). \tag{7.34}$$

The Markov property of X is reflected in the fact that $\lambda_t(C)$, which is of course also $\mathcal{F}_t^X$-predictable, depends on the past $(X_s)_{0 \le s < t}$ through X_{t-} only. Note that the two expressions are the same, no matter what the deterministic structure (the ϕ_{st} or ϕ_t) for X is. For X a Markov chain, (7.33) is simply the same as (7.14).

It is quite easy to generalize Theorem 7.3.1 to the case where the $P^{(n)}$ do not have densities: imitating the proof one only has to verify that the quantities in (7.22) and (7.23) depend on the past through X_s only. The result is

Theorem 7.3.2 (a) *Suppose $\phi : D \to G$ is a measurable function which satisfies (7.25), (7.26) and is such that $t \mapsto \phi_{st}(y)$ is continuous on $[s, \infty[$ for all $s \in \mathbb{R}_0$, $y \in G$. Then the piecewise deterministic process X given by $X_0 \equiv x_0$ and*

$$X_t = \phi_{T_{\langle t \rangle}, t}(Y_{\langle t \rangle})$$

is a Markov process with transitions that do not depend on x_0, provided the $\pi^{(n)}$ are as in Theorem 7.3.1 and together with the $P^{(n)}$ yield a stable RCM for arbitrary $x_0 \in G$ with the $P^{(n)}$ of the form

$$\overline{P}^{(n)}_{z_n | x_0}(t) = \overline{F}_{t_n, y_n}(t) \quad (t \ge t_n),$$

where for each s, y, $F_{s,y}$ is the distribution function for a probability on $]s, \infty]$ with survivor function $\overline{F}_{s,y} = 1 - F_{s,y}$, and the $\overline{F}_{s,y}$ satisfy that for $s \le t \le u$, $y \in G$,

$$\overline{F}_{s,y}(u) = \overline{F}_{s,y}(t) \overline{F}_{t, \phi_{st}(y)}(u). \tag{7.35}$$

(b) *The Markov process X from (a) is time-homogeneous if $\phi_{st} = \phi_{t-s}$, $r_t(y, C) = r(y, C)$ and if in addition $\overline{F}_{s,y}(t) = \overline{F}_y(t - s)$ depends on s and t through $t - s$ only.*

Remark 7.3.2 Theorem 7.3.1 is the special case of Theorem 7.3.2 corresponding to

$$\overline{F}_{s,y}(t) = \exp\left(-\int_s^t q_u(\phi_{su}(y))\, du\right).$$

Exercise 7.3.1 Let X be a PDMP as described in Theorem 7.3.2, and let $h : (G, \mathcal{G}) \to (G', \mathcal{G}')$ be a bimeasurable bijection (1-1, onto with h and h^{-1} measurable). Define $T'_n = T_n$, $Y_{n'} = h(Y_n)$, $\mu' = \sum_{n:T'_n<\infty} \varepsilon_{(T'_n,Y'_n)}$ and $X'_t = h(X_t)$ for $t \geq 0$. Show that X' is a PDMP (which is piecewise continuous if G' is a topological space and h is a continuous, but well defined in general) corresponding to the following choice of piecewise deterministic behaviour ϕ'_{st}, survivor functions $\overline{F}'_{s,y'}$ and transition probabilities r'_t on G':

$$\phi'_{st} = h \circ \phi_{st} \circ h^{-1}, \quad \overline{F}'_{s,y'} = \overline{F}_{s,h^{-1}(y')}, \quad r'_t\left(y', C'\right) = r_t\left(h^{-1}\left(y'\right), h^{-1}\left(C'\right)\right).$$

The Markov properties (7.12) and (7.13) established for Markov chains have complete analogues for PDMPs: for $s \geq 0$, $x \in G$ define the probability $Q^{s,x}$ on $(\mathcal{M}, \mathcal{H})$ to be generated from the Markov kernels

$$\begin{aligned}
\overline{P}^{(0)}_{|s,x}(t) &= \begin{cases} 1 & (t < s) \\ \overline{F}_{s,x}(t) & (t \geq s), \end{cases} \qquad \pi^{(0)}_{t|s,x} = r_t\left(\phi_{st}(x), \cdot\right) \\
\overline{P}^{(n)}_{z_n|s,x}(t) &= \overline{F}_{t_n,y_n}(t), \quad (n \geq 1, t \geq t_n) \\
\pi^{(n)}_{z_n,t|s,x} &= r_t\left(\phi_{t_n t}(y_n), \cdot\right), \quad (n \geq 1, t > t_n),
\end{aligned} \tag{7.36}$$

and in the homogeneous case, where $\phi_{st} = \phi_{t-s}$, $\overline{F}_{s,y}(t) = \overline{F}_y(t-s)$ and $r_t = r$, define Q^x as the probability obtained from this when $s = 0$, i.e., Q^x is generated by the Markov kernels

$$\begin{aligned}
\overline{P}^{(0)}_{|x}(t) &= \overline{F}_x(t), \quad \pi^{(0)}_{t|x} = r\left(\phi_t(x), \cdot\right) \\
\overline{P}^{(n)}_{z_n|x}(t) &= \overline{F}_{y_n}(t - t_n), \quad (n \geq 1, t \geq t_n) \\
\pi^{(n)}_{z_n,t|x} &= r\left(\phi_{t-t_n}(y_n), \cdot\right), \quad (n \geq 0, t > t_n).
\end{aligned} \tag{7.37}$$

Then for all $H \in \mathcal{H}$,

$$\mathbb{P}\left(\theta_t \mu \in H \,\middle|\, \mathcal{F}^\mu_t\right) = Q^{t,X_t}(H) \tag{7.38}$$

in general, and for the homogeneous case,

$$\mathbb{P}\left(\theta^*_t \mu \in H \,\middle|\, \mathcal{F}^\mu_t\right) = Q^{X_t}(H). \tag{7.39}$$

We shall call (7.38) the *general Markov property* and (7.39) the *homogeneous Markov property.*

The transition probabilities for the non-homogeneous (resp. homogeneous) processes

$$p_{st}(x, C) = Q^{s,x}\left(X^\circ_t \in C\right), \quad \left(p_t(x, C) = Q^x\left(X^\circ_t \in C\right)\right)$$

where now $X^\circ_t = \phi_{\tau_{\langle t\rangle},t}\left(\eta_{\langle t\rangle}\right)$ with $\tau_0 \equiv s$, $\eta_0 \equiv x$ when $\overline{N}^\circ_t = 0$ (resp. $X^\circ_t = \phi_{t-\tau_{\langle t\rangle}}\left(\eta_{\langle t\rangle}\right)$ with $\tau_0 \equiv 0$, $\eta_0 \equiv x$) is the canonical Markov process on $\mathcal{M}$, satisfies the Chapman–Kolmogorov equations (7.4) exactly. Equivalently to this, the transition operators

$$\mathrm{P}_{st} f(x) = \int f\left(X_t^\circ\right) dQ^{s,x}, \quad \left(\mathrm{P}_t f(x) = \int f\left(X_t^\circ\right) dQ^x\right)$$

form a semigroup of operators, $\mathrm{P}_{su} = \mathrm{P}_{st}\mathrm{P}_{tu}$ for $0 \le s \le t \le u$, ($\mathrm{P}_{s+t} = \mathrm{P}_s\mathrm{P}_t$ for $s, t \ge 0$) on the space of bounded and measurable functions $f : G \to \mathbb{R}$.

Note that both (7.38) and (7.39) hold simultaneously for all H,

$$\mathbb{P}\left(\theta_t\mu \in \cdot \,\middle|\, \mathcal{F}_t^\mu\right) = Q^{t,X_t} \tag{7.40}$$

and in the homogeneous case,

$$\mathbb{P}\left(\theta_t^*\mu \in \cdot \,\middle|\, \mathcal{F}_t^\mu\right) = Q^{X_t},$$

meaning e.g. in the case of (7.40) that for ω outside a $\mathbb{P}$-null set, $\mathbb{P}\left(\theta_t\mu \in \cdot \,\middle|\, \mathcal{F}_t^\mu\right)(\omega)$ and $Q^{t,X_t(\omega)}$ are the same probability measure on $(\mathcal{M}, \mathcal{H})$.

We shall discuss further the structure of PDMPs and begin with a discussion of the solutions to the homogeneous semigroup equation (7.28).

It is immediately checked that a general form of solutions are obtained by considering a continuous injective map $\Phi : G \to V$, where V is a topological vector space, and defining

$$\phi_t(y) = \Phi^{-1}\left(\Phi(y) + tv_0\right) \tag{7.41}$$

for some $v_0 \in V$. In particular, if $G \subset \mathbb{R}^d$ one may take $V = \mathbb{R}^d$.

Different choices for Φ in (7.41) may lead to the same $\phi_t(y)$: replacing $\Phi(y)$ by $\Phi(y) + v$ for an arbitrary $v \in V$ does not change anything. Also changing $\Phi(y)$ to $K\Phi(y)$ and v_0 to Kv_0 for an arbitrary $K \neq 0$ does not affect the resulting $\phi_t(y)$.

Assuming that $G = \mathbb{R}$, it is also possible to obtain partial differential equations for solutions of (7.28). By assumption $t \mapsto \phi_t(y)$ 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 (7.28)

$$\begin{aligned} D_t\phi_t(y) &= \lim_{s\downarrow 0}\frac{1}{s}(\phi_{s+t}(y) - \phi_t(y)) \\ &= \lim_{s\downarrow 0}\frac{1}{s}(\phi_s(y) - y)\frac{\phi_t(\phi_s(y)) - \phi_t(y)}{\phi_s(y) - y} \\ &= a(y)D_y\phi_t(y), \end{aligned}$$

so, assuming that the partial derivatives exist, we arrive at the first order linear partial differential equation

$$D_t\phi_t(y) = a(y)D_y\phi_t(y) \tag{7.42}$$

with the boundary condition $\phi_0(y) = y$.

Because the ϕ_t commute under composition, a differential equation different from (7.42) 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

$$D_t\phi_t(y) = a(\phi_t(y)). \tag{7.43}$$

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 (7.41) with Φ satisfying $v_0/\Phi' = a$ (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 $\phi_t(y) = ye^{Kt}$ corresponding to piecewise exponential processes.
(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, using Theorem 7.3.2 it may still be possible to obtain a homogeneous PDMP by forcing a jump every time $\phi_t(y)$ reaches the critical value 0).
(v) More generally, if $y > 0$ and $a(y) = Ky^\beta$, where $\beta \neq 1$, then $\phi_t(y) = \left(y^{1-\beta} + K(1-\beta)t\right)^{1/(1-\beta)}$.
(vi) If $a(y) = Ke^{-y}$, then $\phi_t(y) = \log(e^y + Kt)$.

If $G \subset \mathbb{R}^d$ (with $d \geq 2$) analogues of (7.42) and (7.43) are readily obtained: write $\phi_t(y)$ as a column vector and assume that

$$a(y) = \lim_{s\downarrow 0} \frac{1}{s}(\phi_s(y) - y) \in \mathbb{R}^{d\times 1}$$

exists for all y and, assuming the partial derivatives do exist, deduce that

$$D_t\phi_t(y) = \left(D_y\phi_t(y)\right)a(y) = a\left(\phi_t(y)\right) \tag{7.44}$$

where $D_y\phi_t(y)$ is the $d\times d$-matrix with elements $\left(D_y\phi_t(y)\right)_{ij} = D_{y_j}\phi_t^i(y)$ with ϕ_t^i the ith coordinate of ϕ_t. With $V = \mathbb{R}^d$ and $\phi_t(y)$ of the form (7.41) with Φ differentiable, one finds

$$a(y) = \left(D_y\Phi(y)\right)^{-1} v_0$$

where $\left(D_y\Phi(y)\right)_{ij} = D_{y_j}\Phi^i(y)$.

Example 7.3.1 Let $\beta_i \in \mathbb{R}$ for $1 \le i \le d$ be given constants and define $\overline{\beta} = \sum_1^d \beta_i$. The functions

$$\phi_t^i(y) = \begin{cases} y_i \left(1 + \frac{\overline{\beta}}{\prod_1^d y_j} t\right)^{\beta_i/\overline{\beta}} & \textit{if } \overline{\beta} \neq 0, \\ y_i \exp\left(\frac{\beta_i}{\prod_1^d y_j} t\right) & \textit{if } \overline{\beta} = 0, \end{cases}$$

are well defined for all $y_i > 0$ and all $t \ge 0$ if $\overline{\beta} \ge 0$, and for all $y_i > 0$ and $t < -\overline{\beta}^{-1} \prod_1^d y_i$ if $\overline{\beta} < 0$. It is easily verified directly that the $\phi_t = \left(\phi_t^i\right)_{1 \le i \le d}$ have the semigroup property (7.28) (as long as they are well defined) with

$$a^i(y) = \frac{\beta_i y_i}{\prod_1^d y_j}.$$

Thus the ϕ_t may be used to define a homogeneous PDMP with state space $\mathbb{R}_+^d$ as long as in the case $\overline{\beta} < 0$ the jumps occur early enough.

A particularly interesting choice of the jump intensities $q(y)$ and the jump probabilities $r(y, \cdot)$ is to take

$$q(y) = \frac{\lambda}{\prod_1^d y_j} \tag{7.45}$$

for some $\lambda > 0$ and

$$r(y, \cdot) = \text{law of } (y_i U_i)_{1 \le i \le d}$$

with $(U_i)_{1 \le i \le d}$ some d-dimensional random variable with all $U_i > 0$. Then one may show that the PDMP has the following *multi-self-similarity* property: if $X^{(x)} = \left(X^{i,(x_i)}\right)_{1 \le i \le d}$ denotes the PDMP with initial state $x = (x_i) \in \mathbb{R}_+^d$, it holds for all constants $c_i > 0$ that

$$\left(c_i X_t^{i,(x_i/c_i)}\right)_{1 \le i \le d, t \ge 0} \stackrel{(d)}{=} \left(X_{ct}^{(x)}\right)_{t \ge 0}, \tag{7.46}$$

writing $c = \prod_1^d c_i$ and where $\stackrel{(d)}{=}$ stands for equality in distribution (between two processes). Furthermore, using $X^{(x)}$ to define $\widehat{X}^{(\hat{x})} = \prod_1^d X^{i,(x_i)}$ where $\widehat{x} = \prod_1^d x_i$, and if

$$A_t := \int_0^t \frac{1}{\widehat{X}_s^{(\widehat{x})}} \, ds,$$

satisfies $\lim_{t \to \infty} A_t = \infty$ a.s., then defining a new process $\varsigma^{(a)} = \left(\varsigma_u^{i,(a_i)}\right)_{1 \le i \le d, u \ge 0}$ where $a = (a_i)$, $a_i = \log x_i$, by a *time-change* through A,

$$\varsigma_{A_t}^{i,(a_i)} = \log X_t^{i,(x_i)} \quad (1 \le i \le d, t \ge 0),$$

it holds that $\varsigma^{(a)}$ is a d-dimensional compound Poisson with drift: $\varsigma^{(a)}$ has the same distribution as the process with ith coordinate $(1 \le i \le d)$

$$a_i + \beta_i u + \sum_{n=1}^{N_u} V_n^i \qquad (u \ge 0)$$

where N is a homogeneous Poisson process with intensity λ and the $V_n = \left(V_n^i\right)_{1 \le i \le d}$ are iid and independent of N such that $\left(V_n^i\right) \stackrel{(\mathrm{d})}{=} (\log U_i)$.

Exercise 7.3.2 Show the multi-self-similarity property (7.46). Also show that if $\overline{\beta} < 0$, the jumps for $X^{(x)}$ come early enough, i.e., show that with the jump intensity function q given by (7.45),

$$\mathbb{P}\left(T_1 < -\frac{1}{\overline{\beta}} \prod_1^d x_i\right) = 1$$

where T_1 is the time of the first jump for $X^{(x)}$.

Show that the product process $\widehat{X}^{(\widehat{x})}$ is a (one-dimensional) piecewise linear PDMP. More generally, show that if

$$\{1, \dots, d\} = \bigcup_{k=1}^{d'} I_k$$

where the I_k are non-empty and disjoint, then the d'-dimensional process $\widetilde{X}^{(\widetilde{x})} = \left(\widetilde{X}^{k,(\widetilde{x}_k)}\right)_{1 \le k \le d'}$ where $\widetilde{x}_k = \prod_{I_k} x_i$ and

$$\widetilde{X}^{k,(\widetilde{x}_k)} = \prod_{i \in I_k} X^{i,(x_i)},$$

is a multi-self-similar process in d' dimensions, of the same type as $X^{(x)}$. And of course the transition probabilities for $\widetilde{X}^{(\widetilde{x})}$ do not depend on the choice of $x = (x_i)$.

A fairly general form of solutions to the functional equation (7.25) 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 $\widetilde{\Phi}$ an injective map from the state space $\mathbb{R}_0 \times G$ for (t, X_t) to $\mathbb{R}_0 \times V$, with V a topological vector space, this makes it natural to look for $\widetilde{\Phi}$ of the form $\widetilde{\Phi}(t, y) = (t, \Phi_t(y))$ with the deterministic part of (t, X_t) given by (7.41), using $\widetilde{\Phi}$ instead of Φ, (s, y) instead of y, and $(1, v_0)$ instead of v_0. The end result is that with $(\Phi_t)_{t \ge 0}$ a family of injections from G to V, the functions $(\phi_{st})_{0 \le s \le t}$ given by

$$\phi_{st}(y) = \Phi_t^{-1}\left(\Phi_s(y) + (t - s)\, v_0\right)$$

for some $v_0 \in V$ satisfy the non-homogeneous equation (7.25) and the boundary condition (7.26).

7.4 Examples of PDMPs

7.4.1 Renewal processes

An SPP $\mathcal{T} = (T_n)_{n\geq 1}$ is a (0-delayed) renewal process with renewal times T_n (see Example 3.1.3) if the waiting times $V_n = T_n - T_{n-1}$ (with $T_0 \equiv 0$) are independent and identically distributed (with the possibility $P(V_n = \infty) > 0$ allowed). Write $N_t = \sum_{n=1}^{\infty} 1_{(T_n \leq t)}$.

Defining the *backward recurrence time process* X by

$$X_t = t - T_{\langle t \rangle},$$

(the time since the most recent renewal), we claim that X is time-homogeneous Markov (with respect to the filtration $(\mathcal{F}_t^X)$). Note that $X_0 \equiv 0$, meaning that the process is 0-*delayed.*

To verify this, we use Theorem 7.3.1 and assume that the distribution of the V_n has hazard function u,

$$\mathbb{P}(V_n > v) = \exp\left(-\int_0^v u(s)ds\right).$$

(X will be time-homogeneous Markov even without this assumption, which is made to make the example fit into the framework of Theorem 7.3.1. For the more general result, use Theorem 7.3.2).

As state space we use $G = [0, t^\dagger[$, with $t^\dagger$ the termination point for the distribution of the V_n. X is then identified with the MPP $(\mathcal{T}, \mathcal{Y}) = ((T_n), (Y_n))$, where the T_n are as above, and $Y_n = X_{T_n}$ is seen to be always 0 on $(T_n < \infty)$; in particular X can be identified with the counting process $N_t = \sum 1_{(T_n \leq t)}$. Hence, when determining the distribution of $(\mathcal{T}, \mathcal{Y})$ 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 Example 3.1.3

$$\overline{P}^{(n)}_{t_1\cdots t_n,0\cdots 0}(t) = \exp\left(-\int_0^{t-t_n} u(s)\,ds\right)$$

and since $\pi^{(n)}_{t_1\cdots t_n t,0\cdots 0} = \varepsilon_0$, and the deterministic behavior of X is described by $X_t = \phi_{t-T_{\langle t\rangle}}(Y_{\langle t\rangle})$ where $\phi_t(y) = y + t$ satisfies (7.28), it follows from Theorem 7.3.1 that X is time-homogeneous Markov with

$$\phi_t(y) = y + t, \quad q(y) = u(y), \quad r(y, \cdot) = \varepsilon_0.$$

With X as defined above, $X_0 \equiv 0$. 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 7.3.1 (b), with $q(y) = u(y)$ and $\phi_t(y) = y + t$, we find

$$\overline{P}^{(0)}_{|x_0}(t) = \exp\left(-\int_0^t u(x_0 + s)\,ds\right), \quad \pi^{(0)}_{t|x_0} = \varepsilon_0, \tag{7.47}$$

while the $P^{(n)}_{z_n|x_0}$ and $\pi^{(n)}_{z_n,t|x_0}$ for $n \geq 1$ do not depend on x_0, hence are exactly as before. In the case where $x_0 > 0$, the sequence (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 $U > x_0$.

Consider again the 0-delayed renewal process $\mathcal{T} = (T_n)$ but assume now that $\mathbb{P}(V_n < \infty) = 1$. The *forward recurrence time process* $\widetilde{X}$ is defined by

$$\widetilde{X}_t = T_{\langle t \rangle + 1} - t,$$

the time until the next renewal. The state space for $\widetilde{X}$ is $\left]0, t^\dagger\right[$ if $\mathbb{P}\left(V_n = t^\dagger\right) = 0$ and $\left]0, t^\dagger\right]$ if $\mathbb{P}\left(V_n = t^\dagger\right) > 0$. 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 $\mathbb{P}(\cdot | T_1 = \widetilde{x}_0)$.

The MPP describing $\widetilde{X}$ is $((T_n), (Y_n))$ with T_n as always the time of the nth 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 $\widetilde{X}_t = \phi_{t-s}(Y_{\langle t \rangle})$, where $\phi_t(y) = y - t$. Since also

$$P^{(0)}_{|\widetilde{x}_0} = \varepsilon_{\widetilde{x}_0}, \qquad P^{(n)}_{z_n|\widetilde{x}_0} = \varepsilon_{t_n + y_n} \quad (n \geq 1),$$

$$\pi^{(n)}_{z_n,t|\widetilde{x}_0}\left(\left]y, \infty\right[\right) = \mathbb{P}(V_1 > y)$$

it follows from Theorem 7.3.2, that $\widetilde{X}$ is time-homogeneous Markov with respect to $(\mathcal{F}^{\widetilde{X}}_t)$: the only problem is with the $P^{(n)}_{z_n|\widetilde{x}_0}$, but writing $\overline{F}_{t_n,y_n}(t) = \overline{P}_{z_n|\widetilde{x}_0}(t)$ we have for $s \leq t$ and all y,

$$\overline{F}_{s,y}(t) = \begin{cases} 1 & \text{if } t < s + y \\ 0 & \text{if } t \geq s + y \end{cases}$$

which depends on s, t through $t - s$ only and satisfies the required functional equation $\overline{F}_{s,y}(u) = \overline{F}_{s,y}(t)\overline{F}_{t,y-(t-s)}(u)$ for $s \leq t \leq u$ and all y. 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: $T_{\langle t \rangle + 1}$ is $\mathcal{F}^{\widetilde{X}}_t$- but not $\mathcal{F}^X_t$-measurable (except of course in the uninteresting case where the the V_n are degenerate, $V_n \equiv v_0$ for some $v_0 > 0$).

Exercise 7.4.1 Assume that $\mathbb{P}(V_n < \infty) = 1$ and consider the two-dimensional process $\left(X, \widetilde{X}\right)$. Discuss what the state space is for this process, and show that it is time-homogeneous Markov.

The assumption $\mathbb{P}(V_n < \infty) = 1$ is made to ensure that $\widetilde{X}$ is well defined as a process with finite values. If $\mathbb{P}(V_n = \infty) > 0$, adjoin ∞ as an extra state, and define $\widetilde{X}_t = T_{\langle t \rangle + 1} - t$ as before with $T_n = \infty$ now possible. The state ∞ is then an absorbing state, reached with certainty (why?) in finite time. Show that $\widetilde{X}$ is still time-homogeneous Markov.

7.4.2 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, independent of N, such that $\mathbb{P}(U_n = 0) = 0$. Let P denote the distribution of the U_n. (This setup was used to describe a compound Poisson process on p.135).

Let $\phi : D \to G = \mathbb{R}^d$ satisfy the conditions of Theorem 7.3.1 (a) and consider the piecewise deterministic process $X = (X_t)_{t\geq 0}$ given by $X_0 \equiv x_0$,

$$X_t = \phi_{T_{\langle t\rangle},t}(Y_{\langle t\rangle})$$

where the U_n are used to determine the jump sizes for X so that for $n \geq 1$,

$$\Delta X_{T_n} = U_n, \quad Y_n = X_{T_n} = \phi_{T_{n-1},T_n}(Y_{n-1}) + U_n$$

with $Y_0 \equiv x_0$. Identifying X with the MPP $((T_n), (Y_n))$, we find

$$\overline{P}^{(n)}_{z_n|x_0}(t) = e^{-\lambda(t-t_n)}$$

and

$$\begin{aligned}\pi^{(n)}_{z_n,t|x_0}(C) &= \mathbb{P}(\phi_{t_n t}(y_n) + U_{n+1} \in C)\\ &= \mathrm{P}(C - \phi_{t_n t}(y_n)).\end{aligned}$$

It follows from Theorem 7.3.1 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 6.2.

7.4.3 A PDMP that solves an SDE

Let N be a homogeneous Poisson process with parameter $\lambda > 0$ and consider the stochastic differential equation (SDE)

$$dX_t = a(X_t)\,dt + \sigma(X_{t-})\,dN_t, \quad X_0 \equiv x_0 \tag{7.48}$$

for an $\mathbb{R}$-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

$$D_t X_t = a(X_t),$$

which should be compared with (7.43), 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 $t \mapsto \phi_{t+s}(x)$, $t \mapsto \phi_t(\phi_s(x))$ defined for $t \geq 0$, it is immediately checked that both solve the differential equation $f'(t) = a(f(t))$, $f(0) = \phi_s(x)$, hence, as expected, $\phi_{t+s} = \phi_t \circ \phi_s$, $\phi_0 = \text{id}$, i.e., (7.28) holds.

It is now clear that (7.48) has a unique solution given by

$$X_t = \phi_{t-T_{\langle t\rangle}}\left(Y_{\langle t\rangle}\right), \quad \Delta X_{T_{n+1}} = \sigma\left(\phi_{T_{n+1}-T_n}(Y_n)\right)$$

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 (7.48)). It is now an easy matter to see that X is a time-homogeneous PDMP: the Markov kernels for $((T_n), (Y_n))$ are

$$\overline{P}^{(n)}_{z_n|x_0}(t) = e^{-\lambda(t-t_n)} \quad (t \geq t_n),$$
$$\pi^{(n)}_{z_n,t|x_0} = \varepsilon_{y_n^*} \quad (t > t_n),$$

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 7.3.1 (b) with ϕ_t as above, $q(y) = \lambda$ for all y, and

$$r(y, \cdot) = \varepsilon_{y+\sigma(y)}. \tag{7.49}$$

The reasoning above works particularly well because N is Poisson, but this implies that the jump times for X are those of N and hence that X is a rather special type of homogeneous PDMP. To obtain solutions X with a more general structure for the jump times one must (apparently) let the driving counting process N in (7.48) depend on X which is a most unpleasant prospect, but nevertheless it is still possible in a certain sense to solve the SDE: suppose q is a given intensity function and let N be given by the Markov kernels $P^{(0)}_{|x_0}$ and $P^{(n)}_{|x_0}$ (for $n \geq 1$ and yet to be determined) such that

$$\overline{P}^{(0)}_{|x_0}(t) = \exp\left(-\int_0^t q(\phi_s(x_0))\right) ds.$$

This distribution generates the first jump time T_1 for N, which is enough to determine the solution process X on the interval $[0, T_1]$, including the state Y_1 reached by the jump at T_1 if T_1 is finite (and determines X completely if $T_1 = \infty$). But then we should obviously use

$$\overline{P}^{(1)}_{T_1|x_0}(t) = \exp\left(-\int_0^{t-T_1} q(\phi_s(Y_1))\,ds\right) \quad (t \geq T_1)$$

for generating T_2 and thereafter X on $]T_1, T_2]$ if $T_2 < \infty$ (if $T_2 = \infty$ we are done). The end result will be a homogeneous PDMP with $q(x)$ the total intensity for a jump from x to occur, solving (7.48) with N the counting process (not Poisson in general) determined by the sequence (T_n) of jump times. Of course X still has deterministic jumps as in (7.49) – for more general jump structures, it is necessary to replace N by an RCM μ with some mark space E, and the term $\sigma(X_{t-})\, dN_t$ in (7.48) by a term $\int_E S_t^y\, \mu(dt, dy)$ for an $\mathcal{F}_t^\mu$-predictable field S for example of the form $S_t^y = \sigma(X_{t-}, y)$: if $\mu = \sum \varepsilon_{(T_n, Y_n)}$ is a homogeneous Poisson random measure, X is then a PDMP with Poisson jump times and, since $\Delta X_{T_n} = \sigma\left(X_{T_n -}, Y_n\right)$,

$$r(y, \cdot) = \text{ law of} by + \sigma(y, Y)$$

where Y has the same distribution as the iid random variables Y_n.

7.5 The strong Markov property

Let X be a PDMP with an arbitrary initial state $X_0 \equiv x_0$, constructed from an RCM μ as in Theorem 7.3.2 using functions ϕ_{st} for the deterministic part, survivor functions $\overline{F}_{s,x}$ for the jump time distributions and transition probabilities r_t for the jumps, with the ϕ_{st} forming a semigroup and the $\overline{F}_{s,x}$ satisfying the functional equation (7.35). With $Q^{s,x}$ as in (7.36), we then have the Markov property (7.40)

$$\mathbb{P}\left(\theta_s \mu \in \cdot \,\middle|\, \mathcal{F}_s^\mu\right) = Q^{s, X_s}, \tag{7.50}$$

and in the homogeneous case, where $\phi_{st} = \phi_{t-s}$, $\overline{F}_{s,x}(t) = \overline{F}_x(t-s)$ and $r_t = r$ and Q^x is defined as $Q^{0,x}$,

$$\mathbb{P}\left(\theta_s^* \mu \in \cdot \,\middle|\, \mathcal{F}_s^\mu\right) = Q^{X_s}. \tag{7.51}$$

The strong Markov property, which we shall now prove, states that in (7.50) and (7.51), the fixed time point s may be replaced by an arbitrary $\mathcal{F}_t^\mu$-stopping time T. Recall (Appendix B) that T is an $\mathcal{F}_t^\mu$-stopping time if $(\mathrm{T} < t) \in \mathcal{F}_t^\mu$ for all t and the pre-T σ-algebra is defined as

$$\mathcal{F}_\mathrm{T}^\mu = \left\{F \in \mathcal{F} : F \cap (\mathrm{T} < t) \in \mathcal{F}_t^\mu \text{ for all } t\right\}.$$

With T an $\mathcal{F}_t^\mu$-stopping time, define the shift $\theta_\mathrm{T}\mu : (\mathrm{T} < \infty) \to \mathcal{M}$ by

$$\theta_\mathrm{T}\mu(\omega) = \sum_{n:\mathrm{T}(\omega) < T_n(\omega) < \infty} \varepsilon_{(T_n(\omega), Y_n(\omega))}$$

and the translated shift $\theta_\mathrm{T}^*\mu : (\mathrm{T} < \infty) \to \mathcal{M}$ by

$$\theta_\mathrm{T}^*\mu(\omega) = \sum_{n:\mathrm{T}(\omega) < T_n(\omega) < \infty} \varepsilon_{(T_n(\omega) - \mathrm{T}(\omega), Y_n(\omega))}.$$

We write $T_{\mathrm{T},n}$ and $Y_{\mathrm{T},n}$ for the jump times and marks determining $\theta_\mathrm{T}\mu$. Thus, if $\mathrm{T}(\omega) < \infty$,

$$\theta_{\mathrm{T}}\mu(\omega) = \sum_{n:T_{\mathrm{T},n}(\omega)<\infty} \varepsilon_{(T_{\mathrm{T},n}(\omega),Y_{\mathrm{T},n}(\omega))}$$

with $\mathrm{T}(\omega) < T_{\mathrm{T},1}(\omega) \leq T_{\mathrm{T},2}(\omega) \leq \cdots$. Note that for $n \geq 1$,

$$T_{\mathrm{T},n} = \tau_n(\theta_{\mathrm{T}}\mu), \quad Y_{\mathrm{T},n} = \eta_n(\theta_{\mathrm{T}}\mu) \tag{7.52}$$

on the set $(\mathrm{T} < \infty)$.

It is straightforward to show that each $T_{\mathrm{T},n}$ is an $\mathcal{F}_t^\mu$-stopping time and that $Y_{\mathrm{T},n}$ is $\mathcal{F}_{T_{\mathrm{T},n}}^\mu$-measurable, as will be used below.

Theorem 7.5.1 (The strong Markov property.) *Let* T *be an* $\mathcal{F}_t^\mu$*-stopping time.*

(i) *It holds* $\mathbb{P}$*-a.s. on* $(\mathrm{T} < \infty)$ *that*

$$\mathbb{P}\left(\theta_{\mathrm{T}}\mu \in \cdot \,\middle|\, \mathcal{F}_{\mathrm{T}}^\mu\right) = Q^{\mathrm{T},X_{\mathrm{T}}}. \tag{7.53}$$

(ii) *If X is time-homogeneous, it holds* $\mathbb{P}$*-a.s. on* $(\mathrm{T} < \infty)$ *that*

$$\mathbb{P}\left(\theta_{\mathrm{T}}^*\mu \in \cdot \,\middle|\, \mathcal{F}_{\mathrm{T}}^\mu\right) = Q^{X_{\mathrm{T}}}. \tag{7.54}$$

Note. Both (7.53) and (7.54) should be understood as identities between probabilities on $(\mathcal{M}, \mathcal{H})$.

Proof. We shall prove (i), (ii) is then an immediate consequence. In order to show (7.53), it suffices to show that for all $H \in \mathcal{H}$,

$$\mathbb{P}\left(\theta_{\mathrm{T}}\mu \in H \,\middle|\, \mathcal{F}_{\mathrm{T}}^\mu\right) = Q^{\mathrm{T},X_{\mathrm{T}}}(H)$$

$\mathbb{P}$-a.s. on $(\mathrm{T} < \infty)$. This in turn follows if we show that for $n \geq 1$ and all measurable and bounded functions $f_i : \overline{\mathbb{R}}_+ \times \overline{G} \to \mathbb{R}$, $1 \leq i \leq n$, it holds that

$$\mathbb{E}\left[\prod_{i=1}^n f_i(\tau_i, \eta_i) \circ \theta_{\mathrm{T}}^\mu \,\middle|\, \mathcal{F}_{\mathrm{T}}^\mu\right] = E^{\mathrm{T},X_{\mathrm{T}}}\left[\prod_{i=1}^n f_i(\tau_i, \eta_i)\right]. \tag{7.55}$$

This we shall prove by induction on n, and here the initial step for $n = 1$ is critical – once that has been done the induction step is easy: suppose (7.55) has been established for $n - 1$ with $n \geq 2$ and all $\mathcal{F}_t^\mu$-stopping times T. Recalling (7.52) and the fact that $T_{\mathrm{T},n-1}$ is a stopping time, applying (7.55) for $n = 1$ gives

$$\mathbb{E}\left[\left(U \circ \theta_{\mathrm{T}}^\mu\right) f_n\left(T_{\mathrm{T},n}, Y_{\mathrm{T},n}\right) \middle| \mathcal{F}_{T_{\mathrm{T},n-1}}^\mu\right] = \left(U \circ \theta_{\mathrm{T}}^\mu\right) E^{T_{\mathrm{T},n-1},Y_{\mathrm{T},n-1}}\left(f_n(\tau_1, \eta_1)\right)$$

where $U = \prod_{i=1}^{n-1} f_i(\tau_i, \eta_i)$. Therefore

$$\begin{aligned}\mathbb{E}\left[\left(U \circ \theta_{\mathrm{T}}^\mu\right) f_n\left(T_{\mathrm{T},n}, Y_{\mathrm{T},n}\right) \middle| \mathcal{F}_{\mathrm{T}}^\mu\right] &= \mathbb{E}\left[\left(U \circ \theta_{\mathrm{T}}^\mu\right) E^{T_{\mathrm{T},n-1},Y_{\mathrm{T},n-1}}\left(f_n(\tau_1, \eta_1)\right) \middle| \mathcal{F}_{\mathrm{T}}^\mu\right] \\ &= E^{\mathrm{T},X_{\mathrm{T}}}\left[U E^{\tau_{n-1},\eta_{n-1}}\left(f_n(\tau_1, \eta_1)\right)\right],\end{aligned}$$

with the last equality following from the induction hypothesis. But by Lemma 4.3.3(bii) and the definition of the $Q^{s,x}$,

$$E^{\tau_{n-1},\eta_{n-1}}\left[f_n\left(\tau_1,\eta_1\right)\right] = E^{\mathrm{T},X_\mathrm{T}}\left[f_n\left(\tau_n,\eta_n\right)|\xi_{n-1}\right]$$

since for $n \geq 2$ both quantities equal

$$\int_{]\tau_{n-1},\infty]} F_{\tau_{n-1},\eta_{n-1}}\left(dt\right) \int_G r_t\left(\phi_{\tau_{n-1}t}\left(\eta_{n-1}\right), dy\right)\ f_n\left(t,y\right),$$

and we see that (7.55) holds for n.

It remains to establish (7.55) for $n = 1$. First approximate T by a sequence of discrete valued stopping times T_K, where for $K \geq 1$,

$$\mathrm{T}_K = \sum_{k=1}^{\infty} \tfrac{k}{2^K} 1_{\left((k-1)2^{-K} \leq \mathrm{T} < k2^{-K}\right)} + \infty 1_{(\mathrm{T}=\infty)}.$$

That (7.53) holds with T replaced by T_K is a direct consequence of the Markov property (7.50): let $H \in \mathcal{H}$, $F \in \mathcal{F}^{\mu}_{\mathrm{T}_K}$ and write $F_{K,k} = F \cap \left(\mathrm{T}_K = k2^{-K}\right) \in \mathcal{F}^{\mu}_{k2^{-K}}$. Then, with $t_{Kk} = k2^{-K}$,

$$\begin{aligned}
\int_{F_{K,k}} Q^{\mathrm{T}_K,X_{\mathrm{T}_K}}\left(H\right)\, d\mathbb{P} &= \int_{F_{K,k}} Q^{t_{Kk},X_{t_{Kk}}}\left(H\right) d\mathbb{P} \\
&= \int_{F_{K,k}} 1_{\left(\theta_{t_{Kk}}\mu \in H\right)}\, d\mathbb{P} \\
&= \mathbb{P}\left(F_{k,K} \cap \left(\theta_{\mathrm{T}_K} \in H\right)\right)
\end{aligned}$$

and

$$\mathbb{P}\left(\theta_{\mathrm{T}_K} \in H \,\middle|\, \mathcal{F}^{\mu}_{\mathrm{T}_K}\right) = Q^{\mathrm{T}_K,X_{\mathrm{T}_K}}\left(H\right)$$

follows (on the set $(\mathrm{T}_K < \infty) = (\mathrm{T} < \infty)$).

We now prove (7.55) for $n = 1$ by first noting that on $(\mathrm{T} < \infty)$,

$$E^{\mathrm{T},X_\mathrm{T}}\left(f_1\left(\tau_1,\eta_1\right)\right) = \int_{]\mathrm{T},\infty]} F_{\mathrm{T},X_\mathrm{T}}\left(dt\right) \int_G r_t\left(\phi_{\mathrm{T}t}\left(X_\mathrm{T}\right), dy\right)\ f_1\left(t,y\right)$$

with a similar expression for $E^{\mathrm{T}_K,X_{\mathrm{T}_K}}\left(f_1\left(\tau_1,\eta_1\right)\right)$. But if $\overline{N}_{\mathrm{T}_K} = \overline{N}_\mathrm{T}$ we have for $t \geq \mathrm{T}_K > \mathrm{T}$,

$$\phi_{\mathrm{T}t}\left(X_\mathrm{T}\right) = \phi_{\mathrm{T}_K t}\left(\phi_{\mathrm{T}\mathrm{T}_K}\left(X_\mathrm{T}\right)\right) = \phi_{\mathrm{T}_K t}\left(X_{\mathrm{T}_K}\right)$$

and

$$\overline{F}_{\mathrm{T},X_\mathrm{T}}\left(t\right) = \overline{F}_{\mathrm{T},X_\mathrm{T}}\left(\mathrm{T}_K\right) \overline{F}_{\mathrm{T}_K,X_{\mathrm{T}_K}}\left(t\right)$$

whence, on the set $\left(\overline{N}_{\mathrm{T}_K} = \overline{N}_\mathrm{T}, \mathrm{T} < \infty\right)$,

$$
\begin{aligned}
E^{\mathrm{T}_K, X_{\mathrm{T}_K}}\left(f_1\left(\tau_1, \eta_1\right)\right) &= \int_{]\mathrm{T}_K, \infty]} F_{\mathrm{T}_K, X_{\mathrm{T}_K}}(dt) \int_G r_t\left(\phi_{\mathrm{T}_K t}\left(X_{\mathrm{T}_K}\right), dy\right) f_1(t, y) \\
&= \frac{1}{\overline{F}_{\mathrm{T}, X_{\mathrm{T}}}\left(\mathrm{T}_K\right)} \int_{]\mathrm{T}_K, \infty]} F_{\mathrm{T}, X_{\mathrm{T}}}(dt) \int_G r_t\left(\phi_{\mathrm{T}t}\left(X_{\mathrm{T}}\right), dy\right) f_1(t, y) \\
&\to E^{\mathrm{T}, X_{\mathrm{T}}}\left(f_1\left(\tau_1, \eta_1\right)\right)
\end{aligned}
$$

as $K \uparrow \infty$ since $\mathrm{T}_K \downarrow \mathrm{T}$. Since also $1_{\left(\overline{N}_{\mathrm{T}_K}=\overline{N}_{\mathrm{T}}, \mathrm{T}<\infty\right)} \to 1_{(\mathrm{T}<\infty)}$, we finally deduce using dominated convergence that for $F \in \mathcal{F}_{\mathrm{T}}^{\mu} \cap (\mathrm{T} < \infty)$, writing $F_K = F \cap \left(\overline{N}_{\mathrm{T}_K} = \overline{N}_{\mathrm{T}}\right)$,

$$
\begin{aligned}
\int_F E^{\mathrm{T}, X_{\mathrm{T}}}\left(f_1\left(\tau_1, \eta_1\right)\right) d\mathbb{P} &= \lim_{K \to \infty} \int_{F_K} E^{\mathrm{T}_K, X_{\mathrm{T}_K}}\left(f_1\left(\tau_1, \eta_1\right)\right) d\mathbb{P} \\
&= \lim_{K \to \infty} \int_{F_K} \left(f_1\left(\tau_1, \eta_1\right)\right) \circ \left(\theta_{\mathrm{T}_K} \mu\right) d\mathbb{P} \\
&= \lim_{K \to \infty} \int_{F_K} f_1\left(T_{\mathrm{T}_K,1}, Y_{\mathrm{T}_K,1}\right) d\mathbb{P} \\
&= \int_F f_1\left(T_{\mathrm{T},1}, Y_{\mathrm{T},1}\right) d\mathbb{P}
\end{aligned}
$$

using that $\left(T_{\mathrm{T}_K,1}, Y_{\mathrm{T}_K,1}\right)(\omega) = \left(T_{\mathrm{T},1}, Y_{\mathrm{T},1}\right)(\omega)$ for $K = K(\omega)$ sufficiently large for the last equality. □

Example 7.5.1 If we apply the strong Markov property to the $\mathcal{F}_t^{\mu}$-stopping time T_{k_0}, the result is a special case of Lemma 4.3.3(b) (special because now we have the additional structure arising from the PDMP X).

7.6 Itô's formula for homogeneous PDMPs

In this section we shall discuss in more detail the structure of $\mathbb{R}^d$-valued piecewise deterministic Markov processes, which are time-homogeneous.

The processes will be of the type described in Theorem 7.3.1 (b). Thus, if $X = (X)_{t \geq 0}$ denotes the process with a subset $G \subset \mathbb{R}^d$ as state space, X is completely specified by its initial value $X_0 \equiv x_0$ and the MPP $((T_n), (Y_n))$, where T_n is the time of the nth discontinuity (nth jump) for X, and $Y_n = X_{T_n} \in \mathbb{R}^d$ is the state reached by the nth jump, cf. (7.27), and

$$X_t = \phi_{t - T_{\langle t \rangle}}(Y_{\langle t \rangle}),$$

with ϕ describing the deterministic behavior of X between jumps so that $t \mapsto \phi_t(y)$ 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. (7.28). Recall also that the distribution of $((T_n), (Y_n))$ is determined by

$$\overline{P}^{(n)}_{z_n|x_0}(t) = \exp\left(-\int_0^{t-t_n} q(\phi_s(y_n))\, ds\right),$$
$$\pi^{(n)}_{z_n,t|x_0}(C) = r(\phi_{t-t_n}(y_n), C)$$

with $q \geq 0$ and such that $t \mapsto q(\phi_t(y))$ is a hazard function for every y, and with r a Markov kernel on G such that, cf. (7.29),

$$r(y, \{y\}) = 0 \quad (y \in G). \tag{7.56}$$

If μ is the RCM determined by $((T_n), (Y_n))$, we recall that μ has $\mathcal{F}_t^{\mu}$- and $\mathcal{F}_t^X$-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 (7.34),

$$L(ds, dy) = q(X_{s-})\, ds\, r(X_{s-}, dy). \tag{7.57}$$

Finally, we shall assume that $t \mapsto \phi_t(y)$ is continuously differentiable, and writing

$$a(y) = D_t\phi_t(y)|_{t=0}$$

as a column vector with coordinates $a_i(y)$, we then have the differential equation (7.44),

$$D_t\phi_t(y) = a(\phi_t(y)).$$

Suppose now that $f : \mathbb{R}_0 \times G \to \mathbb{R}$ is a continuous function, and consider the process $(f(t, X_t))_{t\geq 0}$. We may then use Itô's formula for MPPs, Section 4.7, to obtain a decomposition of $f(\cdot, X)$ into a local martingale and a predictable process. To formulate the result, introduce the *space-time generator* $\mathcal{A}$ acting on the space $\mathcal{D}(\mathcal{A})$ of functions f that are *bounded on finite time intervals* and continuously differentiable in t and y and with $\mathcal{A}f$ given by

$$\mathcal{A}f(t, y) = D_t f(t, y) + \sum_{i=1}^{d} D_{y_i} f(t, y) a_i(y) + q(y) \int_G r(y, dy')\,(f(t, y') - f(t, y)). \tag{7.58}$$

That f is bounded on finite time intervals means that f is bounded on all sets $[0, t] \times G$ for $t \in \mathbb{R}_0$.

Theorem 7.6.1 (a) *For $f \in \mathcal{D}(\mathcal{A})$ the process $f(\cdot, X)$ may be written*

$$f(t, X_t) = f(0, x_0) + U_t + \int_{]0,t]\times G} S_s^y\, M(ds, dy), \tag{7.59}$$

where M is the martingale measure $\mu - L$, $(S_s^y)_{s\geq 0, y\in G}$ is the $\mathcal{F}_t^X$-predictable field given by

$$S_t^y = f(t, y) - f(t, X_{t-}) \tag{7.60}$$

and U is continuous and $\mathcal{F}_t^X$-predictable with

$$U_t = \int_0^t \mathcal{A} f(s, X_s)\, ds. \tag{7.61}$$

The term

$$M_t(S) = \int_{]0,t]\times G} S_s^y\, M(ds, dy)$$

in (7.59) defines a local $\mathcal{F}_t^\mu$-martingale with reducing sequence (T_n), and the decomposition of $f(t, X_t)$ into a local martingale and an $\mathcal{F}_t^\mu$-predictable process is unique up to indistinguishability.

(b) *If $f \in \mathcal{D}(\mathcal{A})$ is such that $\mathcal{A}f$ is bounded on finite time intervals, then the process $M(S)$ is an $\mathcal{F}_t^\mu$-martingale and for all t,*

$$\begin{aligned} \mathbb{E}_{|x_0} f(t, X_t) &= f(0, x_0) + \mathbb{E}_{|x_0} \int_0^t \mathcal{A} f(s, X_s)\, ds \\ &= f(0, x_0) + \int_0^t \mathbb{E}_{|x_0} \mathcal{A} f(s, X_s)\, ds. \end{aligned} \tag{7.62}$$

If in particular

$$\mathcal{A} f \equiv 0, \tag{7.63}$$

then the process $(f(t, X_t))_{t\geq 0}$ is an $\mathcal{F}_t^\mu$-martingale and

$$\mathbb{E}_{|x_0} f(t, X_t) = f(0, x_0) \quad (t \geq 0).$$

Note. We write $\mathbb{E}_{|x_0}$ rather than $\mathbb{E}$ for expectation in order to emphasize the fixed, but arbitrary value of the arbitrary initial state x_0.

Proof. (a). As in Section 4.7 we identify S by identifying the jumps in (7.59). 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 (7.59) 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 (7.60).

Note that because f and therefore also S is bounded on finite time intervals, with $|S| \leq K$ we have

$$\int_{]0,t]\times G} \left| S_s^y \right| L(ds, dy) \leq K \overline{\Lambda}_t < \infty$$

$\mathbb{P}$-a.s. The convergence of this stochastic integral is used without comment below.

Having found S, we define U by solving (7.59) for U_t. It then follows that between jumps U is differentiable in t (at least in the sense that increments between jumps, $U_t - U_s$, have the form $\int_s^t D_u U_u \, du$) and that $D_t U_t$ is given by

$$D_t U_t = D_t f(t, X_t) + D_t \int_{]0,t]\times G} S_s^y \, L(ds, dy).$$

But

$$D_t f(t, X_t) = D_t f(t, X_t) + \sum_{i=1}^{d} D_{y_i} f(t, X_t) \, D_t X_t^i$$

with X_t^i the ith coordinate of X^i, and by (7.44), between jumps

$$D_t X_t = D_t \phi_{t-T_{\langle t\rangle}} \left(Y_{\langle t\rangle}\right) = a\left(\phi_{t-T_{\langle t\rangle}}\left(Y_{\langle t\rangle}\right)\right) = a(X_t).$$

Since also, by (7.57)

$$\begin{aligned} D_t \int_{]0,t]\times G} S_s^y \, L(ds, dy) &= D_t \int_0^t ds \, q(X_{s-}) \int_G r(X_{s-}, dy) \, S_s^y \\ &= q(X_t) \int_G r(X_t, dy) \, S_t^y, \end{aligned}$$

using (7.60) it follows that between jumps

$$D_t U_t = \mathcal{A} f(t, X_t)$$

and we have shown that (7.59) holds with S, U given by (7.60), (7.61).

Because S is bounded on all sets $[0, t]\times G$, by Theorem 4.6.1 (iii1) $M(S)$ is an $\mathcal{F}_t^\mu$-local martingale with reducing sequence (T_n). The uniqueness of the decomposition follows from Proposition 4.5.1.

(b) (7.59) together with the boundedness of f and $\mathcal{A}f$ on finite time intervals shows the local martingale $M(S)$ to be uniformly bounded on $[0, t]$ for any t, hence $M(S)$ is a true martingale with, necessarily, expectation 0. The remaining assertions are now obvious. □

A generalization of the theorem appears as Proposition 7.7.1 below.

Remark 7.6.1 For (a) the assumption that f be bounded on finite time intervals may be relaxed: it suffices that for all t

$$\sup_{s\le t, y} \int_G r\left(y, dy'\right) \left|f(s, y') - f(s, y)\right| < \infty.$$

The conclusion in (b) may of course hold even without assuming that f and $\mathcal{A}f$ are bounded on finite time intervals, but then great care should be used when verifying first that $M(S)$ is a local martingale and second that it is a true martingale.

Suppose that $h : G \to \mathbb{R}$ 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

$$Ah(y) = \sum_{i=1}^{d} D_{y_i} h(y) a_i(y) + q(y) \int_G r(y, d\widetilde{y}) \left(h(\widetilde{y}) - h(y)\right), \tag{7.64}$$

while (7.59) takes the form

$$h(X_t) = h(x_0) + \int_0^t Ah(X_s)\, ds + \int_{]0,t]\times G} S_s^y\, M(ds, dy)$$

with

$$S_s^y = h(y) - h(X_{s-}).$$

The operator A, acting e.g., on a suitable subspace of the space of bounded h, such that $D_y h$ exists and is continuous with Ah continuous, is a version of the *infinitesimal generator* for the time-homogeneous Markov process X. If also Ah is bounded (as is usually assumed when defining the domain of the generator), by (7.62)

$$\mathbb{E}_{|x_0} h(X_t) = h(x_0) + \int_0^t \mathbb{E}_{|x_0} Ah(X_s)\, ds,$$

from which one obtains, when using the continuity of Ah, the familiar formula for the generator,

$$\begin{aligned} Ah(x_0) &= \lim_{t\to 0} \frac{1}{t}\left(\mathbb{E}_{|x_0} h(X_t) - h(x_0)\right) \\ &= \lim_{t\to 0} \frac{1}{t}\left(\mathrm{P}_t h(x_0) - h(x_0)\right), \end{aligned} \tag{7.65}$$

which is valid for arbitrary initial states x_0.

In the next section we shall discuss in detail a specific version of the generator, what we refer to as the full generator, where one essential aspect is to extend the domain to allow for certain functions that need not be continuously differentiable.

Remark 7.6.2 It is not difficult to generalize the version (7.59) of Itô's formula to non-homogeneous PDMPs: suppose that

$$X_t = \phi_{T_{\langle t\rangle},t}\left(Y_{\langle t\rangle}\right)$$

where the ϕ_{st} satisfy the semigroup property (7.25) together with (7.26) and that furthermore the Markov kernels generating the jump times and jumps for X are of the form

$$\begin{aligned} \overline{P}^{(n)}_{z_n|x_0}(t) &= \exp\left(-\int_{t_n}^t q_s(\phi_{t_n s}(y_n))\, ds\right), \\ \pi^{(n)}_{z_n,t|x_0}(C) &= r_t(\phi_{t_n t}(y_n), C). \end{aligned}$$

Then provided $t \mapsto \phi_{st}(x)$ is continuously differentiable on $[s, \infty[$ for all x, one may obtain, for functions f that are bounded on finite time intervals and otherwise smooth enough, decompositions of the form

$$f(t, X_t) = f(0, x_0) + \int_0^t \mathcal{A}_s f(s, X_s)\, ds + \int_{]0,t]\times G} S_s^y\, M(ds, dy)$$

with S as in (7.60). Here

$$\begin{aligned}\mathcal{A}_t f(t, y) = D_t f(t, y) &+ \sum_{i=1}^d D_{y_i} f(t, y) a_i(t, y) \\ &+ q_t(y) \int_G r_t(y, dy')\,(f(t, y') - f(t, y))\end{aligned}$$

where $a_i(t, y)$ is the ith element of the column vector

$$D_u \phi_{tu}(y)\,|_{u=t}\ .$$

We shall conclude with a brief discussion of Itô's formula and the structure of the generator for general homogeneous PDMPs constructed as in Theorem 7.3.2, i.e., the conditional distributions $P^{(n)}_{z_n|x_0}$ of the jump times may have atoms and need certainly not have densities, as has been otherwise assumed so far in this section.

For such a general PDMP X it is still possible to establish an Itô formula, e.g., for $h : G \to \mathbb{R}$ one attempts to write

$$h(X_t) = h(X_0) + U_t + \text{local martingale}_t \tag{7.66}$$

with U $\mathcal{F}_t^X$-predictable, $U_0 \equiv 0$. Since the compensating measure L for $((T_n), (Y_n))$ may now have jumps, typically also U will have jumps, and to find the decomposition (7.66) one must argue as in Examples 4.7.3 and 4.7.4. The decomposition is particularly nice if U is continuous, even absolutely continuous, so that for some function Ah,

$$h(X_t) = h(X_0) + \int_0^t Ah(X_s)\, ds + \text{local martingale}_t. \tag{7.67}$$

This requires that h have a special structure and one would then call the space of such h with h, Ah both bounded and the local martingale a true martingale, for the domain of the *extended generator* with of course Ah the value of the generator applied to h.

We shall not pursue this general discussion here but illustrate what happens by an example.

Example 7.6.1 Let X be the forward recurrence time process from Subsection 7.4.1, i.e., $X_t = T_{\langle t\rangle+1} - t$ where $T_n = V_1 + \cdots + V_n$ is the sum of independent waiting times with the V_k identically distributed for $k \geq 2$ according to an arbitrary given distribution F on $\mathbb{R}_+$ and where $N_t = \sum_1^\infty 1_{(T_n \leq t)}$. Starting X at a given state $x_0 \in G$ with $x_0 > 0$ (where $G = \left]0, t^\dagger\right[$ if $\mathbb{P}\left(V_n = t^\dagger\right) = 0$ and $G = \left]0, t^\dagger\right]$ if $\mathbb{P}\left(V_n = t^\dagger\right) > 0$,

$t^\dagger$ denoting the termination point for the distribution of the V_n) corresponds to the PDMP-description $\phi_t(y) = y - t$, $P^{(0)}_{|x_0} = \varepsilon_{x_0}$, $P^{(n)}_{z_n|x_0} = \varepsilon_{t_n + y_n}$ for $n \geq 1$ and $\pi^{(n)}_{z_n,t|x_0}(]y,\infty[) = \overline{F}(y)$, as shown in Subsection 7.4.1.

Suppose now that $h : G \to \mathbb{R}$ is continuous and such that the limit $h(0) = \lim_{x\downarrow 0} h(x)$ exists. We wish to determine for which h we can write $h(X_t)$ in the form (7.67) and to identify Ah. Note first that with $X_0 \equiv x_0$, the compensating measure for the MPP $((T_n),(Y_n))$ (with $Y_n = T_{n+1} - T_n = V_{n+1}$ a.s.) is given by

$$L(dt,dy) = \sum_{n=0}^{\infty} \varepsilon_{T_n+Y_n}(dt)\ F(dy)$$
$$= \sum_{n=0}^{\infty} \varepsilon_{T_{n+1}}(dt)\ F(dy)$$

a.s., and consequently, any $\mathcal{F}^X_t$-martingale has the form

$$M_t(S) = \int_{]0,t]\times G} S^y_s\,(\mu(ds,dy) - L(ds,dy))$$
$$= \sum_{n=1}^{\overline{N}_t}\left(S^{Y_n}_{T_n} - \int_G S^y_{T_n}\ F(dy)\right)$$

for some predictable field $\left(S^y_s\right)$. In particular $M(S)$ is constant between jumps and

$$\Delta M_t(S) = \Delta\overline{N}_t\left(S^{Y\langle t\rangle}_t - \int_G S^y_t\ F(dy)\right). \tag{7.68}$$

Now, because h is continuous $h(X)$ jumps only when $\overline{N}$ does and

$$\Delta h(X_t) = \Delta\overline{N}_t\left(h\left(Y_{\overline{N}_t}\right) - h(0)\right) \tag{7.69}$$

since for all n, $X_{T_n-} = 0$. Matching the jumps from (7.68) and (7.69) show that they are equal if $S^y_t := h(y)$ and $h(0) = \int_G h(y)\ F(dy)$. Since for $t \in \left[T_n, T_{n+1}\right[$, $h(X_t) = h(Y_n - t)$ it is now clear that the following decomposition holds: suppose h belongs to $\mathcal{D}(A)$ defined as the space of bounded continuously differentiable functions h with bounded derivative h' and such that $h(0) = \int_G h(y)\ F(dy)$. Then, under any $\mathbb{P}_{|x_0}$,

$$h(X_t) = h(x_0) - \int_0^t h'(X_s)\ ds + M_t(S)$$

with $M(S)$ a martingale. $\mathcal{D}(A)$ is the domain of the extended generator and $Ah = -h'$ for $h \in \mathcal{D}(A)$. In particular for $h \in \mathcal{D}(A)$ and all $x \in G$,

$$\mathrm{P}_t h(x) = h(x) - \int_0^t \mathrm{P}_s h'(x)\ ds. \tag{7.70}$$

If instead of the forward recurrence time process we consider the joint process $\mathbf{X} = \left(X, \widetilde{X}\right)$ with X the backward and $\widetilde{X}$ the forward recurrence time process (see Exercise 7.4.1) a similar reasoning applies: the state space is $G^{(2)} = \left[0, t^{\dagger}\right[\times \left]0, t^{\dagger}\right[$ (if $\mathbb{P}\left(V_n = t^{\dagger}\right) = 0$) or $G^{(2)} = \left[0, t^{\dagger}\right] \times \left]0, t^{\dagger}\right]$ (if $\mathbb{P}\left(V_n = t^{\dagger}\right) > 0$), we have $\phi_t\left(x, \widetilde{x}\right) = \left(x + t, \widetilde{x} - t\right)$ and, corresponding to an initial state $\mathbf{x}_0 = \left(x_0, \widetilde{x}_0\right) \in G^{(2)}$ with $\widetilde{x}_0 > 0$, we find writing

$$\mathbf{z}_n = \left(t_1, \ldots, t_n; \mathbf{y}_1, \ldots, \mathbf{y}_n\right)$$

that

$$P^{(0)}_{|\mathbf{x}_0} = \varepsilon_{\widetilde{x}_0}, \qquad P^{(n)}_{\mathbf{z}_n|\mathbf{x}_0} = \varepsilon_{t_n + \widetilde{y}_n}, \qquad \pi^{(n)}_{\mathbf{z}_n, t|\mathbf{x}_0}\left(\{0\} \times \left]y, \infty\right[\right) = \overline{F}\left(y\right)$$

with the only relevant marks $\mathbf{y}_n = \left(y_n, \widetilde{y}_n\right)$ having the form $\left(0, \widetilde{y}_n\right)$.

The local martingales are essentially as those above,

$$M_t\left(S\right) = \sum_{n=1}^{\overline{N}_t}\left(S^{\left(0, \widetilde{Y}_n\right)}_{T_n} - \int_G S^{\left(0, \widetilde{y}\right)}_{T_n} F\left(d\widetilde{y}\right)\right)$$

with

$$\Delta M_t(S) = \Delta \overline{N}_t\left(S^{\left(0, \widetilde{Y}_n\right)}_t - \int_G S^{\left(0, \widetilde{y}\right)}_t F\left(d\widetilde{y}\right)\right)$$

and for $h : G^{(2)} \to \mathbb{R}$ continuous,

$$\Delta h\left(X_t, \widetilde{X}_t\right) = \Delta \overline{N}_t\left(h\left(0, \widetilde{Y}_{\overline{N}_t}\right) - h\left(X_{t-}, 0\right)\right).$$

Consequently, taking $S^{\left(0, \widetilde{y}\right)}_t = h\left(0, \widetilde{y}\right)$ the jumps for $M\left(S\right)$ and $h\left(X, \widetilde{X}\right)$ will match if $h\left(x, 0\right) = h\left(0, 0\right)$ for all x and $h\left(0, 0\right) = \int_G h\left(0, \widetilde{y}\right)\ F\left(d\widetilde{y}\right)$. It follows that

$$h\left(X_t, \widetilde{X}_t\right) = h\left(x_0, \widetilde{x}_0\right) + \int_0^t Ah\left(X_s, \widetilde{X}_s\right) ds + M_t\left(S\right)$$

with $M\left(S\right)$ a martingale provided h belongs to the domain $\mathcal{D}\left(A^{(2)}\right)$ consisting of all bounded and continuously differentiable h such that

$$A^{(2)}h = \left(D_x - D_{\widetilde{x}}\right) h$$

is bounded and $h\left(x, 0\right) = h\left(0, 0\right) = \int_G h\left(0, \widetilde{y}\right)\ F\left(d\widetilde{y}\right)$ for all $x \in \left[0, t^{\dagger}\right[$.

7.7 The full infinitesimal generator

We continue the study of a time-homogeneous PDMP X with state space $G \subset \mathbb{R}^d$ of the form considered in the previous section: $X_t = \phi_{t - T_{\langle t \rangle}}\left(Y_{\langle t \rangle}\right)$ with the ϕ_t forming a semigroup under composition, $q(y)$ the intensity for a jump from y and $r(y, \cdot)$ the distribution of the state reached by that jump. In particular all the T_n have a distribution with a density with respect to Lebesgue measure on $\mathbb{R}_+$ (plus possibly an atom at ∞).

We shall assume that the function $t \mapsto \phi_t(y)$ is continuous for all y, but it need not be differentiable, as assumed in the discussion of the generator in the previous section. To ease the notation we shall often write $\phi_t y$ instead of $\phi_t(y)$.

We shall call a measurable function $h : G \to \mathbb{R}$ *path-continuous* if the function

$$t \mapsto h(\phi_t y) \tag{7.71}$$

is continuous for all y, and *path-differentiable* if the function given by (7.71) is continuously differentiable.

If h is continuous, it is of course also path-continuous since by assumption $\phi_t(y)$ is continuous in t. In the one-dimensional case, $d = 1$, path-continuity may well (but not always, see below) be equivalent to continuity, but if $d \geq 2$ there will typically be functions that are path-continuous but not continuous: path-continuity of h amounts only to requiring $h(y_n) \to h(y)$ when $y_n \to y$ in a specified manner, e.g., in the form $y_n = h\left(\phi_{t_n} y\right)$ with $t_n \downarrow 0$.

The concept of path-continuity depends on the semigroup (ϕ_t). The special case of Markov chains, i.e., $\phi_t(y) = y$ for all t, y, is exceptional: by (7.71) *any* measurable function h is path-continuous!

If h is continuously differentiable and $t \mapsto \phi_t(y)$ is continuously differentiable so that, see (7.44), $D_t \phi_t(y) = a(\phi_t y)$, it follows that h is path-differentiable with

$$D_t h(\phi_t y) = \sum_{i=1}^{d} D_{y_i} h(\phi_t y)\, a_i(\phi_t y). \tag{7.72}$$

Again, for $d = 1$ path-differentiability of h often implies differentiability, but for $d \geq 2$ it does not. And the Markov chain case $\phi_t(y) = y$ is still exceptional: any measurable h is path-differentiable.

Suppose now that h is path-differentiable. Then for all $t \geq 0$,

$$\begin{aligned} D_t h(\phi_t y) &= \lim_{s \downarrow 0} \frac{1}{s}\left(h(\phi_{s+t} y) - h(\phi_t y)\right) \\ &= \lim_{s \downarrow 0} \frac{1}{s}\left(h(\phi_s(\phi_t y)) - h(\phi_t y)\right). \end{aligned}$$

For $t = 0$ we obtain the limit

$$\delta_\phi(h)(y) := \lim_{s \downarrow 0} \frac{1}{s}\left(h(\phi_s y) - h(y)\right) \tag{7.73}$$

and then see that

$$D_t h(\phi_t y) = \delta_\phi h(\phi_t y) \quad (t \geq 0). \tag{7.74}$$

For the remainder of this section, assume that the PDMP X has the following path-continuity property: *for any bounded and path-continuous h it holds that*

$$t \mapsto q(\phi_t y) \int_G r(\phi_t y, d\tilde{y})\, h(\tilde{y}) \tag{7.75}$$

is continuous. Taking $h \equiv 1$ it follows in particular that then q *is path-continuous.*

Definition 7.7.1 The *full infinitesimal generator* for the homogeneous PDMP X, is the linear operator A given by (7.76) below acting on the *domain* $\mathcal{D}(\mathrm{A})$ of bounded measurable functions $h : G \to \mathbb{R}$ such that

(i) h is path-differentiable with $\delta_\phi h$ path-continuous;
(ii) the function $\mathrm{A}h : G \to \mathbb{R}$ given by

$$\mathrm{A}h(y) = \delta_\phi h(y) + q(y) \int_G r(y, d\widetilde{y})\, (h(\widetilde{y}) - h(y)) \tag{7.76}$$

is bounded.

Note that if $h \in \mathcal{D}(\mathrm{A})$, then by (i) and (7.75), $\mathrm{A}h$ is path-continuous.

Suppose that h is bounded, continuously differentiable and that $t \mapsto \phi_t(y)$ is continuously differentiable. Then $h \in \mathcal{D}(\mathrm{A})$ provided $\mathrm{A}h$ is bounded and $\mathrm{A}h = Ah$ with Ah given by (7.64), cf. (7.74) and (7.72).

Note that always $h \equiv 1$ belongs to $\mathcal{D}(\mathrm{A})$ with $\mathrm{A}h \equiv 0$.

Example 7.7.1 Suppose X is a Markov chain so that $\phi_t(y) = y$. Then both (7.75) and (i) from Definition 7.7.1 are automatic for any bounded and measurable $h,$ and it follows that $h \in \mathcal{D}(\mathrm{A})$ iff h is bounded and measurable with $\mathrm{A}h$ bounded, where

$$\mathrm{A}h(y) = q(y) \int_G r(y, d\widetilde{y})\, (h(\widetilde{y}) - h(y)). \tag{7.77}$$

The general theory of (time-homogeneous) Markov processes contains various definitions of the generator and in particular of its domain. For most standard definitions the domain would consist entirely of continuous functions, and one would presuppose that the transition operators

$$\mathrm{P}_t h(x) = \mathbb{E}_{|x}\, h(X_t) = \int h\left(X_t^\circ\right) dQ^x$$

(see (7.37) for the Markov kernels defining Q^x) for $t \geq 0$ should map any bounded and continuous function h into a continuous function $\mathrm{P}_t h$. Here (see Theorem 7.7.4 below) we shall only use that path-continuity is preserved. The difference between the two approaches is of course particularly startling in the Markov chain case: as we saw in Example 7.7.1, for our definition of $\mathcal{D}(\mathrm{A})$ it is certainly not required that $h \in \mathcal{D}(\mathrm{A})$ be continuous.

It is of course advantageous to have the domain of the generator as large as possible. At the same time though, it is important that the domain should have the property that if h belongs to the domain, so should $\mathrm{P}_t h$ for any t. This fact we establish for the domain $\mathcal{D}(\mathrm{A})$ in Theorem 7.7.4 below. First, however, we give a generalization of Itô's formula (7.59). Recall that $f : \mathbb{R}_0 \times G \to \mathbb{R}$ is said to be bounded on finite time intervals if it is bounded on all sets $[0, t] \times G$ for $t \in \mathbb{R}_0$.

Proposition 7.7.1 *Suppose that $f : \mathbb{R}_0 \times G \to \mathbb{R}$ is bounded on finite time intervals and measurable with $t \mapsto f(t, x)$ continuously differentiable for all x and $x \mapsto f(t, x)$ path-differentiable for all t. Then*

$$f(t, X_t) = f(0, X_0) + \int_0^t \mathcal{A} f(s, X_s)\, ds + \int_{]0,t] \times G} S_s^y\, M(ds, dy) \tag{7.78}$$

where

$$\mathcal{A} f(t, x) = D_t f(t, x) + \delta_{\phi,x} f(t, x) + q(x) \int_G r(x, dy)\, (f(t, y) - f(t, x))$$

and where $\left(S_t^y\right)$ is the predictable field

$$S_t^y = f(t, y) - f(t, X_{t-})$$

and M is the martingale measure determined from the MPP $((T_n), (Y_n))$. In the decomposition (7.78), the stochastic integral $M(S)$ is a local $\mathcal{F}_t^{\mu}$-martingale.

Furthermore, if $h \in \mathcal{D}(\mathrm{A})$, then

$$\mathrm{P}_t h(x) = h(x) + \int_0^t \mathrm{P}_s(\mathrm{A}h)(x)\, ds. \tag{7.79}$$

Note. $\delta_{\phi,x} f(t, \cdot)$ is the function obtained by path-differentiating $f(t, x)$ as a function of x.

Proof. For the first assertion, just repeat the proof of Theorem 7.6.1(a), using that by the assumptions on f, the process $(f(t, X_t))_{t \geq 0}$ jumps only when the RCM μ does and that between jumps $t \mapsto f(t, X_t) = f\left(t, \phi_{t-T_{\langle t \rangle}}\left(Y_{\langle t \rangle}\right)\right)$ is continuously differentiable with derivative

$$D_t f(t, X_t) + \delta_{\phi,x} f(t, X_t).$$

Next, arguing exactly as in the proof of Theorem 7.6.1(b) one sees that when $f(t, x) = h(x)$ with $h \in \mathcal{D}(\mathrm{A})$, the last term on the right of (7.78) is a martingale under $\mathbb{P}_{|x}$ for any x and (7.79) then follows taking expectations and using that

$$\mathbb{E}_{|x} \int_0^t \mathrm{A}h(X_s)\, ds = \int_0^t \mathbb{E}_{|x}\, \mathrm{A}h(X_s)\, ds. \qquad \square$$

Before proceeding we need two simple auxiliary results.

Lemma 7.7.2 *Suppose h is bounded and path-continuous. Then for all $x \in G$ the function $t \mapsto \mathrm{P}_t h(x)$ is continuous, and if $h \in \mathcal{D}(\mathrm{A})$ the function is continuously differentiable with*

$$D_t(\mathrm{P}_t h)(x) = \mathrm{P}_t(\mathrm{A}h)(x). \tag{7.80}$$

Proof. Fix t and consider $s \mapsto h(X_s) = h\left(\phi_{s-T_{\langle s\rangle}}\left(Y_{\langle s\rangle}\right)\right)$. Because of the structure assumed for X the distribution of any of the jump times T_n has a density under any $\mathbb{P}_{|x}$ and hence it holds $\mathbb{P}_{|x}$-a.s. that $T_n \neq t$ for all n and therefore also, since h is path-continuous, that $s \mapsto h(X_s)$ is continuous at t $\mathbb{P}_{|x}$-a.s. Now use dominated convergence to see that $s \mapsto \mathrm{P}_s h(x)$ is continuous at t. Applying the first assertion of the lemma to $\mathrm{A}h$ shows that $s \mapsto \mathrm{P}_s(\mathrm{A}h)(x)$ is continuous so (7.80) follows immediately from (7.79). □

Note. Since there are PDMPs such that for some x_0, n_0 and t_0 it holds that $\mathbb{P}_{|x_0}\left(T_{n_0} = t_0\right) > 0$, it is not true in general that $\mathrm{P}_t h(x)$ is continuous in t if h is bounded and path-continuous. But the continuity does hold if for all x and n, T_n has a continuous distribution under $P_{|x}$, i.e., there need not be densities as is otherwise assumed for the PDMPs considered in this section.

The next useful result is easily shown by conditioning on (T_1, Y_1) and using the strong Markov property, Theorem 7.5.1.

Lemma 7.7.3 *For any bounded and measurable $h : G \to \mathbb{R}$ it holds for any $x \in G$, $t \geq 0$ that*

$$\mathrm{P}_t h(x) = h(\phi_t x)\exp\left(-\int_0^t q(\phi_u x)\,du\right)$$
$$+\int_0^t du\, q(\phi_u x)\exp\left(-\int_0^u q(\phi_v x)\,dv\right)\int_G r(\phi_u x, dy)\,\mathrm{P}_{t-u}h(y). \quad (7.81)$$

Note. (7.81) is the *backward integral equation* for the functions $\mathrm{P}_t h$.

We can now prove the main result of this section.

Theorem 7.7.4 *Suppose that $h \in \mathcal{D}(\mathrm{A})$. Then $\mathrm{P}_t h \in \mathcal{D}(\mathrm{A})$ for all $t \geq 0$ and*

$$\mathrm{A}(\mathrm{P}_t h) = \mathrm{P}_t(\mathrm{A}h). \quad (7.82)$$

Proof. Fix $h \in \mathcal{D}(\mathrm{A})$, $t > 0$ and consider $\mathrm{P}_t h$. We must show that $\mathrm{P}_t h$ satisfies (i) and (ii) of Definition 7.7.1 and start by showing that $s \mapsto \mathrm{P}_t h(\phi_s x)$ is continuously differentiable. Using (7.81) with x replaced by $\phi_s(x)$ together with the semigroup property of the ϕ_s gives

$$\begin{aligned}\mathrm{P}_t h(\phi_s x) &= h(\phi_{t+s}x)\exp\left(-\int_0^t q(\phi_{u+s}x)\,du\right)\\ &\quad+\int_0^t du\, q(\phi_{u+s}x)\exp\left(-\int_0^u q(\phi_{v+s}x)\,dv\right)\int_G r(\phi_{u+s}x, dy)\,\mathrm{P}_{t-u}h(y)\\ &= h(\phi_{t+s}x)\exp\left(-\int_s^{t+s} q(\phi_u x)\,du\right)\end{aligned}$$

$$
\begin{aligned}
&+\exp\left(\int_0^s q\left(\phi_v x\right)\, dv\right)\int_s^{t+s} du\, q\left(\phi_u x\right)\exp\left(-\int_0^u q\left(\phi_v x\right)\, dv\right)\\
&\times\int_G r\left(\phi_u x, dy\right)\, \mathrm{P}_{t+s-u} h\left(y\right). \qquad (7.83)
\end{aligned}
$$

From this one verifies first that $\mathrm{P}_t h$ is path-continuous, using in particular that $s \mapsto \mathrm{P}_{t+s-u} h\left(y\right)$ is continuous by Lemma 7.7.2. Consider next the two terms after the last equality in (7.83). The first term and the first factor from the second term are continuously differentiable as functions of s because q is path-continuous by assumption and because $h \in \mathcal{D}\left(\mathrm{A}\right)$. By Lemma 7.7.2 also $s \mapsto \mathrm{P}_{t+s-u} h\left(y\right)$ is continuously differentiable with derivative $\mathrm{P}_{t+s-u}\left(\mathrm{A}h\right)\left(x\right)$. Therefore the integral from s to $t+s$ in the final term of (7.83) is differentiable with respect to s with derivative

$$
\begin{aligned}
&q\left(\phi_{t+s} x\right)\exp\left(-\int_0^{t+s} q\left(\phi_v x\right)\, dv\right)\int_G r\left(\phi_{t+s} x, dy\right)\, h\left(y\right)\\
&-q\left(\phi_s x\right)\exp\left(-\int_0^{s} q\left(\phi_v x\right)\, dv\right)\int_G r\left(\phi_{s} x, dy\right)\, \mathrm{P}_t h\left(y\right)\\
&+\int_s^{t+s} du\, q\left(\phi_u x\right)\exp\left(-\int_0^{u} q\left(\phi_v x\right)\, dv\right)\int_G r\left(\phi_{u} x, dy\right)\, \mathrm{P}_{t+s-u}\left(\mathrm{A}h\right)\left(y\right).
\end{aligned}
$$

That this is continuous in s follows from the path-continuity of $\mathrm{P}_t h$ established above, (7.75) and the simple observation that for all s, t, u and y, $\left|\mathrm{P}_{t+s-u}\left(\mathrm{A}h\right)\left(y\right)\right| \le \sup_{y'}\left|\mathrm{A}h\left(y'\right)\right| < \infty$.

Having shown that $\mathrm{P}_t h$ is path-differentiable, we finally compute $\delta_\phi \mathrm{P}_t h$ directly (the expression for the path-derivative resulting from the calculations above is useless for proving (7.82)): consider

$$
\begin{aligned}
&\frac{1}{s}\left(\mathrm{P}_t h\left(\phi_s x\right) - \mathrm{P}_t h\left(x\right)\right)\\
&= \frac{1}{s}\left(\left(\mathrm{P}_{s+t} h\left(x\right) - \mathrm{P}_t h\left(x\right)\right) + \left(\mathrm{P}_t h\left(\phi_s x\right) - \mathrm{P}_{s+t} h\left(x\right)\right)\right). \qquad (7.84)
\end{aligned}
$$

By Lemma 7.7.2,

$$
\lim_{s\downarrow 0}\frac{1}{s}\left(\mathrm{P}_{s+t} h\left(x\right) - \mathrm{P}_t h\left(x\right)\right) = \mathrm{P}_t\left(\mathrm{A}h\right)\left(x\right). \qquad (7.85)
$$

Next, write

$$
\mathrm{P}_{s+t} h\left(x\right) = \mathbb{E}_{|x}\, h\left(X_{s+t}\right) 1_{\left(T_1 > s\right)} + \mathbb{E}_{|x}\, h\left(X_{s+t}\right) 1_{\left(T_1 \le s\right)}
$$

and use the Markov property to see that $\mathbb{E}_{|x}\left[h\left(X_{s+t}\right)\middle|\mathcal{F}_s^X\right] = \mathrm{P}_t h\left(X_s\right) = \mathrm{P}_t h\left(\phi_s x\right)$ on the set $\left(T_1 > s\right)$. Thus

$$
\begin{aligned}
&\frac{1}{s}\left(\mathrm{P}_t h\left(\phi_s x\right) - \mathrm{P}_{s+t} h\left(x\right)\right)\\
&= \frac{1}{s}\mathrm{P}_t h\left(\phi_s x\right)\left(1-\exp\left(-\int_0^s q\left(\phi_u x\right)\, du\right)\right) - \frac{1}{s}\mathbb{E}_{|x}\, h\left(X_{s+t}\right) 1_{\left(T_1 \le s\right)}.
\end{aligned}
$$

As $s \downarrow 0$, the first term converges to $q(x)\,\mathrm{P}_t h(x)$. Then rewrite the second term as

$$\frac{1}{s}\int_0^s du\, q(\phi_u x)\exp\left(-\int_0^u q(\phi_v x)\,dv\right)\int_G r(\phi_u x, dy)\,\mathrm{P}_{s+t-u}h(y). \tag{7.86}$$

Because by (7.79)

$$\begin{aligned}\left|\mathrm{P}_{s+t-u}h(y)-\mathrm{P}_t h(y)\right| &\leq \int_t^{s+t-u}\left|\mathrm{P}_v(\mathrm{A}h)(y)\right|dv\\ &\leq (s-u)\sup_{y'}\left|\mathrm{A}h\left(y'\right)\right|,\end{aligned}$$

the expression (7.86) has the same limit as $s \downarrow 0$ as

$$\begin{aligned}&\frac{1}{s}\int_0^s du\, q(\phi_u x)\exp\left(-\int_0^u q(\phi_v x)\,dv\right)\int_G r(\phi_u x, dy)\,\mathrm{P}_t h(y)\\ &\to q(x)\int_G r(x, dy)\,\mathrm{P}_t h(y),\end{aligned}$$

the convergence following from (7.75). We now have

$$\lim_{s\downarrow 0}\frac{1}{s}\left(\mathrm{P}_t h(\phi_s x)-\mathrm{P}_{s+t}h(x)\right)=q(x)\int_G r(x,dy)\,\left(\mathrm{P}_t h(x)-\mathrm{P}_t h(y)\right)$$

and combining this with (7.85) and inserting into (7.84) we finally get

$$\delta_\phi(\mathrm{P}_t h)(x)=\mathrm{P}_t(\mathrm{A}h)(x)+q(x)\int_G r(x,dy)\,\left(\mathrm{P}_t h(x)-\mathrm{P}_t h(y)\right),$$

i.e.,

$$\mathrm{A}(\mathrm{P}_t h)(x)=\delta_\phi(\mathrm{P}_t h)(x)+q(x)\int_G r(x,dy)\,\left(\mathrm{P}_t h(y)-\mathrm{P}_t h(x)\right)$$

equals $\mathrm{P}_t(\mathrm{A}h)(x)$ and is bounded. □

Combining Theorem 7.7.4 with Lemma 7.7.2 we obtain

Corollary 7.7.5 *For $h \in \mathcal{D}(\mathrm{A})$ and any $x \in G$, the function $t \mapsto \mathrm{P}_t h(x)$ is continuously differentiable and satisfies the* backward differential equation

$$D_t\mathrm{P}_t h(x)=\mathrm{A}(\mathrm{P}_t h)(x) \tag{7.87}$$

and the forward differential equation

$$D_t\mathrm{P}_t h(x)=\mathrm{P}_t(\mathrm{A}h)(x). \tag{7.88}$$

Example 7.7.2 If X is a Markov chain we saw in Example 7.7.1 that any bounded and measurable h belongs to $\mathcal{D}(\mathrm{A})$. By (7.77), (7.87) takes the form

$$D_t \mathrm{P}_t h(x) = q(x) \int_G r(x, dy)\ (\mathrm{P}_t h(y) - \mathrm{P}_t h(x))$$

which in fact also follows directly from (7.81) when $\phi_u(x) = x$.

If G is at most countable, X is a standard Markov chain on a finite or countably infinite state space. Labelling the states by i, j rather than x, y and writing $p_{ij}(t) = \mathbb{P}_{|i}(X_t = j)$ for the transition probabilities and $q_i = q(i)$ and $q_i r(i, \{j\}) = q_{ij}$ for $i \neq j$ for the transition intensities, taking $h(i) = 1_{\{j\}}(i) = \delta_{ij}$ for some given j we see that (7.87) becomes

$$D_t p_{ij}(t) = -q_i p_{ij}(t) + \sum_{k:k\neq i} q_{ik} p_{kj}(t), \tag{7.89}$$

while because for any $h \in \mathcal{D}(\mathrm{A})$,

$$\begin{aligned} \mathrm{P}_t(\mathrm{A}h)(i) &= \mathbb{E}_{|i}\, \mathrm{A}h(X_t) \\ &= \mathbb{E}_{|i} \left(\sum_{\ell:\ell\neq X_t} q_{X_t \ell} h(\ell) - q_{X_t} h(X_t) \right), \end{aligned}$$

(7.88) reduces to

$$D_t p_{ij}(t) = -q_j p_{ij}(t) + \sum_{k:k\neq j} q_{kj} p_{ik}(t). \tag{7.90}$$

Here (7.89) and (7.90) are respectively the classical backward and forward Feller–Kolmogorov equations for the transition probabilities (cf. (7.15) and (7.18)). When solving them one uses the obvious boundary conditions $p_{ij}(0) = \delta_{ij}$.

7.8 Stationarity

Suppose X is a time-homogeneous Markov process with state space $G \subset \mathbb{R}^d$ (or G e.g., a subset of a Polish space) such that X is right-continuous. As usual p_t denotes the transition probabilities and P_t the transition operators for X; cf. Section 7.1.

Definition 7.8.1 A probability measure ρ on $(G, \mathcal{G})$ is *invariant* for X if for all $t \geq 0$ and all $A \in \mathcal{G}$,

$$\int_G \rho(dx)\ p_t(x, A) = \rho(A). \tag{7.91}$$

An invariant probability is also called a *stationary distribution* for X.

If $h : G \to \mathbb{R}$ is bounded and measurable and ρ is a probability on $(G, \mathcal{G})$, write $\rho(h) = \int_G \rho(dx)\ h(x)$. The condition (7.91) for ρ to be invariant is then equivalent to

$$\rho(\mathrm{P}_t h) = \rho(h) \quad (t \geq 0) \tag{7.92}$$

for all bounded or measurable h, or just for all h in a *determining class,* i.e., a class $\mathcal{C}$ of bounded and measurable functions h such that the condition $\rho_1(h) = \rho_2(h)$ for all $h \in \mathcal{C}$, where ρ_1 and ρ_2 are probabilities on G, implies that $\rho_1 = \rho_2$.

Apart from the transition probabilities p_t we shall below also use the Markov kernels ψ_t on G defined by

$$\psi_t(x, A) = \frac{1}{t}\int_0^t p_s(x, A)\, ds \tag{7.93}$$

for $t > 0,\ A \in \mathcal{G}$. Clearly, if ρ is invariant also

$$\int_G \rho(dx)\ \psi_t(x, A) = \rho(A) \quad (t > 0, A \in \mathcal{G})\,. \tag{7.94}$$

The converse follows because X is right-continuous:

Lemma 7.8.1 *If ρ is a probability on $(G, \mathcal{G})$ such that (7.94) holds, then ρ is invariant.*

Proof. (7.94) implies that, writing $\psi_t(x, h) = \int_G \psi_t(x, dy)\ h(y)$,

$$\int_G \rho(dx)\ \psi_t(x, h) = \rho(h)$$

for, in particular, all bounded and continuous $h : G \to \mathbb{R}$. Equivalently

$$\int_0^t ds \int_G \rho(dx)\ \mathrm{P}_s h(x) = t\rho(h)\,.$$

But $t \mapsto \mathrm{P}_t h(X_t) = \mathbb{E}_{|x}\, h(X_t)$ is right-continuous since h is continuous and X is right-continuous, so differentiating from the right with respect to t we obtain $\rho(\mathrm{P}_t h) = \rho(h)$ for all $t \geq 0$. Since the bounded and continuous h form a determining class it follows that ρ is invariant. □

That ρ is invariant for X is equivalent to saying that if X has initial distribution ρ, then X_t for all t also has distribution ρ. But because X is Markov it then also follows that the joint distribution of $\left(X_{t_1+t}, \ldots, X_{t_n+t}\right)$ for any n and any $0 \leq t_1 < \cdots < t_n$ is the same for all t and, a fortiori since X is right-continuous, that the $\mathbb{P}_\rho$-distribution of the process $(X_{s+t})_{s\geq 0}$ does not depend on t: X with initial distribution ρ is a *strictly stationary process.* (To verify that the distribution of the vector $\left(X_{t_1+t}, \ldots, X_{t_n+t}\right)$ does not depend on t, proceed by induction on n and use the Markov property, e.g., for h_k bounded and measurable one writes

$$\mathbb{E}_\rho \prod_{k=1}^n h_k\left(X_{t_k+t}\right) = \mathbb{E}_\rho \left(\prod_{k=1}^{n-1} h_k\left(X_{t_k+t}\right)\right) \mathrm{P}_{t_n - t_{n-1}} h_n\left(X_{t_{n-1}+t}\right)).$$

In general the problems concerning the existence and uniqueness of invariant probabilities are very difficult. And even should a unique invariant probability exist, it may be quite impossible to find it! We shall here consider the class of homogeneous PDMPs also treated in Section 7.7 and for these processes establish what is the standard equation for finding the invariant probabilities. Then we continue with a discussion of some of the recurrence properties shared by the PDMPs that do have an invariant probability. At the end of the section we give some examples related to renewal processes where there is a unique invariant probability, and where it can be determined explicitly.

So suppose now that $t \mapsto \phi_t(x)$ is continuous for all x and that X is otherwise determined by the path-continuous jump intensity $x \mapsto q(x)$ and the jump probabilities $r(x, \cdot)$ satisfying (7.75). Consider the full infinitesimal generator A for X, defined on the domain $\mathcal{D}(\mathrm{A})$ (Definition 7.7.1).

Theorem 7.8.2 *Suppose that $\mathcal{D}(\mathrm{A})$ is a determining class. Then a probability ρ on $(G, \mathcal{G})$ is invariant if and only if*

$$\rho(\mathrm{A}h) = 0 \quad (h \in \mathcal{D}(\mathrm{A}))\,. \tag{7.95}$$

In particular, if there is only one probability ρ that satisfies (7.95), then the invariant probability ρ is uniquely determined.

Proof. Suppose ρ is invariant. Then for $h \in \mathcal{D}(\mathrm{A})$, since $\rho(\mathrm{P}_t h) = \rho(h)$ and $\rho(\mathrm{P}_s(\mathrm{A}h)) = \rho(\mathrm{A}h)$ it follows directly from (7.79) that $\rho(\mathrm{A}h) = 0$.

If conversely $\rho(\mathrm{A}h) = 0$ for all $h \in \mathcal{D}(\mathrm{A})$, by Theorem 7.7.4 also $\mathrm{P}_s h \in \mathcal{D}(\mathrm{A})$ for every s and $\rho(\mathrm{P}_s(\mathrm{A}h)) = \rho(\mathrm{A}(\mathrm{P}_s h)) = 0$. Inserting this in (7.79) gives $\rho(\mathrm{P}_t h) = \rho(h)$ for all t; since $\mathcal{D}(\mathrm{A})$ is a determining class this implies that ρ is invariant. □

By repeating the second part of the proof one immediately obtains

Corollary 7.8.3 *Suppose $\mathcal{D}_0 \subset \mathcal{D}(\mathrm{A})$ is a determining class such that $\mathrm{P}_t : \mathcal{D}_0 \to \mathcal{D}_0$ for all t. If ρ is a probability on $(G, \mathcal{G})$ such that*

$$\rho(\mathrm{A}h) = 0 \quad (h \in \mathcal{D}_0)\,,$$

then ρ is invariant.

Example 7.8.1 Suppose X is a Markov chain. From Example 7.7.1 we know that $\mathcal{D}(\mathrm{A})$ consists of all bounded and measurable h so $\mathcal{D}(\mathrm{A})$ is trivially a determining class. Retaining the notation from Example 7.7.1 it follows that a probability ρ is invariant iff

$$\int_G \rho(dx)\; q(x) \int_G r(x, dy)\; (h(y) - h(x)) = 0$$

for all bounded and measurable h. By Corollary 7.8.3 it e.g., suffices that this identity holds for all bounded and continuous h, provided for any t, $\mathrm{P}_t h$ is continuous whenever h is bounded and continuous.

It is well known that the existence of an invariant probability implies some sort of positive recurrence property on at least part of the state space for the process. Also, uniqueness of the invariant probability forces a certain structure on the communication between different parts of the state space. We shall comment briefly on this and shall to begin with consider an arbitrary right-continuous homogeneous Markov process.

With ψ_t defined as in (7.93), for $x \in G$, $A \in \mathcal{G}$ write $x \rightsquigarrow A$ if $\psi_t(x, A) > 0$ for some $t > 0$ (equivalently, for all sufficiently large $t > 0$). Introducing the *occupation measures*

$$I_t(A) = \int_0^t 1_{(X_s \in A)}\, ds, \quad I(A) = \int_0^\infty 1_{(X_s \in A)}\, ds \tag{7.96}$$

(so $t \mapsto I_t(A)$ is increasing with limit $I(A)$ as $t \uparrow \infty$, and $I_t(A) \le t$ while $I(A) = \infty$ is possible and relevant), since by Fubini's theorem

$$\mathbb{E}_{|x} I_t(A) = \int_0^t p_s(x, A)\, ds = t\psi_t(x, A),$$
$$\mathbb{E}_{|x} I(A) = \int_0^\infty p_s(x, A)\, ds,$$

it is clear that

$$x \rightsquigarrow A \quad \text{iff} \quad \mathbb{E}_{|x} I(A) > 0. \tag{7.97}$$

Call a set $A \in \mathcal{G}$ *transition-closed* if $\psi_t(x, A) = 1$ for all $x \in A$, $t > 0$. In particular A is transition-closed if $p_t(x, A) = 1$ for all $x \in A$, $t \ge 0$.

Proposition 7.8.4 *Suppose that $A_0 \in \mathcal{G}$ is transition-closed and that ρ is an invariant probability. Then*

(i) $\rho(A') = 0$ *where* $A' = \{x \in G \backslash A_0 : x \rightsquigarrow A_0\}$*;*
(ii) *if $\rho(A_0) > 0$ the conditional probability $\widetilde{\rho} := \rho(\cdot\, | A_0)$ is invariant. In particular, if $\rho(A_0) > 0$ and $\rho(G \backslash A_0) > 0$ the invariant probability is not unique;*
(iii) *if the set $G \backslash (A_0 \cup A')$ is non-empty it is transition-closed.*

Proof. (i). By invariance and because A_0 is transition-closed,

$$\rho(A_0) = \int_G \rho(dx)\, \psi_t(x, A_0) = \rho(A_0) + \int_{G \backslash A_0} \rho(dx)\, \psi_t(x, A_0),$$

so the last term vanishes and in particular

$$\rho\{x \in G \backslash A_0 : \psi_t(x, A_0) > 0\} = 0,$$

which implies for $t \uparrow \infty$ that

$$\rho(A') = \rho\left\{x \in G \backslash A_0 : \int_0^\infty p_s(x, A_0)\, ds > 0\right\} = 0;$$

cf. (7.97).

(ii). We have for $A \in \mathcal{G}$,

$$\begin{aligned}\widetilde{\rho}(A) &= \frac{1}{\rho\,(A_0)} \int_G \rho\,(dx)\ \psi_t\,(x, A \cap A_0) \\ &= \frac{1}{\rho\,(A_0)} \int_{G \backslash A'} \rho\,(dx)\ \psi_t\,(x, A \cap A_0)\end{aligned}$$

by (i). But if $x \in G \backslash \left(A_0 \cup A'\right)$ we have by the definition of A' that $\psi_t\,(x, A \cap A_0) \leq \psi_t\,(x, A_0) = 0$ for all $t > 0$ and therefore

$$\begin{aligned}\widetilde{\rho}(A) &= \frac{1}{\rho\,(A_0)} \int_{A_0} \rho\,(dx)\ \psi_t\,(x, A \cap A_0) \\ &= \int_G \widetilde{\rho}\,(dx)\ \psi_t\,(x, A)\end{aligned}$$

for all $t > 0$ since $\widetilde{\rho}$ is concentrated on A_0 and $\psi_t\,(x, G \backslash A_0) = 0$ for $x \in A_0$.

Thus $\widetilde{\rho}$ is invariant and clearly, if $\rho\,(G \backslash A_0) > 0$, $\widetilde{\rho} \neq \rho$ and the invariant probability is not unique.

Finally, suppose that $x \in G \backslash \left(A_0 \cup A'\right)$. By the definition of A', $\psi_t\,(x, A_0) = 0$ for all $t > 0$ so in order to show that $G \backslash \left(A_0 \cup A'\right)$ is transition-closed, it suffices to show that $p_t\left(x, A'\right) = 0$ for all $t > 0$.

But for any $t > 0$,

$$\begin{aligned}0 &= \int_0^\infty p_s\,(x, A_0)\ ds \\ &\geq \int_t^\infty p_s\,(x, A_0)\ ds \\ &\geq \int_t^\infty ds \int_{A'} p_t\,(x, dy)\ p_{s-t}\,(y, A_0) \\ &= \int_{A'} p_t\,(x, dy) \int_0^\infty ds\ p_s\,(y, A_0)\,,\end{aligned}$$

and since for all $y \in A'$, $\int_0^\infty ds\ p_s\,(y, A_0) > 0$ by definition, $p_t\left(x, A'\right) = 0$ follows. □

Proposition 7.8.5 *Suppose that ρ is invariant. If $A \in \mathcal{G}$ is such that $\rho\,(A) = 0$, then also $\rho\left(A'\right) = 0$ where $A' = \{x \in G : x \rightsquigarrow A\}$.*

If $A \in \mathcal{G}$ is such that $x \rightsquigarrow A$ for all $x \in G$, then $\rho\,(A) > 0$.

Proof. Since necessarily $\rho\,(G) = 1 > 0$ the second assertion follows from the first. To prove the first, note that by invariance,

$$\rho\,(A) \geq \int_{A'_t} \rho\,(dx)\ \psi_t\,(x, A)$$

where $A'_t = A' \cap \{x : \psi_t\,(x, A) > 0\}$. So if $\rho\,(A) = 0$ also $\rho\left(A'_t\right) = 0$, and since $A'_t \uparrow A'$ as $t \uparrow \infty$, $\rho\left(A'\right) = 0$ follows. □

Now let X be a homogeneous PDMP such that all $t \mapsto \phi_t(x)$ are continuous so that in particular X is right-continuous. For certain sets A there is then a useful alternative description of the relationship $x \rightsquigarrow A$. Call $A \in \mathcal{G}$ *path-open* if for all $x \in A$ it holds that $\phi_t(x) \in A$ for all sufficiently small $t > 0$.

Example 7.8.2 If X is a Markov chain, $\phi_t(x) = x$, then all $A \in \mathcal{G}$ are path-open. If G is an interval $\subset \mathbb{R}$ and $\phi_t(x) = x + t$, then all half-open intevals $[a, b[\subset G$ are path-open. If $G = \mathbb{R}^2$ (or G is a two-dimensional interval) and $\phi_t(x_1, x_2) = (x_1 + t, x_2 + t)$, then any set of the form $\{(x_1, x_1 + c) : a \le x_1 < b\}$ for some c and $a < b$ is path-open, illustrating that path-open sets in higher dimensions can be one-dimensional and need therefore not be (ordinarily) open.

In general any open set is path-open provided all $\phi_t(x)$ are continuous in t.

Let A be a path-open set and define the hitting time

$$\mathrm{T}_A := \inf\{t \ge 0 : X_t \in A\}$$

with $\inf \emptyset = \infty$. Then for all $t > 0$,

$$(\mathrm{T}_A < t) = \bigcup_{q \in \mathbb{Q}_0 : q < t} \left(X_q \in A\right) \tag{7.98}$$

since the inclusion $\supset$ is trivial, while if $\mathrm{T}_A(\omega) = t_0 < t$ we have $X_{t_1}(\omega) \in A$ for some $t_1 \in [t_0, t[$ and therefore also that $X_s(\omega) = \phi_{s-t_1}\left(X_{t_1}(\omega)\right) \in A$ if only $s > t_1$ is sufficiently close to t_1, in particular $X_q(\omega) \in A$ for $q \in]t_1, t[$ rational and sufficiently close to t_1.

Note that because A is path-open,

$$\mathrm{T}_A := \inf\{t > 0 : X_t \in A\}.$$

From (7.98) it follows in particular that T_A is an $\mathcal{F}_t^X$-stopping time. We have also seen that if $\mathrm{T}_A(\omega) < \infty$, then the set $\{t : X_t(\omega) \in A\}$ contains a non-degenerate interval, in particular $I(A, \omega) > 0$, cf. (7.96). With this in mind it is easy to show

Proposition 7.8.6 *For A path-open, $x \rightsquigarrow A$ if and only if* $\mathbb{P}_{|x}(\mathrm{T}_A < \infty) > 0$.

Proof. We have

$$(\mathrm{T}_A < \infty) = (I(A) > 0),$$

the inclusion $\subset$ having been argued above and the opposite inclusion being trivial from the definition of $I(A)$ and T_A. Hence $\mathbb{P}_{|x}(\mathrm{T}_A < \infty) > 0$ iff

$$0 < \mathbb{E}_{|x}\, I(A) 1_{(I(A) > 0)} = \mathbb{E}_{|x}\, I(A)$$

which by (7.97) is precisely to say that $x \rightsquigarrow A$. $\square$

We shall now in some detail investigate the behaviour of an invariant probability on the path-open sets. For A path-open, define

$$\mathrm{T}_{t,A} = \inf\{s \geq t : X_s \in A\},$$

the first time X enters A on the time interval $[t, \infty[$, and note that

$$\mathrm{T}_{t,A} = t + \widetilde{\theta}_t\,(\mathrm{T}_A)$$

where

$$\widetilde{\theta}_t\,(\mathrm{T}_A) := \inf\{s \geq 0 : X_{t+s} \in A\}$$

is the time it takes the post-t process $(X_{t+s})_{s\geq 0}$ to enter A. Clearly $\mathrm{T}_{t,A} \leq \mathrm{T}_{t',A}$ if $t \leq t'$ and

$$\mathrm{T}_{t,A} = \mathrm{T}_A \quad \text{on} \quad (\mathrm{T}_A \geq t)\,. \tag{7.99}$$

Also, by the Markov property,

$$\mathbb{P}\left(\widetilde{\theta}_t\,(\mathrm{T}_A) > s \,\middle|\, \mathcal{F}_t^X\right) = \mathbb{P}_{|X_t}\,(\mathrm{T}_A > s)\,. \tag{7.100}$$

All of these facts will be used repeatedly below without any further comment.

Theorem 7.8.7 *Suppose that there exists an invariant probability ρ for X. Then for any $A \in \mathcal{G}$ path-open, and any $t \geq 0$,*

$$\mathbb{E}_\rho\left[\mathrm{T}_{t,A}; \mathrm{T}_A \leq t\right] < \infty. \tag{7.101}$$

In particular, if $\rho\,(A) > 0$ it holds for ρ-almost all $x \in A$ simultaneously for all $t \geq 0$ that

$$\mathbb{E}_{|x}\,\mathrm{T}_{t,A} < \infty. \tag{7.102}$$

Proof. We show (7.101) by showing that

$$\mathbb{E}_\rho\left[\widetilde{\theta}_t\,(\mathrm{T}_A)\,; \mathrm{T}_A \leq t\right] < \infty$$

for all t. Define

$$g(t) = \mathbb{P}_\rho\,(\mathrm{T}_A > t) \quad (t \geq 0).$$

Because ρ is invariant, for any $s > 0$, $t \geq 0$,

$$\begin{aligned} g(t) &= \mathbb{P}_\rho\left(\widetilde{\theta}_s\,(\mathrm{T}_A) > t\right) \\ &= \mathbb{P}_\rho\left(\widetilde{\theta}_s\,(\mathrm{T}_A) > t, \mathrm{T}_A > s\right) + \mathbb{P}_\rho\left(\widetilde{\theta}_s\,(\mathrm{T}_A) > t, \mathrm{T}_A \leq s\right). \end{aligned}$$

Here the first term equals $P_\rho\,(\mathrm{T}_A > t + s)$ by (7.99) and so

$$g(t) = g(t+s) + \mathbb{P}_\rho\left(\widetilde{\theta}_s(\mathrm{T}_A) > t, \mathrm{T}_A \le s\right).$$

Taking $t = ks$ and summing on k from 0 to $n-1$ gives

$$g(0) - g(ns) = \sum_{k=0}^{n-1} \mathbb{P}_\rho\left(\widetilde{\theta}_s(\mathrm{T}_A) > ks, \mathrm{T}_A \le s\right).$$

It follows that

$$\sum_{k=0}^{\infty} \mathbb{P}_\rho\left(\widetilde{\theta}_s(\mathrm{T}_A) > ks, \mathrm{T}_A \le s\right) \le g(0) = \mathbb{P}_\rho(\mathrm{T}_A > 0) \tag{7.103}$$

(with equality if $\mathbb{P}_\rho(\mathrm{T}_A = \infty) = 0$ since then $g(ns) \to 0$ as $n \to \infty$). In particular the infinite series converges and since the kth term is

$$\ge \mathbb{P}_\rho\left(\widetilde{\theta}_s(\mathrm{T}_A) = \infty, \mathrm{T}_A \le s\right)$$

we deduce that

$$\mathbb{P}_\rho\left(\widetilde{\theta}_s(\mathrm{T}_A) = \infty, \mathrm{T}_A \le s\right) = 0.$$

Thus the infinite series becomes

$$\begin{aligned}
&\sum_{k=0}^{\infty} \mathbb{P}_\rho\left(ks < \widetilde{\theta}_s(\mathrm{T}_A) < \infty, \mathrm{T}_A \le s\right) \\
&= \sum_{k=0}^{\infty}\sum_{n=k}^{\infty} \mathbb{P}_\rho\left(ns < \widetilde{\theta}_s(\mathrm{T}_A) \le (n+1)s, \mathrm{T}_A \le s\right) \\
&= \sum_{n=0}^{\infty} (n+1)\,\mathbb{P}_\rho\left(ns < \widetilde{\theta}_s(\mathrm{T}_A) \le (n+1)s, \mathrm{T}_A \le s\right) \\
&\ge \frac{1}{s}\sum_{n=0}^{\infty} \mathbb{E}_\rho\left[\widetilde{\theta}_s(\mathrm{T}_A); ns < \widetilde{\theta}_s(\mathrm{T}_A) \le (n+1)s, \mathrm{T}_A \le s\right] \\
&= \frac{1}{s}\mathbb{E}_\rho\left[\widetilde{\theta}_s(\mathrm{T}_A); \mathrm{T}_A \le s\right];
\end{aligned}$$

using (7.103) therefore

$$\mathbb{E}_\rho\left[\widetilde{\theta}_s(\mathrm{T}_A); \mathrm{T}_A \le s\right] \le s\mathbb{P}_\rho(\mathrm{T}_A > 0)$$

and (7.101) follows.

To prove the second assertion of the theorem, note that by the definition of T_A, $\mathrm{T}_A = 0$ on $(X_0 \in A)$ and consequently, for $t \ge 0$

$$\int_A \rho(dx)\ \mathbb{E}_{|x}\left[\mathrm{T}_{t,A}; \mathrm{T}_A \le t\right] = \int_A \rho(dx)\ \mathbb{E}_{|x}\,\mathrm{T}_{t,A}$$

with the integral finite since it is $\le \mathbb{E}_\rho\left[\mathrm{T}_{t,A}; \mathrm{T}_A \le t\right]$. So for any given t, $\mathbb{E}_{|x}\,\mathrm{T}_{t,A} < \infty$ for ρ-a.a. $x \in A$. But since $t \mapsto \mathbb{E}_{|x}\,\mathrm{T}_{t,A}$ is increasing, this implies that for ρ-a.a. $x \in A$, $\mathbb{E}_{|x}\,\mathrm{T}_{t,A} < \infty$ simultaneously for all t — a statement that although true for all path-open A is of course of interest only if $\rho(A) > 0$. □

That (7.102) holds expresses in particular that if the process starts from a ρ-typical $x \in A$, even after arbitraily large t, the expected extra time beyond t required to reenter A is finite: in this sense the set A is *positive recurrent.*

Suppose now that $A_0 \in \mathcal{G}$ is a given path-open set. For any $x \in G$, $\mathbb{P}_{|x}\left(\mathrm{T}_{A_0} = 0\right) = \mathbb{P}_{|x}\left(\mathrm{T}_{A_0} = 0, T_1 > 0\right)$ and the only possible values for this probability are 0 and 1: the probability $= 1$ iff either (i) $x \in A_0$ or (ii) $x \in G \backslash A_0$ is such that $\phi_{t_k}(x) \in A_0$ for some sequence (t_k) of time points $t_k > 0$ with $t_k \downarrow 0$ as $k \uparrow \infty$. Define

$$\overline{A}_0 = \left\{x \in G : \mathbb{P}_{|x}\left(\mathrm{T}_{A_0} = 0\right) = 1\right\} \tag{7.104}$$

and

$$B_0 = \left\{x \in G \backslash \overline{A}_0 : x \rightsquigarrow A_0\right\}.$$

In particular, $\mathbb{P}_{|x}\left(\mathrm{T}_{A_0} > 0\right) = 1$ for $x \in B_0$.

Theorem 7.8.8 *Suppose that ρ is invariant for X.*

(a) *The set B_0 is path-open.*
(b) *For ρ-almost all $x \in B_0$ it holds that*

$$\mathbb{P}_{|x}\left(\mathrm{T}_{A_0} < \infty\right) = 1.$$

(c) *The set $G \backslash \left(\overline{A}_0 \cup B_0\right)$ is transition-closed and if $\rho\left(\overline{A}_0 \cup B_0\right) > 0$, the set $\overline{A}_0 \cup B_0$ is weakly transition-closed in the sense that for ρ-a.a. $x \in \overline{A}_0 \cup B_0$, $\psi_t\left(x, \overline{A}_0 \cup B_0\right) = 1$ simultaneously for all $t > 0$. In particular, if $\rho\left(\overline{A}_0 \cup B_0\right) > 0$ and $\rho\left(G \backslash \left(\overline{A}_0 \cup B_0\right)\right) > 0$ the invariant probability ρ is not unique.*

Proof. (a) Let $x \in B_0$. Since $\left(t < \mathrm{T}_{A_0} < \infty, T_1 > t\right) \downarrow \left(0 < \mathrm{T}_{A_0} < \infty\right)$ as $t \downarrow 0$ and since $\mathbb{P}_{|x}\left(\mathrm{T}_{A_0} > 0\right) = 1$ it follows that

$$\mathbb{P}_{|x}\left(t < \mathrm{T}_{A_0} < \infty, T_1 > t\right) > 0$$

for $t > 0$ sufficiently small. But by the Markov property and (7.100) this probability equals

$$\begin{aligned} &\mathbb{E}_{|x}\left[\mathbb{P}_{|X_t}\left(0 < \mathrm{T}_{A_0} < \infty\right); \mathrm{T}_{A_0} > t, T_1 > t\right] \\ &= \mathbb{P}_{|\phi_t x}\left(0 < \mathrm{T}_{A_0} < \infty\right)\mathbb{P}_{|x}\left(\mathrm{T}_{A_0} > t, T_1 > t\right) \end{aligned} \tag{7.105}$$

and thus $\mathbb{P}_{|\phi_t x}\left(0 < \mathrm{T}_{A_0} < \infty\right) > 0$ for $t > 0$ small, i.e., $\mathbb{P}_{|\phi_t x}\left(\mathrm{T}_{A_0} > 0\right) = 1$ and $\mathbb{P}_{|\phi_t x}\left(\mathrm{T}_{A_0} < \infty\right) > 0$ which is precisely to say that $\phi_t(x) \in B_0$ for $t > 0$ small.

(b) By Proposition 7.8.6, $B_0 = \left\{x \in G \backslash \overline{A}_0 : \mathbb{P}_{|x}\left(\mathrm{T}_{A_0} < \infty\right) > 0\right\}$. Defining

$$B_n = \left\{x \in B_0 : \mathbb{P}_{|x}\left(\mathrm{T}_{A_0} < \infty\right) > \tfrac{1}{n}\right\}$$

for $n \in \mathbb{N}$, clearly $B_n \uparrow B_0$ as $n \uparrow \infty$.

We first show that each B_n is path-open, which because $B_0 = \bigcup B_n$ implies that B_0 is path-open. If $x \in B_n$, arguing as in the proof of (a) above it is clear that

$$\mathbb{P}_{|x}\left(t < \mathrm{T}_{A_0} < \infty, T_1 > t\right) > \tfrac{1}{n}$$

for $t > 0$ small enough. But then (7.105) directly implies that

$$\mathbb{P}_{|\phi_t x}\left(0 < \mathrm{T}_{A_0} < \infty\right) > \frac{1}{n}$$

for $t > 0$ small, i.e., $\phi_t(x) \in B_n$ for $t > 0$ small.

We now complete the proof of (b) by verifying that for each n we have that $\mathbb{P}_{|x}\left(\mathrm{T}_{A_0} < \infty\right) = 1$ for ρ-a.a. $x \in B_n$ — since $B_n \uparrow B_0$ this will suffice.

Since B_n is path-open, by Theorem 7.8.7 for ρ-a.a. $x \in B_n$,

$$\mathbb{P}_{|x}\left(\mathrm{T}_{t,B_n} < \infty\right) = \mathbb{P}_{|x}\left(\widetilde{\theta}_t\left(\mathrm{T}_{B_n}\right) < \infty\right) = 1$$

simultaneously for all t. For such x therefore, defining $g(y) = \mathbb{P}_{|y}\left(\mathrm{T}_{A_0} = \infty\right)$ for any $y \in G$, we find that for every t,

$$\begin{aligned}
g(x) &= \mathbb{P}_{|x}\left(\widetilde{\theta}_t\left(\mathrm{T}_{A_0}\right) = \infty, \mathrm{T}_{A_0} > t\right) \\
&= \mathbb{P}_{|x}\left(\widetilde{\theta}_t\left(\mathrm{T}_{B_n}\right) < \widetilde{\theta}_t\left(\mathrm{T}_{A_0}\right), \widetilde{\mathrm{T}}_{n,A_0} = \infty, \mathrm{T}_{A_0} > t\right) \\
&\leq \mathbb{P}_{|x}\left(\widetilde{\theta}_t\left(\mathrm{T}_{B_n}\right) < \infty, \widetilde{\mathrm{T}}_{n,A_0} = \infty, \mathrm{T}_{A_0} > t\right)
\end{aligned}$$

where

$$\widetilde{\mathrm{T}}_{n,A_0} = \inf\left\{s \geq 0 : X_{\mathrm{T}_{t,B_n}+s} \in A_0\right\}$$

is the extra time beyond T_{t,B_n} required for X to reenter A_0. Thus by the strong Markov property,

$$\begin{aligned}
g(x) &\leq \mathbb{E}_{|x}\left[g\left(X_{\mathrm{T}_{t,B_n}}\right); \widetilde{\theta}_t\left(\mathrm{T}_{B_n}\right) < \infty, \mathrm{T}_{A_0} > t\right] \\
&\leq \mathbb{E}_{|x}\left[g\left(X_{\mathrm{T}_{t,B_n}}\right); \mathrm{T}_{A_0} > t\right].
\end{aligned} \tag{7.106}$$

We now claim that

$$g\left(X_{\mathrm{T}_{t,B_n}}\right) \leq 1 - \tfrac{1}{n} \quad \mathbb{P}_{|x}\text{-a.s.} \tag{7.107}$$

This is obvious if $X_{\mathrm{T}_{t,B_n}} \in B_n$ or $\in \overline{A}_0$. And if $y = X_{\mathrm{T}_{t,B_n}} \notin B_n \cup \overline{A}_0$, by the definition of T_{t,B_n}, $\phi_{s_k}(y) = X_{s_k} \in B_n$ for some sequence (s_k) with $s_k > 0$, $s_k \downarrow 0$ as $k \uparrow \infty$. (The sequence (s_k) is random but completely determined by $y = X_{\mathrm{T}_{t,B_n}}$). Thus

$$\begin{aligned}
g(y) &= \mathbb{P}_{|y}\left(\widetilde{\theta}_{s_k}\left(\mathrm{T}_{A_0}\right) = \infty, \mathrm{T}_{A_0} > s_k\right) \\
&\leq \mathbb{P}_{|y}\left(\widetilde{\theta}_{s_k}\left(\mathrm{T}_{A_0}\right) = \infty, \mathrm{T}_{A_0} > s_k, T_1 > s_k\right) + \mathbb{P}_{|y}\left(T_1 \leq s_k\right).
\end{aligned}$$

Here the second term $\downarrow 0$ as $k \uparrow \infty$ while the first equals

$$g\left(\phi_{s_k} y\right) \mathbb{P}_{|y}\left(\mathrm{T}_{A_0} > s_k, T_1 > s_k\right) \leq 1 - \tfrac{1}{n}$$

since $\phi_{s_k}(y) \in B_n$. Because $y \notin \overline{A}_0$, $\mathbb{P}_{|y}\left(\mathrm{T}_{A_0} > 0\right) = 1$ so

$$\lim_{k\to\infty} \mathbb{P}_{|y}\left(\mathrm{T}_{A_0} > s_k, T_1 > s_k\right) = 1$$

and the proof of (7.107) is complete. (7.106) then yields

$$g(x) \leq \left(1 - \tfrac{1}{n}\right) \mathbb{P}_{|x}\left(\mathrm{T}_{A_0} > t\right)$$

which for $t \uparrow \infty$ forces $g(x) \leq \left(1 - \frac{1}{n}\right) g(x)$ so $g(x) = 0$, as desired.

(c) We show that $G \backslash \left(\overline{A}_0 \cup B_0\right)$ is transition-closed by showing that

$$p_t\left(x, \overline{A}_0 \cup B_0\right) = 0$$

for all $t > 0$, $x \in G \backslash \left(\overline{A}_0 \cup B_0\right)$. But if $t > 0$, $x \in G$ (cf. the last part of the proof of Proposition 7.8.4), by the definition of B_0, for any $t > 0$

$$\begin{aligned} 0 &= \int_0^\infty ds\; p_s\left(x, A_0\right) \\ &\geq \int_t^\infty ds\; p_s\left(x, A_0\right) \\ &= \int_t^\infty ds \int_G p_t\left(x, dy\right)\; p_{s-t}\left(y, A_0\right) \\ &\geq \int_{\overline{A}_0 \cup B_0} p_t\left(x, dy\right) \int_0^\infty ds\; p_s\left(y, A_0\right). \end{aligned}$$

But the inner integral is > 0 for $y \in B_0$ by the definition of B_0 and > 0 for $y \in \overline{A}_0$ because A_0 is path-open. Thus

$$p_t\left(x, \overline{A}_0 \cup B_0\right) = 0 \quad \left(x \in G \backslash \left(\overline{A}_0 \cup B_0\right)\right), \tag{7.108}$$

as desired.

Next, by stationarity for any $t > 0$,

$$\begin{aligned} \rho\left(\overline{A}_0 \cup B_0\right) &= \mathbb{P}_\rho\left(X_t \in \overline{A}_0 \cup B_0\right) \\ &= \int_G \rho\left(dx\right)\; p_t\left(x, \overline{A}_0 \cup B_0\right) \\ &= \int_{\overline{A}_0 \cup B_0} \rho\left(dx\right)\; p_t\left(x, \overline{A}_0 \cup B_0\right) \end{aligned}$$

using (7.108) for the last equality. But then also

$$\rho\left(\overline{A}_0 \cup B_0\right) = \int_{\overline{A}_0 \cup B_0} \rho\left(dx\right)\; \psi_t\left(x, \overline{A}_0 \cup B_0\right)$$

so for any given $t > 0$, $\psi_t\left(x, \overline{A}_0 \cup B_0\right) = 1$ for ρ-a.a. $x \in \overline{A}_0 \cup B_0$. Since $\psi_t\left(x, \overline{A}_0 \cup B_0\right)$ is continuous in t, we conclude that for for ρ-a.a. $x \in \overline{A}_0 \cup B_0$, $\psi_t\left(x, \overline{A}_0 \cup B_0\right) = 1$ simultaneously for all t.

The last assertion in (c) is immediate from Proposition 7.8.4 (ii). □

Remark 7.8.1 The theorem is mainly of interest when $\rho(B_0) > 0$, in which case automatically also $\rho(A_0) > 0$ by Proposition 7.8.5.

The next result shows how one may construct an invariant probability from having a suitable form of *local stationarity* within a given path-open set.

For $A_0 \in \mathcal{G}$ path-open, recall the definition (7.104) of $\overline{A}_0$ and note that with respect to any x, $X_{\mathrm{T}_{t_0,A_0}} \in \overline{A}_0$ $\mathbb{P}_{|x}$-a.s. on the set $\left(\mathrm{T}_{t_0,A_0} < \infty\right)$: by the definition of T_{t_0,A_0}, $\widetilde{\theta}_{\mathrm{T}_{t_0,A_0}}\left(\mathrm{T}_{A_0}\right) = 0$ on $\left(\mathrm{T}_{t_0,A_0} < \infty\right)$, so by the strong Markov property

$$\begin{aligned}\mathbb{P}_{|x}\left(\mathrm{T}_{t_0,A_0} < \infty\right) &= \mathbb{P}_{|x}\left(\mathrm{T}_{t_0,A_0} < \infty, \widetilde{\theta}_{\mathrm{T}_{t_0,A_0}}\left(\mathrm{T}_{A_0}\right) = 0\right) \\ &= \mathbb{E}_{|x}\left[\mathbb{P}_{|X_{\mathrm{T}_{t_0,A_0}}}\left(\mathrm{T}_{A_0} = 0\right); \mathrm{T}_{t_0,A_0} < \infty\right]\end{aligned}$$

so that $\mathbb{P}_{|X_{\mathrm{T}_{t_0,A_0}}}\left(\mathrm{T}_{A_0} = 0\right) = 1$ $\mathbb{P}_{|x}$-a.s. on $\left(\mathrm{T}_{t_0,A_0} < \infty\right)$, which is precisely to say that $X_{\mathrm{T}_{t_0,A_0}} \in \overline{A}_0$ $\mathbb{P}_{|x}$-a.s. This observation is relevant for (7.110) below, ensuring that this identity is automatic for $B = \overline{A}_0$.

Theorem 7.8.9 *Suppose that $A_0 \in \mathcal{G}$ is path-open and that for some $t_0 > 0$ it holds that there is a positive bounded measure ζ_{t_0} on $\left(\overline{A}_0, \mathcal{G} \cap \overline{A}_0\right)$ with $\zeta_{t_0}\left(\overline{A}_0\right) > 0$ such that*

$$C := \int_{\overline{A}_0} \zeta_{t_0}(dx)\, \mathbb{E}_{|x}\, \mathrm{T}_{t_0,A_0} < \infty \tag{7.109}$$

and

$$\int_{\overline{A}_0} \zeta_{t_0}(dx)\, \mathbb{P}_{|x}\left(X_{\mathrm{T}_{t_0,A_0}} \in B\right) = \zeta_{t_0}(B) \quad \left(B \in \mathcal{G} \cap \overline{A}_0\right). \tag{7.110}$$

Then the probability ρ on $(G, \mathcal{G})$ given by

$$\rho(B) = \frac{1}{C} \int_{\overline{A}_0} \zeta_{t_0}(dx)\, \mathbb{E}_{|x} \int_0^{\mathrm{T}_{t_0,A_0}} 1_{(X_s \in B)}\, ds \quad (B \in \mathcal{G}) \tag{7.111}$$

satisfies that

$$\rho(\mathrm{A}h) = 0$$

for all h in the domain $\mathcal{D}(\mathrm{A})$ for the full infinitesimal generator. In particular, if $\mathcal{D}(\mathrm{A})$ is a determining class, then ρ is an invariant probability for X.

Proof. First note that by (7.111), for any $f : G \to \mathbb{R}$ bounded and measurable,

$$\rho(f) = \frac{1}{C} \int_{\overline{A}_0} \zeta_{t_0}(dx)\, \mathbb{E}_{|x} \int_0^{\mathrm{T}_{t_0,A_0}} f(X_s)\, ds. \tag{7.112}$$

Next take $h \in \mathcal{D}(\mathrm{A})$. By Itô's formula (7.78),

$$h(X_t) = h(X_0) + \int_0^t \mathrm{A}h(X_s)\, ds + M_t(S) \quad (t \geq 0)$$

with $M(S)$ a mean-zero martingale with respect to $\mathbb{P}_{|x}$ for all x. By optional sampling therefore, for all t,

$$\mathbb{E}_{|x}\, h\left(X_{\mathrm{T}_{t_0,A_0} \wedge t}\right) = h(x) + \mathbb{E}_{|x} \int_0^{\mathrm{T}_{t_0,A_0} \wedge t} \mathrm{A}h(X_s)\, ds$$

and letting $t \to \infty$ and using dominated convergence, since

$$\left| \int_0^{\mathrm{T}_{t_0,A_0} \wedge t} \mathrm{A}h(X_s)\, ds \right| \leq \left(\sup_G |\mathrm{A}h| \right) \mathrm{T}_{t_0,A_0},$$

we deduce that for all x such that $\mathbb{E}_{|x}\, \mathrm{T}_{t_0,A_0} < \infty$,

$$\mathbb{E}_{|x}\, h\left(X_{\mathrm{T}_{t_0,A_0}}\right) = h(x) + \mathbb{E}_{|x} \int_0^{\mathrm{T}_{t_0,A_0}} \mathrm{A}h(X_s)\, ds. \tag{7.113}$$

By the assumption (7.109) this is valid in particular for ζ_{t_0}-a.a. $x \in \overline{A}_0$, so applying (7.112) with $f = \mathrm{A}h$ gives

$$\rho(\mathrm{A}h) = \frac{1}{C} \int_{\overline{A}_0} \zeta_{t_0}(dx) \left(\mathbb{E}_{|x}\, h\left(X_{\mathrm{T}_{t_0,A_0}}\right) - h(x) \right) = 0$$

by (7.110). The proof is completed by referring to Theorem 7.8.2. □

The condition (7.110) is the condition for what could be called local stationarity on A_0. If satisfied, one then obtains from (7.111) a true global invariant probability provided only that $\mathcal{D}(\mathrm{A})$ is a determining class. A useful result is obtained from Theorem 7.8.9 by considering what is only formally a special case corresponding to taking A_0 a one-point set (which is path-open only for Markov chains and which forces ζ_{t_0} to be a point mass at x_0) and $t_0 = 0$: for $x \in G$ define

$$\mathrm{T}_x = \inf \{ t > 0 : X_t = x \}.$$

This description of the hitting time T_x will be used below on the set $(X_0 = x)$, hence the qualification '$t > 0$' in the definition which allows the possibility that $\mathrm{T}_x > 0$ on $(X_0 = x)$: this possibility occurs iff $\phi_t(x) \neq x$ for $t \in\,]0, \delta[$ for some $\delta > 0$ and in that case $\mathbb{P}_{|x_0}\left(\mathrm{T}_{x_0} > 0\right) = 1$.

Corollary 7.8.10 *Suppose there exists $x_0 \in G$ such that $\mathbb{P}_{|x_0}\left(\mathrm{T}_{x_0} > 0\right) = 1$ and $C := \mathbb{E}_{|x_0}\mathrm{T}_{x_0} < \infty$. Then the probability measure ρ on $(G, \mathcal{G})$ given by*

$$\rho(B) = \frac{1}{C}\mathbb{E}_{|x_0}\int_0^{\mathrm{T}_{x_0}} 1_{(X_s \in B)}\, ds \quad (B \in \mathcal{G}) \tag{7.114}$$

satisfies that

$$\rho(\mathrm{A}h) = 0$$

for all h in the domain $\mathcal{D}(\mathrm{A})$ for the full infinitesimal generator. In particular, if $\mathcal{D}(\mathrm{A})$ is a determining class, then ρ is an invariant probability for X.

Proof. For f bounded and measurable,

$$\rho(f) = \frac{1}{C}\mathbb{E}_{|x_0}\int_0^{\mathrm{T}_{x_0}} f(X_s)\, ds. \tag{7.115}$$

Now just follow the steps in the proof of Theorem 7.8.9, using $x = x_0$ only and replacing T_{t_0, A_0} by T_{x_0}. The analogue of (7.113) is then, for $h \in \mathcal{D}(\mathrm{A})$,

$$\mathbb{E}_{|x_0} h\left(X_{\mathrm{T}_{x_0}}\right) = h(x_0) + \mathbb{E}_{|x_0}\int_0^{\mathrm{T}_{x_0}} \mathrm{A}h(X_s)\, ds,$$

and since $h\left(X_{\mathrm{T}_{x_0}}\right) = h(x_0)$ $\mathbb{P}_{|x_0}$-a.s., $\rho(\mathrm{A}h) = 0$ follows from (7.115). □

We conclude this section with a discussion of the invariant probabilities for the backward and forward recurrence time processes associated with renewal processes.

Example 7.8.3 Consider the backward recurrence time process from Subsection 7.4.1 so that $X_t = t - T_{\langle t \rangle}$ with $T_n = V_1 + \cdots + V_n$ the sum of independent waiting times such that the V_k for $k \geq 2$ are identically distributed with a distribution with hazard function u. We have $G = \left[0, t^\dagger\right[$ and

$$\phi_t(x) = x + t, \quad q(x) = u(x), \quad r(x, \cdot) = \varepsilon_0.$$

It follows that a function h is path-continuous iff it is continuous and path-differentiable iff it is continuously differentiable.

In order for the process X to satisfy the assumption involving the continuity of (7.75) we assume that u is continuous and then see that

$$\begin{aligned} t \mapsto q(\phi_t x) &\int r(\phi_t x, dy)\, h(y) \\ &= u(x+t)h(0) \end{aligned}$$

is also continuous. It is now clear that the domain $\mathcal{D}(\mathrm{A})$ for the full infinitesimal generator consists of all continuously differentiable h such that

$$\mathrm{A}h(x) = h'(x) + u(x)\,(h(0) - h(x)) \tag{7.116}$$

is bounded. Note that $\mathcal{D}\,(\mathrm{A})$ is a determining class since e.g., all continuously differentiable h with compact support in G are included.

Suppose that an invariant probability ρ exists. Then for some $0 < x_0 < t^\dagger$ we have $\rho\,(A_0) > 0$, where A_0 is the path-open set $[0, x_0[$, so by Theorem 7.8.7, for ρ-a.a. $x \in A_0$ we have for all t that $\mathbb{E}_{|x}\,\mathrm{T}_{t,A_0} < \infty$, a fact we shall use for some $t \in \left]x_0 - x, t^\dagger - x\right[$: when starting from $x < x_0$, if $T_1 > t > x_0 - x$ the first return to A_0 occurs at time $\mathrm{T}_{t,A_0} = T_1$ and thus

$$\mathbb{E}_{|x}\,\mathrm{T}_{t,A_0} \geq \mathbb{E}_{|x}\,[T_1; T_1 > t]\,.$$

Using (7.47) it is easily seen that the expression on the right is $< \infty$ iff $\xi := \mathbb{E}V_2 = \mathbb{E}_{|0}\,T_1 < \infty$ and we have shown that in order for X *to have an invariant probability it is necessary that* $\xi = \int_0^{t^\dagger} v f_V(v)\,dv < \infty$, f_V denoting the density for the V_n. Hence from now on we assume that $\xi < \infty$.

We now use Theorem 7.8.2 to find what turns out to be the unique invariant probability ρ. Considering $h \in \mathcal{D}\,(\mathrm{A})$ with compact support $\subset \left[0, t^\dagger\right[$ and using (7.116) this gives the condition

$$\begin{aligned}\int_{[0,t^\dagger[} \rho\,(dx)\;h'(x) &= -\int_{[0,t^\dagger[} \rho\,(dx)\;u(x)\,(h(0) - h(x))\\ &= \int_{[0,t^\dagger[} \rho\,(dx)\;u(x)\int_0^x dy\,h'(y)\\ &= \int_0^{t^\dagger} dy\,h'(y)\int_{[y,t^\dagger[} \rho\,(dx)\;u(x)\end{aligned}$$

forcing

$$\rho\,(dx) = \left(\int_{[x,t^\dagger[} \rho\,(dy)\;u(y)\right)dx.$$

In particular, ρ is absolutely continuous with a density f_ρ satisfying

$$f_\rho(x) = \int_x^{t^\dagger} dy\;f_\rho\,(y)\,u\,(y)$$

with respect to Lebesgue measure. We deduce that f_ρ is continuously differentiable with

$$f_\rho'(x) = -f_\rho(x)u(x),$$

so

$$f_\rho(x) = K\exp\left(-\int_0^x u(y)\,dy\right)$$

with the normalizing constant K determined so that $\int f_\rho = 1$, i.e.,

$$f_\rho(x) = \frac{1}{\xi}\exp\left(-\int_0^x u(y)\,dy\right) = \frac{1}{\xi}\overline{F}_V(x) \tag{7.117}$$

with $\overline{F}_V$ the survivor function for the distribution of the V_n.

We are now ready to show that $\rho\,(dx) = f_\rho(x)\,dx$ with f_ρ given by (7.117) *is the unique invariant probability for X*: as just argued, this ρ is the only possible candidate and to see that it succeeds we verify that $\rho\,(\mathrm{A}h) = 0$ for all h in the determining class $\mathcal{D}\,(\mathrm{A})$. Taking limits to avoid problems with integrability we find

$$\begin{aligned}\xi\rho\,(\mathrm{A}h) &= \lim_{a\uparrow t^\dagger}\left[\int_0^a dx\,\overline{F}_V(x)h'(x) - \int_0^a dx\,\overline{F}_V(x)u(x)\int_0^x dy\,h'(y)\right]\\ &= \lim_{a\uparrow t^\dagger}\left[\int_0^a dx\,\overline{F}_V(x)h'(x) - \int_0^a dy\,h'(y)\int_y^a dx\,\overline{F}_V(x)u(x)\right],\end{aligned}$$

which since $f_V(x) = \overline{F}_V(x)u(x)$

$$\begin{aligned}&= \lim_{a\uparrow t^\dagger}\left[\int_0^a dx\,\overline{F}_V(x)h'(x) + \int_0^a dy\,h'(y)\left(\overline{F}_V(a) - \overline{F}_V(y)\right)\right]\\ &= \lim_{a\uparrow t^\dagger}\overline{F}_V(a)\,(h(a) - h(0)) = 0,\end{aligned}$$

since h is bounded and $\overline{F}_V(a) \to 0$.

Corollary 7.8.10 provides an alternative and quick way of finding the density f_ρ: taking $x_0 = 0$ in the corollary it is clear that $\mathbb{P}_{|0}$-a.s, $\mathrm{T}_0 = T_1 > 0$, in particular $\mathbb{E}_{|0}\,\mathrm{T}_0 = \xi < \infty$. Thus ρ given by (7.114) is invariant, i.e., for $x \in G$,

$$\begin{aligned}\rho\,([0,x]) &= \frac{1}{\xi}\mathbb{E}_{|0}\int_0^{T_1} 1_{(X_s\le x)}\,ds = \frac{1}{\xi}\mathbb{E}_{|0}\,(T_1 \wedge x)\\ &= \frac{1}{\xi}\left(\int_0^x v f_V(v)\,dv + x\overline{F}_V(x)\right),\end{aligned}$$

which by differentiation gives

$$f_\rho(x) = \frac{1}{\xi}\overline{F}_V(x).$$

It should be emphasized that the results shown in this example hold in greater generality: any backward recurrence time process such that $\xi < \infty$ has a unique invariant probability ρ with density $\frac{1}{\xi}\overline{F}_V(x)$ — the distribution of the V_n can be anything and need certainly not have a continuous density.

Example 7.8.4 In this example we look first at the forward recurrence time process $X_t = T_{\langle t\rangle+1} - t$ with $T_n = V_1 + \cdots + V_n$ the sum of independent waiting times, the V_k for $k \geq 2$ being identically distributed according to an arbitrary distribution F_V on $\mathbb{R}_+$.

No matter what F_V is, the main result Theorem 7.8.2 cannot be used: for that result and for the definition of the full infinitesimal generator it was assumed throughout that the conditional jump time distributions $P_{z_n}^{(n)}$ should have a density (in order for the jump intensities $q(y)$ to exist) and now we have that $P_{z_n}^{(n)} = \varepsilon_{t_n+y_n}$ is degenerate. However, the discussion of the extended generator in Example 7.6.1 makes it possible to find the invariant probability ρ when it exists, as we shall now see.

It may be shown that if ρ exists it is unique and that it exists iff $\xi = \mathbb{E}V_n < \infty$, just as in the case for the backward process treated in the preceding example.

Suppose now that $h \in \mathcal{D}(A)$, the domain defined in Example 7.6.1: h is bounded, continuously differentiable with h' bounded and

$$h(0) = \int h(y)\, F_V\,(dy)\,. \tag{7.118}$$

Let $0 < x < t^\dagger$ and consider $h \in \mathcal{D}(A)$ such that $h(y)$ (and therefore also $h'(y)$) vanishes for $x \le y < t^\dagger$. By partial integration from (7.118) we find

$$h(0) = -\left[h(y)\overline{F}_V(y)\right]_0^x + \int_0^x h'(y)\overline{F}_V(y)\,dy$$

or, equivalently, that

$$\int_0^x h'(y)\overline{F}_V(y)\,dy = 0.$$

If ρ is invariant, by (7.70) $\rho\left(h'\right) = 0$,

$$\int_{[0,x[} \rho\,(dy)\; h'\,(y) = 0,$$

and letting h and also x and h to vary, it emerges that

$$\rho\,(dx) = \frac{1}{\xi}\overline{F}_V\,(x)\;dx \tag{7.119}$$

is an invariant probability and, in fact, the only one. Thus, in particular the backward and forward recurrence time processes have the same invariant probability.

As a special case, suppose that $F_V = \varepsilon_{x_0}$ is degenerate. Subject to any $\mathbb{P}_{|x}$, where $0 < x \le x_0$, the forward process X is purely deterministic (non-random), yet it has a non-degenerate invariant probability ρ_{det} given by (7.119): we have $\xi = x_0$ and $\overline{F}_V(x) = 1$ for all $x < x_0$ so

$$\rho_{\text{det}} \text{ is the uniform distribution on }]0, x_0]\,. \tag{7.120}$$

Consider next the joint backward and forward process $\mathbf{X} = \left(X, \widetilde{X}\right)$; cf. Example 7.6.1. From the discussion there it follows that if ρ is invariant for $\mathbf{X}$, then $\rho\left(A^{(2)}h\right) = 0$ where

$$A^{(2)}h = (D_x - D_{\widetilde{x}})\,h$$

and h apart from being bounded and continuously differentiable must also satisfy the boundary conditions

$$h\,(x,0) = h(0,0) = \int h\,(0,\widetilde{y})\;F_V\,(d\widetilde{y}) \tag{7.121}$$

for all x.

We shall now indicate how ρ may be found, assuming of course that $\xi < \infty$ but also that F_V has a continuous density f_V. Then in fact ρ is unique with a smooth density $g\,(x,\widetilde{x})$ which we proceed to determine.

Let $0 < x, \widetilde{x} < t^\dagger$ and assume that $h \in \mathcal{D}\left(A^{(2)}\right)$ has compact support $\subset\ ]0,x[\ \times\ ]0,\widetilde{x}[$. Using $\rho\left(A^{(2)}h\right) = 0$ and partial integration yields

$$\iint \left(D_y g - D_{\widetilde{y}} g\right) h\,dy\,d\widetilde{y} = 0,$$

so necessarily $\left(D_y - D_{\widetilde{y}}\right) g \equiv 0$, i.e., g must have the form

$$g\,(y,\widetilde{y}) = g^*\,(y+\widetilde{y})$$

for some continuously differentiable g^*. In order to exploit the boundary conditions (7.121) suppose next that h has compact support $\subset [0,x[\ \times [0,\widetilde{x}[$, i.e., $h\,(y,\widetilde{y})$ need no longer vanish when y or $\widetilde{y} = 0$. The condition $\rho\left(A^{(2)}h\right) = 0$ reads

$$\int_0^x dy \int_0^{\widetilde{x}} d\widetilde{y}\,g^*\,(y+\widetilde{y})\left(D_y - D_{\widetilde{y}}\right) h = 0$$

or, by partial integration

$$\begin{aligned} &\int_0^{\widetilde{x}} d\widetilde{y}\left\{\left[g^*\,(y+\widetilde{y})\,h\,(y,\widetilde{y})\right]_{y=0}^{x} - \int_0^x dy\,g^{*\prime}\,(y+\widetilde{y})\,h\,(y,\widetilde{y})\right\} \\ &= \int_0^x dy\left\{\left[g^*\,(y+\widetilde{y})\,h\,(y,\widetilde{y})\right]_{\widetilde{y}=0}^{\widetilde{x}} - \int_0^{\widetilde{x}} d\widetilde{y}\,g^{*\prime}\,(y+\widetilde{y})\,h\,(y,\widetilde{y})\right\}. \end{aligned}$$

Equivalently, since $h\,(x,\widetilde{y}) = h\,(y,\widetilde{x}) = 0$ by the assumptions on h,

$$\int_0^{\widetilde{x}} d\widetilde{y}\,g^*\,(\widetilde{y})\,h\,(0,\widetilde{y}) = \int_0^x dy\,g^*\,(y)\,h\,(y,0)\,.$$

Now we invoke the boundary conditions: by (7.121) $h\,(y,0) = h(0,0)$ must not depend on y and since $h(x,0) = 0$ by assumption, necessarily $h\,(y,0) = 0$ for all y. We are left with

$$\int_0^{\widetilde{x}} d\widetilde{y}\,g^*\,(\widetilde{y})\,h\,(0,\widetilde{y}) = 0 \tag{7.122}$$

and the requirement from (7.121) that

$$0 = h(0,0) = \int F_V\,(d\widetilde{y})\; h\,(0,\widetilde{y}) = \int_0^{\widetilde{x}} d\widetilde{y}\, f_V\,(d\widetilde{y})\, h\,(0,\widetilde{y})\,. \tag{7.123}$$

Letting $\widetilde{x}$ and h vary and comparing (7.122) and (7.123) finally gives $g^* = Cf_V$ for some normalising constant C, i.e., the invariant probability ρ is given by

$$\rho\,(dx, d\widetilde{x}) = \frac{1}{\xi} f_V\,(x+\widetilde{x})\; dx\, d\widetilde{x}. \tag{7.124}$$

Note that if $\left(U, \widetilde{U}\right)$ has distribution ρ, U and $\widetilde{U}$ have the same marginal distribution with density $\frac{1}{\xi}\overline{F}_V$ in agreement with the results found for respectively the backward and forward recurrence time process.

Suppose that $\mathbf{X}$ is started according to the invariant distribution ρ with density g determined as in (7.124). Defining $T_0 = -X_0$ and $V_1 = T_1 - T_0$ we have for any $t \geq 0$ that $X_t = t - T_{\langle t\rangle}$, $\widetilde{X}_t = T_{\langle t\rangle+1} - t,$ and because of the stationarity, we obtain that the joint $\mathbb{P}_\rho$-distribution of $\left(t - T_{\langle t\rangle}, T_{\langle t\rangle+1} - t\right)$ and therefore also the $\mathbb{P}_\rho$-distribution of $\left(t - T_{\langle t\rangle}\right) + \left(T_{\langle t\rangle+1} - t\right) = V_{\langle t\rangle+1}$ is the same for all t. A priori one might expect this distribution of the $V_t = V_{\langle t\rangle+1}$ to be the same as that of the V_k for $k \geq 2$, but this is not the case. The density for V_t is that of $X_t + \widetilde{X}_t$, i.e., the density is

$$f_{V_t}\,(z) = \int_0^z g\,(x, z-x)\; dx = \frac{z}{\xi} f_V\,(z) \quad (z > 0)$$

which is certainly not f_V. It even holds (see the last paragraph of this example for a proof) that for all z,

$$F_{V_t}(z) \leq F_V\,(z) \tag{7.125}$$

with F_{V_t}, F_V the distribution functions determined from f_{V_t} and f_V, i.e., V_t is stochastically larger than the V_k. This is the so-called *waiting time paradox* which is particularly striking if the renewal counting process $\overline{N}$ is a Poisson process. If $\lambda > 0$ is the Poisson parameter we have $f_V(x) = \lambda e^{-\lambda x}$ with $\xi = 1/\lambda$ and find from (7.124) that $g\,(x,\widetilde{x}) = \lambda^2 e^{-\lambda(x+\widetilde{x})}$, i.e., the two components are iid exponential with rate λ and thus the distribution of V_t is that of the sum $V_2 + V_3$ say, which certainly shows V_t to be larger (in the stochastic sense) than the V_k.

In one special case the V_t have the same distribution of the V_k. Suppose F_V is degenerate at $x_0 > 0$. Although there is no density f_V as was assumed above, one may still show that $\mathbf{X}$ has an invariant probability ρ which is the distribution of $(x_0 - U, U)$ with U uniformly distributed on $]0, x_0];$ cf. (7.120). But now of course all $V_k \equiv x_0$ so also $V_t \equiv x_0$.

Finally we give the proof of (7.125). If V is an $\mathbb{R}$-valued random variable with (an arbitrary) distribution function F_V, (7.125) may be written

$$\mathbb{E}\,[V; V \leq z] \leq (\mathbb{E}V)\,\mathbb{P}\,(V \leq z)\,,$$

i.e., it is claimed that V and $1_{(V \le z)}$ are negatively correlated for all $z \in \mathbb{R}$. But for $z \le \xi = \mathbb{E}V$,

$$\mathbb{E}[V; V \le z] \le z\mathbb{P}(V \le z) \le \xi\mathbb{P}(V \le z)$$

and for $z \ge \xi$,

$$\begin{aligned}\mathbb{E}[V; V \le z] &= \xi - \mathbb{E}[V; V > z] \\ &\le \xi - z\mathbb{P}(V > z) \\ &\le \xi(1 - \mathbb{P}(V > z)).\end{aligned}$$

The argument also shows that in (7.125) the inequality is sharp if either $z < \xi$ is such that $\mathbb{P}(V \le z) > 0$ or $z \ge \xi$ is such that $\mathbb{P}(V > z) > 0$. In particular equality holds for all z iff V is degenerate.

7.9 Likelihood processes for PDMPs

Using the results of Section 5.1 together with the construction from Section 7.3, it is an easy matter to derive likelihood processes for observation of PDMPs.

Suppose given an RCM $\mu = \sum_{n:T_{n<\infty}} \varepsilon_{(T_n, Y_n)}$ on some measurable space $(\Omega, \mathcal{F})$, and let $\mathbb{P}, \widetilde{\mathbb{P}}$ be two probability measures on $(\Omega, \mathcal{F})$. Write $\mathbb{P}_t, \widetilde{\mathbb{P}}_t$ for the restrictions of $\mathbb{P}, \widetilde{\mathbb{P}}$ to $\mathcal{F}_t^\mu$ and write $\widetilde{\mathbb{P}} \ll_{\text{loc}} \mathbb{P}$ if $\widetilde{\mathbb{P}}_t \ll \mathbb{P}_t$ for all $t \in \mathbb{R}_0$. The likelihood process

$$\mathcal{L}^\mu = \left(\mathcal{L}_t^\mu\right)_{t \ge 0} = \left(\frac{d\widetilde{\mathbb{P}}_t}{d\mathbb{P}_t}\right)_{t \ge 0}$$

for observing μ is then given by the results from Section 5.1. More specifically, if $Q, \widetilde{Q}$ are the distributions of μ under $\mathbb{P}, \widetilde{\mathbb{P}}$ respectively, then $\widetilde{\mathbb{P}} \ll_{\text{loc}} \mathbb{P}$ iff $\widetilde{Q} \ll_{\text{loc}} Q$ and

$$\mathcal{L}_t^\mu = \mathcal{L}_t \circ \mu,$$

where $\mathcal{L}_t = \frac{d\widetilde{Q}_t}{dQ_t}$ is the likelihood process from Section 5.1.

Now, suppose further that $X = (X_t)_{t \ge 0}$ is a $(G, \mathcal{G})$-valued PDMP under both $\mathbb{P}$ and $\widetilde{\mathbb{P}}$, defined by $X_0 \equiv x_0$ and

$$X_t = \phi_{T_{\langle t \rangle}, t}\left(Y_{\langle t \rangle}\right),$$

where each $t \mapsto \phi_{st}(y)$ is continuous and the ϕ_{st} satisfy the semigroup property (7.25) and the boundary condition (7.26); cf. Theorem 7.3.1. To emphasize the value of the initial state we shall write $\mathbb{P}_{|x_0}, \widetilde{\mathbb{P}}_{|x_0}$ instead of $\mathbb{P}, \widetilde{\mathbb{P}}$. Finally we shall assume that the Markov kernels generating $\mathbb{P}_{|x_0}$ are as in Theorem 7.3.1, see (7.30) and (7.31), while those generating $\widetilde{\mathbb{P}}_{|x_0}$ have a similar structure,

$$\widetilde{\overline{P}}^{(n)}_{z_n|x_0}(t) = \exp\left(-\int_{t_n}^{t} \widetilde{q}_s\left(\phi_{t_n,s}(y_n)\right) ds\right)$$

$$\widetilde{\pi}^{(n)}_{z_n,t|x_0}(C) = \widetilde{r}_t\left(\phi_{t_n,t}(y_n), C\right),$$

with $y_0 = x_0$, where $\widetilde{q}_s \geq 0$ and each $\widetilde{r}_t$ is a Markov kernel on G such that $\widetilde{r}_t(y, \{y\}) = 0$ for all t, y, cf. (7.29). It is also assumed that the RCMs generated by the kernels $P^{(n)}, \pi^{(n)}$ and $\widetilde{P}^{(n)}, \widetilde{\pi}^{(n)}$ respectively, are stable for any choice of x_0.

Thus X is a PDMP under both $\mathbb{P}_{|x_0}$ and $\widetilde{\mathbb{P}}_{|x_0}$ and from Theorem 5.1.1 we immediately obtain the following result:

Theorem 7.9.1 *A sufficient condition for* $\widetilde{\mathbb{P}}_{|x_0} \ll_{\text{loc}} \mathbb{P}_{|x_0}$ *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* $\widetilde{r}_t(y, \cdot) \ll r_t(y, \cdot)$ *for all* t, y.

In that case an $\left(\mathcal{F}_t^X\right)$*-adapted version of the likelihood process* $\mathcal{L}^{\mu}_{|x_0}$ *is* $\mathcal{L}^X_{|x_0} = \left(\mathcal{L}^X_{t|x_0}\right)_{t\geq 0}$ *given by*

$$\mathcal{L}^X_{t|x_0} = \exp\left(-\int_0^t (\widetilde{q}_s - q_s)(X_s)\, ds\right) \prod_{n=1}^{\overline{N}_t^X} \frac{\widetilde{q}_{T_n^X}\left(Y_{n-}^X\right)}{q_{T_n^X}\left(Y_{n-}^X\right)} \frac{d\widetilde{r}_{T_n^X}\left(Y_{n-}^X, \cdot\right)}{dr_{T_n^X}\left(Y_{n-}^X, \cdot\right)} \left(Y_n^X\right). \tag{7.126}$$

Notation. Of course $\overline{N}_t^X$ is the total number of jumps for X on $[0, t]$, T_n^X is the time of the nth jump for X and $Y_n^X = X_{T_n^X}$ is the state reached by that jump; finally Y_{n-}^X is short for $X_{T_n^X-}$. Note that $\overline{N}^X$ is $\mathbb{P}_{|x_0}$-indistinguishable from $\overline{N}$ and that for all n, $T_n^X = T_n$, $Y_{n-}^X = Y_n$ and $Y_{n-}^X = X_{T_n^X-}$ $\mathbb{P}_{|x_0}$-a.s.

We have not given Theorem 7.9.1 in its most general form, which would amount to a direct translation of Theorem 5.1.1. 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 PDMPs, 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 in the case above, while the q_t, r_t can change into $\widetilde{q}_t, \widetilde{r}_t$, the ϕ_{st} must remain unchanged when switching from one process to the other.

Note that the expression (7.126) for the likelihood, depends on x_0 only through the fact that $X_0 = x_0$ $\mathbb{P}_{|x_0}$- and $\widetilde{\mathbb{P}}_{|x_0}$-a.s.

If X is time-homogeneous under both $\mathbb{P}_{|x_0}$ and $\widetilde{\mathbb{P}}_{|x_0}$, i.e., when $\phi_{st} = \phi_{t-s}$ and neither of $q_t, \widetilde{q}_t, r_t, \widetilde{r}_t$ depends on t, we have $\widetilde{\mathbb{P}}_{|x_0} \ll_{\text{loc}} \mathbb{P}_{|x_0}$ if $q(y) > 0$ for all y, $\int_0^t q(\phi_s(y))\, ds < \infty$ for all t, y and $\widetilde{r}(y, \cdot) \ll r(y, \cdot)$ for all y with

$$\mathcal{L}^X_{t|x_0} = \exp\left(-\int_0^t (\widetilde{q} - q)(X_s)\, ds\right) \prod_{n=1}^{\overline{N}_t^X} \frac{\widetilde{q}\left(Y_{n-}^X\right)}{q\left(Y_{n-}^X\right)} \frac{d\widetilde{r}\left(Y_{n-}^X, \cdot\right)}{dr\left(Y_{n-}^X, \cdot\right)} \left(Y_n^X\right). \tag{7.127}$$

The likelihood processes (7.126) and (7.127) have an important multiplicative structure which we shall now discuss.

Suppose that the conditions of Theorem 7.9.1 hold and consider for $s > 0$ and $x \in G$ given, the $\mathbb{P}_{|x_0}$- and $\widetilde{\mathbb{P}}_{|x_0}$-conditional distribution of $\theta_s\mu$ given $X_s = x$. These conditional distributions are $Q^{s,x}$ and $\widetilde{Q}^{s,x}$ where e.g., $Q^{s,x}$ is generated by the Markov kernels (see (7.36) and (7.38)),

$$\overline{P}^{(n)}_{z_n|s,x}(t) = \exp\left(-\int_{t_n}^{t} q_u\left(\phi_{t_n u}(y_n)\right) du\right) \quad (t \geq t_n),$$
$$\pi^{(n)}_{z_n,t|s,x} = r_t\left(\phi_{t_n t}(y_n), \cdot\right) \quad (t > t_n)$$

using $t_0 = s$ (!) and $y_0 = x$, and similarly for $\widetilde{Q}^{s,x}$. But it is then clear from Theorem 5.1.1 (or in fact also from Theorem 7.9.1) that $\widetilde{Q}^{s,x} \ll_{\text{loc}} Q^{s,x}$ with

$$\mathcal{L}_t^{\circ s,x} := \frac{d\widetilde{Q}_t^{s,x}}{dQ_t^{s,x}}$$

given by

$$\mathcal{L}_t^{\circ s,x} = 1_{(\tau_1 > s)} \exp\left(-\int_s^t (\widetilde{q}_u - q_u)\left(X_u^\circ\right) du\right) \prod_{n=1}^{\overline{N}_t^\circ} \frac{\widetilde{q}_{\tau_n}\left(X^\circ_{\tau_n-}\right)}{q_{\tau_n}\left(X^\circ_{\tau_n-}\right)} \frac{d\widetilde{r}_{\tau_n}\left(X^\circ_{\tau_n-}, \cdot\right)}{dr_{\tau_n}\left(X^\circ_{\tau_n-}, \cdot\right)}\left(X^\circ_{\tau_n}\right)$$

for $t \geq s$ and with $X^\circ = \left(X_u^\circ\right)_{u \geq s}$ the canonical PDMP after time s defined on $(\mathcal{M}, \mathcal{H})$ by

$$X_u^\circ = \phi_{\tau_{\langle u\rangle},u}\left(\eta_{\langle u\rangle}\right) \circ \theta_s \quad (u \geq s) \tag{7.128}$$

using $\tau_0 \equiv s$ and $\eta_0 \equiv x$. (Because $Q^{s,x}(\tau_1 > s) = 1$ all jumps before time s are irrelevant, which explains the appearance of the shift θ_s in the definition of X_u° and the free of charge inclusion of the indicator $1_{(\tau_1 > s)}$ in the expression for $\mathcal{L}_t^{\circ s,x}$. Note that X° here depends on both s and x. This dependence is used systematically below, see (7.130)).

Comparing (7.128) with (7.126) it follows using that $\theta_s \circ \theta_s = \theta_s$, $\overline{N}_t^\circ \circ (\theta_s\mu) = \overline{N}_t - \overline{N}_s$, $\tau_n \circ (\theta_s\mu) = T_{n+\langle s\rangle}$, $\eta_n \circ (\theta_s\mu) = Y_{n+\langle s\rangle}$ (both for $n \geq 1$ only!), $X_u^\circ \circ (\theta_s\mu) = X_u$ when $x = X_s$ and $1_{(\tau_1 > s)} \circ (\theta_s\mu) \equiv 1$, that for $s \leq t$,

$$\mathcal{L}^X_{t|x_0} = \mathcal{L}^X_{s|x_0}\,\mathcal{L}_t^{\circ s,X_s} \circ (\theta_s\mu).$$

In other words, the likelihood process at time t factors into the value of the process at time s, which depends on the 'past' $(X_u)_{u \leq s}$, and a factor depending on the 'future' $\theta_s\mu$, with the two factors connected only through the 'present' X_s. Furthermore, the second factor is the conditional likelihood for observing $(X_u)_{s \leq u \leq t}$ given X_s — or what by the Markov property is the same thing — given $\mathcal{F}_s^X$.

In the homogeneous case, with X time-homogeneous under both $\mathbb{P}_{|x_0}$ and $\widetilde{\mathbb{P}}_{|x_0}$, one may instead consider the conditional distributions Q^x and $\widetilde{Q}^x$ of $\theta_s^*\mu$ given $X_s = x$, (see (7.37) and (7.39)), i.e., Q^x is generated by the Markov kernels

$$\overline{P}^{(n)}_{z_n|x}\,(t) = \exp\left(-\int_0^{t-t_n} q\left(\phi_u\left(y_n\right)\right)\,du\right) \quad (t \ge t_n)\,,$$
$$\pi^{(n)}_{z_n,t|x} = r\left(\phi_{t-t_n}\left(y_n\right),\cdot\right) \quad (t > t_n)\,,$$

using $t_0 = 0$, $y_0 = x$, and similarly for $\widetilde{\mathbb{P}}_{|x_0}$. Thus $\widetilde{Q}^x \ll_{\mathrm{loc}} Q^x$ with (now $X^\circ_u = \phi_{u-\tau_{\langle u\rangle}}\left(\eta_{\langle u\rangle}\right)$ for all $u \ge 0$ and $\tau_0 \equiv 0$, $\eta_0 \equiv x$),

$$\mathcal{L}^{\circ x}_t := \frac{d\widetilde{Q}^x_t}{dQ^x_t} = \exp\left(-\int_0^t \left(\widetilde{q}-q\right)\left(X^\circ_u\right)\,du\right)\prod_{n=1}^{\overline{N}^\circ_t}\frac{\widetilde{q}\left(X^\circ_{\tau_n-}\right)}{q\left(X^\circ_{\tau_n-}\right)}\frac{d\widetilde{r}\left(X^\circ_{\tau_n-},\cdot\right)}{dr\left(X^\circ_{\tau_n-},\cdot\right)}\left(X^\circ_{\tau_n}\right);$$

it follows that we have the factorization

$$\mathcal{L}^X_{s+t|x_0} = \mathcal{L}^X_{s|x_0}\,\mathcal{L}^{\circ X_s}_t \circ \left(\theta^*_s\mu\right) \quad (s,t \ge 0)$$

of the likelihood (7.127).

Having described the multiplicative structure of the likelihood process when both $\mathbb{P}_{|x_0}$ and $\widetilde{\mathbb{P}}_{|x_0}$ make X Markov, we shall now only suppose that $\mathbb{P}_{|x_0}$ is given and assume that X is Markov under $\mathbb{P}_{|x_0}$.

Suppose that under $\mathbb{P}_{|x_0}$, $M = (M_t)_{t\ge 0}$ is a non-negative $\mathcal{F}^\mu_t$-martingale with $M_0 \equiv 1$. By Theorem 5.2.1 there is a unique probability $\widetilde{\mathbb{P}}_{|x_0}$ on $\left(\Omega, \mathcal{F}^\mu_t, \mathcal{F}^\mu\right)$ such that for all $t \ge 0$ and all $F \in \mathcal{F}^\mu_t$,

$$\widetilde{\mathbb{P}}_{|x_0}\,(F) = \int_F M_t\,d\mathbb{P}_{|x_0}\,. \tag{7.129}$$

We shall now discuss conditions on M that ensure that X is Markov also under $\widetilde{\mathbb{P}}_{|x_0}$, and, when X is time-homogeneous under $\mathbb{P}_{|x_0}$, conditions on M that make X homogeneous Markov under $\widetilde{\mathbb{P}}_{|x_0}$.

For this, consider first the canonical space $(\mathcal{M},\mathcal{H})$. The canonical version of X is X° defined by

$$X^\circ_u = \phi_{\tau_{\langle u\rangle},u}\left(\eta_{\langle u\rangle}\right) \quad (u \ge 0,\ \tau_0 \equiv 0,\ \eta_0 \equiv x_0)\,.$$

For $s \ge 0$, $x \in G$ define the canonical version $X^{\circ s,x}$ of X beyond s starting from x as in (7.128),

$$X^{\circ s,x}_u = \phi_{\tau_{\langle u\rangle},u}\left(\eta_{\langle u\rangle}\right)\circ\theta_s \quad (u \ge s,\ \tau_0 \equiv s,\ \eta_0 \equiv x) \tag{7.130}$$

with $X^\circ = X^{\circ 0,x_0}$ a special case. The semigroup property (7.25) entails, as is quite essential, that for $s \le t \le u$,

$$X^{\circ s,x}_u = X^{\circ t,x'}_u \circ \theta_t \text{ for } x' = X^{\circ s,x}_t\,,$$

in particular (take $s = 0$, $x = x_0$ and replace t by s),

$$X^{\circ s,x}_u \circ \left(\theta_s\mu\right) = X_u \text{ for } x = X_s. \tag{7.131}$$

In the homogeneous case, $\phi_{st} = \phi_{t-s}$, introducing $X^{\circ x} \equiv X^{\circ 0,x}$ (so $X^\circ \equiv X^{\circ x_0}$) we similarly have for $s, t \geq 0$ that

$$X_t^{\circ x} \circ \left(\theta_s^* \mu\right) = X_{s+t} \text{ for } x = X_s. \tag{7.132}$$

Next, consider a family $\left(\Phi_t^{s,x}\right)_{0 \leq s \leq t, x \in G}$ of $\mathbb{R}$-valued random variables defined on $(\mathcal{M}, \mathcal{H})$ such that each $\Phi_t^{s,x}$ is $\mathcal{H}_t^s$-measurable, where $\mathcal{H}_t^s$ is the σ-algebra generated by the restriction of the shifted RCM $\theta_s \mu^\circ$ to $[s, t]$; equivalently $\mathcal{H}_t^s$ is the σ-algebra generated by $\left(N_u^\circ (A) - N_s^\circ (A)\right)_{s \leq u \leq t, A \in \mathcal{E}}$. In particular $\Phi_t^x := \Phi_t^{0,x}$ is $\mathcal{H}_t$-measurable. Call $\left(\Phi_t^{s,x}\right)$ a *multiplicative X -functional* if $(s, t, x, m) \mapsto \Phi_t^{s,x}(m)$ is measurable, and if $\Phi_s^{s,x} \equiv 1$ for all s, x and for all $s \leq t$, $x \in G$

$$\Phi_t^x = \Phi_s^x \cdot \left(\Phi_t^{s,x'} \circ \theta_s\right)\Big|_{x' = X_s^{\circ 0,x}}. \tag{7.133}$$

Similarly, let $\left(\Psi_t^x\right)_{t \geq 0, x \in G}$ be a family of $\mathbb{R}$-valued, $\mathcal{H}_t$-measurable random variables defined on $(\mathcal{M}, \mathcal{H})$. Call $\left(\Psi_t^x\right)$ a *homogeneous multiplicative X-functional* if $(t, x, m) \mapsto \Psi_t^x(m)$ is measurable, $\Psi_0^x \equiv 1$ for all x and for all $s, t \geq 0$, $x \in G$,

$$\Psi_{s+t}^x = \Psi_s^x \cdot \left(\Psi_t^{x'} \circ \theta_s^*\right)\Big|_{x' = X_s^{\circ 0,x}}.$$

Now return to the setup with X a PDMP under $\mathbb{P}_{|x_0}$ as above and M with $M_0 \equiv 1$ a non-negative $\mathcal{F}_t^\mu$-martingale under $\mathbb{P}_{|x_0}$. Let $\widetilde{\mathbb{P}}_{|x_0}$ be the probability on $\left(\Omega, \mathcal{F}_t^\mu, \mathcal{F}^\mu\right)$ defined by (7.129).

Proposition 7.9.2 (a) *A sufficient condition for X to be a PDMP under* $\widetilde{\mathbb{P}}_{|x_0}$ *is that there exists a non-negative multiplicative X-functional* $\left(\Phi_t^{s,x}\right)$ *such that*

$$M_t = \Phi_t^{x_0}(\mu) \quad (t \geq 0).$$

(b) *Suppose that X is a time-homogeneous PDMP under* $\mathbb{P}_{|x_0}$ *with, in particular,* $\phi_{st} = \phi_{t-s}$. *Then a sufficient condition for X to be a time-homogenous PDMP under* $\widetilde{\mathbb{P}}_{|x_0}$ *is that there exists a non-negative homogeneous multiplicative X-functional* $\left(\Psi_t^x\right)$ *such that*

$$M_t = \Psi_t^{x_0}(\mu) \quad (t \geq 0).$$

Proof. (a) Let $s < t$, $F \in \mathcal{F}_s^\mu$, $C \in \mathcal{G}$ and consider

$$\begin{aligned}
\widetilde{\mathbb{P}}_{|x_0}(F \cap (X_t \in C)) &= \int_{F \cap (X_t \in C)} M_t \, d\mathbb{P}_{|x_0} \\
&= \int_{F \cap (X_t \in C)} \Phi_s^{x_0}(\mu) \left(\Phi_t^{s,x}(\theta_s \mu)\right)\big|_{x = X_s} \, d\mathbb{P}_{|x_0} \\
&= \int_F \Phi_s^{x_0}(\mu) \, \mathbb{E}_{|x_0}\left[\Phi_t^{s,x}(\theta_s \mu); X_t \in C \,\middle|\, \mathcal{F}_s^\mu\right]\big|_{x = X_s} \, d\mathbb{P}_{|x_0}.
\end{aligned}$$

But since by (7.131) $X_t = X_t^{\circ s,x} \circ (\theta_s \mu)$ for $x = X_s$, because of the Markov property (7.38) for X under $\mathbb{P}_{|x_0}$, the conditional expectation reduces to

$$\int \Phi_t^{s,x}(m) \, 1_{\left(X_t^{\circ s,x}(m) \in C\right)} \, Q^{s,x}(dm) = \widetilde{p}_{st}(x, C),$$

say, with $x = X_s$, and we obtain since $M_s = \Phi_s^{x_0}(\mu)$ that

$$\widetilde{\mathbb{P}}_{|x_0}\left(F \cap (X_t \in C)\right) = \int_F \widetilde{p}_{st}(X_s, C) \, d\widetilde{\mathbb{P}}_{|x_0}.$$

This shows that X is Markov under $\widetilde{\mathbb{P}}_{|x_0}$ with transition probabilities $\widetilde{p}_{st}$,

$$\widetilde{\mathbb{P}}_{|x_0}\left(X_t \in C \,\middle|\, \mathcal{F}_s^{\mu}\right) = \widetilde{p}_{st}(X_s, C).$$

(b) Let $s, t \geq 0$, $F \in \mathcal{F}_s^{\mu}$, $C \in \mathcal{G}$ and consider

$$\begin{aligned} \widetilde{\mathbb{P}}_{|x_0}\left(F \cap (X_{s+t} \in C)\right) &= \int_{F \cap (X_{s+t} \in C)} M_{s+t} \, d\mathbb{P}_{|x_0} \\ &= \int_{F \cap (X_{s+t} \in C)} \Psi_s^{x_0}(\mu) \left(\Psi_t^x\left(\theta_s^* \mu\right)\right)\Big|_{x = X_s} \, d\mathbb{P}_{|x_0}. \end{aligned}$$

Now, by (7.132) and proceeding as in the proof of (a) it follows using the homogeneous Markov property (7.39) that X is time-homogenous Markov under $\widetilde{\mathbb{P}}_{|x_0}$ with transition probabilities $\widetilde{p}_t$,

$$\widetilde{\mathbb{P}}_{|x_0}\left(X_{s+t} \in C \,\middle|\, \mathcal{F}_s^{\mu}\right) = \widetilde{p}_t(X_s, C),$$

where, cf. (7.39),

$$\widetilde{p}_t(x, C) = \int \Psi_t^x(m) \, 1_{\left(X_t^{\circ x}(m) \in C\right)} \, Q^x(dm)$$

with $x = X_s$. □

The likelihood process (7.126) corresponds to the multiplicative X-functional

$$\Phi_t^{s,x} = \exp\left(-\int_s^t \left(\widetilde{q}_u - q_u\right)\left(X_u^{\circ s,x}\right) du\right) \prod_{n = \overline{N}_s^{\circ}+1}^{\overline{N}_t^{\circ}} \frac{\widetilde{q}_{\tau_n}\left(X_{\tau_n -}^{\circ s,x}\right)}{q_{\tau_n}\left(X_{\tau_n -}^{\circ s,x}\right)} \frac{d\widetilde{r}_{\tau_n}\left(X_{\tau_n -}^{\circ s,x}, \cdot\right)}{dr_{\tau_n}\left(X_{\tau_n -}^{\circ s,x}, \cdot\right)}\left(X_{\tau_n}^{\circ s,x}\right).$$

The likelihood (7.127) corresponds to the homogeneous multiplicative X-functional

$$\Psi_t^x = \exp\left(-\int_0^t \left(\widetilde{q} - q\right)\left(X_s^{\circ 0,x}\right) ds\right) \prod_{n=1}^{\overline{N}_t^{\circ}} \frac{\widetilde{q}\left(X_{\tau_n -}^{\circ 0,x}\right)}{q\left(X_{\tau_n -}^{\circ 0,x}\right)} \frac{d\widetilde{r}\left(X_{\tau_n -}^{\circ 0,x}, \cdot\right)}{dr\left(X_{\tau_n -}^{\circ 0,x}, \cdot\right)}\left(X_{\tau_n}^{\circ 0,x}\right).$$

Usually multiplicative functionals are assumed to be non-negative. A fairly general class of multiplicative X-functionals is obtained by considering

$$\Phi_t^{s,x} = \frac{h\left(X_u^{\circ t,x}\right)}{h\left(x\right)} \exp\left(\int_s^t a\left(u, X_u^{\circ s,x}\right) du + \int_{]s,t]\times G} b\left(u, X_{u-}^{\circ s,x}, y\right) \mu^\circ\left(du, dy\right)\right)$$

for given functions h, a and b, or, in order to allow the value 0,

$$\Phi_t^{s,x} = \frac{h\left(X_u^{\circ t,x}\right)}{h\left(x\right)} \exp\left(\int_s^t a\left(u, X_u^{\circ s,x}\right) du\right) \prod_{n=\overline{N}_s^\circ+1}^{\overline{N}_t^\circ} c\left(\tau_n, X_{\tau_n-}^{\circ s,x}, \eta_n\right).$$

In the homogeneous case, $\phi_{st} = \phi_{t-s}$, the corresponding forms of the homogeneous X-functionals are

$$\Psi_t^x = \frac{h\left(X_u^{\circ x}\right)}{h\left(x\right)} \exp\left(\int_0^t a\left(X_u^{\circ x}\right) du + \int_{]0,t]\times G} b\left(X_{u-}^{\circ x}, y\right) \mu^\circ\left(du, dy\right)\right)$$

and

$$\Psi_t^x = \frac{h\left(X_u^{\circ x}\right)}{h\left(x\right)} \exp\left(\int_0^t a\left(X_u^{\circ x}\right) du\right) \prod_{n=1}^{\overline{N}_t^\circ} c\left(X_{\tau_n-}^{\circ x}, \eta_n\right).$$

The multiplicative property (7.133) is quite strong: for any $s > 0$ it forces

$$\Phi_{s+ds}^x = \Phi_s^x \Phi_{s+ds}^{s,x'}\Big|_{x'=X_s^{\circ 0,x}}$$

suitably understood, where the last factor must depend only on $X_s^{\circ 0,x}$ and the behaviour of μ on $]s, s+ds]$, i.e., it must depend on $X_s^{\circ 0,x}$ alone. It is this important instantaneous X°-dependence, together with the fact that the $X^{\circ s,x}$-processes are determined from a semigroup (ϕ_{st}) and μ° as in (7.130), that makes the factorizations we have considered at all possible.

Example 7.9.1 Suppose that X with $X_0 \equiv x_0$ is a time-homogeneous PDMP under $\mathbb{P}_{|x_0}$ (so that $X_t = \phi_{t-\langle t\rangle}\left(Y_{\langle t\rangle}\right)$ with $\tau_0 \equiv 0$, $Y_0 \equiv x_0$) with jump intensity function $q\left(x\right)$ and jump probabilities $r\left(x,\cdot\right)$. Suppose also that X satisfies (7.75). Let h be a function in $\mathcal{D}\left(\mathrm{A}\right)$, the domain of the full infinitesimal generator A for X, see Definition 7.7.1, and assume that h is strictly positive and bounded away from 0. Then consider the process

$$M_t^h = \frac{h\left(X_t\right)}{h\left(x_0\right)} U_t^h$$

where

$$U_t^h = \exp\left(-\int_0^t \frac{\mathrm{A}h}{h}\left(X_s\right) ds\right).$$

In particular $M_0^h \equiv 1$ $\mathbb{P}_{|x_0}$-a.s. and $M_t^h = \Psi_t^{x_0} \circ \mu$ where Ψ is the homogeneous multiplicative functional

$$\Psi_t^x = \frac{h\left(X_t^{\circ x}\right)}{h(x)} \exp\left(-\int_0^t \frac{\mathrm{A}h}{h}\left(X_s^{\circ x}\right) ds\right).$$

The jumps for M^h satisfy

$$\Delta M_t^h = \Delta \overline{N}_t S_t^{X_t}$$

with S the $\mathcal{F}_t^\mu$-predictable field

$$S_t^y = M_{t-}^h \left(\frac{h(y)}{h(X_{t-})} - 1\right), \tag{7.134}$$

while between jumps, $t \mapsto M_t^h$ is differentiable with (cf. (7.74) and Definition 7.7.1)

$$\begin{aligned} D_t M_t^h &= \frac{U_t^h}{h(x_0)}\left(\delta_\phi h(X_t) - \mathrm{A}h(X_t)\right) \\ &= -\frac{U_t^h}{h(x_0)} q(X_t) \int_G r(X_t, dy)\ (h(y) - h(X_t)). \end{aligned} \tag{7.135}$$

But then, with

$$M(dt, dy) = \mu(dt, dy) - q(X_{t-})\ dt\, r(X_{t-}, dy)$$

the fundamental martingale measure for μ, it follows that

$$\begin{aligned} M_t^h &= 1 + \int_0^t D_s M_s^h\, ds + \int_{]0,t]\times G} S_s^y\ \mu(ds, dy) \\ &= 1 + \int_{]0,t]\times G} S_s^y\ M(ds, dy), \end{aligned}$$

simply because using (7.135) and (7.134) we see that

$$-\int_0^t ds\, q(X_{s-}) \int_G r(X_{s-}, dy)\ S_s^y = \int_0^t D_s M_s^h\, ds.$$

The assumptions on h ensure that the predictable field S is uniformly bounded on finite time intervals, hence by Theorem 4.6.1(iii), M^h is a local martingale. But M^h is bounded on finite time intervals, hence it is a true martingale under $\mathbb{P}_{|x_0}$ and we can define the probability $\widetilde{\mathbb{P}}_{|x_0}$ on $\mathcal{F}^\mu$ by (7.129). By Proposition 7.9.2(b) it follows that X under $\widetilde{\mathbb{P}}_{|x_0}$ is a homogeneous PDMP, determined by some jump intensity function $\widetilde{q}(x)$ and some jump probabilities $\widetilde{r}(x, \cdot)$ (and of course with deterministic behaviour given by the original ϕ_t). In order to identify $\widetilde{q}$ and $\widetilde{r}$, we use the homogeneous version (7.127) of the likelihood $\mathbb{P}_{|x_0}$-martingale from Theorem 7.9.1.

Introduce the bounded function

$$H(x) = \int_G r(x, dy)\, h(y).$$

By differentiation between jumps

$$D_t \log \mathcal{L}^X_{t|x_0} = -\left(\widetilde{q}(X_t) - q(X_t)\right)$$

which must match

$$D_t \log M^h_t = -\frac{q(X_t)}{h(X_t)}\left(H(X_t) - h(X_t)\right)$$

as found from (7.135) when dividing by M^h_t. Consequently the choice

$$\widetilde{q}(x) = \frac{q(x)\,H(x)}{h(x)} \tag{7.136}$$

ensures that

$$D_t \mathcal{L}^X_{t|x_0} = D_t M^h_t, \tag{7.137}$$

and it then remains to match the jumps of $\mathcal{L}^X_{|x_0}$ and M^h. But

$$\begin{aligned}
\Delta \mathcal{L}^X_{t|x_0} &= \Delta \overline{N}_t \mathcal{L}^X_{t-|x_0}\left(\frac{\widetilde{q}(X_{t-})}{q(X_{t-})}\frac{d\widetilde{r}(X_{t-},\cdot)}{dr(X_{t-},\cdot)}(X_t) - 1\right)\\
&= \Delta \overline{N}_t \mathcal{L}^X_{t-|x_0}\left(\frac{H(X_{t-})}{h(X_{t-})}\frac{d\widetilde{r}(X_{t-},\cdot)}{dr(X_{t-},\cdot)}(X_t) - 1\right)
\end{aligned}$$

which we want to equal

$$\Delta M^h_t = \Delta \overline{N}_t S^{X_t}_t$$

with S given by (7.134). Since $\mathcal{L}^X_{0|x_0} = M^h_0 \equiv 1$, (7.137) gives that $\mathcal{L}^X_{|x_0}$ and M^h agree on $[0, T_1[$, hence the first jump for $\mathcal{L}^X_{|x_0}$ is the same as that for M^h provided

$$\frac{H(X_{t-})}{h(X_{t-})}\frac{d\widetilde{r}(X_{t-},\cdot)}{dr(X_{t-},\cdot)}(y) = \frac{h(y)}{h(X_{t-})},$$

i.e., if

$$\frac{d\widetilde{r}(x,\cdot)}{dr(x,\cdot)}(y) = \frac{h(y)}{H(x)}. \tag{7.138}$$

With this condition satisfied we have that $\mathcal{L}^X_{|x_0} \equiv M^h$ $\mathbb{P}_{|x_0}$-a.s. on $[0, T_1]$ and then deduce from (7.137) that $\mathcal{L}^X_{|x_0} \equiv M^h$ on $[T_1, T_2[$, and then, with (7.138) satisfied, that the identity extends to the closed interval $[T_1, T_2]$. Continuing, it is clear that under $\widetilde{\mathbb{P}}_{|x_0}$, X is a homogeneous PDMP with $\widetilde{q}$ and $\widetilde{r}$ given by (7.136) and (7.138) respectively.

The $\mathbb{P}_{|x_0}$-martingale M^h has a particularly simple form if h is *harmonic* for X under $\mathbb{P}_{|x_0}$, i.e., if in addition to the previous assumptions on h, it also holds that $\mathrm{A}h \equiv 0$. In that case

$$M^h_t = \frac{h(X_t)}{h(x_0)}.$$

Part II

Applications

Introduction to Part II

The main purpose of this second part of the book is to show how the theory for marked point processes and random counting measures developed in the earlier chapters may be used to discuss concrete examples of models relevant for applications in statistics and topics from applied probability: (i) the use of martingale estimators and the description of partially specified statistical models in survival analysis; (ii) the description using a PDMP of a branching process with age-dependent birth- and death-rates, and the proof of the basic branching property using martingales; (iii) in risk theory, the use of Itô's formula and optional sampling for computing the distribution of the time to ruin and ruin probabilities; (iv) how to use change of measure techniques to obtain a model for the dynamics of a soccer game, based on a model for the final outcome of the game. (v) in mathematical finance, a PDMP model for the development of the price of several risky assets together with a discussion of basic issues such as self-financing strategies, no arbitrage, the existence of equivalent martingale measures and completeness of markets for this particular model; (vi) how to turn standard non-Markovian models in queueing theory into PDMPs, a discussion of PDMP versions of Jackson networks and examples of how to treat stability of queueing models by finding invariant measures of relevant PDMPs.

8

The Basic Models from Survival Analysis

This chapter deals with the problem of estimating an unknown survival distribution based on observation of an iid sample subjected to right-censoring. Also treated is the estimation problem for the Cox regression model.

References. The book [2] by Andersen, Borgan, Gill and Keiding is indispensable. Hougaard [50] is an important reference for the analysis of multivariate survival data.

8.1 Independent survival times, right-censoring

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, i.e., the survivor function for the distribution of X_i is

$$\mathbb{P}(X_i > t) = \exp\left(-\int_0^t u_i(s)ds\right).$$

Define the counting process N^i by

$$N_t^i = 1_{(X_i \le t)}.$$

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

$$\begin{aligned}\lambda_s^i &= u_i(s)1_{(X_i \ge s)}\\ &= u_i(s)1_{(N_{s-}^i = 0)}.\end{aligned}$$

The aggregate (see Section 6.1) 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, (6.3) is satisfied).

By Theorem 6.1.1 (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,

$$R_{s-} = \sum_{i=1}^r 1_{(N_{s-}^i = 0)} = \sum_{i=1}^r (1 - N_{s-}^i) = r - \overline{N}_{s-}.$$

Notice in particular that $\overline{M} := \overline{N} - \overline{\Lambda}$ is an $\mathcal{F}_t^\mu$-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

$$\lambda_s^\circ = u(s)(r - N_{s-}^\circ)^+ \tag{8.1}$$

with the $^+$-sign added to give an intensity process defined on all of W, which is everywhere ≥ 0.

From the discussion in Section 7.2 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 $\overline{N}_0 \equiv 0$, total intensity function

$$q_t(k) = (r - k)\, u\,(t)$$

for $k \le r$ and jump probabilities

$$r_t\,(k, \cdot) = \varepsilon_{k+1}$$

for $k \le r - 1$. The transition probabilities are

$$p_{st}(j, k) = \mathbb{P}(\overline{N}_t = k | \overline{N}_s = j),$$

non-zero only if $0 \le j \le k \le r,\ 0 \le s \le t,\ s < t^\dagger$ (where $t^\dagger$ is the termination point for the distribution of the X_i) 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 $j \le k \le r,\ s \le t,\ s < t^\dagger$

$$\begin{aligned}
p_{st}(j,k) &= \mathbb{P}\left(X_i \in]0,t] \text{ for } k \text{ values of } i \,|X_i \in]0,s] \text{ for } j \text{ values of } i\right) \\
&= \frac{\binom{r}{j,k-j,r-k}(F(s))^j(F(t)-F(s))^{k-j}(\overline{F}(t))^{r-k}}{\binom{r}{j}(F(s))^j(\overline{F}(s))^{r-j}} \\
&= \binom{r-j}{k-j}\left(1-\frac{\overline{F}(t)}{\overline{F}(s)}\right)^{k-j}\left(\frac{\overline{F}(t)}{\overline{F}(s)}\right)^{r-k}.
\end{aligned}$$

Thus, conditionally on $\overline{N}_s = j$, $\overline{N}_t - j$ follows a binomial distribution $(r-j,\ 1-\overline{F}(t)/\overline{F}(s))$.

Consider now the statistical model when the hazard function u is unknown. Writing

$$\beta(t) = \int_0^t u(s)\,ds,$$

one then defines a martingale estimator of $\beta(t)$ by

$$\widehat{\beta}_t = \int_{]0,t]} \frac{1}{R_{s-}}\,d\overline{N}_s, \tag{8.2}$$

well defined since $R_{s-} \geq 1$ for each jump time s of $\overline{N}$ — these jump times are the X_i ordered according to size and

$$\widehat{\beta}_t = \sum_{k=1}^{\overline{N}_t} \frac{1}{r-k+1}. \tag{8.3}$$

Defining

$$S_s = \begin{cases} \frac{1}{R_{s-}} & \text{if } R_{s-} \geq 1, \\ 0 & \text{if } R_{s-} = 0, \end{cases}$$

S is an $\mathcal{F}_t^{\overline{N}}$-predictable process, and since both $S \geq 0$ and $\overline{N}$ are bounded, by Theorem 4.6.1(ii),

$$\overline{M}(S) = \overline{N}(S) - \overline{\Lambda}(S)$$

is an $\mathcal{F}_t^{\overline{N}}$-martingale. But $\overline{N}_t(S) = \widehat{\beta}_t$ and

$$\beta_t^* := \overline{\Lambda}_t(S) = \int_0^t u(s)1_{(R_{s-}\geq 1)}\,ds, \tag{8.4}$$

so $\widehat{\beta} - \beta^*$ is a martingale and in particular

$$\begin{aligned}
\mathbb{E}\widehat{\beta}_t = \mathbb{E}\beta_t^* &= \int_0^t u(s)\mathbb{P}\left(R_s \geq 1\right)\,ds \\
&= \int_0^t u(s)\left(1-(F(s))^r\right)\,ds
\end{aligned}$$

since $\mathbb{P}(R_s \geq 1) = 1 - \mathbb{P}(R_s = 0) = 1 - (F(s))^r$. Thus

$$\left(1 - (F(t))^r\right) \beta(t) \leq \mathbb{E}\widehat{\beta}_t \leq \beta(t)$$

and for $t < t^\dagger$ we see that asymptotically, as $r \to \infty$, $\lim \mathbb{E}\widehat{\beta}_t = \beta(t)$ with rapid convergence: for $t < t^\dagger$ and r large, $\widehat{\beta}_t$ is almost an unbiased estimator of $\beta(t)$.

To assess the quality of the estimator we compute the variance of $\widehat{\beta}_t - \beta_t^*$ (which is not quite the variance of $\widehat{\beta}_t$),

$$\mathbb{E}\left(\widehat{\beta}_t - \beta_t^*\right)^2 = \mathbb{E}\left(\overline{M}_t(S)\right)^2 = \mathbb{E}\left\langle \overline{M}(S)\right\rangle_t,$$

see p. 83 for the definition of the quadratic characteristic. But

$$\begin{aligned}\left\langle \overline{M}(S)\right\rangle_t &= \overline{\Lambda}_t(S^2) \\ &= \int_0^t u(s)\frac{1}{R_{s-}} 1_{(R_{s-}\geq 1)}\, ds\end{aligned}$$

(with the integrand 0 if $R_{s-} = 0$), a quantity that since $\frac{1}{r}R_{s-} \overset{\text{prob}}{\to} \overline{F}(s)$ for $r \to \infty$ by the law of large numbers may be shown to satisfy that

$$r\left\langle \overline{M}(S)\right\rangle_t \overset{\text{prob}}{\to} \int_0^t \frac{u(s)}{\overline{F}(s)}\, ds = \frac{F(t)}{\overline{F}(t)}$$

for $t < t^\dagger$. It may then further be argued, using a functional central limit theorem for martingales, that for $t < t^\dagger$ fixed and $r \to \infty$, the sequence

$$\left(\sqrt{r}\left(\widehat{\beta}_s - \beta_s^*\right)_{0\leq s\leq t}\right)_{r\geq 1}$$

of martingales on $[0, t]$ and the sequence $\left(\sqrt{r}\left(\widehat{\beta}_s - \beta(s)\right)_{0\leq s\leq t}\right)$ of processes both converge in distribution to the Gaussian martingale $(W_s)_{0\leq s\leq t}$, where

$$W_s = B_{\sigma^2(s)} \tag{8.5}$$

with B a standard Brownian motion and the variance function (and deterministic time change) σ^2 given by $\sigma^2(s) = F(s)/\overline{F}(s)$. (For the convergence in distribution, the processes are considered as random variables with values in the Skorohod space $D[0, t]$ of real-valued, cadlag paths defined on $[0, t]$, with convergence in distribution meaning weak convergence of the distributions when $D[0, t]$ is equipped with the Skorohod topology).

Using that $\overline{M}\left(S^2\right)$ is a martingale, it is natural to 'estimate' $\left\langle \overline{M}(S)\right\rangle_s = \overline{\Lambda}_s\left(S^2\right)$ by $\overline{N}_s\left(S^2\right)$, and therefore estimate $\sigma^2(s)$ by

$$\widehat{\sigma}^2(s) = r\overline{N}_s\left(S^2\right) = r\int_{]0,s]} \frac{1}{R_{u-}^2}\, d\overline{N}_u, \tag{8.6}$$

an estimator that can be shown to be uniformly consistent on $[0, t]$,

$$\sup_{s:s\leq t}\left|\widehat{\sigma}^2(s) - \sigma^2(s)\right| \xrightarrow{\text{prob}} 0$$

for any $t < t^\dagger$.

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,

$$N_t^i = 1_{(X_i\leq t\wedge V_i)}.$$

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 σ-algebra generated by the failures and censorings observed to occur in $[0, t]$. Formally,

$$\begin{aligned}\mathcal{F}_t &= \sigma(S_i 1_{(S_i\leq s)},\ \delta_i 1_{(S_i\leq s)})_{1\leq i\leq r, s\leq t}\\ &= \sigma(S_i 1_{(S_i\leq t)},\ \delta_i 1_{(S_i\leq t)})_{1\leq i\leq r}.\end{aligned}$$

(The second equality is easy to show: if $S_i \leq t$ and the values of S_i and δ_i are known, so are the values of $S_i 1_{(S_i\leq s)}$ and $\delta_i 1_{(S_i\leq s)}$ for any $s \leq t$; and if $S_i > t$ also $S_i > s$ for $s \leq t$. Note that $(S_i > t) \in \mathcal{F}_t$: since $S_i > 0$, $S_i 1_{(S_i\leq t)} = 0$ iff $S_i > t$). What is usually assumed about a model for right-censored survival data is that for every i, M^i is an $\mathcal{F}_t$-martingale, where

$$M_t^i = N_t^i - \int_0^t \lambda_s^i ds, \quad \lambda_s^i = u^i(s) I_s^i, \tag{8.7}$$

that $(M^i)^2 - \int_0^\cdot \lambda^i$ is a martingale, and finally that $M^i M^j$ is a martingale for $i \neq j$. Here I^i in (8.7) is the $\mathcal{F}_t$-predictable indicator

$$I_s^i = 1_{(S_i\geq s)},$$

which is one when item i is still at *risk at time* s, i.e., just before s item i has as yet neither been censored nor observed to fail.

Note that with the joint distribution of all S_i and δ_i described through an MPP as indicated below, the assumptions about the martingale structure of $(M^i)^2$ and $M^i M^j$ presented after (8.7) are direct consequences of (8.7) and Proposition 4.5.3 (b), so are not new conditions.

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 an 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

$$E := \{(i, A) : 1 \le i \le r,\ A \subset \{1, \ldots, r\} \backslash \{i\}\} \cup \{(c, A) : \emptyset \ne A \subset \{1, \ldots, r\}\},$$

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 (8.7) 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 (8.7), much less of course the joint distribution of all (X_i, V_i). The message is that with (8.7) 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 a full specification of the compensators Λ^y for all N^y, $y \in E$, must respect the special structure of the N^y, i.e., that each N^y has at most one jump and that jump is possible at time s only if all items listed in y are at risk at time s. Thus, in case $\Lambda_t^y = \int_0^t \lambda_s^y \, ds$ with an $\mathcal{F}_t$-predictable intensity process λ^y (which need not exist in general), λ^y must satisfy

$$\lambda_s^{(i,A)} = \lambda_s^{(i,A)} 1_{(S_j \ge s, j \in A \cup \{i\})}, \quad \lambda_s^{(c,A)} = \lambda_s^{(c,A)} 1_{(S_j \ge s, j \in A)}.$$

Consider now the special case where all $u_i = u$. Even though (8.7) is only a partial specification of the compensator structure, it is good enough to enable the estimation procedure from the iid case without censoring to go through: define $\beta(t) = \int_0^t u$ as before and still use (8.2) to give

$$\widehat{\beta}_t = \int_{]0,t]} \frac{1}{R_{s-}} \, d\overline{N}_s, \tag{8.8}$$

the celebrated *Nelson–Aalen estimator*, with now $\overline{N} = \sum_i N^i$ and

$$R_{s-} = \sum_{i=1}^{r} I_s^i = \sum_{i=1}^{r} 1_{(S_i \geq s)},$$

the number of items at risk just before s, (so (8.3) is no longer valid). With β_t^* still given by (8.4), $\widehat{\beta} - \beta^*$ is an $\mathcal{F}_t$-martingale, and if $\frac{1}{r} R_s \overset{\text{prob}}{\to} \rho(s)$ for all s with ρ some continuous function (necessarily decreasing and $\leq \overline{F}(s)$), the convergence in distribution of $\left(\sqrt{r}\left(\widehat{\beta}_s - \beta_s^*\right)\right)_{0\leq s\leq t}$ to a Gaussian martingale $(W_s)_{0\leq s\leq t}$ (see (8.5)) still holds for $t < \inf\{s : \rho(s) = 0\}$, only now the variance function for W is

$$\sigma^2(s) = \int_0^s u(v)\frac{1}{\rho(v)}\, dv.$$

Finally, a uniformly consistent estimator of $\sigma^2(s)$ on $[0, t]$ is still given by (8.6).

Writing $\widehat{\nu}\,(]0, t]) = \widehat{\beta}_t$, it is seen that $\widehat{\nu}$ is the hazard measure for a discrete distribution on $\overline{\mathbb{R}}_+$, concentrated on the timepoints $S_i < \infty$ for which $\delta_i = 1$ and possibly also with a mass at ∞. The corresponding distribution has survivor function

$$\widehat{\overline{F}}(t) = \prod_{0<s\leq t}\left(1 - \frac{\Delta \overline{N}_s}{R_{s-}}\right),$$

the classical *Kaplan–Meier estimator* for the survivor function

$$\overline{F}(t) = \exp\left(-\int_0^t u\,(s)\; ds\right),$$

which is often used rather than the natural alternative $\exp\left(-\widehat{\beta}_t\right)$. Note that $\widehat{\overline{F}}$ has finite termination point iff $S_{i_0} = \max_i S_i < \infty$ and $\delta_{i_0} = 1$. Note also that in the simple model without censoring (all $V_i \equiv \infty$), $\widehat{\overline{F}}$ is the survivor function for the empirical distribution on the $S_i = X_i$: each finite X_i is given mass $1/r$.

We shall now give some examples of censoring models, where (8.7) holds. It is assumed always that the X_i are independent with hazard functions u_i.

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,

$$\mathbb{P}(T_1^i > t) = \overline{F}_i(t \wedge v_i),$$

$\overline{F}_i$ 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

$$\lambda_t^i = u_i(t)1_{[0,v_i]}(t)1_{(N_{t-}^i=0)} = u_i(t)I_t^i,$$

i.e., (8.7) 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 hazard function u_i^c. (It is not required that the distribution of V_i have a density, but it is a convenient assumption for the calculations below). Identify any given observable pair (S_i, δ_i) with an MPP described by the two counting processes corresponding to the two marks i and c, i,

$$N_t^i = 1_{(X_i \le t \wedge V_i)}, \quad N_t^{c,i} = 1_{(V_i \le t, V_i < X_i)},$$

(which makes sense since $\mathbb{P}(X_i = V_i < \infty) = 0$). 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 for the mark i,

$$\rho_i(t) := \pi_t^{i,(0)}(\{i\}) = 1 - \pi_t^{i,(0)}(\{c,i\}) = \mathbb{P}(X_i \le V_i | S_i = t).$$

But, cf. Proposition 4.1.3(ii)

$$u^{i,(0)}(t) = u_i(t) + u_i^c(t), \tag{8.9}$$

while ρ_i is determined by the condition

$$\begin{aligned}\int_B \rho_i(t)\,\mathbb{P}(S_i \in dt) &= \int_B \rho_i(t)\left(u_i(t) + u_i^c(t)\right)\exp\left(-\int_0^t \left(u_i + u_i^c\right)\right)dt \\ &= \mathbb{P}(X_i \le V_i,\ S_i \in B) \quad (B \in \mathcal{B}_0).\end{aligned}$$

Using the expressions for the density for X_i and the survivor function for V_i respectively, it follows, conditioning on the value of X_i, that

$$\begin{aligned}\mathbb{P}(X_i \le V_i,\ S_i \in B) &= \mathbb{P}(X_i \le V_i,\ X_i \in B) \\ &= \int_B u_i(t)\exp\left(-\int_0^t u_i\right)\exp\left(-\int_0^t u_i^c\right)dt,\end{aligned}$$

and consequently

$$\rho_i = \frac{u_i}{u_i + u_i^c}.$$

Comparing with (8.9) 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 (8.7). Because of the independence of the pairs (X_i, V_i), using Theorem 6.1.1 (a) it follows that (8.7) 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 (8.7) 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 (8.7) 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

$$\int_0^t \widetilde{\lambda}_s^i \, ds, \qquad \widetilde{\lambda}_s^i = u_i(s) I_i(s)$$

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 (8.7) is satisfied.

Finally we give an example of a model for right-censoring for which (8.7) holds but where the V_i are not independent of the 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). It is easy to see that (8.7) 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 $\{1, \dots, r\}$ 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

$$N_t^i = 1_{(X_i \le t, X_i \le X_{(m)})},$$

corresponding to an RCM $\mu = \sum_{n: T_n < \infty} \varepsilon_{(T_n, Y_n)}$ with mark space $\{1, \dots, r\}$. Clearly $(T_k, Y_k) = (\widetilde{T}_k, \widetilde{Y}_k)$ for $k \le m$, $T_{m+1} = T_{m+2} = \cdots = \infty$. So the Markov kernels determining μ agree with those of $\widetilde{\mu}$ up to and including the time of the mth jump. We know the compensators for $\widetilde{\mu}$ and deduce the desired (8.7) structure for those of μ.

We leave this discussion of censoring models by quoting an intriguing question: suppose only that the joint distribution of the (S_i, δ_i) satisfies the martingale property (8.7) for all i, but does 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? The answer is indeed ‘yes’, but the proof is not a triviality.

8.2 The Cox regression model

It was always assumed above that the failure times X_i were independent. One of the most versatile of all models used in survival analysis, where no independence assumptions are made, 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 $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ and let for $1 \leq i \leq r$, ζ^i be an $\mathbb{R}^p$-valued $\mathcal{F}_t$-predictable process (the covariate process for item 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,

$$N_t^i = 1_{(X_i \leq t \wedge V_i)}$$

is $\mathcal{F}_t$-adapted, and that the $\{0, 1\}$-valued process I^i,

$$I_t^i = 1_{(S_i \geq t)},$$

where $S_i = X_i \wedge V_i$, 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,

$$\lambda_t^i = u(t) e^{\gamma^T \zeta_t^i} I_t^i, \tag{8.10}$$

where u is the hazard function for some distribution on $\overline{\mathbb{R}}_+$ and $\gamma^T = (\gamma_1, \ldots, \gamma_p)$ is a row vector of regression parameters. The 'model' arises by allowing the *baseline hazard* u and the γ-parameters to vary.

Note that if $\gamma = 0$ 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

$$u_i(t) = u(t) e^{\gamma^T \varsigma^i(t)}. \tag{8.11}$$

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 4.4.2),

$$u(t+) e^{\gamma^T \zeta_{t+}^i} = \lim_{h \downarrow 0} \frac{1}{h} \mathbb{P}(t < X_i \leq t + h | \mathcal{F}_t) \quad \text{on the set } (S_i > t)$$

describing the imminent risk of failure for item i.

With only the intensities (8.10) 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 RCMs, a full specification of the model could be given involving not only the failure intensities (8.10) 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.

The Cox regression model has proved immensely useful in survival analysis, primarily because even though the model is only partailly specified, it is still possible to estimate in an efficient manner both the baseline hazard and the regression parameters, and thereby also to analyse in practice the effect on survival of any of the covariates.

In order to estimate the integrated baseline hazard,

$$\beta(t) = \int_0^t u(s)\, ds,$$

one imitates the idea leading to the Nelson–Aalen estimator (8.8) and defines

$$\widehat{\beta}_{\gamma,t} = \int_0^t \frac{1}{\sum_{i \in \mathcal{R}_s} e^{\gamma^T \zeta_s^i}}\, d\overline{N}_s$$

with $\overline{N} = \sum_i N^i$ and

$$\mathcal{R}_s = \{i : 1 \leq i \leq r, S_i \geq s\}$$

the set of items at risk immediately before time s. Then, with u the true baseline hazard and γ the true value of the vector of regression parameters, $\widehat{\beta}_\gamma - \beta^*$ is an $\mathcal{F}_t$-martingale, where as before

$$\beta_t^* = \int_0^t u(s)\, 1_{(\mathcal{R}_s \neq \emptyset)}\, ds.$$

The estimator $\widehat{\beta}_\gamma$ depends on the unknown regression parameter, so ultimately $\beta(t)$ is estimated by $\widehat{\beta}_{\widehat{\gamma},t}$, where $\widehat{\gamma}$ is the estimator of γ that we shall now discuss.

Let C_t for $t \geq 0$ denote the *partial likelihood* process

$$C_t(\gamma) = \prod_{n=1}^{\overline{N}_t} \frac{e^{\gamma^T \zeta_{T_n}^{J_n}}}{\sum_{i \in \mathcal{R}_{T_n}} e^{\gamma^T \zeta_{T_n}^i}}, \tag{8.12}$$

defined for arbitrary $\gamma \in \mathbb{R}^p$. Here the T_n are the jump times for $\overline{N}$ ordered according to size, $T_1 < T_2 < \cdots$ (at most r), i.e., the T_n are the ordered S_i-values for which $X_i \leq V_i$, while $J_n = j$ if it is item j that is observed to fail at time T_n.

Suppose e.g., that everything (i.e., all the N^i, I^i and ζ^i) is observed on the interval $[0, t_0]$ for some $t_0 > 0$, the estimator used for γ is then $\widehat{\gamma} = \widehat{\gamma}_{t_0}$, the value of γ that maximizes $\gamma \mapsto C_{t_0}(\gamma)$, and $\beta(t)$ for $t \leq t_0$ is then estimated by $\widehat{\beta}_{\widehat{\gamma}_{t_0},t}$.

We shall give some comments on the structure of the partial likelihood (8.12). Since always $J_n \in \mathcal{R}_{T_n}$, each factor in the product is > 0 and ≤ 1. Furthermore it may be shown that $\gamma \mapsto \log C_t(\gamma, \omega)$ (for a given value of $t > 0$ and $\omega \in \Omega$ — corresponding to a concrete set of observations on $[0, t]$) is concave and strictly concave iff the contrast covariate vectors

$$\rho_n^i(\omega) = \zeta_{T_n}^{J_n}(\omega) - \zeta_{T_n}^i(\omega)$$

for $1 \le n \le \overline{N}_t(\omega)$, $i \in \mathcal{R}_{T_n}(\omega)$ span all of $\mathbb{R}^p$. And $\log C_t(\gamma, \omega)$ attains its maximal value at a uniquely determined point iff it is impossible to find a row vector $v^T \neq 0$ such that $v^T \rho_n^i(\omega) \ge 0$ for all choices of n and i as above.

With the interval of observations $[0, t_0]$ as before, it may be shown under regularity conditions that as $r \to \infty$, $\sqrt{r}\left(\widehat{\gamma}_{t_0} - \gamma\right)$ converges in distribution to a p-dimensional Gaussian distribution with mean vector 0 and a covariance matrix that itself can be estimated consistently. A more refined analysis even leads to results on the joint asymptotic behaviour of $\widehat{\gamma} = \widehat{\gamma}_{t_0}$ and $\left(\widehat{\beta}_{\widehat{\gamma},t}\right)_{0 \le t \le t_0}$. To discuss the asymptotics in any detail is beyond the scope of this book, but martingale central limit theorems are essential and a key observation for the study of the asymptotics of $\widehat{\gamma}_{t_0}$ is then the following:

Exercise 8.2.1 With γ the true value of the vector of regression parameters and assuming e.g., that all the covariate processes ζ^i are bounded, show that the *partial score process* is an $\mathcal{F}_t$-martingale, i.e., show that

$$D_{\gamma_k} \log C_t(\gamma) \quad (0 \le t \le t_0)$$

is an $\mathcal{F}_t$-martingale for $1 \le k \le p$. (Hint: the task is to compute the derivatives $D_{\gamma_k} \log C_t(\gamma)$ and then rewrite them in terms of stochastic integrals of predictable integrands with respect to the basic martingales

$$M_t^i = N_t^i - \int_0^t \lambda_s^i \, ds).$$

We shall conclude with a brief discussion of the terminology 'partial likelihood' and how it relates to the theory of MPPs and RCMs. Suppose there is no censoring and that the $\zeta_t^i = \varsigma^i(t)$ are given functions of t. Then the X_i are independent with hazard functions u_i given by (8.11), we have

$$N_t^i = 1_{(X_i \le t)}$$

and with $(\mathcal{F}_t)$ the filtration generated by the N^i for $1 \le i \le r$, find that the $\mathcal{F}_t$-compensator for N^i is

$$\Lambda_t^i = \int_0^t u_i(s) \, I_s^i \, ds$$

with $I_s^i = 1_{(X_i \ge s)}$. But then one finds that the distribution of the aggregate μ of $\left(N^1, \dots, N^r\right)$ is determined by the Markov kernels

$$\overline{P}_{z_n}^{(n)}(t) = \exp\left(-\int_{t_n}^t \sum_{i \in \mathcal{R}(z_n)} u(s)\, e^{\gamma^T \varsigma^i(s)} \, ds\right)$$

for $t \ge t_n$ with $\mathcal{R}(z_n)$ the set of items at risk after the time t_n of the nth failure, i.e., the items remaining when all items listed in $z_n = (t_1, \dots, t_n; j_1, \dots, j_n)$ have been removed, and

$$\pi^{(n)}_{z_n,t}\left(\{i_0\}\right) = \frac{e^{\gamma^T \varsigma^{i_0}(t)}}{\sum_{i\in\mathcal{R}(z_n)} e^{\gamma^T \varsigma^{i}(t)}} 1_{(i_0\in\mathcal{R}(z_n))}, \tag{8.13}$$

cf. (4.47) and (4.48). Comparing this with the expression for the likelihood process $\mathcal{L}_t$ from Theorem 5.1.1(b) (using for reference measure the case $\gamma = 0$, $u(t) \equiv 1$) it emerges that $C_t(\gamma)$ is, apart from a factor not depending on γ, the contribution to $\mathcal{L}_t$ that consists of the product of the 'π-derivatives',

$$C_t(\gamma) \propto \prod_{n=1}^{\overline{N}_t} \frac{d\pi^{(n-1)}_{Z_{n-1},T_n}}{d\pi^{0,(n-1)}_{Z_{n-1},T_n}}(J_n)$$

with $Z_n = (T_1, \ldots, T_n; J_1, \ldots, J_n)$, the $\pi^{(n)}$ as in (8.13) and $\pi^{0,(n)}$ referring to the reference measure, i.e.

$$\pi^{0,(n)}_{z_n,t}\left(\{i_0\}\right) = \frac{1}{r-n} 1_{(i_0\in\mathcal{R}(z_n))}.$$

In a more general setup with random censorings and covariates, the nth factor in $C_t(\gamma)$ may be interpreted as the conditional probability that 'it is item J_n that fails at time T_n, the time of the n'th observed failure' given the past up to time T_n. In a complete MPP description of the model with censorings and piecewise deterministic ζ^i so that jumps corresponding to observed censoring times and jumps for the ζ^i are included, such a probability would arise from a $\pi^{(l)}_{z_l,t}$-probability (the $\pi^{(l)}$ and z_l now referring to the full model), by conditioning furthermore on the event that the jump that is to occur at time t should correspond to a failure and then asking for the probability that it is a given item that fails. As we have just seen, such factors appear naturally in the expression for the likelihood process, so the terminology 'partial' is appropriate in the sense that $C_t(\gamma)$ picks out parts of the total likelihood $\mathcal{L}_t$ involving the π-kernels and referring to special types of jumps, i.e., the occurrence of failures.

9

Branching, Ruin, Soccer

In this chapter three quite different models are presented: a branching process for the evolution of a population with age-dependent birth and death intensities; a classical model from risk theory with a discussion of the problem of calculating the probability of ruin and the distribution of the time to ruin; a model for how a soccer game develops over time, using a simple multiplicative Poisson model as starting point.

References. Branching processes: Jagers [68] and the collection [5] of papers edited by Athreya and Jagers deal with rather more advanced models than the one presented here. Ruin probabilities: important books for further reading include Rolski et al [105] and Asmussen [4].

9.1 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

$x + t = \{a_1 + t, \ldots, a_k + t\}$. (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 — see Exercise 9.1.1 below).

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 $\{a_1, \ldots, a_k\} \cup 0$ corresponding to a birth or $\{a_1, \ldots, a_k\} \backslash a_i$ for some $1 \le i \le k$ 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,

$$X_t = \phi_{t - T_{\langle t \rangle}} \left(Y_{\langle t \rangle} \right)$$

with T_n the time of the nth jump, Y_n the state reached by that jump, and the deterministic behaviour given by

$$\phi_t(x) = x + t$$

if $|x| \ge 1$ and $\phi_t(\emptyset) = \emptyset$. The Markov kernels determining the distribution of the MPP $((T_n), (Y_n))$ are given by, for $t \ge t_n$, with $y_n = \left\{ a_1^{(n)}, \ldots, a_{k_n}^{(n)} \right\}$,

$$\overline{P}^{(n)}_{z_n|x_0}(t) = \exp\left(-\int_0^{t-t_n} [\beta + \delta]_{y_n}(s)\, ds \right),$$

$$\pi_{z_n,t|x_0}\left((y_n + t - t_n) \cup 0 \right) = \frac{\sum_{i=1}^{k_n} \beta \left(a_i^{(n)} + t - t_n \right)}{[\beta + \delta]_{y_n}(t - t_n)}$$

$$\pi_{z_n,t|x_0}\left((y_n + t - t_n) \backslash \left(a_i^{(n)} + t - t_n \right) \right) = \frac{\delta \left(a_i^{(n)} + t - t_n \right)}{[\beta + \delta]_{y_n}(t - t_n)} \quad (1 \le i \le k_n),$$

using the notation

$$[\beta + \delta]_{y_n}(s) := \sum_{i=1}^{k_n} \left(\beta(a_i^{(n)} + s) + \delta(a_i^{(n)} + s) \right).$$

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 X defined above does not explode — this is related to the fact that the birth process from Example 3.1.4 with $\lambda_n = \beta n$ for some $\beta > 0$ does not explode; in particular, if the age-dependent birth rate $\beta(a)$ is bounded as a function of a, it is easy to show that X does not explode using Corollary 4.4.4.

Clearly the ϕ_t satsify (7.28) and it is then immediately verified from Theorem 7.3.1 (b) that X is time-homogeneous Markov with

$$q(x) = \sum_{i=1}^{k} (\beta(a_i) + \delta(a_i)),$$

$$r(x, x \cup 0) = \frac{\sum_{i=1}^{k} \beta(a_i)}{\sum_{i=1}^{k} (\beta(a_i) + \delta(a_i))},$$

$$r\left(x, x \backslash a_{i_0}\right) = \frac{\delta\left(a_{i_0}\right)}{\sum_{i=1}^{k} (\beta(a_i) + \delta(a_i))} \quad (1 \leq i_0 \leq k),$$

where $x = \{a_1, \ldots, a_k\}$, and where only the exhaustive non-zero point probabilities for $r(x, \cdot)$ have been specified. The corresponding predictable intensities, $\lambda_t(C)$ (for special choices of C but enough to determine all $\lambda_t(C)$), are given by

$$\begin{aligned} \overline{\lambda}_t &= \sum_{a \in X_{t-}} (\beta(a) + \delta(a)), \\ \lambda_t(X_{t-} \cup 0) &= \sum_{a \in X_{t-}} \beta(a), \\ \lambda(X_{t-} \backslash a_0) &= \sum_{a \in X_{t-}} \delta(a_0)\, 1_{(a = a_0 \in X_{t-})}, \end{aligned} \tag{9.1}$$

justifying the interpretation of β, δ as age-dependent birth and death intensities. (Note that if all ages $a \in X_{t-}$ are different,

$$\lambda(X_{t-} \backslash a_0) = \delta(a_0)$$

for $a_0 \in X_{t-}$).

A particularly simple case of the model above is obtained when both β and δ are constants. In this 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, for $k \geq 1$,

$$q_{k,k+1} = k\beta, \quad q_{k,k-1} = k\delta.$$

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 the same age-dependent birth- and death intensities β and δ respectively such that $X_0^i \equiv \left\{a_i^{(0)}\right\}$. Then $\widetilde{X} := \bigcup_{i=1}^{k_0} X^i$ has the same distribution as X. To see this, let μ^i be the RCM describing the jump times T_n^i and jumps $Y_n^i = X_{T_n^i}^i$ for X^i, and let μ be the aggregate given by (6.1) which is well defined since the T_n^i have continuous distributions. Assuming (after discarding a null set if necessary) that for all ω, $T_k^i(\omega) \neq T_n^j(\omega)$ whenever $(i, k) \neq (j, n)$ (cf. (6.5)), from Theorem 6.1.1 it follows that μ has $\mathcal{F}_t^\mu$-compensating measure L with $\Lambda_t(A) = L([0,t] \times A)$ for $A = \bigcup_{i=1}^r \{i\} \times A^i$ given by

$$\Lambda_t(A) = \int_0^t \lambda_s(A)\,ds$$

with

$$\lambda_t(A) = \sum_{i=1}^r \sum_{a^i \in X_{t-}^i} \left(\beta\left(a^i\right) 1_{(X_{t-}^i \cup 0 \in A^i)} + \delta\left(a^i\right) 1_{(X_{t-}^i \setminus a^i \in A^i)}\right). \tag{9.2}$$

Now let $\widetilde{\mu} = \sum_{\widetilde{T}_n < \infty} \varepsilon_{(\widetilde{T}_n, \widetilde{Y}_n)}$ be the RCM determined by the jump times $\widetilde{T}_n$ and jumps $\widetilde{Y}_n = \widetilde{X}_{\widetilde{T}_n}$ for $\widetilde{X}$. The $\widetilde{T}_n$ are the same as the jump times T_n for the aggregate μ, and since obviously $\widetilde{X}$ can only jump by either a birth or a death occurring for precisely one of the populations X^i, the $\mathcal{F}_t^\mu$-predictable intensity processes for $\widetilde{\mu}$ are given by the total intensity

$$\overline{\widetilde{\lambda}}_t = \sum_{i=1}^r \sum_{a^i \in X_{t-}^i} (\beta + \delta)\left(a^i\right) = \sum_{\widetilde{a} \in \widetilde{X}_{t-}} (\beta + \delta)(\widetilde{a}),$$

the birth intensity (take $A = \widetilde{X}_{t-} \cup 0$ in (9.2))

$$\widetilde{\lambda}_t\left(\widetilde{X}_{t-} \cup 0\right) = \sum_{i=1}^r \sum_{a^i \in X_{t-}^i} \beta\left(a^i\right) = \sum_{\widetilde{a} \in \widetilde{X}_{t-}} \beta(\widetilde{a})$$

and the death intensity, for $\widetilde{a}_0 \in \widetilde{X}_{t-}$, (take $A = \widetilde{X}_{t-} \setminus \widetilde{a}_0$ in (9.2))

$$\widetilde{\lambda}_t\left(\widetilde{X}_{t-} \setminus \widetilde{a}_0\right) = \sum_{i=1}^r \delta(\widetilde{a}_0)\, 1_{(\widetilde{a}_0 \in X_{t-}^i)} = \sum_{\widetilde{a} \in \widetilde{X}_{t-}} \delta(\widetilde{a}_0)\, 1_{(\widetilde{a} = \widetilde{a}_0 \in \widetilde{X}_{t-})}.$$

Thus the $\mathcal{F}_t^\mu$-predictable intensity processes for $\widetilde{\mu}$ are $\mathcal{F}_t^{\widetilde{\mu}}$-predictable; hence by Theorem 4.8.1 they determine the distribution of $\widetilde{\mu}$. Since the $\widetilde{\lambda}_t(C)$ agree in form exactly with the intensities $\lambda_t(C)$ for X, cf. the three equations (9.1), and obviously the deterministic behaviour between jumps of $\widetilde{X}$ is the same as that of X while also $\widetilde{X}_0 \equiv X_0$, it follows that $\widetilde{X}$ and X have the same distribution.

Exercise 9.1.1 Consider a model for a population process X^*, where each X_t^* not only reveals the current ages of the individuals alive at time t, but also for each individual, the ages at which she gave birth. Thus a state x^* for X^* is an object of the form $\{a_1^*, \ldots, a_k^*\}$ with each $a_i^* = \left(a_i; b_1^{(i)}, \cdots, b_m^{(i)}\right)$, a_i denoting the age of i, and $0 < b_1^{(i)} < \cdots < b_m^{(i)} \le a_i$ the age prior to her present age at which i gave birth (with $m = 0$ possible). To get back to the process X discussed above, it is natural to consider a function of X^*, $\left(g(X_t^*)\right)_{t \ge 0}$, where

$$g\left(x^*\right) = \{a_1, \ldots, a_k\}.$$

Give a construction of X^* that makes it a homogeneous PDMP and which fits with the construction of X above in the sense that the process $g(X^*)$ has the same distribution as X when the initial states $X_0^* \equiv x_0^*$, $X_0 \equiv x_0$ are fixed and $g\left(x_0^*\right) = x_0$.

A quantity that is easily calculated (although formally the construction from Exercise 9.1.1 is needed) 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 $N_t^d = 1_{(\zeta \leq t)}$. then $\left(N^b, N^d\right)$ has $\mathcal{F}_t^{\left(N^b, N^d\right)}$-compensator $\left(\Lambda^b, \Lambda^d\right)$ where

$$\Lambda_t^b = \int_0^t 1_{(N_{s-}^d = 0)} \beta(s)\, ds, \quad \Lambda_t^d = \int_0^t 1_{(N_{s-}^d = 0)} \delta(s)\, ds.$$

Since Λ^d is $\mathcal{F}_t^{N^d}$-predictable we recognize that

$$\mathbb{P}(\zeta > t) = \exp\left(-\int_0^t \delta(s)\, ds\right);$$

it then follows that

$$\mathbb{E}N_t^b = \mathbb{E}\Lambda_t^b = \int_0^t \exp\left(-\int_0^s \delta(u)\, du\right) \beta(s)\, ds,$$

and defining $N_\infty^b = \lim_{t\to\infty} N_t^b$ that

$$\gamma = \mathbb{E}N_\infty^b = \int_0^\infty \exp\left(-\int_0^s \delta(u)\, du\right) \beta(s)\, ds.$$

Assuming that $\int_0^\infty \delta = \infty$ (i.e., that individuals have finite lifetimes), one would expect the population to become extinct almost surely (no matter what the value of X_0), i.e., the *extinction probability* is 1 if $\gamma \leq 1$, 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. These facts follow directly from the classical theory of Galton–Watson processes in discrete time from which one knows that only finitely many individuals are ever born (the population becomes extinct) iff the expected number of offspring from each individual is ≤ 1. But for the continuous time model considered here, the total number of individuals ever born is exactly that of a Galton–Watson process with offspring distribution equal to that of N_∞^b. Note that for the linear birth- and death process $\gamma \gtreqless 1$ according as $\beta \gtreqless \delta$.

9.2 An example involving ruin probabilities

If X is a homogeneous PDMP, from Itô's formula Theorem 7.6.1 (b) we know that if the function f solves the equation $\mathcal{A}f \equiv 0$ (see (7.63)) with $\mathcal{A}$ the space-time generator given by (7.58), then if f and $\mathcal{A}f$ are bounded on finite intervals of time, the process $(f(t, X_t))_{t\geq 0}$ is a martingale. In general there is not much hope of solving an equation like (7.63) 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 $\mathbb{P}(U_n = 0) = 0$, independent of N. Finally, let $x_0 > 0$, let $\alpha \in \mathbb{R}$ and define

$$X_t = x_0 + \alpha t + \sum_{n=1}^{N_t} U_n, \tag{9.3}$$

i.e., X is a compound Poisson process (p.135) with a linear drift $t \mapsto \alpha t$ added on. In particular X has stationary independent increments, cf. Section 6.2, 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

$$r(y, \cdot) = \text{the distribution of } y + U_1.$$

We define the *time to ruin* as

$$\mathrm{T_r} = \inf\{t : X_t \leq 0\}.$$

The problem is then to find the *ruin probability* $p_\mathrm{r} = \mathbb{P}(\mathrm{T_r} < \infty)$ and, if possible, the distribution of $\mathrm{T_r}$.

We shall focus on two different setups: (i) the *simple ruin problem* corresponding to the case where $\mathbb{P}(U_1 > 0) = 1,\ \alpha < 0$, where X decreases linearly between the strictly positive jumps, (ii) the *difficult ruin problem* where $\mathbb{P}(U_1 < 0) = 1,\ \alpha > 0$ with X increasing linearly between the strictly negative jumps.

It is clear that for Problem (i),

$$X_{\mathrm{T_r}} = 0 \quad \text{on} \quad (\mathrm{T_r} < \infty), \tag{9.4}$$

and it is this property that makes (i) simple, while for Problem (ii) it may well happen that

$$X_{\mathrm{T_r}} < 0,$$

and it is this possibility of *undershoot*, which makes (ii) difficult.

There are simple necessary and sufficient conditions for the case when $p_\mathrm{r} = 1$. A preliminary argument leading to these conditions is as follows: in both Problem (i) and (ii), $\xi = \mathbb{E}U_1$ is well defined, but may be infinite. Now with T_n the time of the nth jump for N,

$$X_{T_n} = x_0 + \alpha T_n + \sum_{k=1}^{n} U_k,$$

and so, by the strong law of large numbers

$$\frac{1}{n} X_{T_n} \overset{\text{a.s.}}{\to} \frac{\alpha}{\lambda} + \xi. \tag{9.5}$$

Consequently, $X_{T_n} \to -\infty$ if $\frac{\alpha}{\lambda} + \xi < 0$, implying that $p_{\mathrm{r}} = 1$, while$X_{T_n} \to \infty$ if $\frac{\alpha}{\lambda} + \xi > 0$, suggesting that $p_{\mathrm{r}} < 1$ if $\frac{\alpha}{\lambda} + \xi > 0$. In fact, and as will be shown below, for both Problem (i) and (ii) it holds that

$$p_{\mathrm{r}} = 1 \quad \text{iff} \quad \frac{\alpha}{\lambda} + \xi \le 0. \tag{9.6}$$

(For (i) this is possible only if $\xi < \infty$. For (ii) (9.6) is satisfied if in particular $\xi = -\infty$).

Let P denote the distribution of the U_n and introduce

$$\psi(\theta) = \mathbb{E} e^{-\theta U_1} = \int e^{-\theta u}\, \mathrm{P}(du).$$

In the case of Problem (i), this is finite if $\theta \ge 0$, and ψ is the Laplace transform for U_1. For Problem (ii), ψ is finite if $\theta \le 0$ and $\psi(-\theta)$ then defines the Laplace transform for $-U_1$. (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 $\theta \ge 0$ and when treating Problem (ii), assume $\theta \le 0$.

By a simple calculation

$$\begin{aligned} \mathbb{E} \exp(-\theta(X_t - x_0)) &= \sum_{n=0}^{\infty} \mathbb{P}(\overline{N}_t = n) e^{-\theta\alpha t} \mathbb{E} \exp\left(-\theta \textstyle\sum_1^n U_k\right) \\ &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} e^{-\theta\alpha t} \psi^n(\theta) \\ &= \exp(t\rho(\theta)) \end{aligned}$$

where

$$\rho(\theta) = \lambda\psi(\theta) - \lambda - \theta\alpha. \tag{9.7}$$

Because X has stationary, independent increments, also

$$\mathbb{E} \exp\left[-\theta(X_{s+t} - X_s)) | \mathcal{F}_s^X\right] = \exp(t\rho(\theta))$$

for any $s \ge 0$, so defining

$$M_t(\theta) = \exp(-\theta(X_t - x_0) - t\rho(\theta)),$$

the following result follows immediately:

Proposition 9.2.1 *For Problem (i),* $(M_t(\theta), \mathcal{F}_t^X)$ *is a martingale for each* $\theta \geq 0$. *For Problem (ii),* $(M_t(\theta), \mathcal{F}_t^X)$ *is a martingale for each* $\theta \leq 0$.

We proved this without using (7.63), but it is quite instructive to verify that (7.63) holds: we have

$$f(t, y) = \exp(-\theta(y - x_0) - t\rho(\theta))$$

and $a(y) = \alpha$ for all y, so (referring to (7.58) for the form of $\mathcal{A}f$)

$$D_t f(t, y) + a(y) D_y f(t, y) = f(t, y)(-\rho(\theta) - \alpha\theta)$$

while

$$\begin{aligned} q(y) \int_{\mathbb{R}} r(y, d\widetilde{y}) \left(f(t, \widetilde{y}) - f(t, y)\right) &= \lambda \left(\int f(t, y + u)\, \mathrm{P}(du) - f(t, y) \right) \\ &= f(t, y)(\lambda\psi(\theta) - \lambda). \end{aligned}$$

Using (9.7) we see that (7.63) 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

$$\psi'(\theta) = -\mathbb{E} U_1 e^{-\theta U_1}.$$

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. (9.6), and that $\theta_0 > 0$ iff $\alpha/\lambda + \xi > 0$, (iff $\rho'(0) < 0$).

Since $M_0(\theta) \equiv 1$, the martingales $M(\theta)$ for $\theta \geq 0$ have constant expectation 1, so applying optional sampling to the bounded stopping times $t \wedge \mathrm{T_r}$ for $t \geq 0$ we obtain

$$1 = \mathbb{E} M_{t \wedge \mathrm{T_r}}(\theta) = \mathbb{E}\left[e^{\theta x_0 - \mathrm{T_r}\rho(\theta)}; \mathrm{T_r} \leq t\right] + \mathbb{E}\left[e^{-\theta(X_t - x_0) - t\rho(\theta)}; \mathrm{T_r} > t\right]$$

using (9.4) on the way. Now take $\theta > \theta_0$ and consider the last term. Since $t < \mathrm{T_r}$, $X_t > 0$ 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

$$1 = \mathbb{E}\left[e^{\theta x_0 - \mathrm{T_r}\rho(\theta)}; \mathrm{T_r} < \infty\right] \quad (\theta > \theta_0).$$

By monotone convergence

$$\mathbb{E}\left[e^{-\mathrm{T}_{\mathrm{r}}\rho(\theta)};\ \mathrm{T}_{\mathrm{r}} < \infty\right] \uparrow p_{\mathrm{r}} \quad \text{as} \quad \theta \downarrow \theta_0,$$

and consequently,

$$p_{\mathrm{r}} = e^{-\theta_0 x_0}.$$

Furthermore, again since $\rho(\theta) > 0$ for $\theta > \theta_0$,

$$\mathbb{E}e^{-\rho(\theta)\mathrm{T}_{\mathrm{r}}} = \mathbb{E}\left[e^{-\mathrm{T}_{\mathrm{r}}\rho(\theta)};\ \mathrm{T}_{\mathrm{r}} < \infty\right] = e^{-\theta x_0}$$

and we have shown the following result:

Proposition 9.2.2 (a) *For Problem (i), the ruin probability is*

$$p_{\mathrm{r}} = e^{-\theta_0 x_0},$$

where $\theta_0 = \sup\{\theta \geq 0 : \rho(\theta) = 0\}$ *and* $\theta_0 = 0$ *iff* $\alpha/\lambda + \xi \leq 0$.

(b) *For* $\vartheta > 0$, *the Laplace transform of the (possibly infinite) random variable* T_{r} *is given by*

$$\mathbb{E}e^{-\vartheta \mathrm{T}_{\mathrm{r}}} = e^{-\rho^{-1}(\vartheta)x_0}, \tag{9.8}$$

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 (9.8) is valid only for $\vartheta > 0$, but that by monotone convergence

$$p_{\mathrm{r}} = e^{-\theta_0 x_0} = \lim_{\vartheta \downarrow 0, \vartheta > 0} e^{-\rho^{-1}(\vartheta)x_0} = \lim_{\vartheta \downarrow 0, \vartheta > 0} \mathbb{E}e^{-\vartheta \mathrm{T}_{\mathrm{r}}}.$$

Problem (ii)

For the solution of Problem (i), (9.4) was used in an essential manner. With the possibility of undershoot occurring in Problem (ii), (9.4) no longer holds. The basic idea now is to consider instead the stopped process $\widetilde{X} = X^{\mathrm{T}_{\mathrm{r}}}$ so that

$$\widetilde{X}_t = \begin{cases} X_t & \text{if } t < \mathrm{T}_{\mathrm{r}} \\ X_{\mathrm{T}_{\mathrm{r}}} & \text{if } t \geq \mathrm{T}_{\mathrm{r}}. \end{cases}$$

Clearly

$$\widetilde{X}_{\mathrm{T}_{\mathrm{r}}} \leq 0 \quad \text{on} \quad (\mathrm{T}_{\mathrm{r}} < \infty).$$

We certainly have that $\widetilde{X}$ is a piecewise deterministic process with state space $\mathbb{R}$ and initial state $\widetilde{X}_0 \equiv x_0 > 0$. If $\widetilde{T}_n$ is the time of the nth 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:

$$\overline{\widetilde{P}}^{(n)}_{z_n|x_0}(t) = \begin{cases} \exp(-\lambda(t-t_n)) & \text{if } y_n > 0 \\ 1 & \text{if } y_n \leq 0 \end{cases}$$

for $t \geq t_n$, while for $t > t_n$, $y_n > 0$,

$$\widetilde{\pi}^{(n)}_{z_n,t|x_0} = \text{the distribution of } y_n + \alpha(t-t_n) + U_1.$$

The expression for $\widetilde{P}^{(n)}$ and $\widetilde{\pi}^{(n)}$ show that $\widetilde{X}$ is a piecewise deterministic time-homogeneous Markov process with

$$\widetilde{\phi}_t(y) = \begin{cases} y+\alpha t & \text{if } y > 0, \\ y & \text{if } y \leq 0, \end{cases}$$

$$\widetilde{q}(y) = \begin{cases} \lambda & \text{if } y > 0, \\ 0 & \text{if } y \leq 0, \end{cases}$$

$$\widetilde{r}(y, \cdot) = \text{the distribution of } y + U_1.$$

From now on assume that $-U_1$ follows an exponential distribution,

$$\mathbb{P}(U_1 \leq u) = e^{\beta u} \quad (u \leq 0),$$

where $\beta > 0$. Thus $\xi = -\frac{1}{\beta}$, so the claim (9.6) amounts to

$$p_{\mathrm{r}} = 1 \quad \text{iff} \quad \alpha\beta \leq \lambda.$$

Let $\vartheta > 0$ and $\zeta \geq 0$ be given. We then claim that there are constants $\gamma = \gamma(\vartheta) < 0$ and $K = K(\vartheta, \zeta)$ (not depending on x_0) such that the process $\left(f_0\left(\mathrm{T_r} \wedge t, \widetilde{X}_t\right)\right)_{t \geq 0}$ with

$$f_0(t,x) = \begin{cases} e^{-\vartheta t+\gamma x} & \text{if } x > 0, \\ Ke^{-\vartheta t+\zeta x} & \text{if } x \leq 0, \end{cases} \tag{9.9}$$

is an $\mathcal{F}_t^{\widetilde{X}}$-martingale. To see this and determine γ and K, we first note that the space-time generator for $\widetilde{X}$ has the form

$$\begin{aligned} \mathcal{A}f(t,x) &= D_t f(t,x) + \alpha D_x f(t,x) \\ &\quad + \widetilde{q}(x) \int \widetilde{r}(x, dy)\, (f(t,y) - f(t,x)) \end{aligned}$$

for $x > 0$, so that by Itô's formula (7.59)

$$\begin{aligned} f\left(t, \widetilde{X}_t\right) &= f(0, x_0) + \int_0^t \mathcal{A}f\left(s, \widetilde{X}_s\right) ds \\ &\quad + \int_{]0,t]\times\mathbb{R}} \left(f(s,y) - f\left(s, \widetilde{X}_{s-}\right)\right) \widetilde{M}(ds, dy) \end{aligned} \tag{9.10}$$

for $t \le T_r$ provided only that for $x > 0$, $f(t,x)$ is continuously differentiable in t and in x and *no matter how $f(t,x)$ is defined for $x < 0$*: (9.10) follows directly from (7.59) if $t < \mathrm{T_r}$ and it then remains only to verify that the jumps on the left- and right-hand side of (9.10) match at the time $t = \mathrm{T_r}$ if $\mathrm{T_r} < \infty$, which is immediate. (In (9.10), $\widetilde{M} = \widetilde{\mu} - \widetilde{L}$ is of course the martingale measure for the RCM $\widetilde{\mu}$ determined by $\left(\widetilde{T}_n, \widetilde{Y}_n\right)$, with $\widetilde{L}$ the $\mathcal{F}_t^{\widetilde{X}}$-compensating measure for $\widetilde{\mu}$).

Using (9.10) with $f = f_0$ leads to the identity

$$f_0\left(\mathrm{T_r} \wedge t, \widetilde{X}_t\right) = f_0(0, x_0) + \int_0^{\mathrm{T_r}\wedge t} \mathcal{A} f_0\left(s, \widetilde{X}_s\right) ds + \int_{]0,t]\times\mathbb{R}} \left(f_0(s,y) - f_0\left(s, \widetilde{X}_{s-}\right)\right) \widetilde{M}(ds, dy)$$

for all $t \ge 0$. (Since neither $\widetilde{\mu}$ nor $\widetilde{L}$ have any mass beyond $\mathrm{T_r}$, the last integral equals the integral over $]0, \mathrm{T_r} \wedge t] \times \mathbb{R}$). If f_0 is bounded, i.e., $\gamma \le 0$, the last term is a martingale, hence if also $\mathcal{A} f_0(t,x) \equiv 0$ for $t \ge 0$, $x > 0$, we find that $f_0\left(\mathrm{T_r} \wedge t, \widetilde{X}_t\right)$ defines a martingale and that

$$\mathbb{E} f_0\left(\mathrm{T_r} \wedge t, \widetilde{X}_t\right) = f(0, x_0) \quad (t \ge 0). \tag{9.11}$$

The condition $\mathcal{A} f_0(t,x) \equiv 0$ also involves $f_0(t,x)$ for $x < 0$! By computation

$$\begin{aligned}\mathcal{A} f_0(t,x) &= e^{-\vartheta t + \gamma x}(-\vartheta + \alpha\gamma - \lambda) \\ &\quad + e^{-\vartheta t}\lambda\left(\int_{-\infty}^{-x} \beta e^{\beta y} K e^{\zeta(x+y)}\, dy + \int_{-x}^{-0} \beta e^{\beta y} e^{\gamma(x+y)}\, dy\right) \\ &= e^{-\vartheta t}\left[\left(-\vartheta + \alpha\gamma - \lambda + \tfrac{\lambda\beta}{\beta+\gamma}\right) e^{\gamma x} + \lambda\beta\left(\tfrac{K}{\beta+\zeta} - \tfrac{1}{\beta+\gamma}\right) e^{-\beta x}\right]\end{aligned}$$

for $x > 0$, so in order for this to vanish we must take $K = K_0$ where

$$K_0(\vartheta, \zeta) = \frac{\beta + \zeta}{\beta + \gamma_0} \tag{9.12}$$

with $\gamma_0 \le 0$ solving the *Lundberg equation*

$$-\vartheta + \alpha\gamma - \lambda + \frac{\lambda\beta}{\beta + \gamma} = 0. \tag{9.13}$$

Rewriting this as an equation in γ for the roots of a polynomial of degree 2, which for $\vartheta > 0$ is easily seen to have one root < 0 and one root > 0, it follows that

$$\gamma_0(\vartheta) = \frac{1}{2\alpha}\left(\vartheta + \lambda - \alpha\beta - \sqrt{(\vartheta + \lambda - \alpha\beta)^2 + 4\alpha\beta\vartheta}\right). \tag{9.14}$$

So now let f_0 be given by (9.9) with $\gamma = \gamma_0 = \gamma_0(\vartheta)$, $K = K_0 = K_0(\vartheta, \zeta)$. Then (9.11) holds, i.e.,

$$e^{\gamma_0 x_0} = \mathbb{E}\left[K_0 e^{-\vartheta \mathrm{T_r} + \zeta X_{\mathrm{T_r}}}; \mathrm{T_r} \le t\right] + \mathbb{E}\left[e^{-\vartheta t + \gamma_0 X_t}; \mathrm{T_r} > t\right].$$

For $t \to \infty$, the last term vanishes by dominated convergence since $\vartheta > 0$ and $\gamma_0 X_t \le 0$ when $t < \mathrm{T_r}$. Thus

$$e^{\gamma_0 x_0} = \mathbb{E}\left[K_0 e^{-\vartheta \mathrm{T_r} + \zeta X_{\mathrm{T_r}}}; \mathrm{T_r} < \infty\right]. \tag{9.15}$$

For $\zeta = 0$ and with $\vartheta \downarrow 0$ we obtain a formula for the ruin probability $p_\mathrm{r} = \mathbb{P}(\mathrm{T_r} < \infty)$. By inspection

$$\lim_{\vartheta \downarrow 0} \gamma_0(\vartheta) = \begin{cases} 0 & \text{if } \alpha\beta \le \lambda, \\ \frac{\lambda}{\alpha} - \beta & \text{if } \alpha\beta > \lambda, \end{cases}$$

and we have a complete proof of the following result:

Proposition 9.2.3 *For Problem (ii) with initial state $x_0 > 0$, assuming that the $-U_n$ are exponential at rate $\beta > 0$, the (possibly defective) joint Laplace transform for the time to ruin $\mathrm{T_r}$ and the size of the undershoot $\mathrm{Y_r} = -X_{\mathrm{T_r}}$ is given by the expression*

$$\mathbb{E}\left[e^{-\vartheta \mathrm{T_r} - \zeta \mathrm{Y_r}}; \mathrm{T_r} < \infty\right] = \frac{\beta + \gamma_0(\vartheta)}{\beta + \zeta} e^{\gamma_0(\vartheta) x_0} \tag{9.16}$$

with γ_0 as in (9.14). The ruin probability p_r is given by the expression

$$p_\mathrm{r} = \begin{cases} 1 & \text{if } \alpha\beta \le \lambda, \\ \frac{\lambda}{\alpha\beta} e^{(\lambda/\alpha - \beta)x_0} & \text{if } \alpha\beta > \lambda. \end{cases}$$

Exercise 9.2.1 Show that

$$\mathbb{E}\left[e^{-\zeta \mathrm{Y_r}} | \mathrm{T_r} < \infty\right] = \frac{\beta}{\beta + \zeta} \quad (\zeta \ge 0),$$

i.e., that conditionally on the event of ruin, the size of the undershoot is exponential at rate β and thus has the same distribution as $-U_n$. Explain the result.

Exercise 9.2.2 For Problem (ii) it was originally assumed that $\alpha > 0$, but of course there is also a model given by (9.3) even if $\alpha \le 0$ and all U_n are < 0 with $-U_n$ exponential at rate β. Show that if $\alpha = 0$, then $p_\mathrm{r} = 1$, and that the joint Laplace transform of $\mathrm{T_r}$ and $\mathrm{Y_r}$ is given as in (9.16) with

$$\gamma_0(\vartheta) = -\frac{\beta\vartheta}{\vartheta + \lambda}.$$

Consider also the case $\alpha < 0$. Then there are two types of ruin: ruin by jump as before, but also ruin by continuity, e.g., because X_t follows the straight line $x_0 + \alpha t$ without jumps until the line hits the level 0 at time $-\frac{x_0}{\alpha}$. Of course $p_\mathrm{r} = 1$ and ruin is bound to happen before time $-\frac{x_0}{\alpha}$. Show that for $\alpha < 0$, the Lundberg equation (9.13) has two strictly negative roots $\gamma_-(\vartheta)$ and $\gamma_+(\vartheta)$ when $\vartheta > 0$. Show also that with $\gamma = \gamma_-$ or $= \gamma_+$ it holds (cf. (9.15)), with A_j the event that ruin occurs by jump, A_c the event that ruin occurs by continuity, that

$$e^{\gamma x_0} = \mathbb{E}\left[K_\gamma e^{-\vartheta \mathrm{T_r} + \zeta X_{\mathrm{T_r}}}; A_\mathrm{j}\right] + \mathbb{E}\left[e^{-\vartheta \mathrm{T_r}}; A_\mathrm{c}\right]$$

where (see (9.12)) $K_\gamma = \frac{\beta+\zeta}{\beta+\gamma}$. Use the two resulting equations to find the partial Laplace transforms

$$\mathbb{E}\left[e^{-\vartheta \mathrm{T_r} + \zeta X_{\mathrm{T_r}}}; A_\mathrm{j}\right] \quad \text{and} \quad \mathbb{E}\left[e^{-\vartheta \mathrm{T_r}}; A_\mathrm{c}\right].$$

9.3 The soccer model

The fact that likelihood processes are martingales, see Section 5.1, may be used to describe the time dynamics of a process on a finite time interval $[0, T]$ from knowledge only of the distribution of the process at a final time point T. To illustrate this we shall describe an MPP model for the development over time of a football (here meaning soccer) game.

In the national soccer leagues in Europe, each team plays each other team in a prescribed number of matches, some at home and some away. Typically, in each match a winner is awarded 3 points, there is 1 point for each team if the match is drawn and a losing team gets 0 points. The league champion is the team with most points at the end of the season, with goal difference or other criteria used if the top teams end up with the same number of points.

Consider a match where team i plays at home to team j — we shall assume that the results of different games are stochastically indpendent and henceforth focus on the outcome of a single game. A simple model, which allows for the comparison of all teams, is obtained by attaching to each team k an attack parameter α_k and a defense parameter δ_k such that the end result of a game between i and j is x goals scored by i, y goals scored by j with probability

$$p(x, y; \alpha_i, \delta_i, \alpha_j, \delta_j) = \frac{1}{x!}\left(\alpha_i \delta_j\right)^x \frac{1}{y!}\left(\alpha_j \delta_i\right)^y e^{-\alpha_i \delta_j - \alpha_j \delta_i} \quad (x, y \in \mathbb{N}_0), \tag{9.17}$$

i.e., the two scores are independent and Poisson, fitting into a multiplicative Poisson model. (In this model the best team is the one with the highest value of α_k / δ_k. In statistical terms, the model is an exponential family model with the total number of goals scored by each team and the total number of goals scored against each team as a minimal sufficient statistic. As it stands, the model is over-parametrized — replacing all α_k by $c\alpha_k$, all δ_k by δ_k / c for some $c \neq 0$, does not change the probability (9.17)).

The model described by (9.17) is too simple and does not take into account that it is (usually) an advantage to play at home, or that it is particularly important to get points (win or draw) from a game. The one advantage the model does have is that there is a natural way to describe the evolution of the game between i and j over time: if the score at time t is X_t goals by i against Y_t goals for j, just assume that X and Y are independent, homogeneous Poisson processes with intensity parameters $\alpha_i \delta_j / T$ and $\alpha_j \delta_i / T$ respectively, with T denoting the duration of a soccer game, $T = 90$ minutes. In particular, the final score (X_T, Y_T) has the distribution given by (9.17).

It is easy to adjust for home court advantage: let the attack and defense parameters depend on whether the match is played at home or away. Writing a, d instead of α, δ when playing away, the game between i at home against j (in a still too simple model) would result in the score x for i, y for j with a probability $p(x, y; \alpha_i, \delta_i, a_j, d_j)$ given by (9.17) with α_j and δ_j replaced by a_j and d_j respectively. Again, the evolution of the game is described by two independent Poisson processes in the obvious manner.

Finally, to incorporate into the model something that describes the ability of each team to get points from a game, we attach to each team k one further parameter $w_k > 0$, so that the end result of a game between i at home against j is x goals for i, y goals for j with probability

$$\begin{aligned} & p(x, y; \alpha_i, \delta_i, w_i, a_j, d_j, w_j) \\ & = \frac{1}{C}\left(w_i^3 1_{x>y} + w_i w_j 1_{x=y} + w_j^3 1_{x<y}\right) p(x, y; \alpha_i, \delta_i, a_j, d_j), \end{aligned} \tag{9.18}$$

where $C = C(\alpha_i, \delta_i, w_i, a_j, d_j, w_j)$ is a normalising constant so that the probabilities add to 1, $\sum_{x\geq 0, y\geq 0} p(x, y) = 1$. Thus, if $w_i = w_j = 1$ we are back to probabilities of the form (9.17), while if, say, $w_i > w_j$, then results with $x > y$ or $x = y$ are favoured more. (The full model for the entire tournament is again an exponential family model with the minimal sufficient statistic comprising (i) the total number of goals scored by any k at home, (ii) the total number of goals scored by any k away, (iii) the total number of goals scored against k at home, (iv) the total number of goals scored against k away, (iv) the total number of points obtained by any k).

Remark 9.3.1 It is possible to express C using the modified Bessel function I_0 of order 0. One expression for I_0 is

$$I_0\left(2\sqrt{\lambda}\right) = \sum_{x=0}^{\infty} \frac{\lambda^x}{(x!)^2}. \tag{9.19}$$

Defining

$$F(\lambda, \mu) = \sum_{y=0}^{\infty} \sum_{x=y+1}^{\infty} \frac{\lambda^x}{x!} \frac{\mu^y}{y!}$$

one finds using (9.19) that

$$D_\lambda F(\lambda, \mu) = I_0\left(2\sqrt{\lambda\mu}\right) + F(\lambda, \mu),$$

and since $F(0, \mu) = 0$ this gives

$$F(\lambda, \mu) = e^\lambda \int_0^\lambda e^{-s} I_0\left(2\sqrt{s\mu}\right) ds.$$

It follows that

$$e^{\alpha_i d_j + a_j \delta_i} C = w_i^3 F\left(\alpha_i d_j, a_j \delta_i\right) + w_i w_j I_0\left(2\sqrt{\alpha_i d_j a_j \delta_i}\right) + w_j^3 F\left(a_j \delta_i, \alpha_i d_j\right).$$

It remains to describe the time evolution of a game, the final outcome of which has the distribution (9.18). Fix $\alpha_i, \delta_i, a_j, d_j$ and let Q denote the distribution of the two independent homogeneous Poisson processes X and Y on the interval $[0, T]$ described above, which in the final model corresponds to $w_i = w_j = 1$. Also, let for $w_i, w_j > 0$ given but arbitrary, $\widetilde{Q}$ denote the distribution of X and Y, when the full set of parameters is $\alpha_i, \delta_i, a_j, d_j, w_i, w_j$ – $\widetilde{Q}$ is the distribution we are looking for. (Formally $\widetilde{Q}$ is the joint distribution of the two counting processes X and Y, trivially identified with an RCM with a mark space having two elements, i (goals scored by i) and j (goals scored by j). Below we think of $X = N^{\circ i}$ and $Y = N^{\circ j}$ as defined on the canonical space $\mathcal{M}$).

We shall construct $\widetilde{Q}$ by finding the likelihood process $\mathcal{L} = (\mathcal{L}_t)_{0 \le t \le T}$ where $\mathcal{L}_t = d\widetilde{Q}_t / dQ_t$ (cf. (5.1)), assuming that the *final value* $\mathcal{L}_T$ *depends on the final scores* X_T, Y_T *and the parameters only* and more precisely define

$$\begin{aligned}
\mathcal{L}_T &= \frac{p(X_T, Y_T; \alpha_i, \delta_i, w_i, a_j, d_j, w_j)}{p(X_T, Y_T; \alpha_i, \delta_i, a_j, d_j)} \\
&= \frac{1}{C}\left(w_i^3 1_{(X_T > Y_T)} + w_i w_j 1_{(X_T = Y_T)} + w_j^3 1_{(X_T < Y_T)}\right).
\end{aligned}$$

We next define $\widetilde{Q}$ on $\mathcal{H}_T$ by

$$\widetilde{Q}(H) = \int_H \mathcal{L}_T \, dQ \quad (H \in \mathcal{H}_T)$$

and by the martingale property of likelihood processes, see Section 5.1, are then forced to take

$$\mathcal{L}_t = E_Q\left(\mathcal{L}_T \,|\mathcal{H}_t\right) \quad (0 \le t \le T).$$

Since under Q, X and Y are independent and homogeneous Poisson, this reduces to

$$\mathcal{L}_t = \frac{1}{C}\left(w_i^3 \psi_{\text{win}}(t; Y_t - X_t) + w_i w_j \psi_{\text{draw}}(t; Y_t - X_t) + w_j^3 \psi_{\text{loose}}(t; Y_t - X_t)\right), \tag{9.20}$$

using the notation

$$\begin{aligned}
\psi_{\text{win}}(t; z) &= Q\left(X_{T-t} - Y_{T-t} > z\right) \\
\psi_{\text{draw}}(t; z) &= Q\left(X_{T-t} - Y_{T-t} = z\right) \\
\psi_{\text{loose}}(t; z) &= Q\left(X_{T-t} - Y_{T-t} < z\right)
\end{aligned}$$

for $z \in \mathbb{Z}$, i.e., each $\psi(t; z)$ is determined by the distribution of the difference between two independent Poisson random variables with parameters $\alpha_i d_j \left(1 - \frac{t}{T}\right)$ and $a_j \delta_i \left(1 - \frac{t}{T}\right)$ respectively.

From (9.20) it is easy to find the intensity processes for $\widetilde{Q}$, i.e., the intensities $\widetilde{\lambda}^{\circ i}$ and $\widetilde{\lambda}^{\circ j}$ for i, respectively j scoring a goal: if $\Delta \overline{N}_t^{\circ} = 1$, we have, cf. (5.7),

$$\mathcal{L}_t = \mathcal{L}_{t-} \frac{\widetilde{\lambda}_t^{\circ \eta\langle t\rangle}}{\lambda_t^{\circ \eta\langle t\rangle}};$$

see (5.7). But here $\lambda_t^{\circ i} = \alpha_i d_j / T$ and $\lambda_t^{\circ j} = a_j \delta_i / T$, hence

$$\begin{aligned}
\widetilde{\lambda}_t^{\circ i} &= \frac{\alpha_i d_j}{T} \frac{\varphi\,(t; Y_{t-} - X_{t-} - 1)}{\varphi\,(t; Y_{t-} - X_{t-})}, \\
\widetilde{\lambda}_t^{\circ j} &= \frac{a_j \delta_i}{T} \frac{\varphi\,(t; Y_{t-} + 1 - X_{t-})}{\varphi\,(t; Y_{t-} - X_{t-})}
\end{aligned}$$

where φ is the function

$$\varphi\,(t; z) = w_i^3 \psi_{\text{win}}\,(t; z) + w_i w_j \psi_{\text{draw}}\,(t; z) + w_j^3 \psi_{\text{loose}}\,(t; z)\,.$$

It follows in particular from Theorem 7.2.1 (see also (7.14)) that (X, Y) under $\widetilde{Q}$ is a (non-homogeneous) Markov chain, and that in fact the difference process $X - Y$ is also a Markov chain.

If, in a soccer match, one team is a strong favourite to win, but is nevertheless behind on goals shortly before the finish, the final minutes may well prove intense with the supposedly stronger team very much on the offensive. To see if such a match evolution is captured by the model, consider $\lim_{t\to T} \widetilde{\lambda}_t^{\circ k}$ for $k = i, j$. When $t \to T$, the Q-distribution of $Y_{T-t} - X_{T-t}$ becomes degenerate at 0 so that

$$\begin{aligned}
\lim_{t\to T} \psi_{\text{win}}\,(t; z) &= 1_{z \le -1}, \\
\lim_{t\to T} \psi_{\text{draw}}\,(t; z) &= 1_{z=0}, \\
\lim_{t\to T} \psi_{\text{loose}}\,(t; z) &= 1_{z \ge 1}.
\end{aligned}$$

Thus e.g.,

$$\widetilde{\lambda}_T^{\circ i} = \lim_{t\to T} \widetilde{\lambda}_t^{\circ i} = \begin{cases} \frac{\alpha_i d_j}{T} \frac{w_i}{w_j^2} & \text{if } Y_{T-} = X_{T-} + 1 \\ \frac{\alpha_i d_j}{T} \frac{w_i^2}{w_j} & \text{if } Y_{T-} = X_{T-} \\ \frac{\alpha_i d_j}{T} & \text{otherwise} \end{cases} \tag{9.21}$$

and similarly for $\lim \widetilde{\lambda}_t^{\circ j}$. We see that if i is stronger than j in the sense that w_i is (much) larger than w_j, then the intensity for i scoring in the final minutes is large if the score is either level or i is behind by just one goal! If all $w_i = w > 0$ and the common value is large, the intensity for a goal (either way) from a level score is large, but small for i scoring if i is behind by just one goal. If, on the other hand, w is small, the intensity for a goal is small if the score is level, but high that i scores if i is behind by one goal.

Exercise 9.3.1 Not everything was argued carefully. In particular you should check that (5.7) indeed may be used to derive the expressions for $\widetilde{\lambda}_t^{\circ i}$ and $\widetilde{\lambda}_t^{\circ j}$. You should also give all details in the reasoning leading to (9.21).

10

A Model from Finance

A simple specific PDMP model is set up to describe the price of a finite number of risky assets. For this model we discuss self-financing trading strategies, arbitrage and the concept of equivalent martingale measures, and the fair pricing of contingent claims.

References. Pliska [99] treats models in discrete time and was used as inspiration for the example given here. Other major references on mathematical finance, mainly in continuous time, include the monographs Björk [9], Cont and Tankov [20], Föllmer and Schied [42], Karatzas and Shreve [74], Melnikov [91], Musiela and Rutkowski [96] and Shiryaev [113].

10.1 The model

We shall consider a PDMP-model for the development of the price of $d \geq 1$ risky assets and where an investor may also place money in a safe bank account or bond. The price of asset i at time t is denoted $X_t^i > 0$ and $X = \left(X^i\right)_{1 \leq i \leq d}$ is the $\mathbb{R}_+^d$-valued price process. Investing in the bank account yields a fixed interest rate $r > 0$, so that if one unit of currency is invested at time 0, the value of that investment at time t is

$$X_t^0 \equiv e^{rt}.$$

Note. Throughout, index i, $0 \leq i \leq d$ is written as an upper index and $i = 0$ always refers to the riskless asset. The assumption $r > 0$ is standard in finance but is not required formally: everything below is true also if $r \leq 0$.

To descibe the price process, let $U = \left(U^i\right)_{1 \leq i \leq d}$ be an $\mathbb{R}^d$-valued compound Poisson process,

$$U_t = \sum_{n=1}^{\overline{N}_t} Y_n, \quad U_0 \equiv 0, \tag{10.1}$$

with jump times T_n, jump sizes $Y_n = \left(Y_n^i\right)_{1 \le i \le d} \in \mathbb{R}^d_{\setminus 0}$, $\overline{N}_t = \sum_{n=1}^{\infty} 1_{(T_n \le t)}$ a homogeneous Poisson process with rate $\lambda > 0$ independent of (Y_n) and the Y_n themselves iid with distribution κ. Equivalently, U is an $\mathbb{R}^d$-valued Lévy process with bounded Lévy measure $\lambda\kappa$, cf. Section 6.2.

Now assume given an initial fixed price $X_0^i \equiv x_0^i > 0$ for each risky asset and assume that each X^i satisfies a stochastic differential equation of the form

$$dX_t^i = a^i X_t^i \, dt + \sigma^i X_{t-}^i \, dU_t^i, \tag{10.2}$$

where $a^i \in \mathbb{R}$ and $\sigma^i > 0$ for $1 \le i \le d$ are given constants, i.e., for $1 \le i \le d$ and $t \ge 0$ we have

$$X_t^i = x_0^i + a^i \int_0^t X_s^i \, ds + \sigma^i \int_{]0,t]} X_{s-}^i \, dU_s^i. \tag{10.3}$$

The model is a multidimensional analogue of the model from Subsection 7.4.3 when taking $a(x) = ax$, $\sigma(x) = \sigma x$ in (7.48). The process X jumps precisely at the jump times T_n for U,

$$\Delta X_{T_n}^i = X_{T_n-}^i \sigma^i Y_n^i,$$

and between jumps

$$D_t X_t^i = a^i X_t^i,$$

i.e., for all $n \ge 1$,

$$X_t^i = e^{a^i (t - T_{n-1})} X_{T_{n-1}}^i \qquad (T_{n-1} \le t < T_n).$$

Thus

$$\begin{aligned} X_{T_n}^i &= X_{T_n-}^i + \Delta X_{T_n}^i \\ &= e^{a^i (T_n - T_{n-1})} X_{T_{n-1}}^i \left(1 + \sigma^i Y_n^i\right) \\ &= e^{a^i T_n} \left(\prod_{k=1}^{n} \left(1 + \sigma^i Y_k^i\right) \right) x_0^i, \end{aligned}$$

and consequently, for all i and t,

$$\begin{aligned} X_t^i &= e^{a^i (t - T_{\langle t \rangle})} X_{T_{\langle t \rangle}}^i \\ &= e^{a^i t} \left(\prod_{k=1}^{\overline{N}_t} \left(1 + \sigma^i Y_k^i\right) \right) x_0^i. \end{aligned} \tag{10.4}$$

Since we are requiring that all $X_t^i > 0$, it is assumed from now on that the Y_n satisfy that

$$Y_n^i > -\frac{1}{\sigma^i} \quad (1 \le i \le d)\,, \tag{10.5}$$

a condition that expressed in terms of κ reads

$$\kappa\left\{y \in \mathbb{R}^d_{\backslash 0} : y^i > -\tfrac{1}{\sigma^i}, 1 \le i \le d\right\} = 1. \tag{10.6}$$

With this assumption in force we may take logarithms in (10.4) and obtain

$$\log X_t^i = \log x_0^i + a^i t + \sum_{k=1}^{\overline{N}_t} \log\left(1 + \sigma^i Y_k^i\right) \tag{10.7}$$

which shows that the process $\log X = \left(\log X^i\right)_{1 \le i \le d}$ is a d-dimensional Lévy process with drift vector $\left(a^i\right)$ and Lévy measure $\lambda\widetilde{\kappa}$, where $\widetilde{\kappa}$ is the distribution of the vector $\left(\log\left(1 + \sigma^i Y_n^i\right)\right)_{1 \le i \le d}$ for any n.

The only restriction we shall for the time being place on the model is that the distribution κ of the jump sizes for U satisfy (10.6). In particular, for $d \ge 2$ it is allowed that two different coordinates of U jump simultaneously, which in turn allows for two different X^i to be correlated.

In the definition of the model, the parameters σ^i are really redundant: X can also be described by the equation (10.2) using all $\sigma^i = 1$ by simply adjusting the distribution κ of the Y_n. However, we prefer to keep the σ^i as some kind of volatility parameters.

Exercise 10.1.1 With X given by (10.4), show that it is always true that each $\log X^i$ is a one-dimensional Lévy process with the constant jump intensity $\widetilde{\lambda}^i = \lambda\kappa^i\left(\mathbb{R}_{\backslash 0}\right)$ and the jump probability distribution which is the conditional distribution $\kappa^i/\kappa^i\left(\mathbb{R}_{\backslash 0}\right)$ of the Y_n^i given that $Y_n^i \ne 0$. Here κ^i denotes the marginal distribution of Y_n^i — so of course $\kappa^i\left(\mathbb{R}_{\backslash 0}\right) < 1$ is possible for $d \ge 2$. (*Hint*: it should be clear that $\log X^i$ is a Lévy process. To find $\widetilde{\lambda}^i$, just determine $\mathbb{E}N_t^i$ where N^i counts the number of jumps for X^i).

Exercise 10.1.2 Show that the coordinate processes X^i for $1 \le i \le d$ are independent iff $\kappa\left(\bigcup_{i=1}^d A^i\right) = 1$, where

$$A^i = \left\{\left(y^j\right) \in \mathbb{R}^d_{\backslash 0} : y^j = 0 \text{ for all } j \ne i\right\}.$$

From (10.4) and (10.5) one immediately obtains the following result:

Proposition 10.1.1 *The price process $X = \left(X^i\right)_{1 \le i \le d}$ is a homogeneous PDMP with state space $G = \mathbb{R}^d_+$ determined as follows: writing $\widetilde{Y}_n = X_{T_n}$ for the state reached by X at time T_n, $X_t = \phi_{t - T_{\langle t \rangle}}\left(\widetilde{Y}_{\langle t \rangle}\right)$ where*

$$\phi_t(x) = \left(e^{a^i t} x^i\right)_{1 \le i \le d},$$

and, with $\widetilde{Z}_n = \left(T_1, \ldots, T_n; \widetilde{Y}_1, \ldots, \widetilde{Y}_n\right)$, $\widetilde{Y}_0 = x_0$,

$$\mathbb{P}_{|x_0}\left(T_{n+1} > t \,\middle|\, \widetilde{Z}_n\right) = \exp\left(-\int_0^{t-T_n} q\left(\phi_s\left(\widetilde{Y}_n\right)\right) ds\right),$$
$$\mathbb{P}_{|x_0}\left(\widetilde{Y}_{n+1} \in \cdot \,\middle|\, \widetilde{Z}_n, T_{n+1}\right) = r\left(\phi_{T_{n+1}-T_n}\left(\widetilde{Y}_n\right), \cdot\right)$$

with the jump intensity function $q(x)$ *and the jump probabilities* $r(x,\cdot)$ *given by*

$$q(x) \equiv \lambda, \quad r(x,\cdot) = \textit{the distribution of } \left(\left(1+\sigma^i Y_n^i\right) x^i\right)_{1\le i\le d}.$$

Before proceeding it is useful to note the following: if $(\Omega, \mathcal{F}, \mathbb{P})$ is the probability space on which U and X are defined, without loss of generality we may and shall assume that $Y_n(\omega) \neq 0$ for *all* $\omega \in \Omega$. Then, with μ the RCM determined by $((T_n), (Y_n))$, it is clear that for all t,

$$\mathcal{F}_t^X = \mathcal{F}_t^\mu. \tag{10.8}$$

(The inclusion $\subset$ is trivial. For the converse just use that $\overline{N}_t = \sum_{0<s\le t} 1_{(\Delta X_s \neq 0)}$ and $Y_n^i = \Delta X_{T_n}^i / \sigma^i X_{T_n-}^i$).

Remark 10.1.1 The description (10.2) is of course very much inspired by the basic model

$$dX_t^i = a^i X_t^i\, dt + \sigma^i X_t^i\, dB_t^i, \quad X_0^i \equiv x_0^i \tag{10.9}$$

for the *continuous* random development of the price of each asset i. Here $B = \left(B^i\right)_{1\le i\le d}$ is a d-dimensional Brownian motion with mean vector 0, $\mathbb{E}B_t^i = 0$ for all i and t, and some covariance matrix Γ, $\mathrm{Cov}\left(B_t^i, B_t^j\right) = t\Gamma_{ij}$ for all i, j and t (with $\Gamma = I_{d\times d}$ a particularly important case). It is worth noting that the model (10.2) with jumps can be made to approximate the diffusion model (10.9) arbitrarily well. By Itô's formula for continuous processes (10.9) implies

$$\log X_t^i = \log x_0^i + \left(a^i - \tfrac{1}{2}\Gamma_{ii}\right) t + B_t^i$$

so the idea is simply to take a sequence of jump processes $X^K = \left(X^{i,K}\right)_{1\le i\le d}$ of the form (10.2) (driven by a compound Poisson process U^K) such that for every t, the vector $\left(\log X_t^{i,K}\right)_{1\le i\le d}$ converges in distribution to the vector $\left(\log X_t^i\right)_{1\le i\le d}$ as $K \to \infty$. But by (10.7)

$$\log X_t^{i,K} = \log x_0^i + a^{i,K} t + \sum_{k=1}^{\overline{N}_t^K} \log\left(1 + \sigma^{i,K} Y_k^{i,K}\right),$$

so the desired weak convergence will in fact follow by taking e.g., $a^{i,K} = a^i - \frac{1}{2}\Gamma_{ii}$, $\sigma^{i,K} = \sigma^i$, letting λ^K, the jump rate for X^K, satisfy that $\lambda^K \to \infty$ and finally requiring that the distribution of the vector $Y_1^K = \Delta U_{T_1}^K$ equal that of $\frac{1}{\sqrt{\lambda^K}} Y'$, where

Y' is an arbitrary d-dimensional random variable with mean vector 0 and covariance matrix Γ. With this setup it even holds that the sequence X^K of processes converges in distribution to X (with X^K, X viewed as random variables with values in the Skorohod space $D^d(\mathbb{R}_0)$ of $\mathbb{R}^d$-valued cadlag functions).

(For an easy proof of the convergence in distribution of $\log X_t^K$ to $\log X_t$ for a given t, just compute characteristic functions).

10.2 Portfolios and self-financing strategies

By a *trading strategy* we shall mean a $d+1$-dimensional process $\psi = \left(\psi^i\right)_{0 \le i \le d}$ with, as is vital, each ψ^i $\mathcal{F}_t^X$-predictable. In particular, the initial values ψ_0^i are just some constants. The interpretation is that ψ_t^i is the number of units the investor holds in asset i at time t, and $\left(\psi_t^i\right)_{0 \le i \le d}$ is then the investor's *portfolio* at time t. Changes in ψ_t^i when t varies reflect that the investor has been trading in asset i — this may be done so that the change is continuous in t, but also by a substantial trade at some point in time corresponding to a jump for ψ_t^i which however, due to the predictability of the ψ^i, can never coincide with a jump for any of the X^i. Note that ψ_t^0 is the amount sitting in the bank account at time t.

In the discussion that follows we shall permit borrowing and shortselling of assets, i.e., $\psi_t^i \le 0$ is allowed.

The *value process* $V = (V_t)_{t \ge 0}$ determined by the trading strategy ψ is the $\mathbb{R}$-valued process given by

$$V_t = \psi_t^0 e^{rt} + \sum_{i=1}^d \psi_t^i X_t^i. \tag{10.10}$$

The *cumulated gains process* $G = (G_t)_{t \ge 0}$ is the $\mathbb{R}$-valued cadlag process defined by

$$G_t = \int_0^t \psi_s^0 r e^{rs}\, ds + \int_{]0,t]} \sum_{i=1}^d \psi_s^i\, dX_s^i, \tag{10.11}$$

provided the integrals exist, the intention being that the (possibly negative) gain incurred by the investor over the infinitesimal interval $[t, t+dt[$ amounts to

$$dG_t = \psi_t^0\, de^{rt} + \sum_{i=1}^d \psi_t^i\, dX_t^i \tag{10.12}$$

because the predictability of the ψ_t^i does not allow the ψ_t^i-values to change across the interval and also ψ_t^i is the same as the value of ψ^i just before t.

The next concept is fundamental:

Definition 10.2.1 A trading strategy ψ is *self-financing* if the integrals (10.11) are well defined and finite for all t and

$$V_t \equiv V_0 + G_t. \tag{10.13}$$

To understand what it means for a trading strategy ψ to be self-financing, consider a ψ that is smooth in the sense that between the jump times for X each ψ_t^i is continuous and differentiable in t. Then if there is no jump for X at time t, we have that all $\Delta\psi_t^i = 0$ and $\Delta V_t = 0$ and from (10.10) and (10.2) see that

$$dV_t = \left(\psi_t^0 re^{rt} + \sum_{i=1}^d \psi_t^i a^i X_t^i\right) dt + \left(\left(D_t\psi_t^0\right) e^{rt} + \sum_{i=1}^d \left(D_t\psi_t^i\right) X_t^i\right) dt.$$

Hence $dV_t = dG_t$ between jumps iff

$$\left(D_t\psi_t^0\right) e^{rt} + \sum_{i=1}^d \left(D_t\psi_t^i\right) X_t^i \equiv 0, \tag{10.14}$$

i.e., the infinitesimal changes in the portfolio during the interval $[t, t+dt[$ are paid for exclusively by the earned interest in the bank account and the current value of the price process. If, on the other hand, $\Delta X_t \neq 0$, the condition for ψ to be self-financing is automatically fulfilled in the sense that the predictability of ψ (cf. Proposition 4.2.1 (biv)) implies that

$$\Delta V_t = \sum_{i=1}^d \psi_t^i \Delta X_t^i = \Delta G_t \tag{10.15}$$

provided V is cadlag at time t — so that it makes sense to talk about ΔV_t and compare it to the jump of the cadlag process G — which however is not always the case; see the discussion following (10.24) below.

The *discounted* value process V^* is defined by

$$V_t^* = e^{-rt} V_t = \psi_t^0 + \sum_{i=1}^d \psi_t^i X_t^{*i}$$

where X^* is the discounted price process,

$$X_t^{*i} = e^{-rt} X_t^i. \tag{10.16}$$

The discounted gains process G^* is defined analogously to G, see (10.12), but using X^* instead of X and replacing de^{rt} by $d1 = 0\,dt$,

$$G_t^* = \int_{]0,t]} \sum_{i=1}^d \psi_s^i \, dX_s^{*i}, \tag{10.17}$$

whenever the integral exists.

To avoid problems with integrability in the sequel it is often convenient to assume that the ψ^i (or other relevant processes) are *pathwise bounded*, i.e., that for ($\mathbb{P}$-almost) all $\omega \in \Omega$ and $t \geq 0$,

$$\sup_{s \leq t} \left|\psi_s^i(\omega)\right| < \infty. \tag{10.18}$$

Proposition 10.2.1 (i) *Suppose that ψ is self-financing with all ψ^i for $0 \leq i \leq d$ pathwise bounded. Then*

$$V_t^* = V_0^* + G_t^*. \tag{10.19}$$

(ii) *Suppose that ψ^i for $1 \leq i \leq d$ are predictable, pathwise bounded processes, let $V_0^* = V_0$ be a given constant, and define*

$$\psi_t^0 = V_0^* + G_t^* - \sum_{i=1}^d \psi_t^i X_t^{*i} \tag{10.20}$$

with G_t^* given by (10.17). Then ψ^0 is predictable, pathwise bounded and the trading strategy $\psi = \left(\psi^i\right)_{0 \leq i \leq d}$ is self-financing.

Proof. (i) Note first that $V_0^* \equiv V_0$. Then use (10.16) and (10.2) to obtain

$$\begin{aligned} dX_t^{*i} &= -re^{-rt} X_t^i \, dt + e^{-rt} \, dX_t^i && (10.21)\\ &= (a_i - r) \, X_t^{*i} \, dt + \sigma^i X_{t-}^{*i} \, dU_t^i, && (10.22) \end{aligned}$$

which shows in particular that X^* is a process of the same type as X itself with $X_0^* \equiv X_0$ obtained by replacing a^i by $a^i - r$ and, when including the discounted bank account $e^{-rt} e^{rt} \equiv 1$, replacing the interest rate r by 0. But

$$\begin{aligned} dV_t^* &= d\left(e^{-rt} V_t\right)\\ &= -rV_t^* \, dt + e^{-rt} \, dV_t\\ &= -rV_t^* \, dt + e^{-rt} \, dG_t \end{aligned}$$

using (10.13), and since by (10.21) we find

$$\begin{aligned} dG_t^* &= \sum_{i=1}^d \psi_t^i \, dX_t^{*i}\\ &= \sum_{i=1}^d \psi_t^i \left(-rX_t^{*i} \, dt + e^{-rt} \, dX_t^i\right)\\ &= -rV_t^* \, dt + e^{-rt} \, dG_t, \end{aligned}$$

$dV_t^* = dG_t^*$ follows.

(ii) The process G^* is cadlag and satisfies

$$\Delta G_t^* = \sum_{i=1}^d \psi_t^i \, \Delta X_t^{*i},$$

and this implies that

$$\psi_t^0 = V_0^* + G_{t-}^* - \sum_{i=1}^d \psi_t^i X_{t-}^{*i}$$

which shows that ψ^0 is predictable. It is pathwise bounded because the ψ^i for $i \geq 1$ are pathwise bounded by assumption, the X^{*i} are pathwise bounded automatically and G^* is then pathwise bounded as a simple consequence. With

$$V_t = \psi_t^0 e^{rt} + \sum_{i=1}^{d} \psi_t^i X_t^i \tag{10.23}$$

it then remains to show that, cf. (10.12),

$$dV_t = \psi_t^0 r e^{rt}\, dt + \sum_{i=1}^{d} \psi_t^i\, dX_t^i.$$

Defining $V_t^* = V_0^* + G_t^*$, (10.20) shows that

$$V_t = e^{rt} V_t^*$$

so

$$\begin{aligned}
dV_t &= rV_t\, dt + e^{rt}\, dG_t^* \\
&= r\left(\psi_t^0 e^{rt}\, dt + \sum_{i=1}^{d} \psi_t^i X_t^i\, dt\right) \\
&\quad + e^{rt} \sum_{i=1}^{d} \psi_t^i \left(-re^{-rt} X_t^i\, dt + e^{-rt}\, dX_t^i\right) \\
&= \psi_t^0 r e^{rt}\, dt + \sum_{i=1}^{d} \psi_t^i\, dX_t^i,
\end{aligned}$$

as desired using (10.23), (10.17) and (10.21) for the second but last equality. $\square$

We shall now in more detail discuss sufficient conditions for a trading strategy ψ to be self-financing. By (10.13) and since by the definition of G, G is cadlag, it follows that if ψ is self-financing, V is necessarily cadlag. As already remarked, in general V need not be cadlag, and to understand this better we resort to the representation provided by Proposition 4.2.1 of the $\mathcal{F}_t^X$-predictable processes ψ^i and write

$$\psi_t^i = f_{Z_{\langle t-\rangle}|x_0}^{i,\langle t-\rangle}(t) \tag{10.24}$$

where $Z_n = (T_1, \ldots, T_n; Y_1, \ldots, Y_n)$; (we retain as marks the jump sizes for U and do not use the $\widetilde{Y}_n = X_{T_n}$ appearing in Proposition 10.1.1). Assuming that all $t \mapsto f_{z_n|x_0}^{i,(n)}(t)$ are continuous, it is clear that for every $n \geq 1$,

$$V_{T_n-} = f^{0,(n-1)}_{Z_{n-1}|x_0}(T_n)\, e^{rT_n} + \sum_{i=1}^{d} f^{i,(n-1)}_{Z_{n-1}|x_0}(T_n)\, X^i_{T_n-},$$

$$V_{T_n} = f^{0,(n-1)}_{Z_{n-1}|x_0}(T_n)\, e^{rT_n} + \sum_{i=1}^{d} f^{i,(n-1)}_{Z_{n-1}|x_0}(T_n)\, X^i_{T_n},$$

$$V_{T_n+} = f^{0,(n)}_{Z_n|x_0}(T_n)\, e^{rT_n} + \sum_{i=1}^{d} f^{i,(n)}_{Z_n|x_0}(T_n)\, X^i_{T_n}$$

where $V_{T_n+} := \lim_{t\downarrow T_n, t>T_n} V_t$. In general all three values may differ, but V is cadlag at T_n iff the two last values agree. With all $f^{i,(n)}_{z_n|x_0}$ continuous, V is continuous on all the open intervals $]T_{n-1}, T_n[$; hence V is cadlag provided it is cadlag at all T_n and using (10.4) to obtain the value of X_{T_n} on each of the sets $(Z_n = z_n)$ and combining the preceding discussion with (10.14), we obtain the following sufficient condition for a trading strategy to be self-financing:

Proposition 10.2.2 *Let ψ be a trading strategy and consider its representation (10.24). Assume that ψ is smooth in the sense that for all n, i and z_n, $t \mapsto f^{i,(n)}_{z_n|x_0}(t)$ is differentiable on $[t_n, \infty[$. Then ψ is self-financing provided for every $0 < t_1 < \cdots < t_n$, and κ-almost all $y_1, \ldots, y_n \in \mathbb{R}^d_{\setminus 0}$ it holds for every $n \geq 0$, $z_n = (t_1, \ldots, t_n; y_1, \ldots, y_n)$ that*

$$\left(D_t f^{0,(n)}_{z_n|x_0}(t)\right) e^{rt} + \sum_{i=1}^{d} \left(D_t f^{i,(n)}_{z_n|x_0}(t)\right) e^{a^i t} \left(\prod_{k=1}^{n} \left(1 + \sigma^i y^i_k\right)\right) x^i_0 = 0 \quad (t \geq t_n) \tag{10.25}$$

and, for $n \geq 1$,

$$\begin{aligned} & f^{0,(n-1)}_{z_{n-1}|x_0}(t_n)\, e^{rt_n} + \sum_{i=1}^{d} f^{i,(n-1)}_{z_{n-1}|x_0}(t_n)\, e^{a^i t_n} \left(\prod_{k=1}^{n} \left(1 + \sigma^i y^i_k\right)\right) x^i_0 \\ &= f^{0,(n)}_{z_n|x_0}(t_n)\, e^{rt_n} + \sum_{i=1}^{d} f^{i,(n)}_{z_n|x_0}(t_n)\, e^{a^i t_n} \left(\prod_{k=1}^{n} \left(1 + \sigma^i y^i_k\right)\right) x^i_0 . \end{aligned} \tag{10.26}$$

At time T_n, $\psi^i_{T_n} = f^{i,(n-1)}_{Z_{n-1}|x_0}(T_n)$, while using the values just after T_n, $\psi^i_{T_n+} = f^{i,(n)}_{Z_n|x_0}(T_n)$. If the two values are different, the interpretation is that a (discontinuous) trading in asset i has taken place immediately after time T_n. Note that with the $f^{i,(n)}_{z_n}|x_0$ continuous, the predictable processes ψ^i are in fact left-continuous with limits from the right.

It is quite easy to find smooth trading strategies that satisfy (10.25) and (10.26): if $V_0 \equiv v_0$ is known and the $f^{i,(0)}_{|x_0}$ for $i \geq 1$ are given, use (10.25) to solve uniquely for $f^{0,(0)}_{|x_0}$ subject to the boundary condition

$$f^{0,(0)}_{|x_0}(0) + \sum_{i=1}^{d} f^{i,(0)}_{|x_0}(0)\, x^i_0 = v_0.$$

Next, assuming that the $f^{i,(1)}_{z_1|x_0}$ are given for $i \geq 1$, use (10.26) to determine the initial value $f^{0,(1)}_{z_1|x_0}(t_1)$ of $f^{0,(1)}_{z_1|x_0}$ and then use (10.25) to determine the function uniquely. Proceeding, it is clear that given e.g., v_0 and all $f^{i,(n)}_{z_n|x_0}$ for $1 \leq i \leq d$, all the functions $f^{0,(n)}_{z_n|x_0}$ are determined uniquely by (10.26) and (10.25), resulting in some self-financing strategy.

Example 10.2.1 The following self-financing strategies will be important later: suppose that $V_0 \equiv 0$ and that the $f^{i,(0)}_{|x_0}$ for $i \geq 1$ are continuously differentiable, and that immediately after time T_1 all the risky assets are sold and the proceeds put into the bank account with no more trading taking place. The choice of $f^{0,(0)}_{|x_0}$ is then forced by (10.25),

$$\begin{aligned} f^{0,(0)}_{|x_0}(t) &= -\sum_{i=1}^{d} x^i_0 f^{i,(0)}_{|x_0}(0) - \sum_{i=1}^{d} x^i_0 \int_0^t e^{(a_i-r)s} D_s f^{i,(0)}_{|x_0}(s)\, ds \\ &= \sum_{i=1}^{d} x^i_0 \Big((a_i - r)\, h^i_{|x_0}(t) - D_t h^i_{|x_0}(t) \Big) \end{aligned} \tag{10.27}$$

by partial integration, where

$$h^i_{|x_0}(t) = \int_0^t e^{(a_i-r)s} f^{i,(0)}_{|x_0}(s)\, ds.$$

For $n \geq 1$, $i \geq 1$, we demand that $f^{i,(n)}_{z_n|x_0} \equiv 0$ and then find from (10.25) that $D_t f^{0,(n)}_{z_n|x_0}(t) = 0$ so that by (10.26) for $n \geq 2$,

$$f^{0,(n)}_{z_n|x_0}(t) = f^{0,(n)}_{z_n|x_0}(t_n) = f^{0,(n-1)}_{z_{n-1}|x_0}(t_n) = f^{0,(n-1)}_{z_{n-1}|x_0}(t_{n-1}) = \cdots = f^{0,(1)}_{z_1|x_0}(t_1)$$

so that

$$f^{0,(n)}_{z_n|x_0}(t) = f^{0,(0)}_{|x_0}(t_1) + \sum_{i=1}^{d} f^{i,(0)}_{|x_0}(t_1)\, e^{(a_i-r)t_1} \Big(1 + \sigma^i y^i_1\Big) x^i_0 \quad (t \geq t_n, n \geq 1), \tag{10.28}$$

again using (10.26). Then (10.25) and (10.26) are satisfied for $n \geq 1$ and the strategy is self-financing.

Instead of specifying the $f^{i,(0)}_{|x_0}$ for $i \geq 1$, one may of course start with sufficiently smooth $h^i_{|x_0}$ and then define

$$f^{i,(0)}_{|x_0}(t) = e^{(r-a^i)t} D_t h^i_{|x_0}(t)$$

for $i \geq 1$. Note that the $h^i_{|x_0}$ must always satisfy that $h^i_{|x_0}(0) = 0$.

Proposition 10.2.2 deals only with smooth trading strategies. We shall not discuss the structure of general self-financing strategies but only note that if ψ is self-financing, by (10.13) and (10.11) not only must V_t be cadlag, but it must also be absolutely continuous as a function of t on all the intervals $\left[T_{n-1}, T_n\right[$. In particular, a discontinuity in some $f^{i,(n)}_{z_n|x_0}$ signifying a discontinuous trade in asset i at a time that is not one of the T_n, must be compensated by a simultaneous discontinuous trade in at least one of the other risky assets or the bank account.

10.3 Arbitrage and martingale measures

Consider a self-financing trading strategy $\psi = \left(\psi^i\right)_{0 \le i \le d}$ and let V and V^* denote the associated value process and discounted value process respectively. Then ψ is an *arbitrage opportunity* if

$$\exists t_0 > 0 : V_0 \equiv 0,\ \ V_{t_0} \ge 0 \text{ and } \mathbb{P}\left(V_{t_0} > 0\right) > 0. \tag{10.29}$$

Sometimes there are arbitrage opportunities that are *global* in the sense that

$$\forall t_0 > 0 : V_0 \equiv 0,\ \ V_{t_0} \ge 0 \text{ and } \mathbb{P}\left(V_{t_0} > 0\right) > 0. \tag{10.30}$$

Any realistic model used in finance must never allow arbitrage, and we shall now discuss when there is no arbitrage for the models given by (10.2).

It is very easy to give examples of models where arbitrage *is* possible: just assume that there is an asset $i_0 \ge 1$ such that the i_0th coordinate of the discounted price process $X_t^{*i_0} = e^{-rt} X_t^{i_0}$ is monotone in t and not constant. In particular, assume that there exists $i_0 \ge 1$ with $a^{i_0} \ge r$ and $\mathbb{P}\left(Y_1^{i_0} \ge 0\right) = 1$ and either $a^{i_0} > r$ or $\mathbb{P}\left(Y_1^{i_0} > 0\right) > 0$. From (10.4) it is clear that $X_t^{*i_0}$ is increasing in t, strictly increasing between jumps if $a^{i_0} > r$, while if $a_0^i = r$ from some (random) timepoint onwards, one has $X_t^{*i_0} > x_0^{i_0}$ since by the assumptions made, eventually a jump with $Y_n^{i_0} > 0$ will occur. The trivial trading strategy $\psi_t^{i_0} \equiv 1$, $\psi_t^0 \equiv -x_0^{i_0}$ and $\psi_t^i \equiv 0$ for $i \ne 0, i_0$ is obviously self-financing, $V_0 \equiv 0$ and

$$V_t^* = -x_0^{i_0} + X_t^{*i_0}$$

is increasing in t and for an arbitrary $t_0 > 0$, $V_{t_0}^* > 0$ if $a^{i_0} > r$, while if $a^{i_0} = r$,

$$\mathbb{P}\left(V_{t_0}^* > 0\right) \ge \mathbb{P}\left(T_1 \le t_0, Y_1^{i_0} > 0\right) = \left(1 - e^{-\lambda t_0}\right) \mathbb{P}\left(Y_0^{i_0} > 0\right) > 0,$$

hence there is arbitrage using the strategy ψ.

Similarly, if there is an $i_0 \ge 1$ such that $a^{i_0} \le r$ and $\mathbb{P}\left(Y_1^{i_0} \le 0\right) = 1$ and either $a^{i_0} < r$ or $\mathbb{P}\left(Y_1^{i_0} < 0\right) > 0$, then $X_t^{*i_0}$ is decreasing in t, eventually (possibly from the start) $< x_0^{i_0}$ and the self-financing strategy $\psi_t^{i_0} \equiv -1$, $\psi_t^0 \equiv x_0^{i_0}$ and all other

$\psi_t^i \equiv 0$ is an arbitrage opportunity at each $t_0 > 0$. As we shall see presently, for $d \geq 2$ these trivial examples of arbitrage are not the only ones that may appear.

The following discussion of arbitrage is based on the fundamental Definition 10.3.1 of risk-neutral measures. In order to formulate this, recall that X is assumed defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{F}^X = \mathcal{F}^\mu$ denote the sub σ-algebra of $\mathcal{F}$ generated by X or, equivalently, the RCM μ determined by the T_n and Y_n. The natural filtration to use is of course $\left(\mathcal{F}_t^X\right)$ which is the same as $\left(\mathcal{F}_t^\mu\right)$; cf. (10.8).

Call a probability $\widetilde{\mathbb{P}}$ on $\left(\Omega, \mathcal{F}^X\right)$ locally equivalent to $\mathbb{P}$ and write $\widetilde{\mathbb{P}} \sim_{\mathrm{loc}} \mathbb{P}$ if

$$\widetilde{\mathbb{P}} \ll_{\mathrm{loc}} \mathbb{P} \quad \text{and} \quad \mathbb{P} \ll_{\mathrm{loc}} \widetilde{\mathbb{P}}$$

where $\ll_{\mathrm{loc}}$ refers to local absolute continuity as discussed in 5.1, but here using the restrictions $\mathbb{P}_t$ and $\widetilde{\mathbb{P}}_t$ of $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ to $\mathcal{F}_t^X$ for every $t \in \mathbb{R}_0$. Thus $\mathbb{P}$ and $\widetilde{\mathbb{P}}$ are locally equivalent if $\widetilde{\mathbb{P}}_t \ll \mathbb{P}_t$ and $\mathbb{P}_t \ll \widetilde{\mathbb{P}}_t$ for all $t \in \mathbb{R}_0$.

Under $\mathbb{P}$, μ has κ-intensity process (λ^y) with all $\lambda_t^y \equiv \lambda$. From Theorem 5.1.1 and Remark 5.1.1 it then follows that if $\widetilde{\mathbb{P}} \ll_{\mathrm{loc}} \mathbb{P}$, under $\widetilde{\mathbb{P}}$ μ also has a κ-intensity process $\left(\widetilde{\lambda}^y\right)$, a fact used in the proof of Proposition 10.3.1 below.

Definition 10.3.1 A probability measure $\widetilde{\mathbb{P}}$ on $\left(\Omega, \mathcal{F}^X\right)$ is called an *equivalent martingale measure* for $\mathbb{P}$ or a *risk-neutral measure* if $\widetilde{\mathbb{P}} \sim_{\mathrm{loc}} \mathbb{P}$ and if under $\widetilde{\mathbb{P}}$ each coordinate X^{*i} for $1 \leq i \leq d$ of the discounted price process is an $\mathcal{F}_t^X$-martingale.

In the sequel we shall refer to $\mathbb{P}$ as the *true measure* and we shall denote by $\mathcal{PM}$ the possibly empty set of equivalent martingale measures. We shall first establish conditions for $\mathcal{PM} \neq \emptyset$, but rather than treating general X as in (10.2), we shall from now on assume that the probability distribution κ (which is the distribution under the true measure) of the jump sizes Y_n for U (see (10.1)) has *bounded support*, i.e., there exists $R > 0$ such that $\kappa\left([-R, R]^d\right) = 1$. With this assumption we shall show (with ^{i} as always referring to an index, not a power):

Proposition 10.3.1 *There exists at least one $\widetilde{\mathbb{P}} \in \mathcal{PM}$ iff there exists a κ-integrable function $\widetilde{\lambda}(y) > 0$ such that*

$$\sigma^i \int_{\mathbb{R}^d_{\backslash 0}} y^i \widetilde{\lambda}(y)\, \kappa(dy) = r - a^i \quad (1 \leq i \leq d). \tag{10.31}$$

If this condition holds, one $\widetilde{\mathbb{P}} \in \mathcal{PM}$ is defined by letting U given by (10.1) under $\widetilde{\mathbb{P}}$ be the compound Poisson process with $\overline{N}$ homogeneous Poisson at rate $\widetilde{\lambda}$ and the Y_n iid and independent of $\overline{N}$ with distribution $\widetilde{\kappa}$, where

$$\widetilde{\lambda} = \int_{\mathbb{R}^d_{\backslash 0}} \widetilde{\lambda}(y)\, \kappa(dy), \quad \widetilde{\kappa}(dy) = \frac{1}{\widetilde{\lambda}} \widetilde{\lambda}(y)\, \kappa(dy). \tag{10.32}$$

Remark 10.3.1 (10.31) is an example of what in finance is called a *drift equation.*

Proof. Initially it should be noted that the price process X is given by (10.2) under $\mathbb{P}$ as well as under any $\widetilde{\mathbb{P}} \in \mathcal{PM}$ — it is only the distribution of the driving process

U that changes when switching from $\mathbb{P}$ to $\widetilde{\mathbb{P}}$. And of course (10.16) does not change either under the switch.

Now assume first that (10.31) holds and define $\widetilde{\mathbb{P}}$ by requiring that under $\widetilde{\mathbb{P}}$, U has the properties given in the last statement of the proposition. Then the basic martingale measure for μ under $\widetilde{\mathbb{P}}$ is

$$\widetilde{M}(dt, dy) = \mu(dt, dy) - \widetilde{\lambda}(y)\, dt\, \kappa(dy)$$

and using (10.22) and (10.31) we obtain

$$\begin{aligned} X_t^{*i} &= x_0^i + \int_0^t \left(a^i - r\right) X_s^{*i}\, ds + \sigma^i \int_{]0,t]\times\mathbb{R}^d_{\backslash 0}} X_{s-}^{*i} y^i\, \mu(ds, dy) \\ &= x_0^i + \sigma^i \int_{]0,t]\times\mathbb{R}^d_{\backslash 0}} X_{s-}^{*i} y^i\, \widetilde{M}(ds, dy). \end{aligned}$$

That X^{*i} is a $\widetilde{\mathbb{P}}$-martingale will follow therefore from Theorem 4.6.1(iii) if we show that (writing $\widetilde{\mathbb{E}}$ for the expectation $\mathbb{E}_{\widetilde{\mathbb{P}}}$ under $\widetilde{\mathbb{P}}$)

$$\widetilde{\mathbb{E}} \int_{]0,t]\times\mathbb{R}^d_{\backslash 0}} X_{s-}^{*i} \left|y^i\right| \mu(ds, dy) = \widetilde{\mathbb{E}} \int_0^t ds\, X_s^{*i} \int_{\mathbb{R}^d_{\backslash 0}} \left|y^i\right| \widetilde{\lambda}(y)\, \kappa(dy) < \infty.$$

Because κ is assumed to have bounded support and $\widetilde{\lambda}(y)$ is κ-integrable, the inner integral is finite and we only need

$$\widetilde{\mathbb{E}} \int_0^t X_s^{*i}\, ds = \int_0^t \widetilde{\mathbb{E}} X_s^{*i}\, ds < \infty. \tag{10.33}$$

But by (10.4), since all $\left|Y_n^i\right| \le R$ for some constant $R > 0$,

$$X_s^{*i} \le e^{(a^i - r)s} \left(1 + \sigma^i R\right)^{\overline{N}_s} \tag{10.34}$$

and since $\overline{N}_s$ under $\widetilde{\mathbb{P}}$ is Poisson $\widetilde{\lambda}$, (10.33) follows.

Suppose now that $\widetilde{\mathbb{P}} \in \mathcal{PM}$ with predictable κ-intensity process $\left(\widetilde{\lambda}^y\right)$. Thus the basic martingale measure under $\widetilde{\mathbb{P}}$ is

$$\widetilde{M}(dt, dy) = \mu(dt, dy) - \widetilde{\lambda}_t^y\, dt\, \kappa(dy)$$

and using (10.22) we get for the $\widetilde{\mathbb{P}}$-martingale X^{*i} the expression

$$\begin{aligned} X_t^{*i} &= x_0^i + \int_0^t \left(a^i - r\right) X_s^{*i}\, ds + \int_{]0,t]\times\mathbb{R}^d_{\backslash 0}} \sigma^i X_{s-}^{*i} y^i\, \mu(ds, dy) \\ &= x_0^i + \int_0^t \left(a^i - r + \sigma^i \int_{\mathbb{R}^d_{\backslash 0}} \widetilde{\lambda}_s^y y^i\, \kappa(dy)\right) X_s^{*i}\, ds \\ &\quad + \int_{]0,t]\times\mathbb{R}^d_{\backslash 0}} \sigma^i X_{s-}^{*i} y^i\, \widetilde{M}(ds, dy). \end{aligned} \tag{10.35}$$

By Theorem 4.6.1(iii), the last term is a local $\widetilde{\mathbb{P}}$-martingale (T_n) provided

$$\widetilde{\mathbb{E}} \int_{]0,T_n \wedge t]} X^{*i}_{s-} \left| y^i \right| \mu(ds, dy) < \infty \tag{10.36}$$

for all t and n. By (10.34)

$$X^{*i}_{s-} \leq e^{|a^i - r| t} \left(1 + \sigma^i R\right)^{n-1} x^i_0 = K_n$$

say, for $s \leq T_n \wedge t$. Hence the expectation in (10.36) is

$$\leq K_n R \mathbb{E}_{\widetilde{\mathbb{P}}} \overline{N}_{T_n \wedge t} \leq K_n R n < \infty,$$

so the last term in (10.35) is a local $\widetilde{\mathbb{P}}$-martingale. But from (10.35) and the assumption $\widetilde{\mathbb{P}} \in \mathcal{PM}$ it then follows that the continuous, hence predictable process

$$\int_0^t \left(a^i - r + \sigma^i \int_{\mathbb{R}^d_{\backslash 0}} \widetilde{\lambda}^y_s y^i \kappa(dy) \right) X^{*i}_s \, ds$$

is a local $\widetilde{\mathbb{P}}$-martingale. By Proposition 4.5.1 this process is therefore constant and since $X^{*i}_0 \equiv x^i_0$, it is identically equal to 0, i.e., with $\widetilde{\mathbb{P}}$-probability 1

$$\int_0^t \left(a^i - r + \sigma^i \int_{\mathbb{R}^d_{\backslash 0}} \widetilde{\lambda}^y_s y^i \kappa(dy) \right) X^{*i}_s \, ds = 0$$

simultaneously for all t. Since $X^{*i}_s > 0$, (10.31) follows. □

Example 10.3.1 It is of particular interest to look for equivalent martingale measures to the true measure $\mathbb{P}$ in the case where κ is discrete. Suppose that κ has $p \geq 1$ atoms $\alpha_1, \ldots, \alpha_p \in \mathbb{R}^d_{\backslash 0}$ and write $\alpha_q = \left(\alpha^i_q\right)_{1 \leq i \leq d}$ for $1 \leq q \leq p$. With $\kappa\left(\alpha_q\right) = \kappa\left(\{\alpha_q\}\right) > 0$, by (10.31) the existence of an $\widetilde{\mathbb{P}} \in \mathcal{PM}$ is equivalent to the existence of real numbers $\ell_q > 0$ such that

$$\sigma^i \sum_{q=1}^{p} \alpha^i_q \kappa\left(\alpha_q\right) \ell_q = r - a^i \quad (1 \leq i \leq d);$$

in other words, with A the $d \times p$-matrix $\left(\sigma^i \alpha^i_q \kappa\left(\alpha_q\right)\right)_{1 \leq i \leq d, 1 \leq q \leq p}$ and v the vector $\left(r - a^i\right)_{1 \leq i \leq d}$, $\mathcal{PM} \neq \emptyset$ iff the linear equation

$$A\ell = v \tag{10.37}$$

has a solution vector ℓ with all $\ell_q > 0$.

If $\mathrm{rank}(A) < p$ (in particular if $d < p$), (10.37) may not have any solutions at all in which case $\mathcal{PM} = \emptyset$. If $d = p$ and A is non-singular, (10.37) has the unique solution $\ell = A^{-1}v$ and $\mathcal{PM} \neq \emptyset$ precisely when for this particular ℓ, all $\ell_q > 0$. If $d > p$ and $\mathrm{rank}(A) \geq p$, (10.37) has infinitely many solutions ℓ and $\mathcal{PM} \neq \emptyset$ iff one of these satisfies $\ell_q > 0$ for all q.

It is quite instructive to consider the case $d = p = 1$: then $\kappa = \varepsilon_\alpha$ for some $\alpha \in \mathbb{R}_{\backslash 0}$ and (10.37) has the unique solution (with the unique index i ignored in the notation)

$$\ell = \frac{r - a}{\sigma\alpha}$$

which is > 0 iff $r \neq a$ and $r - a$ and α have the same sign. Thus $\mathcal{PM} = \emptyset$ iff either $r = a$ or $r \neq a$ with $(r - a)\,\alpha < 0$, which means precisely that $t \mapsto X_t^*$ is monotone increasing ($a \geq r$, $\alpha > 0$) or monotone decreasing ($a \leq r$, $\alpha < 0$), a situation with a trivial form of arbitrage as discussed p. 257.

Exercise 10.3.1 Show directly for general κ, by studying the drift equation (10.31), that if some $X_t^{*i_0}$ is monotone and not constant as described p. 257, then the equation has no solution $\widetilde{\lambda}\,(y) > 0$.

We shall call any $\widetilde{\mathbb{P}} \in \mathcal{PM}$ for which U is compound Poisson as described in Proposition 10.3.1 for a *natural* equivalent martingale measure. So if $\mathcal{PM} \neq \emptyset$, there is always one $\widetilde{\mathbb{P}} \in \mathcal{PM}$ that is natural and if the equivalent martingale measure is unique, it is automatically natural. If there are many equivalent natural martingale measures there are also many with a random intensity process $\left(\widetilde{\lambda}^y\right)$: just switch from one (deterministic) solution to (10.31) to another in an ω-dependent fashion, e.g., define $\widetilde{\lambda}_t^y = \widetilde{\lambda}^{(n)}\,(y)$ for $T_n < t \leq T_{n+1}$ with each $\widetilde{\lambda}^{(n)}\,(y)$ a solution to (10.31).

One of the main results in mathematical finance states that 'no arbitrage' is the same as 'the existence of an equivalent martingale measure'. To formulate our version of this result and in order to avoid various technicalities, we shall not only assume that the Y_n are bounded (κ has bounded support as assumed above) but also restrict ourselves to considering only *bounded* trading strategies ψ, i.e., self-financing ψ such that

$$\sup_{s \leq t, \omega \in \Omega} \left|\psi_s^i\,(\omega)\right| < \infty \quad (t \geq 0, 1 \leq i \leq d)\,. \tag{10.38}$$

A bounded arbitrage opportunity is then an arbitrage opportunity generated by a bounded, self-financing strategy. Note that if ψ is bounded, it is in particular pathwise bounded; cf. (10.18).

Theorem 10.3.2 (i) *There are no bounded arbitrage opportunities if and only if there exists an equivalent martingale measure for the true measure $\mathbb{P}$: $\mathcal{PM} \neq \emptyset$.*
(ii) *If $\mathcal{PM} = \emptyset$, then there exists a global arbitrage opportunity.*

Proof. Suppose first that $\mathcal{PM} \neq \emptyset$ and suppose that ψ is a bounded, self-financing strategy with $V_0^* \equiv 0$. Choose a $\widetilde{\mathbb{P}} \in \mathcal{PM}$ which is natural, determined using a particular solution $\widetilde{\lambda}\,(y)$ to (10.31). Then

$$V_t^* = \int_{]0,t]} \sum_{i=1}^{d} \psi_s^i \, dX_s^{*i}$$
$$= \int_{]0,t]\times\mathbb{R}^d_{\backslash 0}} \sum_{i=1}^{d} \psi_s^i \sigma^i X_{s-}^{*i} y^i \, \widetilde{M}(ds, dy),$$

cf. (10.35), with $\widetilde{M}(ds, dy) = \widetilde{\lambda}(y)\, ds\, \kappa(dy)$ the martingale measure for μ under $\widetilde{\mathbb{P}}$. With the ψ^i uniformly bounded on $[0, t]$ and because κ has bounded support (all $\left|Y_n^i\right| \le R$) and

$$\widetilde{\mathbb{E}} \sup_{s\le t} X_s^{*i} \le \widetilde{\mathbb{E}} e^{\left|a^i - r\right| t} \left(1 + \sigma^i R\right)^{\overline{N}_t} x_0^i < \infty$$

(using that $\overline{N}_t$ is Poisson under $\widetilde{\mathbb{P}}$ for the last inequality), it follows from Theorem 4.6.1(iii) that V^* is a $\widetilde{\mathbb{P}}$-martingale and consequently

$$\widetilde{\mathbb{E}} V_t^* = \widetilde{\mathbb{E}} V_0^* = 0. \tag{10.39}$$

If a bounded arbitrage opportunity ψ existed, for some $t_0 > 0$ we would have $V_{t_0}^* \ge 0$ $\mathbb{P}$-a.s. and $\mathbb{P}\left(V_{t_0}^* > 0\right) > 0$. Since $\widetilde{\mathbb{P}} \sim_{\text{loc}} \mathbb{P}$ then also $V_{t_0}^* \ge 0$ $\widetilde{\mathbb{P}}$-a.s. and $\widetilde{\mathbb{P}}\left(V_{t_0}^* > 0\right) > 0$ which contradicts (10.39), and we conclude that there are no bounded arbitrage opportunities.

For the converse and the proof of (ii), assume that $\mathcal{PM} = \emptyset$. The task is then to produce a global bounded arbitrage opportunity (see (10.30)), and for this we need in fact only consider the simple self-financing strategies from Example 10.2.1. Therefore let $\psi_t^i = f_{Z_{\langle t-\rangle}|x_0}^{i,\langle t-\rangle}(t)$ where for $i \ge 1$, $n \ge 1$, $f_{z_n|x_0}^{i,(n)} \equiv 0$, while with the $f_{|x_0}^{i,(0)}$ for $i \ge 1$ given and continuous, cf. (10.27),

$$f_{|x_0}^{0,(0)}(t) = \sum_{i=1}^{d} x_0^i \left(\left(a^i - r\right) h_{|x_0}^i(t) - D_t h_{|x_0}^i(t)\right)$$

where for $i \ge 1$,

$$h_{|x_0}^i(t) = \int_0^t e^{\left(a^i - r\right)s} f_{|x_0}^{i,(0)}(s)\, ds.$$

Finally, for $n \ge 1$, $f_{z_n|x_0}^{0,(n)}$ is as in (10.28).

As shown in Example 10.2.1, ψ is self-financing, $V_0^* \equiv 0$ and with trading stopping immediately after T_1, using only that the $f_{|x_0}^{i,(0)}$ are continuous and Y_1 is bounded, it is verified directly that ψ is a bounded strategy.

Now, for $t < T_1$,

$$V_t^* = f_{|x_0}^{0,(0)}(t) + \sum_{i=1}^{d} f_{|x_0}^{i,(0)}(t)\, e^{\left(a^i - r\right)t} x_0^i$$
$$= \sum_{i=1}^{d} x_0^i \left(a^i - r\right) h_{|x_0}^i(t)$$

which defines a non-negative increasing function of t provided

$$\sum_{i=1}^{d} x_0^i \left(a^i - r\right) D_t h^i_{|x_0}(t) \geq 0 \quad (t \geq 0). \tag{10.40}$$

Also,

$$\begin{aligned} V^*_{T_1} &= f^{0,(0)}_{|x_0}(T_1) + \sum_{i=1}^{d} f^{i,(0)}_{|x_0}(T_1)\, e^{(a^i - r)T_1}\left(1 + \sigma^i Y_1^i\right) x_0^i \\ &= \sum_{i=1}^{d} x_0^i \left(a^i - r\right) h^i_{|x_0}(T_1) + \sum_{i=1}^{d} x_0^i \sigma^i Y_1^i D_t h^i_{|x_0}(T_1) \end{aligned}$$

and $V^*_t = V^*_{T_1}$ for $t > T_1$. It follows that ψ is a global arbitrage opportunity if (10.40) holds and

$$\sum_{i=1}^{d} x_0^i \sigma^i Y_1^i D_t h^i_{|x_0}(T_1) \geq 0$$

with a sharp inequality somewhere. Specializing further to $D_t h^i_{|x_0}(t) \equiv c^i(x_0)$ (i.e., $h^i_{|x_0}(t) = tc^i(x_0)$) we conclude that there exists a bounded arbitrage opportunity if constants δ^i can be found such that

$$\sum_{i=1}^{d} \left(a^i - r\right) \delta^i \geq 0, \qquad \sum_{i=1}^{d} \sigma^i Y_1^i \delta^i \geq 0 \quad \mathbb{P}\text{-a.s.} \tag{10.41}$$

with either the first inequality sharp or

$$\mathbb{P}\left(\sum_{i=1}^{d} \sigma^i Y_1^i \delta^i > 0\right) > 0. \tag{10.42}$$

The proof is therefore concluded by showing that such δ^i exist when knowing that $\mathcal{PM} = \emptyset$, i.e., (see Proposition 10.3.1) that there is no κ-integrable function $\widetilde{\lambda}(y) > 0$ such that

$$\sigma^i \int_{\mathbb{R}^d_{\backslash 0}} y^i \widetilde{\lambda}(y)\, \kappa(dy) = r - a^i \quad (1 \leq i \leq d). \tag{10.43}$$

The proof is quite long and is performed in two main steps: κ finite support or not. Suppose first that κ has finite support with $p \geq 1$ atoms $\alpha_1, \ldots, \alpha_p$ as in Example 10.3.1. As there, writing

$$A = \left(\sigma^i \alpha_q^i \kappa\left(\alpha_q\right)\right)_{1 \leq i \leq d, 1 \leq q \leq p}$$

and v for the column vector $\left(r - a^i\right)_{1 \leq i \leq d}$, the assumption $\mathcal{PM} = \emptyset$ is then that the equation

$$A\ell = v \tag{10.44}$$

has no solution ℓ with all $\ell_q > 0$, and since the only possible values for Y_1^i are the α_q^i for $1 \le q \le p$, we want to show that there exists constants δ^i such that

$$\sum_{i=1}^d \delta^i \left(a^i - r\right) \ge 0, \qquad \sum_{i=1}^d \delta^i A_q^i \ge 0 \quad (1 \le q \le p)$$

with a sharp inequality somewhere. Equivalently, with $A_{|q}$ denoting the qth column of A and writing $\langle \cdot, \cdot \rangle$ for the inner product on $\mathbb{R}^d$, $\delta = \left(\delta^i\right) \in \mathbb{R}^d$ must be found so that

$$\langle \delta, v \rangle \le 0, \qquad \left\langle \delta, A_{|q} \right\rangle \ge 0 \quad (1 \le q \le p) \tag{10.45}$$

with at least one sharp inequality.

Consider first the case where rank $(A) = d$ (in particular $p \ge d$). Let $\varepsilon > 0$ and define

$$S_\varepsilon = \left\{ Aw : w \in \mathbb{R}^{p \times 1}, \text{ all } w_q \ge \varepsilon \right\}, \tag{10.46}$$

which is a closed convex subset of $\mathbb{R}^d$. By the assumption $\mathcal{PM} = \emptyset$, $v \notin S_\varepsilon$ (for any ε) and hence there exists a separating hyperplane (if $d = 1$ a single point separating the point v from the closed interval S_ε),

$$H_\varepsilon = \left\{ u \in \mathbb{R}^d : \langle \delta_\varepsilon, u \rangle = \gamma_\varepsilon \right\}$$

for some $\delta_\varepsilon \in \mathbb{R}^d_{\backslash 0}$, $\gamma_\varepsilon \in \mathbb{R}$ such that either

$$\langle \delta_\varepsilon, v \rangle > \gamma_\varepsilon, \quad \langle \delta_\varepsilon, Aw \rangle < \gamma_\varepsilon$$

when all $w_q \ge \varepsilon$ or

$$\langle \delta_\varepsilon, v \rangle < \gamma_\varepsilon, \quad \langle \delta_\varepsilon, Aw \rangle > \gamma_\varepsilon \tag{10.47}$$

when all $w_q \ge \varepsilon$. Since H_ε may be represented by any pair $(c\delta_\varepsilon, c\gamma_\varepsilon)$ with $c \ne 0$ arbitrary, we may and shall assume that we are always in the second case (10.47) and furthermore that $\|\delta_\varepsilon\| + |\gamma_\varepsilon| = 1$ (of course $\|u\|^2 = \langle u, u \rangle$).

The second inequality in (10.47) states that

$$\sum_{q=1}^p \left\langle \delta_\varepsilon, A_{|q} \right\rangle w_q > \gamma_\varepsilon \tag{10.48}$$

whenever all $w_q \ge \varepsilon$, and this forces

$$\left\langle \delta_\varepsilon, A_{|q} \right\rangle \ge 0 \tag{10.49}$$

for all q: if for some q_0, $\langle \delta_\varepsilon, A_{|q_0} \rangle < 0$ just choose $w_{q_0} \geq \varepsilon$ large and $w_q = \varepsilon$ for $q \neq q_0$ to obtain a contradiction to (10.48).

Let now $\varepsilon_n \downarrow 0$ as $n \uparrow \infty$ with all $\varepsilon_n > 0$ and determine δ_{ε_n}, γ_{ε_n} as above with (10.47) valid with ε replaced by ε_n and $\left\|\delta_{\varepsilon_n}\right\| + \left|\gamma_{\varepsilon_n}\right| = 1$. By compactness and by choosing a suitable subsequence, we may assume that

$$\left(\delta_{\varepsilon_n}, \gamma_{\varepsilon_n}\right) \to (\delta, \gamma) \tag{10.50}$$

as $n \to \infty$, where $\delta \in \mathbb{R}^{d \times 1}$, $\gamma \in \mathbb{R}$ and

$$\|\delta\| + |\gamma| = \lim_{n \to \infty} \left(\left\|\delta_{\varepsilon_n}\right\| + \left|\gamma_{\varepsilon_n}\right|\right) = 1. \tag{10.51}$$

(10.49) implies that

$$\left\langle \delta, A_{|q} \right\rangle = \lim_{n \to \infty} \left\langle \delta_{\varepsilon_n}, A_{|q} \right\rangle \geq 0. \tag{10.52}$$

Also, using (10.48) with all $w_q = \varepsilon_n$ we find that

$$\varepsilon_n \sum_{q=1}^{p} \left\langle \delta_{\varepsilon_n}, A_{|q} \right\rangle > \gamma_{\varepsilon_n},$$

implying when $n \to \infty$ that $\gamma \leq 0$. Since the first inequality in (10.47) gives that

$$\langle \delta, v \rangle = \lim_{n \to \infty} \left\langle \delta_{\varepsilon_n}, v \right\rangle \leq \lim_{n \to \infty} \gamma_{\varepsilon_n} = \gamma \leq 0, \tag{10.53}$$

we have shown that the vector δ, determined by (10.50), satisfies the inequalities (10.45) and it only remains to show that one of them is sharp. There is nothing to show if $\langle \delta, v \rangle < 0$, so assume that $\langle \delta, v \rangle = 0$. In this case (10.53) forces $\gamma = 0$ and (10.51) then shows that $\delta \neq 0$. Since we assumed that $\operatorname{rank}(A) = d$, δ cannot be orthogonal to all the columns in A and therefore (cf. (10.52)) $\left\langle \delta, A_{|q} \right\rangle > 0$ for some q, as desired.

Consider next the case where $d_A = \operatorname{rank}(A) < d$. In particular $d \geq 2$ since always $\operatorname{rank}(A) \geq 1$ (all the elements of A are $\neq 0$). The argument above still applies, but the problem is that now the vector δ found may be orthogonal to v and all the $A_{|q}$, which means that there is no sharp inequality in (10.45). We shall here treat two subcases and first assume that $v \in L_A$, the linear subspace spanned by the vectors $A_{|q}$. But then the argument from above may be copied to produce for any $\varepsilon > 0$ a vector $\delta_{\varepsilon,A} \in L_A$ and a scalar $\gamma_{\varepsilon,A}$ such that

$$H_{\varepsilon,A} = \left\{u_A \in L_A : \left\langle \delta_{\varepsilon,A}, u_A \right\rangle = \gamma_{\varepsilon,A}\right\}$$

separates $v \in L_A$ from $S_\varepsilon \subset L_A$ (S_ε as defined in (10.46)). Pursuing the reasoning from before one finds $\delta_A \in L_A$ and $\gamma_A \in \mathbb{R}$ with $\|\delta_A\| + |\gamma_A| = 1$ such that

$$\langle \delta_A, v \rangle \leq \gamma_A \leq 0 \leq \left\langle \delta_A, A_{|q} \right\rangle$$

for all q. Since clearly $\delta_A \neq 0$, it is impossible for δ_A to be orthogonal to v and all the $A_{|q}$; hence (10.45) holds with at least one sharp inequality.

Finally, if $d_A < d$ and v does not belong to the linear subspace L_A, (10.44) has no solutions at all and it is possible to find $\delta \in \mathbb{R}^d_{\setminus 0}$ such that $\delta \perp A_{|q}$ for all q, but δ is not orthogonal to v. Replacing δ by $-\delta$ if necessary we obtain that

$$\langle \delta, v \rangle < 0, \quad \left\langle \delta, A_{|q} \right\rangle = 0$$

for all q, i.e., (10.45) holds with the first inequality sharp.

We have now shown that if κ has finite support, then there exists a vector δ such that (10.41) and (10.42) hold when $\mathcal{PM} = \emptyset$. So suppose now that κ does not have finite support, and to avoid an annoying special case briefly discussed at the end of the proof, that the support, $\mathrm{supp}\,(\kappa)$, spans all of $\mathbb{R}^d$. (Formally, $\mathrm{supp}\,(\kappa)$ consists of all $y \in \mathbb{R}^d_{\setminus 0}$ such that $\kappa\,(O) > 0$ for any open set O containing y).

Because (10.43) has no solution $\widetilde{\lambda}\,(y) > 0$, it holds in particular for any finite partitioning $\mathbb{R}^d_{\setminus 0} = \bigcup_{q=1}^{p} B_q$ of $\mathbb{R}^d_{\setminus 0}$ into disjoint sets B_q that the equation

$$\sigma^i \sum_{q=1}^{p} \int_{B_q} y^i\,\kappa\,(dy)\;\ell_q = v^i \quad (1 \le i \le d) \tag{10.54}$$

has no solution ℓ with all $\ell_q > 0$. Defining

$$A^i_q = \sigma^i \int_{B_q} y^i\,\kappa\,(dy)\,,$$

(10.54) becomes $A\ell = v$ and we are back to the situation with κ having finite support. Therefore there exists $\delta \in \mathbb{R}^d_{\setminus 0}$ such that

$$\langle \delta, v \rangle \le 0, \qquad \left\langle \delta, A_{|q} \right\rangle \ge 0 \quad (1 \le q \le p) \tag{10.55}$$

with sharp inequality somewhere, at least if all $\kappa\left(B_q\right) > 0$. Now it is unproblematic, when using the assumption that $\mathrm{supp}\,(\kappa)$ spans all of $\mathbb{R}^d$, to define a sequence of finite partitionings $\mathbb{R}^d_{\setminus 0} = \bigcup_{q=1}^{p_n} B_{q,n}$ for $n \ge 1$ such that

(i) for $n \ge 2$, each $B_{q,n}$ is a subset of some $B_{q',n-1}$;
(ii) all $\kappa\left(B_{q,n}\right) > 0$ and all $A^i_{q,n} = \sigma^i \int_{B_{q,n}} y^i\,\kappa\,(dy) \ne 0$;
(iii) for all $y_0 = \left(y^i_0\right) \in \mathrm{supp}\,(\kappa)$, if $B_{q_0,n}$ is that member of the n'th partitioning containing y_0, it holds that $\bigcap_n B_{q_0,n} \cap \mathrm{supp}\,(\kappa) = \{y_0\}$.

(If $d = 1$ one would e.g., consider the intervals $\left]\frac{q-1}{2^n}, \frac{q}{2^n}\right]$ and $\left[-\frac{q}{2^n}, -\frac{q-1}{2^n}\right[$ for $1 \le q \le n2^n$ and define the partitioning by attaching any interval I with $\kappa\,(I) = 0$ to one with $\kappa\,(I) > 0$. Since within each interval $y = y^1$ always has the same sign it is clear that in particular $A^i_{q,n} \ne 0$).

For each n, we can find $\delta_n \ne 0$ such that (10.55) holds with δ replaced by δ_n and we may normalize and assume that $\|\delta_n\| = 1$ for all n. Passing to a subsequence if necessary, we may also assume that $\delta_n \to \delta$ as $n \to \infty$, with obviously $\|\delta\| = 1$. Also

$$\langle \delta, v \rangle = \lim_{n\to\infty} \langle \delta_n, v \rangle \le 0. \tag{10.56}$$

For any $y_0 \in \operatorname{supp}(\kappa)$, determining $B_{q_0,n}$ as in (iii) above, since

$$y_0^i = \lim_{n\to\infty} \frac{1}{\kappa\left(B_{q_0,n}\right)} \int_{B_{q_0,n}} y^i \,\kappa\,(dy)$$

it follows that

$$\begin{aligned}
\sum_{i=1}^d \sigma^i y_0^i \delta^i &= \lim_{n\to\infty} \sum_{i=1}^d \frac{\sigma^i}{\kappa\left(B_{q_0,n}\right)} \int_{B_{q_0,n}} y^i \,\kappa\,(dy)\ \delta_n^i \\
&= \lim_{n\to\infty} \frac{1}{\kappa\left(B_{q_0,n}\right)} \left\langle \delta_n, A_{|q_0,n} \right\rangle \\
&\ge 0.
\end{aligned}$$

Combining this with (10.56) gives (10.41), and (10.42) will then hold because $\operatorname{supp}(\kappa)$ spans $\mathbb{R}^d$ and it is therefore possible to find $y_0 \in \operatorname{supp}(\kappa)$ such that $\sum_{i=1}^d \sigma^i y_0^i \delta^i > 0$, an inequality then also true for y in a small open neighborhood of y_0 and hence with κ-probability > 0.

Finally, if $\operatorname{supp}(\kappa)$ spans a linear subspace L_κ of $\mathbb{R}^d$ of dimension $d_\kappa < d$, one may proceed as in the preceding argument, but now considering partitionings of L_κ only. □

Remark 10.3.2 *Farkas' Lemma* from the theory of linear inequalities states that if $A \in \mathbb{R}^{d\times p}$ is a matrix and $v \in \mathbb{R}^{d\times 1}$ a column vector, then either

$$A\ell = v, \quad \text{all } \ell_q \ge 0 \tag{10.57}$$

has a solution $\ell \in \mathbb{R}^{p\times 1}$, or

$$\langle \delta, v \rangle < 0, \qquad \left\langle \delta, A_{|q} \right\rangle \ge 0 \quad (1 \le q \le p) \tag{10.58}$$

has a solution $\delta \in \mathbb{R}^d$, but there is no solution to both (10.57) and (10.58). The assertion that if there is no solution to (10.57), then (10.58) has a solution is close to, but not quite the same as the statement that if there is no solution to (10.44), then (10.45) has a solution with at least one inequality sharp. The argument in the proof above when $\operatorname{supp}(\kappa)$ is finite is a variation on a standard proof of Farkas' Lemma.

10.4 Contingent claims and pricing

In this section we shall assume that $\mathcal{PM} \neq \emptyset$ and will be particularly interested in models where there is exactly one equivalent martingale measure, $|\mathcal{PM}| = 1$.

From Example 10.3.1 we saw that if κ has finite support and is concentrated on p atoms $\alpha_1, \ldots, \alpha_p$ with $\kappa\left(\alpha_q\right) > 0$ for $1 \le q \le p$, then $\mathcal{PM} \neq \emptyset$ iff the equation (10.37),

$$A\ell = v, \tag{10.59}$$

has a solution $\ell = \left(\ell_q\right)_{1 \le q \le p}$ with all $\ell_q > 0$. Here, as before, $v = \left(r - a^i\right)_{1 \le i \le d}$ and $A \in \mathbb{R}^{d \times p}$ is the matrix with elements

$$A^i_q = \sigma^i \alpha^i_q \kappa\left(\alpha_q\right). \tag{10.60}$$

Now assuming that $\mathcal{PM} \neq \emptyset$ and that ℓ with all $\ell_q > 0$ satisfies (10.59), it is clear that if the equation $A\ell' = 0$ has a solution $\ell' \neq 0$, then also $\widetilde{\ell} = \ell + \varepsilon\ell'$ will solve (10.59) with all $\widetilde{\ell}_q > 0$ provided $\varepsilon \neq 0$ is sufficiently close to 0. Each such $\widetilde{\ell}$ defines a new member of $\mathcal{PM}$, and since by linear algebra an $\ell' \neq 0$ solving $A\ell' = 0$ exists iff $\operatorname{rank}(A) < p$, we conclude that if $\mathcal{PM} \neq \emptyset$, then $|\mathcal{PM}| = \infty$ or $|\mathcal{PM}| = 1$ with $|\mathcal{PM}| = \infty$ whenever $p > \operatorname{rank}(A)$ and $|\mathcal{PM}| = 1$ precisely when $p = \operatorname{rank}(A) \le d$.

The most natural condition for $|\mathcal{PM}| = 1$ is that $p = d$ and A is non-singular with all elements of the vector $A^{-1}v$ strictly positive, but $|\mathcal{PM}| = 1$ may also occur if $p = \operatorname{rank}(A) < d$ in which case v necessarily must belong to the subspace spanned by the columns in A.

If $\mathcal{PM} \neq \emptyset$ and κ does not have finite support, it is easy to see that $|\mathcal{PM}| = \infty$: let $\widetilde{\mathbb{P}} \in \mathcal{PM}$ be natural with deterministic κ-intensity $\widetilde{\lambda}(y) > 0$ satisfying (10.31). Because κ has infinite support, one may pick $\delta_1, \ldots, \delta_{d+1} > 0$ and disjoint subsets $B_1, \ldots, B_{d+1}$ of $\mathbb{R}^d_{\backslash 0}$ with all $\kappa\left(B_q\right) > 0$ and $\widetilde{\lambda}(y) \ge \delta_q$ for all $y \in B_q$, $1 \le q \le d+1$. But then there exists a vector $(\beta_1, \ldots, \beta_{d+1}) \in \mathbb{R}^{d+1}_{\backslash 0}$ such that

$$\sum_{q=1}^{d+1} \sigma^i \beta_q \int_{B_q} y^i\, \kappa(dy) = 0 \quad (1 \le i \le d)$$

and for $\varepsilon \neq 0$ sufficiently close to 0, the function

$$y \mapsto \widetilde{\lambda}(y) + \varepsilon \sum_{q=1}^{d+1} \beta_q 1_{B_q}(y)$$

will be > 0 and satisfy (10.31). We have shown

Proposition 10.4.1 *Suppose that $\mathcal{PM} \neq \emptyset$. Then either there is exactly one equivalent martingale measure relative to the true measure $\mathbb{P}$ or there are infinitely many. There is exactly one if and only if the support for κ is a finite set with $p \le d$ elements and furthermore* $\operatorname{rank}(A) = p$ *where A is the matrix defined by (10.60).*

If there are infinitely many equivalent martingale measures, then there are in particular infinitely many natural ones.

Note that if $|\mathcal{PM}| = 1$, necessarily $v \neq 0$, i.e. $a^{i_0} \neq r$ for some $1 \le i_0 \le d$. Finding the unique $\widetilde{\mathbb{P}} \in \mathcal{PM}$ requires us in particular to find the unique solution ℓ to (10.59) and if $v = 0$ we get $\ell = 0$, violating the constraint that all $\ell_q > 0$. For this reason and also for technical considerations it is assumed from now on that

$$\exists i_0 : a^{i_0} \neq r. \tag{10.61}$$

Let $t_0 > 0$ be given but arbitrary. A t_0*-contingent claim* is defined as an $\mathcal{F}^X_{t_0}$-measurable random variable. The t_0-claim C is called *attainable* if there exists a self-financing trading strategy $\psi = (\psi_t)_{0 \le t \le t_0}$ defined on the time-interval $[0, t_0]$ only, such that C is the value at time t_0 of the portfolio resulting from ψ, i.e.,

$$C = V_0 + \int_0^{t_0} \psi_s^0 r e^{rs}\, ds + \int_{]0,t_0]} \sum_{i=1}^d \psi_s^i\, dX_s^i \quad \mathbb{P}\text{-a.s.},$$

cf. (10.13) and (10.11). The strategy then *replicates* the claim C.

At first sight it seems obvious that the natural price to pay at time 0 for the claim C should be

$$V_0 = \psi_0^0 + \sum_{i=1}^d \psi_0^i x_0^i,$$

the value of the replicating portfolio at time 0. But as we shall presently see, there are always infinitely many self-financing strategies replicating C leading in fact to infinitely many prices. The remainder of this section is now devoted to a discussion of this disturbing fact and how it is still possible to price at least certain contingent claims fairly and uniquely.

To find some of the self-financing strategies replicating C, consider ψ represented as in (10.24) with the special structure

$$f^{0,(n)}_{z_n|x_0}(t) = \alpha^{0,(n)}_{z_n|x_0} + \beta^{0,(n)}_{z_n|x_0} t, \tag{10.62}$$

$$f^{i,(n)}_{z_n|x_0}(t) = \alpha^{i,(n)}_{z_n|x_0} + \beta^{i,(n)}_{z_n|x_0} e^{\left(r-a^i\right)t} \quad (1 \le i \le d), \tag{10.63}$$

for $n \ge 0$, $t_n \le t \le t_0$ where only $z_n = (t_1, \ldots, t_n; y_1, \ldots, y_n)$ with $t_n < t_0$ and y_k belonging to a suitable set of κ-measure 1 are considered. (With $\mathbb{P}$-probability 1 there will be no jumps at t_0, hence the condition $t_n < t_0$ is allowed). The conditions from Proposition 10.2.2 for ψ to be self-financing now become

$$\beta^{0,(n)}_{z_n|x_0} + \sum_{i=1}^d \left(r - a^i\right) \beta^{i,(n)}_{z_n|x_0} \eta^{i,(n)}_{z_n|x_0} = 0 \quad (n \ge 0), \tag{10.64}$$

and

$$\begin{aligned} &\alpha^{0,(n)}_{z_n|x_0} + \beta^{0,(n)}_{z_n|x_0} t_n + \sum_{i=1}^d \beta^{i,(n)}_{z_n|x_0} \eta^{i,(n)}_{z_n|x_0} \\ &= \alpha^{0,(n-1)}_{z_{n-1}|x_0} + \beta^{0,(n-1)}_{z_{n-1}|x_0} t_n + \sum_{i=1}^d \beta^{i,(n-1)}_{z_{n-1}|x_0} \eta^{i,(n)}_{z_n|x_0} \quad (n \ge 1), \end{aligned} \tag{10.65}$$

where we write

$$\eta_{z_n|x_0}^{i,(n)} = \left(\prod_{k=1}^{n}\left(1+\sigma^i y_k^i\right)\right) x_0^i$$

(which in fact depends on z_n through the y_k only). In particular $\eta_{|x_0}^{i,(0)} = x_0^i$.

By Proposition 4.2.1(b), since C is $\mathcal{F}_{t_0}^X$-measurable we may write

$$C = \sum_{n=0}^{\overline{N}_{t_0}} g_{Z_n|x_0}^{(n)} 1_{(\overline{N}_{t_0}=n)}$$

for some measurable $\mathbb{R}$-valued functions $(x_0, z_n) \mapsto g_{z_n|x_0}^{(n)}$. The requirement that ψ replicates C on the set $\left(\overline{N}_{t_0} = n, Z_n = z_n\right)$ then becomes

$$\alpha_{z_n|x_0}^{0,(n)} + \beta_{z_n|x_0}^{0,(n)} t_0 + \sum_{i=1}^{d} \beta_{z_n|x_0}^{i,(n)} \eta_{z_n|x_0}^{i,(n)} = e^{-rt_0} g_{z_n|x_0}^{(n)} \qquad (n \geq 0) \tag{10.66}$$

when $t_n < t_0$.

For $n \geq 1$, working recursively in n, we thus have the three linear equations (10.64), (10.65) and (10.66) to solve for the $d+2$ unknowns $\alpha_{z_n|x_0}^{0,(n)}$ and $\left(\beta_{z_n|x_0}^{i,(n)}\right)_{0\leq i\leq d}$, while for $n = 0$ there are only the two equations (10.64) and (10.66). So start for $n = 0$ by choosing $\alpha_{|x_0}^{0,(0)}$ and the $\beta_{|x_0}^{i,(0)}$ for $i \geq 1$, $i \neq i_0$ (with i_0 as in (10.61)) arbitrarily and then solving for $\beta_{|x_0}^{0,(0)}$ and $\beta_{|x_0}^{i_0,(0)}$ which is possible if the matrix of coefficients to the two unknowns,

$$\begin{pmatrix} 1 & \left(r - a^{i_0}\right) x_0^{i_0} \\ t_0 & x_0^{i_0} \end{pmatrix}$$

is non-singular, i.e., if $t_0\left(r - a^{i_0}\right) \neq 1$ as we assume from now on. Then proceed by induction, for each n choosing the $\alpha_{z_n|x_0}^{0,(n)}$ and the $\beta_{z_n|x_0}^{i,(n)}$ for $i \geq 1$, $i \neq i_0$ arbitrarily and then solving for $\alpha_{z_n|x_0}^{0,(n)}$, $\beta_{z_n|x_0}^{0,(n)}$ and $\beta_{z_n|x_0}^{i_0,(n)}$ which is possible because the matrix (where $\eta^{(n)}$ is short for $\eta_{z_n|x_0}^{i_0,(n)}$)

$$\begin{pmatrix} 0 & 1 & \left(r - a^{i_0}\right)\eta^{(n)} \\ 1 & t_n & \eta^{(n)} \\ 1 & t_0 & \eta^{(n)} \end{pmatrix} \tag{10.67}$$

is non-singular since $a^{i_0} \neq r$ and $t_n < t_0$.

Thus, provided $t_0\left(r - a^{i_0}\right) \neq 1$, for $d \geq 2$ there is for each n the free choice of at least one $\beta_{z_n|x_0}^{i,(n)}$. For $d = 1$ (in which case $a = a^{i_0} \neq r$ by assumption), given $\alpha_{|x_0}^{0,(0)}$, $\beta_{|x_0}^{0,(0)}$ and $\beta_{|x_0}^{1,(0)}$ the $\alpha_{z_n|x_0}^{0,(n)}$, $\beta_{z_n|x_0}^{0,(n)}$ and $\beta_{z_n|x_0}^{1,(n)}$ are uniquely determined, but there is still a free choice for $\alpha_{|x_0}^{0,(0)}$. Thus, for all d we have infinitely many self-financing trading strategies of the special form (10.62) and (10.63) that replicate C, and since

$$V_0 = \alpha^{0,(0)}_{|x_0} + \sum_{i=1}^{d} \beta^{i,(0)}_{|x_0} x^i_0$$

where $\alpha^{0,(0)}_{|x_0}$ can certainly be chosen freely, different strategies suggest different prices for C! To get around this unpleasant fact and find a fair price for some claims C, we shall focus on and exploit the probabilistic structure of the model.

Consider a natural $\widetilde{\mathbb{P}} \in \mathcal{PM}$ with deterministic κ-intensity $\widetilde{\lambda}(y) > 0$ satisfying (10.31). Suppose further that C is a t_0-contingent claim such that

$$\widetilde{\mathbb{E}}\,|C| < \infty,$$

e.g.,

$$|C| \le K\,(1+L)^{\overline{N}_{t_0}}$$

for some constants $K, L \ge 0$ (in which case also $\mathbb{E}_{\mathbb{P}}\,|C| < \infty$ and $\mathbb{E}_{\mathbb{P}'}\,|C| < \infty$ for all natural $\mathbb{P}' \in \mathcal{PM}$ since under $\mathbb{P}, \mathbb{P}', \overline{N}_{t_0}$ is Poisson).

Now define $M(C) = (M_t(C))_{0 \le t \le t_0}$ as the cadlag version of the martingale

$$M_t(C) = \widetilde{\mathbb{E}}\left[e^{-rt_0} C \,\middle|\, \mathcal{F}^X_t\right] \quad (0 \le t \le t_0)\,, \tag{10.68}$$

in particular

$$M_0(C) \equiv \widetilde{\mathbb{E}} e^{-rt_0} C.$$

By the martingale representation theorem, Theorem 4.6.1, there exists a predictable field $(S^y)_{y \in \mathbb{R}^d_{\backslash 0}}$ such that

$$M_t(C) = \widetilde{\mathbb{E}} e^{-rt_0} C + \int_{]0,t_0] \times \mathbb{R}^d_{\backslash 0}} S^y_s \, \widetilde{M}(ds, dy)\,, \tag{10.69}$$

where $\widetilde{M}(ds, dy)$ is the fundamental $\widetilde{\mathbb{P}}$-martingale random measure,

$$\widetilde{M}(ds, dy) = \mu(ds, dy) - \widetilde{\lambda}(y)\, ds\, \kappa(dy)\,.$$

The idea is now to look for predictable and e.g., pathwise bounded (see (10.38)) processes ψ^i for $i \ge 1$ such that

$$M^* \equiv M(C) \quad \mathbb{P}\text{-a.s.} \tag{10.70}$$

on $[0, t_0]$ where

$$M^*_t = V^*_0 + \int_{]0,t]} \sum_{i=1}^{d} \psi^i_s \, dX^{*i}_s, \tag{10.71}$$

since then in particular

$$e^{-rt_0}C = M^*_{t_0} = V^*_0 + \int_{]0,t_0]} \sum_{i=1}^{d} \psi^i_s \, dX^{*i}_s \quad \mathbb{P}\text{-a.s.}$$

Recalling Proposition 10.2.1 and defining ψ^0 by (10.20) we obtain a self-financing trading strategy ψ replicating C, which is such that the corresponding discounted value process $M^* = M(C)$ is a $\widetilde{\mathbb{P}}$-martingale.

For (10.70) to hold, obviously

$$V^*_0 \equiv \widetilde{\mathbb{E}} e^{-rt_0} C \tag{10.72}$$

which is then the fair price for C suggested by the martingale measure $\widetilde{\mathbb{P}}$. Furthermore, identifying the jumps at T_n for M^* and $M(C)$ gives the condition

$$\sum_{i=1}^{d} \psi^i_{T_n} \sigma^i X^{*i}_{T_n-} Y^i_n = S^{Y_n}_{T_n} \quad \text{on } (T_n < t_0), \tag{10.73}$$

while the identification between jumps yields

$$\sum_{i=1}^{d} \psi^i_t \sigma^i X^{*i}_{t-} \int_{\mathbb{R}^d_{\backslash 0}} y^i \widetilde{\lambda}(y)\, \kappa(dy) = \int_{\mathbb{R}^d_{\backslash 0}} S^y_t \widetilde{\lambda}(y)\, \kappa(dy) \quad (0 \le t < t_0), \tag{10.74}$$

as is seen using the identity

$$X^{*i}_t = x^i_0 + \int_{]0,t]} \int_{\mathbb{R}^d_{\backslash 0}} \sigma^i X^{*i}_{s-} y^i \, \widetilde{M}(ds, dy),$$

cf. (10.22).

Using (10.73) one would like the ψ^i to satisfy (since each T_n can be anything between 0 and t_0 and the Y_n are arbitrary jump sizes for U)

$$\sum_{i=1}^{d} \psi^i_t \sigma^i X^{*i}_{t-} y^i = S^y_t \quad (0 \le t < t_0) \tag{10.75}$$

simultaneously for κ-a.a. $y \in \mathbb{R}^d_{\backslash 0}$. This may not be possible, but if it is, multiplying by $\widetilde{\lambda}(y)$ and integrating with respect to κ yields (10.74). Thus (10.75) is critical if there is to exist a pathwise bounded self-financing strategy replicating C, so we shall now in some detail discuss when (10.75) can be solved. (It may be noted that it is possible to deduce (10.75) from (10.73): the idea is that if the ψ^i for $i \ge 1$ are pathwise bounded predictable processes satisfying (10.73), then one can show using only (10.73) that with $\widetilde{S}^y_t$ defined by the expression on the left of (10.75) it holds that the stochastic integrals $\widetilde{M}(S)$ and $\widetilde{M}\big(\widetilde{S}\big)$ are $\widetilde{\mathbb{P}}$-indistinguishable on $[0, t_0]$, i.e., the representation (10.69) of the martingale $M(C)$ is not affected if S is replaced by $\widetilde{S}$).

Suppose first that $|\mathcal{PM}| = 1$. Then by Proposition 10.4.1 κ has finite support $\alpha_1, \ldots, \alpha_p$ with $p \le d$ and $\text{rank}(A) = p$ where A is given by (10.60). (10.75) now becomes

$$\sum_{i=1}^{d} \psi_t^i \sigma^i X_{t-}^{*i} \alpha_q^i = S_t^{\alpha_q} \qquad (1 \le q \le p)$$

or equivalently,

$$A^T w = u \tag{10.76}$$

where

$$w = \left(\psi_t^i X_{t-}^{*i}\right)_{1 \le i \le d}, \qquad u = \left(S_t^{\alpha_q} \kappa\left(\alpha_q\right)\right)_{1 \le q \le p}.$$

If $p < d$, (10.76) has infinitely many solutions w, while if $p = d$, there is a unique solution w. Then $\psi_t^i = w^i / X_{t-}^{*i}$ for $i \ge 1$ together with ψ^0 given by (10.20) defines the desired replicating strategy provided e.g., that the ψ^i are pathwise bounded, something that may of course depend on the structure of S. But if the ψ found is pathwise bounded, M_t^* is well defined by (10.71), M^* is a $\widetilde{\mathbb{P}}$-martingale and the fair price for the claim C is given by (10.72) with $\widetilde{\mathbb{P}}$ the unique risk-neutral measure.

Suppose next that $|\mathcal{PM}| = \infty$ with $\widetilde{\mathbb{P}} \in \mathcal{PM}$ given and natural. Now it may well happen that (10.75) has no solution, e.g., for $d = 1$ it is required that

$$\psi_t^1 = \frac{S_t^y}{\sigma X_{t-}^* y}$$

simultaneously for κ-a.a. $y \in \mathbb{R}_{\backslash 0}$, so that S_t^y must be a linear function in y. But if there exist pathwise bounded predictable ψ^i for $i \ge 1$ that do satisfy (10.75), then defining ψ^0 as in (10.20) we see that the self-financing strategy ψ replicates C with M^* a $\widetilde{\mathbb{P}}$-martingale and (10.72) the price for C corresponding to the given $\widetilde{\mathbb{P}}$. This of course defines a unique price for C iff $\mathbb{E}_{\mathbb{P}'} e^{-rt_0} C$ is the same for all $\mathbb{P}' \in \mathcal{PM}$ and at least we can give a simple sufficient condition for this: suppose that (10.75) corresponding to the original $\widetilde{\mathbb{P}}$ has a bounded (in the sense of (10.38)) and predictable solution ψ^i for $i \ge 1$. Then M^* given by (10.71) will be a $\mathbb{P}'$-martingale for *any* $\mathbb{P}' \in \mathcal{PM}$ and consequently

$$\mathbb{E}_{\mathbb{P}'} e^{-rt_0} C = M_0^* = V_0^* = \widetilde{\mathbb{E}} e^{-rt_0} C.$$

In mathematical finance one calls a market *complete* if 'any contingent claim can be hedged' and one shows that the market is complete 'iff there is a unique risk-neutral measure'. That a t_0-claim C can be *hedged* means that there is a self-financing strategy such that the value of the corrsponding portfolio at time t_0 is precisely C. For our model we cannot quite state results like this, but the discussion above at least gives Theorem 10.4.2 below.

In order to state this result we shall only consider the class $\mathcal{C}_{t_0}$ of t_0-claims C (for an arbitrary $t_0 > 0$) such that $V_0^* = \widetilde{\mathbb{E}} e^{-rt_0} C$ is the same for all $\widetilde{\mathbb{P}} \in \mathcal{PM}$ and then say that C can be hedged if there exists a pathwise bounded self-financing trading strategy $\psi = \left(\psi^i\right)_{i \ge 0}$ such that

$$e^{-rt_0}C = V_0^* + \int_{]0,t_0]} \sum_{i=1}^{d} \psi_s^i \, dX_s^{*i}. \tag{10.77}$$

In that case V_0^* is the time 0 price for the claim C.

Theorem 10.4.2 (i) *Suppose that $|\mathcal{PM}| = 1$ and let $\widetilde{\mathbb{P}}$ denote the unique equivalent martingale measure. Then $\mathcal{C}_{t_0}$ is the class of $\widetilde{\mathbb{P}}$-integrable $\mathcal{F}_{t_0}^X$-measurable random variables.*

Let $C \in \mathcal{C}_{t_0}$, consider the $\widetilde{\mathbb{P}}$-martingale $M(C)$ given by (10.68) and let S be determined as in (10.69). Then (10.75) always has a predictable solution ψ^i for $i \geq 1$ and a sufficient condition for C to be hedgeable is that this solution be pathwise bounded.

(ii) *Suppose that $|\mathcal{PM}| = \infty$ and let $\widetilde{\mathbb{P}} \in \mathcal{PM}$ be given. Then $\mathcal{C}_{t_0}$ is a genuine linear subspace of the class of $\widetilde{\mathbb{P}}$-integrable $\mathcal{F}_{t_0}^X$-measurable random variables.*

Let $C \in \mathcal{C}_{t_0}$, consider the $\widetilde{\mathbb{P}}$-martingale $M(C)$ given by (10.68) and let S be determined as in (10.69). Then (10.75) may not have a predictable solution ψ^i for $i \geq 1$, but if it does, a sufficient condition for C to be hedgeable is that this solution be bounded.

(iii) *In both (i) and (ii), if ψ^i for $i \geq 1$ is the required solution to (10.75), then $\psi = \left(\psi^i\right)_{0 \leq i \leq d}$ with ψ^0 defined by (10.20) is a self-fiancing strategy replicating C.*

The only assertion not proved already is the claim from (ii) that $\mathcal{C}_{t_0}$ is a genuine subspace of $L^1\left(\Omega, \mathcal{F}_{t_0}^X, \widetilde{\mathbb{P}}\right)$. To see this, take $\widetilde{\mathbb{P}} \neq \mathbb{P}' \in \mathcal{PM}$ both natural and just note that $\widetilde{\mathbb{P}} \neq \mathbb{P}'$ because they are natural simply means that the $\widetilde{\mathbb{P}}$-distribution of (T_1, Y_1) is different from the $\mathbb{P}'$-distribution, implying since T_1 and Y_1 are independent under both $\widetilde{\mathbb{P}}$ and $\mathbb{P}'$, that there exists $B \in \mathcal{B}_{\backslash 0}^d$ such that

$$\widetilde{\mathbb{P}}\left(T_1 < t_0, Y_1 \in B\right) \neq \mathbb{P}'\left(T_1 < t_0, Y_1 \in B\right),$$

i.e., the t_0-claim

$$1_{(T_1 < t_0, Y_1 \in B)}$$

does not belong to $\mathcal{C}_{t_0}$.

Remark 10.4.1 In case (ii), observe that $\mathcal{C}_{t_0}$ is always a linear space of random variables containing the constants and all the $X_{t_0}^i$.

Let C be any t_0-contingent claim. We saw earlier that there is a host of simple self-financing strategies ψ that replicate C, see (10.62) and (10.63) and that these strategies yield different prices. How can this fact be reconciled with the findings from Theorem 10.4.2 according to which certain claims have a well-defined unique price? Suppose that $|\mathcal{PM}| = 1$, write $\widetilde{\mathbb{P}}$ for the unique risk-neutral measure, let $C \in \mathcal{C}_{t_0}$ and assume that C can be hedged using the self-financing strategy ψ described in Theorem 10.4.2(i). Then if $\widetilde{\psi}$ is another replicating strategy of the form (10.62) and (10.63) with

$$\widetilde{V}_0 = \widetilde{\psi}_0^0 + \sum_{i=1}^{d} \widetilde{\psi}_0^i x_0^i \neq V_0^* = \widetilde{\mathbb{E}} e^{-rt_0} C,$$

we certainly have

$$e^{-rt_0} C = \widetilde{V}_0 + \int_{]0,t_0]} \sum_{i=1}^{d} \widetilde{\psi}_t^i \, dX_t^{*i}$$

and since $\widetilde{V}_0 \neq V_0^*$ are forced to conclude that the stochastic integral

$$\int_{]0,t]} \sum_{i=1}^{d} \widetilde{\psi}_s^i \, dX_s^{*i} \tag{10.78}$$

involving the $\widetilde{\mathbb{P}}$-martingales X^{*i} cannot itself be a $\widetilde{\mathbb{P}}$-martingale. To substantiate this further, go back to the construction of $\widetilde{\psi}$, which on the set $\left(\overline{N}_{t_0} = n, Z_n = z_n\right)$ when $t_n < t_0$ and $n \geq 1$ involved the solution of three linear equations with three unknowns with coefficient matrix (10.67). By inspection one sees that for t_n close to t_0 this matrix is nearly singular, implying that the trading strategy may well explode when t_n is close to t_0: following the strategy $\widetilde{\psi}$, if a jump occurs close to time t_0 violent adjustments in the portfolio may be needed in order to replicate C and the martingale property (integrability) of (10.78) may be destroyed in contrast to what happens with the strategies ψ described in Theorem 10.4.2.

Example 10.4.1 Some of the simplest t_0-contingent claims imagineable have the form

$$C = h\left(X_{t_0}^{i_1}\right)$$

for some $i_1 \geq 1$ and some function h. If $|\mathcal{PM}| = 1$ such C are of course hedgeable provided $\widetilde{\mathbb{E}}\,|C| < \infty$ with $\widetilde{\mathbb{P}}$ the unique member of $\mathcal{PM}$. A compact explicit expression for the price $\widetilde{\mathbb{E}} e^{-rt_0} C$ is usually not available, but if $d = 1$, $\kappa = \varepsilon_\alpha$ and $r \neq a$, we have that the total Poisson-intensity for a jump under $\widetilde{\mathbb{P}}$ is

$$\widetilde{\lambda} = \frac{r - a}{\sigma\alpha} > 0,$$

cf. Example 10.3.1, and find that

$$\widetilde{\mathbb{E}} e^{-rt_0} C = e^{-rt_0} \sum_{n=0}^{\infty} \frac{\left(\widetilde{\lambda} t_0\right)^n}{n!} e^{-\widetilde{\lambda} t_0} h\left(e^{at_0} (1 + \sigma\alpha)^n x_0\right).$$

Special cases include the *European call option* with exercise price K, expiry date t_0 and the *European put option* corresponding to

$$h(x) = (x - K)^+, \quad h(x) = (K - x)^+$$

respectively.

We have seen that if $|\mathcal{PM}| = \infty$ not all contingent claims are hedgeable. By definition hedgeable claims must have the form (10.77) (which need not even be the case in general for the claims $h\left(X_{t_0}^{i_1}\right)$ considered in Example 10.4.1), so in order to construct a hedgeable t_0-claim C one may start with a collection $\left(\psi^i\right)_{i\geq 1}$ of predictable and, say, bounded processes, define ψ^0 by (10.20) to obtain a self-financing strategy and then consider the t_0-claim

$$C = e^{rt_0}\left(V_0^* + \int_{]0,t_0]} \sum_{i=1}^{d} \psi_s^i \, dX_s^{*i}\right).$$

Then using (10.75) to define S,

$$S_t^y = \sum_{i=1}^{d} \psi_t^i \sigma^i X_{t-}^{*i} y^i$$

we have in fact for any $\widetilde{\mathbb{P}} \in \mathcal{PM}$ with κ-intensity process $\left(\widetilde{\lambda}_t^y\right)$ that

$$\int_{]0,t]} \sum_{i=1}^{d} \psi_s^i \, dX_s^{*i} = \int_{]0,t]\times\mathbb{R}^d_{\setminus 0}} S_s^y \, \widetilde{M}(ds, dy) \quad (t \leq t_0) \tag{10.79}$$

where $\widetilde{M}(ds, dy) = \widetilde{\lambda}_s^y \, ds \, \kappa(dy)$ is the martingale measure for $\widetilde{\mathbb{P}}$. Thus the representation (10.69) of the $\widetilde{\mathbb{P}}$-martingale $\left(\widetilde{\mathbb{E}}\left[e^{-rt_0} C \,\middle|\, \mathcal{F}_t^X\right]\right)_{0\leq t\leq t_0}$ holds for all $\widetilde{\mathbb{P}} \in \mathcal{PM}$ with a predictable field S that does not depend on $\widetilde{\mathbb{P}}$.

Exercise 10.4.1 With an arbitrary $\widetilde{\mathbb{P}} \in \mathcal{PM}$ as above, it follows from Proposition 10.3.1 and its proof that

$$\sigma^i \int_{\mathbb{R}^d_{\setminus 0}} y^i \widetilde{\lambda}_t^y \, \kappa(dy) = r - a^i$$

for all t and all $i \geq 1$. Use this to show (10.79) by identifying the jumps and the behaviour between the jumps.

11

Examples of Queueing Models

In the first section the classical GI/G/1 model is treated with the emphasis of finding when there is equilibrium (stationarity). For the simplest case, the M/M/1 model, stationarity is discussed in detail, not just for the length of the queue but also involving the time since most the recent arrival and the time since the present service started. The second section deals with the description of some PDMP models of queueing networks that are not just (homogeneous) Markov chains.

References. Some important books on queueing theory and networks are Asmussen [3], Brémaud [15], Baccelli and Brémaud [7], Kalashnikov [72], Kelly [79], Robert [103], Serfozo [110] and Walrand [119].

11.1 The GI/G/1 queue

The simplest and most basic model in queueing theory is the single server queue, where customers arrive at one service station, are serviced one at a time on a first come first serve basis and when service is completed the customers leave the system.

Here we shall discuss the *GI/G/1 queue* where arrivals occur according to a renewal process with interarrival times that have a given hazard function $\alpha(a)$ and where the service times are iid and independent of the arrival process with a given hazard function $\beta(s)$. Particularly simple and important is the *M/M/1 queue* where $\alpha(a) \equiv \alpha$, $\beta(s) \equiv \beta$ so that arrivals occur according to a homogeneous Poisson process (assuming that the first customer arrives at time 0) and the service times are exponential.

A question of particular importance for the study of any queueing system is whether it is stable in the sense that the number of customers in the system does not explode as time $t \to \infty$. Phrased differently, it is of interest to determine conditions for a queueing system to be in equlibrium meaning more precisely that if the system can be described using a time-homogeneous Markov process, one wants conditions for this process to have an invariant distribution (see Section 7.8).

Let K_t denote the number of customers in the GI/G/1 queue at time t, i.e., either $K_t = 0$ and the queue is empty, or $K_t \geq 1$ and K_t counts the customer presently being serviced as well as all customers that have arrived but not yet started service.

For the M/M/1 queue (K_t) is a time-homogeneous Markov chain on $\mathbb{N}_0$ and one of the simplest and classical results in queueing theory states that this Markov chain has an invariant distribution $\rho = (\rho_p)_{p \geq 0}$ iff $\alpha < \beta$ and that in that case ρ is a geometric distribution, see (11.12) below.

For the general GI/G/1 queue, (K_t) is not Markov. However, as we shall now see, it is still possible to view the GI/G/1 queue as a time-homogeneous PDMP — and hence to discuss conditions for equlibrium of the GI/G/1 queue.

Introduce the arrival process (A_t) and the service process (S_t), where A_t is the time elapsed since the most recent arrival and S_t is the time elapsed since the customer presently being serviced started service (so S_t is defined only if $K_t \geq 1$). Note in particular that (A_t) is simply the backward recurrence time process for a renewal process.

Now consider the joint process $X = (K_t, A_t, S_t)_{t \geq 0}$. The state space consists of pairs $x = (0, a)$ (the queue is empty, time a elapsed since the most recent arrival) and triples $x = (p, a, s)$ with $p \in \mathbb{N}$, $a \geq 0$, $s \geq 0$ (p customers in the queue, time a gone since latest arrival, the customer being serviced started service s time units ago).

We formally make X into a time-homogeneous PDMP, with jump times T_n and $Y_n = X_{T_n}$ the state reached by the nth jump, by defining the deterministic behaviour

$$\phi_t(0, a) = (0, a + t), \qquad \phi_t(p, a, s) = (p, a + t, s + t) \quad (p \geq 1),$$

the total intensity $q(x)$ for a jump from an arbitrary state x,

$$q(0, a) = \alpha(a), \qquad q(p, a, s) = \alpha(a) + \beta(s) \quad (p \geq 1),$$

and the distribution $r(x, \cdot)$ for a jump from x,

$$r((0, a), \cdot) = \varepsilon_{(1,0,0)},$$

$$r((p, a, s), \cdot) = \frac{\alpha(a)}{\alpha(a) + \beta(s)} \varepsilon_{(p+1,0,s)} + \frac{\beta(s)}{\alpha(a) + \beta(s)} \varepsilon_{(p-1,a,0)} \quad (p \geq 1)$$

(with $\varepsilon_{(p-1,a,0)}$ replaced by $\varepsilon_{(0,a)}$ if $p = 1$), reflecting in particular the facts that if a new customer arrives, K_t increases by 1, A_t is reset to 0 and S_t is unchanged, while if sevice is completed for a customer, K_t decreases by 1, S_t is reset to 0 if there is a customer waiting to start service and A_t is retained.

Some special features of the process X should be noted: no matter where X starts, it is perfectly possible that for a strictly positive amount of time one has $A_t = S_t$, but for $t \geq T_1$ this can only happen if also $K_t = 1$ and the single customer involved arrived at an empty queue. Also for $t \geq T_1$, if $K_t = 1$ necessarily $S_t \leq A_t$.

Let

$$\xi_A = \int_0^\infty e^{-\int_0^t \alpha(s)\, ds}\, dt, \quad \xi_S = \int_0^\infty e^{-\int_0^t \beta(s)\, ds}\, dt$$

denote the mean of the interarrival time distribution and the service time distribution respectively. Then it may be shown that a necessary and sufficient condition for X to have an invariant distribution is that

$$\xi_S < \xi_A < \infty. \tag{11.1}$$

Here we just indicate why this condition is necessary: if X has an invariant distribution there is in particular an invariant distribution for the backward recurrence time process (A_t) and therefore $\xi_A < \infty$ as shown in Example 7.8.3. And for n large the time of the nth arrival will take place approximately at time $n\xi_A$ while the time required to service these n customers is close to $n\xi_S$, i.e., $K_t \to \infty$ a.s. if $\xi_S > \xi_A$ which is not possible if X is started according to an invariant distribution which would imply in particular that all K_t have the same distribution. (If $\xi_S = \xi_A$ one may show that $\limsup_{t\to\infty} K_t = \infty$ a.s.).

Suppose that (11.1) holds. In order to study what the invariant distribution for X looks like we shall use the results from Section 7.8. What follows uses Theorem 7.8.2 but it should be noted that Corollary 7.8.10 can also be applied with the choice

$$x_0 = (1, 0, 0)$$

corresponding to T_{x_0} from the corollary being the first time a customer arrives at an empty queue: no matter what is the initial state of the queue, it is clear that $\mathbb{P}_{|x_0}\left(\mathrm{T}_{x_0} < \infty\right) > 0$ as is a minimal requirement if the corollary is to be of any use. The corollary shows in particular that the condition $\mathbb{E}_{|x_0} \mathrm{T}_{x_0} < \infty$ is sufficient (it is in fact also necessary) for the existence of an invariant distribution.

For functions $h_p(a, s)$ that are continuously differentiable in a and s, the infinitesimal generator for X has the form

$$(Ah)(0, a) = D_a h_0(a) + \alpha(a)\left(h_1(0, 0) - h_0(a)\right) \tag{11.2}$$

and for $p \geq 1$,

$$\begin{aligned}(Ah)(p, a, s) = D_a h_p(a, s) + D_s h_p(a, s) + \alpha(a)\left(h_{p+1}(0, s) - h_p(a, s)\right)\\ + \beta(s)\left(h_{p-1}(a, 0) - h_p(a, s)\right).\end{aligned} \tag{11.3}$$

We shall also need the full infinitesimal generator A (with $\mathrm{A}h = Ah$ for the smooth h just considered) described in Section 7.7 and note in particular that any function h of the form

$$h_p(a, s) = 1_{p\geq 1, a=s}\, \widetilde{h}_p(a)$$

belongs to the domain $\mathcal{D}(\mathrm{A})$ provided h is bounded and each $\widetilde{h}_p(a)$ is continuously differentiable in a with $\mathrm{A}h$ bounded where

$$(\mathrm{A}h)(0, a) = \alpha(a)\, \widetilde{h}_1(0) \tag{11.4}$$

and for $p \geq 1$,

$$(\mathrm{A}h)(p,a,s) = 1_{a=s} D_a \widetilde{h}(a) + \alpha(a)\left(1_{s=0}\widetilde{h}_{p+1}(0) - 1_{a=s}\widetilde{h}_p(a)\right) + \beta(s)\left(1_{a=0}\widetilde{h}_{p-1}(0) - 1_{a=s}\widetilde{h}_p(a)\right), \tag{11.5}$$

cf. Definition 7.7.1.

We shall discuss what is the form of the invariant distribution $\Phi(dp, da, ds)$ when it exists and first note that it is a good guess that for each given value of p, $\Phi(\{p\}, da, ds)$ has a density as follows:

$$\begin{aligned} \Phi(\{0\}, da) &= \varphi_0(a)\, da, \\ \Phi(\{1\}, da, ds) &= \varphi_1(a,s)\, da\, ds + 1_{s=a}\,\psi_1(a)\, da, \\ \Phi(\{p\}, da, ds) &= \varphi_p(a,s)\, da\, ds \quad (p \geq 2) \end{aligned} \tag{11.6}$$

with the contribution from ψ_1 taking care of the possibility that there is only one customer present who arrived when the system was empty.

By Theorem 7.8.2, for identifying the φ_p and ψ_1 one studies the equation

$$0 = \int_0^\infty da\, \varphi_0(a)(Ah)(0,a) + \sum_{p=1}^{\infty} \int_0^\infty da \int_0^\infty ds\, \varphi_p(a,s)(Ah)(p,a,s) + \int_0^\infty da\, \psi_1(a)(Ah)(1,a,a)$$

for $h_p(a,s)$ that are bounded and differentiable in a and s with Ah given by (11.2) and (11.3) bounded. Here it is natural to consider

$$h_p(a,s) = 1_{p=p_0} f(a,s) \tag{11.7}$$

for any p_0 (omitting the argument s if $p_0 = 0$) with f smooth enough and having compact support. Doing this for $p = p_0 \geq 3$ gives

$$\begin{aligned} 0 = &\int_0^\infty da \int_0^\infty ds\, \varphi_p(a,s)\left(D_a f + D_s f - (\alpha(a) + \beta(s)) f\right)(a,s) \\ &+ \int_0^\infty da \int_0^\infty ds\, \varphi_{p-1}(a,s)\,\alpha(a)\, f(0,s) \\ &+ \int_0^\infty da \int_0^\infty ds\, \varphi_{p+1}(a,s)\,\beta(s)\, f(a,0), \end{aligned}$$

and by partial integration, since f has compact support this implies that

$$\begin{aligned} 0 = &\int_0^\infty da \int_0^\infty ds\, f(a,s)\left(-D_a\varphi_p - D_s\varphi_p - (\alpha(a) + \beta(s))\varphi_p\right)(a,s) \\ &- \int_0^\infty ds\, f(0,s)\,\varphi_p(0,s) - \int_0^\infty da\, f(a,0)\,\varphi_p(a,0) \\ &+ \int_0^\infty da \int_0^\infty ds\, \varphi_{p-1}(a,s)\,\alpha(a)\, f(0,s) \\ &+ \int_0^\infty da \int_0^\infty ds\, \varphi_{p+1}(a,s)\,\beta(s)\, f(a,0). \end{aligned}$$

Choosing f to vanish everywhere except in a small neighborhood of a given point (a, s) one deduces that if $a > 0,\ s > 0$,

$$D_a \varphi_p (a, s) + D_s \varphi_p (a, s) = -\left(\alpha (a) + \beta (s)\right) \varphi_p (a, s), \tag{11.8}$$

while if $a > 0,\ s = 0$,

$$\varphi_p (a, 0) = \int_0^\infty ds\, \varphi_{p+1} (a, s)\, \beta (s) \tag{11.9}$$

and by analogy if $a = 0,\ s > 0$

$$\varphi_p (0, s) = \int_0^\infty ds\, \varphi_{p-1} (a, s)\, \alpha (a) .$$

Similar reasoning with h as in (11.7) and $p_0 = 1, 2$ or with $p_0 = 0$ and $h_p (a, s) = 1_{p=0} f (a)$ shows that (11.8) and (11.9) are valid also for $p = 1, 2$ and that in addition

$$D_a \varphi_0 (a) + \alpha (a)\, \varphi_0 (a) = \int_0^\infty ds\, \varphi_1 (a, s)\, \beta (s) + \psi_1 (a)\, \beta (a)$$

and

$$\varphi_1 (0, s) = 0, \quad \varphi_2 (0, s) = \int_0^\infty ds\, \varphi_1 (a, s)\, \alpha (a) + \psi_1 (s)\, \alpha (s) .$$

Finally, for $p_0 = 1$ and f vanishing outside a neighborhood of a point of the form (a, a) with $a > 0$ one finds that

$$D\psi_1 (a) = -\left(\alpha (a) + \beta (a)\right) \psi_1 (a) \tag{11.10}$$

and for $p_0 = 0$ with f vanishing away from a neighborhood of $(0, 0)$ that

$$\varphi_0 (0) = 0, \quad \int_0^\infty da\, \varphi_0 (a)\, \alpha (a) = \psi_1 (0) .$$

If the equations from (11.8) and below can be solved for $\varphi_p \geq 0,\ \psi_1 \geq 0$ with

$$\int_0^\infty da\, \varphi_0 (a) + \sum_{p=1}^\infty \int_0^\infty da \int_0^\infty ds\, \varphi_p (a, s) + \int_0^\infty da \psi_1 (a) = 1,$$

Theorem 7.8.2 shows that X has an invariant distribution given by (11.6). Of course, solving the equations explicitly is not possible, but it may be noted that the first order partial differential equation (11.8) has the complete solution

$$\varphi_p (a, s) = g_p (s - a) \exp\left(-\int_0^a \alpha - \int_0^s \beta\right)$$

for some functions g_p, while (11.10) immediately gives

$$\psi_1(a) = C_1 \exp\left(-\int_0^a (\alpha + \beta)\right).$$

It should also be remembered that since for $t \geq T_1,\ S_t \leq A_t$ whenever $K_t = 1$, one has $\varphi_1(a, s) = 0$ if $a < s$.

We shall now proceed to describe Φ (see (11.6)) explicitly in the M/M/1 case when $\alpha(a) \equiv \alpha$ and $\beta(a) \equiv \beta$ with $\alpha < \beta$ (the condition for an invariant distribution to exist, see (11.1)). Even then the equations above for the φ_p and ψ_1 are not easy to solve so instead we shall find the partial Laplace transforms (for any given p and $\theta \geq 0,\ \vartheta \geq 0$)

$$\begin{aligned} L_0(\theta) &= \int_0^\infty e^{-\theta a}\varphi_0(a)\, da, \\ L_p(\theta, \vartheta) &= \int_0^\infty da \int_0^\infty ds\, e^{-\theta a - \vartheta s}\varphi_p(a, s) \quad (p \geq 1), \\ L_1^D(\theta) &= \int_0^\infty e^{-\theta a}\psi_1(a)\, da. \end{aligned}$$

In particular the Laplace transforms determine the point probabilities $\rho_p = \Phi(\{p\} \times \mathbb{R}_0 \times \mathbb{R}_0)$,

$$\rho_p = \begin{cases} L_0(0) & \text{for } p = 0, \\ L_1(0,0) + L_1^D(0) & \text{for } p = 1, \\ L_p(0,0) & \text{for } p \geq 2. \end{cases} \tag{11.11}$$

Theorem 11.1.1 *The invariant distribution Φ for the M/M/1 queue (K_t, A_t, S_t) exists if and only if $\alpha < \beta$ and is then determined by*

$$\rho_p = \left(1 - \frac{\alpha}{\beta}\right)\left(\frac{\alpha}{\beta}\right)^p \quad (p \geq 0) \tag{11.12}$$

and for $\theta \geq 0,\ \vartheta \geq 0$,

$$L_0(\theta) = \rho_0 \frac{\alpha\beta}{(\theta + \alpha)(\theta + \beta)}, \tag{11.13}$$

$$L_1(\theta, \vartheta) = \rho_1 \frac{\alpha\beta}{(\theta + \vartheta + \alpha + \beta)(\theta + \beta)}, \tag{11.14}$$

$$L_1^D(\theta) = \rho_1 \frac{\beta}{\theta + \alpha + \beta}, \tag{11.15}$$

and for $p \geq 2$,

$$L_p(\theta, \vartheta) = \rho_p \left(\frac{\alpha\beta}{(\theta + \beta)(\vartheta + \alpha)} + \frac{\beta\vartheta}{(\theta + \vartheta + \alpha + \beta)(\vartheta + \alpha)}\left(\frac{\beta}{\vartheta + \alpha + \beta}\right)^{p-1}\right). \tag{11.16}$$

Proof. We shall use that

$$\int \Phi\,(dp, da, ds)\ (\mathrm{A}h)\,(p, a, s) = 0 \tag{11.17}$$

for all $h \in \mathcal{D}\,(\mathrm{A})$, the domain of the full infinitesimal generator; cf. 7.8.2. Starting with

$$h_p\,(a, s) = 1_{p=0}\, e^{-\theta a}$$

gives using (11.2), since (11.17) contributes only for $p = 0$ or 1,

$$-\,(\theta + \alpha)\, L_0\,(\theta) + \beta \left(L_1\,(\theta, 0) + L_1^D\,(\theta) \right) = 0. \tag{11.18}$$

Proceeding with

$$h_p\,(a, s) = 1_{p=p_0}\, e^{-\theta a - \vartheta s}$$

for a given $p_0 \geq 1$, using (11.3) and noting that (11.17) contributes only for $p = p_0 - 1$ or $p_0 + 1$ gives first for $p_0 = 1$,

$$-\,(\theta + \vartheta + \alpha + \beta) \left(L_1\,(\theta, \vartheta) + L_1^D\,(\theta + \vartheta) \right) + \alpha L_0\,(0) + \beta L_2\,(\theta, 0) = 0, \tag{11.19}$$

and for $p_0 = 2$,

$$-\,(\theta + \vartheta + \alpha + \beta)\, L_2\,(\theta, \vartheta) + \alpha \left(L_1\,(0, \vartheta) + L_1^D\,(\vartheta) \right) + \beta L_3\,(\theta, 0) = 0, \tag{11.20}$$

and finally for $p = p_0 \geq 3$,

$$-\,(\theta + \vartheta + \alpha + \beta)\, L_p\,(\theta, \vartheta) + \alpha L_{p-1}\,(0, \vartheta) + \beta L_{p+1}\,(\theta, 0) = 0. \tag{11.21}$$

Taking $\theta = \vartheta = 0$ (cf. (11.11)) gives the classical equations for determining the ρ_p,

$$\rho_1 = \delta \rho_0, \qquad \rho_{p+1} - \delta \rho_p = \rho_p - \delta \rho_{p-1} \quad (p \geq 1)$$

where $\delta = \alpha/\beta$. This forces $\rho_p = \delta^p \rho_0$, and since we need $\sum_0^\infty \rho_p = 1$ it follows that necessarily $\alpha < \beta$ and that under this condition ρ_p is given by (11.12).

The equations (11.19), (11.20) and (11.21) show that for $p \geq 1$, $L_p\,(\theta, \vartheta)$ is determined once $L_p\,(\theta, 0)$, $L_p\,(0, \vartheta)$ for $p \geq 1$ and $L_1^D\,(\theta)$ have been found. So the next step is of course to take $\theta = 0$, respectively $\vartheta = 0$ in the three equations. This gives

$$-\,(\vartheta + \alpha + \beta) \left(L_1\,(0, \vartheta) + L_1^D\,(\vartheta) \right) + \alpha \rho_0 + \beta \rho_2 = 0, \tag{11.22}$$

$$-\,(\theta + \alpha + \beta) \left(L_1\,(\theta, 0) + L_1^D\,(\theta) \right) + \alpha \rho_0 + \beta L_2\,(\theta, 0) = 0, \tag{11.23}$$

$$-\,(\vartheta + \alpha + \beta)\, L_2\,(0, \vartheta) + \alpha \left(L_1\,(0, \vartheta) + L_1^D\,(\vartheta) \right) + \beta \rho_3 = 0, \tag{11.24}$$

$$-\,(\vartheta + \alpha + \beta)\, L_p\,(0, \vartheta) + \alpha L_{p-1}\,(0, \vartheta) + \beta \rho_{p+1} = 0, \tag{11.25}$$

$$-\,(\theta + \alpha + \beta)\, L_p\,(\theta, 0) + \alpha \rho_{p-1} + \beta L_{p+1}\,(\theta, 0) = 0. \tag{11.26}$$

with (11.25) valid for $p \geq 3$ and (11.26) valid for $p \geq 2$. Here (11.22) immediately gives

$$L_1(0,\vartheta) + L_1^D(\vartheta) = \frac{\alpha\rho_0 + \beta\rho_2}{\vartheta + \alpha + \beta} = \rho_1 \frac{\alpha + \beta}{\vartheta + \alpha + \beta} \tag{11.27}$$

and (11.24) and (11.25) can then be used to determine $L_p(0,\vartheta)$ recursively leading to

$$L_p(0,\vartheta) = \rho_p \left(\frac{\alpha}{\vartheta + \alpha} + \frac{\beta\vartheta}{(\vartheta + \alpha + \beta)(\vartheta + \alpha)} \left(\frac{\beta}{\vartheta + \alpha + \beta} \right)^{p-1} \right) \quad (p \geq 2)$$

(with the expression on the right for $p = 1$ in fact equal to $L_1(0,\vartheta) + L_1^D(\vartheta)$ as given by (11.27)).

In order to determine $L_p(\theta, 0)$ we rewrite (11.26) as

$$L_{p+1}(\theta,0) = \frac{\theta + \alpha + \beta}{\beta} L_p(\theta,0) - \rho_p \quad (p \geq 2),$$

which implies that (recall that $\rho_p = (1-\delta)\delta^p$)

$$\begin{aligned} L_p(\theta,0) &= \gamma^{p-2} L_2(\theta,0) - (1-\delta)\gamma^{p-1} \sum_{k=2}^{p-1} \left(\frac{\delta}{\gamma}\right)^k \\ &= \gamma^{p-2} \left(L_2(\theta,0) - \frac{\rho_2}{\gamma} \frac{1 - \left(\frac{\delta}{\gamma}\right)^{p-2}}{1 - \frac{\delta}{\gamma}} \right) \quad (p \geq 2), \end{aligned}$$

where $\gamma = (\theta + \alpha + \beta)/\beta$, $\delta = \alpha/\beta$. But $\gamma > 1$ and since $L_p(\theta,0) \leq \rho_p \leq 1$ and $\delta/\gamma < 1$, as $p \to \infty$ this forces

$$L_2(\theta,0) = \rho_2 \frac{1}{\gamma - \delta} = \rho_2 \frac{\beta}{\theta + \beta} \tag{11.28}$$

and

$$L_p(\theta,0) = \rho_p \frac{1}{\gamma - \delta} = \rho_p \frac{\beta}{\theta + \beta} \quad (p \geq 3).$$

It still remains to find $L_1(0,\vartheta)$, $L_1(\theta,0)$, $L_1^D(\theta)$ and $L_0(\theta)$ (with the latter directly obtainable from (11.18) once the others have been found) and for this the unused equations (11.22) and (11.23) are simply not enough! But using the full generator (11.4) and (11.5) on the function

$$h_p(a,s) = 1_{p=1,a=s}\, e^{-\theta a}$$

and integrating Ah with respect to Φ gives

$$-(\theta+\alpha+\beta) L_1^D(\theta)+\alpha\rho_0=0$$

proving (11.15). Then (11.22) gives

$$L_1(0,\vartheta)=\rho_1\frac{\alpha}{\vartheta+\alpha+\beta}$$

while inserting the expressions (11.28) for $L_2(\theta,0)$ and (11.15) for $L_1^D(\theta)$ into (11.23) yields

$$L_1(\theta,0)=\rho_1\frac{\alpha\beta}{(\theta+\alpha+\beta)(\theta+\beta)}.$$

Inserting this together with (11.15) into (11.18) implies (11.13).

It is now straightforward to complete the proof using (11.19), (11.20) and (11.21) and the findings so far to obtain (11.14) and (11.16). □

Remark 11.1.1 The partial Laplace transforms found in Theorem 11.1.1 are easily recognizable in particular when $\vartheta=0$. For instance one finds the following properties of the distribution of the time since last arrival A_{t_0} for an arbitrary t_0 when the system is in equilibrium (was started with initial distribution Φ):

(i) given $K_{t_0}=0$, the distribution of A_{t_0} is the convolution of two exponential distributions with rates α and β respectively;
(ii) given $K_{t_0}=p\geq 1$, the distribution of A_{t_0} is exponential at rate β (the service rate, not the arrival rate!);
(iii) combining (i) and (ii) it follows that A_{t_0} has Laplace transform

$$(1-\delta)\frac{\alpha\beta}{(\theta+\alpha)(\theta+\beta)}+\delta\frac{\beta}{\theta+\beta}=\frac{\alpha}{\theta+\alpha},$$

i.e., A_{t_0} is exponential at rate α (the arrival rate). Since (A_t) is the backward recurrence time process for a renewal process with rate α exponential interarrival times, this could also have been seen directly from Example 7.8.3.

Taking $\theta=0$ in (11.15) one finds

(iv) given $K_{t_0}=1$, the probability that this single customer arrived at an empty queue is $\frac{\beta}{\alpha+\beta}$.

The service time S_{t_0} in equilibrium is not so simple, but

(v) given $K_{t_0}=1$, S_{t_0} is exponential at rate $\alpha+\beta$;

and after some calculations

(vi) given $K_{t_0}\geq 1$, S_{t_0} is exponential at rate β.

(For the verification of (vi) and some of the other results above, remember that $p = 1$ is special and that $L_1(\theta, 0) + L_1^D(\theta)$ or $L_1(0, \vartheta) + L_1^D(\vartheta)$ should be used).

The description of the PDMP determining the GI/G/1 queue described above looked back in time: time since most recent arrival and time since present service started. But it is also possible to define another homogeneous PDMP $\widetilde{X}$ that looks forward in time by using as marks (the states reached by each jump) the time to the next arrival whenever a new arrival occurs and the time to the completion of service whenever a new service starts; cf. the discussion in Subsection 7.4.1 of the forward recurrence time process associated with a renewal process.

The states for $\widetilde{X}$ are either pairs $(0, \widetilde{a})$ (the queue empty, time $\widetilde{a} > 0$ left until the next arrival) or triples $(p, \widetilde{a}, \widetilde{s})$ for $p \geq 1$, $\widetilde{a} > 0$, $\widetilde{s} > 0$ (p customers in the queue including the customer presently being serviced, time $\widetilde{a}$ remaining until the next arrival, time $\widetilde{s}$ remaining of the service time for the customer being serviced). Formally $\widetilde{X}$ is described as the homogeneous PDMP with piecewise deterministic behaviour

$$\widetilde{\phi}_t(0, \widetilde{a}) = (0, \widetilde{a} - t) \quad (0 \leq t < \widetilde{a}),$$
$$\widetilde{\phi}_t(p, \widetilde{a}, \widetilde{s}) = (p, \widetilde{a} - t, \widetilde{s} - t) \quad (p \geq 1, 0 \leq t < \widetilde{a} \wedge \widetilde{s})$$

and with jump times $\widetilde{T}_n$ and marks $\widetilde{Y}_n = \widetilde{X}_{\widetilde{T}_n}$ whose joint distribution is determined by the following Markov kernels corresponding to an arbitrary initial state $\widetilde{x}_0$ where we write $\widetilde{z}_n = \left(\widetilde{t}_1, \ldots, \widetilde{t}_n; \widetilde{y}_1, \ldots, \widetilde{y}_n\right)$ with $\widetilde{y}_n = (0, \widetilde{a}_n)$ or $\widetilde{y}_n = (p_n, \widetilde{a}_n, \widetilde{s}_n)$ for $p \geq 1$: for $\widetilde{z}_n$ such that all $\widetilde{t}_{k+1} = \widetilde{t}_k + (\widetilde{a}_k \wedge \widetilde{s}_k)$ (no other $\widetilde{z}_n$ are relevant),

$$P^{(n)}_{\widetilde{z}_n|\widetilde{x}_0} = \begin{cases} \varepsilon_{\widetilde{t}_n + \widetilde{a}_n} & \text{if } p_n = 0, \\ \varepsilon_{\widetilde{t}_n + (\widetilde{a}_n \wedge \widetilde{s}_n)} & \text{if } p_n \geq 1, \end{cases}$$

and with $\widetilde{z}_n$ as above and $t = \widetilde{t}_n + (\widetilde{a}_n \wedge \widetilde{s}_n)$,

$$\pi^{(n)}_{\widetilde{z}_n,t|\widetilde{x}_0} = \begin{cases} \text{the distribution of } (1, U, V) & \text{if } p_n = 0, \\ \varepsilon_{(0, \widetilde{a}_n - \widetilde{s}_n)} & \text{if } p_n = 1,\ \widetilde{s}_n < \widetilde{a}_n, \\ \text{the distribution of } (p_n + 1, U, \widetilde{s}_n - \widetilde{a}_n) & \text{if } p_n \geq 1,\ \widetilde{a}_n < \widetilde{s}_n, \\ \text{the distribution of } (p_n - 1, \widetilde{a}_n - \widetilde{s}_n, V) & \text{if } p_n \geq 2,\ \widetilde{s}_n < \widetilde{a}_n, \end{cases}$$

where U and V are independent random variables such that U has hazard function α and V has hazard function β. Note that cases with $\widetilde{a}_n = \widetilde{s}_n$ are ignored as may be done since with continuous distributions for the interarrival and service times there is probability 0 that the time to next arrival ever equals the remaining service time (unless of course one deliberately chooses $\tilde{x}_0$ with $p_0 \geq 1$ and $\widetilde{a}_0 = \widetilde{s}_0$, so for convenience, assume that this is not the case). That $\widetilde{X}$ is indeed a homogeneous PDMP follows from Theorem 7.3.2 (b).

For the model just discussed, the service time for a customer was attached to that customer only when service started. Yet another version of a forward looking PDMP describing the GI/G/1 queue may be obtained by associating the service time for each customer with the customer immediately upon arrival. The state space and mark space now consists of pairs $(0, \widetilde{a})$ as above and, for $p \geq 1$, of vectors $\left(p, \widetilde{a}, \widetilde{s}_1, \ldots, \widetilde{s}_p\right)$ with p the number of customers in the queue, $\widetilde{a}$ the time remaining until the next arrival and $\left(\widetilde{s}_1, \ldots, \widetilde{s}_p\right)$ the remaining service times for the p customers in the queue listed in their order of arrival so that $\tilde{s}_1$ is the remaining time of service for the customer being serviced — and the $\widetilde{s}_k$ for $k \geq 2$ is the complete service time required to handle customer k. One now finds

$$\widetilde{\phi}_t\,(0, \widetilde{a}) = \widetilde{a} - t, \quad \widetilde{\phi}_t\left(p, \widetilde{a}, \widetilde{s}_1, \ldots, \widetilde{s}_p\right) = \left(p, \widetilde{a} - t, \widetilde{s}_1 - t, \widetilde{s}_2, \ldots, \widetilde{s}_p\right)$$

and, with $\widetilde{y}_n = \left(p_n, \widetilde{a}_n, \widetilde{s}_{1,n}, \ldots, \widetilde{s}_{p_n,n}\right)$,

$$P^{(n)}_{\widetilde{z}_n|\widetilde{x}_0} = \begin{cases} \varepsilon_{\widetilde{t}_n + \widetilde{a}_n} & \text{if } p_n = 0, \\ \varepsilon_{\widetilde{t}_n + (\widetilde{a}_n \wedge \widetilde{s}_{1,n})} & \text{if } p_n \geq 1, \end{cases}$$

whenever all $\widetilde{t}_{k+1} = \widetilde{t}_k + \left(\widetilde{a}_k \wedge \widetilde{s}_{1,k}\right)$, and when $t = \widetilde{t}_n + \left(\widetilde{a}_n \wedge \widetilde{s}_{1,n}\right)$,

$$\pi^{(n)}_{\widetilde{z}_n,t|\widetilde{x}_0} = \begin{cases} \text{the distribution of } (1, U, V) & \text{(i)}, \\ \varepsilon_{(0, \widetilde{a}_n - \widetilde{s}_{1,n})} & \text{(ii)}, \\ \text{the distribution of } \left(p_n + 1, U, \widetilde{s}_{1,n} - \widetilde{a}_n, \widetilde{s}_{2,n}, \ldots, \widetilde{s}_{p_n,n}, V\right) & \text{(iii)}, \\ \varepsilon_{(p_n - 1, \widetilde{a}_n - \widetilde{s}_n, \widetilde{s}_{2,n}, \ldots, \widetilde{s}_{p_n,n})} & \text{(iv)}, \end{cases}$$

with U and V as above and the four cases corresponding to (i) $p_n = 0$, (ii) $p_n = 1$, $\widetilde{s}_{1,n} < \widetilde{a}_n$, (iii) $p_n \geq 1$, $\widetilde{a}_n < \widetilde{s}_{1,n}$ and (iv) $p_n \geq 2$, $\widetilde{s}_{1,n} < \widetilde{a}_n$ respectively.

A particularly relevant feature of the corresponding PDMP $\widetilde{X}$ with

$$\widetilde{X}_t = \left(K_t, \widetilde{A}_t, \widetilde{S}_{1,t}, \ldots, \widetilde{S}_{K_t,t}\right)$$

is that it in a natural fashion incorporates the socalled *virtual waiting time*,

$$V_t = \sum_{k=1}^{K_t} \widetilde{S}_{k,t},$$

the remaining amount of work required to handle the customers in the system at time t.

11.2 Network models

An open *Jackson network* consists of a number $J \geq 1$ of nodes, each with a single server, with new (external) customers arriving at a randomly chosen node and moving randomly from node to node before leaving the system. Each visit to a node requires the customer to queue there before being serviced and moving on to the next node or exiting the network.

In its classical form, the network is modeled as a time-homogeneous Markov chain $\mathbf{K} = (K_1, \ldots, K_J)$ with state space $\mathbb{N}_0^J$. Here $K_{j,t}$ is the number of customers waiting at node j (including the customer being serviced) and the intensity matrix $Q = \left(q_{\mathbf{pp}'}\right)$ has the following structure: $q_{\mathbf{pp}} = -\sum_{\mathbf{p}' \neq \mathbf{p}} q_{\mathbf{pp}'}$ and for $\mathbf{p}, \mathbf{p}' \in \mathbb{N}_0^J$ with $\mathbf{p} \neq \mathbf{p}'$ the only non-zero off-diagonal elements of Q are

$$q_{\mathbf{p}\mathbf{p}'} = \begin{cases} \alpha_j & \text{if } \mathbf{p}' = \left[p_j + 1\right], \\ & \text{(external arrival at } j\text{)}, \\ \beta_j \gamma_j & \text{if } p_j \geq 1 \text{ and } \mathbf{p}' = \left[p_j - 1\right], \\ & \text{(completion of service and exit from } j\text{)}, \\ \beta_j \gamma_{jk} & \text{if } j \neq k,\ p_j \geq 1 \text{ and } \mathbf{p}' = \left[p_j - 1, p_k + 1\right], \\ & \text{(completion of service at } j,\ \text{move to } k\text{)}. \end{cases} \tag{11.29}$$

Notation. The bracket notation $[\cdot]$ is used to denote a vector $\mathbf{p}' = \left(p'_1, \ldots, p'_J\right)$ that differs from $\mathbf{p} = (p_1, \ldots, p_J)$ only on the coordinates indicated and in the way described within the brackets, e.g., $\mathbf{p}' = \left[p_j - 1, p_k + 1\right]$ is given by $p'_j = p_j - 1$, $p'_k = p_k + 1$ and all other $p'_l = p_l$.

In (11.29) α_j is the (external) arrival intensity at node j, β_j is the service intensity at j and γ_{jk} for $1 \leq j, k \leq J$ is the probability that a customer that has completed service at j moves on to k, while γ_j is the probability that a customer having completed service at j exits from the network. We shall require the parameters of the model to satisfy that all $\alpha_j \geq 0$, all $\beta_j > 0$, all $\gamma_j \geq 0$, $\gamma_{jk} \geq 0$, $\gamma_{jj} = 0$ with all $\gamma_j + \sum_k \gamma_{jk} = 1$ together with the free movement conditions that it is possible to exit eventually with the customer having visited any given node,

$$\forall j\ \exists n, j_1, \ldots, j_n : \gamma_{jj_1}\gamma_{j_1 j_2} \cdots \gamma_{j_{n-1} j_n}\gamma_{j_n} > 0, \tag{11.30}$$

and that it is possible for externally arriving customers to visit any node,

$$\forall j\ \exists n, j_0, \ldots, j_n : \alpha_{j_0}\gamma_{j_0 j_1}\gamma_{j_1 j_2} \cdots \gamma_{j_{n-1} j_n}\gamma_{j_n j} > 0.$$

The assumption (11.30) implies that the matrix $I - \Gamma$, where $\Gamma = \left(\gamma_{jk}\right)$, is non-singular. In particular the set

$$\zeta_j = \alpha_j + \sum_{k=1}^{J} \zeta_k \gamma_{kj} \quad (1 \leq j \leq J) \tag{11.31}$$

of equations has a unique solution vector $\left(\zeta_j\right)$ of so-called *throughput rates* with, in fact, all $\zeta_j > 0$. The following remarkably simple result then holds:

Proposition 11.2.1 *The open Jackson network* $\mathbf{K}$ *has an invariant distribution* ρ *if and only if* $\zeta_j < \beta_j$ *for all* $1 \leq j \leq J$, *and in this case*

$$\rho_{\mathbf{p}} = \prod_{j=1}^{J} \left(1 - \frac{\zeta_j}{\beta_j}\right) \left(\frac{\zeta_j}{\beta_j}\right)^{p_j} \quad \left(\mathbf{p} = (p_1, \ldots, p_J) \in \mathbb{N}_0^J\right). \tag{11.32}$$

In other words, with the network in equlibrium the distribution of any $\mathbf{K}_t$ is such that the coordinates $K_{j,t}$ are independent, each following a geometric distribution. (Taking $J = 1$, the result states that the M/M/1 queue has an invariant distribution iff $\alpha < \beta$ in which case ρ is geometric; cf. (11.12)).

We shall not give a detailed proof of Proposition 11.2.1, but various aspects of the result will be discussed below in a more general setup.

Exercise 11.2.1 Show that with (ζ_j) the solution to (11.31) and assuming that all $\zeta_j < \beta_j$, then ρ defined by (11.32) is invariant, i.e., show that

$$\sum_{\mathbf{p}' \in \mathbb{N}_0^J} \rho_{\mathbf{p}'} q_{\mathbf{p}'\mathbf{p}} = 0$$

for all $\mathbf{p} \in \mathbb{N}_0^J$.

In the network model discussed so far, external arrivals at each node j occurs according to a Poisson process with rate α_j with the J arrival processes independent. Also, all service times are independent and independent of the arrival processes, with the service times at node j exponential at rate β_j. We shall now discuss more general models that become homogeneous PDMPs when incorporating for each j, 'the time since last arrival' and 'the time since service began'; cf. the discussion of the GI/G/1 queue in the preceding section. Thus we shall obtain a process $\mathbf{X} = (\mathbf{K}, \mathbf{A}, \mathbf{S})$ with $\mathbf{K} = (K_j)_{1 \le j \le J}$ the queue lengths at the different nodes as before, and $\mathbf{A} = (A_j)_{1 \le j \le J}$ and $\mathbf{S} = (S_j)_{1 \le j \le J}$ where A_j is the process giving the time since the most recent 'arrival' at j, S_j is the process giving the time since the present service (if any, i.e., if $K_{j,t} \ge 1$) began. The state space G consists of elements $\mathbf{x} = (\mathbf{p}, \mathbf{a}, \mathbf{s})$ with $\mathbf{p} \in \mathbb{N}_0^J$, $\mathbf{a} \in \mathbb{R}_0^J$ and $\mathbf{s} \in \mathbb{R}_0^J$ when all $p_j \ge 1$ and s_j omitted if $p_j = 0$.

It is obvious that the deterministic behaviour of $\mathbf{X}$ is given by

$$\phi_t(\mathbf{p}, \mathbf{a}, \mathbf{s}) = (\mathbf{p}, \mathbf{a} + t\mathbf{1}, \mathbf{s} + t\mathbf{1})$$

writing $\mathbf{1}$ for the appropriate vector of 1's. Trivially the ϕ_t satisfy the semigroup property (7.28), so to complete the PDMP description it suffices to define the total intensity $q(\mathbf{p}, \mathbf{a}, \mathbf{s})$ for a jump and the jump probabilities $r((\mathbf{p}, \mathbf{a}, \mathbf{s}), \cdot)$.

The interpretation of $\mathbf{K}$, $\mathbf{A}$ and $\mathbf{S}$ makes it clear that from any $(\mathbf{p}, \mathbf{a}, \mathbf{s})$ it is only possible to jump to finitely many other states:

(i) if an external customer arrives at j,

$$(\mathbf{p}, \mathbf{a}, \mathbf{s}) \longrightarrow \begin{cases} \left(\left[p_j + 1\right], \left[0_j\right], \mathbf{s}\right) & \text{if } p_j \ge 1, \\ \left(\left[1_j\right], \left[0_j\right], \left[0_j\right]\right) & \text{if } p_j = 0; \end{cases}$$

(ii) if a customer completes service at j and then exits the network, provided $p_j \ge 1$,

$$(\mathbf{p}, \mathbf{a}, \mathbf{s}) \longrightarrow \begin{cases} \left(\left[p_j - 1\right], \mathbf{a}, \left[0_j\right]\right) & \text{if } p_j \ge 2, \\ \left(\left[0_j\right], \mathbf{a}, \mathbf{s}_{\setminus j}\right) & \text{if } p_j = 1; \end{cases}$$

(iii) if a customer completes service at j and then moves to $k \ne j$, provided $p_j \ge 1$, and if $p_k \ge 1$ (if $p_k = 0$ the service time at k in (11.33) and (11.34) should be reset to 0)

$$(\mathbf{p}, \mathbf{a}, \mathbf{s}) \longrightarrow \begin{cases} \left(\left[p_j - 1, p_k + 1\right], \mathbf{a}, \left[0_j\right]\right) & \text{if } p_j \ge 2, \\ \left(\left[0_j, p_k + 1\right], \mathbf{a}, \mathbf{s}_{\setminus j}\right) & \text{if } p_j = 1, \end{cases} \tag{11.33}$$

assuming that only external arrivals count as 'arrivals' when adjusting $\mathbf{a}$, while if all arrivals are included

$$(\mathbf{p}, \mathbf{a}, \mathbf{s}) \longrightarrow \begin{cases} \left(\left[p_j - 1, p_k + 1\right], [0_k], \left[0_j\right]\right) & \text{if } p_j \geq 2, \\ \left(\left[0_j, p_k + 1\right], [0_k], \mathbf{s}_{\setminus j}\right) & \text{if } p_j = 1. \end{cases} \tag{11.34}$$

Notation. The bracket notation used earlier is now used for each of the three components $\mathbf{p}$, $\mathbf{a}$ and $\mathbf{s}$. Also, $\mathbf{s}_{\setminus j}$ is the vector obtained from $\mathbf{s}$ by omitting s_j and retaining $s_{j'}$ for all $j' \neq j$.

With the jump structure prescribed, the q and r may be chosen quite arbitrarily (one may have to worry about stability though), but to simplify we shall now focus on the model where external arrivals at the nodes occur according to independent renewal processes with all the service times independent and independent of all the arrival processes, while the moves from node to node or from node to exit occurs using time-independent transition probabilities γ_{jk} and γ_j (not depending on $\mathbf{p}$ nor $\mathbf{a}$ nor $\mathbf{s}$) exactly as in the simple network discussed earlier. Apart from the γ_{jk} and γ_j, the model is described by the hazard functions $\alpha_j\left(a_j\right)$ for the distribution of the interarrival times at the nodes j and the hazard functions $\beta_j\left(s_j\right)$ for the distributions of the service times at the various j. Since in particular only external arrivals should affect the value of $\mathbf{A}$, we use (11.33) and not (11.34) and obtain

$$q(\mathbf{p}, \mathbf{a}, \mathbf{s}) = \sum_{j=1}^{J} \left(\alpha_j\left(a_j\right) + \beta_j\left(s_j\right) 1_{p_j \geq 1}\right) \tag{11.35}$$

and, with $\mathbf{x}_{\text{(i)}}$, $\mathbf{x}_{\text{(ii)}}$, $\mathbf{x}_{\text{(iii)}}$ the states reached by the jump described in cases (i), (ii) and (iii) above, for given values of j and $k \neq j$,

$$\begin{aligned} r\left((\mathbf{p}, \mathbf{a}, \mathbf{s}), \left\{\mathbf{x}_{\text{(i)}}\right\}\right) &= \frac{\alpha_j\left(a_j\right)}{q(\mathbf{p}, \mathbf{a}, \mathbf{s})}, \\ r\left((\mathbf{p}, \mathbf{a}, \mathbf{s}), \left\{\mathbf{x}_{\text{(ii)}}\right\}\right) &= \frac{\beta_j\left(s_j\right)\gamma_j}{q(\mathbf{p}, \mathbf{a}, \mathbf{s})} 1_{p_j \geq 1}, \\ r\left((\mathbf{p}, \mathbf{a}, \mathbf{s}), \left\{\mathbf{x}_{\text{(iii)}}\right\}\right) &= \frac{\beta_j\left(s_j\right)\gamma_{jk}}{q(\mathbf{p}, \mathbf{a}, \mathbf{s})} 1_{p_j \geq 1}. \end{aligned} \tag{11.36}$$

Exercise 11.2.2 Discuss the model using (11.34) instead of (11.33) but otherwise retaining (11.35) and (11.36). Discuss in particular how the occurrence of external arrivals is affected by the change. In the basic case discussed earlier where all the functions α_j and β_j are constant, show that $\mathbf{K}$ is the same Markov chain whether (11.33) or (11.34) is used.

Exercise 11.2.3 It is intuitively clear that with the model just defined (using (11.33)), $\mathbf{A} = \left(A_j\right)$ is itself a time-homogeneous PDMP with, assuming e.g., that $\mathbf{X}$ started from a given (but arbitrary) state, the coordinate processes A_j independent and each A_j a backward recurrence time process as in Subsection 7.4.1. Give a formal proof of this by finding the $\mathcal{F}_t^{\mathbf{X}}$-compensating measure for the RCM describing the jump times and jumps for $\mathbf{A}$.

Using (11.35) and (11.36) one finds that the infinitesimal generator for $\mathbf{X} = (\mathbf{K}, \mathbf{A}, \mathbf{S})$ has the form (cf. (7.64))

$$\begin{aligned} Ah(\mathbf{p}, \mathbf{a}, \mathbf{s}) = & \sum_{j=1}^{J} \left(D_{a_j} h(\mathbf{p}, \mathbf{a}, \mathbf{s}) + D_{s_j} h(\mathbf{p}, \mathbf{a}, \mathbf{s}) \right) \\ & + \sum_{j=1}^{J} \alpha_j (a_j) \left(h\left([p_j + 1], [0_j], \mathbf{s}^* \right) - h(\mathbf{p}, \mathbf{a}, \mathbf{s}) \right) \\ & + \sum_{j=1}^{J} \beta_j (s_j) \gamma_j \left(h\left([p_j - 1], \mathbf{a}, \mathbf{s}^{**} \right) - h(\mathbf{p}, \mathbf{a}, \mathbf{s}) \right) 1_{p_j \geq 1} \\ & + \sum_{j,k=1}^{J} \beta_j (s_j) \gamma_{jk} \left(h\left([p_j - 1, p_k + 1], \mathbf{a}, \mathbf{s}^{***} \right) - h(\mathbf{p}, \mathbf{a}, \mathbf{s}) \right) 1_{p_j \geq 1} \end{aligned} \tag{11.37}$$

where

$$\mathbf{s}^* = \begin{cases} \mathbf{s} & \text{if } p_j \geq 1 \\ [0_j] & \text{if } p_j = 0, \end{cases} \qquad \mathbf{s}^{**} = \begin{cases} [0_j] & \text{if } p_j \geq 2 \\ \mathbf{s}_{\backslash j} & \text{if } p_j = 1, \end{cases}$$

and

$$\mathbf{s}^{***} = \begin{cases} [0_j] & \text{if } p_j \geq 2,\ p_k \geq 1, \\ [0_j, 0_k] & \text{if } p_j \geq 2,\ p_k = 0, \\ \mathbf{s}_{\backslash j} & \text{if } p_j = 1,\ p_k \geq 1, \\ (\mathbf{s}_{\backslash j}, [0_k]) & \text{if } p_j = 1,\ p_k = 0. \end{cases}$$

The description of the network model $\mathbf{X} = (\mathbf{K}, \mathbf{A}, \mathbf{S})$ makes it possible to discuss when the network is in equilibrium, i.e., one may look for conditions ensuring that there is an invariant distribution. By Exercise 11.2.3 $\mathbf{A}$ is itself a PDMP, hence a necessary condition for an invariant distribution to exist for $\mathbf{X}$ is that $\mathbf{A}$ on its own has one, and from Example 7.8.3 it follows that this happens iff

$$\xi_{A,j} = \int_0^\infty \exp\left(- \int_0^a \alpha_j \right) da < \infty$$

for all j, as we assume from now on. Also, cf. Exercise 11.2.3, the invariant distribution for $\mathbf{A}$ is the product of the invariant distributions for the bcakward recurrence time processes A_j so that in equlibrium each $A_{j,t}$ for $t \geq 0$ has density

$$\frac{1}{\xi_{A,j}} \exp\left(- \int_0^a \alpha_j \right). \tag{11.38}$$

Now use (11.37) to compute Ah for the very special case $h(\mathbf{p}, \mathbf{a}, \mathbf{s}) = p_{j_0}$ for an arbitrary j_0. The end result is

$$Ah(\mathbf{p}, \mathbf{a}, \mathbf{s}) = \alpha_{j_0}(a_{j_0}) - \beta_{j_0}(s_{j_0}) 1_{p_{j_0} \geq 1} + \sum_{j=1}^{J} \beta_j (s_j) \gamma_{j j_0} 1_{p_j \geq 1}$$

and believing that if an invariant distribution ρ for $\mathbf{X}$ exists, then

$$\int_G \rho\,(d\mathbf{p},d\mathbf{a},d\mathbf{s})\ Ah\,(\mathbf{p},\mathbf{a},\mathbf{s}) = 0$$

(since h is not bounded and Ah need not be bounded, we cannot use Theorem 7.8.2 directly), one finds that for any t, writing

$$\zeta_j = \mathbb{E}_\rho\left[\beta_j\left(S_{j,t}\right); K_{j,t} \geq 1\right], \tag{11.39}$$

it holds that

$$\zeta_{j_0} = \frac{1}{\xi_{A,j_0}} + \sum_{j=1}^{J} \zeta_j \gamma_{jj_0}, \tag{11.40}$$

where it was used that by (11.38),

$$\mathbb{E}_\rho \alpha_{j_0}\left(A_{j_0,t}\right) = \int_0^\infty \alpha_{j_0}(a)\,\frac{1}{\xi_{A,j_0}} \exp\left(-\int_0^a \alpha_{j_0}\right) da = \frac{1}{\xi_{A,j_0}}.$$

The equations (11.40) for j_0 varying have exactly the same structure as (11.31), hence assuming a suitably adjusted version of the conditions for free movement on p. 288, there is a unique solution vector $\left(\zeta_j\right)$ of throughput rates with all $\zeta_j > 0$.

In the special case with all the functions α_j and β_j constant, we have $\xi_{j_0} = 1/\alpha_{j_0}$ and recover (11.31) and furthermore recognize, using (11.39), that the ζ_j from (11.31) satisfy

$$\zeta_j = \beta_j \mathbb{P}_\rho\left(K_{j,t} \geq 1\right).$$

Starting from any fixed initial state $\mathbf{x} = (\mathbf{p},\mathbf{a},\mathbf{s})$ it is easy to see that $\mathbb{P}_{|\mathbf{x}}\left(K_{j,t} = 0\right) > 0$ for any $t > 0$; hence we deduce as part of the assertions from Proposition 11.2.1 that if ρ exists necessarily $\zeta_j < \beta_j$ and in this case

$$\mathbb{P}_\rho\left(K_{j,t} = 0\right) = 1 - \frac{\zeta_j}{\beta_j}$$

in agreement with (11.32).

In the general case, ζ_j has an interesting interpretation: suppose that $\mathbf{X}$ starts from a given state $\mathbf{x}$, let $\mu = \sum_n \varepsilon_{(T_n,Y_n)}$ be the RCM that registers the jump times T_n and the marks (jumps)

$$Y_n = \mathbf{X}_{T_n} = \left(\mathbf{K}_{T_n}, \mathbf{A}_{T_n}, \mathbf{S}_{T_n}\right)$$

for $\mathbf{X}$, and consider for an arbitrary j, the counting process N_j^{arr} counting all arrivals (external and internal) at node j, i.e.

$$N_{j,t}^{\mathrm{arr}} = \int_{]0,t]\times G} 1_{(p_j = K_{j,u-}+1)}\, \mu\,(du,(d\mathbf{p},d\mathbf{a},d\mathbf{s}))\,.$$

By (7.32), (11.35) and (11.36) the $\mathcal{F}_t^{\mathbf{X}}$-compensator for N_j^{arr} is

$$\Lambda_{j,t}^{\text{arr}} = \int_0^t \left(\alpha_j \left(A_{j,u} \right) + \sum_{k=1}^{J} \beta_k \left(S_{k,u} \right) \gamma_{kj} 1_{(K_{k,u} \geq 1)} \right) du ,$$

and hence

$$\mathbb{E}_{|\mathbf{x}} N_{j,t}^{\text{arr}} = \int_0^t \left(\mathbb{E}_{|\mathbf{x}} \alpha_j \left(A_{j,u} \right) + \sum_{k=1}^{J} \mathbb{E}_{|\mathbf{x}} \left[\beta_k \left(S_{k,u} \right) ; K_{k,u} \geq 1 \right] \gamma_{kj} \right) du.$$

Assuming that the invariant distribution ρ exists and integrating $\mathbf{x}$ with respect to ρ and using (11.40) now yields the identity

$$\mathbb{E}_\rho N_{j,t}^{\text{arr}} = t \zeta_j$$

with the ζ_j given by (11.39). Thus ζ_j is the average time rate (throughput rate) at which customers, external as well as internal, arrive at j, and if these customers are to be handled successfully in equilibrium it is at least intuitively clear that in order for ρ to exist it is necessary (and with constant α_j, β_j, also sufficient according to Proposition 11.2.1) that

$$\zeta_j < \frac{1}{\xi_{S,j}}$$

for all j, where

$$\xi_{S,j} = \int_0^\infty \exp\left(- \int_0^s \beta_j \right) ds$$

is the expected service time for customers at node j.

We shall not discuss further the question of stationarity for the general network model, only note in passing that for establishing necessary and sufficient conditions for stationarity in a rigorous manner, Theorem 7.8.9 might be useful (but in contrast with the case for the GI/G/1 queue, the simpler Corollary 7.8.10 cannot be used). The main point of the discussion has been to emphasize that because the model is formulated as a homogeneous PDMP, it is immediately clear what it means for it to be in equilibrium, and to indicate at least, how the results from Section 7.8 may be used to study properties of the invariant probability when it exists.

Part III

Appendices

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 $K \in \mathbb{N}$, write $I_{k,K} = \left](k-1)/2^K, k/2^K\right]$ for $k = 1, 2, \ldots$. Also, for any $\mathbb{R}$-valued function g on $\mathbb{R}_0$, define for $k \in \mathbb{N}_0$, $K \in \mathbb{N}$, $g_{k,K} = g\left(\frac{k}{2^K}\right)$ and g_K as the function

$$g_K(t) = \sum_{k=1}^{\infty} 1_{I_{k,K}}(t) \frac{g_{k,K} - g_{k-1,K}}{A_{k,K} - A_{k-1,K}}$$

with the convention $\frac{0}{0} = 0$.

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 $\mathcal{G}_K = \sigma\left(I_{k,K}\right)_{k\geq 1}$. Then $\mathrm{E}_A\left[f \,|\mathcal{G}_K\right] = F_K$, hence $(F_K)_{K\geq 1}$ is a uniformly integrable martingale on $(\mathbb{R}_0, \mathcal{B}_0, A)$ converging A-a.e. (and in $L^1(A)$) to $\mathrm{E}_A\left[f \,|\mathcal{G}_\infty\right]$ where $\mathcal{G}_\infty$ is the smallest σ-algebra containing all the $\mathcal{G}_K$,i.e., $\mathcal{G}_\infty = \mathcal{B}_0$ so that $\mathrm{E}_A\left[f \,|\mathcal{G}_\infty\right] = f$.

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. $\square$

Now let $F : \mathbb{R}_0 \to \mathbb{R}$ be a cadlag function, but not a priori 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 $f = D_A F$. In particular, Proposition A.0.1 states that the integral (A.1) is differentiable with respect to A with derivative 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 there is a function* $g \geq 0$, *bounded on finite intervals, such that*

$$|F(t) - F(s)| \leq \int_{]s,t]} |g|\, dA \tag{A.3}$$

for all $s < t$.

Note. Clearly (A.3) is necessary for (A.2) to hold: take $g = f$.

Proof. Given t, write $t_K = \left[2^K t\right]/2^K$ (where $[x]$ is the integer part of x), $t'_K = t_K + \frac{1}{2^K}$. Using (A.3) we find for $s \leq t$,

$$|F_K(s)| \leq \sum_{k=1}^{\infty} 1_{I_{k,K}}(s) \frac{\int_{I_{k,K}} |g|\, dA}{A_{k,K} - A_{k-1,K}} \leq \sup_{[0,t]} |g|$$

and hence, by dominated convergence,

$$\int_{]0,t]} f\, dA = \lim_{K\to\infty} \int_{]0,t]} F_K\, dA.$$

But

$$
\begin{aligned}
\int_{]0,t]} F_K \, dA &= \sum_{k=1}^{\infty} \left(F_{k,K} - F_{k-1,K}\right) \frac{A\left(\frac{k}{2^K} \wedge t\right) - A\left(\frac{k-1}{2^K} \wedge t\right)}{A_{k,K} - A_{k-1,K}} \\
&= \sum_{k=1}^{[2^K t]} \left(F_{k,K} - F_{k-1,K}\right) + \left(F\left(t_K'\right) - F\left(t_K\right)\right) \frac{A(t) - A\left(t_K\right)}{A\left(t_K'\right) - A\left(t_K\right)} \\
&= F\left(t_K\right) - F(0) + \left(F\left(t_K'\right) - F\left(t_K\right)\right)\left(1 - \frac{A(t_K') - A\left(t\right)}{A\left(t_K'\right) - A\left(t_K\right)}\right) \\
&= F\left(t_K'\right) - F(0) - \left(F\left(t_K'\right) - F\left(t_K\right)\right) \frac{A\left(t_K'\right) - A(t)}{A\left(t_K'\right) - A\left(t_K\right)}
\end{aligned}
$$

where (i) above has been used for the second equality. Since F is right-continuous, $F\left(t_K'\right) \to F(t)$ 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

$$
\left| \left(F\left(t_K'\right) - F\left(t_K\right)\right) \frac{A\left(t_K'\right) - A(t)}{A\left(t_K'\right) - A\left(t_K\right)} \right| \le 2c \frac{A\left(t_K'\right) - A(t)}{\Delta A(t)} \to 0
$$

to again obtain convergence to $F(t) - F(0)$. $\square$

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

$$
D_A\left(F_1 F_2\right)(t) = \left(D_A F_1\right)(t) \, F_2(t) + F_1(t-) \left(D_A F_2\right)(t). \tag{A.4}
$$

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

$$
\left(F_1 F_2\right)(t) = \left(F_1 F_2\right)(0) + \int_{]0,t]} \left(\left(D_A F_1\right)(s) \, F_2(s) + F_1(s-) \left(D_A F_2\right)(s)\right) \, A(ds).
$$

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 $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $(\mathcal{F}_t)_{t \geq 0}$, the *filtration*, is an increasing family of sub σ-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 $\mathbb{P}$-null set is measurable: if $F_0 \in \mathcal{F}$, $\mathbb{P}(F_0) = 0$, then $F \in \mathcal{F}$ for any $F \subset F_0$.

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ may always be completed: define $\mathcal{N} = \{N \subset \Omega : \exists N_0 \in \mathcal{F}$ with $\mathbb{P}(N_0) = 0$ such that $N \subset N_0\}$, and let $\overline{\mathcal{F}}$ be the smallest σ-algebra containing $\mathcal{F}$ and $\mathcal{N}$. Then $\overline{\mathcal{F}} = \{F \cup N : F \in \mathcal{F},\ N \in \mathcal{N}\}$ and $\mathbb{P}$ extends uniquely to a probability $\overline{\mathbb{P}}$ on $(\Omega, \overline{\mathcal{F}})$ using the definition $\overline{\mathbb{P}}(\overline{F}) = \mathbb{P}(F)$ for any $\overline{F} \in \overline{\mathcal{F}}$ and any representation $\overline{F} = F \cup N$ of $\overline{F}$ with $F \in \mathcal{F}$, $N \in \mathcal{N}$. The probability space $(\Omega, \overline{\mathcal{F}}, \overline{\mathbb{P}})$ is complete and is called the *completion* of $(\Omega, \mathcal{F}, \mathbb{P})$.

A filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ satisfies the *usual conditions* if $(\Omega, \mathcal{F}, \mathbb{P})$ is complete, if $\mathcal{N} \subset \mathcal{F}_0$ where now $\mathcal{N} = \{N \in \mathcal{F} : \mathbb{P}(N) = 0\}$, 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. *In this book we do not make this assumption* and unless stated explicitly, it is not required in this appendix either. In the setup of the main text filtrations are automatically right-continuous, but they are not then completed: this is relevant in particular for Section 4.2 and Proposition 4.2.1 there, where the measurable structure on the canonical spaces of counting process paths and discrete counting measures is discussed.

Let $(G, \mathcal{G})$ be a measurable space. A *stochastic process* (in continuous time) with state space $(G, \mathcal{G})$, defined on $(\Omega, \mathcal{F}, \mathbb{P})$, 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

$$(t, \omega) \mapsto X_t(\omega)$$

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 σ-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 $t \mapsto X_t(\omega)$ is right-continuous for *all* ω. Similarly, X is *left-continuous, cadlag, increasing, continuous* if for all ω, $t \mapsto X_t(\omega)$ is respectively left-continuous, cadlag (right-continuous with left limits), increasing (in each of the $d-$ coordinates), continuous. We shall say that X is e.g., a $\mathbb{P}$-a.s. defined cadlag process if there exists $N \in \mathcal{F}$ with $\mathbb{P}(N) = 0$ such that $t \mapsto X_t(\omega)$ is cadlag for all $\omega \in \Omega' := \Omega \backslash N$. In that case X may be viewed as a true cadlag process in our sense when defined as a process on $\left(\Omega', \mathcal{F}', \mathcal{F}', \mathbb{P}'\right)$ where $\mathcal{F}' = \mathcal{F} \cap \Omega' = \left\{F \cap \Omega' : F \in \mathcal{F}\right\}$, $\mathcal{F}_t' = \mathcal{F}_t \cap \Omega'$ and $\mathbb{P}' = \mathbb{P}\left(\cdot \cap \Omega'\right)$.

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space. A process X defined on $(\Omega, \mathcal{F}, \mathbb{P})$ 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 X_0 is $\mathcal{F}_0$-measurable and the map $(t, \omega) \mapsto X_t(\omega)$ from $(\mathbb{R}_+ \times \Omega, \mathcal{B}_+ \otimes \mathcal{F})$ to $(G, \mathcal{G})$ is $\mathcal{P}$-measurable, where $\mathcal{P}$, the σ-algebra of *predictable sets*, is the sub σ-algebra of $\mathcal{B}_+ \otimes \mathcal{F}$ generated by the subsets of the form

$$]s, \infty[\times F \quad (s \in \mathbb{R}_0, F \in \mathcal{F}_s).$$

A predictable process is always measurable. Note that all sets of the form $]s, t] \times F$ or $]s, t[\times F$ for $0 \leq s < t$, $F \in \mathcal{F}_s$ are predictable. For the definition it is critically important that the intervals above are open at the left endpoint s.

If X, X' are two processes on $(\Omega, \mathcal{F}, \mathbb{P})$ with state space $(G, \mathcal{G})$, they are *versions* of each other if for all t, $\mathbb{P}(X_t = X_t') = 1$. They are *indistinguishable* if $F_0 := \bigcap_{t\geq 0}\left(X_t = X_t'\right) \in \mathcal{F}$ and $\mathbb{P}(F_0) = 1$.

Proposition B.0.3 *Let X be a process on $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ with state space $(G, \mathcal{G}) \subset (\mathbb{R}^d, \mathcal{B}^d)$.*

(i) *If X is right-continuous or left-continuous, then X is measurable.*
(ii) *If X is right-continuous and each X_t is $\mathcal{F}_t$-measurable, then X is adapted.*
(iii) *If X is left-continuous and X_t is $\mathcal{F}_t$-measurable for all t, then X is predictable.*

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 $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space and let M, X be real valued processes (state spaces $\subset (\mathbb{R}, \mathcal{B})$).

Definition B.0.1 *M is a* martingale *if for all t, $\mathbb{E}|M_t| < \infty$, M_t is $\mathcal{F}_t$-measurable and*

$$\mathbb{E}\left[M_t|\mathcal{F}_s\right] = M_s \quad (0 \leq s \leq t).$$

X is a submartingale *if for all t, $\mathbb{E}|X_t| < \infty$, X_t is $\mathcal{F}_t$-measurable and*

$$\mathbb{E}\,[X_t|\mathcal{F}_s] \geq X_s \quad (0 \leq s \leq t).$$

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 $(M_t, \mathcal{F}_t)$ 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 $\mathbb{E}|\varphi(M_t)| < \infty$ 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 $\mathbb{E}|\varphi(X_t)| < \infty$ for all t, then $(\varphi(X_t), \mathcal{F}_t)$ is a submartingale.*

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 $x > 0$*

$$\mathbb{P}\left(\sup_{s \in D} X_s > x\right) \leq \frac{1}{x}\mathbb{E}X_t^+$$
$$\mathbb{P}\left(\inf_{s \in D} X_s < -x\right) \leq \frac{1}{x}\left(\mathbb{E}X_t^+ - \mathbb{E}X_0\right).$$

(b) *If in addition X is right-continuous or left-continuous, for all $t \geq 0,\ x > 0$*

$$\mathbb{P}\left(\sup_{s \leq t} X_s > x\right) \leq \frac{1}{x}\mathbb{E}X_t^+$$
$$\mathbb{P}\left(\inf_{s \leq t} X_s < -x\right) \leq \frac{1}{x}(\mathbb{E}X_t^+ - \mathbb{E}X_0).$$

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

$$\begin{aligned}\beta_D(f; a, b) = \sup\{n \in \mathbb{N}_0 : \exists t_1 < t_2 < \cdots < t_{2n} \in D \\ \text{with } f(t_{2k-1}) < a < b < f(t_{2k}),\ 1 \leq k \leq n\}\end{aligned}$$

with $\sup \emptyset = 0$. The following analytic fact is a basic tool for establishing the main theorem on continuous time martingales. Here and in the statement of the theorem, the argument q is always understood to be rational, $q \in \mathbb{Q}_0$, such that e.g., $\lim_{q \to t}$ is a limit through $q \in \mathbb{Q}_0$.

Lemma B.0.6 *Let $f : \mathbb{Q}_0 \to \mathbb{R}$.*

(a) *The following two conditions are equivalent:*

(i) *the limits*

$$f(t+) := \lim_{q \to t,\ q > t} f(q), \quad f(t-) := \lim_{q \to t,\ q < t} f(q)$$

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) $\beta_{\mathbb{Q}_0 \cap [0,t]}(f; a, b) < \infty \quad (t \in \mathbb{R}_0,\ a < b \in \mathbb{R})$.

(b) *If (i), (ii) are satisfied, then the function* $t \mapsto f(t+)$ *from* $\mathbb{R}_0$ *to* $\overline{\mathbb{R}}$ *is cadlag.*

Note that (ii) holds for all $t \in \mathbb{R}_0,\ a < b \in \mathbb{R}$ iff it holds for all $t = N \in \mathbb{N},\ a < b \in \mathbb{Q}$, the latter imposing only a countable infinity of conditions.

In order to show that there are finitely many upcrossings, one uses the classical *upcrossing lemma:*

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 an* $\mathcal{F}_t$*-measurable random variable and*

$$\mathbb{E}\beta_D(X; a, b) \le \frac{1}{b-a}\mathbb{E}(X_t - a)^+.$$

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* $\mathbb{P}$*-almost all* ω *the limits*

$$M_{t+}(\omega) := \lim_{q \to t, q > t} M_q(\omega), \quad M_{t-}(\omega) := \lim_{q \to t, q < t} M_q(\omega)$$

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,

$$\mathbb{E}|M_{t+}| < \infty, \quad \mathbb{E}|M_{t-}| < \infty.$$

(ii) *For all* $t \ge 0,\ \mathbb{E}\left[M_{t+}|\mathcal{F}_t\right] = M_t$ *a.s., and for all* $0 \le s < t,\ \mathbb{E}\left[M_t|\mathcal{F}_{s+}\right] = M_{s+}$ *a.s. Moreover, for a given* $t \ge 0,\ M_{t+} = M_t$ *holds a.s. if one of the two following conditions is satisfied:*

$$\begin{aligned} &(*)\quad \mathcal{F}_{t+} = \mathcal{F}_t, \\ &(**)\ M \text{ is right-continuous in probability at } t. \end{aligned} \tag{B.1}$$

(iii) *The process* $M_+ := (M_{t+})_{t \ge 0}$ *defined by*

$$M_{t+}(\omega) = \begin{cases} \lim_{q \to t, q > t} M_q(\omega) & \text{if the limit exists and is finite} \\ 0 & \text{otherwise} \end{cases}$$

for all $\omega \in \Omega,\ t \ge 0$, *satisfies that all* M_{t+} *are* $\mathcal{F}_{t+}$*-measurable and that* $(M_{t+}, \mathcal{F}_{t+})$ *is a* $\mathbb{P}$*-a.s. defined cadlag martingale. Furthermore, if condition (B.2) (** in (B.1) is satisfied for every* $t \ge 0$, *then* M_+ *is a version of* M.

(iv) *If* $\sup_{t\geq 0}\mathbb{E}M_t^+ < \infty$ *or if* $\sup_{t\geq 0}\mathbb{E}M_t^- < \infty$, *then*

$$M_\infty := \lim_{q\to\infty} M_q = \lim_{t\to\infty} M_{t+}$$

exists a.s. and $\mathbb{E}|M_\infty| < \infty$.

(b) *Let* $X = (X_t, \mathcal{F}_t)$ *be a submartingale.*

(i) *For* $\mathbb{P}$*-almost all* ω *the limits*

$$X_{t+}(\omega) := \lim_{q\to t, q>t} X_q(\omega), \quad X_{t-}(\omega) := \lim_{q\to t, q<t} X_q(\omega)$$

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,

$$\mathbb{E}|X_{t+}| < \infty, \quad \mathbb{E}|X_{t-}| < \infty.$$

(ii) *For all* $t \geq 0$, $\mathbb{E}\left[X_{t+}|\mathcal{F}_t\right] \geq X_t$ *a.s., and for all* $0 \leq s < t$, $\mathbb{E}\left[X_t|\mathcal{F}_{s+}\right] \geq X_{s+}$ *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* $X_+ := (X_{t+})_{t\geq 0}$ *defined by*

$$X_{t+}(\omega) = \begin{cases} \lim_{q\to t, q>t} X_q(\omega) & \text{if the limit exists and is finite} \\ 0 & \text{otherwise} \end{cases}$$

for all $\omega \in \Omega$, $t \geq 0$, *satisfies that all* X_{t+} *are* $\mathcal{F}_{t+}$*-measurable and that* $(X_{t+}, \mathcal{F}_{t+})$ *is a* $\mathbb{P}$*-a.s. defined cadlag submartingale. Furthermore, if condition (B.2) (**) is satisfied for every* $t \geq 0$, *then* X_+ *is a version of* X.

(iv) *If* $\sup_{t\geq 0}\mathbb{E}X_t^+ < \infty$, *then*

$$X_\infty := \lim_{q\to\infty} X_q = \lim_{t\to\infty} X_{t+}$$

exists a.s. and $\mathbb{E}|X_\infty| < \infty$.

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.

From Theorem B.0.8 (a), one immediately obtains the following result:

Corollary B.0.9 *Let* $(M_t, \mathcal{F}_t)$ *be a cadlag martingale. Then* $(M_t, \mathcal{F}_{t+})$ *is also a cadlag martingale.*

A classical and important result from the general theory of processes is

Theorem B.0.10 (The Doob–Meyer decomposition theorem.) *Let us suppose that* $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ *satisfies the usual conditions (!) and let X be a cadlag local submartingale. Then there is up to indistinguishability a unique decomposition*

$$X = X_0 + M + A$$

where M is a cadlag local martingale with $M_0 = 0$ *a.s. and A is predictable, cadlag and increasing with* $A_0 = 0$ *a.s.*

Local martingale and local submartingales are defined in Definition B.0.2 below.

We next proceed with a brief discussion of stopping times and the optional sampling theorem.

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ 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 σ-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 $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.

In view of the importance of the right-continuous filtration $(\mathcal{F}_{t+})$ it should also be noted that τ is a stopping time iff

$$(\tau < t) \in \mathcal{F}_{t+} \quad (t \in \mathbb{R}_+),$$

and that

$$\mathcal{F}_\tau := \{F \in \mathcal{F} : F \cap (\tau < t) \in \mathcal{F}_{t+} \text{ for all } t \in \mathbb{R}_+\}.$$

(To see this, just use that $(\tau < t) = \bigcup_{n=1}^{\infty} (\tau < t - \frac{1}{n})$).

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.

If σ, τ are stopping times, so are $\sigma \wedge \tau$, $\sigma \vee \tau$, $\sigma + \tau$ and $h(\sigma, \tau)$ for any measurable $h : \overline{\mathbb{R}}_0^2 \to \overline{\mathbb{R}}_0$ satisfying $h(s, t) \geq s \vee t$. Also if $(\tau_n)_{n \geq 1}$ is an increasing or decreasing (everywhere on Ω) sequence of stopping times, then $\tau_\infty = \lim_{n\to\infty} \tau_n$ is also a stopping time.

Now let X be an $\mathbb{R}$-valued right-continuous process defined on $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ and let τ be a stopping time. Define the $\overline{\mathbb{R}}$-valued random variable X_τ by

$$X_\tau(\omega) = \begin{cases} X_{\tau(\omega)}(\omega) & \text{if } \tau(\omega) < \infty, \\ \lim\limits_{t\to\infty} X_t(\omega) & \text{if } \tau(\omega) = \infty \text{ and the limit exists in } \overline{\mathbb{R}}, \\ 0 & \text{if } \tau(\omega) = \infty \text{ and } \liminf_{t\to\infty} X_t(\omega) < \limsup_{t\to\infty} X_t(\omega). \end{cases}$$

For stopping times τ such that $\mathbb{P}(\tau = \infty) > 0$ the definition is used mainly for processes X such that the limit $X_\infty = \lim_{t\to\infty} X_t(\omega)$ exists a.s. and in that case

$$X_\tau = X_\infty \text{ a.s.} \quad \text{on} \quad (\tau = \infty) .$$

Lemma B.0.11 *If X is an $\mathbb{R}$-valued process which is right-continuous and adapted and τ is a stopping time, then X_τ is $\mathcal{F}_\tau$-measurable.*

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) $\sup_{i\in I} \mathbb{E}|U_i| < \infty$ and (ii) $\lim_{x\to\infty} \sup_{i\in I} \int_{(|U_i|>x)} |U_i|\, d\mathbb{P} = 0$. In particular, (U_i) is uniformly integrable if either (a) there exists a random variable $U \in L^1(\mathbb{P})$ such that $\mathbb{P}(|U_i| \le |U|) = 1$ for all i, or (b) there exists $p > 1$ such that (U_i) is bounded in $L^p(\mathbb{P})$: $\sup_{i\in I} \mathbb{E}|U_i|^p < \infty$.

Theorem B.0.12 (Optional sampling).

(a) *Let $(M_t, \mathcal{F}_t)$ be a cadlag martingale and let $\sigma \le \tau$ be stopping times.*
If either of the following two conditions (i), (ii) is satisfied, then $\mathbb{E}|M_\sigma| < \infty$, $\mathbb{E}|M_\tau| < \infty$ and

$$\mathbb{E}[M_\tau | \mathcal{F}_\sigma] = M_\sigma; \tag{B.3}$$

(i) *τ is bounded.*
(ii) *$(M_t)_{t\ge 0}$ is uniformly integrable.*

(b) *Let $(X_t, \mathcal{F}_t)$ be a cadlag submartingale and let $\sigma \le \tau$ be stopping times.*
If either of the following two conditions (i), (ii) are satisfied, then $\mathbb{E}|X_\sigma| < \infty$, $\mathbb{E}|X_\tau| < \infty$ and

$$\mathbb{E}[X_\tau | \mathcal{F}_\sigma] \ge X_\sigma; \tag{B.4}$$

(i) *τ is bounded.*
(ii) *$(X_t^+)_{t\ge 0}$ is uniformly integrable.*

In view of Corollary B.0.9 one might in part (a) have assumed from the start that $(M_t, \mathcal{F}_{t+})$ is a cadlag martingale, in which case (B.3) if $\sigma \equiv t_0$ would read $\mathbb{E}\left[M_\tau | \mathcal{F}_{t_0+}\right] = M_{t_0}$ with the a.s. identity $\mathbb{E}\left[M_\tau | \mathcal{F}_{t_0}\right] = M_{t_0}$ an immediate consequence.

Note that if (aii) (or (bii)) holds, then $M_\infty = \lim_{t\to\infty} M_t$ ($X_\infty = \lim_{t\to\infty} X_t$) exists a.s. (and in the case of M_∞ also as a limit in $L^1(\mathbb{P})$). 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 $\sigma \le \tau$. In particular, if (aii) holds we may take $\tau \equiv \infty$ in (B.3) and obtain

$$\mathbb{E}[M_\infty | \mathcal{F}_\sigma] = M_\sigma$$

for all stopping times σ.

We finally need to discuss stopped processes and local martingales.

Let X be an $\mathbb{R}$-valued, adapted, cadlag process and let τ be a stopping time. Then X^τ, *X stopped at τ*, is the process given by

$$X_t^\tau := X_{\tau \wedge t}.$$

Stopping preserves martingales: if M is a cadlag $\mathcal{F}_t$-martingale and τ is a stopping time, then M^τ is also an $\mathcal{F}_t$-martingale. (Note that by optional sampling, for $s < t$

$$\mathbb{E}\left[M_t^\tau \,|\mathcal{F}_{\tau \wedge s}\right] = M_s^\tau .$$

To obtain the stronger result

$$\mathbb{E}\left[M_t^\tau \,|\mathcal{F}_s\right] = M_s^\tau ,$$

one shows that if $F \in \mathcal{F}_s$, then $F \cap (\tau > s) \in \mathcal{F}_{\tau \wedge s}$ and therefore

$$\begin{aligned}\int_F M_t^\tau \, d\mathbb{P} &= \int_{F\cap(\tau>s)} M_t^\tau \, d\mathbb{P} + \int_{F\cap(\tau\le s)} M_\tau \, d\mathbb{P}\\ &= \int_{F\cap(\tau>s)} M_s^\tau \, d\mathbb{P} + \int_{F\cap(\tau\le s)} M_s^\tau \, d\mathbb{P}\\ &= \int_F M_s^\tau \, d\mathbb{P}).\end{aligned}$$

Definition B.0.2 *An adapted,* $\mathbb{R}$*-valued cadlag process* M *is a* local $\mathcal{F}_t$-martingale *(an adapted,* $\mathbb{R}$*-valued cadlag process*X *is a* local $\mathcal{F}_t$-submartingale*) if there exists a sequence* $(\sigma_n)_{n\ge 1}$ *of stopping times, increasing to* ∞ *a.s., such that for every* n, M^{σ_n} *is a cadlag* $\mathcal{F}_t$ *-martingale (*X^{σ_n} *is a cadlag* $\mathcal{F}_t$*-submartingale).*

That (σ_n) increases to ∞ a.s. means that for all n, $\sigma_n \le \sigma_{n+1}$ a.s. and that $\lim_{n\to\infty} \sigma_n = \infty$ a.s. The sequence (σ_n) is called a *reducing sequence* for the local martingale M, and we write that M is a local $\mathcal{F}_t$ -martingale (σ_n) .

Clearly any martingale is also a local martingale (use $\sigma_n \equiv \infty$ for all n). If (σ_n) is a reducing sequence and (ρ_n) is a sequence of stopping times increasing to ∞ a.s., since $(M^{\sigma_n})^{\rho_n} = M^{\sigma_n \wedge \rho_n}$ it follows immediately that $(\sigma_n \wedge \rho_n)$ 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.0.13 *Let* M *be a local* $\mathcal{F}_t$*-martingale. For* M *to be an* $\mathcal{F}_t$*-martingale it is sufficient that for all* t,

$$\mathbb{E} \sup_{s:s\le t} |M_s| < \infty. \tag{B.5}$$

Warning. A local martingale M need not be a martingale even though $\mathbb{E}\,|M_t| < \infty$ for all t. There are even examples of local martingales M that are not martingales although the exponential moments $\mathbb{E}\exp(\theta\,|M_t|)$ are $< \infty$ for *all* $t, \theta \ge 0$. 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.

Bibliographical Notes

General

Doob [35] is a major classical and still relevant reference on stochastic process theory. Also classical is Karlin and Taylor [75] and [76] with discussion of a host of specific types of models. For 'le théorie generale des processus' which although not used in this book is at the core of the modern treatment of processes, Dellacherie and Meyer [33] (the first eight chapters of their four volume work) is basic; a very fine overview is given in Chapter I of Jacod and Shiryaev [65]. More recent comprehensive treatments of stochastic processes are Gikhman and Skorokhod [45] and Rogers and Williams [104] and, at a more basic level, Resnick [102] and Borovkov [13]. Two important monographs focusing on general semimartingale theory are Liptser and Shiryaev [85] and Protter [100].

Chapters 2 and 3

The construction of canonical simple and marked point processes given here follows and extends the construction from Jacobsen [53] and [56]. An important exposition of MPP theory, relying more on the general process theory than is done here, is Last and Brandt [81]. Major references on general point processes (not necessarily indexed by time) and random measures are Kallenberg [73] and Daley and Vere-Jones [27]. Reiss [101] demonstrates the relevance of point process theory to various applications in statistics and stochastic geometry.

Chapter 4

Hazard measures are mostly discussed and used for absolutely continuous distributions, described by their hazard function. The general version appears e.g. in Chou and Meyer [18], Jacod [60] (Lemma 3.5) or Appendix A5 of Last and Brandt [81]. It was used also in Jacobsen [53]. The structure of adapted and predictable processes on the canonical spaces presented in Proposition 4.2.1 is certainly part of the folklore, but is not widely used. Compensators and compensating measures appeared with the discussion of Doob-Meyer decomposition of counting processes (which are trivial examples of submartingales), see e.g. Boel et al [12] and the very important paper Jacod [60] where compensating measures for marked point processes on general filtered spaces are introduced. The key lemma 4.3.3 is certainly well known, but perhaps not often used as explicitly and to the same extent as is done in this book. Major references on general coupling constructions such as those underlying Proposition 4.3.5 and Corollary 4.4.4 are Lindvall [83] and Thorisson [118]; for constructions referring directly to point processes and processes with jumps, see e.g. Whitt [120] and Kwieciński and Szekli [80]; also Section 9.3 of Last and Brandt [81] contains explicit coupling constructions of 'jump processes'. The technique used in Sections 4.5 and 4.6 to show that e.g. the 'basic martingales' are indeed (local) martingales, is taken from Jacobsen [53] but has its origin in the study of 'one-jump martingales', see Section 1, 'le cas elementaire' of Chou and Meyer [18]. The martingale representation theorem, Theorem 4.6.1, goes back to the mid 70s, e.g. Chou and Meyer [18], Jacod [60], Boel et al [12] and Davis [28]; more general representations are in Jacod's papers [61] and [62]. Theorem 4.7.1 (Itô's formula) relies on the explicit structure of adapted and predictable processes on the canonical spaces, and is therefore rarely seen in this form. From the point of view of the general theory of stochastic integration, the integrals from Section 4.6 are elementary since they can be defined directly. Excellent references for the general theory, including general forms of Itô's formula, are Rogers and Williams [104] and Protter [100]. Section 4.8 on MPPs on general filtered probability spaces is quite brief, for a much more extensive treatment see Last and Brandt [81].

There are several important topics from general point process theory that are not treated in this book: (i) Palm probabilities, i.e. the description of the 'conditional' distribution of an MPP say, given that one of the jump times has a given value; this is non-trivial since there is no interpretation within the world of ordinary conditional distributions; see e.g. Section 1.3 of Franken et al [43], Chapter 10 of Kallenberg [73], Chapter 12 (first edition) of Daley and Vere-Jones [27] or Section 1.3 of Baccelli and Brémaud [7]. (ii) stationarity (which is discussed for piecewise deterministic Markov processes in Chapter 7), see e.g. Sigman [114] and Thorisson [118]. (iii) thinning of point processes, i.e. the omission of points from a point process by random selection, see e.g. Chapter 8 of Kallenberg [73] or Section 9.3 of Daley and Vere-Jones [27]. (iv) convergence of point processes, see e.g. Chapter 4 of Kallenberg [73] or Chapter 9 of Daley and Vere-Jones [27] or Chapters V and VIII of Jacod and Shiryaev [65].

Chapter 5

Change of measure formulas originated with Girsanov [46] with van Schuppen and Wong [107] giving an early generalisation of Girsanov's 'change of drift' formula. For MPPs, early references are Boel et al [12] and Jacod [60], and for certain processes with jumps, Segall and Kailath [109]. More general results were given by Jacod and Mémin [64] and in a series of papers [69], [70] and [71] by Kabanov, Liptser and Shiryaev; see also Section III.5 of Jacod and Shiryaev [65]. Most of the formulas are examples of the socalled Doleans-Dadé exponentials, Doleans-Dadé [34], see also

Subsection I.4f of Jacod and Shiryaev [65]. It is well known how to change measure using non-negative martingales as in Theorem 5.2.1(i); the version involving local martingales, Theorem 5.2.1(ii), is possibly new.

Chapter 6

The discussion of independence in Section 6.1 is not standard and traditionally the emphasis is more on superposition of independent point processess, e.g. Section 9.2 of Daley and Vere-Jones [27] or Section 1.5.1 of Baccelli and Brémaud [7]. Poisson processes and their compensators are thoroughly studied and used as prime examples of point processes in any textbook, but the curtailed Poisson processes (Definition 6.2.3(ii)) are certainly less well known, as is their relevance for the understanding of counting processes with deterministic compensators. The Lévy processes appearing in Section 6.2 are elementary in the sense that they have bounded Lévy measure. Two recent and essential references on the theory of general Lévy processes are Bertoin [8] and Sato [108]. Infinitely divisible point processes is a related topic not treated in this book but is the theme of Matthes, Kerstan and Mecke [89], see also Chapters 6 and 7 of Kallenberg [73], Section 8.4 of Daley and Vere-Jones [27].

Chapter 7

Classical expositions of the theory of homogeneous Markov processes are [38] and Blumenthal and Getoor [11]. Sharpe [111] is more recent and exploits general process theory to the full. Non-homogeneous processes are not given the same attention, but an early reference is Dynkin [37]. Chung [19] is the classic on homogeneous Markov chains on a countable state space. Non-homogeneous chains are well studied when they have transition intensities; for chains that do not, see Jacobsen [52]. The fundamental references for the definition and theory of PDMPs are the paper [29] and the monograph [30] by Mark Davis. An even earlier little known dissertation Wobst [122] contains possibly the first definition of homogeneous PDMPs. Davis' work deals only with homogeneous PDMPs where, apart from certain forced jumps, all conditional jump time distributions have densities. The paper Jacod and Skorokhod [66], which contains a vast amount of information on homogeneous PDMPs in particular, dispense with the density assumption for their discussion of 'jumping Markov processes' and 'jumping filtrations' and for homogeneous PDMPs Theorem 7.3.2(b) is contained in their Theorems 5 and 6. Example 7.3.1 is from Jacobsen and Yor [59] who introduced the concept of 'multi-self-similarity'. The strong Markov property is contained in Section 25 of Davis [30]. The 'full infinitesimal generator' of Section 7.7 is essentially the same as the 'extended generator' characterized by Davis [30], Theorem (26.14), but he does not use the terminology 'path-continuity' and 'path-differentiability'. Jacod and Skorokhod [66] also include a subsection on general infinitesimal generators, martingale structures etc. for the processes they consider. Itô's formula for homogeneous PDMPs is found in Theorem (31.3) of Davis [30]. Theorem 7.8.2 generalizes with the proper definition of the generator and its domain to virtually all homogeneous Markov processes. The remainder of the material in Section 7.8 has been prepared explicitly for this book, but is inspired by the technique of proving stationarity for processes in general if they have a suitable regenerative structure, see e.g. Asmussen [3], Sigman [114] and Thorisson [118]. For other results on stationarity of PDMPs, see e.g. Costa [21] and Dufour and Costa [36]. Convergence to stationarity is not treated in Section 7.8, for how to do this using coupling methods, see the references just quoted and Lindvall [83]. For stationarity of general state space homogeneous Markov chains in discrete time, the indispensable reference is Meyn and Tweedie [93]; in [94] and [95] they transfer some of their results to continuous time processes. The change of measure martingale from Example 7.9.1 is taken from Subsection 5.2 of Palmowski and Rolski [98].

Among all piecewise deterministic processes, in this book the focus has been on PDMPs. Other classes of processes with few jumps, and which are piecewise deterministic, include Markov renewal processes and semi-Markov processes, see e.g. Asmussen [3] 2nd edn, Section VII.4, and Markov additive models, ibid. Chapter XI. The latter includes certain hidden Markov processes in continuous time; a very good reference on hidden Markov models in discrete time is MacDonald and Zucchini [86].

The different chapters in Part II of the book serve to illustrate how the theory from Part I may be used to study and understand different topics in mathematical statistics and applied probability. This means that the contents appear more as examples rather than surveys of the different topics, hence the references below are primarily aimed at further reading.

Chapter 8, Survival Analysis

The use of martingale methods to define and study estimators of e.g. integrated hazard functions go back to the fundamental paper Aalen [1]. Standard references on the use of these and other probabilistic methods in survival analysis are Fleming and Harrington [41] and the very comprehensive book [2] by Andersen, Borgan, Gill and Keiding. For a detailed discussion of censoring patterns that agree with a nice martingale structure, see Jacobsen [55]. The Cox regression model, which is absolutely essential for survival analysis, is due to Cox [24]. An excellent reference for the analysis of multivariate data is Hougaard [50], where more elaborate and sophisticated models are treated than the ones discussed here, including models for longitudinal data, multistate models and frailty models. For the use of point process methods and martingale techniques in the analysis of longitudinal data, see e.g. Scheike [106] and Martinussen and Scheike [87] and [88].

A main reference for the theory of statistical inference for point processes on general spaces is Karr [78] and, for stochastic processes in general, Liptser and Shiryaev [84].

Chapter 9

A branching process

The model discussed in Section 9.1 is a simple and special case of the classical type of branching processes referred to as Crump-Mode-Jagers processes, where in particular a life history of a very general nature is attached to each individual, see e.g. Jagers [68]. Such a life history may include all sorts of characteristics of an individual and not only the ages at which she gives birth and the age at death as in the example in this book. Of particular interest are the socalled Malthusian populations which exhibit an exponential expected growth, and where martingales may be used to evaluate the a.s. asymptotic behaviour of the population when time t tends to ∞. Modern branching process theory deals not only with the classical type of models, but also has close relations with the theory of random fields, random trees and particle systems. The collection of papers edited by Athreya and Jagers [5] shows the variety of topics treated by the modern theory. Among the most important classical monographs on branching processes are Athreya and Ney [6] and Jagers [67].

Ruin probabilities

Ruin probabilities are now studied for very general risk models, see e.g. the comprehensive monographs Rolski et al [105] and Asmussen [4]. The results often involve approximations and only a few exact results are known, especially when it comes to the discussion of the distribution of the time to ruin. Combining martingales with optional sampling is a standard technique and the first to exploit this for a model with undershoot was Gerber [44], who found the ultimate ruin probability in the basic risk model with Poisson arrivals and iid exponential claims, cf. Proposition 9.2.3. Recent and much more general exact results based on martingale techniques can be found in Jacobsen [57] and [58]. Both monographs Rolski et al [105] and Asmussen [4] include discussions of the use of martingale techniques for solving ruin problems, but the focus is on other methods yielding approximate or exact results; in particular, for risk models where claims arrive according to a renewal process, random walk techniques including Wiener-Hopf factorizations prove especially powerful. Also important are renewal theory techniques, which can be used to derive integro-differential equations for quantities relating to ruin time distributions, and may also lead to exact expressions for the socalled 'double Laplace transform' for the time to ruin (with one integration performed with respect to the initial state of the risk process).

The soccer model

This is, apart possibly from the special case of the multiplicative Poisson model (see (9.17)), definitely the author's own invention! The multiplicative model was used as an example in Jacobsen [54], and with an analysis of data from the English premier league as Example 10.21 in Davison [31] (with the model refined to accomodate for home team advantage).

Chapter 10, Finance

Most of the literature on mathematical finance deals with diffusion models or diffusions with jumps. Cont and Tankov [20] is a very recent monograph on jump processes and finance with the focus on models involving Lévy processes (with, in general, unbounded Lévy measures) and allowing also for a continuous part. The model from Chapter 10 was chosen to illustrate how the MPP and PDMP theory could be used for understanding concepts from mathematical finance in a simple jump process model with the exposition essentially self-contained. The book Pliska [99] on discrete time models served as the main inspiration. An early study of a model from finance with jumps (diffusion with jumps) is given by Björk et al [10]. Otherwise all the definitions and main results from Chapter 10 are of course standard with the results valid in much greater generality, see e.g. Harrison and Pliska [49] and Delbaen and Schachermayer [32]. Apart from these names, other good sources are the work of e.g. Björk, Eberlein, Föllmer, Madan, Runggaldier and Schweizer. Recent monographs on mathematical finance include Björk [9], Föllmer and Schied [42], Karatzas and Shreve [74], Melnikov [91], Musiela and Rutkowski [96] and Shiryaev [113]. And with the emphasis more on the economic aspects, Merton [92] must be mentioned.

Chapter 11, Queueing

Some pertinent references on the huge literature on queueing theory are Franken et al [43], Brémaud [15], Asmussen [3], Kalashnikov [72], Baccelli and Brémaud [7] and Robert [103]. Jackson networks were introduced by Jackson [51]. More recent treatments of Jackson and other networks in queueing theory include Kelly [79], Walrand [119], Serfozo [110], Chao et al [16] and Chen and Yao [17]. A particularly important issue in queueing theory is concerned with stability, and here one often uses methods from random walk theory, regenerative processes (Asmussen [3]) and coupling techniques (Baccelli and Brémaud [7]) rather than, as was done in Chapter 11, attempting to view a given queueing system as a homogeneous Markov process and then checking for the existence of an invariant distribution. A particularly interesting technique for studying stability involves the existence of the socalled ‘fluid limits’, see e.g. Dai [25] and Dai [26] which are also discussed in Robert [103], Chapter 9. Theorem 11.1.1 on the three-dimensional invariant distribution for the M/M/1 queue is most likely known, but the author does not have a reference.

References

1. Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6,** 701–726.
2. Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1993). *Statistical Models based on Counting Processes.* Springer, New York.
3. Asmussen, S. (1987). *Applied Probability and Queues.* Wiley, Chichester. 2nd edition, Springer, New York (2003).
4. Asmussen, S. (2000). *Ruin Probabilities.* World Scientific, Singapore.
5. Athreya, K.B. and Jagers, P. eds. (1997). *Classical and Modern Branching Processes.* Springer, New York.
6. Athreya, K.B. and Ney, P.E. (1972). *Branching Processes.* Springer, Berlin.
7. Baccelli, F. and Brémaud, P. (1994). *Elements of Queueing Theory.* Springer, New York.
8. Bertoin, J. (1996). *Lévy Processes.* Cambridge University Press, Cambridge.
9. Björk, T. (1998). *Arbitrage Theory in Continuous Time.* Oxford University Press, Oxford.
10. Björk, T., Kabanov, Y. and Runggaldier, W. (1997). Bond market structure in the presence of marked point processes. *Math. Finance* **7**, 211–239.
11. Blumenthal, R.M. and Getoor, R.K. (1968). *Markov Processes and Potential Theory.* Academic Press, New York.
12. 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.
13. Borovkov, K. (2003). *Elements of Stochastic Modelling.* World Scientific, River Edge.
14. Brandt, A. and Last, G. (1994). On the pathwise comparison of jump processes driven by stochastic intensities. *Math. Nachr.* **167,** 21–42.
15. Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics.* Springer, Berlin.
16. Chao, X., Miyazawa, M. and Pinedo, M. (1999). *Queueing Networks. Customers, Signals and Product Form Solutions.* Wiley.
17. Chen, H. and Yao, D. (2001). *Fundamentals of Queueing Networks: Performance, Asymptotics and Optimization.* Springer, New York.
18. Chou, C.S. and Meyer, P.A. (1975). Sur la répresentation des martingales comme intégrales stochastique dans les processus ponctuels. In *Séminaire de Probabilités IX.* Lecture Notes in Mathematics 465, Springer, Berlin, 226–236.
19. Chung, K.L. (1960). *Markov Chains with Stationary Transition Probabilities.* Springer, New York. 2nd edition (1967).
20. Cont, R. and Tankov, P. (2004). *Financial Modelling with Jump Processes.* Chapman & Hall/CRC, Boca Raton.

21. Costa, O.L.V. (1990). Stationary distributions for piecewise-deterministic Markov processes. *J. Appl. Probab.* **27,** 60–73.
22. Cox, D.R. (1955). The analysis of non-Markovian stochastic processes by the inclusion of supplementary variables. *Proc. Camb. Phil. Soc.* **51,** 433–441.
23. Cox, D.R. (1962). *Renewal Theory.* Methuen, London.
24. Cox, D.R. (1972). Regression models and lifetables. *J. Roy. Statist. Soc. Ser. B* **34,** 187–220 (with discussion).
25. Dai, J. (1995a). On positive Harris recurrence of multiclass queueing networks: A unified approach via fluid limits. *Ann. Appl. Probab.* **5**, 49–77.
26. Dai, J. (1995b). Stability of open queueing networks via fluid models. In *Stochastic Networks* (F.P. Kelly and R.J. Williams, eds.). Springer, New York, 71–90.
27. Daley, D.J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes.* Springer, New York. 2nd edition Vol. I, (2003).
28. Davis, M.H.A. (1976). The representation of martingales of jump processes. *SIAM J. Control* **14,** 623–638.
29. 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.
30. Davis, M.H.A. (1993). *Markov Models and Optimization.* Chapman & Hall, London.
31. Davison, A.C. (2003). *Statistical Models.* Cambridge University Press, Cambridge.
32. Delbaen, F. and Schachermeyer, W. (1995). The existence of absolutely continuous local martingale measures. *Ann. Appl. Probab.* **5**, 926–945.
33. Dellacherie, C. and Meyer, P.A. (1976,1980). *Probabilités et potentiel, Ch. I-IV, Ch. V-VIII (Théorie des martingales).* Hermann, Paris.
34. Doléans-Dade, C. (1970). Quelques applications de la formule de changement de variable pour les semimartingales. *Z. Wahrsch. verw. Gebiete* **16,** 181–194.
35. Doob, J.L. (1953). *Stochastic Processes.* Wiley, New York.
36. Dufour, F. and Costa, O.L.V. (2002). Stability of piecewise deterministic Markov processes. *SIAM J. Control Optim.* **37,** 1483–1502.
37. Dynkin, E.B. (1961). *Die Grundlagen der Theorie der Markoffschen Prozesse.* Springer, Berlin.
38. Dynkin, E.B. (1965). *Markov Processes I-II.* Springer, Berlin.
39. Elliott, R.S. (1982). *Stochastic Calculus and Applications.* Springer, New York.
40. Ethier, S.N. and Kurtz, T.G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.
41. Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis.* Wiley, New York.
42. Föllmer, H. and Schied, A. (2002). *Stochastic Finance.* de Gruyter, Berlin.
43. Franken, P., König, D., Arndt, U. and Schmidt, V. (1982). *Queues and Point Processes,* Akademie-Verlag, Berlin and Wiley, Chichester.
44. Gerber, H.E. (1979). *An Introduction to Mathematical Risk Theory.* Huebner Foundation Monographs, Wharton School, Philadelphia.
45. Gikhman, I.I. and Skorokhod, A.V. (1974,1975,1979). *The Theory of Stochastic Processes I-III.* Springer, Berlin. 2nd edition, Vol. I, (1980).
46. Girsanov, I.V. (1960). On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory Probab. Appl.* **5,** 285–301.
47. Grandell, J. (1991). *Aspects of Risk Theory.* Springer, Berlin.
48. Grigelionis, B. (1973). On non-linear filtering theory and absolute continuity of measures, corresponding to stochastic processes. In *Proc. 2nd Japan-USSR Symposium on Probability.* Lecture Notes in Mathematics 330, 80–94, Springer, Berlin.

49. Harrison, J.M. and Pliska, S.R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stoch. Proc. Appl.* **11,** 215–260.
50. Hougaard, P. (2000). *Analysis of Multivariate Survival Data.* Springer, New York.
51. Jackson, J.R. (1957). Networks of waiting lines. *Operations Research* **5**, 518–521.
52. Jacobsen, M. (1972). A characterization of minimal Markov jump processes. *Z. Wahrsch. verw. Gebiete* **23,** 32–46.
53. Jacobsen, M. (1982). *Statistical Analysis of Counting Processes.* Lecture Notes in Statistics 12, Springer, New York.
54. Jacobsen, M. (1989a). Existence and unicity of MLEs in discrete exponential family distributions. *Scand. J. Statist.* **16**, 335–349.
55. Jacobsen, M. (1989b). Right censoring and martingale methods for failure time data. *Ann. Statist.* **17,** 1133–1156.
56. Jacobsen, M. (1999). *Marked Point Processes and Piecewise Deterministic Processes.* Lecture Notes no 3, Centre for Mathematical Physics and Stochastics (MaPhySto), Aarhus.
57. Jacobsen, M. (2003). Martingales and the distribution of the time to ruin. *Stoch. Proc. Appl.* **107,** 29–51.
58. Jacobsen, M. (2005). The time to ruin for a class of Markov additive processes. *Adv. Appl. Probab.* (to appear).
59. Jacobsen, M. and Yor, M. (2003). Multi-self-similar Markov processes on $\mathbb{R}_+^n$ and their Lamperti representations. *Prob. Theory Relat. Fields* **126,** 1–28.
60. Jacod, J. (1975). Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrsch. verw. Gebiete* **31,** 235–253.
61. Jacod, J. (1976). Un théorème de représentation pour les martingales discontinues. *Z. Wahrsch. verw. Gebiete* **34,** 225–244.
62. Jacod, J. (1977). A general theorem of representation of martingales. *Proc. Symp. Pure Math.* **31,** 37–53.
63. Jacod, J. (1979). *Calcul stochastique et problèmes de martingales.* Lecture Note in Mathematics 714, Springer, Berlin.
64. Jacod, J. and Mémin, J. (1976). Caractéristiques locales et conditions de continuité absolue pour les semi-martingales. *Z. Wahrsch. verw. Gebiete* **35,** 1–37.
65. Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes.* Springer, Berlin. 2nd edition (2003).
66. Jacod, J. and Skorokhod, A.V. (1996). Jumping Markov processes. *Ann. Inst. H. Poincaré Probab. Statist.* **32,** 11–67.
67. Jagers, P. (1975). *Branching Processes with Biological Applications.* Wiley, London.
68. Jagers, P. (1989). General branching processes as random fields. *Stoch. Proc. Appl.* **32,** 183–212.
69. 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, 232–252.
70. Kabanov, Y., Liptser, R.S. and Shiryaev, A.N. (1977). On the question of absolute continuity and singularity of probability measures. *Math. USSR Sb.* **33,** 203–221.
71. Kabanov, Y., Liptser, R.S. and Shiryaev, A.N. (1979,1980). Absolute continuity and singularity of locally absolutely continuous probability distributions I-II. *Math. USSR Sb.* **35,** 631–680, *ibid.* **36,** 31–58.
72. Kalashnikov, V.V. (1994). *Mathematical Methods in Queueing Theory.* Kluwer, Dordrecht.
73. Kallenberg, O. (1983). *Random Measures.* Akademie-Verlag, Berlin and Academic Press, London.

74. Karatzas, I. and Shreve, S.E. (1998). *Methods of Mathematical Finance.* Springer, New York.
75. Karlin, S. and Taylor, H.M. (1966). *A First Course in Stochastic Processes.* Academic Press, New York. 2nd edition (1975).
76. Karlin, S. and Taylor, H.M. (1981). *A Second Course in Stochastic Processes.* Academic Press, New York.
77. El Karoui, N. and Lepeltier, J.P. (1977). Représentations des processus ponctuels multivariés à l'aide d'un processus de Poisson. *Z. Wahrsch. verw. Gebiete* **39,** 111–134.
78. Karr, A. (1991). *Point Processes and their Statistical Inference.* Marcel Dekker, New York.
79. Kelly, F.P. (1979). *Reversibility and Stochastic Networks.* Wiley, Chichester.
80. Kwieciński, A. and Szekli, R. (1991). Compensator conditions for stochastic ordering of point processes. *J. Appl. Probab.* **28,** 751–761.
81. Last, G. and Brandt, A. (1995). *Marked Point Processes on the Real Line. The Dynamic Approach.* Springer, New York.
82. Last, G. (1994). Predictable projections for point process filtrations. *Probab. Theory Relat. Fields,* **99,** 361–388.
83. Lindvall, T. (1988). *Lectures on the Coupling Method.* Wiley, New York.
84. Liptser, R.S. and Shiryaev, A.N. (1978). *Statistics of Random Processes, Vol. I General Theory, Vol. II Applications.* Springer, Berlin. 2nd edition (2001).
85. Liptser, R.S. and Shiryaev, A.N. (1986). *Theory of Martingales.* Kluwer, Dordrecht.
86. MacDonald, I.L. and Zucchini, W. (1997). *Hidden Markov and Other Models for Discrete-valued Time Series.* Chapman & Hall, London.
87. Martinussen, T. and Scheike, T. (1999). A semi-parametric additive regression model for longitudinal data. *Biometrika,* **86**, 691–702.
88. Martinussen, T. and Scheike, T. A nonparametric additive regression model for longitudinal data. *Ann. Statist.* **28**, 1000–1025.
89. Matthes, K., Kerstan, J. and Mecke, J. (1978). *Infinitely Divisible Point Processes.* Wiley, New York.
90. Mecke, J. (1965). Erneuerungseigenschaften bei Punktprozessen. *Math. Nachr.* **30,** 275–294.
91. Melnikov, A.V. (1997). *Financial Markets. Stochastic Analysis and the Pricing of Derivative Securities.* American Mathematical Society, Providence.
92. Merton, R.C. (1992, revised edition). *Continuous–Time Finance.* Blackwell, Cambridge (MA).
93. Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability.* Control and Communication in Engineering, Springer, New York.
94. Meyn, S.P. and Tweedie, R.L. (1993). Stability of Markovian Processes II: Continuous time processes and sampled chains. *Adv. Appl. Probab.* **25,** 487–517.
95. Meyn, S.P. and Tweedie, R.L. (1993). Stability of Markovian Processes III: Foster-Lyapunov criteria for continuous time processes. *Adv. Appl. Probab.* **25,** 518–548.
96. Musiela, M. and Rutkowski, M. (1997). *Martingale Methods in Financial Modelling.* Springer, Berlin.
97. Neveu, J. (1977). Processus ponctuels. *École d'Eté de Probabilités de Saint-Flour VI.* Lecture Notes in Mathematics 598, 249–445, Springer, Berlin.
98. Palmowski, Z. and Rolski, T. (2002). A technique for exponential change of measure for Markov processes. *Bernoulli* **8,** 767-785.
99. Pliska, S.R. (1997). *Introduction to Mathematical Finance. Discrete Time Models.* Blackwell, Malden.
100. Protter, P. (1990). *Stochastic Integration and Differential Equations. A New Approach.* Springer, Berlin. 2nd edition (2004).

101. Reiss, R.-D. (1993). *A Course on Point Processes.* Springer, New York.
102. Resnick, S. (1992). *Adventures in Stochastic Processes.* Birkhäuser, Boston.
103. Robert, P. (2003). *Stochastic Networks and Queues.* Springer, Berlin.
104. Rogers, L.C.G. and Williams, D. (vol. 1 2nd edition 1994, vol. 2 1987). *Diffusions, Markov Processes and Martingales 1-2.* Wiley, Chichester.
105. Rolski, T., Schmidli, H., Schmidt, V. and Teugels, J. (1999). *Stochastic Processes for Insurance and Finance.* Wiley, New York.
106. Scheike, T. (1994). Parametric regression for longitudinal data with counting process measurement times. *Scand. J. Statist.* **21**, 245–263.
107. van Schuppen, J.H. and Wong, E. (1974). Transformations of local martingales under a change of law. *Ann. Probab.* **2,** 879–888.
108. Sato, K-I. (1999). *Lévy Processes and Infintely Divisible Distributions.* Cambridge University Press, Cambridge.
109. Segall, A. and Kailath, T. (1975). Radon-Nikodym derivatives withrespect to measures induced by discontinuous independent increment processes. *Ann. Probab.* **3,** 449–464.
110. Serfozo, R. (1999). *Introduction to Stochastic Networks.* Springer, New York.
111. Sharpe, M. (1988). *General Theory of Markov Processes.* Academic Press, San Diego.
112. Shiryaev, A.N. (1984). *Probability.* Springer, New York. 2nd edition (1996).
113. Shiryaev, A.N. (2000). *Essentials of Stochastic Finance. Facts, Models, Theory.* World Scientific, River Edge.
114. Sigman, K. (1995). *Stationary Marked Point Processes. An Intuitive Approach.* Chapman & Hall, New York.
115. Snyder, D.L. (1975). *Random Point Processes.* Wiley, New York.
116. Stoyan, D. (1983). *Comparison Methods for Queues and Other Stochastic Models.* Wiley, Chichester.
117. Szekli, R. (1995). *Stochastic Ordering and Dependence in Applied Probability.* Lecture Notes in Statistics 97, Springer, New York.
118. Thorisson, H. (2000). *Coupling, Stationarity and Regeneration.* Springer, New York.
119. Walrand, J. (1988). *An Introduction to Queueing Networks.* Prentice–Hall.
120. Whitt, W. (1981). Comparing counting processes and queues. *Adv. Appl. Probab.* **13,** 207–220.
121. Williams, D. (1990). *Probabilities with martingales.* Cambridge University Press, Cambridge.
122. Wobst, R. (1983). *On jump processes with drift.* Dissertationes Math. (Rozprawy Mat.), Polska Akademia Nauk, Inst. Matematyczny, Warszawa.

Notation Index

Index