An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods, Second Edition

D.J. Daley
D. Vere-Jones

Springer

Probability and its Applications

A Series of the Applied Probability Trust

Editors: J. Gani, C.C. Heyde, T.G. Kurtz

Springer
New York
Berlin
Heidelberg
Hong Kong
London
Milan
Paris
Tokyo

D.J. Daley D. Vere-Jones

An Introduction to the Theory of Point Processes

Volume I: Elementary Theory and Methods

Second Edition

Springer

D.J. Daley
Centre for Mathematics and its Applications
Mathematical Sciences Institute
Australian National University
Canberra, ACT 0200, Australia
daryl@maths.anu.edu.au

D. Vere-Jones
School of Mathematical and Computing Sciences
Victoria University of Wellington
Wellington, New Zealand
David.Vere-Jones@mcs.vuw.ac.nz

Library of Congress Cataloging-in-Publication Data
Daley, Daryl J.
An introduction to the theory of point processes / D.J. Daley, D. Vere-Jones.
p. cm.
Includes bibliographical references and index.
Contents: v. 1. Elementary theory and methods
ISBN 0-387-95541-0 (alk. paper)
1. Point processes. I. Vere-Jones, D. (David) II. Title
QA274.42.D35 2002
519.2´3—dc21 2002026666

ISBN 0-387-95541-0 Printed on acid-free paper.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1 SPIN 10885680

Typesetting: Photocomposed pages prepared by the authors using plain TeX files.

www.springer-ny.com

Springer-Verlag New York Berlin Heidelberg
A member of BertelsmannSpringer Science+Business Media GmbH

◆

To Nola,
and in memory of Mary

◆

This page intentionally left blank

Preface to the Second Edition

In preparing this second edition, we have taken the opportunity to reshape the book, partly in response to the further explosion of material on point processes that has occurred in the last decade but partly also in the hope of making some of the material in later chapters of the first edition more accessible to readers primarily interested in models and applications. Topics such as conditional intensities and spatial processes, which appeared relatively advanced and technically difficult at the time of the first edition, have now been so extensively used and developed that they warrant inclusion in the earlier introductory part of the text. Although the original aim of the book—to present an introduction to the theory in as broad a manner as we are able—has remained unchanged, it now seems to us best accomplished in two volumes, the first concentrating on introductory material and models and the second on structure and general theory. The major revisions in this volume, as well as the main new material, are to be found in Chapters 6–8. The rest of the book has been revised to take these changes into account, to correct errors in the first edition, and to bring in a range of new ideas and examples.

Even at the time of the first edition, we were struggling to do justice to the variety of directions, applications and links with other material that the theory of point processes had acquired. The situation now is a great deal more daunting. The mathematical ideas, particularly the links to statistical mechanics and with regard to inference for point processes, have extended considerably. Simulation and related computational methods have developed even more rapidly, transforming the range and nature of the problems under active investigation and development. Applications to spatial point patterns, especially in connection with image analysis but also in many other scientific disciplines, have also exploded, frequently acquiring special language and techniques in the different fields of application. Marked point processes, which were clamouring for greater attention even at the time of the first edition, have acquired a central position in many of these new applications, influencing both the direction of growth and the centre of gravity of the theory.

We are sadly conscious of our inability to do justice to this wealth of new material. Even less than at the time of the first edition can the book claim to provide a comprehensive, up-to-the-minute treatment of the subject. Nor are we able to provide more than a sketch of how the ideas of the subject have evolved. Nevertheless, we hope that the attempt to provide an introduction to the main lines of development, backed by a succinct yet rigorous treatment of the theory, will prove of value to readers in both theoretical and applied fields and a possible starting point for the development of lecture courses on different facets of the subject. As with the first edition, we have endeavoured to make the material as self-contained as possible, with references to background mathematical concepts summarized in the appendices, which appear in this edition at the end of Volume I.

We would like to express our gratitude to the readers who drew our attention to some of the major errors and omissions of the first edition and will be glad to receive similar notice of those that remain or have been newly introduced. Space precludes our listing these many helpers, but we would like to acknowledge our indebtedness to Rick Schoenberg, Robin Milne, Volker Schmidt, Günter Last, Peter Glynn, Olav Kallenberg, Martin Kalinke, Jim Pitman, Tim Brown and Steve Evans for particular comments and careful reading of the original or revised texts (or both). Finally, it is a pleasure to thank John Kimmel of Springer-Verlag for his patience and encouragement, and especially Eileen Dallwitz for undertaking the painful task of rekeying the text of the first edition.

The support of our two universities has been as unflagging for this endeavour as for the first edition; we would add thanks to host institutions of visits to the Technical University of Munich (supported by a Humboldt Foundation Award), University College London (supported by a grant from the Engineering and Physical Sciences Research Council) and the Institute of Mathematics and its Applications at the University of Minnesota.

Daryl Daley
Canberra, Australia

David Vere-Jones
Wellington, New Zealand

Preface to the First Edition

This book has developed over many years—too many, as our colleagues and families would doubtless aver. It was conceived as a sequel to the review paper that we wrote for the Point Process Conference organized by Peter Lewis in 1971. Since that time the subject has kept running away from us faster than we could organize our attempts to set it down on paper. The last two decades have seen the rise and rapid development of martingale methods, the surge of interest in stochastic geometry following Rollo Davidson's work, and the forging of close links between point processes and equilibrium problems in statistical mechanics.

Our intention at the beginning was to write a text that would provide a survey of point process *theory* accessible to beginning graduate students and workers in applied fields. With this in mind we adopted a partly historical approach, starting with an informal introduction followed by a more detailed discussion of the most familiar and important examples, and then moving gradually into topics of increased abstraction and generality. This is still the basic pattern of the book. Chapters 1–4 provide historical background and treat fundamental special cases (Poisson processes, stationary processes on the line, and renewal processes). Chapter 5, on finite point processes, has a bridging character, while Chapters 6–14 develop aspects of the general theory.

The main difficulty we had with this approach was to decide when and how far to introduce the abstract concepts of functional analysis. With some regret, we finally decided that it was idle to pretend that a general treatment of point processes could be developed without this background, mainly because the problems of existence and convergence lead inexorably to the theory of measures on metric spaces. This being so, one might as well take advantage of the metric space framework from the outset and let the point process itself be defined on a space of this character: at least this obviates the tedium of having continually to specify the dimensions of the Euclidean space, while in the context of completely separable metric spaces—and this is the greatest

generality we contemplate—intuitive spatial notions still provide a reasonable guide to basic properties. For these reasons the general results from Chapter 6 onward are couched in the language of this setting, although the examples continue to be drawn mainly from the one- or two-dimensional Euclidean spaces $\mathbb{R}^1$ and $\mathbb{R}^2$. Two appendices collect together the main results we need from measure theory and the theory of measures on metric spaces. We hope that their inclusion will help to make the book more readily usable by applied workers who wish to understand the main ideas of the general theory without themselves becoming experts in these fields. Chapter 13, on the martingale approach, is a special case. Here the context is again the real line, but we added a third appendix that attempts to summarize the main ideas needed from martingale theory and the general theory of processes. Such special treatment seems to us warranted by the exceptional importance of these ideas in handling the problems of inference for point processes.

In style, our guiding star has been the texts of Feller, however many light-years we may be from achieving that goal. In particular, we have tried to follow his format of motivating and illustrating the general theory with a range of examples, sometimes didactical in character, but more often taken from real applications of importance. In this sense we have tried to strike a mean between the rigorous, abstract treatments of texts such as those by Matthes, Kerstan and Mecke (1974/1978/1982) and Kallenberg (1975, 1983), and practically motivated but informal treatments such as Cox and Lewis (1966) and Cox and Isham (1980).

Numbering Conventions. Each chapter is divided into sections, with consecutive labelling within each of equations, statements (encompassing Definitions, Conditions, Lemmas, Propositions, Theorems), examples, and the exercises collected at the end of each section. Thus, in Section 1.2, (1.2.3) is the third equation, **Statement 1.2.III** is the third statement, EXAMPLE 1.2(c) is the third example, and Exercise 1.2.3 is the third exercise. The exercises are varied in both content and intention and form a significant part of the text. Usually, they indicate extensions or applications (or both) of the theory and examples developed in the main text, elaborated by hints or references intended to help the reader seeking to make use of them. The symbol □ denotes the end of a proof. Instead of a name index, the listed references carry page number(s) where they are cited. A general outline of the notation used has been included before the main text.

It remains to acknowledge our indebtedness to many persons and institutions. Any reader familiar with the development of point process theory over the last two decades will have no difficulty in appreciating our dependence on the fundamental monographs already noted by Matthes, Kerstan and Mecke in its three editions (our use of the abbreviation MKM for the 1978 English edition is as much a mark of respect as convenience) and Kallenberg in its two editions. We have been very conscious of their generous interest in our efforts from the outset and are grateful to Olav Kallenberg in particular for saving us from some major blunders. A number of other colleagues, notably

David Brillinger, David Cox, Klaus Krickeberg, Robin Milne, Dietrich Stoyan, Mark Westcott, and Deng Yonglu, have also provided valuable comments and advice for which we are very grateful. Our two universities have responded generously with seemingly unending streams of requests to visit one another at various stages during more intensive periods of writing the manuscript. We also note visits to the University of California at Berkeley, to the Center for Stochastic Processes at the University of North Carolina at Chapel Hill, and to Zhongshan University at Guangzhou. For secretarial assistance we wish to thank particularly Beryl Cranston, Sue Watson, June Wilson, Ann Milligan, and Shelley Carlyle for their excellent and painstaking typing of difficult manuscript.

Finally, we must acknowledge the long-enduring support of our families, and especially our wives, throughout: they are not alone in welcoming the speed and efficiency of Springer-Verlag in completing this project.

Daryl Daley
Canberra, Australia

David Vere-Jones
Wellington, New Zealand

This page intentionally left blank

Contents

Chapter Titles for Volume II

This page intentionally left blank

Principal Notation

Very little of the general notation used in Appendices 1–3 is given below. Also, notation that is largely confined to one or two sections of the same chapter is mostly excluded, so that neither all the symbols used nor all the uses of the symbols shown are given. The repeated use of some symbols occurs as a result of point process theory embracing a variety of topics from the theory of stochastic processes. Where they are given, page numbers indicate the first or significant use of the notation. Generally, the particular interpretation of symbols with more than one use is clear from the context.

Throughout the lists below, N denotes a point process and ξ denotes a random measure.

Spaces

$\mathbb{C}$	complex numbers	
$\mathbb{R}^d$	d-dimensional Euclidean space	
$\mathbb{R} = \mathbb{R}^1$	real line	
$\mathbb{R}_+$	nonnegative numbers	
$\mathbb{S}$	circle group and its representation as $(0, 2\pi]$	
$\mathbb{U}_{2\alpha}^d$	d-dimensional cube of side length 2α and vertices $(\pm\alpha, \ldots, \pm\alpha)$	
$\mathbb{Z}$, $\mathbb{Z}_+$	integers of $\mathbb{R}$, $\mathbb{R}_+$	
$\mathcal{X}$	state space of N or ξ; often $\mathcal{X} = \mathbb{R}^d$; always $\mathcal{X}$ is c.s.m.s. (complete separable metric space)	
Ω	space of probability elements ω	
$\emptyset$, $\emptyset(\cdot)$	null set, null measure	
$\mathcal{E}$	measurable sets in probability space	
$(\Omega, \mathcal{E}, \mathcal{P})$	basic probability space on which N and ξ are defined	158
$\mathcal{X}^{(n)}$	n-fold product space $\mathcal{X} \times \cdots \times \mathcal{X}$	123
$\mathcal{X}^{\cup}$	$= \mathcal{X}^{(0)} \cup \mathcal{X}^{(1)} \cup \cdots$	129

$\mathcal{B}(\mathcal{X})$	Borel σ-field generated by open spheres of c.s.m.s. $\mathcal{X}$	34
$\mathcal{B}_{\mathcal{X}}$	$= \mathcal{B}(\mathcal{X})$, $\mathcal{B} = \mathcal{B}_{\mathbb{R}} = \mathcal{B}(\mathbb{R})$	34, 374
$\mathcal{B}_{\mathcal{X}}^{(n)} = \mathcal{B}(\mathcal{X}^{(n)})$	product σ-field on product space $\mathcal{X}^{(n)}$	129
$\mathrm{BM}(\mathcal{X})$	measurable functions of bounded support	161
$\mathrm{BM}_+(\mathcal{X})$	measurable nonnegative functions of bounded support	161
$\mathcal{K}$	mark space for marked point process (MPP)	194
$\mathcal{M}_{\mathcal{X}}$ ($\mathcal{N}_{\mathcal{X}}$)	totally finite (counting) measures on c.s.m.s. $\mathcal{X}$	158, 398
$\mathcal{M}_{\mathcal{X}}^{\#}$	boundedly finite measures on c.s.m.s. $\mathcal{X}$	158, 398
$\mathcal{N}_{\mathcal{X}}^{\#}$	boundedly finite counting measures on c.s.m.s. $\mathcal{X}$	131
$\mathcal{P}^+$	p.p.d. (positive positive-definite) measures	359
$\mathcal{S}$	infinitely differentiable functions of rapid decay	357
$\mathcal{U}$	complex-valued Borel measurable functions on $\mathcal{X}$ of modulus ≤ 1	144
$\mathcal{U} \otimes \mathcal{V}$	product topology on product space $\mathcal{X} \times \mathcal{Y}$ of topological spaces $(\mathcal{X},\mathcal{U})$, $(\mathcal{Y},\mathcal{V})$	378
$\mathcal{V} = \mathcal{V}(\mathcal{X})$	$[0,1]$-valued measurable functions $h(x)$ with $1 - h(x)$ of bounded support in $\mathcal{X}$	149, 152

General

Unless otherwise specified, $A \in \mathcal{B}_{\mathcal{X}}$, k and $n \in \mathbb{Z}_+$, t and $x \in \mathbb{R}$, $h \in \mathcal{V}(\mathcal{X})$, and $z \in \mathbb{C}$.

$\tilde{}$	$\tilde{\nu}, \widetilde{F}$ = Fourier–Stieltjes transforms of measure ν or d.f. F	411–412
	$\tilde{\phi}$ = Fourier transform of Lebesgue integrable function ϕ for counting measures	357
$\breve{}$	reduced (ordinary or factorial) (moment or cumulant) measure	160
$\#$	extension of concept from totally finite to boundedly finite measure space	158
$\|\mu\|$	variation norm of measure μ	374
a.e. μ, μ-a.e.	almost everywhere with respect to measure μ	376
a.s., $\mathcal{P}$-a.s.	almost sure, $\mathcal{P}$-almost surely	376
$A^{(n)}$	n-fold product set $A \times \cdots \times A$	130
$\mathcal{A}$	family of sets generating $\mathcal{B}$; semiring of bounded Borel sets generating $\mathcal{B}_{\mathcal{X}}$	31, 368
B_u (T_u)	backward (forward) recurrence time at u	58, 76
c_k, $c_{[k]}$	kth cumulant, kth factorial cumulant, of distribution $\{p_n\}$	116
$c(x) = c(y, y+x)$	covariance density of stationary mean square continuous process on $\mathbb{R}^d$	160, 358

$C_{[k]}(\cdot),\ c_{[k]}(\cdot)$	factorial cumulant measure and density	147
$\breve{C}_2(\cdot), \breve{c}(\cdot)$	reduced covariance measure of stationary N or ξ	292
$\breve{c}(\cdot)$	reduced covariance density of stationary N or ξ	160, 292
$\delta(\cdot)$	Dirac delta function	
$\delta_x(A)$	Dirac measure, $= \int_A \delta(u-x)\,\mathrm{d}u = I_A(x)$	382
$\Delta F(x) = F(x) - F(x-)$	jump at x in right-continuous function F	107
$\mathrm{e}_\lambda(x) = (\frac{1}{2}\lambda)^d \exp\big(-\lambda\sum_{i=1}^d \lvert x_i\rvert\big)$	two-sided exponential density in $\mathbb{R}^d$	359
F	renewal process lifetime d.f.	67
F^{n*}	n-fold convolution power of measure or d.f. F	55
$F(\cdot\,;\cdot)$	finite-dimensional (fidi) distribution	158–161
$\boldsymbol{\mathcal{F}}$	history	236, 240
$\Phi(\cdot)$	characteristic functional	15
$G[h]$	probability generating functional (p.g.fl.) of N,	15, 144
$G[h \mid x]$	member of measurable family of p.g.fl.s	166
$G_c[\cdot],\ G_m[\cdot \mid x]$	p.g.fl.s of cluster centre and cluster member processes N_c and $N_m(\cdot \mid x)$	178
$G,\ G_I$	expected information gain (per interval) of stationary N on $\mathbb{R}$	280, 285
$\Gamma(\cdot),\ \gamma(\cdot)$	Bartlett spectrum, its density when it exists	304
$H(\mathcal{P};\mu)$	generalized entropy	277, 283
$\boldsymbol{\mathcal{H}},\ \boldsymbol{\mathcal{H}}^*$	internal history of ξ on $\mathbb{R}_+$, $\mathbb{R}$	236
$I_A(x) = \delta_x(A)$	indicator function of element x in set A	
$I_n(x)$	modified Bessel function of order n	72
$J_n(A_1 \times \cdots \times A_n)$	Janossy measure	124
$j_n(x_1, \ldots, x_n)$	Janossy density	125
$J_n(\cdot \mid A)$	local Janossy measure	137
K	compact set	371
$K_n(\cdot),\ k_n(\cdot)$	Khinchin measure and density	146
$\ell(\cdot)$	Lebesgue measure in $\mathcal{B}(\mathbb{R}^d)$, Haar measure on σ-group	31, 408–409
$L_u = B_u + T_u$	current lifetime of point process on $\mathbb{R}$	58, 76
$L[f]$ $(f \in BM_+(\mathcal{X}))$	Laplace functional of ξ	161
$L_\xi[1-h]$	p.g.fl. of Cox process directed by ξ	170
$L_2(\xi^0),\ L_2(\Gamma)$	Hilbert spaces of square integrable r.v.s ξ^0, and of functions square integrable w.r.t. measure Γ	332
$L_A(x_1, \ldots, x_n), = j_N(x_1, \ldots, x_N \mid A)$	likelihood, local Janossy density, $N \equiv N(A)$	22, 212
λ	rate of N, especially intensity of stationary N	46
$\lambda^*(t)$	conditional intensity function	231
$m_k\ (m_{[k]})$	kth (factorial) moment of distribution $\{p_n\}$	115

$\breve{m}_2$, $\breve{M}_2$	reduced second-order moment density, measure, of stationary N	289
m_{g}	mean density of ground process N_{g} of MPP N	198, 323
$N(A)$	number of points in A	42
$N(a,b]$	number of points in half-open interval $(a,b]$,	19
	$= N((a,b])$	42
$N(t)$	$= N(0,t] = N((0,t])$	42
N_c	cluster centre process	176
$N(\cdot \mid x)$	cluster member or component process	176
$\{(p_n, \Pi_n)\}$	elements of probability measure for finite point process	123
$P(z)$	probability generating function (p.g.f.) of distribution $\{p_n\}$	10, 115
$P(x,A)$	Markov transition kernel	92
$P_0(A)$	avoidance function	31, 135
$\mathcal{P}_{jk}$	set of j-partitions of $\{1,\ldots,k\}$	121
$\mathcal{P}$	probability measure of stationary N on $\mathbb{R}$,	53
	probability measure of N or ξ on c.s.m.s. $\mathcal{X}$	158
$\{\pi_k\}$	batch-size distribution	28, 51
$q(x) = f(x)/[1-F(x)]$	hazard function for lifetime d.f. F	2, 106
$Q(z)$	$= -\log P(z)$	27
$Q(\cdot)$, $Q(t)$	hazard measure, integrated hazard function (IHF)	109
$\rho(x,y)$	metric for x, y in metric space	370
$\{S_n\}$	random walk, sequence of partial sums	66
$S(x) = 1 - F(x)$	survivor function of d.f. F	2, 109
$S_r(x)$	sphere of radius r, centre x, in metric space $\mathcal{X}$	35, 371
$t(x) = \prod_{i=1}^{d}(1-\lvert x_i\rvert)_+$	triangular density in $\mathbb{R}^d$	359
T_u	forward recurrence time at u	58, 75
$\mathcal{T} = \{S_1(\mathcal{T}), \ldots, S_j(\mathcal{T})\}$	a j-partition of k	121
$\mathcal{T} = \{\mathcal{T}_n\} = \{\{A_{ni}\}\}$	dissecting system of nested partitions	382
$U(A) = \mathrm{E}[N(A)]$	renewal measure	67
$U(x)$	$= U([0,x])$, expectation function,	61
	renewal function ($U(x) = 1 + U_0(x)$)	67
$V(A)$	$= \operatorname{var} N(A)$, variance function	295
$V(x) = V((0,x])$	variance function for stationary N or ξ on $\mathbb{R}$	80, 301
$\{X_n\}$	components of random walk $\{S_n\}$,	66
	intervals of Wold process	92

Concordance of Statements from the First Edition

The table below lists the identifying number of formal statements of the first edition (1988) of this book and their identification in this volume.

1988 edition	this volume
2.2.I–III	2.2.I–III
2.3.III	2.3.I
2.4.I–II	2.4.I–II
2.4.V–VIII	2.4.III–VI
3.2.I–II	3.2.I–II
3.3.I–IX	3.3.I–IX
3.4.I–II	3.4.I–II
3.5.I–III	3.5.I–III
3.6.I–V	3.6.I–V
4.2.I–II	4.2.I–II
4.3.I–III	4.3.I–III
4.4.I–VI	4.4.I–VI
4.5.I–VI	4.5.I–VI
4.6.I–V	4.6.I–V
5.2.I–VII	5.2.I–VII
5.3.I–III	5.3.I–III
5.4.I–III	5.4.I–III
5.4.IV–VI	5.4.V–VII
5.5.I	5.5.I
7.1.XII–XIII	6.4.I(a)–(b)
8.1.II	6.1.II, IV
8.2.I	6.3.I
8.2.II	6.3.II, (6.3.6)
8.3.I–III	6.3.III–V
8.5.I–III	6.2.II
11.1.I–V	8.6.I–V
11.2.I–II	8.2.I–II
11.3.I–VIII	8.4.I–VIII
11.4.I–IV	8.5.I–IV
11.4.V–VI	8.5.VI–VII
13.1.I–III	7.1.I–III
13.1.IV–VI	7.2.I–III
13.1.VII	7.1.IV
13.4.III	7.6.I
A1.1.I–5.IV	A1.1.I–5.IV
A2.1.I–III	A2.1.I–III
A2.1.IV	A1.6.I
A2.1.V–VI	A2.1.IV–V
A2.2.I–7.III	A2.2.I–7.III
A3.1.I–4.IX	A3.1.I–4.IX

This page intentionally left blank

CHAPTER 1

Early History

The ancient origins of the modern theory of point processes are not easy to trace, nor is it our aim to give here an account with claims to being definitive. But any retrospective survey of a subject must inevitably give some focus on those past activities that can be seen to embody concepts in common with the modern theory. Accordingly, this first chapter is a historical indulgence but with the added benefit of describing certain fundamental concepts informally and in a heuristic fashion prior to possibly obscuring them with a plethora of mathematical jargon and techniques. These essentially simple ideas appear to have emerged from four distinguishable strands of enquiry—although our division of material may sometimes be a little arbitrary. These are

(i) life tables and the theory of self-renewing aggregates;
(ii) counting problems;
(iii) particle physics and population processes; and
(iv) communication engineering.

The first two of these strands could have been discerned in centuries past and are discussed in the first two sections. The remaining two essentially belong to the twentieth century, and our comments are briefer in the remaining section.

1.1. Life Tables and Renewal Theory

Of all the threads that are woven into the modern theory of point processes, the one with the longest history is that associated with intervals between events. This includes, in particular, renewal theory, which could be defined in a narrow sense as the study of the sequence of intervals between successive replacements of a component that is liable to failure and is replaced by a new

component every time a failure occurs. As such, it is a subject that developed during the 1930s and reached a definitive stage with the work of Feller, Smith, and others in the period following World War II. But its roots extend back much further than this, through the study of 'self-renewing aggregates' to problems of statistical demography, insurance, and mortality tables—in short, to one of the founding impulses of probability theory itself. It is not easy to point with confidence to any intermediate stage in this chronicle that recommends itself as the natural starting point either of renewal theory or of point process theory more generally. Accordingly, we start from the beginning, with a brief discussion of life tables themselves. The connection with point processes may seem distant at first sight, but in fact the theory of life tables provides not only the source of much current terminology but also the setting for a range of problems concerning the evolution of populations in time and space, which, in their full complexity, are only now coming within the scope of current mathematical techniques.

In its basic form, a life table consists of a list of the number of individuals, usually from an initial group of 1000 individuals so that the numbers are effectively proportions, who survive to a given age in a given population. The most important parameters are the number ℓ_x surviving to age x, the number d_x dying between the ages x and $x+1$ ($d_x = \ell_x - \ell_{x+1}$), and the number q_x of those surviving to age x who die before reaching age $x+1$ ($q_x = d_x/\ell_x$). In practice, the tables are given for discrete ages, with the unit of time usually taken as 1 year. For our purposes, it is more appropriate to replace the discrete time parameter by a continuous one and to replace numbers by probabilities for a single individual. Corresponding to ℓ_x we have then the survivor function

$$S(x) = \Pr\{\text{lifetime} > x\}.$$

To d_x corresponds $f(x)$, the density of the lifetime distribution function, where

$$f(x)\,\mathrm{d}x = \Pr\{\text{lifetime terminates between } x \text{ and } x + \mathrm{d}x\},$$

while to q_x corresponds $q(x)$, the *hazard function*, where

$$q(x)\,\mathrm{d}x = \Pr\{\text{lifetime terminates between } x \text{ and } x + \mathrm{d}x \mid \text{it does not terminate before } x.\}$$

Denoting the lifetime distribution function itself by $F(x)$, we have the following important relations between the functions above:

$$S(x) = 1 - F(x) = \int_x^\infty f(y)\,\mathrm{d}y = \exp\Big(-\int_0^x q(y)\,\mathrm{d}y\Big), \tag{1.1.1}$$

$$f(x) = \frac{\mathrm{d}F}{\mathrm{d}x} = \frac{\mathrm{d}S}{\mathrm{d}x}\,, \tag{1.1.2}$$

$$q(x) = \frac{f(x)}{S(x)} = \frac{\mathrm{d}}{\mathrm{d}x}[\log S(x)] = -\frac{\mathrm{d}}{\mathrm{d}x}\{\log[1 - F(x)]\}. \tag{1.1.3}$$

The first life table appeared, in a rather crude form, in John Graunt's (1662) *Observations on the London Bills of Mortality*. This work is a landmark in the early history of statistics, much as the famous correspondence between Pascal and Fermat, which took place in 1654 but was not published until 1679, is a landmark in the early history of formal probability. The coincidence in dates lends weight to the thesis (see e.g. Maistrov, 1967) that mathematical scholars studied games of chance not only for their own interest but for the opportunity they gave for clarifying the basic notions of chance, frequency, and expectation, already actively in use in mortality, insurance, and population movement contexts.

An improved life table was constructed in 1693 by the astronomer Halley, using data from the smaller city of Breslau, which was not subject to the same problems of disease, immigration, and incomplete records with which Graunt struggled in the London data. Graunt's table was also discussed by Huyghens (1629–1695), to whom the notion of expected length of life is due. A. de Moivre (1667–1754) suggested that for human populations the function $S(x)$ could be taken to decrease with equal yearly decrements between the ages 22 and 86. This corresponds to a uniform density over this period and a hazard function that increases to infinity as x approaches 86. The analysis leading to (1.1.1) and (1.1.2), with further elaborations to take into account different sources of mortality, would appear to be due to Laplace (1747–1829). It is interesting that in *A Philosophical Essay on Probabilities* (1814), where the classical definition of probability based on equiprobable events is laid down, Laplace gave a discussion of mortality tables in terms of probabilities of a totally different kind. Euler (1707–1783) also studied a variety of problems of statistical demography.

From the mathematical point of view, the paradigm distribution function for lifetimes is the exponential function, which has a constant hazard independent of age: for $x > 0$, we have

$$f(x) = \lambda \mathrm{e}^{-\lambda x}, \quad q(x) = \lambda, \quad S(x) = \mathrm{e}^{-\lambda x}, \quad F(x) = 1 - \mathrm{e}^{-\lambda x}. \tag{1.1.4}$$

The usefulness of this distribution, particularly as an approximation for purposes of interpolation, was stressed by Gompertz (1779–1865), who also suggested, as a closer approximation, the distribution function corresponding to a power-law hazard of the form

$$q(x) = A\mathrm{e}^{\alpha x} \qquad (A > 0,\ \alpha > 0,\ x > 0). \tag{1.1.5}$$

With the addition of a further constant [i.e. $q(x) = B + A\mathrm{e}^{\alpha x}$], this is known in demography as the *Gompertz–Makeham* law and is possibly still the most widely used function for interpolating or graduating a life table.

Other forms commonly used for modelling the lifetime distribution in different contexts are the *Weibull*, *gamma*, and *log normal* distributions, corresponding, respectively, to the formulae

$$q(x) = \beta\lambda x^{\beta-1} \qquad \text{with} \quad S(x) = \exp(-\lambda x^{\beta}) \qquad (\lambda > 0,\ \beta > 0), \tag{1.1.6}$$

$$f(x) = \lambda\alpha x^{\alpha-1}\mathrm{e}^{-\lambda x}/\Gamma(\alpha), \tag{1.1.7}$$

$$f(x) = (\sigma x\sqrt{2\pi}\,)^{-1}\mathrm{e}^{-[(\log x-\mu)/\sigma]^2/2}. \tag{1.1.8}$$

The Weibull distribution was introduced by Weibull (1939a, b) as a model for brittle fracture. Both this and the preceding distribution have an interpretation in terms of extreme value theory (see e.g. Exercise 1.1.2), but it should be emphasized that as a general rule the same distribution may arise from several models (see Exercise 1.1.3).

The gamma distribution has a long history and arises in many different contexts. When $\alpha = \frac{1}{2}k$ and $\lambda = \frac{1}{2}$, it is nothing other than the chi-squared distribution with k degrees of freedom, with well-known applications in mathematical statistics. When $\alpha = 1$, it reduces to the exponential distribution, and when $\alpha = \frac{3}{2}$, it reduces to the Maxwell distribution for the distribution of energies of molecules in a perfect gas. The most important special cases in the context of life tables arise when α is a positive integer, say $\alpha = k$. It then has an interpretation as the sum of k independent random variables, each having an exponential distribution. Although commonly known as the Erlang distribution, after the Danish engineer and mathematician who introduced it as a model for telephone service and intercall distributions in the 1920s, this special form and its derivation were known much earlier. One of the earliest derivations, if not the first, is due to the English mathematician R.C. Ellis (1817–1859) in a remarkable paper in 1844 that could well be hailed as one of the early landmarks in stochastic process theory, although in fact it is rarely quoted. In addition to establishing the above-mentioned result as a special case, Ellis studied a general renewal process and in that context established the asymptotic normality of the sum of a number of independent nonnegative random variables. It is particularly remarkable in that he used Fourier methods; in other words, essentially the modern characteristic function proof (with a few lacunae from a modern standpoint) of the central limit theorem.

An equally interesting aspect of Ellis' paper is the problem that inspired the study. This takes us back a century and a half to an even less familiar statistician in the guise of Sir Isaac Newton (1642–1728). For much of his later life, Newton's spare time was devoted to theological problems, one of which was to reconcile the ancient Greek and Hebrew chronologies. In both chronologies, periods of unknown length are spanned by a list of successive rulers. Newton proposed to estimate such periods, and hence to relate the two chronologies, by supposing each ruler to reign for a standard period of 22 years. This figure was obtained by a judicious comparison of averages from a miscellany of historical data for which more or less reliable lengths of reigns were known. It is a statistical inference in the same sense as many of Graunt's inferences from the London Bills of Mortality: a plausible value based on the best or only evidence available and supported by as many cross-checks as can be devised. How far it was explicitly present in Newton's mind that he was dealing with a statistical problem and whether he made any attempts

to assess the likely errors of his results himself are questions we have not been able to answer with any certainty. In an informal summary of his work, Newton (1728) wrote: "I do not pretend to be exact to a year: there may be errors of five or ten years, and sometimes twenty, and not much above." However, it appears unlikely that these figures were obtained by any theory of compounding of errors. It is tempting to conjecture that he may have discussed the problems with such friends and Fellows of the Royal Society as Halley, whose paper to the Royal Society would have been presented while Newton was president, and de Moivre, who dedicated the first edition of *The Doctrine of Chances* to Newton, but if records of such discussions exist, we have not found them.

Up until the middle of the nineteenth century, as will be clear even from the brief review presented above, mathematical problems deriving from life tables not only occupied a major place in the subject matter of probability and statistics but also attracted the attention of many leading mathematicians of the time. From the middle of the nineteenth century onward, however, actuarial mathematics (together, it may be added, with many other probabilistic notions), while important in providing employment for mathematicians, became somewhat disreputable mathematically, a situation from which it has not fully recovered. (How many elementary textbooks in statistics, for example, even mention life tables, let alone such useful descriptive tools as the hazard function?) The result was that when, as was inevitably the case, new applications arose that made use of the same basic concepts, the links with earlier work were lost or only partially recognized. Moreover, the new developments themselves often took place independently or with only a partial realization of the extent of common material.

In the twentieth century, at least three such areas of application may be distinguished. The first, historically, was queueing theory, more specifically the theory of telephone trunking problems. Erlang's (1909) first paper on this subject contains a derivation of the Poisson distribution for the number of calls in a fixed time interval. It is evident from his comments that even before that time the possibility of using probabilistic methods in that context was being considered by engineers in several countries. The work here appears to be quite independent of earlier contributions. In later work, the analysis was extended to cover queueing systems with more general input and service distributions.

Mathematical interest in actuarial problems as such re-emerged in the 1910s and 1920s in connection with the differential and integral equations of population growth. Here at least there is a bridge between the classical theory of life tables on the one hand and the modern treatments of renewal processes on the other. It is provided by the theory of 'self-renewing aggregates' [to borrow a phrase from the review by Lotka (1939), which provides a useful survey of early work in this field], a term that refers to a population (portfolio in the insurance context) of individuals subject to death but also able to regenerate themselves so that a stable population can be achieved.

As a typical illustration, consider the evolution of a human population for which it is assumed that each female of age x has a probability $\phi(x)\,\mathrm{d}t$ of giving birth to a daughter in a time interval of length $\mathrm{d}t$, independently of the behaviour of other females in the population and also of any previous children she may have had. Let $S(x)$ denote the survivor function for the (female) life distribution and $n(t)$ the expected female birth rate at time t. Then $n(t)$ satisfies the integral equation

$$n(t) = \int_0^t n(t-x)S(x)\phi(x)\,\mathrm{d}x,$$

which represents a breakdown of the total female birth rate by age of parent. If the population is started at time zero with an initial age distribution having density $r(x)$, the equation can be rewritten in the form

$$n(t) = n_0(t) + \int_0^t n(t-x)S(x)\phi(x)\,\mathrm{d}x,$$

where

$$n_0(t) = \int_0^\infty r(x)\,\frac{S(t+x)}{S(x)}\,\phi(t+x)\,\mathrm{d}x$$

is the contribution to the birth rate at time t from the initial population. In this form, the analogy with the integral equation of renewal theory is clear. Indeed, the latter equation corresponds to the special case where at death each individual is replaced by another of age zero and no other 'births' are possible. The population size then remains constant, and it is enough to consider a population with just one member. In place of $n(t)$, we then have the *renewal density* $m(t)$, with $m(t)\,\mathrm{d}t$ representing the probability that a replacement will be required in the small time interval $(t, t+\mathrm{d}t)$; also, $\phi(x)$ becomes the hazard function $h(x)$ for the life distribution, and the combination $S(x)h(x)$ can be replaced by the probability density function $f(x)$ as in (1.1.3). Thus, we obtain the renewal equation in the form

$$m(t) = n_0(t) + \int_0^t m(t-u)f(u)\,\mathrm{d}u.$$

If, finally, the process is started with a new component in place at time 0, then $n_0(t) = f(t)$ and we have the standard form

$$m_s(t) = f(t) + \int_0^t m_s(t-u)f(u)\,\mathrm{d}u.$$

The third field to mention is reliability theory. A few problems in this field, including Weibull's discussion of brittle fracture, appeared before World War II, but its systematic development relates to the post-war period and the rapid growth of the electronics industry. Typical problems are the calculation

of lifetime distributions of systems of elements connected in series ('weakest link' model) or in parallel. Weibull's analysis is an example of the first type of model, which typically leads to an extreme-value distribution with a long right tail. An early example of a parallel model is Daniels' (1945) treatment of the failure of fibre bundles; the distributions in this case have an asymptotically normal character. In between and extending these two primary cases lie an infinite variety of further failure models, in all of which the concepts and terminology invented to cover the life table problem play a central role.

In retrospect, it is easy to see that the three fields referred to are closely interconnected. Together, they provide one of the main areas of application and development of point process theory. Of course, they do not represent the only fields where life table methods have been applied with success. An early paper by Watanabe (1933) gives a life table analysis of the times between major earthquake disasters, a technique that has been resurrected by several more recent writers under the name of *theory of durability*. An important recent field of application has been the study of trains of nerve impulses in neurophysiology. In fact, the tools are available and relevant for any phenomenon in which the events occur along a time axis and the intervals between the time points are important and meaningful quantities.

Exercises and Complements to Section 1.1

1.1.1 A nonnegative random variable (r.v.) X with distribution function (d.f.) F has an *increasing failure rate* (abbreviated to IFR) if the conditional d.f.s

$$F_x(u) = \Pr\{X \le x + u \mid X > x\} = \frac{F(x+u) - F(x)}{1 - F(x)} \qquad (u, x \ge 0)$$

are increasing functions of x for every fixed u in $0 < u < \infty$. It has a *decreasing mean residual life* (DMRL) if $\mathrm{E}(X - x \mid X > x)$ decreases with increasing x, and it is *new better than used in expectation* (NBUE) if $\mathrm{E}(X - x \mid X > x) \le \mathrm{E}X$ (all $x > 0$). Show that IFR implies DMRL, DMRL implies NBUE, and NBUE implies that $\operatorname{var} X \le (\mathrm{E}X)^2$ [see Stoyan (1983, Section 1.6)].

1.1.2 Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed r.v.s with d.f. $F(\cdot)$. Then, for any fixed nonnegative integer n,

$$\Pr\Big\{ \max_{1 \le j \le n} X_j \le u \Big\} = \big(F(u)\big)^n.$$

Replacing n by a Poisson-distributed r.v. N with mean μ yields

$$G(u) \equiv \Pr\Big\{ \max_{1 \le j \le N} X_j \le u \Big\} \equiv \mathrm{e}^{-\mu} \sum_{k=0}^{\infty} \mu^k (k!)^{-1} \big(F(u)\big)^k = \mathrm{e}^{-\mu(1-F(u))}.$$

When $F(u) = 1 - \mathrm{e}^{-\lambda u}$, G is the Gumbel d.f., while when $F(u) = 1 - \lambda u^{-\alpha}$, G is the Weibull d.f. [In the forms indicated, these extreme-value distributions include location and/or scale parameters; see e.g. Johnson and Kotz (1970, p. 272).]

1.1.3 Let $X_1, X_2, \ldots$ be as in the previous exercise with $F(u) = 1 - \mathrm{e}^{-\lambda u}$. Show that $Y \equiv \max(X_1, \ldots, X_n)$ has the same distribution as $\sum_{j=1}^{n} X_j/j$.
[*Hint*: Regard $X_1, \ldots, X_n$ as lifetimes in a linear death process with death rate λ, so that y is the time to extinction of the process. Exercise 2.1.2 gives more general properties.]

1.1.4 Suppose that the lifetimes of rulers are independent r.v.s with common d.f. F and that conditional on reaching age 21 years, a ruler has a son (with lifetime d.f.s F) every 2 years for up to six sons, with the eldest surviving son succeeding him. Conditional on there being a succession, what is the d.f. of the age at succession and the expected time that successor reigns (assuming a reign terminated by death from natural causes)?

What types of error would be involved in matching chronologies from a knowledge of the orders of two sets of rulers (see the reference to Newton's work in the text)? How would such chronologies be matched in the light of developments in statistical techniques subsequent to Newton?

1.1.5 Investigate the integral equation for the stationary age distribution in a supercritical age-dependent branching process. Using a suitable metric, evaluate the difference between this stationary age distribution and the backward recurrence time distribution of a stationary renewal process with the same lifetime distribution as a function of the mean of the offspring distribution. Note that Euler worked on the age distribution in exponentially growing populations.

1.2. Counting Problems

The other basic approach to point process phenomena, and the only systematic approach yet available in spaces of higher dimension, is to count the numbers of events in intervals or regions of various types. In this approach, the machinery of discrete distributions plays a central role. Since in probability theory discrete problems are usually easier to handle than continuous problems, it might be thought that the development of general models for a discrete distribution would precede those for a continuous distribution, but in fact the reverse seems to be the case. Although particular examples, such as the Bernoulli distribution and the negative binomial distribution, occurred at a very early stage in the discussion of games of chance, there seems to be no discussion of discrete distributions as such until well into the nineteenth century.

We may take as a starting point Poisson's (1837) text, which included a derivation of the Poisson distribution by passage to the limit from the binomial (the claim that he was anticipated in this by de Moivre is a little exaggerated in our view: it is true that de Moivre appends a limit result to the discussion of a certain card problem, but it can hardly be said that the resulting formula was considered by de Moivre as a distribution, which may be the key point). Even Poisson's result does not seem to have been widely noted at the time, and it is not derived in a counting process context. The first discussions of counting problems known to us are by Seidel (1876) and Abbé (1879),

who treated the occurrence of thunderstorms and the number of blood cells in haemocytometer squares, respectively, and both apparently independently of Poisson's work. Indeed, Poisson's discovery of the distribution seems to have been lost sight of until attention was drawn to it in Von Bortkiewicz's (1898) monograph *Vas Gesetz der kleinen Zahlen*, which includes a systematic account of phenomena that fit the Poisson distribution, including, of course, the famous example of the number of deaths from horse kicks in the Prussian army.

Lyon and Thoma (1881), on Abbé's data, and Student (1907) gave further discussions of the blood cell problem, the latter paper being famous as one of the earliest applications of the chi-square goodness-of-fit test. Shortly afterward, the Poisson process arose simultaneously in two very important contexts. Erlang (1909) derived the Poisson distribution for the number of incoming calls to a telephone trunking system by supposing the numbers in disjoint intervals to be independent and considering the limit behaviour when the interval of observation is divided into an increasing number of equally sized subintervals. This effectively reproduces the Poisson distribution as the limit of the binomial, but Erlang was not aware of Poisson's work at the time, although he corrected the omission in later papers. Then, in 1910, Bateman, brought in as mathematical consultant by Rutherford and Geiger in connection with their classical experiment on the counting of α particles, obtained the Poisson probabilities as solutions to the family of differential equations

$$\begin{aligned} p_n'(t) &= -\lambda p_n(t) + p_{n-1}(t) \qquad (n \geq 1), \\ p_0'(t) &= -\lambda p_0(t). \end{aligned}$$

[Concerning the relation $p_0(t) = \mathrm{e}^{-\lambda t}$, Bateman (1910) commented that it "has been known for some time (Whitworth's *Choice and Chance*, 4th Ed., Proposition LI)," while Haight (1967) mentioned the result as a theorem of Boltzmann (1868) and quoted the reference to Whitworth, who does not indicate the sources of his results; in a Gresham lecture reproduced in Whitworth (1897, p. xxxiii), he wrote of Proposition LI as "a general theorem which I published in 1886, which met with rather rough treatment at the hands of a reviewer in *The Academy*." Whitworth's (1867) book evolved through five editions. It is easy to envisage repeated independent discovery of his Proposition LI.]

These equations represent a formulation in terms of a pure birth process and the first step in the rapid development of the theory of birth and death processes during the next two decades, with notable early papers by McKendrick (1914, 1926) and Yule (1924). This work preceded the general formulation of birth and death processes as Markov processes (themselves first studied by Markov more than a decade earlier) in the 1930s and is not of immediate concern, despite the close connection with point process problems. A similar remark can be made about branching processes, studied first by Bienaymé (see Heyde and Seneta, 1977) and of course by Galton and Watson

(1874). There are close links with point processes, particularly in the general case, but the early studies used special techniques that again lie a little outside the scope of our present discussion, and it was only from the 1940s onward that the links became important.

Closer in line with our immediate interests is the work on alternatives to the Poisson distribution. In many problems in ecology and elsewhere, it is found that the observed distribution of counts frequently shows a higher dispersion (i.e. a higher variance for a given value of the mean) than can be accounted for satisfactorily by the Poisson distribution, for which the variance/mean ratio is identically unity. The earliest and perhaps still the most widely used alternative is the *negative binomial distribution*, which figures in early papers by Student (1907), McKendrick (1914), and others. A particularly important paper for the sequel was the study by Greenwood and Yule (1920) of accident statistics, which provided an important model for the negative binomial, and in so doing sparked a controversy, still not entirely resolved, concerning the identifiability of the model describing accident occurrence. Since the accident process is a kind of point process in time, and since shades of the same controversy will appear in our own models, we briefly paraphrase their derivation. Before doing so, however, it is convenient to summarize some of the machinery for handling discrete distributions.

The principal tool is the *probability generating function* (p.g.f.) defined for nonnegative integer-valued random variables X by the equation

$$P(z) = \sum_0^\infty p_n z^n,$$

where $p_n = \Pr\{X = n\}$. It is worth mentioning that although generating functions have been used in connection with difference equations at least since the time of Laplace, their application to this kind of problem in the 1920s and 1930s was hailed as something of a technological breakthrough.

In Chapter 5, relations between the p.g.f., factorial moments, and cumulants are discussed. For the present, we content ourselves with the observation that the negative binomial distribution can be characterized by the form of its p.g.f.,

$$P(z) = \left(\frac{\mu}{1+\mu-z}\right)^\alpha \qquad (\alpha > 0, \mu > 0), \tag{1.2.1}$$

corresponding to values of the probabilities themselves,

$$p_n = \frac{(\alpha-1+n)!}{(\alpha-1)!\,n!}\left(\frac{\mu}{1+\mu}\right)^\alpha \left(\frac{1}{1+\mu}\right)^n.$$

[1]Note that there is a lack of agreement on terminology. Other authors, for example Johnson and Kotz (1969), would label this as a compound Poisson and would call the distribution we treat below under that name a generalized Poisson. The terminology we use is perhaps more common in texts on probability and stochastic processes; the alternative terminology is more common in the statistical literature.

Greenwood and Yule derived this distribution as an example of what we call a *mixed Poisson*[1] *distribution*; that is, it can be obtained from a Poisson distribution $p_n = \mathrm{e}^{-\lambda}\lambda^n/n!$ by treating the parameter λ as a random variable. If, in particular, λ is assumed to have the gamma distribution

$$\mathrm{d}F(\lambda) = \mu^\alpha \lambda^{\alpha-1} \big(\Gamma(\alpha)\big)^{-1} \mathrm{e}^{-\mu\lambda}\,\mathrm{d}\lambda,$$

then the resultant discrete distribution has p.g.f.

$$P(z) = \int_0^\infty \mathrm{e}^{\lambda(z-1)}\,\mathrm{d}F(\lambda) = \left(\frac{\mu}{1+\mu-z}\right)^\alpha,$$

$\mathrm{e}^{\lambda(z-1)}$ being the p.g.f. of the Poisson distribution with parameter λ.

It is not difficult to verify that the mean and variance of this negative binomial distribution equal α/μ and $(\alpha/\mu)(1+\mu^{-1})$, so that the variance/mean ratio of the distribution equals $1+\mu^{-1}$, exceeding by μ^{-1} the corresponding ratio for a Poisson distribution. Greenwood and Yule interpreted the variable parameter λ of the underlying Poisson distribution as a measure of individual 'accident proneness,' which was then averaged over all individuals in the population.

The difficulty for the sequel is that, as was soon recognized, many other models also give rise to the negative binomial, and these may have quite contradictory interpretations in regard to accidents. Lüders (1934) showed that the same distribution could be derived as an example of a *compound* Poisson distribution, meaning a random sum of independent random variables in which the number of terms in the sum has a Poisson distribution. If each term is itself discrete and has a *logarithmic distribution* with p.g.f.

$$P(z) = \frac{\log(1+\mu-z)}{\log\mu}, \tag{1.2.2}$$

and if the number of terms has a Poisson distribution with parameter α, then the resultant distribution has the identical p.g.f. (1.2.1) for the negative binomial (see Exercise 1.2.1). The interpretation here would be that all individuals are identical but subject to accidents in batches. Even before this, Eggenberger and Pólya (1923) and Pólya (1931) had introduced a whole family of distributions, for which they coined the term 'contagious distributions' to describe situations where the occurrence of a number of events enhances the probability of the occurrence of a further event, and had shown that the negative binomial distribution could be obtained in this way. If the mixed and compound models can be distinguished in principle by examining the joint distributions of the number of accidents in nonoverlapping intervals of a person's life, Cane (1974, 1977) has shown that there is no way in which the mixed Poisson and Pólya models can be distinguished from observations on individual case histories, for they lead to identical conditional distributions (see Exercise 1.2.2).

Another important contribution in this field is the work of Neyman (1939), who introduced a further family of discrete distributions, derived from consideration of a cluster model. Specifically, Neyman was concerned with distributions of beetle larvae in space, supposing these to have crawled some small distance from their initial locations in clusters of eggs. Further analysis of this problem resulted in a series of papers, written by Neyman in collaboration with E.L. Scott and other writers, which treated many different statistical questions relating to clustering processes in ecology, astronomy, and other subjects (see e.g. Neyman and Scott, 1958).

Many of these questions can be treated most conveniently by the use of generating functionals and moment densities, a theory that had been developing simultaneously as a tool for describing the evolution of particle showers and related problems in theoretical physics. The beginnings of such a general theory appear in the work of the French physicist Yvon (1935), but the main developments relate to the post-war period, and we therefore defer a further discussion to the following section.

Exercises and Complements to Section 1.2

1.2.1 *Poisson mixture of logarithmic distributions is negative binomial.* Verify that if $X_1, X_2, \ldots$ are independent r.v.s with the logarithmic distribution whose p.g.f. is in (1.2.2), and if N, independent of $X_1, X_2, \ldots$, is a Poisson r.v. with mean α, then $X_1 + \cdots + X_N$ has the negative binomial distribution in (1.2.1).

1.2.2 *Nonidentifiability in a model for accident proneness.* Suppose that an individual has n accidents in the time interval $(0, T)$ at $t_1 < t_2 < \cdots < t_n$. Evaluate the likelihood function for these n times for the two models:

(i) accidents occur at the epochs of a Poisson process at rate λ, where λ is fixed for each individual but may vary between individuals;

(ii) conditional on having experienced j accidents in $(0, t)$, an individual has probability $(k + j)\mu\,\mathrm{d}t/(1 + \mu t)$ of an accident in $(t, t + \mathrm{d}t)$, independent of the occurrence times of the j accidents in $(0, t)$; each individual has probability $k\mu\,\mathrm{d}t$ of an accident in $(0, \mathrm{d}t)$.

Show that the probabilities of n events in $(0, T)$ are Poisson and negative binomial, respectively, and deduce that the conditional likelihood, given n, is the same for (i) and (ii). See Cane (1974) for discussion.

1.2.3 The negative binomial distribution can also arise as the limit of the Pólya–Eggenberger distribution defined for integers n and $\alpha, \beta > 0$ by

$$p_k = \binom{n}{k} \frac{\Gamma(\alpha + k)\Gamma(\beta + n - k)\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + n)\Gamma(\alpha)\Gamma(\beta)} = \binom{-\alpha}{k} \frac{\Gamma(\alpha + \beta)n!\Gamma(\beta + n - k)}{\Gamma(\beta)(n - k)!\Gamma(\beta + n + \alpha)}\,.$$

When β and $n \to \infty$ with $n/\beta \to \mu$, a constant, and α fixed, show that $\{p_k\}$ has the p.g.f. in (1.2.1). [For further properties, see Johnson and Kotz (1969) and the papers cited in the text.]

1.2.4 *Neyman's Type A distribution* (e.g. Johnson and Kotz, 1969) has a p.g.f. of the form

$$\exp\left(\mu \sum_i \alpha_i \Big(\exp[-\lambda_i(1-z)] - 1 \Big)\right),$$

where $\alpha_i \geq 0$, $\sum_i \alpha_i = 1$, $\lambda_i > 0$, and $\mu > 0$, and arises as a cluster model. Give such a cluster model interpretation for the simplest case $\alpha_i = 1$ for $i = 1, \alpha_i = 0$ otherwise, and general $\lambda \equiv \lambda_1$ and μ.

1.2.5 Suppose that a (large) population evolves according to a one-type Galton–Watson branching process in which the distribution of the number of children has p.g.f. $P(z)$. Choose an individual at random in a particular generation. Show that the distribution of the number of sibs (sisters, say) of this randomly chosen individual has p.g.f. $P'(z)/P'(1)$ and that this is the same as for the number of aunts, or great-aunts, of this individual.
[*Hint*: Attempting to estimate the offspring distribution by using the observed family size distribution, when based on sampling via the children, leads to the distribution with p.g.f. $zP'(z)/P(1)$ and is an example of length-biased sampling that underlies the waiting-time paradox referred to in Sections 3.2 and 3.4. The p.g.f. for the number of great-aunts is used in Chapter 11.]

1.3. Some More Recent Developments

The period during and following World War II saw an explosive growth in theory and applications of stochastic processes. On the one hand, many new applications were introduced and existing fields of application were extended and deepened; on the other hand, there was also an attempt to unify the subject by defining more clearly the basic theoretical concepts. The monographs by Feller (1950) and Bartlett (1955) (preceded by mimeographed lecture notes from 1947) played an important role in stressing common techniques and exploring the mathematical similarities in different applications; both remain remarkably succinct and wide-ranging surveys.

From such a busy scene it is difficult to pick out clearly marked lines of development, and any selection of topics is bound to be influenced by personal preferences. Bearing such reservations in mind, we can attempt to follow through some of the more important themes into the post-war period.

On the queueing theory side, a paper of fundamental importance is Connie Palm's (1943) study of intensity fluctuations in traffic theory, a title that embraces topics ranging from the foundation of a general theory of the input stream to the detailed analysis of particular telephone trunking systems. Three of his themes, in particular, were important for the future of point processes. The first is the systematic description of properties of a renewal process, as a first generalization of the Poisson process as input to a service system. The notion of a *regeneration point*, a time instant at which the system reverts to a specified state with the property that the future evolution is independent of how the state was reached, has proved exceptionally fruitful in many different applications. In Palm's terminology, the Poisson process

is characterized by the property that every instant is a regeneration point, whereas for a general renewal process only those instants at which a new interval is started form regeneration points. Hence, he called a Poisson process a process *without aftereffects* and a renewal process a process with *limited aftereffects*. Another important idea was his realization that two types of distribution function are important in describing a stationary point process—the distribution of the time to the next event from a fixed but arbitrary origin and the distribution of the time to the next event from an arbitrary event of the process. The relations between the two sets of distributions are given by a set of equations now commonly called the Palm–Khinchin equations, Palm himself having exhibited only the simplest special case. A third important contribution was his (incomplete) proof of the first limit theorem for point processes: namely, that superposition of a large number of independent sparse renewal processes leads to a Poisson process in the limit. Finally, it may be worth mentioning that it was in Palm's paper that the term 'point processes' (Punktprozesse) was first used as such—at least to the best of our knowledge.

All these ideas have led to important further development. H. Wold (1948, 1949), also a Swedish mathematician, was one of the first to take up Palm's work, studying processes with Markov-dependent intervals that, he suggested, would form the next most complex alternative to the renewal model. Bartlett (1954) reviewed some of this early work. Of the reworkings of Palm's theory, however, the most influential was the monograph by Khinchin (1955), which provided a more complete and rigorous account of Palm's work, notably extended it in several directions, and had the very important effect of bringing the subject to the attention of pure mathematicians. Thus, Khinchin's book became the inspiration of much theoretical work, particularly in the Soviet Union and Eastern Europe. Ryll-Nardzewski's (1961) paper set out fundamental properties of point processes and provided a new and more general approach to Palm probabilities. Starting in the early 1960s, Matthes and co-workers developed many aspects concerned with infinitely divisible point processes and related questions. The book by Kerstan, Matthes and Mecke (1974) represented the culmination of the first decade of such work; extensive revisions and new material were incorporated into the later editions in English (1978) (referred to as MKM in this book) and in Russian (1982).

In applications, these ideas have been useful not only in queueing theory [for continuing development in this field, see the monographs of Franken *et al.* (1981) and Brémaud (1981)] but also in the study of level-crossing problems. Here the pioneering work was due to Rice (1944) and McFadden (1956, 1958). More rigorous treatments, using some of the Palm–Khinchin theory, were given by Leadbetter and other writers [see e.g. Leadbetter (1972) and the monographs by Cramér and Leadbetter (1967) and Leadbetter, Lindgren and Rootzen (1983)].

On a personal note in respect of much of this work, it is appropriate to remark that Belyaev, Franken, Grigelionis, König, Matthes, and one of us,

among others, were affected by the lectures and personal influence of Gnedenko (see Vere-Jones, 1997), who was a student of Khinchin.

Meanwhile, there was also rapid development on the theoretical physics front. The principal ideas here were the characteristic and generating functionals and product densities. As early as 1935, Kolmogorov suggested the use of the characteristic functional

$$\Phi(\xi) = \mathrm{E}(\mathrm{e}^{i\langle X,\xi\rangle})$$

as a tool in the study of random elements X from a linear space L; ξ is then an element from the space of linear functionals on L. The study of probability measures on abstract spaces remained a favourite theme of the Russian school of probability theory and led to the development of the weak convergence theory for measures on metric spaces by Prohorov (1956) and others, which in turn preceded the general study of random measures [e.g. Jiřina (1966) and later writers including the Swedish mathematicians Jagers (1974) and Kallenberg (1975)]. After the war, the characteristic functional was discussed by LeCam (1947) for stochastic processes and Bochner (1947) for random interval functions. Bochner's (1955) monograph, in particular, contains many original ideas that have only partially been followed up, for example, by Brillinger (1972). Kendall (1949) and Bartlett and Kendall (1951) appear to be the first to have used the characteristic functional in the study of specific population models.

Of more immediate relevance to point processes is the related concept of a *probability generating functional* (p.g.fl.) defined by

$$G[h] = \mathrm{E}\left[\prod_i h(x_i)\right] = \mathrm{E}\left[\exp\left(\int \log h(x)\, N(\mathrm{d}x)\right)\right],$$

where $h(x)$ is a suitable test function and the x_i are the points at which population members are located, that is, the atoms of the counting measures $N(\cdot)$. The p.g.fl. is the natural extension of the p.g.f., and, like the p.g.f., it has an expansion, when the total population is finite, in terms of the probabilities of the number of particles in the population and the probability densities of their locations. There is also an expansion, analogous to the expansion of the p.g.f. in terms of factorial moments, in terms of certain factorial moment density functions, or *product densities* as they are commonly called in the physical literature. Following the early work of Yvon noted at the end of Section 1.2, the p.g.fl. and product densities were used by Bogoliubov (1946), while properties of product densities were further explored in important papers by Bhabha (1950) and Ramakrishnan (1950). Ramakrishnan, in particular, gave formulae expressing the moments of the number of particles in a given set in terms of the product densities and Stirling numbers. Later, these ideas were considerably extended by Ramakrishnan, Janossy, Srinivasan, and others; an extensive literature exists on their application to cosmic ray showers summarized in the monographs by Janossy (1948) and Srinivasan (1969, 1974).

This brings us to another key point in the mathematical theory of point processes, namely the fundamental paper by Moyal (1962a). Drawing principally on the physical and ecological contexts, Moyal for the first time set out clearly the mathematical constructs needed for a theory of point processes on a general state space, clarifying the relations between such quantities as the product densities, finite-dimensional distributions, and probability generating functionals and pointing out a number of important applications. Independently, Harris (1963) set out similar ideas in his monograph on branching processes, subsequently (Harris, 1968, 1971) contributing important ideas to the general theory of point processes and the more complex subject of interacting particle systems.

In principle, the same techniques are applicable to other contexts where population models are important, but in practice the discussions in such contexts have tended to use more elementary, *ad hoc* tools. In forestry, for example, a key problem is the assessment of the number of diseased or other special kinds of trees in a given region. Since a complete count may be physically very difficult to carry out and expensive, emphasis has been on statistical sampling techniques, particularly of transects (line segments drawn through the region) and nearest-neighbour distances. Matérn's (1960) monograph brought together many ideas, models, and statistical techniques of importance in such fields and includes an account of point process aspects. Ripley's (1981) monograph covers some more recent developments.

On the statistical side, Cox's (1955) paper contained seeds leading to the treatment of many statistical questions concerning data generated by point processes and discussing various models, including the important class of doubly stochastic Poisson processes. A further range of techniques was introduced by Bartlett (1963), who showed how to adapt methods of time series analysis to a point process context and brought together a variety of different models. This work was extended to processes in higher dimensions in a second paper (Bartlett, 1964). Lewis (1964a) used similar techniques to discuss the instants of failure of a computer. The subsequent monograph by Cox and Lewis (1966) was a further important development that, perhaps for the first time, showed clearly the wide range of applications of point processes as well as extending many of the probabilistic and statistical aspects of such processes.

In the 1970s, perhaps the most important development was the rapid growth of interest in point processes in communications engineering (see e.g. Snyder, 1975). It is a remarkable fact that in nature, for example in nerve systems, the transfer of information is more often effected by pulse signals than by continuous signals. This fact seems to be associated with the high signal/noise ratios that it is possible to achieve by these means; for the same reason, pulse techniques are becoming increasingly important in communication applications. For such processes, just as for continuous processes, it is meaningful to pose questions concerning the prediction, interpolation, and estimation of signals, and the detection of signals against background noise (in this context, of random pulses). Since the signals are intrinsically nonnega-

tive, the distributions cannot be Gaussian, so linear models are not in general appropriate. Thus, the development of a suitable theory for point processes is closely linked to the development of nonlinear techniques in other branches of stochastic process theory. As in the applications to processes of diffusion type, martingale methods provide a powerful tool in the discussion of these problems, yielding, for example, structural information about the process and its likelihood function as well as more technical convergence results. Amongst other books, developments in this area were surveyed in Liptser and Shiryayev (1974; English translation 1977, 1978; 2nd ed. 2000), Brémaud (1981), and Jacobsen (1982).

The last quarter-century has seen both the emergence of new fields of applications and the consolidation of older ones. Here we shall attempt no more than a brief indication of major directions, with references to texts that can be consulted for more substantive treatments.

Spatial point processes, or *spatial point patterns* as they are often called, have become a burgeoning subject in their own right. The many fields of application include environmental studies, ecology, geography, astrophysics, fisheries and forestry, as well as substantially new topics such as image processing and spatial epidemic theory. Ripley (1981) and Diggle (1983) discuss both models and statistical procedures, while Cressie (1991) gives a broad overview with the emphasis on applications in biology and ecology. Image processing is discussed in the now classical work of Serra (1982). Theoretical aspects of spatial point patterns link closely with the fields of stereology and *stochastic geometry*, stemming from the seminal work of Roger Miles and, particularly, Rollo Davidson (see Harding and Kendall, 1974) and surveyed in Stoyan, Kendall and Mecke (1987, 2nd ed. 1995) and Stoyan and Stoyan (1994). There are also close links with the newly developing subject of random set theory; see Mathéron (1975) and Molchanov (1997). The broad-ranging set of papers in Barndorff-Nielsen *et al.* (1998) covers many of these applications and associated theory.

Time, space–time, and marked space–time point processes have continued to receive considerable attention. As well as in the earlier applications to queueing theory, reliability, and electrical engineering, they have found important uses in geophysics, neurophysiology, cardiology, finance, and economics. Applications in queueing theory and reliability were developed in the 1980s by Brémaud (1981) and Franken *et al.* (1981). Baccelli and Brémaud (1994) contains a more recent account. Second-order methods for the statistical analysis of such data, including spectral theory, are outlined in the now classic text of Cox and Lewis (1966) and in Brillinger (1975b). Snyder and Miller (1991) describe some of the more recent applications in medical fields. Extreme-value ideas in finance are discussed, from a rather different point of view than in Leadbetter *et al.* (1983) and Resnick (1987), in Embrechts *et al.* (1997). Prediction methods for point processes have assumed growing importance in seismological applications, in which context they are reviewed in Vere-Jones (1995).

Survival analysis has emerged as another closely related major topic, with applications in epidemiology, medicine, mortality, quality control, reliability, and other fields. Here the study of a single point process is usually replaced by the study of many individual processes, sometimes with only a small number of events in each, evolving simultaneously. Starting points include the early papers of Cox (1972b) and Aalen (1975). Andersen *et al.* (1993) give a major survey of modelling and inference problems in this field; their treatment includes an excellent introduction to point process concepts in general, emphasizing martingale concepts for inference, and the use of product-integral formulae.

The growing range of applications has led to an upsurge of interest in inference problems for point process models. Many of the texts referred to above devote a substantial part of their discussion to the practical implementation of inference procedures. General principles of inference for point processes are treated in the text by Liptser and Shiryayev already mentioned and in Kutoyants (1980, 1984), Karr (1986, 2nd ed. 1991), and Kutoyants (1998).

Theoretical aspects have also continued to flourish, particularly in the connections with statistical mechanics and stochastic geometry. Recent texts on basic theory include Kingman's (1993) beautiful discussion of the Poisson process and Last and Brandt's (1995) exposition of marked point processes. There are close connections between point processes and infinite particle systems (Liggett, 1999), while Georgii (1988) outlines ideas related to spatial processes and phase changes. Branching processes in higher-dimensional spaces exhibit many remarkable characteristics, some of which are outlined in Dawson *et al.* (2000). Very recently, Coram and Diaconis (2002), exploiting Diaconis and Evans (2000, 2001), have studied similarities between finite point processes of n points on the unit circle constructed from the eigenvalues of random unitary matrices from the unitary group U_n, and blocks of n successive zeros of the Riemann zeta function, where n depends on the distance from the real axis of the block of zeros.

CHAPTER 2

Basic Properties of the Poisson Process

The archetypal point processes are the Poisson and renewal processes. Their importance is so great, not only historically but also in illustrating and motivating more general results, that we prefer to give an account of some of their more elementary properties in this and the next two chapters before proceeding to more complex examples and the general theory of point processes.

For our present purposes, we shall understand by a point process some method of randomly allocating points to intervals of the real line or (occasionally) to rectangles or hyper-rectangles in a d-dimensional Euclidean space $\mathbb{R}^d$. It is intuitively clear and will be made rigorous in Chapters 5 and 9 that a point process is completely defined if the joint probability distributions are known for the number of events in all finite families of disjoint intervals (or rectangles, etc.). We call these joint or *finite-dimensional* distributions *fidi* distributions for short.

2.1. The Stationary Poisson Process

With the understanding just enunciated, the stationary Poisson process on the line is completely defined by the following equation, in which we use $N(a_i, b_i]$ to denote the number of events of the process falling in the half-open interval $(a_i, b_i]$ with $a_i < b_i \le a_{i+1}$:

$$\Pr\{N(a_i, b_i] = n_i, i = 1, \ldots, k\} = \prod_{i=1}^{k} \frac{[\lambda(b_i - a_i)]^{n_i}}{n_i!} \, \mathrm{e}^{-\lambda(b_i - a_i)}. \tag{2.1.1}$$

This definition embodies three important features:

(i) the number of points in each finite interval $(a_i, b_i]$ has a Poisson distribution;

(ii) the numbers of points in disjoint intervals are independent random variables; and

(iii) the distributions are stationary: they depend only on the lengths $b_i - a_i$ of the intervals.

Thus, the joint distributions are multivariate Poisson of the special type in which the variates are independent.

Let us first summarize a number of properties that follow directly from (2.1.1). The mean $M(a, b]$ and variance $V(a, b]$ of the number of points falling in the interval $(a, b]$ are given by

$$M(a, b] = \lambda(b - a) = V(a, b]. \tag{2.1.2}$$

The constant λ here can be interpreted as the *mean rate* or *mean density* of points of the process. It also coincides with the *intensity* of the process as defined following Proposition 3.3.I.

The facts that the mean and variance are equal and that both are proportional to the length of the interval provide a useful diagnostic test for the stationary Poisson process: estimate the mean $M(a, b]$ and the variance $V(a, b]$ for half-open intervals $(a, b]$ over a range of different lengths, and plot the ratios $V(a, b]/(b - a)$. The estimates should be approximately constant for a stationary Poisson process and equal to the mean rate. Any systematic departure from this constant value indicates some departure either from the Poisson assumption or from stationarity [see Exercise 2.1.1 and Cox and Lewis (1966, Section 6.3) for more discussion].

Now consider the relation, following directly from (2.1.1), that

$$\Pr\{N(0, \tau] = 0\} = \mathrm{e}^{-\lambda\tau} \tag{2.1.3}$$

is the probability of finding no points in an interval of length τ. This may also be interpreted as the probability that the random interval extending from the origin to the point first appearing to the right of the origin has length exceeding τ. In other words, it gives nothing other than the *survivor function* for the length of this interval. Equation (2.1.3) therefore shows that the interval under consideration has an exponential distribution. From stationarity, the same result applies to the length of the interval to the first point of the process to the right of any arbitrarily chosen origin and then equally to the interval to the first point to the left of any arbitrarily chosen origin. In this book, we follow queueing terminology in calling these two intervals the *forward* and *backward recurrence times*; thus, for a Poisson process both forward and backward recurrence times are exponentially distributed with mean $1/\lambda$.

Using the independence property, we can extend this result to the distribution of the time interval between any two consecutive points of the process, for the conditional distribution of the time to the next point to the right of the origin, given a point in $(-\Delta, 0]$, has the same exponential form, which, being independent of Δ, is therefore the limiting form of this conditional distribution as $\Delta \to 0$. When such a unique limiting form exists, it can be

identified with the distribution of the time interval between two arbitrary points of the process (see also Section 3.4 in the next chapter). Similarly, by considering the limiting forms of more complicated joint distributions, we can show that successive intervals are independently distributed as well as having exponential distributions (see Exercises 2.1.2–4 and, for extensions to $\mathbb{R}^2$ and $\mathbb{R}^3$, Exercises 2.1.7–8).

On the other hand, the particular interval containing the origin is not exponentially distributed. Indeed, since it is equal to the sum of the forward and backward recurrence times, and each of these has an exponential distribution and is independent of the other, its distribution must have an Erlang (or gamma) distribution with density $\lambda^2 x e^{-\lambda x}$. This result has been referred to as the 'waiting-time paradox' because it describes the predicament of a passenger arriving at a bus stop when the bus service follows a Poisson pattern. The intuitive explanation is that since the position of the origin (the passenger's arrival) is unrelated to the process governing the buses, it may be treated as effectively uniform over any given time interval; hence, it is more likely to fall in a large rather than a small interval. See Sections 3.2 and 3.4 for more detail and references.

Now let t_k, $k = 1, 2, \ldots$, denote the time from the origin $t_0 = 0$ to the kth point of the process to the right of the origin. Then we have

$$\{t_k > x\} = \{N(0, x] < k\} \tag{2.1.4}$$

in the sense that the expressions in braces describe identical events. Hence, in particular, their probabilities are equal. But the probability of the event on the right is given directly by (2.1.1), so we have

$$\Pr\{t_k > x\} = \Pr\{N(0, x] < k\} = \sum_{j=0}^{k-1} \frac{(\lambda x)^j}{j!} \, \mathrm{e}^{-\lambda x}. \tag{2.1.5}$$

Differentiating this expression, which gives the survivor function for the time to the kth point, we obtain the corresponding density function

$$f_k(x) = \frac{\lambda^k x^{k-1}}{(k-1)!} \, \mathrm{e}^{-\lambda x}, \tag{2.1.6}$$

which is again an Erlang distribution. Since the time to the kth event can be considered as the sum of the lengths of the k random intervals $(t_0, t_1]$, $(t_1, t_2], \ldots, (t_{k-1}, t_k]$, which as above are independently and exponentially distributed, this provides an indirect proof of the result that the sum of k independent exponential random variables has the Erlang distribution.

In much the same vein, we can obtain the *likelihood* of a finite realization of a Poisson process. This may be defined as the probability of obtaining the given number of observations in the observation period, times the joint conditional density for the positions of those observations, given their number. Suppose that there are N observations on $(0, T]$ at time points $t_1, \ldots, t_N$. From (2.1.1), we can write down immediately the probability of obtaining

single events in $(t_i - \Delta, t_i]$ and no points on the remaining part of $(0, T]$: it is just

$$\mathrm{e}^{-\lambda T} \prod_{j=1}^{N} \lambda \Delta.$$

Dividing by Δ^N and letting $\Delta \to 0$, to obtain the density, we find as the required likelihood function

$$L_{(0,T]}(N; t_1, \ldots, t_N) = \lambda^N \mathrm{e}^{-\lambda T}. \tag{2.1.7}$$

Since the probability of obtaining precisely N events in $(0, T]$ is equal to $[(\lambda T)^N / N!]\mathrm{e}^{-\lambda T}$, this implies *inter alia* that the conditional density of obtaining points at $(t_1, \ldots, t_N)$, given N points in the interval, is just $N!/T^N$, corresponding to a uniform distribution over the hyperoctant

$$0 \le t_1 \le \cdots \le t_N \le T.$$

One point about this result is worth stressing. It corresponds to treating the points as *indistinguishable* apart from their locations. In physical contexts, however, we may be concerned with the positions of N physically distinguishable particles. The factor $N!$, which arises in the first instance as the volume of the unit hyperoctant, can then be interpreted also as the combinatorial factor representing the number of ways the N distinct particles can be allocated to the N distinct time points. The individual particles are then to be thought of as uniformly and independently distributed over $(0, T]$. It is in this sense that the conditional distributions for the Poisson process are said to correspond to the distributions of N particles laid down uniformly at random on the interval $(0, T]$ (see Exercise 2.1.5). Furthermore, either from this result or directly from (2.1.1), we obtain

$$\begin{aligned} \Pr\{N(0,x] = k \mid N(0,T] = N\} &= \frac{\Pr\{N(0,x] = k, N(x,T] = N-k\}}{\Pr\{N(0,T] = N\}} \\ &= \binom{N}{k} \big(p_{x,T}\big)^k \big(1 - p_{x,T}\big)^{N-k}, \end{aligned} \tag{2.1.8}$$

where $p_{x,T} = x/T$, representing a binomial distribution for the number in the subinterval $(0, x]$, given the number in the larger interval $(0, T]$.

Most of the results in this section extend both to higher dimensions and to nonstationary processes (see Exercises 2.1.6–8). We conclude the present section by mentioning the simple but important extension to a Poisson process with time-varying rate $\lambda(t)$, commonly called the nonhomogeneous or *inhomogeneous* Poisson process. The process can be defined exactly as in (2.1.1), with the quantities $\lambda(b_i - a_i) = \int_{a_i}^{b_i} \lambda \, \mathrm{d}x$ replaced wherever they occur by quantities

$$\Lambda(a_i, b_i] = \int_{a_i}^{b_i} \lambda(x) \, \mathrm{d}x.$$

Thus, the joint distributions are still Poisson, and the independence property still holds. Furthermore, conditional distributions now correspond to particles

independently distributed on $(0, T]$ with a common distribution having density function $\lambda(x)/\Lambda(0, T]$ $(0 \le x \le T)$. The construction of sample realizations is described in Exercise 2.1.6, while the likelihood function takes the more general form

$$L_{(0,T]}(N; t_1, \ldots, t_N) = \mathrm{e}^{-\Lambda(0,T]} \prod_{i=1}^{N} \lambda(t_i) = \exp\left(-\int_0^T \lambda(t)\,\mathrm{d}t + \int_0^T \log \lambda(t)\, N(\mathrm{d}t) \right). \tag{2.1.9}$$

From this expression, we can see that results for the nonstationary Poisson process can be derived from those for the stationary case by a deterministic time change $t \mapsto u(t) \equiv \Lambda(0, t]$. In other words, if we write $N(t) = N(0, t]$ (all $t \ge 0$) and define a new point process by

$$\widetilde{N}(t) = N\big(u^{-1}(t)\big),$$

then $\widetilde{N}(t)$ has the rate quantity $\tilde{\Lambda}(0, t) = u(u^{-1}(t)) = t$ and is therefore a stationary Poisson process at unit rate.

In Chapters 7 and 14, we shall meet a remarkable extension of this last result, due to Papangelou (1972a, b): any point process satisfying a simple continuity condition can be transformed into a Poisson process if we allow a random time change in which $\Lambda[0, t]$ depends on the past of the process up to time t. Papangelou's result also implies that (2.1.9) represents the typical form of the likelihood for a point process: in the general case, all that is needed is to replace the absolute rate $\lambda(t)$ in (2.1.9) by a conditional rate that is allowed to depend on the past of the process.

Other extensions lead to the class of mixed Poisson processes (see Exercise 2.1.9) and Cox processes treated in Chapter 6.

Exercises and Complements to Section 2.1

2.1.1 Let $N_1, \ldots, N_n$ be i.i.d. like the Poisson r.v. N with mean $\mu = \mathrm{E}N$, and write $\overline{N} = (N_1 + \cdots + N_n)/n$ for the sample mean. When μ is sufficiently large, indicate why the sample index of dispersion

$$Z = \sum_{j=1}^{n} \frac{(N_j - \overline{N})^2}{\overline{N}}$$

has a distribution approximating that of a χ^2_{n-1} r.v. Darwin (1957) found approximations to the distribution of Z for a general distribution for N based on its cumulants, illustrating his work via the Neyman, negative binomial, and Thomas distributions (see also Kathirgamatamby, 1953).

2.1.2 *Exponential distribution order properties.* Let $X_1, \ldots, X_n$ be i.i.d. exponential r.v.s on $(0, \infty)$ with $\Pr\{X_1 > x\} = \mathrm{e}^{-\lambda x}$ $(x \ge 0)$ for some positive finite λ.

(a) Let $X_{(1)} < \cdots < X_{(n)}$ be the order statistics of $X_1, \ldots, X_n$. Then $(X_{(1)}, \ldots, X_{(n)})$ has the same distribution as the vector whose kth component is

$$\frac{X_n}{n} + \frac{X_{n-1}}{n-1} + \cdots + \frac{X_{n-k+1}}{n-k+1}.$$

(b) Write $Y = X_1 + \cdots + X_n$ and set $Y_{(k)} = (X_1 + \cdots + X_k)/Y$. Then $Y_{(1)}, \ldots, Y_{(n-1)}$ are the order statistics of $n-1$ i.i.d. r.v.s uniformly distributed on $(0,1)$.

2.1.3 *Exponential r.v.s have no memory.* Let X be exponentially distributed as in Exercise 2.1.2, and for any nonnegative r.v. Y that is independent of X, define an r.v. X_Y as any r.v. whose d.f. has as its tail

$$R(z) \equiv \Pr\{X_Y > z\} = \Pr\{X > Y + z \mid X > Y\}.$$

Then X_Y and X have the same d.f.
[There exist innumerable characterizations of exponential r.v.s via their lack of memory properties; many are surveyed in Galambos and Kotz (1978).]

2.1.4 A process satisfying (2.1.1) has

$$\Pr\{N(t-x-\Delta, t-\Delta] = 0,\ N(t-\Delta, t] = 1,\ N(t, t+y] = 0 \mid N(t-\Delta, t] > 0\} \\ \to \mathrm{e}^{-\lambda x}\mathrm{e}^{-\lambda y} \quad (\Delta \to 0),$$

showing the stochastic independence of successive intervals between points of the process.

2.1.5 *Order statistics property of Poisson process.* Denote the points of a stationary Poisson process on $\mathbb{R}_+$ by $t_1 < t_2 < \cdots < t_{N(T)} < \cdots$, where for any positive T, $t_{N(T)} \le T < t_{N(T)+1}$. Let $u_{(1)} < \cdots < u_{(n)}$ be the order statistics of n i.i.d. points uniformly distributed on $[0,T]$. Show that, conditional on $N(T) = n$, the distributions of $\{u_{(i)}: i = 1, \ldots, n\}$ and $\{t_i: i = 1, \ldots, n\}$ coincide.

2.1.6 *Conditional properties of inhomogeneous Poisson processes.* Given a finite measure $\Lambda(\cdot)$ on a c.s.m.s. $\mathcal{X}$, let $\{t_1, \ldots, t_{N(\mathcal{X})}\}$ be a realization of an inhomogeneous Poisson process on $\mathcal{X}$ with parameter measure $\Lambda(\cdot)$.

(a) *I.i.d. property.* Let r.v.s $U_1, \ldots, U_n$ be i.i.d. on $\mathcal{X}$ with probability distribution $\Lambda(\cdot)/\Lambda(\mathcal{X})$. Show that the joint distributions of $\{U_i\}$ coincide with those of $\{t_i\}$ conditional on $N(\mathcal{X}) = n$.

(b) *Binomial distribution.* When $\mathcal{X} = (0,T]$, show that (2.1.8) still holds for the process $N(\cdot)$ with $p_{x,T} = \Lambda(x)/\Lambda(T)$.

(c) *Thinning construction.* To construct a realization on $(0,T]$ of an inhomogeneous Poisson process Π_1 for which the local intensity $\lambda(\cdot)$ satisfies $0 \le \lambda(u) \le \lambda_{\max}$ $(0 < u \le T)$ for some finite positive constant $\lambda_{\max}$, first construct a realization of a stationary Poisson process with rate $\lambda_{\max}$ (using the fact that successive intervals are i.i.d. exponential r.v.s with mean $1/\lambda_{\max}$), yielding the points $0 < t_l < t_2 < \cdots$, say. Then, independently for each $k = 1, 2, \ldots$, retain t_k as a point of Π_1 with probability $\lambda(t_k)/\lambda_{\max}$ and otherwise delete it. Verify that the residual set of points satisfies the independence axiom and that

$$\mathrm{E}(\#\{j: 0 < t_j < u, t_j \in \Pi_1\}) = \int_0^u \lambda(v)\,\mathrm{d}v.$$

[See also Lewis and Shedler (1976) and Algorithm 7.5.II.]

2.1.7 *Avoidance functions of Poisson process in* $\mathbb{R}^d$. The distance X of the point closest to the origin of a Poisson process in $\mathbb{R}^d$ with rate λ satisfies

$$\Pr\{X > y\} = \exp\big(-\lambda v_d(y)\big),$$

where $v_d(y) = y^d v_d(1)$ is the volume of a sphere of radius y in $\mathbb{R}^d$. In particular,

(i) in $\mathbb{R}^1$, $\Pr\{X > y\} = \mathrm{e}^{-2\lambda y}$;

(ii) in $\mathbb{R}^2$, $\Pr\{X > y\} = \mathrm{e}^{-\pi\lambda y^2}$;

(iii) in $\mathbb{R}^3$, $\Pr\{X > y\} = \mathrm{e}^{-(4\pi/3)\lambda y^3}$.

These same expressions also hold for the nearest-neighbour distance of an arbitrarily chosen point of the process.

2.1.8 *Simulating a Poisson process in* $\mathbb{R}^d$. Using the notation of Exercise 2.1.6, we can construct a realization of a Poisson process Π_d in a neighbourhood of the origin in $\mathbb{R}^d$ by adapting Exercises 2.1.6 and 2.1.7 to give an inhomogeneous Poisson process on $(0, T)$ with intensity $\lambda(\mathrm{d}/\mathrm{d}y)v_d(y)$ and then, denoting these points by $r_1, r_2, \ldots$, taking the points of Π_d as having polar coordinates (r_j, θ_j), where θ_j are points independently and uniformly distributed over the surface of the unit sphere in $\mathbb{R}^d$.

[An alternative construction for r_j is to use the fact that $\lambda(v_d(r_j) - v_d(r_{j-1}))$, with $r_0 = 0$, are i.i.d. exponential r.v.s with unit mean. See also Quine and Watson (1984). The *efficient* simulation of a Poisson process in a d-dimensional hypersphere, at least for small d, is to choose a point at random in a d-dimensional hypercube containing the hypersphere and use a rejection method of which Exercise 2.1.6(c) is an example.]

2.1.9 (a) *Mixed Poisson process.* A point process whose joint distributions are given by integrating λ in the right-hand side of (2.1.1) with respect to some d.f. defines a *mixed Poisson process* since the distributions come from regarding λ as a random variable. Verify that

$$\begin{aligned} N(0,t]/t &\to_{\text{a.s.}} \lambda \qquad (t \to \infty), \\ \mathrm{E}N(0,t] &= (\mathrm{E}\lambda)t, \\ \operatorname{var} N(0,t] &= (\mathrm{E}\lambda)t + (\operatorname{var}\lambda)t^2 \geq \mathrm{E}N(0,t], \end{aligned}$$

with strict inequality unless $\operatorname{var}\lambda = 0$.

(b) *Compound Poisson process.* Let $Y, Y_1, Y_2, \ldots$ be i.i.d. nonnegative integer-valued r.v.s with probability generating function $g(z) = \mathrm{E}z^Y$ $(|z| \leq 1)$, and let them be independent of a Poisson process N_c at rate λ; write $N_c(t) = N_c(0,t]$. Then

$$N(0,t] \equiv \sum_{i=1}^{N_c(t)} Y_i$$

defines the counting function of a *compound Poisson process* for which

$$\begin{aligned} \mathrm{E}z^{N(0,t]} &= \exp\big[-\lambda t(1-g(z))\big], \\ \mathrm{E}N(0,t] &= \lambda(\mathrm{E}Y)t, \\ \operatorname{var} N(0,t] &= \lambda(\operatorname{var} Y)t + \lambda(\mathrm{E}Y)^2 t = \lambda[\mathrm{E}(Y^2)]t \\ &= [\mathrm{E}N_c(t)](\operatorname{var} Y) + [\operatorname{var} N_c(t)](\mathrm{E}Y)^2 \geq \mathrm{E}N(0,t], \end{aligned}$$

with strict inequality unless $\mathrm{E}[Y(Y-1)] = 0$, i.e. $Y = 0$ or 1 a.s.

[Both the mixed and compound Poisson processes are in general overdispersed compared with a Poisson process in the sense that $(\operatorname{var} N(0,t])/\mathrm{E}N(0,t] \geq 1$, with equality holding only in the exceptional cases as noted.]

2.1.10 For a Poisson process with the cyclic intensity function

$$\lambda(t) = \lambda \exp[\kappa \sin(\omega_0 t + \theta)]/I_0(\kappa) \qquad (\kappa \geq 0,\ \omega_0 > 0,\ 0 \leq \theta < 2\pi,\ \lambda > 0),$$

where $I_0(\kappa) = \int_0^{2\pi} \exp(\kappa \sin u)\,\mathrm{d}u$ is the modified Bessel function of the first kind of zero order, the likelihood [see (2.1.9) above] of the realization $t_1, \ldots, t_N$ on the interval $(0, T)$ where, for convenience of simplifying the integral below, T is a multiple of the period $2\pi/\omega_0$, equals

$$\exp\left(-\int_0^T \frac{\lambda \exp[\kappa \sin(\omega_0 t + \theta)]}{I_0(\kappa)}\,\mathrm{d}t\right)\left(\frac{\lambda}{I_0(\kappa)}\right)^N \exp\left(\kappa \sum_{i=1}^N \sin(\omega_0 t_i + \theta)\right)$$

$$= \mathrm{e}^{-\lambda T/2\pi}\left(\frac{\lambda}{I_0(\kappa)}\right)^N \exp\left(\kappa \sum_{i=1}^N \sin(\omega_0 t_i + \theta)\right).$$

Consequently, N is a sufficient statistic for λ, and, when the frequency ω_0 is known,

$$\left(N, \sum_{i=1}^N \sin \omega_0 t_i, \sum_{i=1}^N \cos \omega_0 t_i\right) \equiv (N, S, C) \quad \text{say},$$

are jointly sufficient statistics for the parameters $(\lambda, \kappa, \theta)$, the maximum likelihood estimates $(\hat\lambda, \hat\kappa, \hat\theta)$ being determined by $\hat\lambda = 2\pi N/T$, $\tan\hat\theta = C/S$, and $(\mathrm{d}/\mathrm{d}\kappa) \log I_0(\kappa)\big|_{\kappa=\hat\kappa} = S/(N\cos\hat\theta) = (S^2 + C^2)^{1/2}/N$ (the constraints that $\hat\kappa \geq 0$ and that S and $\cos\theta$ are of the same sign determine which root $\hat\theta$ is taken). [See Lewis (1970) and Kutoyants (1984, Chapter 4) for more details.]

2.2. Characterizations of the Stationary Poisson Process: I. Complete Randomness

In applications, the Poisson process is sometimes referred to simply as a *random distribution* of points on a line (as if there were no alternative random processes!) or slightly more specifically as a *purely random* or *completely random* process. In all these terminologies, what is in view is the fundamental independence property referred to in (ii) under (2.1.1). We start our discussion of characterizations by examining how far this property alone is capable of characterizing the Poisson process. More precisely, let us assume that we are given a point process satisfying the assumptions below and examine how far the distributions are determined by them.

Assumptions 2.2.I.

(i) *The number of points in any finite interval is finite and not identically zero.*

(ii) *The numbers in disjoint intervals are independent random variables.*

(iii) *The distribution of $N(a + t, b + t]$ is independent of t.*

For brevity, we speak of a process satisfying (i) as *boundedly finite and non-null*, while property (ii) may be referred to as *complete independence* and (iii) as *(crude) stationarity*.

Theorem 2.2.II. *Under Assumptions* 2.2.I, *the probability generating function (p.g.f.)*

$$P(z,\tau) = \mathrm{E}(z^{N(0,\tau]})$$

can be written uniquely in the form

$$P(z,\tau) = \mathrm{e}^{-\lambda\tau[1-\Pi(z)]}, \tag{2.2.1}$$

where λ *is a positive constant and* $\Pi(z) = \sum_{n=1}^{\infty} \pi_n z^n$ *is the p.g.f. of a discrete distribution having no zero term.*

Remark. From the stationarity and independence assumptions, all the joint distributions can be written down once the form of (2.2.1) is given, so that (2.2.1) is in fact sufficient to specify the process completely. Hence, the assumption of crude stationarity suffices in the case of the Poisson process to ensure its (complete) stationarity (see Definition 3.2.I below).

Proof. Since $N(a,b]$ is a monotonically increasing function of b, it is clear that $P(z,\tau)$ is a monotonically *decreasing* function of τ for any fixed z with $0 \le z \le 1$, while $Q(z,\tau) = -\log P(z,\tau)$, finite because of Assumption 2.2.I(i), is a monotonically *increasing* nonnegative function of τ. Also, since

$$N(0,\tau_1+\tau_2] = N(0,\tau_1] + N(\tau_1,\tau_1+\tau_2],$$

it follows from the stationarity and independence assumptions that

$$\begin{aligned} P(z,\tau_1+\tau_2) &= P(z,\tau_1)P(z,\tau_2), \\ Q(z,\tau_1+\tau_2) &= Q(z,\tau_1) + Q(z,\tau_2). \end{aligned} \tag{2.2.2}$$

Now it is well known (see e.g. Lemma 3.6.III) that the only monotonic solutions of the functional equation (2.2.2) are of the form

$$Q(z,\tau) = \text{constant} \times \tau,$$

where in this case the constant is a function of z, $C(z)$ say. Thus, for all $\tau > 0$ we can write

$$P(z,\tau) = \mathrm{e}^{-\tau C(z)} \tag{2.2.3}$$

for some uniquely determined function $C(z)$.

Consider first the case $z = 0$. From Assumption 2.2.I(i), $N(0,\tau] \not\equiv 0$, so $P(0,\tau) \not\equiv 1$, and hence $C(0) \neq 0$. Now

$$\{N(0,1] \ge n\} \supseteq \bigcap_{k=1}^{n} \Big\{ N\Big(\frac{k-1}{n}, \frac{k}{n}\Big] \ge 1 \Big\},$$

so using the independence assumption and (2.2.3), we have

$$\Pr\{N(0,1] \ge n\} \ge \big(\Pr\{N(0,1/n] \ge 1\}\big)^n = \big(1 - \mathrm{e}^{-C(0)/n}\big)^n.$$

If now $C(0) = \infty$, then $\Pr\{N(0,1] \geq n\} = 1$ (all $n = 1, 2, \ldots$), contradicting Assumption 2.2.I(i) that $N(0,1]$ is a.s. finite. Thus, we conclude that

$$0 < C(0) < \infty. \tag{2.2.4}$$

Define quantities λ and $\Pi(z)$ by

$$\lambda = C(0) \qquad \text{and} \qquad \Pi(z) = \frac{C(0) - C(z)}{C(0)} = \frac{\log P(z,\tau) - \log P(0,\tau)}{-\log P(0,\tau)},$$

the finiteness and nonnegativity of $\Pi(z)$ on $0 \leq z \leq 1$ being ensured by the monotonicity in z of $P(z,\cdot)$. From (2.2.3) and (2.2.4), it follows that $P(z,\tau) \to 1$ $(\tau \to 0)$ for every fixed z in $0 \leq z \leq 1$, so from (2.2.3) we have

$$\tau C(z) = 1 - P(z,\tau) + o(\tau) \qquad (\tau \downarrow 0),$$

from which also

$$\Pi(z) = \lim_{\tau \downarrow 0} \frac{P(z,\tau) - P(0,\tau)}{1 - P(0,\tau)}.$$

This representation expresses $\Pi(\cdot)$ as the limit of p.g.f.s, namely the p.g.f.s of the conditional probabilities

$$\pi_{k|\tau} \equiv \Pr\{N(0,\tau] = k \mid N(0,\tau] > 0\}.$$

The definition of $\Pi(z)$ shows that it inherits from $P(z,\tau)$ the property of continuity as $z \uparrow 1$, and therefore the continuity theorem for p.g.f.s (see e.g. Feller, 1968, Section XI.6) ensures that $\Pi(z)$ must also be a p.g.f., $\Pi(z) = \sum \pi_k z^k$ say, where

$$\pi_k = \lim_{\tau \downarrow 0} \pi_{k|\tau} = \lim_{\tau \downarrow 0} \Pr\{N(0,\tau] = k \mid N(0,\tau] > 0\} \qquad (k = 0, 1, \ldots). \tag{2.2.5}$$

In particular, $\pi_0 = \Pi(0) = 0$.

We have thus established the required form of the representation in (2.2.1). Uniqueness follows from the uniqueness of $P(z,\tau)$, which defines $C(z)$ by (2.2.3), and $C(z)$ in turn defines λ and $\Pi(z)$. □

The process defined by Assumptions 2.2.I is clearly more general than the Poisson process, to which it reduces only in the case $\pi_1 = 1$, $\pi_k = 0$ $(k \neq 1)$. The clue to its interpretation comes from the limit relation (2.2.5), which suggests that $\{\pi_k\}$ should be interpreted as a 'batch-size' distribution, where 'batch' refers to a collection of points of the process located at the same time point. None of our initial assumptions precludes the possibility of such batches. The distribution of the *number* of such batches in $(0,1)$ is found by replacing $\Pi(z)$ by z in (2.2.1), and therefore it is Poisson with rate λ. Thus, the general process defined by Assumptions 2.2.I can be described as consisting of a succession of batches, the successive batch sizes or *multiplicities* being independent random variables [as follows readily from Assumption 2.2.I(ii)] having the common distribution $\{\pi_k\}$, and the number of batches following

a Poisson process with constant rate λ. Recognizing that (2.2.1) specifies the p.g.f. of a compound Poisson distribution, we refer to the process as the *compound Poisson process* [see the footnote on p.10 regarding terminology].

Processes with batches represent an extension of the intuitive notion of a point process as a random placing of points over a region. They are variously referred to as nonorderly processes, processes with multiple points, compound processes, processes with positive integer marks, and so on. For a general proof of the existence of a batch-size distribution for stationary point processes, see Proposition 3.3.VII. It should be noted that the uniqueness of the representation (2.2.1) breaks down once we drop the convention $\pi_0 = 0$. Indeed, given any p.g.f. $\Pi(\cdot)$ as in (2.2.1), let π_0^* be any number in $0 \le \pi_0^* < 1$, and define $\lambda^* = \lambda/(1-\pi_0^*)$, $\pi_n^* = (1-\pi_0^*)\pi_n$. Then $\Pi^*(z) \equiv \sum_{n=0}^{\infty} \pi_n^* z^n = \pi_0^* + (1-\pi_0^*)\Pi(z)$, and

$$\lambda^*\big(1-\Pi^*(z)\big) = \lambda(1-\pi_0^*)^{-1}\{(1-\pi_0^*)[1-\Pi(z)]\} = \lambda\big(1-\Pi(z)\big).$$

The interpretation of this nonuniqueness is that if we increase the rate of occurrence of batches, we may compensate for this increase by observing only those batches with nonzero batch size.

We obtain an alternative interpretation of the process by writing (2.2.1) in the form

$$P(z,\tau) = \prod_{k=1}^{\infty} \exp[-\lambda\pi_k\tau(1-z^k)],$$

corresponding to a representation of the total as the sum of independent contributions from a countable family of simpler processes, the kth of which may be regarded as a modified Poisson process in which the rate of occurrence of points is equal to $\lambda\pi_k$ and each such point is treated as a batch of fixed size k. In this representation, the process is regarded as a superposition of independent component processes, each of Poisson type but with fixed batch size. Since both interpretations lead to the same joint distributions and hence to the same probability structures, they must be regarded as equivalent.

Theorem 2.2.II may also be regarded as a special case of the more general theorem of Lévy on the structure of processes with stationary independent increments (see e.g. Loève, 1963, Section 37). In our case, there can be no Gaussian component (since the realizations are monotonic), no drift component (since the realizations are integer-valued), and the Poisson components must have positive integral jumps. Because a process has independent increments if and only if the distributions of the increment over any finite interval are infinitely divisible, (2.2.1) also gives the general form of an infinitely divisible distribution taking values only on the nonnegative integers [see Exercise 2.2.2 and Feller (1968, Section XII.2)].

Analytically, the condition corresponding to the requirement of no batches, or points occurring one at a time, is clearly $\pi_1 = 1$, or equivalently

$$\begin{aligned}\Pr\{N(0,\tau] > 1\} &= o(\Pr\{N(0,\tau] > 0\}) \\ &= o(1-\mathrm{e}^{-\lambda\tau}) = o(\tau) \qquad \text{for } \tau \downarrow 0. \end{aligned} \tag{2.2.6}$$

More generally, a stationary process satisfying this condition was called by Khinchin (1955) an orderly process (Russian *ordinarnii*), and we follow this terminology for the time being, as contrasted with the sample path terminology of a *simple* point process. The relations between analytical and sample path properties are discussed later in Section 3.3 and Chapter 9. For the present, suffice it to be noted that the analytical condition (2.2.6) is equivalent to the absence of batches with probability 1 (see Exercise 2.2.4). Using the notion of an orderly process, we obtain the following characterization of the Poisson process as a corollary to Theorem 2.2.II.

Theorem 2.2.III. *A stationary point process satisfying Assumption* 2.2.I(i) *is a Poisson process if and only if* (a) *it has the complete independence property* 2.2.I(ii) *and* (b) *it is orderly.*

Exercises and Complements to Section 2.2

2.2.1 In equation (2.2.3), $P(z,\tau) \to 1$ $(z \to 1)$ for every finite τ (why?), and equation (2.2.2) and $\lambda\tau > 0$ suffice to check that $\Pi(1) = 1$. (A general proof, using only stationarity and not the Poisson assumption, is given in Proposition 3.3.VIII below.)

2.2.2 Call the p.g.f. $P(z)$ *infinitely divisible* when for $0 \le z \le 1$ its uniquely defined nonnegative kth root $P_{1/k}(z) \equiv (P(z))^{1/k}$ is a p.g.f. for every positive integer. Then show that unless $P(z) = 1$ for all $0 \le z \le 1$:

(a) $p_0 = P(0) > 0$;

(b) $(P(z)/p_0)^{1/k} \to 1 \quad (k \to \infty)$;

(c) $$\frac{\log P(z) - \log P(0)}{-\log P(0)} = \lim_{k\uparrow\infty} \frac{P_{1/k}(z) - P_{1/k}(0)}{1 - P_{1/k}(0)};$$

(d) the left-hand side of (c) represents a p.g.f. on $\{1, 2, \ldots\}$.

Hence, deduce that every nontrivial infinitely divisible p.g.f. is of the form $\exp[-\lambda(1-\Pi(z))]$ for finite λ (in fact, $p_0 = \mathrm{e}^{-\lambda}$), and p.g.f. $\Pi(z) = \sum_{n=1}^{\infty} \pi_n z^n$ [for details see e.g. Feller (1968, Section XII.2)].

2.2.3 (*Continuation*). Show that an r-variate p.g.f. $P(z_1, \ldots, z_r)$, which is nontrivial in the sense that $P(z_1, \ldots, z_r) \not\equiv 1$ in $\sum_{j=1}^{r} |1 - z_j| > 0$, is infinitely divisible if and only if it is expressible in the form $\exp[-\lambda(1 - \Pi(z_1, \ldots, z_r))]$ for some p.g.f.

$$\Pi(z_1, \ldots, z_r) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_r=0}^{\infty} \pi_{n_1,\ldots,n_r} z^{n_1} \cdots z_r^{n_r}$$

for which $\pi_{0\ldots0} = 0$.

2.2.4 If a point process N has $N((k-1)/n, k/n] \le 1$ for $k = 1, \ldots, n$, then there can be no batches on $(0, 1]$. Use the complete independence property in Assumption 2.2.I(ii) and the fact that $(1 - o(1/n))^n \to 1$ $(n \to \infty)$ to show that a Poisson process satisfying the analytic orderliness property in (2.2.6) has a.s. no batches on the unit interval, and hence on $\mathbb{R}$.

2.3. Characterizations of the Stationary Poisson Process: II. The Form of the Distribution

The discussion to this point has stressed the independence property, and it has been shown that the *Poisson* character of the finite-dimensional distributions is really a consequence of this property. To what extent is it possible to work in the opposite direction and derive the independence property from the Poisson form of the distributions? Observe that for any partition $A_1, \ldots, A_r$ of a Borel set A, the *avoidance probability* $P_0(A)$ of a Poisson process satisfies

$$P_0(A) = \Pr\{N(A) = 0\} = \exp(-\lambda\ell(A)) = \prod_{i=1}^{r} \exp(-\lambda\ell(A_i)) = \prod_{i=1}^{r} P_0(A_i), \tag{2.3.1}$$

so the events $\{N(A_i) = 0\}$ are independent [in (2.3.1), $\ell(\cdot)$ denotes Lebesgue measure]. Rényi (1967) weakened this assumption by requiring (2.3.1) to hold merely on all sets A that are finite unions of finite intervals, and then, adding the requirement that N be orderly, he deduced that N must be Poisson.

In the converse direction, it is *not* enough to take $\mathcal{A}$ to be the class of unions of any *fixed* number of intervals: in particular, it is not enough to know that $N(A)$ has a Poisson distribution for all single intervals $A = [a, b]$, as shown in a series of counterexamples provided by Shepp in Goldman (1967), Moran (1967, 1976a, b), Lee (1968), Szasz (1970), and Oakes (1974); two such counterexamples are described in Exercises 2.3.1 and 4.5.12.

Theorem 2.3.I. *Let N be an orderly point process on $\mathbb{R}$. Then, for N to be a stationary Poisson process it is necessary and sufficient that for all sets A that can be represented as the union of a finite number of finite intervals,*

$$P_0(A) = \mathrm{e}^{-\lambda\ell(A)}. \tag{2.3.2}$$

It is as easy to prove a more general result for a Poisson process that is not necessarily stationary. To this end, define a simple Poisson process in d-dimensional space $\mathbb{R}^d$ as a point process N for which the joint distributions of the counts $N(A_i)$ on bounded disjoint Borel sets A_i satisfy [see equation (2.1.1)]

$$\Pr\{N(A_i) = k_i \ (i = 1, \ldots, r)\} = \prod_{i=1}^{r} \frac{[\mu(A_i)]^{k_i}}{k_i!}\, \mathrm{e}^{-\mu(A_i)} \qquad (r = 1, 2, \ldots)$$

for some nonatomic measure $\mu(\cdot)$ that is bounded on bounded sets. Thus, the $N(A_i)$ are Poisson-distributed and independent, $\mathrm{E}[N(A)] = \mu(A)$, and μ being nonatomic, $\mu(A_n) \to 0$ for any monotonic sequence of bounded sets $A_n \downarrow \emptyset$ or $\{x'\}$ for any singleton set $\{x'\}$ (see Lemma A1.6.II). It is an elementary property of the Poisson distribution that this then implies that $\Pr\{N(A_n) \geq 2\}/\Pr\{N(A_n) \geq 1\} \to 0$ for the same sequence $\{A_n\}$; thus, N has the property of orderliness noted below (2.2.6).

Theorem 2.3.II. *Let μ be a nonatomic measure on $\mathbb{R}^d$, finite on bounded sets, and suppose that the simple point process N is such that for any set A that is a finite union of rectangles,*

$$\Pr\{N(A) = 0\} = \mathrm{e}^{-\mu(A)}. \tag{2.3.3}$$

Then N is a Poisson process with mean $\mu(A)$.

Proof. We use the idea of a dissecting system (see Appendix A1.6). For any set A as in (2.3.3), let the set $\mathcal{T}_n$ of rectangles $\{A_{ni}: i = 1, \dots, r_n\}$ be an element of a dissecting system $\{\mathcal{T}_n\}$ of partitions for A [so, for given n, the union of the A_{ni} is A, A_{ni} and A_{nj} are disjoint for $i \neq j$, and each A_{nj} is the union of some subset A_{n+1,i_s} $(s = 1, \dots, r'_{n,i})$ of $\mathcal{T}_{n+1}$, and for any $x \in A$, there is a sequence $\{A_n(x)\}$, $A_n(x) \in \mathcal{T}_n$ with $\bigcap_n A_n(x) = \{x\}$]. Since μ is nonatomic, $\mu(A_n(x)) \to 0$ as $n \to \infty$.

Given a partition $\mathcal{T}_n$, define the indicator random variables

$$I_{ni} = \begin{cases} 1 & \text{if } N(A_{ni}) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

and set $N_n(A) = \sum_{i=1}^{r_n} I_{ni}$. Because the sets A_{ni} are disjoint, the random variables of the set $\{I_{ni_j}: j = 1, \dots, s\}$ are mutually independent because they are $\{0, 1\}$-valued and

$$\begin{aligned} \Pr\{I_{ni_j} = 0\ (j = 1, \dots, s)\} &= \Pr\{N(A_{ni_j}) = 0\ (j = 1, \dots, s)\} \\ &= \Pr\{N\big(\textstyle\bigcup_{j=1}^{s} A_{ni_j}\big) = 0\} \\ &= \exp\big[-\mu\big(\textstyle\bigcup_{j=1}^{s} A_{ni_j}\big)\big] \\ &= \prod_{j=1}^{s} \exp[-\mu(A_{ni_j})]\,. \end{aligned}$$

Also, $\mathrm{E}(z^{I_{ni}}) = 1 - (1 - z)(1 - \mathrm{e}^{-\mu(A_{ni})})$, so $N_n(A)$ has p.g.f.

$$\mathrm{E}(z^{N_n(A)}) = \prod_i \mathrm{E}(z^{I_{ni}}) = \prod_i \big[1 - (1 - z)(1 - \mathrm{e}^{-\mu(A_{ni})})\big].$$

Because μ is nonatomic, $\sup_i \mu(A_{ni}) \equiv \epsilon_n \to 0$ as $n \to \infty$ (see Lemma A1.6.II), and thus, using $1 - \delta < \mathrm{e}^{-\delta} < 1 - \delta + \delta^2$ for all δ sufficiently small, the p.g.f. of $N_n(A)$ converges to $\exp[-(1 - z)\mu(A)]$ as $n \to \infty$.

Since N is simple, for each realization there exists n_0 such that, for all $n \geq n_0$, each of the $N(A)$ points x_j is in a distinct set A_{nj}, say. Then, for $n \geq n_0$, $N_n(A) = N(A)$. Also, the random variables $N_n(A)$ are monotonically increasing in n and thus have the a.s. limit $N(A)$. It follows that $\mathrm{E}(z^{N(A)}) = \exp[-(1 - z)\mu(A)]$; i.e. $N(A)$ is Poisson-distributed with mean $\mu(A)$ for sets A as in the theorem.

Next, let $\{A_j\}$ be a finite family of disjoint sets that are unions of rectangles. Repeating the argument above shows that the random variables $\{N(A_j)\}$ are mutually independent Poisson random variables with means $\mu(A_j)$.

Now let A be an open set. Then there is a sequence of families $\mathcal{T}_n'$ of rectangles A_{ni}' that are disjoint, as for $\mathcal{T}_n$, with union a subset of A and the unions converging monotonically to A. Analysis similar to that just given shows that $N(A)$ is Poisson distributed with mean $\mu(A)$.

Similarly, for a finite family of disjoint open sets A_j, the random variables $N(A_j)$ are independent.

Finally, we extend these properties to arbitrary disjoint bounded Borel sets A_j by using generating functionals (see Definition 5.5.I) with functions that equal 1 on open sets contained by A_j, vanish on a closed set containing A_j, and are continuous (and between 0 and 1). Such approximating functions yield generating functions that are of Poisson variables and that decompose into products of the separate functions (for each distinct A_j), so the $N(A_j)$ are Poisson-distributed and independent. □

Theorem 2.3.II is due to Rényi (1967); the proof above is adapted from Kingman (1993). This result includes Theorem 2.3.I as a special case, while in the other direction, it is a corollary of a more general result, proved in Chapter 9 and due to Kurtz, that for a simple point process N, it is enough to know the avoidance probabilities $P_0(A)$ on a sufficiently rich class of sets A in order to determine its distribution. In turn, this leads to a characterization of those set functions $P_0(A)$ that can be avoidance functions.

Exercises and Complements to Section 2.3

2.3.1 (see Theorem 2.3.II). Let $N(\cdot)$ be a point process on $\mathbb{R}$ having as its fidi distributions those of a stationary Poisson process of unit rate except for the following eight probabilities relating to the interval $(0, 4]$:

$$p_{0010} = p_{0101} = p_{1011} = p_{1100} = \mathrm{e}^{-4} + \epsilon,$$
$$p_{0100} = p_{1010} = p_{1101} = p_{0011} = \mathrm{e}^{-4} - \epsilon,$$

where $p_{ijkl} = \Pr\{N(0,1] = i,\ N(1,2] = j,\ N(2,3] = k,\ N(3,4] = l\}$, $0 < \epsilon < \mathrm{e}^{-4}$, and, conditional on $N(a, a+1] = 1$ for $a = 0, 1, 2, 3$, that point is uniformly distributed over that unit interval. Verify that $N(I)$ is Poisson-distributed for any interval I, but $N(\cdot)$ is *not* a Poisson process (Lee, 1968).

2.3.2 (a) *Raikov's theorem.* Let Z be a Poisson r.v. expressible as the sum $Z = X + Y$ of independent nondegenerate, nonnegative r.v.s X and Y. Then X and Y are Poisson r.v.s [see e.g. Loève (1963, Section 19.2) or Moran (1968, p. 408)].

(b) Let N be a Poisson process for which $N = N' + N''$ for nontrivial independent point processes N', N''. Show that each of N' and N'' is a Poisson process.

2.3.3 (see Theorem 2.3.III). Suppose a stationary orderly point process satisfies (2.3.1). Since orderliness implies that

$$\begin{aligned}\Pr\{N((0,1]\setminus((k-1)/n,k/n])=0\}-\Pr\{N(0,1]=0\}\\ =\Pr\{N((0,1]\setminus((k-1)/n,k/n])=0,\ N((k-1)/n,k/n]=1\}+o(1/n),\end{aligned}$$

deduce that $\Pr\{N(0,1]=1\}=\lim_{n\to\infty} n(\mathrm{e}^{-\lambda(1-1/n)}-\mathrm{e}^{-\lambda}-o(1/n))=\lambda\mathrm{e}^{-\lambda}$. Extend this argument to show that $\Pr\{N(0,1]=j\}=\lambda^j\mathrm{e}^{-\lambda}/j\,!$

2.3.4 (a) *Random thinning.* Let $N(\cdot)$ be an orderly inhomogeneous Poisson process on $\mathbb{R}^d$ with rate $\lambda(\cdot)$. Form a new process $N'(\cdot)$ by treating each point of a realization $\{x_i\}$ independently of all other points; namely $(*)$ either retain x_i with probability $p(x_i)$ or delete it with probability $1-p(x_i)$, where $p(\cdot)$ is a measurable function with $0\le p(x)\le 1$ for all x. Show that $N'(\cdot)$ is a Poisson process with rate $p(x)\lambda(x)$.

(b) *Random translation.* Repeat part (a) but instead of $(*)$ use $(\dagger)$: translate x_i to x_i+Y_i, where Y_i are independent identically distributed random variables with distribution function $F(\cdot)$. Show that the resulting point process, $N''(\cdot)$ say, is Poisson with rate $\int_{\mathbb{R}^d}\lambda(x-y)\,F(\mathrm{d}y)$.

(c) What conditions on $\lambda(\cdot)$ and $p(\cdot)$ make $N'(\cdot)$ stationary? What conditions make $N''(\cdot)$ stationary?

2.4. The General Poisson Process

We suppose in this section that the point process takes its values in a complete separable metric space (c.s.m.s.) $\mathcal{X}$, thereby anticipating the context of Chapter 9, and without necessarily being stationary, homogeneous, or isotropic. The cases of frequent occurrence are those in which $\mathcal{X}$ is two- or three-dimensional Euclidean space (see the exercises), while the setting includes spatial point processes as in Section 5.3 and Chapter 15, for example.

We suppose throughout that $N(A)$, the number of points in the set A, is defined and finite for every bounded set A in the Borel σ-field $\mathcal{B}(\mathcal{X})\equiv\mathcal{B}_{\mathcal{X}}$ generated by the open spheres of $\mathcal{X}$. We may express this more succinctly by saying that (with probability 1) the trajectories $N(\cdot)$ are *boundedly finite* [recall Assumption 2.2.I(i)]. The Poisson process can then be defined by assuming that there exists a boundedly finite Borel measure $\Lambda(\cdot)$ such that for every finite family of disjoint bounded Borel sets $\{A_i, i=1,\dots,k\}$

$$\Pr\{N(A_i)=n_i, i=1,\dots,k\}=\prod_{i=1}^{k}\frac{[\Lambda(A_i)]^{n_i}}{n_i!}\,\mathrm{e}^{-\Lambda(A_i)}. \tag{2.4.1}$$

The measure $\Lambda(\cdot)$ is called the *parameter measure* of the process. Note that when $\mathcal{X}$ is the real line, (2.4.1) includes as special cases the two examples given in Section 2.1: for the homogeneous process $\Lambda(A)=\lambda\ell(A)$, and for the inhomogeneous process, $\Lambda(A)=\int_A\lambda(x)\,\mathrm{d}x$. Equation (2.4.1) embraces a

nontrivial increase in generality because, in general, the parameter measure may have both a discrete (or atomic) component and a continuous singular component.

In this general setting, we first clarify the role of the discrete component of $\Lambda(\cdot)$. Suppose, in particular, that $\Lambda(\cdot)$ has an atom of mass λ_0 at the point x_0. Since the single-point set $\{x_0\}$ is a Borel set, it follows at once from (2.4.1) that $N\{x_0\} \equiv N(\{x_0\})$ must have a Poisson distribution with parameter λ_0. We say that any point x_0 with the property $\Pr\{N\{x_0\} > 0\} > 0$ is a *fixed atom* of the process. Thus, we conclude that every atom of $\Lambda(\cdot)$ is a fixed atom of $N(\cdot)$. Conversely, if x_0 is a fixed atom of $N(\cdot)$, then $N\{x_0\}$ must have a Poisson distribution with nonzero parameter λ_0, say. From this, it follows that x_0 is an atom of $\Lambda(\cdot)$ with mass λ_0. Hence, the following is true.

Lemma 2.4.I. *The point x_0 is an atom of the parameter measure Λ if and only if it is a fixed atom of the process.*

Note that whether a given point x_0 represents a fixed atom of the process is not discernible from a single realization: any point of the process is an atom of its particular realization. For x_0 to constitute a fixed atom, there must be positive probability of it recurring over a whole family of realizations. Thus, the fixed atoms relate to the probability structure of the process, not to the structure of individual realizations.

In the Poisson case, the fixed atoms are also the key to the question of *orderliness*. The definition given earlier in (2.2.6) is most naturally extended to the present context by requiring

$$\Pr\{N(S_\epsilon(x)) > 1\} = o(\Pr\{N(S_\epsilon(x)) > 0\}) \qquad (\epsilon \to 0), \tag{2.4.2}$$

for each $x \in \mathcal{X}$, where $S_\epsilon(x)$ denotes the open sphere with radius ϵ and centre x. In the case of a Poisson process, $N(S_\epsilon(x))$ has a Poisson distribution, with parameter $\Lambda(S_\epsilon(x)) = \Lambda_\epsilon$, say, so that

$$\begin{aligned}\Pr\{N(S_\epsilon(x)) > 0\} &= 1 - \mathrm{e}^{-\Lambda_\epsilon},\\ \Pr\{N(S_\epsilon(x)) > 1\} &= 1 - \mathrm{e}^{-\Lambda_\epsilon} - \Lambda_\epsilon \mathrm{e}^{-\Lambda_\epsilon}.\end{aligned}$$

Now if x is a fixed atom of Λ, $\Lambda_\epsilon \to \Lambda_0 = \Lambda\{x\} > 0$ as $\epsilon \downarrow 0$, whereas if x is not a fixed atom, $\Lambda_\epsilon \to 0$. In the first case, the ratio $\Pr\{N(S_\epsilon(x)) > 1\}/\Pr\{N(S_\epsilon(x)) > 0\}$ tends to the positive constant $1 - \Lambda_0/(\mathrm{e}^{\Lambda_0} - 1)$, whereas in the second case it tends to zero. Thus, the process is orderly, in the sense of (2.4.2), if and only if $\Lambda(\cdot)$ has no atoms.

Theorem 2.4.II. *The Poisson process defined by* (2.4.1) *is orderly if and only if it has no fixed atoms; equivalently, if and only if the parameter measure has no discrete component.*

When $\mathcal{X}$ is the real line, the distribution function $F_\Lambda(x) \equiv \Lambda(0, x]$ is continuous if and only if Λ has no discrete component, so in this case Λ itself could

be called continuous. One should beware of claiming any such conclusions for more general $\mathcal{X}$, however, for even though $\Lambda(\cdot)$ may have no atoms, it may well have concentrations on lines, surfaces, or other lower-dimensional subsets that may cause an associated distribution function to be discontinuous. In such situations, in contrast to the case of a homogeneous Poisson process, there will be some positive probability of points of the process appearing on such lines, surfaces, and so on.

We turn next to the slightly more difficult problem of extending the characterizations based on the complete independence property stated below.

Assumption 2.4.III. *For each finite family of bounded, disjoint Borel sets* $\{A_i, i = 1, \ldots, k\}$, *the random variables* $N(A_1), \ldots, N(A_k)$ *are mutually independent.*

The most important result is contained in the following lemma.

Lemma 2.4.IV. *Suppose* (i) N *is boundedly finite* a.s. *and has no fixed atoms, and* (ii) N *has the complete independence property of Assumption* 2.4.III. *Then, there exists a boundedly finite nonatomic Borel measure* $\Lambda(\cdot)$ *such that*

$$P_0(A) = \Pr\{N(A) = 0\} = \mathrm{e}^{-\Lambda(A)} \qquad \text{(all bounded Borel sets } A\text{)}.$$

Proof. Set $Q(A) = -\log P_0(A)$, observing immediately that $Q(A) \geq 0$ and that by (ii) it is finitely additive. Countable additivity is equivalent to having $Q(A_n) \to 0$ for any decreasing sequence $\{A_n\}$ of bounded Borel sets for which $Q(A_n) < \infty$ and $A_n \downarrow \emptyset$. For $A_n \downarrow \emptyset$, we must have $N(A_n) \to 0$ a.s., and thus $\mathrm{e}^{-Q(A_n)} = P_0(A_n) = \Pr\{N(A_n) = 0\} \to 1$, establishing $Q(A_n) \to 0$ as required. To show that $Q(\cdot)$ is nonatomic, observe that, by (i),

$$0 = \Pr\{N\{x\} > 0\} = 1 - \mathrm{e}^{-Q(\{x\})},$$

so that $Q(\{x\}) = 0$ for every x.

It remains to show that $Q(\cdot)$ is boundedly finite, which is equivalent to $P_0(A) > 0$ for any bounded Borel set A. Suppose the contrary for some set A, which without loss of generality we may assume to be closed, for if not, $0 \leq P_0(\bar{A}) \leq P_0(A) = 0$, whence $P_0(\bar{A}) = 0$. Since $\mathcal{X}$ is separable, A can be covered by a countable number of disjoint Borel sets A_n, each with diameter less than 1, so $A = \bigcup_{n=1}^{\infty} A_n$. Let $p_n = \Pr\{N(A_n) > 0\}$, so that $N(A) = 0$ only if $N(A_n) = 0$ for all n, and thus $0 = P_0(A) = \prod_{n=1}^{\infty}(1 - p_n)$. This infinite product vanishes only if $p_n = 1$ for some n, or else $\sum_{n=1}^{\infty} p_n$ diverges. In the latter event, the Borel–Cantelli lemma implies that a.s. infinitely many $N(A_n)$ are nonzero, and hence $N(A) = \infty$ a.s., contradicting the assumption that $N(\cdot)$ is boundedly finite. Consequently, we must have $p_n = 1$ for some set A_n, $A_{(1)}$ say, and $A_{(1)}$ has diameter less than 1 and as with A may be assumed to be closed. By repeating the argument, we can find a closed set $A_{(2)}$ with diameter less than 2^{-1} such that $P_0(A_{(2)}) = 0$. Proceeding by induction, a

sequence $\{A_{(n)}\}$ of nested closed sets is constructed with diameters $\to 0$, and $P_0(A_{(n)}) = 0$ (all n). Choose $x_n \in A_{(n)}$, so that $\{x_n\}$ is a Cauchy sequence, $x_n \to x_0$ say, and, each $A_{(n)}$ being closed, $x_0 \in A_{(n)}$, and therefore $A_n \downarrow \{x_0\}$. Then $N(A_{(n)}) \downarrow N(\{x_0\})$, and by monotone convergence, $P_0(\{x_0\}) = \lim_{n\to\infty} P_0(A_{(n)}) = 0$. Equivalently, $\Pr\{N\{x_0\} > 0\} = 1$, so that x_0 is a fixed atom of the process, contradicting (i). □

Now suppose that the process is orderly in addition to satisfying the conditions of Lemma 2.4.IV. Then, it follows from Theorem 2.3.II that we have a Poisson process without fixed atoms. Thus, the following theorem, due to Prekopa (1957a, b), is true.

Theorem 2.4.V. *Let $N(\cdot)$ be a.s. boundedly finite and without fixed atoms. Then $N(\cdot)$ is a Poisson process if and only if*

(i) *it is orderly, and*

(ii) *it has the complete independence property of Assumption* 2.4.III.

To extend this result to the nonorderly case, consider for fixed real z in $0 \le z \le 1$ the set function

$$Q_z(A) \equiv -\log \mathrm{E}(z^{N(A)}) \equiv -\log P_z(A)$$

defined over the Borel sets A. It follows immediately that

$$0 \le Q_z(A) < Q(A),$$

and using also the argument of Lemma 2.4.VI, it follows that $Q_z(\cdot)$ is a measure, absolutely continuous with respect to $Q(\cdot)$. Consequently, there exists a density, $q_z(x)$ say, such that

$$Q_z(A) = \int_A q_z(x)\, Q(\mathrm{d}x) \tag{2.4.3}$$

and for Q-almost-all x

$$q_z(x) = \lim_{\epsilon\downarrow 0} \frac{Q_z(S_\epsilon(x))}{Q(S_\epsilon(x))}\,,$$

where $S_\epsilon(x)$ is as in (2.4.2); see also e.g. Lemma A1.6.III for this property of Radon–Nikodym derivatives. If we continue to assume that the process has no fixed atoms, $Q(S_\epsilon(x))$ and hence also $Q_z(S_\epsilon(x))$ both $\to 0$ as $\epsilon \to 0$, for then $S_\epsilon(x) \to \{x\}$. We can then imitate the argument leading to Theorem 2.2.II and write for Q-almost-all x

$$\Pi_z(x) = 1 - q_z(x) = \lim_{\epsilon\downarrow 0} \frac{P_z(S_\epsilon(x)) - P_0(S_\epsilon(x))}{1 - P_0(S_\epsilon(x))}\,. \tag{2.4.4}$$

Now, for fixed A, $Q_z(A)$ is monotonically decreasing in z for $0 \le z \le 1$, so by taking a countably dense set of z values in $[0,1]$, (2.4.4) holds for such z except possibly on a Q-null set formed by the union of the Q-null sets where it may fail for the separate values of z.

For each ϵ, (2.4.4) is the p.g.f. of the conditional distribution

$$\Pr\{N(S_\epsilon(x)) = k \mid N(S_\epsilon(x)) > 0\}.$$

Now a sequence of p.g.f.s converging on a countably dense set of z values in $[0,1)$ converges for all $0 \le z < 1$, with the limit being a p.g.f. of a possibly dishonest distribution. In the present case, the limit is in fact Q-a.e. honest because by monotone convergence and (2.4.3),

$$0 = \log P_1(A) = \lim_{z\uparrow 1} Q_z(A) = \int_A \Big(\lim_{z\to 1} q_z(x) \Big)\, Q(\mathrm{d}x),$$

implying that $\lim_{z\to 1} q_z(x) = 0 \;\; Q$-a.e.

Consequently, except for a Q-null set, (2.4.4) holds for all $0 \le z \le 1$, and for the limit $q_z(x)$, $1 - q_z(x)$ is the p.g.f. of a proper distribution, $\{\pi_k(x)\}$ say, for which

$$\pi_0(x) = 0, \qquad \Pi_z(x) = \sum_{k=1}^{\infty} \pi_k(x) z^k,$$

and

$$P_z(A) = \exp\bigg(- \int_A [1 - \Pi_z(x)]\, Q(\mathrm{d}x) \bigg). \tag{2.4.5}$$

There is the alternative form for (2.4.5),

$$P_z(A) = \exp\big(- Q(A)[1 - \Pi_z(A)] \big),$$

in which there appears the p.g.f. $\Pi_z(A)$ of the 'averaged' probabilities

$$\pi_k(A) = \frac{1}{Q(A)} \int_A \pi_k(x)\, Q(\mathrm{d}x).$$

Thus, the distributions in this process still have the compound Poisson form.

Finally, suppose we reinstate the fixed atoms of the process. Note that these are also atoms of $Q(\cdot)$ and can therefore be at most countable in number, and also that the number of points of the process at each fixed atom must be a discrete random variable independent of the rest of the process. We thus arrive at the following structure theorem for the general point process satisfying the complete independence property.

Theorem 2.4.VI. *Let $N(\cdot)$ be a point process that has the complete independence property of Assumption* 2.4.III. *Then $N(\cdot)$ can be written in the form of a superposition $N = N_1 + N_2$, where N_1 and N_2 are independent and*

(i) *N_1 consists of a finite or countable family of fixed atoms, $\{x_1, x_2, \ldots\}$, where for each i, $N_1\{x_i\}$ has a proper, discrete distribution and is independent of the rest of the process; and*

(ii) *N_2 is a process without fixed atoms, which can be represented in the compound Poisson form* (2.4.5), *where $Q(\cdot)$ is a fixed, boundedly finite, nonatomic measure, and for Q-almost-all x, $\Pi_z(x)$ is the p.g.f. of a proper discrete distribution, satisfying $\Pi_0(x) = 0$.*

We remark that, analogously to the situation described by Theorem 2.2.II, the realizations of N_2 consist a.s. of random batches of points, where the number of batches is governed by a Poisson process with parameter measure $Q(\cdot)$ and, conditional on a batch occurring at x, its probability distribution is given by $\{\pi_k(x)\}$. These sample-path results can be established directly for this special case, but we prefer to treat them as special cases of the theorems established in Chapter 3.

Exercises and Complements to Section 2.4

2.4.1 Let N_1, N_2 be independent Poisson processes with parameter measures Λ_1, Λ_2. Show that $N_1 + N_2$ is a Poisson process with parameter measure $\Lambda_1 + \Lambda_2$.

2.4.2 *Poisson process on the surface of a sphere.* There is an area-preserving map of the surface of a sphere of radius r onto the curved surface of a cylinder of radius r and height $2r$. Conclude that a homogeneous Poisson process on the surface of such a sphere can be represented as a Poisson process on a rectangle with side-lengths $2r$ and $2\pi r$. How may a homogeneous Poisson process on the surface of an oblate or prolate elliptical spheroid be constructed?
[*Hint*: An oblate spheroid is the solid of revolution obtained by rotating an ellipse with major and minor axes of lengths $2a$ and $2b$, respectively, about its minor axis, so it has the same surface area as the curved surface of a cylinder of radius a and height $2\int_0^{\pi/2} \cos\theta \sqrt{a^2 \sin^2\theta + b^2 \cos^2\theta}\, \mathrm{d}\theta$. For a prolate spheroid, use a cylinder of radius b and height $2\int_0^{\pi/2} \sin\theta \sqrt{a^2 \sin^2\theta + b^2 \cos^2\theta}\, \mathrm{d}\theta$.]

2.4.3 *Poisson process on a lattice.* A homogeneous Poisson process with density λ on a given (countably infinite) lattice of points, $\{z_i\}$ say, is a sequence of i.i.d. Poisson r.v.s, $\{N_i\}$ say, with common mean λ.

A *homogeneous binary process* on such a lattice is a sequence, $\{Y_i\}$ say, of i.i.d. $\{0,1\}$-valued r.v.s $\{Y_i\}$ for which $\Pr\{Y_i = 1\} = p$ for some $p \in (0,1)$. It is only approximately Poisson, and then only for small p.

2.4.4 Define a homogeneous Poisson process on a cylinder of unit radius as a Poisson process of points $\{(x_i, \theta_i)\}$ on the doubly infinite strip $\mathbb{R} \times (0, 2\pi]$ at rate $\lambda\, \mathrm{d}x\, \mathrm{d}\theta$. Such a point process can also be interpreted as a Poisson process of directed lines in the plane since any such line is specified by its orientation relative to a given direction and its distance from the origin (negative if the origin is to the left of the line rather than the right).

(a) In this line-process interpretation, check that the largest circle that can be drawn around a randomly chosen point in the plane without intersecting a line has radius R with distribution $\Pr\{R > y\} = \Pr\{\text{strip of width } 2y \text{ has no point } (x_i, \theta_i)\} = \exp(-\lambda\, 2\pi y)$.

(b) Show that the expected number of intersections lying within the circle $S_R(0)$ between the line $(x, 0)$ and lines of the process, where $0 < x < R$, equals $4\int_x^R \operatorname{arsin}(y/R)\, 2\lambda\, \mathrm{d}y$. Deduce that the expected number of intersections between any two lines of the process and lying in a circle of radius R equals

$$2\pi \int_0^R 2\lambda\, \mathrm{d}x \int_x^R 8\lambda \operatorname{arsin}(y/R)\, \mathrm{d}y = (2\lambda\, \pi R)^2.$$

Observe that such a point process (from line intersections) cannot be Poisson because with probability 1, given any two points, there are infinitely many other points collinear with the two given points.

2.4.5 *Poisson process in Hilbert space.*

(i) Find an example of a Hilbert-space-valued random variable that does not have its distribution concentrated on a finite-dimensional subspace. [*Hint*: Consider a series of the form $Y = \sum a_k U_k e_k$, where the a_k form a scalar series, the U_k are i.i.d., and e_k is the unit vector in the kth dimension. Other examples follow from the Hilbert-space Gaussian measures discussed in Chapter 9.] By combining copies of this probability measure suitably, build up examples of σ-finite measures.

(ii) Using the measures above, construct examples of well-defined Poisson processes on a Hilbert space. Discuss the nature of the realizations in increasing sequences of spheres or cubes.

(iii) Show that if a σ-finite measure is invariant under Hilbert-space translations, then it cannot be boundedly finite. Hence, show that no Poisson process can exist that is invariant under the full set of Hilbert-space translations.

CHAPTER 3

Simple Results for Stationary Point Processes on the Line

The object of this chapter is to give an account of some of the distinctive aspects of stationary point processes on the line without falling back on the measure-theoretic foundations that are given in Chapter 9. Some aspects that are intuitively reasonable and that can in fact be given a rigorous basis are taken at face value in order that the basic ideas may be exposed without the burden of too much mathematical detail. Thus, the results presented in this chapter may be regarded as being made logically complete when combined with the results of Chapter 9.

Ideas introduced here concerning second-order properties are treated at greater length in Chapters 8 and 12, and Palm theory in Chapter 13.

3.1. Specification of a Point Process on the Line

A point process on the line may be taken as modelling the occurrences of some phenomenon at the time epochs $\{t_i\}$ with i in some suitable index set. For such a process, there are four equivalent descriptions of the sample paths:

(i) counting measures;
(ii) nondecreasing integer-valued step functions;
(iii) sequences of points; and
(iv) sequences of intervals.

In describing a point process as a counting measure, it does not matter that the process is on the real line. However, for the three other methods of describing the process, the order properties of the reals are used in an essential way. While the methods of description may be capable of extension into higher dimensions, they become less natural and, in the case of (iv), decidedly artificial.

In Chapters 1 and 2, we mostly used the intuitive notion of a point process as a counting measure. To make this notion precise, take any subset A of the real line and let $N(A)$ denote the number of occurrences of the process in the set A; i.e.

$$\begin{aligned} N(A) &= \text{number of indices } i \text{ for which } t_i \text{ lies in } A \\ &= \#\{i\colon t_i \in A\}. \end{aligned} \tag{3.1.1}$$

When A is expressed as the union of the disjoint sets $A_1, \ldots, A_r$, say, that is,

$$A = \bigcup_{i=1}^{r} A_i \quad \text{where } A_i \cap A_j = \emptyset \text{ for } i \neq j,$$

it is a consequence of (3.1.1) that

$$N\left(\bigcup_{i=1}^{r} A_i\right) = \sum_{i=1}^{r} N(A_i) \quad \text{for mutually disjoint } A_1, \ldots, A_r. \tag{3.1.2}$$

It also follows from (3.1.1) that

$$N(A) \text{ is nonnegative integer-(possibly } \infty\text{-)valued.} \tag{3.1.3}$$

In order that we may operate conveniently on $N(A)$ for different sets A—in particular, in order that the probability of events specified in terms of $N(A)$ may be well defined—we must impose a restriction on the sets A that we are prepared to consider. Since we want to include intervals and unions thereof, the usual constraint is that

$$N(A) \text{ is defined for all Borel subsets } A \text{ of the real line.} \tag{3.1.4}$$

Finally, in order to exclude the possibility of 'too many' points occurring 'too close' together, we insist that, for the point processes we consider,

$$N(A) \text{ is finite for bounded sets } A. \tag{3.1.5}$$

The assumptions in (3.1.2–5) with (3.1.2) extended to allow $r = \infty$ are precisely those that make $N(\cdot)$ a counting measure on the σ-field $\mathcal{B}_{\mathbb{R}}$ of all Borel subsets of the real line $\mathbb{R}$. The constraint in (3.1.3) that $N(\cdot)$ be integer-valued distinguishes it from other more general nonnegative measures as a *counting* measure.

To be consistent with $N(\cdot)$ being a set function, we ought to write, for example, $N((a, b])$ when A is the half-open interval $(a, b]$; our preference for the less cumbersome abbreviation $N(a, b]$ should lead to no confusion.

We have already used in Chapters 1 and 2 the further contraction

$$N(t) = N(0, t] = N((0, t]) \qquad (0 < t \leq \infty); \tag{3.1.6}$$

the difference in argument should suffice to distinguish the real function $N(t)$ $(t > 0)$ from the set function $N(A)$. This function $N(t)$ is nondecreasing, right-continuous, and integer-valued, and hence a step function. For point processes on the positive half-line, knowledge of $N(t)$ for all $t \geq 0$ suffices to determine $N(A)$ for Borel sets $A \subset (0, \infty)$ in precisely the same manner as a distribution function determines a probability measure on Borel sets. When the point process is defined on the whole line, we extend the definition (3.1.6) to

$$N(t) = \begin{cases} N((0,t]) & (t > 0), \\ 0 & (t = 0), \\ -N((t,0]) & (t < 0). \end{cases} \tag{3.1.7}$$

In this way, $N(t)$ retains the properties of being a right-continuous integer-valued function on the whole line. Moreover, $N(t)$ determines $N(A)$ for all Borel sets A and hence describes the point process via a step function. Thus, instead of starting with $N(A)$ (all $A \in \mathcal{B}$), we could just as well have specified the sample path as a right-continuous function $N(t)$ $(-\infty < t < \infty)$ that is nonnegative and integer-valued for $t > 0$, nonpositive and integer-valued for $t < 0$, and has $N(0) = 0$.

The simplest case of the third method listed above occurs where the process is defined on the half-line $t > 0$. Setting

$$t_i = \inf\{t > 0 \colon N(t) \geq i\} \qquad (i = 1, 2, \ldots), \tag{3.1.8}$$

it follows that for $i = 1, 2, \ldots$, we have the seemingly obvious but most important relation

$$t_i \leq t \text{ if and only if } N(t) \geq i. \tag{3.1.9}$$

This relation makes it clear that specifying the sequence of points $\{t_i\}$ is equivalent to specifying the function $N(t)$ in the case where $N(-\infty, 0] = 0$. It should be noted that the set of points $\{t_i\}$ in (3.1.8) is in increasing order; such a restriction is not necessarily implied in talking of a set of time epochs $\{t_i\}$ as at the beginning of the present section.

If the point process has points on the whole line and not just the positive axis, the simplest extension consistent with (3.1.8) is obtained by defining

$$\begin{aligned} t_i &= \inf\{t \colon N(t) \geq i\} \\ &= \begin{cases} \inf\{t > 0 \colon N(0,t] \geq i\} & (i = 1, 2, \ldots), \\ -\inf\{t > 0 \colon N(-t, 0] \geq -i + 1\} & (i = 0, -1, \ldots). \end{cases} \end{aligned} \tag{3.1.10}$$

Such a doubly infinite sequence of points has the properties that

$$t_i \leq t_{i+1} \text{ (all } i) \qquad \text{and} \qquad t_0 \leq 0 < t_1. \tag{3.1.11}$$

Finally, by setting

$$\tau_i = t_i - t_{i-1} \qquad \text{with } \{t_i\} \text{ as in (3.1.10)} \tag{3.1.12}$$

[or else, in the case of only a half-line as in (3.1.8), with the added conventions that $t_0 = 0$ and τ_i is defined only for $i = 1, 2, \dots$], the process is fully described by the sequence of *intervals* $\{\tau_i\}$ and one of the points $\{t_i\}$, usually t_0. Observe that $\tau_i \geq 0$ and that if $N(t) \to \infty$ as $t \to \infty$, then $\sum_{i=1}^{n} \tau_i \to \infty$ as $n \to \infty$, while if $N(t) \not\to \infty$ as $t \to \infty$, then τ_i is not defined for $i > \lim_{t\to\infty} N(t)$.

We now make the intuitively plausible assumption that there exists a probability space on which the functions $N(A)$, $N(t)$, t_i, τ_i are well-defined random variables and furthermore that we can impose various constraints on these random variables in a manner consistent with that assumption. The question of the existence of such a probability space is discussed in Chapter 9.

Exercises and Complements to Section 3.1

3.1.1 Suppose that the r.v.s $\{t_i\}$ in (3.1.8) are such that $\Pr\{t_{i+1} > t_i\} = 1$, and define

$$G_i(x) = \Pr\{t_i \leq x\}.$$

(a) Show that $\lim_{x\to 0} G_i(x) = 0$ for all integers $i > 0$.

(b) Show that the assumption in (3.1.5) of $N(\cdot)$ being boundedly finite implies that, for all real $x > 0$,

$$\lim_{i\to\infty} G_i(x) = 0.$$

3.1.2 (*Continuation*). Show that for $x > 0$, $M(x) \equiv \mathrm{E}N(x) = \sum_{i=1}^{\infty} G_i(x)$ and, more generally, that

$$\mathrm{E}\big([N(x)]^r\big) = \sum_{i=1}^{\infty} \big(i^r - (i-1)^r\big) G_i(x) = \sum_{i=1}^{\infty} i^r \big(G_i(x) - G_{i+1}(x)\big)$$

in the sense that either both sides are infinite or, if one is finite, so is the other and the two sides are equal.

3.1.3 (*Continuation*). Show that for $|z| \leq 1$ and $x > 0$,

$$P(x; z) \equiv \mathrm{E}z^{N(x)} = 1 + (z-1) \sum_{i=0}^{\infty} G_{i+1}(x) z^i.$$

3.2. Stationarity: Definitions

The notion of stationarity of a point process at first sight appears to be a simple matter: at the very least, it means that the distribution of the number of points lying in an interval depends on its length but not its location; that is,

$$p_k(x) \equiv \Pr\{N(t, t+x] = k\} \qquad (x > 0,\ k = 0, 1, \dots)$$

depends on the length x but not the location t. Lawrance (1970) called this property *simple stationarity*, while we follow Chung (1972) in calling it *crude stationarity*. It is in fact weaker than the full force of the definition below (see Exercise 3.2.1).

Definition 3.2.I. *A point process is* stationary *when for every* $r = 1, 2, \ldots$ *and all bounded Borel subsets* $A_1, \ldots, A_r$ *of the real line, the joint distribution of*

$$\{N(A_1 + t), \ldots, N(A_r + t)\}$$

does not depend on t $(-\infty < t < \infty)$.

In the case where the point process is defined only on the positive half-line, the sets A_i must be Borel subsets of $(0, \infty)$ and we require $t > 0$.

There is also the intuitive feeling that the intervals $\{\tau_i\}$ should be stationary, and accordingly we introduce the following definition.

Definition 3.2.II. *A point process is* interval stationary *when for every* $r = 1, 2, \ldots$ *and all integers* $i_i, \ldots, i_r$, *the joint distribution of* $\{\tau_{i_1+k}, \ldots, \tau_{i_r+k}\}$ *does not depend on* k $(k = 0, \pm 1, \ldots)$.

Note that this definition makes no reference to the point t_0 required to complete the specification of a sample path as below (3.1.12). It is most natural to take $t_0 = 0$ [see (3.1.11)]. Such processes may then be regarded as a generalization of renewal processes in that the intervals between occurrences, instead of being mutually independent and identically distributed, constitute merely a stationary sequence.

The relation that exists between the probability distributions for interval stationarity on the one hand and stationarity on the other is taken up in Section 3.4 and elsewhere, notably Chapter 13, under its usual heading of Palm–Khinchin theory.

Some authors speak of *arbitrary times* and *arbitrary points* in connection with point processes. A probability distribution with respect to an arbitrary time epoch of a stationary point process is one that is stationary as under Definition 3.2.I; a probability distribution with respect to an arbitrary point of a point process is one determined by the interval stationary distributions as under Definition 3.2.II.

The importance of maintaining a distinction between interval stationarity and ordinary stationarity is underlined by the waiting-time paradox. If in some town buses run exactly on schedule every Δ minutes and a stranger arrives at a random time to wait for the next bus, then his expected waiting time $\mathrm{E}W$ is $\frac{1}{2}\Delta$ minutes. If, on the other hand, buses run haphazardly according to a Poisson process with an average time Δ between buses, then the expected waiting time of the same stranger is Δ. The core of the so-called paradox lies in the use of Δ as an average interval length from the arrival of one bus to the next, and the waiting time $\mathrm{E}W$ being half the mean interval between bus arrivals when the probabilities of different intervals being chosen are proportional to their lengths. In renewal theory, the resolution of the paradox is known as *length-biased sampling* [see Feller (1966, Section I.4), Exercise 1.2.5 above, and (3.4.17) below].

Exercises and Complements to Section 3.2

3.2.1 (a) Construct an example of a crudely stationary point process that is not stationary (for one example, see Exercise 2.3.1).

(b) Let $N(\cdot)$ be crudely stationary. Is it necessarily true that

$$\Pr\{N(\{t\}) \geq 2 \text{ for some } t \text{ in } (-1, 0]\} = \Pr\{N(\{t\}) \geq 2 \text{ for some } t \text{ in } (0, 1]\} \text{ ?}$$

[See the proof of Proposition 3.3.VI, where equality is shown to hold when the probabilities equal zero.]

3.3. Mean Density, Intensity, and Batch-Size Distribution

A natural way of measuring the average density of points of a point process is via its mean, or in the case of a stationary point process, its mean density, which we define as

$$m = \mathrm{E}(N(0, 1]). \tag{3.3.1}$$

Defining the function

$$M(x) = \mathrm{E}(N(0, x]), \tag{3.3.2}$$

it is a consequence of the additivity properties of $N(\cdot)$ as in (3.1.2) and of expectations of sums, and of the crude stationarity property in (3.2.1), that for $x, y \geq 0$,

$$\begin{aligned} M(x+y) &= \mathrm{E}\big(N(0, x+y]\big) = \mathrm{E}\big(N(0, x] + N(x, x+y]\big) \\ &= \mathrm{E}\big(N(0, x]\big) + \mathrm{E}\big(N(x, x+y]\big) = \mathrm{E}\big(N(0, x]\big) + \mathrm{E}\big(N(0, y]\big) \\ &= M(x) + M(y). \end{aligned}$$

In other words, $M(\cdot)$ is a nonnegative function satisfying Cauchy's functional equation

$$M(x+y) = M(x) + M(y) \qquad (0 \leq x, y < \infty).$$

Consequently, by Lemma 3.6.III,

$$M(x) = M(1)x = mx \qquad (0 \leq x < \infty), \tag{3.3.3}$$

irrespective of whether $M(x)$ is finite or infinite for finite $x > 0$.

There is another natural way of measuring the rate of occurrence of points of a stationary point process, due originally to Khinchin (1955).

Proposition 3.3.I (Khinchin's Existence Theorem)**.** *For a stationary (or even crudely stationary) point process, the limit*

$$\lambda = \lim_{h \downarrow 0} \frac{\Pr\{N(0, h] > 0\}}{h} \tag{3.3.4}$$

exists, though it may be infinite.

PROOF. Introduce the function

$$\phi(x) = \Pr\{N(0, x] > 0\}. \tag{3.3.5}$$

Then $\phi(x) \downarrow 0$ as $x \downarrow 0$, and $\phi(\cdot)$ is subadditive on $(0, \infty)$ because for $x, y > 0$,

$$\begin{aligned}\phi(x+y) &= \Pr\{N(0, x+y] > 0\} \\ &= \Pr\{N(0, x] > 0\} + \Pr\{N(0, x] = 0, N(x, x+y] > 0\} \\ &\le \Pr\{N(0, x] > 0\} + \Pr\{N(x, x+y] > 0\} \\ &= \phi(x) + \phi(y).\end{aligned}$$

The assertion of the proposition now follows from the subadditive function Lemma 3.6.I. □

The parameter λ is called the *intensity* of the point process, for when it is finite, it makes sense to rewrite (3.3.4) as

$$\begin{aligned}\Pr\{N(x, x+h] > 0\} &= \Pr\{\text{there is at least one point in } (x, x+h]\} \\ &= \lambda h + o(h) \qquad (h \downarrow 0).\end{aligned} \tag{3.3.6}$$

Examples of a point process with $\lambda = \infty$ are given in Exercises 3.3.2–3.

These two measures of the 'rate' of a stationary point process coincide when the point process has the following property.

Definition 3.3.II. *A point process is* simple *when*

$$\Pr\{N(\{t\}) = 0 \text{ or } 1 \text{ for all } t\} = 1. \tag{3.3.7}$$

Daley (1974) called this sample-path property *almost sure orderliness* to contrast it with the following analytic property due to Khinchin (1955).

Definition 3.3.III. *A crudely stationary point process is* orderly *when*

$$\Pr\{N(0, h] \ge 2\} = o(h) \qquad (h \downarrow 0). \tag{3.3.8}$$

Notice that stationarity plays no role in the definition of a simple point process, nor does it matter whether the point process is defined on the real line or even a Euclidean space. While orderliness can be defined for point processes that either are nonstationary or are on some space different from the real line, the defining equation (3.3.8) must then be suitably amended [see Exercise 3.3.1, Chapter 9, and Daley (1974) for further discussion and references].

It is a consequence of Korolyuk's theorem and Dobrushin's lemma, given below, that for stationary point processes with finite intensity, Definitions 3.3.II and 3.3.III coincide.

Proposition 3.3.IV (Korolyuk's Theorem)**.** *For a crudely stationary simple point process,*

$$\lambda = m, \quad \textit{finite or infinite.}$$

Remark. In Khinchin's (1955, Section 11) original statement of this proposition, the point process was assumed to be orderly rather than simple. In view of the possible generalizations of the result to nonstationary point processes and to processes on spaces other than the real line where any definition of orderliness may be more cumbersome, it seems sensible to follow Leadbetter (1972) in connecting the present result with Korolyuk's name.

PROOF. We use a sequence of nested intervals that in fact constitute a dissecting system (see Section A1.6 and the proof of Theorem 2.3.II). For any positive integer n and $i = 1, \ldots, n$, define indicator random variables

$$I_{ni} = \begin{cases} 1 \\ 0 \end{cases} \quad \text{according as} \quad N\Big(\frac{i-1}{n}, \frac{i}{n}\Big] \begin{cases} > 0, \\ = 0. \end{cases} \tag{3.3.10}$$

Then, as $n \to \infty$ through the integers 2^p, $p = 1, 2, \ldots$,

$$\sum_{i=1}^{n} I_{ni} \uparrow N(0, 1] \tag{3.3.11}$$

for those realizations $N(\cdot)$ for which $N(0, 1] < \infty$ and $N(\{t\}) = 0$ or 1 for all $0 < t \le 1$; that is, in view of (3.1.5) and (3.3.7), (3.3.11) holds a.s. Then

$$\begin{aligned} m = \mathrm{E}\big(N(0,1]\big) &= \mathrm{E}\left(\lim_{n\to\infty} \sum_{i=1}^{n} I_{ni} \right) \\ &= \lim_{n\to\infty} \mathrm{E}\left(\sum_{i=1}^{n} I_{ni} \right) \qquad \text{by Lebesgue's monotone convergence theorem,} \\ &= \lim_{n\to\infty} n\phi(n^{-1}) \qquad \text{by (3.3.5), (3.3.10), and crude stationarity,} \\ &= \lambda \qquad \text{by Khinchin's existence theorem.} \end{aligned}$$

□

Proposition 3.3.V (Dobrushin's Lemma)**.** *A crudely stationary simple point process of finite intensity is orderly.*

PROOF. For any positive integer n, $\mathrm{E}(N(0,1]) = n\,\mathrm{E}(N(0, n^{-1}])$ by crude stationarity, so

$$\begin{aligned} m = \mathrm{E}(N(0,1]) &= n \sum_{j=1}^{\infty} \Pr\{N(0, n^{-1}] \ge j\} \\ &\ge n\phi(n^{-1}) + n \Pr\{N(0, n^{-1}] \ge 2\}. \end{aligned} \tag{3.3.12}$$

Being crudely stationary, Khinchin's existence theorem applies, so $n\phi(n^{-1}) \to \lambda$ as $n \to \infty$, and being simple also, Korolyuk's theorem applies, so $\lambda = m$. Combining these facts with (3.3.12), $n \Pr\{N(0, n^{-1}] \ge 2\} \to 0$ as $n \to \infty$, which by (3.3.8) is the same as orderliness. □

Dobrushin's lemma is a partial converse of the following result in which there is no finiteness restriction on the intensity.

Proposition 3.3.VI. *A crudely stationary orderly point process is simple.*

PROOF. Simpleness is equivalent to

$$0 = \sum_{r=-\infty}^{\infty} \Pr\big\{N(\{t\}) \geq 2 \text{ for some } t \text{ in } (r, r+1]\big\},$$

which in turn is equivalent to

$$0 = \Pr\big\{N(\{t\}) \geq 2 \text{ for some } t \text{ in } (r, r+1]\big\} \qquad (r = 0, \pm 1, \ldots). \quad (3.3.13)$$

For every positive integer n,

$$\begin{aligned}\Pr\{N(\{t\}) \geq 2 \text{ for some } t \text{ in } (0,1]\} &\leq \sum_{i=1}^{n} \Pr\Big\{N\Big(\frac{i-1}{n}, \frac{i}{n}\Big] \geq 2\Big\} \\ &= n \Pr\{N(0, n^{-1}] \geq 2\} \qquad \text{by crude stationarity,} \\ &\to 0 \quad (n \to \infty) \qquad \text{when } N(\cdot) \text{ is orderly,}\end{aligned}$$

so (3.3.13) holds for $r = 0$ and, by trite changes, for all r. □

In the results just given, a prominent role is played by orderliness, which stems from the notion that the points $\{t_i\}$ can indeed be ordered; that is, in the notation of (3.1.10), we have $t_i < t_{i+1}$ for all i. Without orderliness, we are led to the idea of batches of points: we proceed as follows.

Proposition 3.3.VII. *For a crudely stationary point process, the limits*

$$\lambda_k = \lim_{h \downarrow 0} \frac{\Pr\{0 < N(0,h] \leq k\}}{h} \quad (3.3.14)$$

exist for $k = 1, 2, \ldots,$ *and*

$$\lambda_k \uparrow \lambda \qquad (k \to \infty), \quad \textit{finite or infinite}; \quad (3.3.15)$$

when λ *is finite,*

$$\pi_k \equiv \frac{\lambda_k - \lambda_{k-1}}{\lambda} = \lim_{h \downarrow 0} \Pr\{N(0,h] = k \mid N(0,h] > 0\} \quad (3.3.16)$$

is a probability distribution on $k = 1, 2, \ldots$.

PROOF. Define, by analogy with (3.3.5),

$$\phi_k(x) = \Pr\{0 < N(0,x] \leq k\} \qquad (x > 0,\ k = 1, 2, \ldots). \quad (3.3.17)$$

Then, like $\phi(\cdot)$, $\phi_k(x) \to 0$ for $x \downarrow 0$ and it is subadditive on $(0, \infty)$ because, for $x, y > 0$,

$$\begin{aligned}\phi_k(x+y) &= \Pr\{0 < N(0,x] \leq k, N(x, x+y] = 0\} \\ &\quad + \Pr\{N(0,x] \leq k - N(x, x+y],\ 0 < N(x, x+y] \leq k\} \\ &\leq \Pr\{0 < N(0,x] \leq k\} + \Pr\{0 < N(x, x+y] \leq k\} \\ &= \phi_k(x) + \phi_k(y),\end{aligned}$$

invoking crude stationarity at the last step. Thus, (3.3.14) follows from the subadditive function lemma, which is also invoked in writing

$$\lambda = \sup_{h>0} \sup_{k>0} \frac{\phi_k(h)}{h} = \sup_{k>0} \sup_{h>0} \frac{\phi_k(h)}{h} = \sup_{k>0} \lambda_k.$$

The monotonicity of λ_k in k is obvious from (3.3.14), so (3.3.15) is now proved. Equation (3.3.16) follows from (3.3.14), (3.3.15), and (3.3.17). □

The limit of the conditional probability in (3.3.16) can be rewritten in the form

$$\Pr\{N(0,h] = k\} = \lambda \pi_k h + o(h) \qquad (h \downarrow 0,\ k = 1, 2, \ldots). \tag{3.3.18}$$

This equation and (3.3.16) suggest that the points $\{t_i\}$ of sample paths occur in batches of size $k = 1, 2, \ldots$ with respective intensities $\lambda\pi_k$. To make this idea precise, recall that for bounded Borel sets A we have assumed $N(A)$ to be integer-valued and finite so that we can define

$$N_k(A) = \#\{\text{distinct } t \in A\colon N(\{t\}) = k\} \qquad (k = 1, 2, \ldots)$$

and thereby express $N(A)$ as

$$N(A) = \sum_{k=1}^{\infty} k N_k(A). \tag{3.3.19}$$

By definition, these point processes $N_k(\cdot)$ are simple and stationary, and for them we can define indicator random variables $I_{ni}^{(k)}$, analogous to I_{ni} in (3.3.10), by

$$I_{ni}^{(k)} = \begin{cases} 1 \\ 0 \end{cases} \quad \text{according as} \quad N\left(\frac{i-1}{n}, \frac{i}{n}\right] \begin{cases} = k, \\ \neq k. \end{cases} \tag{3.3.20}$$

By letting $n \to \infty$ through $n = 2^p$ for $p = 1, 2, \ldots$, it follows from (3.3.20) and the construction of $N_k(\cdot)$ that

$$N_k(0,1] = \lim_{n\to\infty} \sum_{i=1}^{n} I_{ni}^{(k)} \quad \text{a.s.} \tag{3.3.21}$$

Now $I_{ni}^{(k)} \le I_{ni}$, so when $\lambda < \infty$, it follows from (3.3.21) by using dominated convergence that $\mathrm{E}(N_k(0,1]) < \infty$, being given by

$$\begin{aligned} \mathrm{E}(N_k(0,1]) &= \lim_{n\to\infty} \mathrm{E}\left(\sum_{i=1}^{n} I_{ni}^{(k)} \right) \\ &= \lim_{n\to\infty} n[\phi_k(n^{-1}) - \phi_{k-1}(n^{-1})] \\ &= \lambda \pi_k \,. \end{aligned} \tag{3.3.22}$$

The sample-path definition of $N_k(\cdot)$ having intensity $\lambda\pi_k$ as in (3.3.22) warrants the use of the term *batch-size distribution* for the probability distribution $\{\pi_k\}$. Note that a stationary orderly point process has the degenerate batch-size distribution for which $\pi_1 = 1$, $\pi_k = 0$ (all $k \neq 1$). Otherwise, the sample paths are appropriately described as having *multiple points*; this terminology is reflected in the frequently used description of a simple point process as one *without multiple points*.

The moments of the distribution $\{\pi_k\}$ can be related to those of $N(\cdot)$ as in the next two propositions, in which we call equation (3.3.23) a *generalized Korolyuk equation*.

Proposition 3.3.VIII. *For a crudely stationary point process of finite intensity,*

$$m = \mathrm{E}(N(0,1]) = \lambda \sum_{k=1}^{\infty} k\pi_k, \quad \textit{finite or infinite.} \tag{3.3.23}$$

Proof. Take expectations in (3.3.19) with $A = (0,1]$ and use Fubini's theorem and (3.3.22) to deduce (3.3.23). □

Proposition 3.3.IX. *For a crudely stationary point process of finite intensity λ and finite γth moment, $\gamma \geq 1$,*

$$\lim_{h \downarrow 0} \frac{\mathrm{E}\big([N^\gamma(0,h]]^\gamma\big)}{h} \quad \textit{exists and equals} \quad \lambda \sum_{k=1}^{\infty} k^\gamma \pi_k \,. \tag{3.3.24}$$

Proof. Introduce

$$M_\gamma(x) = \mathrm{E}(N^\gamma(0,x]),$$

and observe that for $x, y > 0$, using $\gamma \geq 1$,

$$\begin{aligned} M_\gamma(x+y) &= \mathrm{E}\big((N(0,x] + N(x,x+y])^\gamma\big) \\ &\geq \mathrm{E}(N^\gamma(0,x]) + \mathrm{E}(N^\gamma(x,x+y]) \\ &= M_\gamma(x) + M_\gamma(y); \end{aligned}$$

that is, the function $M_\gamma(x)$ is superadditive for $x > 0$. When $M_\gamma(x)$ is finite for $0 < x < \infty$, $M_\gamma(x) \to 0$ $(x \downarrow 0)$, so the subadditive function Lemma 3.6.IV applied to $-M_\gamma(x)$ proves the existence part of (3.3.24). Since

$$N^\gamma(0,1] \geq \sum_{i=1}^{n} \left(\sum_{k-1}^{\infty} k^\gamma I_{ni}^{(k)} \right) \to \sum_{k=1}^{\infty} k^\gamma N_k(0,1] \quad \text{a.s.} \quad (n \to \infty),$$

we can use dominated convergence and crude stationarity to conclude that

$$\lim_{n\to\infty} nM_\gamma(n^{-1}) = \mathrm{E}\left(\sum_{k=1}^{\infty} k^\gamma N_k(0,1] \right) = \lambda \sum_{k=1}^{\infty} k^\gamma \pi_k \,. \qquad \square$$

Exercises and Complements to Section 3.3

3.3.1 Verify that a simple point process (Definition 3.3.II) can be defined equivalently as one for which the distances between points of a realization are a.s. positive. [*Hint*: When the realization consists of the points $\{t_n\}$, (3.3.7) is equivalent (Vasil'ev, 1965) to the relation

$$\Pr\{|t_i - t_j| > 0 \quad (\text{all } i \neq j)\} = 1. \,]$$

3.3.2 Show that a mixed Poisson process for which

$$\Pr\{N(0,t] = j\} = \int_1^\infty \frac{\mathrm{e}^{-\lambda t}(\lambda t)^j}{j!}\, \tfrac{1}{2}\lambda^{-3/2}\, \mathrm{d}\lambda$$

is simple but not orderly. A mixed Poisson process with

$$\Pr\{N(0,t] = j\} = \int_1^\infty \frac{\mathrm{e}^{-\lambda t}(\lambda t)^j}{j!}\, \lambda^{-2}\, \mathrm{d}\lambda$$

also has infinite intensity, but it does satisfy the orderliness property (3.3.8).

3.3.3 (a) Let the r.v. X be distributed on $(0,\infty)$ with distribution function $F(\cdot)$ and, conditional on X, let the r.v. Y be uniformly distributed on $(0,X)$. Now define a point process to consist of the set of points $\{nX + Y : n = 0, \pm 1, \ldots\}$. Verify that such a process is stationary and that

$$\Pr\{N(0,h] = 0\} = \int_h^\infty \left(1 - \frac{h}{x}\right) \mathrm{d}F(x) = 1 - h\int_h^\infty x^{-2}F(x)\,\mathrm{d}x,$$

$$\Pr\{N(0,h] \geq 2\} = h\int_{(1/2)h}^h x^{-2}F(x)\,\mathrm{d}x.$$

When $F(x) = x$ for $0 < x < 1$, show that

(i) the intensity $\lambda = \infty$;

(ii) the process is not orderly; and

(iii) it has the Khinchin orderliness property [Khinchin (1956); see also Leadbetter (1972) and Daley (1974)]

$$\Pr\{N(0,h] \geq 2 \mid N(0,h] \geq 1\} \to 0 \qquad (h \to 0). \tag{3.3.25}$$

(b) Let the realizations of a stationary point process come, with probability $\frac{1}{2}$ each, either from a process of doublets consisting of two points at each of $\{n + Y : n = 0, \pm 1, \ldots\}$, where Y is uniformly distributed on $(0,1)$, or from a simple point process as in part (a). Then $\Pr\{N(\{t\}) \leq 1 \text{ for all } t\} = \frac{1}{2}$, so the process is not simple, but it does have the Khinchin orderliness property in (3.3.25).

3.3.4 Suppose that $N(\cdot)$ is a simple point process on $(0,\infty)$ with finite first moment $M(x) = \mathrm{E}N(x)$, and suppose that $M(\cdot)$ is absolutely continuous in the sense that $M(x) = \int_0^x m(y)\,\mathrm{d}y \quad (x > 0)$ for some density function $m(\cdot)$. Show that the distribution functions $G_i(\cdot)$ of Exercise 3.1.1 are also absolutely continuous with density functions $g_i(\cdot)$, where

$$G_i(x) = \int_0^x g_i(y)\,\mathrm{d}y, \qquad \text{and} \qquad m(x) = \sum_{i=1}^\infty g_i(x) \quad \text{a.e.}$$

3.3.5 (*Continuation*). Now define $G_i(x;t)$ as the d.f. of the ith forward recurrence time after t, i.e. $G_j(x;t)$ is the d.f. of $\inf\{u > t: N(t,u] \geq i\}$. Supposing that $N(\cdot)$ has finite first moment and is absolutely continuous in the sense of Exercise 3.3.4, show that when $N(\cdot)$ is simple,

$$g_1(0;t) = m(t), \quad g_i(0;t) = 0 \qquad (i \geq 2).$$

Use these results to give an alternative proof of Korolyuk's Theorem 3.3.IV.

Show also that when the rth moment of $N(\cdot)$ is finite,

$$\lim_{h \downarrow 0} \frac{\mathrm{E}[(N(t,t+h])^r]}{h} = m(t).$$

3.3.6 Given any point process with sample realizations N, define another point process with sample realization N^* by means of

$$N^*(A) = \#\{\text{distinct } \ x \in A: N(\{x\}) \geq 1\} \qquad (\text{all Borel sets } A)$$

(in the setting of marked point processes in Section 6.4 below, N^* here is an example of a ground process, denoted N_{g} there). Show that if, for any real finite $s > 0$,

$$\mathrm{E}(\mathrm{e}^{-sN(A)}) \geq \mathrm{E}(\mathrm{e}^{-sN^*(A)}) \qquad (\text{all Borel sets } A),$$

then N is simple. Irrespective of whether or not it is simple, $N(A) = 0$ iff $N^*(A) = 0$.

Show that if N is a compound Poisson process as in Theorem 2.2.II, then N^* is a stationary Poisson process with rate λ.

3.3.7 Consider a compound Poisson process as in Theorem 2.2.II, and suppose that the mean batch size $\Pi'(1) = \sum k\pi_k$ is infinite. Let the points of the process be subject to independent shifts with a common distribution that has no atoms. The resulting process is no longer Poisson, is simple, and has infinite intensity.

When the shifts are i.i.d. and uniform on $(0,1)$, show that, for $0 < h < 1$,

$$\Pr\{N(0,h] = 0\} = \exp\left(-\lambda(1+h) + \lambda(1-h)\Pi(1-h) + 2\lambda \int_0^h \Pi(1-u)\,\mathrm{d}u\right).$$

3.4. Palm–Khinchin Equations

Throughout this section, we use $\mathcal{P}$ to denote the probability measure of a stationary point process (Definition 3.2.I). Our aim is to describe an elementary approach to the problem raised by the intuitively reasonable idea that the stationarity of a point process as in Definition 3.2.I should imply some equivalent interval stationarity property as in Definition 3.2.II. For example, for positive x and y and small positive h, stationarity of the point process $N(\cdot)$ implies that

$$\begin{aligned}
&\mathcal{P}\{N(t,t+h] = N(t+x,t+x+h] = N(t+x+y,t+x+y+h] = 1,\\
&\qquad\qquad N(t,t+x+y+h] = 3\}\\
&= \mathcal{P}\{N(-h,0] = N(x-h,x] = N(x+y-h,x+y] = 1, N(-h,x+y] = 3\}\\
&\equiv \mathcal{P}\{A_{x,y,h}\}, \quad \text{say.} \qquad (3.4.1)
\end{aligned}$$

Now the event $A_{x,y,h}$ describes a sample path with a point near the origin

and intervals of about x and y, respectively, to the next two points. Our intuition suggests that, as far as the dependence on the variables x and y is concerned, $\mathcal{P}\{A_{x,y,h}\}$ should be related to the probability measure $\mathcal{P}_0(\cdot)$ for an interval stationary point process; that is, there should be a simple relation between $\mathcal{P}\{A_{x,y,h}\}$ and $\mathcal{P}_0\{\tau_1 \simeq x, \tau_2 \simeq y\}$. We proceed to describe the partial solution that has its roots in Khinchin's monograph (1955) and that connects $\mathcal{P}\{N(0,x] \le j\}$ to what we shall show is a distribution function

$$R_j(x) = \lim_{h\downarrow 0} \mathcal{P}\{N(0,x] \ge j \mid N(-h,0] > 0\} \qquad (j = 1, 2, \ldots). \tag{3.4.2}$$

What emerges from the deeper considerations of Chapter 13 is that, granted orderliness, there exists an interval stationary point process $\{\tau_j\}$ with probability measure $\mathcal{P}_0$, so $\mathcal{P}_0\{t_0 = 0\} = 1$, for which we can indeed set

$$\mathcal{P}_0(\cdot) = \lim_{h\downarrow 0} \mathcal{P}(\cdot \mid N(-h,0] > 0).$$

It then follows, for example, that

$$\mathcal{P}_0\{\tau_1 + \cdots + \tau_j \le x\} = R_j(x) \tag{3.4.3}$$

[see (3.4.2) and (3.1.9)], thereby identifying a random variable having $R_j(\cdot)$ as its distribution function.

Instead of the expression in (3.4.1), we consider first the probability

$$\psi_j(x,h) \equiv \mathcal{P}\{N(0,x] \le j,\ N(-h,0] > 0\} \tag{3.4.4}$$

and prove the following proposition.

Proposition 3.4.I. *For a stationary point process of finite intensity, the limit*

$$Q_j(x) = \lim_{h\downarrow 0} \mathcal{P}\{N(0,x] \le j \mid N(-h,0] > 0\} \tag{3.4.5}$$

exists for $x > 0$ and $j = 0, 1, \ldots$, being right-continuous and nonincreasing in x with $Q_j(0) = 1$.

Proof. Observe that for $u, v > 0$,

$$\begin{aligned}\psi_j(x, u+v) = {} & \mathcal{P}\{N(0,x] \le j,\ N(-u,0] > 0\} \\ & + \mathcal{P}\{N(0,x] \le j,\ N(-u,0] = 0,\ N(-u-v,-u] > 0\}.\end{aligned}$$

In the last term,

$$\begin{aligned}\{N(0,x] \le j,\ N(-u,0] = 0\} &= \{N(-u,x] \le j,\ N(-u,0] = 0\} \\ &\subseteq \{N(-u,x] \le j\} \\ &\subseteq \{N(-u, x-u] \le j\},\end{aligned}$$

and then using stationarity of $\mathcal{P}(\cdot)$, we have

$$\psi_j(x, u+v) \le \psi_j(x,u) + \psi_j(x,v).$$

Consequently, the subadditivity lemma implies that the limit as $h \to 0$ of $\psi_j(x,h)/h$ exists, being bounded by λ [because $\psi_j(x,h) \le \phi_j(h)$], so by writing

$$\mathcal{P}\{N(0,x] \le j \mid N(-h,0] > 0\} = \frac{\psi_j(x,h)}{\phi(h)} = \frac{\psi_j(x,h)/h}{\phi(h)/h},$$

we can let $h \to 0$ to prove the assertion in (3.4.5) concerning existence. By subadditivity, and right-continuity and monotonicity in x of $\psi_j(x,h)$,

$$Q_j(x) = \sup_{h>0} \frac{\psi_j(x,h)}{\lambda h} = \sup_{h>0} \sup_{y>x} \frac{\psi_j(y,h)}{\lambda h} = \sup_{y>x} Q_j(y),$$

so $Q_j(x)$ is right-continuous and nonincreasing in x, with $Q_j(0) = 1$ since $\psi_j(0,h) = \phi(h)$. □

It follows from this result that every

$$R_j(x) \equiv 1 - Q_{j-1}(x) \qquad (j = 1, 2, \ldots) \tag{3.4.6}$$

is a d.f. on $(0,\infty)$ except for the possibility, to be excluded later under the conditions of Theorem 3.4.II, that $\lim_{x\to\infty} R_j(x)$ may be less than 1. The plausible interpretation of (3.4.5), or equivalently, of (3.4.6), is that $R_j(x)$ represents the conditional probability (in which the conditioning event has zero probability)

$$\mathcal{P}\big\{N(0,x] \ge j \mid N(\{0\}) > 0\big\} = \mathcal{P}\{\tau_1 + \cdots + \tau_j \le x \mid t_0 = 0, t_1 > 0\}. \tag{3.4.7}$$

Example 3.4(a) *Renewal process.* Consistent with (3.4.7), for a renewal process starting at 0 with lifetime d.f. F for which $F(0+) = 0$, $R_j(x) = F^{j*}(x)$, where $F^{n*}(\cdot)$ is the n-fold convolution of F. In this case then, $R_j(\cdot)$ is the d.f. of the sum of j random variables that are not merely stationary but also independent. On the other hand, if we have a renewal process with a point at 0 and having lifetime d.f. F for which $0 < F(0+) < 1$, then the constraint in (3.4.7) that $\tau_1 = t_1 - t_0 > 0$ means that τ_1 has d.f. $F_+(x) = (F(x) - F(0+))/(1 - F(0+))$, while $\tau_2, \tau_3, \ldots$ have d.f. F and

$$R_j(x) = \int_0^x F^{(j-1)*}(x-u)\, \mathrm{d}F_+(u) \qquad (j = 1, 2, \ldots).$$

Thus, $R_j(x)$ is here the d.f. of the sum of nonstationary r.v.s, and so for a renewal process we have the stationarity property at (3.4.3) only when $F(0+) = 0$; that is, when the process is orderly (or equivalently, simple). □

This last assumption is also what enables us to proceed simply in general [but, note the remarks around (3.4.12) below].

Theorem 3.4.II. *For an orderly stationary point process of finite intensity λ and such that*

$$\mathcal{P}\{N(-\infty,0] = N(0,\infty) = \infty\} \;=\; 1, \tag{3.4.8}$$

$$\mathcal{P}\{N(0,x] \le j\} = \lambda \int_x^\infty q_j(u)\,\mathrm{d}u \qquad (j = 0,1,\ldots), \tag{3.4.9}$$

where

$$q_j(x) = \lim_{h\downarrow 0} \mathcal{P}\{N(0,x] = j \mid N(-h,0] > 0\}, \tag{3.4.10}$$

and $R_j(x) = 1 - \sum_{k=0}^{j-1} q_k(x)$ is a distribution function on $(0,\infty)$ with mean $j\lambda^{-1}$ for each $j = 1,2,\ldots$.

Proof. Set

$$P_j(x) = \mathcal{P}\{N(0,x] \le j\}$$

and observe by Proposition 3.4.I and the assumption of orderliness that

$$\begin{aligned} P_j(x+h) &= \sum_{i=0}^{j} \mathcal{P}\{N(0,x] \le j-i, N(-h,0] = i\} \\ &= \mathcal{P}\{N(0,x] \le j\} \;-\; \mathcal{P}\{N(0,x] \le j, N(-h,0] > 0\} \\ &\quad +\; \mathcal{P}\{N(0,x] \le j-1, N(-h,0] = 1\} + o(h). \end{aligned}$$

Thus,

$$\begin{aligned} P_j(x+h) - P_j(x) &= \mathcal{P}\{N(0,x] \le j-1,\; N(-h,0] \ge 1\} \\ &\quad - \mathcal{P}\{N(0,x] \le j,\; N(-h,0] > 0\} + o(h) \\ &= -\lambda h q_j(x) + o(h), \end{aligned}$$

where the existence of $q_j(x)$ in (3.4.10) is assured by (3.4.5) directly for $j = 0$ and then by induction for $j = 1,2,\ldots$. Using D_+ to denote the right-hand derivative operator, it follows that

$$D_+ P_j(x) = -\lambda q_j(x).$$

Setting $Q_{-1}(x) \equiv 0$, the nonnegative function $q_j(x) = Q_j(x) - Q_{j-1}(x)$ is the difference of two bounded nonincreasing functions and hence is integrable on bounded intervals with

$$P_j(x) - P_j(y) = \lambda \int_x^y q_j(u)\,\mathrm{d}u. \tag{3.4.11}$$

The assumption in (3.4.8) implies that $P_j(y) \to 0$ for $y \to \infty$, so (3.4.9) now follows from (3.4.11).

Letting $x \downarrow 0$ in (3.4.9), it follows that

$$\lambda^{-1} = \int_0^\infty q_j(u)\,\mathrm{d}u \qquad (j = 0, 1, \ldots),$$

and hence, using (3.4.6) as well, that for $j = 1, 2, \ldots,$

$$\int_0^\infty \big(1 - R_j(u)\big)\,\mathrm{d}u = \int_0^\infty Q_{j-1}(u)\,\mathrm{d}u = j\lambda^{-1}. \qquad \square$$

There is a most instructive heuristic derivation of (3.4.9) as follows. By virtue of (3.4.8), if we look backward from a point x, there will always be some point $u < x$ for which $N(u, x] \le j$ and $N[u, x] > j$. In fact, because of orderliness, we can write (with probability 1)

$$\{N(0, x] \le j\} = \bigcup_{u \le 0} \{N(u, x] = j,\ N(\{u\}) = 1\},$$

in which we observe that the right-hand side is the union of the mutually exclusive events that the $(j + 1)$th point of $N(\cdot)$ looking backward from x occurs at some $u \le 0$. Consequently, we can add their 'probabilities', which by (3.4.7), (3.3.4), and orderliness equal $q_j(x - u)\lambda\,\mathrm{d}u$, yielding the Palm–Khinchin equation (3.4.9) in the form

$$P_j(x) = \lambda \int_{-\infty}^0 q_j(x - u)\,\mathrm{d}u.$$

Without the orderliness assumption, made from (3.4.8) onward above, we can proceed as follows. First (see Proposition 3.4.I), we show that the function

$$\psi_{j|i}(x, h) \equiv \mathcal{P}\{N(0, x] \le j,\ 0 < N(-h, 0] \le i\} \tag{3.4.12}$$

is subadditive in h and so deduce that, for those i for which $\pi_i > 0$ [see (3.3.16)], there exists the limit

$$Q_{j|i}(x) = \lim_{h \downarrow 0} \mathcal{P}\{N(0, x] \le j \mid N(-h, 0] = i\}, \tag{3.4.13}$$

with

$$\mathcal{P}\{N(0, x] \le j,\ N(-h, 0] = i\} = \lambda\pi_i Q_{j|i}(x)h + o(h) \qquad (h \downarrow 0)$$

irrespective of $\pi_i >$ or $= 0$ by setting $Q_{j|i}(x) \equiv 0$ when $\pi_i = 0$. Then, the argument of the proof of Theorem 3.4.II can be mimicked in establishing that

$$P_j(x) = \lambda \int_x^\infty \sum_{i=1}^\infty \pi_i [Q_{j|i}(u) - Q_{j-i|i}(u)]\,\mathrm{d}u, \tag{3.4.14}$$

setting $Q_{k|i}(u) \equiv 0$ for $k < 0$, and it can also be shown that, when $\pi_i > 0$,

$$R_{j|i}(x) \equiv 1 - Q_{j-1|i}(x) \equiv 1 - \sum_{k=0}^{j-1} q_{k|i}(x)$$

is a proper distribution function on $(0, \infty)$.

For any point process N, the random variable

$$T_u \equiv \inf\{t > 0 \colon N(u, y + t] > 0\} \tag{3.4.15}$$

is the *forward recurrence time r.v.* For a stationary point process, $T_u =_d T_0$ for all u, and we can study its distribution via the Palm–Khinchin equations since $\{T_0 > x\} = \{N(0, x] = 0\}$. Assuming that (3.4.8) holds,

$$\mathcal{P}\{T_0 > x\} = \lambda \int_x^\infty q_0(u)\,\mathrm{d}u \tag{3.4.16}$$

when $N(\cdot)$ is orderly as in Theorem 3.4.II. Recall that $q_0(\cdot)$ is the tail of the d.f. $R_1(\cdot)$, which can be interpreted as the d.f. of the length τ_1 of an arbitrarily chosen interval. Then, still assuming that (3.4.8) holds,

$$\begin{aligned} \mathrm{E}T_0 &= \int_0^\infty \mathcal{P}\{T_0 > x\}\,\mathrm{d}x = \lambda \int_0^\infty u q_0(u)\,\mathrm{d}u \\ &= \lambda \int_0^\infty u\big(1 - R_1(u)\big)\,\mathrm{d}u \;=\; \tfrac{1}{2}\lambda(\mathrm{E}\tau_1^2). \end{aligned} \tag{3.4.17}$$

When all intervals are of the same length, Δ say, $\lambda = \Delta^{-1}$ and $\mathrm{E}T_0 = \frac{1}{2}\Delta$, whereas for a Poisson process, τ_1 has mean Δ and second moment $\mathrm{E}\tau_1^2 = 2\Delta^2$, so then $\mathrm{E}T_0 = \Delta$. These remarks amplify the comments on the waiting-time paradox at the end of Section 3.2.

In both Theorem 3.4.II and the discussion of the forward recurrence time r.v. T_u, the caveat that $\mathcal{P}\{N(0, \infty) = \infty\} = 1$ has been added. This is because stationary point processes on the real line $\mathbb{R}$ have the property that

$$\mathcal{P}\{N(0, \infty) = \infty = N(-\infty, 0)\} \;=\; 1 - \mathcal{P}\{N(\mathbb{R}) = 0\}, \tag{3.4.18}$$

which is equivalent to

$$\mathcal{P}\{0 < N(\mathbb{R}) < \infty\} = 0. \tag{3.4.19}$$

A similar property in a more general setting is proved in Chapter 12.

Inspection of the statements onward from (3.4.8) shows that they are either conditional probability statements (including limits of such statements), which in view of (3.4.18) reduce to being conditional also on $\{N(\mathbb{R}) = \infty\}$, or unconditional statements, which without (3.4.8) need further elaboration. This is quickly given: (3.4.8) is equivalent by (3.4.18) to $\mathcal{P}\{T_0 < \infty\} = 1$, and without (3.4.8), equations (3.4.16) and (3.4.17) must be replaced by assertions

of the form

$$\mathcal{P}\{T_0 > x\} = \lambda\varpi \int_x^\infty q_0(u)\,\mathrm{d}u + 1 - \varpi, \tag{3.4.20}$$

$$\mathrm{E}(T_0 \mid T_0 < \infty) = \tfrac{1}{2}\lambda \mathrm{E}(\tau_1^2), \tag{3.4.21}$$

where $\varpi = \mathcal{P}\{N(\mathbb{R}) = \infty\} = \mathcal{P}\{T_0 < \infty\}$.

Exercises and Complements to Section 3.4

3.4.1 Analogously to (3.4.15), define a *backward recurrence time* r.v. $B_u \equiv \inf\{t > 0: N(u-t, u] > 0\}$ (assuming this to be finite a.s.). Show that when $N(\cdot)$ is a stationary point process, $B_u =_d B_0 =_d T_0$. The r.v. $L_u = B_u + T_u$ denotes the *current lifetime* r.v.; when N is orderly and stationary, show that $\mathrm{E}L_0 = (\mathrm{E}\tau_1^2)/(\mathrm{E}\tau_1)$ [see (3.4.16)] and that

$$\mathcal{P}\{L_0 < x\} = \lambda \int_0^x [q_0(u) - q_0(x)]\,\mathrm{d}u = \lambda \int_0^x u\,\mathrm{d}R_1(u).$$

3.4.2 Use Palm–Khinchin equations to show that when the hazard functions q and r of the interval and forward recurrence r.v.s τ_0 and T_0, respectively, are such that $r(x) = r(0) + \int_0^x r'(u)\,\mathrm{d}u$ for some density function r', then q and r are related by

$$r(x) = q(x) + r'(x)/r(x) \qquad (x > 0).$$

3.4.3 Show that for an orderly point process,

$$\mathrm{E}N(0,1] = \int_0^1 \mathcal{P}\{N(\mathrm{d}x) \geq 1\},$$

where the right-hand side is to be interpreted as a Burkill integral [see Fieger (1971) for further details].

3.4.4 For a point process N on $\mathbb{R}$, define the event

$$B_k \equiv B_k((x_i, j_i): i = 1, \ldots, k) = \{N(0, x_i] \leq j_i \ (i = 1, \ldots, k)\}$$

for positive x_i, nonnegative integers j_i $(i = 1, \ldots, k)$, and any fixed finite positive integer k.

(a) When N is stationary with finite intensity λ,

$$\psi(B_k, h) = \mathcal{P}(B_k \cap \{N(-h, 0] > 0\})$$

is subadditive in $h > 0$, the limit $Q(B_k) = \lim_{h\downarrow 0} \mathcal{P}(B_k \mid \{N(-h, 0] > 0\})$ exists finite, is right-continuous and nonincreasing in each x_i and nondecreasing in j_i, is invariant under permutations of $(x_1, j_1), \ldots, (x_k, j_k)$, satisfies the consistency conditions

$$\begin{aligned} Q(B_k) &= Q(B_{k+1}((0, j_{k+1}),\ (x_i, j_i)\ (i = 1, \ldots, k))) \\ &= Q(B_{k+1}((x_{k+1}, \infty),\ (x_i, j_i)\ (i = 1, \ldots, k))), \end{aligned}$$

and

$$Q(B_k) = \lim_{h\downarrow 0} \psi(B_k, h)/\lambda h = \sup_{h>0} \psi(B_k, h\}/\lambda h.$$

(b) Define a shift operator S_h $(h > 0)$ and a difference operator Δ on B_k by

$$S_h B_k = B_k\big((x_i + h, j_i)\ (i = 1, \ldots, k)\big),$$
$$\Delta B_k = B_k\big((x_i, j_i - 1)\ (i = 1, \ldots, k)\big),$$

and put $q(B_k) = Q(B_k) - Q(\Delta B_k)$, with the convention that if any $j_i = 0$, then ΔB_k is a null set with $Q(\Delta B_k) = 0$. Under the condition (3.4.8) of Theorem 3.4.II, the right-hand derivative $D_+\mathcal{P}(B_k)$ exists in the sense that $D_+\mathcal{P}(S_h B_k)\,|_{h=0} = -\lambda q(B_k)$, and

$$\mathcal{P}(B_k) - \mathcal{P}(S_x B_k) = \lambda \int_0^x q(S_u B_k)\,\mathrm{d}u.$$

[See Daley and Vere-Jones (1972, Section 7) and Slivnyak (1962, 1966). Note that Slivnyak used a slightly different operator S_h^0 defined by

$$S_h^0 B_k = B_{k+1}\big((h, 0), (x_i + h, j_i)\ (i = 1, \ldots, k)\big),$$

so that $\psi(B_k, h) = \mathcal{P}(B_k) - \mathcal{P}(S_h^0 B_k)$, and deduced the existence of a derivative in h of $\mathcal{P}(S_h^0 B_k)$ from the convexity in h of this function, assuming stationarity of N but not necessarily that it has finite intensity.]

3.5. Ergodicity and an Elementary Renewal Theorem Analogue

Let $N(\cdot)$ be a stationary point process with finite mean density $m = \mathrm{E}N(0, 1]$. Then, the sequence $\{X_n\}$ of random variables defined by

$$X_n = N(n - 1, n] \qquad (n = 0, \pm 1, \ldots)$$

is stationary with finite first moment $m = \mathrm{E}X_n$ (all n), and by the strong law for stationary random sequences,

$$\frac{N(0, n]}{n} = \frac{X_1 + \cdots + X_n}{n} \to \xi \quad \text{a.s.}$$

for some random variable ξ for which $\mathrm{E}\xi = m$. Using $\lfloor x \rfloor$ to denote the largest integer $\le x$, it then follows on letting $x \to \infty$ in the inequalities

$$\frac{N\big(0, \lfloor x \rfloor\big]}{\lfloor x \rfloor} \cdot \frac{\lfloor x \rfloor}{x} \le \frac{N\big(0, \lfloor x \rfloor\big]}{x} \le \frac{N\big(0, \lfloor x \rfloor + 1\big]}{\lfloor x \rfloor + 1} \cdot \frac{\lfloor x \rfloor + 1}{x} \qquad (x \ge 1)$$

that we have proved the following proposition.

Proposition 3.5.I. *For a stationary point process with finite mean density $m = \mathrm{E}N(0, 1]$, $\zeta \equiv \lim_{x\to\infty} N(0, x]/x$ exists a.s. and is a random variable with $\mathrm{E}\zeta = m$.*

In our discussion of limit properties of stationary point processes we shall have cause to use various concepts of ergodicity; for the present we simply use the following definition.

Definition 3.5.II. *A stationary point process with finite mean density m is* ergodic *when*

$$\mathcal{P}\{N(0,x]/x \to m \ \ (x \to \infty)\} = 1.$$

Suppose that in addition to being ergodic, the second moment $\mathrm{E}[(N(0,1])^2]$ is finite, so by stationarity and the Cauchy–Schwarz inequality, $\mathrm{E}[(N(0,x])^2] < \infty$ for all finite positive x. Then, we can use an argument similar to that leading to Proposition 3.5.I to deduce from the convergence in mean square of $(X_1 + \cdots + X_n)/n = N(0,n]/n$ to the same limit [see e.g. (2.15) of Doob (1953, p. 471) or Chapter 12 below] that

$$\mathrm{var}(N(0,x]/x) = \mathrm{E}(N(0,x]/x - m)^2 \to 0 \qquad (x \to \infty) \tag{3.5.1}$$

when $N(\cdot)$ is ergodic with finite second moment. This is one of the key probabilistic steps in the proof of the next theorem, in which the asymptotic result in (3.5.3), combined with the remarks that follow, is an analogue of the elementary renewal theorem [see Exercise 4.1.1(b) and Section 4.4 below]. The function $U(\cdot)$, called the *expectation function* in Daley (1971), is the analogue of the renewal function.

Theorem 3.5.III. *For a stationary ergodic point process with finite second moment and mean density m, the second-moment function*

$$M_2(x) \equiv \mathrm{E}[(N(0,x])^2] = \int_0^x \big(2U(u) - 1\big)m \,\mathrm{d}u \tag{3.5.2}$$

for some nondecreasing function $U(\cdot)$ for which

$$U(x)/x \to m \qquad (x \to \infty); \tag{3.5.3}$$

when the process is orderly,

$$U(x) = \sum_{j=0}^{\infty} R_j(x). \tag{3.5.4}$$

Remarks. (1) It is consistent with the interpretation of $R_j(\cdot)$ in (3.4.3) as the d.f. of the sum $S_j = \tau_1 + \cdots + \tau_j$ that

$$U(x) = \lim_{h \downarrow 0} \mathrm{E}(N(0,x] + 1 \mid N(-h,0] > 0)$$

in the case where $N(\cdot)$ is orderly. In the nonorderly case, it emerges that, given an ergodic stationary sequence $\{\tau_j\}$ of nonnegative random variables

with $\mathrm{E}\tau_j = 1/m$ and partial sums $\{S_n\}$ given by $S_0 = 0$ and

$$S_n = \tau_1 + \cdots + \tau_n,\ S_{-n} = -(\tau_0 + \cdots + \tau_{-(n-1)}) \qquad (n = 1, 2, \ldots),$$

we can interpret $U(\cdot)$ as

$$2U(x) - 1 = \mathrm{E}\#\{n = 0, \pm 1, \ldots : |S_n| \le x\} = \sum_{n=-\infty}^{\infty} \Pr\{|S_n| \le x\}. \quad (3.5.5)$$

In the case where the random variables $\{\tau_j\}$ are independent and identically distributed,

$$U(x) = \sum_{n=0}^{\infty} F^{n*}(x) \quad (3.5.6)$$

and hence $U(\cdot)$ is then the renewal function.

(2) It follows from (3.5.2) that

$$\operatorname{var} N(0, x] = \int_0^x \big(2[U(u) - mu] - 1\big) m \, \mathrm{d}u. \quad (3.5.7)$$

(3) It is a simple corollary of (3.5.3) that for every fixed finite y,

$$\frac{U(x + y)}{U(x)} \to 1 \qquad (x \to \infty). \quad (3.5.8)$$

Proof of Theorem 3.5.III. From the definition in (3.5.2) with $N(x) = N(0, x]$,

$$\begin{aligned} M_2(x) = \mathrm{E}\big([N(x)]^2\big) &= \operatorname{var} N(x) + \big(\mathrm{E}N(x)\big)^2 \\ &= x^2[\operatorname{var}(N(x)/x) + m^2] \sim m^2 x^2 \qquad (x \to \infty) \end{aligned}$$

when $N(\cdot)$ is ergodic, by (3.5.1). If we can assume that $M_2(\cdot)$ is absolutely continuous and that the function $U(\cdot)$, which can then be defined as in (3.5.2), is monotonically nondecreasing, we can appeal to a Tauberian theorem (e.g. Feller, 1966, p. 421) and conclude that (3.5.3) holds.

It remains then to establish (3.5.2), for which purpose we assume first that $N(\cdot)$ is orderly so that the representation (3.4.9) is at our disposal. It is a matter of elementary algebra that

$$\begin{aligned} M_2(x) + mx = \mathrm{E}\big(N(x)(N(x) + 1)\big) &= \sum_{j=1}^{\infty} j(j+1)\mathcal{P}\{N(x) = j\} \\ &= 2\sum_{k=1}^{\infty} k\mathcal{P}\{N(x) \ge k\} \end{aligned}$$

$$= 2\sum_{k=1}^{\infty}(k+1)\int_0^x q_k(u)\lambda\,\mathrm{d}u$$
$$= 2\int_0^x \left(1+\sum_{j=0}^{\infty}\big(1-Q_j(u)\big)\right)\lambda\,\mathrm{d}u \;=\; 2\int_0^x \sum_{j=0}^{\infty} R_j(u)\lambda\,\mathrm{d}u,$$

where $R_0(u)\equiv 1$. Thus, we have (3.5.2) in the case of orderly $N(\cdot)$ with the additional identification that

$$U(x)=\sum_{j=0}^{\infty} R_j(x), \tag{3.5.9}$$

of which (3.5.6) is a special case. Note in (3.5.9) that the nondecreasing nature of each $R_j(\cdot)$ ensures the same property for $U(\cdot)$.

When $N(\cdot)$ is no longer orderly, we must appeal to (3.4.14) in writing

$$M_2(x)+mx = 2\sum_{k=0}^{\infty}(k+1)\big(1-P_k(x)\big)$$
$$= 2\sum_{k=0}^{\infty}(k+1)\int_0^x \sum_{i=1}^{\infty}\pi_i\big(Q_{k|i}(u)-Q_{k-i|i}(u)\big)\lambda\,\mathrm{d}u. \tag{3.5.10}$$

Without loss of generality, we may set $Q_{k|i}(x)\equiv 1$ when $\pi_i=0$. Fubini's theorem is then applicable as before in the manipulations below:

$$2\sum_{k=0}^{\infty}(k+1)\sum_{i=1}^{\infty}\pi_i\sum_{j=(k-i+1)_+}^{k} q_{j|i}(u) = 2\sum_{i=1}^{\infty}\pi_i\sum_{k=0}^{\infty}(k+1)\sum_{j=(k-i+1)_+}^{k} q_{j|i}(u)$$
$$= \sum_{i=1}^{\infty}\pi_i\sum_{j=0}^{\infty} i(2j+i+1)q_{j|i}(u)$$
$$= \sum_{i=0}^{\infty} i\pi_i\left[i+1+2\sum_{j=0}^{\infty}\big(1-Q_{j|1}(u)\big)\right]. \tag{3.5.11}$$

Substitute (3.5.11) in (3.5.10) and recall that $Q_{j|i}(u)$ is nonincreasing; this establishes the existence of nondecreasing $U(\cdot)$ in (3.5.2) as required. □

Exercises and Complements to Section 3.5

3.5.1 (see Theorem 3.5.III). Use the Cauchy–Schwarz inequality to show that, when $M_2(x)\equiv \mathrm{E}N^2(0,x]<\infty$ for finite x, $(M_2(x))^{1/2}$ is subadditive in $x>0$ and hence that there is then a finite constant $\lambda_2\ge m^2$ such that $M_2(x)\sim \lambda_2 x^2$ $(x\to\infty)$.

3.5.2 Let $N(\cdot)$ be a stationary mixed Poisson process with $\mathcal{P}\{N(0,t]=j\}=\frac{1}{2}\mathrm{e}^{-t}t^j/j!+\frac{1}{2}\mathrm{e}^{-2t}(2t)^j/j!$. Show that $\lambda=\frac{3}{2}=m<U(t)/t=\frac{5}{3}$ (all $t>0$) (cf. Theorem 3.5.III; this process is not ergodic) and that $N(0,t]/t\to\xi$ $(t\to\infty)$, where $\xi=1$ or 2 with probability $\frac{1}{2}$ each.

3.6. Subadditive and Superadditive Functions

We have referred earlier in this chapter to properties of subadditive and superadditive functions, and for convenience we now establish these properties in a suitable form. For a more extensive discussion of such functions, see Hille and Phillips (1957).

A function $g(x)$ defined for $0 \le x < a \le \infty$ is *subadditive* when

$$g(x+y) \le g(x) + g(y) \tag{3.6.1}$$

holds throughout its domain of definition; similarly, a function $h(x)$ for which

$$h(x+y) \ge h(x) + h(y) \tag{3.6.2}$$

holds is *superadditive*. A function $f(x)$ for which

$$f(x+y) = f(x) + f(y) \tag{3.6.3}$$

holds is *additive*, and (3.6.3) is known as *Cauchy's functional equation* or (see e.g. Feller, 1966, Section IV.4) the *Hamel equation*.

Lemma 3.6.I. *For a subadditive function $g(\cdot)$ that is bounded on finite intervals, $\mu \equiv \inf_{x>0} g(x)/x$ is finite or $-\infty$, and*

$$\frac{g(x)}{x} \to \mu \qquad (x \to \infty). \tag{3.6.4}$$

Proof. There exists y for which $g(y)/y < \mu'$ for any $\mu' > \mu$. Given any x, there is a unique integer n for which $x = ny + \eta$, where $0 \le \eta < y$, and $n \to \infty$ as $x \to \infty$. Then

$$\begin{aligned} \frac{g(x)}{x} &\le \frac{g(ny) + g(\eta)}{x} \le \frac{ng(y)}{ny+\eta} + \frac{g(\eta)}{x} \\ &= \frac{g(y)}{y + \eta/n} + \frac{g(\eta)}{x} \to \frac{g(y)}{y} \qquad (x \to \infty). \end{aligned}$$

Thus, $\limsup_{x\to\infty} g(x)/x \le \mu'$, and μ' being an arbitrary quantity $> \mu$, this proves the lemma. □

The function $-h(x)$ is subadditive when $h(\cdot)$ is superadditive, and an additive function is both subadditive and superadditive, so Lemma 3.6.I implies both of the following results.

Lemma 3.6.II. *For a superadditive function $h(\cdot)$ that is bounded on finite intervals, $\mu \equiv \sup_{x>0} h(x)/x$ is finite or $+\infty$ and*

$$\frac{h(x)}{x} \to \mu \qquad (x \to \infty). \tag{3.6.5}$$

Lemma 3.6.III. *An additive function $f(\cdot)$ that is bounded on finite intervals satisfies*

$$f(x) = f(1)x \qquad (0 \le x < \infty). \tag{3.6.6}$$

In passing, note that there do exist additive functions that do not have the linearity property (3.6.6): they are unbounded on every finite interval and moreover are not measurable (see e.g. Hewitt and Zuckerman, 1969).

Observe also that nonnegative additive functions satisfy (3.6.6) with the understanding that $f(1) = \infty$ is allowed.

The behaviour near 0 of subadditive and superadditive functions requires the stronger condition of continuity at 0 in order to derive a useful result [a counterexample when $f(\cdot)$ is not continuous at 0 is indicated in Hille and Phillips (1957, Section 7.11)].

Lemma 3.6.IV. *Let $g(x)$ be subadditive on $[0, a]$ for some $a > 0$, and let $g(x) \to 0$ as $x \to 0$. Then $\lambda \equiv \sup_{x>0} g(x)/x$ is finite or $+\infty$, and*

$$\frac{g(x)}{x} \to \lambda \qquad (x \to 0). \tag{3.6.7}$$

Proof. The finiteness of $g(x)$ for some $x > 0$ precludes the possibility that $\lambda = -\infty$. Consider first the case where $0 < \lambda < \infty$, and suppose that $g(a_n)/a_n < \lambda - 2\epsilon$ for some $\epsilon > 0$ for all members of a sequence $\{a_n\}$ with $a_n \to 0$ as $n \to \infty$. For any given $x > 0$, we can find a_n sufficiently small that $\sup_{0 \le \delta < a_n} g(\delta) < \epsilon x$. Write $x = k_n a_n + \delta_n$ for some nonnegative integer k_n and $0 \le \delta_n < a_n$. Then

$$\begin{aligned} \frac{g(x)}{x} &\le \frac{k_n g(a_n) + g(\delta_n)}{k_n a_n + < \delta_n} \le \frac{g(a_n)/a_n}{1 + (\delta_n/a_n)/k_n} + \frac{g(\delta_n)}{x} \\ &\le \lambda - \epsilon \qquad (\text{all } n \text{ sufficiently large}). \end{aligned}$$

Thus, $\sup_{x>0} g(x)/x \le \lambda - \epsilon$, contradicting the definition of λ. The case $-\infty < \lambda \le 0$ is established by considering $g_1(x) \equiv g(x) + \lambda' x$ for some finite $\lambda' > -\lambda$. Finally, the case $\lambda = \infty$ is proved by contradiction starting from the supposition that $g(a_n)/a_n \to \lambda'' < \infty$ for some $\{a_n\}$ with $a_n \to 0$. □

Lemma 3.6.V. *Let $h(x)$ be superadditive on $[0, a]$ for some $a > 0$, and let $h(x) \to 0$ as $x \to 0$. Then $\lambda \equiv \inf_{x>0} h(x)/x$ is finite or $-\infty$, and*

$$\frac{h(x)}{x} \to \lambda \qquad (x \to 0). \tag{3.6.8}$$

CHAPTER 4

Renewal Processes

The renewal process and variants of it have been the subject of much study, both as a model in many fields of application (see e.g. Cox, 1962; Cox and Lewis, 1966; Cox and Isham, 1980) and as a source of important theoretical problems. It is not the aim of this chapter to repeat much of the material that is available, for example, in Volume II of Feller (1966); rather, we have selected some features that are either complementary to Feller's treatment or relevant to more general point processes.

The first two sections are concerned with basic properties, setting these where possible into a point process context. The third section is concerned with some characterization theorems and the fourth section with aspects of the renewal theorem, a topic so important and with such far-reaching applications that it can hardly be omitted. Two versions of the theorem are discussed, corresponding to different forms of convergence of the renewal measure to Lebesgue measure. Some small indication of the range of applications is given in Section 4.5, which is concerned with 'neighbours' of the renewal process, notably the Wold process of correlated intervals.

A final section is concerned with the concept of a hazard measure for the lifetime distribution, a topic that is of interest in its own right and of central importance to the discussion of compensators and conditional intensity functions in Chapters 7 and 14.

4.1. Basic Properties

Let $X, X_1, X_2, \ldots$ be independent identically distributed nonnegative random variables, and define the partial sums

$$S_0 = 0, \qquad S_n = S_{n-1} + X_n = X_1 + \cdots + X_n \qquad (n = 1, 2, \ldots). \quad (4.1.1)$$

For Borel subsets A of $(0,\infty)$, we attempt to define the counting measure of a point process by setting

$$N(A) = \#\{n\colon S_n \in A\}. \tag{4.1.2}$$

Even if we exclude the trivial case $X = 0$ a.s., as we do throughout this chapter, it may not be completely obvious that (4.1.2) is finite. To see that this is so, observe that for $X \neq 0$ a.s. there must exist positive ε, δ such that

$$\Pr\{X > \varepsilon\} > \delta$$

so that with probability 1 the event $\{X_n > \varepsilon\}$ must occur infinitely often (by the Borel–Cantelli lemmas) and hence $S_n \to \infty$ a.s. It follows that the right-hand side of (4.1.2) is a.s. finite whenever A is bounded, thus justifying the definition (4.1.2). (Here we ignore measurability aspects, for which see Chapter 9.) The process so defined is the (*ordinary*) *renewal process*.

In the notation and terminology of Chapter 3, provided $X_1 > 0$, we have $t_i = S_i$ and $\tau_i = X_i$ for $i = 1, 2, \ldots$, while the assumption that the $\{X_n\}$ are i.i.d. implies that $N(\cdot)$ is interval stationary. *Orderliness* of the process here means $S_{n+1} > S_n$ for $n = 0, 1, \ldots$; that is, $X_n > 0$ for all $n \geq 0$, all with probability 1. But the probability that $X_n > 0$ for $n = 0, 1, \ldots, N-1$ is equal to $(\Pr\{X > 0\})^N \to 0$ as $N \to \infty$ unless $\Pr\{X > 0\} = 1$. Thus, the process is orderly if and only if $\Pr\{X > 0\} = 1$; that is, if and only if the lifetime distribution has zero mass at the origin.

Taking expectations of (4.1.2) yields the renewal measure

$$U(A) = \mathrm{E}(\#\{n\colon S_n \in A, n = 0, 1, 2, \ldots\}) = \mathrm{E}[N(A)], \tag{4.1.3}$$

an equation that remains valid even if A includes the origin. $U(A)$ is just the *first moment or expectation measure* of $N(\cdot)$.

Writing $F(\cdot)$ for the common lifetime distribution and F^{k*} for its k-fold convolution (which is thus the distribution function for S_k), and immediately abusing the notation by writing $F(\cdot)$ for the measure induced on the Borel sets of $\mathcal{B}_{\mathbb{R}}$ by F, we have

$$U(A) = \mathrm{E}\left(\sum_{k=0}^{\infty} I_{\{S_k \in A\}}\right) = \delta_0(A) + \sum_{k=1}^{\infty} F^{k*}(A). \tag{4.1.4}$$

We note in passing that the higher moments of $N(A)$ can also be expressed in terms of $U(\cdot)$ (see Exercise 4.1.2). The quantity most commonly studied is the cumulative function, commonly called the *renewal function*,

$$U(x) \equiv U([0,x]) = 1 + \sum_{k=1}^{\infty} F^{k*}(x) \qquad (x \geq 0). \tag{4.1.5}$$

Again, $U(x)$ is always finite. To see this, choose any $\delta > 0$ for which $F(\delta) < 1$ (possible since we exclude the case $X = 0$ a.s.). Then, since $F(0-) = 0$, we have for any positive integers i, j and x, $y > 0$,

$$1 - F^{(i+j)*}(x+y) \geq \big(1 - F^{i*}(x)\big)\big(1 - F^{j*}(y)\big),$$

and for $0 < y < x$,

$$F^{i*}(x-y)F^{j*}(y) \le F^{(i+j)*}(x) \le F^{i*}(x)F^{j*}(x).$$

Thus, $F^{k*}(\delta) \le (F(\delta))^k < 1$, and therefore the series in (4.1.5) certainly converges for $x < \delta$. For general x in $0 < x < \infty$, there exists finite positive k for which $x/k < \delta$. For given x and such k, $1 - F^{k*}(x) > [1 - F(x/k)]^k > 0$, so

$$\begin{aligned} U(x) &\le \big(1 + F(x) + \cdots + F^{(k-1)*}(x)\big) \sum_{n=0}^{\infty} F^{nk*}(x) \\ &\le \big(1 + F(x) + \cdots + F^{(k-1)*}(x)\big) / \big(1 - F^{k*}(x)\big) < \infty. \end{aligned}$$

Thus, (4.1.5) converges for all $x > 0$.

Taking Laplace–Stieltjes transforms in (4.1.5), we have for $\mathrm{Re}(\theta) > 0$

$$\chi(\theta) \equiv \int_0^{\infty} \mathrm{e}^{-\theta x}\, \mathrm{d}U(x) = \sum_{k=0}^{\infty} \big(\psi(\theta)\big)^k = \frac{1}{1 - \psi(\theta)}, \tag{4.1.6}$$

where $\psi(\theta) = \int_0^{\infty} \mathrm{e}^{-\theta x}\, \mathrm{d}F(x)$. Equivalently, for $\mathrm{Re}(\theta) > 0$,

$$\psi(\theta) = 1 - 1/\chi(\theta),$$

which shows (using the uniqueness theorem for Laplace–Stieltjes transforms) that U determines F uniquely and hence that there is a one-to-one correspondence between lifetime distributions F and renewal functions U.

From (4.1.5), we have for $x > 0$

$$U(x) = 1 + \int_0^x U(x-y)\, \mathrm{d}F(y), \tag{4.1.7}$$

this being the most important special case of the *general renewal equation*

$$Z(x) = z(x) + \int_0^x Z(x-y)\, \mathrm{d}F(y) \qquad (x > 0), \tag{4.1.8}$$

where the solution function Z is generated by the initial function z. If the function $z(x)$ is measurable and bounded on finite intervals, one solution to (4.1.8) is given by

$$Z_0(x) = z(x) + \sum_{k=1}^{\infty} \int_0^x z(x-y)\, \mathrm{d}F^{k*}(y) = \int_0^x z(x-y)\, \mathrm{d}U(y), \tag{4.1.9}$$

the convergence of the series in the middle member being justified by comparison with (4.1.5).

Using the monotonicity of the relation $z \to Z_0$, we easily see that if $z \ge 0$, (4.1.9) is the minimal nonnegative solution to (4.1.8). In fact, considerably more is true, for if $z(x)$ is merely measurable and bounded on finite inter-

vals, the difference $D(x)$ between any two solutions of (4.1.8) with the same property satisfies

$$D(x) = \int_0^x D(x-y)\,\mathrm{d}F^{k*}(y) \qquad \text{for each } \ k = 1, 2, \ldots;$$

hence, $D(x) \equiv 0$ from the fact that $F^{k*}(x) \to 0$ as $k \to \infty$ and the assumed boundedness of D. We summarize as follows.

Lemma 4.1.I (Renewal Equation Solution)**.** *When $z(x)$ is measurable and bounded on finite intervals, the general renewal equation* (4.1.8) *has a unique measurable solution that is also bounded on finite intervals, and it is given by* (4.1.8)*. In particular, $U(x)$ is the unique monotonic and finite-valued solution of* (4.1.7)*.*

Example 4.1(a) *Exponential intervals.* The lack of memory property of the exponential distribution bequeaths on the renewal process that it generates the additional independence properties of the Poisson process. Suppose specifically that

$$F(x) = 1 - \mathrm{e}^{-\lambda x} \qquad (\lambda > 0,\ 0 \le x < \infty).$$

The renewal function for the corresponding Poisson process is $U(x) = 1 + \lambda x$, as can be checked either by using the transform equation in (4.1.6), by summing the convolution powers as in (4.1.5), or by direct verification in the integral equation in (4.1.7). □

Example 4.1(b) *Forward recurrence time.* We gave below (3.4.15) an expression for the distribution of the forward recurrence time r.v. T_u of a stationary point process. The definition at (3.4.15) does not require stationarity, and in the present case of a renewal process, it can be written as

$$\begin{aligned} T_u &= \inf\{S_n \colon S_n > u\} - u = \inf\{S_n - u \colon S_n - u > 0\} \\ &= \begin{cases} X_1 - u & \text{if } X_1 > u, \\ \inf\{S_n - X_1 \colon S_n - X_1 > u - X_1\} - (u - X_1) & \text{otherwise.} \end{cases} \end{aligned}$$

Now when $X_1 \le u$, T_u has the same distribution as the forward recurrence time r.v. T'_{u-X_1}, defined on the renewal process with lifetime r.v.s $\{X'_n\} \equiv \{X_{n+1}\}$, so

$$\Pr\{T_u > y\} = \Pr\{X_1 > y + u\} + \int_0^u \Pr\{T_{u-v} > y\}\,\mathrm{d}F(v). \tag{4.1.10}$$

But this equation is of the form (4.1.8), with $z(x) = \Pr\{X_1 > y + x\} = 1 - F(y + x)$, so by (4.1.9)

$$\Pr\{T_u > y\} = \int_{0-}^{u} [1 - F(y + u - v)]\,\mathrm{d}U(v). \tag{4.1.11}$$

In particular, putting $y = 0$, we recover the identity that is implicit in (4.1.5),

$$1 = \int_{0-}^{x} [1 - F(x - v)]\,\mathrm{d}U(v) \qquad (\text{all } x \ge 0). \tag{4.1.12}$$

□

EXAMPLE 4.1(c) *Renewal equation with linear solution.* As another important application of (4.1.8), consider the generator $z(\cdot)$ that corresponds to the solution $Z(x) = \lambda x$ (all $x > 0$), assuming such a solution function exists, and that $\lambda^{-1} = \mathrm{E}X_n = \int_0^\infty [1 - F(x)]\,\mathrm{d}x$ is finite. Rearranging (4.1.8) yields

$$z(x) = \lambda x - \lambda \int_0^x (x - y)\,\mathrm{d}F(y) = \lambda \int_0^x [1 - F(y)]\,\mathrm{d}y.$$

We can recognize this expression as the distribution function of the forward recurrence time of a stationary point process. This argument identifies the only initial distribution for which the delayed renewal function is linear. □

We conclude this section with a few brief remarks concerning the more general case where the random variables X_n are not necessarily nonnegative or even one-dimensional; thus we admit the possibility that the X_n are d-dimensional vectors for some integer $d > 1$. In such cases, the sequence $\{S_n\}$ constitutes a *random walk*. Such a walk is said to be *transient* if (4.1.2) is finite for all bounded Borel sets A; otherwise, it is *recurrent*, in which case the walk revisits any nonempty open set infinitely often. Thus, it is only for transient random walks that (4.1.2) can be used to define a point process, which we shall call the *random walk point process*. In $\mathbb{R}^1$, it is known that a random walk is transient if the mean $\mathrm{E}(X)$ is finite and nonzero; if $\mathrm{E}(X)$ exists but $\mathrm{E}(X) = 0$, the random walk is recurrent. If the expectation is not defined (the integral diverges), examples of both kinds can occur. In $\mathbb{R}^2$, the random walk can be transient even if $\mathrm{E}(X) = 0$, but only if the variance is infinite. In higher dimensions, every random walk is transient unless perhaps it is concentrated on a one- or two-dimensional subspace. Proofs and further details are given, for example, in Feller (1966).

Most of the renewal equation results also carry over to this context with only nominal changes of statement but often more difficult proofs. Thus, the expectation or renewal measure may still be defined as in (4.1.4), namely

$$U(A) = \delta_0(A) + \sum_{k=1}^{\infty} F^{k*}\{A\}, \tag{4.1.4′}$$

and is finite for bounded Borel sets whenever the random walk is transient (but not otherwise, at least if A has nonempty interior). Furthermore, if $z(x)$ is bounded, measurable, and vanishes outside a bounded set, we may consider the function

$$Z_0(x) = z(x) + \sum_{k=1}^{\infty} \int_{\mathbb{R}^d} z(x - y)\,F^{k*}(\mathrm{d}y) = \int_{\mathbb{R}^d} z(x - y)\,U(\mathrm{d}y), \tag{4.1.13}$$

which is then a solution, bounded on finite intervals, of the *generalized renewal equation*

$$Z(x) = z(x) + \int_{\mathbb{R}^d} Z(x - y)\,F(\mathrm{d}y). \tag{4.1.14}$$

Note that in (4.1.8) we were constrained not only to distributions $F(\cdot)$ concentrated on the half-line but also to functions $z(x)$ and solutions $Z(x)$ that could be taken as zero for $x < 0$. Without such constraints, the proof of uniqueness becomes considerably more subtle: one possible approach is outlined in Exercise 4.1.4. Note too that both (4.1.13) and (4.1.14) remain valid on replacing the argument x by a bounded Borel set A, provided $Z(\cdot)$ is then a set function uniformly bounded under translation for such A.

EXAMPLE 4.1(d) *Random walks with symmetric stable distributions.* Here we define the symmetric stable distributions to be those distributions in $\mathbb{R}$ with characteristic functions of the form

$$\phi_\alpha(s) = \exp(-c|s|^\alpha) \qquad 0 < \alpha \leq 2.$$

Let us consider the associated random walks for the cases $\alpha \leq 1$ for which the first moment does not exist. The case $\alpha = 1$ corresponds to the Cauchy distribution with density function for some finite positive c

$$f(x) = \frac{c}{\pi(c^2 + x^2)} \qquad (-\infty < x < \infty).$$

The nth convolution is again a Cauchy distribution with parameter $c_n = nc$. If the renewal measure were well defined, we would expect it to have a *renewal density*

$$u(x) = \sum_{n=1}^{\infty} f^{n*}(x) = \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{cn}{c^2n^2 + x^2}\,.$$

The individual terms are $O(n^{-1})$ as $n \to \infty$, so the series diverges. It follows readily that the first-moment measure is infinite, so the associated random walk is *recurrent*.

For $\alpha < 1$, it is difficult to obtain a convenient explicit form for the density, but standard results for stable distributions imply that f^{n*} and f differ only by a scale factor,

$$f_\alpha^{n*}(x) = n^{-1/\alpha} f_\alpha(xn^{-1/\alpha}),$$

so that, assuming f_α is continuous at zero,

$$f_\alpha^{n*}(x) \sim xn^{-1/\alpha} f_\alpha(0).$$

Thus, the series is convergent for $1/\alpha > 1$ (i.e. for $\alpha < 1$), and divergent otherwise, so the associated random walk is transient only for $\alpha < 1$. □

EXAMPLE 4.1(e) *A renewal process in two dimensions.* We consider independent pairs (X_n, Y_n) where each pair has a bivariate exponential distribution with density vanishing except for $x \geq 0$, $y \geq 0$, where

$$f(x,y) = \frac{\lambda_1\lambda_2}{1-\rho} \exp\left(\frac{\lambda_1 x + \lambda_2 y}{1-\rho}\right) I_0\left(\frac{2(\rho\lambda_1\lambda_2 xy)^{1/2}}{1-\rho}\right),$$

λ_1, λ_2, and ρ are positive constants, $0 \le \rho < 1$, and $I_n(x)$ is the modified Bessel function of order n defined by the series

$$I_n(x) = \sum_{k=0}^{\infty} \frac{(x/2)^{2k+n}}{k!\,(k+n)!}\,. \tag{4.1.15}$$

The marginal distributions are exponential with parameters λ_1, λ_2; ρ is the correlation between X_1 and Y_1; and the joint distribution has bivariate Laplace–Stieltjes transform

$$\psi(\theta,\phi) = \{(1+\theta/\lambda_1)(1+\phi/\lambda_2) - \rho\theta\phi/\lambda_1\lambda_2\}^{-1}.$$

Much as in the one-dimensional case, the renewal function can be defined as

$$U(x,y) = \mathrm{E}(\#\{n\colon S_n \le x,\ T_n \le y\}),$$

where $S_n = \sum_{k=1}^n X_k$ and $T_n = \sum_{k=1}^n Y_k$ and has Laplace–Stieltjes transform $\chi(\theta,\phi)$ given by

$$\chi(\theta,\phi) = \frac{1}{1-\psi(\theta,\phi)}\,.$$

Substituting for $\psi(\theta,\phi)$ and simplifying, we obtain

$$\chi(\theta,\phi) - 1 = [\theta/\lambda_1 + \phi/\lambda_2 + (1-\rho)\theta\phi/\lambda_1\lambda_2]^{-1},$$

corresponding to the renewal density

$$u(x,y) = \frac{\lambda_1\lambda_2}{1-\rho}\,\exp\left(-\frac{\lambda_1 x + \lambda_2 y}{1-\rho}\right) I_0\left(\frac{2(\lambda_1\lambda_2 xy)^{1/2}}{1-\rho}\right) \qquad (x>0,\ y>0).$$

It should be noted that while the renewal density has uniform marginals, corresponding to the fact that each marginal process is Poisson, the bivariate renewal density is far from uniform, and in fact as $x \to \infty$ and $y \to \infty$, it becomes relatively more and more intensely peaked around the line $\lambda_1 x = \lambda_2 y$, as one might anticipate from the central limit theorem.

The example is taken from Hunter (1974a, b), where more general results can be found together with a bibliography of earlier papers on bivariate renewal processes. See also Exercise 4.1.5. □

Exercises and Complements to Section 4.1

4.1.1 (a) Using a sandwich argument and the strong law of large numbers for the i.i.d. sequence of lifetimes, prove that $N(x)/x \to \lambda$ a.s. as $x \to \infty$.

(b) Deduce from (a) the *Elementary Renewal Theorem*: The renewal function $U(x)$ satisfies $U(x)/x \to \lambda$ as $x \to \infty$, i.e. $U(x) \sim \lambda x$. [*Hint*: See Smith (1958) and Doob (1948). This is not the only possible proof.]

(c) Similarly, if the lifetime distribution has finite second moment with variance σ^2, deduce from the central limit theorem for the X_n that as $x \to \infty$, $(N(x) - \lambda x)/\lambda\sigma\sqrt{\lambda x}$ converges in distribution to a standard $N(0,1)$ random variable. [*Hint*: $N(x) \ge n$ if and only if $S_n \le x$, and if $n, x \to \infty$ such that $(x - n/\lambda)/(\sigma\sqrt{n}\,) \to z$ for finite z, then $\lambda x/n \to 1$.]

4.1.2 *Higher moments of the number of renewals.*

(a) Show that for $0 < x < y < \infty$,

$$\mathrm{E}[N(\mathrm{d}x)\,N(\mathrm{d}y)] = U(\mathrm{d}x)\,U(\mathrm{d}y - x),$$

where U is the renewal measure. Similarly, for any finite sequence $0 < x_1 < x_2 < \cdots < x_k < \infty$,

$$\mathrm{E}[N(\mathrm{d}x_1)\cdots N(\mathrm{d}x_k)] = U(\mathrm{d}x_1)\,U(\mathrm{d}x_2 - x_1)\cdots U(\mathrm{d}x_k - x_{k-1}).$$

[These are differential forms for the moment measures. When the densities exist, they reduce to the moment or product densities as discussed in Chapter 5; see, in particular, Example 5.4(b).]

(b) Prove directly that $\mathrm{E}[(N(0,x])^{[k]}] \le k!\,[U_0(x)]^k < \infty$, where $n^{[k]} = n(n-1)\cdots(n-k+1)$ and $U_0(x) = U(x) - 1$.

(c) In terms of the renewal function $U(x)$, use (a) to show that

$$\mathrm{E}[(N[0,x])^2] = U(x) + 2\int_{0-}^{x} U_0(x-y)\,\mathrm{d}U(y)$$

and hence that when the renewal process is simple,

$$\operatorname{var} N[0,x] = \operatorname{var} N(0,x] = U_0(x) + 2\int_{0+}^{x} [U_0(x-y) - U_0(y)]\,\mathrm{d}U_0(y).$$

Check that in the case of a Poisson process at rate λ, $\mathrm{E}[(N[0,x])^2] = 1 + 3\lambda x + \lambda^2 x^2$ and $\operatorname{var} N(0,x] = \lambda x$.

4.1.3 Let $Q(z;x) = \sum_{n=0}^{\infty} z^n \Pr\{N[0,x] \ge n\}$. Show that

$$Q(z;x) = 1 + z\int_0^x Q(z;x-y)\,\mathrm{d}F(y)$$

and hence that the Laplace–Stieltjes transform is given by

$$\widetilde{Q}(z;\theta) = \int_{0-}^{\infty} \mathrm{e}^{-\theta x}\,\mathrm{d}_x Q(z;x) = \frac{1}{1 - z\psi(\theta)}\,,$$

where $\psi(\theta)$ is the Laplace–Stieltjes transform of F. Obtain corresponding results for the p.g.f. $P(z;x) = \sum_{n=0}^{\infty} z^n \Pr\{N[0,x] = n\}$. Deduce that the factorial moment $\mathrm{E}[(N[0,x])^{[k]}]$ is the k-fold convolution of $U(x) - 1$.

4.1.4 For the one-dimensional random walk with nonlattice step distribution F, prove that the only bounded measurable solutions of the equation

$$D(x) = \int_{-\infty}^{\infty} D(x-y)\,F(\mathrm{d}y)$$

are constant. An outline of one method is as follows.

(1°) Let $Y_n = D(-S_n)$, where $S_n = \sum_{i=1}^{n} X_1$. Use the equation to show that for any bounded measurable solution D, the random variables $\{Y_n\}$ constitute a bounded martingale (see Appendix 3) and hence converge a.s. to some limit random variable Y_∞.

(2°) Since Y_∞ is defined on the tail σ-algebra of the i.i.d. sequence $\{X_n\}$, it must be degenerate; that is, $Y_\infty = c$ for some finite real number c.

(3°) Since for all X_1' independent of S_n, $D(-X_1' - S_n) =_d D(-S_{n+1}) \to c$ a.s., deduce that

$$\mathrm{E}(D(-X_1' - S_n) \mid X_1') \to c$$

and hence, using the equation again, that $D(-X_1') = c$ a.s., whence also $D(-S_n) = c$ a.s. for $n = 1, 2, \ldots$. Thus, finally, $D(x) = c$ a.e. whenever X has a nonlattice distribution.

[*Hint*: See Doob, Snell and Williamson (1960); for an alternative proof, see Feller (1966, Section XI.2), and for a review, see Rao and Shanbhag (1986).]

4.1.5 *Two-dimensional renewal process.* In the context of Example 4.1(e), let $N(x, y) = \#\{n : S_n \le x,\ T_n \le y\}$, where $S_n = \sum_{i=1}^n X_i$ and $T_n = \sum_{i=1}^n Y_i$, and put

$$Q(z; x, y) = \sum_{n=0}^{n} z^n \Pr\{N(x, y) \ge n\},$$

$$P(z; x, y) = \sum_{n=0}^{\infty} z^n \Pr\{N(x, y) = n\}.$$

Extend the result of Exercise 4.1.3 to show that the double Laplace–Stieltjes transform of $P(z; x, y)$ is given by

$$\tilde{P}(z; \theta, \phi) = \frac{1 - \psi(\theta, \phi)}{1 - z\psi(\theta, \phi)}, \qquad \psi(\theta, \phi) = \int_0^\infty \int_0^\infty \mathrm{e}^{-\theta x - \phi y} \, \mathrm{d}_{x,y} F(x, y).$$

For the particular bivariate exponential distribution in Example 4.1(e), the renewal measure has the density $\sum_{n=1}^\infty f^{n*}$, where for $x, y > 0$,

$$f^{n*}(x, y) = f(x, y) \left(\frac{\zeta}{\rho}\right)^{n-1} \frac{I_{n-1}(2\zeta/(1-\rho))}{I_0(2\zeta/(1-\rho))}, \qquad \zeta = \sqrt{\rho\lambda_1\lambda_2 xy}\,.$$

4.2. Stationarity and Recurrence Times

A *modified* or *delayed renewal process*, $\{S_n'\}$ say, is defined much as in (4.1.1) but with X_1 replaced by X_1', which is independent of, but not necessarily identically distributed with, the remaining variables $X_2, X_3, \ldots$. Let $F_1(x) = \Pr\{X_1' \le x\}$. Then, in terms of a forward recurrence time r.v. T_u for a renewal process as in Example 4.1(b), the forward recurrence time r.v. T_u' for such a process $\{S_n'\}$ is defined by $T_u' = \inf\{S_n' : S_n' > u\} - u$ and satisfies

$$T_u' =_d \begin{cases} X_1' - u & \text{if } X_1' > u, \\ T_{u - X_1'} & \text{otherwise,} \end{cases} \tag{4.2.1}$$

hence (see (4.1.10))

$$\Pr\{T_u' > y\} = 1 - F_1(y + u) + \int_0^u \Pr\{T_{u-v} > y\} \, \mathrm{d}F_1(v). \tag{4.2.2}$$

The most important delayed renewal process arises when X_1' has the probability density function

$$f_1(x) = \lambda\big(1 - F(x)\big) \qquad \big(x \geq 0,\ \lambda^{-1} = \mathrm{E}(X)\big), \tag{4.2.3}$$

for then the resulting point process in $(0, \infty)$, with counting measure

$$N'(A) = \#\{n\colon S_n' \in A\},$$

is stationary, as we might anticipate from (3.4.16) and Example 4.1(c). Note that here we are dealing with stationarity on the half-line, in the sense that Definition 3.2.I is required to hold only for Borel subsets of $(0, \infty)$ and for shifts $t \geq 0$.

To establish this stationarity property more formally, define another delayed renewal process, $\{S_n''\}$ say, with initial lifetime r.v. $X_1'' = T_u'$ that is followed by a further sequence of i.i.d. random variables with common d.f. F. Stationarity of $\{S_n'\}$ is proved by showing that the distributions of the two sequences $\{S_n'\}$ and $\{S_n''\}$ coincide. From the assumed independence and distributional properties, it is enough to show that the distributions of the two initial intervals X_1' and X_1'' coincide; i.e. $\Pr\{X_1' > y\} = \Pr\{T_u' > y\}$ for all nonnegative u and y. Using (4.2.2) and (4.1.11), $\Pr\{T_u' > y\}$ equals

$$\lambda \int_{y+u}^{\infty} [1 - F(x)]\,\mathrm{d}x + \int_0^u \left[\int_{0-}^{u-v} [1 - F(y + u - v - w)]\,\mathrm{d}U(w) \right] \lambda[1 - F(v)]\,\mathrm{d}v, \tag{4.2.4}$$

and the last term here equals

$$\begin{aligned}
&\lambda \int_{0-}^{u} \mathrm{d}U(w) \int_0^{u-w} \big(1 - F(v)\big)\big(1 - F(y + u - v - w)\big)\,\mathrm{d}v \\
&\quad = \lambda \int_0^u \mathrm{d}U(w) \int_0^{u-w} \big(1 - F(u - w - v)\big)\big(1 - F(y + v)\big)\,\mathrm{d}v \\
&\quad = \lambda \int_0^u \big(1 - F(y + v)\big)\,\mathrm{d}v \int_0^{u-v} \big(1 - F(u - v - w)\big)\,\mathrm{d}U(w) \\
&\quad = \lambda \int_0^u \big(1 - F(y + v)\big)\,\mathrm{d}v, \qquad \text{using (4.1.12).}
\end{aligned}$$

Substituting back in (4.2.4) and simplifying leads by (4.2.3) to $\Pr\{T_u' > y\}$ $= \lambda \int_y^{\infty} [1 - F(x)]\,\mathrm{d}x = \Pr\{X_1' > y\}$, as required.

These remarks prove the first part of the following proposition (see Exercise 4.2.2 for an alternative proof of this part).

Proposition 4.2.I. *If the lifetime d.f. has finite first moment λ^{-1}, then the delayed renewal process with initial density* (4.2.3) *is stationary, and for all $u > 0$ the forward recurrence time T_u' has this density. If the mean of the lifetime distribution is infinite, then no delayed renewal process with this lifetime distribution can be stationary.*

Proof. To prove the last statement, start by noting from the key renewal theorem, proved later in Proposition 4.4.II, that the forward recurrence time r.v. T_u for a renewal process $\{S_n\}$ whose lifetime distribution has infinite

mean satisfies (see also Example 4.4(a))

$$\text{for every finite } y, \qquad \lim_{u\to\infty} \Pr\{T_u \le y\} = 0.$$

Then, by dominated convergence, letting $u \to \infty$ in (4.2.2) shows that, irrespective of the distribution F_1 of X_1', $\Pr\{T_u' > y\} \to 1$ for every y, so no stationary form for the distribution of T_u' is possible. □

The intuitive interpretation of the last somewhat paradoxical limit statement is that if $\lambda^{-1} = \infty$, we shall spend an ever greater proportion of time traversing intervals of exceptional length and find ourselves in a situation where the current interval has a length greater than y still to run.

Now recall from Exercise 3.4.1 the definition of a *backward recurrence time* r.v. B_u as a companion to the forward recurrence time r.v. T_u:

$$T_u = \inf\{y\colon N(u, u+y] > 0\}, \qquad B_u = \inf\{x\colon N(u-x, u] > 0\}. \tag{4.2.5}$$

Note that there is an asymmetry in the definitions of B_u and T_u: because $N(\cdot)$ is a.s. finite on bounded intervals, $T_u > 0$ a.s. but it is quite possible to have $\Pr\{B_u = 0\} > 0$. The current lifetime r.v. L_u can then be defined by

$$L_u \equiv B_u + T_u\,.$$

The joint distribution of any two of these r.v.s thus gives the distribution of all three: the simplest is that of B_u and T_u for which, when $N(\cdot)$ is stationary and orderly,

$$\begin{aligned}\Pr\{B_u > x,\, T_u > y\} &= \Pr\{N(u-x, u+y] = 0\}\\ &= \Pr\{N(u, u+x+y] = 0\}\\ &= \Pr\{T_u > x+y\} = \lambda \int_{x+y}^{\infty} \big(1 - F(v)\big)\,\mathrm{d}v.\end{aligned} \tag{4.2.6}$$

Note that under stationarity and orderliness, B_u has the same marginal d.f. as T_u, while

$$\begin{aligned}\Pr\{L_u > z\} &= \int_0^z \Pr\{T_u > z - x,\ B_u \in (x, x+\mathrm{d}x)\} + \Pr\{B_u > z\}\\ &= \int_0^z \lambda\big(1 - F(x+z-x)\big)\,\mathrm{d}x + \int_z^{\infty} \lambda\big(1 - F(v)\big)\,\mathrm{d}v\\ &= \lambda \int_0^{\infty} \big(1 - F(\max(v, z))\big)\,\mathrm{d}v.\end{aligned} \tag{4.2.7}$$

Thus,

$$\mathrm{E}L_u = 2\mathrm{E}T_u = 2\mathrm{E}B_u = \lambda \mathrm{E}X^2 = \mathrm{E}X^2/\mathrm{E}X \ge \mathrm{E}X, \tag{4.2.8}$$

with equality only in the case where $X = \mathrm{E}X$ a.s.; that is, all lifetimes are equal to the same constant, when the renewal process is variously called a *deterministic renewal process* or a *process of equidistant points*.

By identifying $1 - F(\cdot)$ with $q_0(\cdot)$ in (3.4.9), equations (4.2.6–8) continue to hold for any stationary orderly point process as discussed in Section 3.4.

Without the assumption of stationarity, we may use the alternative definition for B_u,

$$B_u = u - \sup\{S_n : S_n \le u\} \qquad (u \ge 0\}.$$

Arguing as in (4.1.10), it is not difficult to show (see Exercise 4.2.1) that for the basic renewal process $\{S_n\}$,

$$\Pr\{B_u > x,\, T_u > y\} = \int_0^{(u-x)_+} \big(1 - F(u+y-v)\big)\, \mathrm{d}U(v). \tag{4.2.9}$$

In the case of a Poisson process, we have $F(x) = 1 - \mathrm{e}^{-\lambda x}$, and it is then not difficult to check from these relations that

$$\mathrm{E}X < \infty \text{ and the distribution of } T_u \text{ is independent of } u; \tag{4.2.10a}$$

$$\mathrm{E}X < \infty \text{ and } B_u \text{ and } T_u \text{ are independent for each } u > 0; \tag{4.2.10b}$$

$$\mathrm{E}T_u < \infty \text{ (all } u\text{) and is independent of } u. \tag{4.2.10c}$$

Properties such as (4.2.10) have been used to characterize the Poisson process amongst renewal processes, as detailed in part in Galambos and Kotz (1978). For example, when $\mathrm{E}T_u < \infty$, integration of (4.1.10) shows that

$$\mathrm{E}T_u = \int_u^\infty \big(1 - F(y)\big)\, \mathrm{d}y + \int_0^u \mathrm{E}(T_{u-v})\, \mathrm{d}F(v),$$

so that when (4.2.10c) holds,

$$\big(1 - F(u)\big)\mathrm{E}T_u = \big(1 - F(u)\big)\mathrm{E}T_0 = \int_u^\infty \big(1 - F(y)\big)\, \mathrm{d}y \qquad (\text{all } u > 0).$$

Thus, $F(y) = 1 - c\,\mathrm{e}^{-\lambda y}$ for some constant $c = 1 - F(0+)$; since $F(0+) = 0$ for an orderly renewal process, $c = 1$. The proof of the rest of Proposition 4.2.II is indicated in Exercises 4.2.3–4.

Proposition 4.2.II. *Any one of the statements* (4.2.10a), (4.2.10b), *and* (4.2.10c) *characterizes the Poisson process amongst orderly renewal processes.*

Exercises and Complements to Section 4.2

4.2.1 By following the argument leading to (4.2.3), show that for an orderly renewal process $N(\cdot)$ for which $N(\{0\}) = 1$ a.s.,

$$\begin{aligned}\Pr\{B_u > x,\, T_u > y\} &= \Pr\{N(u-x, u+y] = 0\} \\ &= \int_{0-}^{(u-x)_+} \big[1 - F(y+u-v)\big]\, \mathrm{d}U(v), \\ \Pr\{L_u > z\} &= \int_{0-}^{u} \big[1 - F\big(\max(z, u-v)\big)\big]\, \mathrm{d}U(v).\end{aligned}$$

4.2.2 Suppose that the delayed renewal process $\{S_n'\}$ with counting function $N(\cdot)$ and lifetime distribution $F(\cdot)$ with finite mean λ^{-1} is stationary. Show that X_1' must have the density (4.2.3). [*Hint*: Stationarity implies that $\mathrm{E}N(0, x] = \lambda x$ (all $x > 0$); now use Example 4.1(c).]

4.2.3 Use (4.1.10) to show that (4.2.10a) characterizes the Poisson process among orderly renewal processes.

4.2.4 Use (4.2.9) with $x \uparrow u$ to deduce that when (4.2.10b) holds,

$$\Pr\{T_u > y\} = \frac{1 - F(y+u)}{1 - F(u)}$$

for each u and $y \geq 0$. Consequently, for all v in the support of $U(\cdot)$,

$$[1 - F(0+)][1 - F(y+v)] = [1 - F(y)][1 - F(v)],$$

so that $F(\cdot)$ is either geometric or exponential. If $F(x)$ is constant for $0 < x < \delta$, then B_u and T_u cannot be independent—hence the characterization in Proposition 4.2.II via (4.2.10b).

4.2.5 For a renewal process with lifetime d.f. $F(x) = 1 - (1 + \mu x)\mathrm{e}^{-\mu x}$, evaluate the renewal function as

$$U(x) = 1 + \tfrac{1}{2}\mu x - \tfrac{1}{4}(1 - \mathrm{e}^{-2\mu x})$$

and hence derive the d.f.s of the forward and backward recurrence time r.v.s T_u and B_u. Verify their asymptotic properties for $u \to \infty$.

4.3. Operations and Characterizations

Because a single d.f. F suffices to describe a renewal or stationary renewal process, it is of interest to ask in various contexts involving the manipulation of point processes what conditions lead again to a renewal process as a result of the transformation or operation concerned. More often than not, the solution to such a question is a characterization of the Poisson process, a conclusion that can be disappointing when it might otherwise be hoped that more general renewal processes could be realized. Roughly speaking, when such a Poisson process characterization solution holds, it indicates that the interval independence property of a renewal process can be preserved only as a corollary of the stronger lack-of-memory property of the Poisson process. We have already given examples of characterizations of the Poisson process in Proposition 4.2.II. The three operations considered in this section concern thinning, superposition, and infinite divisibility.

Example 4.3(a) *Thinning of renewal processes.* Given a renewal process $\{S_n\}$, let each point S_n for $n = 1, 2, \ldots$ be omitted from the sequence with probability $1 - \alpha$ and retained with probability α for some constant α in $0 < \alpha < 1$, each such point S_n being treated independently. This independence property means that if $\{S_{n(r)}, r = 1, 2, \ldots\}$ is the sequence of retained points with $0 = n(0) < n(1) < n(2) < \ldots$, then $N_r \equiv n(r) - n(r-1)$ is a family of i.i.d. positive integer-valued r.v.s with $\Pr\{N_r = j\} = \alpha(1-\alpha)^{j-1}$ for $j = 1, 2, \ldots$, and hence

$$\{Y_r\} \equiv \{S_{n(r)} - S_{n(r-1)}\} \tag{4.3.1}$$

is a family of i.i.d. r.v.s with d.f.

$$\Pr\{Y_r \le x\} = \sum_{j=1}^{\infty} \alpha(1-\alpha)^{j-1} F^{j*}(x).$$

Consequently, $\{S_{n(r)}\}$ is still a renewal process, and it is not hard to verify that its renewal function, U_α say, is related to that of $\{S_n\}$ by rescaling as in

$$U_\alpha(x) - 1 = \alpha\big(U(x) - 1\big). \tag{4.3.2}$$

It is readily seen that whenever $\{N_r\}$ here is a family of i.i.d. positive integer-valued r.v.s, $\{S_{n(r)}\}$ is a renewal process, but it is only for the geometric distribution for N_r that (4.3.2) holds. In connection with this equation, the converse question can be asked as to when it can be taken as defining a renewal function for $\alpha > 1$. In general, for a given renewal function U, there is a finite largest $\alpha \ge 1$ for which $1+\alpha(U(x)-1)$ is a renewal function, although there is a class of lifetime d.f.s, including the exponential and others besides, for which $1 + \alpha(U(x) - 1)$ is a renewal function for all finite positive α [Daley (1965); see also van Harn (1978) and Exercise 4.3.1].

Any renewal function U satisfies $U(x)/\lambda x \to 1$ as $x \to \infty$, and consequently the renewal function U_α of the thinned renewal process $\{S_{n(r)}\}$, when rescaled so as to have the same mean lifetime, becomes U_α^s, say, defined by

$$U_\alpha^s(x) - 1 = \alpha\big(U(x/\alpha) - 1\big) \to \lambda x \qquad (\alpha \downarrow 0).$$

Thus, if U_α^s is independent of α, it must equal the renewal function of a Poisson process, which is therefore the only renewal process whose renewal function is preserved under thinning and rescaling, i.e. $U_\alpha^s = U$ (all $0 < \alpha < 1$). □

Example 4.3(b) *Superposition of renewal processes.* Let $N_1, \ldots, N_r$ be independent nontrivial stationary renewal processes. When is the superposed process

$$N = N_1 + \cdots + N_r \tag{4.3.3}$$

again a renewal process? Certainly, N is a renewal process (indeed a Poisson process) when each of the components $N_1, \ldots, N_r$ is a Poisson process. Conversely, since by Raikov's theorem (e.g. Lukacs, 1970) independent random variables can have their sum Poisson-distributed only if every component of the sum is Poisson-distributed also, it follows from writing $N(A) = N_1(A) + \cdots + N_r(A)$ (all Borel sets A) and appealing to Renyi's characterization in Theorem 2.3.II that if N is a Poisson process, then so also is each N_j. Because a renewal process is characterized by its renewal function, and this is linear only if the process is Poisson, one way of proving each of the two assertions below is to show that the renewal function concerned is linear. □

Proposition 4.3.I. *A stationary renewal process is the superposition of two independent nontrivial stationary renewal processes only if the processes are Poisson.*

Proposition 4.3.II. *A stationary renewal process is the superposition of $r \geq 2$ independent identically distributed stationary renewal processes only if the processes are Poisson.*

PROOF. We start by allowing the renewal processes N_j to have possibly different lifetime d.f.s F_j, denoting each mean by λ_j^{-1}, so by Proposition 4.1.I, each λ_j is finite and positive. Write $\lambda = \lambda_1 + \cdots + \lambda_r$, $p_j = \lambda_j/\lambda$, $\pi_j = F_j(0+)$, and $\pi = F(0+)$, where F is the lifetime d.f. of the superposed process N. For any such renewal process, we have, for small $h > 0$ and $|z| \leq 1$,

$$\begin{aligned}\mathrm{E}\big(z^{N(0,h)}\big) &= 1 - \frac{\lambda h(1-z)}{(1-\pi)(1-z\pi)} + o(h) \\ &= \prod_{j=1}^{r} \mathrm{E}\big(z^{N_j(0,h)}\big) = \prod_{j=1}^{r} \left(1 - \frac{\lambda_j h(1-z)}{(1-\pi_j)(1-z\pi_j)} + o(h)\right).\end{aligned}$$

It follows by equating powers of z that for $i = 1, 2, \ldots,$

$$\lim_{h\downarrow 0} \Pr\{N(0,h] = i \mid N(0,h] > 0\} = \pi^{i-1}(1-\pi) = (1-\pi)\lambda^{-1} \sum_{j=1}^{r} \lambda_j \pi_j^{i-1}\,.$$

All these equations can hold for nonzero π and π_j (and nonzero λ) only if $\pi = \pi_j$ for $j = 1, \ldots, r$; that is, only if all renewal processes concerned have the same probability of zero lifetimes. Consequently, it is enough to establish the propositions in the orderly case, which we assume to hold from here on.

In place of the renewal function U in (4.1.5), we use

$$H(x) = \sum_{n=1}^{\infty} F^{n*}(x), \quad \text{so } H(x) = \lambda x \ \text{ for a Poisson process.} \tag{4.3.4}$$

Then, from (3.5.3), for a stationary renewal process N,

$$\begin{aligned}\operatorname{var} N(0,x) = \operatorname{var} N(0,x] &= \lambda \int_0^x [2H(u) + 1]\,\mathrm{d}u - (\lambda x)^2 \\ &= \lambda \int_0^x \big(2[H(u) - \lambda u] + 1\big)\,\mathrm{d}u \equiv V(x)\end{aligned}$$

and thus

$$\begin{aligned}\operatorname{cov}\big(N[-x,0),\, N(0,y]\big) &= \tfrac{1}{2}\big(V(x+y) - V(x) - V(y)\big) \\ &= \lambda \int_0^y \big(G(x+u) - G(u)\big)\,\mathrm{d}u,\end{aligned}$$

where $G(x) = H(x) - \lambda x$. It is convenient to write below, for r.v.s Y for which the limits exist,

$$\mathrm{E}^0(Y) = \lim_{h\downarrow 0} \mathrm{E}(Y \mid N(0,h] > 0).$$

Since $p_j = \lim_{h\downarrow 0} \Pr\{N_j(0,h] > 0 \mid N(0,h] > 0\}$,

$$H(x) = \mathrm{E}^0(N(0,x] \mid N(\{0\}) > 0)$$
$$= \lim_{h\to 0} \sum_{j=1}^{r} \mathrm{E}^0\Bigg(\sum_{i=1}^{r} N_i(0,x] \;\Big|\; N_j(\{0\}) > 0 \Bigg) \frac{\Pr\{N_j(-h,0] > 0\}[1+o(1)]}{\Pr\{N(-h,0] > 0\}}$$
$$= \sum_{j=1}^{r} \Bigg(p_j H_j(x) + p_j \sum_{i\neq j} \lambda_i x \Bigg), \tag{4.3.5}$$

so $G(x) = \sum_{j=1}^{r} p_j G_j(x)$. Similar, somewhat lengthier, algebra leads to

$$G(x,y) \equiv \lim_{h\to 0} \mathrm{E}^0\big(\big(N(-x,0) - \lambda x\big)\big(N(0,y) - \lambda y\big) \mid N(\{0\}) > 0\big)$$
$$= \sum_{j=1}^{r} p_j G_j(x,y) + \lambda \int_0^y \Bigg(G(x+u) - G(u) - \sum_{j=1}^{r} p_j^2 \big(G_j(x+u) - G_j(u)\big) \Bigg)\, \mathrm{d}u.$$

Thus, when $N_1, \ldots, N_r$ are identically distributed, $p_j = 1/r$, $G_j(x) = G_1(x)$ (all j), and $G_1(x) = G(x)$. Also, for a renewal process, $G(x,y) = G(x)G(y)$, so

$$G(x)G(y) = G(x)G(y) + \lambda(1-1/r) \int_0^y \big(G(x+u) - G(u)\big)\, \mathrm{d}u.$$

It follows that $G(x+y) = G(y) = G(0)$ (all $x, y > 0$). Thus, $H(x) = \lambda x$, and Proposition 4.3.II is proved.

On the other hand, for $r = 2$ and possibly different F_1 and F_2, replacing $G(x,y)$ by $G(x)G(y)$ with $G(x) = p_1 G_1(x) + p_2 G_2(x)$, $p_1 + p_2 = 1$, leads to

$$-p_1 p_2 \big(G_1(x) - G_2(x)\big)\big(G_1(y) - G_2(y)\big)$$
$$= \lambda p_1 p_2 \int_0^y \big(G_1(x+u) + G_2(x+u) - G_1(u) - G_2(u)\big)\, \mathrm{d}u.$$

The function $K(y) \equiv G_1(y) - G_2(y)$ thus has a right-derivative $k(\cdot)$ given by

$$-K(x)k(y) = \lambda\big(G_1(x+y) + G_2(x+y) - G_1(y) - G_2(y)\big).$$

Either $K(x) = 0$, in which case $G_1 = G_2$ and the earlier argument shows that $G(x) = 0$, or else by letting $y \downarrow 0$ and using $G_1(0) = G_2(0) = 0$, it follows that $G_1(x)$ is proportional to $G_2(x)$, with $G_1(x)$ having the derivative $g_1(x)$, say. Consequently,

$$g_1(x)g_1(y) = \alpha g_1(x+y)$$

for some nonzero α, so $g_1(x) = \alpha \mathrm{e}^{-\beta x}$ for some $0 < \beta < \infty$ because $G_1(x)/x \to 0$ as $x \to \infty$. Transform calculus now shows that each $1 - F_j(u) = \mathrm{e}^{-b_j u}$.□

An earlier version of Proposition 4.3.I is in McFadden and Weissblum (1963), and a different proof is in Mecke (1969). Another argument is used in Mecke (1967) to prove the following result (the proof is omitted here).

Proposition 4.3.III. *Let the stationary renewal process N be the superposition of the independent stationary point processes N_1 and N_2 with N_1 renewal. If the lifetime d.f.s F and F_1 of N and N_1 have density functions that are continuous on $(0,\infty)$ and right-continuous at 0, then N_1 is a Poisson process.*

By taking N_1 to be Poisson with rate parameter λ and N_2 to be an alternating renewal process with exponential distributions for the alternating lifetime d.f.s, their parameters α and β being such that $\lambda^2 = \alpha\beta$, Daley (1973a) furnished an example showing that Mecke's result cannot characterize N_2 as a Poisson process. If only the differentiability assumptions could be omitted, the restriction in Proposition 4.3.II that the components N_j of the sum N at (4.3.3) should be identically distributed could be dropped.

EXAMPLE 4.3(c) *Infinite divisibility.* A natural complement to Example 4.3(b) is to ask whether there are any stationary renewal processes other than the Poisson that are infinitely divisible. Here we ask whether for (any or all) integers r, the stationary renewal process N in (4.3.3) is expressible as the superposition of i.i.d. stationary point processes $N_1, \ldots, N_r$. Assuming that the lifetime distribution concerned has a density function, [MKM] state that Häberlund (1975) proved that the Poisson process is the only one, while under the additional assumption of the existence of density functions for all the joint distributions of the component process N_1, Ito (1980) has asserted the stronger result that if N is expressible as $N = N_1 + \cdots + N_r$ for one integer $r \geq 2$, then it is Poisson and hence infinitely divisible. □

There are innumerable characterizations of the exponential distribution and Poisson processes (see reviews in Galambos and Kotz (1978) and Johnson and Kotz (1994, Section 19.8)). Fosam and Shanbhag (1997) has a useful list of papers exploiting variants of the Choquet–Deny functional equation approach.

Exercises and Complements to Section 4.3

4.3.1 (a) When $F(x) = 1 - (1 + x)\mathrm{e}^{-x}$, show (e.g. by using Laplace–Stieltjes transforms) that $1 + \alpha(U(x) - 1)$ is a renewal function if and only if $0 < \alpha \leq 1$.

(b) Let $\{X(t): t \geq 0\}$ be a stochastic process with $X(0) = 0$ and stationary nonnegative independent increments, with Lévy–Khinchin representation $\mathrm{E}(\mathrm{e}^{-\theta X(t)}) = \mathrm{e}^{t\psi(\theta)}$, where

$$\psi(\theta) = -\theta\mu_0 + \int_{(0,\infty)} (\mathrm{e}^{-\theta x} - 1)\, \mu(\mathrm{d}x),$$

with $\mu_0 \geq 0$ and $\mu(\cdot)$ a nonnegative measure on $(0, \infty)$ satisfying $\int_{(0,\infty)} \min(x, 1)\ \mu(\mathrm{d}x) < \infty$, and $\mu(0, \infty) = \infty$ if $\mu_0 = 0$. Let $0 = t_0 < t_1 < \cdots$ be the successive epochs of a Poisson process in $(0, \infty)$ with unit intensity so that the r.v.s $X(t_n) - X(t_{n-1})$ are i.i.d. with d.f. $F(x) = \int_0^\infty F(x, t)\mathrm{e}^{-t}\, \mathrm{d}t$, where $F(x, t) = \Pr\{X(t) \leq x\}$. Show that with $U(\cdot)$ the renewal function corresponding to F and $U_0(x) = U(x) - 1$, $1 + \alpha U_0(x)$ is a renewal function for all $0 < \alpha < \infty$, and that $U_0(x)$ is subadditive (see Kingman, 1972, p. 100).

4.3.2 Let the stationary point process N_1 arise as the jump epochs of a Markov process on countable state space, and let N_2 be a stationary Poisson process independent of N_1. Daley (1975b) showed that for $N \equiv N_1 + N_2$ to be a stationary renewal process different from Poisson, not only must the Markov chain transition rates underlying N_1 have a particular structure but also there is a unique rate λ for N_2 for which N can have the renewal property.

4.4. Renewal Theorems

Considerable effort has been expended in the mathematics of renewal theory on establishing Theorem 4.4.I below and its equivalents; they are stronger statements than the elementary renewal theorem [i.e. the property $U(x) \sim \lambda x$ given in Exercise 4.1.1(b) of which there is a generalization in (3.5.3)]. Theorem 4.4.I is variously known as Blackwell's renewal theorem or the key renewal theorem, depending basically on how it is formulated.

Theorem 4.4.I (Blackwell's Renewal Theorem)**.** *For fixed positive y, restricted to finite multiples of the span of the lattice when the lifetime d.f. is lattice, and otherwise arbitrary,*

$$U(x+y) - U(x) \to \lambda y \qquad (x \to \infty). \tag{4.4.1}$$

Equation (4.4.1) says roughly that the renewal measure ultimately behaves like a multiple of Lebesgue measure. To make this more precise, let $S_t U$ denote the shifted version of the renewal measure U so that

$$S_t U(A) = U(t + A).$$

Then (4.4.1) implies that on any finite interval $(0, M)$, $S_t U$ converges weakly to the multiple $\lambda\ell$ of Lebesgue measure $\ell(\cdot)$ (or, equivalently, $S_t U$ as a whole converges vaguely to $\lambda\ell$; see Section A2.3 for definitions and discussion of weak and vague convergence). Blackwell's theorem represents the 'set' form of the criterion for weak convergence, while the key renewal theorem (Theorem 4.4.II below) represents a strengthened version of the corresponding 'function' form, the strengthening taking advantage of the special character of the limit measure and its approximants.

On the other hand, the theorem is not so strong as to assert anything concerning a density $u(\cdot)$ for U. Such results require further assumptions about the lifetime distributions and are explored, together with further strengthenings of Blackwell's theorem, following Theorem 4.4.II.

Proof of Theorem 4.4.I. The proof given here is probabilistic and uses a coupling method [see Lindvall (1977, 1992) and Thorisson (2000, Section 2.8)]. We compare each sample path $\{S_n\}$ with the sample path $\{S'_n\}$ of a stationary renewal process as defined in Section 4.2, $\{S_n\}$ and $\{S'_n\}$ being defined on a common probability space $(\Omega, \mathcal{F}, P)$ so as to be mutually independent. For each $\omega \in \Omega$, and every integer $i \geq 0$, define for $\{S'_n\}$ the forward recurrence time r.v.s $Z_i\omega = T'_{S_i(\omega)}$ so that

$$Z_i(\omega) = \min\{S'_j(\omega) - S_i(\omega) \colon S'_j(\omega) > S_i(\omega)\}.$$

Because the sequence $\{S_{i+n} - S_i\}$ has a distribution independent of i and is independent of $\{S'_n\}$, and because T'_u is stationary, it follows that the sequence $\{Z_i\}$ is also stationary. Thus, the events

$$A_i \equiv \{Z_j < \delta \ \text{ for some } \ j \geq i\},$$

which we define for any fixed $\delta > 0$, have the same probability for each $i = 0, 1, \ldots$, and in particular therefore $P(A_0) = P(A_\infty)$, where

$$A_0 \supseteq A_1 \supseteq \cdots \supseteq A_\infty \equiv \bigcap_{i=1}^{\infty} A_i = \{Z_i < \delta \text{ i.o.}\}.$$

Now A_∞ is a tail event on the conditional σ-field (namely, conditional on X_1') of the i.i.d. r.v.s $\{X_1, X_1', X_2, X_2', \ldots\}$ and therefore by the zero–one law for tail events (see e.g. Feller, 1966, Section IV.6), for ℓ-a.e. x,

$$P(A_\infty \mid X_1' = x) = 0 \text{ or } 1 \qquad (0 < x < \infty).$$

Because F is nonlattice, $P\{u - x < S_j' - X_1' < u - x + \delta \text{ for some } j\}$ is positive for all sufficiently large u for fixed $\delta > 0$ (see Feller, 1966, Section V.4a, Lemma 2), and hence $P(A_0 \mid X_1' = x) > 0$ for every x. Thus, the equations

$$\begin{aligned} 0 < \lambda \int_0^\infty P(A_0 \mid X_1' = x)[1 - F(x)]\,\mathrm{d}x &= P(A_0) \\ = P(A_\infty) = \lambda \int_0^\infty P(A_\infty \mid X_1' = x)[1 - F(x)]\,\mathrm{d}x & \end{aligned}$$

force $P(A_\infty \mid X_1' = x) = 1$ for every x for which $F(x) < 1$. Hence, $P(A_\infty) = 1 = P(A_0)$, so that for every $\delta > 0$,

$$P\{Z_i < \delta \text{ for some } i\} = 1.$$

To establish (4.4.1), it is enough to show that, for any $\delta > 0$, we can find x_0 such that $x \geq x_0$ implies that $|\mathrm{E}N(x, x+y] - \lambda y| \leq \delta$. Observe that $\lambda y = \mathrm{E}N'(x, x+y]$, where N' is the counting function for the stationary renewal process with intervals $\{X_n'\}$. Let $I_\delta = \inf\{i\colon Z_i < \delta\}$, so that $P\{I_\delta < \infty\} = 1$. Defining

$$J \equiv \inf\{j\colon S_j'(\omega) > S_{I_\delta}(\omega)\},$$

we then have $0 < Z_{I_\delta}(\omega) = S_J'(\omega) - S_{I_\delta}(\omega) < \delta$. Define a new point process by means of the sequence of intervals

$$\{X_1, \ldots, X_{I_\delta}, X_{J+1}', X_{J+2}', \ldots\},$$

and denote its counting function by N'' so that for any Borel set A,

$$\begin{aligned} N''(A) &= N\big(A \cap (0, S_{I_\delta})\big) + N'\big((A + Z_{I_\delta}) \cap (S_j', \infty)\big) \\ &= N\big(A \cap (0, S_{I_\delta})\big) + N'(A + Z_{I_\delta}) - N'\big((A + Z_{I_\delta}) \cap (0, S_j')\big). \end{aligned}$$

When A is the interval $(x, x+y]$, the shifted interval $A + Z_{I_\delta}$ has $\mathrm{E}N'(A + Z_{I_\delta})$ lying between $\lambda(y - \delta)$ and $\lambda(y + \delta)$ because

$$(x + \delta, x + y] \subseteq (x + Z_{I_\delta}, x + y + Z_{I_\delta}] \subseteq (x, x + y + \delta].$$

For every x, the r.v.s $N(x, x+y]$ are stochastically dominated by the r.v. $1 + N(0, y]$, and since this has finite expectation, $\{N(x, x+y]\colon x \geq 0\}$ is a

uniformly integrable family of r.v.s. This ensures that

$$\mathrm{E}\big(N(x,x+y]I_{\{x<S_{I_\delta}\}}\big)\to 0 \qquad \text{as} \quad x\to\infty$$

since then $P\{x<S_{I_\delta}\}\to 0$. Similarly, $N'(x+Z_{I_\delta},x+y+Z_{I_\delta}]$ is stochastically dominated by $1+N(0,y]$ and $P\{x<S'_j\}\to 0$ as $x\to\infty$, so

$$\mathrm{E}(N'(x+Z_{I_\delta},x+y+Z_{I_\delta}]I_{\{x<S'_j\}})\to 0.$$

Consequently, for x sufficiently large, $U(x+y)-U(x)=\mathrm{E}N''(x,x+y]$ is arbitrarily close to $\mathrm{E}N'(A+Z_{I_\delta})$, and since δ is arbitrarily positive, (4.4.1) is established. □

We now turn to an equivalent but very important form of Theorem 4.4.I for nonlattice lifetimes. A function $g(\cdot)$ defined on $[0,\infty)$ is *directly Riemann integrable* there when, for any $h>0$, the normalized sums

$$h\sum_{n=1}^{\infty} g_-^h(nh) \qquad \text{and} \qquad h\sum_{n=1}^{\infty} g_+^h(nh)$$

converge to a common finite limit as $h\to 0$; here,

$$g_-^h(x)=\inf_{0\le\delta\le h} g(x-\delta), \qquad g_+^h(x)=\sup_{0\le\delta\le h} g(x-\delta).$$

Exercise 4.4.1 states sufficient conditions for g to be directly Riemann integrable. For such a function, with $U(x)\equiv 0$ for $x<0$ and monotonically increasing on $x\ge 0$,

$$\int_0^x g(x-y)\,\mathrm{d}U(y) \begin{matrix}\le\\ \ge\end{matrix} \sum_{n=1}^{\infty} g_\pm^h(nh)\big(U\big(x-(n-1)h\big)-U(x-nh)\big).$$

These sums can be truncated to finite sums with truncation error bounded by

$$\begin{aligned}\int_0^{x-C} &|g(x-y)|\,\mathrm{d}U(y)\\ &\le \sum_{n=1}^{[x-C]} |g|_+^1(C+n)\big(U(x+1-C-n)-U(x-C-n)\big)\\ &\le U(1)\sum_{n=1}^{\infty}|g|_+^1(C+n),\end{aligned}$$

which can be made arbitrarily small, uniformly in $x>0$, by taking C sufficiently large. Thus, the sums are approximated by

$$\begin{aligned}\int_{x-C}^{x} g(x-y)\,\mathrm{d}U(y) &\begin{matrix}\le\\ \ge\end{matrix} \sum_{n=1}^{[C/h]} g_\pm^h(nh)[U(x-nh+h)-U(x-nh)]\\ &\to \lambda h\sum_{n=1}^{[C/h]} g_\pm^h(nh) \qquad (x\to\infty)\\ &\to \lambda\int_0^C g(u)\,\mathrm{d}u \qquad (h\to 0).\end{aligned}$$

The following equivalent form of Theorem 4.4.I can now be given.

Theorem 4.4.II (Key Renewal Theorem)**.** *For nonlattice lifetime distributions and directly Riemann integrable functions* $g(\cdot)$,

$$\int_0^x g(x-y)\,\mathrm{d}U(y) \to \lambda \int_0^\infty g(y)\,\mathrm{d}y \qquad (x \to \infty). \tag{4.4.2}$$

Some results for monotonically decreasing but not necessarily integrable functions $g(\cdot)$ are sketched in Exercise 4.4.5(c).

The following examples may serve as prototypes for the application of the renewal theorem to problems of convergence to equilibrium.

EXAMPLE 4.4(a) *Convergence of the forward recurrence time distribution.* Our starting point is (4.1.11), which after subtracting from (4.1.12) can be written

$$F_u(y) \equiv \Pr\{T_u \le y\} = \int_{0-}^{u} [F(y+u-v) - F(u-v)]\,\mathrm{d}U(v). \tag{4.4.3}$$

This is in the form (4.4.2) with $g(x) = F(y+x) - F(x)$. This function is integrable and of bounded variation over the whole half-line; it then follows easily (see Exercise 4.4.1) that the function is directly Riemann integrable, so that the theorem can be applied. It asserts that, provided the lifetime distribution is nonlattice,

$$F_u(y) \;\to\; \lambda \int_0^\infty [F(y+x) - F(x)]\,\mathrm{d}x \;=\; \lambda \int_0^y [1 - F(v)]\,\mathrm{d}v \qquad (u \to \infty).$$

If $\lambda^{-1} < \infty$, this is the usual form of the *length-biased distribution* associated with F, the fact that the distribution is proper following from the identity $1 = \lambda \int_0^\infty \big(1 - F(v)\big)\,\mathrm{d}v$. In this case, (4.4.2) asserts directly that the forward recurrence time distribution converges weakly to its limit form. The extension of this result to a delayed renewal process with arbitrary initial distribution follows then from (4.4.4).

When $\lambda^{-1} = \infty$, $F_u(y) \to 0$ for all y and no stationary form can exist. □

EXAMPLE 4.4(b) *Convergence of the renewal density.* As a further corollary, we shall prove (see Feller, 1966, Section XI.4) that *if the lifetime distribution* F *has finite mean and bounded density* $f(t)$*, then* $U(t)$ *has density* $u(t)$ *such that*

$$u(t) - f(t) \to \lambda. \tag{4.4.4}$$

This follows from the fact that $u(t)$, when it exists, satisfies the renewal equation in its traditional form

$$u(t) = f(t) + \int_0^t u(t-x) f(x)\,\mathrm{d}x.$$

[To check this, note that equation (4.1.9) implies that the solution has the form $u(s) = \int_0^s f(s-x)\,\mathrm{d}U(x)$, which on integrating yields $\int_0^t u(s)\,\mathrm{d}s = U(t) - 1$.]

Moreover, the function

$$u(t) - f(t) = \sum_{k=2}^{\infty} f^{k*}(t)$$

satisfies the renewal equation

$$u(t) - f(t) = f^{2*}(t) + \int_0^t [u(t-x) - f(t-x)]f(x)\,\mathrm{d}x. \tag{4.4.5}$$

Now if $f(t)$ is bounded, $f^{2*}(t)$ is directly Riemann integrable. Indeed, as the convolution of a bounded and an integrable function, it is uniformly continuous (Exercise 4.4.2), while the inequality

$$\begin{aligned} f^{2*}(t) &= \int_0^{t/2} f(t-y)f(y)\,\mathrm{d}y + \int_{t/2}^t f(t-y)f(y)\,\mathrm{d}y \\ &= 2\int_0^{t/2} f(t-y)f(y)\,\mathrm{d}y \;\le\; 2C[1 - F(\tfrac{1}{2}t)], \end{aligned}$$

where $C = \sup|f(t)|$, shows that when $\mu = \lambda^{-1} < \infty$, $f^{2*}(t)$ is also bounded above by an integrable monotonic function and is therefore directly Riemann integrable by Exercise 4.4.1(c). Thus, Proposition 4.4.II applies, yielding (4.4.4).

The argument can be extended to the case where, if not f itself, at least one of its convolution powers has bounded density (see Exercise 4.4.3). □

Even a partial assumption of absolute continuity allows the conclusions of the renewal theorems to be substantially strengthened—for example, from local weak convergence of the renewal measure to local convergence in variation norm, namely

$$\|S_t U - \lambda\ell\|_M \to 0, \tag{4.4.6}$$

where $\|\mu\|_M$ is the variation norm of the (signed) measure μ over $[0, M]$. Equation (4.4.6) would imply that, in Blackwell's theorem, $U(t+A) \to \lambda\ell(A)$ not only for A an interval, as in (4.4.1), but for any bounded Borel A, a strengthening considered by Breiman (1965) [see Feller (1966, Section XI.1) for counterexamples].

An appropriate condition is embodied in the following definition.

Definition 4.4.III. *A probability distribution F is* spread out *if there exists a positive integer n_0 such that F^{n_0*} has a nonzero absolutely continuous component with respect to Lebesgue measure.*

The definition implies that F^{n_0*} can be written in the form

$$F^{n_0*} = \Sigma + A, \tag{4.4.7}$$

where Σ is singular and A is absolutely continuous with respect to Lebesgue measure, and A has a nonzero density $a(x)$, so that

$$\sigma = \|\Sigma\| = 1 - \int_0^\infty a(x)\,\mathrm{d}x < 1.$$

Since the convolution of A with any power of F or Σ is again absolutely continuous, it follows that the total masses of the absolutely continuous components F^{n*} can only increase as $n \to \infty$, and in fact must approach 1, since $\|\Sigma^{k*}\| = \sigma^k \to 0$. Thus, we might anticipate that the asymptotic behaviour of the renewal measure for a spread out distribution would approximate the behaviour to be expected when a density exists. This is the broad content of the following proposition (see Stone, 1966) from which our further results will follow as corollaries.

Proposition 4.4.IV. *Let F be spread out, U the renewal measure associated with F, and $U_G = G * U$ the renewal measure associated with the corresponding delayed renewal process with initial distribution G. Then U_G can be written in the form*

$$U_G = U_{1G} + U_{2G}, \tag{4.4.8}$$

where U_{1G} is absolutely continuous with density $u_{1G}(x)$ satisfying

$$u_{1G}(x) \to \lambda, \qquad \lambda^{-1} = \int_0^\infty x\,\mathrm{d}F(x), \tag{4.4.9}$$

and U_{2G} is totally finite.

Proof. Consider first the ordinary renewal measure U associated with F. Since the convolution of A with itself can always be taken to dominate a uniformly continuous function (Exercise 4.4.2), there is no loss of generality in supposing that the density $a(x)$ of A in (4.4.6) is continuous, bounded, and vanishes outside some finite interval $(0, M)$.

With this understanding, let U_3 denote the renewal measure associated with the distribution F^{n_0*} so that we may write

$$u_3 = \delta_0 + F^{n_0*} + F^{2n_0*} + \cdots$$

and

$$U = [\delta_0 + F + F^{2*} + \cdots + F^{(n_0-1)*}] * U_3 = \rho * U_3,$$

where ρ has total mass n_0. Also, since U_3 satisfies the renewal equation

$$U_3 = \delta_0 + F^{n_0*} * U_3 = \delta_0 + (\Sigma + A) * U_3,$$

we have $U_3 * (\delta_0 - \Sigma) = \delta_0 + A * U_3$. Since $\delta_0 - \Sigma$ has total mass less than unity, this factor may be inverted to yield

$$U_3 = U_\sigma + A * U_\sigma * U_3,$$

where $U_\sigma = \delta_0 + \Sigma + \Sigma^{2*} + \cdots$ has total mass $(1-\sigma)^{-1}$. Thus, we obtain for U, and then for U_G,

$$U_G = G * \rho * U_\sigma + A * G * \rho * U_\sigma * U_3.$$

This will serve as the required decomposition, with $U_{2G} = G * \rho * U_\sigma$ totally finite and $U_{1G} = A * G * \rho * U_\sigma * U_3$ absolutely continuous, since it is a

convolution in which one of the terms is absolutely continuous. To show that its density has the required properties, we note first that the key renewal theorem applies to U_3 in the form

$$(U_3 * g)(t) \to \frac{\lambda}{n_0} \int_0^\infty g(x)\,\mathrm{d}x$$

whenever g is directly Riemann integrable. But then a similar result applies also to $H = G*\rho*U_\sigma*U_3$, which is simply a type of delayed renewal measure in which the initial 'distribution' $G*\rho*U_\sigma$ has total mass $1 \times n_0 \times (1-\sigma)^{-1}$, so that

$$(H * g)(t) \to \frac{\lambda}{1-\sigma} \int_0^\infty g(x)\,\mathrm{d}x \qquad (t \to \infty).$$

Finally, since the density of A is continuous and vanishes outside a bounded set, we can take $g(t) = a(t)$, in which case the left-hand side of the last equation reduces to $u_{1G}(t)$ and we obtain

$$u_{1G}(t) \to \frac{\lambda}{1-\sigma} \int_0^\infty a(x)\,\mathrm{d}x = \lambda. \qquad \square$$

We have the following corollary (see Arjas, Nummelin and Tweedie, 1978).

Corollary 4.4.V. *If F is spread out and $g \geq 0$ is bounded, integrable, and satisfies $g(x) \to 0$ as $x \to \infty$, then*

$$\lim_{t\to\infty} \sup_{|f|\leq g} \left| (U_G * f)(t) - \lambda \int_0^\infty f(x)\,\mathrm{d}x \right| \to 0. \tag{4.4.10}$$

Proof. We consider separately the convolution of g with each of the two components in the decomposition (4.4.8) of U_G. Taking first the a.c. component, and setting $u_G(x) = 0$ for $x < 0$, we have

$$\sup_{|f|\leq g} \left| \int_0^t u_{1G}(t-x) f(x)\,\mathrm{d}x - \lambda \int_0^\infty f(x)\,\mathrm{d}x \right| \leq \int_0^\infty \left| u_{1G}(t-x) - \lambda \right| g(x)\,\mathrm{d}x.$$

Now $u_{1G}(t) \to \lambda$ so it is bounded for sufficiently large t, $|u_{1G}(t) - \lambda| \leq C$ say, for $t > T$, and we can write the last integral as

$$\int_0^{t-T} g(x) \left| u_{1G}(t-x) - \lambda \right| \mathrm{d}x + \int_0^T \left| u_{1G}(s) - \lambda \right| g(t-s)\,\mathrm{d}s,$$

where the first integral tends to zero by dominated convergence because $|u_{1G}(t-x) - \lambda|$ is bounded, $u_{1G}(t-x) \to \lambda$ for each fixed x, and $g(x)$ is integrable, while the second tends to zero by dominated convergence since $|u_{1G}(s) - \lambda|$ has finite total mass over $(0,T)$ and by assumption $g(t-s) \to 0$ for each fixed s.

Similarly, the integral against the second component is dominated for all $|f| \le g$ by

$$\int_0^t g(t-x)\,\mathrm{d}U_{2G}(x),$$

where again the integrand is bounded and tends to zero for each fixed x, while U_{2G} has finite total mass, so the integral tends to zero by dominated convergence. □

Corollary 4.4.VI. *If F is spread out, then for each finite interval $(0, M)$*

$$\|S_t U_G - \lambda\ell\|_M \to 0.$$

The version of the renewal theorem summarized by these results has the double advantage of not only strengthening the form of convergence but also replacing the rather awkward condition of direct Riemann integrability by the simpler conditions of Proposition 4.4.IV. Further variants are discussed in Exercise 4.4.4 and in the paper by Arjas *et al.* (1978). With further conditions on the lifetime distributions—for example, the existence of moments—it is possible to obtain bounds on the rate of convergence in the renewal theorem. For results of this type, see Stone (1966), Schäl (1971), and Bretagnolle and Dacunha-Castelle (1967); for a very simple case, see Exercise 4.4.5(a).

Exercises and Complements to Section 4.4

4.4.1 *Conditions for direct Riemann integrability.* Let $z(x)$ be a measurable function defined on $[0, \infty)$. Show that each of the following conditions is sufficient to make $z(\cdot)$ directly Riemann integrable (see also Feller, 1966).

(a) $z(x)$ is nonnegative, monotonically decreasing, and Lebesgue integrable.

(b) $z(x)$ is continuous, and setting $\alpha_n = \sup_{n<x\le n+1} |z(x)|$, $\Sigma\alpha_n < \infty$. [*Hint*: $z(x)$ is Riemann integrable on any finite interval, and the remainder term outside this interval provides a contribution that tends to zero.]

(c) $z(x) \ge 0$, $z(x)$ is uniformly continuous and bounded above by a monotonically decreasing integrable function.

4.4.2 (a) If g is bounded and continuous and f is integrable, then their convolution product $f * g = \int_{\mathbb{R}} g(t-x)f(x)\,\mathrm{d}x$ is uniformly continuous.

(b) Extend this to the case where g is any bounded measurable function by approximating g by bounded continuous functions. In particular, therefore, $\int_A f(t-x)\,\mathrm{d}x$ is uniformly continuous whenever A is a measurable set.

(c) Let F have a.c. component f; show from (b) that $F * F$ has an a.c. component f_2, which dominates a uniformly continuous function and hence a bounded function that vanishes outside a bounded set and is twice continuously differentiable.

4.4.3 Apply the key renewal theorem as around (4.4.5) to show that if F has density f with f^{k*} bounded, and if $\lambda^{-1} < \infty$, then the renewal density $u(x)$ exists and satisfies

$$u(x) - \sum_{j=1}^{2k-1} f^{j*}(x) \to \lambda.$$

[*Hint*: $u(x) - \sum_{j=1}^{2k-1} f^{j*}(x) = \sum_{j=2k}^{\infty} f^{j*}(x)$ satisfies the renewal equation with $z(x) = f^{2k*}(x)$, which is uniformly continuous and bounded above by an integrable function. Necessary and sufficient conditions for $u(x)$ itself to converge are given in Smith (1962); see also Feller (1966, Section XI.4).]

4.4.4 *Strong convergence counterexample.* Let G_u denote the distribution of the forward recurrence time at $t = u$ and G_∞ its limit, if it exists, of a renewal process $N(\cdot)$ with lifetime distribution F with mean $1/\lambda$.

(a) Suppose that F has discrete support but is nonlattice. Show that $G_u(x) \to G_\infty(x) = \lambda \int_0^x [1-F(u)]\,\mathrm{d}u$, but that $\|G_u - G_\infty\| = 2$ (all finite u). [Hence, G_u does not converge in variation norm $\|\cdot\|$, i.e. strong convergence fails.]

(b) Show that $\|G_u - G_\infty\| \to 0$ $(u \to \infty)$ when F is spread out.

4.4.5 *Rate of convergence in renewal theorems.*

(a) Consider (4.1.8) with $z(t) = \lambda \int_t^\infty \overline{F}(y)\,\mathrm{d}y$, where $\overline{F}(y) = 1 - F(y)$ and F has second moment $\sigma^2 + \mu^2$. Deduce that Z, the solution of (4.1.8) with such z, equals $\phi(t) \equiv U(t) - \lambda t$. Use the key renewal theorem to conclude that for nonlattice F,

$$0 \le \phi(t) = \lambda \int_0^t \left(\int_{t-u}^\infty \overline{F}(v)\,\mathrm{d}v \right) \mathrm{d}U(u) \to \tfrac{1}{2}\lambda^2(\sigma^2 + \mu^2) \qquad (0 \le t \to \infty).$$

(b) Let the r.v.s T_1, T_2 be independent with $\Pr\{T_1 > t\} = z(t)$ as in (a). Use the subadditivity of the renewal function $U(\cdot)$ to give, for all $t \ge 0$,

$$U(2t) \le 2\mathrm{E}U(t + T_1 - T_2),$$

and hence deduce from $\mathrm{E}U(t - T_1) = \lambda t$ (cf. Example 4.1(c) and Proposition 4.2.I) that

$$2\lambda t \le U(2t) \le 2\lambda t + \lambda^2\sigma^2 + 1.$$

[See Carlsson and Nerman (1986) for details and earlier references.]

(c) Suppose that the generator $z(\cdot)$ in the general renewal equation (4.1.8) is positive and decreases monotonically. Show that $J_1(t) \equiv \int_0^t z(u)\lambda\,\mathrm{d}u \to \infty$ $(t \to \infty)$ if and only if $J_2(t) \equiv \int_0^t z(t-u)\,\mathrm{d}U(u) \to \infty$ $(t \to \infty)$ and that then $\lim_{t\to\infty} J_1(t)/J_2(t) = 1$.

Deduce that, when $F(\cdot)$ has infinite second moment, $U(t) - \lambda t \sim \int_0^\infty \lambda^2 \min(v,t)\overline{F}(v)\,\mathrm{d}v \equiv G(t)$ (Sgibnev, 1981).

For an alternative proof, show that $\phi(t) \le \int_0^\infty U(\min(v,t))\lambda\overline{F}(v)\,\mathrm{d}v \equiv G_U(t) \ge G(t)$ by the elementary renewal theorem. Use Blackwell's theorem to show that $\limsup_{t\to\infty} G_U(t)/G(t) \le 1$.

When $F(\cdot)$ has finite second moment and is nonarithmetic, show that $\lim_{t\to\infty}[J_1(t) - J_2(t)] = 0$.

(d) Use the asymptotics of $\phi(\cdot)$ to deduce that for a stationary orderly renewal process $N(\cdot)$, $\mathrm{var}\,N(0,t] \sim (\mathrm{var}\,\lambda X)(\lambda t)$ when the lifetime d.f. has finite second moment, and $\mathrm{var}\,N(0,t] \sim \lambda^2 t^2 - \lambda^3 \int_0^t (t-v)^2\overline{F}(v)\,\mathrm{d}v$ otherwise. [*Hint*: First, find $\mathrm{var}\,N(0,t]$ from (3.5.2) and (3.5.6).]

4.5. Neighbours of the Renewal Process: Wold Processes

The specification of a renewal process via independent identically distributed intervals raises the possibility of specifying other point processes via intervals that are one step removed from independence. In this section, we consider point processes for which the successive intervals $\{X_n\}$ form a Markov chain so that the distribution of X_{n+1} given $X_n, X_{n-1}, \ldots$ in fact depends only on X_n. Such processes seem to have been considered first by Wold (1948); accordingly, we call them *Wold processes*.

EXAMPLE 4.5(a) *A first-order exponential autoregressive process.* Suppose that the family $\{X_n\}$ of intervals satisfy the relation

$$X_{n+1} = \rho X_n + \epsilon_n \tag{4.5.1}$$

for some $0 \le \rho < 1$ and family $\{\epsilon_n\}$ of i.i.d. nonnegative random variables (note $\{X_n\}$ is itself i.i.d. if $\rho = 0$). For the particular distribution given by

$$\Pr\{\epsilon_n = 0\} = \rho \qquad \text{and} \qquad \Pr\{\epsilon_n > y\} = (1-\rho)\mathrm{e}^{-y} \quad (y > 0),$$

taking Laplace transforms of (4.5.1) shows that if a stationary sequence of intervals is to exist, the common distribution F of the $\{X_n\}$ must have its Laplace–Stieltjes transform $\widetilde{F}$ satisfy the functional equation

$$\widetilde{F}(s) = \frac{\widetilde{F}(\rho s)(1+\rho s)}{1+s}\,.$$

The only solution of this equation for which $\widetilde{F}(0) = \widetilde{F}(0+) = 1$ is $\widetilde{F}(s) = (1+s)^{-1}$. Thus, a stationary version of the Markov chain exists and the marginal distribution for the intervals is exponential as for a Poisson process. The parameter ρ controls the degree of association between the intervals. For $\rho > 0$, a realization of the process consists of a sequence of intervals each one of which is an exact fraction of the preceding one, followed by an interval independently chosen from the same exponential distribution. The construction can be extended to more general types of gamma distribution and has been studied extensively by P.A.W. Lewis and co-authors: see, for example, Gaver and Lewis (1980). They have advocated its use as an alternative to the Poisson process, partly on the grounds of the very simple behaviour of the spectrum of the interval process. Other aspects are more intractable, however, and from a point process viewpoint its partly deterministic behaviour gives it a rather special character (see Exercises 4.5.2 and 4.5.9). □

In general, the interval structure of a Wold process is determined by a Markov transition kernel $P(x, A)$; that is, a family $\{P(x,\cdot)\colon 0 \le x < \infty\}$ of probability measures in $[0,\infty)$, and the distribution, $P_0(\cdot)$ say, of the initial interval X_0, with $P(\cdot, A)$ measurable for each fixed Borel set $A \subseteq [0,\infty)$. When the chain $\{X_n\}$ is irreducible [see e.g. Harris (1956), Orey (1971) or Meyn and

Tweedie (1993) for discussions of the precise meaning of irreducibility] and admits a stationary distribution, $\pi(\cdot)$ say, so that for all such Borel subsets A

$$\pi(A) = \int_{0-}^{\infty} P(x, A)\, \pi(\mathrm{d}x), \tag{4.5.2}$$

an interval sequence $\{X_n\}$ with a stationary distribution can be specified. The following construction then leads to a counting process $N(\cdot)$ that is stationary in the sense of Definition 3.2.I.

First, let $\{X_0, X_1, \ldots\}$ be a realization of the Markov chain for which X_0 has the initial distribution

$$P_0(\mathrm{d}x) \equiv \Pr\{X_0 \in (x, x + \mathrm{d}x)\} = \frac{x\, \pi(\mathrm{d}x)}{\int_{0-}^{\infty} u\, \pi(\mathrm{d}u)}\,, \tag{4.5.3a}$$

where we suppose both $\pi\{0\} = 0$ and finiteness of the normalizing factor; i.e.

$$\lambda^{-1} \equiv \int_{0-}^{\infty} x\, \pi(\mathrm{d}x) = \int_{0}^{\infty} \pi(u, \infty)\, \mathrm{d}u \;<\; \infty. \tag{4.5.3b}$$

Next, conditional on X_0, let X_0' be uniformly distributed on $(0, X_0)$, and determine N by

$$N(0, x] \;=\; \#\{n \colon S_n' \le x\},$$

where

$$S_1' = X_0', \qquad S_{n+1}' = S_n' + X_n \qquad (n = 1, 2, \ldots).$$

The relation (4.5.3), in conjunction with the definition of S_n', states that the origin is located uniformly at random within an interval selected according to the length-biased distribution with increment around x proportional to $x\, \pi(\mathrm{d}x)$. Since $\pi\{0\} = 0$, the normalizing constant λ is just the intensity of the process. Note that the distributions here are consistent with the relations found in Exercise 3.4.1 for the stationary distributions for the forward recurrence time and the length of the current interval. Indeed, the construction here can be rephrased usefully in terms of the bivariate, continuous-time Markov process

$$\mathbf{X}(t) = \big(L(t), R(t)\big), \tag{4.5.4}$$

where $L(t)$ is the length of the interval containing t and $R(t)$ is the forward recurrence time at time t. The Markovian character of $\mathbf{X}(t)$ follows readily from that of the sequence of intervals. Moreover, it is clear that the process $N(t)$ is uniquely determined by $\mathbf{X}(t)$ and vice versa. By starting the Markov process with its stationary distribution, we ensure that it remains stationary in its further evolution, and the same property then holds for the point process.

An immediate point of contrast to the ordinary point process is that it is not necessary, in (4.5.2), to have $\int_{\mathbb{R}_+} \pi(\mathrm{d}x) < \infty$. If the underlying Markov chain is null recurrent, a stationary regime can exist for the point process (though not for its intervals) in which, because of the dependence between the lengths

of successive intervals, long runs of very short intervals intervene between the occurrences of longer intervals; in such situations, divergence of $\int_{\mathbb{R}_+} \pi(\mathrm{d}x)$ can coexist with convergence of $\int_{\mathbb{R}_+} x\,\pi(\mathrm{d}x)$ (i.e. near the origin, π may integrate x but not 1). This leads to the possibility of constructing stationary Wold processes with infinite intensity but finite mean interval length. One such construction is given in Daley (1982); another is outlined in Exercise 4.5.1.

With such examples in mind, it is evident that the problem of formulating analogues of the renewal theorems for the Wold process needs to be approached with some care. One possible approach is through the family of renewal measures

$$U(A \mid x) = \mathrm{E}[\#\{n\colon S_n \in A\} \mid X_0 = x]$$

and their associated cumulative processes $U(t \mid x) \equiv U([0,t] \mid x)$. The latter functions satisfy the renewal-type equations

$$U(t \mid x) = I_{\{t \geq x\}}(t) + \int_0^\infty U(t - x \mid y)\, P(x, \mathrm{d}y). \tag{4.5.5}$$

Unfortunately, these equations seem rather intractable in general. The analogy with the renewal equations of Section 4.4 becomes clearer on taking Laplace–Stieltjes transforms of (4.5.5) with respect to t. Introducing the integral operator T_θ with kernel

$$t_\theta(\mathrm{d}y, x) = \mathrm{e}^{-\theta x} P(x, \mathrm{d}y),$$

the transform versions of equation (4.5.5) become

$$U_\theta(x) \equiv \int_0^\infty \mathrm{e}^{-\theta t}\, U(\mathrm{d}t \mid x) \;=\; \mathrm{e}^{-\theta x} + (T_\theta U_\theta)(x)$$

with the formal solution $U_\theta = (1 - T_\theta)^{-1} e_\theta$, where $(e_\theta)(x) \equiv \mathrm{e}^{-\theta x}$, which may be compared to equation (4.1.6).

Example 4.5(b) *Discrete Wold processes.* Consider a simple point process ($\{0,1\}$-valued process) on the lattice of integers $\{0, 1, \ldots\}$; the kernel $P(x, \mathrm{d}y)$ here becomes a matrix p_{ij} and in place of the cumulative form in (4.5.5) it is more natural to consider the renewal functions $u(j \mid i) = \Pr\{N\{j\} = 1 \mid X_0 = i\}$. Then

$$u(j \mid i) = \delta_{ij} + \sum_{k=1}^{\infty} p_{ik} u(j - i \mid k),$$

taking the right-hand side here to be zero for $j < i$. By introducing the transforms $u_i(z) = \sum_{k=i}^\infty z^k u(k \mid i)$, these become

$$u_i(z) = z^i + \sum_{k=1}^{\infty} p_{ik} z^i u_k(z),$$

or in matrix-vector form

$$\mathbf{u}(z) = \zeta + \mathbf{P}_z \mathbf{u}(z),$$

where $\mathbf{P}_z = \{p_{ik}z^i\}$, $\mathbf{u}(z) = \{u_i(z)\}$, and $\zeta = (1, z, z^2, \ldots)$. The asymptotic behaviour of $u(j \mid i)$ as $j \to \infty$ is therefore related to the behaviour of the resolvent-type matrix $(\mathbf{I} - \mathbf{P}_z)^{-1}$ as $z \to 1$. When $\mathbf{P}$ is finite, this can be discussed in classical eigenvector/eigenvalue terms; see Exercise 4.5.4 and for further details Vere-Jones (1975). A particular question that arises relates to *periodicity* of the process: nonzero values of $u(j \mid i)$ may be restricted to a sublattice of the integers. This phenomenon is *not* directly related to periodicity of the underlying Markov chain; again, see Exercise 4.5.4 for some examples. □

A more general approach, which can be extended to the denumerable case and anticipates the general discussion to be given below, is to consider the discrete version of the Markov chain $\mathbf{X}(t)$ in (4.5.4). When this bivariate chain is aperiodic and recurrent, returns to any given state pair—for example, time points at which an interval of specified length i_0 is just commencing—constitute an imbedded renewal process for $\mathbf{X}(t)$ and allow standard renewal theory results to be applied.

Example 4.5(c) *Transition kernels specified by a diagonal expansion.* Lancaster (1963) investigates the class of bivariate probability densities that can be represented by an expansion of the kind

$$f(x,y) = f_X(x)f_Y(y)\left(1 + \sum_{n=1}^{\infty} \rho_n L_n(x) M_n(y)\right),$$

where $f_X(\cdot)$, $f_Y(\cdot)$ are the marginal densities and $L_n(x)$, $M_n(y)$ are families of complete orthonormal functions defined with respect to the marginal distributions $f_X(\cdot)$, $f_Y(\cdot)$, respectively. When f_X and f_Y coincide (so $L_n = M_n$), the bivariate density can be used to define the density of the transition kernel of a stationary Markov chain with specified stationary distribution $f_X(x)$: just put

$$p(x,y) = \frac{f(x,y)}{f_X(x)} = f_X(y)\left(1 + \sum_{n=1}^{\infty} \rho_n L_n(x) L_n(y)\right).$$

For many of the standard distributions, this leads to expansions in terms of classical orthogonal polynomials (see e.g. Tyan and Thomas, 1975). In particular, when $f_X(x)$ and $f_Y(y)$ are both taken as gamma distributions,

$$f_X(x) = x^{\alpha-1}\mathrm{e}^{-x}/\Gamma(\alpha), \qquad \text{say,}$$

the $L_n(x)$ become the Laguerre polynomials of order α. The bivariate exponential density of Example 4.1(e) is a case in point when $\alpha = 1$ and $\rho_n = \rho^n$. The resulting Wold process then has exponential intervals, but in contrast to Example 4.5(a), the realizations have no deterministic properties but simply appear as clustered groups of small or large intervals, the degree of clustering being controlled by the parameter ρ. Lampard (1968) describes an electrical counter system that produces correlated exponential intervals. More gener-

ally, when $\alpha = \frac{1}{2}d$, such correlated gamma distributions can be simulated from bivariate normal distributions with random variables in common; this leads to the possibility of simulating Wold processes with correlated gamma intervals starting from a sequence of i.i.d. normal variates (see Exercise 4.5.7).

Even in such a favourable situation, the analytic study of the renewal functions remains relatively intractable. Lai (1978) studies the exponential case in detail and provides a perturbation expansion for the renewal function and (count) spectral density of the process in terms of the parameter ρ. □

As such examples illustrate, explicit computations for the Wold process are often surprisingly difficult. However, a useful and general approach to the asymptotic results can be developed by identifying a sequence of regeneration points within the evolution of the process and by applying to this sequence the renewal theorems of Section 4.4. It is by no means obvious that any such sequence of regeneration points exists, but the 'splitting' techniques developed for Markov chains with general state space by Nummelin (1978) and Athreya and Ney (1978) allow such a sequence to be constructed for a wide class of examples. The essence of this idea is to identify a particular set A_0 in the state space and a particular distribution ϕ on A_0 such that whenever the process enters A_0, it has a certain probability of doing so 'according to ϕ', when its future evolution will be just the same as when it last entered A_0 'according to ϕ'. In effect, returns to A_0 according to ϕ can be treated as if they are returns to a fixed atom in the state space and provide the regeneration points we seek. The following conditions summarize the requirements on the transition kernel for this to be possible (see Athreya and Ney, 1978).

Conditions 4.5.I. (Regenerative Homing Set Conditions). *For the Markov chain $\{X_n\}$ on state space $S \subseteq [0,\infty) \equiv \mathbb{R}_+$, there exists a homing set $A_0 \in \mathcal{B}(\mathbb{R}_+)$, $A_0 \subseteq S$, a probability measure ϕ on A_0, and a positive constant c such that for all $x \in S$,*

(i) $\Pr\{X_n \in A_0 \text{ for some } n = 1, 2, \ldots \mid X_0 = x\} = 1$; *and*

(ii) *for every Borel subset B of A_0, $P(x, B) \geq c\phi(B)$.*

The first of these conditions embodies a rather strong recurrence condition; indeed Athreya and Ney call a chain satisfying Condition 4.5.I 'strongly aperiodic recurrent' since the conditions imply aperiodicity as well as recurrence. The second condition is more akin to an absolute continuity requirement on the transition kernel. In particular, it is satisfied whenever the following simpler but more stringent condition holds.

Condition 4.5.I′. (ii)′ *For all $x \in A_0$, $P(x, B)$ has density $p(x, y)$ on A_0 with respect to ϕ such that $p(x, y) \geq c > 0$ for all $y \in A_0$.*

Typically, A_0 is a set with positive Lebesgue measure and ϕ the uniform distribution on A_0 (i.e. a multiple of Lebesgue measure scaled to give A_0 total mass unity). In the discrete case, 4.5.I(ii) is equivalent to the assumption that the matrix of transition probabilities has at least one positive diagonal element.

Conditions 4.5.I are trivially satisfied in the independent (renewal) case if we take S to be the support of the lifetime distribution F and put $A_0 = S$, $\phi = F$, and $c = 1$.

Under Conditions 4.5.I, Athreya and Ney (1978) show that the chain is recurrent in the sense of Harris (1956) and admits a unique finite invariant measure $\pi(\cdot)$. The important feature for our purposes is not so much the existence of the invariant measure as its relation to the sequence $\{\nu_k\}$ of 'returns to A_0 according to ϕ'. This aspect is made explicit in the following proposition [see Athreya and Ney (1978) and Nummelin (1978) for proof].

Proposition 4.5.II. *Conditions* 4.5.I *imply that for the Markov chain* $\{X_n\}$,

(a) *there exists a stopping time* $\nu \geq 1$ *with respect to the* σ*-fields generated by* $\{X_n\}$ *such that for Borel subsets* B *of* A_0

$$\Pr\{X_\nu \in B \mid X_0 \cdots X_{\nu-1}; \nu\} = \phi(B); \tag{4.5.6}$$

(b) $\{X_n\}$ *has an invariant measure* $\pi(\cdot)$ *related to* ϕ *by*

$$\pi(B) = \mathrm{E}_\phi\left(\sum_{n=0}^{\nu-1} I_B(X_n)\right) \qquad \big(\text{all } B \in \mathcal{B}(\mathbb{R}_+)\big), \tag{4.5.7}$$

where E_ϕ *refers to expectations under the initial condition that* X_0 *has distribution* ϕ *on* A_0, *i.e.* $\Pr\{X_0 \in B\} = \phi(B \cap A_0)$ *for* $B \in \mathcal{B}(\mathbb{R}_+)$.

Equation (4.5.7) can be extended by linearity and approximation by simple functions to

$$\int_{\mathbb{R}_+} f(x)\,\pi(\mathrm{d}x) = \mathrm{E}_\phi\left(\sum_{n=0}^{\nu-1} f(X_n)\right) \tag{4.5.8}$$

whenever f is Borel-measurable and either nonnegative or π-integrable. Special cases of (4.5.8) include

$$\mathrm{E}_\phi(\nu) = \int_{\mathbb{R}_+} \pi(\mathrm{d}x) \tag{4.5.9a}$$

and

$$\mathrm{E}_\phi(X_0 + X_1 + \cdots + X_{\nu-1}) = \int_{\mathbb{R}_+} x\,\pi(\mathrm{d}x). \tag{4.5.9b}$$

Now let $S_n = \sum_{i=1}^n X_i$, and let $\{T_k\} = \{S_{\nu_k} - 1\}$ denote the sequence of times at which the process returns to A_0 according to ϕ. These T_k form the regeneration points that we seek. If $G(\cdot)$ denotes the distribution function of the successive differences $T_k - T_{k-1}$ so that in particular

$$G(u) = \mathrm{E}_\phi\{I_{S_{\nu-1}} \leq u\} = \Pr_\phi\{S_{\nu-1} \leq u\}, \tag{4.5.10}$$

then the T_k form the instants of a renewal process with lifetime distribution G. We apply this fact, with the theorems of Section 4.4, to determine the asymptotic behaviour of the Wold process.

The results are stated for the renewal function

$$U_\phi(C \times T_t B) = \mathrm{E}_\phi \#\{n\colon X_n \in C, S_n \in T_t B\}, \tag{4.5.11}$$

where $T_t B$ is the translate of B through time t. If the process is started from a general distribution κ for X_0, we write $U_\kappa(\cdot)$ for the corresponding renewal function. The analogue of Blackwell's renewal theorem for this function reads, for $B = (0, h)$ and λ as in (4.5.3b),

$$U_\phi(C \times T_t B) \to \lambda \pi(C) \ell(B).$$

We approach these results through an extended version of the key renewal theorem, fixing a bounded measurable function $h(x, y)$ with support in the positive quadrant $x \geq 0, y \geq 0$, and setting for $t > 0$

$$Z(t) = \mathrm{E}_\phi \left(\sum_{n=0}^{N(t)} h(X_n, t - S_n) \right) = \int_0^\infty \int_0^t h(x, t-u)\, U_\phi(\mathrm{d}x \times \mathrm{d}u). \tag{4.5.12}$$

Considering the time T_1 to the first return to A_0 according to ϕ, we find that $Z(t)$ satisfies the renewal equation $Z(t) = z(t) + \int_0^t Z(t-u)\, \mathrm{d}G(x)$, where

$$z(t) = \mathrm{E}_\phi \left(\sum_{n=0}^{\nu-1} h(X_n, t - S_n) \right) = \mathrm{E}_\phi \left(\int_0^T h(X_{N(u)}, t-u)\, \mathrm{d}N(u) \right). \tag{4.5.13}$$

If then we can show that $z(t)$ satisfies the condition of direct Riemann integrability (for Feller's form of the key renewal theorem in 4.4.II) or the conditions in 4.4.III for the Breiman form of the theorem, we shall be able to assert that

$$Z(t) \to \lambda \int_0^\infty z(t)\, \mathrm{d}t \qquad (t \to \infty).$$

To evaluate the integral, we make use of (4.5.8) so that formally

$$\begin{aligned} \int_0^\infty z(t)\, \mathrm{d}t &= \int_0^\infty \mathrm{E}_\phi \left(\sum_{n=0}^{\nu-1} h(X_n, t - S_n) \right) \mathrm{d}t \\ &= \mathrm{E}_\phi \left(\sum_{n=0}^{\nu-1} \int_{S_n}^\infty h(X_n, t - S_n)\, \mathrm{d}t \right) = \mathrm{E}_\phi \left(\sum_{n=0}^{\nu-1} \int_0^\infty h(X_n, u)\, \mathrm{d}u \right) \\ &= \int_0^\infty \int_0^\infty h(x, t)\, \pi(\mathrm{d}x)\, \mathrm{d}t, \end{aligned} \tag{4.5.14}$$

the formal operations being justified by Fubini's theorem whenever $h \geq 0$ or h is $(\pi \times \ell)$-integrable.

Direct Riemann integrability can be established directly in simple cases, to which we add the following general sufficient condition. For $\delta > 0$, any α in $0 \le \alpha < \delta$, and $I_j(\delta) \equiv (j\delta, (j+1)\delta]$, define

$$\overline{m}_\delta(x,\alpha) = \sum_{j=0}^{\infty} \sup_{t \in I_j(\delta)} h(x,t) \qquad \text{and} \qquad \overline{m}_\delta(x) = \sup_{0 \le \alpha < \delta} \overline{m}_\delta(x,\alpha),$$

and similarly $\underline{m}_\delta(x,\alpha)$ and $\underline{m}_\delta(x)$ by replacing sup by inf. For any y, there is a unique $\alpha_\delta(y)$ in $[0,\delta)$ such that $y = j'\delta + \alpha_\delta(y)$ for some integer j'. Then

$$\sum_{j=0}^{\infty} \sup_{t \in I_j(\delta)} h(x, t-y) = \overline{m}_\delta\big(x, \alpha_\delta(y)\big).$$

Using first Fatou's lemma and then Fubini's theorem,

$$\begin{aligned} \sum_{j=0}^{\infty} \sup_{t \in I_j(\delta)} z(t) &\le \mathrm{E}_\phi\left(\sum_{n=0}^{\nu-1} \overline{m}_\delta\big(X_n, \alpha_\delta(-S_n)\big) \right) \\ &\le \mathrm{E}_\phi\left(\sum_{n=0}^{\nu-1} \overline{m}_\delta(X_n) \right) = \int_0^\infty \overline{m}_\delta(x)\, \pi(\mathrm{d}x). \end{aligned}$$

A similar lower bound with sup and $\overline{m}_\delta$ replaced by inf and $\underline{m}_\delta$, respectively, holds. Thus, a sufficient condition for the direct Riemann integrability of $z(t)$ is that, as $\delta \downarrow 0$,

$$\delta \int_0^\infty [\overline{m}_\delta(x) - \underline{m}_\delta(x)]\, \pi(\mathrm{d}x) \to 0. \tag{4.5.15}$$

If, alternatively, G is spread out, then it is enough to show that $z(t)$ is integrable and tends to zero as $t \to \infty$. Simple sufficient conditions for the latter (not the most general possible) are that

$$h(x,t) \to 0 \quad \text{as} \quad t \to \infty \quad \text{for each fixed } x \tag{4.5.16a}$$

and

$$|h(x,t)| \le h_0(x), \tag{4.5.16b}$$

where $h_0(x)$ is π-integrable. This follows readily from (4.5.13) and an application of the dominated convergence theorem.

Summarizing these results, we have the following theorem.

Theorem 4.5.III. *Suppose that the Markov transition kernel associated with a Wold process satisfies the regenerative homing set Conditions* 4.5.I *and that its invariant measure π has a finite normalizing factor λ^{-1} as in* (4.5.3b). *Also let $h(x,t)$ be a fixed measurable function, vanishing outside the positive quadrant in $\mathbb{R}^2$ and $(\pi \times \ell)$-integrable in $\mathbb{R}_+ \times \mathbb{R}_+$, and define G, U_ϕ, Z_ϕ, and z_ϕ by* (4.5.10–13), *respectively. If either*

(i) *G is nonlattice and z_ϕ is directly Riemann integrable, or*
(ii) *G is spread out and $z(t)$ is bounded and $\to 0$ as $t \to \infty$,*
then

$$Z_\phi(t) = \int_0^\infty \int_0^t h(x, t-u)\, U_\phi(\mathrm{d}x \times \mathrm{d}u)$$
$$\to \lambda \int_0^\infty \int_0^\infty h(x,u)\, \pi(\mathrm{d}x)\, \mathrm{d}u \qquad (t \to \infty). \tag{4.5.17}$$

In particular, (4.5.15) *implies Condition* (i) *and* (4.5.16) *Condition* (ii).

We now apply this theorem to some important special cases. Consider first the Blackwell-type result, where

$$h(x,t) = I_A(x) I_{(0,h)}(t).$$

In general, $h(x,t)$ is only $(\pi \times \ell)$-integrable if A is bounded away from zero. Then, since $I_{(0,h)}(t)$ has only two points of discontinuity, each of unit height, it is easy to see that for all $x \in \mathbb{R}_+$,

$$\overline{m}_\delta(x) - \underline{m}_\delta(x) \le 2 I_A(x),$$

so that both (4.5.15) and (4.5.16) are satisfied. Equation (4.5.16) also holds if the interval $(0,h)$ is replaced by any bounded Borel set B. Finally, if $\pi(\cdot)$ is totally finite, the condition on A can be dropped and the same results hold. Thus, we have the following corollary.

Corollary 4.5.IV. *Let A, B be Borel subsets of $\mathbb{R}_+$. If G is nonlattice, then*

$$U_\phi(A \times T_t B) \to \lambda \pi(A) \ell(B) \qquad (t \to \infty) \tag{4.5.18}$$

whenever B is a finite interval $(0,h)$ and $A \subseteq [\epsilon, \infty)$ for some $\epsilon > 0$. If G is spread out, the same result holds for B any bounded Borel set.

If $\pi(\cdot)$ is totally finite, these results hold without any further condition on A.

We next extend the results to an arbitrary initial distribution, κ say, for X_0. If we denote the corresponding renewal functions by U_κ, Z_κ, then Z_κ satisfies

$$Z_\kappa(t) = z_\kappa(t) + \int_0^t Z_\phi(t-u)\, G(\mathrm{d}u) \tag{4.5.19}$$

with

$$z_\kappa(t) = \mathrm{E}_\kappa \left(\sum_{n=0}^{\nu'-1} h(X'_n, t - S'_n) \right), \tag{4.5.20}$$

where X'_1, S'_n refer to the sequence of interval lengths and renewals for the process with initial distribution κ, and ν' is the time of the first entry to A_0 according to ϕ, again starting from X_0 distributed according to κ. It follows

from Condition 4.5.I(i) that this entry is certain, so ν' is finite with probability 1. It then follows from (4.5.19) that

$$Z_\kappa(t) - Z_\phi(t) = z_\kappa(t) - z_\phi(t)$$

so that we need conditions to ensure the convergence of the right-hand side to zero. This will follow from (4.5.20) if $\mathrm{E}_\kappa(\nu') < \infty$ and h is bounded and satisfies (4.5.16a).

Corollary 4.5.V. *Suppose that* (4.5.17) *holds for* U_ϕ *and that* κ *is an arbitrary initial distribution for* X_0. *Then* (4.5.17) *continues to hold with* U_κ *in place of* U_ϕ *if and only if* $z_\kappa(t) - z_\phi(t) \to 0$, *in particular if* h *is bounded and satisfies* (4.5.16a), *and* $\mathrm{E}_\kappa(\nu') < \infty$, $\mathrm{E}_\phi(\nu) = \int_{\mathbb{R}_+} \pi(\mathrm{d}x) < \infty$.

Finally, we turn to the question of the weak convergence of the process $\mathbf{X}(t)$ in (4.5.4). It somewhat simplifies the algebraic details to work with the bivariate process $\mathbf{Y}(t) = (L(t), L(t) - R(t))$, i.e. with the backward recurrence time $L(t) - R(t)$ in place of the forward one. If then $\xi(x, y)$ is any bounded continuous function of x, y in $\mathbb{R}_+ \times \mathbb{R}_+$, we consider $\xi(\mathbf{Y}(t))$, which we may write in the form

$$\xi\big(\mathbf{Y}(t)\big) = \sum_{n=0}^{\infty} h(L_n, t - S_n),$$

where

$$h(x,t) = \begin{cases} \xi(x,t) & (0 \le t \le x), \\ 0 & (t > x), \end{cases}$$

since in fact only the term with $n = N(t)$ contributes to the sum. Suppose first that G is nonlattice, and define the *modulus of continuity* $\omega(x, \delta)$ of $h(\cdot)$ by

$$\omega(x,\delta) = \sup_{0 \le t \le x-\delta} \sup_{0 \le u \le \delta} |h(x,t) - h(x,t+u)|.$$

Then, for the particular choice of h given above,

$$\overline{m}_\delta(x) - \underline{m}_\delta(x) \le (x/\delta)\omega(x,\delta)$$

so that

$$\delta \int_{\mathbb{R}_+} [\overline{m}_\delta(x) - \underline{m}_\delta(x)]\, \pi(\mathrm{d}x) \le \int_{\mathbb{R}_+} x\, \omega(x,\delta)\, \pi(\mathrm{d}x).$$

For each fixed $x > 0$, $h(x,t)$ is continuous and nonvanishing on a finite closed interval so it is uniformly continuous, and hence $\omega(x,\delta) \to 0$. Also, $\omega(x,\delta)$ is uniformly bounded in x and δ, so by dominated convergence, the integral on the right converges to zero as $\delta \to 0$; that is, (4.5.15) holds. Also,

$$|z_\kappa(t)| \le \mathrm{E}_\kappa\big[\, \big|\xi\big(\mathbf{Y}(t)\big)\big|; T > t \,\big] \le CP_\kappa\{T > t\},$$

where the last term tends to zero from the recurrence property assumed in Condition 4.5.I(i). Consequently, the conditions for Corollary 4.5.V hold. If, furthermore, G is spread out, then this result alone is sufficient to ensure the truth of the Riemann-type theorem. This means the continuity condition on ξ can be dropped, implying that the weak convergence of $\mathbf{Y}(t)$ to its limit can be replaced by convergence in variation norm.

Proposition 4.5.VI. *Let $P_{\kappa,t}$ denote the distribution of $\mathbf{X}(t)$ supposing X_0 has initial distribution κ, and π_∞ the stationary distribution for $\mathbf{X}(t)$ with elementary mass $\lambda\,\pi(\mathrm{d}x)\,\mathrm{d}y$ over the region $0 \le y \le x < \infty$. If G is nonlattice and $\lambda^{-1} = \int_{\mathbb{R}_+} x\,\pi(\mathrm{d}x) < \infty$, then $P_{\kappa,t} \to \pi_\infty$ weakly. If, furthermore, G is spread out, then $P_{\kappa,t} \to \pi_\infty$ in variation norm.*

Throughout our discussion, we have assumed finiteness of the mean λ^{-1} [see (4.5.3b)]. When the mean is infinite, further types of behaviour are possible, some of which are sketched in Athreya, Tweedie and Vere-Jones (1980).

Exercises and Complements to Section 4.5

4.5.1 *A Wold process with infinite intensity.* Consider a symmetric random walk $\{X_n\}$ with reflecting barrier at the origin, supposing the walk to have density and be null recurrent; for example, the single-step distribution could be $N(0,1)$. Then, the invariant measure for X_n is Lebesgue measure on $(0,\infty)$. Now transform the state space by setting $Y_n = T(X_n)$, where for $y > 0$

$$x = T^{-1}(y) = y^{-\beta}(1+y)^{-\alpha} \qquad (\alpha > 0,\ \beta > 0);$$

note that under T the origin is mapped into the point at infinity and vice versa. Then, the transformed process Y_n is Markovian with invariant measure having density $\pi(y)$, where near the origin $\pi(y) \sim y^{-(1+\beta)}$ and near infinity $\pi(y) \sim y^{-(\alpha+\beta+1)}$. Choose α and β so that $0 < \beta < 1$, $\alpha+\beta > 1$; then $\int_0^\infty y\,\pi(y)\,\mathrm{d}y < \infty$ but $\int_0^1 \pi(y)\,\mathrm{d}y = \infty$. Complete the construction of a stationary version of the corresponding Wold process by using the joint distribution of the current interval and forward recurrence time as indicated in the text following (4.5.4).

4.5.2 *Infinitely divisible autoregressive process.* Let $X \ge 0$ have an infinitely divisible distribution with representation of the form

$$\psi(\theta) = \mathrm{E}(\mathrm{e}^{-\theta X}) = \exp\left(-\int_0^\infty [1 - \mathrm{e}^{-\theta x}]\,M(\mathrm{d}x)\right) \qquad (\mathrm{Re}(\theta) > 0),$$

where $\int_{0,\infty)} \min(x,1)\,M(\mathrm{d}x) < \infty$. Show that there exists a stationary sequence $\{X_n\}$, satisfying the autoregressive equation

$$X_{n+1} = \rho X_n + \epsilon_n \qquad (\epsilon_n \text{ independent of } X_n)$$

and having marginal distribution with Laplace–Stieltjes transform $\psi(\theta)$, whenever M is absolutely continuous with monotonically decreasing density $m(x)$, hence in particular whenever the X_n are gamma distributed.
[*Hint*: If ϵ_n is also infinitely divisible, its Laplace–Stieltjes transform, $\phi(\theta)$ say, must satisfy $\phi(\theta) = \psi(\theta)/\psi(\rho\theta) = \exp\big(\int_0^\infty (\mathrm{e}^{-\theta x} - 1)\,[M(\mathrm{d}x) - M(\rho^{-1}\mathrm{d}x)]\big)$.]

4.5.3 Let $F(t;x,y)$ be the distribution function of the bivariate process $\mathbf{Y}(t) = (L(t), L(t) - R(t))$, conditional on an event at the origin and $L(0-) = s$. Then, if F has a density $f(t;x,y) \equiv f(t;x,y \mid s)$, it satisfies for $0 < y < \min(x,t)$

$$\frac{\partial F}{\partial t} + \frac{\partial F}{\partial y} = \int_0^t f(t;u,u)P(u,(0,x])\,\mathrm{d}u - \int_0^y f(t;u,u)\,\mathrm{d}u,$$

and if also the density function is sufficiently regular, then for the same x, y, t,

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial y} = 0.$$

Argue on probabilistic grounds that $f(t;x,y) = f(t-v;x,y-v)$ for $0 < y-v < \min(x, t-v)$, so $f(t;x,x) = f(t-x;x,0+)$ for $0 < x < t$, and that

$$f(t;x,0+) = p(s,t)p(t,x) + \int_0^t f(t;u,u)p(u,x)\,\mathrm{d}u. \tag{4.5.21}$$

When the p.d.f.s $p(u,x)$ are independent of u, this reduces to the renewal density function equation.

Assuming that the conditions for the limits of Theorem 4.5.III and its corollaries are satisfied, identify $f(x,y) \equiv \lim_{t\to\infty} f(t;x,y)$ with the density function $\pi(x)$ for the stationary measure $\pi(\cdot)$ of the theorem, and deduce the density version of equation (4.5.2) by taking the limit in (4.5.21).

Now let $L(0-)$ be an r.v. with p.d.f. $\lambda s\pi(s)$ with λ as in the theorem. Interpret $\int_0^t \mathrm{d}x \int_0^\infty y f(t;x,y \mid s)\lambda s\pi(s)\,\mathrm{d}s$ as the density of the expectation function $U(\cdot)$ of the Wold process. [Lai (1978) has other discussion and references.]

4.5.4 *Discrete Wold processes.*

(a) Suppose integer-valued intervals are generated by a finite Markov chain on $\{1,2,3\}$ with transition matrices of the forms

$$\text{(i) } P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}; \quad \text{(ii) } P = \begin{pmatrix} 0 & 0 & 1 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}; \quad \text{(iii) } P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

For which of these P do the corresponding Wold processes show lattice behaviour? What is the relation of periodicity of P to lattice behaviour of the associated Wold process?

(b) Define $m_{ij}(n) = \Pr\{\text{interval of length } j \text{ starts at } n \mid X_0 = i\}$ and show that, for $n \geq 0$,

$$m_{ij}(n) = \delta_{ij}\delta_{0n} + \sum_k m_{ik}(n-k)p_{kj} = \delta_{ij}\delta_{0n} + \sum_k p_{ik}m_{kj}(n-i),$$

where we interpret $m_{ij}(n) = 0$ for $n < 0$. In matrix form, the p.g.f.s are given by

$$\mathbf{M}(z) = \{\widetilde{m}_{ij}(z)\} \equiv \left\{ \sum_{n=0}^{\infty} m_{ij}(n) z^n \right\} = \big(\mathbf{I} - \mathbf{H}(z)\big)^{-1},$$

where $\mathbf{H}(z) = \big(h_{ij}(z)\big) \equiv (z^i p_{ij})$.

(c) If the Wold process is nonlattice and P is irreducible,

$$(1-z)[\mathbf{I} - \mathbf{H}(z)]^{-1} = \lambda\mathbf{\Pi} + (1-z)\mathbf{Q}(z),$$

where $\mathbf{\Pi}$ is the one-dimensional projection onto the null space of $\mathbf{I}-\mathbf{P}$ and $\mathbf{Q}(z)$ is analytic within some disk $|z| \leq 1+\epsilon$, $\epsilon > 0$ (see Vere-Jones, 1975).

4.5.5 *Denumerable discrete Wold processes.* Consider the bivariate process $\mathbf{X}(n) = (L(n), R(n))$ [or $\mathbf{Y}(n) = (L(n), L(n) - R(n))$] as a Markov chain with an augmented space. Show that the Wold process is nonlattice if and only if this augmented chain is aperiodic, and that if the original Markov chain is

positive recurrent with stationary distribution $\{\pi_j\}$, having finite mean, the augmented chain $\mathbf{X}(n)$ is positive recurrent with stationary distribution

$$\pi(h,j) = \Pr\{L_n = j, R_n = h\} = \begin{cases} \lambda\pi_j & (h = 1, \ldots, j), \\ 0 & \text{otherwise,} \end{cases}$$

where $\lambda^{-1} = \sum j\pi_j < \infty$ as before.

4.5.6 *Markov chains with kernels generated by a power diagonal expansion.*
(a) If $\{X_n\}$ is generated by a kernel with the structure

$$p(x,y) = f(y)\sum_{n=1}^{\infty} \rho^n L_n(x) L_n(y)$$

for an orthogonal family of functions $L_n(\cdot)$, then the m-step transition kernel $p^{(m)}(x,y)$ is generated by a kernel with similar structure and ρ replaced by $\rho_m = \rho^m$.

(b) In the particular case where $f(\cdot)$ is exponential and the $\{L_n(x)\}$ are Laguerre polynomials, a key role is played by the Hille–Hardy formula

$$\sum_{n=0}^{\infty} L_n(x)L_n(y)\rho^n \;=\; \frac{\mathrm{e}^{-(x+y)\rho/(1+\rho)}}{1-\rho}\, I_0\Big(\frac{2\sqrt{xy\rho}}{1-\rho}\Big).$$

Use this to show the following [see Lai (1978) for details]:
(i) Convergence to the stationary limit as $m \to \infty$ is *not* uniform in x.
(ii) For every $x > 0$, the conditional d.f.s $F(h \mid x) = \int_0^h p(x,y)\,\mathrm{d}y$ are bounded by a common function $\alpha(h)$, where $\alpha(h) < 1$ for $h < \infty$.
(iii) If $A(\theta)$ is the integral operator on $L_1[0,\infty)$ with kernel $p(x,y)\mathrm{e}^{-\theta x}$, then for all θ with $\mathrm{Re}(\theta) \geq 0$, $\theta \neq 0$, $\|A^2(\theta)\| < 1$, so the inverse $[I - A(\theta)]^{-1}$ exists and is defined by an absolutely convergent series of powers of $A(\theta)$.

4.5.7 *Simulation of Wold process with χ^2 interval distribution.* Let $Z_0, Z_1, \ldots$ be a sequence of i.i.d. $N(0,\sigma^2)$ variables; define successively $Y_1 = Z_0/\sqrt{1-\rho^2}$ and $Y_{i+1} = \rho Y_1 + Z_i$ $(i = 1, 2, \ldots)$. Then $\{Y_i\}$ is a stationary sequence of normal r.v.s with first-order autoregressive structure. Construct d independent realizations of such autocorrelated normal series, $\{Y_{1i}, \ldots, Y_{di}; i = 1, 2, \ldots\}$ say, and generate a stationary sequence of autocorrelated gamma r.v.s $\{X_i\}$ by setting

$$X_i = \sum_{k=1}^{d} Y_{ki}^2$$

so $\mathrm{E}X_i = d\sigma^2/(1-\rho^2) \equiv \lambda^{-1}$, $\mathrm{var}\, X_i = 2d\sigma^4/(1-\rho^2)^2$, and $\mathrm{cov}(X_i, X_{i+1}) = d\sigma^4(1+\rho^2)/(1-\rho^2)^2$. These X_i can be used as the intervals of a point process, but the process so obtained is not initially stationary: to obtain a stationary version, the length-biased distribution may be approximated by choosing $T \gg \lambda^{-1}$, selecting a time origin uniformly on $(0,T)$ and taking the initial interval to be the one containing the origin so selected, and the subsequent intervals to be X_1, X_2 and so on.

4.5.8 *Wold processes with intervals conditionally exponentially distributed.* Let $p(x, y)$ be of the form $\lambda(x)\mathrm{e}^{-\lambda(x)y}$.

(a) When $\lambda(x) = \lambda x^{-1/2}$, the marginal density $\pi(x)$ can be found via Mellin transforms (Wold, 1948).

(b) When $\lambda(x) = \lambda + \alpha x$, the density $\pi(x)$ is given by

$$\pi(x) = c(\lambda + \alpha x)^{-1}\mathrm{e}^{-\lambda x}$$

for finite $c > 0$ [see Cox (1955), Cox and Isham (1980, pp. 60–62), and Daley (1982); the model has a simple form of likelihood function and has been used to illustrate problems of inference for Poisson processes when the alternative is a Wold process, in particular of the type under discussion].

4.5.9 *Time-reversed exponential autoregression.* Let the intervals Y_n of a point process be stationary and satisfy

$$Y_{n+1} = \min(Y_n/\rho, \eta_n)$$

for i.i.d. nonnegative η_n and $0 < \rho < 1$. Show that when η_n is exponentially distributed, so also is Y_n, with $\mathrm{corr}(Y_0, Y_n) = \rho^{|n|}$. Furthermore, $\{Y_n\} =_d \{X_{-n}\}$, where X_n are as in Example 4.5(a) with $\Pr\{\epsilon_n > y\} = (1-\rho)\mathrm{e}^{-y}$ [see Chernick *et al.* (1988), where it is also shown that this identification of $\{X_n\}$ as the time-reversed process of $\{Y_n\}$ characterizes the exponential distribution].

4.5.10 *Lampard's reversible counter system* [see Lampard (1968) and Takacs (1976)]. Consider a system with two counters, one of which is initially empty but accumulates particles according to a Poisson process of rate λ, the other of which has an initial content $\xi_0 + r$ particles and loses particles according to a Poisson process of rate μ until it is empty. At that point, the roles of the two counters are reversed; an additional r particles are added to the number ξ_1 accumulated in the first counter, which then begins to lose particles at rate μ, while the second counter begins to accumulate particles again at rate λ. We take $X_0, X_1, \ldots$ to be the intervals between successive reversals of the counters. Then, the $\{X_i\}$ form a Markov chain that has a stationary distribution if and only if $\mu > \lambda$.

4.5.11 *mth-order dependence.* Suppose that the intervals $\{X_i\}$ of a point process form an mth-order Markov chain. Then, in place of the process $(L(t), R(t))$, we may consider the process $\mathbf{X}(t) = (L_{-m+1}(t), \ldots, L_{-1}(t), L(t), R(t))$, where the state is defined as the set of $m-1$ preceding intervals, the current interval, and the forward recurrence time. The regenerative homing set conditions can be applied to the discrete time vector process with state $\mathbf{U}_n = (X_{n-m+1}, \ldots, X_{n-1}, X_n)$, which is Markovian in the simple sense. Establish analogues to Theorem 4.5.III and its corollaries. [See Chong (1981) for details.]

4.5.12 *A non-Poisson process with exponentially distributed intervals.* Let the intervals $\tau_1, \tau_2, \ldots$ of a point process on $\mathbb{R}_+$ be defined pairwise by i.i.d. pairs $\{(\tau_{2n-1}, \tau_{2n})\}$, $n = 1, 2, \ldots$ as follows. For each pair, the joint density function $f(u, v) = \mathrm{e}^{-u-v} + f_\epsilon(u, v)$, where $f_\epsilon(u, v) = 0$ except for (u, v) in the

set $A = \{0 < u < 2 \text{ and } 2 < v < 4, \text{ or } 0 < v < 2 \text{ and } 2 < u < 4\}$, where it equals ϵ for $u \in (0,1)$ and $v \in (2,3)$; $u \in (1,2)$ and $v \in (3,4)$; $v \in (0,1)$ and $u \in (3,4)$; and $v \in (1,2)$ and $u \in (2,3)$; and $f_\epsilon = -\epsilon$ on the complement in A of these four unit squares. Check that τ_{2n-1} and τ_{2n} are not independent, that each τ_i is exponentially distributed with unit mean, and that *every* pair (τ_i, τ_{i+1}) has $\Pr\{\tau_i + \tau_{i+1} \le y\} = \int_0^y we^{-w}\,dw$. Conclude that for any $k = 1, 2, \ldots$, the length of k consecutive intervals has the same distribution as for a Poisson process at unit rate and hence that $N(a,b]$ for $a < b$ is Poisson-distributed with mean $b - a$. [This counterexample to Theorem 2.3.II is due to Moran (1967).]

4.5.13 A stationary point process N with finite second moment is *long-range dependent* when

$$\limsup_{x\to\infty} \frac{\operatorname{var} N(0,x]}{x} = \infty.$$

(a) A renewal process is long-range dependent if and only if the lifetime distribution has infinite second moment (Teugels, 1968; Daley, 1999).

(b) Construct an example of a stationary Wold process that is long-range dependent but for which the marginal distribution of intervals has finite second moment. [Daley, Rolski and Vesilo (2000) note two examples.]

4.6. Stieltjes-Integral Calculus and Hazard Measures

The results in this section can be regarded as being a prelude to the general discussion of conditional intensities and compensators in Chapters 7 and 14. The simplest case concerns a renewal process whose lifetime distribution function $F(\cdot)$ is absolutely continuous with density $f(\cdot)$. An important role is played by the hazard function $q(x) = f(x)/S(x)$ [see (1.1.3)], particularly in applications to forecasting because we can interpret $q(x)$ as the risk of an event occurring in the next short time interval, given the time elapsed since the last renewal; that is,

$$q(x)\,dt = \Pr\{\text{event in } t, t + dt \mid \text{last event at } t - x\}.$$

EXAMPLE 4.6(a) *Prediction of the time to the next event in a renewal process.* Suppose a renewal process has hazard function $q(\cdot)$ as just described and that at time t the time back to the last event is observed to be x. Then, the distribution of the time to the next event has hazard function

$$q_x(y) = q(x + y) \qquad (y \ge 0),$$

corresponding to a d.f. with tail (i.e. conditional survivor function)

$$S_x(y) = 1 - F_x(y) = \exp\left(- \int_0^y q(x+u)\,du \right) = \frac{1 - F(x+y)}{1 - F(x)}\,.$$

Note that x here denotes an observation, and that for a stationary Poisson process, the risk $q_x(y)$ is everywhere constant. □

What of the nonabsolutely continuous case in this example? An appropriate extension of the hazard function is the *hazard measure* $Q(\cdot)$ in Definition 4.6.IV below. Our discussion of $Q(\cdot)$ is facilitated by two results for Lebesgue–Stieltjes integrals. The first is just the formula for integration by parts in the Lebesgue–Stieltjes calculus. The second is much more remarkable: it is the *exponential formula*, which has been used mainly in connection with martingale theory without its being in any sense a martingale result; it is in fact a straightforward (if unexpected) theorem in classical real analysis.

Lemma 4.6.I (Integration-by-Parts Formula)**.** *Let $F(x)$ and $G(x)$ be monotonically increasing right-continuous functions of $x \in \mathbb{R}$. Then*

$$\int_a^b F(x)\,\mathrm{d}G(x) = F(b)G(b) - F(a)G(a) - \int_a^b G(x-)\,\mathrm{d}F(x). \tag{4.6.1}$$

This is a standard result on Lebesgue–Stieltjes integrals; it can be proved directly from first principles or as an application of Fubini's theorem (see e.g. Brémaud 1981, p. 336). Note that the last term of (4.6.1) contains the left-continuous function $G(x-)$; also, recall the convention for Lebesgue–Stieltjes integrals that

$$\int_a^b u(x)\,\mathrm{d}G(x) = \int_{-\infty}^{\infty} I_{(a,b]}(x)u(x)\,\mathrm{d}G(x);$$

if we wish to include the contribution from a jump of G at a itself, then we write the integral as

$$\int_{a-}^{b} u(x)\,\mathrm{d}G(x);$$

similarly, $\int_a^{b-} u(x)\,\mathrm{d}G(x)$ excludes the effect of any jump of G at b. □

Lemma 4.6.II (Exponential Formula)**.** *Suppose $F(x)$ is a monotonically increasing right-continuous function of $x \in \mathbb{R}$ and that $u(x)$ is a measurable function for which $\int_0^t |u(x)|\,\mathrm{d}F(x) < \infty$ for each $t > 0$. Let $\{x_i\}$ be the set of discontinuities of F in $[0,\infty)$; set $\Delta F(x_i) = F(x_i) - F(x_i-)$ and write $F_c(x) = F(x) - \sum_{0<x_i\le t} \Delta F(x_i)$ for the continuous part of $F(\cdot)$. Then, the function*

$$H(t) = H(0)\exp\left(\int_0^t u(x)\,\mathrm{d}F_c(x)\right)\prod_{0<x_i\le t}\big(1 + u(x_i)\Delta F(x_i)\big) \tag{4.6.2}$$

is the unique solution in $t \ge 0$ of the integral equation

$$H(t) = H(0) + \int_0^t H(x-)u(x)\,\mathrm{d}F(x) \tag{4.6.3}$$

satisfying $\sup_{0\le s\le t}|H(s)| < \infty$ for each $t > 0$.

PROOF. We outline a proof (see Brémaud, 1981, pp. 336–339; Andersen *et al.*, 1993, Theorem II.6.1). Write

$$G_1(t) = H(0) \prod_{0<x_i\le t} \big(1 + u(x_i)\big)\Delta F(x_i)$$

and

$$G_2(t) = \exp\bigg(\int_0^t u(x)\,\mathrm{d}F_c(x) \bigg).$$

Then, the relation between (4.6.2) and (4.6.3) is just an application of the integration-by-parts formula to obtain an expression for $G_1(t)G_2(t)$, noting that $G_1(\cdot)$ increases by jumps only at the points $t = x_i$, where in fact the jump is equal to

$$G_1(x_i) - G_1(x_i-) = \big(1 + u(x_i)\big)G_1(x_i-) - G_1(x_i-) = u(x_i)G_1(x_i-).$$

To show that (4.6.2) is the unique bounded solution to (4.6.3), let

$$D(t) = H_1(t) - H_2(t)$$

be the difference between any two bounded solutions. Then $D(t)$ itself is bounded in every finite interval, and we can form the estimate, using (4.6.3) and for fixed finite s and t with $0 < s < t$,

$$|D(s)| \le \int_0^s |D(x-)|\,|u(x)|\,\mathrm{d}F(x) \le M \int_0^s |u(x)|\,\mathrm{d}F(x),$$

where $M = \sup_{0\le s\le t} |D(s)|$. Now feeding this estimate back into (4.6.3) yields

$$|D(s)| \le M \int_0^s \bigg(\int_0^x |u(y)|\,\mathrm{d}F(y) \bigg) |u(x)|\,\mathrm{d}F(x) \le \frac{M}{2} \bigg(\int_0^s |u(x)|\,\mathrm{d}F(x) \bigg)^2.$$

Evidently, this iteration may be continued and yields for general $n \ge 1$

$$|D(s)| \le \frac{M}{n!} \bigg(\int_0^s |u(x)|\,\mathrm{d}F(x) \bigg)^n.$$

This last expression converges to zero as $n \to \infty$, so $D(s) \equiv 0$. □

Corollary 4.6.III. *Lemmas* 4.6.I *and* 4.6.II *remain true when the functions F and G are of bounded variation on finite intervals.*

PROOF. For Lemma 4.6.I, use the fact that any function of bounded variation is the difference of two monotonically increasing right-continuous functions. For Lemma 4.6.II, observe that the argument depends only on the use of the formula for integration by parts and the estimate, for any bounded interval A,

$$\bigg| \int_A u(x)\,\mathrm{d}F(x) \bigg| \le \int_A |u(x)|\,\mathrm{d}V_F(x),$$

where V_F is the total variation of F. □

We now specialize these results to the case where F is a distribution function of a positive random variable, so $F(0+) = 0$, $F(\infty) = \lim_{x\to\infty} F(x) \le 1$.

Definition 4.6.IV. *The* hazard measure $Q(\cdot)$ *associated with the distribution* F *on* $[0,\infty)$ *is the measure on* $[0,\infty)$ *for which*

$$Q(\mathrm{d}x) = \frac{F(\mathrm{d}x)}{S(x-)} = \frac{F(\mathrm{d}x)}{1-F(x-)}\,;$$

in integrated form, the integrated hazard function (IHF) *is the function*

$$Q(t) = \int_0^t \frac{\mathrm{d}F(x)}{1-F(x-)}\,.$$

In the case where F has a density f, we have simply

$$Q(t) = \int_0^t q(x)\,\mathrm{d}x = -\log S(t),$$

where $q(x) = f(x)/S(x)$ is the hazard function and $S(x) = 1 - F(x)$ the survivor function of F. However, this logarithmic relation holds only in the continuous case; in the discrete case, it must be replaced by a relation analogous to (4.6.2) (Kotz and Shanbhag (1980) or Andersen *et al.* (1993, Theorem II.6.6)].

Proposition 4.6.V. *The IHF of a right-continuous d.f.* F *is monotonically increasing and right continuous, and at each discontinuity* x_i *of* F *it has a jump of height*

$$\Delta Q(x_i) = \frac{\Delta F(x_i)}{S(x_i-)} \le 1.$$

Conversely, any monotonically increasing right-continuous nonnegative function Q *with discontinuities of magnitude* < 1*, except perhaps for a final discontinuity of size* 1*, can be the IHF of some d.f.* F *given by the inversion formula*

$$S(t) = 1 - F(t) = \prod_{0\le x_i\le t} \big(1 - \Delta Q(x_i)\big) \exp\left(-\int_0^t \mathrm{d}Q_c(x)\right), \qquad (4.6.4)$$

where $\Delta Q(x_i)$ *is the jump of* Q *at its discontinuity* x_i *and* Q_c *the continuous part of* Q*.*

Proof. Given a d.f. F on $[0,\infty)$, observe first that when F has a jump $\Delta F(x_i)$ at the discontinuity x_i, the corresponding jump in the IHF is $\Delta F(x_i)/S(x_i-)$ by Definition 4.6.IV. Since $\Delta F(x_i) = F(x_i) - F(x_i-) \le 1 - F(x_i-) = S(x_i-)$ with equality if and only if $F(x_i) = 1$—that is, x_i is a discontinuity of F and is the supremum of the support of F—we must have $\Delta Q(x_i) \le 1$ with equality possible only for such x_i.

The inversion formula (4.6.4) is an immediate application of the exponential formula. To see this, we have from Definition 4.6.IV

$$\mathrm{d}F(x_i) = S(x_i-)\,\mathrm{d}Q(x_i)$$

with

$$S(t) = 1 - F(t) = 1 - \int_0^t \mathrm{d}F(x) = 1 - \int_0^t S(x-)\,\mathrm{d}Q(x).$$

Taking $u(x) = -1$ in (4.6.3), $S(\cdot)$ is the unique solution of the equation satisfying $\int_0^t |S(x)|\,\mathrm{d}Q(x) < \infty$ for $t < \infty$, so (4.6.4) holds. □

Corollary 4.6.VI. *The d.f. F is uniquely determined by its IHF and conversely.*

This corollary is simply a formalization and extension of the fact that a renewal process is determined entirely by its lifetime d.f. The fact that the hazard measure is also the central concept in estimating the time to the next renewal has been shown already in Example 4.6(a) which we now continue but without any assumption of absolute continuity.

Example 4.6(a) (continued). Recall the setting leading to the density $q_x(y)$ earlier. If the lifetime has a jump at x, then we should think of the risk as having a δ-function component at x, the weight associated with the δ-function being given by $\Delta Q(x)$ as above. Then, in place of the survivor function $S_x(y)$ given earlier, we now appeal to the corresponding modification of (4.6.4), namely

$$S_x(y) = \prod_{x \le x_i \le x+y} \big(1 - \Delta Q(x_i)\big) \exp\Big(- \int_x^{x+y} \mathrm{d}Q_c(u)\Big). \qquad \square$$

In a Wold process, the risk has to be conditioned not only by the time since the last event but also by the length of the most recently observed complete interval as in the following example.

Example 4.6(b) *Wold process with exponential conditional distributions* (see Exercise 4.5.8). Wold (1948) and Cox (1955) both considered processes with Markov-dependent intervals, where the transition kernel has the form

$$P(x, \mathrm{d}y) = p(x,y)\,\mathrm{d}y = \lambda(x) \exp[-\lambda(x)y]\,\mathrm{d}y \qquad (x, y > 0),$$

corresponding to the assumption that, conditional on the length x of the last interval, the current interval is exponentially distributed with parameter $\lambda(x)$.

In this case, if we observe the process at time t and the length of the last completed interval is x, the risk is constant at $\lambda(x)$ until the occurrence of the next event. As a stochastic process, the conditional risk appears as a step function, constant over intervals, the constant for any one interval being a function of the length of the preceding interval. □

Clearly, the ideas in these two examples can be generalized to situations where the dependence on the past extends to more than just the time since the last event or the length of the last completed interval. Such extensions and further examples are explored in Chapters 7 and 14.

CHAPTER 5

Finite Point Processes

The Poisson process can be generalized in many directions. We have already discussed some consequences of relaxing the independency assumptions while retaining those of stationarity and orderliness of a point process on the line. In this chapter we examine generalizations in another direction, stemming from the observation in Chapter 2 that, for a Poisson process, conditional on the total number of points in a bounded region of time or space, the individual points can be treated as independently and identically distributed over the region. This prompts an alternative approach to specifying the structure of point processes in a bounded domain or, more generally, of any point process in which the total number of points is finite with probability 1. Such a process is called a *finite point process*.

Such finite point processes arise naturally as models for populations of animals, insects, and plants in the ecological field and as models for particle processes in physics, which was also the context of the first general theory of point processes given by Moyal (1962a) following earlier work by Yvon (1935), Bogoliubov (1946), Janossy (1950), Bhabha (1950) and Ramakrishnan (1950). More recently, spatial point processes have been extensively studied with an emphasis on finite models. Useful reviews can be found in Ripley (1981), Diggle (1983), Stoyan, Kendall and Mecke (1987, 1995), Baddeley and Møller (1989), Cressie (1991), Stoyan and Stoyan (1994), Baddeley *et al.* (1996), and Barndorff-Nielsen (1998), amongst others.

In this chapter, we give a somewhat informal introduction to concepts and structure theorems for finite point processes, with a sketch of some of their applications. In contrast to the methods of the previous two chapters, the order properties of the real line here play no role in the discussion, and the theory can be developed as easily for a general state space as it can for the real line. In this sense, the present chapter serves as a precursor to the general theory developed more systematically in Volume Two.

The approach we take is first to specify the distribution of the total number N of points, and then, given N, to specify the joint distribution of the N points over the region. This leads to a treatment of point process probabilities as probability measures over the space $\mathcal{X}^{\cup}$ introduced formally above Proposition 5.3.II and of the associated battery of Janossy measures, moment measures, cumulant measures, etc., all of which are recurrent themes in the development of the general theory.

A special feature of the treatment of finite point processes is its dependence on combinatorial arguments. The reader may find it helpful to brush up on the definitions of binomial and multinomial coefficients and their relation to the number of ways of sorting a set of objects into various subsets. Closely related to these ideas are the results collected together in Section 5.2 concerning some basic tools for handling discrete distributions: factorial moments and cumulants and their relation with probability generating functions. The importance of this material for the theory of point processes would be hard to overemphasize. Most of the results of this chapter, and much of the general theory also, may be seen as extensions of the results for discrete distributions summarized in that section.

5.1. An Elementary Example: Independently and Identically Distributed Clusters

We start with an elementary example that may help to illustrate and motivate the more general discussion. Let a random number N of particles be independently and identically distributed (i.i.d.) over a Euclidean space $\mathcal{X}$ according to some common probability measure $F(\cdot)$ on the Borel sets of $\mathcal{X}$. Then, given N, the number of particles in any subregion A is found by 'binomial sampling': each particle, independently of the others, may fall in A with probability $p = F(A)$, so, conditional on N, the number of particles in A has the binomial distribution

$$p(n; A \mid N) = \binom{N}{n} (F(A))^n (1 - F(A))^{N-n}.$$

Similarly, given any finite partition $A_1, \ldots, A_k$ of $\mathcal{X}$, the joint distribution of the number of particles is given by the multinomial probability

$$p(n_1, \ldots, n_k; A_1, \ldots, A_k \mid N) = \binom{N}{n_1 \ \cdots \ n_k} (F(A_1))^{n_1} \ldots (F(A_k))^{n_k}.$$

Unconditionally, the joint distribution of the numbers $N(A_1), \ldots, N(A_k)$ of particles in $A_1, \ldots, A_k$ is found by averaging over N:

$$\Pr\{N(A_i) = n_i \, (i = 1, \ldots, k)\} = \sum_{n=0}^{\infty} \Pr\{N = n\} \, p(n_1, \ldots, n_k; A_1, \ldots, A_k \mid n).$$

The procedure just outlined is most readily carried out in terms of probability generating functions (p.g.f.s). Let $P_N(z) = \mathrm{E}(z^N)$, and write for convenience $p_i = F(A_i)$. Then, the joint p.g.f. of $N(A_i)$ $(i = 1, \ldots, k)$ is

$$\begin{aligned} P(A_1, \ldots, A_k; z_1, \ldots, z_k) &\equiv \mathrm{E}(z_1^{N(A_1)} \cdots z_k^{N(A_k)}) \\ &= P_N(p_1 z_1 + \cdots + p_k z_k). \end{aligned} \tag{5.1.1}$$

More generally, for $A_1, \ldots, A_k$ just a set of mutually disjoint subregions,

$$P(A_1, \ldots, A_k; z_1, \ldots, z_k) = P_N(p_1 z_1 + \cdots + p_k z_k + (1 - p_1 - \cdots - p_k)); \tag{5.1.2}$$

in effect, we have introduced a further subset $A_{k+1} = (A_1 \cup \cdots \cup A_k)^c$ and set $z_{k+1} = 1$ on A_{k+1}.

As special cases, when N is Poisson-distributed with parameter λ, the $N(A_i)$ are independent Poisson random variables with parameters $\lambda F(A_i)$. In this case, (5.1.1) reduces to the identity

$$\begin{aligned} P(A_1, \ldots, A_k; z_1, \ldots, z_k) &= \exp\left(\lambda \left[\sum_{i=1}^{k} z_i f(A_i) - 1\right]\right) \\ &= \prod_{i=1}^{k} \exp[\lambda F(A_i)(z_i - 1)]. \end{aligned}$$

When N has a negative binomial distribution on $\{0, 1, \ldots\}$ so that $P_N(z) = (1 + \mu(1 - z))^{-\alpha}$ for some μ, $\alpha > 0$, $\{N(A_i)\}$ is a set of mutually correlated binomial random variables with joint p.g.f.

$$P(A_1, \ldots, A_k; z_1, \ldots, z_k) = \left(1 + \mu \sum_{i=1}^{k} F(A_i)(1 - z_i)\right)^{-\alpha}.$$

In particular, from (5.1.2), the distribution of $N(A_i)$ itself has the p.g.f.

$$P(A_i; z) = [1 + \mu F(A_i)(1 - z)]^{-\alpha}$$

and is again negative binomial with parameters $\mu F(A_i)$, α.

It is not only the distributions of the $N(A_i)$ that may be of interest but also their moments. Consider, for example, the problem of finding the covariance of the number of points in two complementary subsets A_1, $A_2 = A_1^c$. For any given N, we have from the binomial sampling property that

$$\mathrm{E}[N(A_1)N(A_2) \mid N] = N(N-1)F(A_1)(1 - F(A_1)) = N(N-1)F(A_1)F(A_2).$$

Hence,

$$\mathrm{E}(N(A_1)N(A_2)) = m_{[2]} F(A_1) F(A_2) \tag{5.1.3}$$

and

$$\mathrm{cov}(N(A_1), N(A_2)) = c_{[2]} F(A_1) F(A_2), \tag{5.1.4}$$

where $m_{[2]}$ is the second factorial moment, and $c_{[2]}$ the second factorial cumulant, of the total number N of points. In the Poisson case, the covariance is

zero, and in the negative binomial case it is positive; both contrast with the more familiar case of fixed N when the covariance is clearly negative.

Note that both the second moment and the covariance have the form of a measure evaluated on the product set $A_1 \times A_2$. This is also the case in general and anticipates the introduction of the factorial moment and cumulant measures in Section 5.4.

5.2. Factorial Moments, Cumulants, and Generating Function Relations for Discrete Distributions

Factorial moments and cumulants are natural tools for handling nonnegative integer-valued random variables, a characteristic they bequeath to their offspring, the factorial moment and cumulant measures, in the point process context. We begin by recalling some basic definitions.

For any integers n and r, the *factorial powers* of n, written $n^{[r]}$, may be defined by

$$n^{[r]} = \begin{cases} n(n-1)\cdots(n-r+1) & (r = 0, \ldots, n), \\ 0 & (r > n). \end{cases}$$

We then have the following definition.

Definition 5.2.I. *For $r = 0, 1, \ldots,$ the* rth factorial moment $m_{[r]}$ *of the nonnegative integer-valued random variable N is $m_{[r]} \equiv \mathrm{E}(N^{[r]})$.*

Thus, when N has probability distribution $\{p_n\} = \{\Pr\{N = n\}\}$,

$$m_{[r]} = \sum_{n=0}^{\infty} n^{[r]} p_n. \tag{5.2.1}$$

Consequently, when the distribution is concentrated on a finite range $0, 1, \ldots, n_0$, all factorial moments of order larger than n_0 are zero.

It is useful to be able to convert from factorial moments to ordinary moments and back again. The coefficients that arise in these conversions are the *Stirling numbers of the first and second kinds*, defined, respectively, as the coefficients arising in the expansion of $x^{[r]}$ and x^r in powers or factorial powers of x, where, by analogy with the definition of $n^{[r]}$,

$$x^{[r]} = x(x-1)\cdots(x-r+1)$$

for any real x and positive integer r. We follow the notation of David and Barton (1962) in denoting them by $D_{j,r}$ and $\Delta_{j,r}$.

Definition 5.2.II. *The* Stirling numbers of the first kind $D_{j,r}$ and second kind $\Delta_{j,r}$ *are defined by the relations*

$$n^{[r]} = \sum_{j=1}^{r} D_{j,r}(-1)^{r-j} n^j \qquad (n \geq r) \tag{5.2.2}$$

and

$$n^r = \sum_{j=1}^{r} \Delta_{j,r} n^{[j]} \qquad (n \geq r). \tag{5.2.3}$$

Replacing n in (5.2.2) and (5.2.3) by the random variable N and taking expectations, we obtain the corresponding relations between moments:

$$m_{[r]} = \sum_{j=1}^{r} D_{j,r} m_j (-1)^{r-j}, \tag{5.2.4}$$

$$m_r \equiv \mathrm{E}(N^r) = \sum_{j=1}^{r} \Delta_{j,r} m_{[j]}. \tag{5.2.5}$$

It is clear that, for a nonnegative random variable, the rth factorial moment is finite if and only if the ordinary rth moment is finite.

Some useful recurrence relations for the Stirling numbers are given in Exercise 5.2.1. For further properties, relation to Bernoulli numbers, and so on, see David and Barton (1962, Chapter 15) and texts on finite differences.

The factorial moments of the random variable N are related to the Taylor series expansion of the p.g.f.

$$P(z) = \mathrm{E}(z^N) \qquad (|z| \leq 1)$$

about $z = 1$ in much the same way as the ordinary moments arise in the expansion of the characteristic or moment generating function about the origin.

Proposition 5.2.III. *For a nonnegative integer-valued random variable N whose k*th *factorial moment is finite, the p.g.f. is expressible as*

$$P(1+\eta) = 1 + \sum_{r=1}^{k} \frac{m_{[r]} \eta^r}{r!} + o(\eta^k) \tag{5.2.6}$$

for all η such that $|1+\eta| \leq 1$. The complete Taylor series expansion of the p.g.f.,

$$P(1+\eta) = 1 + \sum_{r=1}^{\infty} \frac{m_{[r]} \eta^r}{r!}\,, \tag{5.2.7}$$

is valid for some nonzero η if and only if all moments exist and the series in (5.2.7) *has nonzero radius of convergence in η; equivalently, if and only if the p.g.f. $P(z)$ is analytic in a disk $|z| < 1 + \epsilon$ for some $\epsilon > 0$. Equation* (5.2.7) *then holds for $|\eta| < \epsilon$.*

Proof. To establish (5.2.6), write

$$(1+\eta)^N = 1 + \sum_{r=1}^{k} \frac{N^{[r]} \eta^r}{r!} + R_k(N, \eta) \qquad (k = 1, 2, \ldots)$$

for remainder terms $R_k(N, \eta)$ that we now investigate. For $k = 0$, set

$$R_0(N, \eta) = (1+\eta)^N - 1$$

and observe that $|R_0(N,\eta)| \le 2$ under the condition of the theorem that $|1+\eta| \le 1$. For general $k = 1, 2, \ldots$, repeated integration of $R_0(N,\cdot)$ shows that

$$\big|R_k(N,\eta)/\eta^k\big| \le 2N^{[k]}/k! \qquad (|1+\eta| \le 1).$$

Since the left-hand side of this inequality $\to 0$ ($\eta \to 0$) for each fixed N and the right-hand side has finite expectation under the assumption of the theorem, it follows by dominated convergence that $\mathrm{E}\big(R_k(N,\eta)\big) = o(\eta^k)$, which is the result required.

To establish (5.2.7), consider the binomial expansion

$$(1+\eta)^N = 1 + \sum_{r=1}^{\infty} \frac{N^{[r]}\eta^r}{r!}\,.$$

For $\eta > 0$, the finiteness of the expectation on the left is equivalent to requiring the p.g.f. to be analytic for $|z| < 1+\eta$. When this condition is satisfied, it follows from Fubini's theorem that for such η the expectation can be taken inside the summation on the right, leading to the right-hand side of (5.2.7).

Conversely, suppose all moments exist and that the sum on the right-hand side of (5.2.7) is at least conditionally convergent for some nonzero η_0. Then $m_{[r]}\eta_0^r/r! \to 0$ as $r \to \infty$, and it follows from a standard power series argument that the series in (5.2.7) is absolutely convergent for $|\eta| < |\eta_0|$ and so defines an analytic function of η there. Since each $m_{[r]} = \mathrm{E}(N^{[r]})$ is nonnegative, we can now take any positive $\eta < |\eta_0|$ and use Fubini's theorem to reverse the argument used earlier to deduce that because (5.2.7) holds for all $0 \le \eta \le |\eta_0|$, $P(z)$, being a power series with nonnegative coefficients, has its first singularity on the positive half-line outside $|z| < 1 + |\eta_0|$. □

In the sequel, we also require the version of Proposition 5.2.III in which the remainder term is bounded by a term proportional to the $(k+1)$th moment. The proof, which is along similar lines, is left to the reader. An alternative approach is indicated in Exercise 5.2.2.

A similar expansion holds for $\log P(1+\eta)$, the coefficients of $\eta^r/r!$ being the *factorial cumulants* $c_{[r]}$ ($r = 1, 2, \ldots$). If $P(\cdot)$ is analytic in a disk as below (5.2.7), then the infinite expansion

$$\log P(1+\eta) = \sum_{r=1}^{\infty} \frac{c_{[r]}\eta^r}{r!} \tag{5.2.8a}$$

is valid, while under the more limited assumption that $m_k < \infty$, we have the finite Taylor series expansion

$$\log P(1+\eta) = \sum_{r=1}^{k} \frac{c_{[r]}\eta^r}{r!} + o(\eta^k) \qquad (\eta \to 0), \tag{5.2.8b}$$

valid for $|1+\eta| < 1$; verification is left to the reader.

The factorial cumulants are related to the factorial moments by the same relations as hold between the ordinary cumulants and moments. The first few relations between the ordinary cumulants c_r, central moments m'_r, and factorial moments and cumulants are useful to list as below:

$$c_{[1]} = c_1 = \mu = m_{[1]}\,, \tag{5.2.9a}$$

$$c_{[2]} = c_2 - c_1 = \sigma^2 - \mu = m_{[2]} - m_{[1]}^2\,, \tag{5.2.9b}$$

$$c_{[3]} = c_3 - 3c_2 + 2c_1 = m'_3 - 3\sigma^2 + 2\mu = m_{[3]} - 3m_{[2]}m_{[1]} + 2m_{[1]}^3. \tag{5.2.9c}$$

Generally, the factorial moments and cumulants provide a much simpler description of the moment properties of a discrete distribution than do the ordinary moments. In particular, for the Poisson distribution $\{p_n(\lambda)\}$,

$$m_{[r]} = \lambda^r, \qquad c_{[1]} = \lambda, \qquad c_{[r]} = 0 \quad (r = 2, 3, \ldots).$$

This vanishing of the factorial cumulants of the Poisson distribution is reminiscent of the vanishing of the ordinary cumulants of the normal distribution and is perhaps one indication of why the Poisson process plays such an outstanding role in the theory of point processes.

There are in fact four expansions of the p.g.f. of possible interest, according to whether we expand $P(z)$ itself or its logarithm and whether the expansion is about $z = 0$ or $z = 1$. The expansions about $z = 1$ yield the factorial moments and factorial cumulants, and the expansion of $P(z)$ about $z = 0$ yields the probability distribution $\{p_n\}$. This leaves the expansion of $\log P(z)$ about $z = 0$, an expansion that, while rarely used, has an important interpretation in the case of an infinitely divisible (compound Poisson) distribution. Since the analogous expansion for the probability generating functional (p.g.fl.) of a point process is also important, again in the context of infinite divisibility, we now consider the last case in some detail.

Proposition 5.2.IV. *If $p_0 > 0$, the p.g.f. $P(\cdot)$ can be written in the form*

$$\log P(z) = -q_0 + \sum_{n=1}^{\infty} q_n z^n \qquad (|z| < R) \tag{5.2.10}$$

where $p_0 = \mathrm{e}^{-q_0}$ and R is the distance from the origin to the nearest zero or singularity of $P(z)$. When $P(\cdot)$ is the p.g.f. of a compound Poisson distribution, the terms q_n are nonnegative and $q_0 = \sum_{n=1}^{\infty} q_n$, so the sequence $\{\pi_n : n = 1, 2, \ldots\} \equiv \{q_n/q_0\}$ can be interpreted as the probability distribution of the cluster size, given that the cluster is nonempty; in this case, (5.2.10) *can be rewritten as*

$$\log P(z) = -q_0 \sum_{n=1}^{\infty} \pi_n (1 - z^n) \qquad (|z| < R).$$

Proof. The structure of the compound Poisson distribution follows from analysis in Chapter 2 (see Theorem 2.2.II and Exercise 2.2.2). The other remarks are standard properties of power series expansions of analytic functions. □

EXAMPLE 5.2(a) *Negative binomial distribution and generating functions.* To illustrate these various expansions consider the p.g.f. of the negative binomial distribution,

$$P(z) = [1 + \mu(1 - z)]^{-\alpha} \qquad (\mu > 0, \ \ \alpha > 0, \ \ |z| \le 1).$$

Putting $z = 1 + \eta$, we find

$$P(1 + \eta) = (1 - \mu\eta)^{-\alpha} = 1 + \sum_{r=1}^{\infty} \binom{\alpha + r - 1}{r} \mu^r \eta^r$$

so that

$$m_{[r]} = \alpha(\alpha + 1) \cdots (\alpha + r - 1)\mu^r.$$

Taking logarithms,

$$\log P(1 + \eta) = -\alpha \log(1 - \mu\eta) = \alpha \sum_{r=1}^{\infty} \frac{\mu^r \eta^r}{r},$$

and hence

$$c_{[r]} = (r - 1)!\, \alpha\mu^r.$$

For the expansions about $z = 0$, we have

$$P(z) = \frac{1}{(1 + \mu)^\alpha} \left(1 - \frac{\mu z}{1 + \mu}\right)^{-\alpha} = \frac{1}{(1 + \mu)^\alpha} \sum_{n=0}^{\infty} \binom{\alpha + n - 1}{n} \left(\frac{\mu z}{1 + \mu}\right)^n,$$

so

$$p_n = \binom{\alpha + n - 1}{n} \frac{1}{(1 + \mu)^\alpha} \left(\frac{\mu}{1 + \mu}\right)^n,$$

and

$$\begin{aligned} \log P(z) &= -\alpha \log(1 + \mu) - \alpha \log\left(1 - \frac{\mu z}{1 + \mu}\right) \\ &= -[\alpha \log(1 + \mu)]\left(1 - \sum_{n=1}^{\infty} \pi_n z^n\right), \end{aligned}$$

where $\pi_n = [n \log(1 + \mu)]^{-1}[\mu/(1 + \mu)]^n$. Clearly, these $\{\pi_n\}$ constitute a probability distribution, namely the logarithmic distribution, illustrating the well-known fact that the negative binomial is infinitely divisible and hence must be expressible as a compound Poisson distribution. □

Corresponding to the four possible expansions referred to above, there are twelve sets of conversion relations between the different coefficients. One of these, the expression for factorial moments in terms of the probabilities, is a matter of definition: what can be said about the others?

Formally, either expansion about $z = 1$ can be converted to an expansion about $z = 0$ by a change of variable and expansion, for example, in (formally) expressing the probabilities in terms of the factorial moments via

$$P(z) = 1 + \sum_{r=1}^{\infty} \frac{m_{[r]}(z - 1)^r}{r!};$$

expanding $(z-1)^r$ and equating coefficients of z^n, we obtain

$$p_n = \sum_{r \geq n} (-1)^{r-n} \frac{m_{[r]}}{r!} \binom{r}{n}$$

or, in the more symmetrical form,

$$n!\, p_n = \sum_{r=n}^{\infty} (-1)^{r-n} \frac{m_{[r]}}{(r-n)!} = \sum_{r=0}^{\infty} (-1)^r \frac{m_{[n+r]}}{r!}\,. \tag{5.2.11}$$

This relation may be compared with its converse

$$m_{[r]} = \sum_{n=r}^{\infty} n^{[r]} p_n = \sum_{n=0}^{\infty} \frac{J_{r+n}}{n!}\,, \tag{5.2.12}$$

where $J_{n+r} = (n+r)!\, P_{n+r}$. Thus, to display the symmetry in these (formal) relations to best advantage, we need to use the quantities J_n, which are analogues of the Janossy measures to be introduced in Section 5.3.

Under what circumstances can the converse relation (5.2.11) be established rigorously? For the derivation above to be valid, we must be able to expand $P(z)$ about $z = 1$ in a disk $|z-1| < 1+\epsilon$ for some $\epsilon > 0$, requiring $P(z)$ itself to be analytic at all points on the line segment $(-\epsilon, 2+\epsilon)$. Since $P(z)$ has nonnegative coefficients, its radius of convergence is determined by the first singularity on the positive real axis. Consequently, in order for (5.2.11) to hold for all $r = 1, 2, \ldots$, it is sufficient that $P(z)$ should be analytic in the disk $|z| < 2+\epsilon$ for some $\epsilon > 0$.

A finite version of (5.2.11) with remainder term is due to Fréchet (1940); extensions are given in Takacs (1967) and Galambos (1975) (see also Daley and Narayan, 1980). We give a simple result in the theorem below, with some extensions left to Exercises 5.2.2–4.

Proposition 5.2.V. *If the distribution $\{p_n\}$ has all its moments finite and its p.g.f. $P(z)$ is convergent in a disk $|z| < 2+\epsilon$ for some $\epsilon > 0$, then* (5.2.11) *holds. Without assuming such analyticity, the finiteness of $m_{[k]}$ ensures that for integers $n = 0, 1, \ldots, k-1$,*

$$n!\, p_n = \sum_{r=n}^{k-1} (-1)^{r-n} \frac{m_{[r]}}{(r-n)!} + R_k^{(n)}, \tag{5.2.13a}$$

where

$$0 \leq (-1)^{k-n} R_k^{(n)} \leq m_{[k]}/(k-n)!\,. \tag{5.2.13b}$$

If all moments are finite and for some integer n_0

$$m_{[k]} = o\big((k-n_0)!\big) \qquad (k \to \infty), \tag{5.2.14a}$$

then

$$\lim_{k\to\infty} \sum_{r=n}^{k} (-1)^{r-n} m_{[r]}/(r-n)! \tag{5.2.14b}$$

exists for $n = 0, 1, \ldots, n_0$ and the formal relation (5.2.11) *holds for such n.*

PROOF. When $P(z)$ is analytic for $|z| < 2 + \epsilon$, the expansion

$$P(z) = \sum_{r=0}^{\infty} \frac{m_{[r]}(z-1)^r}{r!}$$

is valid for $|z-1| < 1+\epsilon$, within which region, and at $z = 0$ in particular, it can be differentiated n times, leading at once to (5.2.11).

Under the weaker condition that $m[k] < \infty$, n-fold differentiation in the definition $P(z) = \mathrm{E}(z^N)$ is possible for all $|z| \le 1$ for $n = 1, \ldots, k$, leading to $P^{(n)}(z) = \mathrm{E}(N^{(n)} z^{N-n})$. Now $P^{(n)}(z)$ is $(k-n)$ times differentiable in $|z| \le 1$, so the Taylor series expansion

$$P^{(n)}(z) = \sum_{r=0}^{k-n-1} \frac{(z-1)^r P^{(n+r)}(1)}{r!} + \frac{(z-1)^{k-n} P^{(k)}(1+(z-1)\nu)}{(k-n)!}$$

holds for real z in $|z| \le 1$ for some $\nu \equiv \nu(z)$ in $(0,1)$. In particular, (5.2.13a) results on putting $z = 0$ with

$$R_k^{(n)} = (-1)^{k-n} \frac{\mathrm{E}(N^{(k)}(1-\nu)^{N-k})}{(k-n)!}\,,$$

from which relation the inequalities in (5.2.13b) follow. When (5.2.14) holds, $R_k^{(n)} \to 0$ $(k \to \infty)$ for each fixed n, and hence (5.2.11) holds in the sense indicated. □

Special cases of (5.2.13) give the Bonferroni inequalities (see Exercise 5.2.5).

Similar relations can be obtained between the factorial cumulants and the quantities π_n of Proposition 5.2.IV. Thus, when $\log P(z)$ is analytic in a disk $|z| < 1 + \epsilon$ for some $\epsilon > 0$, r-fold differentiation of (5.2.10) and then setting $z = 1$ yields

$$c_{[r]} = \sum_{n=r}^{\infty} q_n n^{[r]} = q_0 \mu_{[r]}\,, \tag{5.2.15}$$

where $\mu_{[r]}$ in the case of a compound Poisson process is the rth factorial moment of the cluster-size distribution. Reversing the exercise, when $\log P(z)$ is analytic in the disk $|z| < 2 + \epsilon$, we have [see the derivation of (5.2.11)]

$$n!\, q_n = \sum_{r=n}^{\infty} (-1)^{r-n} \frac{c_{[r]}}{(r-n)!}\,. \tag{5.2.16}$$

The most difficult relations to treat in a general form are those between the moments and cumulants, or between the $\{p_n\}$ and the $\{q_n\}$; these arise from taking exponentials or logarithms of a given series and expanding it by formal manipulation. The feature of these relations is that they involve partitions. For given positive integers j and k with $j \le k$, we define a *j-partition of k* as a partition of the set of k numbers $\{1, \ldots, k\}$ into j nonempty subsets.

Let $\mathcal{P}_{jk}$ denote the collection of all such j-partitions and write $\mathcal{T} = \{S_1(\mathcal{T}), \ldots, S_j(\mathcal{T})\}$ for an element of $\mathcal{P}_{jk}$, noting that the order in which the subsets $S_\cdot(\mathcal{T})$ are labelled or written is immaterial. Thus, for example, the collection of sets $\{1,2,4\}$, $\{3,5\}$, $\{6,8\}$, $\{7\}$ is a 4-partition of 8 and is the same as $\{1,2,4\}$, $\{6,8\}$, $\{7\}$, $\{3,5\}$. The following lemma is basic (see e.g. Andrews, 1976); in it, $|S_j(\mathcal{T})|$ denotes the number of elements in $S_j(\mathcal{T}) \subset \{1,\ldots,k\}$.

Lemma 5.2.VI. *Let* $\{c_j : j = 1,2,\ldots\}$ *be a sequence satisfying* $\sum_{j=1}^{\infty} |c_j|/j! < \infty$. *Then*

$$\exp\left(\sum_{j=1}^{\infty} \frac{c_j z^j}{j!}\right) = \sum_{k=0}^{\infty} \frac{d_k z^k}{k!} \qquad (\text{all } |z| \le 1), \tag{5.2.17}$$

where $d_0 = 1$ *and for* $k = 1,2,\ldots,$

$$d_k = \sum_{j=1}^{k} \sum_{\mathcal{T}\in\mathcal{P}_{jk}} \prod_{i=1}^{j} c_{|S_i(\mathcal{T})|}\,, \tag{5.2.18}$$

$$c_k = \sum_{j=1}^{k} (-1)^{j-1}(j-1)! \sum_{\mathcal{T}\in\mathcal{P}_{jk}} \prod_{i=1}^{j} d_{|S_i(\mathcal{T})|}\,. \tag{5.2.19}$$

Proof. Establishing (5.2.18) and (5.2.19) is essentially a matter of counting terms. For (5.2.18), consider the expansion $1+\Sigma+\Sigma^2/2!+\cdots$ of the exponential function in (5.2.17) (here, $\Sigma = \sum_{j=1}^{\infty} c_j z^j/j!$), and concentrate attention on all the terms in a specified product of coefficients such as $c_3c_2^2c_1$. Observe first that such terms involve z to the power of the sum of the indices, here $3+2+2+1 = 8$, and thus they contribute to the term d_8. Second, if we transfer the coefficient $1/k!$ of $d_k z^k$ to the multiplier $k!$ on the opposite side, each particular term $c_3c_2^2c_1$ is then multiplied by the ratio of factorials $8!/3!\,2!\,2!\,1!$ arising from the factorials associated with the c_j and d_k. Third, the number of such terms obtained from expanding Σ^4 equals the multinomial coefficient $4!/1!\,2!\,1!$, which on division by the factorial $4!$ from the expansion of $\exp(\Sigma)$ leaves the factor $1!/1!\,2!\,1!$. Thus, altogether the contribution of the coefficient of $c_3c_2^2c_1$ to d_8 is $8!/\{(3!\,2!\,2!\,1!)\,(1!\,2!\,1!)\}$.

On the other hand, in the expression asserted for d_k in (5.2.18), we have to look at 4-partitions of 8 into subsets of sizes 3, 2, 2, 1. The number of such subsets is just $8!/3!\,2!\,2!\,1!$, which must be divided by $2!$ because there are two subsets of size 2. Thus, the coefficient of $c_3c_2^2c_1$ is of the form implied by (5.2.18).

Arguing this way in general establishes (5.2.18), and a similar kind of argument leads to (5.2.19). □

We remark that the advantage of working with j-partitions, rather than with additive partitions as in David and Barton (1962), is that the counting procedure automatically takes into account repeated terms without requiring explicit notation for the number of repetitions; such notation would make (5.2.18) and (5.2.19) appear much more cumbersome. Examples of full expansions are given in Exercises 5.2.6–8.

Corollary 5.2.VII.

(a) *Factorial moments* $m_{[k]}$ *and factorial cumulants* $c_{[k]}$ *are related as in* (5.2.18) *and* (5.2.19) *via the substitutions* $c_j = c_{[j]}$ *and* $d_k = m_{[k]}$.

(b) *In equation* (5.2.10), *the probabilities* p_n *and* q_n *are also related as at* (5.2.18) *and* (5.2.19) *with* $c_j = j!\, q_j/(-\log p_0)$ *and* $d_k = k!\, p_k/p_0$.

Exercises and Complements to Section 5.2

5.2.1 *Recurrence relations for Stirling numbers.* Use $n^{[r+1]} = (n-r)n^{[r]}$ to show that

$$\Delta_{j,r+1} = j\Delta_{j,r} + \Delta_{j-1,r}, \quad \Delta_{1r} = 1 \quad (r \geq 1), \qquad \Delta_{j0} = 0 \quad (j \geq 1),$$
$$D_{j,r+1} = rD_{j,r} + D_{j-1,r}, \quad D_{0r} = 0 \quad (r \geq 1), \quad D_{11} = 1,\ D_{j1} = 0 \ \ (j \geq 2).$$

5.2.2 Show that when $P(z)$ is any p.g.f. with finite first moment $P'(1)$, the function $(1-P(z))/P'(1)(1-z)$ is also a p.g.f. Use this fact in an induction argument to show that (see Proposition 5.2.III) when $m_{[k]} = P^{(k)}(1) < \infty$, the function $m_k(z)$ in the expansion

$$P(z) = 1 + \sum_{r=1}^{k-1} \frac{(z-1)^r m_{[r]}}{r!} + \frac{(z-1)^k m_k(z)}{k!}$$

equals $m_{[k]}$ times a p.g.f. Since $m_k(z) = m_{[k]} + o(1)$ as $z \to 1$ through values $|z| \leq 1$, (5.2.6) follows, as well as the alternative version with remainder bounded by $m_{[k]}$. Equations (5.2.13) can also be derived by n-fold differentiation of an expansion to $k-n$ terms (e.g. Daley and Narayan, 1980).

5.2.3 Let the nonnegative integer-valued r.v. N have all factorial moments $m_{[r]}$ finite and $\limsup_{r\to\infty}(m_{[r]}/r!)^{1/r} = 1/\epsilon$ for some $\epsilon > 0$. Show that the p.g.f. $P(z)$ of N has radius of convergence $1+\epsilon$, and hence deduce that the moments $m_{[r]}$ determine the distribution of N uniquely. Relate $P(z)$ to a moment generating function and deduce that $1+\epsilon = \exp\epsilon'$, where $1/\epsilon' \equiv \limsup_{r\to\infty}(m_r/r!)^{1/r}$.

5.2.4 (*Continuation*). By using an analytic continuation technique (see Takacs, 1965), show that when $\epsilon > 0$ and for any nonnegative $z > \epsilon^{-2} - 1$,

$$p_n = \sum_{r=n}^{\infty} \binom{r}{n} \frac{1}{(1+z)^{r+1}} \sum_{s=n}^{r} (-1)^{s-n} \binom{r-n}{s-n} z^{r-s} \frac{m_{[s]}}{s!}\,.$$

5.2.5 *Bonferroni inequalities.* Let the r.v. N count the number of occurrences amongst a given set of ν events $A_1, \ldots, A_\nu$. Show that

$$S_r \equiv \sum_{(r)} \Pr(A_i \cap A_j \cap \cdots) = \mathrm{E}(N^{(r)})/r!\,,$$

where the summation $\sum_{(r)}$ extends over all $\binom{\nu}{r}$ distinct subsets $\{i, j, \ldots\}$ of size r from the index set $\{1, \ldots, \nu\}$. [*Hint*: Using indicator r.v.s, write

$$N^{(r)} = r! \sum_{(r)} I(A_i \cap A_j \cap \cdots),$$

where the term $r!$ arises from the $r!$ ordered subsets of $\{1, \ldots, \nu\}$ yielding the same (unordered) subset $\{i, j, \ldots\}$ containing r indices.] Deduce from (5.2.13) the Bonferroni inequalities

$$0 \le S_n - \binom{n+1}{1} S_{n+1} + \cdots + \binom{n+k}{k} S_{n+k} - p_n \le \binom{n+k+1}{k+1} S_{n+k+1},$$

where k is an even integer (see e.g. Moran, 1968, pp. 25–31).

5.2.6 For given positive integers j and k with $j \le k$, define $\mathcal{P}(j,k) = \{$positive integers $\{r_1, \ldots, r_p\}$ and $\{\pi_1, \ldots, \pi_p\}$ such that $\sum_{i=1}^{p} \pi_i = j$, $\sum_{i=1}^{p} \pi_i r_i = k\}$ = set of all j-partitions of k. Write the series (5.2.7) in the form $P = 1 + \Sigma$ so that

$$\log P(z) = \Sigma - \Sigma^2/2 + \Sigma^3/3 - \cdots,$$

and expand the series Σ^n as a multinomial expansion. By equating coefficients of z^k, show formally that the factorial cumulants in (5.2.8) are given by

$$c_{[k]} = k! \sum_{j=1}^{k} (-1)^{j-1} (j-1)! \sum_{\mathcal{P}(j,k)} \frac{1}{\pi_1!} \left(\frac{m_{[r_1]}}{r_1!} \right)^{\pi_1} \cdots \frac{1}{\pi_p!} \left(\frac{m_{[r_p]}}{r_p!} \right)^{\pi_p}.$$

5.2.7 Apply Lemma 5.2.VI to show that

$$c_{[4]} = m_{[4]} - 4m_{[3]}m_{[1]} - 3m_{[2]}^2 + 12m_{[2]}m_{[1]}^2 - 6m_{[1]}^4,$$
$$m_{[4]} = c_{[4]} + 4c_{[3]}c_{[1]} + 3c_{[2]}^2 + 6c_{[2]}c_{[1]}^2 + c_{[1]}^4.$$

5.2.8 Investigate the use of Lemma 5.2.VI in deriving explicit expressions for probabilities of

(i) the 'doubly Poisson' compound Poisson distribution with p.g.f. $P(z) = \exp\{-\mu[1 - \exp(-\lambda(1-z))]\}$;

(ii) the Hermite distribution with p.g.f. $P(z) = \exp(az + bz^2)$ for appropriate constants a and b (see Milne and Westcott, 1993).

5.3. The General Finite Point Process: Definitions and Distributions

We now drop any special assumptions and suppose only that the following conditions hold concerning a finite point process.

Conditions 5.3.I. (a) *The points are located in a complete separable metric space (c.s.m.s.) $\mathcal{X}$, as, for example, $\mathcal{X} = \mathbb{R}^d$.*
(b) *A distribution $\{p_n\}$ $(n = 0, 1, \ldots)$ is given determining the total number of points in the population, with $\sum_{n=0}^{\infty} p_n = 1$.*
(c) *For each integer $n \ge 1$, a probability distribution $\Pi_n(\cdot)$ is given on the Borel sets of $\mathcal{X}^{(n)} \equiv \mathcal{X} \times \cdots \times \mathcal{X}$, and it determines the joint distribution of the positions of the points of the process, given that their total number is n.*

Such a definition is both natural and powerful. In particular, it provides a *constructive* definition that could be used to simulate the process: first, generate a random number N according to the distribution $\{p_n\}$ (and note that $\Pr\{0 \le N < \infty\} = 1$), and then, supposing $N = n$ and excepting the case $n = 0$ in which case there is nothing else to do, generate a random vector $(x_1, \ldots, x_n)$ according to the distribution $\Pi_n(\cdot)$.

At this stage, the distinction between ordered and unordered sets of points should be clarified. In talking of stochastic point processes, we make the tacit assumption that we are dealing with *unordered* sets of points: points play the role of *locations* at which a given set of particles might be found. We talk of the probability of finding a given number k of points in a set A: we do not give names to the individual points and ask for the probability of finding k specified individuals within the set A. Nevertheless, this latter approach is quite possible (indeed, natural) in contexts where the points refer to individual particles, animals, plants, and so on. Moreover, it is actually this latter point of view that is implicit in Conditions 5.3.I, for as yet there is nothing in them to prevent x_1, say—that is, the first point or particle named—from taking its place preferentially in some part of the space, leaving the other particles to distribute themselves elsewhere.

To be consistent with treating point processes as a theory of unordered sets, we stipulate that the distributions $\Pi_n(\cdot)$ should give equal weight to all $n!$ permutations of the coordinates $(x_1, \ldots, x_n)$, i.e. $\Pi_n(\cdot)$ *should be symmetric*. If this is not already the case in Condition 5.3.I(c), it is easily achieved by introducing the symmetrized form for any partition $(A_1, \ldots, A_n)$ of $\mathcal{X}$,

$$\Pi_n^{\mathrm{sym}}(A_1 \times \cdots \times A_n) = \frac{1}{n!} \sum_{\mathrm{perm}} \Pi_n(A_{i_1} \times \cdots \times A_{i_n}), \tag{5.3.1}$$

where the summation $\sum_{\mathrm{perm}}$ is taken over all $n!$ permutations $(i_1, \ldots, i_n)$ of the integers $(1, \ldots, n)$ and the normalizing factor $1/n!$ ensures that the resulting measure still has total mass unity.

An alternative notation, which has some advantages in simplifying combinatorial formulae, utilizes the nonprobability measures

$$\begin{aligned} J_n(A_1 \times \cdots \times A_n) &= p_n \sum_{\mathrm{perm}} \Pi_n(A_{i_1} \times \cdots \times A_{i_n}) \\ &= n!\, p_n \Pi_n^{\mathrm{sym}}(A_1 \times \cdots \times A_n). \end{aligned} \tag{5.3.2}$$

We follow Srinivasan (1969) in referring to these as *Janossy measures* after their introduction by Janossy (1950) in the context of particle showers. By contrast, Yvon (1935), Bogoliubov (1946) and Bhabha (1950) worked with the form (5.3.1), as have also Macchi (1975) and co-workers, who refer to quantities such as $\Pi_n^{\mathrm{sym}}(\cdot)$ in (5.3.1) as *exclusion probabilities*.

An important feature of Janossy measures is their simple interpretation when derivatives exist. If $\mathcal{X} = \mathbb{R}^d$ and $j_n(x_1, \ldots, x_n)$ denotes the density of

$J_n(\cdot)$ with respect to Lebesgue measure on $(\mathbb{R}^d)^{(n)}$ with $x_i \neq x_j$ for $i \neq j$, then

$$j_n(x_1, \ldots, x_n)\, \mathrm{d}x_1 \cdots \mathrm{d}x_n = \Pr\left\{ \begin{array}{c} \text{there are exactly } n \text{ points in the} \\ \text{process, one in each of the } n \text{ distinct} \\ \text{infinitesimal regions } (x_i, x_i + \mathrm{d}x_i) \end{array} \right\}.$$

This interpretation gives the Janossy densities a fundamental role in the structural description and likelihood analysis of finite point processes. Thus, they appear as likelihoods in Chapter 7, where they play a key role in the study of spatial point patterns (see also Chapter 15 and references there) and also in pseudolikelihoods. They are well adapted to describing the behaviour on observational regions, which, being finite, are typically bounded.

Example 5.3(a) *I.i.d. clusters* (continued from Section 5.1). In this case, $\mathcal{X} = \mathbb{R}^d$ and, assuming $F(A) = \int_A f(x)\, \mathrm{d}x$ for some density function $f(\cdot)$, the joint density function for the ordered sequence of n points at $x_1, \ldots, x_n$ is

$$\pi_n(x_1, \ldots, x_n) = f(x_1) \cdots f(x_n),$$

which is already in symmetric form. Here

$$j_n(x_1, \ldots, x_n) = p_n n!\, f(x_1) \cdots f(x_n),$$

and it is $j_n(\cdots)$, not $\pi_n(\cdots)$, that gives the probability density of finding one particle at each of the n points $(x_1, \ldots, x_n)$, the factorial term giving the number of ways the particles can be allocated to these locations. □

Example 5.3(b) *Finite renewal processes and random walks.* Suppose $\mathcal{X} = \mathbb{R}^1$ and that, given $N = n$, the points of the process are determined by the successive points $S_1, \ldots, S_n$ of a simple renewal process for which the common distribution of the lifetimes $S_j - S_{j-1}$ (where $S_0 \equiv 0$ and $j = 1, \ldots, n$) has a density function $f(\cdot)$. Then

$$\pi_n(S_1, \ldots, S_n) = \prod_{j=1}^{n} f(S_j - S_{j-1}). \tag{5.3.3}$$

In moving to the symmetrized form, some care is needed. For any $x_1, \ldots, x_n$, we have, formally,

$$\pi_n^{\mathrm{sym}}(x_1, \ldots, x_n) = \frac{1}{n!} \sum_{\mathrm{perm}} f(x_{i_1}) f(x_{i_2} - x_{i_1}) \cdots f(x_{i_n} - x_{i_{n-1}}).$$

Let $x_{(1)}, \ldots, x_{(n)}$ denote the set $\{x_1, \ldots, x_n\}$ in ascending order. Then, at least one term in each product in $\sum_{\mathrm{perm}}$ will vanish (since $f(x) = 0$ for $x < 0$) unless we already have $x_1, \ldots, x_n$ ordered; that is, $x_j = x_{(j)}$ for $j = 1, \ldots, n$. Hence,

$$\pi_n^{\mathrm{sym}}(x_1, \ldots, x_n) = \frac{1}{n!}\, f(x_{(1)}) f(x_{(2)} - x_{(1)}) \cdots f(x_{(n)} - x_{(n-1)}). \tag{5.3.4}$$

Comparing (5.3.3) and (5.3.4), $1/n!$ in the latter is seemingly a discrepant factor. The reconciliation lies in the fact that (5.3.3) vanishes outside the hyperoctant $x_1 < x_2 < \cdots < x_n$, whereas (5.3.4) repeats itself symmetrically in all $n!$ hyperoctants.

Finally, the Janossy densities are given by

$$j_n(x_1, \ldots, x_n) = p_n \, f(x_{(1)}) f(x_{(2)} - x_{(1)}) \cdots f(x_{(n)} - x_{(n-1)}), \qquad (5.3.5a)$$

where as before p_n is the probability that the process contains just n points. Again it is to be noted that (5.3.3) vanishes outside the first hyperoctant, whereas (5.3.5) gives positive measure to all hyperoctants.

Once the unidirectional character of each step is lost, these simplifications do not occur. What is then available for a general random walk is confined to the forms (5.3.3) and the corresponding expression

$$j_n(x_1, \ldots, x_n) = p_n \sum_{\text{perm}} f(x_{i_1}) f(x_{i_2} - x_{i_1}) \cdots f(x_{i_n} - x_{i_{n-1}}). \qquad (5.3.5b)$$

The simplest renewal example occurs when f has an exponential density. The joint density (5.3.3) then reduces to

$$\pi_n(x_1, \ldots, x_n) = \begin{cases} \lambda^n \exp(-\lambda x_n) & (0 \le x_1 \le x_n), \\ 0 & \text{otherwise}, \end{cases}$$

or in terms of (5.3.5),

$$j_n(x_1, \ldots, x_n) = p_n \lambda^n \mathrm{e}^{-\lambda x_{(n)}}.$$

Remarkably, the joint distribution depends only on the position of the extreme value $x_{(n)}$; given this value, the other points are distributed uniformly over $(0, x_{(n)})$.

The simplest example of a symmetric random walk is probably that for which the individual steps are normally distributed $N(0, 1)$. The successive S_i are then the partial sums of a sequence of independent normal variates

$$S_i = \sum_{j=1}^{i} Z_j$$

and for any given n are therefore jointly normally distributed with zero mean vector and covariance matrix having elements

$$\sigma_{ij} = \min(i, j) \qquad (1 \le i, j \le n).$$

No dramatic simplifications seem possible, but some further details are given in Exercise 5.3.1. □

Example 5.3(c) *Gibbs processes: processes generated by interaction potentials.* A fundamental class of point processes arising in statistical physics is described by means of forces acting on and between particles. The total

potential energy corresponding to a given configuration of particles is assumed to be decomposable into terms representing the interactions between the particles taken in pairs, triples, and so on; first-order terms representing the potential energies of the individual particles due to the action of an external force field may also be included. This leads to a representation of the total potential energy for a configuration of n particles at $x_1, \ldots, x_n$ by a series of the form

$$U(x_1, \ldots, x_n) = \sum_{r=1}^{n} \sum_{1 \le i_1 < \cdots < i_r \le n} \psi_r(x_{i_1}, \ldots x_{i_r}), \tag{5.3.6}$$

where $\psi(\cdot)$ is the rth-order interaction potential. Frequently, it is supposed that only the first- and second-order terms need be included, so that the process is determined by the *point pair potentials*, and

$$U(x_1, \ldots, x_n) = \sum_{i=1}^{n} \psi_1(x_i) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \psi_2(x_i, x_j). \tag{5.3.7}$$

It is then one of the fundamental principles of statistical mechanics that in equilibrium the probability density of a particular configuration is inversely proportional to the exponential of the potential energy. In terms of Janossy densities, this means that

$$j_n(x_1, \ldots, x_n) = C(\theta) \exp[-\theta U(x_1, \ldots, x_n)] \tag{5.3.8}$$

for some constant of proportionality $C(\theta)$ and parameter θ related to the temperature of the system. The normalizing constant is referred to as the *partition function*. The major difficulty in handling processes of this type lies in expressing the partition function as a function of θ (or, indeed, of any other parameters that may occur in the description of the system).

It is important to note that for finite point processes for which the Janossy densities exist, there is a converse to equation (5.3.8) where the densities $j_n(\cdot)$ are expressed in terms of the interaction potentials $\psi_r(\cdot)$ via the function $U(\cdot)$. Specifically, Exercise 5.3.7 describes $\psi_k(\cdot)$ in terms of $j_r(\cdot)$ $(r = 1, \ldots, k)$.

In fact, two slightly different situations may be considered. In the first of these, the *canonical ensemble*, the number n of particles is regarded as fixed and the normalizing constant is chosen to satisfy

$$\frac{1}{C(\theta)} = \int_{\mathcal{X}^{(n)}} \exp[-\theta U(x_1, \ldots, x_n)] \, \mathrm{d}x_1 \cdots \mathrm{d}x_n.$$

In the second, the *grand canonical ensemble*, both the number of particles and their locations are regarded as variable, and the partition function has to be chosen to satisfy (5.3.9) below.

Here we examine two special cases; further discussion is given around Examples 7.1(c)–(f).

(i) *No interactions (ideal gas).* Here,

$$j_n(x_1,\dots,x_n) = C(\theta)\exp\left(-\theta\sum_{i=1}^{n}\psi(x_i)\right) = C(\theta)\prod_{i=1}^{n}\exp[-\theta\psi(x_i)].$$

Integrating over $(x_1,\dots,x_n)\in\mathcal{X}^{(n)}$ and summing, using (5.3.9), we obtain

$$1 = C(\theta)\sum_{n=0}^{\infty}\frac{[\Lambda(\theta)]^n}{n!} = C(\theta)\mathrm{e}^{\Lambda(\theta)},$$

setting $j_0 = J_0 = C(\theta)$ and $\Lambda(\theta) = \int_{\mathcal{X}}\mathrm{e}^{-\theta\psi(x)}\,\mathrm{d}x$. Thus, $C(\theta) = \mathrm{e}^{-\Lambda(\theta)}$ and the process is just an inhomogeneous Poisson process with intensity $\mathrm{e}^{-\theta\psi(x)}$.

(ii) *Repulsive interactions.* Consider next the case of a homogeneous process in which the potential is specified entirely by the pairwise interactions $\psi_2(x,y)$, which are assumed to be a function $\phi(r)$ of the distance $r = |x-y|$ between the pair of points. A large variety of special forms have been considered for the function $\phi(\cdot)$ both in the statistical mechanics literature (e.g. Ruelle, 1969; Preston, 1976) and more recently as models for spatial point processes in other contexts (see e.g. Ripley, 1977; Ogata and Tanemura, 1984). Typical examples include

$$\begin{aligned}
\phi_1(r) &= -\log(1-\mathrm{e}^{-(r/\sigma)^2}),\\
\phi_2(r) &= (\sigma/|r|)^n \qquad (n = 4, 6, \text{etc.}),\\
\phi_3(r) &= \infty \ \text{ or } \ 0 \quad \text{as } r \le \text{ or } > \sigma.
\end{aligned}$$

The function $\phi_1(\cdot)$ represents relatively weak repulsive forces, even for r near zero, and it is therefore described as a 'soft-core' model. $\phi_3(\cdot)$ corresponds to the 'hard-core' model; every point pair has a separation $> \sigma$, and no other interaction occurs. The second model is of intermediate type, approximating the behaviour of the hard-core model for large n. None of these models is easy to handle analytically, and special expansion techniques have been developed to approximate the partition functions. □

For the subsequent discussions, we use mainly the Janossy measures. In this formulation, the normalization condition $\sum p_n = 1$ takes the form

$$\sum_{n=0}^{\infty}\frac{J_n(\mathcal{X}^{(n)})}{n!} = 1 \tag{5.3.9}$$

since we may interpret $J_0(\mathcal{X}^{(0)}) = p_0$ and, for $n \ge 1$, we have

$$J_n(\mathcal{X}^{(n)}) = p_n\sum_{\text{perm}}\Pi_n(\mathcal{X}^{(n)}) = p_n\, n!\,.$$

It is clear that from any family of symmetric measures $J_n(\cdot)$ satisfying (5.3.9), we can construct a probability distribution $\{p_n\}$ and a set of symmetric probability measures $\{\Pi_n^{\mathrm{sym}}(\cdot)\}$ satisfying Conditions 5.3.I, and conversely.

Either specification is equivalent to specifying a global probability measure $\mathcal{P}$ on the Borel sets $\mathcal{A}$ of the countable union (with $\mathcal{X}^{(0)}$ interpreted as an isolated point)

$$\mathcal{X}^{\cup} = \mathcal{X}^{(0)} \cup \mathcal{X}^{(1)} \cup \mathcal{X}^{(2)} \cup \cdots ; \tag{5.3.10}$$

Moyal (1962a) takes $(\mathcal{X}^{\cup}, \mathcal{P})$ as the canonical probability space of a finite point process. Given such a measure $\mathcal{P}$, the measure $p_n \Pi_n^{\mathrm{sym}}$, or equivalently, $(n!)^{-1} J_n$, appears as the restriction of $\mathcal{P}$ to the component $\mathcal{X}^{(n)}$. The situation is summarized in the following proposition.

Proposition 5.3.II. *Let $\mathcal{X}$ be a complete separable metric space, and let $\mathcal{B}_{\mathcal{X}}^{(n)}$ be the product σ-field on $\mathcal{X}^{(n)}$i, with the added convention that the set $\mathcal{X}^{(0)}$ denotes an ideal point such that $\mathcal{X}^{(0)} \times \mathcal{X} = \mathcal{X} \times \mathcal{X}^{(0)} = \mathcal{X}$. Then, the following specifications are equivalent, and each suffices to define a finite point process on $\mathcal{X}$:*

(i) *a probability distribution $\{p_n\}$ on the nonnegative integers and a family of symmetric probability distributions $\Pi_n^{\mathrm{sym}}(\cdot)$ on $\mathcal{B}_{\mathcal{X}}^{(n)}$, $n \geq 1$;*

(ii) *a family of nonnegative symmetric measures $J_n(\cdot)$ on $\mathcal{B}_{\mathcal{X}}^{(n)}$, $n \geq 1$, satisfying the normalization condition* (5.3.9) *and with $J_0(\mathcal{X}^{(0)}) = p_0$;*

(iii) *a symmetric probability measure $\mathcal{P}$ on the symmetric Borel sets of the countable union in* (5.3.10).

There is one point of principle to be noted here concerning the canonical choice of state space for a finite point process. To be consistent with treating a point process as a set of *unordered* points, a realization with, say, k points should be thought of not as a point in $\mathcal{X}^{(k)}$ but as a point in the quotient space of $\mathcal{X}^{(k)}$ with respect to the group of permutations amongst the k coordinates. For example, when $\mathcal{X} = \mathbb{R}$ and $k = 2$, then in place of all pairs (x_1, x_2), with (x_1, x_2) and (x_2, x_1) being treated as equivalent, we should consider some representation of the quotient space such as the set $\{(x_1, x_2) : x_1 \leq x_2\}$. The difficulty with this approach in general is that it is often hard to find a convenient concrete representation of the quotient space (consider for example the case just cited with $\mathbb{R}$ replaced by the unit circle or sphere), with the attendant problems of visualizing the results and bringing geometric intuition to bear. We have therefore preferred the redundant representation, which allows a distinction between the points but then gives all permutations amongst the labelling of the points equal weight in the measure. It must be borne in mind that there is then a many–one relation between the points in the space $\mathcal{X}^{\cup}$ and the set of all totally finite counting measures.

Another way of treating the same problem is to introduce the σ-algebra of *symmetric sets* in $\mathcal{X}^{(k)}$, that is, the sets invariant under permutations of the coordinate axes. A symmetric set in $\mathcal{X}^{\cup}$ is a set whose projections onto $\mathcal{X}^{(k)}$ are symmetric for each positive integer k.

Then, any event defined on the point process represents a symmetric set in $\mathcal{X}^{\cup}$, and thus the natural σ-algebra to use in discussing point process properties is this σ-algebra of symmetric sets. We do not emphasize this

approach because our main development in Chapter 9 is given in terms of counting measures; we merely refer the reader seeking details to Moyal (1962a) and Macchi (1975) (see also Exercises 5.3.4–6).

Now let us turn to the problem of expressing in terms of Janossy measures (or one of their equivalents) the probability distributions of the random variables $N(A_i)$. If $(A_1, \ldots, A_k)$ represents a finite partition of $\mathcal{X}$, the probability of finding exactly n_i points in A_i $(i = 1, \ldots, k)$ can be written, with $n_1 + \cdots + n_k = n$, as

$$P_k(A_1, \ldots, A_k; n_1, \ldots, n_k) = \frac{J_n(A_1^{(n_1)} \times \cdots \times A_k^{(n_k)})}{n_1! \cdots n_k!}$$
$$= p_n \binom{n}{n_1 \cdots n_k} \Pi_n^{\mathrm{sym}}(A_1^{(n_1)} \times \cdots \times A_k^{(n_k)}), \tag{5.3.11}$$

where the multinomial coefficient can be interpreted as the number of ways of grouping the n points so that n_i lie in A_i $(i = 1, \ldots, k)$.

It is important in (5.3.11) both that the sets A_i are disjoint and that they have union $\mathcal{X}$ (i.e. they are a partition of $\mathcal{X}$). For any i for which $n_i = 0$, the corresponding term is omitted from the right-hand side.

From (5.3.11), it follows in particular that the probability of finding n points in A, irrespective of the number in its complement A^c, is given by

$$n!\, P_1(A; n) = \sum_{r=0}^{\infty} \frac{J_{n+r}(A^{(n)} \times (A^c))^{(r)})}{r!}\,. \tag{5.3.12}$$

Similarly, if $A_1, \ldots, A_k$ are any k disjoint Borel sets, $C = (A_1 \cup \cdots \cup A_k)^c$, and $n = n_1 + \cdots + n_k$, the probability of finding just n_i points in A_i, $i = 1, \ldots, k$, is given by

$$n_1! \cdots n_k!\, P_k(A_1, \ldots, A_k; n_1, \ldots, n_k) = \sum_{r=0}^{\infty} \frac{J_{n+r}(A_1^{(n_1)} \times \cdots \times A_k^{(n_k)} \times C^{(r)})}{r!}. \tag{5.3.13}$$

These probabilities are in fact the joint distributions of the random variables $N(A_i)$, $i = 1, \ldots, k$. The fact that they do form a *consistent* set of finite-dimensional (fidi) distributions is implicit in their derivation, but it can also be verified directly, as we show following the discussion of such conditions in Chapter 9.

An alternative approach, following Moyal (1962a), starts from the observation that each realization can be represented as a random vector $Y \in \mathcal{X}^{(n)}$ for some $n \geq 0$. Any such vector defines a counting measure on $\mathcal{X}$, through

$$N(A) = \#\{i : y_i \in A\},$$

where the y_i are the components of the random vector Y. The random vector thus gives rise to a mapping from $\mathcal{X}^{(n)}$ into the space $\mathcal{N}_{\mathcal{X}}^{\#}$ of all counting measures on $\mathcal{X}$. It is easy to see that this mapping is measurable so it defines a point process (see Chapter 9). This being true for every n, the whole

process is a point process, and since (5.3.13) are its fidi distributions, they are necessarily consistent. As Moyal pointed out, this approach to the existence of finite point processes can be extended to more general cases by considering the restrictions of the process to an increasing family of Borel sets (spheres, say) chosen so that they expand to fill the whole space but with probability 1 have only a finite number of points in each. The main difficulty with this approach from our point of view is that it does not extend readily to random measures, which we require for their own sake and for applications in later chapters.

We conclude this section with a lemma that will play a useful role in simplifying the relations amongst various measures introduced in the sequel. It is needed in particular in checking that the distributions defined by (5.3.13) satisfy the consistency conditions of Chapter 9.

Lemma 5.3.III. *Let A be a Borel subset of $\mathcal{X}$ and S a symmetric measure defined on $\mathcal{X}^{(n)}$ for some $n > 0$. Then, for any partition $\{A_1, \ldots, A_k\}$ of A,*

$$S(A^{(n)}) = \sum \binom{n}{j_1 \cdots j_k} S(A_1^{(j_1)} \times \cdots \times A_k^{(j_k)}), \tag{5.3.14}$$

where the summation extends over all nonnegative integers $j_1, \ldots, j_k$ for which $j_1 + \cdots + j_k = n$.

Proof. Equation (5.3.14) expresses the fact that the partitioning of A induces a partitioning of $A^{(n)}$ into k^n subsets, which are grouped together into classes that are identified by vectors $(j_1, \ldots, j_k)$: within any given class, each constituent subset has A_i appearing as a coordinate or 'edge' j_i times. The symmetry of S implies that all subsets in the same class have the same S measure; hence, (5.3.14) follows. □

Exercises and Complements to Section 5.3

5.3.1 [see Example 5.3(b)]. For a finite random walk with normally distributed $N(0,1)$ steps, show that

$$\pi_2^{\mathrm{sym}}(x,y) = \frac{(\mathrm{e}^{-x^2/2} + \mathrm{e}^{-y^2/2})\mathrm{e}^{-(x-y)^2/2}}{4\pi}$$

and

$$\pi_3^{\mathrm{sym}}(x,y,z) = \frac{f(x,y,z) + f(y,z,x) + f(z,x,y)}{12\pi(2\pi)^{1/2}},$$

where $f(x,y,z) = \mathrm{e}^{-(x^2+(y-z)^2)/2}(\mathrm{e}^{-(y-x)^2/2} + \mathrm{e}^{-(z-x)^2/2})$.

5.3.2 Check Proposition 5.3.II in detail.

5.3.3 Show that, by a suitable choice of metric, $\mathcal{X}^{\cup}$ in (5.3.10) becomes a c.s.m.s. [Recall the assumption, made in Condition 5.3.I(a), that $\mathcal{X}$ is a c.s.m.s.]

5.3.4 Let $A^{(k)}$ denote the k-fold product $A \times \cdots \times A$. Show that a symmetric measure on the Borel sets of $\mathcal{X}^{(2)}$ is determined by its values on sets of the form $A^{(2)}$ but that the corresponding statement for $\mathcal{X}^{(k)}$ with $k \geq 3$ is false. [*Hint*: Consider first $\mathcal{X} = \{1, 2\}$ and $k = 2, 3$.]

5.3.5 (*Continuation*). Let $\mathcal{B}^{(k)}_{\mathrm{sym}}$ be the smallest σ-algebra containing the sets $A^{(k)}$ for Borel subsets A of $\mathcal{X}$. Show that $\mathcal{B}^{(k)}_{\mathrm{sym}}$ consists of all symmetric Borel subsets of $\mathcal{X}^{(k)}$ and that any symmetric measure μ on $\mathcal{B}^{(k)}$ is completely determined by its values on $\mathcal{B}^{(k)}_{\mathrm{sym}}$. Show also that a symmetric measure μ on $\mathcal{B}^{(k)}$ is completely determined by integrals of the form

$$\int_{\mathcal{X}^{(k)}} \zeta(x_1)\cdots\zeta(x_k)\,\mu(\mathrm{d}x_1\times\cdots\times\mathrm{d}x_k)$$

for functions ζ in the class $\mathcal{U}$ of Definition 5.5.I.

5.3.6 Let $\mathcal{X}_0^{(n)}$ denote the quotient space $\mathcal{X}^{(n)}/\Pi^{(n)}$, where $\Pi^{(n)}$ is the permutation group over the coordinates of a point in $\mathcal{X}^{(n)}$. Prove that there is a one-to-one correspondence between measures on the Borel subsets of $\mathcal{X}_0^{(n)}$ and symmetric measures on the Borel subsets of $\mathcal{X}^{(n)}$. [Macchi (1975) uses $\bigcup_{n=0}^{\infty}\mathcal{X}_0^{(n)}$ in place of $\mathcal{X}^{\cup}$ in (5.3.10) as the sample space for finite point processes.]

5.3.7 Let $\{j_k(\cdot)\colon k=1,2,\ldots\}$ be a family of positive Janossy densities for an a.s. finite point process. Define functions $\psi_1(x)=-\log j_1(x)$,

$$\psi_k(x_1,\ldots,x_k)=-\log j_k(x_1,\ldots,x_k)-\sum_{r=1}^{k-1}\sum_{1\le i_1<\cdots<i_r\le k}\psi_r(x_{i_1},\ldots,x_{i_r}).$$

Show that $\{j_k(\cdot)\}$ thereby defines recursively a unique family of interaction potentials for a Gibbs process [see Example 5.3(c), especially (5.3.8)].

5.3.8 Let $f(\cdot)$ be a bounded or nonnegative functional of an a.s. finite point process with Janossy measures $J_n(\cdot)$. Show that

$$\mathrm{E}[f(N)]=\sum_{n=0}^{\infty}\frac{1}{n!}\int_{\mathcal{X}^{(n)}} f(\delta_{x_1}+\cdots+\delta_{x_n})\,J_n(\mathrm{d}x_1\times\cdots\times\mathrm{d}x_n).$$

5.4. Moment Measures and Product Densities

We now investigate the moment structure of finite point processes, extending to counting measures the notions of ordinary and factorial moments and cumulants developed for nonnegative integer-valued r.v.s in Section 5.2. In fact, because we require a general point process to be finite a.s. on bounded sets, the definitions can be extended almost immediately to the general case (these extensions are treated in Chapter 9).

Suppose then that the total population has finite kth moment $\mu_k = \mathrm{E}\big([N(\mathcal{X})]^k\big)$ for some $k=1,2,\ldots$. Then, for any Borel set $A\in\mathcal{B}_{\mathcal{X}}$, define

$$M_k(A^{(k)})=\mathrm{E}\big([N(A)]^k\big), \tag{5.4.1}$$

where we choose to regard the left-hand side as the value on the product set $A^{(k)}$ of a set function defined on the product σ-field $\mathcal{B}^{(k)}_{\mathcal{X}}$ in $\mathcal{X}^{(k)}$. In

particular, if the total population has finite mean $\mu_1 = \mathrm{E}[N(\mathcal{X})]$, we can define the *expectation measure* $M(\cdot)$ by

$$M(A) \equiv M_1(A) = \mathrm{E}[N(A)] \qquad (A \in \mathcal{B}_{\mathcal{X}}). \tag{5.4.2}$$

Here it is clear from Fubini's theorem that $M(\cdot)$ inherits countable additivity from $N(\cdot)$ so that it does in fact define a measure on $\mathcal{B}_{\mathcal{X}}$.

For $k > 1$, we can extend the definition of M_k to arbitrary rectangle sets of the form

$$A_1^{(k_1)} \times \cdots \times A_r^{(k_r)},$$

where $\{k_1, \cdots, k_r\}$ is a partition of k (so $k_r \geq 1$ and $k_1 + \cdots + k_r = k$) and the A_i are disjoint sets of $\mathcal{B}_{\mathcal{X}}$, by setting

$$M_k(A_1^{(k_1)} \times \cdots \times A_r^{(k_r)}) = \mathrm{E}\big([N(A_1)]^{k_1} \cdots [N(A_r)]^{k_r}\big). \tag{5.4.3}$$

It is not difficult to check that M_k is countably additive on these k-dimensional rectangle sets and hence can be extended to a measure on the Borel sets $\mathcal{B}_{\mathcal{X}}^{(k)}$. In fact, M_k can be regarded as the expectation measure of a point process on $\mathcal{X}^{(k)}$: the point process consists of all k-tuples (allowing repetitions and distinguishing the order in this k-tuple) of points from the original realization; that is, it consists of the k-fold product $N^{(k)}$ of N with itself. Thus, M_k gives the expected number of such k-tuples in arbitrary sets from $\mathcal{B}_{\mathcal{X}}^{(k)}$. Since $N^{(k)}$ is a *symmetric* measure on $\mathcal{X}^{(k)}$, so too is its expectation measure M_k. We call M_k the kth *moment measure of* N.

Similarly, we can introduce the kth *factorial moment measure* $M_{[k]}$. Here, $M_{[1]} = M_1 = M$, and for $k > 1$ the ordinary powers inside the expectation in (5.4.3) are replaced by factorial powers: with A_i and k_i as in (5.4.3), we set

$$M_{[k]}(A_1^{(k_1)} \times \cdots \times A_r^{(k_r)}) = \mathrm{E}\big([N(A_1)]^{[k_1]} \cdots [N(A_r)]^{[k_r]}\big). \tag{5.4.4}$$

As for M_k, the set function on the left-hand side of this defining relation is countably additive on rectangle sets in $\mathcal{X}^{(k)}$ and can be interpreted as the expectation measure of a certain point process in $\mathcal{X}^{(k)}$. In this case, the realizations of the new process consist of all k-tuples of *distinct* points from the original process, still distinguishing the order within the k-tuple but not allowing repetitions. (Note that if the original process N has multiple points, each such point is to be enumerated according to its multiplicity: for example, a double point of N should be regarded as two distinct points having the same coordinates when constructing the k-tuples.) Then $M_{[k]}(A)$ represents the expected number of such k-tuples falling in $A \in \mathcal{B}_{\mathcal{X}}^{(k)}$.

Proposition 5.4.I. *If $\mu_k = \mathrm{E}([N(\mathcal{X})]^k) < \infty$, the set functions M_k and $M_{[k]}$ defined by* (5.4.3) *and* (5.4.4) *are countably additive on rectangle sets and have unique extensions to symmetric measures M_k and $M_{[k]}$, respectively, on $\mathcal{B}_{\mathcal{X}}^{(k)}$.*

Using the identities (5.2.2) and (5.2.3) that relate ordinary and factorial powers, it is possible to write down explicit expressions for M_k on certain sets in terms of $\{M_{[j]},\, j = 1, \ldots, k\}$ and for $M_{[k]}$ in terms of $\{M_j,\, j = 1, \ldots, k\}$. Directly from (5.2.5), we have the important special case

$$\mathrm{E}\big([N(\mathcal{X})]^k\big) = M_k(A^{(k)}) = \sum_{j=1}^{k} \Delta_{j,k} M_{[j]}(A^{(j)}). \tag{5.4.5}$$

Such relations are particularly useful when the factorial moment measures are absolutely continuous so that the right-hand side of (5.4.5) can be expressed as a sum of integrals of the product densities introduced below Lemma 5.4.III. Note also relations such as

$$\begin{aligned} M_{[2]}(A \times B) &= \mathrm{E}[N(A)N(B)] - \mathrm{E}[N(A \cap B)] \\ &= M_2(A \times B) - M(A \cap B) \qquad (A,\, B \in \mathcal{B}_{\mathcal{X}}) \end{aligned} \tag{5.4.6}$$

(see Exercises 5.4.1–6 for a more systematic exposition of such relations).

Applications of these moment measures appear in subsequent chapters; here we explore their relation to the Janossy measures and their interpretation in terms of product densities.

Since (5.4.4) is simply the factorial moment of a fidi distribution, which can be expressed in terms of the Janossy measures by means of (5.3.11), we can obtain an expression for $M_{[k]}(\cdot)$ in terms of Janossy measures. To examine this expression, we return to the case where $A_1, \ldots, A_r$ is a partition of $\mathcal{X}$. Assuming $\mathrm{E}([N(\mathcal{X})]^{[k]}) < \infty$, we have directly from the definitions, when $k_1 + \cdots + k_r = k$, that

$$\begin{aligned} M_{[k]}(A_1^{(k_1)} \times \cdots \times A_r^{(k_r)}) &= \sum_{j_i \ge k_i,\, i=1,\ldots,r} j_1^{[k_1]} \cdots j_r^{[k_r]} P_r(A_1, \ldots, A_r;\, j_1, \ldots, j_r) \\ &= \sum_{j_i \ge k_i} \frac{J_{j_1+\cdots+j_r}(A_1^{(j_1)} \times \cdots \times A_r^{(j_r)})}{\prod_{i=1}^{r}(j_i - k_i)!}\,. \end{aligned}$$

To simplify the last sum, put $n_i = j_i - k_i$ and group together the terms for which $n_1 + \cdots + n_r = n$. Setting $k = k_1 + \cdots + k_r$, we obtain

$$\begin{aligned} &M_{[k]}(A_1^{(k_1)} \times \cdots \times A_r^{(k_r)}) \\ &\qquad = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{\sum n_i = n} \binom{n}{n_1 \cdots n_r} J_{k+n}(A_1^{(k_1+n_1)} \times \cdots \times A_r^{(k_r+n_r)}). \end{aligned}$$

The inner sum can be reduced by Lemma 5.3.III, taking $A = \mathcal{X}$ and defining S by

$$S(B) = J_{k+n}(A_1^{(k_1)} \times \cdots \times A_r^{(k_r)} \times B) \qquad (B \in \mathcal{B}_{\mathcal{X}}^{(n)}),$$

thereby yielding the equation

$$M_{[k]}(A_1^{(k_1)} \times \cdots \times A_r^{(k_r)}) = \sum_{n=0}^{\infty} \frac{J_{k+n}(A_1^{(k_1)} \times \cdots \times A_r^{(k_r)} \times \mathcal{X}^{(n)})}{n!}\,.$$

Using the countable additivity of both sides, this extends to the following elegant generalization of (5.2.12),

$$M_k(B) = \sum_{n=0}^{\infty} \frac{J_{k+n}(B \times \mathcal{X}^{(n)})}{n!} \qquad (\text{all } B \in \mathcal{B}_{\mathcal{X}}^{(k)}). \tag{5.4.7}$$

To obtain the inverse relation, suppose that all factorial moments $\mu_{[k]}$ of $N(\mathcal{X})$ exist and that the p.g.f.

$$P(1+\eta) = \sum_{k=0}^{\infty} \frac{\mu_{[k]}\eta^k}{k!} \tag{5.4.8}$$

is convergent in a disk $|\eta| < 1 + \epsilon$ for some $\epsilon > 0$ [equivalently, that $P(z) = \mathrm{E}(z^{N(\mathcal{X})})$ is analytic in some disk $|z| < 2+\epsilon$]. Then, the inverse relation (5.2.1) can be applied to yield, with the same notation as in (5.4.7) and following a parallel route,

$$\begin{aligned} J_n(A_1^{(k_1)} \times \cdots \times A_r^{(k_r)}) &= \sum_{k=0}^{\infty} (-1)^k \frac{M_{[n+k]}(A_1^{(k_1)} \times \cdots \times A_r^{(k_r)} \times \mathcal{X}^{(k)})}{k!} \\ &= \sum_{j_i \geq k_i} \prod_{i=1}^{r} (-1)^{j_i - k_i} \frac{M_{[j_1+\cdots+j_r]}(A_1^{(j_1)} \times \cdots \times A_r^{(j_r)})}{(j_i - k_i)!} \end{aligned}$$

so that for general $B \in \mathcal{B}_{\mathcal{X}}^{(n)}$,

$$J_n(B) = \sum_{k=0}^{\infty} (-1)^k \frac{M_{[n+k]}(B \times \mathcal{X}^{(k)})}{k!}\,. \tag{5.4.9}$$

These results may be summarized for reference in the following theorem.

Theorem 5.4.II. *If the total population size has finite k*th *moment, then the k*th *factorial moment measure is defined and finite and can be represented in terms of the Janossy measures by* (5.4.7). *Conversely, if all moments are finite and for some $\epsilon > 0$ the p.g.f.* (5.4.8) *is convergent for $|\eta| < 1 + \epsilon$, then the Janossy measures can be represented in terms of the factorial moment measures by* (5.4.9).

Example 5.4(a) *Avoidance function.* To illustrate the application of Theorem 5.4.II, consider the set function

$$P_0(A) \equiv \Pr\{N(A) = 0\} = P_1(A; 0);$$

that is, the probability of finding no points in a given subset A of $\mathcal{X}$, or, equivalently, the probability that the support of N avoids A. Taking $n = 0$ in (5.4.9) and restricting $\mathcal{X}$ to A itself, we obtain immediately

$$P_0(A) = J_0(A) = \sum_{k=0}^{\infty} (-1)^k \frac{M_{[k]}(A^{(k)})}{k!}\,. \tag{5.4.10}$$

An important feature of (5.4.10) is that it is not necessary to know anything about the nature of the moment measure outside A to determine the probability. In the case $\mathcal{X} = \mathbb{R}$ and A equal to the interval $(0, t]$, the result in (5.4.10) gives the survivor function for the forward recurrence time in terms of the moment measures on $(0, t]$. Of course, from another point of view, (5.4.10) is just a special case of equation (5.2.11) giving the probabilities of a discrete distribution in terms of the factorial moments. □

We now turn and consider densities for the moment measures, assuming $\mathcal{X}$ to be a real Euclidean space (or well-behaved subset thereof). Recall the standard result, which follows from Fubini's theorem, that if a totally finite measure can be represented as the superposition of a finite or countably infinite family of component measures, then it is absolutely continuous with respect to a given measure if and only if each component is absolutely continuous, the density of the superposition being represented a.e. by the sum of the densities. Applied to the representation (5.4.7), this yields immediately the following lemma.

Lemma 5.4.III. *If the* kth *factorial moment measure* $M_{[k]}(\cdot)$ *exists, then it is absolutely continuous if and only if the Janossy measures* $J_n(\cdot)$ *are absolutely continuous for all* $n \geq k$*, in which case the densities* $m_{[k]}(\cdot)$ *and* $j_n(\cdot)$ *are related by the equations, for* $k = 1, 2, \ldots,$

$$m_{[k]}(x_1, \ldots, x_k) = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} j_{k+n}(x_1, \ldots, x_k, y_1, \ldots, y_n)\, \mathrm{d}y_1 \cdots \mathrm{d}y_n.$$

The inverse relation follows in a similar way: if all the factorial moment measures exist and are absolutely continuous, and if the series (5.4.9) is absolutely convergent, then the corresponding Janossy measure is absolutely continuous with density given by

$$j_n(x_1, \ldots, x_n) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} m_{n+k}(x_1, \ldots, x_n, y_1, \ldots, y_k)\, \mathrm{d}y_1 \cdots \mathrm{d}y_k. \tag{5.4.11}$$

Historically, the introduction of factorial moment densities, also referred to as *product densities* in Bhabha (1950) and Ramakrishnan (1950) and as *coincidence densities* in Macchi (1975), considerably preceded the more general treatment as above using factorial moment measures. This is easily understood in view of the simple physical interpretation of the densities: equations (5.4.7) and (5.3.9) imply that if $m_{[k]}(x_1, \ldots, x_k)$ is bounded in a neighbourhood of $(x_1, \ldots, x_k)$, then we can write

$$\begin{aligned} m_{[k]}(x_1, \ldots, x_k)\, \mathrm{d}x_1 \cdots \mathrm{d}x_k &= \sum_{n=0}^{\infty} \frac{J_{k+n}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k \times \mathcal{X}^{(n)})}{n!} \\ &= \Pr\left\{ \begin{array}{c} \text{one particle located in each of the} \\ \text{infinitesimal subsets } \mathrm{d}x_i \ \ (i = 1, \ldots, k) \end{array} \right\}, \end{aligned} \tag{5.4.12}$$

where $\mathrm{d}x_i$ denotes both the infinitesimal set $(x_i, x_i + \mathrm{d}x_i)$ and its Lebesgue

measure. This interpretation may be contrasted with that for the density

$$j_k(x_1,\ldots,x_k)\,\mathrm{d}x_1\cdots\mathrm{d}x_k = \Pr\left\{\begin{array}{c}\text{exactly } k \text{ points in realization, one in each}\\ \text{subset } \mathrm{d}x_i \ \ (i=1,\ldots,k), \text{ and none elsewhere}\end{array}\right\}. \quad (5.4.13)$$

From an experimental point of view, (5.4.12) can be estimated from the results of k observations at specific times or places, whereas the Janossy measure requires indefinitely many observations to determine the exact (total) number of occurrences. For this reason, the densities (5.4.12) are in principle amenable to experimental determination (through 'coincidence' experiments, hence the name *coincidence densities*) in a way that Janossy measures are not, at least in the context of counting particles. However, as Macchi (1975) has stressed, the Janossy measures, and hence the joint distributions, can be determined by the converse relations (5.4.9) and (5.4.11).

Moment measures also have the important feature, in common with relations such as (5.4.10), that they are *global* in character, in contrast to the *local* character of the Janossy measures. We mean by this that the form of the moment measures is not influenced by the nature of the region of observations: if two observation regions overlap, the moment measures coincide over their common region. On the other hand, the Janossy measures depend critically on the observation regions: just as the number of points observed in the region depends on its size and shape, so also the Janossy measures are exactly tailored to the particular region. This feature lends further importance to the converse relations (5.4.9) and (5.4.11): knowing the moment densities, the Janossy densities for any observation region A can be calculated by taking $\mathcal{X} = A$ in (5.4.11), a remark that continues to have force even when the point process is not totally finite over the whole of $\mathcal{X}$. Thus, the one set of moment measures suffices to determine the Janossy measures for as many observation regions as one cares to nominate. When the region of interest is indeed a bounded subset A of the space $\mathcal{X}$ where the point process is defined, we introduce the following definition.

Definition 5.4.IV (Local Janossy Measures and Densities)**.** *Given any bounded Borel set A, the* Janossy measures localized to A *are the measures $J_n(\cdot \mid A)$ $(n = 1, 2, \ldots)$ satisfying, for locations $x_i \in A$ $(i = 1, \ldots, n)$,*

$$J_n(\mathrm{d}x_1 \times \cdots \mathrm{d}x_n \mid A) = \Pr\left\{\begin{array}{c}\text{exactly } n \text{ points in } A \text{ at}\\ \text{locations } \mathrm{d}x_1,\ldots,\mathrm{d}x_n\end{array}\right\}.$$

When these measures have densities, they define the local Janossy densities.

Such local functions have particular importance when the process is no longer a.s. finite-valued on the whole space $\mathcal{X}$. For these local functions the identities in (5.4.9) and (5.4.11) continue to hold with $\mathcal{X}^{(k)}$ replaced by $A^{(k)}$

(and the local functions on the respective left-hand sides), as for example

$$j_n(x_1, \ldots, x_n \mid A) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_A \cdots \int_A m_{n+k}(x_1, \ldots, x_n, y_1, \ldots, y_k)\, \mathrm{d}y_1 \cdots \mathrm{d}y_k. \quad (5.4.14)$$

What is remarkable about such a relation is that by merely changing the range of integration of a function defined globally, we can recover the local probabilistic structure when all the moments exist [see Example 5.5(b)].

Local Janossy densities $j_n(x_1, \ldots, x_n \mid A)$ feature prominently in the discussion of point process likelihoods in Section 7.1.

The existence of densities is closely linked to the concept of orderliness, or more properly, simplicity, in the sense of Chapter 3, that with probability 1 there are no coincidences amongst the points. Suppose on the contrary that, for some population size n, the probability that two points coincide is positive. In terms of the measure $J_n(\cdot)$, the necessary and sufficient condition for this probability to be positive is that $J_n(\cdot)$ should allot nonzero mass to at least one (and hence all) of the diagonal sets $\{x_i = x_j\}$, where x_i is a point in the ith coordinate space. Thus, we have the following proposition.

Proposition 5.4.V. (a) *A necessary and sufficient condition for a point process to be simple is that, for all $n = 1, 2, \ldots$, the associated Janossy measure $J_n(\cdot)$ allots zero mass to the 'diagonals' $\{x_i = x_j\}$.*
(b) *When $\mathcal{X} = \mathbb{R}^d$, the process is simple if for all such n the Janossy measures have densities $j_n(\cdot)$ with respect to (nd)-dimensional Lebesgue measure.*

It is more convenient to frame an analogous condition in terms of the moment measures (assuming they exist). From the preceding result and the representation (5.4.7), we have immediately the following proposition.

Proposition 5.4.VI. *Suppose the second factorial moment measure $M_{[2]}(\cdot)$ exists. Then, a necessary and sufficient condition for the point process to be simple is that $M_{[2]}(\cdot)$ allots zero mass to the 'diagonal' set $\{x_i = x_j\}$. In particular, for $\mathcal{X} = \mathbb{R}^d$, the process is simple whenever $M_{[2]}(\cdot)$ has a density $m_{[2]}(\cdot)$ with respect to $2d$-dimensional Lebesgue measure.*

An alternative approach to this proposition can be given in the context of random measures: for the stationary case, see Proposition 8.1.IV and its Corollary 8.1.V.

In some applications, we may wish to verify that a given family of densities constitutes the product densities of some point process. The following result gives a simple sufficient condition, which, however, is far from necessary (see remarks after the proof).

Proposition 5.4.VII. *Let $m_{[k]}(\cdot)$ on $\mathcal{X}^{(k)}$ $(k = 1, 2, \ldots)$ be a family of symmetric nonnegative functions with finite total integrals*

$$\mu_{[k]} = \int_{\mathcal{X}^{(k)}} m_{[k]}(x)\, \mathrm{d}x,$$

and suppose that for some $\epsilon > 0$ *the series* $\sum_{k=1}^{\infty} \mu_{[k]} z^k$ *is convergent for* $|z| < 1+\epsilon$. *Then, a necessary and sufficient condition for the family* $\{m_{[k]}(\cdot)\}$ *to be factorial moment densities of a finite point process is that the integrals in* (5.4.11) *should be nonnegative for every* $n = 1, 2, \ldots$ *and every vector* $x = (x_1, \ldots, x_n)$. *These factorial moment densities then determine the process uniquely.*

Proof. The integrals are convergent by assumption and clearly define a family of nonnegative symmetric functions. The only other requirement needed for them to form a set of Janossy functions is the normalization condition (5.4.9). On integrating (5.4.11) over $x_1, \ldots, x_n$, the required condition is seen to be equivalent to demanding that if we define $\{p_n\}$ by

$$N!\, p_n = \sum_{k=0}^{\infty} (-1)^k \, \frac{\mu_{[k+n]}}{k!} \,,$$

then the $\{p_n\}$ should sum to unity. But this reduces to the condition $\mu_{[0]} = m_{[0]} = 1$, which may be assumed without loss of generality. □

Remarks. The constraint that $\sum_{k=1}^{\infty} \mu_{[k]}/k!$ converges for $|z| < 1+\epsilon$ is stronger than is needed: it is enough that $\limsup_{r\to\infty} (\mu_{[r]}/r!)^{1/r} < \infty$, but a more complicated definition of p_n may then be needed (see Exercises 5.4.3–4). Also, for the product densities to define a point process that is not necessarily a finite point process, it is enough for the result to hold (with either the given or modified conditions on $\{\mu_{[r]}\}$) with the state space $\mathcal{X}$ replaced by a sequence $\{A_n\}$ of bounded sets for which $A_n \uparrow \mathcal{X}$ as $n \to \infty$.

Example 5.4(b) *Moment densities of a renewal process* (Macchi, 1971a). It is well known (see Chapter 4) that the moment properties of a renewal process are completely specified by the renewal function. Although the renewal process is not a finite point process, the machinery developed in this section can be carried over to give a particularly succinct formulation of this result in terms of the factorial moment densities, where for ease of exposition it is assumed that the renewal density exists, $u(\cdot)$ say. In these terms, and assuming stationarity, the renewal density is just a multiple of the second-moment density since for $s < t$ and with $m = M_{[1]}((0, 1])$,

$$\begin{aligned} m_{[2]}(s,t)\,\mathrm{d}s\,\mathrm{d}t &= \Pr\{\text{renewals in } (s, s+\mathrm{d}s) \text{ and } (t, t+\mathrm{d}t)\} \\ &= m\,\mathrm{d}s\, u(t-s)\,\mathrm{d}t. \end{aligned}$$

Similarly, exploiting the regenerative property, we have for $t_1 < \cdots < t_k$ that

$$\begin{aligned} m_{[k]}(t_1, \ldots, t_k)\,\mathrm{d}t_1 \cdots \mathrm{d}t_k &= \Pr\{\text{renewals in } (t_i, t_i + \mathrm{d}t_i),\ 1 \le i \le k\} \\ &= m\,\mathrm{d}t_1\, u(t_2 - t_1)\,\mathrm{d}t_2 \cdots u(t_k - t_{k-1})\,\mathrm{d}t_k. \quad (5.4.15) \end{aligned}$$

Thus, when the moment densities exist, a necessary condition for a point process to be a stationary renewal process is that the densities be expressible in the product form (5.4.15).

This condition is also sufficient. To see this, assume (5.4.15) holds for some constant m and some function $u(\cdot)$ for each $k = 1, 2, \ldots$. From the cases $k = 1$, 2, first the constant m and then the function $u(\cdot)$ are identified in terms of first- and second-moment densities. From (5.4.11), we can obtain an expression for the density of the interval distribution by taking $\mathcal{X} = [0, t]$ and requiring exactly two events, one at 0 and one at t, thus yielding for the lifetime density $f(\cdot)$ the relation

$$
\begin{aligned}
m\, f(t) &= m \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_{[0,t]^{(k)}} \cdots \int u(x_1) u(x_2 - x_1) \cdots u(t - x_k)\, \mathrm{d}x_1 \cdots \mathrm{d}x_k \\
&= m \sum_{k=0}^{\infty} (-1)^k \int_{0<x_1<\cdots<x_k<t} \cdots \int u(x_1) u(x_2 - x_1) \cdots u(t - x_k)\, \mathrm{d}x_1 \cdots \mathrm{d}x_k
\end{aligned}
$$

This identifies $f(\cdot)$ as the solution to an inverse of the renewal equation in the form $f = u - f * u$. Finally, uniqueness follows from the fact that the moment measures, which coincide with those constructed from a renewal process with this density $f(\cdot)$, determine the process uniquely. □

Example 5.4(c) *The fermion process* (Macchi, 1975). The renewal process of the previous example generally produces a spacing or 'antibunching' effect, at least if its lifetime distribution has its coefficient of variation less than unity. Such behaviour is characteristic of fermions (e.g. electrons) as distinct from bosons (e.g. photons) in the elementary particle context. Benard and Macchi (1973) and Macchi (1975) developed a remarkable dual theory for both types of particles. This theory, while derived in the the first instance from considerations of quantum mechanics, leads to a dual family of point processes of considerable general interest. The first family coincides with the family of renewal processes under suitable conditions; we describe a typical member shortly. The dual family is described in Example 6.2(b) and consists of a doubly stochastic processes.

A striking application concerns the zeros of the Riemann zeta function. Coram and Diaconis (2002) provide statistical tests that illustrate aspects of a considerable literature on close connections between blocks of n 'adjacent' zeros and eigenvalues of random unitary matrices in the unitary group U_n furnished with Haar measure, for suitably chosen n. This statistical work includes comparisons of spacings (between adjacent zeros and eigenvalues), traces (of blocks of n zeros and eigenvalues of random elements of U_n), and correlation studies of points in intervals. D.E. Littlewood's immanants (e.g. Littlewood, 1950, Chapter 6), of which permanents and determinants as linear forms of all n-fold products over n points of the kernel $C(\cdot, \cdot)$ below are extremes, can be viewed as interpolating between boson and fermion point processes, respectively, via the group characters (Diaconis and Evans, 2001). Given $C(\cdot, \cdot)$ and a character group, the immanant, if positive, is proportional to the Janossy density of a simple point process with n points.

Our state space $\mathcal{X}$ is a general d-dimensional Euclidean space, and we use A to denote a closed bounded subset (e.g. a rectangle) within $\mathcal{X}$. Let $C(x,y)$ be a covariance function defined on $\mathcal{X}$ so that with

$$\widetilde{C}\begin{pmatrix} x_1 & \cdots & x_k \\ y_1 & \cdots & y_k \end{pmatrix} = \det \begin{pmatrix} C(x_1,y_1) & \cdots & C(x_1,y_k) \\ \vdots & \ddots & \vdots \\ C(x_k,y_1) & \cdots & C(x_k,y_k) \end{pmatrix},$$

the symmetric determinant

$$\widetilde{C}\begin{pmatrix} x_1 & \cdots & x_k \\ x_1 & \cdots & x_k \end{pmatrix} \geq 0.$$

In general, $C(\cdot,\cdot)$ may be complex-valued and therefore Hermitian so that $C(x,y) = \overline{C(x,y)}$, but for ease of writing we assume here that $C(\cdot,\cdot)$ is real.

It follows from nonnegativity that for $\lambda > 0$ the function

$$m_{[k]}(x_1,\ldots,x_k) = \lambda^k \widetilde{C}\begin{pmatrix} x_1 & \cdots & x_k \\ x_1 & \cdots & x_k \end{pmatrix} \tag{5.4.16}$$

is at least a possible candidate for the kth factorial moment density of some orderly point process on $\mathcal{X}$. To decide whether this is a legitimate choice, we need to investigate whether the corresponding Janossy densities, given formally by (5.4.11), are well defined and nonnegative.

In fact, the Janossy densities have a representation parallel to (5.4.16) in terms of the solution $R_\lambda(x,y)$ of the resolvent equation

$$R_\lambda(x,y) - \lambda \int_A C(x,y) R_\lambda(u,y)\,\mathrm{d}u = C(x,y). \tag{5.4.17}$$

It is well known in the theory of integral equations (see e.g. Pogorzelski, 1966, p. 47) that $R_\lambda(x,y)$ can be expressed as a series in λ with terms involving (5.4.16); specifically, $\lambda R_\lambda(x,y)$ equals

$$\frac{1}{d(\lambda)}\left[\lambda C(x,y) + \lambda \sum_{j=1}^{\infty} \frac{(-\lambda)^j}{j!} \int_A \cdots \int_A \widetilde{C}\begin{pmatrix} x & x_1 & \cdots & x_j \\ y & x_1 & \cdots & x_j \end{pmatrix} \mathrm{d}x_1 \cdots \mathrm{d}x_j\right],$$

where

$$d(\lambda) = 1 + \sum_{j=1}^{\infty} \frac{(-\lambda)^j}{j!} \int_A \cdots \int_A \widetilde{C}\begin{pmatrix} u_1 & \cdots & u_j \\ u_1 & \cdots & u_j \end{pmatrix} \mathrm{d}u_1 \cdots \mathrm{d}u_j$$

is the Fredholm determinant associated with equation (5.4.17). More generally, the $k \times k$ 'Fredholm minor' associated with this equation, obtained by replacing C by R_λ in the basic determinant (5.4.16), is given by

$$\begin{aligned} &\lambda^k \widetilde{R}_\lambda \begin{pmatrix} x_1 & \cdots & x_k \\ y_1 & \cdots & y_k \end{pmatrix} \\ &\quad = \frac{1}{d(\lambda)}\left[\lambda^k \widetilde{C}\begin{pmatrix} x_1 & \cdots & x_k \\ y_1 & \cdots & y_k \end{pmatrix}\right. \\ &\qquad \left. + \lambda^k \sum_{j=1}^{\infty} \frac{(-1)^j}{j!} \int_A \cdots \int_A \widetilde{C}\begin{pmatrix} x_1 & \cdots & x_k & u_1 & \cdots & u_j \\ y_1 & \cdots & y_k & u_1 & \cdots & u_j \end{pmatrix} \mathrm{d}u_1 \cdots \mathrm{d}u_j\right] \end{aligned} \tag{5.4.18}$$

(see e.g. Pogorzelski, 1966, p. 52). Now (5.4.18) has the same form as (5.4.11) if we identify the factorial moment densities by (5.4.16) and the Janossy measures by

$$j_n(x_1, \ldots, x_k) = \lambda^k d(\lambda)\, \widetilde{R}_\lambda \begin{pmatrix} x_1 & \cdots & x_k \\ x_1 & \cdots & x_k \end{pmatrix}. \tag{5.4.19}$$

The convergence of (5.4.18) is ensured by the general theory, using the Hadamard inequality to bound the determinants appearing therein. Thus, only the nonnegativity of the functions (5.4.19) needs to be checked. While these functions need not be nonnegative in general, an appropriate sufficient condition can easily be stated in terms of λ and the eigenvalues of (5.4.17); that is, the values of λ for which the homogeneous equation corresponding to (5.4.17) [i.e. (5.4.17) with the right-hand side replaced by zero] admits a nontrivial solution. In fact, the determinant $\widetilde{R}_\lambda$ in (5.4.19) is nonnegative if the function R_λ is itself a covariance function, for which it suffices that the eigenvalues $\mu_i(\lambda)$ of R_λ be nonnegative. Now these eigenvalues are related to those of C by the equation

$$\mu_i(\lambda) = \lambda_i - \lambda,$$

so a necessary and sufficient condition for R_λ to be a covariance function is that $\lambda < \min\{\lambda_i\}$, in which case $d(\lambda)$ is also nonnegative. It is now easy to check that this condition is necessary and sufficient for the existence of a well-defined point process with factorial moments and Janossy densities given by (5.4.16) and (5.4.19).

A great virtue of this model is that it provides a rather general model for 'antibunching' with repulsive rather than attractive points for which moment and probability densities can be given explicitly, or at least be computed numerically, and is not restricted to the state space $\mathbb{R}$.

Further details of the process, including a discussion of the corresponding discrete process in which the integral operator is replaced by a matrix, are given in Exercises 5.4.7–10. □

Exercises and Complements to Section 5.4

5.4.1 (see Proposition 5.4.I). Show that for disjoint sets A and B,

$$M_{[2]}((A \cup B)^{(2)}) = M_{[2]}(A^{(2)}) + M_{[2]}(B^{(2)}) + 2M_{[2]}(A \times B).$$

5.4.2 Establish the analogues below of (5.4.6), where $\sum^*$ denotes summation over all distinct terms of like kind:

$$\begin{aligned} M_{[3]}(A_1 \times A_2 \times A_3) &= \mathrm{E}[N(A_1)N(A_2)N(A_3)] \\ &\quad - \textstyle\sum^* \mathrm{E}[N(A_1)N(A_2 \cap A_3)] + 2\mathrm{E}[N(A_1 \cap A_2 \cap A_3)], \\ M_{[4]}(A_1 \times A_2 \times A_3 \times A_4) &= E[N(A_1)N(A_2)N(A_3)N(A_4)] \\ &\quad - \textstyle\sum^* \mathrm{E}[N(A_1)N(A_2)N(A_3 \cap A_4)] + \sum^* \mathrm{E}[N(A_1 \cap A_2)N(A_3 \cap A_4)] \\ &\quad + 2\textstyle\sum^* \mathrm{E}[N(A_1)N(A_2 \cap A_3 \cap A_4)] - 6\mathrm{E}[N(A_1 \cap A_2 \cap A_3 \cap A_4)]. \end{aligned}$$

5.4.3 (*Continuation*). Find the generalization for $M_{[k]}(A_1 \times \cdots \times A_k)$ for general k, and discuss the relation to the Stirling numbers $D_{j,k}$. Observe that the relation is essentially one between the ordinary product counting measure $N^{(k)}$ and the modified product counting measure consisting of distinct ordered k-tuplets.

5.4.4 Show that $M_3(\mathrm{d}x_1 \times \mathrm{d}x_2 \times \mathrm{d}x_3)$ equals

$$M_{[3]}(\mathrm{d}x_1 \times \mathrm{d}x_2 \times \mathrm{d}x_3) + \sum\nolimits^* M_{[2]}(\mathrm{d}x_1 \times \mathrm{d}x_2)\delta(x_2, x_3) + M_{[1]}(\mathrm{d}x_1)\delta(x_1, x_2, x_3),$$

where $\delta(x_1, x_2)$ and $\delta(x_1, x_2, x_3)$ vanish outside the hyperplane $x_1 = x_2$ and $x_1 = x_2 = x_3$, respectively, and $\sum^*$ is as in Exercise 5.4.2.

5.4.5 (*Continuation*). Show that in general

$$M_k(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k) = \sum_{j=1}^{k} \sum_{\mathcal{V}} M_{[j]}\left(\prod_{i=1}^{j} \mathrm{d}y_i(\mathcal{V}) \right) \delta(\mathcal{V}),$$

where the inner sum is taken over all partitions $\mathcal{V}$ of the k coordinates into j nonempty subsets, the $y_i(\mathcal{V})$ constitute an arbitrary selection of one coordinate from each subset, and $\delta(\mathcal{V})$ is a δ-function that equals zero unless equality holds among the coordinates in each of the nonempty subsets of $\mathcal{V}$ (see Krickeberg, 1974).

5.4.6 (*Continuation*). Show that if a point process is simple, the moment measure M_k completely determines M_j for $j \le k$. [*Hint*: Consider the representation of M_k in terms of the factorial moment measures $M_{[j]}$ with $j \le k$. If the process is simple, each diagonal term for M_k can be identified with one of the $M_{[j]}$.] Provide a counterexample showing that for point processes that are not simple, two distinct processes may have the same M_2 but different M_1 (see Krickeberg, 1974, Theorem 3, Corollary 3).

5.4.7 *Discrete fermion process.* As an analogue of Example 5.4(c), let $\mathcal{X}$ be a discrete space of K points labelled $1, \ldots, K$, and for $k \ge 1$ set

$$m_k(i_1, \ldots, i_k) = \mathrm{E}[N\{i_1\} \cdots N\{i_k\}] \equiv \lambda^k \widetilde{C} \begin{pmatrix} i_1 & \cdots & i_k \\ i_1 & \cdots & i_k \end{pmatrix},$$

where $\widetilde{C} = (c_{ij})$ is a $k \times k$ covariance matrix. Observe that the determinant on the right vanishes if an index is repeated (and hence, in particular, if $k > K$), so that the function $m_k(\cdot)$ is nonzero only for combinations of distinct indices.

Define

$$\begin{aligned} P(1 + \eta_1, \ldots, 1 + \eta_K) &= 1 + \sum_{k=1}^{K} \lambda^k \sum_{\mathrm{comb}} \widetilde{C} \begin{pmatrix} i_1 & \cdots & i_k \\ i_1 & \cdots & i_k \end{pmatrix} \eta_{i_1} \cdots \eta_{i_k} \\ &= \det(I + \lambda D_\eta C), \end{aligned}$$

where $D_\eta = \mathrm{diag}(\eta_1, \ldots, \eta_K)$, C is the $K \times K$ matrix with elements c_{ij}, and $\sum_{\mathrm{comb}}$ is taken over all distinct combinations of k indices from K. Show that, with $z_i = 1 + \eta_i$, $P(\cdot)$ is a proper multivariate p.g.f. [*Hint*: Use the identity

$$(I + \lambda D_z R_\lambda)(I - \lambda C) = I + \lambda D_\eta C,$$

where $R_\lambda = C(I - \lambda C)^{-1}$, leading to $P(z_1, \ldots, z_K) = d(\lambda) \det(I + \lambda D_z R_\lambda)$, where $d(\lambda) = \det(I - \lambda C)$ and thus

$$j_k(i_1, \ldots, i_k) = d(\lambda)\, \lambda^k \widetilde{R}_\lambda \begin{pmatrix} i_1 & \cdots & i_k \\ i_1 & \cdots & i_k \end{pmatrix}$$
$$= \Pr\{N\{i_1\} = \cdots = N\{i_k\} = 1,\ N\{j\} = 0\ \ (j \notin \{i_1, \ldots, i_k\})\}.$$

Check that this expression is nonnegative provided $0 < \lambda < \min\{\lambda_i\}$, where the λ_i solve $d(\lambda) = 0$.]

5.4.8 (*Continuation*). Show that the process of Exercise 5.4.7 satisfies the following:
(i) The process is simple, i.e. $\Pr\{N\{j\} = 0 \text{ or } 1 \text{ for } j = 1, \ldots, K\} = 1$;
(ii) $\mathrm{E}[N\{i\}N\{j\}] = \lambda^2(c_{ii}c_{jj} - |c_{ij}|^2) < \lambda^2 c_{ii}c_{jj} = \mathrm{E}[N\{i\}]\,\mathrm{E}[N\{j\}]$, and hence the values are negatively correlated for all i, j;
(iii) $N(\mathcal{X})$, the total number of points on $\mathcal{X}$, has p.g.f. $d(\lambda(z-1))$;
(iv) $\Pr\{N(\mathcal{X}) = 0\} = d(\lambda)$.
For a dual model, see Exercises 6.2.3–5.

5.4.9 Derive the results asserted in Example 5.4(c) by a passage to the limit from the discrete analogue described in the preceding exercises assuming $C(x, y)$ is bounded and continuous on A and imitating the proofs of the Fredholm theory approach to integral equations. For a dual model, see Exercise 6.2.6.

5.4.10 For the special case of Example 5.4(c) with $\mathcal{X} = \mathbb{R}$ and $C(x, y) = \rho\, \mathrm{e}^{-|x-y|/L}$, the fermion process reduces to a stationary renewal process with interval distributions having density

$$f(x) = \frac{2\rho}{\sqrt{1 - 2\rho L}}\, \mathrm{e}^{-x/L} \sinh\left[(x/L)\sqrt{1 - 2\rho L}\,\right]$$

(see Macchi, 1971b). More generally, a reduction to a renewal process is possible whenever

$$C(x, y)C(y, z) = C(x, z)C(y, y) \qquad (x \le y \le z).$$

For a dual model, see Exercise 6.2.7.

5.5. Generating Functionals and Their Expansions

The factorial moment densities are closely linked, as are the factorial moments in the univariate and finite multivariate cases, to an appropriate version of the generating function concept. In the point process context, the appropriate generalization is the *probability generating functional*, which we introduce as follows. Let $\zeta(\cdot)$ be any bounded complex-valued Borel measurable function; then, for a realization $\{X_i : i = 1, \ldots, N\}$ of a finite point process, the (random) product $\prod_{i=1}^{N} \zeta(x_i)$ is well defined, and on imposing the further requirement that $|\zeta(x)| \le 1$ (all $x \in \mathcal{X}$), its expectation will exist and be finite.

When p.g.fl.s return in Chapter 9, they are defined first much as here and then extended.

Definition 5.5.I. *Let $\mathcal{U}: \mathcal{X} \mapsto \mathbb{C}$ be the class of complex-valued Borel measurable functions satisfying the condition $|\zeta(x)| \leq 1$. Then, for a finite point process, the* probability generating functional (p.g.fl.) *is defined for $\zeta \in \mathcal{U}$ by*

$$G[\zeta] = \mathrm{E}\left(\prod_{i=1}^{N} \zeta(x_i) \right), \tag{5.5.1}$$

where the product is zero if $N > 0$ and $\zeta(x_i) = 0$ for some i, and is unity if $N = 0$.

We can get some feel for the p.g.fl. by taking $A_1, \ldots, A_r$ to be a measurable partition of $\mathcal{X}$ and setting

$$\zeta(x) = \sum_{i=1}^{r} z_i I_{A_i}(x), \tag{5.5.2}$$

where $I_A(x)$ is the indicator function of the set A and $|z_i| \leq 1$ for $i = 1, \ldots, r$. The function ζ in (5.5.2) belongs to $\mathcal{U}$, and substitution in (5.5.1) leads to

$$G\left[\sum_{i=1}^{r} z_i I_{A_i}(\cdot) \right] = \mathrm{E}\left(\prod_{i=1}^{r} z_i^{N(A_i)} \right),$$

which is just the multivariate p.g.f. of the number of points in the sets of the given partition. The case of a general function $\zeta \in \mathcal{U}$ may be regarded as a limiting form of this result, where every infinitesimal region $\mathrm{d}x$ is treated as a separate set in a grand partition of $\mathcal{X}$, and $\zeta(x)$ is the coefficient (z value) of the corresponding indicator function in (5.5.2). In this way, the p.g.fl. provides a portmanteau description of the p.g.f. of all possible finite or infinite families of counting r.v.s $N(\cdot)$. As in the case of an ordinary discrete distribution, the p.g.fl. provides a useful way of summarizing and illuminating the complex combinatorial results associated with the moments and a convenient formal tool for deriving relations between them.

In further analogy to the univariate case, there are two useful expansions of the p.g.fl., the first about $\zeta \equiv 0$ and the second about $\zeta \equiv 1$. The first results directly from the definition (5.5.1) when the expectation is written out in terms of the elements $\{(p_n, \Pi_n)\}$ of the point process or, equivalently, in terms of the Janossy measures $J_n(\cdot)$ [see Conditions 5.3.I and equation (5.3.11)]. For all $\zeta \in \mathcal{U}$, we have

$$G[\zeta] = p_0 + \sum_{n=1}^{\infty} p_n \int_{\mathcal{X}^{(n)}} \zeta(x_1) \cdots \zeta(x_n)\, \Pi_n(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_n) \tag{5.5.3a}$$

$$= J_0 + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathcal{X}^{(n)}} \zeta(x_1) \cdots \zeta(x_n)\, J_n(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_n). \tag{5.5.3b}$$

The second expansion can be derived as a generalization from the case where ζ has the particular form (5.5.2) when the p.g.fl. reduces to a multivariate

p.g.f., and the expansion can be expressed in terms of the multivariate factorial moments. Assuming as in (5.4.8) that the series $\sum_{k=0}^{\infty} \mu_{[k]} z^k$ is convergent for $|z| < \epsilon$ for some $\epsilon > 0$ and expressing the factorial moments of the counting r.v.s in terms of the factorial moment measures (5.4.4), we obtain

$$G\left[\sum_{i=1}^{r} z_i I_{A_i}\right] = G\left[1 + \sum_{i=1}^{r} (z_i - 1) I_{A_i}\right]$$
$$= 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \sum_{k_1 + \cdots + k_r = k} \binom{k}{k_1 \ \cdots \ k_r} \prod_{i=1}^{r} (z_i - 1)^{k_i} M_{[k]}(A_1^{(k_1)} \times \cdots \times A_r^{(k_r)}).$$

The final sum here can be identified with the integral with respect to $M_{[k]}(\cdot)$ of the product $\prod_{i=1}^{r} (z_i - 1)^{k_i} I_{A_i}(x_j)$ so we have

$$G[1 + \eta] = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{\mathcal{X}^{(k)}} \eta(x_1) \cdots \eta(x_k)\, M_{[k]}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k), \qquad (5.5.4)$$

where $\eta(x) = \sum_{i=1}^{r} (z_i - 1) I_{A_i}(x)$ in the special case considered. Since any Borel measurable function can be approximated by simple functions such as η, the general result follows by familiar continuity arguments, using the dominated convergence theorem and the assumed convergence of $\sum \mu_{[k]} z^k$ in $|z| < \epsilon$, supposing that $|\eta(x)| < \epsilon$ for $x \in \mathcal{X}$.

By taking logarithms of the expansions in (5.5.3) and (5.5.4), we can obtain expansions analogous to those in (5.2.10) and (5.2.8). The first of these takes the form, under the condition that $J_0 > 0$,

$$\log G[\zeta] = -K_0 + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathcal{X}^{(n)}} \zeta(x_1) \cdots \zeta(x_n)\, K_n(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_n), \quad (5.5.5)$$

where $J_0 = \exp(-K_0)$ and the $K_n(\cdot)$ $(n = 1, 2, \ldots)$ are symmetric signed measures, which, following Bol'shakov (1969), we call *Khinchin measures*. This expansion is important when the point process is infinitely divisible and can be given a cluster interpretation generalizing that of the compound Poisson distribution (see Section 6.3). Here we note that in this case the measures $K_n(\cdot)/K_0$ can be identified as the Janossy measures of the process characterizing the clusters, so $K_0 = \sum_{n=1}^{\infty} K_n(\mathcal{X}^{(n)})/n!$, and the expansion can be rewritten in the form

$$\log G[\zeta] = \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathcal{X}^{(n)}} [\zeta(x_1) \cdots \zeta(x_n) - 1]\, K_n(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_n). \qquad (5.5.6)$$

Taking logarithms of the expansions (5.5.4) leads to a development in terms of *factorial cumulant measures* $C_{[k]}$, namely

$$\log G[1 + \eta] = \sum_{k=1}^{\infty} \frac{1}{k!} \int_{\mathcal{X}^{(k)}} \eta(x_1) \cdots \eta(x_k)\, C_{[k]}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k). \qquad (5.5.7)$$

This expansion converges under the same conditions as (5.5.4) itself, namely that the factorial moments $\mu_{[k]}$ of the total population size should satisfy $\sum \mu_{[k]}\epsilon^k < \infty$ for some $\epsilon > 0$ or, equivalently, that the p.g.f. of the total population size should be analytic within a disk $|z| < 1 + \epsilon$. Note that the scope of application of these results can be increased considerably by recalling that $\mathcal{X}$ itself can be deliberately restricted to a subspace such as a finite interval or rectangle of the original space in which the process may not even be finite.

Relations between the factorial cumulant measures and factorial moment measures can be derived from the expansions (5.5.4) and (5.5.7) by formal substitution or by recalling that the measures appearing in those expansions are symmetric: without this restriction, they are not uniquely defined by integral representations such as (5.5.7). For example, by comparing the linear and quadratic terms of ζ, we have

$$\int \zeta(x_1)\, C_{[1]}(\mathrm{d}x_1) = \int \zeta(x_1)\, M_{[1]}(\mathrm{d}x_1), \tag{5.5.8a}$$

$$\begin{aligned}\int_{\mathcal{X}^{(2)}} \zeta(x_1)\zeta(x_2)\, C_{[2]}(\mathrm{d}x_1 \times \mathrm{d}x_2) = \\ \int_{\mathcal{X}^{(2)}} \zeta(x_1)\zeta(x_2)\, M_{[2]}(\mathrm{d}x_1 \times \mathrm{d}x_2) - \int_{\mathcal{X}} \zeta(x_1)\, M_{[1]}(\mathrm{d}x_1) \int_{\mathcal{X}} \zeta(x_2)\, M_{[1]}(\mathrm{d}x_2),\end{aligned} \tag{5.5.8b}$$

which can be abbreviated to

$$C_{[1]}(\mathrm{d}x_1) = M_{[1]}(\mathrm{d}x_1), \tag{5.5.8c}$$

$$C_{[2]}(\mathrm{d}x_1 \times \mathrm{d}x_2) = M_{[2]}(\mathrm{d}x_1 \times \mathrm{d}x_2) - M_{[1]}(\mathrm{d}x_1)\, M_{[1]}(\mathrm{d}x_2). \tag{5.5.8d}$$

The latter statement follows because any Borel measure on $\mathcal{X}^{(2)}$ is determined by its values on rectangles $A \times B$, which in the case of a symmetric measure may be taken to be squares $A \times A$ for which the indicator functions have the form $\zeta(x_1)\zeta(x_2)$. In the sequel, we repeatedly use such infinitesimal notation to represent equality of measures on product spaces. Using this notation, the general relation between $C_{[k]}$ and the factorial moment measures $M_{[j]}$ for $j \le k$ is most conveniently written in the form, analogous to (5.2.19),

$$\begin{aligned}&C_{[k]}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k) \\ &\quad = \sum_{j=1}^{k} (-1)^{j-1}(j-1)! \sum_{\mathcal{T} \in \mathcal{P}_{jk}} \prod_{i=1}^{j} M_{[|S_i(\mathcal{T})|]}(\mathrm{d}x_{i1} \times \cdots \times \mathrm{d}x_{i,|S_i(\mathcal{T})|}).\end{aligned} \tag{5.5.9}$$

To check that (5.5.9) holds, apply Lemma 5.2.VI to the expansions (5.5.4) for the p.g.fl. and (5.5.7) for its logarithm. Note that in (5.5.9), unlike (5.2.19), here we must take explicit note of the elements $x_{i1}, \ldots, x_{i,|S_i(\mathcal{T})|}$ of each constituent set $S_i(\mathcal{T})$ in each partition $\mathcal{T}$ in $\mathcal{P}_{jk}$.

In practice, it is convenient to group together those partitions $\mathcal{T}$ in $\mathcal{P}_{jk}$ that have common numbers of elements in their subsets: using $\sum^*$ to denote summation over such groups, (5.5.9) then yields, for example when $k = 4$,

$$\begin{aligned} C_{[4]}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_4) = {} & M_{[4]}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_4) \\ & - \textstyle\sum^* M_{[1]}(\mathrm{d}x_1) M_{[3]}(\mathrm{d}x_2 \times \mathrm{d}x_3 \times \mathrm{d}x_4) \\ & - \textstyle\sum^* M_{[2]}(\mathrm{d}x_1 \times \mathrm{d}x_2) M_{[2]}(\mathrm{d}x_3 \times \mathrm{d}x_4) \\ & + 2 \textstyle\sum^* M_{[1]}(\mathrm{d}x_1) M_{[1]}(\mathrm{d}x_2) M_{[2]}(\mathrm{d}x_3 \times \mathrm{d}x_4) \\ & - 6 M_{[1]}(\mathrm{d}x_1) \cdots M_{[1]}(\mathrm{d}x_4). \end{aligned} \tag{5.5.10}$$

Here, the first two $\sum^*$ terms come from $\mathcal{P}_{24}$, with four terms in the former sum and three terms in the latter, while the other $\sum^*$ term comes from $\mathcal{P}_{34}$ and has six terms. This expression then compares immediately with the relation in Exercise 5.2.7.

Inverse relations can be derived in the same way and take the form

$$M_{[k]}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k) = \sum_{j=1}^{k} \sum_{\mathcal{T} \in \mathcal{P}_{jk}} \prod_{i=1}^{j} C_{[|S_i(\mathcal{T})|]}(\mathrm{d}x_{i1} \times \cdots \times \mathrm{d}x_{i,|S_i(\mathcal{T})|}). \tag{5.5.11}$$

Just as with integer-valued r.v.s, expansions such as (5.4.9) and (5.5.11) can in principle be combined to provide expressions for the Janossy measures in terms of the factorial cumulant measures and vice versa. While they may appear to be too clumsy to be of any great practical value, when one or more of the entities concerned has a relatively simple structure, as occurs for example with the Poisson process, they can in fact provide a usable theoretical tool (see e.g. Proposition 7.1.III). Similar comments apply to the relations between the Khinchin measures and the factorial moment measures.

For ease of reference, we give at the end of this section a summary of the various expansions of the p.g.fl. $G[\cdot]$ of an a.s. finite point process N, together with the corresponding relations between the associated families of measures. First, we illustrate uses of the p.g.fl. in three examples; for the third of these, concerning branching processes, it is convenient to present here a range of results needed later in the book.

Example 5.5(a) *I.i.d. clusters* [continued from Section 5.1 and Example 5.3(a)]. Returning to our initial example, we see that equation (5.1.1) for the joint p.g.f. of this example is a special case of the general form for the p.g.fl.

$$G[\zeta] = P_N\Big(\textstyle\int_{\mathcal{X}} \zeta(x)\, F(\mathrm{d}x) \Big), \tag{5.5.12}$$

where as before $P_N(\cdot)$ is a p.g.f. of the cluster size and $F(\cdot)$ is the distribution of the individual cluster members about the origin.

The case where $P_N(\cdot)$ has the compound Poisson form (see Theorem 2.2.II)

$$P_N(z) = \mathrm{e}^{-\lambda[1-\Pi(z)]}$$

and $\Pi(\cdot)$ is the p.g.f. of the compounding distribution, is of interest. Expanding $\log G[\zeta]$, we have

$$\log G[\zeta] = \lambda\left[\Pi\Big(\textstyle\int_{\mathcal{X}} \zeta(x)\, F(\mathrm{d}x)\Big) - 1\right] = \lambda \sum_{n=1}^{\infty} \pi_n \left[\left(\int_{\mathcal{X}} \zeta(x)\, F(\mathrm{d}x)\right)^n - 1\right];$$

hence, $K_0 = \lambda$ and for $n = 1, 2, \ldots,$

$$K_n(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_n) = \lambda \pi_n n!\, F(\mathrm{d}x_1) \cdots F(\mathrm{d}x_n).$$

This can be compared with the form for the Janossy measures for which $J_0 = \mathrm{e}^{-\lambda}$ and for $n = 1, 2, \ldots,$

$$J_n(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_n) = \pi_n n!\, F(\mathrm{d}x_1) \cdots F(\mathrm{d}x_n),$$

the interpretation being as follows. The process can be regarded as the superposition of ν i.i.d. nonempty subclusters, where ν has a Poisson distribution with mean λ, and for each subcluster, $K_n(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_n)/K_0$ is the probability that the subcluster consists of n points and that they are located at $\{x_1, \ldots, x_n\}$. The Janossy measure yields as $J_n(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_n)$ the probability that the superposition of the ν subclusters results in n points in all, with these points being located at $\{x_1, \ldots, x_n\}$.

In this particular case, the measures $J_n(\cdot)$ and $K_n(\cdot)$ for $n = 1, 2, \ldots$ differ only by a scale factor that depends on n: this is a consequence of the i.i.d. nature of the locations of the points. In the more complex examples studied in Chapters 6 and 10, this no longer need hold [see also Example 7.1(e)]. □

Example 5.5(b) *P.g.fl. for the local process on A.* Let $\mathcal{V}(A)$ denote the space of all measurable functions h on A satisfying $0 \le h \le 1$, and for $h \in \mathcal{V}(A)$ extend h to all $\mathcal{X}$ by putting $h^*(x) = h(x) I_A(x)$. Then, the p.g.fl. $G_A[h]$ of the local process on A is defined in terms of the global p.g.fl. G by the equation

$$G_A[h] = G[1 - I_A + h^*] \qquad (h \in \mathcal{V}(A)). \tag{5.5.13}$$

This representation follows immediately from the interpretation of the p.g.fl. as the expectation

$$G_A[h] = \mathrm{E}\left[\prod_{x_i \in A} h(x_i)\right] = \mathrm{E}\left[\prod_{x_i \in \mathcal{X}} [1 - I_A(x_i) + h^*(x_i)]\right].$$

Thus, the local Janossy measures can be obtained from an expansion of the p.g.fl. about the function $1 - I_A(\cdot)$ rather than about 0. Specifically,

$$\begin{aligned} G_A[\rho h] &= G[1 - I_A + \rho h^*] \\ &= p_0(A) + \sum_{n=1}^{\infty} \frac{\rho^n}{n!} \int_{A^{(n)}} h(x_1) \cdots h(x_n)\, J_n(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_n \mid A). \end{aligned} \tag{5.5.14}$$

A similar comment applies to the Khinchin measures arising from the expansion of the log p.g.fl. We can introduce local Khinchin measures, $K_n(\cdot \mid A)$ say, via the expansion [see equation (5.5.5)] of $\log G_A[\rho h]$ as

$$\log G[1 - I_A + \rho h^*] = -K_0(A) - \sum_{n=1}^{\infty} \frac{\rho^n}{n!} \int_{A^{(n)}} h(x_1) \cdots h(x_n)\, K_n(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_n \mid A), \tag{5.5.15}$$

where $p_0(A) = \exp[-K_0(A)]$. □

Example 5.5(c) *General branching processes; multiplicative population chains.* This basic model stimulated much of the early discussion of generating functionals and moment measures (see e.g. Bartlett and Kendall, 1951; Moyal, 1962a, b) and may be described as follows. A population evolves in discrete time or generations $t = 0, 1, \ldots$. The members of each generation are characterized by both their total number and their locations in the state space $\mathcal{X}$ in such a way that the population consisting of the tth generation can be described by a finite point process on $\mathcal{X}$. The fundamental *multiplicative* property of the process expresses the fact that the population at the $(t+1)$th generation is built up as the sum or, more properly, the superposition of the contributing processes representing the offspring from each of the members of the tth generation. Here we shall assume that, given the number Z_t and the locations $\{x_{ti}: i = 1, \ldots, Z_t\}$ of the members of the tth generation, the contributing processes to the $(t+1)$th generation are mutually independent and independent of both Z_t and all generations prior to t. This relation is then expressible in the form

$$N_{t+1}(A) = \sum_{i=1}^{Z_t} N(A \mid x_{ti}) \qquad (A \in \mathcal{B}_{\mathcal{X}},\ \ t = 0, 1, \ldots), \tag{5.5.16}$$

where the Z_t finite point processes $\{N(\cdot \mid x_{ti}): i = 1, \ldots, Z_t\}$ are mutually independent. The distributions of the contributing or offspring processes $N(\cdot \mid x)$ may depend on the location x of the parent. They can be specified by probability distributions $\{p_n(x): n = 0, 1, \ldots\}$ and symmetric distributions $\Pi_n(\cdot \mid x)$ as in Conditions 5.3.I with the additional requirement that, for fixed values of their other arguments, the $p_n(x)$ and $\Pi_n(\cdot \mid x)$ are all assumed to be measurable functions of x for each $n = 0, 1, \ldots$. Then, the offspring p.g.fl., $G[\zeta \mid x]$ say, will also be a measurable function, and the relation (5.5.16) can be expressed as

$$G_{t+1}[\zeta \mid N_t] = \prod_{i=1}^{Z_t} G[\zeta \mid x_{ti}], \tag{5.5.17}$$

where the left-hand side represents the conditional p.g.fl. for the $(t+1)$th generation given the number and locations of the members of the tth generation

as specified by the point process N_t. It is clear that the right-hand side is a measurable function of $\{Z_t, x_{ti}\ (i = 1, \ldots, Z_t)\}$ and hence that the left-hand side is a measurable function of the finite process N_t. We may therefore take expectations over the left-hand side with respect to N_t, thus obtaining the relation

$$G_{t+1}[\zeta] = G_t\big[G[\zeta \mid \cdot]\big], \tag{5.5.18}$$

where $G[\zeta \mid \cdot]$ is to be treated as the argument of G_t (note that $G[\zeta \mid \cdot] \in \mathcal{U}$ whenever $\zeta \in \mathcal{U}$). Equation (5.5.18) is a far-reaching generalization of the functional iteration relation for the p.g.f.s of the number of offspring in successive generations of the Galton–Watson process (see also Exercise 5.5.3).

Analogous formulae for the factorial moment measures can be established by similar conditioning arguments or else more formally by expanding the p.g.fl. in powers of ζ and equating like terms. We illustrate these procedures for the expectation measures, denoting by $M(\cdot \mid x)$ the expectation measure for the offspring process $N(\cdot \mid x)$ with a parent at x and by $M_{(t)}(\cdot)$ the expectation measure for the population at the tth generation. Corresponding to (5.5.17), we have

$$M_{(t+1)}(A \mid N_t) = \sum_{i=1}^{Z_t} M(A \mid x_{ti}) = \int_{\mathcal{X}} M(A \mid x)\, N_t(\mathrm{d}x), \tag{5.5.19}$$

where again the measurability of $M(A \mid x)$ as a function of x is clear from the assumptions. Taking expectations with respect to N_t, we then have

$$M_{(t+1)}(A) = \int_{\mathcal{X}} M(A \mid x)\, M_{(t)}(\mathrm{d}x), \tag{5.5.20}$$

showing that the expectation measures for successive generations are obtained by operating on $M_{(0)}(\cdot)$ by successive powers of the integral operator with kernel $M(\cdot \mid x)$. As in the case of a multitype Galton–Watson process (which indeed is the special case when the state space consists of a finite number of discrete points), this operator governs the asymptotic behaviour of the process. In particular, its maximum eigenvalue determines the asymptotic rate of growth (or decay) of the mean population size.

These and many other properties are discussed in standard references on general branching processes (see e.g. Moyal, 1962b; Harris, 1963; Athreya and Ney, 1972; Jagers, 1975). Most attention has been given to the case where $\mathcal{X}$ is compact, which results in behaviour similar to that of the finite multitype case. New types of behaviour occur in the noncompact case: for example, $M(A \mid \cdot)$ may be the kernel of a transient Markov chain, in which case the total mass is preserved but, in contrast to the compact case, the population need not necessarily become extinct—it may continue 'moving' indefinitely across the state space as a kind of population wave. Some further aspects and examples are taken up in the exercises [see also Chapter 12 of MKM (1978) and Liemant *et al.* (1988)]

For an alternative derivation of (5.5.20), write $\zeta = 1 + \eta$ in (5.5.18) and expand the two sides. We have

$$1 + \int_{\mathcal{X}} \eta(x)\, M_{(t+1)}(\mathrm{d}x) + \cdots = 1 + \int_{\mathcal{X}} (G[1+\eta(x)] - 1) M_{(t)}(\mathrm{d}x) + \cdots$$
$$= 1 + \int_{\mathcal{X}} M_{(t)}(\mathrm{d}x) \left(\int_{\mathcal{X}} \eta(u)\, M(\mathrm{d}u \mid x) + \cdots \right) + \cdots,$$

where all terms omitted involve product terms in η. Equating the measures with respect to which η is integrated on each side of the equation, we obtain (5.5.20). This brief illustration is a typical example of the fact that the p.g.fl. acts as a portmanteau device for condensing a broad range of formulae (see also Exercise 5.5.4). □

We conclude this section with a summary of the various expansions of the p.g.fl. $G[\cdot]$ of an a.s. finite point process N, together with the corresponding relations between the associated families of measures. For brevity of notation, the latter are written in density form: they can easily be translated into measure notation [for example, equation (5.5.11) is an analogue of (5.5.28) both for measure notation and analogous expansions]. For point processes that are not a.s. finite, the expansions must be applied to the local process on A, $N(\cdot \cap A)$ say, for any bounded $A \in \mathcal{B}_{\mathcal{X}}$ [see Example 5.5(b)].

Some statements below have already been proved; proofs of the rest are left to the reader.

(I) $G[h]$
Janossy measures

(II) $G[1+\eta]$
Factorial moment measures

(III) $\log G[h]$
Khinchin measures

(IV) $\log G[1+\eta]$
Factorial cumulant measures

(A) Definitions, Ranges of Validity

For suitable measurable functions h and family of measures $\{\mu_n : n = 0, 1, \ldots\}$ with μ_0 a constant and μ_n defined on $\mathcal{B}(\mathcal{X}^{(n)})$, write

$$Y[h, \{\mu_n\}] = \sum_{n=1}^{\infty} \frac{1}{n!} \int_{\mathcal{X}^{(n)}} h(x_1) \cdots h(x_n)\, \mu_n(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_n), \qquad (5.5.21)$$

where $\mathcal{V}$ denotes the class of measurable functions $h : \mathcal{X} \mapsto [0, 1]$ such that $h(x) = 1$ for x outside some bounded Borel set. R denotes the radius of convergence of the p.g.f. $P(z) = \sum_{n=0}^{\infty} p_n z^n = \mathrm{E}(z^{N(\mathcal{X})})$. Always, $R > 1$.

(I) *Janossy Measures* $\{J_n\}$.

$$G[h] = J_0 + Y[h, \{J_n\}], \tag{5.5.22}$$

valid for $h \in \mathcal{V}$ and subject to $\{J_n\}$ satisfying the normalizing condition

$$1 = G[1] = J_0 + \sum_{n=1}^{\infty} \frac{J_n(\mathcal{X}^{(n)})}{n!}\,. \tag{5.5.23}$$

$\{J_n(\cdot)/n!\}$ is a probability measure on $\mathcal{X}^{\cup} = \bigcup_{n=0}^{\infty} \mathcal{X}^{(n)}$, with $p_n = J_n(\mathcal{X}^{(n)})/n!$ $(n = 0, 1, \dots)$.

(II) *Factorial Moment Measures* $\{M_{[n]}\}$.

$$G[1 + \eta] = 1 + Y[\eta, \{M_{[n]}\}], \tag{5.5.24}$$

valid for $|1 + \eta| \in \mathcal{V}$ for which $|\eta(x)| < \epsilon$ (all x) provided $R \geq 1 + \epsilon > 1$, imply that all $M_{[n]}(\mathcal{X}^{(n)}) < \infty$, $M_{[0]} = 1$.

(III) *Khinchin Measures* $\{K_n\}$.

$$\log G[h] = -K_0 + Y[h, \{K_n\}], \tag{5.5.25}$$

valid for $h \in \mathcal{V}$ with $K_0 > 0$ and $\{K_n\}$ satisfying the normalizing condition

$$K_0 = \sum_{n=1}^{\infty} \frac{K_n(\mathcal{X}^{(n)})}{n!}\,. \tag{5.5.26}$$

For $n \geq 1$, $K_n(\cdot)$ need not necessarily be nonnegative; if every $K_n(\cdot) \geq 0$, then N is infinitely divisible.

(IV) *Factorial Cumulant Measures* $\{C_{[n]}\}$.

$$\log G[1 + \eta] = Y[\eta, \{C_{[n]}\}], \tag{5.5.27}$$

valid for η as in (II), with $R \geq 1 + \epsilon > 1$ implying that $|C_{[n]}(\mathcal{X}^{(n)})| < \infty$ for all n, $C_{[0]} = 0$.

(B) Relations Between Measures in Different Expansions

The conditions given for validity are sufficient but not always necessary.

(I) $\to$ (II). This is a matter of definition! For n such that $M_{[n]}(\mathcal{X}^{(n)}) < \infty$,

$$m_{[n]}(x_1, \dots, x_n) = \sum_{r=0}^{\infty} \frac{1}{r!} \int_{\mathcal{X}^{(r)}} j_{n+r}(x_1, \dots, x_n, y_1, \dots, y_r)\, \mathrm{d}y_1 \cdots \mathrm{d}y_r. \tag{5.5.28}$$

(II) $\to$ (I). For $R > 2$,

$$j_n(x_1, \dots, x_n) = \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} \int_{\mathcal{X}^{(r)}} m_{[n+r]}(x_1, \dots, x_n, y_1, \dots, y_r)\, \mathrm{d}y_1 \cdots \mathrm{d}y_r. \tag{5.5.29}$$

(I) → (III). $K_0 = -J_0$ (and hence needs $J_0 > 0$) and $R > 1$.

$$k_n(x_1, \ldots, x_n) = \sum_{r=1}^{n} (-1)^{r-1}(r-1)! \sum_{\mathcal{T} \in \mathcal{P}_{rn}} \prod_{i=1}^{r} j_{|S_i(\mathcal{T})|}(x_{i1}, \ldots, x_{i,|S_i(\mathcal{T})|}). \tag{5.5.30}$$

(III) → (I). $J_0 = \exp(-K_0)$ (and hence needs $K_0 < \infty$) and $R > 1$.

$$j_n(x_1, \ldots, x_n) = J_0 \left(\sum_{r=0}^{n} \sum_{\mathcal{T} \in \mathcal{P}_{rn}} \prod_{i=1}^{r} k_{|S_i(\mathcal{T})|}(x_{i1}, \ldots, x_{i,|S_i(\mathcal{T})|}) \right). \tag{5.5.31}$$

(III) → (IV) and (IV) → (III). These are the direct analogues of the relations between (I) and (II), noting that $C_{[0]} = 0$. Valid for $R > 2$.

(II) → (IV) and (IV) → (II). These are the direct analogues of the relations between (I) and (III), noting that $M_{[0]} = 1$. Valid for $R > 2$.

Exercises and Complements to Section 5.5

5.5.1 [Section 5.1 and Examples 5.3(a) and 5.5(a)]. Derive (5.1.1) from (5.5.12) by putting $\xi(x) = \sum_{i=1}^{j} z_i I_{A_i}(x)$, where $\{A_1, \ldots, A_j\}$ is a finite partition of $\mathcal{X}$. Put $\xi = 1 + \eta$ to establish the formal relation

$$G[1+\eta] = 1 + \sum_{k=1}^{\infty} \frac{\mu_{[k]}}{k!} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \eta(x_1) \cdots \eta(x_k)\, \Pi(\mathrm{d}x_1) \cdots \Pi(\mathrm{d}x_k),$$

and hence, when $\mu_{[k]} = \mathrm{E}(N^{(k)}) < \infty$,

$$M_{[k]}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k) = \mu_{[k]} \Pi(\mathrm{d}x_1) \cdots \Pi(\mathrm{d}x_k),$$

of which the case $k = 2$ appears in (5.1.3).

5.5.2 For a Gibbs process as in Example 5.3(c), express the Khinchin densities in terms of the interaction potentials $\psi_r(\cdot)$.

More generally, for finite point processes for which the Janossy densities exist, explore the relationship between Khinchin densities and the interaction potentials $\psi_r(\cdot)$ (see Exercise 5.3.7).

5.5.3 *Branching process* [continued from Example 5.5(c)]. Let $G_t[\zeta \mid x]$ denote the p.g.fl. for the point process $N_t(\cdot \mid x)$ describing the points that constitute the tth generation of the process of Example 5.5(c) starting from a single ancestor at x; so, $G_1[\zeta \mid x] = G[\zeta \mid x]$. Show that for all $k = 1, \ldots, t-1$,

$$G_t[\zeta \mid x] = G_{t-k}[G_k[\zeta \mid \cdot\,] \mid x] = G^{(t)}[\zeta \mid x],$$

where $G^{(t)}[\zeta \mid x]$ is the tth functional iterate of $G[\cdot \mid \cdot\,]$ [see (5.5.18)].

5.5.4 (*Continuation*). Let $q_t(x)$ denote the probability of extinction within t generations starting from a single ancestor at x, so that $q_t(x) = \Pr\{N_t(\mathcal{X} \mid x) = 0\}$. Show that for each fixed $x \in \mathcal{X}$, $\{q_t(x)\colon t = 0, 1, \ldots\}$ is a monotonically decreasing sequence and that, for $k = 1, \ldots, t-1$,

$$q_t(x) = G_{t-k}[q_k(\cdot) \mid x],$$

so, in particular, $q_{t+1}(x) = G[q_t(\cdot) \mid x]$. Deduce that the probability of ultimate extinction starting from an initial ancestor at x, $q(x)$ say, is the smallest nonnegative solution of the equation $q(x) = G[q(\cdot) \mid x]$.

5.5.5 (*Continuation*). Show that the first-moment measure $M_{(t)}(\cdot \mid x)$ of $N_t(\cdot \mid x)$ and the second factorial cumulant measure, $C^{(t)}_{[2]}(A \times B \mid x)$ say, of $N_t(\cdot \mid x)$ satisfy the recurrence relations (with $M \equiv M_{(1)}$)

$$\begin{aligned} M_{(t+1)}(A \mid x) &= \int_{\mathcal{X}} M_{(t)}(A \mid y)\, M(\mathrm{d}y \mid x), \\ C^{(t+1)}_{[2]}(A \times B \mid x) &= \int_{\mathcal{X}^{(2)}} M_{(t)}(A \mid y) M_{(t)}(B \mid z)\, C_{[2]}(\mathrm{d}y \times \mathrm{d}z) \\ &\qquad + \int_{\mathcal{X}} C^{(t)}_{[2]}(A \times B \mid y)\, M(\mathrm{d}y \mid x). \end{aligned}$$

[*Hint*: Use $N_{t+1}(A \mid X) =_d \sum_{x_i} N_t(A \mid x_i)$, where the $\{x_i\}$ denote the individuals of the first generation; see also equations (6.3.3–5).]

5.5.6 (*Continuation*). Let $H_t[\zeta \mid x]$ denote the p.g.fl. for all individuals up to and including those in the tth generation starting from an initial ancestor at x. Show that these p.g.fl.s satisfy the recurrence relations

$$H_{t+1}[\zeta \mid x] = \zeta(x) G[H_t[\zeta \mid \cdot\,] \mid x].$$

Show also that, if extinction is certain, the total population over all generations has p.g.fl. $H[\zeta \mid \cdot\,]$, which for $0 < \zeta < 1$ is the smallest nonnegative solution to the functional equation

$$H[\zeta \mid x] = \zeta(x) G[H[\zeta \mid \cdot\,] \mid x],$$

and find equations for the corresponding first two moment measures.

5.5.7 *Model for the spread of infection.* Take $\mathcal{X} = \mathbb{R}^d$, and suppose that any individual infected at x in turn gives rise to infected individuals according to a Poisson process with parameter measure $\mu(\cdot \mid x) = \mu(\cdot - x \mid 0) \equiv \mu(\cdot - x)$, where $\int_{\mathcal{X}} \mu(\mathrm{d}u) = \nu < 1$. Show that the total number $N(\mathcal{X} \mid 0)$ of infected individuals, starting from one individual infected at 0, is finite with probability 1 and that the p.g.fl. $H[\cdot \mid \cdot\,]$ for the entire population of infected individuals satisfies the functional equation

$$H[\zeta \mid 0] = \zeta(0) \exp\left(- \int_{\mathcal{X}} (1 - H[\zeta \mid u])\, \mu(\mathrm{d}u) \right),$$

where $H[\zeta \mid u] = H[T_u \zeta \mid 0]$ and $T_u \zeta(v) = \zeta(v + u)$.

Deduce, in particular, the following:

(i) The p.g.f. of $N(\mathcal{X} \mid 0)$ satisfies $f(z) \equiv \mathrm{E}z^{N(\mathcal{X}|0)} = z\exp[-\nu(1-f(z))]$.

(ii) The expectation measure $M(\cdot \mid 0)$ for the total population of infected individuals, given an initial infected individual at the origin, satisfies

$$\begin{aligned} M(A \mid 0) &= \delta_0(A) + \int_{\mathcal{X}} M(A-u \mid 0)\,\mu(\mathrm{d}u) \\ &= \delta_0(A) + \mu(A) + \mu^{2*}(A) + \cdots . \end{aligned}$$

(iii) The second factorial moment measure $M_{[2]}(A \times B \mid 0)$ of $N(\cdot \mid 0)$ satisfies

$$\begin{aligned} M_{[2]}(A \times B \mid 0) &= M(A \mid 0)M(B \mid 0) \\ &\quad + \int_{\mathcal{X}} M_{[2]}(A-u, B-u \mid 0)\,\mu(\mathrm{d}u) - \delta_0(A)\delta_0(B). \end{aligned}$$

(iv) The Fourier transforms for $M(\cdot \mid 0)$ and $M_{[2]}(\cdot \mid 0)$ are expressible in terms of $\tilde{\mu}(\theta) = \int_{\mathcal{X}} \mathrm{e}^{i\theta\cdot x}\,\mu(\mathrm{d}x)$ thus:

$$\widetilde{M}(\theta \mid 0) = \int_{\mathcal{X}} \mathrm{e}^{i\theta\cdot x}\,M(\mathrm{d}x \mid 0) = \frac{1}{1-\tilde{\mu}(\theta)}\,,$$

$$\widetilde{M}_{[2]}(\theta,\phi \mid 0) = \iint \mathrm{e}^{i(\theta\cdot x+\phi\cdot y)}\,M_{[2]}(\mathrm{d}x \times \mathrm{d}y \mid 0) = \frac{\widetilde{M}(\theta \mid 0)\widetilde{M}(\phi \mid 0)-1}{1-\tilde{\mu}(\theta+\phi)}\,.$$

5.5.8 *Age-dependent branching process.* Let $\mathcal{X} = \mathbb{R}$, and suppose that an individual born at time u produces offspring according to a Poisson process with parameter measure $\mu(\cdot \mid u) = \mu(\cdot - u \mid 0) \equiv \mu(\cdot - u)$ for some boundedly finite measure $\mu(\cdot)$ that vanishes on $(-\infty, 0]$. Let $G_t[h \mid 0]$ denote the p.g.fl. for the *ages* of individuals present in the population at time t starting from a single newly born individual at time 0.

(a) Show that G_t satisfies the equation

$$G_t[h \mid 0] = h(t)\,\exp\left(-\int_0^t (1-G_t[h \mid u])\,\mu(\mathrm{d}u)\right),$$

where $G_t[h \mid u] = G_{t-u}[h \mid 0]$ for $0 < u < t$.

(b) When $\mu(A) = \mu\ell(A \cap \mathbb{R}_+)$, show that

$$G_t[h \mid 0] = h(t)\left[1 + \mu\int_0^t [1-h(u)]\mathrm{e}^{\mu(t-u)}\,\mathrm{d}u\right]^{-1}.$$

5.5.9 Equation (5.5.29) expresses Janossy densities in terms of factorial moment densities when $R > 2$. Investigate whether the relation in Exercise 5.2.4 has an analogue for densities valid when only $R > 1$.

CHAPTER 6

Models Constructed via Conditioning: Cox, Cluster, and Marked Point Processes

In this chapter, we bring together a number of the most widely used classes of point process models. Their common theme is the generation of the final model by a two-stage construction: first, the generation of an indexed family of processes, and then an operation applied to members of the family to produce the final process. The first two classes (Cox and cluster processes) extend the simple Poisson process in much the same way that the mixed and compound Poisson distributions extend the basic Poisson distribution. Independence plays a central role and leads to elegant results for moment and generating functional relationships. Both processes are used typically in contexts where the realizations are stationary and therefore define infinite collections of points. To deal with these issues, we anticipate the transition from finite to general point processes to be carried out in Chapter 9 and present in Section 6.1 a short review of some key results for more general point processes and random measures.

The third class of processes considered in this chapter represents a generalization in a different direction. In many situations, events are characterized by both a location and a weight or other distinguishing attribute. Such processes are already covered formally by the general theory, as they can be represented as a special type of point process on a product space. However, marked point processes are deserving of study in their own right because of their wide range of applications, such as in queueing theory, and their conceptual importance in contexts such as Palm theory (see [MKM] especially).

6.1. Infinite Point Families and Random Measures

Although the framework developed for finite point processes in Chapter 5 needs to be extended, it nevertheless contains the essential ingredients of the

more general theory. We retain the assumption that the points are located within a complete, separable metric space (c.s.m.s.) $\mathcal{X}$, and will generally interpret $\mathcal{X}$ as either $\mathbb{R}^1$ or $\mathbb{R}^2$.

The space $\mathcal{X}^{\cup}$ as in (5.3.10) is no longer the appropriate space for defining the realizations; instead we move to a description of the realizations in terms of *counting measures*, meaning measures whose values on Borel sets are non-negative integers. The interpretation is that the value of the measure on such a set counts the number of points falling inside that set.

A basic assumption, which really defines the extent of current point process theory, is that the measures are *boundedly finite*: only a finite number of points fall inside any bounded set (i.e. there are no finite accumulation points). In the martingale language of Chapters 7 and 14, this is equivalent to requiring the realizations to be 'nonexplosive'. The space $\mathcal{X}^{\cup}$ is then replaced by the space[1] $\mathcal{N}_{\mathcal{X}}^{\#}$ of all *boundedly finite counting measures on* $\mathcal{X}$. A remarkable feature is that a relatively simple and natural distance between counting measures can be defined and allows $\mathcal{N}_{\mathcal{X}}^{\#}$ to be interpreted as a metric space in its own right. It then acquires a natural topology and a natural family of Borel sets $\mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#})$ that can be used to define measures on $\mathcal{N}_{\mathcal{X}}^{\#}$. We shall not give details here but refer to Chapter 9 and Appendix A2.6.

Thus, the way is open to formally introducing a point process on $\mathcal{X}$ as a *random counting measure on* $\mathcal{X}$, meaning technically a measurable mapping from a probability space $(\Omega, \mathcal{E}, \mathcal{P})$ into the space $(\mathcal{N}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#}))$. Often, the latter space itself is taken as the *canonical probability space* for a point process on $\mathcal{X}$. Every distinct probability measure on $(\mathcal{N}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#}))$ defines a distinct point process.

As in the finite case, specific examples of point processes are commonly specified by their *finite-dimensional distributions*, or *fidi* distributions for short. These can no longer be defined globally, as was done through the Janossy measures for a finite point process, but are introduced by specifying consistent joint distributions

$$P_k(A_1, \ldots, A_k; n_1, \ldots, n_k) = \Pr\{N(A_1) = n_1, \ldots, N(A_k) = n_k\} \qquad (6.1.1)$$

for the number of points in finite families of bounded Borel sets. Indeed, this was the way we introduced the Poisson process in Chapter 2.

Consistency here combines conditions of two types: first, the usual conditions (analogous to those for any stochastic process) for consistency of marginal distributions and invariance under simultaneous permutation of the sets and the numbers falling into them; second, conditions to ensure that the realizations are almost surely measures, namely that

$$N(A \cup B) = N(A) + N(B) \quad \text{a.s.} \qquad \text{and} \qquad N(A_n) \to 0 \quad \text{a.s.} \qquad (6.1.2)$$

[1] In this edition, we use $\mathcal{M}_{\mathcal{X}}^{\#}$ (and $\mathcal{N}_{\mathcal{X}}^{\#}$) to denote spaces of boundedly finite (counting) measures on $\mathcal{X}$ where in the first edition we used $\widehat{\mathcal{M}}_{\mathcal{X}}$ (and $\widehat{\mathcal{N}}_{\mathcal{X}}$), respectively.

for (respectively) all disjoint Borel sets A, B, and all sequences $\{A_n\}$ of Borel sets with $A_n \downarrow \emptyset$. These two conditions reduce to the requirements on the fidi distributions that, for all finite families of *disjoint* bounded Borel sets, $(A_1, \ldots, A_k)$,

$$\sum_{r=0}^{n} P_k(A_1, A_2, A_3, \ldots, A_k;\, n-r, r, n_3, \ldots, n_k) = P_{k-1}(A_1 \cup A_2, A_3, \ldots, A_k;\, n, n_3, \ldots, n_k), \tag{6.1.3}$$

and

$$P_1(A_k; 0) \to 1 \tag{6.1.4}$$

for all sequences of bounded Borel sets $\{A_k\}$ with $A_k \downarrow \emptyset$. Moreover, for point processes defined on Euclidean spaces, it is enough for these relationships to hold when the sets are bounded intervals.

Example 6.1(a) *Simple Poisson process on* $\mathbb{R}$. Recall equation (2.2.1):

$$\Pr\{N(a_i, b_i] = n_i,\ i = 1, \ldots, k\} = \prod_{i=1}^{k} \frac{[\lambda(b_i - a_i)]^{n_i}}{n_i!}\, \mathrm{e}^{-\lambda(b_i - a_i)}. \tag{6.1.5}$$

Consistency of the marginals means that if one of the variables, say $N(a_1, b_1]$, is integrated out (by summing over n_1), the resulting quantity is the joint probability corresponding to the remaining variables. Invariance under permutations of the variables means that if the sets and the number of points falling into them are written down in a different order, the resulting probability is not affected. In the present example, both conditions are obvious from the product form of the joint distributions. The additivity requirement (6.1.3) comes from the additivity property of the Poisson distribution: for Poisson random variables N_1 and N_2 that are independent (as is implied here by the product form of the distributions), their sum again has a Poisson distribution. Finally, (6.1.4) follows from the property $\mathrm{e}^{-\delta_n} \to 1$ when $\delta_n \to 0$. □

Moment measures, factorial moment measures, and probability generating functionals can be defined as in Sections 5.4 and 5.5. The main differences are that in defining the moment measures we should restrict ourselves to bounded sets and that in defining the p.g.fl. we should confine ourselves to functions h in $\mathcal{V}(\mathcal{X})$, the space of nonnegative, measurable functions bounded by unity and such that $1 - h(x)$ vanishes outside some bounded set. Within these constraints, the relations between generating functionals, moment measures, and all the various quantities derived from these in Chapter 5 hold much as they did there. A more detailed account, examining existence and convergence conditions, is given in Chapter 9.

For many of the examples that we consider, the point processes will be defined on a Euclidean space and *stationary*, meaning that their fidi distributions are invariant under simultaneous shifts of their arguments: writing

$A + u = \{x + u, x \in A\}$, stationarity means that, for all real u,

$$P_k(A_1, \ldots, A_k; n_1 \ldots n_k) = P_k(A_1 + u, \ldots, A_k + u; n_1, \ldots, n_k). \qquad (6.1.6)$$

The full consequences of this assumption are quite profound (see the foretaste in Chapter 3), but for the present it is enough to note the following.

Proposition 6.1.I (Stationarity Properties)**.**

(i) *A point process with p.g.fl. $G[h]$ is stationary if and only if for all real u, $G[(S_u h)] = G[h]$, where $(S_u h)(x) = h(x - u)$.*

(ii) *If a point process is stationary and the first-moment measure M_1 exists, then M_1 reduces to a multiple of the uniform measure (Lebesgue measure), $M_1(\mathrm{d}x) = m\,\ell(\mathrm{d}x) = m\,\mathrm{d}x$, say.*

(iii) *If a point process is stationary and the second-moment measure M_2 exists, then M_2 reduces to the product of a Lebesgue component along the diagonal $x = y$ and a reduced component*[2]*, $\breve{M}_2(\mathrm{d}u)$ say, where $u = x - y$, orthogonal to the diagonal.*

Proof. The fidi distributions as above are determined by the p.g.fl. and can be evaluated by taking h to be the sum of simple functions on disjoint sets; conversely, the fidi distributions determine the p.g.fl., which has the shift-invariance properties under stationarity. Property (ii) can be proved from Cauchy's functional equation (see Section 3.6), while property (iii) is the measure analogue of the familiar fact that the covariance function of a stationary time series is a function of the difference in the arguments only:

$$c(x, y) = \breve{c}(x - y).$$

Similar expressions for the moment densities follow from property (iii) whenever the moment measures have densities, but in general they have a singular component along the diagonal $x = y$, which reappears as an atom at the origin in the reduced measure $\breve{M}_2(\cdot)$ (see also Section 8.1). General routes to these reduced measures are provided by the factorization theorems in Section A2.7 or by the disintegration theory outlined in Section A1.4 (see Chapter 8 for further discussion and examples). Estimation of these reduced moment measures and their Fourier transforms (spectral measures) is a key issue in the statistical analysis of point process data and will be taken further in Chapter 8 and in more detail in Chapter 12. □

We shall also need the idea of a *random measure*, so we note some elementary properties. The general theory of random measures is so closely interwoven with point process theory that the two can hardly be separated. Point processes are indeed only a special class (integer-valued) of the former,

[2] In this edition, we use $\breve{M}_2(\cdot)$ and $\breve{C}_2(\cdot)$ to denote reduced second moment and covariance measures (and $\breve{m}$ and $\breve{c}$ for their densities) where in the first edition we wrote $\widehat{M}_2(\cdot)$ and $\widehat{C}(\cdot)$, etc.

and much of the general theory runs in parallel for both cases, a fact exploited more systematically in Chapter 9. Here we provide just sufficient background to handle some simple applications.

The formal definition of a random measure $\xi(\cdot)$ proceeds much as in the discussion for point processes given above. Once again, the realizations $\xi(\cdot)$ are required to be a.s. *boundedly finite* and *countably additive*, and their distributional properties are completely specified by their finite-dimensional distributions. Since the values of the measure are no longer integer-valued in general (although still nonnegative), these take the more general form

$$F_k(A_1, \ldots, A_k; x_1, \ldots, x_k) = \Pr\{\xi(A_i) \le x_i,\ i = 1, \ldots, k\}. \tag{6.1.7}$$

The moment measures are defined as for point processes, although the special role played by the factorial moment measures is not sustained, particularly when the realizations are continuous. In place of the p.g.fl., the most useful transform is the *Laplace functional*, defined for $f \in \mathrm{BM}_+(\mathcal{X})$, the space of all nonnegative $f \in \mathrm{BM}(\mathcal{X})$, by

$$L[f] \equiv L_\xi[f] = \mathrm{E}\big[\exp\big(-\textstyle\int_{\mathcal{X}} f(x)\,\xi(\mathrm{d}x)\big)\big]. \tag{6.1.8}$$

[We sometimes write L_ξ as a reminder of the random measure ξ to which the Laplace functional L relates and $\int f\,\mathrm{d}\xi$ as shorthand for the integral in (6.1.8).] Of course, the Laplace functional can also be defined for point processes and is therefore the natural tool when both are discussed together.

Although L_ξ defines (the fidi distributions of) a random measure ξ uniquely, via appropriate inversion theorems, there is no easy counterpart to the expansion of the p.g.fl. about the zero function as in equations (5.5.3). There is, however, a Taylor series expansion for the Laplace functional about $f \equiv 0$, corresponding to the p.g.fl. expansion about $h \equiv 1$. It takes the form

$$\begin{aligned} L[sf] = 1 - s\int_{\mathcal{X}} f(x)\,M_1(\mathrm{d}x) + \frac{s^2}{2!}\int_{\mathcal{X}^{(2)}} f(x_1)f(x_2)\,M_2(\mathrm{d}x_1 \times \mathrm{d}x_2) - \cdots \\ + \frac{(-s)^r}{r!}\int_{\mathcal{X}^{(r)}} f(x_1)\ldots f(x_r)\,M_r(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_r) + \cdots. \end{aligned} \tag{6.1.9}$$

This expression is just the expectation of the expansion of the ordinary Laplace transform of the linear functional $Y = \int_{\mathcal{X}} f(x)\,\xi(\mathrm{d}x)$. Its validity depends first on the existence of all moments of the random measure ξ and second on the convergence, typically in a disk around the origin $s = 0$ with radius determined by the length of the largest interval $(0, r)$ within which the Laplace transform is analytic. Finite Taylor series expansions, when just a limited number of moment measures exist, are possible for imaginary values of s, corresponding to the use of the *characteristic functional*, and are set out in Chapter 9.

EXAMPLE 6.1(b) *Gamma random measures (stationary case).* Suppose that the random variables $\xi(A_i)$ in (6.1.7) are independent for disjoint Borel sets A_i in $\mathbb{R}^d$ and have the gamma distributions with Laplace–Stieltjes transforms

$$\mathrm{E}(\mathrm{e}^{-s\xi(A_i)}) = \psi(A_i, s) = (1 + \lambda s)^{-\alpha\ell(A_i)} \qquad (\lambda > 0,\ \alpha > 0,\ \mathrm{Re}(s) \geq 0), \tag{6.1.10}$$

where $\ell(\cdot)$ denotes Lebesgue measure. By inspection, $\psi(A_i, s) \to 1$ as $s \to 0$, showing that $\xi(A)$ is a.s. finite for any fixed bounded set A. Then, since $\mathcal{X}$ is separable, it can be represented as a denumerable union $\bigcup A_i$ of such sets and

$$\Pr\{\text{at least one } \xi(A_i) \text{ is infinite}\} \leq \sum_{i=1}^{\infty} \Pr\{\xi(A_i) = \infty\} = 0.$$

As in the case of a Poisson process, additivity of ξ is a consequence of independence and the additivity property of the gamma distribution. Also, $\psi(A_i, s) \to 1$ as $\ell(A_i) \to 0$, implying the equivalent of (6.1.4), which guarantees *countable* additivity for ξ and is equivalent to stochastic continuity of the cumulative process $\xi((0,t])$ when the process is on $\mathbb{R}^1$.

The Laplace functional of ξ can be found by extending (6.1.10) to the case where f is a linear combination of indicator functions and generalizing: it takes the form

$$L[f] = \exp\Big(- \int_{\mathcal{X}} \log[1 + \lambda f(x)]\, \alpha\, \ell(\mathrm{d}x) \Big).$$

Expanding this expression as in (6.1.9) and examining the first and second coefficients, we find

$$\begin{aligned} \mathrm{E}\big(\xi(\mathrm{d}x)\big) &= \lambda\alpha\, \ell(\mathrm{d}x), \\ \mathrm{E}\big(\xi(\mathrm{d}x)\, \xi(\mathrm{d}y)\big) &= \lambda^2\alpha^2\, \ell(\mathrm{d}x)\, \ell(\mathrm{d}y) + \delta(x-y)\lambda^2\alpha\, \ell(\mathrm{d}x). \end{aligned} \tag{6.1.11}$$

Thus, the covariance measure for $\xi(\cdot)$ vanishes except for the diagonal component along $x = y$, or, equivalently, the reduced covariance measure is just an atom of mass $\lambda^2\alpha$ at the origin. These features are consequences of the independence of the increments and the purely atomic nature of the sample paths $\xi(\cdot)$, equivalent when $\mathcal{X} = \mathbb{R}^1$ to the pure jump character of the cumulative process (see Section 8.3 for further discussion). From these results, we can also confirm the expressions for the moments as follow directly from (6.1.10), namely

$$\mathrm{E}\xi(A) = \lambda\alpha\, \ell(A) \qquad \text{and} \qquad \mathrm{var}\, \xi(A) = \lambda^2\alpha\, \ell(A).$$

Exercise 6.1.1 gives a more general version of a gamma random measure. □

EXAMPLE 6.1(c) *Quadratic random measure.* Let $Z(t)$ be a Gaussian process with a.s. continuous trajectories, and consider, for any Borel set A, the set function

$$\xi(A) = \int_A Z^2(u)\, \mathrm{d}u.$$

Since Z is a.s. continuous, so is Z^2, so the integral is a.s. well defined and is additive on disjoint sets. In particular, when Z has zero mean, each value $Z^2(t)$ is proportional to a chi-square random variable, so $\xi(A)$ for suitably 'small' sets A is also approximately a chi-square r.v. Generally, $\xi(A)$ can be defined (being an integral) as a limit of linear combinations of $Z^2(t_i)$ for points t_i that become dense in A, and this is quadratic in the Z, hence the name. The random measure properties of ξ are discussed in more detail in Chapter 9. See Exercise 6.1.3 for the first two moments of ξ. □

The next example has a long history. It was originally introduced in early work by Campbell (1909) to describe the properties of thermionic noise in vacuum tubes. Moran (1968, pp. 417–423) gives further details and references. In his work, Campbell developed formulae for the moments, such as

$$\mathrm{E}\left[\int g(x)\, N(\mathrm{d}x)\right] = \int g(x)\, M(\mathrm{d}x),$$

which led Matthes *et al.* (1978) to adopt the term *Campbell measure* for the concept that underlies their treatment of moments and Palm distributions (see also Chapter 13). Since that time, the ideas have appeared repeatedly in applications [see e.g. Vere-Jones and Davies (1966), where the model is referred to as a 'trigger process' and used to describe earthquake clustering]. Here we introduce it as a prelude to the major theme of this chapter. It is, like the other models in the chapter, a two-stage model, for which we consider here only the first stage.

Example 6.1(d) *Intensity of a shot-noise process.* A model for a shot-noise process is that the observations are those of a Poisson point process with a random intensity $\lambda(\cdot)$ with the following structure. A stochastic process $\lambda(t)$ is formed as a filtered version of a simple stationary Poisson process $N(\cdot)$ on $\mathbb{R}$ at rate ν with typical realization $\{t_i\}$, the filtering being effected by

(1) a nonnegative function g that integrates to unity and vanishes on $(-\infty, 0]$, and
(2) random 'multiplier' effects, $\{Y_i\}$, a series of i.i.d. nonnegative random variables with common distribution $F(\cdot)$.

We then define $\lambda(t)$ by

$$\lambda(t) = \sum_{i:t_i<t} Y_i g(t - t_i) = \int_0^\infty Y(u) g(t-u)\, N(\mathrm{d}u), \tag{6.1.12}$$

where $Y(u)$ is a (fictitious) process of i.i.d. variables with distribution F.

Since $\lambda(t)$, when finite, is stationary in t and is measurable, it is locally integrable: indeed, since its arguments are nonnegative, if it has finite expectation it must be finite a.s. For Borel sets A, the integral

$$\xi(A) \equiv \int_A \lambda(u)\, \mathrm{d}u = \sum_i Y_i \int_{A+t_i} g(u)\, \mathrm{d}u$$

is then well defined, though possibly infinite (see Exercise 6.1.4).

The Laplace functional of ξ can be evaluated as follows. We require

$$L[f] = \mathrm{E}\Big[\exp\Big(-\int_{\mathbb{R}} f(u)\lambda(u)\,\mathrm{d}u\Big)\Big].$$

Now, from (6.1.12), the integral can be written as a sum of terms

$$\int_{\mathbb{R}} f(u)\lambda(u)\,\mathrm{d}u = \sum_i Y_i \int_{\mathbb{R}} f(u)g(u-t_i)\,\mathrm{d}u \equiv \sum_i Z_i, \quad \text{say.}$$

If the points t_i are treated as given (i.e. fixed), then the Z_i are independent and, with $\phi(\cdot)$ denoting the common Laplace–Stieltjes transform of the Y_i, $Z_i = Y_i \int_{\mathbb{R}} f(u)g(u-t_i)\,\mathrm{d}u$ has the transform

$$\begin{aligned}\mathrm{E}(\mathrm{e}^{-Z_i}) &= \mathrm{E}\big[\exp\big(-Y_i \textstyle\int_{\mathbb{R}} f(u)g(u-t_i)\,\mathrm{d}u\big)\big] \\ &= \phi\big[\textstyle\int_{\mathbb{R}} f(u)g(u-t_i)\,\mathrm{d}u\big] \;\equiv\; \zeta(t_i),\end{aligned}$$

say, which lies in $(0,1]$ because f, g and the Y_i are all nonnegative. Proceeding formally, the last three equations give us

$$\begin{aligned}L[f] &= \mathrm{E}\big[\textstyle\prod_{t_i\in N}\zeta(t_i)\big] \\ &= G_N[\zeta], \qquad\qquad \text{by definition of a p.g.fl.,} \\ &= \exp\big[\nu\textstyle\int_{\mathbb{R}}[\zeta(t)-1]\,\mathrm{d}t\big], \qquad G_N \text{ is the p.g.fl. of a Poisson process,} \\ &= \exp\big\{\nu\textstyle\int_{\mathbb{R}}\big[\phi\big(\int_{\mathbb{R}} f(u)g(u-t)\,\mathrm{d}u\big)-1\big]\,\mathrm{d}t\big\}.\end{aligned}$$

It is clear from the random measure analogue of Proposition 6.1.I that the random measure $\xi(\cdot)$ here is stationary (we can easily check that $L[S_u f] = L[f]$). With a view to applying the expansion (6.1.9), we find after some manipulation that $L[f]-1$ equals

$$\begin{aligned}&\nu\int\Big[-\mu_1\int f(u)g(u-t)\,\mathrm{d}u + \tfrac{1}{2}\mu_2\int f(u)g(u-t)\,\mathrm{d}u\int f(v)g(v-t)\,\mathrm{d}v - \cdots\Big]\mathrm{d}t \\ &\quad + \tfrac{1}{2}\nu^2\iint\Big[\mu_1^2\int f(u)g(u-t)\,\mathrm{d}u\int f(v)g(v-s)\,\mathrm{d}v + \cdots\Big]\,\mathrm{d}t\,\mathrm{d}s + \cdots,\end{aligned}$$

where $\mu_j = \mathrm{E}(Y^j)$ for $j = 1, 2$. Collect terms, identify the measures associated with first and second powers of $f(\cdot)$, and recall that $\int_{-\infty}^{\infty} g(u)\,\mathrm{d}u = 1$ and $g(u) = 0$ for $u < 0$; then

$$\begin{aligned}M_1(\mathrm{d}t) &= \nu\mu_1\,\mathrm{d}t, \\ M_2(\mathrm{d}s\times\mathrm{d}t) &= \Big[\nu^2\mu_1^2 + \nu\mu_2\int_{-\infty}^{\min(s,t)} g(s-u)g(t-u)\,\mathrm{d}u\Big]\,\mathrm{d}s\,\mathrm{d}t,\end{aligned}$$

so that M_1 has constant density $\nu\mu_1$ and M_2 has the density

$$m(s,t) = \breve{m}_2(v) = \nu^2\mu_1^2 + \nu\mu_2\int_0^{\infty} g(y)g(y+|v|)\,\mathrm{d}y, \qquad \text{where } v = s-t.$$

The fact that M_2 is absolutely continuous stems from the absolute continuity of the trajectories. The appearance of the reduced density $\breve{m}_2$ here is characteristic of the stationary form of the moment measures (see Proposition 8.1.I and onward). □

While these arguments appear intuitively reasonable, to make them rigorous we must check two further points. First, we must establish that the random measure ξ is well defined in the sense that, despite the infinite sums in the definition, the realizations are a.s. boundedly finite; see Exercise 6.1.4.

Second, the implicit conditioning step, consisting here of being given a realization $\{t_i\}$ of the Poisson process and then taking expectations over such realizations, needs to be justified. In a more general context, this task hinges on the technical concept of measurability and is the subject of the next proposition; it appears repeatedly in this and later chapters.

As in Example 6.1(d), the models considered in this chapter are defined in two steps: first, an initial process is laid down and then a secondary process is defined, with distributions conditional on the realization of the initial process. The existence and other properties of such processes depend on extensions of standard theorems concerning the structure of bivariate distributions. Because a realization of a point process (or indeed a more general random measure) can be thought of as a point in a metric space, the same basic apparatus for describing the distributions conditional on the realization of a random measure is available as for dealing with bivariate distributions in $\mathbb{R}^2$. A general discussion of conditions for a bivariate random system in which each component takes its value in a c.s.m.s. is in Proposition A1.5.II. To apply the concepts in a point process context, the key idea we utilize is that of a *measurable family* of point processes or random measures.

Suppose there is given a family $\{N(\cdot \mid y): y \in \mathcal{Y}\}$ of point processes taking their values in the c.s.m.s. $\mathcal{X}$ and indexed by the elements y of the c.s.m.s. $\mathcal{Y}$. This family forms a *measurable family* if, for each set A in $\mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#})$, the function $\mathcal{P}(A \mid y)$ is $\mathcal{B}(\mathcal{Y})$-measurable, where

$$\mathcal{P}(A \mid y) = \Pr\{N(\cdot \mid y) \in A\}. \tag{6.1.13}$$

As in Proposition A1.5.II, we average across a measurable family of point processes to form a new point process as a mixture of the originals.

Proposition 6.1.II. *Suppose there is given*

(a) *a measurable family of point processes $\mathcal{P}(A \mid y)$, defined on the c.s.m.s. $\mathcal{X}$ and indexed by elements of $\mathcal{Y}$, and*

(b) *a $\mathcal{Y}$-valued random variable Y with distribution Π on $\mathcal{B}(\mathcal{Y})$.*

Then the integrals

$$\mathcal{P}(A) = \mathrm{E}[\mathcal{P}(A \mid Y)] = \int_{\mathcal{Y}} \mathcal{P}(A \mid y)\, \Pi(\mathrm{d}y) \tag{6.1.14}$$

define a probability measure $\mathcal{P}$ on $\mathcal{B}(\mathcal{X})$ and hence a point process on $\mathcal{X}$.

Corresponding concepts can readily be defined for random measures and are set out in Exercise 6.1.5.

The next lemma gives simple sufficient conditions for checking whether an indexed family of point processes forms a measurable family.

Lemma 6.1.III. *Each of the following conditions is necessary and sufficient to define a measurable family of point processes on a Euclidean space:*
(a) *for all choices of positive integer k, finite unions of disjoint intervals $(B_1, \dots, B_k)$, and nonnegative integers $(n_1, \dots, n_k)$, the fidi probabilities $P_k(B_1 \dots, B_k; n_1, \dots, n_k \mid y)$ are $\mathcal{B}(\mathcal{Y})$-measurable functions of y;*
(b) *for all functions h in the space $\mathcal{V}(\mathcal{X})$, the p.g.fl. $G[h \mid y]$ is a $\mathcal{B}(\mathcal{Y})$-measurable function of y.*

Proof. Denote by $\mathcal{A}$ the class of subsets A of $\mathcal{N}_\mathcal{X}$ for which $\mathcal{P}(A \mid y)$ is measurable in y with respect to $\mathcal{B}(\mathcal{Y})$. If (a) holds, then $\mathcal{A}$ contains the cylinder sets used in defining the fidi probabilities. It follows from the closure properties of families of measurable functions (see Appendix A1.4) that the class $\mathcal{A}$ is closed under monotone limits and therefore contains the σ-field of all subsets of $\mathcal{X}$ generated by the cylinder sets; that is, $\mathcal{A} \supseteq \mathcal{B}(\mathcal{X})$. Hence the given family of point processes forms a measurable family. If, alternatively, (b) holds, then by taking h to be a linear combination of indicator functions and differentiating, we can recover the fidi distributions. Differentiation and the other operations involved preserve measurability so that the result follows from (a). The necessity of (a) is obvious, and that of (b) follows on observing that $G[h \mid y]$ for a general $h \in \mathcal{V}(\mathcal{X})$ can be obtained from the case where h is a linear combination of indicator functions by operations that preserve the measurability in y. □

We can immediately apply this lemma to give sufficient conditions that are simpler to check than those of Proposition 6.1.II.

Corollary 6.1.IV. *Suppose there is given a $\mathcal{Y}$-valued random variable Y with distribution Π on $\mathcal{B}(\mathcal{Y})$ and either*
(a) *a family of fidi probabilities $P_k(B_1 \dots, B_k; n_1, \dots, n_k \mid y)$ satisfying condition* (a) *of Lemma* 6.1.III *or*
(b) *a family of p.g.fl.s $G[h \mid y]$ satisfying condition* (b) *of Lemma* 6.1.III.
For each of these cases, there exists a well-defined point process on $\mathcal{X}$ for which in case (a) *the fidi probabilities are given by*

$$P_k(B_1, \dots, B_k; n_1, \dots, n_k) = \mathrm{E}\big[P_k(B_1, \dots, B_k; n_1, \dots, n_k \mid Y)\big] = \int_\mathcal{Y} P_k(B_1, \dots, B_k; n_1, \dots, n_k \mid y)\, \Pi(\mathrm{d}y) \tag{6.1.15a}$$

and in case (b) *the p.g.fl. is given by*

$$G[h] = \mathrm{E}\big(G[h \mid Y]\big) = \int_\mathcal{Y} G[h \mid y]\, \Pi(\mathrm{d}y). \tag{6.1.15b}$$

The following is perhaps the simplest example to which these ideas apply; their applications will be explored more systematically in the next two sections.

Example 6.1(e) *Mixed Poisson process.* Take the distributions (6.1.5) as a candidate for a measurable family, with the role of y played by λ and that of $\mathcal{Y}$ played by the half-line $\mathbb{R}_+ = [0, \infty)$. For a fixed set of half-open intervals, the function (6.1.5) is a continuous and hence a measurable function of λ so that condition (a) of Lemma 6.1.III is satisfied. Thus, the simple Poisson processes form a measurable family with respect to the real variable λ. Consequently, we can mix (average) them with respect to a distribution Π for λ to obtain the fidi distributions of a new point process. If, for example, Π is the exponential distribution with density $\mu e^{-\mu\lambda}\,d\lambda$, then the number of points falling into any given set A has a geometric distribution $p_n = qp^n$ with parameter $p = \mu/(\mu + |A|)$, $q = 1 - p$. Moreover, the locations of the points in A, given the number of events in A, are uniformly distributed over A.

Alternatively, we could work from the p.g.fl. for the Poisson process, namely $G[h] = \exp\big(-\lambda \int [1-h(u)]\,du\big)$, and take expectations over λ using condition (b) of the lemma and Corollary 6.1.IV. The resultant process has p.g.fl.

$$G[h] = \int_0^\infty \exp\big(-\lambda \textstyle\int [1 - h(u)]\,du\big)\,\Pi(d\lambda) = \Pi^*\big(\textstyle\int [1 - h(u)]\,du\big), \quad (6.1.16)$$

where $\Pi^*(\theta) = \mathrm{E}(e^{-\theta Y})$ is the Laplace–Stieltjes transform of an r.v. Y with distribution Π. In particular, when Π is exponential with mean $1/\mu$, the p.g.fl. reduces to

$$G[h] = \frac{\mu}{\mu + \int [1 - h(u)]\,du}.$$

This reduces to the p.g.f. $\mu/\big[\mu + |A|(1 - z)\big]$ of the geometric distribution described above when we set $h(u) = 1 - (1 - z)I_A(u)$. □

Exercises and Complements to Section 6.1

6.1.1 A *general gamma random measure* on the c.s.m.s. $\mathcal{X}$ can be constructed as a process with independent nonnegative increments for which the increment $\xi(A)$ on the bounded Borel set A has a gamma distribution with Laplace transform

$$\mathrm{E}(e^{-s\xi(A)}) = (1 + \lambda s)^{-\alpha(A)},$$

where the *scale* parameter λ is finite and positive and the *shape* parameter measure $\alpha(\cdot)$ is a boundedly finite measure on $\mathcal{B}_\mathcal{X}$.

(a) Verify that these marginal distributions, coupled with the independent increment property, lead to a well-defined random measure.

(b) In the case $\mathcal{X} = \mathbb{R}$, show that $\xi(\cdot)$ may be regarded as the increments of an underlying nondecreasing stochastic process $X(t)$, which with positive probability is discontinuous at t if and only if $\alpha(\{t\}) > 0$.

(c) Show that ξ has as its Laplace functional

$$L[f] = \exp\left(-\int_\mathcal{X} \log(1 + \lambda f(x))\,\alpha(dx)\right) \qquad (f \in \mathrm{BM}_+(\mathcal{X})).$$

[*Hint*: See Chapter 9 for more detail, especially parts (b) and (c).]

6.1.2 *Stable random measure.* Consider a random measure ξ for which $\mathrm{E}(\mathrm{e}^{-s\xi(A)}) = (1+[\exp(-s^\alpha)])^{-\Lambda(A)}$ for some fixed measure $\Lambda(\cdot)$ and that has independence properties as in Example 6.1(a). Verify that for $0 < \alpha < 1$, there is a well-defined random measure with marginal distributions as stated.

6.1.3 Let ξ be the quadratic random measure of Example 6.1(c) in which the Gaussian process Z is stationary with zero mean, variance σ^2 and $\mathrm{cov}(Z(s), Z(t)) = c(s-t)$. Show that for bounded Borel sets A and B,

$$\mathrm{E}[\xi(A)] = \sigma^2 \ell(A),$$

$$\mathrm{cov}(\xi(A), \xi(B)) = 2\int_A \int_B c^2(u-t)\,\mathrm{d}u\,\mathrm{d}t.$$

6.1.4 *Random measure and shot noise.* Denote by $\{x_i\}$ the points of a stationary Poisson process on $\mathbb{R}$ with rate parameter ν, and let $\{Y_j : j = 0, \pm 1, \ldots\}$ denote a sequence of i.i.d. r.v.s independent of $\{x_j\}$. Let the function g be as in Example 6.1(d). Investigate conditions under which the formally defined process

$$Y(t) = \sum_{x_j \le t} Y_j g(t - x_j)$$

is indeed well defined (e.g. by demanding that the series is absolutely convergent a.s.). Show that sufficient conditions are that

(a) $\mathrm{E}|Y| < \infty$, or else

(b) $g(\cdot)$ is nonincreasing on $\mathbb{R}_+$ and there is an increasing nonnegative function $\tilde{g}(\cdot)$ with $\tilde{g}(t) \to \infty$ as $t \to \infty$ such that $\int_0^\infty \tilde{g}(t)g(t)\,\mathrm{d}t < \infty$ and whose inverse $\tilde{g}^{-1}(\cdot)$ satisfies $\mathrm{E}\tilde{g}^{-1}(|Y|) < \infty$ [see also Daley (1981)].

6.1.5 Write down conditions, analogous to (6.1.13), for a measurable family of random measures, and establish the analogue of Proposition 6.1.II for random measures. Frame sufficient conditions for the existence of a two-stage process similar to those in Lemma 6.1.III and Corollary 6.1.IV but using the Laplace functional in place of the p.g.fl.

6.1.6 Let ξ be a random measure on $\mathcal{X} = \mathbb{R}^d$. For a nonnegative bounded measurable function g, define $G(A) = \int_A g(x)\,\ell(\mathrm{d}x)$ $(A \in \mathcal{B}_\mathcal{X})$, where ℓ denotes Lebesgue measure on $\mathbb{R}^d$, and

$$\eta(A) = \int_\mathcal{X} G(A - x)\,\xi(\mathrm{d}x).$$

(a) Show that $\eta(A)$ is an a.s. finite-valued r.v. for bounded $A \in \mathcal{B}_\mathcal{X}$ and that it is a.s. countably additive on $\mathcal{B}_\mathcal{X}$. Then, the existence theorems in Chapter 9 can be invoked to show that η is a well-defined random measure.

(b) Show that if ξ has moment measures up to order k, so does η, and find the relation between them. Verify that the kth moment measure of η is absolutely continuous with respect to Lebesgue measure on $(\mathbb{R}^d)^{(k)}$.

(c) Denoting the characteristic functionals of ξ and η by $\Phi_\xi[\cdot]$ and $\Phi_\eta[\cdot]$, show that, for $f \in \mathrm{BM}_+(\mathcal{X})$,

$$h(x) = \int_\mathcal{X} f(y)g(y-x)\,\mathrm{d}y$$

is also in $\mathrm{BM}_+(\mathcal{X})$, and $\Phi_\eta[f] = \Phi_\xi[h]$.

6.1.7 (*Continuation*). By its very definition, η is a.s. absolutely continuous with respect to Lebesgue measure, and when ξ is completely random, its density

$$Y(t) \equiv \int_{\mathcal{X}} g(t-x)\,\xi(\mathrm{d}x)$$

is called a *linear process*. [The shot-noise process noted in (6.1.12) is an example; for other references, see e.g. Westcott (1970).] Find the characteristic functional of Y when ξ is a stationary gamma random measure.

6.2. Cox (Doubly Stochastic Poisson) Processes

The doubly stochastic Poisson process—or, more briefly, the *Cox process*, so named in recognition of its appearance in a seminal paper of Cox (1955)—is obtained by randomizing the parameter measure in a Poisson process. It is thus a direct generalization of the mixed Poisson process in Example 6.1(e). We first give a definition, then discuss the consequences of the structural features it incorporates, and finally in Proposition 6.2.II give a more mathematical definition together with a list of properties.

Definition 6.2.I. *Let ξ be a random measure on $\mathcal{X}$. A point process N on $\mathcal{X}$ is a* Cox process directed by ξ *when, conditional on ξ, realizations of N are those of a Poisson process $N(\cdot \mid \xi)$ on $\mathcal{X}$ with parameter measure ξ.*

We must check that such a process is indeed well defined. The probabilities in the Poisson process $N(\cdot \mid \xi)$ are readily seen to be measurable functions of ξ; for example, $P(A;n) = [\xi(A)]^n \mathrm{e}^{-\xi(A)}/n!$ is a measurable function of $\xi(A)$, which in turn is a measurable function of ξ as an element in the metric space $\mathcal{M}_{\mathcal{X}}^{\#}$ of boundedly finite measures on $\mathcal{X}$; hence, we can apply Corollary 6.1.IV(a) and take expectations with respect to the distribution of ξ to obtain a well-defined 'mixed' point process on $\mathcal{X}$.

The finite-dimensional (i.e. fidi) distributions are easily obtained in terms of the distributions of the underlying directing measure ξ and are all of mixed Poisson type. Thus, for example,

$$P(A;k) = \Pr\{N(A)=k\} = \mathrm{E}\left(\frac{[\xi(A)]^k}{k!}\,\mathrm{e}^{-\xi(A)}\right) = \int_0^\infty \frac{x^k}{k!}\,\mathrm{e}^{-x}\,F_A(\mathrm{d}x), \tag{6.2.1}$$

where F_A is the distribution function for the random mass $\xi(A)$.

The factorial moment measures of the Cox process turn out to be the ordinary moment measures of the directing measure; this is because the factorial moment measures for the Poisson process are powers of the directing measure. Thus, denoting by μ_k and γ_k the ordinary moment and cumulant measures for ξ, we have for $k=2$,

$$M_{[2]}(A\times A) = \mathrm{E}\big(\mathrm{E}[N(A)(N(A)-1)\mid\xi]\big) = \mathrm{E}\big([\xi(A)]^2\big) = \mu_2(A\times A)\,,$$

and similarly for the covariance measures

$$C_{[2]}(A \times A) = \gamma_2(A \times A)\,.$$

The algebraic details are most easily handled via the p.g.fl. approach outlined in Corollary 6.1.IV(b). As a function of the parameter measure ξ, the p.g.fl. of the Poisson process can be written, for $h \in \mathcal{V}(\mathcal{X})$, as

$$G[h \mid \xi] = \exp\big(-\textstyle\int_{\mathcal{X}}[1-h(x)]\,\xi(\mathrm{d}x)\big). \tag{6.2.2}$$

For fixed h, this is a measurable function of ξ as an element of $\mathcal{M}_{\mathcal{X}}$. Thus, the family of p.g.fl.s (6.2.2) is a measurable family in the sense of Corollary 6.1.IV(b), which implies that we can indeed construct the p.g.fl of a point process by taking expectations in (6.2.2) with respect to any probability measure for ξ in $\mathcal{M}_{\mathcal{X}}$. The expectation

$$\mathrm{E}\left[\exp\left(-\int_{\mathcal{X}}[1-h(x)]\,\xi(\mathrm{d}x)\right)\right],$$

however, can be identified with the Laplace functional [see (6.1.8)] of the random measure ξ, evaluated at the function $[1-h(x)]$. This establishes the first part of the proposition below. The remaining parts are illustrated above for particular cases and are left for the reader to check in general.

Proposition 6.2.II. *Let ξ be a random measure on the c.s.m.s. $\mathcal{X}$ and L_ξ its Laplace functional. Then, the p.g.fl. of the Cox process directed by the random measure ξ is given by*

$$G[h] = \mathrm{E}\big[\exp\big(\textstyle\int_{\mathcal{X}}[h(x)-1]\,\xi(\mathrm{d}x)\big)\big] = L_\xi[1-h]. \tag{6.2.3}$$

The fidi distributions of a Cox process are of mixed Poisson type, as in (6.2.1); *its moment measures exist up to order n if and only if the same is true for ξ. When finite, the kth factorial moment measure $M_{[k]}$ for the Cox process equals the corresponding ordinary moment measure μ_k for ξ. Similarly, the kth factorial cumulant measure $C_{[k]}$ of the Cox process equals the corresponding ordinary cumulant measure γ_k for ξ.*

Note that this last result implies that the second cumulant measure of a Cox process is nonnegative-definite (see Chapter 8). Also, for bounded $A \in \mathcal{B}_{\mathcal{X}}$,

$$\begin{aligned}\operatorname{var} N(A) &= M_{[1]}(A) + C_{[2]}(A \times A)\\ &= M_{[1]}(A) + \operatorname{var}\big(\xi(A)\big) \;\geq\; M_{[1]}(A) \;=\; \mathrm{E}N(A),\end{aligned}$$

so a Cox process, like a Poisson cluster process, is overdispersed relative to the Poisson process.

Example 6.2(a) *Shot-noise or trigger process* [see Example 6.1(d) and Lowen and Teich (1990)]. We continue the discussion of this example by supposing the (random) function

$$\lambda(t) = \sum_{i:x_i<t} Y_i g(t-x_i) \tag{6.2.4}$$

to be the density of the random measure directing the observed Poisson process. In more picturesque language, the epochs $\{x_i\}$ are trigger events with respective sizes (or weights) $\{Y_i\}$ that decay according to the function g. Note that in the definition it is not necessary to assume that g decays monotonically: integrability is sufficient (see also Exercise 6.1.4).

Now we use the generating function formalism to obtain some elementary properties of the shot-noise process. Conditional on the sequence $\{(x_i, Y_i)\}$, we can appeal to (6.2.2) and write

$$G[h \mid \{(x_i, Y_i)\}] = \exp\left(\sum_i -Y_i \int_{x_i}^{\infty} [1 - h(t)]\, g(t - x_i)\, \mathrm{d}t \right). \tag{6.2.5}$$

Write $\phi(\theta) = \mathrm{E}(\mathrm{e}^{-\theta Y_1})$ for the common Laplace–Stieltjes transform of the $\{Y_i\}$. Taking expectations in (6.2.5) first with respect to $\{Y_i\}$ and then with respect to $\{x_i\}$, we have for the p.g.fl. of the process

$$\begin{aligned} G[h] &= \mathrm{E}\left(\prod_i \phi\Big(\textstyle\int_{x_i}^{\infty} [1 - h(t)]\, g(t - x_i)\, \mathrm{d}t \Big) \right) \\ &= \exp\left(\nu \int_{\mathbb{R}} \Big[\phi\Big(\textstyle\int_x^{\infty} [1 - h(t)]\, g(t - x)\, \mathrm{d}t \Big) - 1 \Big]\, \mathrm{d}x \right). \end{aligned} \tag{6.2.6}$$

By taking logarithms in this expression and expanding, it follows that the point process has factorial cumulant measures existing to as many orders as the r.v.s Y_i have finite moments, as is consistent with Proposition 6.2.II. It also follows that these moment measures are absolutely continuous with densities

$$\begin{aligned} m_1 &= \nu\mu_1 \int_0^{\infty} g(u)\, \mathrm{d}u, \\ c_{[2]}(t_1, t_2) &= \breve{c}_{[2]}(t_1 - t_2) \equiv \breve{c}_{[2]}(t_1') = \nu\mu_2 \int_0^{\infty} g(u) g(t_1' + u)\, \mathrm{d}u, \\ c_{[k]}(t_1, \ldots, t_k) &= \breve{c}_{[k]}(t_1', \ldots, t_{k-1}') \\ &= \nu\mu_k \int_0^{\infty} g(u) g(t_1' + u) \cdots g(t_{k-1}' + u)\, \mathrm{d}u, \end{aligned}$$

where $t_j' = t_j - t_k$ $(j = 1, \ldots, k-1)$ and $\mu_k = \mathrm{E}(Y^k)$. These relations are analogues of Campbell's formulae in the theory of shot noise (see references preceding Example 6.1(c)), while the first two illustrate the proposition insofar as the right-hand sides represent the ordinary cumulants of the directing shot-noise process. The fact that they are absolutely continuous reflects the same property in the realizations of ξ.

The representation (6.2.6) shows that the process can equally be regarded as a Neyman–Scott Poisson cluster process [see Example 6.3(a)]. The fact that the shot-noise process and the associated Neyman–Scott process have

the same p.g.fl. means that they are identical as point processes: no measurements on the point process can distinguish the clustering and doubly stochastic (or Cox) interpretations. This ambiguity of interpretation is an extension of the corresponding ambiguity concerning the dual interpretation of contagious distributions alluded to in Exercise 1.2.3. The possibility of such dual interpretations is not restricted to cluster processes: for example, Exercise 6.2.1 sketches a nontrivial characterization of the class of renewal processes that can be represented as Cox processes. □

EXAMPLE 6.2(b) *Boson processes* (Macchi, 1971a, 1975) [see Example 5.4(c)]. In optical problems concerning light beams of low density, the particulate aspects of light are important, and the emission or reception of individual photons (or more generally bosons) can be treated as a point process in time, or space, or both. A standard approach to modelling this situation is to treat the photon process as a Cox process directed by the fluctuating intensity of the light beam, with this latter phenomenon modelled as the squared modulus of a complex Gaussian process. Thus, for the (density of the) random intensity, we take the function

$$\lambda(t) = \lambda|X(t)|^2 \qquad (\lambda > 0), \tag{6.2.7}$$

where $X(\cdot)$ is a complex Gaussian process with zero mean and complex covariance function $C(s,t)$. The process $\lambda(\cdot)$ is similar to the quadratic random measure discussed in Example 6.1(c) with appropriate attention given to the conventions regarding a complex Gaussian process. These require that $X(t) = U(t) + iV(t)$, where $U(\cdot)$ and $V(\cdot)$ are real Gaussian processes such that

$$\begin{aligned}
\mathrm{E}\big(U(s)U(t)\big) &= \mathrm{E}\big(V(s)V(t)\big) = C_1(s,t),\\
\mathrm{E}\big(U(s)V(t)\big) &= -\mathrm{E}\big(U(t)V(s)\big) = C_2(s,t),\\
C(s,t) &= \mathrm{E}\big(\overline{X}(s)X(t)\big) = 2\big(C_1(s,t) + iC_2(s,t)\big).
\end{aligned}$$

Here it is to be understood that C_1 is real, symmetric, and nonnegative-definite, while C_2 is antisymmetric (so, in particular, $C_2(s,s) = 0$, and $\mathrm{E}[X(s)X(t)] = 0$ for all s,t).

The moments of the process $\lambda(\cdot)$ are given by a classical result concerning the even moments of a complex Gaussian process (see e.g. Goodman and Dubman, 1969)

$$\begin{aligned}
\mathrm{E}\big(\overline{X}(s_1)\cdots\overline{X}(s_k)X(t_1)\cdots X(t_k)\big) &= {}^{+}\begin{vmatrix} C(s_1,t_1) & \cdots & C(s_1,t_k)\\ \vdots & \ddots & \vdots\\ C(s_k,t_1) & \cdots & C(s_k,t_k)\end{vmatrix}^{+}\\
&= C^{+}\begin{pmatrix} s_1,\ldots,s_k\\ t_1,\ldots,t_k\end{pmatrix},
\end{aligned} \tag{6.2.8}$$

where the permanent per $B \equiv {}^{+}|B|^{+}$ of a matrix B contains the same terms as the corresponding determinant $\det B$ but with constant positive signs for each

product of matrix elements in place of the alternating positive and negative signs of the determinant, so, for example,

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix}^{+} = ad + bc.$$

It can be shown (see Minc, 1978) that for any nonnegative-definite Hermitian matrix B, $\operatorname{per} B \geq \det B$.

Equations (6.2.7) and (6.2.8), taken together with Proposition 6.2.I, show that the factorial moment densities for the boson process are given by

$$m_{[k]}(t_1, \ldots, t_k) = \mathrm{E}\big(\lambda(t_1) \cdots \lambda(t_k)\big) = \lambda^k \, C^{+} \begin{pmatrix} t_1, \ldots, t_k \\ t_1, \ldots, t_k \end{pmatrix}. \tag{6.2.9}$$

This result paves the way for a discussion that exactly parallels the discussion of the fermion process of Example 5.4(c). In place of the expansion of the Fredholm determinant $d(\lambda)$ used there, we have here an analogous expansion of the function

$$d^{+}(\lambda) = 1 + \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \int_A \cdots \int_A C^{+} \begin{pmatrix} u_1, \ldots, u_k \\ u_1, \ldots, u_k \end{pmatrix} \mathrm{d}u_1 \cdots \mathrm{d}u_k \,,$$

where as before the observation region A is a closed, bounded set in a general Euclidean space $\mathbb{R}^{\mathrm{d}}$. Corresponding to the expression (5.4.18) for the Fredholm minor is the expression

$$\begin{aligned} &\lambda^k R^{+}_{-\lambda} \begin{pmatrix} x_1, \ldots, x_k \\ y_1, \ldots, y_k \end{pmatrix} \\ &\quad = \frac{1}{d^{+}(\lambda)} \Bigg\{ \lambda^k C^{+} \begin{pmatrix} x_1, \ldots, x_k \\ y_1, \ldots, y_k \end{pmatrix} \\ &\qquad + \lambda^k \sum_{j=1}^{\infty} (-\lambda)^j \int_A \cdots \int_A C^{+} \begin{pmatrix} x_1, \ldots, x_k, \, u_1, \ldots, u_j \\ y_1, \ldots, y_k, \, u_1, \ldots, u_j \end{pmatrix} \mathrm{d}u_1 \cdots \mathrm{d}u_j \Bigg\}. \end{aligned} \tag{6.2.10}$$

This shows that the Janossy measures for the photon process have densities

$$j_k(x_1, \ldots, x_k) = \lambda^k d^{+}(\lambda) R^{+}_{-\lambda} \begin{pmatrix} x_1, \ldots, x_k \\ x_1, \ldots, x_k \end{pmatrix} \qquad (k = 1, 2, \ldots). \tag{6.2.11}$$

Macchi (1971a) established (6.2.11) directly by evaluating the expectation

$$j_k(x_1, \ldots, x_k) = \mathrm{E}\bigg(\lambda(x_1) \cdots \lambda(x_k) \exp\bigg(- \int_A \lambda(u) \, \mathrm{d}u \bigg) \bigg)$$

[see also Grandell (1976) and Exercises 6.2.5–6 for further discussion]. □

EXAMPLE 6.2(c) *A pseudo-Cox process: the Gauss–Poisson process.* The Gauss–Poisson process will be introduced as a two-point cluster process in Example 6.3(d) in the next section. Here we wish only to point out that the p.g.fl. $G[h]$ in (6.3.30) for such a process, if the measures Q_1 and Q_2 there are absolutely continuous with respect to Lebesgue measure, equals

$$\exp\left(\int_{\mathcal{X}}[1-h(x)]\,m(x)\,\mathrm{d}x - \tfrac{1}{2}\int_{\mathcal{X}}\int_{\mathcal{X}}[1-h(x)]\,[1-h(y)]\,c(x,y)\,\mathrm{d}x\,\mathrm{d}y\right),$$

where, in the notation of (6.3.30), in which $Q_2(\cdot)$ is symmetric,

$$m(x)\,\mathrm{d}x = Q_1(\mathrm{d}x) + 2Q_2(\mathrm{d}x\times\mathcal{X}) \quad \text{and} \quad c(x,y)\,\mathrm{d}x\,\mathrm{d}y = 2Q_2(\mathrm{d}x\times\mathrm{d}y).$$

This expression is identical in form with the expression $L^*[1-h]$ for the Laplace functional of a Gaussian process, $\{X(t)\colon t\in\mathbb{R}\}$ say, with mean $m(t) = \mathrm{E}X(t)$ and covariance $c(t,u) = \mathrm{cov}(X(t),X(u))$, provided only that the function $c(t,u)$ is positive-definite. On the other hand, the process is not an example of the construction described in Definition 6.2.I because, a.s., a realization of a Gaussian process takes both positive and negative values, so the notion of a Poisson process with parameter measure with density equal to the realization of such a Gaussian process is void. Newman (1970) coined the name 'Gauss–Poisson' because of this formal property of the p.g.fl. □

This example also serves to illustrate that while the conditions of 6.2.II are sufficient for a functional $L^*[1-h]$ to represent the p.g.fl. of a point process, they are not necessary because the functional displayed at the outset of Example 6.2(c) is not the Laplace functional of a random measure.

Exercises and Complements to Section 6.2

6.2.1 Let $\{I_n\} = \{(a_n,b_n]\colon n=1,2,\ldots\}$ be a sequence of random intervals on $\mathbb{R}_+$ of lengths $X_n = b_n - a_n > 0$ a.s. and having gaps $Y_n = a_{n+1} - b_n > 0$ a.s., with $\{X_n\}$ i.i.d. exponential r.v.s, $\{Y_n\}$ i.i.d. r.v.s independent of $\{X_n\}$ and with finite mean, and $a_1 = 0$. Let a Cox process N on $\mathbb{R}_+$ be directed by a random measure ξ, which has density λ on the set $\bigcup_{n=1}^{\infty} I_n$ and zero elsewhere. Show that $N(\cdot) + \delta_0(\cdot)$ is a renewal process.
[The points of the set $\{a_n,b_n\colon n=1,2,\ldots\}$ are those of an alternating renewal process with exponential lifetimes for one of the underlying lifetime distributions. Kingman (1964) showed, effectively, that any stationary Cox process that is also a stationary renewal process must be directed by the stationary version of the random measure described.]

6.2.2 *Discrete boson process.* Let $C \equiv (c_{ij})$ be a (real or complex) covariance matrix. The discrete counterpart of Example 5.4(c) and its associated exercises is the mixed Poisson process obtained by taking $N(i)$ $(i=1,\ldots,K)$ to be Poisson with random parameter $\lambda|Z_i|^2$, where $Z = (Z_1,\ldots,Z_K)$ has the multivariate normal distribution $N(0,C)$. For $K=1$, this reduces to a geometric distribution with p.g.f. $P(1+\eta) = 1/(1-\lambda c_{11}^2\eta)$. For $K>1$, the multivariate p.g.f. has the form

$$P(1+\eta_1,\ldots,1+\eta_K) = \frac{1}{\det(I-\lambda D_\eta C)}\,, \qquad (6.2.12)$$

where $D_\eta = \mathrm{diag}(\eta_1,\ldots,\eta_K)$.

The factorial moment relations corresponding to (6.2.9) may be written down as follows. For any $k > 0$, let $r_1, \ldots, r_K$ be nonnegative integers such that $r_1 + \cdots + r_K = k$; here, r_j is to be interpreted as the number of repetitions of the index j in defining the factorial moment

$$m_{[k]}(i_1, \ldots, i_k) = \mathrm{E}(N(1)^{[r_1]} \cdots N(K)^{[r_K]}),$$

where the set $(i_1, \ldots, i_k)$ consists of the index j repeated r_j times ($j = 1, \ldots, K$). We then have

$$m_{[k]}(i_1, \ldots, i_k) = \lambda^k C^+ \begin{pmatrix} i_1, \ldots, i_k \\ i_1, \ldots, i_k \end{pmatrix}. \tag{6.2.13}$$

6.2.3 (*Continuation*). The relations (6.2.12) and (6.2.13) of Exercise 6.2.2 are together equivalent to the identity for the reciprocal of the characteristic polynomial

$$\frac{1}{\det(I - \lambda D_\eta C)} = 1 + \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} \sum_{\text{perm}} C^+ \begin{pmatrix} i_1, \ldots, i_k \\ i_1, \ldots, i_k \end{pmatrix} \eta_{i_1} \cdots \eta_{i_k},$$

where the inner summation extends over all distinct permutations of k indices from the set $i_1, \ldots, i_k$ allowing repetitions [this is related to the Master Theorem of MacMahon (1915, Sections 63–66); see also Vere-Jones (1984, 1997)].

6.2.4 (*Continuation*). Using (6.2.12), we have also

$$P(z_1, \ldots, z_K) = d^+(\lambda) \det(I - \lambda D_z R_{-\lambda}),$$

where $R_{-\lambda} = C(I + \lambda C)^{-1}$ and $d^+(\lambda) = \det(I + \lambda C)$. From this p.g.f., we obtain the multivariate probabilities in the form (using the notation of preceding exercises)

$$\begin{aligned} \pi_k(i_1, \ldots, i_k) &= \Pr\{N(j) = r_j \ (j = 1, \ldots, K)\} \\ &= \lambda^k d^+(\lambda) \cdot \frac{R^+_{-\lambda} \begin{pmatrix} i_1, \ldots, i_k \\ i_1, \ldots, i_k \end{pmatrix}}{r_1! \cdots r_k!}. \end{aligned}$$

6.2.5 (*Continuation*). Derive the results of Example 6.2(b) by a suitable passage to the limit of the last three exercises.
[An alternative route to these results uses the expansion of $Z(t)$ in an orthogonal series over A: see Macchi (1971a) and Grandell (1976).]

6.2.6 (*Continuation*). When $C(s,t) = \sigma^2 \mathrm{e}^{-\alpha|s-t|}$ in Example 6.2(b), show that with $\beta = \sqrt{\alpha(\alpha - 2\sigma^2)}$,

$$\Pr\{N(0,T] = 0\} = \mathrm{e}^{\alpha T} \big(\cosh \beta T + (\alpha + 2\sigma^2)\beta^{-1} \sinh \beta T \big)^{-1}.$$

6.3. Cluster Processes

Cluster processes form one of the most important and widely used models in point process studies, whether applied or theoretical. They are natural

models for the locations of objects in the plane or in three-dimensional space, in a remarkable range of contexts: for example, plants, molecules, protozoa, human settlements, stars, galaxies, and earthquake epicentres. Along the time axis, they have been used to model photoelectric emissions, volcano eruptions, arrivals and departures at queueing systems, nerve signals, faults in computer systems, and many other phenomena. The cluster mechanism is also a natural way to describe the locations of individuals from consecutive generations of a branching process, an application with unexpectedly rich mathematical structure as well as its obvious practical applications.

The intuitive motivation of such processes involves two components: the locations of clusters and the locations of elements within a cluster. The superposition of the latter constitutes the 'observed' process. To model the cluster elements, we specify a countable family of point processes $N(\cdot \mid y_i)$ indexed by the cluster centres $\{y_i\}$ (a 'cluster field' in [MKM]). To model the cluster locations, we suppose there is given a process N_c of cluster centres, often unobserved, whose generic realization consists of the points $\{y_i\} \subset \mathcal{Y}$. More often than not, we have $\mathcal{Y} = \mathcal{X}$; it is useful to preserve the notational distinction as a reminder of the structure of the process. The centres y_i act as the germs (= ancestors in the branching process context) for the clusters they generate; it is supposed in general that there are no special features attaching to the points of a given cluster that would allow them to be distinguished from the points in some other cluster. More formally, we have the following definition.

Definition 6.3.I. *N is a* cluster process *on the c.s.m.s. $\mathcal{X}$, with centre process N_c on the c.s.m.s. $\mathcal{Y}$ and component processes the measurable family of point processes $\{N(\cdot \mid y): y \in \mathcal{Y}\}$, when for every bounded $A \in \mathcal{B}_{\mathcal{X}}$,*

$$N(A) = \int_{\mathcal{Y}} N(A \mid y)\, N_c(\mathrm{d}y) = \sum\nolimits_{y_i \in N_c(\cdot)} N(A \mid y_i) < \infty \quad \text{a.s.} \tag{6.3.1}$$

The definition requires the superposition of the clusters to be almost surely boundedly finite. There is, however, no requirement in general that the individual clusters must themselves be a.s. finite [i.e. the condition $N(\mathcal{X} \mid y) < \infty$ a.s. is not necessary], although it is a natural constraint in many examples. A general cluster random measure can be introduced in the same way by allowing the component processes to be random measures (see Exercise 6.3.1).

For the remainder of this section, we require the component processes to be mutually independent. We shall then speak of the component processes as coming from an *independent measurable family* and thereby defining an *independent cluster process*. In this definition, it is to be understood that multiple independent copies of $N(\cdot \mid y)$ are taken when $N_c\{y\} > 1$. If $\mathcal{Y} = \mathcal{X}$ (i.e. the cluster centre process and the component processes are all defined on the same space $\mathcal{X}$ and $\mathcal{X}$ admits translations), then the further constraint that the translated components $N(A - y \mid y)$ are identically distributed may be added, thus producing a natural candidate for a stationary version of the process.

Conditions for the existence of the resultant point process are not so easily obtained as for the Cox process, even though the superposition of the cluster member processes involves only operations that are clearly measurable. The difficulty revolves around the finiteness requirement embodied in equation (6.3.1). The number of clusters that are potentially able to contribute points to a given bounded set soars as the dimension of the state space increases, imposing delicate constraints that have to be met by any proposed existence theorem. For independent cluster processes, the finiteness condition can be rephrased somewhat more formally as follows.

Lemma 6.3.II. *An independent cluster process exists if and only if, for any bounded set $A \in \mathcal{B}_{\mathcal{X}}$,*

$$\int_{\mathcal{Y}} p_A(y)\, N_c(\mathrm{d}y) = \sum\nolimits_{y_i \in N_c} p_A(y_i) < \infty \quad \Pi_c\text{-a.s.}, \tag{6.3.2}$$

where $p_A(y) = \Pr\{N(A \mid y) > 0\}$ for $y \in \mathcal{Y}$ and $A \in \mathcal{B}_{\mathcal{X}}$, and Π_c is the probability measure for the process of cluster centres.

Proof. The sum (6.3.2) is required to converge a.s. as part of the definition of a cluster process. The converse, for given N_c, is an application of the second Borel–Cantelli lemma to the sequence of events

$$E_i = \{\text{cluster } i \text{ contributes at least one point to the set } A\}. \qquad \square$$

The condition of Lemma 6.3.II can alternatively be rephrased in terms of generating functionals (see Exercise 6.3.2). When the components of the process are stationary (i.e. their cluster centre process is stationary and the distribution of the cluster members depends only on their positions relative to the cluster centre), a simple sufficient condition for the resultant cluster process to exist is that the mean cluster size be finite; even in the Poisson case, however, this condition is not necessary (see Exercise 6.3.5 for details).

The moments are easier to handle. Thus, taking expectations conditional on the cluster centres yields

$$\mathrm{E}[N(A) \mid N_c] = \sum\nolimits_{y_i \in N_c} M_1(A \mid y_i) = \int_{\mathcal{Y}} M_1(A \mid y)\, N_c(\mathrm{d}y),$$

where $M_1(\cdot \mid y)$ denotes the expectation measure of the cluster member process with centre at y, assuming this latter exists. From the assumption that the cluster member processes form a measurable family, it follows also that whenever $M_1(A \mid y)$ exists, it defines a measurable kernel (a measure in A for each y and a measurable function of y for each fixed Borel set $A \in \mathcal{B}_{\mathcal{X}}$). Then we can take expectations with respect to the cluster centre process to obtain

$$\mathrm{E}[N(A)] = \int_{\mathcal{Y}} M_1(A \mid y)\, M^c(\mathrm{d}y), \tag{6.3.3}$$

finite or infinite, where $M^c(\cdot) = \mathrm{E}[N_c(\cdot)]$ is the expectation measure for the process of cluster centres. From this representation, it is clear that the first-

moment measure of the resultant process exists if and only if the integral in (6.3.3) is finite for all bounded Borel sets A.

Similar representations hold for the higher-order moment measures. In the case of the second factorial moment measure, for example, we need to consider all possible ways in which two distinct points from the superposition of clusters could fall into the product set $A \times B$ $(A, B \in \mathcal{B}_{\mathcal{X}})$. Here there are two possibilities: either both points come from the same cluster or they come from distinct clusters. Incorporating both cases, supposing the cluster centre process is given, we obtain

$$\begin{aligned}\mathrm{E}[N^{[2]}(A \times B \mid N_c)] = &\int_{\mathcal{Y}} M_{[2]}(A \times B \mid y)\, N_c(\mathrm{d}y) \\ &+ \int_{\mathcal{Y}^{(2)}} M_1(A \mid y_1) M_1(B \mid y_2)\, N_c^{[2]}(\mathrm{d}y_1 \times \mathrm{d}y_2),\end{aligned}$$

where the superscript in $N^{[2]}$ denotes the process of distinct pairs from N and in the second integral we have used the assumption of *independent* clusters. Taking expectations with respect to the cluster centre process, we obtain for the second factorial moment of the cluster process

$$\begin{aligned}M_{[2]}(A \times B) = &\int_{\mathcal{Y}} M_{[2]}(A \times B \mid y)\, M^c(\mathrm{d}y) \\ &+ \int_{\mathcal{Y}^{(2)}} M_1(A \mid y_1) M_1(A \mid y_2)\, M_{[2]}^c(\mathrm{d}y_1 \times \mathrm{d}y_2). \qquad (6.3.4)\end{aligned}$$

Again, the second factorial moment measure of the cluster process exists if and only if the component measures exist and the integrals in (6.3.4) converge. Restated in terms of the factorial cumulant measure, equation (6.3.4) reads

$$\begin{aligned}C_{[2]}(A \times B) = &\int_{\mathcal{Y}^{(2)}} M(A \mid y_1)\, M(B \mid y_2)\, C_{[2]}^c(\mathrm{d}y_1 \times \mathrm{d}y_2) \\ &+ \int_{\mathcal{Y}} M_{[2]}(A \times B \mid y)\, M^c(\mathrm{d}y). \qquad (6.3.5)\end{aligned}$$

Many of these relationships are derived most easily, if somewhat mechanically, from the portmanteau relation for the probability generating functionals, which takes the form, for $h \in \mathcal{V}(\mathcal{X})$ and exploiting the independent cluster assumptions,

$$\begin{aligned}G[h] \;&=\; \mathrm{E}\big(G[h \mid N_c]\big) \;=\; \mathrm{E}\Big[\exp\Big(-\int_{\mathcal{Y}} \big(-\log G_m[h \mid y]\big)\, N_c(\mathrm{d}y)\Big)\Big] \\ &=\; G_c\big[G_m[h \mid \cdot\,]\big], \qquad (6.3.6)\end{aligned}$$

where $G_m[h \mid y]$ for $h \in \mathcal{V}(\mathcal{X})$ is the p.g.fl. of $N(\cdot \mid y)$, and

$$G[h \mid N_c] = \prod_{y_i \in N_c} G_m[h \mid y_i] = \exp\Big[-\int_{\mathcal{Y}} \big(-\log G_m[h \mid y]\big)\, N_c(\mathrm{d}y)\Big] \qquad (6.3.7)$$

is the conditional p.g.fl. of N given N_c. The a.s. convergence of the infinite product in (6.3.7) is equivalent to the a.s. convergence of the sum in Lemma 6.3.II by Exercise 6.3.2. The measurable family requirements of the family of p.g.fl.s for the cluster centres follow from the initial assumptions for the process. Thus, the p.g.fl. representation is valid whenever the cluster process exists.

One class of cluster processes occurs so frequently in applications, and is so important in the theory, that it warrants special attention. In this class, (1°) the cluster centres are the points of a Poisson process, and (2°) the clusters are independent and finite with probability 1. Whenever condition (1°) holds, we speak of a *Poisson cluster process*. The basic existence and moment results for Poisson cluster processes are summarized in the proposition below.

Proposition 6.3.III. *Suppose that the cluster centre process is Poisson with parameter measure $\mu_c(\cdot)$ and that the cluster member processes form an independent measurable family. Then, using the notation above,*

(i) *a necessary and sufficient condition for the existence of the resultant process is the convergence for each bounded $A \in \mathcal{B}_{\mathcal{X}}$ of the integrals*

$$\int_{\mathcal{Y}} p_A(y)\, \mu_c(\mathrm{d}y); \tag{6.3.8}$$

(ii) *when the process exists, its p.g.fl. is given by the expression*

$$G[h] = \exp\left(-\int_{\mathcal{Y}} \big(1 - G_m[h \mid y]\big)\, \mu_c(\mathrm{d}y)\right); \tag{6.3.9}$$

(iii) *the resultant process has first and second factorial moment measures and second factorial cumulant measure given, respectively, for $A, B \in \mathcal{B}_{\mathcal{X}}$, by*

$$M_1(A) = M_{[1]}(A) = \int_{\mathcal{Y}} M_{[1]}(A \mid y)\, \mu_c(\mathrm{d}y), \tag{6.3.10}$$

$$M_{[2]}(A \times B) = \int_{\mathcal{Y}} M_{[2]}(A \times B \mid y)\, \mu_c(\mathrm{d}y) + M_1(A)M_1(B), \tag{6.3.11}$$

$$C_{[2]}(A \times B) = \int_{\mathcal{Y}} M_{[2]}(A \times B \mid y)\, \mu_c(\mathrm{d}y); \tag{6.3.12}$$

(iv) *when $\mathcal{X} = \mathbb{R}^d$, the distribution function F of the distance from the origin to the nearest point of the process is given by*

$$1 - F(r) = \exp\left(-\int_{\mathcal{Y}} p_{S_r(0)}(y)\, \mu_c(\mathrm{d}y)\right), \tag{6.3.13}$$

where $S_r(0)$ is the sphere in $\mathcal{X} = \mathbb{R}^d$ of radius r and centre at 0.

Proof. Since $\mathrm{E}[N_c(\mathrm{d}y)] = M^c(\mathrm{d}y) = \mu_c(\mathrm{d}y)$ for a Poisson cluster process, condition (6.3.8) implies the a.s. convergence of (6.3.2) and hence the existence of the process. If the process exists, then since for $\bar{h} \in \mathcal{V}(\mathcal{Y})$, $G_c[\bar{h}] =$

$\exp\big(-\int[1-\bar{h}(y)]\,\mu_c(\mathrm{d}y)\big)$, equation (6.3.9) is just the appropriate special form of (6.3.6) with $\bar{h}(y) = G_m[h \mid y]$ for $h \in \mathcal{V}(\mathcal{X})$ and so it holds. Putting $h(x) = 1 - I_A(x)$, the integral in (6.3.9) reduces to

$$1 - G_m[1 - I_A(\cdot) \mid y] = p_A(y)\,,$$

from which the necessity of (6.3.8) is obvious.

The moment relations are just restatements of equations (6.3.3–5) for the special case of the Poisson process, where $M^c(\mathrm{d}y) = \mu_c(\mathrm{d}y)$ and $C_{[2]}(\mathrm{d}y_1 \times \mathrm{d}y_2) \equiv 0$. The final equation (6.3.13) is a consequence of the fact that if R is the distance from the origin to the nearest point of the process, then $R > r$ if and only if the sphere $S_r(0)$ contains no point of the process, which yields (6.3.13) as the special case of (6.3.9) with $h(x) = 1 - I_{S_r(0)}(x)$. □

If $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and the process is stationary, and the factorial measures entering into equations (6.3.10–12) have densities, then the latter equations simplify further. In this case, the cluster centre process reduces to a Poisson process with constant intensity μ_c, say, and the first-moment density for the cluster member process can be written

$$m_1(x \mid y) = m_1(x - y \mid 0) \equiv \rho_1(x - y), \quad \text{say.}$$

Similarly, the second factorial moment and cumulant densities can be written

$$m_{[2]}(x_1, x_2 \mid y) = m_{[2]}(x_1 - y, x_2 - y) \equiv \rho_{[2]}(x_1 - y, x_2 - y),$$
$$c_{[2]}(x_1, x_2 \mid y) = c_{[2]}(x_1 - y, x_2 - y) \equiv \gamma_{[2]}(x_1 - y, x_2 - y).$$

Substituting, we obtain simplified forms for the corresponding densities of the cluster process:

$$\begin{aligned} m &= \mu_c \int_{\mathcal{X}} \rho_1(u)\,\mathrm{d}u = \mu_c M_1(\mathcal{X} \mid 0) = \mu_c \mathrm{E}[N_m(\mathcal{X} \mid 0)], \\ \breve{m}_{[2]}(u) &= m_{[2]}(y, y+u) = \mu_c \int_{\mathcal{X}} \rho_{[2]}(w, u+w)\,\mathrm{d}w + m^2, \qquad (6.3.14) \\ \breve{c}_{[2]}(u) &= \mu_c \int_{\mathcal{X}} \rho_{[2]}(w, u+w)\,\mathrm{d}w. \end{aligned}$$

A more systematic treatment of such reduced densities $\breve{m}_{[2]}$ and $\breve{c}_{[2]}$ is given in Section 8.1.

The particularly simple form of these expressions means that it is often possible to obtain explicit expressions for the second moments of the counting process in such examples. Note also that since the cumulant density $\breve{c}_{[2]}(u)$ is everywhere nonnegative, the resultant process is generally *overdispersed* relative to a Poisson process with the same first-moment measure (i.e. it shows greater variance in the number of counts). The alternative terms in the first line of (6.3.14) illustrate the sufficient condition for the existence of the process mentioned earlier and in Exercise 6.3.5: if the mean cluster size $M_1(\mathcal{X} \mid 0)$

is finite, then the first-moment measure of the resultant process exists, and *a fortiori* the resultant process itself exists.

Other aspects of the process, such as interval properties, are generally less easy to obtain. Nevertheless, some partial results may be obtained in this direction via equation (6.3.13). Suppose that $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. Then, from (6.3.13) but using the half-interval $(0, t)$ in place of the 'sphere' $(-t, t)$, the survivor function $S(t)$ [see below (2.1.3)] for the length of the interval from 0 to the first point of the process in $\mathbb{R}_+$ is given by

$$S(t) = \exp\left(- \int_{\mathbb{R}} p(t \mid y)\, \mu_c(\mathrm{d}y) \right), \tag{6.3.15}$$

where $p(t \mid y) = p_{(0,t)}(y)$, a special case of the function $p_A(y)$ in (6.3.2). Taking logarithms of (6.3.15) and differentiating, we see that the hazard function $r(t)$ for this first interval is given by

$$r(t) = - \int_{\mathbb{R}} \frac{\partial p(t \mid y)}{\partial t}\, \mu_c(\mathrm{d}y)\,.$$

When the process is stationary, a further differentiation gives the hazard function $q(\cdot)$ of the distribution of the interval between two consecutive points of the process, as in Exercise 3.4.2.

In higher dimensions, a similar approach may be used for the nearest-neighbour distributions, although explicit expressions here seem harder to determine (see Chapter 15).

In all of Examples 6.3(a)–(e) below, the spaces $\mathcal{X}$ and $\mathcal{Y}$ of Definition 6.3.I are the same.

Example 6.3(a) *The Neyman–Scott process: centre-satellite process; process of i.i.d. clusters* (Neyman and Scott, 1958, 1972; Thompson, 1955; Warren, 1962, 1971). Suppose that the individual cluster members are independently and identically distributed; that is, we are dealing with i.i.d. clusters as in Section 5.1 [see also Examples 5.3(a) and 5.5(a)]. Write $F(\mathrm{d}x \mid y)$ for the probability distribution of the cluster members with cluster centre at y and $Q(z \mid y)$ for the p.g.f. of the total cluster size (assumed finite). Then, the cluster member p.g.fl. is given by (5.5.12), which in the notation above becomes

$$G_m[h \mid y] = Q\left(\int_{\mathcal{X}} h(x)\, F(\mathrm{d}x \mid y) \;\middle|\; y \right), \tag{6.3.16}$$

while the corresponding factorial measures take the form

$$M_{[k]}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k \mid y) = \mu_{[k]}(y) \prod_{i=1}^{k} F(\mathrm{d}x_i \mid y), \tag{6.3.17}$$

where $\mu_{[k]}(y)$ is the kth factorial moment for the cluster size distribution when the cluster centre is at y. Note that if F is degenerate at y, we obtain the compound Poisson process discussed in Example 2.1.10(b) and again in the next section, while if every cluster has exactly one point [so $Q(z \mid y) = z$], we have random translations, first mentioned above at Exercise 2.3.4(b).

In many practical applications with $\mathcal{X} = \mathbb{R}^d$, the cluster centre process is stationary Poisson at rate μ_c, $Q(z \mid y)$ and $\mu_{[k]}(y)$ are independent of y, and $F(\mathrm{d}x \mid y)$ is a function of the vector distance $x - y$ alone and has density function $f(x \mid y) = \breve{f}(x - y) = (\mathrm{d}/\mathrm{d}x)\breve{F}(x - y)$. With these simplifying assumptions, the resultant p.g.fl. takes the compact form

$$G[h] = \exp\left\{\mu_c \int_{\mathbb{R}^d} \left[Q\left(\int_{\mathbb{R}^d} h(y+x)\,\breve{F}(\mathrm{d}x)\right) - 1\right] \mathrm{d}y\right\}, \tag{6.3.18}$$

while from the densities in (6.3.14), the mean rate and second factorial cumulant measures for the resultant process are given by $m = \mu_c\mu_{[1]}$ and

$$\breve{c}_{[2]}(u) = \mu_c\mu_{[2]} \int_{\mathbb{R}^d} \breve{f}(y+u)\,\breve{f}(y)\,\mathrm{d}y, \tag{6.3.19}$$

respectively. Also, for the survivor function $S(t)$ of the interval to the first point in the case $d = 1$, we obtain

$$-\log S(t) = \mu_c \int_{\mathbb{R}} \left[1 - Q\big(1 - \breve{F}(y+t) + \breve{F}(y)\big)\right] \mathrm{d}y \tag{6.3.20}$$

with a pleasing simplification when $\breve{F}(\cdot)$ is the exponential distribution (see Exercise 6.3.7). Exercise 6.3.10 sketches a two-dimensional extension. □

EXAMPLE 6.3(b) *Bartlett–Lewis model: random walk cluster process; Poisson branching process* (Bartlett, 1963; Lewis, 1964a, b). In this example, we take $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ and suppose that the points in a cluster are the successive end points in a finite random walk, starting from and including the cluster centre. The special case where the random walk has unidirectional steps in $\mathbb{R}^1$ (i.e. forms a finite renewal process), was used as a road traffic model in Bartlett (1963) and studied in depth by Lewis (1964a) as a model for computer failures.

A closed-form expression for $G_m[h \mid y]$ does not appear to exist, although for the special case where both the step lengths and the number of steps are independent of the positions of the cluster centre, it can be represented in the form

$$\begin{aligned} h(y)\Bigg(q_0 + q_1 \int_{\mathcal{X}} h(y+x_1)\,F(\mathrm{d}x_1) \\ + q_2 \int_{\mathcal{X}^{(2)}} h(y+x_1)h(y+x_1+x_2)\,F(\mathrm{d}x_1)\,F(\mathrm{d}x_2) + \cdots\Bigg), \end{aligned} \tag{6.3.21}$$

where q_j is the probability that the walk terminates after j steps and F is the common step-length distribution.

Assuming also a constant intensity μ_c for the Poisson process of cluster centres, the mean density takes the form

$$m = \mu_c \sum_{j=0}^{\infty} (j+1)q_j = \mu_c(1 + m_{[1]}), \tag{6.3.22}$$

while the reduced form for the second factorial cumulant measure is given by

$$\breve{C}_{[2]}(\mathrm{d}u) = \mu_c \sum_{j=1}^{\infty} q_j \sum_{k=1}^{j} (j-k+1)\big(F^{k*}(\mathrm{d}u) + F^{k*}(-\mathrm{d}u)\big). \tag{6.3.23}$$

Expressions for the nearest point and nearest-neighbour distance can be obtained at least for the case $\mathcal{X} = \mathbb{R}$ and unidirectional $F(\cdot)$. Under these conditions, the probability $p(t \mid y)$ that a cluster with centre at y has a point in the interval $(0, t)$ is given by

$$p(t \mid y) = \begin{cases} 0 & \text{for } y > t, \\ 1 & \text{for } 0 \le y \le t, \\ \displaystyle\sum_{i=0}^{\infty} r_{i+1} \int_0^{|y|} \big[F(|y| + t - x) - F(|y| - x)\big]\, \mathrm{d}F^{i*}(x) & \text{for } y < 0, \end{cases}$$

where $r_i = \sum_{j=1}^{\infty} q_j$. Substituting in (6.3.17) and simplifying, we obtain for the log survivor and hazard functions

$$-\log S(t) = \mu_c t + \mu_c m_{[1]} \int_0^t [1 - F(x)]\, \mathrm{d}x = mt - \mu_c m_{[1]} \int_0^t F(x)\, \mathrm{d}x, \tag{6.3.24a}$$

$$r(t) = \mu_c + \mu_c m_{[1]}\big(1 - F(t)\big), \tag{6.3.24b}$$

where $1 + m_{[1]} = m/\mu_c$ as in (6.3.22) (see also Exercise 6.3.9). □

The next model, the Hawkes process, figures widely in applications of point processes to seismology, neurophysiology, epidemiology, and reliability. It is also an important model from the theoretical point of view and will figure repeatedly in later sections of this book. One reason for its versatility and popularity is that it combines in the one model both a cluster process representation and a simple conditional intensity representation, which is moreover linear. It comes closest to fulfilling, for point processes, the kind of role that the autoregressive model plays for conventional time series. However, the class of processes that can be approximated by Hawkes processes is more restricted than the class of time series models that can be approximated by autoregressive models. In particular, its representation as a cluster process means that the Hawkes process can only be used in situations that are overdispersed relative to the Poisson model.

In introducing the model, Hawkes (1971a, b, 1972) stressed the linear representation aspect from which the term 'self-exciting' derives. Here we derive its cluster process representation, following Hawkes and Oakes (1974), mainly because this approach leads directly to extensions in higher dimensional spaces but also because it simplifies study of the model.

Example 6.3(c) *Hawkes process: self-exciting process; infectivity model* [see also Examples 6.4(c) (marked Hawkes process), 7.2(b) (conditional intensity representation), 8.2(e) (Bartlett spectrum), 8.5(d) (mutually exciting point

processes) and 8.3(c) (linear prediction formulae)]. The points $\{x_i\}$ of a Hawkes process are of two types: 'immigrants' without extant parents in the process, and 'offspring' that are produced by existing points. An evolutionary construction of the points is as follows. Immigrants $\{y_j\}$, say, arrive according to a Poisson process at constant rate μ_c, while the offspring arise as elements of a finite Poisson process that is associated with some point already constructed. Any point of the process, located at x', say, has the potential to produce further points whose locations are those of a (finite) Poisson process with intensity measure $\mu(A - x')$; we assume that $\mu(\cdot)$ has total mass $\nu \equiv \mu(\mathcal{X}) < 1$ and that all these finite Poisson processes are mutually independent and, given the point that generates them, identically distributed (modulo the shift as noted) and independent of the immigrant process as well. Consequently, each immigrant has the potential to produce descendants whose numbers in successive generations constitute a Galton–Watson branching process with Poisson offspring distribution whose mean is ν. Since $\nu < 1$, this branching process is subcritical and therefore of finite total size with mean $1/(1-\nu) < \infty$ if we include the initial immigrant member. Regard the totality of all progeny of a given immigrant point y_j as a cluster; then the totality of all such immigrant points and their clusters constitutes a Hawkes process.

An important task is to find conditions that ensure the existence of a stationary Hawkes process (i.e. of realizations of point sets $\{x_i\}$ on the whole space $\mathcal{X} = \mathbb{R}^d$ having the structure above and with distributions invariant under translation). Since the immigrant process is stationary, a sufficient condition, by Exercise 6.3.5, is that the mean cluster size be finite [or else, since the immigrant process is Poisson, Proposition 6.3.III(i) can be invoked].

The cluster centres may be regarded as 'infected immigrants' from outside the system and the clusters they generate as the process of new infections they produce. Then, $\mu(\mathrm{d}x)$ is a measure of the infectivity at the point x due to an infected individual at the origin.

The key characteristics of any cluster are the first- and second-moment measures for the total progeny. From Exercise 5.5.6, the first of these is given by

$$M_1(A \mid 0) = \delta_0(A) + \mu(A) + \mu^{2*}(A) + \cdots \qquad (\text{bounded } A \in \mathcal{B}_{\mathcal{X}}),$$

while the second satisfies the integral equation

$$\begin{aligned}\int_{\mathcal{X}} & M_{[2]}(\mathrm{d}y, y + A \mid 0) \\ &= \int_{\mathcal{X}} M_1(y + A \mid 0)\, M_1(\mathrm{d}y \mid 0) - \delta_0(A) + \int_{\mathcal{X}} M_{[2]}(\mathrm{d}u, u + A \mid 0) \int_{\mathcal{X}} \mu(\mathrm{d}v),\end{aligned}$$

so that

$$(1-\nu) \int_{\mathcal{X}} M_{[2]}(\mathrm{d}y, y+A \mid 0) = \int_{\mathcal{X}} M_1(y+A \mid 0)\, M_1(\mathrm{d}y \mid 0) - \delta_0(A). \quad (6.3.25)$$

From the general results (6.3.10–12), it now follows that the mean density of the resultant cluster process is given by

$$m = \lambda M_1(\mathcal{X} \mid 0) = \mu_c/(1-\nu), \tag{6.3.26}$$

while for its factorial covariance measure we have

$$\begin{aligned} \breve{C}_{[2]}(A) &= \mu_c \int_{\mathcal{X}} M_{[2]}(\mathrm{d}y, y + A \mid 0) \\ &= \frac{\mu_c}{1-\nu}\left[\int_{\mathcal{X}} M_1(y + A \mid 0)\, M_1(\mathrm{d}y \mid 0) - \delta_0(A)\right]. \end{aligned} \tag{6.3.27}$$

This corresponds to the reduced density

$$\breve{c}_{[2]}(x) = \frac{\mu_c}{1-\nu}\left[\int_{\mathcal{X}} m_1(y)\, m_1(x+y)\, \mathrm{d}y - \delta_0(x)\right]$$

when $M_1(A \mid 0)$ is absolutely continuous with density $m_1(x)$, say, apart from the δ-function at the origin. An important feature of these formulae is that they lead to simple Fourier transforms, and we exploit this fact later in illustrating the spectral theory in Example 8.2(e).

For a parametric example, with $\mathcal{X} = \mathbb{R}$ and $\mu(\cdot)$ with support in $\mathbb{R}_+$, suppose that for some $\alpha > 0$ and $0 < \nu < 1$

$$\mu(\mathrm{d}x) = \begin{cases} \nu\alpha \mathrm{e}^{-\alpha x}\mathrm{d}x & \text{for } x \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then $M_1(\cdot)$ is absolutely continuous apart from an atom at the origin; for its density $m_1(\cdot)$, we find on $x \geq 0$ that

$$m_1(x) = \delta(x) + \nu\alpha \mathrm{e}^{-\alpha(1-\nu)x}.$$

It follows that $\breve{C}_{[2]}(\cdot)$ is absolutely continuous also, and by substituting in (6.3.26) and (6.3.27), we find that the covariance density of the stationary process is given by

$$\breve{c}_{[2]}(y) = \frac{\mu_c \alpha\nu(1 - \frac{1}{2}\nu)}{(1-\nu)^2} \mathrm{e}^{-\alpha(1-\nu)|y|}. \tag{6.3.28}$$

□

Example 6.3(d) *The Gauss–Poisson process: process of correlated pairs* (Bol'shakov, 1969; Newman, 1970; Milne and Westcott, 1972). This process has the curious distinction of being simultaneously a Neyman–Scott process, a Bartlett–Lewis process, and a pseudo-Cox process [Example 6.2(c)]. Its essential characteristic is that the clusters contain either one or two points (so it exists if and only if the cluster centre process exists). Let one point be taken as the cluster centre, let $F(\mathrm{d}x \mid y)$ denote the distribution of the second point relative to the first, and let $q_1(y)$, $q_2(y)$ be the probabilities of 1 and 2 points, respectively, when the centre is at y. Then, we may regard the process as a special case of the Example 6.3(b) with

$$G_m[h \mid y] = q_1(y)h(y) + q_2(y)h(y) \int_{\mathcal{Y}} h(x)\, F(\mathrm{d}x \mid y)$$

so that for the resultant process (and recall that $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$ here),

$$\log G[h] = \int_{\mathcal{X}} \big(h(y) - 1\big) q_1(y)\, \mu(\mathrm{d}y) + \int_{\mathcal{X}} \int_{\mathcal{X}} \big(h(x)h(y) - 1\big) q_2(y)\, \mu(\mathrm{d}y)\, F(\mathrm{d}x \mid y). \tag{6.3.29}$$

This is not quite in standard form because the measure $q_2(y)\, \mu(\mathrm{d}y)\, F(\mathrm{d}x \mid y)$ is not symmetric in general. However, the value of the p.g.fl. is unaltered when we replace this measure by its symmetrized form $Q_2(\mathrm{d}x \times \mathrm{d}y)$, say, so without loss of generality we may write the p.g.fl. in the form

$$\log G[h] = \int_{\mathcal{X}} \big(h(x) - 1\big)\, Q_1(\mathrm{d}x) + \int_{\mathcal{X}^{(2)}} \big(h(x)h(y) - 1\big)\, Q_2(\mathrm{d}x \times \mathrm{d}y), \tag{6.3.30}$$

where Q_1 and Q_2 are boundedly finite and Q_2 is symmetric with boundedly finite marginals. If now we define $\widetilde{Q}_2 = 2Q_2$ and substitute in (6.3.30), we obtain the standard form in (6.3.32) below using Khinchin measures.

Conversely, given any two such measures Q_1 and Q_2, any expression of the form (6.3.30) represents the p.g.fl. of a process of correlated points because we can first define a measure μ by

$$\mu(A) = Q_1(A) + Q_2(A \times \mathcal{X}),$$

then appeal to the Radon–Nikodym theorem to assert the existence μ-a.e. of nonnegative functions $q_1(\cdot)$, $q_2(\cdot)$ with $q_1(x) + q_2(x) = 1$ satisfying, for all bounded $A \in \mathcal{B}_{\mathcal{X}}$,

$$Q_1(A) = \int_A q_1(x)\, \mu(\mathrm{d}x) \quad \text{and} \quad Q_2(A \times \mathcal{X}) = \int_A q_2(x)\, \mu(\mathrm{d}x),$$

and finally use Proposition A1.5.III concerning regular conditional probabilities to define a family of probability measures $\{F(\cdot \mid x) : x \in \mathcal{X}\}$ by

$$Q_2(A \times B) = \int_A F(B \mid x)\, Q_2(\mathrm{d}x \times \mathcal{X}) = \int_A F(B \mid x) q_2(x)\, \mu(\mathrm{d}x)$$

for all bounded A and all $B \in \mathcal{B}_{\mathcal{X}}$. □

This discussion characterizes the p.g.fl. of such two-point cluster processes, but Milne and Westcott (1972) give the following stronger result.

Proposition 6.3.IV. *For* (6.3.30) *to represent the p.g.fl. of a point process, it is necessary and sufficient that*

(i) *Q_1 and Q_2 be nonnegative and boundedly finite, and*

(ii) *Q_2 have boundedly finite marginals.*

Proof. The additional point to be proved is that (6.3.30) fails to be a p.g.fl. if either Q_1 or Q_2 is a signed measure with nontrivial negative part. Exercise 6.3.11 sketches details [see also Example 6.2(c) and Exercises 6.3.12–13]. □

Observe that for the process with p.g.fl. given by (6.3.30), the expectation and second cumulant measures exist and are given, respectively, by

$$M(\mathrm{d}x) = Q_1(\mathrm{d}x) + Q_2(\mathrm{d}x \times \mathcal{X}) + Q_2(\mathcal{X} \times \mathrm{d}x), \quad (6.3.31a)$$

$$C_{[2]}(\mathrm{d}x_1 \times \mathrm{d}x_2) = Q_2(\mathrm{d}x_1 \times \mathrm{d}x_2) + Q_2(\mathrm{d}x_2 \times \mathrm{d}x_1), \quad (6.3.31b)$$

the representation holding whether or not Q_2 is given in its symmetric version.

It appears to be an open problem to determine conditions similar to those in Proposition 6.3.IV for an expansion such as (6.3.30) with just k terms ($k \geq 3$) to represent the log p.g.fl. of a point process [see Milne and Westcott (1993) for discussion].

Example 6.3(e) *A bivariate Poisson process* [see also Examples 7.3(a) (intensity functions and associated martingales), 7.4(e) (random-time transformation to unit-rate Poisson process) and 8.3(a) (spectral properties), and Exercise 8.3.7 (joint forward recurrence time d.f.)]. A bivariate process can be represented as a process on the product space $\mathcal{X} \times \{1, 2\}$, where indices (or marks) $1, 2$ represent the two component processes. The p.g.fl. expansions are most conveniently written out with the integrals over each component space taken separately. Consider, in particular, a Poisson cluster process on $\mathcal{X} \times \{1, 2\}$ in which the clusters may be of three possible types only: a single point in process 1, a single point in process 2, and a pair of points, one from each process. Arguments analogous to those in the preceding example show that the joint p.g.fl. can be written in the form

$$\log G[h_1, h_2] = \int_{\mathcal{X}} \big(h_1(x) - 1\big)\, Q_1(\mathrm{d}x) + \int_{\mathcal{X}} \big(h_2(x) - 1\big)\, Q_2(\mathrm{d}x) + \int_{\mathcal{X}^{(2)}} \big(h_1(x_1)h_2(x_2) - 1\big)\, Q_3(\mathrm{d}x_1 \times \mathrm{d}x_2),$$

where Q_1, Q_2 and Q_3 are boundedly finite and Q_3 has boundedly finite marginals. The marginal p.g.fl. for process 1 can be found by setting $h_2 = 1$ and is therefore a Poisson process with parameter measure

$$\mu_1(\mathrm{d}x) = Q_1(\mathrm{d}x) + Q_3(\mathrm{d}x \times \mathcal{X});$$

similarly, the process with mark 2 is also Poisson with parameter measure

$$\mu_2(\mathrm{d}x) = Q_2(\mathrm{d}x) + Q_3(\mathcal{X} \times \mathrm{d}x).$$

Finally, the superposition of the two processes is of Gauss–Poisson type, with

$$\widetilde{Q}_1(\mathrm{d}x) = Q_1(\mathrm{d}x) + Q_2(\mathrm{d}x)$$

and (taking the symmetric form)

$$\widetilde{Q}_2(\mathrm{d}x_1 \times \mathrm{d}x_2) = \tfrac{1}{2}[Q_3(\mathrm{d}x_1 \times \mathrm{d}x_2) + Q_3(\mathrm{d}x_2 \times \mathrm{d}x_1)].$$

Evidently, this is the most general example of a bivariate Poisson cluster process with Poisson marginals since clusters of any higher order would introduce higher-order clusters in the marginals and hence destroy the Poisson property.

The resulting fidi distributions are infinitely divisible bivariate Poisson distributions of the kind studied by Holgate (1964) and Milne (1974); see also Griffiths, Milne and Wood (1979). The particular bivariate distribution studied by Dwass and Teicher (1957) corresponds to the situation where the pairs must occur for both processes at the same location x; the resultant process is then not only infinitely divisible but also has complete independence. □

Example 6.3(e) appears in many guises—for example as the joint process of the input and output streams of the M/M/∞ queue. It is closely related to the Gauss–Poisson process, which is nothing other than the 'ground process' (see Section 6.4) of the bivariate example above. We shall use it repeatedly to illustrate the structure of multivariate processes—their moments, spectra, conditional intensities, and compensators. See in particular Example 7.3(a).

There are, of course, many examples of bivariate Poisson processes that are not infinitely divisible; one class may be obtained by mixing over the relative proportions of pairs and single points in the example above (see Exercise 6.3.12). A queueing example is given in Daley (1972a).

The previous examples illustrate the point that the same process can be represented in several equivalent ways as a Poisson cluster process: the Gauss–Poisson process, for example, can be represented either as a Neyman–Scott process or as a Bartlett–Lewis type process for appropriately chosen special cases of those models. This same example also points the way to an intrinsic characterization of Poisson cluster processes. In the next result, the measures $K_k(\cdot)$ are extended versions of the Khinchin measures defined for finite processes by (5.5.5).

Proposition 6.3.V. *The p.g.fl. of every Poisson cluster process with a.s. finite clusters can be uniquely represented in the form*

$$\log G[h] = \sum_{k=1}^{\infty} \frac{1}{k!} \int_{\mathcal{X}^{(k)}} \big(h(x_1) \dots h(x_k) - 1\big)\, K_k(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k), \quad (6.3.32)$$

where the $\{K_k\}$ form a family of symmetric, boundedly finite measures on $\mathcal{B}(\mathcal{X}^{(k)})$ such that each $K_k(\cdot)$ has boundedly finite marginals $K_k(\cdot \times \mathcal{X}^{(k-1)})$, and the sum

$$\sum_{k=1}^{\infty} \frac{1}{k!} \sum_{i=1}^{k} \binom{k}{i} K_k\big(A^{(i)} \times (A^c)^{(k-i)}\big) \quad (6.3.33)$$

is finite for bounded $A \in \mathcal{B}_{\mathcal{X}}$.

Conversely, given any such family of measures $\{K_k : k \geq 1\}$, the p.g.fl. (6.3.32) represents the p.g.fl. of a Poisson cluster process.

Proof. Suppose there is given a Poisson cluster process with cluster centres defined on the space $\mathcal{Y}$ and having parameter measure $\mu_c(\cdot)$. Suppose also

that the clusters are a.s. finite, so that they can be represented in terms of a family of Janossy measures $J_k(\cdot \mid y)$ (see Section 5.3), conditioned by the location y of the cluster centre. Note that by definition these measures are symmetric. Consequently, we consider the quantities $K_k(\cdot)$ defined by setting

$$K_k(B) = \int_{\mathcal{Y}} J_k(B \mid y)\, \mu_c(\mathrm{d}y) \qquad \big(B \in \mathcal{B}(\mathcal{X}^{(k)})\big)$$

and check that they are in fact boundedly finite measures. From Proposition 6.3.III, we know that the integral $\int_{\mathcal{Y}} p_A(y)\, \mu(\mathrm{d}y)$ converges for each bounded set $A \in \mathcal{B}_{\mathcal{X}}$. Here, $p_A(y)$ is just the sum over $k \geq 1$ of the probabilities that the cluster has k members of which at least one falls into the set A, so that, referring to (5.3.10), $p_A(y)$ equals

$$\sum_{k=1}^{\infty} \frac{J_k(\mathcal{X}^{(k)} \mid y) - J_k((A^c)^{(k)} \mid y)}{k!} = \sum_{k=1}^{\infty} \sum_{i=1}^{k} \binom{k}{i} \frac{J_k\big(A^{(i)} \times (A^c)^{(k-i)} \mid y\big)}{k!}\,.$$

The finiteness of $K_k(B)$ follows when B is of the form $A^{(k)}$ for bounded A. Similarly, by taking the term in the sum with $i = 1$, we deduce the bounded finiteness of the marginals. Finally, (6.3.33) is just a restatement of the necessary and sufficient condition that (6.3.8) be finite.

We can then obtain the representation (6.3.32) from the standard representation of a Poisson cluster p.g.fl.

$$\log G[h] = \int_{\mathcal{Y}} (G[h \mid y] - 1)\, \mu_c(\mathrm{d}y) \qquad \big(h \in \mathcal{V}(\mathcal{X})\big)$$

by expressing $G[h \mid y]$ in terms of the associated Janossy measures as in equation (5.5.3) and rearranging the integrations. Note that the term with $k = 0$ drops out of the summation. Uniqueness follows from standard results concerning uniqueness of the expression of the p.g.fl. and its logarithm about the origin.

Now suppose conversely that a family of measures K_k satisfying the stated conditions is given. We wish to construct at least one Poisson cluster process that has the p.g.fl. representation (6.3.32). Take $\mathcal{X} = \mathcal{Y}$, and let the measure $\mu_0(\cdot)$ be defined over bounded $A \in \mathcal{B}_{\mathcal{X}}$ by

$$\mu_0(A) = \sum_{k=1}^{\infty} K_k(A \times \mathcal{X}^{(k-1)})/k! \tag{6.3.34}$$

as the parameter measure for the cluster centre process. Note that the finiteness condition (6.3.33) entails the finiteness of (6.3.34) because

$$\begin{aligned} \sum_{i=1}^{k} \binom{k}{i} K_k\big(A^{(i)} \times (A^c)^{(k-i)}\big) &= \sum_{i=1}^{k} \frac{k}{i} \binom{k-1}{i-1} K_k\big(A \times A^{(i-1)} \times (A^c)^{k-i}\big) \\ &\geq K_k(A \times \mathcal{X}^{(k-1)}). \end{aligned}$$

As in the Gauss–Poisson case, we can define μ_0-a.e. a probability distribution $\{q_k(y)\}$ on $k = 1, 2, \ldots$ as the Radon–Nikodym derivatives in

$$\int_A q_k(y)\,\mu_0(\mathrm{d}y) = \frac{K_k(A \times \mathcal{X}^{(k-1)})}{k!}\,,$$

these probabilities $\{q_k(y)\}$ determining the number of points k in a cluster with centre y. The cluster member structure can be defined by taking one point as the cluster centre and locating the positions of the others relative to it through the distribution $P_{k-1}(B \mid y)$ defined μ_0-a.e. over $B \in \mathcal{B}(\mathcal{X}^{(k-1)})$ by

$$\int_A P_{k-1}(B \mid y)\, K_k(\mathrm{d}y \times \mathcal{X}^{(k-1)}) = K_k(A \times B),$$

appealing again to the existence of regular conditional probabilities. We can now check that the process with these components has the p.g.fl. representation (6.3.32) and that the existence condition (6.3.33) is satisfied. □

Note that there are many other processes that could be constructed from the same ingredients. In particular (see below Theorem 2.2.II), we can introduce an arbitrary probability $\tilde{q}_0(y)$ of empty clusters with $0 \le \tilde{q}_0(y) < 1$ (all y) by redefining

$$\tilde{q}_k(y) = \big(1 - \tilde{q}_0(y)\big) q_k(y) \qquad (k = 1, 2, \ldots)$$

and setting

$$\tilde{\mu}_c(\mathrm{d}y) = \big(1 - \tilde{q}_0(y)\big)^{-1} \mu_c(\mathrm{d}y).$$

The p.g.fl. is unaltered by this transformation, and the resultant processes are equivalent; we record this formally.

Corollary 6.3.VI. *The probability of a zero cluster is not an estimable parameter in any Poisson cluster model.*

A similar range of possibilities exists for the way the cluster centre x is defined relative to the joint distributions $P_k(\cdot)$ of the points in the cluster. In the construction above, we have chosen to fix the centre at an arbitrary point of the cluster. The measures $J_k(\cdot \mid y)$ are then related to the $P_k(\cdot \mid y)$ by $J_1(A) = P_1(A)$ and, for $k \ge 2$, the symmetrization relations

$$J_k(A_1 \times A_2 \times \cdots \times A_k \mid y) = k^{-1} \sum_{\mathrm{sym}} \delta_y(A_1) P_{k-1}(A_2 \times \cdots \times A_k \mid y).$$

Alternatively, we might prefer to locate the cluster centre at the multivariate centre of mass of the distribution (assuming this to be defined) or else in some other manner. This can be done without altering the final form of the p.g.fl. If it is necessary to select one particular form of representation for the process, we shall choose that used in the proof above and refer to it as the *regular representation* of the given process. The proposition implies that there is a one-to-one correspondence between measures on $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ induced by Poisson cluster processes and the elements in their regular representations.

Exercises and Complements to Section 6.3

6.3.1 *LeCam's precipitation process.* Formulate a definition for a general cluster random measure ζ analogous to Definition 6.3.I by replacing $\{N(\cdot \mid y)\}$ by a measurable family of random measures $\{\xi(\cdot \mid y)\}$. When these components are independent and $L_\xi[f \mid y]$ denotes the Laplace functional of $\xi(\cdot \mid y)$ defined over $f \in \mathrm{BM}_+(\mathcal{X})$ [see around (6.1.8)], the Laplace functional L_ζ of ζ is related to $\{L_\xi[f \mid y]\}$ and the p.g.fl. G_c of the cluster centre process by

$$L_\zeta = G_c[L_\xi[f \mid \cdot\,]]$$

provided ζ is well defined. [This model is discussed in LeCam (1961), who was motivated by the problem of modelling precipitation.]

6.3.2 Show that an independent cluster process exists if and only if, for each $h \in \mathcal{V}(\mathcal{X})$, the infinite product $G[h \mid N_c] = \prod_i G_m[h \mid y_i]$ converges Π_c-a.s.

6.3.3 Frequently, it may be desired specifically to include the cluster centres with the points generated by the cluster member processes with p.g.fl. $G_m[h \mid y]$. Show that the modified process has p.g.fl. $G_c[h(\cdot)G_m[h \mid \cdot\,]]$.

6.3.4 *Moment measures for a cluster process.* For a cluster process, the r.v. $X_f \equiv \int_{\mathcal{X}} f(y)\, N(\mathrm{d}y)$ can be expressed as the sum $\sum_i Y_f(y_i)$, where the y_i are the cluster centres and $Y_f(y) = \int_{\mathcal{X}} f(x)\, N_m(\mathrm{d}x \mid y)$ is the potential contribution to X_f from a cluster member with centre at y. Assume that for $f \in \mathrm{BM}_+(\mathcal{X})$

$$M_{1,f}(y) \equiv \mathrm{E}[Y_f(y)] = \int_{\mathcal{X}} f(x)\, M_1(\mathrm{d}x \mid y) < \infty,$$
$$M_{2,f}(y) \equiv \mathrm{E}[Y_f^2(y)] = \int_{\mathcal{X}^{(2)}} f(x_1) f(x_2)\, M_2(\mathrm{d}x_1 \times \mathrm{d}x_2 \mid y) < \infty.$$

Use a conditioning argument to obtain the basic relations

$$\begin{aligned}
\mathrm{E}X_f &= \int_{\mathcal{Y}} \mathrm{E}[Y_f(y)]\, M^c(\mathrm{d}y) = \int_{\mathcal{Y}} M_{1,f}(y)\, M^c(\mathrm{d}y) \\
&= \int_{\mathcal{Y}} \int_{\mathcal{X}} f(x)\, M_1(\mathrm{d}x \mid y)\, M^c(\mathrm{d}y), \\
\mathrm{E}X_f^2 &= \int_{\mathcal{Y}} V_2(y)\, M^c(\mathrm{d}y) + \int_{\mathcal{Y}^{(2)}} M_{1,f}(y)\, M_{1,f}(z)\, M_2^c(\mathrm{d}y \times \mathrm{d}z), \\
\operatorname{var} X_f &= \int_{\mathcal{Y}} V_2(y)\, M^c(\mathrm{d}y) + \int_{\mathcal{Y}^{(2)}} M_{1,f}(y) M_{1,f}(z)\, C_2^c(\mathrm{d}y \times \mathrm{d}z),
\end{aligned}$$

where $V_2(y) = M_{2,f}(y) - (M_{1,f}(y))^2 = \operatorname{var} Y_f(y)$. Derive equations (6.3.3–5) by considering also $\operatorname{cov}(X_f, X_g)$ and setting $f(\cdot) = I_A(\cdot)$, $g(\cdot) = I_B(\cdot)$. [*Hint*: Take care in passing from ordinary to factorial moments.]

6.3.5 (a) Show that a sufficient condition for the existence of a stationary cluster process is that the mean cluster size be finite.

(b) Show by counterexample that the condition is not necessary, even for a Poisson cluster process.

[*Hint*: For part (a), show first that in the stationary case,

$$M_1(A) = \mu_c \int_{\mathcal{X}} M_1(A \mid x)\,\mathrm{d}x = \mu_c \int_{\mathcal{X}} M_1(A - x \mid 0)\,\mathrm{d}x = m\ell(A),$$

and then observe that $p(A \mid x) \le M_1(A \mid x)$. For part (b), consider a compound Poisson process with infinite mean batch size.]

6.3.6 (a) Show that a stationary Poisson cluster process is simple if and only if each cluster member process is simple.

(b) When this condition is satisfied, show that the d.f. F corresponding to an interval between successive points of the process has coefficient of variation ≥ 1. [*Hint*: Show that $R(t) \equiv -\log S(t)$ in (6.3.8) is subadditive in $t > 0$ and hence that $S(t) \ge \exp(-R'(0+)t)$. Use Korolyuk's theorem to identify $1/R'(0+)$ as the first moment of F, and use a hazard function argument (see Exercise 3.4.2) to identify the second moment of F with $(2/R'(0+)) \int_0^\infty S(t)\,\mathrm{d}t$. Exercise 6.3.9(b) below gives a special case.]

6.3.7 For a Neyman–Scott Poisson cluster process as around (6.3.20) with $\mathcal{Y} = \mathcal{X} = \mathbb{R}$, suppose $F(x)$ has an exponential distribution. Use (6.3.20) to show (see Vere-Jones, 1970) that the hazard function below (6.3.15) for the distance from the origin to the nearest point of the process is given by

$$r(t) = \frac{\mu_c(1 - Q(\mathrm{e}^{-\lambda t}))}{1 - \mathrm{e}^{-\lambda t}}\,.$$

6.3.8 Consider a Neyman–Scott cluster process with cluster centres y_i the points of a Poisson process at rate μ_c and for each such point a Poisson-distributed random number n_i of points, with mean Y_i for an i.i.d. sequence of r.v.s $\{Y_i\}$, are located at $\{y_i + x_{ij} : j = 1, \dots, n_i\}$, where the x_{ij} are i.i.d. with probability density $g(\cdot)$. Show that such a process $\{y_i + x_{ij} : i = 1, \dots, n_i, \text{ all } i\}$ is identical with the shot-noise process of Example 6.2(a).

6.3.9 (a) Evaluate the first-moment measure of the interval $(0, t]$ for a cluster with centre y in a Bartlett–Lewis process as

$$M_c((0,t] \mid y) = \begin{cases} 0 & y > t, \\ 1 + \sum_{i=1}^\infty r_i F^{i*}(t - y) & 0 < y \le t, \\ \sum_{i=1}^\infty r_i [F^{i*}(t + |y|) - F^{i*}(|y|)] & y \le 0. \end{cases}$$

(b) Show that the hazard function for the interval distribution in the process corresponding to (6.3.24) is

$$r(t) = \mu_c + \mu_c m_{[1]}(1 - F(t)) - \frac{m_{[1]} f(t)}{1 + m_{[1]}(1 - F(t))}\,,$$

where $f(t)$ is the density corresponding to $F(t)$. Now verify Exercise 6.3.6(b): the interval distribution has coefficient of variation ≥ 1 (Lewis, 1964a).

6.3.10 Suppose the common d.f. in a Neyman–Scott type process in $\mathbb{R}^2$ is circular normal with density $f(x, y) = (2\pi)^{-1} \exp[-\frac{1}{2}(x^2 + y^2)]$. Show that the probability that a particular point of a given cluster falls in the circle of radius r and centre at the origin, when the cluster centre is at a distance ρ from the

origin, equals

$$P(r \mid \rho) \equiv \mathrm{e}^{-\rho^2/2} \int_0^t u\mathrm{e}^{-u^2/2} I_0(u\rho)\,\mathrm{d}u,$$

where I_0 is the modified Bessel function of zero order. Then the log survivor function of the distance from the origin to the nearest point of such a Neyman–Scott Poisson cluster process, with cluster p.g.f. $Q(z)$, is given by

$$-\log S(r) = 2\pi\mu_c \int_0^\infty [1 - Q(1 - P(r \mid \rho))]\,\rho\,\mathrm{d}\rho.$$

In particular, if the number in each cluster has a Poisson distribution with mean λ,

$$-\log S(r) = 2\pi\mu_c \int_0^\infty (1 - \mathrm{e}^{-\lambda P(r|\rho)})\,\rho\,\mathrm{d}\rho.$$

6.3.11 Show that $P(z) = \exp\{q_1(z-1) + q_2(z^2-1)\}$ is a univariate p.g.f. if and only if $q_1 \geq 0$, $q_2 \geq 0$, and hence complete the proof of Proposition 6.3.IV. [*Hint*: To be a p.g.f., $P(z)$ must have nonnegative coefficients as a power series in z, while by virtue of its representation, $P(z)$ is an entire function. Hence, show that $\log P(z)$ must be well defined and nondecreasing on the whole positive half-line $z > 0$, and deduce that both q_1 and $q_2 \geq 0$.]

6.3.12 Show that a point process N is Gauss–Poisson if and only if the first two Khinchin measures are nonnegative with boundedly finite marginals and all remaining Khinchin measures vanish. [This is a rephrasing of Proposition 6.3.IV and Examples 6.2(c) and 6.3(d).]

6.3.13 Show that the functional of (possibly signed) measures $\overline{Q}_1(\cdot)$ and $\overline{Q}_2(\cdot \times \cdot)$

$$\int_{\mathcal{X}} [h(x) - 1]\,\overline{Q}_1(\mathrm{d}x) + \tfrac{1}{2} \int_{\mathcal{X}^{(2)}} [h(x) - 1]\,[h(y) - 1]\,\overline{Q}_2(\mathrm{d}x \times \mathrm{d}y)$$

equals the logarithm of the p.g.fl. of some point process if and only if $\overline{Q}_1$ is nonnegative and the symmetrized version

$$\overline{Q}_2^s(A \times B) = \tfrac{1}{2}(\overline{Q}_2(A \times B) + \overline{Q}_2(B \times A))$$

is nonnegative and bounded as in $\overline{Q}_2^s(A \times B) \leq \min(\overline{Q}_1(A), \overline{Q}_1(B))$ for bounded A, $B \in \mathcal{B}_{\mathcal{X}}$. [*Hint*: Reduce the functional above to the form of (6.3.30) and appeal to Proposition 6.3.IV. See also Example 6.2(d).]

6.3.14 Proposition 6.3.V represents a Poisson cluster process with a.s. finite clusters. Realize a cluster of size k and choose one of its points, Y say, at random. Show that

$$\Pr\{Y \in A\} = \frac{K_k(A \times \mathcal{Y}^{(k-1)})}{K_k(\mathcal{Y}^{(k)})},$$

but

$$\Pr\left\{\begin{array}{l}\text{a cluster realization of}\\ \text{size } k \text{ has a point in } A\end{array}\right\} = \sum_{i=1}^{k} \binom{k}{i} \frac{K_k(A^{(i)} \times (A^c)^{(k-i)})}{K_k(\mathcal{Y}^{(k)})}.$$

6.3.15 The factorial cumulant measures $C_{[k]}$ of a Gauss–Poisson process vanish for $k = 3, 4, \ldots$. Show in general that for a Poisson cluster process with clusters of size not exceeding k_0, $C_{[k]}$ vanishes for $k > k_0$. [*Hint*: Use (6.3.32) and write $1 + h$ for h.]

6.4. Marked Point Processes

In many stochastic process models, a point process arises not as the primary object of study but as a component of a more complex model; often, the point process is the component that carries the information about the locations in time or space of objects that may themselves have a stochastic structure and stochastic dependency relations. From the point of view of point process theory, many such models can be subsumed under the heading of *marked point processes*. In this section, we provide an initial study of such processes, particularly those with links to the Cox and cluster processes described in the two preceding sections.

For any marked point process, the locations $\{x_i\}$ where the events occur constitute an important process in their own right (the x_i may denote times but could also be two- or three-dimensional, for example). We shall refer to this process as the *ground process* and accordingly denote it by N_{g}.

Definitions 6.4.I.

(a) *A* marked point process (MPP), *with locations in the c.s.m.s.* $\mathcal{X}$ *and marks in the c.s.m.s.* $\mathcal{K}$, *is a point process* $\{(x_i, \kappa_i)\}$ *on* $\mathcal{X} \times \mathcal{K}$ *with the additional property that the ground process* $N_{\mathrm{g}}(\cdot)$ *is itself a point process; i.e. for bounded* $A \in \mathcal{B}_{\mathcal{X}}$, $N_{\mathrm{g}}(A) = N(A \times \mathcal{K}) < \infty$.

(b) *A* multivariate (*or* multitype) point process *is a marked point process with mark space the finite set* $\{1, \ldots, m\}$ *for some finite integer* m.

If a marked point process N is regarded as a process on the product space $\mathcal{X} \times \mathcal{K}$, then the ground process N_{g} is the *marginal process of locations*. However, it is a consequence of Definition 6.4.I(a) that not all point processes on product spaces are marked point processes. For example, the bivariate Poisson process on $\mathbb{R}^2$ with parameter measure $\mu\,\mathrm{d}x\,\mathrm{d}y$ cannot be represented as an MPP on $\mathbb{R} \times \mathbb{R}$ because such a Poisson process has $N(A \times \mathbb{R}) = \infty$ a.s. for Borel sets A of positive Lebesgue measure. However, in the special case of a multivariate point process, the extra condition is redundant since the finiteness of the mark space immediately implies that each *component process* $N_i(\cdot) = N(\cdot \times \{i\})$ is boundedly finite and we can write

$$N_{\mathrm{g}}(\cdot) = N(\cdot \times \{1, \ldots, m\}) = \sum_{i=1}^{m} N_i(\cdot). \tag{6.4.1}$$

In general, an MPP can be regarded either as a point process in the product space $\mathcal{X} \times \mathcal{K}$ subject to the finiteness constraint on the ground process N_{g} as set out above, or as an ordinary (not necessarily simple) point process in $\mathcal{X}$, $\{x_i\}$ say, with an associated sequence of random variables $\{\kappa_i\}$ taking their values in $\mathcal{K}$. Either approach leads to the representation of the MPP as a set of pairs $\{(x_i, \kappa_i)\}$ in the product space. They are equivalent whenever it can be shown that the marks κ_i in an MPP are well-defined random variables, which is certainly the case when the ground process has finite intensity, but there are subtleties in general: see Section 8.3 and Chapter 9 for further discussion.

The class of MPPs is a great deal richer than might at first appear. This is due to the great variety of forms that can be taken by the marks and the variety of dependence relations that can exist between the marks themselves and their locations. When $\mathcal{X} = \mathbb{R}$, for example, many remarkable results can be obtained by taking the mark at an event x_i to represent some feature from the history of the process up to x_i. A careful study of such MPPs lies at the heart of the fundamental researches of Matthes, Mecke, and co-workers.

Extending the concepts of earlier chapters, we define for MPPs the following two classes of point processes.

Definition 6.4.II.
(a) *The MPP N is* simple *if the ground process N_g is simple.*
(b) *The MPP N on $\mathcal{X} = \mathbb{R}^d$ is* stationary (homogeneous) *if the probability structure of the process is invariant under shifts in $\mathcal{X}$.*

The structure of an MPP may be spelled out in a variety of ways. If the ground process N_g is not necessarily simple, it can be thought of as a cluster process in which the cluster centres x_i are the distinct locations in $\mathcal{X}$ and the cluster members are all pairs in $\mathcal{X} \times \mathcal{K}$ of the form (x_i, κ_{ij}), where the κ_{ij} are the marks of the points with common location x_i. Equally, however, the family κ_{ij} could be thought of as a single, compound mark in the space $\mathcal{K}^{\cup}$ defined as in (5.3.8). This last comment implies that *by suitably redefining the marks, any MPP on $\mathcal{X}$ can be represented as an MPP on $\mathcal{X}$ for which the ground process N_g is simple.* For many applications, though not for all, we may therefore assume that the MPPs we encounter are simple.

The next pair of definitions characterize two important types of independence relating to the mark structure of MPPs. Observe in part (b) that a crucial feature is the role of order in the location space: it reflects the evolutionary property that we associate with a time-like dimension.

Definition 6.4.III (Independent marks and unpredictable marks)**.** *Let the MPP $N = \{(x_i, \kappa_i)\}$ on $\mathcal{X} \times \mathcal{K}$ be given.*
(a) *N has* independent marks *if, given the ground process $N_g = \{x_i\}$, the $\{\kappa_i\}$ are mutually independent random variables such that the distribution of κ_i depends only on the corresponding location x_i.*
(b) *For $\mathcal{X} = \mathbb{R}$, N has* unpredictable marks *if the distribution of the mark at x_i is independent of locations and marks $\{(x_j, \kappa_j)\}$ for which $x_j < x_i$.*

The most common case of an MPP with independent marks occurs when the κ_i are in fact i.i.d. Similarly, the most common case of a process with unpredictable marks occurs when the marks are conditionally i.i.d. given the past of the process (but the marks may influence the future of N_g).

The next proposition outlines the basic structure of processes with independent marks, introducing in particular the *mark kernel* $F(\cdot \mid \cdot)$ at a specified location. P.g.fl.s for MPPs are defined over the space $\mathcal{V}(\mathcal{X} \times \mathcal{K})$ of measurable functions $h(x, \kappa)$ that lie between 0 and 1 and for some bounded set A, $h(x, \kappa) = 1$ for all $\kappa \in \mathcal{K}$ and $x \notin A$.

Proposition 6.4.IV (Structure of MPP with independent marks)**.** *Let N be an MPP with independent marks.*

(a) *The probability structure of N is completely defined by the distribution of the ground process N_{g} and the* mark kernel $\{F(K \mid x)\colon K \in \mathcal{B}(\mathcal{K}),\ x \in \mathcal{X}\}$, *representing the conditional distribution of the mark, given the location x.*

(b) *The p.g.fl. for N takes the form*

$$G[h] = G_{\mathrm{g}}[h_F] \qquad (h \in \mathcal{V}(\mathcal{X} \times \mathcal{K})), \tag{6.4.2}$$

where G_{g} is the p.g.fl. of N_{g} and $h_F(x) = \int_{\mathcal{K}} h(x,\kappa)\, F(\mathrm{d}\kappa \mid x)$.

(c) *The moment measure M_k of order k for N exists if and only if the corresponding moment measure M_k^{g} exists for the ground process N_{g}, in which case*

$$\begin{aligned} &M_k(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k \times \mathrm{d}\kappa_1 \times \cdots \times \mathrm{d}\kappa_k) \\ &\qquad = M_k^{\mathrm{g}}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k) \prod_{i=1}^{k} F(\mathrm{d}\kappa_i \mid x_i). \end{aligned} \tag{6.4.3}$$

Similar representations hold for factorial and cumulant measures.

Proof. All the statements above are corollaries of the general results for conditional point processes outlined in Section 6.1. In the present case, we deduce statements for the process of pairs $\{(x_i, \kappa_i)\}$ from their distribution conditional on the process of locations $\{x_i\}$ using the conditional independence of the κ_i.

Because of the independence properties, it is easiest to approach the statements via the p.g.fl. Given the locations x_i, the p.g.fl. of the pairs (x_i, κ_i) takes the form

$$G[h(x,\kappa) \mid N_{\mathrm{g}}] = \prod_i \left[\int_{\mathcal{K}} h(x_i,\kappa)\, F(\mathrm{d}\kappa \mid x_i) \right] = \prod_i h_F(x_i). \tag{6.4.4}$$

Note that $h_F \in \mathcal{V}(\mathcal{X})$ when $h \in \mathcal{V}(\mathcal{X} \times \mathcal{K})$ because for some bounded set A, $h(x,\kappa) = 1$ for $x \notin A$ and all $\kappa \in \mathcal{K}$, and hence for such x, $h_F(x) = \int_{\mathcal{K}} h(x,\kappa)\, F(\mathrm{d}\kappa \mid x) = 1$. Provided then that N_{g} exists, the final product is well defined for $h \in \mathcal{V}(\mathcal{X} \times \mathcal{K})$ and defines a measurable function of N_{g}. We thus have a measurable family satisfying Lemma 6.1.III(b); taking expectations over the locations, we obtain (6.4.2). Since the p.g.fl. is well defined, so are the fidi distributions and hence the probability structure of the process.

To justify the expressions for the moment measures, consider an integral of the form $\int h(x_1, \ldots, x_k, \kappa_1, \ldots, \kappa_k)\, N(\mathrm{d}x_1 \times \mathrm{d}\kappa_1) \cdots N(\mathrm{d}x_k \times \mathrm{d}\kappa_k)$. Conditional on the locations $\{x_i\}$, its expectation can be written

$$\int_{\mathcal{K}} \cdots \int_{\mathcal{K}} h(x_1, \ldots, x_k, \kappa_1, \ldots, \kappa_k)\, F(\mathrm{d}\kappa_1 \mid x_1) \cdots F(\mathrm{d}\kappa_k \mid x_k). \tag{6.4.5}$$

Now taking expectations over the locations, assuming the moment measure to exist for the ground process, we obtain (6.4.3), finite or infinite according to whether the integrals converge. But convergence of the integrals for all appropriate h is the necessary and sufficient condition for the existence of the moment measures, so statement (c) follows. □

In many applications, $\mathcal{K} = \mathbb{R}_+$ and interest centres on the random measure defined by

$$\xi(A) = \int_{A\times\mathcal{K}} \kappa\, N(\mathrm{d}x \times \mathrm{d}\kappa) = \sum_{x_i \in A} \kappa_i\,. \tag{6.4.6}$$

Its properties when ξ has independent marks are summarized below. Observe that if $\kappa_i = \kappa$ a.s. for all i, then $\xi(A) = \kappa N_{\mathrm{g}}(A)$.

Proposition 6.4.V. *If $\mathcal{K} = \mathbb{R}_+$ and the MPP N has independent marks, ξ in (6.4.6) defines a purely atomic random measure on $\mathcal{X}$ with only finitely many atoms on any bounded set $A \in \mathcal{B}_{\mathcal{X}}$. It has Laplace functional*

$$L_\xi[h] = G_{\mathrm{g}}[\phi_h] \qquad (h \in \mathrm{BM}_+(\mathcal{X})), \tag{6.4.7}$$

where $\phi_h(x) = \int_{\mathcal{K}} \mathrm{e}^{-\kappa h(x)}\, F(\mathrm{d}\kappa \mid x)$ and G_{g} is as in (6.4.2).

The moment measure M_k^ξ of order k for ξ exists if

(i) *the moment measure M_k^{g} of order k exists for the ground process N_{g},*

(ii) *the kth moment of the mark distribution, $\mu_k(x) = \int_{\mathbb{R}_+} \kappa^k\, F(\mathrm{d}\kappa \mid x)$ exists M_1^{g}-a.e., and*

(iii) *the integrals defining M_k^ξ in terms of μ_r and M_s^{g} for $r, s = 1, \dots, k$, converge.*

When they exist, the first- and second-moment measures are given, for bounded $A, B \in \mathcal{B}_{\mathcal{X}}$, by

$$M_1^\xi(B) = \int_B \mu_1(x)\, M_1^{\mathrm{g}}(\mathrm{d}x)\,, \tag{6.4.8}$$

$$M_2^\xi(A\times B) = \int_{A\times B} \mu_1(x_1)\mu_1(x_2)\, M_{[2]}^{\mathrm{g}}(\mathrm{d}x_1 \times \mathrm{d}x_2) + \int_{A\cap B} \mu_2(x)\, M_1^{\mathrm{g}}(\mathrm{d}x)\,. \tag{6.4.9}$$

Proof. The statements follow from reasoning similar to that used in Proposition 6.4.IV. The integral in (6.4.6) is a.s. finite when A is bounded (since the sum is then over an a.s. finite number of terms) and is easily seen to have the additivity properties required of a random measure. Its Laplace functional and moment measures can again be found by first conditioning on the locations. Thus, $L_\xi(h \mid N_{\mathrm{g}})$ equals

$$\mathrm{E}\left[\exp\left(-\int_{\mathbb{R}_+} h(x)\,\xi(\mathrm{d}x)\right) \,\middle|\, N_{\mathrm{g}}\right] = \prod_i \left[\int_{\mathbb{R}_+} \mathrm{e}^{-\kappa h(x_i)}\, F(\mathrm{d}\kappa \mid x_i)\right].$$

Equation (6.4.7) follows on taking expectations over the locations. Note that when $h \in \mathrm{BM}_+(\mathcal{X})$, the Laplace–Stieltjes transform $\phi_h \in \mathcal{V}(\mathcal{X})$, as is required for a p.g.fl. Equation (6.4.8) is derived similarly.

To obtain (6.4.9), we have to condition on the location of pairs (x_i, x_j) defined by the product counting measure $N_{\mathrm{g}} \times N_{\mathrm{g}}$. Note the special attention given to the diagonal pairs (x_i, x_i): $M_2^{\xi}(A \times B)$ equals

$$\mathrm{E}\Bigg[\int_A \int_B \Bigg(\int_{\mathcal{K}} \int_{\mathcal{K}} \kappa_1\kappa_2 \, F(\mathrm{d}\kappa_1 \mid x_1)\, F(\mathrm{d}\kappa_2 \mid x_2)\Bigg)\, N_{\mathrm{g}}(\mathrm{d}x_1)\, N_{\mathrm{g}}(\mathrm{d}x_2) + \int_{A\cap B} \Bigg(\int_{\mathcal{K}} \kappa^2 \, F(\mathrm{d}\kappa \mid x)\Bigg)\, N_{\mathrm{g}}(\mathrm{d}x)\Bigg]$$
$$= \int_{A\times B} \mu_1(x_1)\mu_1(x_2)\, M^{\mathrm{g}}_{[2]}(\mathrm{d}x_1 \times \mathrm{d}x_2) + \int_{A\cap B} \mu_2(x)\, M^{\mathrm{g}}_1(\mathrm{d}x)\,.$$

These expressions can be checked by expanding the functionals and transforms concerned (see Exercise 6.4.1 for the case $k = 3$). □

As for cluster processes, the results simplify if the process is stationary, and the relevant factorial moment densities exist. Stationarity implies that the mark kernel is independent of x, $F(\cdot \mid x) = F(\cdot)$ say, so that ϕ_h in (6.4.7) becomes $\phi_h(x) = \int_{\mathcal{K}} \mathrm{e}^{-\kappa h(x)}\, F(\mathrm{d}\kappa)$, the usual Laplace–Stieltjes transform of the distribution F evaluated at $h(x) \in \mathrm{BM}_+(\mathcal{X})$. Given the existence of the reduced densities $\breve{m}^{\mathrm{g}}_{[2]}(\cdot)$ and $\breve{c}^{\mathrm{g}}_{[2]}(\cdot)$, and writing $\mu_k = \int_{\mathcal{K}} \kappa^k\, F(\mathrm{d}\kappa)$, (6.4.8) and (6.4.9) lead to

$$m = \mu_1 m_{\mathrm{g}}\,, \tag{6.4.10}$$
$$\breve{m}_2(u) = (\mu_1)^2 \breve{m}^{\mathrm{g}}_{[2]}(u) + \delta(u)\mu_2 m_{\mathrm{g}}\,, \tag{6.4.11a}$$
$$\breve{c}_2(u) = (\mu_1)^2 \breve{c}^{\mathrm{g}}_{[2]}(u) + \delta(u)\mu_2 m_{\mathrm{g}}\,. \tag{6.4.11b}$$

The appearance of the δ-function in (6.4.11) is a reminder that the ξ process, as well as the process N_{g}, is purely atomic and therefore has a diagonal concentration (see Section 8.1 below). Equation (6.4.11b) leads to the well-known expression for the variance of a random sum of i.i.d. r.v.s,

$$\operatorname{var} \xi(A) = [\mathrm{E}(\kappa)]^2 \operatorname{var} N_{\mathrm{g}}(A) + \mathrm{E}[N_{\mathrm{g}}(A)] \operatorname{var} \kappa. \tag{6.4.12}$$

Extension of the discussion above to the mark space $\mathcal{K} = \mathbb{R}$ is possible but leads to signed measures and requires the use of characteristic functionals in place of Laplace functionals; see Exercise 6.4.2.

An important special case arises when the ground process N_{g} is Poisson. We call such a process a *compound Poisson process*. As such, it extends the compound Poisson process introduced in Section 2.2, where $\mathcal{K} = \mathbb{Z}_+$. For this (generalized) compound Poisson process, the marks often represent a weight associated with the point, such as a monetary value in financial applications, an energy or seismic moment in seismology, a weight or volume in forestry or

geological prospecting, and so on. In such cases, ξ measures the total value, energy, weight, volume, etc., accumulating within a certain time interval or region. We give some examples shortly but first present a simple, important structural property that foreshadows results for more general classes of MPPs.

Lemma 6.4.VI. *A compound Poisson process that has mark kernel $F(\cdot \mid \cdot)$, and for which the Poisson ground process N_g has intensity measure $\mu(\cdot)$, is equivalent to a Poisson process on the product space $\mathcal{X} \times \mathcal{K}$ with intensity measure $\Lambda(\mathrm{d}x \times \mathrm{d}\kappa) = \mu(\mathrm{d}x)\, F(\mathrm{d}\kappa \mid x)$.*

Proof. We examine the p.g.fl.s. Substituting in (6.4.2) for the p.g.fl. of the Poisson process for N_g and rearranging, we have, using notation from (6.4.2),

$$G[h] = \exp\left(\int [h_F(x) - 1]\, \mu(\mathrm{d}x)\right) = \exp\left(\int\int [h(x,\kappa) - 1]\, F(\mathrm{d}\kappa \mid x)\, \mu(\mathrm{d}x)\right),$$

where the last expression can be identified with the p.g.fl. of the Poisson process on the product space. □

Many classical stochastic models are rooted in the compound Poisson process. One famous example is as follows.

Example 6.4(a) *Lundberg's collective risk model* (Lundberg, 1903; Cramér, 1930). Suppose that claims W_i against an insurer are made at times t_i. Let $\xi(t)$ represent the accumulated claims $\sum_{i:0<t_i<t} W_i$ over the period $(0,t)$. If the distribution of a generic claim W is supposed constant (independent of t) and the claim times t_i follow a Poisson process with constant intensity μ, then the pairs $\{(t_i, W_i)\}$ form a compound Poisson process. Typically, in this context, the claim distribution is chosen from the gamma, Weibull, or Pareto families or various modifications of these chosen to fit the specific application. From equations (6.4.10) and (6.4.11), we obtain the elegant special forms

$$\begin{aligned} \mathrm{E}[\xi(t)] &= \mu \mathrm{E}(W)\, t, \\ \operatorname{var} \xi(t) &= \mu \mathrm{E}(W^2)\, t. \end{aligned}$$

The crucial simplification underlying this elegance arises from the location process being Poisson, for then the covariance density $c_{[2]}(\cdot)$ vanishes.

If the insurance company has initial capital U_0 and it is assumed that gross premium income comes in at a constant rate α, then the financial reserve of the company after time t, excluding running costs, depreciation, inflation, income from investment, and other external factors, is equal to $U(t) = U_0 + \alpha t - \xi(t)$. The *classical ruin problem* consists in determining whether, and if so when, $U(t)$ first becomes zero. If $\alpha \leq \mu \mathrm{E}(W)$, ruin is certain, but the time to ruin may still be of importance. If $\alpha > \mu \mathrm{E}(W)$, ruin may be avoided and interest centres around estimating the *probability of ruin*, say η. In both cases, important information may be derived from the observation that, if

$\tau_i = t_i - t_{i-1}$, then the random variables $Z_i = W_i - \alpha\tau_i$ are independent, so that the process

$$U_n - U_0 = \sum_{i=1}^{n} Z_i = \alpha t_n - \xi t_n$$

constitutes a random walk. In particular, this observation, coupled to a standard martingale argument, leads to the classical Cramér bound on the probability of ultimate ruin. The argument is outlined in Exercise 6.4.3 (or else, see e.g. Embrechts *et al.*, 1997, Section 1.1). □

EXAMPLE 6.4(b) *Negative binomial processes.* The negative binomial distribution is a common choice for the count random variables $N(A)$ in applications to processes $N(\cdot)$ where a clustering alternative is preferred to the Poisson process. It is somewhat surprising that the only known examples of processes yielding the negative binomial form for the distributions of $N(A)$ are both extreme cases: a compound Poisson process that has the complete independence property and in which all the clusters are concentrated at single points, and a mixed Poisson process in which the individual realizations are indistinguishable from those of a Poisson process. The usefulness of the negative binomial distribution in practice stems more from its relative simplicity and tractability than its link to organic physical models, although it will of course be true that for long time intervals, when the time scale of clustering is short relative to the time scale of observation, the compound Poisson model may be an adequate approximation. We describe these two models; see also Grégoire (1984) and the review article of Diggle and Milne (1983).

(i) *Compound Poisson process leading to negative binomial distributions.* Suppose there is given a compound Poisson process with constant intensity μ and discrete mark distribution that is independent of the location x. If $N(A)$ is to have a negative binomial distribution, then we know from Example 5.2(a) that the cluster size distribution should have the logarithmic form

$$\pi_n(x) = (\rho^n/n)\log[1/(1-\rho)].$$

Taking this as the mark distribution, we find that the p.g.fl. for the resulting random measure ξ, which in this case is again a point process but nonorderly, now has the form

$$G[h] = \exp\left(\int_{\mathcal{X}} \frac{\log\big([1-\rho h(x)]/(1-\rho)\big)}{\log(1-\rho)}\,\mu(\mathrm{d}x)\right) \qquad (h \in \mathcal{V}(\mathcal{X})).$$

This corresponds to the multivariate p.g.f. for the fidi distributions on disjoint sets $A_1, \ldots, A_k$,

$$P_k(A_1, \ldots, A_k;\, z_1, \ldots, z_k) = \prod_{i=1}^{k} \left[\frac{1-\rho}{1-\rho z_i}\right]^{-\mu(A_i)/\log(1-\rho)},$$

representing one simple type of multivariate negative binomial distribution. The factorial cumulant measures can be obtained from the expansion

$$\begin{aligned}\log G[1+\eta] &= \int_{\mathcal{X}} \frac{\log[1-\rho\eta(x)/(1-\rho)]}{\log(1-\rho)}\,\mu(\mathrm{d}x)\\ &= -\frac{1}{\log(1-\rho)}\sum_{k=1}^{\infty}\frac{1}{k}\left(\frac{\rho}{1-\rho}\right)^k\int_{\mathcal{X}}[\eta(x)]^k\,\mu(\mathrm{d}x)\end{aligned}$$

so that $C_{[k]}(\cdot)$ for $k \geq 2$ is a singular measure with a concentration $c_{[k]}\mu(\cdot)$ on the diagonal $x_1 = \cdots = x_k$, where $c_{[k]}$ is the kth factorial moment of the logarithmic distribution, or, equivalently, $c_{[k]}/\log[1/(1-\rho)]$ is the kth factorial cumulant of the negative binomial distribution.

Recall the p.g.f. of the negative binomial distribution in Example 5.2(a) and the p.g.fl. for a local process on a bounded Borel set A as in Example 5.5(b). The p.g.fl. for the type (i) negative binomial process applied to Example 5.5(b) gives us (since the integral over A^c vanishes)

$$G_A[1-I_A+h^*] = \exp\left(\frac{1}{\log(1-\rho)}\int_A \log\left(\frac{1-\rho h}{1-\rho}\right)\mu(\mathrm{d}x)\right),$$

where $h^*(x) = h(x)I_A(x)$. Thus, the localized process is still a negative binomial process. The local Janossy measures can be found from the expansion

$$\log\left(\frac{1-\rho h}{1-\rho}\right) = -\log(1-\rho) + \sum_{n=1}^{\infty}\frac{\rho^n}{n}\,h^{(n)},$$

from which we deduce that $p_0(A) = \exp[-\mu(A)]$ and

$$\begin{aligned}J_1(\mathrm{d}x \mid A) &= \rho p_0(A)\,\mu(\mathrm{d}x),\\ J_2(\mathrm{d}x_1 \times \mathrm{d}x_2 \mid A) &= \rho^2 p_0(A)[\mu(\mathrm{d}x_1)\mu(\mathrm{d}x_2) + \delta(x_1,x_2)\mu(\mathrm{d}x_1)],\end{aligned}$$

where the two terms in J_2 represent contributions from two single-point clusters at x_1 and x_2 ($x_1 \neq x_2$) and a two-point cluster at $x_1 = x_2$. □

(ii) *Mixed Poisson process leading to negative binomial distributions.* Take the mixing distribution Π with Laplace–Stieltjes transform Π^* as in (6.1.16), now generalized to the nonstationary case, to have the gamma distribution $\Gamma(\alpha,\lambda)$ with Laplace–Stieltjes transform $(1+s/\lambda)^{-\alpha}$. Then

$$G[h] = \Pi^*\left(\int_{\mathcal{X}}[1-h(x)]\,\mu(\mathrm{d}x)\right) = \left(1+\frac{1}{\lambda}\int_{\mathcal{X}}[1-h(x)]\,\mu(\mathrm{d}x)\right)^{-\alpha},$$

so that the multivariate p.g.f. has the form

$$P_k(A_1,\ldots,A_k;\,z_1,\ldots,z_k) = \left(1+\frac{1}{\lambda}\sum_{i=1}^{k}(1-z_i)\mu(A_i)\right)^{-\alpha}.$$

The factorial cumulants can be obtained from the expansion

$$\log G[1+\eta] = -\alpha \log\left(1 - \frac{1}{\lambda}\int_{\mathcal{X}} \eta(x)\,\mu(\mathrm{d}x)\right) = \alpha \sum_{k=1}^{\infty} \frac{1}{k}\left(\int_{\mathcal{X}} \frac{\eta(x)}{\lambda}\,\mu(\mathrm{d}x)\right)^k,$$

so

$$C_{[k]}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k) = \alpha\lambda^{-k}(k-1)!\,\mu(\mathrm{d}x_1)\cdots\mu(\mathrm{d}x_k),$$

where we can recognize the coefficient of the product measure on the right-hand side as the kth cumulant measure of the negative binomial distribution. Note that Example 5.2(a) corresponds to the case where the measure $\mu(\cdot)$ is totally finite, in which case $\mu(\mathcal{X})/\lambda$ here equals the parameter μ there. □

Most of the examples of point processes that we have considered in earlier sections can be adorned with marks in a way similar to the Poisson process in Examples 6.4(a) and (b) above. The choice of underlying model will depend on the context and anticipated dependence structure. The most interesting extensions appear when we drop the assumption of completely independent marks and consider ways in which either the marks can influence the future development of the process or the current state of the process can influence the distribution of marks, or both. Using the Hawkes process of Example 6.3(c) as below illustrates some of the many possible issues that can arise.

Example 6.4(c) *Marked Hawkes process.* Marked versions of the Hawkes process of Example 6.3(c) are best known from Hawkes (1971b, 1972), who considered the multivariate case in detail, with an application in Hawkes and Adamopoulos (1973), though Kerstan (1964) considered them at length. We consider here the case of unpredictable marks; for a more general multivariate extension, see Example 8.3(c). Both extensions have important applications in seismology [see also Example 6.4(d) below], epidemiology, neurophysiology, and teletraffic (see e.g. Brémaud and Massoulié, 1996).

In extending the Hawkes process of Example 6.3(c) to an MPP $\{(x_i, \kappa_i)\}$, we interpret the marks κ_i as the 'type' of an individual in a multitype branching process. Recall that, in the branching process interpretation, points in a Hawkes process are either 'immigrants' without parents or 'offspring' of another point in the process. This (multitype) model now incorporates the following assumptions:

(i) immigrants arrive according to a compound Poisson process $N(\mathrm{d}y \times \mathrm{d}\kappa)$ with constant rate μ_c and fixed mark distribution $F(\mathrm{d}\kappa)$;

(ii) each individual in the process, whether an immigrant or not, has the potential to act as an ancestor and thereby yield first-generation offspring according to an ordinary Poisson process with intensity measure $\mu(\mathrm{d}u \mid \kappa) = \psi(\kappa)\,\mu(\mathrm{d}u)$ that depends only on the mark κ of the ancestor event and the distance u of the offspring from the ancestor; and

(iii) the marks of the offspring form an i.i.d. sequence with the same d.f. F as the immigrants.

The factor $\psi(\kappa)$ determines the relative average sizes of families with different marks, while the measure $\mu(\cdot)$ determines how the family members are spread out along the time axis. For a stable process, $\mu(\mathcal{X})$ must be finite, and for the sake of definiteness, we assume that $\mu(\mathcal{X}) = 1$ so that $\psi(\kappa)$ becomes the expected number of direct offspring with mark κ.

In principle, the analysis of such a process requires the general theory of multiple type branching processes with a continuous range of types. However, the assumption of i.i.d. marks (i.e. offspring types) greatly simplifies the analysis. Indeed, the assumptions above imply that the ground process N_g for this marked point process can be described as an ordinary Hawkes process with immigration rate μ_c and infectivity measure

$$\mu_g(\mathrm{d}u) = \rho\,\mu(\mathrm{d}u), \qquad \text{where } \rho = \mathrm{E}[\psi(\kappa)] = \int_{\mathcal{K}} \psi(\kappa)\,F(\mathrm{d}\kappa) < \infty.$$

If then $\rho < 1$, the total number of progeny is a.s. finite with finite mean $1/(1-\rho)$ so that the ground process is well defined and has a stationary version (see Exercise 6.3.5). Since the overall process may itself be regarded as a Poisson cluster process taking its values in $\mathcal{X} \times \mathcal{K}$, a second application of Exercise 6.3.5 implies that the overall process has a well-defined stationary version. We state this formally for reference.

Proposition 6.4.VII. *Using the notation above, sufficient conditions for the existence of a stationary version of the marked Hawkes process with unpredictable marks are*

(i) *the intensity measure $\mu(\cdot)$ is totally finite (and then taken to be a probability measure); and*

(ii) $\rho = \mathrm{E}[\psi(\kappa)] < 1$.

First- and second-order properties of the process can be obtained by combining results for branching processes with results for cluster processes and are given in Chapter 8. The p.g.fl. is difficult to obtain explicitly; one approach is suggested in Exercise 6.4.4.

Many variations and extensions of this model are possible. Example 7.3(b) will show that the conditional intensity for this process has a very simple and powerful linear form, which lends itself to various types of generalization. The mark can be expanded to include a spatial as well as a size component, as for the spatial ETAS model described below. The assumption of unpredictable marks can also be weakened in several ways, for example by allowing the distributions of the marks of the offspring to depend on either the mark of the ancestor or the offspring's distance from the ancestor, or both. See Example 8.3(e) for a somewhat simpler model illustrating such dependence.

If the branching structure is critical rather than subcritical (i.e. $\rho = 1$), further types of behaviour can occur. For example, if the infectivity function is sufficiently long-tailed, Brémaud and Massoulié (2001) provides examples of stationary Hawkes processes *without* immigration (i.e. of a Hawkes process

whose clusters overlap at such large distances that the process maintains a stationary regime). Further details are given in Chapter 10. □

EXAMPLE 6.4(d) *Ordinary and spatial ETAS models.* Ogata (1988) introduced the ETAS (Epidemic Type After-Shock) model to describe earthquake occurrence, following earlier applications of the Hawkes model to this context by Hawkes and Adamopoulos (1973) and Vere-Jones and Ozaki (1982). It corresponds to the special case of the marked Hawkes process where $\mathcal{X} = \mathcal{K} = \mathbb{R}$, the x_i are interpreted as the occurrence times of the earthquakes and the κ_i as their magnitudes, and the following specific choices are made:

$$\begin{aligned}\psi(\kappa) &= A\mathrm{e}^{\alpha(\kappa-\kappa_0)} I_{\{\kappa>\kappa_0\}}(\kappa),\\ \mu(\mathrm{d}u) &= \frac{K}{(c+u)^{1+p}}\, I_{\{u>0\}}(u)\,\mathrm{d}u,\\ F(\mathrm{d}\kappa) &= \beta\mathrm{e}^{-\beta(\kappa-\kappa_0)} I_{\{\kappa>\kappa_0\}}(\kappa)\,\mathrm{d}\kappa.\end{aligned}$$

These choices are dictated largely by seismological considerations: thus, the mark distribution cited above corresponds to the Gutenberg–Richter frequency–magnitude law, while the power-law form for μ follows the empirical *Omori Law* for aftershock sequences. The free parameters are β, α, c, A and p. $K = p\,c^p$ is a normalizing constant chosen to ensure $\int_0^\infty \mu(\mathrm{d}u) = 1$.

In this case, sufficient conditions for a stationary process are that

$$p > 0, \qquad \beta > \alpha, \qquad \text{and} \qquad \rho = A\beta/(\beta-\alpha) < 1.$$

The last condition in particular is physically somewhat unrealistic since it is well known that the frequency–magnitude distribution cannot retain the pure exponential form indefinitely, but must drop to zero much more quickly for very large magnitudes.

An important extension involves adding locations to the description of the offspring so that the branching structure evolves in both space and time. Then, one obvious way of extending the model is to have the ground process include both space and time coordinates, retaining the same mark space $\mathcal{K}$.

From the computational point of view, however, and especially for the conditional intensity and likelihood analyses to be described in Chapter 7, there are advantages in keeping the ground process to the set of time points and regarding the spatial coordinates as additional dimensions of the mark. The weight (magnitude) component of the mark retains its unpredictable character (so the weights are i.i.d. given the past), but we allow the spatial component of the mark to be affected by the spatial location of its ancestor.

No matter which of these descriptions we adopt, the cluster structure evolves over both space and time, offspring events occurring at various distances away from the initial ancestor, just as they follow it in time. When the branching structure is spatially homogeneous, the infectivity measure $\mu(\mathrm{d}t \times \mathrm{d}x)$ depends both on the time delay $u = t - t_0$ and the displacement $y = x - x_0$ from the time and location of the ancestor (t_0, x_0).

Various branching mechanisms of this type have been proposed in the literature [see e.g. Ogata (1998) for a review]. Thus, Vere-Jones and Musmeci (1992) suggests a space–time diffusion with infectivity density

$$\mu(\mathrm{d}u \times \mathrm{d}y \times \mathrm{d}z) = \frac{\beta \mathrm{e}^{-\beta u}}{2\pi u \sigma_y \sigma_z} \exp\left[-\frac{1}{2u}\left(\frac{y^2}{\sigma_y^2} + \frac{z^2}{\sigma_z^2}\right)\right] \mathrm{d}u\,\mathrm{d}y\,\mathrm{d}z,$$

whereas Ogata's space–time ETAS model uses a simpler product form for the space and time terms. Many choices are possible for the components of the model without affecting the underlying cluster character. In some applications, the assumption of spatial homogeneity may not be appropriate, so the infectivity or mark distribution may depend on the absolute location of the offspring as well as its separation from the ancestor.

In all of this wide diversity of models, the basic sufficient condition for the existence of a stationary version of the model, essentially the subcriticality of the offspring branching process, is affected only insofar as the integral of the infectivity measure needs to be extended over space as well as time. □

We conclude this section with a preliminary foray into the fascinating and also practically important realm of *stochastic geometry*. Marked point processes play an important role here as models for finite or denumerable families of random geometrical objects. The objects may be of many kinds: triplets or quadruplets of points (then, the process would be a special case of a cluster process), circles, line segments, triangles, spheres, and so on.

Definition 6.4.VIII (Particle process)**.** *A* particle process *is a point process with state space $\Sigma_{\mathcal{X}}$ equal to the class of nonempty compact sets in $\mathcal{X}$.*

Thus, a typical realization of a particle process is a sequence, ordered in some way, of compact sets $\{K_1, K_2, \ldots\}$ from the c.s.m.s. $\mathcal{X}$. An underlying difficulty with such a definition is that of finding a convenient metric for the space $\Sigma_{\mathcal{X}}$. One possibility is the *Hausdorff metric* defined by

$$\rho(K_1, K_2) = \inf\{\epsilon\colon K_1 \subseteq K_2^{\epsilon} \text{ and } K_2 \subseteq K_1^{\epsilon}\},$$

where K^{ϵ} is the *halo set* $\bigcup_{x \in K} S_{\epsilon}(x)$ (see Appendix A2.2); for further references and discussion, see Stoyan *et al.* (1995), Stoyan and Stoyan (1994), and Molchanov (1997), amongst others. In special cases, when the elements are more specific geometrical objects such as spheres or line segments, this difficulty does not arise, as there are many suitable metrics at hand. Very often, interest centres on the *union set* or *coverage process*

$$\Xi = \bigcup S_i$$

(see Hall, 1988), which is then an example of a *random closed set* in $\mathcal{X}$.

Now let us suppose that $\mathcal{X} = \mathbb{R}^d$ and that for each compact set $S \subset \mathcal{X}$ we can identify a unique *centre* $y(S)$, for example its centre of gravity. Then, we

may introduce an equivalence relation among the sets in $\Sigma_{\mathcal{X}}$ by defining two compact sets to belong to the same equivalence class if they differ only by a translation. The sets in $\Sigma^o \equiv \Sigma^o_{\mathcal{X}}$, the compact subsets of $\mathcal{X}$ with their centres at the origin, index the equivalence classes so that every set $S \in \Sigma_{\mathcal{X}}$ can be represented as the pair (y, S^o), where $y \in \mathcal{X}$ and $S^o \in \Sigma^o$, and $S = y + S^o$ (set addition). This opens the way to defining the particle process as an MPP $\{y_i, S_i\}$, where the $\{y_i\}$ form a point process in $\mathcal{X}$ and the marks $\{S_i\}$ take their values in Σ^o. Once again, there is the problem of identifying a convenient metric on Σ^o, but this point aside, we have represented the original particle process as an example of a so-called *germ–grain model* in which the $\{y_i\}$ are the *germs* and the $\{S_i\}$ are the *grains*. The next example illustrates one of the most straightforward and widely used models of this type.

Example 6.4(e) *Boolean model.* This is the compound Poisson analogue for germ–grain models. We suppose that the locations $\{y_i\}$ form a Poisson process in $\mathcal{X}$ and that the compact sets S_i^o are i.i.d. and independent of the location process; write $S_i = y_i + S_i^o$. Two derived processes suggest themselves for special attention. One is the random measure $\Upsilon(\cdot)$ formed by superposing the compact sets S_i. With the addition of random weights W_i, this gives the bounded set A the (random) mass

$$\Upsilon(A) = \sum_i W_i \, \ell(A \cap S_i) \qquad (A \in \mathcal{B}_{\mathcal{X}}), \tag{6.4.13}$$

where $\ell(\cdot)$ is the reference measure on $\mathcal{X}$ (e.g. Lebesgue measure, or counting measure on a lattice). The other is the localized measure of the union set Ξ described above, which gives the bounded set A the (random) mass

$$\Psi(A) = \ell(A \cap \Xi) \equiv \ell\big\{ \textstyle\bigcup_i (A \cap S_i) \big\}. \tag{6.4.14}$$

For example, (6.4.13) might represent the total mass of ejected material falling within the set A from a series of volcanic eruptions at different locations; then (6.4.14) would represent the area of A covered by the ejected material.

In both cases, the processes can be represented in terms of densities forming random processes (random fields) on $\mathcal{X}$. Thus, (6.4.13) and (6.4.14) have respective densities

$$v(x) = \sum_i W_i I_{S_i}(x) \tag{6.4.15}$$

and

$$\psi(x) = I_{\{\cup_i S_i\}}(x). \tag{6.4.16}$$

Many aspects of these and related processes are studied in the stochastic geometry literature such as Mathéron (1975), Stoyan *et al.* (1995) and Molchanov (1997). Here we restrict ourselves to a consideration of the mean and covariance functions of (6.4.15) and (6.4.16) under the more explicit assumptions that $\mathcal{X} = \mathbb{R}^2$, that the location process N_{g} of centres $\{y(S_i)\} = \{y_i\}$ is a simple Poisson process with constant intensity λ, and that each S_i is a

disk of random radius R_i and has weight W_i that may depend on R_i but that the pairs (R_i, W_i) are mutually independent and independent also of the centres $\{y_i\}$. Consistent with our earlier description, we thus have an MPP on $\mathbb{R}^2$, with mark space $\mathbb{R}_+ \times \mathbb{R}_+$, and hence a point process N on $\mathbb{R}^2 \times \mathbb{R}_+^2$.

The mean and covariance function for $v(x)$ can be found by first conditioning on the ground process N_{g} as in earlier examples. Thus, writing $v(x)$ as

$$v(x) = \int_{\mathbb{R}^2 \times \mathbb{R}_+^2} w I_{\{r \geq \|y-x\|\}}(r, y)\, N(\mathrm{d}y \times \mathrm{d}r \times \mathrm{d}w) \tag{6.4.17}$$

and taking expectations, the independence assumptions coupled with the stationarity of the Poisson process yield

$$\begin{aligned} \mathrm{E}[v(x)] &= \lambda\, \mathrm{E}\Big[W \int_{\mathbb{R}^2} I_{\{R \geq \|y\|\}}(R, y)\, \mathrm{d}y \Big] = \lambda\, \mathrm{E}\Big[W \int_0^R \int_0^{2\pi} r\, \mathrm{d}r\, \mathrm{d}\theta \Big] \\ &= \lambda\, \pi\, \mathrm{E}(W R^2)\,. \end{aligned}$$

The second moment $\mathrm{E}[v(x_1)v(x_2)]$ can be found similarly by first conditioning on the $\{y_i\}$. Terms involving both pairs of distinct locations and coincident locations (arising from the diagonal term in the second-moment measure of the location process) are involved. However, as for Poisson cluster processes, we find that the covariance $\mathrm{cov}[v(x_1), v(x_2)]$ depends only on the term involving coincident locations: it equals

$$\begin{aligned} &\mathrm{E}\Big[\int_{\mathbb{R}^2 \times \mathbb{R}_+ \times \mathbb{R}_+} w^2 I_{\{r \geq \|y-x_1\|, r \geq \|y-x_2\|\}}(r, y)\, N(\mathrm{d}y \times \mathrm{d}r \times \mathrm{d}w) \Big] \\ &\qquad = \lambda \mathrm{E}\left[W^2 \int_{\mathbb{R}^2} I_{\{R \geq \max(\|y-x_1\|, \|y-x_2\|)\}}(R, y)\, \mathrm{d}y \right] \\ &\qquad = 2\lambda \mathrm{E}\big[W^2 \big(R^2 \arccos(u/R) - u\sqrt{R^2 - u^2} \big) I_{\{R \geq u\}}(R) \big], \end{aligned}$$

where $u = \frac{1}{2}\|x_1 - x_2\|$. Note that the first moment is independent of x and the covariance is a function only of $\|x_1 - x_2\|$, as we should expect from the stationary, isotropic character of the generating process. Note also that if the radius R is fixed, the covariance vanishes for $\|x_1 - x_2\| > 2R$.

The resemblance of these formulae to those for Poisson cluster processes is hardly coincidental. From a more general point of view, the process is a special case of LeCam's precipitation model in Exercise 6.3.1, where the Poisson cluster structure is generalized to cluster random measures. Some details and extensions are indicated in Exercise 6.4.6.

The corresponding formulae for the union process present quite different and, in general, much harder problems since we lose the additive structure for the independent contributions to the sum process. The first moment $\mathrm{E}[\psi(x)]$ represents the *volume fraction* of space (in this case area) occupied

by the union set Ξ. It can be approached by the following argument, which is characteristic for properties of the Boolean model. First, note that

$$1 - \mathrm{E}[\psi(x)] = 1 - \Pr\{\Xi \ni x\} = \Pr\{\Xi \not\ni x\} = \mathrm{E}\Big[\prod_i [1 - I_{S_i}(x)]\Big].$$

Conditioning on the locations $\{y_i\}$ (i.e. on the ground process N_{g}), we can write

$$\Pr\{\Xi \not\ni x \mid N_{\mathrm{g}}\} = \prod_i \Pr\{R_i < \|x - y_i\|\} = \prod_i h(y_i; x)\,,$$

say, where $h(y; x) = \mathrm{E}[I_{[0,\|y-x\|)}(R)]$ and R has the common distribution of the i.i.d. radii R_i. Removing the conditioning, we have

$$1 - \mathrm{E}[\psi(x)] = \mathrm{E}\Big[\prod_i h(y_i; x)\Big] = G_{\mathrm{g}}[h(\cdot\,; x)] = \exp\Big(-\lambda \int_{\mathbb{R}^2} [1 - h(y; x)]\,\mathrm{d}y\Big).$$

Substituting for $h(y; x)$ and simplifying, we obtain for the mean density the constant

$$p^* \equiv \mathrm{E}[\psi(x)] = 1 - \mathrm{e}^{-\lambda\,\mathrm{E}(\pi R^2)}. \tag{6.4.18}$$

For the second product moment, using similar reasoning, we have

$$\begin{aligned} m_2(x_1, x_2) = \mathrm{E}[\psi(x_1)\psi(x_2)] &= \Pr\{\Xi \ni x_1,\, \Xi \ni x_2\} \\ &= \Pr\{\Xi \ni x_1\} + \Pr\{\Xi \ni x_2\} - [1 - \Pr\{\Xi \not\ni x_1 \text{ or } x_2\}] \\ &= \mathrm{E}[\psi(x_1)] + \mathrm{E}[\psi(x_2)] - [1 - \Pr\{\Xi \not\ni x_1,\, \Xi \not\ni x_2\}] \\ &= 2p^* - 1 + G_{\mathrm{g}}[h(\cdot\,; x_1, x_2)], \end{aligned}$$

say, where $h(y; x_1, x_2) = \mathrm{E}[I_{[0,\min(\|y-x_1\|,\|y-x_2\|)]}(R)]$. Substituting for the p.g.fl. of the Poisson ground process, putting $u = \frac{1}{2}\|x_1 - x_2\|$ and simplifying, we find that $m(x_1, x_2)$ equals

$$2p^* - 1 + \exp\left(-\lambda \mathrm{E}\Big[\pi R^2 (1 + I_{\{R<u\}}) + 2\Big(R^2 \arcsin\frac{u}{R} + u\sqrt{R^2 - u^2}\Big) I_{\{R \geq u\}}\Big]\right).$$

Exercise 6.4.10 sketches an extension to higher-order product moments. □

Exercises and Complements to Section 6.4

6.4.1 For the atomic random measure ξ with independent marks as in Proposition 6.4.V, show that the third-order moment measure $M_3^{\xi}(A_1 \times A_2 \times A_3)$ equals

$$\begin{aligned} &\int_{A_1 \times A_2 \times A_3} \mu_1(x_1)\mu_1(x_2)\mu_1(x_3)\, M_{[3]}^{\mathrm{g}}(\mathrm{d}x_1 \times \mathrm{d}x_2 \times \mathrm{d}x_3) \\ &\quad + \Big[\int_{A_1 \times A_{23}} + \int_{A_2 \times A_{31}} + \int_{A_3 \times A_{12}}\Big] \mu_1(x_1)\mu_2(x_2)\, M_{[2]}^{\mathrm{g}}(\mathrm{d}x_1 \times \mathrm{d}x_2) \\ &\quad + \int_{A_1 \cap A_2 \cap A_3} \mu_3(x_1)\, M_1^{\mathrm{g}}(\mathrm{d}x_1), \end{aligned}$$

where $A_{ij} = A_i \cap A_j$ for $i \neq j$.

[*Hint*: Each side is the coefficient of $\frac{1}{6}s^3$ in the respective expansions of (6.4.7) with argument $sh(\cdot)$, using (6.1.9) for the Laplace functional, (5.5.4) [with $\eta(x) = \phi_{sh(x)} - 1 = -sh(x)\mu_1(x) + \frac{1}{2}s^2[h(x)]^2\mu_2(x) - \frac{1}{6}s^3[h(x)]^3\mu_3(x) + \cdots$ and $\mu_r(x) = \int_{\mathcal{K}} \kappa^r F(\mathrm{d}\kappa \mid x)$, $r = 1, 2, 3$] for the p.g.fl., and ϕ_{sh} as in (6.4.7). The general case now follows by appealing to the symmetry (invariance under permutations of the axes) of the moment measures.]

6.4.2 Develop formulae, analogous to those of Proposition 6.4.V, for characteristic functionals of MPPs with marks in $\mathbb{R}$. Use these to extend the results of Proposition 6.4.V to the case where ξ may be a signed measure.

6.4.3 *Cramér bound on probability of ruin.* For the compound risk process, verify the following results [with notation as for Example 6.4(a)].

(i) The sequence $U_n - U_0$ forms a random walk with mean $\alpha/\mu - \mathrm{E}(W)$.

(ii) If ruin occurs, then it does so at the first time point t_n for which $U_n < 0$.

(iii) If $\alpha \le \mu\mathrm{E}(W)$, then ruin is certain, but if $\alpha > \mu\mathrm{E}(W)$, then there is positive probability that ruin will never occur.

(iv) In the latter case, if the Laplace–Stieltjes transform $\mathrm{E}(\mathrm{e}^{-sW})$ is an entire function of s, then there exists positive real s^* such that $\mathrm{E}(\mathrm{e}^{-s^*W}) = 1$.

(v) The sequence $\{\zeta_n\} = \{\exp(-s^*U_n)\}$ constitutes a martingale for which the time of ruin is a stopping time.

(vi) Let p_M denote the probability that ruin occurs before the accumulated reserves reach a large number M. Deduce from the martingale property that

$$p_M\mathrm{E}[\exp(s^*\Delta_0) \mid 0] + (1 - p_M)\mathrm{E}[\exp(-s^*\Delta_M) \mid M] = \exp(-s^*U_0),$$

where $-\Delta_0$ and Δ_M are the respective overshoots at 0 and M.

(vii) Hence, obtain the Cramér bound for the probability of ultimate ruin

$$p = \lim_{M\to\infty} p_M \le \exp(-s^*U_0)\,.$$

6.4.4 Find first and second factorial moment measures for the ground processes of the marked and space–time Hawkes processes described in Example 6.4(c). [*Hint*: Use the cluster process representation much as in Example 6.3(c).]

6.4.5 Study the Laplace functional and moment measures for the random measure ξ for a Hawkes process with unpredictable marks. [*Hint*: Use the cluster representation to get a general form for the p.g.fl. of the process as a process on $\mathcal{X} \times \mathcal{K}$. From it, develop equations for the first and second moments.] Are explicit results available?

6.4.6 Formulate the process $\Upsilon(A)$ in (6.4.13) as an example of a LeCam process (see Exercise 6.3.1). Show that in the special case considered in (6.4.17), when the random sets are spheres [= disks in $\mathbb{R}^2$] with random radii we can write

$$L_\xi[f \mid x] = \mathrm{E}\left[\exp\left(-W\int_{\mathbb{R}^2} f(y)\, I_{\{R\ge\|x-y\|\}}(y)\,\mathrm{d}y\right)\right].$$

Derive expressions for the mean and covariance functions of $v(x)$ as corollaries.

6.4.7 *Higher-order moments of the union set.* In the context of the union set Ξ of the Boolean model of Example 6.4(e), show that the kth product moment

$$\mathrm{E}[\psi(x_1)\cdots\psi(x_k)] = \Pr\{\Xi \ni x_j \ (j = 1, \ldots, x_k)\},$$

for k distinct points $x_1, \ldots, x_k$ in $\mathcal{X} = \mathbb{R}^2$, equals

$$1 + \sum_{r=1}^{k} (-1)^r \sum\nolimits_r{}' q(x_{j_1}, \ldots, x_{j_r}),$$

where $\sum'$ denotes the sum over all distinct r-tuplets of the set $\{x_1, \ldots, x_k\}$, $q(x_1, \ldots, x_r) = G_{\mathrm{g}}[h(\cdot\,; x_1, \ldots, x_r)]$, and the function $h(y\,; x_1, \ldots, x_r) = \Pr\{R < \min_{1\le j\le r}\{\|x_j - y\|\}\}$. [*Hint*: The relation arises from taking expectations in the expansion of products of indicator random variables

$$\begin{aligned} I\{\Xi \ni \text{ all } x_j\} &= \prod\nolimits_j I\{\Xi \ni x_j\} = \prod\nolimits_j (1 - I\{\Xi \not\ni x_j\}) \\ &= 1 + \sum\nolimits_{r=1}^{k} (-1)^r \sum\nolimits_r{}' \prod\nolimits_{\ell=1}^{r} I\{\Xi \not\ni x_{j_\ell}\} \end{aligned}$$

and

$$\prod\nolimits_{\ell=1}^{r} I\{\Xi \not\ni x_{j_\ell}\} = \prod\nolimits_{\ell=1}^{r} \prod\nolimits_i I\{S_i \not\ni x_{j_\ell}\} = \prod\nolimits_i \prod\nolimits_{\ell=1}^{r} I\{S_i \not\ni x_{j_\ell}\},$$

and the conditional expectation of the last product, given the locations $\{y_i\}$, equals $h(y_i; x_{j_1}, \ldots, x_{j_r})$, as indicated.]

CHAPTER 7

Conditional Intensities and Likelihoods

A notable absence from the previous chapter was any discussion of likelihood functions. There is a good reason for this absence: the likelihood functions for most of the processes discussed in that chapter are relatively intractable. This difficulty was a block to the application of general point process models until the late 1960s, when a quite different approach was introduced in papers on filtering theory pioneered by the electrical engineers: see for example Yashin (1970), Snyder (1972), Boel, Varaiya and Wong (1975), Snyder(1975, 2nd ed. Snyder and Miller, 1991), and Kailath and Segall (1975). This approach led to the concept of the *conditional intensity function*. Once recognised, its role in elucidating the structure of point process likelihoods was soon exploited. General definitions of the conditional intensity function were given in Rubin (1972) and especially by Brémaud (1972), in whose work conditional intensity functions were rigorously defined and applied to likelihood and other problems (see also Brémaud, 1981). Even earlier, Gaver (1963) had introduced what is essentially the same concept through his notion of a *random hazard function*.

Many of these ideas came together in the 1971 Point Process Conference (Lewis, 1972), as a result of which the links between likelihoods, conditional intensities, the theoretical work of Watanabe (1964) and Kunita and Watanabe (1967), and the more practical approaches of Gaver, Hawkes (1971a, b) and Cox (1972a) became more evident. Later, Liptser and Shiryayev (1974, 1977, 1978; 2nd ed. 2000) gave a comprehensive theoretical treatment, while Brémaud (1981) gave a more accessible account that emphasises applications to queueing theory; this same emphasis is in Baccelli and Brémaud (1994). The last two decades have seen the systematic development and application of these ideas to applied problems in many fields, perhaps especially in conjunction with techniques for simulating and predicting point processes.

Throughout this chapter runs the theme of delineating classes of models for which the conditional intensity function, and hence the likelihood, has a rel-

atively simple form. A key requirement is that the point process should have an *evolutionary* character: at any time, the current risk—which is just informal terminology for the conditional intensity function—should be explicitly expressible in terms of the past of the process. Many simple point processes in time, including stationary and nonstationary Poisson processes, renewal and Wold processes, and Hawkes processes, fall into this category. So too do many marked point processes in time and also space–time processes, provided that the current distributions of the marks and spatial locations, as well as the current risk, are explicitly expressible in terms of the past.

Purely spatial processes—so-called *spatial point patterns*—cannot be handled so readily this way because *they lack a time-like, evolutionary dimension*. Nor can processes such as the Neyman–Scott cluster process, in which estimation of the current risk requires averaging over complex combinations of circumstances. However, in some cases of this type, filtering and related iterative techniques can sometimes provide a route forward; they are discussed further in Chapters 14 and 15 alongside the more careful theoretical analysis required to handle conditional intensity functions in a general context.

This chapter provides an informal treatment of these issues. We start with a brief introduction to point process likelihoods for a.s. finite point processes, based on the Janossy densities introduced in Chapter 5. In principle the methods can be applied to observations on a general point process observed within a bounded observation region, but in practice the usefulness of this approach is severely curtailed by the difficulty of writing down the Janossy densities for the process within the observation region in terms of a global specification of the process. In Section 7.2, we move to the representation of the likelihood of a simple point process evolving in time. Here the technique of successive conditionings on the past, as the process evolves in time, reduces the difficulty above to that of specifying initial conditions for the process. It leads to a simple and powerful representation of the likelihood in terms of the conditional intensity function. Then, in Section 7.3 we examine the extension of these ideas to marked and space–time point processes, where the process retains an evolutionary character along the time axis.

Section 7.4 is devoted to the discussion of intensity-based random time changes, which have the effect of reducing a general initial process to a simple or compound Poisson process. The time changes are motivated by their applications to goodness-of-fit procedures based on the technique of 'residual point process analysis'. The concluding Sections 7.5 and 7.6 are concerned with uses of the conditional intensity for testing, simulating, and forecasting such processes, and with the links between point process entropy and the evaluation of probability forecasts.

7.1. Likelihoods and Janossy Densities

In the abstract at least, there are no special difficulties involved in the notion of a point process likelihood. Granted a realization $(x_1, \ldots, x_n)$ in some subset

A of the state space $\mathcal{X}$, we require the joint probability density of the x_i with respect to a convenient reference measure, which when $\mathcal{X} = \mathbb{R}^d$ is commonly the n-fold product of Lebesgue measure on $\mathbb{R}^d$. As usual, the likelihood should be considered as a function of the parameters defining the joint density and not as a function of the x_i and n, which are taken as given. The density here is for an unordered set of points; it represents loosely the probability of finding particles at each of the locations x_i and nowhere else within A, and so it is nothing other than the local Janossy density (Definition 5.4.IV) $j_n(x_1, \ldots, x_n \mid A)$ for the point process restricted to A. These considerations are formalized in the following two definitions.

Definition 7.1.I. (a) *Given a bounded Borel set $A \subseteq \mathbb{R}^d$, a point process N on $\mathcal{X} = \mathbb{R}^d$ is* regular on A *if for all integers $k \geq 1$ the local Janossy measures*

$$J_k(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_k \mid A)$$

of Section 5.4 *are absolutely continuous on $A^{(k)}$ with respect to Lebesgue measure in $\mathcal{X}^{(k)}$.*
(b) *It is* regular *if it is regular on A for all bounded $A \in \mathcal{B}(\mathbb{R}^d)$.*

Proposition 5.4.V implies that a *regular point process is necessarily simple.*

Definition 7.1.II. *The* likelihood *of a realization $x_1, \ldots, x_n$ of a regular point process N on a bounded Borel set $A \subseteq \mathbb{R}^d$, where $n = N(A)$, is the local Janossy density*

$$L_A(x_1, \ldots, x_n) = j_n(x_1, \ldots, x_n \mid A). \tag{7.1.1}$$

For convenience, we often abbreviate L_A to L.

When the whole point process is a.s. finite, and the set A coincides with the space $\mathcal{X}$, the situation is particularly simple. In many cases, the likelihood can be written down immediately from the definition; some examples follow.

Example 7.1(a) *Finite inhomogeneous Poisson process in $A \subset \mathbb{R}^d$.* Suppose the process has intensity measure $\Lambda(\cdot)$ with density $\lambda(x)$ with respect to Lebesgue measure on $\mathbb{R}^d$. It follows from the results in Section 2.4 that the total number of points in A has a Poisson distribution with mean $\Lambda(A)$ and that conditional on the number N of such points, the points themselves are i.i.d. on A with common density $\lambda(x)/\Lambda(A)$. Suppose we observe the points $\{x_1, \ldots, x_n\}$ within A, with $n = N(A)$. In this case, we may assume $\mathcal{X} = A$ without any effective loss of generality, as the complete independence property ensures that the behaviour within A is unaffected by realization of the process outside A. Then, taking logs of the Janossy density gives for the log likelihood the formula

$$\log L(x_1, \ldots, x_n) = \sum_{i=1}^{n} \log \lambda(x_i) - \int_A \lambda(x)\, \mathrm{d}x, \tag{7.1.2}$$

of which (2.1.9) is the special case $\mathcal{X} = \mathbb{R}$. This example continues shortly. □

Equation (7.1.2) is basic to the likelihood theory of evolutionary processes. As we shall see in the next section, it extends to a wide range of such processes, provided the rate $\lambda(t)$ is interpreted in a sufficiently broad manner.

Another important use for the likelihood in (7.1.2) is as a reference measure for the more general concept of the *likelihood ratio*. Let N, N' be two point processes defined on a common state space $\mathcal{X}$ and with probability measures $\mathcal{P}$, $\mathcal{P}'$, respectively, on some common probability space $(\Omega, \mathcal{E})$. By a mild abuse of language, we shall say that N is *absolutely continuous* with respect to N', denoting it $N \ll N'$, if $\mathcal{P}$ is absolutely continuous with respect to $\mathcal{P}'$. In talking about a finite point process on a bounded Borel subset A of $\mathbb{R}^d$, the appropriate probability space is $A^{\cup}$ [see (5.3.8)], and an appropriate reference measure is that of a Poisson process on A with constant intensity. In this context, we have the following result.

Proposition 7.1.III. *Let N, N' be point processes defined on the c.s.m.s. $\mathcal{X} = \mathbb{R}^d$, and let A be a bounded Borel set $\subset \mathbb{R}^d$. Then $N \ll N'$ on A if and only if for each $k > 0$ the local Janossy measures $J_k(\cdot \mid A)$ and $J'_k(\cdot \mid A)$ associated with N and N', respectively, satisfy $J_k(\cdot \mid A) \ll J'_k(\cdot \mid A)$. In particular, if N' is the Poisson process with constant intensity $\lambda > 0$, then $N \ll N'$ if and only if N is regular on A.*

Proof. If N' vanishes identically on A, the conclusion is trivial, so we suppose this is not the case. Recall from the discussion around Proposition 5.3.II that an event E from $A^{\cup}$ has the structure $E = \bigcup_0^\infty S_k$, where each S_k is a symmetric set; i.e. an element of $\mathcal{B}_{\mathrm{sym}}^{(k)}(A)$ (see Exercise 5.3.5).

To establish the absolute continuity $N \ll N'$ on A, we have to show that if $\mathcal{P}$, $\mathcal{P}'$ are the probability measures induced on $A^{\cup}$ by N, N', then $\mathcal{P}(E) = 0$ whenever $\mathcal{P}'(E) = 0$. Since N' is not identically zero, $\mathcal{P}'(E) = 0$ only if $S_0 = \emptyset$ and $\mathcal{P}'(S_k) = 0$ for all $k > 0$. It is enough here to suppose that S_k is the symmetrized form of a product set $A_1 \times \ldots \times A_k$, where the A_i form a partition of A, since product sets of this form generate the symmetric sets in $A^{(k)}$. Then, from the definition of the local Janossy measures,

$$k!\,\mathcal{P}(S_k) = J_k(A_1 \times \ldots \times A_k \mid A) = J_k(S_k \mid A).$$

Similarly,

$$k!\,\mathcal{P}'(S_k) = J'_k(A_1 \times \ldots \times A_k \mid A).$$

Thus, if $\mathcal{P}'(E) = 0$, then for each k, $J'_k(S_k \mid A) = 0$, and if $J_k(\cdot \mid A) \ll J'_k(\cdot \mid A)$, then $J_k(S_k \mid A) = \mathcal{P}(S_k) = 0$ as well, so $\mathcal{P}(E) = 0$.

The same equivalences establish the converse relation.

If, in particular, N' is the Poisson process on A with constant intensity λ, then

$$J'_k(S_k \mid A) = k!\,\mathcal{P}'(S_k) = \left(\prod_{i=1}^{k} \lambda \ell(A_i)\right) \mathrm{e}^{-\lambda \ell(A)},$$

where ℓ is Lebesgue measure in $\mathbb{R}^d$. Thus, each local Janossy measure $J'_k(\cdot \mid A)$ is proportional to Lebesgue measure in $(\mathbb{R}^d)^k$, so $J_k(\cdot \mid A) \ll J'_k(\cdot \mid A)$ for all $k > 0$ if and only if N is regular. □

When densities are known explicitly for both processes, the likelihood ratio for a realization $\{x_1, \ldots, x_n\}$ within A is the ratio of the two Janossy densities of order n for the process on A. When the reference measure is that of a Poisson process with unit intensity, $\mathcal{P}^\#$ say, this can be written

$$L_A/L_A^\# = \mathrm{e}^{\ell(A)} j_n(x_1, \ldots, x_n \mid A). \tag{7.1.3a}$$

In other words, it is directly proportional to the Janossy measure itself. Alternatively, (7.1.3a), or more properly the collection of such expressions for all integers n, can be regarded simply as the *density* of the given point process on $A^\cup$ relative to the Poisson process measure as a reference measure. Written out in full, the Radon–Nikodym derivative for the two measures on $A^\cup$ takes the form (see Exercise 5.3.8)

$$\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}\mathcal{P}'}(\omega) = \mathrm{e}^{\lambda\ell(A)} \left(J_0 I_{N(A)=0} + \sum_1^\infty \frac{\lambda^n}{n!} j_n(x_1, \ldots, x_n) I_{N(A)=n} \right). \tag{7.1.3b}$$

We look again at the inhomogeneous Poisson process example in this light.

Example 7.1(a) (continued). As in (7.1.2), $\mathcal{P}_A$ denotes the distribution associated with an inhomogeneous Poisson process with intensity $\lambda(x)$. Then, the log likelihood ratio relative to the unit-rate Poisson takes the form

$$\log(L_A/L_A^\#) = \sum_{i=1}^N \log \lambda(x_i) - \int_A [\lambda(x) - 1] \, \mathrm{d}x.$$

One further manipulation of this equation is worth pointing out. Suppose that $\lambda(x)$ has the form

$$\lambda(x) = C\phi(x),$$

where C is a positive scale parameter and $\phi(x)$ is normalized so that $\int_A \phi(x) \, \mathrm{d}x = 1$. Then (7.1.3) becomes

$$\log(L_A/L_A^\#) = N \log C + \sum_{i=1}^N \log \phi(x_i) - C + \ell(A).$$

Differentiation with respect to C yields the maximum likelihood estimate

$$\widehat{C} = N,$$

and it is clear that here N is a sufficient statistic for C. Moreover, substituting this value back into the likelihood yields $\hat{L}_A$, say, and the ratio becomes

$$\log(\hat{L}_A/L_A^\#) = N \log N - N + \ell(A) + \sum \log \phi(x_i).$$

Apart from a constant term, this is the same expression as would be obtained by first conditioning on N, when the likelihood reduces to that for N independent observations on the distribution with density $\phi(x_i)$. Clearly, in this situation, estimates based on Poisson observations with variable N yield the same results as estimates obtained by first conditioning on N, a statement that is not true with other distributions even asymptotically.

Finally, consider the model with constant but arbitrary (unknown) rate C, so that

$$\lambda(x) = C/\ell(A)$$

with likelihood L_A^0, say. We find as a special case of the above

$$\log(\hat{L}_A^0/L_A^{\#}) = N \log N - N + \ell(A) - N \log \ell(A),$$

from which

$$\log(\hat{L}_A/\hat{L}_A^0) = \sum \log \phi(x_i) + N \log \ell(A).$$

Thus, the term on the right-hand side is the increment to the log likelihood ratio achieved by fitting a model with density proportional to $\phi(x)$ over a model with constant density. This elementary observation often provides a useful reduction in the complexity of numerical computations involving Poisson models. □

The next three examples form some of the key models in representing spatial point patterns within finite regions. Although the likelihoods can be given in more or less explicit form, explicit analytic forms for other characteristics of the process—moment and covariance densities, for example—are not easy to find, mainly because of the intricate links between the numbers and locations of particles within a given region.

Another major problem is that, in many important examples, the characteristics of the process are not given directly in terms of the local Janossy measures for the process on A but in terms of global characteristics from which the local characteristics have to be derived. If the process is defined directly in terms of the local Janossy measures, then it is assumed, either tacitly or otherwise, that any effects from points outside the observation region A have been incorporated into the definitions or ignored. If this is not the case—if, for example, one wishes to fit a stationary version of a process with specified interaction potentials—the situation becomes considerably more complex. Allowing for the influence exerted in an average sense by points outside A amounts to nothing less than a generalized version of the Ising problem, where the issue was first posed in the context of magnetized particles in a one-dimensional continuum. The issue is discussed further around Example 7.1(e) and in Chapter 15. In the next three examples, this difficulty is avoided by assuming that the process is totally finite on $\mathcal{X}$ and that $\mathcal{X} = A$.

EXAMPLE 7.1(b) *Finite Gibbs processes on $\mathcal{X}$; pairwise interaction systems* [see Example 5.3(c)]. An important class of examples from theoretical physics

was introduced in Example 5.3(c), with Janossy densities and hence likelihoods of the form

$$L(x_1, \ldots, x_n) = C(\theta) \exp[-\theta U(x_1, \ldots, x_n)], \tag{7.1.4}$$

where U can be expressed as a sum of interaction potentials, and the partition function $C(\theta)$ is chosen to satisfy the normalization condition of equation (5.3.7). In the practically important case of pairwise interactions, only first-and-second order interaction terms are present, and U takes the form

$$U(x_1, \ldots, x_n) = \sum_{i}^{n} \psi_1(x_i) + \sum_{j<i}^{n} \psi_2(x_i, x_j).$$

Although such models have a valuable flexibility in modelling different types of spatial interactions, their initial attractiveness is somewhat countered by the difficulty of expressing the partition function $C(\theta)$ in terms of the other parameters of the model. In fact, exact expressions for the likelihood do not seem to be available in any cases where the second-order term is nontrivial. Ogata and Tanemura (1981) advocate using the approximations (virial expansions) developed by physicists for this purpose, but even so the computations are laborious and their accuracy uncertain. Diggle *et al.* (1994) compares different numerical approximations. More recent work has focussed on Markov chain Monte Carlo (MCMC) approximations, where the equilibrium solution is obtained numerically as a long-term average of simulations of a Markov chain having the required distribution as its stationary distribution (see e.g. Häggstrøm *et al.*, 1999; Andersson and Britton, 2000, Chapter 11). By judicious choice of the Markov chain transition probabilities, the normalizing constant can be made to disappear from the estimates (e.g. Exercise 7.1.7).

Another technique that obviates the need to explicitly evaluate the normalizing constant is to replace the true likelihood L by the *pseudolikelihood* $L^\dagger$ defined by

$$L^\dagger(x_1, \ldots, x_n) = \prod_{k=1}^{n} \frac{j_n(x_1, \ldots, x_n)}{j_{n-1}(\{x_1, \ldots, x_n\} \setminus x_k)}\,.$$

Since this involves a ratio of Janossy densities, the normalizing constant disappears. It is very much easier, therefore, to derive the pseudolikelihood estimates for a model of this kind than it is to derive the true maximum likelihood estimates. On the other hand, the properties of estimates obtained by maximizing the pseudolikelihood, for example their consistency or asymptotic normality, are currently only partially resolved. In practice, they behave in much the same way as standard maximum likelihood estimates, and it seems likely that in time the theory of both will be subsumed under a more general umbrella. See Baddeley (2001) for examples and further discussion. □

Example 7.1(c) *Strauss processes; hard-core models* (Strauss, 1975; Kelly and Ripley, 1976). Strauss processes are the special cases of the model above

when ψ_1 is a constant α and $\psi_2(x_i, x_j)$ has a fixed value β within the range $\|x_i - x_j\| < R$, for some fixed $R < \infty$, and is zero outside it. In this case, the Janossy density takes the form

$$j_n(x_1, \ldots, x_n) = C(\alpha, \beta, R)\, \alpha^n \beta^m,$$

where $m = m(x_1, \ldots, x_n)$ is the number of distinct pairs x_i, x_j for which $\|x_i - x_j\| < R$. The Janossy density is constant on hypercylinders around the diagonals $x_i = x_j$ and their intersections in $\mathcal{X}^{(n)}$.

For the process to be well defined, the sum of the Janossy measures must converge [see equation (5.3.9)], which occurs if and only if either $\beta < 1$ or $\beta = 1$ and $\alpha \leq 1$ (cf. Exercise 7.1.8). The condition $\beta < 1$ implies some degree of repulsion between points, implying underdispersion relative to the Poisson process. In particular, the choice $\beta = 0$ corresponds to a so-called *hard-core* model, in which points cannot come closer than within a distance R of each other. Other examples of hard-core models appear in Section 8.3.

For other values of α and β, the series of Janossy measures diverges so that they no longer correspond to a well-defined finite point process. Thus, the process cannot be used directly to model clustering, but modified Strauss processes with $\beta > 1$ can be produced by weighting the Janossy densities with a sequence of constants, w_n say, chosen to ensure convergence of the Janossy measures. The most extreme case, corresponding to setting $w_n = 1$ for some selected value of n and to 0 otherwise, corresponds to conditioning on an outcome of fixed size n. See Kelly and Ripley (1976) and Exercise 7.1.8 for details. □

Example 7.1(d) *Markov point processes* (Ripley and Kelly, 1977). In order to introduce some concept of Markovianity into the unordered context of spatial point processes, Ripley and Kelly first assume the existence of a relationship $\sim$ among the points $\{x_i\}$ of a realization. When $x_i \sim x_j$, the points (x_i, x_j) are said to belong to the same *clique* or *neighbourhood class*. Given any realization of the process, the points may be uniquely divided up into cliques, where a point x_i forms a clique by itself if there are no other points x_j in the realization for which $x_i \sim x_j$. Let $\varphi: \mathcal{X}^{\cup} \mapsto \mathbb{R}^+$ be a function defined on cliques $\mathcal{V}$ and taking real positive values. Then, a finite point process is said to be a Markov point process if the Janossy density for a realization with a total of N points coming from V cliques $\mathcal{V}_k$ with N_k points in $\mathcal{V}_k$ takes the form

$$j_N(x_1, \ldots, x_N \mid A) = C \prod_{k=1}^{V} \varphi(\mathcal{V}_k), \tag{7.1.5}$$

where $N = \sum_k N_k$ and C is a normalization constant chosen to ensure the Janossy measures satisfy condition (5.3.7). This is equivalent to requiring that the density relative to a unit-rate Poisson process is always proportional to the product $\prod_{k=1}^{V} \varphi(\mathcal{V}_k)$ no matter how many points the realization may contain.

A common choice is to take $x_i \sim x_j$ if $||x_i - x_j|| < R$. We leave the reader to verify that this leads to a well-defined equivalence relation and that if

$$\varphi(\mathcal{V}) = \begin{cases} 0 & \text{if } N(\mathcal{V}) \geq 2, \\ \alpha & \text{otherwise,} \end{cases}$$

then we recover the hard-core version of the Strauss model.

Many other important examples of spatial point processes may be put into this form, although the appropriate definitions of clique and the function ϕ may take some teasing out. A more extended discussion of Markov point processes is given in Chapter 10. □

In some examples, it is possible to take advantage of a simple expression for the log p.g.fl.; this generally leads to simple expressions for the Khinchin measures, which can then be used to construct the Janossy measures via the combinatorial formulae (5.5.31). The simplest example is the Poisson process, for which only the first Khinchin measure is nonzero, so in the notation of Exercise 5.5.8 we have, say,

$$K_0 = -\log p_0(A) = \int_A \lambda(x)\,\mathrm{d}x = \Lambda(A),$$
$$k_1(x \mid A) = \lambda(x).$$

Then, from (5.5.31) we have $j_n(x_1, \ldots, x_n) = p_0(A) \prod_{i=1}^n \lambda(x_i)$ as used in (7.1.3a).

The next most complicated example of this type is the Gauss–Poisson process described in detail in Example 6.3(d) for which just the first two of the Khinchin measures are nonzero.

At this point, we meet an example of the difficulty referred to in the discussion preceding Example 7.1(b). The defining quantities for the Gauss–Poisson process are the measures $Q_1(\mathrm{d}x)$ and $Q_2(\mathrm{d}x_1 \times \mathrm{d}x_2)$ described in Proposition 6.3.IV. If the process is observed on a bounded set A, then we have to determine whether these quantities are given explicitly for the process on A or quite generally for the process on the whole of $\mathbb{R}$. In the former case the analysis can proceed directly and is outlined in Example 7.1(e)(i) below. In the latter case, however, and specifically in the case where we want to fit a model with densities $q_1(x) \equiv q_1$, $q_2(x_1, x_2) = q(x_1 - x_2)$ corresponding to a stationary version of the process, it is not clear how to allow for the interactions with points of the process lying outside of A and hence unobserved. It turns out that, for this particular model, explicit corrections for the average influence of such outside points can be made and amount to modifying the parameters for the process observed on A. This discussion is outlined in Example 7.1(e)(ii).

Example 7.1(e) (i) *Gauss–Poisson process on a bounded Borel set A.* From (6.3.30) or Exercise 6.3.12, we know that the log p.g.fl. of a Gauss–Poisson process defined on a bounded Borel set A as state space has the expansion

$$-\log G[h] = \int_A [1 - h(x)]\, K_1(\mathrm{d}x) + \int_{A^{(2)}} [1 - h(x)h(y)]\, K_2(\mathrm{d}x \times \mathrm{d}y).$$

Assume that $K_1(\mathrm{d}x) = \mu(x)\,\mathrm{d}x$ and $K_2(\mathrm{d}x \times \mathrm{d}y) = \frac{1}{2}q(x-y)\,\mathrm{d}x\,\mathrm{d}y$ for some function $\mu(\cdot)$ and some symmetric function $q(\cdot)$. Then, the Khinchin densities k_r are given by

$$k_1(x) = \mu(x), \qquad k_2(x,y) = q(x-y), \qquad k_r(\cdot) = 0 \quad (\text{all } r = 3, 4, \ldots),$$

and

$$\begin{aligned} K_0 = -\log p_0(A) &= \int_A \mu(x)\,\mathrm{d}x + \tfrac{1}{2}\int_A\int_A q(x-y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_A k_1(x)\,\mathrm{d}x + \tfrac{1}{2}\int_A\int_A k_2(x,y)\,\mathrm{d}x\,\mathrm{d}y. \end{aligned}$$

We turn to the expansion of the Janossy densities in terms of Khinchin densities given by equation (5.5.31), namely

$$j_n(x_1, \ldots, x_n \mid A) = \exp(-K_0) \sum_{r=1}^{n} \sum_{\mathcal{T} \in \mathcal{P}_{rn}} \prod_{i=1}^{r} k_{|S_i(\mathcal{T})|}(x_{i1}, \ldots, x_{i,|S_i(\mathcal{T})|}),$$

where the inner summation is taken over all partitions $\mathcal{T}$ of $x_1, \ldots, x_n$ into i subsets as described above Lemma 5.2.VI. The only nonzero terms arising in this summation are those relating to partitions into sets of sizes 1 and 2 exclusively. This leads to the form for the Janossy densities

$$\begin{aligned} &j_n(x_1, \ldots, x_n \mid A) \\ &= p_0(A) \sum_{k=0}^{[n/2]} {\sum^{*}}' \mu(x_{i'_1}) \cdots \mu(x_{i'_{n-2k}})\, q(x_{i_1} - x_{i_2}) \cdots q(x_{i_{2k-1}} - x_{i_{2k}}), \end{aligned} \tag{7.1.6}$$

where the summation ${\sum^{*}}'$ extends over the $n!/[(n-2k)!\,2^k]$ distinct sets of k pairs of different indices $(i_1, i_2), \ldots, (i_{2k-1}, i_{2k})$ from $\{1, \ldots, n\}$ satisfying $i_{2j-1} < i_{2j}$ $(j = 1, \ldots, k)$ and $i_1 < i_3 < \cdots < i_{2k-1}$, and $\{i'_1, \ldots, i'_{n-2k}\}$ is the complementary set of indices.

Given a realization $x_1, \ldots, x_n$ of a Gauss–Poisson process on a set A, its likelihood is then $j_n(x_1, \ldots, x_n \mid A)$, which is in principle computable but in practice is somewhat complex as soon as n is of moderate size.

Newman (1970) established (7.1.6) by an induction argument.

(ii) *Stationary Gauss–Poisson process.* In the specific case of a stationary (translation-invariant) Gauss–Poisson process, we can proceed as follows. The global process is defined by two global parameters, a mean density, say m, and a factorial covariance measure $\breve{C}_{[2]}$, which we shall assume to have density $q(x-y)$. From these we can obtain obtain versions of the *local* Khinchin densities from equations, analogous to (5.4.11),

$$k_1(x \mid A) = c_{[1]}(x) + \sum_{i=1}^{\infty} \frac{(-1)^j}{j!} \int_{A^{(j)}} c_{[1+j]}(x, y_1, \ldots, y_j)\,\mathrm{d}y_1 \cdots \mathrm{d}y_j,$$

which here reduces to

$$k_1(x \mid A) = m - \int_A q(x-y)\,\mathrm{d}y \equiv \mu(x) \qquad (x \in A),$$

and

$$k_2(x_1, x_2 \mid A) = q(x_1 - x_2) \qquad (x_1, x_2 \in A),$$

while all higher-order Khinchin measures vanish. Since these two densities define the two measures Q_1, Q_2 characterizing a Gauss–Poisson process [see Example 6.3(d)], we see firstly that the process on A is still a Gauss–Poisson process and secondly that its defining measures, unlike the moment measures, depend explicitly on the locations within the observation set A. In other words, although the local process on A is still a process of correlated pairs, its properties are no longer constant across A but depend in general on the proximity to the boundary of A.

From this discussion, we see that there is no loss of generality in assuming that $\mathcal{X} = A$, although to obviate the need for edge corrections we shall have to assume that the defining measures are not stationary, even though the global process may be so (see also Brix and Kendall, 2002). □

In principle, it is possible to write down expressions even more complicated than (7.1.6) for cluster processes with up to 3, 4, ... points in each cluster. Baudin (1981) developed an equivalent systematic procedure for writing down the likelihood of a Neyman–Scott cluster process, but again it is of substantial combinatorial complexity: see Exercises 7.1.5–6 for details (see also Baddeley, 1998).

The difficulty of finding the local Janossy measures in terms of global parameters of the model varies greatly with the model. In a few simple cases, such as the Poisson and Gauss–Poisson examples just considered, explicit expressions may be obtained. In other examples, finding exact solutions raises difficulties of principle as much as technical difficulty. Only the evolutionary processes, considered in the later sections of this chapter, provide a substantial class of models for which a ready solution exists and then only by taking special advantage of the order properties of the time-like dimension. Further discussion of the general problem is deferred until Chapter 15.

At the practical level, the difficulty can be alleviated to some extent by the use of so-called *plus sampling* or *minus sampling*. This consists of either adding to ('plus') or subtracting from ('minus') the original sampling region A a buffer region in which the points contribute indirectly to the likelihood by virtue of their effects on the probability density of the points in the inner region but are not included as part of the realization as such. Of course, the points in the buffer region do not play their full weight in the analysis, and the corrections so obtained are only approximate. There is clearly some delicacy in choosing the buffer region large enough to improve accuracy by reducing bias (arising from edge effects) but not so large that the improvement is offset by the loss of information due to not making full use of the data points in the buffer region. Edge effects are discussed again at the end of Section 8.1.

Another possible strategy is to introduce 'periodic boundary effects', essentially by wrapping the time interval around a circle, in the case of a one-dimensional problem, or, for a rectangular region in the plane, by repeating the original region (with the original data) at all contiguous positions in a rectangular tiling of the plane with the original region as base set. The rationale behind the procedure is that the missing data in a neighbourhood of the original observation will be replaced by data that may be expected to have similar statistical properties in general terms. Further discussion of these and similar techniques can be found in the texts by Ripley (1981), Cressie (1991), and Stoyan and Stoyan (1994).

EXAMPLE 7.1(f) *Fermion and boson processes* [see Examples 5.4(c) and 6.2(b)]. Each of these processes is completely specified by a global covariance function $c(x, y)$, and the local Janossy densities appear as either determinants [for the fermion process: see (5.4.19)] or permanents [for the boson process: see (6.2.11)]. In each case, the densities are derived from a resolvent kernel of the integral equation on A with kernel $c(\cdot\,,\cdot)$. As for the Gauss–Poisson process, the resulting explicit expressions for the Janossy densities (and thus the likelihoods) incorporate requisite adjustments for boundary effects. □

We conclude this section with an excursion into the realm of hypothesis testing; it has the incidental advantage of illustrating further the role of the Khinchin density functions. A commonly occurring need in practice is to test for the null hypothesis of a Poisson process against some appropriate class of alternatives, and it is then pertinent to enquire as to the form of the optimal or at least locally optimal test statistic for this purpose. This question has been examined by Davies (1977), whose general approach we follow.

The locally optimal test statistic is just the derivative of the log likelihood function, calculated at the parameter values corresponding to the null hypothesis. Davies' principal result is that this quantity has a representation as a sum of orthogonal terms, containing contributions from the factorial cumulants of successively higher orders. The formal statement is as follows (note that we return here to the general case of an observation region $A \subset \mathcal{X} = \mathbb{R}^d$).

Proposition 7.1.IV. *For a bounded Borel subset A of $\mathbb{R}^d$, let the distributions $\{\mathcal{P}_\theta\}$ correspond to a family of orderly point processes on $\mathbb{R}^d$ indexed by a single real parameter θ such that*

(i) *for $\theta = 0$ the process is a Poisson process with constant intensity μ, and*

(ii) *for all θ in some neighbourhood V of the origin, all factorial moment and cumulant densities $m_{[k]}$ and $c_{[k]}$ exist and are differentiable functions of θ and are such that for each $s = 1, 2, \ldots$ the series*

$$\sum_{k=1}^{\infty} \frac{1}{k!} \int_A \cdots \int_A c'_{[k+s]}(x_1, \ldots, x_s, y_1, \ldots, y_k; \theta)\, \mathrm{d}y_1 \cdots \mathrm{d}y_k \qquad (7.1.7)$$

is uniformly convergent for $\theta \in V$, and the series

$$\sum_{k=1}^{\infty} \frac{(1+\delta)^k}{k!} \int_A \cdots \int_A c'_{[k]}(y_1, \ldots, y_k;\, \theta)\, \mathrm{d}y_1 \cdots \mathrm{d}y_k \tag{7.1.8}$$

converges for some $\delta > 0$.

Then, the efficient score statistic $\partial \log L/\partial\theta\big|_{\theta=0}$ can be represented as the sum

$$D \equiv \frac{\partial \log L}{\partial \theta}\bigg|_{\theta=0} = \sum_{k=1}^{\infty} D_k, \tag{7.1.9}$$

where, with $I(y_1, \ldots, y_k) = 1$ if no arguments coincide and $= 0$ otherwise and $Z(\mathrm{d}y) = N(\mathrm{d}y) - \mu\, \mathrm{d}y$,

$$D_k = \frac{1}{\mu^k k!} \int_A \cdots \int_A I(y_1, \ldots, y_k) c'_{[k]}(y_1, \ldots, y_k;\, 0)\, Z(\mathrm{d}y_1) \cdots Z(\mathrm{d}y_k). \tag{7.1.10}$$

Under the null hypothesis $\theta = 0$ and $j > k \geq 1$,

$$\mathrm{E}(D_k) = \mathrm{E}(D_k D_j) = 0, \tag{7.1.11a}$$

$$\operatorname{var} D_k = \frac{1}{\mu^k k!} \int_A \cdots \int_A [c'_{[k]}(y_1, \ldots, y_k;\, 0)]^2\, \mathrm{d}y_1 \cdots \mathrm{d}y_k. \tag{7.1.11b}$$

Proof. We again use the machinery for finite point processes starting with the expression for the likelihood $L \equiv L_\theta = j_n(x_{1(1)n}; \theta)$ of the realization $\{x_1, \ldots, x_n\} \equiv \{x_{1(1)n}\}$ on the set A in the form [see (5.5.31)]

$$L = \exp(-K_0(\theta)) \sum_{j=1}^{n} \sum_{\mathcal{T} \in \mathcal{P}_{jn}} \prod_{i=1}^{j} k_{|S_i(\mathcal{T})|}(x_{i,1}, \ldots, x_{i,|S_i(\mathcal{T})|};\, \theta), \tag{7.1.12}$$

where the $k_r(\cdot)$ denote Khinchin densities and the inner summation extends over the set $\mathcal{P}_{jn}$ of all j-partitions $\mathcal{T}$ of the realization $\{x_{1(1)n}\}$. Because $\theta = 0$ corresponds to a Poisson process, $K_0(0) = \mu\ell(A)$ and $k_r(y_{1(1)n};\, 0) = 0$ unless $r = 1$ when $k_1(y;\, 0) = \mu$. Consequently, (7.1.12) for $\theta = 0$ reduces to $L_0 = \mu^n \exp(-\mu\ell(A))$, as it should. This fact simplifies the differentiation of (7.1.12) because, assuming (as we justify later) the existence of the derivatives

$$k'_r(y_{1(1)r};\, 0) \equiv \frac{\partial}{\partial\theta} k_r(y_{1(1)r};\, \theta)\big|_{\theta=0},$$

in differentiating the product term in (7.1.12), nonzero terms remain on setting $\theta = 0$ only if at most one set $S_i(\mathcal{T})$ has $|S_i(\mathcal{T})| > 1$ and all other $j - 1$ sets have $|S_i(\mathcal{T})| = 1$. Thus,

$$\begin{aligned}(\log L)' \equiv \frac{\partial \log L}{\partial \theta}\bigg|_{\theta=0} &= -K'_0(0) + \sum_{j=1}^{n} \mu^{j-1} {\sum}^{*} \frac{k'_{n-j+1}(x_{r_1}, \ldots, x_{r_{n-j+1}};\, 0)}{\mu^n} \\ &= -K'_0(0) + \sum_{i=1}^{n} \mu^{-i} {\sum}^{*} k'_i(x_{r_1}, \ldots, x_{r_i};\, 0),\end{aligned}$$

where the summation $\sum^*$ extends over all distinct selections of size i from the set $\{x_{1(1)n}\}$. Since this set is a realization of the process $N(\cdot)$ over A, the sum $\sum^*$ is expressible as the integral

$$\frac{1}{i!}\int_A \cdots \int_A I(y_{1(1)i};0)\, N(\mathrm{d}y_1)\cdots N(\mathrm{d}y_i),$$

where the factor $I(y_{1(1)i})$ avoids repeated indices and division by $i!$ compensates for the $i!$ recurrences of the same set of indices in different orders. This leads to the representation

$$(\log L)' = -K_0'(0) + \sum_{i=1}^{\infty} \frac{1}{\mu^i i!}\int_A \cdots \int_A I(y_{1(1)i})k'(y_{1(1)i};0)\, N(\mathrm{d}y_1)\cdots N(\mathrm{d}y_i) \tag{7.1.13}$$

now valid on an infinite range for i as the sum terminates after $N(A)$ terms.

When the Khinchin measures are known explicitly, (7.1.13) can be used directly. Otherwise, use the expansion akin to (5.5.29) of $k(\cdot)$ in terms of factorial cumulant densities

$$k_i(y_{1(1)i};0) = \sum_{j=0}^{\infty} \frac{(-1)^j}{j!}\int_A \cdots \int_A c_{[i+j]}(y_{1(1)i},\, u_{1(1)j};\, \theta)\, \mathrm{d}u_1 \cdots \mathrm{d}u_j,$$

which, in view of the assumption in (7.1.7), both shows that the $k_i(\cdot)$ are differentiable as assumed earlier and justifies term-by-term differentiation. Because of (7.1.12), the same is also true of L_θ. Also, since by (5.5.26) $K_0(\theta)$ is a weighted sum of all other Khinchin measures, substitution for $k_i'(\cdot)$ yields

$$K_0'(\theta) = \sum_{i=1}^{\infty} \frac{1}{i!}\int_A \cdots \int_A \Bigg(\sum_{j=0}^{\infty} \frac{(-1)^j}{j!} \times \int_A \cdots \int_A c'_{[i+j]}(y_{1(1)i}, u_{1(1)j};\theta)\, \mathrm{d}u_1 \cdots \mathrm{d}u_j \Bigg) \mathrm{d}y_1 \cdots \mathrm{d}y_i,$$

which on replacing j by $j-i$, inverting the order of summation, and using $\sum_{i=1}^{j}(-1)^{j-i}/[i!\,(j-i)!] = -(-1)^j/j!$ gives for $\theta = 0$

$$K_0'(0) = -\sum_{j=1}^{\infty} \frac{(-1)^j}{j!}\int_A \cdots \int_A c'_{[j]}(u_{1(1)j};0)\, \mathrm{d}u_1 \cdots \mathrm{d}u_j.$$

Similar substitution after differentiation into (7.1.13), rearrangement of the order of summation, and substitution for $-K_0'(0)$ yields

$$(\log L)' = \sum_{j=1}^{\infty} \frac{1}{\mu^j j!} \sum_{i=0}^{j} \frac{(-\mu)^{j-i} j!}{i!\,(j-i)!} \times \int_A \cdots \int_A c'_{[j]}(y_{1(1)i}, u_{1(1)j-i};0)\, N(\mathrm{d}y_1)\cdots N(\mathrm{d}y_i)\, \mathrm{d}u_1 \cdots \mathrm{d}u_{j-i}.$$

Here we recognize that the inner sum can arise from an expansion of $\prod_{i=1}^{j}[N(\mathrm{d}v_i) - \mu(\mathrm{d}v_i)]$, the symmetry of the densities $c_{[j]}(\cdot)$ implying equality of their integrals with respect to any reordering of the indices in a differential expansion such as $N(\mathrm{d}v_1)\cdots N(\mathrm{d}v_i)\,\mathrm{d}v_{i+1}\cdots\mathrm{d}v_j$. Inserting this product form leads to (7.1.9) and (7.1.10).

Verification of equations (7.1.11a) and (7.1.11b) under the null hypothesis is straightforward. □

EXAMPLE 7.1(g) *Poisson cluster processes with bounded cluster size.* Suppose the size of the clusters is limited to M so that only the first M terms are present in the expansions in terms of Khinchin or cumulant densities; the Gauss–Poisson case of Example 7.1(e) corresponds to $M = 2$. Then, for $\theta > 0$, we may define the process as the superposition of a stationary Poisson process with parameter μ and a Poisson cluster process with clusters of size $2, \ldots, M$ with Khinchin measures with densities $\theta k_j(y_1, \ldots, y_j)$ taken from the p.g.fl. representation (6.3.32) (i.e. k_j is the density of the measure K_j there). Then, the Khinchin densities in the resultant process have the form (identifying the state space $\mathcal{X}$ with the set A)

$$K_0(\theta) = \theta\ell(A) + \theta\sum_{j=1}^{M}\frac{1}{j!}\int_A\cdots\int_A k_j(x_1,\ldots,x_j)\,\mathrm{d}x_1\cdots\mathrm{d}x_j,$$

$$k_1(x;\theta) = \mu + \theta k_1(x), \qquad k_j(x_1,\ldots,x_j) = \theta k_j(x_1,\ldots,x_j) \quad (j = 2,\ldots,M).$$

From (7.1.21), we have the expansion

$$\begin{aligned}\left.\frac{\partial \log L}{\partial\theta}\right|_{\theta=0} &= \sum_{j=1}^{M}\frac{1}{\mu^j}\int_A\cdots\int_A\frac{I(y_1,\ldots,y_j)\,k_j(y_1,\ldots,y_j)}{j!}\,N(\mathrm{d}y_1)\cdots N(\mathrm{d}y_j)\\ &= \sum_{j=1}^{M}\frac{1}{\mu^j}\sum{}^{*}k_j(x_{r_1},\ldots x_{r_j}).\end{aligned}$$

This expression exhibits the efficient score $\partial\log L/\partial\theta\big|_{\theta=0}$ as the sum of first-, second-, …, Mth-order statistics in the observed points $x_1,\ldots,x_N$. In the Gauss–Poisson case, only the first- and second-order terms are needed.

The derivation here implies that the form of the cluster process, up to and including the detailed specification of the K_j, is known *a priori*. The situation if the structure is not known is much more complex but would in effect involve taking a supremum over an appropriate family of functions K_j.

An alternative representation is available through (7.1.9) and (7.1.10). This has the advantage that the cumulant densities can be specified globally so that no implicit assumptions about boundary effects are needed. It follows from (6.3.32) (see Exercise 6.3.17) that only the first M factorial cumulant densities $c_{[j]}$ need be considered and (since the $c_{[j]}$ are derived from linear combinations

of the k_j) that the same kind of structure holds for the $c_{[j]}$, namely

$$c_{[1]}(x;\theta) = \mu + \theta c_{[1]}(x),$$
$$c_{[j]}(x_1,\ldots,x_j;\theta) = \theta c_{[j]}(x_1,\ldots,x_j) \qquad (j=2,\ldots,M).$$

Then (7.1.9) leads to a similar expansion in terms of linear, quadratic, ... statistics, namely

$$D_k = \frac{1}{k!\,\mu^k}\int_A \cdots \int_A I(y_1,\ldots,y_k)\, c_{[k]}(y_1,\ldots,y_k)\, Z(\mathrm{d}y_1)\cdots Z(\mathrm{d}y_k).$$

For further examples, asymptotic behaviour in the stationary case, and the possibility of representing the D_k in terms of spectral measures, see Davies (1977) and Exercises 7.1.8–10. □

Exercises and Complements to Section 7.1

7.1.1 Let N_1, N_2 be two finite Poisson processes with intensity measures Λ_1, Λ_2, respectively. Show that $N_1 \ll N_2$ if and only if $\Lambda_1 \ll \Lambda_2$ (see above Proposition 7.1.III for $N_1 \ll N_2$).

7.1.2 Exercise 2.1.9 discusses the likelihood of a *cyclic Poisson process* with rate parameter

$$\mu(t) = \exp[\alpha + \beta\sin(\omega_0 t + \theta)],$$

though the parametric form is different: e^{α} here equals $\lambda/I_0(\kappa)$ there. The derivation of maximum likelihood estimators given there assumes ω_0 is known; here we extend the discussion to the case where ω_0 is unknown.

(a) Show that the supremum of the likelihood function in general is approached by a sequence of arbitrarily large values of ω_0 for which $\sin\omega_0 t_i \approx$ constant and $\cos\omega_0 t_i \approx$ constant for every t_i of a given realization. A global maximum of the likelihood is attainable if the parameters are constrained to a compact set.

(b) Suppose the observation interval $T \to \infty$, and constrain ω_0 to an interval $[0,\omega_T]$, where $\omega_T/T^{1-\epsilon} \to 0$ $(T\to\infty)$ for some $\epsilon > 0$. Then, the sequence of estimators $\widehat{\omega}_0(T)$ is consistent. [See Vere-Jones (1982) for details.]

7.1.3 Another cyclic Poisson process model assumes $\mu(t) = \alpha + \beta[1+\sin(\omega_0 t+\theta)]$. Investigate maximum likelihood estimators for the parameters [see earlier references and Chapter 4 of Kutoyants (1980, 1984)].

7.1.4 Suppose that the density $\mu(\cdot)$ of an inhomogeneous Poisson process on the bounded Borel set A such as the unit interval (or rectangle or cuboid, etc.) can be expanded as a finite series of polynomials orthogonal with respect to some weight function $w(\cdot)$ so that

$$\mu(x) = \alpha w(x)\left(1 + \sum_{j=1}^{r}\beta_j v_j(x)\right) \equiv \alpha w(x)\psi(x),$$

where $\int_A w(x)\,\mathrm{d}x = 1$, $\int_A w(x)v_j(x)\,\mathrm{d}x = 0$, $\int_A w(x)v_j(x)v_k(x)\,\mathrm{d}x = \delta_{jk}$ $(j,k = 1,\ldots,r)$. Show that the problem of maximizing the log likelihood ratio $\log(L/L_0)$, where L_0 refers to a Poisson process with density $w(x)$, is equivalent to the problem of maximizing $\sum_{i=1}^N \log\psi(x_i)$ subject to the constraint that $\psi(x) \geq 0$ on A. This maximization has to be done numerically; the main difficulty arises from the nonnegativity constraint.

7.1.5 Use the relations in equation (5.5.31) between the Janossy and Khinchin densities to provide a representation of the likelihood of a Poisson cluster process in terms of the Janossy densities of the cluster member process.
[*Hint*: Suppose first that the process is a.s. totally finite. Expand $\log G[h] = \int_{\mathcal{X}}(G[h \mid y] - 1)\,\mu_c(\mathrm{d}y)$ $(h \in \mathcal{V}(\mathcal{X}))$ and obtain

$$k_n(x_1,\ldots,x_n) = \int_{\mathcal{X}} j_n(x_1,\ldots,x_n \mid y)\,\mu_c(\mathrm{d}y).$$

In the general case, proceed from the p.g.fl. expansion of the local process on A as in (5.5.14) and (5.5.15).]

7.1.6 (*Continuation*). When the cluster structure is that of a stationary Neyman–Scott process with $\mu_c(\mathrm{d}y) = \mu_c\,\mathrm{d}y$ as in Example 6.3(a) so that

$$G[h \mid y] = \sum_{j=0}^{\infty} p_j\left(\int_{\mathcal{X}} h(y+u)\,F(\mathrm{d}u)\right)^j \equiv Q\left(\int h(y+u)f(u)\,\mathrm{d}u\right), \quad \text{say,}$$

deduce that the Janossy densities for the local process on A are given by

$$j_n(x_1,\ldots,x_n \mid A) = \exp\left(\mu_c\int_{\mathcal{X}}[Q(1-F(A-y))-1]\,\mathrm{d}y\right)$$
$$\times \sum_{b\in\mathcal{B}_{01}}\prod_{i=1}^{2^n-1}\left[\mu_c\int_{\mathcal{X}} Q^{(|a_i|)}(1-F(A-y))\prod_{j=1}^{n}[f(x_j-y)]^{a_{ij}}\,\mathrm{d}y\right]^{b(a_i)},$$

where $a_i = (a_{i1},\ldots,a_{in})$ is the binary expansion of $i = 1,\ldots,2^n-1$, $|a_i| = \#\{j\colon a_{ij} = 1\}$, and $\mathcal{B}_{01}$ is the class of all $\{0,1\}$-valued functions $b(\cdot)$ defined on $\{a_i\colon i = 1,\ldots,2^n-1\}$ such that $\sum_i b(a_i)a_i = (1,\ldots,1)$. [Thus, any $b(\cdot)$ has $b(a) = 0$ except for at most n subsets of a partition of $\{1,\ldots,n\}$, and $\sum_b\prod_i$ is here equivalent to $\sum_j\sum_{\mathcal{T}}\prod_x$ in (5.5.31). Baudin (1981) used a combinatorial lemma in Ammann and Thall (1979) to deduce the expression above and commented on the impracticality of its use for even a moderate number of points!]

7.1.7 Suppose that for each n the function $U \equiv U_n$ of (7.1.4) satisfies $U_n(x_1,\ldots,x_n) \geq -cn$ for some finite positive constant c. Show that a distribution is well defined (i.e. that a finite normalizing constant exists).

7.1.8 *Clustered version of the Strauss process.* In the basic Strauss model of Example 7.1(c), if $\beta > 1$, the Janossy densities, and hence also their integrals over the observation region, will tend to increase as the number of points in

the region increases. Suppose that the densities are taken proportional to $w_n\alpha^n\beta^{m(n)}$, where $m(n)$ is as defined in the example. Then, the integrals are dominated by the quantities $Cw_n\alpha^n\beta^{n(n-1)}$, and a sufficient condition for the process to be well defined is that

$$\sum w_n\alpha^n\beta^{n(n-1)} < \infty.$$

Show that this condition is *not* satisfied if $w_n \equiv 1$, and investigate conditions on the w_n to make it hold. Note that such modifications will not affect the sampling patterns for fixed n but only the probabilities p_n controlling the relative frequency of patterns with different numbers of events. See Kelly and Ripley (1976) for further discussion.

7.1.9 (a) For a stationary Gauss–Poisson process [see Example 7.1(e)] for which $c_{[1]}(u) = \mu + \theta$ and $c_{[2]}(u, v) = \theta\gamma(u - v)$ for some symmetric p.d.f. $\gamma(\cdot)$ representing the distribution of the signed distance between the points of a two-point cluster, show that its efficient score statistic D (see Proposition 7.1.IV) is expressible as $D = D_1 + D_2$, where

$$D_1 = N(A) - \mu\ell(A) \equiv Z(A),$$
$$D_2 = \int_A \int_A \gamma(x - y)\, Z(\mathrm{d}x)\, Z(\mathrm{d}y).$$

(b) In practice, $\hat{\mu}$ is estimated by $N(A)/\ell(A)$, so D_1 vanishes, and in the second term, Z is replaced by $\widehat{Z}(\cdot) = N(\cdot) - \hat{\mu}\ell(\cdot)$. Davies (1977) shows that the asymptotic results remain valid with this modification, so the efficiency of other second-order statistics can be compared with the locally optimum form D_2. Write the variance estimator in the form

$$(r-1)\sum_{j=1}^{r}[N(\Delta_j) - \hat{\mu}\ell(\Delta_j)]^2,$$

where $\Delta_1 \cup \cdots \cup \Delta_r$ is a partition of the observation region A into subregions of equal Lebesgue measure, in a form similar to D_2, and investigate the variance-to-mean ratio as a test for the Gauss–Poisson alternative to a Poisson process. [Davies suggested that the asymptotic local efficiency is bounded by $\frac{2}{3}$.]

7.1.10 (*Continuation*). In the case of a Neyman–Scott process with Poisson cluster size distribution, all terms D_k in the expansion in (7.1.9) are present, and D_2 dominates D only if the cluster dimensions are small compared with the mean distance between cluster centres.

7.1.11 When the Poisson cluster process of Example 7.1(g) for $\mathcal{X} = \mathbb{R}$ is stationary and $A = (0, t]$,

$$D_j \approx \frac{1}{t^{j+1} j!\, \mu^j} \sum_{l_1+\cdots+l_j=0} \cdots \sum \phi_j(l_1/t, \ldots, l_j/t)\, g_j(\lambda_1, \ldots, \lambda_j;\, t),$$

where

$$\phi_j(\lambda_1,\dots,\lambda_j) = \int_{\mathbb{R}}\cdots\int_{\mathbb{R}} k_j'(t_1,\dots,t_j)\exp\left(2\pi i\sum_{r=1}^{j}\lambda_r t_r\right)\mathrm{d}t_2\cdots\mathrm{d}t_j$$

with $\lambda_1+\cdots+\lambda_j=0$ and $g_j(\lambda_1,\dots,\lambda_j;\,t)$ equals

$$\int_0^t\cdots\int_0^t I(t_1,\dots,t_j)\exp\left(2\pi i\sum_{r=1}^{j}\lambda_r t_r\right)Z(\mathrm{d}t_1)\cdots Z(\mathrm{d}t_j).$$

[*Hint*: Use Parseval-type relations to show that $t^{-1}\mathrm{E}(|\overline{D}_j - D_j|^2)\to 0$ as $t\to\infty$. See also Theorem 3.1 of Davies (1977).]

7.2. Conditional Intensities, Likelihoods, and Compensators

If the discussion in the previous section suggests that there are no easy methods for evaluating point process likelihoods on general spaces, it is all the more remarkable, and fortunate, that in the special and important case $\mathcal{X}=\mathbb{R}$ there is available an alternative approach of considerable power and generality. The essence of this approach is the use of a causal description of the process through successive conditionings. A full development of this approach is deferred to Chapter 14; here we seek to provide an introduction to the topic and to establish its links to representations in terms of Janossy densities. For simplicity, suppose observation of the process occurs over the time interval $A=[0,T]$ so that results may be described in terms of a point process on $\mathbb{R}_+$.

Denote by $\{t_1,\dots,t_{N(T)}\}$ the ordered set of points occurring in the fixed interval $(0,T)$. As in the discussion around equation (3.1.8), the t_i, as well as the intervals $\tau_i = t_i - t_{i-1}$, $i\ge 1$, $t_0=0$, are taken to be well-defined random variables. Suppose also that the point process is regular on $(0,T)$, so that the Janossy densities $j_k(\cdot)$ all exist (recall Definition 7.1.I). We suppose that if there is any dependence on events before $t=0$, it is already incorporated into the Janossy densities. For ease of writing, we use $j_n(t_1,\dots,t_n \mid u)$ for the local Janossy density on the interval $(0,u)$, and $J_0(u)$ for $J_0((0,u))$.

Now introduce the conditional survivor functions $S_k(u \mid t_1,\dots,t_{k-1}) = \Pr\{\tau_k > u \mid t_1,\dots,t_{k-1}\}$ and observe that these can be represented recursively in terms of the (local) Janossy functions through the equations

$$\begin{aligned}
S_1(u) &= J_0(u) && (0<u<T),\\
S_2(u\mid t_1)p_1(t_1) &= j_1(t_1 \mid t_1+u) && (0<t_t<t_1+u<T),\\
S_3(u\mid t_1,t_2)p_2(t_2\mid t_1) &= j_2(t_1,t_2\mid t_2+u) && (0<t_1<t_2<t_2+u<T),
\end{aligned}$$

and so on, where $p_1(t), p_2(t\mid t_1),\dots$ are the probability densities corresponding to the survivor functions $S_1(u), S_2(u\mid t_1),\dots$. The fact that these densities exist is a corollary of the assumed regularity of the process. This can be

seen more explicitly by noting identities such as (for $S_1(\cdot)$)

$$J_0(t) = J_0(T) + \sum_{k=1}^{\infty} \frac{1}{k!} \int_t^T \cdots \int_t^T j_k(u_1, \ldots, u_k \mid T)\, \mathrm{d}u_1 \cdots \mathrm{d}u_k,$$

from which

$$p_1(t) = j_1(t \mid T) + \sum_{k=2}^{\infty} \frac{1}{(k-1)!} \int_t^T \cdots \int_t^T j_k(t, u_2, \ldots, u_k \mid T)\, \mathrm{d}u_2 \cdots \mathrm{d}u_k,$$

an expression that is actually independent of T for $T > t$. Similarly, for S_2 we find (for $t_1 < t < T$)

$$\begin{aligned} p_1(t_1) S_2(t \mid t_1) & \\ &= j_1(t_1 \mid T) + \sum_{k=2}^{\infty} \frac{1}{(k-1)!} \int_t^T \cdots \int_t^T j_k(t_1, u_2, \ldots, u_k \mid T)\, \mathrm{d}u_2 \cdots \mathrm{d}u_k \\ &= j_1(t_1 \mid t), \end{aligned}$$

from which it follows that $p_1(t_1)p_2(t \mid t_1)$ equals

$$j_2(t_1, t \mid T) + \sum_{k=3}^{\infty} \frac{1}{(k-2)!} \int_t^T \cdots \int_t^T j_k(t_1, t, u_3, \ldots, u_k \mid T)\, \mathrm{d}u_3 \cdots \mathrm{d}u_k,$$

again establishing the absolute continuity of $S_2(t \mid t_1)$. Further results follow by an inductive argument, the details of which we leave to the reader. Together they suffice to establish the first part of the following proposition.

Proposition 7.2.I. *For a regular point process on $\mathcal{X} = \mathbb{R}_+$, there exists a uniquely determined family of conditional probability density functions $p_n(t \mid t_1, \ldots, t_{n-1})$ and associated survivor functions*

$$S_n(t \mid t_1, \ldots, t_{n-1}) = 1 - \int_{t_{n-1}}^{t} p_n(u \mid t_1, \ldots, t_{n-1})\, \mathrm{d}u \qquad (t > t_{n-1})$$

defined on $0 < t_1 < \cdots < t_{n-1} < t$ such that each $p_n(\cdot \mid t_1, \ldots, t_{n-1})$ has support carried by the half-line (t_{n-1}, ∞), and for all $n \geq 1$ and all finite intervals $[0, T]$ with $T > 0$,

$$J_0(T) = S_1(T), \tag{7.2.1a}$$

$$\begin{aligned} j_n(t_1, \ldots, t_n \mid T) &\equiv j_n(t_1, \ldots, t_n \mid (0, T)) \\ &= p_1(t_1) p_2(t_2 \mid t_1) \cdots p_n(t_n \mid t_1, \ldots, t_{n-1}) \\ &\qquad \times S_{n+1}(T \mid t_1, \ldots, t_n), \end{aligned} \tag{7.2.1b}$$

where $0 < t_1 < \cdots < t_n < T$ can be regarded as the order statistics of the points of a realization of the point process on $[0, T]$. Conversely, given any such family of conditional densities for all $t > 0$, equations (7.2.1a) *and* (7.2.1b) *specify uniquely the distribution of a regular point process on $\mathbb{R}_+$.*

PROOF. Only the converse requires a brief comment. Given a family of conditional densities p_n, both $J_0(T)$ and symmetric densities $j_k(\cdot \mid T)$ can be defined by (7.2.1), and we can verify that they satisfy

$$J_0(T) + \sum_{n=1}^{\infty} \frac{1}{n!} \int_0^T \cdots \int_0^T j_n(t_1, \ldots, t_n \mid T)\, \mathrm{d}t_1 \cdots \mathrm{d}t_n$$
$$= J_0(T) + \sum_{n=1}^{\infty} \int\cdots\int_{0<t_1<\cdots<t_n<T} j_n(t_1, \ldots, t_n \mid T)\, \mathrm{d}t_1 \cdots \mathrm{d}t_n = 1.$$

It follows from Proposition 5.3.II that there exists a well-defined point process with these densities.

Since the point process is uniquely determined by the Janossy measures and these are equivalent to the conditional densities $p_n(t \mid t_1, \ldots, t_{n-1})$ for a regular point process, there is a one-to-one correspondence between regular point processes and families $p_n(\cdot \mid \cdot)$, as described. □

We now make a seemingly innocuous but critical shift of view. Instead of specifying the conditional densities $p_n(\cdot \mid \cdot)$ directly, we express them in terms of their hazard functions

$$h_n(t \mid t_1, \ldots, t_{n-1}) = \frac{p_n(t \mid t_1, \ldots, t_{n-1})}{S_n(t \mid t_1, \ldots, t_{n-1})}$$

so that

$$p_n(t \mid t_1, \ldots, t_{n-1}) = h_n(t \mid t_1, \ldots, t_{n-1}) \exp\left(-\int_{t_{n-1}}^{t} h_n(u \mid t_1, \ldots, t_{n-1})\, \mathrm{d}u\right). \tag{7.2.2}$$

Given a sequence $\{t_i\}$ with $0 < t_1 < \cdots < t_n < \cdots$, we define an amalgam of the hazard functions by

$$\lambda^*(t) = \begin{cases} h_1(t) & (0 < t \le t_1), \\ h_n(t \mid t_1, \ldots, t_{n-1}) & (t_{n-1} < t \le t_n,\ n \ge 2). \end{cases} \tag{7.2.3}$$

Definition 7.2.II. *The* conditional intensity function *for a regular point process on* $\mathbb{R}_+ = [0, \infty)$ *is the representative function* $\lambda^*(\cdot)$ *defined piecewise by* (7.2.3).

Note on terminology. In the general definition of *conditional intensities*, care must be taken to specify the information on which the conditioning is based. This is conveniently summarized by a σ-algebra of events. In the conditional intensity defined above, the conditioning is taken with respect to the minimal σ-algebra consistent with observations on the process, namely the σ-algebra generated by the observed past of the process. More general versions may include information about exogenous variables or processes, as illustrated around Examples 7.2(d)–(e). The conditional intensity introduced here follows the terminology of Brémaud (1981) and related references in the

electrical engineering literature; it should be carefully distinguished from the conditional intensity used in more recent discussions of spatial point patterns (see e.g. Baddeley and Turner, 2000), where it is a special case of the Papangelou intensity introduced in Chapter 15. This Papangelou conditional intensity relates to the effect of adding an additional point *within* the observation region; Definition 7.2.II refers to adding an additional point within *an extension* of the observation region.

The intuitive content of the notion of a conditional intensity function is well expressed through the suggestive relation

$$\lambda^*(t)\,\mathrm{d}t \approx \mathrm{E}[N(\mathrm{d}t) \mid \mathcal{H}_{t-}], \tag{7.2.3′}$$

where $\mathcal{H}_{t-}$ is the σ-algebra of events occurring at times up to but not including t. Thus, the conditional intensity can be interpreted as the conditional risk of the occurrence of an event at t, given the realization of the process over the interval $[0, t)$. Strictly, the notation should reflect the fact that $\lambda^*(\cdot)$ is a function $\lambda^*(\cdot \mid t_1, \ldots, t_{N(t)})$ of the point history, or, even more generally, that it is itself a stochastic process $\lambda^*(t, \omega)$ depending on ω through the realization $\{t_1(\omega), \ldots, t_N(\omega)\}$ of the history up to time t. The terms *conditional risk* (or *rate* or *hazard*) *function*, or even these terms omitting the word 'conditional', have also been used to describe $\lambda^*(\cdot)$ as defined in (7.2.3). It is the key both to the likelihood analysis and to solving problems of prediction, filtering, and simulating point processes on a half-line.

Just as the density function of a probability distribution can in principle be specified only up to its values on a set of Lebesgue measure zero, so also a lack of uniqueness arises in defining $\lambda^*(\cdot)$. In all practical situations, the densities $p_n(\cdot \mid \cdot)$ will be at least piecewise continuous, and uniqueness can then be ensured by (for example) taking the left-continuous modification $\lambda^*(t-)$ for $\lambda^*(t)$. The reason for using *left* continuity is connected with predictability: if the conditional intensity has a discontinuity at a point of the process, then its value at that point should be defined by the history before that point, not by what happens at the point itself. This is implicit in the way the hazard functions are defined and crucial to the correct definition of the likelihood, since it is the density for the interval preceding a point that figures in the likelihood, not the new density that comes into play once the point has occurred. A rigorous discussion of these issues leads to the concept of a predictable σ-algebra and to the existence of predictable versions of the conditional intensity; see comments later in this chapter and Chapter 14.

In the remainder of this section, unless stated otherwise, it is tacitly assumed that a left-continuous version of $\lambda^*(\cdot)$ exists and is being used.

Proposition 7.2.III. *Let N be a regular point process on $[0, T]$ for some finite positive T, and let $t_1, \ldots, t_{N(T)}$ denote a realization of N over $[0, T]$. Then, the likelihood L of such N is expressible in the form*

$$L = \left[\prod_{i=1}^{N(T)} \lambda^*(t_i)\right] \exp\left(-\int_0^T \lambda^*(u)\,\mathrm{d}u\right), \tag{7.2.4}$$

and its log likelihood ratio relative to the Poisson process on $[0,T]$ *with constant rate* 1 *is expressible as*

$$\log \frac{L}{L_0} = \sum_{i=1}^{N(T)} \log \lambda^*(t_i) - \int_0^T [\lambda^*(u) - 1]\,\mathrm{d}u. \tag{7.2.5}$$

PROOF. To establish (7.2.4), it is enough to express the Janossy densities in terms of the conditional densities $p_n(t \mid t_1, \ldots, t_{n-1})$ and then express each of these in terms of their hazard functions and hence of $\lambda^*(\cdot)$. Details are left to the reader: see Exercise 7.2.1. □

An important consequence of the construction used in the proof above is that the conditional intensity function determines the family of conditional hazard functions at (7.2.3) and that these in turn determine the Janossy densities. This can be summarized as below.

Proposition 7.2.IV. *Let* N *be a regular point process on* $\mathbb{R}_+$. *Then, the conditional intensity function determines the probability structure of the point process uniquely.*

Our first example illustrates these ideas in the context of a Wold process.

EXAMPLE 7.2(a) *Wold process of correlated intervals* (see Section 4.5). Suppose the Markov process of successive interval lengths $\{I_n\} \equiv \{t_n - t_{n-1}\}$ $(n = 1, 2, \ldots)$, with $t_0 \equiv 0$, is governed by the transition kernel with density $p(y \mid x)$ for the length y of the interval I_n given the length x of the interval I_{n-1}. For $n \geq 3$, the conditional distribution has the density

$$p_n(t \mid t_1, \ldots, t_{n-1}) = p(t - t_{n-1} \mid t_{n-1} - t_{n-2}),$$

so that in terms of the hazard function $h(y \mid x) = p(y \mid x)/S(y \mid x)$, where $S(y \mid x) = 1 - \int_0^y p(u \mid x)\,\mathrm{d}u$, we have

$$\lambda^*(t) = h(t - t_{N(t)} \mid t_{N(t)} - t_{N(t)-1}).$$

Here, $t_{N(t)}$ and $t_{N(t)-1}$ are the first and second points to the left of t, and it is assumed that $N(t) \geq 2$. To specify $\lambda^*(\cdot)$ at the beginning of the observation period (i.e. in $\{t > 0 : N(t) \leq 1\}$), some further description of the initial conditions is needed. If observations are started from an event of the process as origin, it is enough to be given the distribution of the initial interval $(0, t_1)$ [e.g. it may be the stationary density $\pi(\cdot)$ satisfying $\pi(y) = \int_0^\infty p(y \mid x)\,\pi(x)\,\mathrm{d}x$, if such $\pi(\cdot)$ exists]. Otherwise, the length of the interval terminating at t_1 may be an additional parameter in the likelihood and we may seek to estimate it, or we may impose further description of both the interval terminating at t_1 and the interval (t_1, t_2). See Exercise 7.2.3 for a particular case. □

EXAMPLE 7.2(b) *Hawkes process* [continued from Example 6.3(c)]. Suppose that the infectivity measure $\mu(\mathrm{d}x)$ has a density $\mu(\mathrm{d}x) = \mu(x)\,\mathrm{d}x$, say. Then,

each event at $t_i < t$ contributes an amount $\mu(t - t_i)$ to the risk at t. There is also a risk, λ say, of a new arrival at t. Assuming no contributions to the risk from the negative half-line, $\lambda^*(\cdot)$ is expressible in the simple form

$$\lambda^*(t) = \lambda + \sum_{0<t_i<t} \mu(t - t_i) = \lambda + \int_0^t \mu(t - u)\, N(\mathrm{d}u). \tag{7.2.6}$$

In applications, it is desirable to give $\mu(\cdot)$ some convenient parametric form. Ogata and Akaike (1982) and Vere-Jones and Ozaki (1982) discuss likelihood estimation for this process using a parametrization of the form

$$\mu(t) = \begin{cases} \sum_{k=0}^K b_k L_k(t) \mathrm{e}^{-\alpha t} & (t > 0), \\ 0 & (t \le 0), \end{cases} \tag{7.2.7}$$

where the functions $L_k(t)$ are Laguerre polynomials defined on $t > 0$; detailed computations are given in the quoted papers. Combinations of exponential terms with different decay parameters could also be considered, but pragmatic problems of estimability arise: even estimating α in (7.2.7) can be difficult.

Initial conditions also pose a problem. It is simplest to suppose that $\lambda^*(0) = 0$ so that any influence from the past is excluded. If this is not the case, then it may be possible to condition on information prior to time $t = 0$; in the technical language of Chapter 14, this means passing from the internal history to a more general *intrinsic* history. If neither of these options is available, then we are faced with a minor version of the Ising problem, as discussed around Examples 7.1(b) and 7.1(e). In principle, we should take the joint distribution of the observations (on $(0, T)$, say) and the entire past and then average over all possible past histories. In simple cases, this may be explicitly possible. For example, if $K = 0$ in (7.2.7), any contribution from events before $t = 0$ decays exponentially at the uniform rate $\exp(-\alpha t)$, and in fact the whole process $\lambda^*(t)$ is Markovian. In the equilibrium case, we can then integrate over the equilibrium distribution of $\lambda^*(0)$ to obtain the appropriate averaged likelihood. Further details on this special case are given in Exercise 7.2.5.

If we assume that $\nu = \int_0^\infty \mu(x)\, \mathrm{d}x < 1$ so that a unique stationary process exists [see Example 6.3(c)], it can be shown that the process converges toward equilibrium as $t \to \infty$ (see Chapter 13). In this case, the conditional intensity approaches the *complete intensity function* $\lambda^\dagger(t)$, which is the analogue of $\lambda^*(t)$ for the process defined on $\mathbb{R}$ and not merely on $\mathbb{R}_+$; that is, events of the process are no longer confined to $t > 0$. Equation (7.2.6) is then replaced by

$$\lambda^\dagger(t) = \lambda + \int_{-\infty}^t \mu(t - u)\, N(\mathrm{d}u).$$

This linear form also arises from second-order theory and suggests that for this example the optimal (least squares) linear predictor coincides with the optimal nonlinear predictor, at least as far as the immediate future is concerned. For further discussion of this issue, see Example 8.5(d). □

Note that in this and similar examples, finding the initial conditions required to make the ensuing process stationary resolves for such a process the problem described in the previous section of expressing the local Janossy densities in terms of the global process. In a one-dimensional point process observed over a finite interval, boundary effects can arise only at the two ends of the interval, while the causal character of the time dimension implies that there are no backward effects from points occurring later than the end of the observation interval. For a stationary point process in time, therefore, the only issue to be resolved is finding the right initial conditions to ensure that the resulting process is stationary.

The form (7.2.7) taken with (7.2.6) gives an example of a *linearly parameterized intensity*. The general usefulness of this model suggests that, in practical applications, it may be more convenient to choose a flexible family of models that are readily amenable to processing in much the same way that ARMA models can be used in conventional time series analysis rather than seeking the conditional intensity of a model that is given *a priori*. To this end, we look for examples in which the conditional intensity has a convenient parametric form. Two broad classes of such models are described below.

EXAMPLE 7.2(c) *Processes with linear or log-linear conditional intensity functions.* The assumption in these models is that the conditional intensity function can be written in one of the forms

$$\lambda^*(t) = \sum_k b_k Q_k^*(t), \tag{7.2.8}$$

$$\log \lambda^*(t) = \sum_k b_k R_k^*(t), \tag{7.2.9}$$

referred to as *linear* and *log-linear* forms, respectively, and where the Q_k^* and R_k^* are known functions. In these two cases, either the likelihood or the log likelihood is a convex function of the parameters so that, *if it exists, the maximum likelihood estimate of* λ^* *is unique* [see Exercise 7.2.6 and Ogata (1978)]. This property is of great importance when the model is highly parameterized; without some safeguard that guarantees convexity, the likelihood function may be extremely irregular, in which case convergence of numerical maximization routines is likely to be the exception rather than the rule.

The known functions $Q_k^*(\cdot)$ or $R_k^*(\cdot)$ may represent many types of dependency: trends or cyclic effects, linear or nonlinear dependence on the lengths of past intervals as in the Wold process, or linear dependence on the occurrence times of past events as in the Hawkes process. It must be admitted, however, that because of the inherent nonlinearity of the algebraic structure of a point process, there has not yet emerged for point processes a single class of parametric models of the same general utility as the ARMA models in conventional time series analysis. Further examples are given in Exercise 7.2.6 together with some indication of the numerical problems of estimation. For a more extended review, see Ogata and Katsura (1986); a deeper theoretical treatment is in Kutoyants (1984). □

So far, we have mainly assumed that the history controlling the conditional intensity is the history of the process itself (i.e. its 'internal history'), or in economics jargon that there are no exogenous variables that may influence the behaviour of the process. In many situations, this is not the case: to define the future progress of the process properly, the observations must include variables over and above the previous points of the process. In the previous example, one can well imagine that some of the terms in the linear combination might depend on external variables in addition to variables defined by the past points of the process itself. Likelihoods and predictions will then depend on just what information is in fact available. In the case of a Cox process, for example, prediction of the process takes on a very different character if the observations available to the predictor include knowledge of the random intensity function.

Ideas of this kind are developed in the *general theory of processes* (see Appendix A3.3 for a brief introduction and further references), in which a *history* (or *filtration*) for the process is defined as a nested, increasing family $\boldsymbol{\mathcal{H}}$ of σ-algebras $\mathcal{H}_t$ such that $N(t)$ is $\mathcal{H}_t$-measurable for all t. $N(t)$ is then said to be $\boldsymbol{\mathcal{H}}$*-adapted*. Conditional intensities can be found for any history of the process and will usually have different forms according to the history chosen.

In such a situation, the *full* likelihood of the process will cover the joint distributions of the point process and also of the additional variables that may influence the process through the dependence on past histories. Often, this is not available or is too complex to be used for practical inference or prediction. In such cases, some kind of *partial likelihood*, treating the observed values of explanatory variables as constants, may still be used for estimation purposes (see e.g. Cox, 1975). Such partial likelihoods have the same structural form as (7.2.4) provided the proper version of the conditional intensity (incorporating the new explanatory variables as they occur) is used.

In this context, where new explanatory variables may arise, it is helpful to view the basic form (7.2.4) as an extension of the likelihood for the Poisson process. Because of the complete independence property of the Poisson process, its likelihood corresponds to a continuous version of the multiplicative property for independent events: for example,

$$\Pr(A \cap B \cap C) = \Pr(A)\Pr(B)\Pr(C).$$

When the events are not independent, this can be replaced by the chain rule formula

$$\Pr(A \cap B \cap C) = \Pr(A)\,\Pr(B \mid A)\,\Pr(C \mid A \cap B),$$

which still represents the joint probability of the three events as a product. Equation (7.2.4), even in the form allowing general histories, can be regarded as an analogous extension of the original Poisson likelihood.

The situation is more transparent for processes in discrete time, as in the simple example below.

EXAMPLE 7.2(d) *Binary processes: discrete-time logistic regression model.* We consider a discrete-time process with realizations of the form $\{0, 0, 1, 0, 0, 0, 1, 1, 0, \dots\}$. In this context, the equivalent of an inhomogeneous Poisson process is a process with independent, nonidentical Bernoulli trials Y_i with success probabilities $p_i = \Pr\{Y_i = 1\}$. The likelihood of a realization $(Y_1, \dots, Y_n)$ with n trials can be written as

$$\log L(Y_1, \dots, Y_n\,;\, p_1, \dots, p_n) = \sum_{i:Y_i=1} \log \frac{p_i}{1-p_i} - \sum_{1}^{n} \log(1-p_i). \quad (7.2.10)$$

Now suppose that the Y_i are no longer independent but have probabilities $p_i^* = \Pr\{Y_i = 1 \mid Y_1, \dots, Y_{i-1}\}$, which can depend on the past history of the process. Then, by the same chain rule argument referred to earlier, (7.2.10) remains valid if the p_i are replaced by the p_i^*.

But there is no essential requirement here to restrict the conditioning to events defined on the previous values of the Y_i. We can add in dependence on additional past variables without affecting the validity of the chain rule formula. This is equivalent to extending the sequence of σ-algebras $\mathcal{H}_i$ (histories) to include all events generated by the relevant random variables before time i, including but not restricted to values of the sequence Y_i itself.

To take a more concrete example, the probabilities p_i^* might depend on the last few values of some explanatory variable U_i. This dependence might be modelled through a logistic regression, such as the explicit representation of $p_i^* = \mathrm{E}(Y_i \mid \mathcal{H}_i) = \mathrm{E}(Y_i \mid U_1, U_2, \dots)$ by an equation of the form

$$\log \frac{p_i^*}{1-p_i^*} = \alpha_0 + \sum_{j=1}^{r} \alpha_j U_{i-j}.$$

This is nothing other than the discrete-time version of a model with log-linear intensity, as described in Example 7.2(c), but with the explanatory variables now a selection of lagged versions of the external variables U_i. The art of the modeller here lies in constructing a form of dependence on the past that captures as much as possible of the true dynamics of the process being modelled. □

EXAMPLE 7.2(e) *Simple and modulated renewal process.* From Example 7.2(a) (see also Exercise 7.2.3), it follows that for a renewal process $N(t)$ denoting the number of renewals in $(0, t)$ and whose lifetime distribution has a hazard function $h(\cdot)$, the conditional intensity has the form $h(t - t_{N(t)})$.

Suppose that in addition to the renewal instants $\{t_i\}$ corresponding to the basic point process $N(t)$, we also observe a (vector) family of stochastic processes

$$\{X(t)\colon 0 < t < \infty\} \equiv \{X_1(t), \dots, X_k(t)\colon 0 < t < \infty\},$$

and suppose that as the defining history for the process we take the σ-algebras $\mathcal{F}_t$ of the form

$$\mathcal{F}_t = \mathcal{H}_t^N \vee \mathcal{H}_t^X,$$

thus combining the internal history of $\{N(t): 0 < t < \infty\}$ with that of $\{X(t): 0 < t < \infty\}$. Now suppose that the hazard function in successive intervals is modified in a multiplicative fashion by some nonnegative function $\psi(X_1(t), \ldots, X_k(t))$ of the current values of the $\{X_i(t)\}$; that is, we take

$$\lambda^*(t) = h(t - t_{N(t)})\, \psi(X_1(t), \ldots, X_k(t)).$$

Cox (1972a) posed the problem of estimating parameters $\beta_1, \ldots, \beta_k$ when $\psi(\cdot)$ has the log-linear form

$$\log \psi(X_1, \ldots, X_k) = \sum_{j=1}^{k} \beta_j X_j \,.$$

There is a close analogy with the problem of estimating the parameters in a model for lifetime distributions when the lifetimes of different individuals may be affected by different values of concomitant variables $X_1, \ldots, X_k$; this is the Cox regression model described in Cox (1972b) and now the subject of a considerable literature (see e.g. Aalen, 1975, 1978; Jacobsen, 1982; and Andersen *et al.*, 1993). Exercise 7.2.7 sketches a specific example. □

Example 7.2(f) *Processes with unpredictable marks* (see Definition 6.4.III). Conditional intensities for marked point processes will be considered more systematically in Section 7.3. In the special case of processes with unpredictable marks, however, the marks occur independently of the past of the process and can be treated as a sequence of independent random variables. Without necessarily assuming stationarity and supposing that the mark distribution at time t has density $f^*(\kappa \mid t)$, the conditional intensity factorizes into the form [see Lemma 7.3.V(iii)]

$$\lambda^*(t, \kappa) = \lambda^*_{\mathrm{g}}(t) f^*(\kappa \mid t).$$

Consequently, the log likelihood can be written as the sum of two terms $\log L = \log L_1 + \log L_2$, where

$$\log L_1 = \sum_{i=1}^{N_{\mathrm{g}}(T)} \log \lambda^*_{\mathrm{g}}(t_i) - \int_0^T \lambda^*_{\mathrm{g}}(u)\, \mathrm{d}u \tag{7.2.11a}$$

and

$$\log L_2 = \sum_{i=1}^{N_{\mathrm{g}}(T)} f^*(\kappa_i \mid t_i). \tag{7.2.11b}$$

The first term is in the standard form for a univariate point process on $(0, T)$ except for the fact that the ground intensity $\lambda^*_{\mathrm{g}}(t)$ may depend on the marks κ_i for events occurring before t as well as on the t_i themselves. In this sense, the ground process has the structure of a point process whose evolution depends on the evolution of a parallel, extrinsic process, namely the process of marks. The second term is the usual sum for a set of independent observations.

If the mark distribution has no parameters in common with the distribution of the ground process, then the two terms can be maximized separately and give the full likelihood estimates. If the marks are treated as a set of given values, about whose structure or distribution we have no information, then the first term could still be maximized as a partial likelihood.

Several of the simpler models for earthquake occurrence and neural impulses are of this form, where the size or strength of the event is treated as an independent mark, but can nevertheless influence the future evolution of the process. A typical example is the ETAS model [see Example 6.4(d) for notation and details], for which the conditional intensity of the ground process has the form

$$\lambda_{\mathrm{g}}^*(t) = \mu_c + D \sum_{i:t_i<t} \mathrm{e}^{\alpha(\kappa_i - \kappa_0)} \frac{1}{(c + t - t_i)^{1+p}} \,.$$

Here $D = AK$ is a constant that controls the criticality of the underlying branching process. This form can be substituted into L_1 above and used to evaluate the parameters μ_c, α, c, p and D without reference to the mark distribution. Conflicts will arise only if there is some departure from the assumption of unpredictable marks or if the mark distribution has some parameter in common with those specified above. See Example 7.3(c) for illustrations and further discussion. □

The stress-release model, considered below, is another example of this general type. It is an example also of a further class of models with the characteristic feature that the conditional intensity is governed by a Markov process that in general is only partially observable. The simplest examples of this type are doubly stochastic processes in which the underlying Markov process governs the stochastic intensity function. Here explicit expressions for the likelihood are not usually available, but an approach to likelihood estimation can nevertheless be made through adaptations of the Baum–Welch or E-M algorithms (see Exercise 7.2.8) or via the general filtering techniques discussed in Chapter 14. In the stress-release model, the occurrence times and marks of the events influence the Markov process itself so that the doubly stochastic character is lost, but in compensation the realization of the Markov process can be reconstructed from the data, given the model parameters and an initial value $X(0)$, so that an explicit form for the likelihood can be obtained.

Example 7.2(g) *Self-correcting or stress-release model.* The model was first investigated by Isham and Westcott (1979) as an example of a process that automatically corrects a deviation from its mean. Motivated by quite different applications in seismology, Knopoff (1971) and Vere-Jones (1978b) introduced essentially the same model as an elementary stochastic version of the so-called *elastic rebound theory* of earthquake formation in which context it has undergone substantial further study and elaboration (e.g. Ogata and Vere-Jones, 1984; Zheng, 1991; Zheng and Vere-Jones, 1994; Lu *et al.*, 1999; Bebbington

and Harte, 2001). Processes analogous to the stress-release model also arise in storage and insurance applications—wherever there is a process of steady accumulation and random release. Vere-Jones (1988) discusses an insurance interpretation.

The model is defined by an unobserved jump-type Markov chain $X(t)$ that increases linearly between events and decreases by a random amount (its mark) when an event occurs. Let the event times and associated marks be denoted by (t_i, κ_i), where it is supposed that the κ_i are nonnegative. Then, for $t \geq 0$, $X(t)$ has the representation

$$X(t) = X(0) + \nu t - \textstyle\sum_{i:0<t_i<t} \kappa_i \, .$$

Now suppose that the risk of an event occurring is an increasing function $\Psi(x)$ of the value x of $X(t)$. Given an initial value $X(0)$, and treating the κ_i as known quantities, the conditional intensity for the ground process (all events $\{t_i\}$) can be written

$$\lambda_{\mathrm{g}}^*(t) = \Psi[X(t)]. \tag{7.2.12}$$

One of the remarkable features of this process is that, apart from the value of $X(0)$, the conditional intensity is fully determined by the parameters of the model and the observations (t_i, κ_i). In other words, (7.2.12) is an $\boldsymbol{\mathcal{H}}$-intensity (internal intensity), in marked contrast to the doubly stochastic models, where one has to distinguish carefully between the internal intensity (conditioning on the observed event times and sizes only) and the intensity with respect to the full history (conditioning on both the events and the realization of the Markov process up to time t), and generally neither is very useful, the former being intractable and the latter inaccessible.

If (as is commonly the case) it is assumed that the event sizes form an i.i.d. sequence, the model again falls into the class of processes with unpredictable marks. The first term of the likelihood, (7.2.11a), is then sufficient to determine the parameter ν and any additional parameters arising in the specification of the function Ψ. In the particularly tractable special case where $\Psi(x) = \exp(\alpha + \rho x)$, the conditional intensity can then be represented in the log-linear form

$$\lambda_{\mathrm{g}}^*(t) = \exp\big(\alpha + \rho[X(0) + \nu t - \textstyle\sum_{i:0<t_i<t} \kappa_i]\big).$$

From this representation, it is immediately apparent that the parameter α is confounded with the initial value $X(0)$ and will not be separately estimable unless $X(0)$ is given. On the other hand, the sum $\alpha + \rho X(0)$ can be treated as a single unknown parameter, α^* say, which *is* estimable and is also sufficient to completely specify the conditional intensity $\lambda_{\mathrm{g}}^*(t)$, although not the process $X(t)$.

Conditions governing the existence of a stationary version of the process have been examined by Zheng (1991) following studies of special cases by Vere-Jones and Ogata (1984), Ogata and Vere-Jones (1984), Hayashi (1986), and Vere-Jones (1988). Assuming that the marks form an i.i.d. sequence with

finite mean μ and that the function Ψ is monotonically increasing, the essential condition (see Zheng, 1991, Proposition 4.3; Vere-Jones, 1988) is that

$$\lim_{x\to-\infty}\Psi(x) < \mu_c/\mu < \lim_{x\to+\infty}\Psi(x). \tag{7.2.13}$$

These two inequalities on μ_c/μ ensure that the process $X(t)$ drifts neither toward $-\infty$ nor toward $+\infty$. Some further properties are developed in Exercises 7.2.8–10. □

The integral of the conditional intensity function over time also plays an important role in the general theory. It is known as the *compensator* of the point process, relative to some given history $\boldsymbol{\mathcal{F}}$, on account of the following key property.

Lemma 7.2.V. *Suppose* $\{N(t)\colon 0 \le t < \infty\}$ *is adapted to the history* $\boldsymbol{\mathcal{F}}$ *and admits a left-continuous* $\boldsymbol{\mathcal{F}}$*-intensity* $\lambda^*(t)$*. Define* $\Lambda^*(t)$ *as the pointwise integral*

$$\Lambda^*(t) = \int_0^t \lambda^*(u)\,\mathrm{d}u.$$

Then, the process $M(t) = N(t) - \Lambda^*(t)$ *is an* $\boldsymbol{\mathcal{F}}$*-martingale: for every* $s > t > 0$*,*

$$\mathrm{E}[M(s) \mid \mathcal{F}_t] = M(t).$$

Proof. The idea behind the proof is simple. Consider the increment in the counting process $N(t)$ over an interval $(t, t+\Delta)$. We have approximately

$$\begin{aligned}\mathrm{E}\big([N(t+\Delta) - N(t)] - [\Lambda^*(t+\Delta) - \Lambda^*(t)] \mid \mathcal{H}_t\big) \\ &\simeq \mathrm{E}[N(t+\Delta) - N(t) \mid \mathcal{H}_t] - \lambda^*(t)\Delta \\ &\simeq \lambda^*(t) - \lambda^*(t) = 0.\end{aligned}$$

However, the simplicity of this argument is deceptive in that the identification $\mathrm{E}[N(\mathrm{d}t) \mid \mathcal{H}_t] = \lambda^*(t)\,\mathrm{d}t$ on which it depends, while intuitively clear, is tantamount to accepting the martingale property as a first premise. When $\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{H}}$, the internal history, the challenge is to derive this seemingly simple statement from the definition of the conditional intensity in terms of a family of hazard functions. Exercise 7.2.2 gives a simple special case.

A formal proof starts from the Doob–Meyer decomposition of a submartingale into an increasing, predictable part and a martingale (see Proposition A3.4.IX). The predictable part is identified with the compensator and shown to equal the integral of the conditional intensity function when such a function exists. See Chapter 14 for details. □

Lemma 7.2.V characterizes the compensator as the process that must be subtracted from the increasing process $N(t)$ to make it a martingale. It is increasing and, as holds for the conditional intensity, it is required to have a predictability property that in practice (at least when a conditional intensity exists) reduces to continuity. It increases continuously even though the process $N(t)$ is a step function with irregularly spaced steps.

By contrast, the martingale component includes jumps and is sometimes referred to as the *innovations process*. It may be compared with the Brownian motion term in a stochastic differential equation. However, it is only in very special situations (notably the Poisson process) that the innovations process for a point process has independent increments. In a renewal process, for example, the compensator is a sum of log survivor functions or, more generally, integrated hazard functions (IHFs) as in Section 4.6, and the martingale component consists of a combination of continuous segments, predictable when the last point is known, and unpredictable jumps (see Exercise 7.2.11).

Another remarkable property of the compensator is embodied in the random time-change theorem outlined in Section 7.4. It provides a far-reaching generalization of the assertion (see Exercise 2.4.4) that a nonstationary Poisson process can be transformed back into a stationary one by stretching the time axis, specifically by setting $\tau = \Lambda(t) = \int_0^t \lambda(u)\,\mathrm{d}u$.

Exercises and Complements to Section 7.2

7.2.1 Complete the details of the proof of Proposition 7.2.III.
[*Hint*: Use (7.2.1b), (7.2.2) and (7.2.3).]

7.2.2 Consider a one-point process with its point t_1 uniformly distributed over $(0, T)$ for some positive T. Show that the conditional intensity is given by

$$\lambda^*(t) = \begin{cases} 1/(T-t) & (0 < t \le t_1), \\ 0 & (t_1 < t \le T). \end{cases}$$

Find also the corresponding compensator $\Lambda^*(t)$ and check that $\mathrm{E}[\Lambda^*(t)] = t/T = \mathrm{E}[N(t)] < 1$ for $0 < t < T$.

7.2.3 (a) For a d.f. F with density f, write $h(x) = f(x)/\bar{F}(x)$ for its hazard function, where $\bar{F}(x) = 1 - F(x)$. Verify that a renewal process with lifetime d.f. F on $\mathbb{R}_+$, with realization $0 = t_0 < t_1 < \cdots < t_n < \cdots$ and $N(t) = \sup\{n: t_n < t\}$ (note that $N(t)$ is then left-continuous), has conditional intensity

$$\lambda^*(t) = h(t - t_{N(t)}) \tag{7.2.14}$$

and likelihood $f(t_1)\, f(t_2 - t_1) \cdots f(t_{N(t)} - t_{N(t)-1})\bar{F}(t - t_{N(t)})$ [see Example 5.3(b)].

(b) Now let $N(\cdot)$ denote the counting function on $\mathbb{R}_+$ of a delayed renewal process in which t_1 has d.f. G with density g and otherwise the lifetime d.f. is F with mean λ^{-1} as in (a). Show that $\lambda^*(t) = g(t)/\bar{G}(t)$ if $N(t) = 0$ and otherwise (7.2.14) holds, and that the likelihood function equals $\bar{G}(t)$ if $N(t) = 0$ and otherwise equals $g(t_1)(\prod_{i=1}^{N(t)-1} f(t_{i+1} - t_i))\bar{F}(t - t_{N(t)})$.

(c) For a *stationary* renewal process, put $g(t) = \lambda\bar{F}(t)$ in (b).

(d) Evaluate the expressions in (a) and (c) when (i) $F(x) = 1 - \mathrm{e}^{-\lambda x}$ $(x > 0)$; (ii) $F(x) = 1 - (1 + \lambda x)\mathrm{e}^{-\lambda x}$ $(x > 0)$.

7.2.4 Let $0 = t_0 < t_1 < \cdots$ be a realization on $(0, t]$ of the Wold process detailed in Exercise 4.5.8. Write down its likelihood function and its hazard function. Investigate both these functions when the process is stationary (so that then $t_0 < 0$ in general). See Lai (1978) for another example.

7.2.5 *Hawkes model with exponential decay.* Consider the model in (7.2.7) with $K = 0$, writing it in the form

$$\lambda^*(t) = \lambda + \nu \int_0^t \alpha \mathrm{e}^{-\alpha(t-u)}\, N(\mathrm{d}u) = \lambda + \nu\alpha \sum\nolimits_{t_i \le t} \mathrm{e}^{-\alpha(t-t_i)},$$

where $\nu = \int_0^\infty \mu(t)\,\mathrm{d}t$. Establish the properties below.

(i) The process $Y(t) = \int_0^t \mathrm{e}^{-\alpha(t-u)}\, N(\mathrm{d}u)$ is Markovian; hence also $\lambda^*(t) = \lambda + \nu\alpha Y(t)$, with infinitesimal transitions and rates

$$Y(t+\mathrm{d}t) = \begin{cases} Y(t)+1 & \text{with probability } [\lambda + \nu\alpha Y(t)]\,\mathrm{d}t, \\ (1-\alpha\,\mathrm{d}t)Y(t) & \text{with probability } 1 - [\lambda + \nu\alpha Y(t)]\,\mathrm{d}t. \end{cases}$$

(ii) The distribution function $F_t(y) = \Pr\{Y(t) \le y\}$ satisfies the forward Kolmogorov equation

$$\frac{\partial F_t(y)}{\partial t} = \alpha y \frac{\partial F_t(y)}{\partial y} - \int_{(y-1)_+}^{y} (\lambda + \nu\alpha u)\, F_t(\mathrm{d}u). \tag{7.2.15}$$

(iii) If $\nu < 1$, an equilibrium distribution exists, with density $\pi(x)$ say, that satisfies

$$\alpha y \pi(y) = \int_{(y-1)_+}^{y} (\lambda + \nu\alpha u)\pi(u)\,\mathrm{d}u,$$

for which $\pi(y) = \pi(1)\mathrm{e}^{\nu(y-1)} y^{(\lambda/\alpha)-1}$ for $0 < y < 1$ and for real $\theta \ge 0$,

$$\phi(\theta) \equiv \int_0^\infty \mathrm{e}^{-\theta y}\pi(y)\,\mathrm{d}y = \frac{\lambda}{\alpha} \int_{\exp(-\theta)}^{1} \frac{(1-w)\,\mathrm{d}w}{\nu(1-w) + \log w}\,.$$

(iv) The likelihood for a set of observations $0 < t_1 < \cdots < t_{N(T)}$ from the equilibrium process on $(0, T)$ is given by $\int_0^{t_1} L_y\, \pi(y)\,\mathrm{d}y$, where L_y is formed in the usual way from the modified conditional intensity

$$\lambda_y^*(t) = y\mathrm{e}^{-\alpha t} + \lambda + \nu \int_0^t \alpha \mathrm{e}^{-\alpha(t-u)}\, N(\mathrm{d}u).$$

7.2.6 (a) For each of the models implied by (7.2.8) and (7.2.9) with r parameters $b_1, \ldots, b_r$, check that

$$\sum_{j=1}^{r}\sum_{k=1}^{r} v_j v_k \frac{\partial^2 \log L}{\partial b_j\, \partial b_k} \le 0 \qquad (\text{all real } v_j, j = 1, \ldots, r).$$

Deduce that if a solution of the equations $\partial L/\partial b_j = 0$ $(j = 1, \ldots, r)$ is found, then it is unique.

(b) For the log-linear model, show that along any ray $\{(\rho b_1, \ldots, \rho b_k)\colon -\infty < \rho < \infty\}$, $\log L \to \infty$ for $|\rho| \to \infty$, so that a maximum on the ray exists, and hence a global maximum for $\log L$ exists.

[See Ogata and Vere-Jones (1984) for an example. In the linear model, there is no guarantee that with any parameters $\{\hat{b}_j\}$ so determined, the likelihood of any other set of observations will necessarily have positive likelihood, nor is it even necessarily the case that the intensity at every point in the realization is positive! In general, it is necessary to treat the problem as one of constrained optimization: see e.g. Ogata (1983) and the discussion by Berman (1983).]

7.2.7 *Poisson process in a random environment* [see Example 7.2(e)]. As a simple example of a modulated renewal process, suppose that the rate $\lambda(t)$ of a simple Poisson process takes different values $\lambda_1, \ldots, \lambda_K$ in response to environmental factors $X(t)$; thus, we can write

$$\lambda^*(t) = \sum_{k=1}^{K} \lambda_k I_{A_k}(X(t)),$$

where A_k denotes the range of values of $X(t)$ on which λ takes on the value λ_k. If $X(t)$ is an observed, continuous function of t but the λ_k are unknown parameters of the process, write down the likelihood conditional on a knowledge of $X(t)$ at time t. Hence, obtain an estimate of λ_k in terms of the proportion of time spent by $X(t)$ in A_k.

Is the result affected if instead of being an external variable, $X(t)$ is a function of the backward recurrence time (i.e. of the age of the 'component' in place at time t)?

7.2.8 *E–M algorithm applied to a Cox process with a Markovian rate function.* In contrast to the previous exercise, suppose that the process $X(t)$ governing the rate of occurrence of points is not observed but is known to be a continuous-time Markov chain with finite state space $\mathcal{K} = \{1, \ldots, K\}$ and Q-matrix $Q = \{q_{kl};\, k, l \in \mathcal{K}\}$, and that when $X(t) = k$, points occur according to a Poisson process with rate λ_k. The aim is to estimate the parameters q_{kl} and λ_k from observations on $N(\cdot)$ alone. Approximate the continuous-time process by a discrete skeleton $X(n\delta)$; then the resulting Markov chain has transition probabilities given approximately (for δ small) by $p_{kk} = 1 - q_{kk}\delta$, $p_{kl} = -q_{kl}\delta$, $k \neq l$. Observations on the process consist of the counts $Y_n = N(n\delta, (n+1)\delta]$, treated as Poisson or even binomial (presence or absence of points). Write down and implement iterative procedures for estimating the parameters of the discrete approximation, and hence of the underlying continuous process, using the E-M methodology.

[*Hint*: This example has been widely discussed in the literature on point process filtering and will be reviewed further in Chapter 14. Since the Markov process is unobserved, the example can be treated as a 'hidden Markov model' and is thus a natural candidate for analysis via the Baum–Welch and E–M algorithms—see Dempster *et al.* (1977), Elliott *et al.* (1995), and MacDonald and Zucchini (1997). The full likelihood is the likelihood for both the realization of the Markov chain and the observed counts; the restricted likelihood is the likelihood for the observed counts only, averaged over the possible realizations of the Markov chain. The references cited give general accounts of the form of the averaging (E-step) and estimation (M-step) procedures that can be employed to pass from the full to the restricted likelihoods and obtain the resulting estimates.]

7.2.9 *Stress-release model: Stationary behaviour.* In Example 7.2(g), let $F(x,t) = \Pr\{X(t) \le x\}$, $S(u) = \Pr\{\kappa > u\}$.

(i) Show, using the notation of the example, that the forward equations for the Markov process $X(t)$ take the form

$$\frac{\partial F}{\partial t} + \nu \frac{\partial F}{\partial x} = \int_x^\infty \Psi(y) S(y-x)\, F(\mathrm{d}y,\, t).$$

(ii) Deduce that, if it exists, the density $\pi(x)$ of the stationary distribution for $X(t)$ satisfies

$$\nu\pi(x) = \int_x^\infty \Psi(y) S(y-x) \pi(y)\, \mathrm{d}y,$$

and that its characteristic function $\varphi(s) = \int \mathrm{e}^{isx}\pi(x)\,\mathrm{d}x = \mathrm{E}(\mathrm{e}^{isX(\cdot)})$ satisfies

$$\varphi(s) = \gamma(s)\varphi_\Psi(s)$$

where, with $\mu = \mathrm{E}(\kappa)$, $\gamma(s)$ and $\varphi_\Psi(s)$ are the characteristic functions of the distributions with densities $S(x)/\mu$ and $(\mu/\nu)\Psi(x)\pi(x)$, respectively, and $\mathrm{E}[\Psi(X(t))] = \mu\nu$ by stationarity.

(iii) If the mark distribution is exponential with mean μ, then

$$\pi(x) = A \exp\left(\frac{x}{\mu} - \frac{1}{\nu} \int_0^x \Psi(u)\mathrm{d}u \right).$$

(iv) If $\Psi(x) = \exp[\beta(x - x_0)]$, the equation for $\varphi(s)$ above takes the form

$$\varphi(s) = c\gamma(s)\varphi(s - i\beta)\,, \quad c = \mathrm{e}^{-\beta x_0 \mu/\nu},$$

which admits the solution in infinite product form

$$\varphi(s) = \mathrm{e}^{isR}\gamma(s) \prod_{k=1}^\infty \mathrm{e}^{is/(\beta k)}\, \frac{\gamma(s - ik\beta)}{\gamma(-ik\beta)}\,,$$

where $R = x_0 + \big(\log(\beta\nu) - \gamma_0\big)/\beta$ and $\gamma_0 = 0.5772\ldots$.

(v) Show that, in the stationary regime, if the jump distribution has moment generating function $m(s)$, the risk $\Psi[X(t)]$ has moments

$$\mathrm{E}\big([\Psi(X)]^k\big) = \begin{cases} \mu\nu & (k=1), \\ \dfrac{(\nu\beta)^k (k-1)!}{\prod_{\ell=0}^{k-1}[1 - m(-\beta\ell)]} & (k = 2, 3, \ldots), \end{cases}$$

$$\mathrm{E}\big([\Psi(X)]^{-k}\big) = \frac{\prod_{\ell=0}^{k}[m(\beta\ell) - 1]}{(\nu\beta)^\ell\, \ell!} \qquad (k = 1, 2, \ldots).$$

[*Hint*: See Vere-Jones (1988) and Borovkov and Vere-Jones (2000).]

7.2.10 (*Continuation*) *Variance properties.*

(a) Let $N(t)$ denote the number of jumps (events in the ground process) for the stress-release model. Show that in the stationary case, if $X(t)$ has finite second moment, then $\operatorname{var} N(t)$ is bounded uniformly in t if and only if the jump distribution is degenerate at a single point.

[*Hint*: In this case, $X(t)$ has bounded variance and the forward result is trivial; for the converse, consider a bivariate version of Wald's identity using the joint characteristic function for the intervals T_i and number of jumps N_i between successive crossings of a fixed level for $X(t)$.]

(b) Under similar conditions, the mean rate and reduced second factorial moment density for the stress-release model can be expressed in the forms

$$m = \int \Psi(x)\pi(x)\,\mathrm{d}x,$$
$$\breve{m}_{[2]}(u) = \iiint \Psi(x)\pi(x)\,\mathrm{d}x\, j(y-x)\,\mathrm{d}y\,\Psi(z)\,F_u(y,\mathrm{d}z),$$

where j is the density of the jump distribution and the transition kernel $F_u(y,z) = \Pr\{X(u) \le z \mid X(0+) = y\}$.

(c) In general, the difficulty of solving the forward equations to obtain the transition kernel $F_u(y,\cdot)$ renders the equations above of relatively academic interest. However, if $\Psi(x) = \sigma$ for $x > 0$ and 0 otherwise, the process alternates between 'periods of prosperity' when $X(t) > 0$ and 'periods of recovery' when $X(t) < 0$, the terminology being suggested by the analogy of a collective risk model. Then, an argument similar to that used for M/G/1 queue and analogous storage problems can be used to show that the reduced covariance density $\breve{c}_{[2]}(u)$ has Laplace transform of the form

$$c^*_{[2]}(s) = [1 + \omega(s)]^{-1},$$

where $\omega(s)$ is the unique solution in $\mathrm{Re}(\theta) > 0$ of the equation $\theta - s = \sigma[1 - j^*(\theta)]$ and j^* is the Laplace transform of the jump density.

7.2.11 *Renewal process compensators.*

(a) By integrating the conditional intensity function in (7.2.14), show that when the lifetime distribution of a renewal process has a density f, the compensator has the form

$$\Lambda^*(t) = -\sum_{n=1}^{N(t)} \log S(T_n - T_{n-1}) - \log S(t - T_{N(t)}),$$

where $S(\cdot)$ is the survivor function for the lifetime d.f. with density f.

(b) Verify directly that $\Lambda^*(t)$ as defined makes $N(t) - \Lambda^*(t)$ a martingale.

(c) Show that (b) continues to hold for a general renewal process whose lifetime r.v.s are positive a.s., provided the log survivor function is replaced by the integrated hazard function (IHF).

7.3. Conditional Intensities for Marked Point Processes

The extension of conditional intensity models to higher dimensions is surprisingly straightforward provided that a causal, time-like character is retained for the principal dimension. When this is present, as in space–time processes, the development of conditional intensities and likelihoods can proceed along

much the same lines as was developed for one-dimensional simple point processes in the preceding sections. When it is absent, as in purely spatial point patterns, analysis is still possible in the finite case (compare the discussions in Chapter 5 and Section 7.1) but raises major problems for nonfinite cases such as occur for homogeneous processes in the plane. In this section, we examine the extension of the ideas of Section 7.2 to MPPs in time and space–time point processes. A more general and rigorous discussion of conditional intensities and related topics, for both simple and marked point processes in time, is given in Chapter 14. An approach to likelihood methods for spatial processes, based on the Papangelou intensity, is in Chapter 15. The ground work for the material in the present section was laid in the basic paper by Jacod (1975); among many other references, Karr (1986) gives both a review of inference procedures for MPPs and a range of examples and applications.

Consider then an MPP on $[0, \infty) \times \mathcal{K}$, where, as in Section 6.4, $\mathcal{K}$ denotes the mark space, which may be discrete (for multivariate point processes), the positive half-line (if the marks represent weights or energies), two- or three-dimensional Euclidean space (for space–time processes), or more general spaces [e.g. for the Boolean model of Example 6.4(d)].

In order to define likelihoods for MPPs, we need first to fix on a measure in the mark space $(\mathcal{K}, \mathcal{B}_{\mathcal{K}})$ to serve as a reference measure in forming densities. We shall denote this reference measure by $\ell_{\mathcal{K}}(\cdot)$, using $\ell(\cdot)$ to denote Lebesgue measure on $\mathbb{R}^d$. When $\mathcal{K}$ is also some Euclidean space, it will often be convenient to take $\ell_{\mathcal{K}}$ to be Lebesgue measure on that space but not always so; for example, in some situations it may be simpler to take $\ell_{\mathcal{K}}$ to be a probability measure on $\mathcal{K}$. Similarly, when the mark space is discrete, it will often be convenient to take the reference measure to be counting measure, but in some situations it may again be more convenient to choose the reference measure to be a probability measure.

Once the reference measure $\ell_{\mathcal{K}}$ has been fixed, we can extend the notion of a regular point process from simple to marked point processes. As in Definition 7.1.I, we shall say that an MPP on $\mathcal{X} = \mathbb{R}^d \times \mathcal{K}$ is *regular on* A for a bounded Borel set $A \in \mathcal{B}_{\mathcal{X}}$ if for all $n \geq 1$ the Janossy measure J_n is absolutely continuous with respect to the n-fold product of $\ell \times \ell_{\mathcal{K}}$ and *regular* if it is regular on A for all bounded $A \in \mathcal{B}_{\mathcal{X}}$. Thus, when the MPP is regular on A, for every $n > 0$ there exists a well-defined Janossy density $j_n(\cdot \mid A \times \mathcal{K})$ with the interpretation

$$j_n(x_1, \ldots, x_n, \kappa_1, \ldots, \kappa_n \mid A \times \mathcal{K})\, \mathrm{d}x_1 \ldots \mathrm{d}x_n\, \ell_{\mathcal{K}}(\mathrm{d}\kappa_1) \ldots \ell_{\mathcal{K}}(\mathrm{d}\kappa_n)$$
$$= \Pr\{\text{points around } (x_1, \ldots, x_n) \text{ with marks around } (\kappa_1, \ldots, \kappa_n)\}.$$

The following equivalences extend to MPPs the discussion around Proposition 7.1.III.

Proposition 7.3.I. *Let $N(\cdot)$ be an MPP on $\mathbb{R}^d \times \mathcal{K}$, let ℓ denote Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ and $\ell_{\mathcal{K}}$ the reference measure on $(\mathcal{K}, \mathcal{B}_{\mathcal{K}})$, and let A be a bounded set in $\mathcal{B}_{\mathbb{R}^d}$. Then, conditions* (i)–(iv) *below are equivalent.*

(i) *$N(\cdot)$ is regular on A.*
(ii) *The probability measure induced by $N(\cdot)$ on $\mathcal{Z}_A^{\cup}$, where $\mathcal{Z}_A = A \times \mathcal{K}$, is absolutely continuous with respect to the measure induced by $\ell \times \ell_{\mathcal{K}}$ on $\mathcal{Z}_A^{\cup}$.*
(iii) *The ground process $N_{\mathrm{g}}(\cdot)$ is regular on A, and for each $n > 0$ the conditional distribution of the marks $(\kappa_1, \ldots, \kappa_n)$, for a given realization $(x_1, \ldots, x_n)$ of the locations within A, is absolutely continuous with respect to $\ell_{\mathcal{K}}^{(n)}$ with density $f_{A,n}(\kappa_1, \ldots, \kappa_n \mid x_1, \ldots, x_n)$, say.*
(iv) *If $\Pi(\cdot)$ is a probability measure equivalent to $\ell_{\mathcal{K}}$ on $(\mathcal{K}, \mathcal{B}_{\mathcal{K}})$, then $N(\cdot)$ is absolutely continuous with respect to the compound Poisson process $N_0(\cdot)$ for which the ground process N_0^{g} has positive intensity λ on A and the marks are i.i.d. with common probability distribution Π.*

PROOF. The four statements are just alternative ways of stating the fact that the Janossy measures $J_n(\cdot)$ in the proposition have appropriate densities on all components of $\mathcal{X}^{\cup}$. □

When any one of the conditions is satisfied, the Radon–Nikodym derivative of the probability measure $\mathcal{P}$ for N with respect to the probability measure $\mathcal{P}_0$ of the compound Poisson process N_0 in (iv) has the form [see (7.1.3b)]

$$\mathrm{e}^{-\lambda\ell(A)}\frac{\mathrm{d}\mathcal{P}}{\mathrm{d}\mathcal{P}_0} = J_0 I_{\{N(T)=0\}} + \sum_{n=1}^{\infty} I_{\{N(T)=n\}} \frac{j_n^{\mathrm{g}}(x_1, \ldots, x_n \mid A)}{\lambda^n} \frac{f_{A,n}(\kappa_1, \ldots, \kappa_n \mid t_1, \ldots, t_n)}{\pi(\kappa_1)\cdots\pi(\kappa_n)}, \tag{7.3.1a}$$

in which $\pi(\kappa) = (\mathrm{d}\Pi/\mathrm{d}\ell_{\mathcal{K}})(\kappa)$ and is itself a portmanteau expression of the statements that, given a realization $(t_1, \kappa_1), \ldots, (t_n, \kappa_n)$ with $N(T) = n$, the likelihood ratio of N with respect to N_0 is given by

$$L/L_0 = j_n^{\mathrm{g}}(x_1, \ldots, x_n \mid A) f_n(\kappa_1, \ldots \kappa_n \mid x_1, \ldots, x_n) / [\lambda^n \pi(\kappa_1)\cdots\pi(\kappa_n)]. \tag{7.3.1b}$$

Much as in the discussion leading to Proposition 7.2.I, we now rewrite the Janossy densities in a way that takes advantage of the directional character of time. Thus, the Janossy densities for the first few pairs may be represented in the form

$$\begin{aligned} J_0(T) &= S_1(T), \\ j_1(t_1, \kappa_1 \mid T) &= p_1(t_1, \kappa_1) = p_1(t_1)\, f_1(\kappa_1 \mid t_1) && (0 < t_1 < T), \\ j_2(t_1, t_2, \kappa_1, \kappa_2 \mid T) &= p_1(t_1)\, f_1(\kappa_1 \mid t_1)\, p_2(t_2 \mid (t_1, \kappa_1))\, f_2(\kappa_2 \mid (t_1, \kappa_1), t_2) \\ & && (0 < t_1 < t_2 < T), \end{aligned}$$

where the $p_i(\cdot)$ refer to the densities, suitably conditioned, for the locations in the ground process, and the $f_i(\cdot)$ refer to the densities, again suitably conditioned, for the marks. There is a subtle difference in the conditioning incorporated into the conditional densities $f_n(\kappa_n \mid (t_1, \kappa_1), \ldots, (t_{n-1}, \kappa_{n-1}), t_n)$ that appear in the equations above and those that appear in the proposition. In the equations above we condition the distribution of the current mark, as

time progresses, on both marks and time points of all preceding events; in the proposition, we condition on the full set of time points in $(0, T)$, irrespective of the marks and of their relative positions in time.

Once again, the dependence of the left-hand side on T is illusory, and the densities for the locations can be expressed in terms of corresponding hazard functions. The conditioning in the hazard functions may now include the values of the preceding marks as well as the length of the current and preceding intervals. All this information is collected into the *internal history* $\boldsymbol{\mathcal{H}} \equiv \{\mathcal{H}_t : t \geq 0\}$ of the process so that the amalgam of hazard functions and mark densities can be represented as a single composite function for the MPP, namely

$$\lambda^*(t,\kappa) = \begin{cases} h_1(t) f_1(\kappa \mid t) & (0 < t \leq t_1), \\ \quad \vdots & \\ h_n\big(t \mid (t_1,\kappa_1),\dots,(t_{n-1},\kappa_{n-1})\big) \times & \\ \quad f_n\big(\kappa \mid (t_1,\kappa_1),\dots,(t_{n-1},\kappa_{n-1}), t\big) & (t_{n-1} < t \leq t_n,\ n \geq 2), \\ \quad \vdots & \end{cases} \tag{7.3.2}$$

where $h_1(t)$ is the hazard function for the location of the initial point, $h_2(t \mid (t_1, \kappa_1))$ the hazard function for the location of the second point conditioned by the location of the first point and the value of the first mark, and so on, while $f_1(\kappa \mid t)$ is the density for the first mark given its location, and so on.

Definition 7.3.II. *Let N be a regular MPP on $\mathbb{R}_+ \times \mathcal{K}$. The* conditional intensity function *for N, with respect to its internal history $\boldsymbol{\mathcal{H}}$, is the representative function $\lambda^*(t,\kappa)$ defined piecewise by* (7.3.2).

Predictability is again important in that the hazard functions refer to the risk at the end of a time interval, not at the beginning of the next time interval, so left-continuity should be preferred where there is a jump in the conditional intensity. Similarly, the conditional mark density refers to the distribution to be anticipated at the end of a time interval, not immediately after the next interval has begun. More formal and more general discussions of predictibility in the MPP context will be given in Chapter 14.

It is often convenient to write

$$\lambda^*(t,\kappa) = \lambda_{\mathrm{g}}^*(t)\, f^*(\kappa \mid t)\,, \tag{7.3.3}$$

where $\lambda_{\mathrm{g}}^*(t)$ is the $\boldsymbol{\mathcal{H}}$-intensity of the ground process (i.e. of the locations $\{t_i\}$ of the events), and $f^*(\kappa \mid t)$ is the conditional density of a mark at t given $\mathcal{H}_{t-}$ (the reader will note that we use the * notation as a reminder that the 'functions' concerned are also random variables dependent in general on the random past history of the process). The two terms in (7.3.3) correspond to the first and second factors in (7.3.2). Heuristically, equations (7.3.2) and (7.3.3) can be summarized in the form

$$\lambda^*(t,\kappa)\,\mathrm{d}t\,\mathrm{d}\kappa \approx \mathrm{E}[N(\mathrm{d}t \times \mathrm{d}\kappa) \mid \mathcal{H}_{t-}] \approx \lambda_{\mathrm{g}}^*(t)\, f^*(\kappa \mid t)\,\mathrm{d}t\,\mathrm{d}\kappa\,. \tag{7.3.4}$$

Notice that the $\boldsymbol{\mathcal{H}}$-intensity $\lambda_{\mathrm{g}}^*(t)$ is *not* in general the same as the conditional intensity $\lambda^{\mathrm{g}}(t)$ of the ground process with respect to its own internal history $\boldsymbol{\mathcal{H}}^{\mathrm{g}}$: $\boldsymbol{\mathcal{H}}$ incorporates information about the values of the marks, whereas $\boldsymbol{\mathcal{H}}^{\mathrm{g}}$ does not. The example below illustrates the difference in a simple special case.

EXAMPLE 7.3(a) *Bivariate Poisson process* [see Example 6.3(e)]. We consider a bivariate Poisson process initiated at time 0 rather than the stationary version considered earlier. We consider also just the process of linked pairs, in which the points $\{t_i\}$ of component I form the 'parents' and arrive according to a simple Poisson process with rate λ while the points $\{s_j\}$ of component II represent the process of 'offspring'. We assume each parent has just one offspring, delayed by nonnegative random times $\{\tau_i\}$ forming an i.i.d. sequence, independent also of the times $\{t_i\}$, with common exponential distribution $1 - \mathrm{e}^{-\mu\tau}$.

We shall treat this process as a special case of an MPP with mark space having two discrete points, corresponding to components I and II. The internal history, $\boldsymbol{\mathcal{H}}$, for the full process records the occurrence times and marks for both types of events but does not record which event in component II is associated with which event in component I. Suppose that, at time t, $N_{\mathrm{I}}(t) = n, N_{\mathrm{II}}(t) = m$, where necessarily $m \le n$. The full $\boldsymbol{\mathcal{H}}$-intensity is given by

$$\lambda^*(t,\kappa) = \begin{cases} \lambda & (\kappa = \mathrm{I}), \\ (n-m)\mu & (\kappa = \mathrm{II}). \end{cases}$$

Let $\boldsymbol{\mathcal{H}}^{\mathrm{I}}, \boldsymbol{\mathcal{H}}^{\mathrm{II}}$, and $\boldsymbol{\mathcal{H}}^{\mathrm{g}}$ denote the internal histories of the component I process, the component II process, and the ground process. The $\boldsymbol{\mathcal{H}}^{\mathrm{I}}$-intensity of component I is clearly equal to its $\boldsymbol{\mathcal{H}}$-intensity $\lambda^{\mathrm{I}} \equiv \lambda$. To find the $\boldsymbol{\mathcal{H}}^{\mathrm{II}}$-intensity of component II, we have to average over the $n \ge m$ points of component I. For a given value of n, the locations t_i may be treated as n i.i.d. variables uniformly distributed over $(0,t)$. The probability that any one such point produces an offspring that appears only *after* time t is given by

$$p(t) = \int_0^t \mathrm{e}^{-\mu(t-s)}\,\frac{\mathrm{d}s}{t} = \frac{1-\mathrm{e}^{-\mu t}}{\mu t}\,.$$

The $k = n - m$ parent points that fail to produce offspring in the interval $(0,t)$ then form a 'thinned' version of the original, Poisson-distributed number n of the component I points in $(0,t)$, the selected and nonselected points forming two independent streams. Independently of the number m of successes, the expected number of points with offspring still pending is thus $\lambda t p(t)$ and we obtain for the $\boldsymbol{\mathcal{H}}^{\mathrm{II}}$-intensity of the component II process

$$\lambda_{\mathrm{II}}(t) = \mathrm{E}[(n-m)\mu \mid N_{\mathrm{II}}(t) = m] = \mu\lambda t(1-\mathrm{e}^{-\mu t})/(\mu t) = \lambda(1-\mathrm{e}^{-\mu t}).$$

This is a nonrandom function of t, and we recognize it as the conditional intensity of a nonstationary Poisson process. Thus, the two components separately are Poisson, and the rate of the component II process approaches that of component I as $t \to \infty$. The ground process has $\boldsymbol{\mathcal{H}}$-intensity $\lambda + (n-m)\mu$

and $\mathcal{H}^{\mathrm{II}}$-intensity $\lambda(2 - \mathrm{e}^{-\mu t})$; its $\mathcal{H}^{\mathrm{g}}$-intensity is that of a Gauss–Poisson process; see Exercise 7.3.1.

Similar distinctions need to be borne in mind with respect to the various compensators and martingales that can be formed with the two component processes. Thus, $N_{\mathrm{I}}(t) - \lambda t$ is both an $\mathcal{H}$- and an $\mathcal{H}^{\mathrm{I}}$-martingale, the process $N_{\mathrm{II}}(t) - \mu(N_{\mathrm{II}}(t) - N_{\mathrm{I}}(t))$ is an $\mathcal{H}$-martingale, and $N_{\mathrm{II}}(t) - \lambda t(1 - \mathrm{e}^{-\mu t})$ is an $\mathcal{H}^{\mathrm{II}}$-martingale. □

We now turn to an MPP extension of Proposition 7.2.III, expressing the likelihood of a simple point process in terms of its conditional intensity. As there, reversing the construction that leads from the point process distributions to the $\mathcal{H}$-intensity in (7.3.2) yields an explicit expression for the Janossy density of the MPP in terms of its conditional intensity (see below). Details of the proof are left to Exercise 7.3.2.

Proposition 7.3.III. *Let N be a regular MPP on $[0,T] \times \mathcal{K}$ for some finite positive T, and let $(t_1, \kappa_1), \ldots, (t_{N_{\mathrm{g}}(T)}, \kappa_{N_{\mathrm{g}}(T)})$ be a realization of N over the interval $[0,T]$. Then, the likelihood L of such a realization is expressible in the form*

$$\begin{aligned} L &= \left[\prod_{i=1}^{N_{\mathrm{g}}(T)} \lambda^*(t_i, \kappa_i) \right] \exp\left(- \int_0^T \int_{\mathcal{K}} \lambda^*(u, \kappa) \, \mathrm{d}u \, \ell_{\mathcal{K}}(\mathrm{d}\kappa) \right) \\ &= \left[\prod_{i=1}^{N_{\mathrm{g}}(T)} \lambda_{\mathrm{g}}^*(t_i) \right] \left[\prod_{i=1}^{N_{\mathrm{g}}(T)} f^*(\kappa_i \mid t_i) \right] \exp\left(- \int_0^T \lambda_{\mathrm{g}}^*(u) \, \mathrm{d}u \right), \end{aligned} \tag{7.3.5}$$

where $\ell_{\mathcal{K}}$ is the reference measure on $\mathcal{K}$. Its log likelihood ratio on $[0,T]$ relative to the compound Poisson process N_0 with constant intensity λ and i.i.d. mark distribution with density $\pi(\cdot)$ is expressible as

$$\begin{aligned} \log \frac{L}{L_0} &= \sum_{i=1}^{N_{\mathrm{g}}(T)} \log \frac{\lambda^*(t_i, \kappa_i)}{\lambda \pi(\kappa)} - \int_0^T \int_{\mathcal{K}} [\lambda^*(u, \kappa) - \lambda \pi(\kappa)] \, \mathrm{d}u \, \ell_{\mathcal{K}}(\mathrm{d}\kappa) \\ &= \sum_{i=1}^{N_{\mathrm{g}}(T)} \log \frac{\lambda_{\mathrm{g}}^*(t_i)}{\lambda} - \int_0^T [\lambda_{\mathrm{g}}^*(u) - \lambda] \, \mathrm{d}u + \sum_{i=1}^{N_{\mathrm{g}}(T)} \log \frac{f^*(\kappa_i \mid t_i)}{\pi(\kappa_i)} \,. \end{aligned} \tag{7.3.6}$$

The second form in equations (7.3.5) and (7.3.6) follows from the assumption that the densities over the mark space are proper (i.e. integrate to unity). The reversibility of the arguments leading to the representation of the conditional intensity function in (7.3.2) (see Exercise 7.3.2) implies the following MPP analogue of Proposition 7.2.IV.

Proposition 7.3.IV. *Let N be a regular MPP as in Proposition* 7.3.III. *Then, the conditional intensity function with respect to the internal history $\mathcal{H}$ determines the probability structure of N uniquely.*

The next proposition gives specific examples of such characterizations, makeing more explicit the distinction between point processes with independent and unpredictable marks introduced already in Section 6.4.

Proposition 7.3.V. *Let N be a regular MPP on $\mathbb{R}_+ \times \mathcal{K}$ with $\mathcal{H}$-intensity expressible as*

$$\lambda^*(t,\kappa) = \lambda_{\mathrm{g}}^*(t) f^*(\kappa \mid t), \tag{7.3.7}$$

where $\lambda_{\mathrm{g}}^*(t)$ is the $\mathcal{H}$-intensity of the ground process. Then N is

(i) a compound Poisson process *if $\lambda_{\mathrm{g}}^*(t) = \lambda(t)$ and $f^*(\kappa \mid t) = f(\kappa \mid t)$ for deterministic functions $\lambda(t)$ and $f(\kappa \mid t)$;*

(ii) a process with independent marks *if $\lambda_{\mathrm{g}}^*(t)$ equals the $\mathcal{H}^{\mathrm{g}}$-intensity for the ground process and $f^*(\kappa \mid t) = f(\kappa \mid t)$ as in* (i); *and*

(iii) a process with unpredictable marks *if $f^*(\kappa \mid t) = f(\kappa \mid t)$ as in* (i).

Proof. In a process with independent marks, the ground process and the marks are completely decoupled (i.e. they are independent processes), whereas for a process with unpredictable marks, the marks can influence the subsequent evolution of the process, though the ground process does not influence the distribution of the marks. The compound Poisson process is the special case of a Poisson process with independent marks. The forms of the conditional intensities follow readily from these comments, which merely reflect the definitions of these three types of MPP given in Definition 6.4.III and preceding Lemma 6.4.VI. The lemma is then a consequence of the uniqueness assertion in Proposition 7.3.IV. Some details and examples are given in Exercise 7.3.5. □

The following nonlinear generalization of the Hawkes process is important for its range of applications. It has been used as a model for neuron firing in Brémaud and Massoulié (1994, 1996), and it also embraces a range of other examples, including both ordinary and space–time versions of the ETAS model [Examples 6.4(d) and 7.2(f)].

Example 7.3(b) *Nonlinear, marked Hawkes processes* [see Example 7.2(b)]. We start by extending the basic Hawkes process N to a nonlinear version with conditional intensity [see (7.2.6)]

$$\lambda^*(t) = \Phi\left(\lambda + \int_0^t \mu(t-u)\, N(\mathrm{d}u)\right), \tag{7.3.8}$$

where the nonnegative function Φ is in general nonlinear but satisfies certain boundedness and continuity conditions; in particular, it is required to be Lipschitz with Lipschitz constant $\alpha \leq 1$.

Such a nonlinear Hawkes process can immediately be extended to a nonlinear marked Hawkes process by giving the points independent marks with density $f(\kappa)$ so that the conditional intensity function for the marked version is

$$\lambda^*(t,\kappa) = \Phi\left(\lambda + \int_0^t \mu(t-u)\, N_{\mathrm{g}}(\mathrm{d}u)\right) f(\kappa) = \lambda_{\mathrm{g}}^*(t) f(\kappa). \tag{7.3.9}$$

The marks here make no contribution to the current risk, nor to the evolution of the ground process, which therefore has the same structure as the process N of (7.3.8). Consequently, in (7.3.9) we have $N_{\mathrm{g}} = N$.

By contrast, generalizing the ETAS model of Example 6.4(d) and using its notation, we may equally well consider extensions in which the conditional intensity has the form

$$\lambda^*(t,\kappa) = \Phi\bigg(\lambda + \int_{(0,t)\times\mathcal{K}} \psi(\chi)\mu(t-u)\,N(\mathrm{d}u\times\mathrm{d}\chi)\bigg) f(\kappa), \tag{7.3.10}$$

where $\psi(\chi)$ modifies the strength of the infectivity density $\mu(\cdot)$ according to the mark χ. In this case, the process has *unpredictable* marks that, depending on the form of $\psi(\cdot)$, can influence substantially the evolution of the ground process.

In both cases, the likelihood for a finite observation period $[0,T]$ decouples and, following the second form in (7.3.6), can be written as

$$\begin{aligned}\log L &= \left[\sum_{i:0\le t_i\le T} \log\lambda_{\mathrm{g}}^*(t_i) - \int_0^T \lambda_{\mathrm{g}}^*(u)\,\mathrm{d}u\right] + \sum_{i:0\le t_i\le T} \log f(\kappa_i) \\ &\equiv \log L_1 \,+\, \log L_2,\end{aligned}$$

where

$$\lambda_{\mathrm{g}}^*(t) = \Phi\bigg(\lambda + \int_{(0,t)\times\mathcal{K}} \psi(\kappa)\mu(t-u)\,N(\mathrm{d}u\times\mathrm{d}\kappa)\bigg).$$

In many parametric models, no parameter appears in both L_1 and L_2, so each term can be maximized separately.

It is not necessary here to limit the mark to a measure of the size of the accompanying event. As suggested in Example 6.4(d), elements in the mark space may comprise both size and spatial components, $\kappa\in\mathcal{K}$ and $y\in\mathcal{Y}$, say. Then we can write, for example,

$$\lambda^*(t,\kappa,x) = \Phi\bigg(\lambda + \int_{(0,t)\times\mathcal{K}\times\mathcal{Y}} \psi(\chi)\mu(t-u)g(x-y)\,N(\mathrm{d}u\times\mathrm{d}\chi\times\mathrm{d}y)\bigg) f(\kappa),$$

where the spatial density $g(\cdot)$, like $f(\cdot)$, has been normalized to have unit integral and determines the positions of the offspring about the ancestor. Because of the independent sizes κ_i here, the log likelihood again separates into two terms, the first of which is analogous to $\log L_1$ above but includes an integration over both space and time.

From a model-building point of view, it is of critical importance to establish conditions for the existence of stationary versions of the process and for convergence to equilibrium. General conditions are given by Brémaud and Massoulié (1996) and discussed further in Chapters 13 and 14. In the special case corresponding to the space–time ETAS model, where the function Φ is linear (and can be taken to be the identity function), the process retains the basic branching structure, and a sufficient condition for the existence of a stationary version is the subcriticality of the underlying branching component, as outlined already in Example 6.4(d). □

It is, of course, quite possible to devise models where the mark distributions *are* dependent on the evolution of the process. A simple example is given below.

EXAMPLE 7.3(c) *Processes governed by a Markovian rate process.* Several models for both simple and marked processes are governed by an underlying Markov process, $X(t)$ say, which both influences and is influenced by the evolving point process. Typically, in the marked case, both the ground process intensity and the mark distribution depend on the current value of $X(t)$. Two simple models of this type are the simple stress-release model in Example 7.2(g) and the Cox process with Markovian rate function considered in Exercise 7.2.7.

To illustrate possible ramifications of such models, consider first a Hawkes process with exponential infectivity density $\mu(x) = \mu e^{-\mu x}$. In this case, the Markovian process $X(t)$ is given by the sum

$$X(t) = \mu \textstyle\sum_{i:\, 0<t_i<t} e^{-\mu(t-t_i)},$$

and we can write

$$\lambda^*(t) = \Phi[X(t)],$$

where in the simplest case, $\Phi(x) = \lambda + \nu x$ for some $\lambda > 0$ and $0 < \nu < 1$ as in Exercise 7.2.5.

Next, we could consider a marked version of such a process, with random event sizes $S_i = \psi(\kappa_i)$, defining $X(t)$ by

$$X(t) = \mu \sum_{i:0<t_i<t} \psi(\kappa_i) e^{-\mu(t-t_i)}. \tag{7.3.11}$$

In the simplest case of independent marks, with common density $f(\kappa)$,

$$\lambda^*(t,\kappa) = \Phi[X(t)] f(\kappa), \tag{7.3.12}$$

corresponding to an ETAS-type model but with exponential rather than power-law decay function.

It might well be natural, however, to suppose that not only the rate $\lambda^*(t)$ but also the density $f(\kappa)$ of the mark distribution could be affected by the value of $X(t)$, in which case $f(\kappa)$ would be replaced by $f(\kappa \mid X(t))$. To take a particular parametric example, let the mark distribution have an exponential density $\beta e^{-\beta\kappa}$, and set $\beta = a + bX(t)$ so that the conditional intensity takes the form

$$\lambda^*(t,\kappa) = e^{-\lambda-\nu X(t)} \cdot [a + bX(t)] e^{-[a+bX(t)]\kappa},$$

with $X(t)$ given by (7.3.11). In this case, the log likelihood can still be written as the sum of two terms, $\log L = \log L_1 + \log L_2$, say, where the second term equals $\sum_i \log f(\kappa_i \mid X(t_i))$, but it is no longer possible to decouple the two parts of the likelihood completely because the parameters relating to $X(t)$ appear in both parts. In the specific example considered, $\log L$ equals

$$\left(\sum_i \log\big(\Phi[X(t_i)]\big) - \int_0^T \Phi[X(u)]\, du \right) + \sum_i \log f(\kappa_i \mid X(t_i)) = \log L_1 + \log L_2,$$

where the parameters λ and ν appear in L_1 only, the parameters a and b appear in L_2 only, but the parameter μ, as well as any parameter involved in the definition of the function ψ, appears in both L_1 and L_2. □

EXAMPLE 7.3(d) *Linked stress-release model.* This is a multivariate version of the basic model outlined in Example 7.2(g). We consider a finite number of distinct regions or components $i = 1, \ldots, I$, say, each with its own stress level $X_i(t)$ and with the property that a proportion θ_{ij} of a stress drop occurring in region i is transferred to region j (but we do not necessarily require either $\theta_{ij} \geq 0$ or $\sum_i \theta_{ij} = 1$). The evolution of stress $X_i(t)$ in the ith region can thus be expressed in the form

$$X_i(t) = X_i(0) + \rho_i t - \sum\nolimits_j \theta_{ij} S^{(j)}(t), \tag{7.3.13}$$

where $S^{(j)}(t)$ is the accumulated stress release in region j over the period $[0, t)$ and ρ_i is the rate of stress input into region i. The process of events is doubly marked: by the region i and by the size of the stress drop κ.

We suppose that both the risk functions (i.e. stress levels) and the jump distributions are functions of the vector $\mathbf{X}(t)$. The assumptions imply that the process $\mathbf{X}(t)$ controls the evolution of the point process and is itself Markovian. They lead to a conditional intensity of the form

$$\lambda^*(t, i, \kappa) = \Psi_i\big(\mathbf{X}(t) f_i[\kappa \mid \mathbf{X}(t)]\big), \tag{7.3.14}$$

where $f_i[\kappa \mid \mathbf{X}(t)]$ is the density for the distribution of stress drop for an event that occurs in region i at a time when the vector of stress levels is $\mathbf{X}(t)$, and Ψ_i gives the risk in region i as a function of the vector of stress levels $\mathbf{X}(t)$. Typically, $\Psi_i(\mathbf{X}) = \exp(\mu_i + \nu_i X_i)$, so that only the stress level in the region under consideration affects the risk in that region. Then, the conditional intensity function can be written in the reparameterized form

$$\lambda^*(t, i, \kappa) = \exp\left[\alpha_i + \nu_i\left(\rho_i t - \sum_j \theta_{ij} S^{(j)}(t)\right)\right] f_i[\kappa \mid \mathbf{X}(t)],$$

where $\alpha_i = \mu_i + \nu_i X_i(t)$, ν_i, ρ_i, and θ_{ij} $(i \neq j)$ are the parameters to be estimated, apart from those involved in the density function for the stress drops, and we set $\theta_{ii} = 1$.

As in Example 7.3(c), the likelihood can be expressed as the sum of two terms, the first relating to the times and the second to the stress drops of the events, but only fully decouples when the stress drops are i.i.d.

In the present context, an appealing candidate for the mark distribution is the tapered Pareto, or Kagan, distribution with survivor function

$$1 - F(\kappa) = \left(\frac{c}{c + \kappa}\right)^{\alpha} \mathrm{e}^{-\beta\kappa}. \tag{7.3.15}$$

Typically, β is taken very small so that for small and intermediate values of κ, the density is close to a power-law form, but for large κ it is dominated by the exponential taper. Distributions of this general type have recently been considered in several contexts where it is desirable for the body of the distribution to have a power-law character but for the moments to remain

finite (see e.g. Kagan, 1999; Kagan and Schoenberg, 2001; Vere-Jones *et al.*, 2001). For the present example, we might take α as fixed and equal to unity and allow β to decrease with the value of $X(t)$ in such a way that the upper turning point $1/\beta$ increases to ∞. In this case, the tail of the distribution would progressively lengthen (admitting larger and larger events) as the stress level increased while its mean approached $+\infty$. For applications of the linked stress-release model to earthquake data, using generally independent marks with exponential distribution, see e.g. Liu *et al.* (1999), Bebbington and Harte (2001), and Lu and Vere-Jones (2000). See Exercise 7.3.6 for stability properties of the model. □

EXAMPLE 7.3(e) *Cumulative processes.* Let $N \equiv \{(t_i, \kappa_i)\}$ be a regular MPP defined over the time interval $(0, T)$, and consider the random measure derived from N as in (6.4.6) and characterized through the cumulative process

$$\xi(t) = \int_{(0,t)\times\mathcal{K}} \kappa\, N(\mathrm{d}u \times \mathrm{d}\kappa).$$

We do not insist here that the process have independent or unpredictable marks. Although $\xi(t)$ corresponds to a random measure rather than a point process, it is still germane to ask questions about its internal history, its likelihood, and its conditional intensity. The following points are straightforward to verify and are left to the reader.

(i) The internal history of ξ coincides with the internal history for the underlying MPP N.

(ii) The likelihood for $\xi(t)$ over an interval $(0, T)$ coincides with the likelihood for N over the same period.

(iii) A conditional intensity $\mu^*(t)$ for $\xi(t)$ can be defined by

$$\mu^*(t)\,\mathrm{d}t \equiv \mathrm{E}[\mathrm{d}\xi(t) \mid \mathcal{H}^{\xi}_{t-}] = \lambda^*(t)\,\mathrm{E}[\kappa \mid \mathcal{H}^{N}_{t-}]\,\mathrm{d}t = \int_{\mathcal{K}} \kappa\,\lambda^*(t,\kappa)\,\mathrm{d}\kappa,$$

where $\lambda^*(t, \kappa)$ is the $\boldsymbol{\mathcal{H}}^N$-conditional intensity for the MPP N. □

Exercises and Complements to Section 7.3

7.3.1 *Further properties of the bivariate Poisson process* [see Example 7.3(a)].

(a) Discuss the $\mathcal{F}_0$-intensity for the process of the example.
[The difficulty here, as for other cluster processes, is in averaging over the different possible ways that parents and offspring may be associated; see the comments on Example 6.2(c) concerning the Gauss–Poisson process.]

(b) Verify the martingale properties asserted at the end of the example.

7.3.2 Write out explicitly the construction leading back from (7.3.2) to the Janossy densities and hence complete the proof of Proposition 7.3.III.

7.3.3 Define a one-point MPP on $(0, T) \times (0, T)$ as follows. For any realization $\{(t_1, \kappa_1)\}$, say, the point t_1 has the density $f(\cdot)$ on $(0, T)$ and, given t_1, κ_1 is uniformly distributed on $(0, T - t_1)$. Find the conditional intensities for this MPP and for the bivariate point process $\{(t_1, t_1 + \kappa_1)\}$. What are the corresponding compensators?

7.3.4 Verify the forms of conditional intensity in Proposition 7.3.V for compound Poisson processes and processes with independent or unpredictable marks.

7.3.5 *Accelerated moment release model.* Let t_f denote the time of a major earthquake, and $\xi(t) = \sum_{i=1}^{N(t)} \kappa_i$ the cumulative release of seismic moments of small or moderate earthquakes up until time $t < t_f$. According to Varnes (1989) and Main (1996), there are physical grounds for supposing that $\xi(t)$ increases hyperbolically before the major event; i.e. $\xi(t) \approx A + B(t_f - t)^{-m}$, where A, B and m are positive constants. Suggest an appropriate conditional intensity model and associated likelihood, assuming that the relationship refers to $\mathrm{E}[\xi(t)]$ and that the increase is due to either

(i) an increase in the frequency of events but not their average size; or
(ii) an increase in the average size of the events but not their frequency; or
(iii) an increase in both frequency and average size.

[*Hint*: In Vere-Jones *et al.* (2001), both exponential and tapered Pareto distributions are used to model the event sizes, and a maximum entropy argument is used to suggest that the increase in moment should be partitioned between the mean size and mean frequency of events in such a way that each takes up the square root of the overall increase.]

7.3.6 *Stability results for linked stress-release model* [see Example 7.3(d)].

(a) Suppose that a stationary regime exists; for events in region i, let $l_i = \mathrm{E}(\Psi_i[\mathbf{X}(t)])$ denote their rate of occurrence and m_i their mean size. Establish the balance equations

$$\rho_i = \sum_j \theta_{ij} l_j m_j \qquad (i = 1, \ldots, I).$$

(b) Let

$$R_i(\mathbf{x}) = \Psi_i(\mathbf{x}\, \mathrm{E}[\kappa_i \mid \mathbf{X}(t) = \mathbf{x}]),$$

and write $\mathbf{R}(\mathbf{x})$ for the vector with components $R_i(\mathbf{x})$, with $\mathbf{x}$ in domain D. Then, a matrix analogue of condition (7.2.13) takes the form

$$\liminf_{\mathbf{x}\in D} \mathbf{R}(\mathbf{x}) \le [2\mathbf{I} - \Theta]^{-1}\boldsymbol{\rho} \le \limsup_{\mathbf{x}\in D} \mathbf{R}(\mathbf{x}),$$

where $\boldsymbol{\rho}$ is the vector of input rates. Investigate possible sufficient conditions for the existence of a stationary version of the process.

7.4. Random Time Change and a Goodness-of-Fit Test

The proposition below has been part of the folklore of point process theory for many years. In essence, it goes back to the work of Watanabe (1964), who first recognised that the Poisson process could be characterized by the form of its compensator (the deterministic function λt), and Meyer (1971). It was first clearly stated and proved by Papangelou (1974), who describes it in the following terms:

"Suppose that, starting at 0 say, we trace $\mathbb{R}_+$ in such a way that at the time we are passing position t our speed is $1/\lambda^*(t)$, which can be ∞. (The value $\lambda^*(t)$ is determined by the past, i.e. by what happened up to t.) Then the time instants at which we shall meet all the points in $\mathbb{R}_+$ of the process form a homogeneous Poisson process."

In other language, the random time transformation $\tau = \Lambda^*(t) = \int_0^t \lambda^*(u)\,\mathrm{d}u$ takes the point process with conditional intensity function $\lambda^*(t)$ into a unit-rate Poisson process.

Theorem 7.4.I. *Let N be a simple point process adapted to a history $\boldsymbol{\mathcal{F}}$ with bounded, strictly positive conditional $\boldsymbol{\mathcal{F}}$-intensity $\lambda^*(t)$ and $\boldsymbol{\mathcal{F}}$-compensator $\Lambda^*(t) = \int_0^t \lambda^*(u)\,\mathrm{d}u$ that is not* a.s.*-bounded. Under the random time change $t \mapsto \Lambda^*(t)$, the transformed process*

$$\widetilde{N}(t) = N\big(\Lambda^{*-1}(t)\big) \tag{7.4.1}$$

is a Poisson process with unit rate.

Conversely, suppose there is given a history $\boldsymbol{\mathcal{G}}$, a $\boldsymbol{\mathcal{G}}$-adapted cumulative process $M(t)$ with a.s. *finite, monotonically increasing and continuous trajectories, and a $\boldsymbol{\mathcal{G}}$-adapted simple Poisson process $N_0(t)$. Let $\boldsymbol{\mathcal{F}}$ denote the history of σ-algebras $\mathcal{F}_t = \mathcal{G}_{M(t)}$. Then $N(t) = N_0\big(M(t)\big)$ is a simple point process that is $\boldsymbol{\mathcal{F}}$-adapted and has $\boldsymbol{\mathcal{F}}$-compensator $M(t)$.*

Proof. The essence of this theorem is a generalization of the well-known result, crucial to many simulation algorithms, that if the random variable X has a continuous distribution function $F(x)$, then $Y = F(X)$ has a uniform distribution on the unit interval. We first restate this result in a form that will make the analogy more transparent.

Lemma 7.4.II. *Let X be a random variable with continuous distribution function $F(\cdot)$ and integrated hazard function $H(x) = -\log[1 - F(x)]$. Then $Y = H(X)$ has a unit exponential distribution (i.e. with unit mean).*

Conversely, if Y is a random variable with unit exponential distribution, then $X = H^{-1}(Y)$ has distribution function $F(\cdot)$.

If, therefore, we have a sequence of interval lengths $X_1, X_2, \ldots$ with continuous distributions $F_1(t), F_2(t), \ldots$, the corresponding sequence of transformed random variables $Y_1 = H_1(X_1)$, $Y_2 = H_2(X_2), \ldots$ is a sequence of unit exponential random variables. Now recall the construction for the conditional intensity function as an amalgam of hazard functions $h_n(u \mid t_1, \ldots, t_{n-1})$ in equation (7.2.3), and set $F_1(x) = 1 - \exp[-H_1(x)]$, $F_2(x) = 1 - \exp[-H_2(x)], \ldots$, where for brevity of notation we have written $H_n(x) = \int_0^x h_n(u \mid t_1, \ldots, t_{n-1})\,\mathrm{d}u$. If the intervals $X_1, X_2, \ldots$ represent the sequence of intervals for a point process with conditional intensity function $\lambda^*(t)$ that can be represented in terms of integrated hazard functions as above, then the joint distribution of any finite sequence of these intervals is the product of the distribution functions $F_i(t)$, and the joint distribution of the corresponding transformed random variables $H_1(X_1), H_2(X_2), \ldots$ is the product of unit ex-

ponential distributions and therefore represents the joint distribution of a set of i.i.d. unit exponential random variables. But such a point process is just a unit-rate Poisson process.

This argument lies behind a possible proof of the direct part of the theorem in the case where $\boldsymbol{\mathcal{F}}$ is the internal history of $N(\cdot)$. The converse part, again for the special case of the internal history, follows by a reversed argument using the converse part of the lemma.

The proof in the general case requires the same kind of attention to questions of predictability that we have mentioned in earlier discussions of the conditional intensity function and its integral. We sketch the general argument below, leaving a fuller discussion to Chapter 14.

Under the stated conditions, $\Lambda^*(t)$ and its inverse are both continuous, so that the process $\widetilde{N}$, like N itself, can increase only by unit jumps. It is also clear that the family of σ-algebras $\mathcal{F}_t$ is mapped into the family of σ-algebras $\mathcal{G}_t = \mathcal{F}_{\Lambda^{*-1}(t)}$, say, for the transformed process. (A rigorous definition of these, and a strict proof, requires use of the optional-sampling theorem as in Appendix A3.3.III.) Furthermore,

$$\begin{aligned} \mathrm{E}[\mathrm{d}\widetilde{N}(t) \mid \mathcal{G}_{t-}] &= \mathrm{E}\big[\mathrm{d}N\big(\Lambda^{*-1}(t)\big) \;\big|\; \mathcal{F}_{\Lambda^{*-1}(t)}\big] \\ &\approx \lambda^*\big(\Lambda^{*-1}(t)\big)\,\mathrm{d}\big(\Lambda^{*-1}(t)\big) = \mathrm{d}t, \end{aligned} \tag{7.4.2}$$

which shows that the process $\widetilde{N}$ has the lack-of-memory property and is therefore the Poisson process (Theorem 2.2.III).

The converse is a further application of the optional-sampling theorem. Since each $T = M(t)$ is a stopping time for $N_0(t)$, the σ-algebras $\mathcal{F}_t = \mathcal{G}_{M(t)}$ are well defined, and $N(t) = N_0\big(M(t)\big)$ is $\boldsymbol{\mathcal{F}}$-adapted. Note the crucial importance that $\mathcal{G}$ should contain the history of the process N_0—indeed, the minimal form of the theorem requires only that $M(t)$ be adapted to the internal history of N_0. $N(t)$ is also a.s. finite and monotonically increasing with unit jumps; hence, it defines a simple point process. The optional-sampling theorem and the martingale property for $N_0(t) - t$ then imply that, for $t > s$, $T = M(t) > S = M(s)$,

$$\mathrm{E}[N(t) - M(t) \mid \mathcal{F}_s] = \mathrm{E}[N_0(T) - T \mid \mathcal{G}_S] = N_0(S) - S = N(s) - M(s).$$

Thus, $N(t) - M(t)$ is an $\boldsymbol{\mathcal{F}}$-martingale, from which it follows that M must be the $\boldsymbol{\mathcal{F}}$-compensator for N. □

Because of this result, a simple point process with continuous compensator is sometimes called *a process of Poisson type*. The theorem implies that all such processes can be derived from a simple Poisson process by a random time transformation.

Example 7.4(a) *Renewal process* [see Exercises 7.2.3(a) and 7.2.11]. We consider an ordinary renewal process started with an event at the origin. We know from Exercise 7.2.11 that the conditional intensity function for this

process is just the hazard function for the interval distribution, evaluated at the backward recurrence time B_t, namely the time elapsed since the most recent event before the present time t. Also, the compensator $A(\cdot)$ satisfies

$$A(t) = A(t - B_t) - \log[1 - F(B_t)]\,.$$

On the transformed time scale, the time interval τ from one event to the next is given by $\tau = -\log[1 - F(X)]$, where X is the length of the interval on the original time scale. As in Lemma 7.4.II, the transformation takes successive intervals into a sequence of i.i.d. exponentially distributed intervals (i.e. into a unit-rate Poisson process). The general case with internal history is a generalization of this argument to the situation where the distributions of successive intervals are conditioned by the previous history of the process. □

The requirement in Theorem 7.4.I that the compensator $\Lambda^*(t)$ should increase without bound ensures that there is no last point in the process. The basic result remains valid without it, except insofar as the final interval is then infinite and so cannot belong to a unit-rate Poisson process. The extreme case in the next example makes the point.

EXAMPLE 7.4(b) *One-point process* (see Exercises 7.2.2 and 7.4.1). Let a point process on $(0, \infty)$ have exactly one point, at t_1, say, where $\Pr\{t_1 \le x\} = F(x)$, and we assume that the d.f. F is continuous. Then

$$\Lambda^*(t) = \int_0^{\min(t,t_1)} \frac{\mathrm{d}F(u)}{1 - F(u)} = -\log\big(1 - F[\min(t, t_1)]\big).$$

The initial interval transforms, as in the previous example, to an interval with unit exponential distribution; the transformed process then terminates. □

The converse part of Theorem 7.4.I contains within it the basis for one general approach to simulating point processes. Using the notation $X_1, X_2, \ldots$, $F_1(\cdot), F_2(\cdot), \cdots$ and $H_1(\cdot), H_2(\cdot)$ as in the proof of that theorem and Lemma 7.4.II, it may be summarized as follows.

Algorithm 7.4.III. *Simulation of point processes by the inverse method.*

1. Simulate a sequence $Y_1, Y_2, \ldots$ of unit exponential random variables (respectively, a sequence $U_1, U_2, \ldots$ of uniform $U(0,1)$ random variables).
2. Transform to the sequence of successive interval lengths $X_1 = H_1^{-1}(Y_1)$, $X_2 = H_2^{-1}(Y_2), \ldots$ (respectively, the sequence $F_1^{-1}(U_1)$, $F_2^{-1}(U_2), \ldots$).
3. Form the point process $(t_1, t_2, \ldots)$ by setting $t_1 = X_1, t_2 = X_1 + X_2, \ldots$.

The use of exponential or uniform random variables to initiate the algorithm is immaterial in that both lead to point processes with identical properties. The use of the exponential variates shows more clearly the relation to the Poisson process and may be marginally more convenient when the process is specified through its conditional intensity function because $t_1, t_2, \ldots$ then solve the successive equations

$$\int_0^{t_1} \lambda^*(u)\mathrm{d}u = Y_1, \qquad \int_{t_1}^{t_2} \lambda^*(u)\mathrm{d}u = Y_2,$$

and so on. The main constraint in the use of this algorithm is the common need to introduce an iterative numerical method to find the inverse of the integrated hazard or distribution function.

In principle, the method may be extended to situations where the interval distributions are conditioned by external as well as internal variables, provided that all the relevant conditioning information is available at the beginning of each new interval. □

A second important application of Theorem 7.4.I is the technique sometimes referred to as *point process residual analysis* (see e.g. Ogata, 1988); it uses the time transformation in testing the goodness-of-fit of a point process model. It depends on the fact that if the compensator used for the transformation is that of the true model, then the transformed process will be unit-rate Poisson, whereas if the wrong compensator is used, the transformed process will show some systematic departure from the unit-rate Poisson process. This means that the problem of testing for goodness-of-fit for a given, perhaps quite complex, model can be reduced to the well-studied and much simpler problem of testing for a unit-rate Poisson process (e.g. Cox and Lewis, 1966).

This device fills what is otherwise something of a gap for point process inference. While estimation and model comparison procedures can be based on standard likelihood methods, and a variety of statistical tests on specific characteristics, such as the interval lengths or the second-order properties of count numbers, are also available [the now classical monograph by Cox and Lewis (1966) remains an excellent introduction to a range of techniques of this kind], the one feature not obviously present there is a general purpose goodness-of-fit test for assessing the adequacy of a model overall.

Before outlining the method, we present a minor rephrasing and extension of the basic theorem.

Proposition 7.4.IV. *Let* $\{0 < t_1 < t_2 < \cdots\}$ *be an unbounded, increasing sequence of time points in the half-line* $(0, \infty)$, N^* *a simple point process with internal history* $\mathcal{H}$, *and monotonic, continuous* $\mathcal{H}$*-compensator* $\Lambda^*(t)$ *such that* $\Lambda^*(t) \to \infty$ a.s. *Then, with probability* 1, *the transformed sequence* $\{\tau_i = \Lambda^*(t_i)\}$ *is a realization of a unit-rate Poisson process if and only if the original sequence* $\{t_i\}$ *is a realization from the point process defined by* $\Lambda^*(t)$.

Proof. This proposition extends Theorem 7.4.I by incorporating the assertion that the character of the transformed process can (with probability 1) be unambiguously determined from a realization on the half-line $\mathbb{R}_+$. This can be regarded as a consequence of the ergodic theorem (see Chapter 12): for a stationary process, the probability of any of the events appearing in the fidi distributions can be recovered as a limiting ratio. If the processes are not identical, there must be at least one such event to which the two processes ascribe different probabilities. Thus, the limiting ratios, and hence the observation sequence, must be able to discriminate between the two processes. Granted this assertion, the result is a corollary of Theorem 7.4.I. □

Now suppose there is given a realization $\{t_1, \ldots, t_{N(T)}\}$ on a finite observation interval $(0, T)$ to which has been fitted a point process model with compensator $\Lambda^*(t)$. The procedure outlined below makes use of Proposition 7.4.II to define a goodness-of-fit test for point process models for which the conditional intensity function, and hence the compensator, is explicitly known.

Algorithm 7.4.V. *Goodness-of-fit test based on the residual point process.*

1. Form the transformed time sequence $\{\tau_i = \Lambda^*(t_i),\ i = 1, \ldots, N(T)\}$.
2. Plot the cumulative step-function $Y(x)$ through the points $(x_i, y_i) = (\tau_i/T, i/N(T))$ in the unit square $0 \le x, y \le 1$.
3. Plot confidence lines $y = x \pm Z_{1-\alpha/2}/\sqrt{T}$, where with Φ denoting the standard normal distribution function, $\Phi(Z_p) = p$.
4. Implement an approximate $100(1-\alpha)\%$ test of the hypothesis that the $\{\tau_i\}$ come from a unit-rate Poisson process by observing whether the empirical process $Y(x)$ falls outside the confidence band drawn in step 3.

At step 4, this procedure uses the maximum deviation from the expected rate curve in the transformed time domain to check for departures from the unit rate expected for the data in the transformed time domain. It is analogous in this context to the Kolmogorov–Smirnov test. The test is approximate in two respects. First, it is a large sample test, based on the Brownian motion approximation to the Poisson process. Second, and perhaps more importantly, it does not take into account the effect of estimating the parameters from the same data as are used to check the model. While both are typical large sample approximations, the bias resulting from the latter in moderate-sized data sets may be considerable, as shown for example in Schoenberg (2002), particularly when the process has strong time-dependence features that reduce the effective amount of information available in the data.

As with any portmanteau test, the test above has the further disadvantage, offset by its wide range of applicability, that its effectiveness (power) against different types of alternatives may be very variable. For more specific alternatives, there are many other tests of Poisson character that could be substituted for the Kolmogorov–Smirnov-type test suggested above [see e.g. Cox and Lewis (1966), as already noted]. Such tests are likely to be more powerful than the test above when the nature of the expected deviation from Poisson character is known. □

One advantage of the residual analysis is that it leads to a visual display (step 2 in the algorithm above) that can be useful in gaining a qualitative impression of the goodness-of-fit whether or not a formal test is applied. Ogata has made ingenious uses of this feature for visually detecting departures from a standard model, as illustrated below.

EXAMPLE 7.4(c) *Use of residual analysis to detect the return to normal background activity.* The rate of occurrence of events in aftershock sequences to (large) earthquakes is traditionally modelled by a Poisson process whose intensity function decays as a power law, known as the modified Omori law

in the seismology literature,

$$\lambda(t) = A/(t+c)^{1+p} \qquad (t > 0),$$

where A, c and p are nonnegative parameters and p is commonly close to zero. It is a delicate question to determine the time at which the aftershocks merge indistinguishably into the general background activity for the region. Leaving aside the problem of defining precisely what is meant by this statement, the visual pattern can be much enhanced by first transforming the time scale by the compensator $\Lambda^*(t) = (A/p)[c^{-p} - (t+c)^{-p}]$, $(t \geq 0)$ of the model above. When the rate of aftershock activity has decayed to about the level of the background activity, the dominant factor in the observed rate changes from the aftershock decay term to the steady background rate, increasing the observed rate above what would be expected from modelling the aftershock sequence. The change point is hard to pinpoint visually on the original time scale, but on the transformed time scale, it shows up relatively clearly as a deviation above the diagonal $y = x$ near the end of the observation sequence. See e.g. Ogata (1988) and Utsu *et al.* (1995) for illustrations and further details. □

Residual analysis can also be adapted to more specific problems as below.

EXAMPLE 7.4(d) *Using the ETAS model to test for relative quiescence in seismic data.* At shallow depths (0–20 km or so), the ETAS model of Example 6.4(d) usually provides a reasonable first approximation to the time–magnitude history of moderate or small-size earthquake events in an observation region. For this reason, departures from the ETAS model, or changes in its apparent parameter values, can be used as an indicator of anomalous seismic activity that may be associated with the genesis of a forthcoming large event. In particular, a reduction in activity below that anticipated by the ETAS model may signify the onset of a period of *seismic quiescence*, a much debated indicator of a larger event. The task of searching for changes in rate is here complicated by the high level of clustering characteristic of earthquake activity, which makes the evaluation of appropriate confidence levels particularly difficult. Again, the task can be much facilitated by first transforming the occurrence times according to the best-fitting ETAS model and then carrying out the change-point test on the transformed data. The problem is then reduced to that of testing for a change point in a constant-rate Poisson process, a relatively straightforward and well-studied problem. Ogata (1988, 1992, 2001) has developed detailed procedures, including a modification to the usual AIC criterion, to take into account the nonstandard character of the change-point problem (the additional parameters are absent in the null hypothesis rather than being fitted to a special numerical value; Davies' (1987) work on the problem of hypothesis testing when parameters vanish under H_0 is pertinent). Some further details are given in the exercises. Exercise 7.4.2 indicates extensions to the marked point process case. □

As with the other procedures we have illustrated in this chapter, the results on random time changes can be generalized relatively straightforwardly to other types of evolutionary point processes (notably multivariate and marked point processes) but only with more difficulty to spatial point patterns (see Chapter 14). We indicate below the extensions to multivariate and marked point processes; for more discussion, see e.g. Brown and Nair (1988). These extensions hinge on the uniqueness of the compensator with respect to the internal history $\boldsymbol{\mathcal{H}}$; see Proposition 7.3.IV for regular MPPs.

Consider first a multivariate point process. Here each component could be transformed by its own compensator, as a result of which we would obtain a multivariate Poisson process in which each component has unit rate. But would these components then be independent?

The answer to this question depends crucially on the histories used to define the compensators. If the full internal history is used for each component, then any dependence between the original components is taken into account and a Poisson process with independent, equally likely components is obtained. On the other hand, if each component is transformed according to its own internal history, the components of the resulting multivariate Poisson process will have equal (unit) rates but in general will not be independent. The next example provides a simple illustration.

EXAMPLE 7.4(e) *Bivariate Poisson process* [see Example 7.3(a)]. The model consists of an initial stream of input points from a Poisson process at constant rate λ and an associated stream of output points formed by delaying the initial points by random times exponentially distributed with mean $1/\mu$ independently for each initial point. Integrating the full $\boldsymbol{\mathcal{H}}$-conditional intensities at (7.3.2), the corresponding compensators are for component I a line of constant slope λ and for component II a broken straight line, with segments whose slopes are nonnegative multiples of μ, the breaks in the line occurring at the points of both processes, the slope increasing by μ whenever a component I point occurs and decreasing by μ whenever a component II point occurs.

The transformed points from component I are identical with the original points apart from an overall linear change of scale. The time transformation for component II is more complex: the distances between points are stretched just after a component I point and shrunk after a component II point. Further, if for any t all points of component I have been cleared (i.e. their associated component II points have already occurred), the transformed time remains fixed until the next component I point arrives. In this way, the dependence between the two components is broken, and both component processes are transformed into unit-rate Poisson processes.

A similar conclusion holds even if either or both components is augmented by the addition of the points from an independent Poisson process or processes: the relative scales of the time changes compensate for any differences in the original component rates, producing always a unit rate in the trans-

formed process, while any dependence between the two components is still broken as explained above. □

Consider now the case of a regular MPP. If the support of the mark distribution is no longer finite, then effectively we have an infinite family of different components; clearly it is not possible to turn them all into unit-rate Poisson processes and hope to retain an MPP as output. To achieve such a result, at least the rates of the components should be adjusted to produce a transformed process with finite ground rate. Here is one way of proceeding.

Suppose that the $\boldsymbol{\mathcal{H}}$-conditional intensity of the original process can be represented in the form

$$\lambda^*(t,\kappa) = \lambda^*_{\mathrm{g}}(t) f^*(\kappa \mid t),$$

where $f^*(\kappa \mid t)$ is a probability density with respect to the reference measure $\ell_{\mathcal{K}}(\cdot)$, which we take here to be itself a probability measure so that $\int_{\mathcal{K}} \ell_{\mathcal{K}}(\mathrm{d}\kappa) = \int_{\mathcal{K}} f(\kappa \mid t)\, \ell_{\mathcal{K}}(\mathrm{d}\kappa) = 1$. Let $A(t,U) = \int_U \int_0^t \lambda^*(s,\kappa)\, \mathrm{d}s\, \ell_{\mathcal{K}}(\mathrm{d}\kappa)$ be the full $\boldsymbol{\mathcal{H}}$-compensator for the process, and write $A_\kappa(t) = \int_0^t \lambda^*(s,\kappa)\, \mathrm{d}s$. To avoid complications in defining the inverse functions, we suppose both $\lambda^*_{\mathrm{g}}(t)$ and $f^*(\kappa \mid t)$ are strictly positive for all t and κ.

Now consider the transformation that takes the pair (t,κ) into the pair $(A_\kappa(t),\kappa)$. We claim that the transformed process is a stationary compound Poisson process with unit ground rate and mark distribution $\ell_{\mathcal{K}}(\cdot)$. To establish this result, we appeal to the uniqueness theorem for compensators (Proposition 7.3.IV). The crucial computation, corresponding to equation (7.4.2), is

$$\mathrm{E}[\widetilde{N}(\mathrm{d}\tau \times \mathrm{d}\kappa)] = \mathrm{E}[N(\mathrm{d}y \times \mathrm{d}\kappa)] \approx \lambda^*(y,\kappa)\, \mathrm{d}y\, \ell_{\mathcal{K}}(\mathrm{d}\kappa) = \mathrm{d}\tau\, \ell_{\mathcal{K}}(\mathrm{d}\kappa),$$

where $y = A_\kappa^{-1}(\tau)$, so that $\mathrm{d}y = \mathrm{d}\tau / \lambda^*(y,\kappa)$. The last form can be identified with the compensator for a stationary compound Poisson process with ground rate $\tilde{\lambda}_{\mathrm{g}} = 1$ and mark distribution $\ell_{\mathcal{K}}(\cdot)$. The uniqueness theorem completes the proof.

The results for both multivariate and marked point processes are summarized in the following proposition (a more careful discussion of the arguments above is given in Chapter 14).

Proposition 7.4.VI. (a) *Let* $\{N_j(t): j = 1,\dots,J\}$ *be a multivariate point process defined on* $[0,\infty)$ *with a finite set of components, full internal history* $\boldsymbol{\mathcal{H}}$*, and left-continuous* $\boldsymbol{\mathcal{H}}$*-intensities* $\lambda^*_j(t)$*. Suppose that for* $j = 1,\dots,J$*, the conditional intensities are strictly positive and that* $\Lambda^*_j(t) = \int_0^t \lambda^*_j(s)\, \mathrm{d}s \to \infty$ *as* $t \to \infty$*. Then, under the simultaneous random time transformations*

$$t \mapsto \Lambda^*_j(t), \qquad (j = 1,\dots,J)$$

the process $\{(N_1(t),\dots,N_J(t)): t \geq 0\}$ *is transformed into a multivariate Poisson process with independent components each having unit rate.*

(b) *Let $N(t,\kappa)$ be an MPP defined on $[0,\infty)\times\mathcal{K}$, where $\mathcal{K}$ is a c.s.m.s. with Borel sets $\mathcal{B}_{\mathcal{K}}$ and reference probability measure $\ell_{\mathcal{K}}(\cdot)$, and let $\boldsymbol{\mathcal{H}}$ denote the full internal history. Suppose that the $\boldsymbol{\mathcal{H}}$-conditional intensity $\lambda^*(t,\kappa) = \lambda_{\mathrm{g}}^*(\kappa)f^*(\kappa \mid t)$ exists, is $\ell_{\mathcal{K}}$-a.e. left-continuous in t and strictly positive on $[0,\infty)\times\mathcal{K}$, and that $\Lambda_\kappa^*(t) = \int_0^t \lambda^*(s,\kappa)\,\mathrm{d}s \to \infty$ as $t\to\infty$ $\ell_{\mathcal{K}}$-a.e. Then, under the random time transformations*

$$(t,\kappa)\mapsto(\Lambda_\kappa^*(t),\kappa),$$

the MPP N is transformed into a compound Poisson process $\widetilde{N}$ with unit ground rate and stationary mark distribution $\ell_{\mathcal{K}}(\cdot)$.

Example 7.4(f) *ETAS model* [see Example 6.4(d)]. This can serve as a typical example of a process with unpredictable marks. The conditional intensity factorizes into the form [see equation (7.3.10)]

$$\lambda^*(t,\kappa) = \left(\lambda_0 + \nu\int_{(-\infty,t)\times\mathcal{K}} \mathrm{e}^{\alpha(\chi-\kappa_0)}g(t-s)\,N(\mathrm{d}s\times\mathrm{d}\chi)\right)f(\kappa) \equiv \lambda_{\mathrm{g}}^*(t)f(\kappa),$$

where $f(\cdot)$, the density of the magnitude distribution, is commonly assumed to have an exponential form on $\mathcal{K}=[0,\infty)$. For stationarity, we require $\rho = \nu\int_0^\infty \mathrm{e}^{\alpha\kappa}f(\kappa)\,\mathrm{d}\kappa < 1$. Under these conditions, it is natural to take the reference measure on $\mathcal{K}$ to be f itself, in which case all the densities relative to the reference measure are equal to unity. Consequently, the multiple time changes here all reduce to the same form:

$$(t,\kappa)\mapsto(\Lambda_{\mathrm{g}}^*(t),\kappa), \quad \text{where } \Lambda_{\mathrm{g}}^*(t) = \int_0^t \lambda_{\mathrm{g}}^*(s)\,\mathrm{d}s.$$

In other words, under the random time change associated with the ground process, the original ETAS process is transformed into a compound Poisson process with unit ground rate and stationary mark density f. Such transformations open the way to corresponding extensions of the procedures described earlier for testing the process. In particular, checking the constancy of the mark distribution simplifies the detection of changes in the relative rates of events of different magnitudes.

Similar remarks apply to other examples with unpredictable marks, such as the stress-release models of Examples 7.2(g) and 7.3(d). □

Schoenberg (1999) gives a random-time change for transforming spatial point processes to Poisson.

Exercises and Complements to Section 7.4

7.4.1 Consider a two-point process t_1, t_2, $(t_1 < t_2)$ on $[0,T]$, where $(t_1, t_2 - t_1)$ has continuous bivariate d.f. $F(t,u)$. Find the compensator and define the random time change explicitly in terms of F. The Poisson process here has to be conditioned on the occurrence of two points within the interval $[0,T]$. [*Hint*: Example 7.4(b) treats the one-point case.]

7.4.2 *Marked point process extension of Algorithm* 7.4.III. Following the discussion around equation (7.3.2), suppose there is given a family of conditional hazard functions $h_n(u \mid (t_1, \kappa_1), \ldots, (t_{n-1}, \kappa_{n-1}))$ and corresponding conditional mark distributions $f_n(\kappa \mid (t_1, \kappa_1), \ldots, (t_{n-1}, \kappa_{n-1}); u)$. Formulate in detail a sequence of simulation steps to solve successively the pairs of equations

$$\int_{t_{n-1}}^{t_n} h_n(u \mid (t_1, \kappa_1), \ldots, (t_{n-1}, \kappa_{n-1}))\, \mathrm{d}u = Y_n\,,$$
$$\int_{0}^{\kappa_n} f_n(\kappa \mid (t_1, \kappa_1), \ldots, (t_{n-1}, \kappa_{n-1}); u)\, \mathrm{d}\kappa = U_n\,.$$

7.4.3 (*Continuation*). Using steps analogous to the simulation argument above, provide an alternative, constructive proof of Proposition 7.4.VI.

7.4.4 *Extension of Ogata's residual analysis to multivariate and marked point processes.* Develop algorithms, analogous to those in Algorithm 7.4.V, for testing multivariate and marked point processes.
[*Hint*: In the multivariate case, test both (a) that the ground process for the transformed process is a unit-rate Poisson process and (b) that the marks are i.i.d. with equal probabilities. In the marked case, take the reference measure to be, say, a unit exponential distribution, and replace (b) with a test for a set of i.i.d. unit exponential variates.]

7.5. Simulation and Prediction Algorithms

In the next two sections, we broach the topics of simulation, prediction, and prediction assessment. In modelling, the existence of a logically consistent simulation algorithm for some process is tantamount to a constructive proof that the process exists. Furthermore, simulation methods have become a key component in evaluating the numerical characteristics of a model, in checking both qualitative and quantitative features of the model, and in the centrally important task of model-based prediction. A brief survey of the principal approaches to point process simulation and of the theoretical principles on which these approaches are based therefore seemed to us an important complement to the rest of the text.

This section provides a brief introduction to simulation methods for evolutionary models; that is, for models retaining a time-like dimension that then dictates the probability structure through the conditional intensity function. Simulation methods can be developed also for spatial point patterns (see Chapter 15), but considerable conceptual simplicity results from the ability to order the evolution of the process in 'time'. The growth in importance of Markov chain Monte Carlo methods for simulating spatial processes is a tacit acknowledgement of the fact that such methods introduce an artificial time dimension even into problems where no such dimension is originally present.

Two general approaches are commonly used for simulating point processes in time. The first we have already considered in Algorithm 7.4.III; it involves simulating the successive intervals, making use of the description of the

conditional intensity function as a family of hazard functions as in equation (7.2.3). Its main disadvantage as a general method is that it requires repeated numerical solution of the equation defining the inverse. The thinning methods outlined in the present section, by contrast, require only evaluations of the conditional intensity function. Although the difference in computational time between these two methods is not huge, it is the main reason why the thinning method is given greater prominence in this section. In addition, the theoretical basis behind thinning methods is of interest in its own right.

The most important theoretical result is a construction, originating in Kerstan (1964) and refined and extended in Brémaud and Massoulié (1996), that has something of the character of a converse to Proposition 7.4.I. There we transformed a point process with general conditional intensity to a Poisson process; here we convert a Poisson process back into a process with general conditional intensity. For this purpose, we use an auxiliary coordinate in the state space, so we consider a unit-intensity Poisson process, $\widetilde{N}$ say, on the product space $\mathcal{X} = \mathbb{R} \times \mathbb{R}_+$. The realizations of $\widetilde{N}$ consist of pairs (x_j, y_j). Also, let $\mathcal{H}_t$ denote the σ-algebra of events defined on a simple point process over the interval $[0, t)$ and $\boldsymbol{\mathcal{H}}$ the history $\{\mathcal{H}_t\}$. The critical assumption below is that λ^* is $\boldsymbol{\mathcal{H}}$-adapted.

Proposition 7.5.I. *Let $\widetilde{N}$, $\boldsymbol{\mathcal{H}}$ be defined as above, let $\lambda^*(t)$ be a nonnegative, left-continuous, $\boldsymbol{\mathcal{H}}$-adapted process, and define the point process N on $\mathbb{R}$ by*

$$N(\mathrm{d}t) = \widetilde{N}\big(\mathrm{d}t \times (0, \lambda^*(t)]\big). \tag{7.5.1}$$

Then N has $\boldsymbol{\mathcal{H}}$-conditional intensity $\lambda^*(t)$.

Proof. Arguing heuristically, it is enough to note that

$$\mathrm{E}[N(\mathrm{d}t) \mid \mathcal{H}_{t-}] = \mathrm{E}\big[\widetilde{N}\big(\mathrm{d}t \times (0, \lambda^*(t-)]\big) \mid \mathcal{H}_{t-}\big] = \lambda^*(t)\,\mathrm{d}t. \qquad \square$$

There is no requirement in this proposition that the conditional intensity be a.s. uniformly bounded as was required in the original Shedler–Lewis algorithm. When such a bound exists, it leads to straightforward versions of the thinning algorithm, as in Algorithm 7.5.II below.

The result can be further extended in various ways, for example to situations where more general histories are permitted or where the initial process is not Poisson but has a conditional intensity function that almost surely bounds that of the process to be simulated; see Exercises 7.5.1–2.

Example 7.5(a) *Standard renewal process on* $[0, \infty)$. We suppose the process starts with an event at $t = 0$. Let $h(u)$ denote the hazard function for the lifetime distribution of intervals between successive points, so that [see Exercise 7.2.3(a)] the conditional intensity function has the form

$$\lambda^*(t) = h(t - t_{N(t)}) \qquad (t \geq 0),$$

where $t_{N(t)}$ is the time of occurrence of the last event before time t. However, $\lambda^*(t)$ should be defined on the history of $\widetilde{N}$ rather than on N. To this end, we first define the sequence of points t_i in terms of $\widetilde{N}$. With $t_0 = 0$, define sequentially

$$t_{n+1} = \min\{x_i : x_i > t_n \text{ and } y_i < h(x_i - t_n)\} \qquad (n = 0, 1, \ldots)$$

and then define $\lambda^*(t)$ as above. Notice that the right-hand side of this expression is $\mathcal{F}_t$-measurable and the whole process is $\boldsymbol{\mathcal{F}}$-adapted. □

Thinning algorithms generally follow much the same lines as in Proposition 7.5.I and the example above. The main difficulty arises from the range of y_i being unbounded, which provides a flexibility that is difficult to match in practice. The original Shedler–Lewis algorithm (Lewis and Shedler, 1976; see also Exercise 2.1.6) was for an inhomogeneous Poisson process in a time interval where the intensity is bounded above by some constant, M say. Then, the auxiliary dimension can be taken as the bounded interval $(0, M)$ rather than the whole of $\mathbb{R}_+$, or equivalently the y_i could be considered i.i.d. uniformly distributed random variables on the interval $(0, M)$. Equivalently again, the time intensity could be increased from unity to M and the y_i taken as i.i.d. uniform on $(0, 1)$, which leads to the basic form of the thinning algorithm outlined in the algorithm below.

In discussing the simulation algorithms below, it is convenient to introduce the term *list-history* to stand for the actual record of times, or times and marks, of events observed or simulated up until the current time t. We shall denote such a list-history by H, or H_t if it is important to record the current time in the notation. Thus, a list-history H is just a vector of times $\{t_1, \ldots, t_{N(t)}\}$ or a matrix of times and marks $\{(t_1, \kappa_1), \ldots, (t_{N(t)}, \kappa_{N(t)})\}$. We shall denote the operation of adding a newly observed or generated term to the list-history by $H \mapsto H \cup t_j$ or $H \mapsto H \cup (t_j, \kappa_j)$. In the discussion of conditioning relations such as occur in the conditional intensity, the list-history H_t bears to the σ-algebra $\mathcal{H}_t$ a relationship similar to that between an observed value x of a random variable X and the random variable X itself.

The algorithms require an extension of Proposition 7.5.I to the situation where the process may depend on an initial history $\mathcal{H}_0$; we omit detail but note the following. Such a history will be reflected in the list-history by a set of times or times and marks of events observed prior to the beginning of the simulation. This is an important feature when we come to prediction algorithms and wish to start the simulation at the 'present', taking into account the real observations that have been observed up until that time. It is also important in the simulation of stationary processes, for which the simulation may be allowed to run for some initial period $(-B, 0)$ before simulation proper begins. The purpose is to allow the effects of any transients from the initial conditions to become negligible. Finding the optimal length of such a preliminary 'burn-in' period is an important question in its own right. Its solution depends on the rate at which the given process converges toward equilibrium

from the initial state, but in general this is a delicate question that is affected by the choice of initial state as well as decay parameters characteristic of the process as a whole.

Suppose, then, that the process to be simulated is specified through its conditional intensity $\lambda^*(t)$, that there exists a finite bound M such that

$$\lambda^*(t) \le M \quad \text{for all possible past histories,}$$

and that the process is to be simulated over a finite interval $[0, A)$ given some initial list-history H_0.

Algorithm 7.5.II. *Shedler–Lewis Thinning Algorithm for processes with bounded conditional intensity.*

1. Simulate $x_1, \ldots, x_i$ according to a Poisson process with rate M (for example, by simulating successive interval lengths as i.i.d. exponential variables with mean $1/M$), stopping as soon as $x_i > A$.
2. Simulate $y_1, \ldots, y_i$ as a set of i.i.d. uniform $(0, 1)$ random variables.
3. Set $k = 1,\ j = 1$.
4. If $x_k > A$, terminate. Otherwise, evaluate $\lambda^*(x_k) = \lambda(x_k \mid H_{x_k})$.
5. If $y_k \le \lambda^*(x_k)/M$, set $t_j = x_k$, update H to $H \cup t_j$, and advance j to $j+1$.
6. Advance k to $k+1$ and return to step 4.
7. The output consists of the list $\{j;\ t_1, \ldots, t_j\}$. □

This algorithm is relatively simple to describe. In the more elaborate versions that appear shortly, it is convenient to include a *termination condition* (or conditions), of which steps 1 and 4 above are simple. In general, we may need some limit on the number of points to be generated that lies outside the *raison d'être* of the algorithm.

While this algorithm works well enough in its original context of fixed intensity functions, its main drawback in applications to processes with random conditional intensities is the need for a bound on the intensity that holds not only over $(0, A)$ but also over all histories of the process up to time A. To meet this difficulty, Ogata (1981) suggested a sequential variant of the algorithm that overcomes this difficulty, requiring only a local boundedness condition on the conditional intensity. A minor variant of his approach is outlined in Algorithm 7.5.IV.

For the sake of clarity, we return to the representation of the conditional intensity function in terms of successive hazard functions, much as in Definition 7.2.II, but allowing all such functions to depend on an initial history $\mathcal{H}_0$, namely

$$h_n(s \mid \mathcal{H}_0, t_1, \ldots, t_{n-1}), \qquad \text{for } 0 < t_1 < \cdots < t_{n-1} < s \le A.$$

For every t in $(0, A)$ and associated σ-algebra $\mathcal{H}_t$, we suppose there are given two quantities, a local bound $M(t \mid \mathcal{H}_t)$ and a time interval of length $L(t \mid \mathcal{H}_t)$, satisfying the following conditions.

Condition 7.5.III. *There exist functions $M(t \mid \mathcal{H}_t)$, $L(t \mid \mathcal{H}_t)$ such that, for all initial histories $\mathcal{H}_0$, all $t \in [0, \infty)$, for every $n = 1, 2, \ldots$, and all sequences $t_1, \ldots, t_{n-1}$ with $0 < t_1 < \cdots < t_{n-1} < t$, the hazard functions satisfy*

$$h_n(t + u \mid \mathcal{H}_0, t_1, \ldots, t_{n-1}) \le M(t \mid \mathcal{H}_t) \qquad (0 \le u < L(t \mid \mathcal{H}_t)\,).$$

Placing the bound on the hazard function is equivalent to placing the bound on the conditional intensity function under the constraint that no additional points of the process occur in the interval $(t, t+u)$ under scrutiny. As soon as a new point does occur, in general the hazard function will change and a new bound will be required. Thus, the bound holds until either the time step $L(\cdot)$ has elapsed or a new point of the process occurs. For the algorithm below, the list-history H_t consists of $\{H_0, t_1, \ldots, t_{N(t)}\}$, where $N(t)$ is the number of points t_i satisfying $0 \le t_i < t$. For brevity, we mostly write $M(t)$ and $L(t)$ for $M(t \mid H_t)$ and $L(t \mid H_t)$. Ogata (1981) gives extended discussion and variants of the procedure.

Algorithm 7.5.IV. *Ogata's modified thinning algorithm.*

1. Set $t = 0$, $i = 0$.
2. Stop if the termination condition is met; otherwise, compute $M(t \mid H_t)$ and $L(t \mid H_t)$.
3. Generate an exponential r.v. T with mean $1/M(t)$ and an r.v. U uniformly distributed on $(0, 1)$.
4. If $T > L(t)$, set $t = t + L(t)$ and return to step 2.
5. If $T \le L(t)$ and $\lambda^*(t + T)/M(t) > U$, replace t by $t + T$ and return to step 2.
6. Otherwise, advance i by 1, set $t_i = t + T$, replace t by t_i, update H to $H \cup t_i$, and return to step 2.
7. The output is the list $\{i;\ t_1, \ldots, t_i\}$. □

The technical difficulties of calculating suitable values for $M(t)$ and $L(t)$ vary greatly according to the character of the process being simulated. In an example such as a Hawkes process, at least when the hazard functions decrease monotonically after an event, it would be enough in principle to consider only $t = t_i$ (i.e. points of the process) and set $M(t_i) = \lambda^*(t_i+)$. This leads to a very inefficient algorithm, however, since the hazard decreases rapidly and a large number of rejected trial points could be generated. A simple modification is to set $M(t) = \lambda^*(t)$ and $L(t) = \frac{1}{2}\lambda^*(t+)$, irrespective of whether or not t is a point of the process. Such a choice gives a reasonable compromise between setting the bound too high, and so generating excessive trial points, and setting it too low, thus requiring too many iterations of step 3.

The next example is a process with an increasing hazard, where the intervention of step 3 is virtually mandatory.

Example 7.5(b) *Self-correcting or stress-release model.* We discuss the simulation of the model of Example 7.2(g). As described there, points $\{t_i\}$ occur

at a rate governed by the conditional intensity function

$$\lambda(t) = \Psi[X(t)],$$

where $X(t)$ is an unobserved Markov jump process that increases linearly between jump times t_i at which it decreases by an amount κ_i, so that

$$X(t) = X(0) + \nu t - \textstyle\sum_{i:t_i<t} \kappa_i.$$

Given an initial history $\mathcal{H}_0$, we can now simulate the process using Algorithm 7.5.IV as, for example, we could take $L(t) = 2/\Psi[X(t)]$ and $M(t) = \Psi[X(t) + \nu L(t)]$. With high probability, the next event would occur within twice the mean interval length at the start of the interval, and because of the increasing nature of the hazard function, a simple bound would be its value at the end of the search interval. □

Algorithm 7.5.IV can be extended to cover the situation where the evolution of the conditional hazard function depends on additional random processes, themselves evolving jointly with the given point process. The immediate requirements are for the existence of explicit algorithms for calculating the intensity and for finding local bounds $L(\cdot)$ and $M(\cdot)$ that take into account current and past values of the auxiliary variables. A deeper difficulty, however, relates to the need to simulate forward not only the point process but also the auxiliary variables on which it depends. For auxiliary variables that change only slowly, this may not be a serious handicap, but for longer-term predictions, a full model is needed from which the point process and auxiliary variables can be jointly simulated.

Extension of the simulation algorithms to marked point processes, including even space–time processes, presents no significant difficulty. Once again, the evolutionary character of the process makes a sequential approach straightforward and natural. First, a candidate for the next time point of the process is selected and either accepted or rejected by the thinning algorithm using the full $\boldsymbol{\mathcal{H}}$-intensity for the overall sequence of time points. Once a new time point is selected, the corresponding mark, whether a weight, a spatial location, or some further characteristic, is simulated, using the conditional density $f^*(\kappa \mid t)$ for the mark distribution. The situation is particularly simple if the process has independent or unpredictable marks, as the mark distribution is then independent of the history of the process. In general, the mark distribution can depend on the past history of the process, including both the past locations and marks, and the simulation will be tractable provided this dependence can be captured in a reasonably simple explicit manner.

For convenience, an outline algorithm is summarized more formally below. In it, we use the same notation as for Algorithm 7.5.IV. The local bounds $M(t)$ and $L(t)$ must be chosen for the full internal intensity $\lambda_g^*(t)$ of the ground process. Subject to replacing $\lambda^*(t)$ by $\lambda_g^*(t)$ at step 5, the first part of the algorithm is just a restatement of the steps in Algorithm 7.5.IV. Note that

we have paid particular attention to the need to update the list-history H. If simulation is to be applied to point process prediction, it is essential to allow the history at time 0 (corresponding to the present) to be nontrivial, in this case including all relevant information on observations of the actual process up to the time when simulation commences.

Algorithm 7.5.V. *Thinning algorithm for marked point processes.*

1. Set $t = 0$, $i = 0$, $H_0 = \emptyset$.
2. Stop if the termination condition is met. Otherwise, calculate $M(t)$, $L(t)$ for the ground intensity $\lambda^{\mathrm{g}}_{\mathcal{H}}(t)$.
3. Generate an exponential r.v. T with mean $1/M(t)$ and an r.v. U uniformly distributed on $(0, 1)$.
4. If $T > L(t)$, set $t = t + L(t)$, update the list-history H, and return to step 2.
5. If $T \leq L(t)$ and $\lambda^*_0(t + T)/M(t) < U$, replace t by $t + T$, update the list-history H, and return to step 2.
6. Advance i by 1, set $t_i = t+T$, replace t by t_i, and generate a mark κ_i from the distribution with density $f(\kappa \mid t_i)$.
7. Update the list-history H to $H \cup (t_i, \kappa_i)$, and return to step 2.
8. The output is the list $\{i;\ (t_1, \kappa_1), \ldots, (t_i, \kappa_i)\}$. □

In Example 7.5(b) above, for example, simulation proceeds as if the process has nonanticipating marks until step 6 is reached, at which point the appropriate value $\phi[X(t)]$ must be read into the simulation routine for producing values according to the tapered Pareto distribution. By way of illustrating Algorithm 7.5.V, we consider the extension of Example 7.5(b) to the linked stress-release model.

Example 7.5(c) *Simulating the linked stress-release model* [see Example 7.3(d)]. In this model, there are two types of marks: the region in which the event occurs (as a surrogate for spatial location) and the size of the event. The basic form of the conditional intensity is given in equation (7.3.14).

A key step in the simulation is updating the list-history. This will consist of a matrix or list type object with one column for each coordinate of the events being described: here the times t_i, their regions K_i, and their magnitudes M_i. When the simulation is started, the list-history may contain information from real or simulated data from the past in order to allow the simulation to join 'seamlessly' onto the past. Each time a new event is simulated, its coordinates are added to the list-history.

Since the simulation of the next event depends only on the form of the conditional intensity, as determined by the current list-history, and additional random numbers, it can proceed on an event-by-event basis. First, the time of the next event in the ground process is simulated, then the region is selected with probabilities proportional to the relative values of the conditional intensities for the different regions at that time, and then a magnitude is selected from the standard magnitude distribution (this distribution is fixed in the standard model, but it can also be made stress- or region-dependent). □

The prediction of point processes, in all but a few very special cases where explicit algorithms are available, goes hand-in-hand with simulation. The quantities that one would like to predict, such as the time to the next event, the probability of an event occurring within a given interval in the future, or the costs caused by events in the future, are commonly nonlinear functionals of the future of the process. They rarely fall into any general category for which analytic expressions are available. Since, on the other hand, simulation of a point process is relatively straightforward once its conditional intensity function is known, and moreover can be extended to situations where an arbitrary initial history can be incorporated into the conditional intensity, it is indeed natural to see prediction as an application and extension of the preceding procedures.

Suppose there is given a realization of the point process on some finite interval (a, b). To link up with the preceding algorithms, we identify the origin $t = 0$ with the end point b of the interval so that, in our earlier notation, the realization on (a, b) forms part of the initial history H_0. Suppose for the sake of definiteness that our aim is to predict a particular quantity V that can be represented as a functional of a finite segment of the future of the process. To fulfil our aim, we estimate the distribution of V.

An outline of a prediction procedure is as follows.

1. Choose a time horizon $(0, A)$ sufficient to encompass the predicted quantity of interest (we need not insist here that A be a fixed number, provided the stopping rule is clearly defined and can be incorporated into the simulation algorithm).
2. Simulate the process forward over $(0, A)$ using the known structure of the conditional intensity function and initial history H_0.
3. Extract from the simulation the value V of the functional that it is required to predict.
4. Repeat steps 2 and 3 sufficiently often to obtain the required precision for the prediction.
5. The output consists of the empirical distribution of the values of V obtained from the successive simulations.

In step 5 above, it is often convenient to summarize the empirical distribution by key characteristics, such as its mean, standard deviation, and selected quantiles. Not all prediction exercises fit exactly into this schema, but many are variations on it.

EXAMPLE 7.5(d) *Prediction of a Wold process with exponential intervals* [see Exercise 4.5.8 and Example 4.6(b)]. In the notation used previously, let an interval preceded by an interval of length x have parameter $\lambda(x)$ [and hence mean $1/\lambda(x)$]. Suppose that we wish to predict the time X_0 to the next event and the length X_1 of the ensuing complete interval, given the current list-history consisting of the times $t_0, t_{-1}, \ldots$ of the preceding events, where 0 denotes the present time so $0 > t_0 > t_{-1} > \cdots$.

The quantity V of the preceding discussion is the pair (X_0, X_1). The particular specification of the model here implies that the joint density function of (X_0, X_1) equals

$$\lambda\big(|t_0 - t_{-1}|\big)\mathrm{e}^{-\lambda(|t_0-t_{-1}|)X_0}\lambda\big(|t_0| + X_0\big)\mathrm{e}^{-\lambda(|t_0|+X_0)X_1};$$

then simulation via the model should lead to a joint histogram that in principle is an approximation to this function.

For pragmatic purposes, we may be satisfied with the first moments

$$\mathrm{E}(X_0 \mid H_0) = \frac{1}{\lambda\big(|t_0 - t_{-1}|\big)}$$

and

$$\mathrm{E}(X_1 \mid H_0) = \int_0^\infty \frac{\lambda(|t_0 - t_{-1}|)}{\lambda(|t_0| + u)}\,\mathrm{e}^{-\lambda(|t_0-t_{-1}|)u}\,\mathrm{d}u.$$

Exercises and Complements to Section 7.5

7.5.1 *Extended form of Proposition* 7.5.I. Let $\boldsymbol{\mathcal{F}}$ be a history on $[0, \infty)$, $\lambda_1(t)$, $\lambda_2(t)$ be two nonnegative, left-continuous (or more generally predictable), history-dependent candidates for conditional intensity functions, and let $N^*(\mathrm{d}t \times \mathrm{d}s)$ be an $\boldsymbol{\mathcal{F}}$-adapted unit-rate Poisson process on $\mathbb{R}_+ \times \mathbb{R}$ that is unpredictable in the sense that its evolution for $s > t$ is independent of the history up to t. Let $N(t)$ on $\mathbb{R}_+$ consist of the time coordinates t_i from those points of N^* lying in the region

$$\min\{\lambda_1(t), \lambda_2(t)\} < s < \max\{\lambda_1(t), \lambda_2(t)\}.$$

Then N is $\boldsymbol{\mathcal{F}}$-adapted and has conditional intensity $|\lambda_1(t) - \lambda_2(t)|$.
[In most cases, as in Proposition 7.5.I, the history will be that generated by the Poisson process itself, but the generalization opens the way to conditioning on external variables. See Brémaud and Massoulié (1996) and Massoulié (1998).]

7.5.2 *Extension of thinning Algorithm* 7.5.II. In the setup for Algorithm 7.5.II, suppose that the x_i are simulated from a process with conditional intensity $\lambda^+(t)$ that satisfies a.s.

$$\lambda^+(t) \geq \lambda^*(t) \qquad (0 < t < T)$$

and that the thinning probability at time t is equal to the ratio $\lambda^*(t)/\lambda^+(t)$. Show that the thinned process is again the point process with intensity $\lambda^*(t)$. [See Ogata (1981).]

7.5.3 *Simulation algorithms for Boolean models.* Devise a simulation procedure for the Boolean model of Example 6.4(d) with a view to describing distributions of functionals such as the intensity function or a joint intensity ('correlation').

7.5.4 Show how Algorithm 7.5.V can be applied to a pure linear birth process.

7.5.5 *Simulation of cluster processes.* Brix and Kendall (2002) describe a technique for the *perfect simulation* of a cluster point process in a given region A (hence, the simulations have no edge effects—this is an analogue of having no 'burn-in' period). The crucial step is to replace the parent process N_c, say, by a process which has at least one offspring point in the observation region.

7.6. Information Gain and Probability Forecasts

We come now to the problem of assessing probability forecasts of the type described in the previous section. A distinction needs to be made here between assessing the probability forecast as such and assessing a decision procedure based on the probability forecast. Commonly, when probability forecasts for weather and other phenomena are being assessed, a threshold probability level is established, and the forecast is counted as a 'success' if either the forecast probability rises above the threshold level and a target event occurs within the forecasting period or region or the forecast probability falls below the threshold level and no event occurs. The assessment is then based on the 2×2 table of observed and forecast successes and failures, and a variety of scores for this purpose have been developed and studied (see e.g. Shi *et al.*, 2001). In effect, such a procedure converts the probability forecast into a decision rule, and it is the decision rule rather than the forecast that is assessed. In fact, many decision rules can be based on the same probability forecast, depending on the application in view. For example, in earthquake forecasts, one relevant decision for a government might be whether or not to issue a public earthquake warning; but other potential users, such as insurance companies, emergency service coordinators, and managers of gas, power, or transport companies, might prefer to initiate actions at quite different probability levels and would therefore score the forecasts quite differently. Our concern is with assessing the probability forecasts as such.

The basic criterion we shall use for this purpose is the binomial or entropy score, in which the forecast is scored by the negative logarithm $-\log \hat{p}_k$ of the forecast probability $\hat{p}_k$ of the outcome k that actually occurs. If outcome k has true probability p_k of occurring, then a 'good' set of forecasts should have $\hat{p}_k \approx p_k$ for outcome k, and therefore the expected score is approximately $-\sum_k p_k \log p_k$, which is just the entropy of the distribution $\{p_k\}$ (up to a multiplicative factor in not using logarithms to base 2). This leads us to a preliminary discussion of the entropy of point process models, a study taken further in Chapter 14. The entropy score itself, summed over a sequence of forecasts based on a specific parametric model, is nothing other than the log likelihood of the model. In this sense, the discussion highlights an alternative interpretation of the likelihood principle. Maximizing the likelihood from within a family of models amounts to finding the model with the best forecast performance in the sense of the entropy score. Equally, testing the model on the basis of its forecasting performance amounts to testing the model on the basis of its likelihood. Other criteria, such as the goodness-of-fit of first- and second-moment properties, may be less relevant to selecting a model for its forecasting ability. In any case, the analysis and assessment of probability forecasts is a topic of importance in its own right, and it is this point of view that motivates the present discussion.

To bring some of the underlying issues into focus, consider first the simpler problem of producing and assessing probability forecasts for a sequence of

i.i.d. multinomial trials in which observation Y_i, for $i = 1, \ldots, N$, may have one of K different outcomes $1, \ldots, K$, say, with respective true probabilities $p_k = \Pr\{\text{outcome is } k\} = \Pr\{Y_i = k\}$ (the trials are i.i.d.) and $\sum_{k=1}^{K} p_k = 1$. Suppose that there is a record available of observations $\{Y_1, \ldots, Y_N\}$ on N independent trials, and write $\hat{p}_k = N^{-1} \sum_{i=1}^{N} \delta_{Y_i\, k} \equiv N_k/N$ for the sample proportion of outcomes equal to k $(k = 1, \ldots, K)$. What should be our forecast for trial $N + 1$?

In accordance with our general prescription, the forecast should be in the form of a set of probabilities based on an assumed model (i.e. a model for which the underlying probabilities are assumed known). In this simple situation, it is intuitively obvious that the $\{p_k\}$ are also the probabilities that we would use to forecast the different possible outcomes of the next event. However, it is also possible to base this choice on somewhat more objective grounds, namely that our choice should maximize some *expected score*, suitably chosen.

Denote the candidate probabilities for the forecast by a_k. In accordance with the discussion above, we consider here the *likelihood ratio score*

$$S_{\mathrm{LR}} = \sum_{i=1}^{N} \log \frac{a_{Y_i}}{\pi_{Y_i}} = N \sum_{k=1}^{K} \hat{p}_k \log \frac{a_k}{\pi_k}, \tag{7.6.1}$$

where $\{\pi_k\}$ is a set of reference probabilities. The use of the logarithm of the ratio a_k/π_k rather than the simple logarithm $\log a_k$ has two benefits: it introduces a natural standard against which the forecasts using the given model can be compared; and it overcomes dimensionality problems in the passage from discrete to continuous contexts (Exercise 7.6.1 gives some further discussion). This score function has the character of a *skill score*, for which higher values show greater skills.

Taking expected values has the effect of replacing the empirical frequencies $\hat{p}_k$ by p_k in the second form of (7.6.1). Elementary computations then show that the score S_{LR} is optimized by the choice $a_k = p_k$; i.e. the procedure that optimizes the expected score is to use the model probabilities as the forecasting probabilities. Specifically, the optimum values achieved by following the procedure above are given by

$$\mathrm{E}(S_{\mathrm{LR}}) = NH(P; \Pi), \tag{7.6.2}$$

where P, Π denote the distributions with elements p_k, π_k, respectively, and $H(\cdot)$ is the *relative* or *generalized entropy* or *Kullback–Leibler distance* between the two distributions. The appearance of the entropy here should not come as a surprise, as it is nothing other than the expected value of (minus) a log probability, or more generally a log likelihood.

In terms of S_{LR}, the distribution that is hardest to predict is the discrete uniform distribution, which has maximum entropy amongst distributions on K points. If we use the uniform as the reference distribution $\{\pi_k\}$, the change in the expected score as the model distribution moves away from the maximum

entropy distribution will be referred to as the *expected information gain*. It represents the improvement in the predictability of the model used relative to the reference model. The greatest expected gains, corresponding to the most effective predictions, will be achieved when the model distribution is largely concentrated on one or a small number of distinguished values. The ratio p_k/π_k of the model probability p_k to the reference probability π_k for any particular distinguished value k is sometimes called the *probability gain for* k.

Now let us examine how these ideas carry over to the point process context. We start with a discrete-time framework, such as would arise if the forecasts were being made regularly, after the elapse of a fixed time interval (weekly, monthly, etc.). We also assume that the process is marked, with the marks taking one of the finite set of values $\{1, \ldots, K\}$. In effect, this merely extends the discussion from the case of independent to dependent trials, with the assumption that the trials are indexed by a time parameter so that the evolutionary character is maintained. Alternatively, and more conveniently for our purposes, we may consider the model as a multivariate point process in discrete time.

Rather than using the sequence of marks Y_n $(n = 1, 2, \ldots)$ as before, introduce $X_{kn} = \delta_{Y_n k}$, and let the K component simple point processes $N_k(n)$ count the number of points with mark k up to 'time' n, with $N_k(0) = 0$ for each k, so $N_k(n) = \sum_{i=1}^{n} X_{ki}$.

An argument similar to that given previously shows that the forecasting probability that optimizes the expected value of the score at step n, given the history $\mathcal{H}_{n-1}$ up to time $n-1$, is $p^*_{kn} = \mathrm{E}(X_{kn} \mid \mathcal{H}_{n-1})$, where $\boldsymbol{\mathcal{H}}$ is the full history of the process, recording information on the marks as well as the occurrence times. If, as a reference process, we take the process of i.i.d. trials having fixed probabilities $\pi_{kn} = f_k$, then the total entropy score over a period of T time units can be written

$$\log \frac{L}{L_0} = \sum_{n=1}^{T} \sum_{k=1}^{K} X_{kn} \log \frac{p^*_{kn}}{\pi_{kn}}\,, \tag{7.6.3}$$

which is just the likelihood ratio for the given process relative to the reference process. This formulation shows clearly that the total entropy score for the multivariate process is the sum of the entropy scores of the component processes. There is no implication here that the component processes are independent; dependence comes through the joint dependence of the components on the full past history.

In the case of a univariate process, for which the only possible outcomes are 0 and 1, the formula in (7.6.3) simplifies to the *binomial score*

$$\log \frac{L}{L_0} = \sum_{n=1}^{T} \left[X_n \log \frac{p^*_n}{\pi_n} + (1 - X_n) \log \frac{1 - p^*_n}{1 - \pi_n} \right]. \tag{7.6.4}$$

Equation (7.6.3) assumes a form closer to that used previously for the likelihood of a multivariate point process if we reserve one mark, 0 say, for

the null event; that is, the event that no event of any other type occurs. Let us assume in addition that the ground process is simple, so that at most one nonnull event can occur in any one time instant, and introduce the notations $p_n^* = \sum_k p_{kn}^*$ for the conditional intensity of the ground process, $f_{k|n}^* = p_{kn}^*/p_n^*$ for the conditional distribution of the mark, given the past history and the occurrence of an event at n, and $X_n = \sum_{k=1}^K X_{kn}$ for the ground process itself. Let us also choose the reference probabilities in the form $\pi_{kn} = f_k \pi_n$ for $k \neq 0$, $\pi_{0n} = 1 - \pi_n$, corresponding to a discrete-time analogue of a continuous-time compound Poisson process. Then we can rewrite (7.6.3) as

$$\begin{aligned}\log \frac{L}{L_0} &= \sum_{n=1}^T \left[\sum_{k=1}^K X_{kn} \log \frac{f_{k|n}^* p_n^*}{f_k \pi_n} + (1 - X_n) \log \frac{1 - p_n^*}{1 - \pi_n} \right] \\ &= \sum_{n=1}^T \left[X_n \log \frac{p_n^*}{\pi_n} + (1 - X_n) \log \frac{1 - p_n^*}{1 - \pi_n} + \sum_{k=1}^K X_{kn} \log \frac{f_{k|n}^*}{f_k} \right]. \quad (7.6.5)\end{aligned}$$

Taking expectations of the nth term, given the past up to time $n-1$, gives the conditional relative entropy or *conditional information gain*

$$\begin{aligned}I_n &= \sum_{k=1}^K p_{kn}^* \log \frac{p_{kn}^*}{p_{kn}} + (1 - p_n^*) \log \frac{1 - p_n^*}{1 - p_n} \\ &= p_n^* \log \frac{p_n^*}{p_n} + (1 - p_n^*) \log \frac{1 - p_n^*}{1 - p_n} + p_n^* \sum_{k=1}^K f_{k|n}^* \log \frac{f_{k|n}^*}{f_k}. \quad (7.6.6)\end{aligned}$$

It is the conditional relative entropy of the nth observation, given the information available prior to the nth step. Note that this quantity is still a random variable since it depends on the random past through the conditioning σ-algebra $\mathcal{H}_{n-1}$. It reduces to the zero random variable when $p_{kn}^* = \pi_{kn}$ but is otherwise positive, as follows from Jensen's inequality. In the special case of a univariate process, it reduces to

$$I_n^* = p_n^* \log \frac{p_n^*}{\pi_n} + (1 - p_n^*) \log \frac{1 - p_n^*}{1 - \pi_n}. \quad (7.6.7)$$

The relation

$$\mathrm{E}[(I_n^* + I_{n-1}^*) \mid \mathcal{H}_{n-1}] = \mathrm{E}(I_n^* \mid \mathcal{H}_{n-1}) + I_{n-1}^*$$

yields the joint conditional entropy of X_n and X_{n+1}, given the information available at the $(n-1)$th step. Continuing in this way, we obtain

$$\mathrm{E}\left[\sum_{n=1}^N I_n^* \,\middle|\, \mathcal{H}_0 \right] = \sum_{n=1}^N \mathrm{E}\big[I_n^* \mid \mathcal{H}_0 \big] = \mathrm{E}\left[\log \frac{L}{L_0} \,\middle|\, \mathcal{H}_0 \right], \quad (7.6.8)$$

the joint entropy of the full set of observations, conditional on the information available at the beginning of the observation period. Dividing this quantity

by N, we obtain the *average expected information gain per time step*. This quantity is of particular interest when the whole setup is stationary and the expectations in (7.6.8) have the same value, namely the expected information gain per unit time. We shall denote this quantity by G. In this situation, we expect the log likelihood to increase roughly linearly with the number of observations, with the expected increment being equal to G. To avoid difficulties with transient effects near $n = 0$, the histories in the stationary case should cover the infinite past rather than the past since some fixed starting time. Following the notation in later chapters, write $p_n^\dagger = \mathrm{E}[X_{n+1} \mid \mathcal{H}_{(-\infty,n]}]$ and set $\pi_n = \mathrm{E}(p_n^\dagger) = \mathrm{E}(X_n) = p$, say. Then, G can be expressed as

$$G = \mathrm{E}\left[p_n^\dagger \log \frac{p_n^\dagger}{p} + (1 - p_n^\dagger) \log \frac{1 - p_n^\dagger}{1 - p} + p_n^\dagger \sum_{k=1}^{K} \log \frac{f_k^\dagger}{f_k}\right]. \tag{7.6.9}$$

The first term represents the information gain from the ground process and the second the additional information gain that comes from predicting the values of the marks, given the ground process. Overall, G represents the expected improvement in forecasting skill, as measured by the entropy score, if we move from using the background probabilities as the forecast to using the time-varying model probabilities.

G ranges from 0, when the trials are i.i.d. and the model probabilities coincide with those of the reference model, to a maximum when the model trials are completely predictable, related to the absolute entropy of the independent trials model. To see this last point, suppose, to take a specific case, that the background model is for i.i.d. trials with equal probabilities $1/K$ for each outcome. Now write G in the form

$$G = \mathrm{E}\Bigg[\left[p_n^\dagger \log p_n^\dagger + (1 - p_n^\dagger) \log(1 - p_n^\dagger)\right] - \left[p_n^\dagger \log p + (1 - p_n^\dagger) \log(1 - p)\right] + p_n^* \sum_{k=1}^{K} f_{k|n}^\dagger \log \frac{f_{k|n}^\dagger}{f_k}\Bigg] \tag{7.6.10}$$

and suppose that with high probability, $p_n^\dagger$ is close to either one or zero and that one of the $f_{k|n}^\dagger$ is also close to one, so that the process is highly predictable. Then, both the first two terms in the first sum above are very small, while in the second sum either $p_n^\dagger$ itself is very small or it is close to one and the remaining sum is close to the value $-\log(1/K)$. After taking expectations, recalling $\mathrm{E}(p_n^\dagger) = p$, G reduces to approximately $-[p \log p + (1-p) \log(1-p) + p \log(1/K)]$, the absolute entropy of the independent trials model with equal probabilities for each outcome. In general, the final term will be of the form $p\,\mathrm{E}[\log f_{k^\dagger}]$, where $f_{k^\dagger}$ is the background probability of the outcome $k^\dagger$ that is successfully predicted.

In summary, we have the following statement.

Proposition 7.6.I. *For a stationary, multivariate, discrete-time process, with full internal history $\boldsymbol{\mathcal{F}}$, overall occurrence rate p, and background model as defined above, G, the expected information gain per time step, is given by* (7.6.9) *above. It is a characteristic of the model and lies in the range*

$$0 \le G \le -[p \log p + (1-p)\log(1-p) + p\,\mathrm{E}(\log f_{k^\dagger})],$$

where $f_{k^\dagger}$ is the background probability of the outcome $k^\dagger$ that is successfully predicted. G takes the lower end point of the range when the increments X_{nk} are independent and the upper end point when perfect prediction is possible.

EXAMPLE 7.6(a) *Discrete Hawkes process: logistic autoregression.* This model defines a univariate process in which p_n^* has the general form

$$\log \frac{p_n^*}{1-p_n^*} = a_0 + \sum_{i=1}^{K} a_i X_{n-i} = a_0 + \sum_{i=1}^{K} I_{\{X_{n-i}=1\}} a_i \,, \tag{7.6.11}$$

where the a_i are parameters and, to accommodate the stationarity requirement, $\boldsymbol{\mathcal{F}}$ is taken to be the *complete history* $\boldsymbol{\mathcal{H}}^\dagger$, so that $\mathcal{H}_n^\dagger$ is generated by the X_i with $-\infty < i \le n$.

For simplicity, we examine just the case of a first-order autoregression; there are then just two parameters, a_0 and a_1, in $1:1$ correspondence with the probabilities $\pi_{1|0} = \Pr\{X_n = 1 \mid X_{n-1} = 0\}$ and $\pi_{1|1} = \Pr\{X_n = 1 \mid X_{n-1} = 1\}$, respectively. Three extreme cases arise. If $\pi_{1|0}$ is close to 0 and $\pi_{1|1}$ is close to 1, then a realization will consist of long sequences of 0s followed by long sequences of 1s, and any prediction should approximate the weatherman's rule: tomorrow's weather will be the same as today's. If $\pi_{1|1}$ is close to 0 and $\pi_{1|0}$ is close to 1, then the realization will be an almost perfect alternation of 0s and 1s, and any prediction rule should approximate the anti-weatherman's rule: tomorrow's weather will be the opposite of today's. In the third case, $\pi_{1|0}$ and $\pi_{1|1}$ are both close to $\frac{1}{2}$, and the sequence will consist of more or less random occurrences of 0s and 1s, and no good prediction rule will be possible.

To examine such effects quantitatively, let us choose the parameters a_0, a_1 so that $\pi_{1|0}$ and $\pi_{1|1}$ can be written

$$\pi_{1|0} = \epsilon \,, \qquad \pi_{1|1} = 1 - \rho\epsilon \,.$$

The stationary probability p solves the equation $p = p\pi_{1|1} + (1-p)\pi_{1|0}$ so $p = 1/(1+\rho)$. Thus, the parameter ϵ controls the mean length of runs of the same digit, and the parameter ρ controls the relative probabilities of 0s and 1s. We examine the behaviour of the predictions for small ϵ. When $X_{n-1} = 0$, we take as our prediction $p_n^* = \pi_{1|0} = \epsilon$, and when $X_{n-1} = 1$ we take $p_n^* = \pi_{1|1} = 1 - \rho\epsilon$.

The information gain when $X_{n-1} = r$ for $r = 0, 1$ is then

$$J_r = \pi_{1|r} \log \frac{\pi_{1|r}}{p} + (1 - \pi_{1|r}) \log \frac{1 - \pi_{1|r}}{1-p} \,.$$

The expected information gain per forecast is $G = pJ_1 + (1-p)J_0$. Substituting for $\pi_{1|0}$, $\pi_{1|1}$ and p, we find that, for small ϵ,

$$G = H_p + 2\rho\epsilon \log \epsilon + O(\epsilon),$$

where H_p is as in Proposition 7.6.I. As ϵ decreases, the expected information gain approaches H_p, whereas if $\epsilon = 1/(1+\rho)$, then $\pi_{1|0} = \pi_{1|1} = 1/(1+\rho)$ and $G = 0$. □

We have stressed that the expected information gain is a function of the model: it is an indicator of its inherent predictability. In practice, other factors may intervene to produce an observed mean information gain that is well below that predicted by the model. This may happen, in particular, if the data are being fitted by a poor model. There would then be substantial long-run discrepancies between the actual data and the data that would be produced by simulation from the model. In such a case, the average information gain over a long sequence of trials could be well below the expected model value. In this sense, the mean information gain, representing the average likelihood per observation, forms the basis for a kind of goodness-of-fit test for the model.

We turn now to the problem of transferring these ideas to the continuous-time, point process context. In practice, forecasts cannot be issued continuously but only after intervals of greater or smaller length. We therefore adopt the following framework.

Suppose there is given a finite interval $(0, T)$ and a *partition* $\mathcal{T}$ into subintervals $\{0 < t_{\mathcal{T},1} < \cdots < t_{\mathcal{T},N} = T\}$. Forecasts are to be made at the end of each subinterval (i.e. at the time points $\{t_{\mathcal{T},k}\}$) for the probability of an event occurring in the next subinterval. Suppose further that the given partition is a member of a *dissecting family of partitions* $\mathcal{T}_n$ in the sense of Appendix A1.6: as $n \to \infty$, the norm $\|\mathcal{T}\| = \max |t_{\mathcal{T},k} - t_{\mathcal{T},k-1}| \to 0$ so that the partitions ultimately distinguish points of $(0, T)$, and the intervals appearing in the partitions are rich enough in total to generate the Borel sets of $(0, T)$. Our aim is to relate the performance of the forecasts on the finite partition to the underlying properties of the point process.

For this purpose, Lemmas A1.6.IV, on convergence to a Radon–Nikodym derivative, and A1.6.V, on the relative entropy of probability measures on nested partitions, play a key role. To apply these lemmas, we must relate the partitions of the interval $(0, T)$ to the partitions of the measurable space $(\Omega, \mathcal{E})$ on which the probabilities are defined. Here it is enough to note that a partition of the interval into N subintervals induces a partition of $(\Omega, \mathcal{E})$ into the $(K+1)^N$ events corresponding to all possible sequences obtained by noting whether or not the subinterval contains a point of the process and, if so, noting the mark of the first point occurring within the subinterval.

From Lemma A1.6.IV, it follows that, as the partitions are refined, the probability gains p^*_{nk}/π_{nk} converge $(\mathcal{P} \times \ell)$-a.e. to the corresponding ratio of intensities $\lambda^*(t,k)/\lambda_0(t,k)$. Lemma A1.6.V then implies that the corresponding relative entropies increase to a limit bounded above by the point process

relative entropy. The latter can be obtained directly by taking expectations of the point process likelihood ratio. Specifically, starting from the MPP log likelihood ratio at (7.6.5), taking expectations when the reference measure corresponds to a compound Poisson process with constant rate λ_0 and mark distribution f_k, the relative entropy $H(\mathcal{P}_T; \mathcal{P}_{0,T})$ equals

$$\mathrm{E}\left[\sum_{k=1}^{K}\int_0^T \lambda^*(t,k)\log\frac{\lambda^*(t,k)}{\lambda_0 f_k}\,\mathrm{d}t - \int_0^T [\lambda_{\mathrm{g}}^*(t)-\lambda_0]\,\mathrm{d}t\right] \tag{7.6.12}$$

$$= \mathrm{E}\left[\int_0^T \lambda_{\mathrm{g}}^*(t)\log\frac{\lambda_{\mathrm{g}}^*(t)}{\lambda_0}\,\mathrm{d}t - \int_0^T [\lambda_{\mathrm{g}}^*(t)-\lambda_0]\,\mathrm{d}t + \int_0^T \lambda_{\mathrm{g}}^*(t)\sum_{k=1}^{K} f_k^*(t)\log\frac{f_k^*(t)}{f_k}\,\mathrm{d}t\right], \tag{7.6.13}$$

where $\lambda_{\mathrm{g}}^*(t)$ is the conditional intensity for the ground process.

A proof of this result for the univariate case when $\boldsymbol{\mathcal{H}}$ is the internal history and the likelihood reduces to the Janossy density is outlined in Exercise 7.6.3. The general case, as well as a more complete discussion of the convergence of the p_{nk}^* to $\lambda^*(t,k)$, is taken up in Chapter 14.

When the process is stationary and λ^* is replaced by $\lambda^\dagger$ (i.e. the conditioning is taken with respect to the infinite past), the relative entropy in (7.6.12) reduces to a multiple of T. If further we assume that $\lambda_0 = \mathrm{E}[\lambda_{\mathrm{g}}^\dagger(0)] \equiv m_{\mathrm{g}}$, then (7.6.12) can be written

$$H(\mathcal{P}_T; \mathcal{P}_{0,T}) = T\left(\mathrm{E}\left[\lambda_{\mathrm{g}}^\dagger(0)\log\frac{\lambda_{\mathrm{g}}^\dagger(0)}{\lambda_0}\right] + m_{\mathrm{g}}\mathrm{E}\left[\sum_{k=1}^{K}\log\frac{f_{k|0}^\dagger}{f_k}\right]\right). \tag{7.6.14}$$

Again, we can write G for the coefficient of T and refer to it as the mean entropy or *expected information gain per unit time*. It is worth noting that here G can be written in the two alternative forms

$$G = \mathrm{E}\left[\lambda_{\mathrm{g}}^\dagger(0)\log\frac{\lambda_{\mathrm{g}}^\dagger(0)}{\lambda_0}\right] + m_{\mathrm{g}}\mathrm{E}\left[\sum_{k=1}^{K}\log\frac{f_{k|0}^\dagger}{f_k}\right] = \sum_{k=1}^{K}\mathrm{E}\left[\lambda_k^\dagger(0)\log\frac{\lambda_k^\dagger(0)}{\lambda_k}\right],$$

where $\lambda_k^\dagger(0) = \lambda_{\mathrm{g}}^\dagger(0) f_{k|0}^\dagger$ and $\lambda_k = m_{\mathrm{g}} f_k$. The first form represents a division of the information gain into components due to forecasting the occurrence times of the points and their marks, while the second represents a division of the information gain into components corresponding to the individual marks. This equality does not hold in general for the approximating discrete-time processes because the two forms then correspond to different ways of scoring situations where more than one point of the process falls into a single time step.

As in the discrete case, the quantity G is a characteristic of the model. It represents an upper bound to the expected information gains per unit time that could be obtained from any approximating discrete model. The results are summarized in the proposition below.

Proposition 7.6.II. *Let $N(t,\kappa)$ be a stationary regular MPP, let*

$$\lambda^\dagger(t,\kappa)\,\mathrm{d}t = \lambda_{\mathrm{g}}^\dagger(t) f_{\kappa|t}^\dagger\,\mathrm{d}t = \mathrm{E}[\mathrm{d}_t N(t,\kappa) \mid \mathcal{H}_{t-}^\dagger]$$

denote its complete $\boldsymbol{\mathcal{H}}^\dagger$-conditional intensity, and suppose that

$$G = \mathrm{E}\left[\lambda_{\mathrm{g}}^\dagger(0)\log\frac{\lambda_{\mathrm{g}}^\dagger(0)}{m_{\mathrm{g}}}\right] < \infty,$$

where $m_{\mathrm{g}} = \mathrm{E}[\lambda_{\mathrm{g}}^\dagger(0)]$. If $\mathcal{T}$ is any finite partition of the interval $(0,T)$ and $G_{\mathcal{T}}$ the associated average expected information gain per unit time, then $G_{\mathcal{T}} \le G$ and, as $\mathcal{T}_n$ increases through any nested sequence of partitions generating the Borel sets in $(0,T)$, $G_{\mathcal{T}_n} \uparrow G^\dagger \equiv \lim_{n\to\infty} G_{\mathcal{T}_n} \le G$.

Proof. The result follows from further applications of Lemmas A1.6.IV and A1.6.V, but a formal proof requires a more careful discussion of conditioning and predictability than given here and is deferred to Chapter 14. □

Since G here is a property of the model, it can be evaluated analytically or numerically (by simulation). The model value of G can then be compared with the mean likelihood $T^{-1}\log L$ obtained by applying the model to a set of data, this latter being just the mean entropy score per unit time for the given model with the given data. If the model is close to the true model for the data, the estimate of G obtained in this way should be close to the model G. When the data do not match the model well, the predictive power of the model should be below that obtained when the model is applied to matching data and hence below the theoretical G of the model. In such a situation, the estimated G from the likelihood will generally come out well below the true G of the model (as well as below the unknown G of the true model). The difference between the model and estimated values of G can therefore serve as a basis for model testing and is in fact so used in contingency table contexts, corresponding roughly to the discrete time-models considered earlier in this section.

Some of these points are illustrated in the following two examples.

Example 7.6(b) *Renewal process.* Consider a stationary renewal process with interval distribution having density $f(x)$, assumed at least left-continuous. Then

$$\lambda^\dagger(t) \;=\; f(B_t)/S(B_t),$$

where B_t has the distribution of a stationary backward recurrence time. For the mean rate and the expected information gain per unit time, we obtain, respectively,

$$m = \mathrm{E}[\lambda^\dagger(t)] = \mathrm{E}\left[\frac{f(B_t)}{S(B_t)}\right],$$

$$G = \mathrm{E}\left[\lambda^\dagger(t)\log\frac{\lambda^\dagger(t)}{m}\right] = \mathrm{E}\left[\frac{f(B_t)}{S(B_t)}\log\frac{f(B_t)}{mS(B_t)}\right], \tag{7.6.15}$$

the two expectations on the extreme right-hand sides being with respect to the distribution of B_t, which has density $\int_y^\infty f(u)\,\mathrm{d}u/\mu$, where μ is the mean interval length [see (4.2.5) or Exercise 3.4.1]. Substituting and simplifying, we find $m = 1/\mu$ and

$$G = m\left[1 + \int_0^\infty f(y) \log \frac{f(y)}{m}\,\mathrm{d}y\right]. \tag{7.6.16}$$

The same result can be obtained from the general result that, for a stationary process, the expected information gain per unit time is just m times the expected information gain per interval, where the latter is defined to be

$$G_I = \mathrm{E}\left[\int_0^\infty f^\dagger(x) \log \frac{f^\dagger(x)}{f_0(x)}\,\mathrm{d}x\right],$$

with $f^\dagger(x)$ the density of the distribution of an interval given the history up to its start, and $f_0(x)$ is the density of an interval under the reference measure. Here, given m, the exponential distribution with mean $1/m$ has maximum entropy so we take $f_0(x) = m\mathrm{e}^{-mx}$ in the expression above, corresponding precisely to the choice of the Poisson process with rate m used in the counting process description.

Now suppose that probability forecasts are made for a forecasting period of length Δ ahead. The probability of an event occurring in the interval $(t, t+\Delta)$, given the past history $\mathcal{F}_t^\dagger$, is given by

$$p^*(\Delta \mid X) = \big[S(X) - S(X+\Delta)\big]/S(X),$$

say, where $S(x)$ is the survivor function for the interval distribution, and X is the backward recurrence time. In the stationary case, writing $p_0 = 1 - \mathrm{e}^{-m\Delta}$ and taking expectations with respect to the stationary form of the backward recurrence time distribution, we consider the quantity

$$\begin{aligned} G_\Delta &= \mathrm{E}[I_\Delta \mid \mathcal{H}_t^\dagger] \\ &= \frac{1}{\Delta}\mathrm{E}\left[p^*(\Delta \mid X) \log \frac{p^*(\Delta \mid X)}{p_0} + [1 - p^*(\Delta \mid X)] \log \frac{1 - p^*(\Delta \mid X)}{1 - p_0}\right]. \end{aligned} \tag{7.6.17}$$

It represents the average expected information gain for forecasts of length Δ, is independent of t and can be shown to satisfy $G_\Delta \le G = \lim_{\Delta\to 0} G_\Delta$. See Exercise 7.6.4 for details and some numerical illustrations. □

The next model both illustrates the ideas of Proposition 7.6.II in a relatively simple context and adds a cautionary note to the discussion of probability forecasts for point processes.

Example 7.6(c) *Marked Hawkes process with exponential infectivity function* [see Example 7.3(b)]. Consider an MPP with complete conditional intensity of the form

$$\lambda^{\dagger}(t,\kappa) = \left[\mu_0 + \sum_{\{i:t_i<t\}} \psi(\kappa_i)\beta \mathrm{e}^{-\beta(t-t_i)}\right] f(\kappa).$$

In common with the ETAS model where the marks κ are commonly denoted by M for magnitudes, it has unpredictable marks, and its ground intensity is just the term in square brackets above. The ground intensity can be written in the form

$$\lambda_{\mathrm{g}}^{\dagger}(t) = \mu_0 + A(t), \qquad A(t) = \sum_{\{i:t_i<t\}} \psi(\kappa_i)\beta \mathrm{e}^{-\beta(t-t_i)}.$$

Now, although the sum defining $A(t)$ goes back into the indefinite past, in fact it is a Markov process, its future evolution depending only on its present value (discounted exponentially in the gaps between events) and the sizes of future events that are chosen independently of the past.

Thus $\mathrm{E}[A(t)] = m_{\mathrm{g}}(1-\mu_0)$ and

$$G = \mathrm{E}\left[\big(\mu_0 + A(t)\big) \log \frac{\mu_0 + A(t)}{m}\right]$$

are both fully determined once the equilibrium distribution for the Markov process $A(t)$ is determined. In this example, the observed performance of predictions based on the true model is likely to be worse than predictions based on a Poisson process with the same mean rate m. This is because the rate in intervals between points is assessed as μ_0 by the model and as m by the Poisson process. When an event occurs, however, it is likely to be followed by several others within the same prediction interval, all of which are likely to be scored (badly). In fact, this is one example where the distinction between the scores S_{LR} at (7.6.1) and S_{Q} at Exercise 7.6.1 makes a crucial difference in the estimation of the performance of the model. A related example with numerical details from simulations is given in Vere-Jones (1999). □

Exercises and Complements to Section 7.6

7.6.1 As a possible alternative to the likelihood score S_{LR} in (7.6.1) for assessing probability forecasts, define the *quadratic score* S_{Q} by

$$S_{\mathrm{Q}} = \sum_{i=1}^{N}\left[\sum_{k=1}^{K}(\delta_{X_i,k} - a_k)^2\right] = N\left[1 - 2\sum_{1}^{K}\hat{p}_k a_k + \sum_{1}^{K}(a_k)^2\right].$$

Show that, just as for S_{LR}, the optimal result is achieved by using the model probabilities as the forecast probabilities. Show also that when these probabilities are used, $\mathrm{E}(S_{\mathrm{Q}}) = N[1 - K^{-1} - \mathrm{var}\, p_X]$, where $\mathrm{var}\, p_X = \sum(p_k - \bar{p})^2$ and $\bar{p} = \sum p_k/K = 1/K$.

7.6.2 (*Continuation*). Consider the effect on S_Q of the limit procedure that passes from a discrete probability to a continuous density. How should a reference measure be introduced so as to secure a meaningful passage to a limit?

7.6.3 *Entropy of a regular finite point process.*

(a) For a regular finite point process, define the point process entropy $H(\mathcal{P})$ as the expected value $\mathrm{E}[\log(L/L_0)]$ of the likelihood ratio. Express L in terms of Janossy densities, and use the representation (i) of Theorem 5.3.II to show (see Rudemo, 1964; McFadden, 1965) that $H(\mathcal{P})$ equals

$$-\sum_k p_k \log p_k - \sum_k \int \pi_k^{\mathrm{sym}}(x_1, \ldots, x_k) \log[k!\, \pi_k(x_1, \ldots, x_k)]\, \mathrm{d}x_1 \cdots \mathrm{d}x_k\,,$$

where $p_k = \Pr\{N(\mathcal{X}) = k\}$.

(b) Now take $\mathcal{X}$ to be the interval $(0, T)$ and represent the Janossy densities in terms of hazard functions and hence the internal conditional intensity. Hence, derive (7.6.14).

7.6.4 *Forecasts for renewal processes* [see Example 7.6(b)].

(a) Recall that the backward recurrence time has density $mS(x)$ in the notation of Example 7.6(b). Hence, simplify the expectation in (7.6.17) and verify the inequality for G_Δ using a convexity argument.

(b) *Uniformly distributed intervals.* Examine the special case

$$S(x) = \begin{cases} 1 - x & (0 < x < 1), \\ 1 & (x \geq 1). \end{cases}$$

Substitute in (7.6.15) and (7.6.16) and investigate the result in (a) numerically.

7.6.5 *Information gain for the Wold process with exponential intervals* [see Exercise 4.5.8 and Example 7.5(c)]. Using the earlier notation, show that the information gain per unit time can be expressed as

$$G = \mathrm{E}\left(\log \frac{\lambda_0}{\lambda(X)} \right),$$

where the expectation is over the stationary distribution for an interval length X, and $\lambda_0 = 1/\mathrm{E}(X)$.

CHAPTER 8

Second-Order Properties of Stationary Point Processes

Second-order properties are extremely important in the statistical analysis of point processes, not least because of the relative ease with which they can be estimated in both spatial and temporal contexts. However, there are several shortcomings when compared with, for example, the second-order properties of classical time series. There are ambiguities in the point process context as to just which second-order aspects of the process are in view. The second-order properties of the intervals, in a point process on $\mathbb{R}$, are far from equivalent to the second-order properties of the counts, as already noted in Chapter 3 and elsewhere. In this chapter, our concern is solely with random measure or counting properties, broadly interpreted.

A more important difficulty, however, is that the defining property of a point process—that its realizations are integer-valued measures—is not clearly reflected in properties of the moment measures. It does imply the presence of diagonal singularities in the moment measures, but this property is shared with other random measures possessing an atomic component. Nor does there seem to exist a class of tractable point processes, analogous to Gaussian processes, whose second-order properties are coextensive with those of point processes in general. Indeed, there are still open questions concerning the class of measures that can appear as moment measures for point processes or for random measures more generally. Gibbs processes defined by point–pair interactions come close to the generality required for a Gaussian process analogue but have neither the same appeal nor the same tractability as the Gaussian processes. Other examples, such as Hawkes processes, also come close to this role without fulfilling it entirely. Ultimately, these problems are related to the nonlinearity of key features of point processes such as positivity and integer counts. Thus, the second-order theory, with its associated toolkit of linear

prediction and filtering methods, although still important, is of less general utility for point processes than for classical time series.

Nevertheless, it seems worthwhile to set out systematically both the aspects of practical importance and their underpinning mathematical properties. Such a programme is the aim of the present chapter, which includes a discussion of both time-domain and frequency-domain techniques for second-order stationary point processes and random measures. Deeper theoretical issues, such as ergodicity, the general structure of moment measures for stationary random measures, and invariance under wider classes of transformations, are taken up in Chapter 12. Spatial processes are treated briefly here, reappearing in Chapters 12 and 15.

To avoid encumbering the main text with tools and arguments that are hardly used elsewhere in the book, the main technical arguments relating to the Fourier transforms of second-moment measures are placed in the final section, Section 8.6.

We shall assume throughout the chapter that the basic point processes are *simple*. For multivariate and marked point processes, we take this to mean that the ground process is simple. As we have already remarked in Chapter 6, there is no significant loss of generality in making this assumption since the batch size in a nonsimple point process can always be treated as an additional mark and the properties of the original process derived from those for marked point processes.

8.1. Second-Moment and Covariance Measures

Second-order properties of stationary processes have already made brief appearances in Section 3.5 and Proposition 6.1.I. Here we take as our starting point the second and third properties listed in Proposition 6.1.I. For the purposes of this chapter, these can be restated as follows.

Proposition 8.1.I (Stationary random measure: Second-order moment structure)**.** *Let ξ be a stationary random measure on $\mathcal{X} = \mathbb{R}^d$ for which the second-order moment measure exists.*

(a) *The first-moment measure $M_1(\cdot)$ is a multiple of Lebesgue measure $\ell(\cdot)$; i.e. $M_1(\mathrm{d}x) = m\,\ell(\mathrm{d}x)$ for a nonnegative constant m, the* mean density.

(b) *The second-moment measure $M_2(\cdot)$ is expressible as the product of a Lebesgue component $\ell(\mathrm{d}x)$ along the diagonal $x = y$ and a reduced measure, $\breve{M}_2(\mathrm{d}u)$ say, along $u = x - y$, or in integral form, for bounded measurable functions f of bounded support,*

$$\int_{\mathcal{X}^{(2)}} f(s,t)\, M_2(\mathrm{d}s \times \mathrm{d}t) = \int_{\mathcal{X}} \int_{\mathcal{X}} f(x, x+u)\, \ell(\mathrm{d}x)\, \breve{M}_2(\mathrm{d}u). \tag{8.1.1a}$$

In particular, by taking $f(x,y) = I_{\mathbb{U}^d}(x) I_B(y-x)$,

$$\breve{M}_2(B) = \mathrm{E}\left[\int_{\mathbb{U}^d} \xi(x+B)\, \xi(\mathrm{d}x)\right]. \tag{8.1.1b}$$

A point process or random measure for which the first- and second-moment measures exist and satisfy (a) and (b) of Proposition 8.1.I will be referred to as being *second-order stationary*. We should note, however, that a point process for which the first- and second-order moments satisfy the stationarity assumptions above is not necessarily stationary: nonstationary processes can have stationary first and second moments (see Exercises 8.1.1 and 8.1.2).

We retain the accent $\breve{}$ to denote reduced measures formed by dropping one component from the moment measures of stationary processes as a consequence of a factorization of the form (8.1.1). Thus, $\breve{M}_{[2]}(\cdot)$, $\breve{C}_2(\cdot)$, and $\breve{C}_{[2]}$ stand, respectively, for the reduced forms of the second factorial moment measure, covariance measure, and factorial covariance measure. A proof of such factorization can be based on the observation that, under stationarity, $M_2(\mathrm{d}x, \mathrm{d}(x+u))$ is independent of x and so should have the form $\ell(\mathrm{d}x)\times Q(\mathrm{d}u)$ for some measure $Q(\cdot)$; see Chapter 12 and Proposition A2.7.III for details and background.

Our principal aim in this section is to study the properties of these reduced measures and the relations between their properties and those of the point processes or random measures from which they derive. We start with a discussion of $\breve{M}_2$, which is arguably the most fundamental if not always the most convenient of the various forms.

Proposition 8.1.II. *Let $\breve{M}_2(\cdot)$ be the reduced second-moment measure of a nonzero, second-order stationary point process or random measure ξ on $\mathbb{R}^d$ with mean density m. Then $\breve{M}_2$ is*

(i) symmetric: $\breve{M}_2(A) = \breve{M}_2(-A)$;

(ii) positive: $\breve{M}_2(A) \geq 0$, *with strict inequality at least when $0 \in A$ and either ξ has an atomic component or A is an open set;*

(iii) positive-definite: *for all bounded measurable functions ψ of bounded support,*

$$\int_{\mathbb{R}^d} (\psi * \psi^*)(x)\, \breve{M}_2(\mathrm{d}x) \geq 0\,, \tag{8.1.2}$$

where

$$\psi * \phi(x) = \int_{\mathbb{R}^d} \psi(y)\overline{\phi(x-y)}\,\mathrm{d}y, \qquad \psi^*(x) = \overline{\psi(-x)};$$

(iv) translation-bounded: *for every bounded Borel set A in $\mathbb{R}^d$, there exists a finite constant K_A such that*

$$\breve{M}_2(x + A) \leq K_A \qquad (\text{all } x \in \mathbb{R}^d). \tag{8.1.3}$$

If also ξ is ergodic and the bounded convex Borel set A increases in such a way that $r(A) = \sup\{r\colon A \supseteq S_r(0)\} \to \infty$, where $S_r(0)$ denotes the ball in $\mathbb{R}^d$ of radius r and centre at 0, then in this limit, for all bounded Borel sets B,

$$\breve{M}_2(A) \,/\, \ell(A) \to m^2 \tag{8.1.4}$$

and

$$\frac{1}{\ell(A)} \int_A \xi(x + B)\, \xi(\mathrm{d}x) \to \breve{M}_2(B) \quad \text{a.s.} \tag{8.1.5}$$

PROOF. Symmetry follows from the symmetry of M_2 so that, in shorthand form,

$$\begin{aligned}\breve{M}_2(\mathrm{d}u)\,\ell(\mathrm{d}x) &= M_2\big(\mathrm{d}x \times \mathrm{d}(x+u)\big)\\ &= M_2\big(\mathrm{d}(x+u) \times \mathrm{d}x\big)\\ &= M_2\big(\mathrm{d}y \times \mathrm{d}(y-u)\big) = \breve{M}_2(-\mathrm{d}u)\,\ell(\mathrm{d}y),\end{aligned}$$

which establishes (i). Nonnegativity of $\breve{M}_2(A)$ follows directly from (8.1.1b). Positivity for $A \ni 0$ when ξ has an atomic component follows from Proposition 8.1.IV below, while for the other case, since A is open so that $A \supseteq S_{2\epsilon}(0)$ for some sphere of radius $2\epsilon > 0$, we can choose $\epsilon < \frac{1}{2}$ and then

$$\begin{aligned}\breve{M}_2\big(A\big) \geq \breve{M}_2\big(S_{2\epsilon}(0)\big) &= \mathrm{E}\bigg[\int_{\mathbb{U}^d} \xi\big(x + S_{2\epsilon}(0)\big)\,\xi(\mathrm{d}x)\bigg]\\ &\geq \mathrm{E}\bigg[\int_{S_\epsilon(0)} \xi\big(S_{2\epsilon}(x)\big)\,\xi(\mathrm{d}x)\bigg] \qquad \text{since } \mathbb{U}^d \supset S_\epsilon(0),\\ &\geq \mathrm{E}\bigg[\int_{S_\epsilon(0)} \xi\big(S_\epsilon(0)\big)\,\xi(\mathrm{d}x)\bigg] \qquad \text{since } S_{2\epsilon}(x) \supset S_\epsilon(0) \text{ for } x \in S_\epsilon(0),\\ &= M_2\big(S_\epsilon(0) \times S_\epsilon(0)\big) \;\geq\; \big[m\ell\big(S_\epsilon(0)\big)\big]^2 \;>\; 0 \qquad \text{since } \epsilon > 0.\end{aligned}$$

Positive-definiteness is a consequence of

$$\begin{aligned}0 \leq \mathrm{E}\bigg(\bigg|\int_{\mathcal{X}} \psi(x)\,\xi(\mathrm{d}x)\bigg|^2\bigg) &= \int_{\mathcal{X}}\int_{\mathcal{X}} \breve{M}_2(du)\overline{\psi(x)}\psi(x+u)\,\ell(\mathrm{d}x)\\ &= \int_{\mathcal{X}} \breve{M}_2(\mathrm{d}u) \int_{\mathcal{X}} \psi^*(u-w)\psi(w)\,\ell(\mathrm{d}w).\end{aligned}$$

Properties (ii) and (iii) together show that $\breve{M}_2$ is a *positive, positive-definite* (p.p.d.) measure; (iv) is then a consequence of general properties of p.p.d. measures, as set out in Section 8.6.

The final two assertions follow from the ergodic theorems developed in Chapter 11. In particular, a simple form of ergodic theorem for point processes and random measures ξ on $\mathbb{R}^d$ asserts that, for sets A satisfying the conditions outlined in (v), as $r(A) \to \infty$, $\xi(A)/\ell(A) \to m$ a.s. and in L_1-norm. If second moments exist, then also $\mathrm{E}\big|\xi(A)/\ell(A) - m\big|^2 \to 0$. From these results, it is easy to show that provided both $r(A)$ and $r(B) \to \infty$, $M_2(A \times A)/[\ell(A)]^2 \to m^2$ and, more generally, $M_2(A \times B)/[\ell(A)\ell(B)] \to m^2$. Approximating further, we find that $M_2(U)/(\ell \times \ell)(U) \to m^2$ for a wide class of sets $U \in \mathcal{X}^{(2)}$ including cylinder sets such as $U(A,r) = \{(x,u)\colon x \in A,\ y \in x + S_r(0)\}$. But

$$\int_{U(A,r)} M_2(\mathrm{d}s \times \mathrm{d}t) = \int_{S_r(0)} \ell(\mathrm{d}u) \int_A \breve{M}_2(\mathrm{d}v) = \ell\big(S_r(0)\big)\breve{M}_2(A),$$

and so (8.1.4) follows after dividing by $(\ell \times \ell)\big(U(A,r)\big) = \ell(A)\ell\big(S_r(0)\big)$.

Equation (8.1.5) can be established by similar arguments and is a simple special case of the higher-order ergodic theorems described in Chapter 11. □

Most of the results above transfer directly or with minor modifications to the other reduced second-order measures. The most important of these is the reduced covariance measure, which can be defined here through the relation

$$\breve{C}_2(\mathrm{d}u) = \breve{M}_2(\mathrm{d}u) - m^2\,\ell(\mathrm{d}u). \tag{8.1.6}$$

The covariance measure itself can be regarded as the second-moment measure of the mean-corrected random *signed* measure

$$\tilde{\xi}(A) \equiv \xi(A) - m\,\ell(A); \tag{8.1.7}$$

note that $\tilde{\xi}$ is a.s. of bounded variation on bounded sets. The reduced form inherits the following properties from $\breve{M}_2(\cdot)$.

Corollary 8.1.III. *The reduced covariance measure $\breve{C}_2(\cdot)$ of a second-order stationary random measure ξ is symmetric, positive-definite, and translation-bounded but in general is a signed measure rather than a measure. If ξ is ergodic, then for A, B and $r(A) \to \infty$ as for* (8.1.5), *and $\tilde{\xi}$ in* (8.1.7),

$$\breve{C}_2(A)/\ell(A) \to 0, \tag{8.1.8}$$

$$\frac{1}{\ell(A)} \int_A \tilde{\xi}(x+B)\,\tilde{\xi}(\mathrm{d}x) \to \breve{C}_2(B) = \mathrm{E}\bigg[\int_{\mathbb{U}^d} \tilde{\xi}(x+B)\,\tilde{\xi}(\mathrm{d}x)\bigg]. \tag{8.1.9}$$

For point processes, a characteristic feature of the reduced forms of both the moment and covariance measures is the atom at the origin. For a simple point process, this is removed by transferring to the corresponding reduced factorial measures $\breve{M}_{[2]}(\cdot)$ and $\breve{C}_{[2]}(\cdot)$. This is not the case, however, for more general point processes and random measures. The situation is summarized in the proposition below and its corollary (see also Kallenberg, 1983, Chapter 2).

Proposition 8.1.IV. *Let ξ be a stationary second-order random measure or point process on $\mathbb{R}^d$ with mean density m and reduced covariance measure $\breve{C}_2$. Then $\breve{C}_2(\mathrm{d}u)$ has a positive atom at $u = 0$ if and only if ξ has a nontrivial atomic component, in which case $\breve{C}_2(\{0\}) = \breve{M}_2(\{0\})$ and both equal*

$$\mathrm{E}\bigg[\int_{\mathbb{U}^d} \xi(\{x\})\,\xi(\mathrm{d}x)\bigg] = \mathrm{E}\bigg[\sum_{i:x_i\in\mathbb{U}^d} [\xi(\{x_i\})]^2\bigg]. \tag{8.1.10}$$

Moreover, there exists a σ-finite measure $\mu(\cdot)$ on $\mathbb{R}_+$ such that

(i) *μ has finite mass outside any neighbourhood of the origin, and for every $b > 0$, the atoms of ξ with mass greater than b can be represented as a stationary marked point process on $\mathcal{X} \times \mathbb{R}_+$ with ground rate $\mu(b,\infty)$ and stationary mark distribution $\Pi_b(\mathrm{d}\kappa) = \mu(\mathrm{d}\kappa]/\mu(b,\infty)$ on $\kappa > b$;*

(ii) *$\mu(\cdot)$ integrates κ on $\mathbb{R}_+$, and $\int_{\mathbb{R}_+} \kappa\,\mu(\mathrm{d}\kappa) \le m$;*

(iii) *ξ is purely atomic* a.s. *if and only if $m = \int_{\mathbb{R}_+} \kappa\,\mu(\mathrm{d}\kappa)$; and*

(iv) *$\mu(\cdot)$ integrates κ^2 on $\mathbb{R}_+$, and $\int_{\mathbb{R}_+} \kappa^2\,\mu(\mathrm{d}\kappa) = \breve{M}_2(\{0\}) = \breve{C}_2(\{0\})$.*

PROOF. Choose any monotonically decreasing sequence of nonempty sets A_n with $\operatorname{diam} A_n \downarrow 0$ and $A_n \downarrow \{0\}$. Then, for any $x \in \mathcal{X}$, $\xi(x + A_n) \downarrow \xi(\{x\})$ a.s. From (8.1.1b) and monotone convergence, we obtain

$$\breve{M}_2(A_n) = \mathrm{E}\left[\int_{\mathbb{U}^d} \xi(x + A_n)\, \xi(\mathrm{d}x)\right] \downarrow \mathrm{E}\left[\int_{\mathbb{U}^d} \xi(\{x\})\, \xi(\mathrm{d}x)\right] = \mathrm{E}\left[\sum_{x_i \in \mathbb{U}^d} \xi(\{x_i\})^2\right].$$

In particular, if ξ is a.s. continuous, it follows that both $\breve{M}_2$ and $\breve{C}_2$ are continuous at the origin, and conversely.

Suppose next that $b > 0$ is given, and consider the atoms from ξ with masses $\xi(\{x\}) > b$. If ξ is second-order stationary, there can be at most a finite number of such atoms in any finite interval. The set of such atoms is therefore denumerable and can be represented as an ordered sequence of pairs $\{(x_i, \kappa_i)\}$, where $x_i < x_j$ for $-\infty < i < j < \infty$ and $b < \kappa_i = \xi(\{x_i\})$. As in Section 6.4, equation (6.4.6), the set of pairs therefore constitutes a marked point process, which we denote by $\xi_b(\cdot)$.

Let m_b^{g} and $\Pi_b(\cdot)$ denote, respectively, the mean density of the ground process for ξ_b and its stationary mark distribution. Consistency of the ergodic limits requires that for $b' < b$ and $B \subseteq (b, \infty)$,

$$m_{b'}^{\mathrm{g}} \Pi_{b'}(B) = m_b^{\mathrm{g}} \Pi_b(B) \equiv \mu(B). \tag{8.1.11}$$

This relation therefore defines μ consistently and uniquely as a σ-finite measure on all of $\mathbb{R}_+$. Taking $B = (b, \infty)$ in (8.1.11) then implies that $\mu(b, \infty) = m_b^{\mathrm{g}} < \infty$, establishing (i). Moreover, the mean density of ξ_b, m_b say, is given by

$$m_b = m_b^{\mathrm{g}} \int_b^\infty \kappa\, \Pi_b(\mathrm{d}\kappa) = \int_b^\infty \kappa\, \mu(\mathrm{d}\kappa) = \int_0^\infty \kappa I_{\{\kappa > b\}}\, \mu(\mathrm{d}\kappa)\,.$$

Since $m_b \le m < \infty$ and for any A, $\xi_b(A) \uparrow \xi_{\mathrm{a}}(A)$ as $b \to 0$, where ξ_{a} denotes the atomic component of ξ, we must have $m_b = \mathrm{E}\big(\xi_b(\mathbb{U}^d)\big) \uparrow \mathrm{E}\big(\xi_{\mathrm{a}}(\mathbb{U}^d)\big) \equiv m_{\mathrm{a}} \le \mathrm{E}(\xi(\mathbb{U}^d)) \equiv m$ as $b \to 0$. Hence,

$$m_{\mathrm{a}} = \lim_{b \to 0} \int_0^\infty \kappa I_{\{\kappa > b\}}\, \mu(\mathrm{d}\kappa) = \int_0^\infty \kappa\, \mu(\mathrm{d}\kappa),$$

establishing (ii). Assertion (iii) is the same as the diffuse measure $\xi - \xi_{\mathrm{a}}$ having zero mean, implying that it is a.s. null.

Finally, for any $b > 0$, consideration of the second moment of ξ_b yields the equations

$$m_b^{\mathrm{g}} \int_b^\infty \kappa^2\, \Pi_b(\mathrm{d}\kappa) = \int_b^\infty \kappa^2\, \mu(\mathrm{d}\kappa) = \mathrm{E}\left[\sum_{x_i \in \mathbb{U}^d\,:\, \xi(\{x_i\}) > b} [\xi(\{x_i\})]^2\right].$$

Since the right-hand side is bounded above by $\breve{M}_2(\{0\}) < \infty$ and converges to $\breve{M}_2(\{0\})$ as $b \to 0$, (iv) follows. □

Condition (iii) above identifies purely atomic stationary random measures (see also Kallenberg, 1983). We would like to be able to use some property of μ to identify point processes (i.e. integer-valued random measures) and then simple point processes. The former identification is tantamount to a version of the moment problem: when do the moments of a measure [here $\mu(\cdot)$] suffice to identify the measure? This has no easy solution for our present purposes. The latter is much simpler.

Corollary 8.1.V. *A second-order stationary point process N with density m is a simple point process if and only if $\breve{C}_2(\{0\}) = \breve{M}_2(\{0\}) = m$, which is equivalent to the reduced second-order factorial moment and covariance measures having no atom at the origin.*

Proof. A stationary random measure ξ is a simple point process if and only if it is integer-valued and all its atoms have mass 1. The latter condition is satisfied if and only if $\int_1^\infty \kappa\, \mu(\mathrm{d}\kappa) = \int_1^\infty \kappa^2\, \mu(\mathrm{d}\kappa)$; i.e. μ has all its mass on $\{1\}$, or equivalently, $m = \breve{M}_2(\{0\})$. The equivalent form of the latter condition follows from the relation $\breve{M}_{[2]}(\{0\}) = \breve{M}_2(\{0\}) - m$. □

Analytical derivations of the relations for $\int \kappa^r\, \mu(\mathrm{d}\kappa)$ for positive integers r and stationary point processes have been given in Propositions 3.3.VIII and 3.3.IX. In Chapter 12, there is an analogue of Corollary 8.1.V for a higher-order reduced factorial measure of a stationary point process to vanish at $\{0\}$ as a condition for the process to have a bounded batch-size distribution or equivalently the factorial moment of the same order of $\mu(\cdot)$ to vanish.

Returning to more general properties, results such as (8.1.4) and (8.1.8) can be rephrased in further equivalent ways. When $\mathcal{X} = \mathbb{R}$, for example, they reduce respectively to

$$\mathrm{E}[\xi^2(0,x)] \sim m^2 x^2\,, \qquad \operatorname{var} \xi(0,x) = o(x^2) \qquad (x \to \infty),$$

results already discussed for ergodic point processes in Section 3.4.

Other useful results follow as special cases of the general representations (8.1.1). These imply, for example, that

$$\operatorname{cov}\left[\int_{\mathbb{R}^d} g(x)\, \xi(\mathrm{d}x),\, \int_{\mathbb{R}^d} h(y)\, \xi(\mathrm{d}y)\right] = \int_{\mathbb{R}^d} \breve{C}_2(\mathrm{d}u) \int_{\mathbb{R}^d} g(x) h(x+u)\, \ell(\mathrm{d}x). \tag{8.1.12}$$

In particular, (8.1.12) leads to the following expressions for the variance:

$$\begin{aligned} V(A) \equiv \operatorname{var} \xi(A) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} I_A(x) I_A(x+u)\, \ell(\mathrm{d}x)\, \breve{C}_2(\mathrm{d}u) \\ &= \int_{\mathbb{R}^d} I_A(x)\, \ell(\mathrm{d}x) \int_{\mathbb{R}^d} I_{A-x}(u)\, \breve{C}_2(\mathrm{d}u) \\ &= \int_A \breve{C}_2(A-x)\, \ell(\mathrm{d}x). \end{aligned} \tag{8.1.13a}$$

When $\mathcal{X} = \mathbb{R}$ and $A = (0, x]$, this becomes

$$V(x) \equiv \operatorname{var} \xi(0, x] = \int_{-x}^{x} (x - |u|)\, \breve{C}_2(\mathrm{d}u) = 2 \int_{0-}^{x} F_c(u)\, \mathrm{d}u, \qquad (8.1.13b)$$

where for $x > 0$, $F_c(x) = \frac{1}{2}\breve{C}_2(\{0\}) + \breve{C}_2(0, x] = \frac{1}{2}\breve{C}_2[-x, x]$ is a symmetrized form of the distribution function corresponding to the reduced covariance measure. Properties of $V(x)$ can be read off rather simply from this last representation: for example, it is absolutely continuous with a density function of which there exists a version that is continuous except perhaps for a countable number of finite discontinuities. Further details and an alternative approach in the point process case are outlined in Exercise 8.1.3. Note that, when it exists, the covariance density is a second derivative in $(0, \infty)$ of $V(x)$. See Exercise 8.1.4 for an analogue of (8.1.13b) in the case of a stationary isotropic point process in $\mathbb{R}^2$.

The variance function $V(A)$ is widely used in applications, often in the form of the ratio to the expected value $M(A)$; for a simple point process, this is just

$$\frac{V(A)}{M(A)} = \frac{\int_A \breve{C}_2(A - x)\, \ell(\mathrm{d}x)}{M(A)} = 1 + \frac{\int_A \breve{C}_{[2]}(A - x)\, \ell(\mathrm{d}x)}{m\ell(A)}. \qquad (8.1.14)$$

This ratio equals 1 for a Poisson process, while values larger than 1 indicate clustering and values less than 1 indicate repulsion or some tendency to regular spacing. For suitably small sets, for which $\operatorname{diam} A \to 0$, $V(A)/M(A) \to 1$; that is, locally the process is like a Poisson process in having the variance-to-mean ratio ≈ 1 (see Exercise 8.1.5). As $\ell(A) \to \infty$, various possibilities for the behaviour of $V(A)/M(A)$ exist and are realizable (see Exercise 8.1.6), but most commonly, the covariance measure is totally finite, in which case

$$V(A)/M(A) \to 1 + m^{-1}\breve{C}_{[2]}(\mathcal{X}) \qquad (A \uparrow \mathcal{X}).$$

A stationary random measure is of *bounded variability* if $V(A)$ itself remains bounded as $\ell(A) \to \infty$ as for (8.1.5) [see Exercises 7.2.10(a) and 8.1.6]. [This terminology is preferred to *controlled variability* (Cox and Isham, 1980, p. 94).]

Example 8.1(a) *Stationary Poisson cluster processes.* For a stationary Poisson cluster process and all values of the cluster centre x, monotone convergence shows that the cluster member process satisfies

$$M_{[2]}(A_n \times A_n \mid x) \to \mathrm{E}[Z(Z - 1)]$$

as $\ell(A_n) \to \infty$ through a convex averaging sequence $\{A_n\}$, where $Z \equiv N_m(\mathcal{X} \mid 0)$ denotes a generic r.v. for the total number of points in a cluster. Then, since (6.3.12) for large A gives $C_{[2]}(A \times A) \sim \mathrm{E}[Z(Z - 1)]\, M^c(A)$, we have $\breve{C}_{[2]}(\mathcal{X}) = \mathrm{E}[Z(Z - 1)]$ and thus

$$V(A)/M(A) \to 1 + \mathrm{E}[Z(Z - 1)]/\mathrm{E}Z = \mathrm{E}Z^2/\mathrm{E}Z. \qquad (8.1.15)$$

Characteristically, therefore, the variance-to-mean ratio for a Poisson cluster process increases from a value approximately equal to 1 for very small sets to a limiting value equal to the ratio of the mean square cluster size to the mean cluster size for very large sets [see the formula for the compound Poisson process in Exercise 2.1.8(b)]. The region of rapid growth of the ratio occurs as A passes through sets with dimensions comparable to those of (the spread of) individual clusters. □

These comments provide the background to diagnostic procedures such as plotting the ratio $V(A)/M(A)$ against $M(A)$ or $\ell(A)$ as $\ell(A) \to \infty$ and to the Greig-Smith method of nested quadrats, which uses a components-of-variance analysis to determine the characteristic dimensions at which clustering effects or local inhomogeneities begin to influence the variance [see Greig-Smith (1964) for further discussion].

The representation (8.1.1b) has important interpretations when ξ is a point process rather than a general random measure, and for the discussion in this section we assume that the process is orderly. In particular, it follows in this case that

$$\breve{M}_2(A) = \mathrm{E}\big[\#\{\text{point-pairs } (x_i, x_j)\colon x_i \in \mathbb{U}^d \text{ and } x_j \in x_1 + A\}\big] \tag{8.1.16a}$$

$$= \mathrm{E}\big[\text{rate of occurrence of point-pairs } (x_i, x_j)\colon x_j - x_i \in A\big]. \tag{8.1.16b}$$

Dividing by the mean density (= intensity = average rate of occurrence) m yields an interpretation of $\breve{M}_2$ in terms of the expectation measure of the *Palm process* (see Section 3.4 and the discussion in Chapter 13) obtained by conditioning on the presence of a point at the origin:

$$\mathrm{E}\big[\#\text{points } x_i \in A \mid \text{point at } x = 0\big] = \breve{M}_2(A)\,/\,m. \tag{8.1.17}$$

It is even more useful to have density versions of (8.1.17), assuming (as we now do) that $\breve{M}_{[2]}$ is absolutely continuous, so $\breve{M}_{[2]}(A) = \int_A m_{[2]}(x)\,\mathrm{d}x$. This density is related to the corresponding covariance density by

$$\breve{m}_{[2]}(x) = \breve{c}_{[2]}(x) + m^2. \tag{8.1.18}$$

When the density exists, the ratio $\breve{m}_{[2]}/m$ has been called the intensity of the process (e.g. Cox and Lewis, 1966, p. 69) or the conditional intensity function (e.g. Cox and Isham, 1980, Section 2.5). We call it the *second-order intensity* and denote it by $\breve{h}_2(\cdot)$ so that

$$\breve{h}_2(x) = \breve{m}_{[2]}(x)/m = m + \breve{c}_{[2]}(x)/m.$$

$\breve{h}_2(x)$ can also be interpreted as the intensity at x of the process conditional on a point at the origin; this is an interpretation taken up further in the discussion of Palm measures in Chapter 13. Notice that, in $d = 1$, we have for

a renewal process as in Chapter 4 with renewal function $U(x)$ $(x > 0)$ that is absolutely continuous, $\breve{h}_2(x) = \breve{h}_2(-x) = U'(|x|)$. We call the ratio

$$r_2(x) \equiv \frac{\breve{h}_2(x)}{m} = \frac{\breve{m}_{[2]}(x)}{m^2} \tag{8.1.19}$$

the *relative second-order intensity* [but note that in Vere-Jones (1978a) it is called the *relative conditional intensity*]. It equals 1 for a stationary Poisson process, while for other stationary processes it provides a useful indication of the strength and character of second-order dependence effects between pairs of points at different separations $x \in \mathbb{R}^d$: for example, when $r_2(x) > 1$, point-pairs separated by the vector x are more common than in the purely random (Poisson) case, while if $r_2(x) < 1$ such point-pairs are less common.

In considering the reduced measures $\breve{M}_2(A)$ and related functions, spheres $S_r(0)$ constitute a natural class of sets to use for A in dimension $d \geq 2$; define

$$\breve{K}_2(r) = \breve{M}_2(S_r(0) \setminus \{0\}) = \breve{M}_{[2]}(S_r(0)), \tag{8.1.20}$$

the equivalent formulation here being a consequence of orderliness. Ripley (1976, 1977) introduced this function, though what is now commonly called *Ripley's K-function* (including Ripley, 1981) is the density-free version

$$K(r) = \frac{\breve{M}_2(S_r(0) \setminus \{0\})}{m^2} = \frac{\breve{K}_2(r)}{m^2}\,, \tag{8.1.21}$$

so, since $\lambda = m$ because of orderliness,

$$\lambda K(r) = \mathrm{E}(\#\text{ of points within } r \text{ of the origin} \mid \text{point at the origin}), \tag{8.1.22}$$

where on the right-hand side the origin itself is excluded from the count. The function $K(r)$ is monotonically nondecreasing on its range of definition $r > 0$ and converges to 0 as $r \to 0$. As can be seen from the examples below and is discussed further in Chapter 12, this function is particularly useful in studying stationary isotropic point processes because it then provides a succinct summary of the second-order properties of the process. For a Poisson process, $K(r) = \ell(S_r(0))$.

Recall the definition of $K(r)$ in terms of the sphere $S_r(0)$. Noting the interpretation in (8.1.22), we see that the derivative $(\mathrm{d}/\mathrm{d}r)\breve{K}_2(r) = K'(r)$ gives the conditional probability of a point on the surface of a spherical shell of radius r, conditional on a point at the centre of the shell. Consequently, for an isotropic process in $\mathbb{R}^2$, the probability density that a point is located at distance r from a given point of the process and in the direction θ equals $K'(r)/(2\pi r)$, independent of θ because of isotropy. In dimension $d \geq 3$, the same equality holds on replacing the denominator $2\pi r$ by the surface area of $S_r(0)$.

For stationary isotropic processes in $\mathbb{R}^2$, the relative second-order intensity $r_2(x)$, which $\to 1$ as $x \to 0$ when it is continuous there, is a function of $|x|$ alone, and

$$\rho(r) = r_2(x) - 1, \qquad \text{where } r = |x|,$$

has been called the *radial correlation function* (see e.g. Glass and Tobler, 1971), though it may lack the positive-definiteness property of a true correlation function. The same quantity can be introduced, irrespective of isotropy, as a derivative of Ripley's K-function $K(r)$ in (8.1.21): write

$$\rho(r) = \frac{\mathrm{d}K(r)}{\mathrm{d}(\pi r^2)} - 1 = \frac{K'(r)}{2\pi r} - 1. \tag{8.1.23}$$

Examples of the use of $\breve{m}_{[2]}(\cdot)$ and $\rho(r)$ are given in Vere-Jones (1978a), Chong (1981) and Ohser and Stoyan (1981), amongst many other references.

Example 8.1(b) *A two-dimensional Neyman–Scott process.* By using the general results of Example 6.3(a), it can be shown that the reduced second factorial cumulant measure is given by

$$\breve{C}_{[2]}(A) \;=\; \mu_c m_{[2]} \int_{\mathbb{R}^2} F(u + A)\, F(\mathrm{d}u) \;=\; \mu_c m_{[2]}\, G(A),$$

where F is the probability distribution for the location of a cluster member about the cluster centre, G is the probability distribution for the difference of two i.i.d. random vectors with distribution F, μ_c is the Poisson density of cluster centres, and $m_{[2]}$ is the second factorial moment of the number of cluster members. For the K-function, we find

$$K(r) = \pi r^2 + [m_{[2]}/(\mu_c m_1^2)]G_1(r),$$

where $G_1(r)$ is the d.f. for the distance between two 'offspring' from the same 'parent', while

$$\rho(r) = [m_{[2]}/(\mu_c m_1^2)]g_1(r),$$

where $g_1(r) = G_1'(r)$ is the probability density function for the distance between two offspring from the same parent. Note that ρ is everywhere positive, an indication of overdispersion or clustering relative to the Poisson process, at all distances from an arbitrarily chosen point of the process.

Some particular results for the case where F is a bivariate normal distribution are given in Exercise 8.1.7. □

Example 8.1(c) *Matérn's Model I for underdispersion* (Matérn, 1960). Let $\{x_n'\}$ denote a realization of a stationary Poisson process N' on the line with intensity λ. Identify the subset $\{x_n''\}$ of those points of the realization that are within a distance R of another such point, i.e.

$$\{x_n''\} = \big\{x \in \{x_n'\}\colon |x - y| < R \;\text{ for some }\; y \in \{x_n'\} \;\text{ with }\; y \neq x\big\},$$

and let $\{x_n'\}\backslash\{x_n''\} \equiv \{x_n\}$ constitute a realization of a new point process N (note that $N = \{x_n\}$ is defined without using any Poisson properties of N'). The probability that any given point x of N' will be absent from N is then the probability, $1 - \mathrm{e}^{-2\lambda R}$, that at least one further point of N' is within a distance R of x. While these events are not mutually independent, they have

the same probability, so the mean density m for the modified process equals

$$m = \lambda \mathrm{e}^{-2\lambda R} \le \mathrm{e}^{-1}/(2R) \qquad \text{for all } \lambda;$$

the inequality is trict except for $\lambda R = \frac{1}{2}$.

To find the second-order properties of N, consider the probability $q(v)$ that for a given pair of points distance v apart in N', both are also in N. Then

$$q(v) = \begin{cases} 0 & (0 < v \le R), \\ \exp\big[-\lambda\big(2R + v\big)\big] & (R < v \le 2R), \\ \exp\big(-4\lambda R\big) & (v > 2R). \end{cases}$$

The factorial moment density of N is thus $\breve{m}_{[2]}(x) = \lambda^2 q(x)$, and the relative second-order intensity [see (8.1.19)] is given by

$$r_2(x) = \begin{cases} 0 & (0 < x \le R), \\ \mathrm{e}^{\lambda(2R-x)_+} & (x > R). \end{cases}$$

Thus, the process shows complete inhibition (as for any hard-core model) up to distance R and then a region of overdispersion for distances between R and $2R$ before settling down to Poisson-type behaviour for distances beyond $2R$.

The process is in fact of renewal type: the results above and others can be deduced from the renewal function for the process [see Exercise 8.1.9(a) for further details].

The model can readily be extended to point processes in the plane or space, but the analogues of the explicit expressions above become more cumbersome as the expression for the area or volume of the common intersection of circles or spheres becomes more complex (see Exercise 8.1.8).

The set of rejected points $\{x_n''\}$ is 'clustered' in the sense that every point has a nearest neighbour within a distance R [see Exercise 8.1.9(c)]. □

We conclude this section with some notes on possible estimates for reduced moment measures, being guided by the interpretations of the model-defined quantities and their interpretation described above. Assume, as is usually the case, that we observe only a finite part of a single realization of an ergodic process. Let B denote a suitable test set, such as an interval on the line or a rectangle or disk in the plane, and A a (larger) observation region. Then, replacing $\mathbb{U}^d$ by A in the right-hand side of (8.1.1b) and allowing for the change to the second factorial moment, we obtain

$$\breve{M}_{[2]}(B) = \frac{1}{\ell(A)} \mathrm{E}\left[\sum_{i:\, x_i \in A} N^*(x_i + B) \right], \tag{8.1.24}$$

where $N^*(x+B) = N(x+B) - \delta_0(B)$, so that $N(x+B)$ is reduced by 1 when B contains the origin.

The corresponding naïve estimate is obtained by dropping the expectation sign in the expression above (i.e. by taking each point x_i in A in turn as origin, counting the number of points in sets $x_i + B$ having a common relative position to x_i but ignoring x_i itself if it happens to lie within the test region, and then dividing by the Lebesgue measure of the observation region); we denote it by

$$\widehat{M}_{[2]}(B; A) = \frac{1}{\ell(A)} \sum_{i: x_i \in A} N^*(x_i + B). \tag{8.1.25}$$

Note that in the case of a process with multiple points, the points at each x_i should be labelled $x_i^{(1)}, \ldots, x_i^{(n_i)}$, and the definition of N^* implies that we omit pairs $(x_i^{(j)}, x_i^{(j)})$ but not any pair $(x_i^{(j)}, x_i^{(k)})$ with $j \neq k$.

In principle, (8.1.1b) implies that this estimate is unbiased, while the assumed ergodicity of the process and the first assertion of (8.1.5) imply that it is consistent. In practice, however, difficulties arise with edge effects since $N^*(x_i + B)$ may not be observable if x_i lies near the boundary of B.

Replacing it by $N^*[(x_i + B) \cap A]$ introduces a bias that may be corrected in a variety of ways. For example, we may subtract an explicit correction factor [see Exercise 8.1.11(b)], or we may take observations over an extended region $A + B$ (plus sampling), thereby ensuring that all necessary information is available but at the expense of the fullest use of the data.

One commonly used correction replaces (8.1.25) by the form

$$M^c_{[2]}(B; A) = \left[\frac{N(A)\ell(B)}{\ell(A)}\right] \frac{\sum_{x_i \in A} N^*[A \cap (x_i + B)]}{\sum_{x_i \in A} \ell[A \cap (x_i + B)]} \tag{8.1.26}$$

so that each observation count $N^*(x_i + B)$ is given a relative weight equal to that fraction of $\ell(x_i + B)$ that remains inside A; see also Exercise 8.1.10(a).

Estimates of the reduced covariance measure, and hence of the variance function, can be obtained by subtracting appropriate multiples of $\ell(B)$ as noted in Exercise 8.1.11(c).

These comments are included to suggest a basis for the systematic treatment of moment estimation for point processes; Krickeberg (1980) and Jolivet (1978) discuss some further issues and special problems, while applications are discussed by Ripley (1976, 1981), Diggle (1983), Vere-Jones (1978a), and many others.

Exercises and Complements to Section 8.1

8.1.1 Consider a nonstationary Poisson cluster process on $\mathbb{R}$ with cluster centres having intensity $\mu_c(t)$ and a cluster with centre t having either a single point at t with probability $p_1(t)$ or two points, one at t and the other at $t + X$, where the r.v. X has d.f. F. Show that $p_1(\cdot)$ and $\mu_c(\cdot)$ can be chosen so that the process is first-order stationary but not second-order stationary.

8.1.2 Construct an example of a point process that has stationary covariance measure but nonstationary expectation measure. [*Hint:* Such a process is necessarily not simple: consider a compound Poisson process in which the rate of occurrence of groups and mean square group size are adjusted suitably.]

8.1.3 Let $V(x) = \operatorname{var}(N(0,x])$ denote the variance function of a second-order stationary point process $N(\cdot)$ on the line, and write $M_2(x) = \mathrm{E}([N(0,x]]^2) = V(x) + (mx)^2$, where $m = \mathrm{E}N(0,1]$.

(a) Show that $M_2(x)$ is superadditive in $x > 0$ and hence that $V'(0+) \equiv \lim_{x\downarrow 0} V(x)/x$ exists, with $V'(0+) \geq m$.

(b) Show that $(M_2(x))^{1/2}$ is subadditive and hence that $\lim_{x\to\infty} V(x)/x^2$ exists and is finite.

(c) When $N(\cdot)$ is crudely stationary (see Section 3.2), show that $V'(0+) = m$ if and only if the process is simple.

(d) Construct an example of a second-order stationary point process for which the set of discontinuities of the left and right derivatives of $V(\cdot)$ is countably dense in $(0,\infty)$.

(e) Writing $M_2(x) = \lambda \int_0^x (1 + 2U(y))\,\mathrm{d}y$, where λ is the intensity of $N(\cdot)$, show that $\lim_{x\to\infty} U(x)/\lambda x$ exists and is ≥ 1.

(f) Show that $\sup_{x>0}(U(x+y) - U(x)) \leq 2U(y) + m/\lambda$.

(g) Use (8.1.13) to show that $V(x) = 2\int_0^x F_c(u)\,\mathrm{d}u$ where, in terms of the reduced covariance measure $\breve{C}_2$, $F_c(u) = \frac{1}{2}\breve{C}_2(\{0\}) + \breve{C}_2(0,u] = \frac{1}{2}\breve{C}_2[-u,u]$. Deduce that, when it exists, the covariance density is a second derivative in $\mathbb{R}_+$ of $V(x)$.

[*Hint*: See Daley (1971) for (a)–(e) and Berbée (1983) for (f).]

8.1.4 Suppose $N(\cdot)$ is a simple stationary isotropic point process in $\mathbb{R}^2$ with intensity λ, finite second-moment measure, and second-order intensity [see (8.1.18)] $\breve{h}_2(x) = \breve{h}(|x|)$, say, for points distance $|x|$ apart. Show that for a sphere S_r of radius r, $V(S_r) \equiv \operatorname{var} N(S_r)$ equals

$$\lambda\pi r^2 + \lambda \int_0^r 2\pi u\,\mathrm{d}u \int_{0+}^{r+u} \arccos\left(\max\left\{-1, \frac{u^2+v^2-r^2}{2uv}\right\}\right) v\breve{h}(v)\,\mathrm{d}v$$

Suppose that $\breve{h}(u) \to 0$ monotonically for u large enough. Deduce that when $\lim_{r\to\infty}\int_1^r u\breve{h}(u)\,\mathrm{d}u < \infty$, $\lim_{r\to\infty} V(S_r)/M(S_r)$ exists [see below (8.1.14)].

8.1.5 (a) If $\{I_n\}$ is a nested decreasing sequence of intervals with $\ell(I_n) \to 0$ as $n \to \infty$, show that for any second-order stationary simple point process on $\mathbb{R}$, $V(I_n)/M(I_n) \to 1$.

(b) Show that replacing $\{I_n\}$ by more general nested sets $\{A_n\}$ may lead to $V(A_n)/M(A_n) \not\to 1$. [*Hint*: Consider a stationary deterministic process at unit rate, and for some fixed integer $j \geq 2$, let $A_n = \bigcup_{i=1}^{j}(i, i+1/n]$.]

(c) Let $\{A_n\}$ be a nested decreasing sequence of sets in $\mathbb{R}^d$ with $\operatorname{diam}(A_n) \to 0$ as $n \to \infty$. Show that $V(A_n)/M(A_n) \to 1$ as $n \to \infty$ for second-order stationary simple point processes on $\mathbb{R}^d$.

8.1.6 *Processes of bounded variability.* Show that for a nontrivial stationary cluster point process on $\mathbb{R}$ with finite second-moment measure to be of bounded variability, the cluster centre process must be of bounded variability and all clusters must be of the same size.

As a special case, suppose the cluster centre process is deterministic and that points are randomly jittered with jitter distribution F, say. What conditions on F are needed for the jittered process to be of bounded variability? [See Cox and Isham (1980, Section 3.5) for more discussion.]

8.1.7 *Isotropic Neyman–Scott process.* In Example 8.1(b), suppose that the d.f. F is the bivariate normal distribution with zero mean and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Then, the symmetrized d.f. G for the vector distance between two offspring from the same parent is bivariate normal also with zero mean vector and covariance matrix 2Σ.

When $\sigma_1^2 = \sigma_2^2 = \sigma^2$, say, and $\rho = 0$, the process is isotropic and

$$K(r) = \pi r^2 + [m_{[2]}/(\mu_c m_1^2)](1 - \mathrm{e}^{-r^2/4\sigma^2}).$$

8.1.8 *$\mathbb{R}^d$-analogue of Matérn's Model I.* Let $v(R, a)$ denote the volume of the intersection of two $\mathbb{R}^d$ hyperspheres of radius R whose centres are distance a apart. Construct a point process in $\mathbb{R}^d$ analogous to the process in $\mathbb{R}$ of Example 8.1(b) and show that this $\mathbb{R}^d$ analogue has

$$M(A) = \lambda \mathrm{e}^{-\lambda v(R,0)} \ell(A),$$

$$\breve{h}_2(x) = \begin{cases} 0 & (0 < |x| \le R), \\ \lambda \exp\big(-\lambda[v(R,0) - v(R,|x|)]\big) & (R < |x| \le 2R), \\ \lambda \exp\big(-\lambda v(R,0)\big) & (2R < |x|). \end{cases}$$

[*Hint*: See Cox and Isham (1980, Exercise 6.3) for the case $d = 2$.]

8.1.9 *Matérn's Model I: Further properties.*

(a) *Renewal process.* Let $\{t_n' : n = 1, 2, \ldots\}$ be the successive epochs in $(1, \infty)$ of a Poisson process on $\mathbb{R}_+$ at rate λ, and attach marks $I(t_n') = 0$ or 1 successively as follows, starting with t_n' initially unmarked. If t_n' is unmarked, then $I(t_n') = 0$ if $t_{n+1}' < t_n' + 1$, in which case $I(t_{n+1}') = 0$ also, or else $t_{n+1}' > t_n' + 1$, $I(t_n') = 1$, and t_{n+1}' is initially unmarked. If $I(t_n') = 0$, then $I(t_{n+1}') = 0$ if $t_{n+1}' < t_n' + 1$, or else $t_{n+1}' > t_n' + 1$ and t_{n+1}' is initially unmarked. Show that $\{t_n : n = 0, 1, \ldots\}$, defined by $t_0 = 0$ and $t_{n+1} = \inf\{t_j' > t_n : I(t_j') = 1\}$ $(n = 0, 1, \ldots)$, are the epochs of a renewal process with a renewal density function $h(\cdot)$ that is ultimately constant, namely

$$h(x)\,\mathrm{d}x = \begin{cases} 0 & (0 < x \le 1), \\ \lambda \mathrm{e}^{-\lambda \min(x,2)} & (x > 1). \end{cases}$$

(b) Show that Example 8.1(c) is a version of the corresponding stationary renewal process.

(c) *The complementary set.* Every point in the complementary set $\{x_n''\}$ of 'rejected points' in the construction of Matérn's Model I in Example 8.1(c) shows clustering characteristics: for one thing, the nearest-neighbour distance of any x_n'' is at most R. Investigate other properties of this process.

[*Hint*: Consider first the case $d = 1$; find its density, cluster structure, nearest-neighbour distribution, and covariance density. Which of these are accessible when $d \geq 2$? What properties of $\{x_n''\}$ can be deduced by complementarity with respect to a Poisson process of the underdispersed process of Example 8.1(c)?]

8.1.10 *Matern's Model II for underdispersion.* Consider an independent marked Poisson process with realization $\{(x_i', \kappa_i)\}$ in which the points $\{x_i'\}$ have intensity λ, say, and the independent marks have a common uniform distribution on $(0, 1)$ (any absolutely continuous distribution will do). A point x_i' is rejected if there is any other point within distance R and with mark larger than κ_i. Show that the retained points $\{x_i\}$, say, have density $(1 - \mathrm{e}^{-2\lambda R})/(2R)$ and that the relative second-order intensity $r_2(x)$ vanishes for $|x| < R$, equals 1 for $|x| > 2R$, and for $R < |x| < 2R$,

$$r_2(x) = \frac{2R + (3R + x)\mathrm{e}^{-\lambda(R+x)} - (5R + x)\mathrm{e}^{-\lambda(3R+x)}}{R(R + x)(3R + x)} > 1.$$

Examine the $\mathbb{R}^d$ analogues of the model (see Exercise 8.1.8).

8.1.11 (a) Show the weighted estimate $M_{[2]}^c(B)$ in (8.1.26) is unbiased.

(b) A simpler but cruder correction subtracts from (8.1.25) the expected bias when the observed process is Poisson with the same mean rate. Express this as a correction to $M_{[2]}^c(B)$. [*Hint*: See e.g. Miles (1974) and Vere-Jones (1978a, p. 80) who give explicit forms.]

(c) Although the cumulative forms given above admit consistent estimates, they are less easy to interpret than smoothed estimates of the corresponding densities. For example, in $\mathbb{R}^2$, estimates of the radial correlation function and related quantities can be obtained by counting the number of points in an annulus about a given point of the realization, dividing by the area of the annulus, subtracting the appropriate mean, and regarding the resultant value as an estimate of $\rho(r)$ at a distance r corresponding to the mid-radius of the annulus. Fill out the details behind these remarks. [*Hint*: See e.g. Vere-Jones (1978a) and Chong (1981) for applications.]

8.2. The Bartlett Spectrum

The spectral theory of point processes has two origins. On the theoretical side, the results can be derived from specializations of Doob's (1949, 1953) theory of processes with stationary increments and related treatments of generalized stochastic processes by Bochner (1955) and Yaglom (1961). The key features relevant to the practical analysis of point process data were identified by Bartlett (1963) and followed up by several authors, as summarized for example in Cox and Lewis (1966) and Brillinger (1972, 1978). The treatment given in this chapter is based on developments of the theory of Fourier transforms of unbounded measures (see e.g. Argabright and de Lamadrid, 1974). As such, it requires an extension, not quite trivial, of the classical Bochner theorem and related results used in standard time series analysis. We describe

this extension, concerned with properties of positive, positive-definite (p.p.d.) measures, in Section 8.6. Here in this section, we summarize and illustrate the properties that are most relevant to the practical analysis of point process models.

We saw in Proposition 8.1.II that the reduced second-moment measure $\breve{M}_2$ of a stationary random measure is a p.p.d. measure so that all the properties developed for such measures in Section 8.6 apply. In particular, $\breve{M}_2$ is *transformable* so that it possesses a well-defined Fourier transform (in the sense of generalized functions), which is again a measure, and for which the explicit versions of the Parseval relation and the inversion theorem, derived in that section, are valid. The reduced covariance measure $\breve{C}_2$ is not itself a p.p.d. measure, but it differs from $\breve{M}_2$ only by the term $m^2\ell$, which is also a p.p.d. measure [its Fourier transform is the multiple $(m^2/2\pi)\delta_0$ of the measure consisting of a single atom at the origin]. Thus, $\breve{C}_2$ can be represented as a difference of two p.p.d. measures, so that the same results (existence of a Fourier transform that is a difference of two p.p.d. measures, Parseval relations, etc.) hold for it also. A similar remark applies to the reduced second factorial moment measure and the corresponding factorial cumulant measure, where it is a matter of subtracting an atom at the origin.

Any one of these four measures could be taken as the basis for further development of the spectral theory. It is convenient, and consistent with the standard convention in time series analysis, to choose as the spectrum of the process ξ the inverse Fourier transform of the (ordinary) covariance measure. The proposition below summarizes the main results pertaining to this transform; (8.2.1) and (8.2.2) are examples of *Parseval relations*.

Proposition 8.2.I. *Let ξ be a second-order stationary point process or random measure on $\mathbb{R}^d$ with reduced covariance measure $\breve{C}_2$. Then*

(a) *there exists a symmetric, translation-bounded measure Γ on $\mathcal{B}_{\mathbb{R}^d}$ such that, for all ψ in the space $\mathcal{S}$ of functions of rapid decay defined below* (8.6.1),

$$\int_{\mathbb{R}^d} \psi(x)\, \breve{C}_2(\mathrm{d}x) = \int_{\mathbb{R}^d} \tilde{\psi}(\omega)\, \Gamma(\mathrm{d}\omega), \tag{8.2.1}$$

where $\tilde{\psi}(\omega) = \int_{\mathbb{R}^d} \mathrm{e}^{i(\omega\cdot u)}\psi(u)\,\mathrm{d}u$ $(\omega \in \mathbb{R}^d)$;

(b) *the inversion relations* (8.6.6–10) *and* (8.6.12) *hold, with μ identified as Γ and ν as $\breve{C}_2$; and*

(c) *for bounded measurable ϕ with bounded support and also for $\phi \in \mathcal{S}$, if $\zeta_\phi = \int_{\mathbb{R}^d} \phi(x)\, \xi(\mathrm{d}x)$, then*

$$\operatorname{var} \zeta_\phi = \int_{\mathbb{R}^d} |\tilde{\phi}(\omega)|^2\, \Gamma(\mathrm{d}\omega) = \int_{\mathbb{R}^d} (\phi * \phi^*)(u)\, \breve{C}_2(\mathrm{d}u) \geq 0, \tag{8.2.2}$$

where $\phi^*(u) = \phi(-u)$.

Proof. The statements all follow from the p.p.d. properties noted in the opening paragraph and the results for p.p.d. measures outlined in Section 8.6. In particular, (8.2.2) follows from Proposition 8.6.IV. □

Definition 8.2.II. *The* Bartlett spectrum *of a second-order stationary point process or random measure ξ on $\mathbb{R}^d$ is the measure $\Gamma(\cdot)$ associated with the reduced covariance measure $\breve{C}_2$ of ξ in Proposition* 8.2.I.

Equations (8.2.1), usually in the form of (8.2.4) below, and (8.2.2) are generally the most convenient results to use in establishing the form of the Bartlett spectrum for a given process. Note in particular the special case for $\mathcal{X} = \mathbb{R}$ and ψ the indicator function for $(0, t]$,

$$\operatorname{var} \xi(0,t] = \int_{\mathbb{R}} \left(\frac{\sin \frac{1}{2}\omega t}{\frac{1}{2}\omega} \right)^2 \Gamma(\mathrm{d}\omega), \tag{8.2.3}$$

which is essentially Daley's (1971) representation for the variance function of a stationary point process or random measure [Daley uses a measure defined on $\mathbb{R}_+$, while in (8.2.3), $\Gamma(\cdot)$ is a symmetric measure on $\mathbb{R}$]. An alternative route to (8.2.3) exploiting a skeleton process, the standard Bochner representation and weak convergence, is sketched in Exercise 8.2.1.

It is clear from Proposition 8.2.I that while the *spectral measure* Γ is positive, it is not in general a p.p.d. measure. However, since the reduced second-moment measure $\breve{M}_2$ is positive and is the Fourier transform of the positive measure $\Gamma(\cdot) + [m^2/(2\pi)^d]\delta_0(\cdot)$, $\Gamma(\cdot)$ can be made into a p.p.d. measure by the addition of a sufficiently large atom at the origin.

In the point process case, the reduced covariance measure has an atom at the origin that transforms into a positive multiple of Lebesgue measure, and consequently the Bartlett spectrum of a point process is never totally finite. On the other hand, the factorial covariance measure is often both absolutely continuous and totally finite, and then $\Gamma(\cdot)$ is absolutely continuous with a density $\gamma(\cdot)$, which can be written (for the case $d = 1$)

$$\begin{aligned} 2\pi\gamma(\omega) &= m + \int_{-\infty}^{\infty} \mathrm{e}^{-i\omega x} c_{[2]}(x)\, \mathrm{d}x \\ &= m + \tilde{c}_{[2]}(-\omega) = m + \tilde{c}_{[2]}(\omega). \end{aligned} \tag{8.2.4}$$

It was in this form that the spectral measure was originally introduced by Bartlett (1963).

It is not known whether every p.p.d. measure can arise as the second-moment measure of some random measure nor, when it does, how to construct a process yielding the given measure as its second-moment measure. The standard construction using Gaussian processes or measures is not available here, as such processes do not have nonnegative trajectories (see Wiener's homogeneous chaos example in Chapter 9). Some partial results arise from the examples considered below and from Exercises 8.2.11–12 and 8.4.6–7. Davidson (1974) provided a construction for identifying the second-moment measures of stationary random measures on the circle (see the further discussion in Chapter 12), but it relies on the finiteness of the invariant measure on a circle, and

it is not obvious how it might be extended to either point processes or random measures on the line. In the very special case of a discrete point process on the four points of the compass (NESW), with translation interpreted as rotation through $\pi/2$, the family of second-moment measures can be identified explicitly and is strictly contained in the class of p.p.d. measures; see Exercise 8.2.5 for details.

We now discuss the Bartlett spectrum for some basic point processes on $\mathbb{R}^d$.

EXAMPLE 8.2(a) *Poisson process with constant intensity on* $\mathbb{R}^d$. Here $\breve{C}_2$ consists only of the atom $m\delta_0(\cdot)$ so Γ is absolutely continuous with density $m/(2\pi)^d$. This 'white-noise' spectrum is consistent with the completely random character of the process. Note that the Parseval relations (8.2.1) and (8.2.2) take, respectively, the special forms, with $\zeta_\phi = \int_{\mathbb{R}^d} \phi(x)\, N(\mathrm{d}x)$,

$$m\psi(0) = \frac{m}{(2\pi)^d} \int_{\mathbb{R}^d} \tilde{\psi}(\omega)\, \mathrm{d}\omega$$

and

$$\operatorname{var} \zeta_\phi = m \int_{\mathbb{R}^d} |\phi(x)|^2\, \mathrm{d}x = \frac{m}{(2\pi)^d} \int_{\mathbb{R}^d} |\tilde{\phi}(\omega)|^2\, \mathrm{d}\omega. \qquad \square$$

EXAMPLE 8.2(b) *Stationary renewal process.* If the renewal density $u(t)$ exists and the process is stationary with mean rate $\lambda = 1/\mu$, where μ is the mean lifetime, we have from Example 5.4(b) that

$$\breve{m}_{[2]}(x) = \lambda u(|x|)$$

and hence

$$\breve{c}_2(x) = \delta_0(x) + \lambda\big[u(|x|) - \lambda\big].$$

If further the difference $u(x) - \lambda$ is integrable on $(0, \infty)$, (8.2.4) yields for $\omega \neq 0$

$$\gamma(\omega) = \frac{\lambda}{2\pi}\left[1 + \frac{\widetilde{F}(\omega)}{1 - \widetilde{F}(\omega)} + \frac{\widetilde{F}(-\omega)}{1 - \widetilde{F}(-\omega)}\right] = \frac{\lambda}{2\pi}\left[\frac{1}{1 - \widetilde{F}(\omega)} + \frac{1}{1 - \widetilde{F}(-\omega)} - 1\right], \tag{8.2.5}$$

where $\widetilde{F}(\omega) = \int_0^\infty \mathrm{e}^{i\omega x}\, \mathrm{d}F(x)$ is the characteristic function of the lifetime distribution. For $\omega = 0$, we obtain from the above or Exercise 4.4.5

$$\gamma(0) = \frac{\lambda}{2\pi}\left(1 + \frac{\sigma^2 + \mu^2}{\mu^2}\right) = \frac{\lambda}{2\pi}\left(1 + 2\int_0^\infty \big[u(x) - \lambda\big]\, \mathrm{d}x\right). \qquad \square$$

Special cases, when lifetime distributions are of 'phase type' for example, yield rational polynomials for $\widetilde{F}$ and hence rational spectral densities (see e.g. Neuts, 1979). Exercise 8.2.6 gives a simple nontrivial example. Since a stationary renewal process has moment measures of all orders whenever it exists, the Bartlett spectrum exists for all such processes, but without the additional restriction it may not be absolutely continuous or (even if it is) $\gamma(0)$ need not be finite as above. The extreme case described in the next example is worth particular mention.

EXAMPLE 8.2(c) *Stationary deterministic process.* Here, points occur on a regular lattice of span a, the whole lattice being randomly shifted so that the first point to the right of the origin is uniformly distributed on $(0, a]$. The measure $\breve{M}_2(\cdot)$ has an infinite sum, with mass $1/a$ at each of the points ka, $(k = 0, \pm 1, \ldots)$. Its Fourier transform has mass $1/a^2$ at each of the points $2\pi j/a$, $(j = 0, \pm 1, \ldots)$. Moving to the Fourier transform of the covariance measure deletes the atom at $j = 0$ so that $\Gamma(\cdot)$ can be written in terms of Dirac measures as

$$\Gamma(A) = \frac{1}{a^2} \sum_{j=1}^{\infty} \big[\delta_{2\pi j/a}(A) + \delta_{-2\pi j/a}(A)\big]. \tag{8.2.6}$$

□

EXAMPLE 8.2(d) *Cluster processes.* For a general cluster process N in $\mathbb{R}^d$, the variance of an integral $\int_{\mathbb{R}^d} \phi(x)\, N(\mathrm{d}x)$ can be written (see Exercise 6.3.4)

$$\operatorname{var}\left(\int_{\mathbb{R}^d} \phi(x)\, N(\mathrm{d}x) \right) = \int_{\mathbb{R}^d} V_\phi(u)\, M^c(\mathrm{d}u) + \int_{(\mathbb{R}^d)^{(2)}} m_\phi(u) m_\phi(v)\, C_2^c(\mathrm{d}u \times \mathrm{d}v), \tag{8.2.7}$$

where

$$m_\phi(u) = \int_{\mathbb{R}^d} \phi(x)\, M_1(\mathrm{d}x \mid u), \qquad V_\phi(u) = \int_{(\mathbb{R}^d)^{(2)}} \phi(s)\phi(t)\, C_2(\mathrm{d}s \times \mathrm{d}t \mid u),$$

and we use the notation $M^c(\cdot)$ and $C_2^c(\cdot)$ from (6.3.4–5). In the stationary case, $M^c(\mathrm{d}u) = m_c\, \mathrm{d}u$, where m_c is the mean density of the cluster centre process, while C_2^c has a reduced form that can be written in terms of the Bartlett spectrum Γ^c of the cluster centre process. Since also $C_2(\mathrm{d}s \times \mathrm{d}t \mid y)$ depends only on the differences $s - y$ and $t - y$, the first term in (8.2.7) can be written in terms of the measure B defined via bounded measurable h by

$$\int_{\mathbb{R}^d} h(y)\, B(\mathrm{d}y) = \int_{(\mathbb{R}^d)^{(2)}} h(s - t)\, C_2(\mathrm{d}s \times \mathrm{d}t \mid 0).$$

Here the measure B is both positive-definite and totally finite (since the mean square cluster size is necessarily finite); it has therefore an ordinary Fourier transform $\widetilde{B}(\omega) = (2\pi)^{-d} \int_{\mathbb{R}^d} \mathrm{e}^{-i(\omega \cdot x)}\, B(\mathrm{d}x)$, which can be written in the symmetric form

$$\widetilde{B}(\omega) = \operatorname{var}\left(\int_{\mathbb{R}^d} \mathrm{e}^{-i(\omega \cdot x)}\, N_m(\mathrm{d}x \mid 0) \right),$$

where, it should be recalled, $\operatorname{var} Z = \mathrm{E}\big(|Z|^2\big) - |\mathrm{E}Z|^2$ for a complex-valued r.v. Z. Thus, writing

$$\widetilde{M}_1(\omega \mid 0) = \int_{\mathbb{R}^d} \mathrm{e}^{-i(\omega \cdot x)}\, M_1(\mathrm{d}x \mid 0) = \mathrm{E} \int_{\mathbb{R}^d} \mathrm{e}^{-i(\omega \cdot x)}\, N_m(\mathrm{d}x \mid 0),$$

we obtain from (8.2.7)

$$\operatorname{var}\left(\int_{\mathbb{R}^d}\phi(x)\,N(\mathrm{d}x)\right)=\int_{\mathbb{R}^d}|\tilde{\phi}(\omega)|^2\left[\tilde{B}(\omega)\frac{m_c}{(2\pi)^d}\,\mathrm{d}\omega+|\widetilde{M}_1(\omega\mid 0)|^2\,\Gamma^c(\mathrm{d}\omega)\right].$$

This relation shows that the Bartlett spectrum of the cluster process N can be identified with the measure

$$\Gamma(\mathrm{d}\omega)=\widetilde{B}(\omega)m_c(2\pi)^{-d}\,\mathrm{d}\omega+|\widetilde{M}_1(\omega\mid 0)|^2\,\Gamma^c(\mathrm{d}\omega). \tag{8.2.8}$$

The first term can be regarded as the contribution to the spectrum from the internal cluster structure; the second term is a filtered version of the spectrum of the cluster centre process with the filtering reflecting the mean distribution of the cluster, as in Daley (1972b).

For a stationary Poisson cluster process, further simplification occurs. Letting μ_c denote the intensity of the Poisson process of cluster centres, we find that Γ has a density γ, which has the simple alternative forms

$$\begin{aligned}\gamma(\omega)&=\frac{\mu_c}{(2\pi)^d}\left[\int_{\mathbb{R}^d}M_1(\mathrm{d}x\mid 0)+\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathrm{e}^{-i(y\cdot\omega)}\,M_{[2]}(\mathrm{d}x\times\mathrm{d}y\mid 0)\right]\\&=\frac{\mu_c}{(2\pi)^d}\mathrm{E}\left(\left|\int_{\mathbb{R}^d}\mathrm{e}^{i(x\cdot\omega)}\,N_m(\mathrm{d}x\mid 0)\right|^2\right),\end{aligned} \tag{8.2.9}$$

which is easily recognized as the transformed version of (6.3.5). Specific results for the Neyman–Scott and Bartlett–Lewis processes follow readily from these equations (see Exercises 8.2.9 and 8.2.10). □

We shall see in Section 8.3 that, for filtering and prediction purposes, a particularly important role is played by point processes having a rational spectral density. Many common and useful examples fall into this class. By suitable specification of the components, both renewal and cluster processes can give rise to spectral measures with rational spectral densities. For example, it is clear from (8.2.5) that this will occur whenever the interval distribution of a renewal process has a rational Laplace transform, that is, whenever the distribution is expressible as a finite convolution or mixture of exponentials. Several types of cluster processes, as well as Cox processes, have rational spectral densities, in particular the Neyman–Scott process with an exponential or Erlang distribution for the distances of the cluster elements from the cluster centre [see also Exercise 8.2.7(b)]. The wide choice of such examples shows not only the richness of the class but also the relative lack of discrimination in the spectrum as a means of distinguishing between processes that in other respects may be quite dissimilar.

One of the most important examples is the Hawkes process with suitably restricted response function (i.e. infectivity measure) as described below.

EXAMPLE 8.2(e) *Hawkes process with rational spectral density.* From Example 6.3(c) and the results on branching processes in Exercise 5.5.6, we see that the Fourier transform $\widetilde{M}_1$ of the first-moment measure of the total offspring process is a rational function of the Fourier–Stieltjes transform $\tilde{\mu}$ of the infectivity measure, namely

$$\widetilde{M}_1(\omega \mid 0) = 1/[1 - \tilde{\mu}(\omega)], \quad \text{where} \quad \tilde{\mu}(\omega) = \int_0^\infty \mathrm{e}^{i\omega x}\, \mu(\mathrm{d}x).$$

Combining this result with the expressions for the mean rate and covariance density given by (6.3.26) and (6.3.27) and with the general form (8.2.8) for cluster processes, we obtain the spectral density for the Hawkes process in the form

$$\gamma(\omega) = \frac{\lambda/(2\pi)}{(1-\nu)\,|1 - \tilde{\mu}(\omega)|^2}\,. \tag{8.2.10}$$

Consequently, when $\tilde{\mu}(\omega)$ is a rational function of ω, so too is $\widetilde{M}_1(\omega)$.

Because the form of (8.2.10) is similar to that of the spectral density of an autoregression in continuous time, one might hope that the Hawkes model could play a role similar to that of autoregressive models in the context of mean square continuous processes. This hope is frustrated by the special probabilistic structure of the Hawkes model, which requires that $\mu(\cdot) \geq 0$. If this condition is violated, it is not clear that there exists any point process with the spectral form (8.2.10), and if such a process does exist, it certainly will not have the Poisson branching structure of a Hawkes process. Despite this difficulty, the possibility of using the Hawkes process to approximate general point process spectra was explored by Hawkes (1971b), Hawkes and Adamopoulos (1973), Ozaki (1979) and, more deliberately, by Ogata and Akaike (1982), with an application in Ogata *et al.* (1982). Ogata and Akaike (1982) suggest taking for μ a measure on $[0, \infty)$ with density function $\mu(t) = \mathrm{e}^{\alpha t} \sum_{k=0}^{K} b_k L_k(t)$ for $\alpha > 0$ and Laguerre polynomials $L_k(t)$. This form leads automatically to processes with rational spectral densities since the Fourier transforms of the Laguerre polynomials are themselves rational. The simplest case occurs when $K = 0$ and $b_0 = \alpha\nu$ for $0 < \nu < 1$, so that $\tilde{\mu}(\omega) = \nu\alpha/(\alpha - i\omega)$ and

$$\gamma(\omega) = \frac{\lambda}{2\pi(1-\nu)} \cdot \frac{\omega^2 + \alpha^2}{\omega^2 + \alpha^2(1-\nu)^2}\,.$$

Note the characteristic feature for point processes with rational spectral density that the numerator and denominator are of equal degree.

Further examples are given in the papers cited and in Vere-Jones and Ozaki (1982). To yield a valid model, the parameters should be constrained to ensure that the density of the infectivity measure (and hence the conditional intensity) is everywhere nonnegative; for stationarity, the infectivity measure should have total mass < 1. These conditions are relatively stringent and quite difficult to impose in estimation procedures. Within these constraints,

however, the Hawkes model is one of the most flexible models available in that it allows both the calculation of the form of the spectrum and the investigation of probabilistic aspects of the process. □

The basic results described so far apply to stationary (translation-invariant) point processes in any general Euclidean space $\mathbb{R}^d$. When $d > 1$, however, additional symmetries such as isotropy (invariance under rotations) become possible and have important implications for the structure of the spectral measures. As an illustration, we conclude this section with a brief discussion of isotropic random measures in $\mathbb{R}^2$, this time looking at the Fourier transforms.

In the stationary, isotropic case, the second-order properties of a random measure in $\mathbb{R}^2$ are fully defined by the mean density m and the function $\breve{K}_2(\cdot)$ defined in (8.1.20). We examine the constraints on the Bartlett spectrum in $\mathbb{R}^2$ implied by this isotropy condition and show how to represent the spectrum in terms of m and $\breve{K}_2(\cdot)$.

Consider first the effect of the double Fourier transform on a function $h\colon \mathbb{R}^2 \mapsto \mathbb{R}$ which, in addition to being bounded, measurable, and of bounded support, is circularly symmetric, i.e.

$$h(x, y) = h(r\cos\theta,\, r\sin\theta) = g(r) \qquad (\text{all } r \in \mathbb{S})$$

for some function g. The transform is given by

$$\begin{aligned}\tilde{h}(\omega, \phi) &\equiv \int_{\mathbb{R}^2} \mathrm{e}^{i(\omega x + \phi y)} h(x, y)\,\mathrm{d}x\,\mathrm{d}y \;=\; \int_0^\infty r g(r)\,\mathrm{d}r \int_0^{2\pi} \mathrm{e}^{ir(\omega\cos\theta + \phi\sin\theta)}\,\mathrm{d}\theta \\ &= \int_0^\infty r g(r)\,\mathrm{d}r \int_0^{2\pi} \mathrm{e}^{ir\rho\cos(\theta-\psi)}\,\mathrm{d}\theta\end{aligned}$$

using (ρ, ψ) as polar coordinates in the (ω, ϕ) plane. Now the integral over θ is simply a Bessel function $J_0(u) = 1/(2\pi)\int_0^{2\pi} \mathrm{e}^{iu\cos\theta}\,\mathrm{d}\theta$, so

$$\tilde{h}(\omega, \phi) = 2\pi \int_0^\infty r J_0(r\rho) g(r)\,\mathrm{d}r \equiv \tilde{g}^B(\rho), \qquad \text{where } \rho = (\omega^2 + \phi^2)^{1/2}. \tag{8.2.11}$$

Consequently, $\tilde{h}(\omega, \phi)$ is again circularly symmetric, reducing to the function $\tilde{g}^B(\cdot)$, which we call the *Bessel transform* of $g(\cdot)$ (we have included the factor 2π—this is a departure from the usual definition) and is also called a Hankel transform—see e.g. Copson (1935, p. 342). By arguing analogously from the inverse Fourier transform

$$h(x, y) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \mathrm{e}^{i(\omega x + \phi y)} \tilde{h}(\omega, \phi)\,\mathrm{d}\omega\,\mathrm{d}\phi,$$

it follows that the Bessel transform is inverted as in

$$g(r) = \frac{1}{2\pi} \int_0^\infty \rho \tilde{g}^B(\rho) J_0(r\rho)\,\mathrm{d}\rho. \tag{8.2.12}$$

From this discussion, we should expect the Bartlett spectral density of a stationary isotropic process to be circularly symmetric in frequency space and to be related to the inverse Bessel transform of the density of $\breve{K}_2(r)$. To cover the situation where densities may not exist, the Bessel transform relation needs to be put into the form of a Parseval relation so that it can be extended to measures, as follows.

Proposition 8.2.III. *Let $\Gamma(\cdot)$ be the Bartlett spectrum on $\mathbb{R}^2$ associated with a simple stationary isotropic point process in $\mathbb{R}^2$. Then $\Gamma(\cdot)$ is circularly symmetric and is expressible via $(\omega_1, \omega_2) = (\rho\cos\psi, \rho\sin\psi)$ as*

$$\Gamma(\mathrm{d}\rho \times \mathrm{d}\psi) = \Big(m\rho\,\frac{\mathrm{d}\rho}{2\pi} + m^2\,\kappa(\mathrm{d}\rho) + 2\pi m^2\,\delta_0(\mathrm{d}\rho)\Big)\frac{\mathrm{d}\psi}{2\pi}\,, \tag{8.2.13}$$

where κ is related to the radial measure $\breve{K}_2(\cdot)$ of (8.1.20) by the Parseval–Bessel equation

$$\int_0^\infty \tilde{g}^B(\rho)\,\kappa(\mathrm{d}\rho) = \int_0^\infty g(r)\,\breve{K}_2(\mathrm{d}r) \tag{8.2.14}$$

for all bounded measurable g of finite support on $\mathbb{R}_+$ and $\tilde{g}^B$ is defined by (8.2.11).

Proof. Recall that the Bartlett spectrum is the Fourier transform in $\mathbb{R}^2$ of the complete covariance measure $\breve{C}_2$, which for disks $S_r(0)$ takes the form

$$\breve{C}_2\big(S_r(0)\big) = m - m^2\pi r^2 + m^2\breve{K}_2(r),$$

where the first term arises from the diagonal concentration associated with a simple point process; the second, the term involving the square of the mean, must be subtracted from the second moment to yield the covariance; and the third is the form of the reduced second factorial moment measure. Using mixed differential notation, this can be rewritten as

$$\breve{C}_2(\mathrm{d}x \times \mathrm{d}y) = m\,\delta_0(\mathrm{d}x \times \mathrm{d}y) - m^2\,\mathrm{d}x\,\mathrm{d}y + m^2\breve{K}_2(\mathrm{d}r)\frac{\mathrm{d}\theta}{2\pi}\,.$$

The first and second terms have the following inverse Fourier transforms, respectively:

$$\frac{m\,\mathrm{d}\omega_1\,\mathrm{d}\omega_2}{(2\pi)^2} = \frac{m\rho\,\mathrm{d}\rho\,\mathrm{d}\psi}{(2\pi)^2} = m\rho\frac{\mathrm{d}\rho}{2\pi}\cdot\frac{\mathrm{d}\psi}{2\pi}\,,$$
$$\frac{4\pi^2 m^2\,\delta_0(\mathrm{d}\omega_1 \times \mathrm{d}\omega_2)}{(2\pi)^2} = 2\pi m^2\,\delta_0(\mathrm{d}\rho)\cdot\frac{\mathrm{d}\psi}{2\pi}\,.$$

Denoting the double Fourier transform of the measure $\breve{K}_2(\mathrm{d}r)\,\mathrm{d}\theta/(2\pi)$ by $L(\mathrm{d}\omega_1 \times \mathrm{d}\omega_2)$, the Parseval relation for such transforms implies that, with h and $\tilde{h}$ as earlier,

$$\int_{\mathbb{R}^2} \tilde{h}(\omega_1,\omega_2)\,L(\mathrm{d}\omega_1 \times \mathrm{d}\omega_2) = \int_0^\infty \breve{K}_2(\mathrm{d}r)\int_0^{2\pi} h(r\cos\theta, r\sin\theta)\,\frac{\mathrm{d}\theta}{2\pi}\,.$$

Now

$$\begin{aligned}\int_0^{2\pi} & h(r\cos\theta, r\sin\theta)\,\frac{\mathrm{d}\theta}{2\pi} \\ &= \frac{1}{(2\pi)^2}\int_0^\infty \mathrm{d}\theta \int_0^\infty \mathrm{d}\rho \int_0^{2\pi} \mathrm{e}^{-i\rho r\cos(\theta-\psi)}\rho\,\tilde{h}(\rho\cos\psi, \rho\sin\psi)\,\mathrm{d}\psi \\ &= \frac{1}{2\pi}\int_0^\infty \mathrm{d}\rho \int_0^{2\pi} \rho J_0(\rho r)\tilde{h}(\rho\cos\psi, \rho\sin\psi)\,\mathrm{d}\psi,\end{aligned}$$

where as before the invariance of integrating θ over any interval of length 2π has been used. If, in particular, we take $\tilde{h}(\omega_1, \omega_2)$ to have the product form $\tilde{g}^B(\rho)f(\psi)$, we obtain from this relation and the Bessel transform equation (8.2.12) that

$$\int_{(0,\infty)\times(0,2\pi)} \tilde{g}^B(\rho)f(\psi)\,L(\mathrm{d}\rho\times\mathrm{d}\psi) = \int_0^{2\pi} f(\psi)\,\frac{\mathrm{d}\psi}{2\pi}\int_0^\infty g(r)\,\breve{K}_2(\mathrm{d}r).$$

Since the integral here depends on f only through its integral over $(0, 2\pi)$, a uniqueness argument implies that $L(\cdot)$ has a disintegration of the form $L(\mathrm{d}\rho\times\mathrm{d}\psi) = \kappa(\mathrm{d}\rho)\,[\mathrm{d}\psi/(2\pi)]$, where $\kappa(\cdot)$ satisfies (8.2.14). □

Note that (8.2.14) defines $(1/r)\breve{K}_2(\mathrm{d}r)\,\mathrm{d}r$ (and not $\breve{K}_2$), in the sense of generalized functions, as the Bessel transform of $(1/\rho)\,\kappa(\mathrm{d}\rho)\,\mathrm{d}\rho$.

Example 8.2(f) *An isotropic Neyman–Scott process.* Consider the circularly symmetric case from Example 8.1(b) and Exercise 8.1.7, for which we have

$$\breve{K}_2(\mathrm{d}r) = 2\pi r\,\mathrm{d}r + \frac{m_{[2]}}{\mu m_1^2} r\mathrm{e}^{-r^2/4\sigma^2}\,\frac{\mathrm{d}r}{2\sigma^2}\,.$$

It is easy to check from (8.2.14) that the measure $2\pi r\,\mathrm{d}r$ on $\mathbb{R}_+$ is the Parseval–Bessel transform of the measure consisting of a unit atom at the origin. The second term is a density, and it can be derived (via the Fourier transform in $\mathbb{R}^2$ or otherwise) as the Parseval–Bessel transform of the density

$$\kappa(\rho) = \frac{\mu m_{[2]}}{2\pi\mu m_1^2}\,\rho\mathrm{e}^{-\sigma^2\rho^2}.$$

Consequently, for this isotropic Neyman–Scott model, the Bartlett spectrum is absolutely continuous with spectral density

$$\gamma(\omega,\phi) = \frac{\mu m_1}{4\pi^2} + \frac{\mu m_{[2]}}{2\pi}\,\mathrm{e}^{-\sigma^2(\omega^2+\phi^2)} \equiv \frac{\beta(\rho)}{2\pi}\,,$$

where the function $\beta(\cdot)$ as just defined exhibits the Bartlett spectrum in the polar form $\beta(\rho)\,\mathrm{d}\rho\,[\mathrm{d}\psi/(2\pi)]$. □

Exercises and Complements to Section 8.2

8.2.1 Given a second-order stationary point process N, the relation $\{X_h(n)\} = \{N(nh,(n+1)h]\}$ defines a second-order stationary discrete time series. Express $\operatorname{var} N(0,nh]$ in terms of the second-moment structure of $\{X_h(n)\}$. Use the standard spectral representation of the second moments of a discrete-time process to give a spectral representation for $\operatorname{var} N(0,nh]$, and argue that for $h \to 0$ there is a weak limit as in (8.2.3).

8.2.2 *Superposition.* Show that if ξ_1, ξ_2 are independent second-order stationary random measures with Bartlett spectra Γ_1, Γ_2, respectively, then $\xi_1 + \xi_2$ has spectrum $\Gamma_1 + \Gamma_2$.

More generally, if $\xi_1, \xi_2, \ldots$ are independent second-order stationary random measures such that the L_2 limit

$$\xi = \xi_1 + \xi_2 + \cdots$$

exists, then ξ has Bartlett spectrum $\Gamma_1 + \Gamma_2 + \cdots$.

8.2.3 *Cox process.* Let ξ be a second-order stationary random measure on $\mathbb{R}^d$ with Bartlett spectrum Γ and mean density m. Show that the Cox process directed by ξ has Bartlett spectrum $\Gamma(\cdot) + m(2\pi)^{-d}\ell(\cdot)$, where $\ell(\cdot)$ denotes Lebesgue measure on $\mathbb{R}^d$.

8.2.4 *Quadratic random measure* [see Example 6.1(c) and Exercise 6.1.3].

(a) Let $Z_i(t) (i = 1, 2)$ be independent mean square continuous second-order stationary random processes on $\mathbb{R}$ with respective spectral d.f.s F_i and zero mean. Show that the product $Z_1 Z_2$ is a mean square continuous second-order stationary process with spectral measure $F_1 * F_2$.

(b) If Z is a mean square continuous stationary Gaussian process with spectral d.f. F and zero mean, then the quadratic random measure whose sample paths have density $Z^2(\cdot)$ has covariance density $2|c(\cdot)|^2$ and Bartlett spectrum $2F * F$, where $c(x) = \operatorname{cov}(Z(0), Z(x))$.

(c) Investigate what changes are needed in (a) and (b) when the zero mean assumption is omitted.

8.2.5 *Cyclic point process on four points.* Consider a $\{0,1\}$-valued process on the four compass points NESW that is stationary (i.e. invariant under cyclic permutations). Denote the probabilities of the six basic configurations 0000, 1000, 1100, 1010, 1110, and 1111 by $\{p_0, p_1, \ldots, p_5\}$, respectively.

(i) Show that the mean density and reduced second-moment measure are given respectively by

$$\begin{aligned} m &= \tfrac{1}{4}p_1 + \tfrac{1}{2}(p_2 + p_3) + \tfrac{3}{4}p_4 + p_5, \\ \breve{M}_2 &= \{a, b, c, d\}, \end{aligned}$$

where $a = m,\ b = d = \frac{1}{4}p_2 + \frac{1}{2}p_4 + p_5,\ c = \frac{1}{2}p_3 + \frac{1}{2}p_4 + p_5$. Show that $\breve{M}_2$ is a p.p.d. measure with Fourier transform proportional to $(a + c + 2b,\ a - c,\ a + c - 2b,\ a - c)$.

(ii) Renormalize the probabilities so that $m = 1$ (equivalent to looking at the Palm measure and its first moment) and the second-moment measure has

standardized form $\{1, \beta, \gamma, \beta\}$. Show that this is a p.p.d. measure if and only if β, γ are nonnegative and $\gamma \le 1$, $1 + \gamma \ge 2\beta$. However, this is the second-moment measure of a point process on NESW if and only if, in addition, $1 + \beta \ge 2\gamma$. [*Hint*: Write $x = \frac{1}{2}p_4 + p_5$, $y = \frac{1}{4}p_1 + \frac{1}{4}p_4$, so that $x < \min(\beta, \gamma)$ and (x, y) lies on the line $y = 3x - K$, where $K = 2\beta + 2\gamma - 1$. Nonnegative solutions x, y exist if and only if $\frac{1}{3}K \le \min(\beta, \gamma)$, which yields both the p.p.d. condition and the additional condition.]

8.2.6 *Stationary renewal process.* Let the lifetime d.f. $F(\cdot)$ of the process as in Example 8.2(b) be the convolution of two exponentially distributed random variables with means $1/\mu_j$ $(j = 1, 2)$. Evaluate (8.2.5) explicitly.

8.2.7 *Random translations.* Let the point process N be second-order stationary with Bartlett spectrum Γ and mean density m. If the points of N are subjected to independent random translation with common d.f. F, show that the resultant point process N_T has Bartlett spectrum [see (8.2.8)]

$$\Gamma_T(\mathrm{d}\omega) = |\widetilde{F}(\omega)|^2\, \Gamma(\mathrm{d}\omega) + m(2\pi)^{-d}(1 - |\widetilde{F}(\omega)|^2)\, \ell(\mathrm{d}\omega).$$

8.2.8 *Iterated random translations.* Let the independent translation of points of N as in Exercise 8.2.7 be iterated n times. Show that the Bartlett spectrum Γ_n of the resulting process satisfies

$$\begin{aligned}\Gamma_n(\mathrm{d}\omega) &= |\widetilde{F}(\omega)|^2\, \Gamma_{n-1}(\mathrm{d}\omega) + m(2\pi)^{-d}(1 - |\widetilde{F}(\omega)|^2)\, \ell(\mathrm{d}\omega) \\ &= |\widetilde{F}(\omega)|^{2n}\, \Gamma(\mathrm{d}\omega) + m(2\pi)^{-d}(1 - |\widetilde{F}(\omega)|^{2n})\, \ell(\mathrm{d}\omega)\end{aligned}$$

and hence give conditions for $\Gamma_n(\cdot)$ to converge weakly to $m(2\pi)^{-d}\ell(\cdot)$. (See Chapter 11).

8.2.9 *Neyman–Scott process* [continued from Example 6.3(a)].

(a) Show that the Bartlett spectrum for a Neyman–Scott process on $\mathbb{R}$, with (Poisson) cluster centre process at rate μ_c, $m_{[1]}$ and $m_{[2]}$ for the first two factorial moments of the cluster size distribution, and common d.f. F for the distances of the points of a cluster from their centre, has density $\gamma_{\mathrm{NS}}(\omega)$ given by

$$\gamma_{\mathrm{NS}}(\omega) = (\mu_c/2\pi)[m_{[1]} + m_{[2]}|\widetilde{F}(\omega)|^2],$$

where $\widetilde{F}(\omega) = \int_{-\infty}^{\infty} \mathrm{e}^{ix\omega}\, F(\mathrm{d}x)$.

(b) In the particular case where $F(x) = 1 - \mathrm{e}^{-\alpha x}$ $(x \ge 0)$, deduce that $\gamma_{\mathrm{NS}}(\cdot)$ is the rational function

$$\gamma_{\mathrm{NS}}(\omega) = \frac{\mu_c m_{[1]}}{2\pi}\Big[1 + \frac{\alpha^2 m_{[2]}/m_{[1]}}{\alpha^2 + \omega^2}\Big].$$

(c) When the Neyman–Scott process is as above on $\mathbb{R}^d$, show that

$$\gamma_{\mathrm{NS}}(\omega) = (\mu_c m_{[1]}/(2\pi)^d)[1 + (m_{[2]}/m_{[1]})|\widetilde{F}(\omega)|^2]$$

with $\widetilde{F}(\omega) = \int_{\mathbb{R}^d} \mathrm{e}^{i x\cdot\omega}\, F(\mathrm{d}x)$. Deduce that when $d = 2$ and $F(\cdot)$ is a bivariate normal d.f. with zero mean and the usual second-moment parameters σ_1^2, σ_2^2 and $\rho\sigma_1\sigma_2$, the spectrum has density

$$\gamma_{\mathrm{NS}}(\omega_1, \omega_2) = \frac{\mu_c m_{[1]}}{4\pi^2}\Big[1 + \frac{m_{[2]}}{m_{[1]}} \exp(-\sigma_1^2\omega_1^2 - 2\rho\sigma_1\sigma_2\omega_1\omega_2 - \rho_2^2\omega_2^2)\Big].$$

(d) Show that if in (a) the cluster structure is modified to include the cluster centre, then

$$\gamma_{\mathrm{NS}}(\omega) = (\mu_c/2\pi)[1 + m_{[1]}(1 + \widetilde{F}(\omega) + \widetilde{F}(-\omega)) + m_{[2]}|\widetilde{F}(\omega)|^2].$$

(e) Show that if in (a) the cluster centre process is a general stationary point process with mean intensity μ_c and Bartlett spectrum $\Gamma_c(\cdot)$, then the Bartlett spectrum $\Gamma_{\mathrm{NS}}(\cdot)$ of the cluster process is given by

$$\Gamma_{\mathrm{NS}}(\mathrm{d}\omega) = |m_{[1]}\widetilde{F}(\omega)|^2\,\Gamma_c(\mathrm{d}\omega) + \frac{\mu_c}{2\pi}[m_{[1]} + (m_{[2]} - m_{[1]}^2)|\widetilde{F}(\omega)|^2]\,\ell(\mathrm{d}\omega).$$

[*Hint*: Except for (d), the results can be derived, first by compounding and then by using random translations as in Exercise 8.2.7; otherwise, see (8.2.8).]

8.2.10 *Bartlett–Lewis model* [continued from Example 6.3(b)].

(a) Use (6.3.23) to show that the Bartlett spectrum has density $\gamma_{\mathrm{BL}}(\cdot)$ given by

$$\gamma_{\mathrm{BL}}(\omega) = \frac{\mu_c}{2\pi}\left[\sum_{j=0}^{\infty}(j+1)q_j + \sum_{j=1}^{\infty}\sum_{k=j}^{\infty}(k+1-j)q_k\big(\widetilde{F}^j(\omega) + \widetilde{F}^j(-\omega)\big)\right].$$

Observe that $\gamma_{\mathrm{BL}}(\omega) = \gamma_{\mathrm{NS}}(\omega)$ as in Exercise 8.2.9(d) in the cases $q_1 = 1$ and $m_{[1]} = 1$, $m_{[2]} = 0$, respectively.

(b) Show that when $q_j = (1-\alpha)\alpha^j \quad (j = 0, 1, \ldots)$ with $0 < \alpha < 1$, so that each cluster is a transient renewal process,

$$\gamma_{\mathrm{BL}}(\omega) = \frac{\mu_c}{2\pi(1-\alpha)}\left[\frac{1}{1-\alpha\widetilde{F}(\omega)} + \frac{1}{1-\alpha\widetilde{F}(-\omega)} - 1\right],$$

while when $q_0 = 0$, $q_j = (1-\alpha)\alpha^{j-1}$ $(j = 1, 2, \ldots)$,

$$\gamma_{\mathrm{BL}}(\omega) = \frac{\mu_c}{2\pi\alpha(1-\alpha)}\left[\frac{1}{1-\alpha\widetilde{F}(\omega)} + \frac{1}{1-\alpha\widetilde{F}(-\omega)} - 1 - (1-\alpha)^2\right].$$

(c) The formulae in parts (a) and (b) assume that the cluster centre is included in the cluster process. Show that omitting the cluster centres leads to

$$\begin{aligned}\gamma_{\mathrm{BL}}(\omega) &= \frac{\mu_c}{2\pi}\left[\sum_{j=1}^{\infty} jq_j + \sum_{k=j+1}^{\infty}(k-j)q_k\big(\widetilde{F}^j(\omega) + \widetilde{F}^j(-\omega)\big)\right] \\ &= \frac{\mu_c}{2\pi}\left[\sum_{j=1}^{\infty} jq_j + \sum_{j=2}^{\infty} q_j \sum_{k=1}^{j-1}(j-k)\big(\widetilde{F}^k(\omega) + \widetilde{F}^k(-\omega)\big)\right].\end{aligned}$$

8.2.11 Let M_2 be a p.p.d. measure on $\mathcal{B}_{\mathbb{R}}$ with density m_2. Show that if $0 < a \le m_2(x) \le b < \infty$ (all x) then there exists a zero-mean Gaussian process $X(t)$ such that $m_2(x) = \mathrm{E}[X^2(t)X^2(t+x)]$ and hence that M_2 is the reduced second-moment measure of the process $\xi(A) = \int_A X^2(t)\,\mathrm{d}t$ $(A \in \mathcal{B}_{\mathbb{R}})$. Deduce that any p.p.d. function $c_2(\cdot)$ can be a reduced covariance density; i.e. there is some $a > 0$ such that $a + c_2(x)$ is the second-moment density of some second-order stationary random measure.

8.2.12 Let F be any totally bounded symmetric measure $\mathbb{R}^d$. Show that F can be a covariance measure. [*Hint*: Construct a Gauss–Poisson process and refer to Proposition 6.3.IV. See Milne and Westcott (1972) for further details.]

8.3. Multivariate and Marked Point Processes

This section provides a first introduction to the wide range of extensions of the previous theory, incorporating both time-domain and frequency-domain aspects. We look first at multivariate and marked point processes, with stationarity in time (i.e. translation invariance) still playing the central role.

The results given thus far for second-order stationary random measures and point processes on $\mathbb{R}^d$ extend easily to multivariate processes on $\mathbb{R}^d$, though for convenience we discuss mostly the case $d = 1$. The first-moment measure in Proposition 8.1.I(a) becomes a vector of first-moment measures

$$M_i(A) = \mathrm{E}[\xi_i(A)] \qquad (i = 1, \ldots, K;\ A \in \mathcal{B}_{\mathbb{R}}),$$

one for each of the K components. Under stationarity, which means translation invariance of the *joint* probability structure, not just of each component separately, this reduces to a vector of mean densities $\{m_i,\ i = 1, \ldots, K\}$.

Similarly, the second-order moment and covariance measures in the univariate case are replaced by matrices $\mathbf{M}$ and $\mathbf{C}$ of auto- and cross-moment (or covariance) measures with elements for $i, j = 1, \ldots, K$ and $A,\ B \in \mathcal{B}_{\mathbb{R}}$,

$$\begin{aligned} M_{ij}(A \times B) &= \mathrm{E}[\xi_i(A)\xi_j(B)], \\ C_{ij}(A \times B) &= M_{ij}(A \times B) - M_i(A)M_j(B). \end{aligned}$$

Under stationarity, the diagonal components M_{ii} are invariant under simultaneous shifts in both coordinates and so possess reduced forms $\breve{M}_{ii}$, which inherit the properties of the reduced moment measures listed in Proposition 8.1.II. More than this is true, however. Since every linear combination $\sum_{i=1}^k \alpha_i \xi_i(A_i)$ is again stationary, we find on taking expectations of the squares that the quadratic forms $\sum_{i=1}^k \sum_{j=1}^k \alpha_i \alpha_j M_{ij}(A_i \times A_j)$ are all stationary under diagonal shifts and therefore possess diagonal factorizations. From this there follows the existence of reduced forms, $\breve{M}_{ij}(\cdot)$, $\breve{C}_{ij}(\cdot)$, say, for the off-diagonal as well as the diagonal components of the matrices.

In the point process case, the off-diagonal components $\breve{M}_{ij}$, $\breve{C}_{ij}$ $(i \neq j)$ will not have the atom at the origin characteristic of the diagonal components unless there is positive probability of pairs of points occurring simultaneously in both the i and j streams. In particular, if the ground process $N_{\mathrm{g}}(\cdot) = \sum_{i=1}^K N_i(\cdot)$ is orderly, both the matrix of reduced factorial moment measures

$$\breve{\mathbf{M}}(A) = \big(\breve{M}_{[i,j]}(A)\big) = \big(\breve{M}_{ij}(A) - [\delta_{ij}\delta_0(A)m_i]\ell(A)\big)$$

and the corresponding matrix of reduced factorial covariance measures with elements

$$\breve{C}_{[i,j]}(A) = \breve{M}_{[i,j]}(A) - m_i\, m_j \ell(A)$$

will be free from atoms at the origin.

Whether or not such atoms exist, the matrix $\breve{\mathbf{M}}$ enjoys matrix versions of the properties listed in Proposition 8.1.II; we state them for clarity.

Proposition 8.3.I (Stationary multivariate random measure: Second-order moment properties)**.**

(i) $\breve{\mathbf{M}}(A) \geq 0$, *with* $\breve{M}_{ii}(A) > 0$ *if* $A \ni 0$ *and either* N_i *has an atomic component or* A *is an open set;*

(ii) $\breve{\mathbf{M}}(A) = \breve{\mathbf{M}}^T(-A)$*;*

(iii) $\breve{\mathbf{M}}$ *is positive-definite: for all finite sequences* $\{f_i\}$ *of bounded measurable complex functions of bounded support,*

$$\sum_{i=1}^{K}\sum_{j=1}^{K}\int_{\mathbb{R}} f_i(x)\overline{f_j(x+u)}\,\breve{M}_{ij}(\mathrm{d}u) \geq 0; \tag{8.3.1}$$

(iv) $\breve{\mathbf{M}}$ *is translation-bounded: for given* A*, there exists a constant* K_A *such that* $||\breve{\mathbf{M}}(x+A)|| = \sum_{i,j=1}^{K} |\breve{M}_{ij}(x+A)| < K_A$*;*

(v) *If also the process is ergodic as for equations* (8.1.4–5)*, then as* $r(A) \to \infty$*,* $\breve{\mathbf{M}}(A)/\ell(A) \to \breve{\mathbf{M}}_\infty \equiv (m_i m_j)$*, and for all bounded Borel sets* B*,*

$$\frac{1}{\ell(A)}\int_A \xi_i(x+B)\,\xi_j(\mathrm{d}x) \to \breve{M}_{ij}(B).$$

The properties follow readily from the same device of applying the univariate results to linear combinations of the components (see Exercise 8.3.1).

Note that property (ii) implies that the diagonal measures are symmetric, while for the off-diagonal measures $\breve{M}_{ij}(A) = \breve{M}_{ji}(-A)$, confirming the importance of order in specifying the cross-moments.

The spectral theory also extends easily to multivariate processes on $\mathbb{R}$. For any linear combination of the components, the basic p.p.d. properties (i) and (iii) above are interchanged by the Fourier transform map, implying that the moment measures can be represented by a matrix of spectral measures, which again enjoys the properties listed above (see Exercise 8.3.2).

For practical purposes, the multivariate extension of the Bartlett spectrum (Definition 8.2.II) is of greatest importance. This comprises the matrix $\mathbf{\Gamma}$ of auto- and cross-spectral measures $\big(\Gamma_{ij}(\cdot)\big)$ in which the diagonal elements $\Gamma_{ii}(\cdot)$ have the properties described in Section 8.2 and the matrix as a whole has the positive-definiteness property in (8.3.1). Indeed, (8.3.1) can be regarded as being derived from the filtered form

$$X(t) = \sum_{i=1}^{k}\int_{-\infty}^{\infty} f_i(t-u)\,\xi_i(\mathrm{d}u) \tag{8.3.2}$$

for which the spectral measure Γ_X has the form

$$\Gamma_X(\mathrm{d}\omega) = \sum_{i=1}^{k}\sum_{j=1}^{k}\tilde{f}_i(\omega)\overline{\tilde{f}}_j(\omega)\,\Gamma_{ij}(\mathrm{d}\omega). \tag{8.3.3}$$

In the generality considered here, the components ξ_i at (8.3.2) may be point processes or random measures. If the latter are absolutely continuous, the appropriate components of the matrix $\mathbf{\Gamma}$ then reduce to the usual spectra

and cross-spectra of the stationary processes formed by their densities. In this way, the theory embraces both point and continuous processes as well as mixed versions. If the continuous process has varying sign, as occurs with a Gaussian process, or is given in the wide sense only, then the appropriate framework is the matrix extension of the wide sense theory summarized after Definition 8.4.VII.

From the practical viewpoint, these remarks mean that the interaction of point process systems, or mixtures of point process and continuous systems, can be studied in the frequency domain very much as if they were all continuous systems. The essential difference is that each point process component leads to a δ-function component in the diagonal term $\breve{C}_{ii}(\cdot)$ to which there is then a corresponding nonzero constant contribution in the spectral measure $\Gamma_{ii}(\cdot)$. Bearing this in mind, all the standard concepts of multivariate spectral theory, such as coherence and phase, or real and quadratic spectra, carry over with minor variations to this more general context and provide valuable tools for the descriptive analysis of multivariate point processes and mixed systems. Brillinger (1975a, b, 1978, 1981) outlines both differences and similarities; for an example studied in depth, see Brillinger (1992).

The next two examples illustrate simple special cases of these ideas.

EXAMPLE 8.3(a) *A bivariate Poisson process* [continued from Example 6.3(e)]. The stationary bivariate point process described earlier is determined by three parameters: rates μ_1 and μ_2 for the occurrence of single points in processes 1 and 2, respectively, and a boundedly finite measure $\widetilde{Q}_3(\mathrm{d}u) = \mu_3 G(\mathrm{d}u)$ on $\mathbb{R}$, in which μ_3 is the rate of occurrence of pairs of points, one in each process and $G(\mathrm{d}u)$ is a probability distribution for the signed distance u from the process 1 point to the other point. It is convenient for the rest of the example to have $G(\mathrm{d}u) = g(u)\,\mathrm{d}u$ for some probability density function $g(\cdot)$ on $\mathbb{R}$.

Since the two component processes are both Poisson, the only nonzero second-order factorial cumulant measure is in the cross-covariance term, with

$$\breve{C}_{[12]}(A) = \mu_3 \int_A g(u)\,\mathrm{d}u = \breve{C}_{[21]}(-A).$$

The matrices $\breve{\mathbf{m}}(u)$, $\breve{\mathbf{c}}(u)$ of densities for the matrices $\breve{\mathbf{M}}$, $\breve{\mathbf{C}}$ of reduced second-moment measures are given respectively by

$$\breve{\mathbf{m}}(u) = \begin{pmatrix} \mu_1+\mu_3 & 0 \\ 0 & \mu_2+\mu_3 \end{pmatrix} \delta_0(u) + \begin{pmatrix} (\mu_1+\mu_3)^2 & (\mu_1+\mu_3)(\mu_2+\mu_3)+\mu_3 g(u) \\ (\mu_1+\mu_3)(\mu_2+\mu_3)+\mu_3 g(-u) & (\mu_2+\mu_3)^2 \end{pmatrix}$$

and

$$\breve{\mathbf{c}}(u) = \begin{pmatrix} \mu_1+\mu_3 & 0 \\ 0 & \mu_2+\mu_3 \end{pmatrix} \delta_0(u) + \begin{pmatrix} 0 & \mu_3 g(u) \\ \mu_3 g(-u) & 0 \end{pmatrix}.$$

The corresponding Bartlett spectra are all absolutely continuous, the densities $\gamma_{ij}(\omega)$ of the matrix $\mathbf{\Gamma}$ being given by

$$\frac{1}{2\pi}\begin{pmatrix} \mu_1+\mu_3 & \mu_3\widetilde{G}(\omega) \\ \mu_3\widetilde{G}(-\omega) & \mu_2+\mu_3 \end{pmatrix}, \tag{8.3.4}$$

where $\widetilde{G}(\omega) = \int_{\mathbb{R}} \mathrm{e}^{-iu\omega} g(u)\,\mathrm{d}u$. The *coherence* of the two processes, at frequency ω, is the ratio

$$\rho_{12}(\omega) = \frac{\mu_3|\widetilde{G}(\omega)|}{\sqrt{(\mu_1+\mu_3)(\mu_2+\mu_3)}},$$

while their *phase* at the same frequency is

$$\theta_{12}(\omega) = \arctan\left(\frac{\mathrm{Im}(\widetilde{G}(\omega))}{\mathrm{Re}(\widetilde{G}(\omega))}\right). \qquad \square$$

Example 8.3(b) *System identification: a special case.* In the previous example, the spectral densities completely determine the parameters of the process. This leads to the more general problem of determining the characteristics of a point process *system*, meaning some mechanism for producing a point process output from a point process input. Deletions (or thinnings), delays (or translations), and triggering of clusters can all be regarded as examples of point process systems. The problem of system identification then consists of determining the mechanism, or at least its main features, from measurements on its input and output. The two components of the previous example can be regarded as the input and output of a system specified as follows: a proportion $\pi_1 = \mu_1/(\mu_1+\mu_3)$ of the input points are randomly deleted while each of the points in the remaining proportion $\pi_2 = 1-\pi_1$ is transmitted after independent delays with d.f. G [such a specification requires $G(\cdot)$ to be concentrated on a half-line], with this transmitted output being contaminated with 'noise' consisting of the points of a Poisson process at rate μ_2. It is evident from the spectral representation in (8.3.4) that the three system parameters π_1, G and μ_2 can be identified by measuring the response of the system to a Poisson input process and finding the joint first- and second-order properties of the input and output. It is equally evident that this identification is impossible on the basis of separate observations of the input and output.

Suppose now that the Poisson input process is replaced by any simple stationary input process with mean density m and spectral density $\gamma(\cdot)$ in place of $(\mu_1+\mu_3)/(2\pi)$. Then, in place of the matrix with components at (8.3.4), we would have the matrix

$$\begin{pmatrix} \gamma(\omega) & \gamma(\omega)\widetilde{G}(\omega) \\ \overline{\gamma(\omega)\widetilde{G}(\omega)} & \dfrac{\mu_2+m\pi_2}{2\pi} + \pi_1^2|\widetilde{G}(\omega)|^2\Big(\gamma(\omega)-\dfrac{m}{2\pi}\Big) \end{pmatrix}. \tag{8.3.5}$$

Once more it is evident that in principle the parameters π_1, G and μ can be identified from this matrix of spectral densities. $\square$

Many applications of multivariate point process models arise as extensions of contingency table models when more precise data become available concerning the occurrence times of the registered events. Typical examples arise in the analysis of medical or epidemiological data collected by different local authorities. If the only data available represent counts of occurrences for each region and within crude (e.g. yearly) time intervals, then methods of categorical data analysis may help to uncover and interpret spatial and temporal dependences. If, however, the data are extended to record the times of each individual occurrence, then marked point process methods may be more appropriate. Several recent books, such as Cressie (1991), Ripley (1988) and Guttorp (1995), provide useful introductions to and examples of such studies. The interpretation of the marks, however, is by no means restricted to such spatial examples. Examples abound in neurophysiology, geology, physics, astronomy, and so on in which interest centres on the evolution and interdependencies of sequences of events involving different types of events.

The first stages in the point process analysis of such data are likely to involve descriptive studies, which have the aim of mapping basic characteristics and dependences. Here, while they may be followed later by model-fitting and testing exercises, nonparametric estimates of the first- and second-order characteristics are of particular importance. Such estimates closely follow the univariate forms described earlier [see in particular (8.1.4–5) and (8.1.16–17)]. They take their cue from (8.1.5) in Proposition 8.1.II. Since we are considering MPPs with time as the underlying dimension, estimates such as (8.1.16) for the reduced moment measures here take the form

$$\widehat{M}_{jk}((0,\tau]) = \frac{1}{T} \sum_{i:0\le t_{ik}<T} N_j(t_{ik}, t_{ik}+\tau]\,. \tag{8.3.6}$$

In the cross terms, the sum is extended over events of type k while the counts are for events of type j. Edge corrections of the type (8.1.26) can be incorporated, or more simply one could apply the plus sampling modification, which in the one-dimensional context would amount to including within the sum the full contributions $N(t_{ik}, t_{ik}+\tau]$ initiated by events of type k with $t_{ik} < T < t_{ik}+\tau$.

Models for such processes typically involve extensions and modifications of the basic univariate models. In particular, it is very easy to develop extensions of the standard cluster models in which the cluster members may be events of different types (see Exercise 8.3.3). More complex versions allow events of any one type to produce 'offspring' of other types. Perhaps the most important such example is the multivariate extension of the Hawkes process considered below.

Example 8.3(c) *Mutually exciting point processes.* Hawkes (1971b, 1972) generalized the model described in Examples 6.4(c) and 7.2(b) to both the multivariate and marked point process cases. We give here the multivariate model but via a cluster process representation, where the branching process

now consists of points of K different types and for each $i,\ j = 1, \ldots, K$ there is a Poisson process of offspring of type j generated by an ancestor of type i at time t governed by the parameter measure $\mu_{ij}(\cdot \mid t)$, all these processes being independent and each new offspring generating its own Poisson process. Assume homogeneity of such offspring processes by setting $\mu_{ij}(s \mid t) = \mu_{ij}(s - t)$ as earlier in Example 6.4(c) and, to ensure that there are a.s. only finitely many descendants to any given individual, that the eigenvalue of largest modulus of the matrix $\big(\mu_{ij}(\mathbb{R})\big)$, which by Perron–Frobenius theory is necessarily positive, is smaller than 1. Finally, suppose that type i points enter the system from outside as ancestors in a Poisson process at rate λ_i $(i = 1, \ldots, K)$.

For notational simplicity, we confine attention to the case where the $\mu_{kl}(\cdot)$ have densities (i.e. $\mu_{kl}(\mathrm{d}v) = \mu_{kl}(v)\,\mathrm{d}v$, say). Then, results from branching processes in Section 5.5 (see e.g. Exercise 5.5.7) show for the cluster member processes first that the first-moment measures $M_{ki}(\cdot)$ have densities $m_{ki}(\cdot)$ for which

$$m_{ki}(x) = \delta_{ik}\delta_0(x) + \sum_{l=1}^{K} \int_{\mathbb{R}} \mu_{kl}(v) m_{li}(x - v)\,\mathrm{d}v, \tag{8.3.7}$$

and for the second-order measures we have the densities

$$m_{k,ij}(x, y) = m_{ki}(x)m_{kj}(y) + \sum_{l=1}^{K} \int_{\mathbb{R}} \mu_{kl}(v) m_{l,ij}(x - v,\ y - v)\,\mathrm{d}v. \tag{8.3.8}$$

The first- and second-moment densities, which incorporate an appropriate δ-function, can be interpreted as

$$m_{ki}(x)\,\mathrm{d}x = \Pr\left\{\begin{array}{c}\text{ancestor of type } k \text{ born at 0 has}\\ \text{type } i \text{ descendant born in } (x, x + \mathrm{d}x)\end{array}\right\},$$

$$m_{k,ij}(x, y)\,\mathrm{d}x\,\mathrm{d}y = \Pr\left\{\begin{array}{c}\text{ancestor of type } k \text{ born at 0 has type } i \text{ and } j \text{ descen-}\\ \text{dants born in } (x, x + \mathrm{d}x) \text{ and } (y, y + \mathrm{d}y), \text{ respectively}\end{array}\right\}.$$

Thus, the mean density of type i points, assuming stationarity, is given by

$$m_i \equiv \sum_{k=1}^{K} \lambda_k \int_{\mathbb{R}} m_{ki}(x)\,\mathrm{d}x. \tag{8.3.9}$$

The integral in (8.3.9) can be found by solving (8.3.7) after integration, but for later use it is better now to introduce the Fourier transforms

$$\widetilde{m}_{ij}(\omega) = \int_{\mathbb{R}} \mathrm{e}^{ix\omega} m_{ij}(x)\,\mathrm{d}x, \qquad \widetilde{\mu}_{ij}(\omega) = \int_{\mathbb{R}} \mathrm{e}^{ix\omega} \mu_{ij}(x)\,\mathrm{d}x,$$

so that $\widetilde{m}(\omega) \equiv \big(\widetilde{m}_{ij}(\omega)\big)$ and $\widetilde{\mu}(\omega) \equiv \big(\widetilde{\mu}_{ij}(\omega)\big)$ are related by

$$\widetilde{m}(\omega) = \big(I - \widetilde{\mu}(\omega)\big)^{-1}, \tag{8.3.10}$$

and the column vector $(m_1, \ldots, m_K)^T = \widetilde{m}(0)(\lambda_1, \ldots, \lambda_K)^T$. The inverse at

(8.3.10) is well defined because the largest eigenvalue of $\big(\mu_{ij}(\mathbb{R})\big) = \big(\widetilde{\mu}_{ij}(0)\big)$ is by assumption less than 1.

Similar lengthier analysis starting from (8.3.8) and using the multitype extension of the relation in (6.3.14) for the reduced covariance density in terms of the second-order cluster member densities leads to

$$\breve{c}_{ij}(u) = \sum_{k=1}^{K} \lambda_k \int_{\mathbb{R}} m_{k,ij}(x,\, x+u)\,\mathrm{d}x,$$

in which the $m_{k,ij}(\cdot)$ are multitype analogues of $\rho_{[2]}(\cdot)$ in (6.3.14). This leads ultimately to the matrix of spectral densities as

$$\begin{aligned}\big(\gamma_{ij}(\omega)\big) &= \left(\frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{e}^{iu\omega}\breve{c}_{ij}(u)\,\mathrm{d}u\right) = \frac{1}{2\pi}\,\widetilde{m}^{T}(-\omega)\,\mathrm{diag}(m_1,\ldots,m_K)\widetilde{m}(\omega)\\ &= \frac{1}{2\pi}\big(I - [\widetilde{\mu}(-\omega)]^{T}\big)^{-1}\,\mathrm{diag}(m_1,\ldots,m_K)\big(I-\widetilde{\mu}(\omega)\big)^{-1}, \qquad (8.3.11)\end{aligned}$$

which generalizes (8.2.10).

Hawkes (1971b) derived (8.3.11) using a Wiener–Hopf argument and the linear intensity structure

$$\lambda_i^*(t) = \lambda_i + \sum_k \int_{-\infty}^{t} \mu_{ki}(t-s)\,\mathrm{d}N_k(s).$$

A range of further models can be obtained by varying the character of the cluster centre process while keeping the mutually exciting form for the cluster members (see Exercise 8.3.4). □

We now turn to the second-order properties of MPPs with general mark space. We consider point processes taking their values in $\mathcal{X} = \mathbb{R}\times\mathcal{K}$ for some c.s.m.s. $\mathcal{K}$ so that the process consists of pairs (t_i,κ_i), where $t_i\in\mathbb{R}$ and $\kappa_i\in\mathcal{K}$. We assume stationarity along the time axis $\mathbb{R}$ and suppose that the first- and second-moment measures exist as boundedly finite measures in $\mathcal{X}$ and $\mathcal{X}^{(2)}$. The main emphasis is on time-domain properties—that is, on the moment and covariance measures themselves—rather than on their Fourier transforms. Much of this theory can be extended immediately to homogeneous point processes in $\mathbb{R}^d$, but mostly we leave such extensions to follow the more systematic analysis of homogeneous processes in Chapter 12.

Although we have met already several examples of processes of this type, particularly in Chapter 6, it may still be helpful to start by listing formally the basic properties of their first- and second-order moment measures.

Proposition 8.3.II (Moment Structure of Stationary MPP)**.** *Let $N(\cdot)$ on $\mathbb{R}\times\mathcal{K}$ be a simple stationary marked point process for which the first- and second-moment measures exist. Then, defining $u = t_2 - t_1$, the first- and second-moment measures have respective factorizations*

$$M_1(\mathrm{d}t\times\mathrm{d}\kappa) = F(\mathrm{d}\kappa)\,\mathrm{d}t, \qquad (8.3.12)$$

$$M_2(\mathrm{d}t_1\times\mathrm{d}t_2\times\mathrm{d}\kappa_1\times\mathrm{d}\kappa_2) = \breve{M}_2(\mathrm{d}u\times\mathrm{d}\kappa_1\times\mathrm{d}\kappa_2)\,\mathrm{d}t_1, \qquad (8.3.13)$$

corresponding, respectively, to the following integral relations, valid for bounded measurable h with bounded support:

$$\int_{\mathbb{R}\times\mathcal{K}} h(t,\kappa)\, M_1(\mathrm{d}t \times \mathrm{d}\kappa) = \int_{\mathbb{R}} \mathrm{d}t \int_{\mathcal{K}} h(t,\kappa)\, F(\mathrm{d}\kappa), \tag{8.3.14}$$

$$\int_{(\mathbb{R}\times\mathcal{K})^{(2)}} h(t_1,t_2,\kappa_1,\kappa_2)\, M_2(\mathrm{d}t_1 \times \mathrm{d}t_2 \times \mathrm{d}\kappa_1 \times \mathrm{d}\kappa_2)$$
$$= \int_{\mathbb{R}} \mathrm{d}t \int_{\mathbb{R}\times\mathcal{K}\times\mathcal{K}} h(t,t+u,\kappa_1,\kappa_2)\, \breve{M}_2(\mathrm{d}u \times \mathrm{d}\kappa_1 \times \mathrm{d}\kappa_2). \tag{8.3.15}$$

Proof. Both statements are straightforward applications of the factorization Lemma A2.7.II, the second after taking coordinates in the space $\mathcal{X}^{(2)}$ so that $(t_1,t_2,\kappa_1,\kappa_2) \mapsto (t_1,t_1+u,\kappa_1,\kappa_2)$ (see Exercise 8.3.5). □

If the ground process has a finite mean density $m_{\mathrm{g}} = \mathrm{E}[N((0,1]\times\mathcal{K})]$, then the measure F is totally finite with $F(\mathcal{K}) = m_{\mathrm{g}}$, and we can thus introduce a probability measure Π on $(\mathcal{K},\mathcal{B}(\mathcal{K}))$ by setting

$$\Pi(A) = F(A)/F(\mathcal{K}) \qquad \big(A \in \mathcal{B}(\mathcal{K})\big). \tag{8.3.16}$$

$\Pi(A)$ can then be interpreted as the *stationary distribution of marks*.

The assumption $m_{\mathrm{g}} < \infty$ is not implied directly by the assumption that the first-moment measure exists (i.e. defines a boundedly finite measure in $\mathbb{R}\times\mathcal{K}$), though to our knowledge all extant counterexamples are nonergodic in character (see Exercise 8.3.6).

The distribution Π has two further important interpretations. First, it is an *ergodic* probability in the sense (see Chapter 12) that, if the process is ergodic and $T \to \infty$,

$$\frac{\#\{(t_i,\kappa_i)\colon 0 < t_i < T, \kappa_i \in A\}}{T} = \frac{N((0,T]\times A)}{T} \to \Pi(A) \quad \text{a.s.}$$

Second, it can be interpreted as the distribution of the mark associated with an arbitrary (loosely, randomly selected) time point (event) t_i of the process. Equivalently, it is the distribution of the mark associated with an event at the origin, given that an event of some kind occurs at the origin. This is the interpretation as a Palm probability, as intimated in Chapter 6 and developed in greater detail in Chapter 13.

The reduced second-moment measure $\breve{M}_2(\mathrm{d}u\times\mathrm{d}\kappa_1\times\mathrm{d}\kappa_2)$ also has a range of important interpretations. For $u \neq 0$, it represents the rate of occurrence of pairs of points u time units apart, the first having its mark in $(\kappa_1,\kappa_1+\mathrm{d}\kappa_1)$ and the second, at the (signed) distance u from the first, having its mark in $(\kappa_2,\kappa_2+\mathrm{d}\kappa_2)$. Note that the order of marks can be distinguished; when $u \neq 0$ and the density $\breve{m}_2(u)$ exists, we have

$$\breve{m}_2(u,\kappa_1,\kappa_2) = \breve{m}_2(-u,\kappa_2,\kappa_1),$$
$$\neq \breve{m}_2(u,\kappa_2,\kappa_1) \qquad \text{in general.}$$

Again, there is an interpretation as an ergodic limit: for $T \to \infty$,

$$\frac{\#\{\text{pairs } (t_i,\kappa_i),(t_j,\kappa_j)\colon 0<t_i<T,\ 0<t_j-t_i<u,\kappa_i\in A,\kappa_j\in B\}}{T} \to \breve{M}_2((0,u]\times A\times B) \quad \text{a.s.}$$

Several different interpretations as a Palm measure are possible, depending on whether one conditions on a point at the origin, without any condition on the mark; on a point at the origin with specified mark; or on two points at a given separation u apart, with the first at the origin. In particular,

$$\begin{aligned}\breve{M}_2(B\mid u,\kappa_1) &= \frac{\breve{M}_2(\mathrm{d}u\times\mathrm{d}\kappa_1\times B)}{\mathrm{d}u\,F(\mathrm{d}\kappa_1)}\,,\\ &= \frac{\int_B \breve{m}_2(u,\kappa_1,\kappa_2)\,\mathrm{d}\kappa_2}{m_{\mathrm{g}}f(\kappa_1)} \qquad \text{if the densities exist,}\end{aligned} \tag{8.3.17}$$

representing the rate of occurrence of points with marks in B conditional on the occurrence of a point with mark κ_1 at a time origin u time units previously. It has the character of a *cross-intensity*. Further variants are set out in Lemma 8.3.III.

The results so far have been stated in terms of the ordinary rather than the factorial moment measures. When the ground process is simple (as we are assuming throughout this chapter), the only differences arise when $u = 0$, in which case the reduced form of the ordinary second-moment measure includes a double δ-function term $\delta(u)\,\delta[\rho(\kappa_1,\kappa_2)]$ (here, $\rho(\cdot)$ represents the distance function in the mark space), a term that is missing from the corresponding factorial moment density. Even if $u = 0$, the complete moment density $\breve{m}_2(0,\kappa_1,\kappa_2)$ can still exist (and is then zero) if $\kappa_1 \neq \kappa_2$.

For $u \neq 0$, the densities $\breve{m}_2(u,\kappa_1,\kappa_2)$ and the corresponding covariance densities $\breve{c}_2(u,\kappa_1,\kappa_2)$ (or normalized versions of them) are usually the main objects of investigation in a second-order analysis of a stationary marked or multivariate point process.

Example 8.3(d) *Stationary process with independent marks* (see Proposition 6.4.IV). Let the simple point process N on $\mathbb{R}$ have mean density m and suppose that marks are allocated independently according to the probability distribution $F(\cdot)$. Then, $F(\cdot)$ coincides with the stationary mark distribution $\Pi(\cdot)$ at (8.3.16) and with the *mark kernel* $F(\cdot\mid t)$ introduced in Proposition 6.4.IV (and here independent of t, from stationarity). For $u \neq 0$, the reduced moment measure $\breve{M}_2$ takes the form

$$\breve{M}_2(\mathrm{d}u\times\mathrm{d}\kappa_1\times\mathrm{d}\kappa_2) = \breve{M}^{\mathrm{g}}_{[2]}(\mathrm{d}u)\times F(\mathrm{d}\kappa_1)\times F(\mathrm{d}\kappa_2)\,,$$

and for the covariance measure,

$$\breve{C}_2(\mathrm{d}u\times\mathrm{d}\kappa_1\times\mathrm{d}\kappa_2) = \breve{C}^{\mathrm{g}}_2(\mathrm{d}u)\times F(\mathrm{d}\kappa_1)\times F(\mathrm{d}\kappa_2),$$

where $\breve{M}_2^{\mathrm{g}}$ and $\breve{C}_2^{\mathrm{g}}$ are the reduced moment and cumulant measures of the initial process N, which here acts as the ground process N_{g}.

Such a simple model may be useful as a null hypothesis in testing for more complex interactions, as, for example, in the discussion of earthquake magnitudes in Vere-Jones (1970). □

Another focus of practical interest is the bivariate distribution of the marks from two points at a given separation from each other. One is typically interested in how the properties of this distribution vary as a function of the distance between the two points. The existence of such distributions, while not a direct corollary of Proposition 8.3.II, does follow from it via a further application of the disintegration theory outlined in Appendix A1.5. We state the result for MPPs with state space $\mathcal{X} = \mathbb{R}$; note that the extensions to stationary (homogeneous) processes on $\mathcal{X} = \mathbb{R}^d$ are immediate (see also Exercise 8.3.7).

Lemma 8.3.III. *Let $N(\cdot)$ satisfy the conditions of Proposition* 8.3.II, *and suppose in addition that for its ground process the second-moment measure exists and has reduced form $\breve{M}_2^{\mathrm{g}}(\cdot)$. Then, there exists a bivariate mark kernel $\Pi_2(K_1 \times K_2 \mid u)$, where $K_1, K_2 \in \mathcal{B}(\mathcal{K})$, such that*

(i) *for $\breve{M}_2^{\mathrm{g}}$-almost-all u, $\Pi_2(\cdot \mid u)$ is a probability distribution on $\mathcal{K}^{(2)}$;*

(ii) *$\Pi_2(K_1 \times K_2 \mid u)$ is a Borel measurable function of u for fixed K_1, K_2;*

(iii) *$\breve{M}_2$ has the factorization*

$$\breve{M}_2(\mathrm{d}u \times \mathrm{d}\kappa_1 \times \mathrm{d}\kappa_2) = \breve{M}_2^{\mathrm{g}}(\mathrm{d}u)\, \Pi_2(\mathrm{d}\kappa_1 \times \mathrm{d}\kappa_2 \mid u),$$

or in integral form, for bounded Borel functions h on $\mathcal{X} \times \mathcal{K}^{(2)}$ with bounded support on $\mathcal{X}$,

$$\int_{\mathbb{R}\times\mathcal{K}^{(2)}} h(u, \kappa_1, \kappa_2)\, \breve{M}_{[2]}(\mathrm{d}u \times \mathrm{d}\kappa_1 \times \mathrm{d}\kappa_2) = \int_{\mathbb{R}} \breve{M}_2^{\mathrm{g}}(\mathrm{d}u) \int_{\mathcal{K}^{(2)}} h(u,\, \kappa_1, \kappa_2)\, \Pi_2(\mathrm{d}\kappa_1 \times \mathrm{d}\kappa_2 \mid u).$$

Proof. The proof is a straightforward application of the disintegration theorems A1.5.II and A1.5.III, starting from the observation that for fixed K_1 and K_2, the measure $\breve{M}_{[2]}(\mathrm{d}u \times K_1 \times K_2)$ is absolutely continuous with respect to the moment measure $\breve{M}_2^{\mathrm{g}}(\mathrm{d}u)$ of the ground process. □

A point to note here is that the univariate mark distributions arising as the marginals in the bivariate distribution above are not in general equal to the stationary mark distribution: the former stem from an analysis of second-order moments, while the latter comes from first-order moments. Nor is it necessarily the case that the bivariate distributions are symmetric. These points are illustrated in Exercise 8.3.8 and Example 8.3(e) below.

Assuming that the conditions of the lemma hold, various characteristics of the bivariate mark kernel $\Pi_2(\cdot \mid u)$ can be studied as functions of u. The

most important are the covariance and the correlation, which we may denote by $\mathrm{cov}^{\mathcal{K}}(u)$ and $\mathrm{corr}^{\mathcal{K}}(u)$, respectively. Exactly parallel concepts can be introduced for spatial processes, with the simplification, when the process is isotropic as well as homogeneous, that the functions depend only on the distance $|u|$.

EXAMPLE 8.3(e) *Marked cluster process with cluster-dependent marks.* We consider cluster processes in which both the cluster centre process and the cluster member processes carry marks and such that the mark, K say, for a given cluster centre controls both the spatial and the mark distributions of the cluster members. In the example that follows, we suppose for simplicity that all marks are nonnegative integers.

Take a Neyman–Scott type MPP in which the cluster centre process has realizations (x_i, K_i), say, where $\{x_i\}$ are the points of a Poisson process at rate λ_c and the marks K_i are i.i.d. with $\Pr\{K_i \geq k\} = s_k$ (all i). For a given cluster centre with mark K, say, let the number of cluster members N_m, say, have a negative binomial distribution with parameters $(\alpha, K/(1+K))$ so that the conditional mean and variance of the cluster size are αK and $\alpha K(K+1)$, respectively. Suppose also that the associated marks for the cluster members, given the parent mark K, are i.i.d. with discrete uniform distribution on the integers $1, \ldots, K$. Thus, the larger the parent mark K, the larger both the number of offspring and their marks. Assume that offspring points are distributed at i.i.d. distances from the parent with common distribution F with density f. The MPP we consider is the collection of all offspring points and associated marks.

Consider first the process of points having a given mark $k \geq 1$. Only clusters with parent mark $K \geq k$ can contribute to this process. Given N_m, the number of cluster members having mark k from such a cluster is found by binomial sampling, with probability of success $1/K$, from the N_m cluster members. The resulting number of cluster members with mark k again has a negative binomial distribution with parameters $(\alpha, \frac{1}{2})$, independent of k, provided $K \geq k$, and with mean α. Overall, the mean density of points with mark k is therefore $\lambda_c \alpha s_k$. For every positive k, the process of points with mark k is well defined. Moreover, the process as a whole is a well-defined point process on $\mathbb{R} \times \mathbb{Z}_+$. On the other hand, in order to be an MPP as defined in Section 6.4, the ground process (meaning the set of all offspring points) must be well defined (i.e. only finitely many points a.s. in bounded sets). Since the cluster centre process is Poisson, and clusters are i.i.d., a sufficient condition for the cluster process to be well defined is that the mean number of events per cluster is finite [see Exercise 6.3.5(a)]. Here the mean number of points per cluster for the ground process is given by $\mathrm{E}(K) = \sum_{k=1}^{\infty} s_k$, which is finite if and only if K has finite first moment. When this condition is satisfied, the stationary distribution of marks overall has the length-biased form $\pi_k = s_k/\mathrm{E}(K)$.

Consider next the process of pairs of points, with marks k_1, k_2, separated by distance $u > 0$. The second-order moment density has the form

$$\breve{m}_2(u;k_1,k_2) \;=\; \lambda^2\alpha^2 s_{k_1}s_{k_2} + \lambda \mathrm{E}\left(\frac{H_{k_1}(K)H_{k_2}(K)N_m(N_m-1)}{K^2}\right)\int_0^\infty \breve{f}(x)\breve{f}(x+u)\,\mathrm{d}x, \quad (8.3.18)$$

where $H_k(j) = 1$ if $j \geq k$, 0 otherwise, and the integral follows the notation of equation (6.3.19). The first term here represents the product of the means, while the second is the contribution to the second moment from pairs belonging to the same cluster. Note that $H_{k_1}(K)\,H_{k_2}(K) = H_{\max(k_1,k_2)}(K)$; taking expectations with respect to the parent cluster mark in the second term yields

$$\breve{m}_2(u;k_1,k_2) = \lambda^2\alpha^2 s_{k_1}s_{k_2} + \lambda\alpha(\alpha+1)s_{\max(k_1,k_2)}\phi(u), \quad (8.3.19)$$

where $\phi(u)$ denotes the integral in (8.3.18). This quantity exists for the marked process without any further restrictions, but the second-moment measure does not exist for the ground process unless the sum $\sum_{k_1}\sum_{k_2} s_{\max(k_1,k_2)} = \sum_k (2k+1)s_k$ converges, equivalent to the existence of a second moment for the parent mark distribution. When this condition is satisfied, the bivariate mark kernel at separation u, $\Pi_2(k_1,k_2 \mid u)$, can be found by renormalizing [i.e. by dividing (8.3.19) by the double sum just described]. Even if we sum out one variable, the marginal distribution of the other does not reduce to the stationary mark distribution because of the intervention of the second term. Expressions for the mark covariance and mark correlation at separation u can be found from the bivariate mark kernel: details are left to the reader.

The assumption of i.i.d. marks within a cluster implies that there is no dependence on the separation u except through the term $\phi(u)$. This implies in particular that the bivariate mark kernel is symmetric in u. It would, however, be quite natural in some modelling situations to incorporate an explicit dependence of the mark distribution on the distance from the cluster centre, in which case a further dependence on u would arise, causing the bivariate distribution to be asymmetric in general. □

MPPs can give rise to a diverse range of second-order characteristics (see e.g. Stoyan, 1984; Isham, 1985): the 'simple' case of a finite mark space in Proposition 8.3.I bears this out. Schlather (2001) gives a valuable survey.

From a theoretical viewpoint, some of the most interesting applications of stationary MPPs are to situations where the marks are not merely statistically dependent on the past evolution of the process but are direct functions of it. As an extreme case, the mark at time t can be taken as the whole past history of the point process up to time t. This idea lies behind one approach to the Palm theory of Chapter 13. The following elementary example gives some insight into this application.

Example 8.3(f) *Forward recurrence times.* Assume there is given a simple stationary point process on $\mathbb{R}$, and associate with any point t_i of the process the length $L_i = \inf\{u\colon N(t_i-u,t_i) \geq 1\}$ of the previous interval. Then, the MPP consisting of the pairs (t_i, L_i) is stationary. Assuming that N has a

finite mean density m, it follows from Proposition 8.3.II and (8.3.16) that a stationary probability distribution $\Pi_L(\cdot)$ exists for the interoccurrence times. The integral relation (8.3.14) then leads to important relations involving $\Pi_L(\cdot)$ as for example in the following deduction of the distribution of the stationary forward recurrence time random variable. The distance of the point nearest to the right of the origin, t_1' say, has this distribution, with $t_1' = \inf\{t_i : t_i > 0\}$. If i' is the index of this point, then $0 < t_1' = t_{i'} \le L_{i'}$. Take any bounded measurable function $g(\cdot)$ of bounded support and define $h(t,\kappa) = g(t)$ if $0 \le \tau \le \kappa$, $h(t,\kappa) = 0$ otherwise. The left-hand side of (8.3.14) equals

$$\begin{aligned}\int_{\mathbb{R}\times\mathbb{R}_+} h(t,\kappa)\, M_1(\mathrm{d}t \times \mathrm{d}\kappa) &= \mathrm{E}\left[\int_{\mathbb{R}\times\mathbb{R}_+} h(t,\kappa)\, N(\mathrm{d}t \times \mathrm{d}\kappa)\right] \\ &= \mathrm{E}\left[\sum_{i:t)i>0} h(t_i,\kappa_i)\right] = \mathrm{E}[g(t_1']\end{aligned}$$

since $h(t,\kappa) = 0$ for $t > t_1'$; evaluating the right-hand side as below gives

$$\mathrm{E}[g(t_1')] = m\int_0^\infty g(u)\,\mathrm{d}u \int_u^\infty \Pi_L(\mathrm{d}\kappa) = m\int_0^\infty [1 - F_L(u)]g(u)\,\mathrm{d}u,$$

where $F_L(t) = \int_0^t \Pi_L\,(\mathrm{d}u)$ is the distribution function for the interval length. Since g is an arbitrary measurable function of bounded support, we can for example choose $g(t) = I_{(0,x]}(t)$ and obtain $\Pr\{t_1' \le x\}$ on the left-hand side, equal to $m\int_0^x [1 - F_L(u)]\,\mathrm{d}u$ from the right-hand side; thus, the distribution for the point t_1' immediately following the origin (i.e. the distribution for the *forward recurrence time*) has the density

$$f_1(x) = m[1 - F_L(x)] = [1 - F_L(x)]/\mu_L\,,$$

where μ_L is the mean interval length [see (4.2.3) and Proposition 4.2.I]. This simple derivation of a Palm–Khinchin relation uses an argument similar to the original work of Palm (1943). □

Example 8.3(g) *Vehicles on a road.* We consider a spatially stationary distribution of cars along a long straight road, the car at x_i having a (constant) velocity v_i, with $v_i \ne v_j$ in general. Our aim is to determine the evolution in time, if any, of characteristics of the process.

The family of transformations that concerns us is given by

$$(x_i, v_i) \mapsto (x_i + tv_i, v_i) \qquad (\text{real } t).$$

Denote by m_t, $\Pi_t(\cdot)$, and $c_t(u, v_1, v_2)$ the mean density, the stationary (in space) velocity distribution, and the spatial covariance density at time t. We can refer moments at time t to moments at time 0 on account of the following

reasoning. From (8.3.14), we have for the space–velocity mean density at time t, $M_t(\mathrm{d}x \times \mathrm{d}v)$ say,

$$\int_{\mathbb{R}\times\mathbb{R}_+} h(x,v)\, M_t(\mathrm{d}x \times \mathrm{d}v) = \int_{\mathbb{R}\times\mathbb{R}_+} h(x+tv,v) m_0\, \mathrm{d}x\, \Pi_0(\mathrm{d}v)$$
$$= \int_{\mathbb{R}\times\mathbb{R}_+} h(y,v) m_0\, \mathrm{d}y\, \Pi_0(\mathrm{d}v),$$

so that the mean vehicle density and velocity distribution remain constant in time whatever their initial forms.

Applying a similar argument to the second-order integrals implies that if the covariance densities $c_t(u, v_1, v_2)$ exist for $t = 0$, they exist for all $t > 0$ and are given by

$$c_t(u, v_1, v_2) = c_0\big(u + t(v_2 - v_1), v_1, v_2\big).$$

The asymptotic covariance properties of $c_t(\cdot)$ at $t \to \infty$ thus depend on the behaviour of $c_0(u, v_1, v_2)$ for large u. In most practical cases, a mixing condition holds and implies that for all v_1, v_2, $c_0(u, v_1, v_2) \to 0$ as $|u| \to \infty$. Under these conditions, any correlation structure tends to die out, this being an illustration of the 'Poisson tendency' of vehicular traffic (Thedeen, 1964).

This example can also be treated as a line process and extended in various ways (see e.g. Bartlett, 1967; Solomon and Wang, 1972). □

Exercises and Complements to Section 8.3

8.3.1 Detail the argument that establishes Proposition 8.3.I by applying Proposition 8.1.I to the linear combinations $\sum a_i \xi_i(\cdot)$.

8.3.2 Let the matrix $\big(\breve{M}_{ij}(\cdot)\big)$ of nonnegative measures be positive-definite as in (8.3.1). Show that the matrix of Fourier transforms $\big(F_{ij}(\cdot)\big)$ consists of nonnegative measures with the same positive-definite property.

8.3.3 Consider a multivariate Neyman–Scott process in which cluster centres occur in time at rate μ_c and cluster members may be of different types with joint density $p(k,u) = \pi_k f_k(u)$, $\sum \pi_k = 1 = \int f_k(u)\mathrm{d}u$ $(k = 1, \ldots, K)$. Find expressions, generalizing those of Example 6.3(c), for the means and covariance densities of the different component streams and the corresponding multivariate Bartlett spectra.

8.3.4 Consider a cluster process in which the cluster centres form a simple stationary point process with mean density λ_c and Bartlett spectrum with density $\gamma_{11}(\cdot)$, while the clusters have the Hawkes branching structure of Example 8.3(c). Regard the resultant process as the output of a system with the cluster centre process the input and the generation of cluster members representing a type of positive feedback with the linear structure characteristic of a Hawkes process.

(a) Arguing from the general relations for the second-order properties of a cluster process, show that the output process here has the spectral density

$$\gamma_{22}(\omega) = \frac{[\lambda_c/(2\pi)]\big((1-\nu)^{-1} - 1\big) + \gamma_{11}(\omega)}{|1 - \widetilde{\mu}(\omega)|^2},$$

where $\nu = \widetilde{\mu}(0)$, which [see (8.3.11)] is a different generalization of (8.2.10). The only contributions to the cross-covariance terms are from the cluster centre to cluster members, leading to $c_{12}(u) = \lambda_c m_1(u \,|\, 0)$ (see the notation in Exercise 5.5.6), and thus

$$\gamma_{12}(\omega) = \frac{\lambda_c/(2\pi)}{\left(1 - \widetilde{\mu}(\omega)\right)^{-1}} = \gamma_{21}(-\omega).$$

(b) By specializing $\gamma_{11}(\cdot)$, more specific examples of input/output systems are obtained. For example, the input may be a Cox process directed by a continuous nonnegative process $X(\cdot)$, in which case we have a continuous input process $X(\cdot)$ causally affecting an output point process. If, moreover, $X(\cdot)$ is itself a shot-noise process generated by some primary point process, we recover a somewhat more general case of mutually exciting point processes.

8.3.5 Explicitly state the mappings and show their use in applying the factorization Lemma A2.7.II to prove Proposition 8.3.II.

8.3.6 *MPPs with infinite mean ground density.* Suppose given a countable infinity of stationary $(\mathbb{R} \times \mathcal{K})$-valued MPPs N_j, $j = 1, 2, \ldots$, defined on some common probability space and $\mathcal{K} \subseteq \mathbb{R}_+$. Suppose that N_j has finite mean density m_j and each point of N_j has the positive-valued mark κ_j, say, and there is a probability distribution $\{\pi_j\}$ with $\pi_j > 0$ for $j = 1, 2, \ldots$ such that $\sum_j \pi_j m_j = \infty$.

(a) Let the MPP N equal N_j with probability π_j for $j = 1, 2, \ldots$. Then N is nonergodic: $\lim_{T\to\infty} N((0,T] \times \mathcal{K})/T = \lim_{T\to\infty} N_j((0,T] \times \mathcal{K})/T = m_j$ with probability π_j. Since each N_j is well defined, so is N, and its mean ground density equals $\sum_j \pi_j m_j = \infty$. Denoting a realization of N by $\{(x_i, \kappa_i)\}$, consider the stationary random measure $\xi(A) = \sum_{x_i \in A} \kappa_i$. Show that $\xi(\cdot)$ is nonergodic unless $m_j \kappa_j$ is independent of j a.s., and that its mean density equals $\sum_j \pi_j m_j \kappa_j$, which can be finite or infinite.

(b) Now suppose that the N_j are mutually independent marked Poisson processes. (i) Show that the superposition of any specified finite collection of the N_j is an MPP with finite mean density. (ii) Let $\mathcal{J}$ be a countably infinite subset of $\{1, 2, \ldots\}$, and consider $N = \sum_{j\in\mathcal{J}} N_j$. Then, N is not an MPP because $N((0,1] \times \mathcal{K}) = \infty$ a.s., contradicting the finiteness condition in Definition 6.4.I(a).

(c) Suppose in (b) that the N_j are mutually independent simple stationary MPPs (not necessarily Poisson). Do the conclusions (i) and (ii) continue to hold?

8.3.7 Let the *bivariate simple Poisson process* model of Example 8.3(a) be stationary so that it can be described in terms of three rate functions μ_1, μ_2, μ_3 and a distribution function $G(\cdot)$ of the signed distance between a pair of related points, taking a type 1 point as the initial point. Show that in terms of these quantities,

$$m_1 = \mu_1 + \mu_3, \qquad m_2 = \mu_2 + u_3,$$
$$\breve{C}_{[2]}(\mathrm{d}u; 1, 2) = \mu_3\, G(\mathrm{d}u) = \breve{C}_{[2]}(-\mathrm{d}u; 2, 1).$$

Use the p.g.fl. or otherwise to show that when $\mathcal{X} = \mathbb{R}$, the joint distribution of the distances T_1 and T_2 from an arbitrary origin to the nearest points of types 1 and 2, respectively, is given by

$$\log \Pr\{T_1 > x, T_2 > y\}$$
$$= -2m_1 x - 2m_2 y + \mu_3 \int_{-x-y}^{x+y} \big(\min(x, y - v) - \max(-x, -y - v) \big)\, G(\mathrm{d}v),$$

while the joint distribution of the forward recurrence times T_1^+, T_2^+ from the origin to the nearest points in the positive direction is given by

$$\log \Pr\{T_1^+ > x, T_2^+ > y\}$$
$$= -m_1 x - m_2 y + \mu_3 \int_{-x}^{y} \big(\min(x, y - v) - \max(0, -v) \big)\, G(\mathrm{d}v).$$

Consider extensions to the case $\mathcal{X} = \mathbb{R}^d$.

8.3.8 *Gauss–Poisson process with asymmetric bivariate mark distribution.* In a marked process of correlated pairs (marked Gauss–Poisson process), suppose that the joint distribution of the marks corresponding to the two points in a pair depends on the separation of the two points and that the mark of the first occurring point in the pair is (say) always the larger. Construct an explicit example for which the bivariate mark distribution at separation u depends explicitly on u and is asymmetric.

8.3.9 *Bivariate forward recurrence time.* Extend the argument of Example 8.3(f) to the case of a bivariate point process by using an MPP in which the mark at a point t_i of the process is of the form $(j_i; L_{1i}, L_{2i})$, where j_i is the type of the point and L_{1i}, L_{2i} are the backward occurrence times to the last points of types 1 and 2, respectively. Obtain a bivariate extension of the Palm–Khinchin equations, and compare these with the extensions to nonorderly point processes discussed in (3.4.14). Hence or otherwise, obtain expressions for the joint distributions of the intervals between an arbitrary point of type i ($i = 1, 2$) and the next occurring points of types 1 and 2 in Example 8.3(a). [Daley and Milne (1975) use a different approach that exploits methods similar to those of Chapter 3].

8.4. Spectral Representation

We take up next the possibility of developing a Cramér-type spectral representation for stationary point processes and random measures. In $\mathbb{R}$, such a representation is essentially a corollary of the spectral representation for processes with stationary increments given by Doob (1949) and for stationary interval functions given by Brillinger (1972). No essentially new points arise, although minor refinements are possible as a result of the additional properties available for p.p.d. measures. We give a brief but essentially self-contained account of the representation theory for random measures in $\mathbb{R}^d$ following the general lines of the approach in Vere-Jones (1974). The relation to spectral representations for stationary generalized processes is discussed in Daley (1971) and Jowett and Vere-Jones (1972).

In order to be consistent with the representation theory for continuous-time processes, we work throughout with the mean-corrected process

$$\xi^0(\mathrm{d}x) = \xi(\mathrm{d}x) - m\,\mathrm{d}x \tag{8.4.1}$$

with zero mean, where ξ is a second-order stationary random measure with mean density m. Thus, we are concerned with properties of the Bartlett spectrum. An equivalent and perhaps slightly more direct theory could be built up from the properties of $\xi(\cdot)$ and the second-moment measure: the differences are outlined in Exercise 8.4.1.

The essence of the Cramér representation is an isomorphism between two Hilbert spaces, one of random variables defined on a probability space and the other of functions on the state space $\mathcal{X} = \mathbb{R}^d$. In the present context, we use the notation $L_2(\xi^0)$ to denote the Hilbert space of (equivalence classes of) random variables formed from linear combinations of the second-order random variables $\xi^0(A)$ (bounded $A \in \mathcal{B}_\mathcal{X}$) and their mean square limits, while $L_2(\Gamma)$ denotes the Hilbert space of (equivalence classes of) measurable functions square integrable with respect to Γ. Since Γ is not in general totally finite, we cannot apply directly the theory for mean square continuous processes. Rather, there are two possible routes to the required representations: we can exploit the results already available for continuous processes by means of smoothing techniques such as those used in Section 8.5, or we can develop the theory from first principles, using appropriate modifications of the classical proofs where necessary. We adopt the latter approach, although we only sketch the arguments where they directly mimic the standard theory.

A convenient starting point is the following lemma in which $\mathcal{S}$ again denotes the space of functions of rapid decay in $\mathbb{R}^d$.

Lemma 8.4.I. *Given any boundedly finite measure Γ in $\mathbb{R}^d$, the space $\mathcal{S}$ is dense in $L_2(\Gamma)$.*

Proof. The result is a minor modification of standard results [see e.g. Kingman and Taylor (1966, p.131) and Exercise 8.4.2]. □

The key step in establishing the isomorphism between the spaces $L_2(\xi^0)$ and $L_2(\Gamma)$ is a special case of Proposition 8.6.IV, which, with the notation

$$\zeta_f = \int_{\mathbb{R}^d} f(x)\,\xi^0(\mathrm{d}x), \tag{8.4.2}$$

where f is a bounded Borel function of bounded support, can be stated in the form

$$\begin{aligned}\|\tilde{f}\|_{L_2(\Gamma)} &= \int_{\mathbb{R}^d} |\tilde{f}(\omega)|^2\,\Gamma(\mathrm{d}\omega) = \int_{\mathbb{R}^d} \breve{C}_2(\mathrm{d}x) \int_{\mathbb{R}^d} f(x) f(x+u)\,\mathrm{d}u \\ &= \mathrm{var}(\zeta_f) = \|\zeta_f\|_{L_2(\xi^0)}\,.\end{aligned} \tag{8.4.3}$$

A first corollary of this equality of norms is the following counterpart of the lemma above.

Lemma 8.4.II. *For $\psi \in \mathcal{S}$, the random integrals $\zeta_\psi = \int_{\mathbb{R}^d} \psi(x)\, \xi^0(\mathrm{d}x)$ are dense in $L_2(\xi^0)$.*

Proof. It is enough to show that for any given bounded $A \in \mathcal{B}(\mathbb{R}^d)$, $\xi^0(A)$ can be approximated in mean square by elements ζ_{ψ_n} with $\psi_n \in \mathcal{S}$. Working from the Fourier transform side, it follows from (8.4.3) that $\tilde{I}_A \in L_2(\Gamma)$ and thus by Lemma 8.4.I that $\tilde{I}_A$ can be approximated by a sequence of functions in $\mathcal{S}$. Now $\mathcal{S}$ is invariant under the Fourier transform map, so this sequence can be written as $\tilde{\psi}_n$ with $\psi_n \in \mathcal{S}$. Applying (8.4.3) with $\psi = I_A \psi_n$ leads to

$$\|\tilde{I}_A - \tilde{\psi}_n\|_{L_2(\Gamma)} = \|\xi^0(A) - \zeta_{\psi_n}\|_{L_2(\xi^0)}\,.$$

By construction, the left-hand side $\to 0$ as $n \to \infty$, and hence also the right-hand side $\to 0$, which from our opening remark is all that is required. □

Lemmas 8.4.I and 8.4.II show that for $\psi \in \mathcal{S}$ there is a correspondence $\tilde{\psi} \leftrightarrow \zeta_\psi$ between elements $\tilde{\psi}$ of a set dense in $L_2(\Gamma)$ and elements ζ_ψ of a set dense in $L_2(\xi^0)$. The correspondence is one-to-one between equivalence classes of functions and is norm-preserving. From this last fact, it follows that the correspondence can be extended to an isometric isomorphism between the full Hilbert spaces $L_2(\Gamma)$ and $L_2(\xi^0)$ (see Exercise 8.4.3 for details), thus establishing the following proposition.

Proposition 8.4.III. *There is an isometric isomorphism between $L_2(\Gamma)$ and $L_2(\xi^0)$ in which, for $\psi \in \mathcal{S}$, the integral ζ_ψ in (8.4.2) $\in L_2(\xi^0)$ and the Fourier transform $\tilde{\psi} \in L_2(\Gamma)$ are corresponding elements.*

The main weakness of this proposition is that it does not give an explicit Fourier representation of the random measure and associated integrals ζ_ψ. To overcome this deficiency, we adopt the standard procedure of introducing a mean square integral with respect to a certain wide-sense random signed measure with uncorrelated values on disjoint sets.

For any bounded $A \in \mathcal{B}(\mathbb{R}^d)$, let $Z(A)$ denote the random element in $L_2(\xi^0)$ corresponding to $\tilde{\psi}(\omega) \equiv I_A(\omega)$ in $L_2(\Gamma)$. For disjoint sets A_1, A_2, it follows from the polarized form of (8.4.2) (obtained by expressing inner products in terms of norms) that

$$\mathrm{E}\big(Z(A_1)\,\overline{Z(A_2)}\,\big) = \int_{\mathbb{R}^d} I_{A_1}(\omega) I_{A_2}(\omega)\, \Gamma(\mathrm{d}\omega) = 0, \tag{8.4.4}$$

so that the $Z(\cdot)$ are indeed uncorrelated on disjoint sets (or, in the setting of the real line, have orthogonal increments). The definition of a mean square integral with respect to such a family is a standard procedure (see e.g. Doob, 1953; Cramér and Leadbetter, 1967) and leads to the conclusion that for every $g \in L_2(\Gamma)$ the integral

$$\int_{\mathbb{R}^d} g(\omega)\, Z(\mathrm{d}\omega)$$

can be defined uniquely as a mean square limit of integrals of simple functions and can be identified with the unique random variable associated with g in the isomorphism theorem described by Proposition 8.4.III. In particular, for $g = \tilde{\psi} \in \mathcal{S}$, the integral below can be identified with the random element ζ_ψ; that is,

$$\int_{\mathbb{R}^d} \tilde{\psi}(\omega)\, Z(\mathrm{d}\omega) = \int_{\mathbb{R}^d} \psi(x)\, \xi^0(\mathrm{d}x).$$

Also, referring to the convergence property displayed in the proof of Lemma 8.4.II (and this defines an equivalence relation as noted), the limit relation can be written as

$$\xi^0(A) = \operatorname*{l.i.m.}_{n\to\infty} \zeta_{\psi_n}$$

(see e.g. Doob, 1953, p. 8). More generally, it follows from Proposition 8.6.IV and (8.4.3) that the same conclusion holds for any bounded ψ of bounded support. Thus, we have the following result, which is a slight strengthening, as well as an extension to $\mathbb{R}^d$, of the corresponding result in Vere-Jones (1974).

Theorem 8.4.IV. *Let ξ be a second-order stationary random measure or point process in $\mathbb{R}^d$ with Bartlett spectrum Γ. Then, there exists a second-order wide-sense random measure $Z(\cdot)$ defined on bounded $A \in \mathcal{B}(\mathbb{R}^d)$ for which*

(i) $\mathrm{E}Z(A) = 0 = \mathrm{E}[Z(A)\overline{Z(B)}\,]$ *for bounded disjoint* $A, B \in \mathcal{B}(\mathbb{R}^d)$; (8.4.4′)

(ii) $\operatorname{var} Z(A) = \mathrm{E}(|Z(A)|^2) = \Gamma(A)$; (8.4.5)

(iii) *for all $g \in L_2(\Gamma)$, the random variable ζ corresponding to g in the isomorphism of Proposition* 8.4.III *is expressible as $\zeta = \int_{\mathbb{R}^d} g(\omega)\, Z(\mathrm{d}\omega)$; and*

(iv) *for all $\psi \in \mathcal{S}$ and all bounded measurable ψ of bounded support,*

$$\zeta_\psi \equiv \int_{\mathbb{R}^d} \psi(x)\, \xi^0(\mathrm{d}x) = \int_{\mathbb{R}^d} \tilde{\psi}(\omega)\, Z(\mathrm{d}\omega) \quad \text{a.s.} \tag{8.4.6}$$

Observe that in the Parseval relation in (8.4.6) the left-hand side represents the usual random integral defined on a realization by realization basis, whereas the right-hand side is a mean square integral that does not have a meaning in this sense. The two most important classes of functions ψ are covered by the theorem. In Exercise 8.4.4, we indicate how (8.4.6) can be extended to somewhat wider classes of functions and, in particular, (8.4.6) continues to hold whenever ψ is Lebesgue integrable and $\tilde{\psi} \in L_2(\Gamma)$.

An alternative approach to the substance of part (iv) of this theorem is simply to *define* the integral on the left-hand side of (8.4.6) to be equal to the right-hand side there for all $\tilde{\psi} \in L_2(\Gamma)$, but this begs the question as to when this definition coincides with the a.s. definition of the integral used until now.

More explicit representation theorems can be obtained as corollaries to (8.4.6). In particular, taking $\psi(x) = I_A(x)$, we have the following.

Corollary 8.4.V. *For all bounded* $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\xi^0(A) = \int_{\mathbb{R}^d} \tilde{I}_A(\omega)\, Z(\mathrm{d}\omega) \quad \text{a.s.} \tag{8.4.7}$$

We cannot immediately obtain an inversion theorem for $Z(\cdot)$ in this form because the corresponding integral $(2\pi)^{-d}\int_{\mathbb{R}^d} \tilde{I}_B(-x)\,\xi^0(\mathrm{d}x)$ need not exist. The finite integral over $\mathbb{U}_T^d$ presents no difficulties, however, and leads to the second corollary.

Corollary 8.4.VI. *For all bounded* $A \in \mathbb{R}^d$ *that are* Γ*-continuity sets,*

$$Z(A) = \operatorname*{l.i.m.}_{T\to\infty} \frac{1}{(2\pi)^d} \int_{\mathbb{U}_T^d} \tilde{I}_A(-x)\,\xi^0(\mathrm{d}x). \tag{8.4.8}$$

Proof. From the theorem, the finite integral in (8.4.8) can be transformed into the expression [for $\theta = (\theta_1, \ldots, \theta_d)$ and $\omega = (\omega_1, \ldots, \omega_d) \in \mathbb{R}^d$]

$$\int_{\mathbb{R}^d} Z(\mathrm{d}\omega) \int_A \left[\prod_{i=1}^d \left(\frac{\sin T(\omega_i - \theta_i)}{\omega_i - \theta_i} \right) \right] \mathrm{d}\theta.$$

Provided A is a continuity set for Γ, the integrand convolved with $I_A(\omega)$ converges in $L_2(\Gamma)$ to $I_A(\omega)$ as $T \to \infty$ (see Exercise 8.4.5: the proof is straightforward for intervals A but not so direct for general bounded A), and hence the integral converges in mean square to $Z(A)$. □

In very simple cases, Corollary 8.4.VI can be used to calculate directly the process $Z(\cdot)$ having orthogonal increments. Such an example is given below, partly to illustrate the potential dangers of using the second-order representation for anything other than second-order properties.

Example 8.4(a) *The Fourier transform of the Poisson process.* Let ξ be a Poisson process on $\mathbb{R}$ with constant rate λ. Then, it follows from (8.4.8) that

$$Z((a,b]) = \operatorname*{l.i.m.}_{T\to\infty} \frac{1}{2\pi i} \int_{-T}^{T} \frac{\mathrm{e}^{ixa} - \mathrm{e}^{ixb}}{x} \big(N(\mathrm{d}x) - \lambda\,\mathrm{d}x\big).$$

Consider in particular the process

$$U_a(\omega) \equiv Z(\omega + a) - Z(\omega - a) = \operatorname*{l.i.m.}_{T\to\infty} \frac{1}{\pi} \int_{-T}^{T} \frac{\mathrm{e}^{-i\omega x} \sin ax}{x} \big(N(\mathrm{d}x) - \lambda\,\mathrm{d}x\big).$$

Using standard results from Chapter 9 for the characteristic functional of the Poisson process, we find

$$\begin{aligned}
\Phi(\omega, s) &\equiv \mathrm{E}\exp\big(isU_a(\omega)\big) \\
&= \exp\left\{ \lambda \int_{-\infty}^{\infty} \Big[\exp\Big(\frac{is\mathrm{e}^{-i\omega x} \sin ax}{x} \Big) - 1 - \frac{is\mathrm{e}^{-i\omega x} \sin ax}{x} \Big]\, \mathrm{d}x \right\} \\
&= \exp\left\{ \lambda \int_{-\infty}^{\infty} \Big[-\tfrac{1}{2}\, s^2 \cos\omega x \Big(\frac{\sin ax}{x} \Big)^2 + O(s^3) \Big(\frac{\sin ax}{x} \Big)^3 \Big]\, \mathrm{d}x \right\} \\
&= \exp\big\{ -\tfrac{1}{2}\pi\lambda a s^2 + O(s^3) \big\}
\end{aligned}$$

uniformly in ω [see e.g. Copson (1935, p. 153) for evaluation of the integral]. It follows that the variance of $U_a(\omega)$ is proportional to the length of the interval and independent of its location, corresponding to the presumption that $Z(\cdot)$ in this case must be a process with orthogonal and second-order stationary increments.

On the other hand, $Z(\cdot)$ clearly does not have strictly stationary increments, for the full form of the characteristic function depends nontrivially on ω. Similarly, it can be checked from the joint characteristic function that Z does not have independent increments. Indeed, as follows from inspecting its characteristic function, $U_a(\omega)$ has an infinitely divisible distribution of pure jump type, with a subtle dependence of the jump distribution on a and ω that produces the requisite characteristics of the second-order properties. □

The spectral representation for stationary random measures and point processes plays a similar role in guiding intuition and aiding computation as it does for classical time series. We illustrate its use below by establishing basic procedures for estimating the Bartlett spectrum in two practically important cases: simple point processes and random (point process) sampling of a stationary continuous process. Further examples arise in Section 8.5, where we examine linear filters and prediction.

Example 8.4(b) *Finite Fourier transform and point process periodogram.* Estimates of the Bartlett spectrum provide a powerful means of checking for periodicity in point process data as well as for investigating other features reflected in the second-order properties. The basic tool for estimating the spectrum is the *point process periodogram*, defined much as in the continuous case through the *finite Fourier transform* of the realization of a point process on a finite time interval $(0, T)$, namely

$$J_T(\omega) = \int_0^T \mathrm{e}^{-i\omega t}\,[N(\mathrm{d}t) - m\,\mathrm{d}t] = \sum_{k=1}^{N(T)} \mathrm{e}^{-i\omega t_k} - m\frac{1-\mathrm{e}^{-i\omega T}}{i\omega}\,, \qquad (8.4.9)$$

in terms of which the periodogram is then defined as

$$I_T(\omega) = \frac{1}{2\pi T}\,|J_T(\omega)|^2 \qquad (\omega \in \mathbb{R}). \qquad (8.4.10)$$

Express $J_T(\omega)$ in the form of the left-hand side of (8.4.6) by setting $\psi(t) = \mathrm{e}^{-i\omega t} I_{(0,T)}(t)$, which is certainly bounded and of bounded support. Then, it follows from Proposition 8.4.IV(iv) that

$$J_T(\omega) = \int_{\mathbb{R}} \frac{\mathrm{e}^{iT(\omega'-\omega)}-1}{i(\omega'-\omega)}\,Z(\mathrm{d}\omega') \quad \text{a.s.}$$

The orthogonality properties of Z now imply that

$$\mathrm{E}[I_T(\omega)] = \frac{1}{2\pi T}\int_{\mathbb{R}} \left|\frac{\mathrm{e}^{iT(\omega'-\omega)}-1}{i(\omega'-\omega)}\right|^2 \Gamma(\mathrm{d}\omega') \qquad (8.4.11a)$$

$$= \frac{T}{2\pi}\int_{\mathbb{R}} \left[\frac{\sin\frac{1}{2}T(\omega'-\omega)}{\frac{1}{2}T(\omega'-\omega)}\right]^2 \Gamma(\mathrm{d}\omega'). \qquad (8.4.11b)$$

If $\Gamma(\cdot)$ has an atom at ω, then it follows from (8.4.11a) that $I_T(\omega) \sim T\,\Gamma(\{\omega\})$. On the other hand, if $\Gamma(\cdot)$ has a continuous density $\gamma(\omega')$ in a neighbourhood of ω, then it follows from (8.4.11b) that $\mathrm{E}[I_T(\omega)] \to \gamma(\omega)$. Thus, the periodogram is an asymptotically unbiased estimate of the spectral density wherever the density exists. The contrast between the two cases is the basis of tests for periodic effects, meaning here some periodic fluctuation in the rate of occurrence of events. Consistency is another story, however, and some degree of smoothing must be introduced to obtain consistent estimates of the spectral density.

The theory here parallels the standard theory except insofar as the observations are not Gaussian and some spectral mass is carried at arbitrarily large frequencies. The latter feature is a consequence of assuming that the points $\{t_k\}$ of the process are observed with complete precision, which is a fiction in any real context: in reality, only limited precision is possible, amounting to some smoothing or rounding of the observations, which then induces a tapering of the spectrum at very high frequencies. Nevertheless, the lack of any natural upper bound to the observed frequency range, even from a finite set of observations, causes difficulties in tackling questions such as the detection and estimation of an unknown periodicity modulating the occurrence times of the observed points. Indeed, the very definition of such a modulation, except for specific models such as the Poisson process (when it can appear as a periodic modulation of the intensity), is a matter of some difficulty. The crux of the matter for the spectral theory is that, whatever the form of modulation may be, it should induce a periodic variation in the reduced covariance measure. Vere-Jones and Ozaki (1982) discuss some of these issues in simple special contexts; the general problem of testing for unknown frequencies in point process models appears to lack any definitive treatment. Brillinger (1978, 1981) gives a systematic overview of the differences between ordinary time series and point process analogues. □

Example 8.4(c) *Random sampling of a random process.* A situation of some practical importance arises when a stationary continuous-time stochastic process $X(t)$ is sampled at the epochs $\{t_i\}$ of a stationary point process. The resultant process can be considered in two ways, either as a discrete-time process $Y_i = X(t_i)$ or as a random measure with jump increments

$$\xi(\mathrm{d}t) = X(t)\,N(\mathrm{d}t).$$

Neither operation is linear, but the second equation is just a multiplication of the two processes and leads to the more tractable results. Neither $N(\cdot)$ nor $\xi(\cdot)$ is a process with zero mean; to express the latter as a process with zero mean, suppose for simplicity that $X(\cdot)$ has zero mean, and then write

$$\xi(\mathrm{d}t) = X(t)\,N^0(\mathrm{d}t) + mX(t)\,\mathrm{d}t,$$

where $N^0(\mathrm{d}t) = N(\mathrm{d}t) - m\,\mathrm{d}t$ and $m = \mathrm{E}N(0,1]$ is the mean rate of the

sampling process. Proceeding formally leads to

$$\int_{\mathbb{R}} \phi(t)\,\xi(\mathrm{d}t) = \int_{\mathbb{R}}\int_{\mathbb{R}} \tilde{\phi}(u-v)\,Z_X(\mathrm{d}u)\,N(\mathrm{d}v) + m\int_{\mathbb{R}} \tilde{\phi}(u)\,Z_X(\mathrm{d}u),$$

corresponding to a representation of the measure Z_ξ as a convolution of Z_X and Z_N with an additional term for the mean. Leaving aside the general case, suppose that the processes $X(\cdot)$ and $N(\cdot)$ are independent. Then we find

$$\begin{aligned}\operatorname{var}&\left(\int_{\mathbb{R}} \phi(t)\,\xi(\mathrm{d}t)\right)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} |\tilde{\phi}(u-v)|^2\,\gamma_X(\mathrm{d}u)\,\gamma_N(\mathrm{d}v) + m^2\int_{\mathbb{R}} |\tilde{\phi}(u)|^2\,\gamma_N(\mathrm{d}u),\end{aligned}$$

from which we deduce that

$$\gamma_\xi(\mathrm{d}\omega) = \int_{\mathbb{R}} \gamma_X(\mathrm{d}\omega - u)\,\gamma_N(\mathrm{d}u) + m^2\,\gamma_X(\mathrm{d}\omega).$$

Hence, for the covariance measures we have

$$\breve{C}_\xi(\mathrm{d}u) = \breve{c}_X(u)\big(m^2\,\mathrm{d}u + \breve{C}_N(\mathrm{d}u)\big) = \breve{c}_X(u)\,\breve{M}_2(\mathrm{d}u).$$

Of course, the last result can easily be derived directly by considering

$$\mathrm{E}\big(X(t)\,N(t,t+\mathrm{d}t]\,X(t+u)\,N(t+u,t+u+\mathrm{d}u]\big).$$

In practice, one generally must estimate the spectrum $\gamma_X(\cdot)$ given a (finite portion of a) realization of $\xi(\cdot)$. When N is a Poisson process at rate m,

$$\gamma_\xi(\mathrm{d}\omega) = (m/2\pi)(\operatorname{var} X)\,\mathrm{d}\omega + m^2\gamma_X(\mathrm{d}\omega),$$

so γ_X can be obtained quite easily from γ_ξ. In general, however, a deconvolution procedure may be needed, and the problem is complicated further by the fact that the spectral measures concerned are not totally finite. Consequently, numerical Fourier transform routines cannot be applied without some further manipulations [see Brillinger (1972) for further details]. □

Only partial results are available for the extension of the spectral theory to random signed measures. One approach, which we outline briefly below, follows Thornett (1979) in defining a second-order random measure as a family of random variables $\{W(A)\}$, indexed by the Borel sets, whose first and second moments satisfy the same additivity and continuity requirements as the first- and second-moment measures of a stationary random measure. The resulting theory may be regarded as a natural generalization to $\mathbb{R}^d$ of the theory of random interval functions developed by Bochner (1955) and extended and applied to a statistical context by Brillinger (1972).

Definition 8.4.VII. A wide-sense second-order stationary random measure *on* $\mathcal{X} = \mathbb{R}^d$ *is a jointly distributed family of real- or complex-valued random variables* $\{\xi(A): A \in \mathcal{B}_{\mathcal{X}}\}$ *satisfying the conditions, for bounded* A, $\{A_n\}$ *and* $B \in \mathcal{B}_{\mathcal{X}}$,

(i) $\mathrm{E}\xi(A) = m\ell(A)$, $\operatorname{var}\xi(A) < \infty$;

(ii) $\operatorname{var}((S_x\xi)(A)) = \operatorname{var}\xi(T_xA) = \operatorname{var}\xi(A)$;

(iii) $\xi(A \cup B) = \xi(A) + \xi(B)$ a.s. *for disjoint* A, B; *and*

(iv) $\xi(A_n) \to 0$ *in mean square when* $A_n \downarrow \emptyset$ *as* $n \to \infty$.

If the random variables $\xi(\cdot)$ here are nonnegative, then (iii) reduces to the first part of (6.1.2) and implies that in (iv) the random variables $\xi(A_n)$ decrease monotonically a.s.; that is, $\xi(A_{n+1}) \le \xi(A_n)$ a.s., so that (iv) can be strengthened to $\xi(A_n) \to 0$ a.s. when $A_n \downarrow \emptyset$ as $n \to \infty$ [see the second part of (6.1.2)]. We then know from Chapter 9 that there exists a strict-sense random measure that can be taken as a version of $\xi(\cdot)$ so that nothing new is obtained. Thus, the essence of the extension in Definition 8.4.VII is to random signed measures.

For the sequel, we work only with the mean corrected version, taking $m = 0$ in the definition. Given such a family then, we can always find a Gaussian family with the same first- and second-moment properties: the construction is standard and needs no detailed explanation (see Doob, 1953; Thornett, 1979). For example, the Poisson process, corrected to have zero mean, has $\operatorname{var}\xi(A) = \lambda\ell(A)$, where λ is the intensity; this function is the same as the variance function for the Wiener chaos process in Chapter 9.

While the definition refers only to variances, covariances are defined by implication from the relation, valid for real-valued $\xi(\cdot)$,

$$2\operatorname{cov}\big(\xi(A), \xi(B)\big) = \operatorname{var}\xi(A \cup B) + \operatorname{var}\xi(A \cap B) - \operatorname{var}\xi(A \setminus B) - \operatorname{var}\xi(B \setminus A),$$

which is readily verified first for disjoint A and B and then for general A and B by substituting in the expansion of

$$\operatorname{cov}\big(\xi(A), \xi(B)\big) = \operatorname{cov}\big(\xi(A \cap B) + \xi(A \backslash B),\ \xi(A \cap B) + \xi(B \backslash A)\big).$$

Although we can obtain in this way a covariance function $C(A \times B)$ defined on products of bounded $A, B \in \mathcal{B}_{\mathcal{X}}$, it is not obvious that it can be extended to a signed measure on $\mathcal{B}(\mathcal{X}^{(2)})$. Consequently, it is not clear whether or not a covariance measure exists for such a family. When it does, the further theory can be developed much as earlier. Irrespective of such existence, it is still possible to define both a spectrum for the process and an associated spectral representation. Thus, for any bounded Borel set A, consider the process

$$X_A(x) \equiv \xi(T_xA).$$

Mean square continuity follows from condition (iv), so $X_A(\cdot)$ has a spectral measure $\Gamma_A(\cdot)$, and we can define

$$\Gamma(d\omega) = |\tilde{I}_A(\omega)|^{-2}\,\Gamma_A(\mathrm{d}\omega)$$

for all ω such that $|\tilde{I}_A(\omega)| \neq 0$. Since we cannot ensure that $|\tilde{I}_A(\omega)| \neq 0$ for all ω, some care is needed in showing that the resultant measure $\Gamma(\cdot)$ can in fact be consistently defined for a sufficiently rich class of sets A [one approach is outlined by Thornett (1979) and given as Exercise 8.4.6]. Just as before, the measure Γ is translation-bounded and hence integrates $(1 + \omega^2)^{-1}$, for example. On the other hand, it is not positive-definite in general and not all the explicit inversion theorems can be carried over. Nevertheless, for all bounded $A \in \mathcal{B}_{\mathcal{X}}$, we certainly have

$$\operatorname{var} \xi(A) = \int |\tilde{I}_A(\omega)|^2 \, \Gamma(\mathrm{d}\omega) \tag{8.4.12}$$

and its covariance extension

$$\operatorname{cov}\big(\xi(A), \xi(B)\big) = \int \tilde{I}_A(\omega) \, \overline{\tilde{I}_B(\omega)} \, \Gamma(\mathrm{d}\omega). \tag{8.4.13}$$

Since the indicator functions are dense in $L_2(\Gamma)$, more general integrals of the form $\int \phi(x) \, \xi(\mathrm{d}x)$ can be defined as mean square limits of linear combinations of the random variables $\xi(A)$, at least when $\tilde{\phi} \in L_2(\Gamma)$. For such integrals, the more general formulae

$$\operatorname{var}\left(\int \phi(x) \, \xi(\mathrm{d}x) \right) = \int |\tilde{\phi}(\omega)|^2 \, \Gamma(\mathrm{d}\omega)$$

and

$$\operatorname{cov}\left(\int \phi(x) \, \xi(\mathrm{d}x), \int \psi(x) \, \xi(\mathrm{d}x) \right) = \int \tilde{\phi}(\omega) \tilde{\psi}(\omega) \, \Gamma(\mathrm{d}\omega)$$

are available, but it is not clear whether the integrals make sense other than in this mean square sense. As noted earlier, it is also an open question as to whether Γ is necessarily the Fourier transform of some measure, which we could then interpret as a reduced covariance measure.

The isomorphism result in Proposition 8.4.III can be extended to this wider context with only minor changes in the argument: it asserts the isomorphism between $L_2(X)$ and $L_2(\Gamma)$ and provides a spectral representation, for bounded $A \in \mathcal{B}_{\mathcal{X}}$,

$$\xi(A) = \int \tilde{I}_A(\omega) \, Z(\mathrm{d}\omega) \qquad \text{a.s.} \tag{8.4.14}$$

just as in the previous discussion.

To summarize, we have the following theorem of which further details of proof are left to the reader.

Theorem 8.4.VIII. *Let $\{\xi(\cdot)\}$ be a wide-sense second-order stationary random measure as in Definition* 8.4.VII. *Then, there exists a spectral measure $\Gamma(\cdot)$ and a process $Z(\cdot)$ of orthogonal increments with* $\operatorname{var} Z(\mathrm{d}\omega) = \Gamma(\mathrm{d}\omega)$ *such that* (8.4.12–14) *hold.*

Exercises and Complements to Section 8.4

8.4.1 *Representation in terms of the second-moment measure.* Show that the effect of working with the Fourier transform of the second moment rather than the Bartlett spectrum would be to set up an isomorphism between the spaces $L_2(\xi)$ generated by all linear combinations of the r.v.s $\xi(A)$ and $L_2(\nu)$, where ν is the inverse Fourier transform of $\breve{M}_2$. Show that the representation

$$\int_{\mathbb{R}^d} \phi(x)\, \xi(\mathrm{d}x) = \int_{\mathbb{R}^d} \tilde{\phi}(\omega)\, Z_1(\mathrm{d}\omega)$$

holds for functions ϕ in a suitably restricted class, where $Z_1(A) = m\delta_0(A) + Z(A)$, and Z and Z_1 differ only by an atom at $\omega = 0$.

8.4.2 Let Γ be a nontrivial boundedly finite measure. Show the following:

(a) Simple functions of the form $\sum_k a_k I_{A_k}$ [bounded $A_k \in \mathcal{B}(\mathbb{R}^d)$] are dense in $L_2(\Gamma)$.

(b) For bounded $A \in \mathcal{B}(\mathbb{R}^d)$, there exist open sets $U_n \in \mathcal{B}(\mathbb{R}^d)$ with $U_n \supseteq A$, $\Gamma(U_n) \downarrow \Gamma(A)$.

(c) Any such U_n is the countable union of hyper-rectangles of the form $\{\alpha_i < x_i \le \beta_i,\ i = 1, \ldots, d\}$.

(d) Indicator functions on such hyper-rectangles can be approximated by sequences of infinitely differentiable functions of bounded support.

Now complete the proof of Lemma 8.4.I.

8.4.3 Given $\tilde{\psi} \in L_2(\Gamma)$, choose $\psi_n \in \mathcal{S}$ such that $\|\tilde{\psi} - \psi\|_{L_2(\Gamma)} \to 0$ $(n \to \infty)$, and deduce that $\{Z_{\psi_n}\}$ is a Cauchy sequence in $L_2(\xi^0)$. Show that there is a unique r.v. $\zeta \in L_2(\xi^0)$ such that $Z_{\psi_n} \to \zeta$ in mean square. Interchange the roles of $L_2(\Gamma)$ and $L_2(\xi^0)$ and deduce the assertion of Proposition 8.4.III.

8.4.4 Show that (8.4.6) can be extended to all L_1 functions ϕ such that $\tilde{\phi} \in L_2(\Gamma)$. [*Hint*: The left- and right-hand sides can be represented, respectively, as an a.s. limit of integrals of bounded functions of bounded support and as a mean square limit. When both limits exist, they must be equal a.s. This argument establishes a conjecture in Vere-Jones (1974).]

8.4.5 Establish the following properties of the function $h_T(\omega) = \omega^{-1} \sin \omega T$ (they are needed in a proof of Corollary 8.4.IV).

(a) $\int_{-\infty}^{\infty} h_T(\omega)\, \mathrm{d}\omega = \pi$.

(b) For any continuous function ϕ with bounded support, the function

$$\phi_T(\omega) \equiv \int_{-\infty}^{\infty} \phi(\omega - u) h_T(u)\, \mathrm{d}u \to \phi(\omega) \quad \text{pointwise as } T \to \infty$$

[this is an application of Fourier's single integral (see Zygmund, 1968, Section 16.1)]. Show that the result still holds if only $\phi \in L_1(\xi)$ and ϕ is of bounded variation in any closed interval contained in its support.

(c) $\phi_T(\omega) \to \phi(\omega)$ in $L_2(\Gamma)$ for any p.p.d. measure (or for any Bartlett spectrum) Γ. [*Hint*: $|\phi_T(\omega)| \le$ constant$/|\omega|$ for large $|\omega|$ while $\sup_\omega |\phi_T(\omega)| < \infty$; these properties are enough to ensure that $|\phi_T(\omega)|^2 \le g(\omega)$ for some Γ-integrable function g.]

(d) Interpret the convergence in (c) as

$$\begin{aligned}\int_{\mathbb{R}} |\phi_T(\omega)|^2\,\Gamma(\mathrm{d}\omega) &= \int_{\mathbb{R}}\int_{\mathbb{R}} h_T(\omega-u)\phi(u)\,\mathrm{d}u \int_{\mathbb{R}} \overline{h_T(\omega-v)\phi(v)}\,\mathrm{d}v\,\Gamma(\mathrm{d}\omega)\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \phi(u)\,\overline{\phi(v)}\,\mathrm{d}u\,\mathrm{d}v \int_{\mathbb{R}} h_T(\omega-u)h_T(\omega-v)\,\Gamma(\mathrm{d}\omega)\\ &= \int_{\mathbb{R}^2} \phi(u)\,\overline{\phi(v)}\,\Gamma_T^*(\mathrm{d}u\times\mathrm{d}v)\\ &\to \int_{\mathbb{R}^2} \phi(u)\,\overline{\phi(v)}\,\Gamma^*(\mathrm{d}u\times\mathrm{d}v)\\ &= \int_{\mathbb{R}} |\phi(\omega)|^2\,\Gamma(\mathrm{d}\omega),\end{aligned}$$

where $\Gamma_T^*(\mathrm{d}u\times\mathrm{d}v)$ and Γ^* are measures in $\mathcal{B}(\mathbb{R}^2)$, the former with density $\int_{\mathbb{R}} h_T(\omega-u)\,\overline{h_T(\omega-v)}\,\Gamma(\mathrm{d}\omega)$, while the latter reduces to Γ along the diagonal $u=v$.

These results are enough to establish that $\Gamma_T^* \to \Gamma^*$ vaguely in $\mathbb{R}^2$ and hence that a similar result holds when $\phi(\cdot)$ is replaced by the indicator function of a bounded Borel set in $\mathbb{R}^1$ that is a continuity set for Γ.

8.4.6 Show that for Γ to be the spectral measure of a wide-sense second-order stationary random measure, it is necessary and sufficient that Γ integrate all functions $|\tilde{I}_A(\omega)|^2$ for bounded Borel sets A. Deduce that any translation-bounded measure can be a spectral measure. [*Hint*: Use a Gaussian construction for the sufficiency; then use Lin's lemma. See also Thornett (1979).]

8.4.7 (a) Show that if a wide-sense second-order stationary process has a reduced covariance measure $\breve{C}(\cdot)$, then $\breve{C}(\{0\}) = \lim_{T\to\infty} \Gamma\big((-T,T]\big)/(2T)$ continues to hold (see Theorem 8.6.III).

(b) Use Exercise 8.2.4 to show that not all spectral measures are transforms; that is, not all wide-sense processes have an associated reduced covariance measure (see also Exercise 8.6.3).

8.5. Linear Filters and Prediction

One of the most important uses of spectral representation theory is to obtain spectral characteristics of processes acted on by a linear filter, meaning here any time-invariant linear combination of values of the process or any mean square limit of such combinations. This use carries over formally unchanged from mean square continuous processes to second-order point processes and random measures and includes the procedures for developing optimal linear predictors for future values of the process. Obtaining the precise conditions for these extensions and their character requires some care, however, and forms the main content of the present section.

Let $\xi(\cdot)$ be a second-order stationary random measure and $\psi \in L_1$ a smoothing function; consider the smoothed process defined by

$$X(t) = \int_{-\infty}^{\infty} \psi(t-u)\,\xi(\mathrm{d}u). \tag{8.5.1}$$

Substituting from the Parseval relation (8.4.6) and recalling that the Fourier transform of the shifted function $\psi(t-u)$ is $\tilde{\psi}(-\omega)\mathrm{e}^{i\omega t}$, we find

$$X(t) = \int_{-\infty}^{\infty} \mathrm{e}^{i\omega t}\tilde{\psi}(-\omega)\,Z(\mathrm{d}\omega). \tag{8.5.2}$$

The spectrum $\Gamma_X(\cdot)$ of the transformed process is

$$\Gamma_X(\mathrm{d}\omega) = |\tilde{\psi}(-\omega)|^2\,\Gamma(\mathrm{d}\omega). \tag{8.5.3}$$

This will be totally finite, which implies that $X(\cdot)$ is a mean square continuous process, provided $\tilde{\psi} \in L_2(\Gamma)$.

The relation (8.5.1) can be interpreted even more broadly; for example, if $A(\cdot)$ is a totally finite measure, the convolution $A * \xi$ still defines a.s. a random measure and (8.5.2) and (8.5.3) continue to hold. Thus, (8.5.1) continues to make sense, with a generalized function interpretation of ψ, provided the outcome defines a.s. a random measure. However, the situation becomes decidedly more complex when, as is often necessary in applications to prediction, signed measures intervene; then at best the wide-sense theory can be used, and the character of the filtered process, in a realization-by-realization sense, has to be ascertained *post hoc*.

Example 8.5(a) *Binning.* A special case of practical importance arises when $\mathcal{X} = \mathbb{R}$ and the measure ξ is 'binned'; that is, integrated over intervals of constant length Δ, say. Considering first the continuous-time process $X(t) \equiv \xi\big(t - \frac{1}{2}\Delta, t + \frac{1}{2}\Delta]$, (8.5.2) yields

$$X(t) = \int_{-\infty}^{\infty} \mathrm{e}^{i\omega t}\,\frac{\sin\frac{1}{2}\omega\Delta}{\frac{1}{2}\omega}\,Z(\mathrm{d}\omega), \qquad \text{hence} \qquad \Gamma_X(\mathrm{d}\omega) = \left(\frac{\sin\frac{1}{2}\omega\Delta}{\frac{1}{2}\omega}\right)^2 \Gamma(\mathrm{d}\omega).$$

It is commonly the case that the binned process is sampled only at the lattice points $\{n\Delta\colon n = 0, \pm1, \ldots\}$. The sampled process can then be represented in the aliased form

$$Y(n) \equiv X(n\Delta) = \int_0^{s\pi/\Delta} \mathrm{e}^{in\theta} \sum_{k=-\infty}^{\infty} Z_X\Big(\frac{2k\pi}{\Delta} + \mathrm{d}\theta\Big).$$

Taking Δ as the unit of time, we see from this representation that the discrete-time process $\{Y(n)\}$ has spectral measure $G_Y(\cdot)$ on $(0, 2\pi]$ given by

$$G_Y(\mathrm{d}\theta) = \sin^2\theta \sum_{k=-\infty}^{\infty} \frac{\Gamma(2k\pi + \mathrm{d}\theta)}{(\theta + 2k\pi)^2}\,. \tag{8.5.4}$$

In the simplest case of a Poisson process, $\Gamma(\mathrm{d}\omega) = [\mu/(2\pi)]\,\mathrm{d}\omega$, so that

$$G_Y(\mathrm{d}\theta) = \sin^2\theta \sum_{k=-\infty}^{\infty} \frac{[\mu/(2\pi)]\,\mathrm{d}\theta}{(\theta + 2k\pi)^2} = \frac{\mu}{2\pi}\,\mathrm{d}\theta$$

since the infinite series is just an expansion of $\operatorname{cosec}^2\theta$. This reduction reflects the fact that the random variables $Y(n)$ are then independent with common variance μ.

Binning is widely used in practical applications of time series methods to point process data, and even where it is not explicitly invoked, it is present implicitly in the rounding of observations to a fixed number of decimal places. Indeed, the point process results themselves can be regarded as the limit when the binsize approaches zero and the character of the process $Y(n)$ approaches that of a sequence of δ-functions in continuous time. See e.g. Vere-Jones and Davies (1966) and Vere-Jones (1970), where these ideas are applied in the earthquake context. □

Perhaps the most important examples of linear filtering come in the form of linear predictions of a time series or point process. By a *linear predictor* we mean a predictor of the form $\int_{-\infty}^{t} f(t-u)\,\xi(\mathrm{d}u)$; that is, a linear functional of the past, with the quantity to be predicted a linear functional of the future. In the point process case, the problem commonly reduces to predicting, as a linear functional of the past, the mean intensity at some time point in the future. When the process has a mean square continuous density, this corresponds exactly to the classical problem of predicting a future value of the process as a linear functional of its past. Thus, our task is essentially to check when the classical procedures can be carried over to random measures and to write out the forms that they take in random measure terms.

It is important to contrast the linear predictors obtained in this way with the conditional intensity functions we described in Chapter 7. The conditional intensity function comprises the best nonlinear predictor of the mean rate at a point just ahead of the present. It is best out of all possible functionals of the past, linear or nonlinear, subject only to the measurability and non-anticipating characteristics described in Chapter 7. The linear predictors are best out of the more restricted class of linear functionals of the past. They are difficult to use effectively in predicting nonlinear features such as a maximum or the time to the next event in a point process. On the other hand, they perform well enough in predicting large-scale features where the law of large numbers tilts the distributions toward normality. They are generally easy to combine and manipulate and can sometimes be obtained when the full conditional intensity is inaccessible.

The Wold decomposition theorem plays an important role in finding the best linear predictor for mean square continuous processes, and we start with an extension of this theorem for random measures. As in Section 8.4, we use ξ and ξ^0 to denote a second-order stationary random measure and its zero mean

form, respectively, with the additional understanding that $\mathcal{X} = \mathbb{R}$. Since the results to be developed depend only on the spectral representation theorems, ξ can be either a strict- or wide-sense random measure. We continue to use $L_2(\xi^0)$ to denote the Hilbert space of equivalence classes of random variables formed from linear combinations of $\xi^0(A)$ for bounded $A \in \mathcal{B}$ and their mean square limits. Similarly, $L_2(\xi^0; t)$ denotes the Hilbert space formed from $\xi^0(A)$ with the further constraint that $A \subset (-\infty, t]$.

Definition 8.5.I. *The second-order strict- or wide-sense stationary random measure ξ is* deterministic *if $\bigcap_{t\in\mathbb{R}} L_2(\xi^0; t) = L_2(\xi^0)$ and* purely nondeterministic *if $\bigcap_{t\in\mathbb{R}} L_2(\xi^0; t) = \{0\}$.*

The following extension of Wold's theorem holds (Vere-Jones, 1974).

Theorem 8.5.II. *For any second-order stationary random measure ξ, the zero mean process ξ^0 can be written uniquely in the form*

$$\xi^0 = \xi_1^0 + \xi_2^0,$$

where ξ_1^0 and ξ_2^0 are mutually orthogonal, stationary, wide-sense zero-mean random measures, and ξ_1^0 is deterministic and ξ_2^0 purely nondeterministic.

Proof. Again we start from the known theorems for mean square continuous processes [see e.g. Cramér and Leadbetter (1967), especially Chapters 5–7] and use smoothing arguments similar to those around (8.5.1) to extend them to the random measure context. To this end, set

$$X(t) = \int_{-\infty}^{t} \mathrm{e}^{-(t-u)}\, \xi^0(\mathrm{d}u), \tag{8.5.5}$$

where the integral can be understood, whether ξ^0 is a strict- or wide-sense random measure, as a mean square limit of linear combinations of indicator functions. These indicator functions can all be taken of sets $\subseteq (-\infty, t]$, so we have $X(t) \in L_2(\xi^0; t)$, and more generally, $X(s) \in L_2(\xi^0; t)$ for any $s \le t$, so $L_2(X; t) \subseteq L_2(\xi^0; t)$. To show that we have equality here, we write

$$\begin{aligned} X(t+h) - \mathrm{e}^{-h}X(t) - \xi^0(t, t+h] &= \int_t^{t+h} [\mathrm{e}^{-(t+h-u)} - 1]\, \xi^0(\mathrm{d}u), \\ &= \int_{-\infty}^{\infty} \mathrm{e}^{i\omega t} \left[\frac{\mathrm{e}^{i\omega h} - \mathrm{e}^{-h}}{1 + i\omega} - \frac{\mathrm{e}^{i\omega h} - 1}{i\omega} \right] Z(\mathrm{d}\omega), \end{aligned}$$

where Z is the process of orthogonal increments associated with ξ^0 as in Theorem 8.4.IV. Subdividing any finite interval $(a, a + \Delta]$ into n subintervals of length $h = \Delta/n$, we obtain

$$\begin{aligned} &\sum_{k=1}^{n} \big[X(a + kh) - \mathrm{e}^{-h} X\big(a + (k+1)h\big)\big] - \xi^0(a, a + \Delta] \\ &\qquad = \int_{-\infty}^{\infty} \left(\sum_{k=1}^{n} \mathrm{e}^{i\omega(a+kh)} \right) \left(\frac{\mathrm{e}^{i\omega h} - \mathrm{e}^{-h}}{1 + i\omega} - \frac{\mathrm{e}^{i\omega h} - 1}{i\omega} \right) Z(d\omega). \end{aligned}$$

The variance of the left-hand side therefore equals

$$\int_{-\infty}^{\infty} \left(\frac{\sin \frac{1}{2}\omega\Delta}{\sin \frac{1}{2}\omega h} \right)^2 \left| 1 - \mathrm{e}^{-h} - \frac{\mathrm{e}^{i\omega} - 1}{i\omega} \right|^2 \frac{\Gamma(\mathrm{d}\omega)}{1+\omega^2} .$$

The measure $(1+\omega^2)^{-1}\,\Gamma(\mathrm{d}\omega)$ is totally finite (see Exercise 8.6.5), the term $|\cdot|^2$ is uniformly bounded in ω by $4h^2$ and for fixed ω it is $o(h^2)$ as $h \to 0$, and the term in braces is bounded by $(\Delta/h)^2$ and for fixed ω equals const. $\times$ $h^{-2}(1+o(1))$ as $h \to 0$. The dominated convergence theorem can therefore be applied to conclude that this variance $\to 0$ as $h \to 0$ and hence that $\xi^0(a,b]$ can be approximated in mean square by linear combinations of $\{X(t)\colon t \le b\}$. This shows that $L_2(\xi^0;t) \subseteq L_2(X;t)$, and thus $L_2(\xi^0;t) = L_2(X;t)$ must hold.

The Wold decomposition for $X(t)$ takes the form

$$X(t) = X_1(t) + X_2(t),$$

where $X_1(\cdot)$ is deterministic and $X_2(\cdot)$ purely nondeterministic. The decomposition reflects an orthogonal decomposition of $L_2(X)$, and hence of $L_2(\xi^0)$ also, into two orthogonal subspaces such that $X_1(t)$ is the projection of $X(t)$ onto one and $X_2(t)$ the projection onto the other. Then $\xi_1^0(A)$ and $\xi_2^0(A)$ may be defined as the projections of $\xi^0(A)$ onto these same subspaces. Furthermore, $\xi_1^0(a,b]$ and $\xi_2^0(a,b]$ can be expressed as mean square limits of linear combinations of $X_1(t)$ and $X_2(t)$ in exactly the same way as $\xi^0(a,b]$ is expressed above in terms of $X(t)$: the deterministic and purely nondeterministic properties of $X_1(\cdot)$ and $X_2(\cdot)$, respectively, carry over to $\xi_1^0(\cdot)$ and $\xi_2^0(\cdot)$. Uniqueness is a consequence of the uniqueness of any orthogonal decomposition. To verify the additivity property of both $\xi_1^0(\cdot)$ and $\xi_2^0(\cdot)$, take a sequence $\{A_n\}$ of disjoint bounded Borel sets with bounded union. From the a.s. countable additivity of ξ^0, which is equivalent to property (iv) of Definition 8.4.VII, we have

$$\xi^0\left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \xi^0(A_n) \quad \text{a.s.;}$$

hence,

$$\xi_1^0\left(\bigcup_{n=1}^{\infty} A_n \right) - \sum_{n=1}^{\infty} \xi^0(A_n) = \xi_2^0\left(\bigcup_{n=1}^{\infty} A_n \right) - \sum_{n=1}^{\infty} \xi(A_n) \quad \text{a.s.}$$

Since the expressions on the two sides of this equation belong to orthogonal subspaces, both must reduce a.s. to the zero random variable. Properties (i)–(iii) of Definition 8.4.VII are readily checked, so it follows that both $\xi_1^0(\cdot)$ and $\xi_2^0(\cdot)$ are wide-sense second-order stationary random measures. But note that even when ξ^0 is known to be a strict-sense random measure, the argument above shows only that ξ_1^0 and ξ_2^0 are wide-sense random measures. □

The classical results that relate the presence of a deterministic component to properties of the spectral measure can also be carried over from $X(\cdot)$ to the random measure $\xi(\cdot)$. They are set out in the following theorem.

Theorem 8.5.III. *Let $\xi(\cdot)$ be a strict- or wide-sense second-order stationary random measure with Bartlett spectrum Γ. Then $\xi(\cdot)$ is purely nondeterministic if and only if Γ is absolutely continuous and its density γ satisfies the condition*

$$\int_{-\infty}^{\infty} \frac{\log \gamma(\omega)\,\mathrm{d}\omega}{1+\omega^2} > -\infty. \tag{8.5.6}$$

This condition is equivalent to the existence of a factorization

$$\gamma(\omega) = |\tilde{g}(\omega)|^2, \tag{8.5.7}$$

where $\tilde{g}(\cdot)$ is the Fourier transform of a (real) generalized function with support on $[0,\infty)$ and can be written in the form $\tilde{g}(\omega) = (1 - i\omega)\tilde{g}_1(\omega)$, where $\tilde{g}_1(\cdot)$ is the Fourier transform of an $L_2(\mathbb{R})$ function with its support in $\mathbb{R}_+$. The function $\tilde{g}(\cdot)$ can be characterized uniquely among all possible factorizations by the requirement that it have an analytic continuation into the upper half-plane $\mathrm{Im}(\omega) > 0$, where it is zero-free and satisfies the normalization condition

$$\tilde{g}(i) = \exp\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\log \gamma(\omega)}{1+\omega^2}\,\mathrm{d}\omega\right). \tag{8.5.8}$$

Proof. Since ξ is purely nondeterministic if and only if X defined at (8.5.5) is purely nondeterministic, the results follow from those for the continuous-time process $X(\cdot)$ as set out, for example, in Hannan (1970, Section 3.4). From Sections 8.2 and 8.6, it follows that the spectral measure Γ_X of $X(\cdot)$ is related to the Bartlett spectrum Γ of ξ by $\Gamma_X(\mathrm{d}\omega) = (1+\omega^2)^{-1}\,\Gamma(\mathrm{d}\omega)$, so Γ_X has a density γ_X if and only if Γ has a density, and the density γ satisfies (8.5.6) if and only if γ_X does because the discrepancy $\int_{-\infty}^{\infty}(1+\omega^2)^{-1}\log(1+\omega^2)\,\mathrm{d}\omega$ is finite.

Similarly, if $\gamma_X(\omega) = |\tilde{g}_X(\omega)|^2$, where $\tilde{g}_X(\cdot)$ is the Fourier transform of an $L_2(\mathbb{R})$ function with support in $\mathbb{R}_+$, we can set $g_1 = g_X$ so that (8.5.7) holds together with the assertions immediately following it.

Finally, (8.5.8) follows from the corresponding relation for g_1 since

$$\begin{aligned}\tilde{g}(i) = 2\tilde{g}_1(i) &= 2\exp\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\log\gamma_X(\omega)}{1+\omega^2}\,\mathrm{d}\omega\right)\\ &= \exp\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\log\gamma(\omega)}{1+\omega^2}\,\mathrm{d}\omega\right)\end{aligned}$$

using the identity

$$\int_{-\infty}^{\infty}\frac{\log(1+\omega^2)}{1+\omega^2}\,\mathrm{d}\omega = 2\pi\log 2. \qquad \square$$

These extensions from Γ_X to Γ are to be expected because the criteria are analytic and relate to the factorization of the function γ rather than to its behaviour as $\omega \to \pm\infty$. We illustrate the results by two examples.

EXAMPLE 8.5(b) *Two-point Poisson cluster process.* Suppose that clusters occur at the instants of a Poisson process with parameter μ and that each cluster contains exactly two members, one at the cluster centre and the other at a fixed time h after the first. Then, the reduced covariance measure has just three atoms, one of mass 2μ at 0 and the others at $\pm h$, each of mass μ. The Bartlett spectrum is therefore absolutely continuous with density

$$\gamma(\omega) = \mu(1 + \cos \omega h)/\pi = \mu(2 \cos^2 \tfrac{1}{2}\omega h)/\pi.$$

In seeking a factorization of the form (8.5.7), it is natural to try $(2\mu/\pi)^{1/2} \times \cos \frac{1}{2}\omega h$ as a candidate, but checking the normalization condition (8.5.8) reveals a discrepancy: using the relation

$$\int_{-\infty}^{\infty} (1 + \omega^2) \log \left(\cos^2 \tfrac{1}{2}\omega h \right) \mathrm{d}\omega = 2\pi \log \left(\tfrac{1}{2}(1 + \mathrm{e}^{-h}) \right)$$

leads to $(2\mu/\pi)^{1/2}(1 + \mathrm{e}^{-h})/2$ for the right-hand side of (8.5.8), while the candidate gives $\tilde{g}(i) = (2\mu/\pi)^{1/2} \cosh \frac{1}{2}\omega h$. It is not difficult to see that the correct factorization is

$$\tilde{g}(\omega) = \sqrt{\frac{2\mu}{\pi}}\, \frac{1 + \mathrm{e}^{i\omega h}}{2} = \sqrt{\frac{2\mu}{\pi}}\, \mathrm{e}^{i\omega h/2} \cos \tfrac{1}{2}\omega h.$$

In this form, we can recognize $\tilde{g}(\cdot)$ as the Fourier transform of a measure with atoms $[\mu/(2\pi)]^{1/2}$ at $t = 0$ and $t = h$, whereas the unsuccessful candidate function is the transform of a measure with atoms of the same mass but at $t = \pm \frac{1}{2}h$; that is, the support is not contained in $[0, \infty)$. □

EXAMPLE 8.5(c) *Random measures with rational spectral density.* When the spectral density is expressible as a rational function, and hence of the form

$$\left(\prod_{j=1}^{m} (\omega^2 + \alpha_j^2) \right) \Big/ \left(\prod_{j=1}^{n} (\omega^2 + \beta_j^2) \right)$$

for nonnegative integers m, n with $m \le n$, and real α_j, β_j, the identification of the canonical factorization is much simpler because it is uniquely determined (up to a constant of unit modulus) by the requirements that $\tilde{g}(\omega)$ be analytic and zero-free in the upper half-plane. Two situations commonly occur according to whether $m < n$ or $m = n$. In the former case, the process has a mean square continuous density $x(\cdot)$ and $\Gamma(\cdot)$ is a totally finite measure. The problem reduces to the classical one of identifying the canonical factorization of the spectrum for the density of the process. For point processes, however, the δ-function in the covariance measure produces a term that does not converge to zero as $|\omega| \to \infty$, implying that $m = n$; the same situation obtains whenever the random measure has a purely atomic component.

As an example of the latter form, recall the comments preceding Example 8.2(e) concerning point process models with spectral densities of the form

$$\gamma(\omega) = \frac{A^2(\alpha^2 + \omega^2)}{\beta^2 + \omega^2}\,.$$

The canonical factorization here takes the form (with A, α, and β real and positive)

$$\tilde{g}(\omega) = \frac{A(\alpha - i\omega)}{\beta - i\omega} = A\left(1 + \frac{\alpha - \beta}{\beta - i\omega}\right)$$

corresponding to the time-domain representation

$$g(t) = A\big(\delta_0(t) + (\alpha - \beta)I_{[0,\infty)}(t)\mathrm{e}^{-\beta t}\big).$$

Similar forms occur in more general point process models, with polynomial a sum of products of exponential and polynomial factors in place of the exponential. □

The main thrust of these factorization results is that they lead to a time-domain representation that can be used to develop explicit prediction formulae. The fact that the canonical factor $\tilde{g}(\omega)$ is in general the transform not of a function but only of a generalized function leads to some specific difficulties. However, much of the argument is not affected by this fact, as we now indicate.

Let $Z(\cdot)$ be the process of orthogonal increments arising in the spectral representation of ξ^0, and $\tilde{g}(\cdot)$ the canonical factor described in Theorem 8.5.III. Introduce a further process $U(\cdot)$ with orthogonal increments by scaling the $Z(\cdot)$ process to have stationary increments as in

$$Z(\mathrm{d}\omega) = \overline{\tilde{g}(\omega)}\, U(\mathrm{d}\omega), \tag{8.5.9}$$

where the invertibility of $\tilde{g}$ implies that for all real ω

$$\mathrm{E}|U(\mathrm{d}\omega)|^2 = |\tilde{g}(\omega)|^{-2}\mathrm{E}|Z(\mathrm{d}\omega)|^2 = \mathrm{d}\omega.$$

Note that the use of the complex conjugate of $\tilde{g}$ in (8.5.9) is purely for convenience: it simplifies the resulting moving average representation in the time domain.

Corresponding to U in the frequency domain, we may, in the usual way, define a new process V in the time domain through the Parseval relations, so

$$\int_{-\infty}^{\infty} \phi(t)\, V(\mathrm{d}t) = \int_{-\infty}^{\infty} \tilde{\phi}(\omega)\, U(\mathrm{d}\omega), \tag{8.5.10}$$

which in this case can be extended to all functions $\phi \in L_2(\mathbb{R})$. It can be verified that $V(\cdot)$ also has orthogonal and stationary increments, with

$$\mathrm{E}|V(\mathrm{d}t)|^2 = 2\pi\,\mathrm{d}t,$$

corresponding to the more complete statement

$$\operatorname{var}\left(\int_{-\infty}^{\infty}\phi(t)\,V(\mathrm{d}t)\right)=2\pi\int_{-\infty}^{\infty}|\phi(t)|^2\,\mathrm{d}t$$
$$=\int_{-\infty}^{\infty}|\phi(\omega)|^2\,\mathrm{d}\omega=\operatorname{var}\left(\int_{-\infty}^{\infty}\tilde{\phi}(\omega)\,U(\mathrm{d}\omega)\right).$$

On the other hand, from the Parseval relation for the ξ^0 process, we have for integrable ϕ, for which $\tilde{\phi}\in L_2(\Gamma)$,

$$\int_{-\infty}^{\infty}\phi(t)\,\xi^0(\mathrm{d}t)=\int_{-\infty}^{\infty}\tilde{\phi}(\omega)\,Z(\mathrm{d}\omega)=\int_{-\infty}^{\infty}\tilde{\phi}(\omega)\,\overline{\tilde{g}(\omega)}\,U(\mathrm{d}\omega).\tag{8.5.11}$$

Thus, if we could identify $\tilde{\phi}\bar{\tilde{g}}$ with the Fourier transform of some function $\phi*g^*$ in the time domain, it would be possible to write

$$\int_{-\infty}^{\infty}\phi(t)\,\xi^0(\mathrm{d}t)=\int_{-\infty}^{\infty}(\phi*g^*)(s)\,V(\mathrm{d}s)=\int_{-\infty}^{\infty}\phi(t)\,\mathrm{d}t\int_{-\infty}^{t}g(t-s)\,V(\mathrm{d}s),$$

corresponding to the moving average representation

$$\xi^0(\mathrm{d}t)=\int_{-\infty}^{t}g(t-s)\,V(\mathrm{d}s)\,\mathrm{d}t.$$

Because $g(\cdot)$ is not, in general, a function, these last steps have a purely formal character. They are valid in the case of a process ξ^0 having a mean square continuous density, but in general we need to impose further conditions before obtaining any meaningful results. In most point process examples, the generalized function $g(\cdot)$ can be represented as a measure, but it is an open question as to whether this is true for all second-order random measures.

We proceed by imposing conditions that, although restrictive, are at least general enough to cover the case of a point process with rational spectral density. They correspond to assuming that the reduced factorial cumulant measure $\widehat{C}_{[2]}$ is totally finite, so that the spectral density can be written in the form

$$\gamma(\omega)=(2\pi)^{-1}\big(m+\tilde{c}_{[2]}(\omega)\big).$$

Specifically, assume that

$$\tilde{g}(\omega)=A\big(1+\tilde{c}(\omega)\big)\tag{8.5.12}$$

for some positive constant A and function $\tilde{c}\in L_2(\mathbb{R})$. Then, the generalized function aspect of $g(\cdot)$ is limited to a δ-function at the origin, and there exists an $L_2(\mathbb{R})$ function $c(\cdot)$ such that

$$g(t)=\begin{cases}A\big(\delta_0(t)+c(t)\big) & (t\ge 0),\\ 0 & (t<0).\end{cases}$$

Under the same conditions, the reciprocal $1/\tilde{g}(\omega)$ can be written

$$1/\tilde{g}(\omega) = A^{-1}\big(1 - \tilde{d}(\omega)\big),$$

where $\tilde{d}(\omega) = \tilde{c}(\omega)/(1 + \tilde{c}(\omega))$, and from

$$\int_{-\infty}^{\infty} |\tilde{d}(\omega)|^2 \, \gamma(\omega) \, \mathrm{d}\omega = A^2 \int_{-\infty}^{\infty} |\tilde{c}(\omega)|^2 \, \mathrm{d}\omega < \infty$$

it follows that $\tilde{d} \in L_2(\gamma)$. Often, we have $L_2(\gamma) \subseteq L_2(\mathbb{R})$, in which case $\tilde{d} \in L_2(\mathbb{R})$, implying the existence of a representation of a Fourier inverse of $1/\tilde{g}(\omega)$ as

$$\begin{cases} A^{-1}\big(\delta_0(t) - d(t)\big) & (t \geq 0) \\ 0 & (t < 0) \end{cases} \tag{8.5.13}$$

for some function $d \in L_2(\mathbb{R})$.

Proposition 8.5.IV (Moving Average and Autoregressive (ARMA) Representations)**.** *Suppose* (8.5.12) *holds for some* $\tilde{c} \in L_2(\mathbb{R})$. *Then, using the notation of* (8.5.12–13), *for* $\phi \in L_1(\mathbb{R})$ *such that* $\tilde{\phi} \in L_2(\mathbb{R})$, *the zero-mean process* $\xi^0(\cdot)$ *is expressible as*

$$\int_{\mathbb{R}} \phi(t) \, \xi^0(\mathrm{d}t) = \int_{\mathbb{R}} \tilde{\phi}(t) \, V(\mathrm{d}t) + \int_{\mathbb{R}} \tilde{\phi}(t) X(t) \, \mathrm{d}t \quad \text{a.s.}, \tag{8.5.14}$$

where $V(\cdot)$ *is a zero-mean process with stationary orthogonal increments such that*

$$\mathrm{E}|V(\mathrm{d}t)|^2 = 2\pi A^2 \, \mathrm{d}t \tag{8.5.15}$$

and $X(\cdot)$ *is a mean square continuous process that can be written in the moving average form*

$$X(t) = \int_{-\infty}^{t} c(t-u) \, V(\mathrm{d}u) \quad \text{a.s.} \tag{8.5.16}$$

or, if furthermore $\tilde{d} \in L_2(\mathbb{R})$, *in the autoregressive form*

$$X(t) = \int_{-\infty}^{t} d(t-u) \, \xi^0(\mathrm{d}u) \quad \text{a.s.} \tag{8.5.17}$$

Proof. Under the stated assumptions, it follows from (8.5.11) that

$$\int_{\mathbb{R}} \phi(t) \, \xi^0(\mathrm{d}t) = A \int_{\mathbb{R}} \tilde{\phi}(\omega) \, U(\mathrm{d}\omega) + A \int_{\mathbb{R}} \tilde{\phi}(\omega) \overline{\tilde{c}(\omega)} \, U(\mathrm{d}\omega) \quad \text{a.s.} \tag{8.5.18}$$

Consider now the process $X(\cdot)$ defined by the spectral representation

$$X(t) = \int_{\mathbb{R}} \mathrm{e}^{it\omega} \tilde{c}(\omega) \, U(\mathrm{d}\omega) = \int_{\mathbb{R}} \mathrm{e}^{it\omega} \, Z_X(\mathrm{d}\omega) \quad \text{a.s.}, \tag{8.5.19}$$

where Z_X has orthogonal increments and satisfies $\mathrm{E}\big(|Z(\mathrm{d}\omega)|^2\big) = \gamma_X(\omega)\,\mathrm{d}\omega = |\tilde{c}(\omega)|^2\,\mathrm{d}\omega$. To ensure that $\int_{\mathbb{R}} X(t)\phi(t)\,\mathrm{d}t$ can be validly interpreted as a mean square integral, it is enough to show that $\tilde{\phi} \in L_2(\gamma_X)$, as in the discussion around (8.5.3). But $\phi \in L_1(\mathbb{R})$ implies that $|\tilde{\phi}(\omega)|$ is bounded for $\omega \in \mathbb{R}$, and then the assumption that $\tilde{c} \in L_2(\mathbb{R})$ implies that

$$\int_{\mathbb{R}} |\tilde{\phi}(\omega)|^2\,|\tilde{c}(\omega)|^2\,\mathrm{d}\omega = \int_{\mathbb{R}} |\tilde{\phi}(\omega)|^2 |\gamma_X(\omega)\,\mathrm{d}\omega < \infty,$$

as required.

The terms on the right-hand side of (8.5.18) can now be replaced by their corresponding time-domain versions. Thus, we have

$$A\int_{\mathbb{R}} \tilde{\phi}(\omega)\,U(\mathrm{d}\omega) = \int_{\mathbb{R}} \phi(t)\,V(\mathrm{d}t),$$

absorbing the constant A into the definition of the orthogonal-increment process V as in (8.5.10), while the discussion above implies that the last term in (8.5.18) can be replaced by $\int_{\mathbb{R}} \tilde{\phi}(t)X(t)\,\mathrm{d}t$, with $X(t)$ defined as in (8.5.16). This establishes the representation (8.5.14).

To establish the autoregressive form in (8.5.17), observe that

$$\begin{aligned} Y(t) \equiv \int_{\mathbb{R}} \mathrm{e}^{it\omega}\tilde{d}(\omega)\,Z(\mathrm{d}\omega) &= A\int_{\mathbb{R}} \mathrm{e}^{it\omega}\tilde{d}(\omega)\big(1+\tilde{c}(\omega)\big)\,U(\mathrm{d}\omega) \\ &= A\int_{\mathbb{R}} \mathrm{e}^{it\omega}\tilde{c}(\omega)\,U(\mathrm{d}\omega) = X(t), \end{aligned}$$

the integrals being well defined and equal a.s. from the assumption that $\tilde{c} \in L_2(\mathbb{R})$, from which it follows that $\tilde{d} \in L_2(\Gamma)$. If ξ^0 is a strict-sense random measure, then the time-domain integral (8.5.17) is well defined for $\phi \in \mathrm{L}_1(\mathbb{R})$ and can be identified a.s. with its frequency-domain version $Y(t)$ above. If ξ^0 is merely a wide-sense process, then (8.5.17) can be defined only as a mean square limit, which will exist whenever $\tilde{d} \in L_2(\Gamma)$. In either case, therefore, $X(t) = Y(t)$ a.s. □

Equation (8.5.14) can be combined with equations (8.5.16) and (8.5.17) to yield the abbreviated but suggestive forms set out below; they embody the essential content of the moving average and autoregressive representations in the present context.

Corollary 8.5.V. *With the same assumptions and notation as in Proposition* 8.5.IV*,*

$$\xi^0(\mathrm{d}t) = V(\mathrm{d}t) + \int_{-\infty}^{t-} c(t-u)\,V(\mathrm{d}u)\,\mathrm{d}t \quad \text{a.s.}, \tag{8.5.20}$$

$$\xi^0(\mathrm{d}t) = V(\mathrm{d}t) + \int_{-\infty}^{t-} d(t-u)\,\xi^0(\mathrm{d}u)\,\mathrm{d}t \quad \text{a.s.} \tag{8.5.21}$$

There is a close analogy between (8.5.20) and the martingale decomposition of the cumulative process outlined in the previous chapter: the first term in (8.5.20) corresponds to the martingale term, or innovation, while the second corresponds to the conditional intensity. The difference lies in the fact that the second term in (8.5.20) is necessarily representable as a linear combination of past values, whereas the conditional intensity, its analogue in the general situation, is not normally a linear combination of this type.

Finally, we can use the results of the proposition to establish the forms of the best linear predictors when the assumptions of Proposition 8.5.IV hold. Consider specifically the problem of predicting forward the integral

$$Q \equiv \int_{\mathbb{R}} \phi(s)\, \xi^0(\mathrm{d}s) \quad \text{a.s.} \tag{8.5.22}$$

from observations on $\xi^0(\cdot)$ up to time t. The best linear predictor, in the mean square sense, is just the projection of ϕ onto the Hilbert space $L_2(\xi^0; t)$. From equations (8.5.14) and (8.5.20), we see that it can be written as

$$\widehat{Q}_t = \int_{-\infty}^{t} \phi(s)\, \xi^0(\mathrm{d}s) + \int_{t}^{\infty} \phi(s) \widehat{X}_t(s)\, \mathrm{d}s \quad \text{a.s.}, \tag{8.5.23}$$

where for $s > t$,

$$\widehat{X}_t(s) = \int_{-\infty}^{t} c(s-u)\, V(\mathrm{d}u) \quad \text{a.s.} \tag{8.5.24}$$

The truncated function

$$c_t^s(u) = \begin{cases} c(u) & (u > s-t), \\ 0 & (u \le s-t), \end{cases}$$

is in $L_2(\mathbb{R})$ when c is, and the same is therefore true of its Fourier transform. Consequently, the random integrals in the definitions of $\widehat{X}_t(s)$ and $\widehat{Q}_t$ are well defined by the same argument as used in proving Proposition 8.5.IV.

Equation (8.5.24) already gives an explicit form for the predictor, but it is not convenient for direct use since it requires the computation of $V(\cdot)$. In practice, the autoregressive representation of $\widehat{X}_t(s)$ is more useful. To find it, observe that

$$\begin{aligned} \widehat{X}_t(s) &= \int_{-\infty}^{t} c_t^s(s-u)\, V(\mathrm{d}u) = \int_{\mathbb{R}} \tilde{c}_t^s(\omega)\, U(\mathrm{d}\omega) = \int_{\mathbb{R}} \tilde{c}_t^s(\omega)[1-\tilde{d}(\omega)]\, Z(\mathrm{d}\omega) \\ &= \int_{-\infty}^{t} \left[c(s-u) - \int_0^{t-u} c(s-u-v) d(v)\, \mathrm{d}v \right] \xi^0(\mathrm{d}u) \quad \text{a.s.} \end{aligned} \tag{8.5.25}$$

The integral is well defined not only in the mean square sense but also in the a.s. sense if $d \in L_1(\mathbb{R})$. In this case, the integrand in (8.5.25) can also be written in the form

$$d(s-u) + \int_{t-u}^{s-u} c(s-u-v) d(v)\, \mathrm{d}v,$$

which is then the sum of two $L_1(\mathbb{R})$ functions, both of which can be integrated against ξ^0.

These arguments are enough to establish the validity of the autoregressive form (8.5.25) as an alternative to (8.5.24). It is important to emphasize that $\widehat{X}_t(s)$ is to be interpreted as the predictor of the intensity of the ξ^0 process at time $s > t$, or in abbreviated notation,

$$\widehat{X}_t(s)\,\mathrm{d}s = \mathrm{E}[\xi^0(\mathrm{d}s) \mid \mathcal{H}_t] = \mathrm{E}[\lambda(s) \mid \mathcal{H}_t], \tag{8.5.26}$$

where both expectations are to be understood only in the sense of Hilbert-space projections.

Thus, the assumptions of Proposition 8.5.IV imply that the intensity is predicted forward as a mean square continuous function of the past. In contrast to the case where the process itself is mean square continuous, when the predictors may involve differentiations, here they are always smoothing operators. The discussion can be summarized as follows.

Proposition 8.5.VI. *Under the conditions of Proposition* 8.5.IV, *the best linear predictor of the functional Q in* (8.5.22), *given the history $\mathcal{H}_t$ of the ξ^0 process on $(-\infty, t]$, is as in* (8.5.23), *in which the mean square continuous process $\widehat{X}_t(s)$ may be regarded as the best linear predictor of the 'intensity' $\xi^0(\mathrm{d}s)/\mathrm{d}s$ for $s > t$ and has the moving average representation* (8.5.24) *and the autoregressive representation*

$$\widehat{X}_t(s) = \int_{-\infty}^{t} h_t(s-u)\,\xi^0(\mathrm{d}u),$$

where

$$\begin{aligned} h_t(s-u) &= c(s-u) - \int_0^{t-u} c(s-u-v)d(v)\,\mathrm{d}v \\ &= d(s-u) + \int_{t-u}^{s-u} c(s-u-w)d(w)\,\mathrm{d}w. \end{aligned} \tag{8.5.27}$$

Returning to the original random measure ξ (as distinct from ξ^0), we obtain the following straightforward corollary, stated in the abbreviated form analogous to (8.5.26).

Corollary 8.5.VII. *The random measure ξ can be predicted forward with predicted intensity at $s > t$ given by*

$$\mathrm{E}[\xi(\mathrm{d}s) \mid \mathcal{H}_t] = \big(m + \widehat{X}_t(s)\big)\,\mathrm{d}s,$$

where the conditional expectation is to be understood in the sense of a Hilbert-space projection.

Example 8.5(d) *A point process with rational spectral density* [continued from Example 8.5(c)]. Consider the case where

$$\gamma(\omega) = A^2(\alpha^2 + \omega^2)/(\beta^2 + \omega^2)\,. \tag{8.5.28}$$

From the form of $\tilde{g}(\omega)$ as earlier, it follows that

$$\tilde{c}(\omega) = \frac{\alpha - \beta}{\beta - i\omega}, \qquad \tilde{d}(\omega) = \frac{\alpha - \beta}{\alpha - i\omega},$$
$$c(t) = (\alpha - \beta)\mathrm{e}^{-\beta t}, \qquad d(t) = (\alpha - \beta)\mathrm{e}^{-\alpha t} \qquad (t \geq 0).$$

Substituting into (8.5.27), we find

$$\begin{aligned} h_t(s-u) &= (\alpha - \beta)\mathrm{e}^{-\beta(s-u)} - (\alpha - \beta)^2 \mathrm{e}^{-\beta(s-u)} \int_0^{t-u} \mathrm{e}^{-(\alpha-\beta)v}\, \mathrm{d}v \\ &= (\alpha - \beta)\mathrm{e}^{-\beta(s-t)}\mathrm{e}^{-\alpha(t-u)}, \end{aligned}$$

so that

$$\widehat{X}_t(s) = (\alpha - \beta)\mathrm{e}^{-\beta(s-t)} \int_{-\infty}^{t} \mathrm{e}^{-\alpha(t-u)}\, \xi^0(\mathrm{d}u) \quad \text{a.s.} \tag{8.5.29}$$

Thus, the predictor here is a form of exponential smoothing of the past. How well it performs relative to the full predictor, based on complete information about the past, depends on the particular process that is under consideration.

The most instructive and tractable example is again the Hawkes process, which, in order to reproduce the second-order properties above, should have a complete conditional intensity of the special form as in Exercise 7.2.5,

$$\lambda^*(t) = \lambda + \nu \int_{-\infty}^{t-} \alpha \mathrm{e}^{-\alpha(t-u)}\, N(\mathrm{d}u) \quad \text{a.s.}, \quad \equiv \lambda + \nu\alpha Y(t), \quad \text{say,} \tag{8.5.30}$$

which leads to (8.5.28) with $A^2 = \lambda/2\pi$, $\beta = \alpha(1 - \nu)$ [see equation (8.2.10)]. The full predictor can be found by taking advantage of the special form of the intensity, which implies that the quantity $Y(t)$ as above and in Exercise 7.2.5 is Markovian. Defining $m(t) = \mathrm{E}[Y(t)] = \int_0^\infty y\, F_t(\mathrm{d}y)$, we find by integrating (7.2.12) that $m(t)$ satisfies the ordinary differential equation

$$\frac{\mathrm{d}m(t)}{\mathrm{d}t} = -\beta m(t) + \lambda$$

with solution

$$m(t) = \frac{\lambda}{\beta} + \left(m(0) - \frac{\lambda}{\beta}\right)\mathrm{e}^{-\beta t}.$$

To apply this result to the nonlinear prediction problem analogous to that solved by $\widehat{X}_t(s)$ in the linear case, we should set $m(0) = Y(t)$ and consider $m(s-t)$, which gives the solution

$$\begin{aligned} \widehat{X}_t^*(s) \equiv \mathrm{E}[\lambda^*(s) \mid \mathcal{H}_t] &= \lambda + \nu\alpha \mathrm{E}[Y(t+s) \mid Y(t)] = \lambda + \nu\alpha m(s-t) \\ &= \frac{\lambda}{1-\nu} + \nu\alpha \left[Y(t) - \frac{\lambda}{\beta}\, \mathrm{e}^{-\beta(s-t)} \right]. \end{aligned}$$

Replacing $Y(t)$ by its representation in terms of the past of the process as in (8.5.30) leads back to (8.5.29).

Thus, *for a Hawkes process with exponential infectivity function, the best linear predictor of the future intensity equals the best nonlinear predictor of the future intensity*. It appears to be an open question whether this result extends to other Hawkes processes or to other stationary point processes.

Linear and nonlinear predictors for an example of a renewal process with rational spectral density are discussed in Exercise 8.5.2. □

EXAMPLE 8.5(e) *Two-point Poisson cluster process* [continued from Example 8.5(b)]. While this example does not satisfy the assumptions of the preceding discussion, it is simple enough to handle directly. From the expression for $\tilde{g}(\omega)$ given earlier, the moving average representation can be written in the form

$$\xi^0(\mathrm{d}t) = (\mu/2\pi)^{1/2}\{V(\mathrm{d}t) + V(\mathrm{d}t - h)\}.$$

The reciprocal has the form

$$1/\tilde{g}(\omega) = (2\pi/\mu)^{1/2}(1 + \mathrm{e}^{i\omega h})^{-1},$$

which, if we proceed formally, can be regarded as being the sum of an infinite series corresponding to the time-domain representation

$$V(\mathrm{d}t) = \sqrt{2\pi/\mu}\,\big[\xi^0(\mathrm{d}t) - \xi^0(\mathrm{d}t - h) + \xi^0(\mathrm{d}t - 2h) - \cdots\big].$$

In fact, the sum is a.s. finite and has the effect of retaining in V only those atoms in ξ^0 that are not preceded by a further atom h time units previously; that is, of retaining the atoms at cluster centres but rejecting their cluster companions. From this, it is clear that the process $V(\cdot)$ is just a scaled version of the zero-mean version of the original Poisson process of cluster centres, and the moving average representation is simply a statement of how the clusters are formed. It is now easy to form linear predictors: we have

$$\xi^0(\mathrm{d}s \mid \mathcal{H}_t) = \begin{cases} 0 & (s - t > h), \\ (\mu/2\pi)^{1/2}\, V(\mathrm{d}s - h) & (0 \le h \le s - t), \end{cases}$$

and on $0 \le h \le s - t$ we also have

$$\hat{\xi}^0(\mathrm{d}s \mid \mathcal{H}_t) = \sum_{j=1}^{\infty} (-1)^j\, \xi^0(\mathrm{d}s - jh).$$

The effect of the last formula is to scan the past to see if there is an atom at $s - h$ not preceded by a further atom at $s - 2h$: the predictor predicts an atom at s when this is the case and nothing otherwise. □

Exercises and Complements to Section 8.5

8.5.1 *Renewal processes with rational spectral density.* Show that the Bartlett spectrum for the renewal process considered in Exercise 4.2.4 with interval density $\mu^2 x e^{-\mu x}$ has the form

$$\gamma(\omega) = \frac{\mu}{4\pi} \frac{\omega^2 + 2\mu^2}{\omega^2 + 4\mu^2} .$$

8.5.2 *Linear and nonlinear prediction of a renewal process.*

(a) Show that for any renewal process the best nonlinear predictor $\mathrm{E}[\lambda^*(t+s) \mid \mathcal{H}_t]$ for the intensity is the renewal density for the delayed renewal process in which the initial lifetime has d.f. $[F(B_t+s)-F(B_t)]/[1-F(B_t)]$, where B_t is the backward recurrence time at time t.

(b) Find explicitly the best predictor for the process in Exercise 8.5.1.

(c) Find the canonical factorization of the spectrum of the renewal process in Exercise 8.5.1, and find the best linear predictor $\widehat{B}_t(s)$, where B_t is the backward recurrence time at t. When does it coincide with the best nonlinear predictor in (b)?

(d) Investigate the expected information gain per event based on the use of the linear and nonlinear predictors outlined above.

8.6. P.P.D. Measures

In this section, we briefly develop the properties of p.p.d. measures required for the earlier sections of this chapter. We follow mainly the work of Vere-Jones (1974) and Thornett (1979); related material, in a more abstract setting, is in Berg and Frost (1975). No significant complications arise in developing the theory for $\mathbb{R}^d$ rather than for the line, so we follow this practice, although most of the examples are taken from the one-dimensional context.

Since the measures we deal with are not totally finite in general, we must first define what is meant by a Fourier transform in this context. As in the theory of generalized functions (see e.g. Schwarz, 1951), we make extensive use of Parseval identities

$$\int_{\mathbb{R}^d} \psi(x)\, \nu(\mathrm{d}x) = \int_{\mathbb{R}^d} \widetilde{\psi}(\omega)\, \mu(\mathrm{d}\omega) \tag{8.6.1}$$

to identify the measure ν as the Fourier transform of the measure μ in (8.6.1). Here

$$\widetilde{\psi}(\omega) = \int_{\mathbb{R}^d} \mathrm{e}^{ix\cdot\omega} \psi(x)\, \mathrm{d}x$$

is the ordinary (d-dimensional) Fourier transform of $\psi(\cdot)$, but such functions must be suitably restricted. A convenient domain for ψ is the space $\mathcal{S}$ of real or complex functions of *rapid decay*; that is, of infinitely differentiable functions that, together with their derivatives, satisfy inequalities of the form

$$\left| \frac{\partial^k \psi(x)}{\partial x_1^{k_1} \cdots \partial x_d^{k_d}} \right| \leq \frac{C(k,r)}{(1+|x|)^r}$$

for some constants $C(k,r) < \infty$, all positive integers r, and all finite families of nonnegative integers $(k_1, \ldots, k_d)$ with $k_1 + \cdots + k_d = k$. The space $\mathcal{S}$ has certain relevant properties, proofs of which are sketched in Exercise 8.6.1:

(i) $\mathcal{S}$ is invariant under the Fourier transformation taking ψ into $\widetilde{\psi}$.

(ii) $\mathcal{S}$ is invariant under multiplication or convolution by real- or complex-valued integrable functions g on $\mathbb{R}^d$ such that both g and $\tilde{g}$ are zero-free.

(iii) Integrals with respect to all functions $\psi \in \mathcal{S}$ uniquely determine any boundedly finite measure on $\mathbb{R}^d$.

The following definitions collect together some properties of boundedly finite measures that are important in the sequel. We use the notation, for complex-valued functions ψ and ϕ,

$$(\psi * \phi)(x) = \int_{\mathbb{R}^d} \psi(y)\overline{\phi(x-y)}\,\mathrm{d}y, \qquad \psi^*(x) = \overline{\psi(-x)},$$

so that

$$(\psi * \psi^*)(x) = \int_{\mathbb{R}^d} \psi(y)\psi(y-x)\,\mathrm{d}y.$$

Definition 8.6.I. *A boundedly finite signed measure $\mu(\cdot)$ on $\mathbb{R}^d$ is*

(i) translation-bounded *if for all $h > 0$ and $x \in \mathbb{R}^d$ there exists a finite constant K_h such that, for every sphere $S_h(x)$ with centre $x \in \mathbb{R}^d$ and radius h,*

$$\big|\mu\big(S_h(x)\big)\big| \le K_h; \tag{8.6.2}$$

(ii) positive-definite *if for all bounded measurable functions ψ of bounded support,*

$$\int_{\mathbb{R}^d} (\psi * \psi^*)(x)\,\mu(\mathrm{d}x) \ge 0; \tag{8.6.3}$$

(iii) transformable *if there exists a boundedly finite measure ν on $\mathbb{R}^d$ such that* (8.6.1) *holds for all $\psi \in \mathcal{S}$;*

(iv) a p.p.d. measure *if it is nonnegative (i.e. a measure rather than a signed measure) and positive-definite.*

A few comments on these definitions are in order. The concept of translation boundedness appears naturally in this context and is discussed further by Lin (1965), Argabright and de Lamadrid (1974), Thornett (1979), and Robertson and Thornett (1984). If μ is nonnegative, then it is clear that if (8.6.2) holds for some $h > 0$ it holds for all such h. The notion of positive-definiteness in (8.6.3) is a direct extension of the same notion for continuous functions; indeed, if μ is absolutely continuous, then it is positive-definite in the sense of (8.6.3) if and only if its density is a positive-definite function in the usual sense. Concerning the Parseval relation in (8.6.1), it is important to note that if the measure μ is transformable, then ν is uniquely determined by μ and conversely. Equation (8.6.1) generalises the relation

$$c(x) = \int_{\mathbb{R}^d} \mathrm{e}^{i\omega\cdot x} F(\mathrm{d}\omega)$$

for the covariance density in terms of the spectral measure F of a mean square continuous process to which it reduces (with the appropriate identifications) when the random measure and associated covariance measure are absolutely continuous.

Our main interest is in the class of p.p.d. measures on $\mathbb{R}^d$, denoted below by $\mathcal{P}^+$. Some examples may help to indicate the scope and character of $\mathcal{P}^+$.

EXAMPLE 8.6(a) *Some examples of p.p.d. measures.*
(1°) *A simple counterexample.* The measure on $\mathbb{R}^d$ with unit mass at each of the two points ± 1 is not a p.p.d. measure because its Fourier transform $2\cos\omega$ can take negative values and it thus fails to be positive-definite. On the other hand, the convolution of this measure with itself (i.e. the measure with unit mass at ± 2 and mass of two units at 0) is a p.p.d. measure, and its Fourier transform is the boundedly finite (but not totally bounded) measure with density $4\cos\omega$. This also shows that the convolution square root measure of a p.p.d. measure need not be p.p.d.
(2°) *Absolutely continuous p.p.d. measures.* Every nonnegative positive-definite function defines the density of an absolutely continuous p.p.d. measure.
(3°) *Counting measure.* Let μ have unit mass at every $2\pi j$ for $j = 0, \pm 1, \ldots$. Then, for $\psi \in \mathcal{S}$, (8.6.1) reduces to the Poisson summation formula (see Exercise 8.6.4 for details)

$$\sum_{n=-\infty}^{\infty} \psi(n) = \sum_{j=-\infty}^{\infty} \widetilde{\psi}(2\pi j);$$

that is, μ has as its Fourier transform the measure ν with unit mass at each of the integers $n = 0, \pm 1, \ldots$. It also shows that ν, and thus μ as well, is positive-definite (take for ψ a function of the form $\phi * \phi^*$ so that the right-hand side becomes $\sum |\tilde{\phi}(2\pi j)|^2 \geq 0$).
(4°) *Closure under product.* Let $\mu_1, \ldots, \mu_d$ be p.p.d. measures on $\mathbb{R}$ with Fourier transforms $\tilde{\mu}_1, \ldots, \tilde{\mu}_d$. Then, the product measure $\mu_1 \times \cdots \times \mu_d$ is a p.p.d. measure on $\mathbb{R}^d$ with Fourier transform $\tilde{\mu}_1 \times \cdots \times \tilde{\mu}_d$. □

A simple and elegant theory for measures in $\mathcal{P}^+$ and their Fourier transforms can be developed by the standard device of approximating μ by a smoothed version obtained by convoluting μ with a suitable smoothing function such as the symmetric probability densities

$$t(x) = (1 - |x|)_+ \qquad \text{(triangular density)},$$
$$\mathrm{e}_\lambda(x) = \tfrac{1}{2}\lambda \mathrm{e}^{-\lambda|x|} \qquad \text{(two-sided exponential density)},$$

and their multivariate extensions

$$t(x) = \prod_{i=1}^{d} (1 - |x_i|)_+ \,, \tag{8.6.4a}$$

$$\mathrm{e}_\lambda(x) = \left(\tfrac{1}{2}\lambda\right)^d \exp\left(-\lambda \sum_{i=1}^{d} |x_i| \right). \tag{8.6.4b}$$

Observe that

$$t(x) = \int_{\mathbb{R}^d} I_{\mathbb{U}^d}(x-y)\, I_{\mathbb{U}^d}(-y)\, \mathrm{d}y. \tag{8.6.4a$'$}$$

We are now in a position to establish the basic properties of $\mathcal{P}^+$.

Proposition 8.6.II. (a) $\mathcal{P}^+$ *is a closed positive cone in* $\mathcal{M}^{\#}(\mathbb{R}^d)$.
(b) *Every p.p.d. measure is symmetric and translation-bounded.*

Proof. In (a), we mean by 'a positive cone' a set closed under the formation of positive linear combinations. Then (a) is just the statement that if a sequence of boundedly finite measures in $\mathbb{R}^d$ converges vaguely to a limit, and if each measure in the sequence is positive-definite, then so is the limit. This follows directly from the definition of vague convergence and the defining relation (8.6.3).

Now let μ be a p.p.d. measure on $\mathbb{R}^d$, and convolve it with $t(\cdot)$ as in (8.6.4a) so that the convolution is well defined. The resultant function

$$c(x) \equiv \int_{\mathbb{R}^d} t(x-y)\, \mu(\mathrm{d}y) \tag{8.6.5}$$

is real-valued, continuous, and for all bounded measurable ψ of bounded support it satisfies, because of (8.6.4a$'$),

$$\int_{\mathbb{R}^d} c(u)(\psi * \psi^*)(u)\, \mathrm{d}u = \int_{\mathbb{R}^d} \big((\psi * I_{\mathbb{U}^d}) * (\psi * I_{\mathbb{U}^d})^*\big)(y)\, \mu(\mathrm{d}y) \geq 0;$$

note that (8.6.3) applies because $\psi * I_{\mathbb{U}^d}$ is measurable and bounded with bounded support whenever ψ is. In other words, the function $c(\cdot)$ is real-valued and positive-definite and hence, from standard properties of such functions, also symmetric and bounded. Since $t(\cdot)$ is symmetric, it is clear that $c(\cdot)$ is symmetric if and only if μ is symmetric, which must therefore hold. Finally, it follows from the positivity of μ and the inequality $t(x) \geq 2^{-d}$ for $\|x\| \leq \frac{1}{4}$ that if K is a bound for $c(\cdot)$,

$$\mu\big(S_{1/4}(x)\big) \leq 2^d \int_{S_{1/4}(x)} c(y)\, \mathrm{d}y \leq 2^d K < \infty.$$

Inequality (8.6.2) is thus established for the case $h = \frac{1}{4}$, and since μ is non-negative, its validity for any other value of h is now apparent. □

The Fourier transform properties can be established by similar arguments, though it is now more convenient to work with the double exponential function $\mathrm{e}_\lambda(\cdot)$ because its Fourier transform

$$\tilde{\mathrm{e}}_\lambda(\omega) = \prod_{i=1}^{d} \frac{\lambda^2}{\lambda^2 + \omega_i^2}$$

has no real zeros. The existence of the convolution $\mu * \mathrm{e}_\lambda$ follows from the translation boundedness just established. The relation

$$d_\lambda(x) = \int_{\mathbb{R}^d} \mathrm{e}_\lambda(x-y)\,\mu(\mathrm{d}y)$$

again defines a continuous positive-definite function. By Bochner's theorem in $\mathbb{R}^d$, it can therefore be represented as the Fourier transform

$$d_\lambda(x) = \int_{\mathbb{R}^d} \mathrm{e}^{i\omega\cdot x}\, G_\lambda(\mathrm{d}\omega)$$

for some totally finite measure $G_\lambda(\cdot)$.

Now let $\psi(\omega)$ be an arbitrary element of $\mathcal{S}$, and consider the function $\tilde{\kappa}(\omega)$ defined by

$$\tilde{\kappa}(\omega) = (1+\omega^2)\psi(-\omega)/(2\pi)^d.$$

Then $\tilde{\kappa} \in \mathcal{S}$ also, and hence $\tilde{\kappa}$ is the Fourier transform of some integrable function κ satisfying

$$\widetilde{\psi}(y) = (\kappa * \mathrm{e}_1)(y).$$

From the Fourier representation of d_1, we have

$$\int_{\mathbb{R}^d} \kappa(x) d_1(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} \tilde{\kappa}(\omega)\, G_1(\mathrm{d}\omega)$$

for all integrable κ and hence in particular for the function κ just constructed. Substituting for κ, we obtain, for all $\psi \in \mathcal{S}$,

$$\begin{aligned}\int_{\mathbb{R}^d} \widetilde{\psi}(y)\,\mu(\mathrm{d}y) &= \int_{\mathbb{R}^d} (\kappa * \mathrm{e}_1)(y)\,\mu(\mathrm{d}y) = \int_{\mathbb{R}^d} \kappa(x) d_1(x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} \tilde{\kappa}(\omega)\, G_1(\mathrm{d}\omega) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \psi(\omega)(1+\omega^2)\, G_1(-\mathrm{d}\omega).\end{aligned}$$

We now *define* the measure ν by

$$\nu(\mathrm{d}\omega) = (2\pi)^{-d}(1+\omega^2)\, G_1(-\mathrm{d}\omega)$$

and observe that ν is boundedly finite and satisfies the equation (8.6.1), which represents ν as the Fourier transform of μ. Thus, we have shown that any p.p.d. measure μ is transformable.

Recall that $\mathcal{S}$ is preserved under the mapping $\psi \to \widetilde{\psi}$. Then, interchanging the roles of ψ and $\widetilde{\psi}$ in (8.6.1) shows that every p.p.d. measure is itself a transform and hence that ν is positive-definite as well as positive; that is, it is itself a p.p.d. measure. Since the determining properties of $\mathcal{S}$ imply that each of the two measures in (8.6.1) is uniquely determined by the other, we have established the principal result of the following theorem.

Theorem 8.6.III. *Every p.p.d. measure $\mu(\cdot)$ is transformable, and the Parseval equation* (8.6.1) *establishes a one-to-one mapping of $\mathcal{P}^+$ onto itself. This mapping can also be represented by the inversion formulae: for bounded ν-continuity sets A,*

$$\nu(A) = \lim_{\lambda\to\infty} \int_{\mathbb{R}^d} \tilde{I}_A(\omega)\tilde{\mathrm{e}}_\lambda(\omega)\,\mu(\mathrm{d}\omega); \tag{8.6.6}$$

for bounded μ-continuity sets B,

$$\mu(B) = \lim_{\lambda\to\infty} -\frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \tilde{I}_B(-x)\tilde{\mathrm{e}}_\lambda(-x)\,\nu(\mathrm{d}x); \tag{8.6.7}$$

$$\nu(\{a\}) = \lim_{T\to\infty} \int_{\mathbb{U}_T^d} \mathrm{e}^{-i\omega\cdot a}\mu(\mathrm{d}\omega); \tag{8.6.8}$$

$$\mu(\{b\}) = \lim_{T\to\infty} \frac{1}{(2\pi T)^d} \int_{\mathbb{U}_T^d} \mathrm{e}^{ix\cdot b}\nu(\mathrm{d}x). \tag{8.6.9}$$

For all Lebesgue integrable ϕ for which $\tilde{\phi}$ is μ-integrable, there holds the extended Parseval relation

$$\int_{\mathbb{R}^d} \phi(x+y)\,\nu(\mathrm{d}y) = \int_{\mathbb{R}^d} \mathrm{e}^{i\omega\cdot x}\tilde{\phi}(\omega)\,\mu(\mathrm{d}\omega) \qquad \text{(a.e. } x\text{)}. \tag{8.6.10}$$

Proof. It remains to establish the formulae (8.6.6–10), all of which are effectively corollaries of the basic identity (8.6.1). Suppose first that A is a bounded continuity set for $\nu(\cdot)$ and hence *a fortiori* for the smoothed version $\nu * \mathrm{e}_\lambda$. Then, for all finite λ, it is a consequence of the Parseval theorem that

$$(\nu * \mathrm{e}_\lambda)(A) = \int_{\mathbb{R}^d} \tilde{I}_\lambda(\omega)\tilde{\mathrm{e}}_\lambda(\omega)\,\mu(\mathrm{d}\omega).$$

Now letting $\lambda \to \infty$, the left-hand side $\to \nu(A)$ by standard properties of weak convergence since it is clear that $\nu * \mathrm{e}_\lambda \to \nu$ weakly on the closure $\bar{A}$ of A. This proves (8.6.6), and a dual argument gives (8.6.7).

To establish (8.6.8), consider again the convolution with the triangular density $t(\cdot)$. Changing the base of the triangle from $(-1,1)$ to $(-h,h)$ ensures that the Fourier transform $\tilde{t}(\omega)$ does not vanish at $\omega = a$ for any given a. Now check via the Parseval identity that the totally finite spectral measure corresponding to the continuous function $c(x)$ in (8.6.5) can be identified with $\tilde{t}(\omega)\nu(\mathrm{d}\omega)$. Then, standard properties of continuous positive-definite functions imply

$$\tilde{t}(a)\nu(\{a\}) = \lim_{T\to\infty} -\frac{1}{(4\pi T)^d} \int_{\mathbb{U}_{2T}^d} \mathrm{e}^{-ia\cdot x}c(x)\,\mathrm{d}x. \tag{8.6.11}$$

Consider

$$D_T \equiv \tilde{t}(a) \int_{\mathbb{U}_{2T}^d} \mathrm{e}^{-ia\cdot x}\mu(\mathrm{d}x) - \int_{\mathbb{U}_{2T}^d} \mathrm{e}^{-ia\cdot x}c(x)\,\mathrm{d}x,$$

which on using the definition of $c(\cdot)$ as the convolution $t * \mu$ yields

$$D_T = \int_{\mathbb{R}^d} \mathrm{e}^{-ia\cdot x}\mu(\mathrm{d}x)\Big\{\tilde{t}(a)I_{\mathbb{U}^d_{2t}}(x) - \int_{-T-x_1}^{T-x_1}\cdots\int_{-T-x_d}^{T-x_d}\mathrm{e}^{-ia\cdot y}t(y)\,\mathrm{d}y\Big\}.$$

The expression inside the braces vanishes both inside the hypercube with vertices $\big(\pm(T-h),\dots,\pm(T-h)\big)$ since the second integral then reduces to $\tilde{t}(a)$ and outside the hypercube with vertices $\big(\pm(T+h),\dots,\pm(T+h)\big)$ since both terms are then zero. Because μ is translation-bounded, there is an upper bound, K_h say, on the mass it allots to any hypercube with edge of length $2h$. The number of such hypercubes needed to cover the region where the integrand is nonzero is certainly bounded by $2d(2+T/h)^{d-1}$, within which region the integrand is bounded by M, say. Thus,

$$\frac{|D_T|}{(4\pi T)^d} \le \frac{2d}{(4\pi)^d}\left(\frac{1}{h}+\frac{2}{T}\right)^{d-1}\frac{MK_h}{T} \to 0 \qquad (T\to\infty).$$

Equation (8.6.8) now follows from (8.6.11), and (8.6.9) follows by a dual argument with the roles of μ and ν interchanged.

It is already evident by analogy with the argument used in constructing $\nu(\cdot)$ that the Parseval relation (8.6.1) holds not only for $\psi\in\mathcal{S}$ but also for any function of the form $(\phi * \mathrm{e}_\lambda)(x)$, where ϕ is integrable. In particular, any function of the form

$$\theta(x) = \int_{\mathbb{R}^d}\phi(y)\overline{\psi(x-y)}\,\mathrm{d}y = (\phi*\psi)(x)$$

has this form for $\psi\in\mathcal{S}$ and ϕ integrable. Hence, for all $\psi\in\mathcal{S}$,

$$\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}x\int_{\mathbb{R}^d}\phi(x+y)\,\nu(\mathrm{d}y) = \int_{\mathbb{R}^d}\tilde{\phi}(\omega)\widetilde{\psi}(\omega)\,\mu(\mathrm{d}\omega).$$

If, furthermore, $\tilde{\phi}$ is μ-integrable, we can rewrite the right-hand side of this equation in the form

$$\int_{\mathbb{R}^d}\psi(x)\,\mathrm{d}x\int_{\mathbb{R}^d}\mathrm{e}^{i\omega\cdot x}\tilde{\phi}(\omega)\,\mu(\mathrm{d}\omega).$$

Since equality holds for all $\psi\in\mathcal{S}$, the coefficients of $\psi(x)$ in the two integrals must be a.e. equal, which gives (8.6.10). □

Many variants on the inversion results given above are possible: the essential point is that μ and ν determine each other uniquely through the Parseval relation (8.6.1). A number of further extensions of this relation can be deduced from (8.6.10), including the following important result.

Proposition 8.6.IV. *For all p.p.d. measures μ with Fourier transform ν as in* (8.6.1)*, and for all bounded functions f of bounded support,*

$$\int_{\mathbb{R}^d}(f*f^*)(x)\,\nu(\mathrm{d}x) = \int_{\mathbb{R}^d}|\tilde{f}(\omega)|^2\,\mu(\mathrm{d}\omega). \tag{8.6.12}$$

PROOF. Examining (8.6.10), we see that the assumed integrability condition implies that the right-hand side there is continuous in x and consequently that the two sides are equal for any value of x at which the left-hand side is also continuous (note that the a.e. condition cannot be dropped in general because altering ϕ at a single point will alter the left-hand side whenever ν has atoms while the right-hand side will remain unchanged). Thus, to check (8.6.12), it is enough to establish the continuity of the left-hand side and the integrability of $|\tilde{f}(\omega)|^2$ with respect to μ on the right-hand side. Appealing to the dominated convergence theorem shows first that $\int_{\mathbb{R}^d} f(u)f(x+u)\,\mathrm{d}u$ is a continuous function of x and second, since this function vanishes outside a bounded set within which $\nu(\cdot)$ is finite, that the integral

$$\int_{\mathbb{R}^d} (f * f^*)(x+y)\,\nu(\mathrm{d}y)$$

also defines a continuous function of x. To establish that $|\tilde{f}(\omega)|^2$ is μ-integrable, we use Lemma 8.6.V given shortly (the lemma is also of interest in its own right). Specifically, express the integral on the right-hand side of (8.6.12) as a sum of integrals over regions B_k as in the lemma. For each term, we then have

$$\int_{B_k} |\tilde{f}(\omega)|^2 \mu(\mathrm{d}\omega) \le b_k \mu(B_k) \le K b_k$$

for some finite constant K using the property of translation boundedness. Finiteness of the integral follows on summing over k and using (8.6.13). □

Lemma 8.6.V (Lin, 1965)**.** *Let A be a bounded set in $\mathbb{R}^d$, h a positive constant, and $\theta(x)$ a square integrable function with respect to Lebesgue measure on A. For $k = (k_1, \ldots, k_d)$, let B_k be the half-open cube $\{k_i h < x_i \le k_i h + h;\ i = 1, \ldots, d\}$, and set*

$$b_k = \sup_{\omega \in B_k} |\tilde{\theta}(\omega)|^2.$$

Then, for all such $\theta(\cdot)$, there exists a finite constant $K(h, A)$ independent of $\theta(\cdot)$ and such that

$$\sum_k b_k \le K(h, A) \int_A |\theta(x)|^2\,\mathrm{d}x, \tag{8.6.13}$$

where summation extends over all integers $k_1, \ldots, k_d = 0, \pm 1, \ldots$.

PROOF. For simplicity, we sketch the proof for $d = 1$, $h = 1$, $A = [-1, 1]$, leaving it to the reader to supply the details needed to extend the result to the general case. Write

$$\alpha_k = \tfrac{1}{2} \int_{-1}^{1} \mathrm{e}^{i\pi k x} \theta(x)\,\mathrm{d}x$$

for the kth Fourier coefficient of θ as a function on the interval $(-1, 1)$. Then, from standard properties of Fourier series, we have

$$\sum_{j=-\infty}^{\infty} |\alpha_j|^2 = \int_{-1}^{1} |\theta(x)|^2 \, \mathrm{d}x < \infty. \tag{8.6.14}$$

Now let ω_k be any point in $B_k = (k, k+1]$, and consider the Taylor series expansion of $\tilde{\theta}(\omega)$ at ω_k. Since A is bounded, $\tilde{\theta}$ is an entire function, and hence the Taylor series about the point k converges throughout B_k, and we can write

$$\begin{aligned}\sum_{k=-\infty}^{\infty} |\tilde{\theta}(\omega_k)|^2 &= \sum_{k=-\infty}^{\infty} \left| \sum_{n=0}^{\infty} \frac{(\omega_k - k)^n}{n!} \, \tilde{\theta}^{(n)}(k) \right|^2 \\ &\leq \sum_{k=-\infty}^{\infty} \left(\sum_{n=0}^{\infty} \frac{|\omega_k - k|^{2n}}{n!} \right) \left(\sum_{n=0}^{\infty} \frac{|\tilde{\theta}^{(n)}(k)|^2}{n!} \right)\end{aligned}$$

from the Cauchy inequality. The first series is dominated by $\sum_{n=0}^{\infty} 1/n! = \mathrm{e}$ for all choices of ω_k; hence, by analogy with (8.6.14), we obtain

$$\begin{aligned}\sum_{k=-\infty}^{\infty} |\tilde{\theta}(\omega_k)|^2 &\leq \mathrm{e} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\sum_{k=-\infty}^{\infty} |\tilde{\theta}^{(n)}(k)|^2 \right) \\ &= \mathrm{e} \sum_{n=0}^{\infty} \frac{1}{n!} \int_{-1}^{1} |x^n \theta(x)|^2 \, \mathrm{d}x \;\leq\; \mathrm{e}^2 \int_{-1}^{1} |\theta(x)|^2 \, \mathrm{d}x.\end{aligned}$$

In particular, choosing ω_k in B_k to maximize $|\tilde{\theta}(\omega_k)|^2$ and so give b_k, (8.6.13) now follows. □

Another integrability result is noted in Exercise 8.6.8.

A simple and characteristic property of a p.p.d. measure is that it remains a p.p.d. measure after addition of an atom of positive mass at the origin. Equally, passing over to the Fourier transforms, it remains a p.p.d. measure after addition of an arbitrary positive multiple of Lebesgue measure. Now suppose that, starting from a given p.p.d. measure μ, we repeatedly subtract multiples of Lebesgue measure in alternation, first from the p.p.d. measure itself and then from its Fourier transform, until one of these measures ceases to be nonnegative. Evidently, certain maximum multiples of Lebesgue measure will be defined by this process, leaving, after subtraction, a p.p.d. measure ν with the additional property that no nonzero multiple of Lebesgue measure can be subtracted from ν or its Fourier transform without destroying the p.p.d. property. Let us call such a measure a *minimal* p.p.d. measure. This leads us to the following elementary structure theorem.

Proposition 8.6.VI. *Every p.p.d. measure μ on $\mathbb{R}^d$ can be uniquely represented as the sum of a minimal p.p.d. measure, a positive multiple of Lebesgue measure on $\mathbb{R}^d$, and an atom of positive mass at the origin.*

Very little is known about the structure of minimal p.p.d. measures, even when $d = 1$. See Exercise 8.6.9.

EXAMPLE 8.6(b). As a simple illustration of (8.6.12), let $f(x)$ be the indicator function of the hyper-rectangle $(0, T_1] \times \cdots \times (0, T_d]$. It then follows that

$$\int_{\mathbb{R}^d} \prod_{i=1}^{d} (T_i - |x_i|)_+ \, \nu(\mathrm{d}x) = \int_{\mathbb{R}^d} \prod_{i=1}^{d} \left(\frac{\sin(\omega_i T_i/2)}{\omega_i/2} \right)^2 \mu(\mathrm{d}\omega). \qquad \square$$

Exercises and Complements to Section 8.6

8.6.1 *The space $\mathcal{S}$.*

(a) Show that if $\mathcal{X} = \mathbb{R}$ and $\psi\colon \mathbb{R} \mapsto \mathbb{R}$ has an integrable kth derivative, then $|\omega^k \widetilde{\psi}(\omega)| \to 0$ as $|\omega| \to \infty$, and that, conversely, if $\int_{-\infty}^{\infty} |x|^k |\psi(x)| \, \mathrm{d}x < \infty$, then $\widetilde{\psi}(\omega)$ is k times differentiable. Deduce that $\mathcal{S}$ is invariant under the Fourier mapping taking ψ into $\widetilde{\psi}$. Extend the result to $\mathbb{R}^d$.

(b) Let $g\colon \mathbb{R}^d \mapsto \mathbb{R}$ be an integrable function with Fourier transform $\tilde{g}$ such that both g and $\tilde{g}$ are zero free on $\mathbb{R}^d$. Show that both the mappings $\psi \mapsto \psi * g$ and $\widetilde{\psi} \mapsto \widetilde{\psi} g$ are one-to-one mappings of $\mathcal{S}$ onto itself. In particular, deduce that this result holds when $\psi(\cdot)$ has the double exponential form $\mathrm{e}_\lambda(\cdot)$ of (8.6.4b).

(c) Show that if μ, ν are boundedly finite measures on $\mathbb{R}$ such that $\int_{\mathbb{R}} \psi \, \mathrm{d}\mu = \int_{\mathbb{R}} \psi \, \mathrm{d}\nu$ for all $\psi \in \mathcal{S}$, then $\mu = \nu$. [*Hint:* Consider $\psi \in \mathcal{S}$ of bounded support and approximate indicator functions.] Extend to $\mathbb{R}^d$.

8.6.2 Let $\{c_n\colon n = 0, \pm 1, \ldots\}$ denote a doubly infinite sequence of reals. Call $\{c_n\}$

(i) *transformable* if $c_n = \int_0^{2\pi} \mathrm{e}^{i\omega n} \nu(\mathrm{d}\omega)$ for some measure ν on $[0, 2\pi]$; and

(ii) *positive-definite* if for all finite families $\{\alpha_1, \ldots, \alpha_k\}$ of complex numbers,

$$\sum_{i=1}^{k} \sum_{j=1}^{k} \alpha_i \bar{\alpha}_j c_{i-j} \geq 0.$$

Let $\mathcal{P}^+(\mathbb{Z})$ denote the class of all p.p.d. sequences and $\mathcal{P}^+(0, 2\pi]$ the class of all p.p.d. measures on $(0, 2\pi]$. Show that every $\{c_n\} \in \mathcal{P}^+(\mathbb{Z})$ is bounded, transformable, and symmetric [i.e. $c_n = c_{-n}$ (all n)] and that a one-to-one mapping between $\mathcal{P}^+(\mathbb{Z})$ and $\mathcal{P}^+(0, 2\pi]$ is defined when the Parseval relation

$$\sum_{j=1}^{k} a_j c_j = \int_0^{2\pi} \tilde{a}(\omega) \, \nu(\mathrm{d}\omega)$$

holds for all $\tilde{a}(\omega) = \sum_{j=1}^{k} a_j \mathrm{e}^{i\omega j}$, with $a_1, \ldots, a_k$ any finite sequence of reals.

8.6.3 Show that not all translation-bounded sequences are transformable.
[*Hint*: Let $\mathcal{X} = \mathbb{R}$ and exhibit a sequence that is bounded but for which $T^{-1} \sum_{j=-T}^{T} c_j$ does not converge to a limit as $T \to \infty$. Use this to define an atomic measure on $\mathbb{R}$ that is not transformable.]

8.6.4 *Poisson summation formula.* Show that if both ψ and $\widetilde{\psi}$ are integrable on $\mathbb{R}$, then

$$\sum_{k=-\infty}^{\infty} \widetilde{\psi}(2\pi k + x) = \sum_{j=-\infty}^{\infty} \psi(j) \mathrm{e}^{-ijx}$$

whenever the left-hand side defines a continuous function of x.
[*Hint*: Under the stated conditions, the left-hand side, $a(x)$ say, is a bounded continuous function of x. Denote by $a_n = (2\pi)^{-1}\int_0^{2\pi} \mathrm{e}^{inx}a(x)\,\mathrm{d}x$ its nth Fourier coefficient, and show by rearrangement that $a_n = \psi(-n)$. Then, the relation is just the representation of $a(\cdot)$ in terms of its Fourier series. Observe that the conditions hold for $\psi \in \mathcal{S}$ and that the formula in Example 8.6(a)(3°) is the special case $x = 0$.]

8.6.5 Show that any p.p.d. measure on $\mathbb{R}$ integrates $(1+\omega^2)^{-\alpha}$ for $\alpha > \frac{1}{2}$ and hence conclude that any p.p.d. measure is a tempered measure in the language of generalized functions.

8.6.6 (a) Let $c(x) = |x|^{-1/2}$ for $(|x| \le 1)$, $c(x) = 0$ elsewhere, and define $g(\omega) = 4 - \int_{-\infty}^{\infty} \mathrm{e}^{i\omega x}c(x)\,\mathrm{d}x$. Show that the measure G with density g is nonnegative and translation-bounded but cannot be made into a p.p.d. measure by adding an atom at the origin.

(b) Show that

$$\nu(A) = \int_A \frac{\mathrm{d}x}{2 - \sin|x|} \qquad \text{(bounded } A \in \mathcal{B}\text{)}$$

defines a measure that is a spectral measure but not a transform (Thornett, 1979).

8.6.7 Show that for $1 < \gamma < 2$ the following functions are densities of p.p.d. measures in $\mathbb{R}^2$, and find their spectral measures:

(a) $c_1(x,y) = \{\sin(\gamma\pi/2)\Gamma(\gamma+1)/2\pi\}^2|xy|^{1-\gamma}$;

(b) $c_2(x,y) = 2^{2(\gamma-2)}\pi^{\gamma-3}(\Gamma(2-\gamma))^{-1}|x^2+y^2|^{1-\gamma}$.

[*Hint*: Both spectral measures are absolutely continuous with densities $g_1(\omega_1,\omega_2) = [\frac{1}{2}\gamma(\gamma-1)]^2|\omega_1\omega_2|^{\gamma-2}$, $g_2(\omega_1,\omega_2) = \pi^{\gamma-2}/[\Gamma(\gamma-1)|\omega_1^2+\omega_2^2|^2]$, respectively. Thornett (1979) has formulae for similar p.p.d. measures in $\mathbb{R}^d$.]

8.6.8 *Translation-boundedness characterization.* A nonnegative Borel measure μ on $\mathcal{B}(\mathbb{R}^d)$ satisfies

$$\int_{\mathbb{R}^d} |\tilde{I}_A(\omega)|^2\,\mu(\mathrm{d}\omega) < \infty$$

for all bounded $A \in \mathcal{B}(\mathbb{R}^d)$, if and only if the measure μ is translation-bounded.
[*Hint*: Establish a converse to Lemma 8.6.V of the form

$$\int |\tilde{f}(\omega)|^2\,\mu(\mathrm{d}\omega) \le K^2 \sup_{x\in A}|f(x)|^2,$$

where f, with Fourier transform $\tilde{f}$, is any bounded measurable function vanishing outside the bounded Borel set A and K is an absolute constant that may depend only on μ. See Robertson and Thornett (1984) for further details. Other results and references for such measures, but on locally compact Abelian groups, are given in Bloom (1984).]

8.6.9 Find the minimal p.p.d. measures corresponding to the Hawkes process with Bartlett spectrum (8.1.10).

APPENDIX 1

A Review of Some Basic Concepts of Topology and Measure Theory

In this appendix, we summarize, mainly without proof, some standard results from topology and measure theory. The aims are to establish terminology and notation, to set out results needed at various stages in the text in some specific form for convenient reference, and to provide some brief perspectives on the development of the theory. For proofs and further details, the reader should refer, in particular, to Kingman and Taylor (1966, Chapters 1–6), whose development and terminology we have followed rather closely.

A1.1. Set Theory

A set A of a space $\mathcal{X}$ is a collection of *elements* or *points* of $\mathcal{X}$. When x is an element of the set A, we write $x \in A$ (x *belongs to* or *is included in* A). The set of points of $\mathcal{X}$ not included in A is the complement of A, written A^c. If A, B are two sets of points from $\mathcal{X}$, their *union*, written $A \cup B$, is the set of points in either A or B or both; their *symmetric difference*, written $A \triangle B$, is the set of points in A or B but *not* both. If every element of B is also an element of A, we say B is *included* in A ($B \subseteq A$) or A *contains* B ($A \supseteq B$). In this case, the *proper difference* of A and B, written either $A - B$ or $A \setminus B$, is the set of points of A but not B. More generally, we use $A - B$ for $A \cap B^c$, so $A - B = A \triangle B$ only when $A \supset B$.

The operations $\cap$ and $\triangle$ on subsets of $\mathcal{X}$ are commutative, associative and distributive. The class of all such subsets thus forms an *algebra* with respect to these operations, where $\emptyset$, the empty set, plays the role of identity for $\triangle$ and $\mathcal{X}$ the role of identity for $\cap$. The special relation $A \cap A = A$ implies that the algebra is *Boolean*. More generally, any class of sets closed under the operations of $\cap$ and $\triangle$ is called a *ring*, or an *algebra* if $\mathcal{X}$ itself is a member of the class. A *semiring* is a class of sets $\mathcal{A}$ with the properties (i) $\mathcal{A}$ is closed under intersections and (ii) every symmetric difference of sets in $\mathcal{A}$ can be

represented as a finite union of disjoint sets in $\mathcal{A}$. The ring generated by an arbitrary family of sets $\mathcal{F}$ is the smallest ring containing $\mathcal{F}$ or, equivalently, the intersection of all rings containing $\mathcal{F}$. Every element in the ring generated by a semiring $\mathcal{A}$ can be represented as a union of disjoint sets of $\mathcal{A}$. If $\mathcal{R}$ is a finite ring, there exists a *basis* of disjoint elements of $\mathcal{R}$ such that every element in $\mathcal{R}$ can be represented uniquely as a union of disjoint elements of the basis.

The notions of union and intersection can be extended to arbitrary classes of sets. If $\{A_n : n = 1, 2, \ldots\}$ is a sequence of sets, write $A_n \uparrow A = \lim A_n$ if $A_n \subseteq A_{n+1}$ $(n = 1, 2, \ldots)$ and $A = \bigcup_{n=1}^{\infty} A_n$; similarly, if $A_n \supseteq A_{n+1}$, write $A_n \downarrow A = \lim A_n$ if $A = \bigcap_{n=1}^{\infty} A_n$. A *monotone class* is a class of sets closed under monotonically increasing sequences. A ring or algebra that is closed under *countable* unions is called a σ-ring or σ-algebra, respectively. The σ-ring generated by a class of sets $\mathcal{C}$, written $\sigma(\mathcal{C})$, is the smallest σ-ring containing $\mathcal{C}$. A σ-ring is *countably generated* if it can be generated by a countable class of $\mathcal{C}$. The following result, linking σ-rings to monotone classes, is useful in identifying the σ-ring generated by certain classes of sets.

Proposition A1.1.I (Monotone Class Theorem)**.** *If $\mathcal{R}$ is a ring and $\mathcal{C}$ is a monotone class containing $\mathcal{R}$, then $\mathcal{C}$ contains $\sigma(\mathcal{R})$.*

A closely related result uses the concept of a *Dynkin system* $\mathcal{D}$ meaning
(i) $\mathcal{X} \in \mathcal{D}$;
(ii) $\mathcal{D}$ is closed under proper differences; and
(iii) $\mathcal{D}$ is closed under monotonically increasing limits.

Proposition A1.1.II (Dynkin System Theorem)**.** *If $\mathcal{S}$ is a class of sets closed under finite intersections, and $\mathcal{D}$ is a Dynkin system containing $\mathcal{S}$, then $\mathcal{D}$ contains $\sigma(\mathcal{S})$.*

A1.2. Topologies

A topology $\mathcal{U}$ on a space $\mathcal{X}$ is a class of subsets of $\mathcal{X}$ that is closed under arbitrary unions and finite intersections and that includes the empty set $\emptyset$ and the whole space $\mathcal{X}$; the members of $\mathcal{U}$ are *open sets*, while their complements are *closed sets*. The pair $(\mathcal{X}, \mathcal{U})$ is a *topological space*. The *closure* of an arbitrary set A from $\mathcal{X}$, written $\bar{A}$, is the smallest closed set (equivalently, the intersection of all closed sets) containing A. The *interior* of A, written A°, is the largest open set (equivalently, the union of all open sets) contained within A. The *boundary* of A, written ∂A, is the difference $\bar{A} \setminus A^\circ$. The following elementary properties of boundaries are needed in the discussion of weak convergence of measures.

Proposition A1.2.I. (a) $\partial(A \cup B) \subseteq \partial A \cup \partial B$;
(b) $\partial(A \cap B) \subseteq \partial A \cup \partial B$;
(c) $\partial A^c = \partial A$.

A *neighbourhood* of the point $x \in \mathcal{X}$ with respect to the topology $\mathcal{U}$ (or, more briefly, a *$\mathcal{U}$-neighbourhood* of x) is an open set from $\mathcal{U}$ containing x. $\mathcal{U}$ is a *Hausdorff* or T_2-topology if the open sets separate points; that is, if for $x \neq y$, x and y possess disjoint neighbourhoods. A family of sets $\mathcal{F}$ forms a *basis* for the topology $\mathcal{U}$ if every $U \in \mathcal{U}$ can be represented as a union of sets in $\mathcal{F}$ and $\mathcal{F} \subseteq \mathcal{U}$. $\mathcal{U}$ is then said to be *generated by* $\mathcal{F}$. $\mathcal{U}$ is *second countable* if it has a *countable basis*.

A sufficient condition for a family of sets to form a basis for some topology is that, if $F_1 \in \mathcal{F}$, $F_2 \in \mathcal{F}$ and $x \in F_1 \cap F_2$, then there exists $F_3 \in \mathcal{F}$ such that $x \in F_3 \subseteq F_1 \cap F_2$. The topology generated by $\mathcal{F}$ is then uniquely defined and consists of all unions of sets in $\mathcal{F}$. Two bases $\mathcal{F}$ and $\mathcal{G}$, say, are *equivalent* if they generate the same topology. A necessary and sufficient condition for $\mathcal{F}$ and $\mathcal{G}$ to be equivalent is that for each $F \in \mathcal{F}$ and $x \in F$, there exists $G \in \mathcal{G}$ with $x \in G \subseteq F$, and similarly for each $G \in \mathcal{G}$ and $y \in G$, there exists $F \in \mathcal{F}$ such that $y \in F \subseteq G$.

Given a topology $\mathcal{U}$ on $\mathcal{X}$, a notion of convergence of sequences (or more generally nets, but we do not need the latter concept) can be introduced by saying $x_n \to x$ in the topology $\mathcal{U}$ if, given any $\mathcal{U}$-neighbourhood of x, U_x, there exists an integer N (depending on the neighbourhood in general) such that $x_n \in U_x$ for $n \geq N$. Conversely, nearly all the important types of convergence can be described in terms of a suitable topology. In this book, the overwhelming emphasis is on *metric topologies*, where the open sets are defined in terms of a *metric* or *distance function* $\rho(\cdot)$ that satisfies the conditions, for arbitrary $x, y, z \in \mathcal{X}$,

(i) $\rho(x, y) = \rho(y, x)$;
(ii) $\rho(x, y) \geq 0$ and $\rho(x, y) = 0$ if and only if $x = y$; and
(iii) (triangle inequality) $\rho(x, y) + \rho(y, z) \geq \rho(x, z)$.

With respect to a given distance function ρ, the *open sphere* $S_\epsilon(x)$ is the set $\{y\colon \rho(x, y) < \epsilon\}$, being defined for any $\epsilon > 0$. For any set A, define its diameter by

$$\operatorname{diam} A = 2 \inf \{r\colon S_r(x) \supseteq A \text{ for some } x\}.$$

The *metric topology* generated by ρ is the smallest topology containing the open spheres; it is necessarily Hausdorff. A set is open in this topology if and only if every point in the set can be enclosed by an open sphere lying wholly within the set. A sequence of points $\{x_n\}$ converges to x in this topology if and only if $\rho(x_n, x) \to 0$. A *limit point* y of a set A is a limit of a sequence of points $x_n \in A$ with $x_n \neq y$; y need not necessarily be in A. The closure of A in the metric topology is the union of A and its limit points. A space $\mathcal{X}$ with topology $\mathcal{U}$ is *metrizable* if a distance function ρ can be found such that $\mathcal{U}$ is equivalent to the metric topology generated by ρ. Two metrics on the same space $\mathcal{X}$ are *equivalent* if they each generate the same topology on $\mathcal{X}$.

A sequence of points $\{x_n\colon n \geq 1\}$ in a metric space is a *Cauchy sequence* if $\rho(x_n, x_m) \to 0$ as $n, m \to \infty$. The space is *complete* if every Cauchy sequence has a limit; i.e. if for every Cauchy sequence $\{x_n\}$ there exists $x \in \mathcal{X}$ such

that $\rho(x_n, x) \to 0$. A set $\mathcal{D}$ is *dense* in $\mathcal{X}$ if, for every $\epsilon > 0$, every point in $\mathcal{X}$ can be approximated by points in $\mathcal{D}$; i.e. given $x \in \mathcal{X}$, there exists $d \in \mathcal{D}$ such that $\rho(x, d) < \epsilon$. The space $\mathcal{X}$ is *separable* if there exists a *countable dense set*, also called a *separability set*. If $\mathcal{X}$ is a separable metric space, the spheres with rational radii and centres on a countable dense set form a countable base for the topology.

Given two topological spaces $(\mathcal{X}_1, \mathcal{U}_1)$ and $(\mathcal{X}_2, \mathcal{U}_2)$, a mapping $f(\cdot)$ from $(\mathcal{X}_1, \mathcal{U}_1)$ to $(\mathcal{X}_2, \mathcal{U}_2)$ is *continuous* if the inverse image $f^{-1}(U)$ of every open set $U \in \mathcal{U}_2$ is an open set in $\mathcal{U}_1$. If both spaces are metric spaces, the mapping is continuous if and only if for every $x \in \mathcal{X}_1$ and every $\epsilon > 0$, there exists $\delta > 0$ such that $\rho_2(f(x'), f(x)) < \epsilon$ whenever $\rho_1(x', x) < \delta$, where ρ_i is the metric in $\mathcal{X}_i$ for $i = 1, 2$; we can express this more loosely as $f(x') \to f(x)$ whenever $x' \to x$. A *homeomorphism* is a one-to-one continuous-both-ways mapping between two topological spaces. A famous theorem of Urysohn asserts that any *complete separable metric space* (*c.s.m.s.*) can be mapped homeomorphically into a countable product of unit intervals. A *Polish space* is a space that can be mapped homeomorphically into an open subset of a c.s.m.s. The theory developed in Appendix 2 can be carried through for an arbitrary Polish space with only minor changes, but we do not seek this greater generality.

A set K in a topological space $(\mathcal{X}, \mathcal{U})$ is *compact* if every covering of K by a family of open sets contains a finite subcovering; i.e. $K \subseteq \bigcup_\alpha U_\alpha$, $U_\alpha \in \mathcal{U}$, implies the existence of $N < \infty$ and $\alpha_1, \ldots, \alpha_N$ such that $K \subseteq \bigcup_{i=1}^N U_{\alpha_i}$. It is *relatively compact* if its closure $\overline{K}$ is compact. In a separable space, every open covering contains a countable subcovering, and consequently it is sufficient to check the compactness property for sequences of open sets rather than general families. More generally, for a c.s.m.s., the following important characterizations of compact sets are equivalent.

Proposition A1.2.II (Metric Compactness Theorem)**.** *Let $\mathcal{X}$ be a c.s.m.s. Then, the following properties of a subset K of $\mathcal{X}$ are equivalent and each is equivalent to the compactness of K.*

(i) (Heine–Borel property) *Every countable open covering of K contains a finite subcovering.*

(ii) (Bolzano–Weierstrass property) *Every infinite sequence of points in K contains a convergent subsequence with its limit in K.*

(iii) (Total boundedness and closure) *K is closed, and for every $\epsilon > 0$, K can be covered by a finite number of spheres of radius ϵ.*

(iv) *Every sequence $\{F_n\}$ of closed subsets of K with nonempty finite intersections* (*i.e.* $\bigcap_{n=1}^N F_n \neq \emptyset$ *for* $N < \infty$, *the* finite intersection property) *has nonempty total intersection* (*i.e.* $\bigcap_{n=1}^\infty F_n \neq \emptyset$).

The space $\mathcal{X}$ itself is compact if the compactness criterion applies with $\mathcal{X}$ in place of K. It is *locally compact* if every point of $\mathcal{X}$ has a neighbourhood with compact closure. A space with a locally compact second countable topology

is always metrizable. In a c.s.m.s., local compactness implies *σ-compactness*: the whole space can be represented as a countable union of compact sets (take the compact closures of the neighbourhoods of any countable dense set). Any finite-dimensional Euclidean space is σ-compact, but the same does not apply to infinite-dimensional spaces such as $C[0,1]$ or the infinite-dimensional Hilbert space ℓ_2.

A useful corollary of Proposition A1.2.II is that any closed subset F of a compact set in a complete metric space is again compact, for by (ii) any infinite sequence of points of F has a limit point in K, and by closure the limit point is also in F; hence, F is compact.

A1.3. Finitely and Countably Additive Set Functions

Let $\mathcal{A}$ be a class of sets in $\mathcal{X}$, and $\xi(\cdot)$ a real- or complex-valued function defined on $\mathcal{A}$. $\xi(\cdot)$ is *finitely additive* on $\mathcal{A}$ if for finite families $\{A_1, \ldots, A_N\}$ of disjoint sets from $\mathcal{A}$, with their union also in $\mathcal{A}$, there holds

$$\xi\Big(\bigcup_{i=1}^{N} A_i\Big) = \sum_{i=1}^{N} \xi(A_i).$$

If a similar result holds for *sequences* of sets $\{A_i : i = 1, 2, \ldots\}$, then ξ is *countably additive* (equivalently, *σ-additive*) on $\mathcal{A}$. A countably additive set function on $\mathcal{A}$ is a *measure* if it is nonnegative; a *signed measure* if it is real-valued but not necessarily nonnegative; and a *complex measure* if it is not necessarily real-valued.

A *determining class* for a particular type of set function is a class of sets with the property that if two set functions of the given type agree on the determining class, then they coincide. In this case, we can say that the set function is determined by its values on the determining class in question. The following proposition gives two simple results on determining classes. The first is a consequence of the representation of any element in a ring of sets as a disjoint union of the sets in any generating semiring; the second can be proved using a monotone class argument and the continuity lemma A1.3.II immediately following.

Proposition A1.3.I. (a) *A finitely additive, real- or complex-valued set function defined on a ring $\mathcal{A}$ is determined by its values on any semiring generating $\mathcal{A}$.*

(b) *A countably additive real- or complex-valued set function defined on a σ-ring $\mathcal{S}$ is determined by its values on any ring generating $\mathcal{S}$.*

Proposition A1.3.II (Continuity Lemma)**.** *Let $\mu(\cdot)$ be a finite real- or complex-valued, finitely additive set function defined on a ring $\mathcal{A}$. Then, μ is countably additive on $\mathcal{A}$ if and only if for every decreasing sequence $\{A_n : n = 1, 2, \ldots\}$ of sets with $A_n \downarrow \emptyset$,*

$$\mu(A_n) \to 0.$$

So far, we have assumed that the set functions take finite values on all the sets for which they are defined. It is frequently convenient to allow a nonnegative set function to take the value $+\infty$; this leads to few ambiguities and simplifies many statements. We then say that a finitely additive set function $\xi(\cdot)$ defined on an algebra or σ-algebra $\mathcal{A}$ is *totally finite* if, for all unions of disjoint sets $A_1, \ldots, A_N$ in $\mathcal{A}$, there exists $M < \infty$ such that

$$\sum_{i=1}^{N} \big|\xi(A_i)\big| \le M.$$

In particular, a nonnegative, additive set function μ is totally finite if and only if $\mu(\mathcal{X}) < \infty$. A finitely additive set function is σ-finite if there exists a sequence of sets $\{A_n : n = 1, 2, \ldots\} \in \mathcal{A}$ such that $\mathcal{X} \subseteq \bigcup_{n=1}^{\infty} A_n$ and for each n the *restriction of* ξ *to* A_n, defined by the equation

$$\hat{\xi}(A) = \xi(A \cap A_n) \qquad (A \in \mathcal{A}),$$

is totally finite, a situation we describe more briefly by saying that ξ is totally finite on each A_n. The continuity lemma extends to σ-finite set functions with the proviso that we consider only sequences for which $|\mu(A_n)| < \infty$ for some $n < \infty$. (This simple condition, extending the validity of Proposition A1.3.II to σ-finite set functions, fails in the general case, however, and it is then better to refer to continuity from below.)

We state next the basic extension theorem used to establish the existence of measures on σ-rings. Note that it follows from Proposition A1.3.I that when such an extension exists, it must be unique.

Theorem A1.3.III (Extension Theorem)**.** *A finitely additive, nonnegative set function defined on a ring $\mathcal{R}$ can be extended to a measure on $\sigma(\mathcal{R})$ if and only if it is countably additive on $\mathcal{R}$.*

As an example of the use of the theorem, we cite the well-known result that a right-continuous monotonically increasing function $F(\cdot)$ on $\mathbb{R}$ can be used to define a measure on the Borel sets of $\mathbb{R}$ (the sets in the smallest σ-ring containing the intervals) through the following sequence of steps.

(i) Define a nonnegative set function on the semiring of half-open intervals $(a, b]$ by setting $\mu_F(a, b] = F(b) - F(a)$.

(ii) Extend μ_F by additivity to all sets in the ring generated by such intervals (this ring consists, in fact, of all finite disjoint unions of such half-open intervals).

(iii) Establish countable additivity on this ring by appealing to compactness properties of finite closed intervals.

(iv) Use the extension theorem to assert the existence of a measure extending the definition of μ_F to the σ-ring generated by the half-open intervals—that is, the Borel sets.

The intrusion of the topological notion of compactness into this otherwise measure-theoretic sequence is a reminder that in most applications there is a

close link between open and measurable sets. Generalizing the corresponding concept for the real line, the *Borel sets* in a topological space are the sets in the smallest σ-ring (necessarily a σ-algebra) $\mathcal{B}_{\mathcal{X}}$ containing the open sets. A *Borel measure* is any measure defined on the Borel sets. The properties of such measures when $\mathcal{X}$ is a c.s.m.s. are explored in Appendix 2.

Returning to the general discussion, we note that no simple generalization of the extension theorem is known for signed measures. However, there is an important result, that shows that in some respects the study of signed measures can always be reduced to the study of measures.

Theorem A1.3.IV (Jordan–Hahn Decomposition)**.** *Let ξ be a signed measure defined on a σ-algebra $\mathcal{S}$. Then, ξ can be written as the difference*

$$\xi = \xi^+ - \xi^-$$

of two measures ξ^+, ξ^- on $\mathcal{S}$, and $\mathcal{X}$ can be written as the union of two disjoint sets U^+, U^- in $\mathcal{S}$ such that, for all $E \in \mathcal{S}$,

$$\xi^+(E) = \xi(E \cap U^+) \qquad \text{and} \qquad \xi^-(E) = -\xi(E \cap U^-),$$

and hence in particular, $\xi^+(U^-) = \xi^-(U^+) = 0$.

The measures ξ^+ and ξ^- appearing in this theorem are called upper and lower variations of ξ, respectively. The total variation of ξ is their sum

$$V_\xi(A) = \xi^+(A) + \xi^-(A).$$

It is clear from Theorem A1.3.IV that

$$V_\xi(A) = \sup_{\mathbb{P}(A)} \sum_{i=1}^{n(\mathbb{P})} |\xi(A_i)|,$$

where the supremum is taken over all finite partitions $\mathbb{P}$ of A into disjoint measurable sets. Thus, ξ is totally bounded if and only if $V_\xi(\mathcal{X}) < \infty$. In this case, $V_\xi(A)$ acts as a norm on the space of totally bounded signed measures ξ on $\mathcal{S}$; it is referred to as the *variation norm* and sometimes written $V_\xi(\mathcal{X}) = \|\xi\|$.

A1.4. Measurable Functions and Integrals

A measurable space is a pair $(\mathcal{X}, \mathcal{F})$, where $\mathcal{X}$ is the space and $\mathcal{F}$ a σ-ring of sets defined on it. A mapping f from a measurable space $(\mathcal{X}, \mathcal{F})$ into a measurable space $(\mathcal{Y}, \mathcal{G})$ is *$\mathcal{G}$-measurable* (or *measurable* for short) if, for all $A \in \mathcal{G}$, $f^{-1}(A) \in \mathcal{F}$. Note that the inverse images in $\mathcal{X}$ of sets in $\mathcal{G}$ form a σ-ring $\mathcal{H} = f^{-1}(\mathcal{G})$, say, and the requirement for measurability is that $\mathcal{H} \subseteq \mathcal{F}$.

By specializing to the case where $\mathcal{Y}$ is the real line $\mathbb{R}$ with $\mathcal{G}$ the σ-algebra of Borel sets generated by the intervals, $\mathcal{B}_{\mathbb{R}}$, the criterion for measurability simplifies as follows.

Proposition A1.4.I. *A real-valued function $f\colon (\mathcal{X}, \mathcal{F}) \mapsto (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is Borel measurable if and only if the set $\{x\colon f(x) \le c\}$ is a set in $\mathcal{F}$ for every real c.*

The family of real-valued (Borel) measurable functions on a measurable space $(\mathcal{X}, \mathcal{F})$ has many striking properties. It is closed under the operations of addition, subtraction, multiplication, and (with due attention to zeros) division. Moreover, any monotone limit of measurable functions is measurable. If $\mathcal{X}$ is a topological space and $\mathcal{F}$ the Borel σ-field on $\mathcal{X}$, then every continuous function on $\mathcal{X}$ is measurable.

The next proposition provides an important approximation result for measurable functions. Here a *simple function* is a finite linear combination of indicator functions of measurable sets; that is, a function of the form

$$s(x) = \textstyle\sum_{k=1}^{N} c_k I_{A_k}(x),$$

where $c_1, \ldots, c_N$ are real and $A_1, \ldots, A_N$ are measurable sets.

Proposition A1.4.II. *A nonnegative function $f\colon (\mathcal{X}, \mathcal{F}) \mapsto (\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$ is measurable if and only if it can be represented as the limit of a monotonically increasing sequence of simple functions.*

Now let μ be a measure on $\mathcal{F}$. We call the triple $(\mathcal{X}, \mathcal{F}, \mu)$ a finite or σ-finite measure space according to whether μ has the corresponding property; in the special case of a *probability space*, when μ has total mass unity, the triple is more usually written $(\Omega, \mathcal{E}, \mathcal{P})$, where the sets of the σ-algebra $\mathcal{E}$ are interpreted as *events*, a measurable function on $(\Omega, \mathcal{E})$ is a *random variable*, and $\mathcal{P}$ is a *probability measure*.

We turn to the problem of defining an *integral* (or in the probability case an *expectation*) with respect to the measure μ. If $s = \sum_{k=1}^{N} c_k I_{A_k}$ is a nonnegative simple function, set

$$\int_{\mathcal{X}} s(x)\, \mu(\mathrm{d}x) = \int_{\mathcal{X}} s \, \mathrm{d}\mu = \sum_{k=1}^{N} c_k \mu(A_k),$$

where we allow $+\infty$ as a possible value of the integral. Next, for any nonnegative measurable function f and any sequence of simple functions $\{s_n\}$ approximating f from below, set

$$\int_{\mathcal{X}} f \, \mathrm{d}\mu = \lim_{n\to\infty} \int_{\mathcal{X}} s_n \, \mathrm{d}\mu$$

and prove that the limit is independent of the particular sequence of simple functions used. Finally, for any measurable function f, write

$$f_+(x) = \big(f(x)\big)^+ = \max\big(f(x), 0\big),$$
$$f_-(x) = f_+(x) - f(x),$$

and if $\int f_+ \, \mathrm{d}\mu$ and $\int f_- \, \mathrm{d}\mu$ are both finite (equivalently, $\int_{\mathcal{X}} |f| \, \mathrm{d}\mu$ is finite), say that f is *integrable* and then define, for any integrable function f,

$$\int_{\mathcal{X}} f \, \mathrm{d}\mu = \int_{\mathcal{X}} f_+ \, \mathrm{d}\mu - \int_{\mathcal{X}} f_- \, \mathrm{d}\mu.$$

The resulting *abstract Lebesgue integral* is well defined, additive, linear, order-preserving, and enjoys strikingly elegant continuity properties. These

last are set out in the theorem below, where we say $f_n \to f$ *μ-almost everywhere* (μ-a.e., or a.e. μ) if the (necessarily measurable) set $\{x: f_n(x) \not\to f(x)\}$ has μ-measure zero. In the probability case, we refer to *almost sure* (a.s.) rather than a.e. convergence.

Theorem A1.4.III (Lebesgue Convergence Theorems)**.** *The following results hold for a sequence of measurable functions $\{f_n: n = 1, 2, \ldots\}$ defined on the measure space $(\mathcal{X}, \mathcal{F}, \mu)$:*

(a) (Fatou's Lemma) *If $f_n \geq 0$,*

$$\int_{\mathcal{X}} \liminf_{n\to\infty} f_n(x)\, \mu(\mathrm{d}x) \leq \liminf_{n\to\infty} \int_{\mathcal{X}} f_n(x)\, \mu(\mathrm{d}x).$$

(b) (Monotone Convergence Theorem) *If $f_n \geq 0$ and $f_n \uparrow f$ μ-a.e., then f is measurable and*

$$\lim_{n\to\infty} \int_{\mathcal{X}} f_n \,\mathrm{d}\mu = \int_{\mathcal{X}} f \,\mathrm{d}\mu$$

in the sense that either both sides are finite, and then equal, or both are infinite.

(c) (Dominated Convergence Theorem) *If $|f_n(x)| \leq g(x)$ where $g(\cdot)$ is integrable, and $f_n \to f$ μ-a.e., then*

$$\lim_{n\to\infty} \int_{\mathcal{X}} f_n \,\mathrm{d}\mu = \int_{\mathcal{X}} f \,\mathrm{d}\mu.$$

If f is an integrable function, the *indefinite integral* of f over any measurable subset can be defined by

$$\xi_f(A) \stackrel{\text{def}}{=} \int_A f \,\mathrm{d}\mu \stackrel{\text{def}}{=} \int_{\mathcal{X}} I_A f \,\mathrm{d}\mu,$$

where I_A is the indicator function of A. It is clear that ξ_f is totally finite and finitely additive on $\mathcal{S}$. Moreover, it follows from the dominated convergence theorem that if $A_n \in \mathcal{S}$ and $A_n \downarrow \emptyset$, then $I_{A_n} f \to 0$ and hence $\xi_f(A_n) \to 0$. Thus, ξ_f is also countably additive; that is, a signed measure on $\mathcal{S}$. This raises the question of which signed measures can be represented as indefinite integrals with respect to a given μ. The essential feature is that the ξ-measure of a set should tend to zero with the μ-measure. More specifically, ξ is *absolutely continuous* with respect to μ whenever $\mu(A) = 0$ implies $\xi(A) = 0$; we then have the following theorem.

Theorem A1.4.IV (Radon–Nikodyn Theorem)**.** *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a σ-finite measure space and ξ a totally finite measure or signed measure on $\mathcal{F}$. Then, there exists a measurable integrable function f such that*

$$\xi(A) = \int_{\mathcal{A}} f(x)\, \mu(\mathrm{d}x) \qquad (\text{all } A \in \mathcal{F}) \tag{A1.4.1}$$

if and only if ξ is absolutely continuous with respect to μ; moreover, f is a.e. *uniquely determined by* (A1.4.1)*, in the sense that any two functions satisfying* (A1.4.1)*, for all $A \in \mathcal{F}$ must be equal* μ-a.e.

The function f appearing in (A1.4.1) is usually referred to as a *Radon–Nikodym derivative* of ξ with respect to μ, written $\mathrm{d}\xi/\mathrm{d}\mu$. Lemma A1.6.III below shows one way in which the Radon–Nikodym derivative can be expressed as a limiting ratio.

There is an obvious extension of Theorem A1.4.IV to the case where ξ is σ-finite; in this extension, (A1.4.1) holds for subsets A of any member of the denumerable family of measurable sets on which ξ is totally finite.

Finally, we consider the relation between a fixed σ-finite measure μ and an arbitrary σ-finite signed measure ξ. ξ is said to be *singular* with respect to μ if there is a set E in $\mathcal{F}$ such that $\mu(E) = 0$ and for all $A \in \mathcal{F}$, $\xi(A) = \xi(E \cap A)$ so that also $\mu(E^c) = 0$ and $\mu(A) = \mu(A \cap E^c)$. We then have the following theorem.

Theorem A1.4.V (Lebesgue Decomposition Theorem)**.** *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a σ-finite measure space and $\xi(\cdot)$ a finite or σ-finite signed measure on $\mathcal{F}$. Then, there exists a unique decomposition of ξ,*

$$\xi = \xi_{\mathrm{s}} + \xi_{\mathrm{ac}},$$

into components that are, respectively, singular and absolutely continuous with respect to μ.

A1.5. Product Spaces

If $\mathcal{X}$, $\mathcal{Y}$ are two spaces, the *Cartesian product* $\mathcal{X} \times \mathcal{Y}$ is the set of ordered pairs $\{(x, y)\colon x \in \mathcal{X},\ y \in \mathcal{Y}\}$. If $\mathcal{X}$ and $\mathcal{Y}$ are either topological or measure spaces, there is a natural way of combining the original structures to produce a structure in the product space. Consider first the topological case. If U, V are neighbourhoods of the points $x \in \mathcal{X}$, $y \in \mathcal{Y}$ with respect to topologies $\mathcal{U}$, $\mathcal{V}$, define a neighbourhood of the pair (x, y) as the product set $U \times V$. The class of product sets of this kind is closed under finite intersections because

$$(U \times V) \cap (A \times B) = (U \cap A) \times (V \cap B).$$

It can therefore be taken as the basis of a topology in $\mathcal{X} \times \mathcal{Y}$; it is called the *product topology* and denoted $\mathcal{X} \otimes \mathcal{Y}$ [we follow e.g. Brémaud (1981) in using a distinctive product sign as a reminder that the product entity here is generated by the elements of the factors]. Most properties enjoyed by the component (or coordinate) topologies are passed on to the product topology. In particular, if $\mathcal{X}$, $\mathcal{Y}$ are both c.s.m.s.s, then $\mathcal{X} \times \mathcal{Y}$ is also a c.s.m.s. with respect to any one of a number of equivalent metrics, of which perhaps the simplest is

$$\rho((x, y), (u, v)) = \max(\rho_{\mathcal{X}}(x, u), \rho_{\mathcal{Y}}(y, v)).$$

More generally, if $\{\mathcal{X}_t\colon t \in \mathcal{T}\}$ is a family of spaces, the Cartesian product

$$\mathcal{X} = \bigtimes_{t \in \mathcal{T}} (\mathcal{X}_t)$$

may be defined as the set of all functions $x\colon \mathcal{T} \mapsto \bigcup_t \mathcal{X}_t$ such that $x(t) \in \mathcal{X}_t$.

A *cylinder set* in this space is a set in which restrictions are placed on a finite subset of the coordinates, on $x(t_1), \ldots, x(t_N)$, say, the values of the other coordinates being unrestricted in their appropriate spaces. A family of basic open sets in $\mathcal{X}$ can be defined by choosing open sets $\{U_t \subseteq \mathcal{X}_{t_i},\, i = 1, \ldots, N\}$ and requiring $x(t_i) \in U_i$, $i = 1, \ldots, N$. The topology generated by the class of cylinder sets of this form is called the *product topology* in $\mathcal{X}$. A remarkable property of this topology is that if the coordinate spaces $\mathcal{X}_t$ are individually compact in their respective topologies, then $\mathcal{X}$ is compact in the product topology. On the other hand, if the individual $\mathcal{X}_t$ are metric spaces, there are again many ways in which $\mathcal{X}$ can be made into a metric space [e.g. by using the supremum of the distances $\rho_t(x(t), y(t))$], but the topologies they generate are not in general equivalent among themselves nor to the product topology defined earlier.

Turning now to the measure context, let $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{G}, \nu)$ be two measure spaces. The product σ-ring $\mathcal{F} \otimes \mathcal{G}$ is the σ-ring generated by the semiring of measurable rectangles $A \times B$ with $A \in \mathcal{F}$, $B \in \mathcal{G}$. The *product measure* $\mu \times \nu$ is the extension to the σ-ring of the countably additive set function defined on such rectangles by

$$(\mu \times \nu)(A \times B) = \mu(A)\, \nu(B)$$

and extended by additivity to the ring of all finite disjoint unions of such rectangles. If μ, ν are both finite, then so is $\mu \times \nu$; similarly, if μ, ν are σ-finite, so is $\mu \times \nu$. The *product measurable space* is the space $(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$, and the *product measure space* is the space $(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G}, \mu \times \nu)$. All the definitions extend easily to the products of finite families of measure spaces. In the probability context, they form the natural framework for the discussion of *independence*. In the context of integration theory, the most important results pertain to the evaluation of *double integrals*, the question we take up next.

Let $\mathcal{H} = \mathcal{F} \otimes \mathcal{G}$ and $\pi = \mu \times \nu$. If C is $\mathcal{H}$-measurable, its *sections*

$$C_x = \{y\colon (x, y) \in C\}, \qquad C^y = \{x\colon (x, y) \in C\}$$

are, respectively, $\mathcal{G}$-measurable for each fixed x and $\mathcal{F}$-measurable for each fixed y. (The converse to this result, that a set whose sections are measurable is $\mathcal{H}$-measurable, is false, however.) Similarly, if $f(x, y)$ is $\mathcal{H}$-measurable, then regarded as a function of y, it is $\mathcal{G}$-measurable for each fixed x, and regarded as a function of x, it is $\mathcal{F}$-measurable for each fixed y. Introducing integrals with respect to μ, ν, write

$$s(x) = \begin{cases} \int_{\mathcal{Y}} f(x, y)\, \nu(\mathrm{d}y) & \text{if the integrand is } \nu\text{-integrable,} \\ +\infty & \text{otherwise;} \end{cases}$$

$$t(y) = \begin{cases} \int_{\mathcal{X}} f(x, y)\, \mu(\mathrm{d}x) & \text{if the integrand is } \mu\text{-integrable,} \\ +\infty & \text{otherwise.} \end{cases}$$

We then have the following theorem.

A1.5.I (Fubini's Theorem)**.** *Let $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{G}, \nu)$ be σ-finite measure spaces, and let $(\mathcal{Z}, \mathcal{H}, \pi)$ denote the product measure space.*

(a) *If f is $\mathcal{H}$-measurable and π-integrable, then $s(x)$ is $\mathcal{F}$-measurable and μ-integrable, $t(y)$ is $\mathcal{G}$-measurable and ν-integrable, and*

$$\int_{\mathcal{Z}} f \, \mathrm{d}\pi = \int_{\mathcal{X}} s \, \mathrm{d}\mu = \int_{\mathcal{Y}} t \, \mathrm{d}\nu.$$

(b) *If f is $\mathcal{H}$-measurable and $f \geq 0$, it is necessary and sufficient for f to be π-integrable that either s be μ-integrable or t be ν-integrable.*

Not all the important measures on a product space are product measures; in the probability context, in particular, it is necessary to study general bivariate probability measures and their relations to the marginal and conditional measures they induce. Thus, if π is a probability measure on $(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$, we define the *marginal probability measures* $\pi_{\mathcal{X}}$ and $\pi_{\mathcal{Y}}$ to be the *projections* of π onto $(\mathcal{X}, \mathcal{F})$ and $(\mathcal{Y}, \mathcal{G})$, respectively; i.e. the measures defined by

$$\pi_{\mathcal{X}}(A) = \pi(A \times \mathcal{Y}) \qquad \text{and} \qquad \pi_{\mathcal{Y}}(B) = \pi(\mathcal{X} \times B).$$

We next investigate the possibility of writing a measure on the product space as an integral (or a mixture of conditional probabilities), say

$$\pi(A \times B) = \int_A Q(B \mid x) \, \pi_{\mathcal{X}}(\mathrm{d}x), \tag{A1.5.1}$$

where $Q(B \mid x)$ may be regarded as the conditional probability of observing the event B given the occurrence of x. Such a family is also known as a disintegration of π.

Proposition A1.5.II. *Given a family $\{Q(\cdot \mid x) \colon x \in \mathcal{X}\}$ of probability measures on $(\mathcal{Y}, \mathcal{G})$ and a probability measure $\pi_{\mathcal{X}}$ on $(\mathcal{X}, \mathcal{F})$, the necessary and sufficient condition that* (A1.5.1) *should define a probability measure on the product space $(\mathcal{Z}, \mathcal{H})$ is that, as a function of x, $Q(B \mid x)$ be $\mathcal{F}$-measurable for each fixed $B \in \mathcal{G}$. When this condition is satisfied, for every $\mathcal{H}$-measurable, nonnegative function $f(\cdot, \cdot)$,*

$$\int_{\mathcal{Z}} f \, \mathrm{d}\pi = \int_{\mathcal{X}} \pi_{\mathcal{X}}(\mathrm{d}x) \int_{\mathcal{Y}} f(x, y) \, Q(\mathrm{d}y \mid x). \tag{A1.5.2}$$

Indeed, the integral in (A1.5.1) is not defined unless $Q(B \mid \cdot)$ is $\mathcal{F}$-measurable. When it is, the right-hand side of (A1.5.2) can be extended to a finitely additive set function on the ring of finite unions of disjoint rectangle sets. Countable additivity and the extension to a measure for which (A1.5.2) holds then follow along standard lines using monotone approximation arguments.

The projection of π onto the space $(\mathcal{Y}, \mathcal{G})$, i.e. the measure defined by

$$\pi_{\mathcal{Y}}(B) = \int_{\mathcal{X}} Q(B \mid x)\, \pi_{\mathcal{X}}(\mathrm{d}x),$$

is known as the *mixture* of $Q(\cdot \mid x)$ with respect to $\pi_{\mathcal{X}}$.

The converse problem, of establishing the existence of a family of measures satisfying (A1.5.1) from a given measure and its marginal, is a special case of the problem of *regular conditional probabilities* (see e.g. Ash, 1972, Section 6.6). For any *fixed* $B \in \mathcal{G}$, $\pi(\cdot \times B)$ may be regarded as a measure on $(\mathcal{X}, \mathcal{F})$, that is clearly absolutely continuous with respect to the marginal $\pi_{\mathcal{X}}$. Hence, there exists a Radon–Nikodym derivative, $Q_R(B \mid x)$ say, that is $\mathcal{F}$-measurable, satisfies (A1.5.1), and should therefore be a candidate for the disintegration of π. The difficulty is that we can guarantee the behaviour of Q_R only for fixed sets B, and it is not clear whether, for x fixed and B varying, the family $Q_R(B \mid x)$ will have the additivity and continuity properties of a measure. If $\{A_1, \ldots, A_N\}$ is a *fixed* family of disjoint sets in $\mathcal{G}$ or if $\{B_n : n \geq 1\}$ is a fixed sequence in $\mathcal{G}$ with $B_n \downarrow \emptyset$, then it is not difficult to show that

$$Q_R\left(\bigcup_{i=1}^{N} A_i \,\middle|\, x\right) = \sum_{i=1}^{N} Q_R(A_i \mid x) \qquad \pi_{\mathcal{X}}\text{-a.e.},$$
$$Q_R(B_n \mid x) \to 0 \quad (n \to \infty) \qquad \pi_{\mathcal{X}}\text{-a.e.},$$

respectively, but because there are uncountably many such relations to be checked, it is not obvious that the exceptional sets of measure zero can be combined into a single such set. The problem, in fact, is formally identical to establishing the existence of random measures and is developed further in Chapter 9. The following result is a partial converse to Proposition A1.5.II.

Proposition A1.5.III (Existence of Regular Conditional Probabilities)**.** *Let $(\mathcal{Y}, \mathcal{G})$ be a c.s.m.s. with its associated σ-algebra of Borel sets, $(\mathcal{X}, \mathcal{F})$ an arbitrary measurable space, and π a probability measure on the product space $(\mathcal{Z}, \mathcal{H})$. Then, with $\pi_{\mathcal{X}}(A) = \pi(A \times \mathcal{Y})$ for all $A \in \mathcal{F}$, there exists a family of kernels $Q(B \mid x)$ such that*

(i) *$Q(\cdot \mid x)$ is a probability measure on $\mathcal{G}$ for each fixed $x \in \mathcal{X}$;*
(ii) *$Q(B \mid \cdot)$ is an $\mathcal{F}$-measurable function on $\mathcal{X}$ for each fixed $B \in \mathcal{G}$; and*
(iii) *$\pi(A \times B) = \int_A Q(B \mid x)\, \pi_{\mathcal{X}}(\mathrm{d}x)$ for all $A \in \mathcal{F}$ and $B \in \mathcal{B}$.*

We consider finally the product of a general family of measurable spaces, $\{(\mathcal{X}_T, \mathcal{F}_t) : t \in \mathcal{T}\}$, where $\mathcal{T}$ is an arbitrary (finite, countable, or uncountable) indexing set. Once again, the cylinder sets play a basic role. A *measurable cylinder set* in $\mathcal{X} = \times_{t \in \mathcal{T}}(\mathcal{X}_t)$ is a set of the form

$$C(t_1, \ldots, t_N;\, B_1, \ldots, B_N) = \{x(t) : x(t_i) \in B_i, i = 1, \ldots, N\},$$

where $B_i \in \mathcal{F}_{t_i}$ is measurable for each $i = 1, \ldots, N$. Such sets form a semiring, their finite disjoint unions form a ring, and the generated σ-ring we denote by

$$\mathcal{F}_\infty = \bigotimes_{t \in \mathcal{T}} \mathcal{F}_t.$$

This construction can be used to define a product measure on $\mathcal{F}_\infty$, but greater interest centres on the *extension problem*: given a system of measures $\pi_{(\sigma)}$ defined on finite subfamilies $\mathcal{F}_{(\sigma)} = \mathcal{F}_{t_1} \otimes \mathcal{F}_{t_2} \otimes \cdots \otimes \mathcal{F}_{t_N}$, where $(\sigma) = \{t_1, \ldots, t_N\}$ is a finite selection of indices from $\mathcal{T}$, when can they be extended to a measure on $\mathcal{F}_\infty$? It follows from the extension theorem A1.3.III that the necessary and sufficient condition for this to be possible is that the given measures must give rise to a countably additive set function on the ring generated by the measurable cylinder sets. As with the previous result, countable additivity cannot be established without some additional assumptions; again it is convenient to put these in topological form by requiring each of the $\mathcal{X}_t$ to be a c.s.m.s. Countable additivity then follows by a variant of the usual compactness argument, and the only remaining requirement is that the given measures should satisfy the obviously necessary consistency conditions stated in the theorem below.

Theorem A1.5.IV (Kolmogorov Extension Theorem)**.** *Let $\mathcal{T}$ be an arbitrary index set, and for $t \in \mathcal{T}$ suppose $(\mathcal{X}_t, \mathcal{F}_t)$ is a c.s.m.s. with its associated Borel σ-algebra. Suppose further that for each finite subfamily $(\sigma) = \{t_1, \ldots, t_N\}$ of indices from $\mathcal{T}$, there is given a probability measure $\pi_{(\sigma)}$ on $\mathcal{F}_{(\sigma)} = \mathcal{F}_{t_1} \otimes \cdots \otimes \mathcal{F}_{t_N}$. In order that there exist a measure π on $\mathcal{F}_\infty$ such that for all (σ), $\pi_{(\sigma)}$ is the projection of π onto $\mathcal{F}_{(\sigma)}$, it is necessary and sufficient that for all (σ), (σ_1), (σ_2),*

(i) *$\pi_{(\sigma)}$ depends only on the choice of indices in (σ), not on the order in which they are written down; and*

(ii) *if $(\sigma_1) \subseteq (\sigma_2)$, then $\pi_{(\sigma_1)}$ is the projection of $\pi_{(\sigma_2)}$ onto $\mathcal{F}_{(\sigma_1)}$.*

Written out more explicitly in terms of distribution functions, condition (i) becomes (in an obvious notation) the condition of *invariance under simultaneous permutations*: if $p_1, \ldots, p_N$ is a permutation of the integers $1, \ldots, N$, then

$$F^{(N)}_{t_1,\ldots,t_N}(x_1, \ldots, x_N) = F^{(N)}_{t_{p_1},\ldots,t_{p_N}}(x_{p_1}, \ldots, x_{p_N}).$$

Similarly, condition (ii) becomes the condition of *consistency of marginal distributions*, namely that

$$F^{(N+k)}_{t_1,\ldots,t_N,s_1,\ldots,s_k}(x_1, \ldots, x_N, \infty, \ldots, \infty) = F^{(N)}_{t_1,\ldots,t_N}(x_1, \ldots, x_N).$$

The measure π induced on $\mathcal{F}_\infty$ by the fidi distributions is called their projective limit. Clearly, if stochastic processes have the same fidi distributions, they must also have the same projective limit. Such processes may be described as being *equivalent* or *versions* of one another.

See Parthasarathy (1967, Sections 5.1–5) for discussion of Theorem A1.5.IV in a slightly more general form and for proof and further details.

A1.6. Dissecting Systems and Atomic Measures

The notion of a *dissecting system* in Definition A1.6.I depends only on topological ideas of separation and distinguishing one point from another by means of distinct sets, though we use it mainly in the context of a metric space where its development is simpler.

If $(\mathcal{X},\mathcal{U})$ is a topological space, the smallest σ-algebra containing the open sets is called the *Borel σ-algebra*. If $f\colon \mathcal{X} \mapsto \mathbb{R}$ is any real-valued continuous function, then the set $\{x\colon f(x) < c\}$ is open in $\mathcal{U}$ and hence measurable. It follows that f is measurable. Thus, every continuous function is measurable with respect to the Borel σ-algebra.

Definition A1.6.I (Dissecting System)**.** *The sequence $\mathcal{T} = \{\mathcal{T}_n\}$ of finite partitions $\mathcal{T}_n = \{A_{ni}\colon i = 1,\dots,k_n\}$ $(n = 1,2,\dots)$ consisting of Borel sets in the space $\mathcal{X}$ is a* dissecting system *for $\mathcal{X}$ when*

(i) (partition properties) $A_{ni} \cap A_{nj} = \emptyset$ *for* $i \neq j$ *and* $A_{n1} \cup \cdots \cup A_{nk_n} = \mathcal{X}$;
(ii) (nesting property) $A_{n-1,i} \cap A_{nj} = A_{nj}$ *or* $\emptyset$; *and*
(iii) (point-separating property) *given distinct x, $y \in \mathcal{X}$, there exists an integer $n = n(x,y)$ such that $x \in A_{ni}$ implies $y \notin A_{ni}$.*

Given a dissecting system $\mathcal{T}$ for $\mathcal{X}$, properties (i) and (ii) of Definition A1.6.I imply that there is a well-defined nested sequence $\{T_n(x)\} \subset \mathcal{T}$ such that

$$\bigcap_{n=1}^{\infty} T_n(x) = \{x\}, \qquad \text{so} \quad \mu(T_n(x)) \to \mu\{x\} \quad (n \to \infty)$$

because μ is a measure and $\{T_n(x)\}$ is a monotone sequence. Call $x \in \mathcal{X}$ an *atom* of μ if $\mu(\{x\}) \equiv \mu\{x\} > 0$. It follows that x is an atom of μ if and only if $\mu(T_n(x)) > \epsilon$ (all n) for some $\epsilon > 0$; indeed, any ϵ in $0 < \epsilon \leq \mu\{x\}$ will do. We use $\delta_x(\cdot)$ to denote *Dirac measure* at x, being defined on Borel sets A by

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

More generally, an atom of a measure μ on a measurable space $(\mathcal{X},\mathcal{F})$ is any nonempty set $F \in \mathcal{F}$ such that if $G \in \mathcal{F}$ and $G \subseteq F$, then either $G = \emptyset$ or $G = F$. However, when $\mathcal{X}$ is a separable metric space, it is a consequence of Proposition A2.1.IV below that the only possible atoms of a measure μ on $(\mathcal{X},\mathcal{F})$ are singleton sets.

A measure with only atoms is *purely atomic*; a *diffuse* measure has no atoms.

Given $\epsilon > 0$, we can identify all atoms of μ of mass $\mu\{x\} \geq \epsilon$, and then using a sequence $\{\epsilon_j\}$ with $\epsilon_j \downarrow 0$ as $j \to \infty$, all atoms of μ can be identified. Because μ is σ-finite, it can have at most countably many atoms, so identifying them as $\{x_j\colon j = 1,2,\dots\}$, say, and writing $b_j = \mu\{x_j\}$, the measure

$$\mu_{\mathrm{a}}(\cdot) \equiv \sum_{j=1}^{\infty} b_j \delta_{x_j}(\cdot),$$

which clearly consists only of atoms, is the *atomic component* of the measure μ. The measure

$$\mu_{\mathrm{d}}(\cdot) \equiv \mu(\cdot) - \mu_{\mathrm{a}}(\cdot) = \mu(\cdot) - \sum_{j=1}^{\infty} b_j \delta_{x_j}(\cdot)$$

has no atoms and is the *diffuse component* of μ. Thus, any measure μ as above has a unique decomposition into atomic and diffuse components.

Lemma A1.6.II. *Let μ be a nonatomic measure and $\{\mathcal{T}_n\}$ a dissecting system for a set A with $\mu(A) < \infty$. Then $\epsilon_n \equiv \sup_i \mu(A_{ni}) \to 0$ as $n \to \infty$.*

Proof. Suppose not. Then there exists $\delta > 0$ and, for each n, some set A_{n,i_n}, say, with $A_{n,i_n} \in \mathcal{T}_n$ and $\mu(A_{n,i_n}) > \delta$. Because $\mathcal{T}_n$ is a dissecting system, the nesting implies that there exists $A_{n-1,i'_{n-1}} \in \mathcal{T}_{n-1}$ and contains A_{n,i_n}, so $\mu(A_{n-1,i'_{n-1}}) > \delta$. Consequently, we can assume there exists a *nested* sequence of sets A_{n,i_n} for which $\mu(A_{n,i_n}) > \delta$, and hence

$$\delta \le \lim_n \mu(A_{n,i_n}) = \mu(\lim_n A_{n,i_n}),$$

equality holding here because μ is a measure and $\{A_{n,i_n}\}$ is monotone. But, because $\mathcal{T}_n$ is a dissecting system, $\lim_n A_{n,i_n}$ is either empty or a singleton set, $\{x'\}$ say. Thus, the right-hand side is either $\mu(\emptyset) = 0$ or $\mu(\{x\}) = 0$ because μ is nonatomic (i.e. $\delta \le 0$), which is a contradiction. □

Dissecting systems can be used to construct approximations to Radon–Nikodym derivatives as follows (e.g. Chung, 1974, Chapter 9.5, Example VIII).

Lemma A1.6.III (Approximation of Radon–Nikodym Derivative)**.** *Let $\mathcal{T} = \{\mathcal{T}_n\} = \{\{A_{ni}: i = 1, \ldots, k_n\}\}$ be a nested family of measurable partitions of the measure space $(\Omega, \mathcal{E}, \mu)$, generating $\mathcal{E}$ and let ν be a measure absolutely continuous with respect to μ, with Radon–Nikodym derivative $\mathrm{d}\nu/\mathrm{d}\mu$. Define*

$$\lambda_n(\omega) = \sum_{i=1}^{k_n} I_{A_{ni}}(\omega) \frac{\nu(A_{ni})}{\mu(A_{ni})} \qquad (\omega \in \Omega).$$

Then, as $n \to \infty$, $\lambda_n \to \dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}$, μ-a.e. and in $L_1(\mu)$ norm.

As a final result involving dissecting systems, given two probability measures P and P_0 on $(\Omega, \mathcal{E})$, define the relative entropy of the restriction of P and P_0 to a partition $\mathcal{T} = \{A_i\}$ of $(\Omega, \mathcal{E})$ by

$$H(P; P_0) = \sum_i P(A_i) \log \frac{P(A_i)}{P_0(A_i)}\,.$$

Additivity of measures, convexity of $x \log x$ on $x > 0$, and the inequality $(a_1 + a_2)/(b_1 + b_2) \le a_1/b_1 + a_2/b_2$, valid for nonnegative a_r and positive b_r, $r = 1, 2$, establishes the result below.

Lemma A1.6.IV. *Let $\mathcal{T}_1, \mathcal{T}_2$ be measurable partitions of $(\Omega, \mathcal{E})$ with $\mathcal{T}_1 \subseteq \mathcal{T}_2$ and P, P_0 two probability measures on $(\Omega, \mathcal{E})$. Then, the relative entropies of the restrictions of P, P_0 to $\mathcal{T}_r$ satisfy $H_1(P; P_0) \le H_2(P; P_0)$.*

APPENDIX 2

Measures on Metric Spaces

A2.1. Borel Sets and the Support of Measures

If $(\mathcal{X},\mathcal{U})$ is a topological space, the smallest σ-algebra containing the open sets is called the *Borel σ-algebra*. If $f\colon \mathcal{X} \mapsto \mathbb{R}$ is any real-valued continuous function, then the set $\{x\colon f(x) < c\}$ is open in $\mathcal{U}$ and hence measurable. It follows that f is measurable. Thus, every continuous function is measurable with respect to the Borel σ-algebra.

It is necessary to clarify the relation between the Borel sets and various other candidates for useful σ-algebras that suggest themselves, such as

(a) the *Baire sets*, belonging to the smallest σ-field with respect to which the continuous functions are measurable;

(b) the *Borelian sets*, generated by the compact sets in $\mathcal{X}$; and

(c) if $\mathcal{X}$ is a metric space, the σ-algebra generated by the open spheres.

We show that, with a minor reservation concerning (b), all three concepts coincide when $\mathcal{X}$ is a c.s.m.s. More precisely, we have the following result.

Proposition A2.1.I. *Let $\mathcal{X}$ be a metric space and $\mathcal{U}$ the topology induced by the metric. Then*

(i) *the Baire sets and the Borel sets coincide;*

(ii) *if $\mathcal{X}$ is separable, then the Borel σ-algebra is the smallest σ-algebra containing the open spheres;*

(iii) *a Borel set is Borelian if and only if it is σ-compact; that is, if it can be covered by a countable union of compact sets. In particular, the Borel sets and the Borelian sets coincide if and only if the whole space is σ-compact.*

PROOF. Part (i) depends on Lemma A2.1.II below, of interest in its own right; (ii) depends on the fact that when $\mathcal{X}$ is separable, every open set can be represented as a countable union of open spheres; (iii) follows from the fact that all closed subsets of a compact set are compact and hence Borelian. □

Lemma A2.1.II. *Let F be a closed set in the metric space $\mathcal{X}$, U an open set containing F, and $I_F(\cdot)$ the indicator function of F. Then, there exists a sequence of continuous functions $\{f_n(x)\}$ such that*

(i) $0 \le f_n(x) \le 1$ $(x \in \mathcal{X})$;
(ii) $f_n(x) = 0$ *outside* U;
(iii) $f_n(x) \downarrow I_F(x)$ *as* $n \to \infty$.

PROOF. Let $f_n(x) = \rho(x, U^c)/[\rho(x, U^c) + 2^n \rho(x, F)]$, where for any set C

$$\rho(x, C) = \inf_{y \in C} \rho(x, y).$$

Then, the sequence $\{f_n(x)\}$ has the required properties. □

It is clear that in a separable metric space the Borel sets are countably generated. Lemma A2.1.III exhibits a simple example of a countable semiring of open sets generating the Borel sets.

Lemma A2.1.III. *Let $\mathcal{X}$ be a c.s.m.s., $\mathcal{D}$ a countable dense set in $\mathcal{X}$, and $\mathcal{S}_0$ the class of all finite intersections of open spheres $S_r(d)$ with centres $d \in \mathcal{D}$ and rational radii. Then*

(i) *$\mathcal{S}_0$ and the ring $\mathcal{A}_0$ generated by $\mathcal{S}_0$ are countable; and*
(ii) *$\mathcal{S}_0$ generates the Borel σ-algebra in $\mathcal{X}$.*

It is also a property of the Borel sets in a separable metric space, and of considerable importance in the analysis of sample-path properties of point processes and random measures, that they include a *dissecting system* defined in Definition A1.6.I.

Proposition A2.1.IV. *Every separable metric space $\mathcal{X}$ contains a dissecting system.*

PROOF. Let $\{d_1, d_2, \ldots\} = \mathcal{D}$ be a separability set for $\mathcal{X}$ (i.e. $\mathcal{D}$ is a countable dense set in $\mathcal{X}$). Take any pair of distinct points $x, y \in \mathcal{X}$; their distance apart equals $2\delta \equiv \rho(x, y) > 0$. We can then find d_m, d_n in $\mathcal{D}$ such that $\rho(d_m, x) < \delta$, $\rho(d_n, y) < \delta$, so the spheres $S_\delta(d_m)$, $S_\delta(d_n)$, which are Borel sets, certainly separate x and y. We have essentially to embed such separating spheres into a sequence of sets covering the whole space.

For the next part of the proof, it is convenient to identify one particular element in each $\mathcal{T}_n$ (or it may possibly be a null set for all n sufficiently large) as A_{n0}; this entails no loss of generality.

Define the initial partition $\{A_{1i}\}$ by $A_{11} = S_1(d_1)$, $A_{10} = \mathcal{X} \setminus A_{11}$. Observe that $\mathcal{X}$ is covered by the countably infinite sequence $\{S_1(d_n)\}$, so the sequence

of sets $\{A'_{n0}\}$ defined by $A'_{n0} = \mathcal{X} \setminus \bigcup_{r=1}^{n} S_1(d_r)$ converges to the null set. For $n = 2, 3, \ldots$ and $i = 1, \ldots, n$, define

$$B_{ni} = S_{1/2^{n-i}}(d_i), \qquad B_{n0} = \left(\bigcup_{i=1}^{n} B_{ni} \right)^c,$$

so that $\{B_{ni} : i = 0, \ldots, n\}$ covers $\mathcal{X}$. By setting $C_{n0} = B_{n0}$, $C_{n1} = B_{n1}$, and $C_{ni} = B_{ni} \setminus \big(B_{n_1} \cup \cdots \cup B_{n,i-1}\big)$, it is clear that $\{C_{ni} : i = 0, 1, \ldots, n\}$ is a partition of $\mathcal{X}$. Let the family $\{A_{ni}\}$ consist of all nonempty intersections of the form $A_{n-1,j} \cap C_{nk}$, setting in particular $A_{n0} = A_{n-1,0} \cap C_{n0} = A'_{n0}$. Then $\big\{\{A_{ni}\} : n = 1, 2, \ldots\big\}$ clearly consists of nested partitions of $\mathcal{X}$ by Borel sets, and only the separation property has to be established.

Take distinct points x, $y \in \mathcal{X}$, and write $\delta = \rho(x, y)$ as before. Fix the integer $r \geq 0$ by $2^{-r} \leq \min(1, \delta) < 2^{-r+1}$, and locate a separability point d_m such that $\rho(d_m, x) < 2^{-r}$. Then $x \in S_{1/2^r}(d_m) = B_{m+r,m}$, and consequently $x \in C_{m+r,j}$ for some $j = 1, \ldots, m$. But by the triangle inequality, for any $z \in C_{m+r,j}$,

$$\rho(x, z) < 2 \quad \text{and} \quad 2^{-(m+r-j)} < 2\delta = \rho(x, y),$$

so the partition $\{C_{m+r,i}\}$, and hence also $\{A_{m+r,j}\}$, separates x and y. □

Trivially, if $\mathcal{T}$ is a dissecting system for $\mathcal{X}$, the nonempty sets of $\mathcal{T} \cap A$ (in an obvious notation) constitute a dissecting system for any $A \in \mathcal{B}_{\mathcal{X}}$. If A is also compact, the construction of a dissecting system for A is simplified by applying the Heine–Borel theorem to extract a finite covering of A from the countable covering

$$\{S_{2^{-n}}(d_r) : r = 1, 2, \ldots\}.$$

Definition A2.1.V. *The ring of sets generated by finitely many intersections and unions of elements of a dissecting system is a dissecting ring.*

A2.2. Regular and Tight Measures

In this section, we examine the extent to which values of a finitely or countably generated set function defined on some class of sets can be approximated by their values on either closed or compact sets.

Definition A2.2.I. (i) *A finite or countably additive, nonnegative set function μ defined on the Borel sets is* regular *if, given any Borel set A and $\epsilon > 0$, there exist open and closed sets G and F, respectively, such that $F \subseteq A \subseteq G$ and*

$$\mu(G - A) < \epsilon \qquad \text{and} \qquad \mu(A_F) < \epsilon.$$

(ii) *It is* compact regular *if, given any Borel set A and $\epsilon > 0$, there exists a compact set C such that $C \subseteq A$ and $\mu(A - C) < \epsilon$.*

We first establish the following.

Proposition A2.2.II. *If $\mathcal{X}$ is a metric space, then all totally finite measures on $\mathcal{B}_\mathcal{X}$ are regular.*

PROOF. Let μ be a totally finite, additive, nonnegative set function defined on $\mathcal{B}_\mathcal{X}$. Call any $A \in \mathcal{B}_\mathcal{X}$ *μ-regular* if $\mu(A)$ can be approximated by the values of μ on open and closed sets in the manner of Definition A2.2.I. The class of μ-regular sets is obviously closed under complementation. It then follows from the inclusion relations

$$\bigcup_\alpha G_\alpha - \bigcup_\alpha F_\alpha \subseteq \bigcup_\alpha (G_\alpha - F_\alpha) \tag{A2.2.1a}$$

and

$$\bigcap_\alpha G_\alpha - \bigcap_\alpha F_\alpha \subseteq \bigcup_\alpha \left(\bigcap_\alpha G_\alpha - F_\alpha \right) \subseteq \bigcup_\alpha (G_\alpha - F_\alpha) \tag{A2.2.1b}$$

that the class is an algebra if μ is finitely additive and a σ-algebra if μ is countably additive. In the latter case, the countable union $\bigcup_\alpha F_\alpha$ in (A2.2.1a) may not be closed, but we can approximate $\mu\big(\bigcup_\alpha F_\alpha\big)$ by $\mu\big(\bigcup_{i=1}^N F_{\alpha_i}\big)$ to obtain a set that is closed and has the required properties; similarly, in (A2.2.1b) we can approximate $\mu\big(\bigcap_\alpha G_\alpha\big)$ by $\mu\big(\bigcap_{i=1}^N A_{\alpha_i}\big)$. Moreover, if μ is σ-additive, the class also contains all closed sets, for if F is closed, the *halo sets*

$$F^\epsilon = \bigcup_{x \in F} S_\epsilon(x) = \{x : \rho(x, F) < \epsilon\} \tag{A2.2.2}$$

form, for a sequence of values of ϵ tending to zero, a family of open sets with the property $F^\epsilon \downarrow F$; hence, it follows from the continuity lemma A1.3.II that $\mu(F^\epsilon) \to \mu(F)$. In summary, if μ is countably additive, the μ-regular sets form a σ-algebra containing the closed sets, and therefore the class must coincide with the Borel sets themselves. □

Note that this proof does not require either completeness or separability.

Compact regularity is a corollary of this result and the notion of a *tight measure*.

Definition A2.2.III (Tightness)**.** *A finitely or countably additive set function μ is* tight *if, given $\epsilon > 0$, there exists a compact set K such that $\mu(\mathcal{X} - K)$ is defined and*

$$\mu(\mathcal{X} - K) < \epsilon.$$

Lemma A2.2.IV. *If $\mathcal{X}$ is a complete metric space, a Borel measure is compact regular if and only if it is tight.*

PROOF. Given any Borel set A, it follows from Proposition A2.2.II that there exists a closed set $C \subseteq A$ with $\mu(A - C) < \frac{1}{2}\epsilon$. If μ is tight, choose K so that $\mu(\mathcal{X} - K) < \frac{1}{2}\epsilon$. Then, the set $C \cap K$ is a closed subset of the compact set K and hence is itself compact; it also satisfies

$$\mu(A - C \cap K) \;\le\; \mu(A - C) + \mu(A - K) \;<\; \epsilon,$$

which establishes the compact regularity of μ. If, conversely, μ is compact regular, tightness follows on taking $\mathcal{X} = K$. □

Proposition A2.2.V. *If $\mathcal{X}$ is a c.s.m.s., every Borel measure μ is tight and hence compact regular.*

PROOF. Let $\mathcal{D}$ be a separability set for $\mathcal{X}$; then for fixed n, $\bigcup_{d\in\mathcal{D}} S_{1/n}(d) = \mathcal{X}$, and so by the continuity lemma A1.3.II, there is a finite set $d_1, \ldots, d_{k(n)}$ such that

$$\mu\left(\mathcal{X} - \bigcup_{i=1}^{k(n)} S_{1/n}(d_i)\right) < \frac{\epsilon}{2^n}.$$

Now consider $K = \bigcap_n \left(\bigcup_{i=1}^{k(n)} S_{1/n}(d_i)\right)$. It is not difficult to see that K is closed and totally bounded, and hence compact, by Proposition A1.2.II and that $\mu(\mathcal{X} - K) < \epsilon$. Hence, μ is tight. □

The results above establish compact regularity as a necessary condition for a finitely additive set function to be countably additive. The next proposition asserts its sufficiency. The method of proof provides a pattern that is used with minor variations at several important points in the further development of the theory.

Proposition A2.2.VI. *Let $\mathcal{A}$ be a ring of sets from the c.s.m.s. $\mathcal{X}$ and μ a finitely additive, nonnegative set function defined and finite on $\mathcal{A}$. A sufficient condition for μ to be countably additive on $\mathcal{A}$ is that, for every $A \in \mathcal{A}$ and $\epsilon > 0$, there exists a compact set $C \subseteq A$ such that $\mu(A - C) < \epsilon$.*

PROOF. Let $\{A_n\}$ be a decreasing sequence of sets in $\mathcal{A}$ with $A_n \downarrow \emptyset$; to establish countable additivity for μ, it is enough to show that $\mu(A_n) \to 0$ for every such sequence. Suppose to the contrary that $\mu(A_n) \geq \alpha > 0$. By assumption, there exists for each n a compact set C_n for which $C_n \subseteq A_n$ and $\mu(A_n - C_n) < \alpha/2^{n+1}$. By (A2.2.1),

$$A_n - \bigcap_k C_k \subseteq \bigcup_k (A_k - C_k).$$

Since $\mathcal{A}$ is a ring, every finite union $\bigcup_{k=1}^n (A_k - C_k)$ is an element of $\mathcal{A}$, so from the finite additivity of μ,

$$\mu\left(A_n - \bigcap_{k=1}^n C_k\right) \leq \sum_{k=1}^n \frac{\alpha}{2^{n+1}} < \tfrac{1}{2}\alpha.$$

Thus, the intersection $\bigcap_{k=1}^n C_k$ is nonempty for each n, and it follows from the finite intersection part of Proposition A1.2.II that $\bigcap_{k=1}^n C_k$ is nonempty. This gives us the required contradiction to the assumption $A_n \downarrow \emptyset$. □

Corollary A2.2.VII. *A finite, finitely additive, nonnegative set function defined on the Borel sets of $\mathcal{X}$ is countably additive if and only if it is compact regular.*

We can now prove an extension of Proposition A2.2.VI that plays an important role in developing the existence theorems of Chapter 9. It is based on the notion of a self-approximating ring and is a generalization of the concept of a covering ring given in Kallenberg (1975).

Definition A2.2.VIII (Self-Approximating Ring)**.** *A ring $\mathcal{A}$ of sets of the c.s.m.s. $\mathcal{X}$ is a* self-approximating ring *if, for every $A \in \mathcal{A}$ and $\epsilon > 0$, there exists a sequence of closed sets $\{F_k(A;\epsilon)\}$ such that*

(i) $F_k(A;\epsilon) \in \mathcal{A}$ $(k = 1, 2, \ldots)$;

(ii) *each set $F_k(A;\epsilon)$ is contained within a sphere of radius ϵ; and*

(iii) $\bigcup_{k=1}^{\infty} F_k(A;\epsilon) = A$.

Kallenberg uses the context where $\mathcal{X}$ is locally compact, in which case it is possible to require the covering to be finite so that the lemma below effectively reduces to Proposition A2.2.VI. The general version is based on an argument in Harris (1968). The point is that it allows checking for countable additivity to be reduced to a denumerable set of conditions.

Lemma A2.2.IX. *Let $\mathcal{A}$ be a self-approximating ring of subsets of the c.s.m.s. $\mathcal{X}$ and μ a finitely additive, nonnegative set function defined on $\mathcal{A}$. In order that μ have an extension as a measure on $\sigma(\mathcal{A})$, it is necessary and sufficient that for each $A \in \mathcal{A}$, using the notation of Definition* A2.2.VIII*,*

$$\lim_{m\to\infty} \mu\left(\bigcup_{i=1}^{m} F_i(A;\epsilon)\right) = \mu(A). \tag{A2.2.3}$$

Proof. Necessity follows from the continuity lemma. We establish sufficiency by contradiction: suppose that μ is finitely additive and satisfies (A2.2.3) but that μ cannot be extended to a measure on $\sigma(\mathcal{A})$. From the continuity lemma, it again follows that there exists $\alpha > 0$ and a sequence of sets $A_n \in \mathcal{A}$, with $A_n \downarrow \emptyset$, such that

$$\mu(A_n) \geq \alpha. \tag{A2.2.4}$$

For each k, use (A2.2.3) to choose a set $F_k = \bigcup_{i=1}^{m_k} F_i(A;k^{-1})$ that is closed, can be covered by a finite number of k^{-1} spheres, and satisfies

$$\mu(A_k - F_k) \leq \alpha/2^{n+1}.$$

From (A2.2.1), we have $A - \bigcap_{j=1}^{k} F_j \subseteq \bigcup_{j=1}^{k}(A_j - F_j)$, which, with the additivity of μ, implies that

$$\mu\left(\bigcap_{j=1}^{k} F_j\right) \geq \tfrac{1}{2}\alpha > 0.$$

Thus, the sets F_j have the finite intersection property.

To show that their complete intersection is nonempty, choose any $x_k \in \bigcap_{j=1}^{k} F_j$. Since F_1 can be covered by a finite number of 1-spheres, there exists a subsequence $\{x'_k\}$ that is wholly contained within a sphere of radius 1. Turning to F_2, we can select a further subsequence x''_k, which for $k \geq 2$ lies wholly within a sphere of radius $\frac{1}{2}$. Proceeding in this way by induction, we finally obtain by a diagonal selection argument a subsequence $\{x_{k_j}\}$ such that for $j \geq j_0$ all terms are contained within a sphere of radius $1/j_0$. This is enough to show that $\{x_{k_j}\}$ is a Cauchy sequence that, since $\mathcal{X}$ is complete, has a limit point $\bar{x}$, say. For each k, the x_j are in $\bigcap_{n=1}^{k} F_n$ for all sufficiently large j. Since the sets are closed, this implies that $\bar{x} \in F_k$ for every k. But this implies also that $\bar{x} \in A_k$ and hence $\bar{x} \in \bigcap_{k=1}^{\infty} A_k$, which contradicts the assumption that $A_n \downarrow \emptyset$. The contradiction shows that (A2.2.4) cannot hold and so completes the proof of the lemma. □

Let us observe finally that self-approximating rings do exist. A standard example, which is denumerable and generating as well as self-approximating, is the ring $\mathcal{C}$ generated by the *closed* spheres with rational radii and centres on a countable dense set. To see this, consider the class $\mathcal{D}$ of all sets that can be approximated by finite unions of closed sets in $\mathcal{C}$ in the sense required by condition (iii) of Definition A2.2.VIII. This class contains all open sets because any open set G can be written as a denumerable union of closed spheres, with their centres at points of the countable dense set lying within G, and rational radii bounded by the nonzero distance from the given point of the countable dense set to the boundary of G. $\mathcal{D}$ also contains all closed spheres in $\mathcal{C}$; for example, suppose ϵ is given, choose any positive rational $\delta < \epsilon$, and take the closed spheres with centres at points of the countable dense set lying within the given sphere and having radii δ. These are all elements of $\mathcal{C}$, and therefore so are their intersections with the given closed sphere. These intersections form a countable family of closed sets satisfying (iii) of Definition A2.2.VIII for the given closed sphere. It is obvious that $\mathcal{D}$ is closed under finite unions and that, from the relation

$$\left(\bigcup_{j=1}^{\infty} F_j\right) \cap \left(\bigcup_{k=1}^{\infty} F'_k\right) = \bigcup_{j=1}^{\infty} \bigcup_{k=1}^{\infty} (F_j \cap F'_k),$$

$\mathcal{D}$ is also closed under finite intersections. Since $\mathcal{D}$ contains all closed spheres and their complements that are open, $\mathcal{D}$ contains $\mathcal{C}$. Thus, every set in $\mathcal{C}$ can be approximated by closed spheres in $\mathcal{C}$, so $\mathcal{C}$ is self-approximating as required.

A2.3. Weak Convergence of Measures

We make reference to the following notions of convergence of a sequence of measures on a metric space (see Section A1.3 for the definition of $\|\cdot\|$).

Definition A2.3.I. *Let $\{\mu_n: n \geq 1\}$ and μ be totally finite measures in the metric space $\mathcal{X}$.*

(i) $\mu_n \to \mu$ weakly *if* $\int f \, d\mu_n \to \int f \, d\mu$ *for all bounded continuous functions* f *on* $\mathcal{X}$.

(ii) $\mu_n \to \mu$ vaguely *if* $\int f \, d\mu_n \to \int f \, d\mu$ *for all bounded continuous functions* f *on* $\mathcal{X}$ *vanishing outside a compact set.*

(iii) $\mu_n \to \mu$ strongly (*or* in variation norm) *if* $\|\mu_n - \mu\| \to 0$.

The last definition corresponds to strong convergence in the Banach space of all totally finite signed measures on $\mathcal{X}$, for which the total variation metric constitutes a genuine norm. The first definition does not correspond exactly to weak convergence in the Banach-space sense, but it reduces to weak star (weak*) convergence when $\mathcal{X}$ is compact (say, the unit interval) and the space of signed measures on $\mathcal{X}$ can be identified with the adjoint space to the space of all bounded continuous functions on $\mathcal{X}$. Vague convergence is particularly useful in the discussion of locally compact spaces; in our discussion, a somewhat analogous role is played by the notion of weak hash convergence ($w^{\#}$-convergence; see around Proposition A2.6.II below); it is equivalent to vague convergence when the space is locally compact.

Undoubtedly, the central concept for our purposes is the concept of weak convergence. Not only does it lead to a convenient and internally consistent topologization of the space of realizations of a random measure, but it also provides an appropriate framework for discussing the convergence of random measures conceived as probability distributions on this space of realizations. In this section, we give a brief treatment of some basic properties of weak convergence, following closely the discussion in Billingsley (1968) to which we refer for further details.

Theorem A2.3.II. *Let $\mathcal{X}$ be a metric space and $\{\mu_n: n \geq 1\}$ and μ measures on $\mathcal{B}_{\mathcal{X}}$. Then, the following statements are equivalent.*

(i) $\mu_n \to \mu$ *weakly.*

(ii) $\mu_n(\mathcal{X}) \to \mu(\mathcal{X})$ *and* $\limsup_{n\to\infty} \mu_n(F) \leq \mu(F)$ *for all closed* $F \in \mathcal{B}_{\mathcal{X}}$.

(iii) $\mu_n(\mathcal{X}) \to \mu(\mathcal{X})$ *and* $\liminf_{n\to\infty} \mu_n(G) \geq \mu(G)$ *for all open* $G \in \mathcal{B}_{\mathcal{X}}$.

(iv) $\mu_n(A) \to \mu(A)$ *for all Borel sets* A *with* $\mu(\partial A) = 0$ *(i.e. all* μ*-continuity sets).*

Proof. We show that (i) $\Rightarrow$ (ii) $\Leftrightarrow$ (iii) $\Rightarrow$ (iv) $\Rightarrow$ (i).

Given a closed set F, choose any fixed $\nu > 0$ and construct a $[0,1]$-valued continuous function f that equals 1 on F and vanishes outside F^ν [see (A2.2.2) and Lemma A2.1.II]. We have for each $n \geq 1$

$$\mu_n(F) \leq \int f \, d\mu_n \leq \mu_n(F^\nu),$$

so if (i) holds,

$$\limsup_{n\to\infty} \mu_n(F) \leq \int f \, d\mu \leq \mu(F^\nu).$$

But $F^\nu \downarrow F$ as $\nu \downarrow 0$, and by the continuity Lemma A1.3.II we can choose ν so that, given any $\epsilon > 0$, $\mu(F^\nu) \le \mu(F) + \epsilon$. Since ϵ is arbitrary, the second statement in (ii) follows, while the first is trivial if we take $f = 1$.

Taking complements shows that (ii) and (iii) are equivalent.

When A is a μ-continuity set, $\mu(A^\circ) = \mu(\bar{A})$, so supposing that (iii) holds and hence (ii) also, we have on applying (ii) to $\bar{A}$ and (iii) to A° that

$$\begin{aligned}\limsup \mu_n(A) \le \limsup \mu_n(\bar{A}) &\le \mu(\bar{A}) = \mu(A^\circ) \\ &\le \liminf \mu_n(A^\circ) \le \liminf \mu_n(A).\end{aligned}$$

Thus, equality holds throughout and $\mu_n(A) \to \mu(A)$ so (iv) holds.

Finally, suppose that (iv) holds. Let f be any bounded continuous function on $\mathcal{X}$, and let the bounded interval $[\alpha', \alpha'']$ be such that $\alpha' < f(x) < \alpha''$ for all $x \in \mathcal{X}$. Call $\alpha \in [\alpha', \alpha'']$ a regular value of f if $\mu\{x\colon f(x) = \alpha\} = 0$. At most a countable number of values can be irregular, while for any α, β that are regular values, $\{x\colon \alpha < f(x) \le \beta\}$ is a μ-continuity set. From the boundedness of f on $\mathcal{X}$, given any $\epsilon > 0$, we can partition $[\alpha', \alpha'']$ by a finite set of points $\alpha_0 = \alpha', \ldots, \alpha_N = \alpha''$ with $\alpha_{i-1} < \alpha_i \le \alpha_{i-1} + \epsilon$ for $i = 1, \ldots, N$, and from the countability of the set of irregular points (if any), we can moreover assume that these α_i are all regular points of f. Defining $A_i = \{x\colon \alpha_{i-1} < f(x) \le \alpha_i\}$ for $i = 1, \ldots, N$ and then

$$f_L(x) = \sum_{i=1}^N \alpha_{i-1} I_{A_i}(x), \qquad f_U(x) = \sum_{i=1}^N \alpha_i I_{A_i}(x),$$

each A_i is a μ-continuity set, $f_L(x) \le f(x) \le f_U(x)$, and by (iv),

$$\begin{aligned}\int f_L \,\mathrm{d}\mu = \sum_{i=1}^N \alpha_{i-1}\mu(A_i) &= \lim_{n\to\infty} \sum_{i=1}^N \alpha_{i-1}\mu_n(A_i) = \lim_{n\to\infty} \int f_L \,\mathrm{d}\mu_n \\ &\le \lim_{n\to\infty} \int f_U \,\mathrm{d}\mu_n = \int f_U \,\mathrm{d}\mu,\end{aligned}$$

the extreme terms here differing by at most $\epsilon\mu(\mathcal{X})$. Since ϵ is arbitrary and $\int f_L \,\mathrm{d}\mu_n \le \int f \,\mathrm{d}\mu_n \le \int f_U \,\mathrm{d}\mu_n$, it follows that we must have $\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu$ for all bounded continuous f; that is, $\mu_n \to \mu$ weakly. □

Since the functions used in the proof that (i) implies (ii) are uniformly continuous, we can extract from the proof the following useful condition for weak convergence.

Corollary A2.3.III. *$\mu_n \to \mu$ weakly if and only if $\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu$ for all bounded and uniformly continuous functions $f\colon \mathcal{X} \mapsto \mathbb{R}$.*

Billingsley calls a class $\mathcal{C}$ of sets with the property that

$$\mu_n(C) \to \mu(C) \quad (\text{all } C \in \mathcal{C}) \qquad \text{implies} \qquad \mu_n \to \mu \quad \text{weakly} \qquad \text{(A2.3.1)}$$

a *convergence-determining class*. In this terminology, (iv) of Theorem A2.3.II asserts that the μ-continuity sets form a convergence-determining class. Any convergence-determining class is necessarily a determining class, but the converse need not be true. In particular circumstances, it may be of considerable importance to find a convergence-determining class that is smaller than the classes in Theorem A2.3.II. While such classes often have to be constructed to take advantage of particular features of the metric space in question, the general result below is also of value. In it, a *covering semiring* is a semiring with the property that every open set can be represented as a finite or countable union of sets from the semiring. If $\mathcal{X}$ is separable, an important example of such a semiring is obtained by first taking the open spheres $S_{r_j}(d_k)$ with centres at the points $\{d_k\}$ of a countable dense set and radii $\{r_j\}$ forming a countable dense set in $(0,1)$, then forming finite intersections, and finally taking proper differences.

Proposition A2.3.IV. *Any covering semiring, together with the whole space $\mathcal{X}$, forms a convergence-determining class.*

Proof. Let G be an open set so that by assumption we have

$$G = \bigcup_{i=1}^{\infty} C_i \qquad \text{for some } C_i \in \mathcal{S},$$

where $\mathcal{S}$ is a generating semiring. Since the limit μ in (A2.3.1) is a measure, given $\epsilon > 0$, we can choose a finite integer K such that

$$\mu\left(G - \bigcup_{i=1}^{K} C_i\right) \leq \tfrac{1}{2}\epsilon, \qquad \text{i.e.} \qquad \mu(G) \leq \mu\left(\bigcup_{i=1}^{K} C_i\right) + \tfrac{1}{2}\epsilon.$$

Further, since $\mathcal{C}$ is a semiring, $\bigcup_{i=1}^{K} C_i$ can be represented as a finite union of disjoint sets in $\mathcal{C}$. From (A2.3.1), it therefore follows that there exists N such that, for $n \geq N$,

$$\mu\left(\bigcup_{i=1}^{K} C_i\right) \leq \mu_n\left(\bigcup_{i=1}^{K} C_i\right) + \tfrac{1}{2}\epsilon.$$

Hence,

$$\mu(G) \leq \liminf_{n\to\infty} \mu_n\left(\bigcup_{i=1}^{K} C_i\right) + \epsilon \leq \liminf_{n\to\infty} \mu_n(G) + \epsilon.$$

Since ϵ is arbitrary, (iii) of Theorem A2.3.II is satisfied, and therefore $\mu_n \to \mu$ weakly. □

We investigate next the preservation of weak convergence under mappings from one metric space into another. Let $\mathcal{X}$, $\mathcal{Y}$ be two metric spaces with associated Borel σ-algebras $\mathcal{B}_{\mathcal{X}}$, $\mathcal{B}_{\mathcal{Y}}$, and f a measurable mapping from $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ into $(\mathcal{Y}, \mathcal{B}_{\mathcal{Y}})$ [recall that f is continuous at x if $\rho_{\mathcal{Y}}\big(f(x'), f(x)\big) \to 0$ whenever $\rho_{\mathcal{X}}(x', x) \to 0$].

Proposition A2.3.V. *Let $(\mathcal{X}, \mathcal{B}_\mathcal{X})$, $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$ be metric spaces and f a measurable mapping of $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ into $(\mathcal{Y}, \mathcal{B}_\mathcal{Y})$. Suppose that $\mu_n \to \mu$ weakly on $\mathcal{X}$ and $\mu(D_f) = 0$; then $\mu_n f^{-1} \to \mu f^{-1}$ weakly.*

PROOF. Let B be any Borel set in $\mathcal{B}_\mathcal{Y}$ and x any point in the closure of $f^{-1}(B)$. For any sequence of points $x_n \in f^{-1}(B)$ such that $x_n \to x$, either $x \in D_f$ or $f(x_n) \to f(x)$, in which case $x \in f^{-1}(\bar{B})$. Arguing similarly for the complement,

$$\partial\{f^{-1}(B)\} \subseteq f^{-1}(\partial B) \cup D_f. \tag{A2.3.2}$$

Now suppose that $\mu_n \to \mu$ weakly on $\mathcal{B}_\mathcal{X}$, and consider the image measures $\mu_n f^{-1}$, μf^{-1} on $\mathcal{B}_\mathcal{Y}$. Let B be any continuity set for μf^{-1}. It follows from (A2.3.2) and the assumption of the proposition that $f^{-1}(B)$ is a continuity set for μ. Hence, for all such B, $(\mu_n f^{-1})(B) = \mu_n(f^{-1}(B)) \to \mu(f^{-1}(B)) = (\mu f^{-1})(B)$; that is, $\mu_n f^{-1} \to \mu f^{-1}$ weakly. □

A2.4. Compactness Criteria for Weak Convergence

In this section, we call a set $\mathcal{M}$ of totally finite Borel measures on $\mathcal{X}$ *relatively compact* for weak convergence if every sequence of measures in $\mathcal{M}$ contains a weakly convergent subsequence. It is shown in Section A2.5 that weak convergence is equivalent to convergence with respect to a certain metric and that if $\mathcal{X}$ is a c.s.m.s., the space of all totally finite Borel measures on $\mathcal{X}$ is itself a c.s.m.s. with respect to this metric. We can then appeal to Proposition A1.2.II and conclude that a set of measures is compact (or relatively compact) if and only if it satisfies any of the criteria (i)–(iv) of that proposition.

This section establishes the following criterion for compactness.

Theorem A2.4.I (Prohorov's Theorem)**.** *Let $\mathcal{X}$ be a c.s.m.s. Necessary and sufficient conditions for a set $\mathcal{M}$ of totally finite Borel measures on $\mathcal{X}$ to be relatively compact for weak convergence are*

(i) *the total masses $\mu(\mathcal{X})$ are uniformly bounded for $\mu \in \mathcal{M}$; and*

(ii) *$\mathcal{M}$ is uniformly tight—namely, given $\epsilon > 0$, there exists a compact K such that, for all $\mu \in \mathcal{M}$,*

$$\mu(\mathcal{X} - K) < \epsilon. \tag{A2.4.1}$$

PROOF. We first establish that the uniform tightness condition is necessary, putting it in the following alternative form.

Lemma A2.4.II. *A set $\mathcal{M}$ of measures is uniformly tight if and only if, for all $\epsilon > 0$ and $\delta > 0$, there exists a finite family of δ-spheres (i.e. of radius δ) $S_1, \ldots, S_N$ such that*

$$\mu\Big(\mathcal{X} - \bigcup_{k=1}^{N} S_k\Big) \leq \epsilon \qquad (\text{all } \mu \in \mathcal{M}). \tag{A2.4.2}$$

PROOF OF LEMMA. If the condition holds, we can find, for every $k = 1, 2, \ldots$, a finite union A_k of spheres of radius $1/k$ such that $\mu(\mathcal{X} - A_k) \leq \epsilon/2^k$ for all $\mu \in \mathcal{M}$. Then, the set $K = \bigcap_{k=1}^{\infty} A_k$ is totally bounded and hence compact, and for every $\mu \in \mathcal{M}$,

$$\mu(\mathcal{X} - K) \leq \sum_{k=1}^{\infty} \mu(\mathcal{X} - A_k) < \epsilon.$$

Thus, $\mathcal{M}$ is uniformly tight. Conversely, if $\mathcal{M}$ is uniformly tight and, given ϵ, we choose a compact K to satisfy (A2.4.1), then for any $\delta > 0$, K can be covered by a finite set of δ-spheres, so (A2.4.2) holds. □

Returning now to the main theorem, suppose if possible that $\mathcal{M}$ is relatively compact but (A2.4.2) fails for some $\epsilon > 0$ and $\delta > 0$. Since we assume $\mathcal{X}$ is separable, we can write $\mathcal{X} = \bigcup_{k=1}^{\infty} S_k$, where each S_k is a δ-sphere. On the other hand, for every finite n, we can find a measure $\mu_n \in \mathcal{M}$ such that

$$\mu_n\Big(\mathcal{X} - \bigcup_{k=1}^{\infty} S_k\Big) \geq \epsilon. \tag{A2.4.3a}$$

If in fact $\mathcal{M}$ is relatively compact, there exists a subsequence $\{\mu_{n_j}\}$ that converges weakly to some limit μ^*. From (A2.4.3a), we obtain via (ii) of Theorem A2.3.II that, for all $N > 0$,

$$\mu^*\Big(\mathcal{X} - \bigcup_{k=1}^{N} S_k\Big) \geq \limsup_{n_j \to \infty} \mu_{n_j}\Big(\mathcal{X} - \bigcup_{k=1}^{N} S_k\Big) \geq \epsilon.$$

This contradicts the requirement that, because $\mathcal{X} - \bigcup_{k=1}^{N} S_k \downarrow \emptyset$, we must have $\mu^*\big(\mathcal{X} - \bigcup_{k=1}^{N} S_k\big) \to 0$. Thus, the uniform tightness condition is necessary. As it is clear that no sequence $\{\mu_n\}$ with $\mu_n(\mathcal{X}) \to \infty$ can have a weakly convergent subsequence, condition (i) is necessary also.

Turning to the converse, we again give a proof based on separability, although in fact the result is true without this restriction. We start by constructing a countable ring $\mathcal{R}$ from the open spheres with rational radii and centres in a countable dense set by taking first finite intersections and then proper differences, thus forming a semiring, and finally taking all finite disjoint unions of such differences.

Now suppose that $\{\mu_n : n \geq 1\}$ is any sequence of measures from $\mathcal{M}$. We have to show that $\{\mu_n\}$ contains a weakly convergent subsequence. For any $A \in \mathcal{R}$, condition (i) implies that $\{\mu_n(A)\}$ is a bounded sequence of real numbers and therefore contains a convergent subsequence. Using a diagonal selection argument, we can proceed to ext ract subsequences $\{\mu_{n_j}\}$ for which the $\mu_n(A)$ approach a finite limit for each of the countable number of sets $A \in \mathcal{R}$. Let us write $\mu^*(A)$ for the limit and for brevity of notation set $\mu_{n_j} = \mu'_j$. Thus, we have

$$\mu'_j(A) \to \mu^*(A) \qquad (\text{all } A \in \mathcal{R}). \tag{A2.4.3b}$$

This might seem enough to set up a proof, for it is easy to see that μ^* inherits finite additivity from the μ'_j, and one might anticipate that the uniform tightness condition could be used to establish countable additivity. The difficulty is that we have no guarantee that the sets $A \in \mathcal{R}$ are continuity sets for μ^*, so (A2.4.3b) cannot be relied on to give the correct value to the limit measure. To get over this difficulty, we have to develop a more elaborate argument incorporating the notion of a continuity set.

For this purpose, we introduce the class $\mathcal{C}$ of Borel sets, which are μ^*-regular in the following sense: given $C \in \mathcal{C}$, we can find a sequence $\{A_n\}$ of sets in $\mathcal{R}$ and an associated sequence of open sets G_n such that $A_n \supseteq G_n \supseteq C$ and similarly a sequence of sets $B_n \in \mathcal{R}$ and closed sets F_n with $C \supseteq F_n \supseteq B_n$, the two sequences $\{A_n\}$, $\{B_n\}$ having the property

$$\liminf \mu^*(A_n) \;=\; \limsup \mu^*(B_n) \;=\; \mu(C), \qquad \text{say.} \tag{A2.4.4}$$

We establish the following properties of the class $\mathcal{C}$.

(1°) *$\mathcal{C}$ is a ring*: Let C, C' be any two sets in $\mathcal{C}$, and consider, for example, the difference $C - C'$. If $\{A_n\}$, $\{G_n\}$, $\{B_n\}$, $\{F_n\}$ and $\{A'_n\}$, $\{G'_n\}$, $\{B'_n\}$, $\{F'_n\}$ are the sequences for C and C', respectively, then $A_n - B'_n \supseteq G_n - F'_n \supseteq C - C' \supseteq F_n - G'_n \supseteq B_n - A'_n$, with $G_n - F'_n$ open, $F_n - G'_n$ closed, and the outer sets elements of $\mathcal{R}$ since $\mathcal{R}$ is a ring. From the inclusion

$$(A_n - B'_n) - (B_n - A'_n) \;\subseteq\; (A_n - B_n) \cup (A'_n - B'_n),$$

we find that $\mu^*(A_n - B'_n)$ and $\mu^*(B_n - A'_n)$ have common limit values, which we take to be the value of $\mu(C - C')$. Thus, $\mathcal{C}$ is closed under differences, and similar arguments show that $\mathcal{C}$ is closed also under finite unions and intersections.

(2°) *$\mathcal{C}$ is a covering ring*: Let d be any element in the countable dense set used to construct $\mathcal{R}$, and for rational values of r define

$$h(r) = \mu^*\big(S_r(d)\big).$$

Then $h(r)$ is monotonically increasing, bounded above, and can be uniquely extended to a monotonically increasing function defined for all positive values of r and continuous at all except a countable set of values of r. It is clear that if r is any continuity point of $h(r)$, the corresponding sphere $S_r(d)$ belongs to $\mathcal{C}$. Hence, for each d, we can find a sequence of spheres $S_{\epsilon_n}(d) \in \mathcal{C}$ with radii $\epsilon_n \to 0$. Since any open set in $\mathcal{X}$ can be represented as a countable union of these spheres, $\mathcal{C}$ must be a covering class.

(3°) *For every $C \in \mathcal{C}$, $\mu'_j(C) \to \mu(C)$*: Indeed, with the usual notation, we have

$$\begin{aligned}\mu^*(A_n) = \lim_{j\to\infty} \mu'_j(A_n) &\geq \limsup_{j\to\infty} \mu'_j(C) \geq \liminf_{j\to\infty} \mu'_j(C) \\ &\geq \lim_{j\to\infty} \mu'_j(B_n) = \mu^*(B_n).\end{aligned}$$

Since the two extreme members can be made as close as we please to $\mu(C)$, the two inner members must coincide and equal $\mu(C)$.

(4°) *μ is finitely additive on $\mathcal{C}$*: This follows from (3°) and the finite additivity of μ'_j.

(5°) *If $\mathcal{M}$ is uniformly tight, then μ is countably additive on $\mathcal{C}$*: Suppose that $\{C_k\}$ is a sequence of sets from $\mathcal{C}$, with $C_k \downarrow \emptyset$ but $\mu(C_k) \geq \alpha > 0$. From the definition of $\mathcal{C}$, we can find for each C_k a set $B_k \in \mathcal{R}$ and a closed set F_k such that $C_k \supseteq F_k \supseteq B_k$ and $\mu^*(B_k) > \mu(C_k) - \alpha/2^{k+1}$. Then

$$\liminf_{j\to\infty} \mu'_j(F_k) \geq \lim_{j\to\infty} \mu'_j(B_k) = \mu^*(B_k) \geq \alpha - \alpha/2^{k+1},$$

and $\mu(C_k) - \liminf_{j\to\infty} \mu'_j\left(\bigcap_{n=1}^{k} F_n\right)$ equals

$$\begin{aligned}\limsup_{j\to\infty} \mu'_j\left(C_k - \bigcap_{n=1}^{k} F_n\right) &\leq \sum_{n=1}^{k} \limsup_{j\to\infty} \mu'_j(C_n - F_n) \\ &\leq \sum_{n=1}^{k} \left[\mu(C_n) - \liminf_{j\to\infty} \mu'_j(F_n)\right] \leq \tfrac{1}{2}\alpha;\end{aligned}$$

hence,

$$\liminf_{j\to\infty} \mu'_j\left(\bigcap_{n=1}^{k} F_n\right) \geq \tfrac{1}{2}\alpha \qquad \text{(all } k\text{)}.$$

If now $\mathcal{M}$ is uniformly tight, there exists a compact set K such that $\mu(\mathcal{X} - K) < \frac{1}{4}\alpha$ for all $\mu \in \mathcal{M}$. In particular, therefore,

$$\mu'_j\left(\bigcap_{n=1}^{k} F_n\right) - \mu'_j\left(\bigcap_{n=1}^{k} (F_n \cap K)\right) < \frac{\alpha}{4}, \quad \text{so} \quad \liminf_{j\to\infty} \mu'_j\left(\bigcap_{n=1}^{k} (F_n \cap K)\right) \geq \frac{\alpha}{4}.$$

But this is enough to show that, for each k, the sets $\left(\bigcap_{n=1}^{k} F_n\right) \cap K$ are nonempty, and since (if $\mathcal{X}$ is complete) each is a closed subset of the compact set K, it follows from Theorem A1.2.II that their total intersection is nonempty. Since their total intersection is contained in $\bigcap_{n=1}^{\infty} C_n$, this set is also nonempty, contradicting the assumption that $C_n \downarrow \emptyset$.

We can now complete the proof of the theorem without difficulty. From the countable additivity of μ on $\mathcal{C}$, it follows that there is a unique extension of μ to a measure on $\mathcal{B}_{\mathcal{X}}$. Since $\mathcal{C}$ is a covering class and $\mu'_j(C) \to \mu(C)$ for $C \in \mathcal{C}$, it follows from Proposition A2.3.III that $\mu'_j \to \mu$ weakly or, in other words, that the original sequence μ_n contains a weakly convergent subsequence, as required. □

A2.5. Metric Properties of the Space $\mathcal{M}_{\mathcal{X}}$

Denote by $\mathcal{M}_{\mathcal{X}}$ the space of all totally finite measures on $\mathcal{B}_{\mathcal{X}}$, and consider the following candidate (the Prohorov distance) for a metric on $\mathcal{M}_{\mathcal{X}}$, where F^ϵ is a halo set as in (A2.2.2):

$$\begin{aligned} d(\mu,\nu) = \inf\{\epsilon\colon \epsilon \geq 0, \text{ and for all closed } F \subseteq \mathcal{X}, \\ \mu(F) \leq \nu(F^\epsilon) + \epsilon \text{ and } \nu(F) \leq \mu(F^\epsilon) + \epsilon\}. \end{aligned} \tag{A2.5.1}$$

If $d(\mu,\nu) = 0$, then $\mu(F) = \nu(F)$ for all closed F, so $\mu(\cdot)$ and $\nu(\cdot)$ coincide. If $d(\lambda,\mu) = \delta$ and $\mathrm{d}(\mu,\nu) = \epsilon$, then

$$\begin{aligned} \lambda(F) &\leq \mu(F^\delta) + \delta \leq \mu(\overline{F^\delta}) + \delta \\ &\leq \nu\big((\overline{F^\delta})^\epsilon\big) + \delta + \epsilon \leq \nu(F^{\delta+\epsilon}) + \delta + \epsilon, \end{aligned}$$

with similar inequalities holding when λ and ν are interchanged. Thus, the triangle inequality holds for d, showing that d is indeed a metric.

The main objects of this section are to show that the topology generated by this metric coincides with the topology of weak convergence and to establish various properties of $\mathcal{M}_{\mathcal{X}}$ as a metric space in its own right. We start with an extension of Theorem A2.3.II.

Proposition A2.5.I. *Let $\mathcal{X}$ be a c.s.m.s. and $\mathcal{M}_{\mathcal{X}}$ the space of all totally finite measures on $\mathcal{B}_{\mathcal{X}}$. Then, each of the following families of sets in $\mathcal{M}_{\mathcal{X}}$ is a basis, and the topologies generated by these three bases coincide:*

(i) *the sets $\{\nu\colon d(\nu,\mu) < \epsilon\}$ for all $\epsilon > 0$ and $\mu \in \mathcal{M}_{\mathcal{X}}$;*

(ii) *the sets $\{\nu\colon \nu(F_i) < \mu(F_i) + \epsilon$ for $i = 1,\dots,k$, $|\nu(\mathcal{X}) - \mu(\mathcal{X})| < \epsilon\}$ for all $\epsilon > 0$, finite families of closed sets $F_1,\dots,F_k$, and $\mu \in \mathcal{M}_{\mathcal{X}}$;*

(iii) *the sets $\{\nu\colon \nu(G_i) > \mu(G_i) - \epsilon$ for $i = 1,\dots,k$, $|\nu(\mathcal{X}) - \mu(\mathcal{X})| < \epsilon\}$ for all $\epsilon > 0$, finite families of open sets $G_1,\dots,G_k$, and $\mu \in \mathcal{M}_{\mathcal{X}}$.*

Proof. Each of the three families represents a family of neighbourhoods of a measure $\mu \in \mathcal{M}_{\mathcal{X}}$. To show that each family forms a basis, we need to verify that, if G, H are neighbourhoods of μ, ν in the given family, and $\eta \in G \cap H$, then we can find a member J of the family such that $\eta \in J \subseteq G \cap H$.

Suppose, for example, that G, H are neighbourhoods of μ, ν in the family (ii) [(ii)-neighbourhoods for short], corresponding to closed sets $F_1,\dots,F_n$, $F'_1,\dots,F'_m$, respectively, and with respective bounds ϵ,ϵ', and that η is any measure in the intersection $G \cap H$. Then we must find closed sets C_i and a bound δ, defining a (ii)-neighbourhood J of η such that, for any $\rho \in J$,

$$\begin{aligned} \rho(F_i) &< \mu(F_i) + \epsilon \qquad (i = 1,\dots,n), \\ \rho(F'_j) &< \mu(F'_j) + \epsilon' \qquad (j = 1,\dots,m), \end{aligned}$$

and $|\rho(\mathcal{X} - \mu(\mathcal{X})| < \epsilon$.

For this purpose, we may take $C_i = F_i, i = 1,\dots,n$; $C_{i+j} = F'_j, j = 1,\dots,m$, and $\delta = \min\{\delta_1,\dots,\delta_n;\ \delta'_1,\dots,\delta'_m;\ \frac{1}{2}\epsilon, \frac{1}{2}\epsilon'\}$, where

$$\delta_i = \mu(F_i) + \epsilon - \eta(F_i) \qquad (i = 1, \ldots, n),$$
$$\delta'_j = \mu(F'_j) + \epsilon' - \eta(F'_j) \qquad (j = 1, \ldots, m).$$

For $\rho \in J$ thus defined, we have, for $i = 1, \ldots, n$,

$$\rho(F_i) < \eta(F_i) + \delta = \eta(F_i) + \mu(F_i) + \epsilon_1 - \eta(F_i) = \mu(F_i) + \epsilon_1,$$

while $|\rho(\mathcal{X}) - \mu(\mathcal{X})| < \epsilon_1$. Thus $J \subseteq G$, and similarly $J \subseteq H$.

The proof for family (iii) follows similar lines, while that for family (i) is standard.

To check that the three topologies are equivalent, we show that for any $\mu \in \mathcal{M}_{\mathcal{X}}$, any (iii)-neighbourhood of μ contains a (ii)-neighbourhood, which in turn contains a (i)-neighbourhood, and that this in turn contains a (iii)-neighbourhood.

Suppose there is given then a (iii)-neighbourhood of μ, as defined in (iii) of the proposition, and construct a (ii)-neighbourhood by setting $F_i = G_i^c$, $i = 1, \ldots, n$, and taking $\frac{1}{2}\epsilon$ in place of ϵ. Then, for any ν in this neighbourhood,

$$\nu(G_i) = \nu(\mathcal{X}) - \nu(G_i^c) > \mu(\mathcal{X}) - \tfrac{1}{2}\epsilon - \mu(G_i^c) - \tfrac{1}{2}\epsilon = \mu(G_i) - \epsilon.$$

Since the condition on $|\mu(\mathcal{X}) - \nu\mathcal{X})|$ carries across directly, this is enough to show that ν lies within the given (iii)-neighbourhood of μ.

Given next a (ii)-neighbourhood, defined as in the proposition, we can find a δ with $0 < \delta < \frac{1}{2}\epsilon$ for which, for $i = 1, \ldots, n$, $\mu(F_i^\delta) < \mu(F_i) + \frac{1}{2}\epsilon$. Consider the sphere in $\mathcal{M}_{\mathcal{X}}$ with centre μ and radius δ, using the weak-convergence metric d. For any ν in this sphere,

$$\nu(F_i) < \mu(F_i^\delta) + \delta < \mu(F_i) + \tfrac{1}{2}\epsilon + \tfrac{1}{2}\epsilon = \mu(F_i) + \epsilon,$$

while taking $F = \mathcal{X}$ in the defining relation for d gives $\nu(\mathcal{X}) - \frac{1}{2}\epsilon < \mu(\mathcal{X}) < \nu(\mathcal{X} + \frac{1}{2}\epsilon$; thus ν also lies within the given (ii)-neighbourhood.

Finally, suppose there is given a (i)-neighbourhood of μ, S_μ say, defined by the relations, holding for all closed F and given $\epsilon > 0$,

$$\{\nu\colon \nu(F) < \mu(F^\epsilon) + \epsilon;\ \mu(F) < \nu(F^\epsilon) + \epsilon\}.$$

We have to construct a (iii)-neighbourhood of μ that lies within S_μ. To this end, we first use the separability of $\mathcal{X}$ to cover $\mathcal{X}$ with a countable union of spheres $S_1, S_2, \ldots$, each of radius $\frac{1}{3}\epsilon$ or less, and each a continuity set for μ. Then, choose N large enough so that $R_N = \mathcal{X} - \cup_1^N S_i$, which is also a continuity set for μ, satisfies $\mu(R_N) < \frac{1}{3}\epsilon$.

We now define a (iii)-neighbourhood of μ by taking the finite family of sets $\mathcal{A}$ consisting of all finite unions of the S_i, $i = 1, \ldots, N$, all finite unions of the closures of their complements $\overline{S_i^c}$, and $\overline{R}_N$, and setting

$$G_\mu = \{\nu : \nu(A) < \mu(A) + \tfrac{1}{3}\epsilon,\ A \in \mathcal{A},\ \ |\nu(\mathcal{X}) - \mu(\mathcal{X})| < \tfrac{1}{3}\epsilon\}.$$

Given an arbitrary closed F in $\mathcal{X}$, denote by F^* the union of all elements of $\mathcal{A}$ that intersect F, so that $F^* \in \mathcal{A}$ and $F^* \subseteq \overline{F^*} \subseteq F^\epsilon$. Then, for $\nu \in G_\mu$,

$$\begin{aligned}\nu(F) &\le \nu(F^*) + \nu(R_N) \;<\; \mu(F^*) + \tfrac{1}{3}\epsilon + \nu(\overline{R_N}) \\ &< \mu(F^*) + \tfrac{1}{3}\epsilon + \mu(\overline{R_N}) + \tfrac{1}{3}\epsilon \;<\; \mu(F^\epsilon) + \epsilon.\end{aligned}$$

Further, $\mu(F) \le \mu(F^*) + \mu(R_N) < \mu(F^*) + \frac{1}{3}\epsilon = \mu(\mathcal{X}) - \mu[(\overline{F^*})^c] + \frac{1}{3}\epsilon$. But $\mu(\mathcal{X}) < \nu(\mathcal{X}) + \frac{1}{3}\epsilon$, and $\mu[(\overline{F^*})^c] \ge \nu[(\overline{F^*})^c] - \frac{1}{3}\epsilon$, so that on substituting

$$\mu(F) < \nu(\mathcal{X}) - \nu[(\overline{F^*})^c] + \epsilon = \nu(\overline{F^*}) + \epsilon < \nu(F^\epsilon) + \epsilon.$$

These inequalities show that $\nu \in S_\mu$ and hence $G_\mu \subseteq S_\mu$. □

The weak convergence of μ_n to μ is equivalent by Theorem A2.3.II to $\mu_n \to \mu$ in each of the topologies (ii) and (iii) and hence by the proposition to $d(\mu_n, \mu) \to 0$. The converse holds, so we have the following.

Corollary A2.5.II. *For μ_n and $\mu \in \mathcal{M}_\mathcal{X}$, $\mu_n \to \mu$ weakly if and only if $d(\mu_n, \mu) \to 0$.*

If A is a continuity set for μ, then we have also $\mu_n(A) \to \mu(A)$. However, it does not appear that there is a basis, analogous to (ii) and (iii) of Proposition A2.5.I, corresponding to this form of the convergence.

Having established the fact that the weak topology is a metric topology, it makes sense to ask whether $\mathcal{M}_\mathcal{X}$ is separable or complete with this topology.

Proposition A2.5.III. *If $\mathcal{X}$ is a c.s.m.s. and $\mathcal{M}_\mathcal{X}$ is given the topology of weak convergence, then $\mathcal{M}_\mathcal{X}$ is also a c.s.m.s.*

Proof. We first establish completeness by using the compactness criteria of the preceding section. Let $\{\mu_n\}$ be a Cauchy sequence in $\mathcal{M}_\mathcal{X}$; we show that it is uniformly tight. Let positive ϵ and δ be given, and choose positive $\eta < \min(\frac{1}{3}\epsilon, \frac{1}{2}\delta)$. From the Cauchy property, there is an N for which $d(\mu_n, \mu_N) < \eta$ for $n \ge N$. Since μ_N itself is tight, $\mathcal{X}$ can be covered by a sequence of spheres $S_1, S_2, \ldots$ of radius η and there is a finite K for which

$$\mu_N(\mathcal{X}) - \mu_N\Big(\bigcup_{i=1}^K S_i\Big) < \eta.$$

For $n > N$, since $d(\mu_n, \mu_N) < \eta$,

$$\mu_n(\mathcal{X}) - \mu_N(\mathcal{X}) < \eta \qquad \text{and} \qquad \mu_N\Big(\bigcup_{i=1}^K S_i\Big) < \mu_n\Big(\big(\bigcup_{i=1}^K S_i\big)^\eta\Big) + \eta,$$

so

$$\begin{aligned}\mu_n(\mathcal{X}) - \mu_n\Big(\big(\bigcup_{i=1}^K S_i\big)^\eta\Big) &< \mu_n(\mathcal{X}) - \mu_n\big(\bigcup_{i=1}^K S_i\big) \\ &\le |\mu_n(\mathcal{X}) - \mu_N(\mathcal{X})| + \big|\mu_N(\mathcal{X}) - \mu_N\big(\bigcup_{i=1}^K S_i\big)\big| + \eta \;\le\; 3\eta < \epsilon.\end{aligned}$$

It follows that for every ϵ and δ we can find a finite family of δ spheres whose union has μ_n measure within ϵ of $\mu_n(\mathcal{X})$, uniformly in n. Hence, the sequence $\{\mu_n\}$ is uniformly tight by Lemma A2.4.II and relatively compact by Theorem A2.4.I [since it is clear that the quantities $\mu_n(\mathcal{X})$ are bounded when $\{\mu_n\}$ is

a Cauchy sequence]. Thus, there exists a limit measure such that $\mu_n \to \mu$ weakly, which implies by Corollary A2.5.II that $d(\mu_n, \mu) \to 0$.

Separability is easier to establish, as a suitable dense set is already at hand in the form of the measures with finite support (i.e. those that are purely atomic with only a finite set of atoms). Restricting the atoms to the points of a separability set $\mathcal{D}$ for $\mathcal{X}$ and their masses to rational numbers, we obtain a countable family of measures, $\mathcal{D}'$ say, which we now show to be dense in $\mathcal{M}_{\mathcal{X}}$ by proving that any sphere $S'_\epsilon(\mu) \subseteq \mathcal{M}_{\mathcal{X}}$ contains an element of $\mathcal{D}'$. To this end, first choose a compact set K such that $\mu(\mathcal{X} \setminus K) < \frac{1}{2}\epsilon$, which is possible because μ is tight. Now cover K with a finite family of disjoint sets $A_1, \ldots, A_n$, each with nonempty interior and of radius ϵ or less. [One way of constructing such a covering is as follows. First, cover K with a finite family of open spheres $S_1, \ldots, S_m$, say, each of radius ϵ. Take $A_1 = \bar{S}_1$, $A_2 = \bar{S}_2 \cap A_1^c$, $A_3 = \bar{S}_3 \cap (A_1 \cup A_2)^c$, and so on, retaining only the nonempty sets in this construction. Then $S_2 \cap A_1^c$ is open and either empty, in which case $S_2 \subseteq A_1$ so $\bar{S}_2 \subseteq \bar{A}_1$ and A_2 is empty, or has nonempty interior. It is evident that each A_i has radius ϵ or less and that they are disjoint.] For each i, since A_i has nonempty interior, we can choose an element x_i of the separability set for $\mathcal{X}$ with $x_i \in A_i$, give x_i rational mass μ_i such that

$$\mu(A_i) \geq \mu_i \geq \mu(A_i) - \epsilon/(2N),$$

and let μ' denote a purely atomic measure with atoms at x_i of mass μ_i. Then, for an arbitrary closed set F, with $\sum'$ denoting $\sum_{i:x_i \in F}$,

$$\mu'(F) = \sum{}' \mu_i \leq \sum{}' \mu(A_i) < \mu(F^\epsilon) + \epsilon,$$

where we have used the fact that $\bigcup_{i:x_i \in F} A_i \subseteq F^\epsilon$ because A_i has radius at most ϵ. Furthermore,

$$\mu(F) < \mu(K \cap F) + \tfrac{1}{2}\epsilon \leq \sum{}'' \mu(F \cap A_i) + \tfrac{1}{2}\epsilon,$$

where $\sum''$ denotes $\sum_{i:A_i \cap F \neq \emptyset}$, so

$$\mu(F) \leq \sum{}'' \mu'(A_i) + \tfrac{1}{2}\epsilon + \tfrac{1}{2}\epsilon < \mu(F^\epsilon) + \epsilon.$$

Consequently, $d(\mu, \mu') < \epsilon$, or equivalently, $\mu' \in S'_\epsilon(\mu)$, as required. □

Denote the Borel σ-algebra on $\mathcal{M}_{\mathcal{X}}$ by $\mathcal{B}(\mathcal{M}_{\mathcal{X}})$ so that from the results just established it is the smallest σ-algebra containing any of the three bases listed in Proposition A2.5.I. We use this fact to characterize $\mathcal{B}(\mathcal{M}_{\mathcal{X}})$.

Proposition A2.5.IV. *Let $\mathcal{S}$ be a semiring generating the Borel sets $\mathcal{B}_{\mathcal{X}}$ of $\mathcal{X}$. Then $\mathcal{B}(\mathcal{M}_{\mathcal{X}})$ is the smallest σ-algebra of subsets of $\mathcal{M}_{\mathcal{X}}$ with respect to which the mappings $\Phi_A : \mathcal{M}_{\mathcal{X}} \mapsto \mathbb{R}$ defined by*

$$\Phi_A(\mu) = \mu(A)$$

are measurable for $A \in \mathcal{S}$. In particular, $\mathcal{B}(\mathcal{M}_{\mathcal{X}})$ is the smallest σ-algebra with respect to which the Φ_A are measurable for all $A \in \mathcal{B}_{\mathcal{X}}$.

PROOF. Start by considering the class $\mathcal{C}$ of subsets A of $\mathcal{X}$ for which Φ_A is $\mathcal{B}(\mathcal{M}_{\mathcal{X}})$-measurable. Since $\Phi_{A\cup B} = \Phi_A + \Phi_B$ for disjoint A and B, and the sum of two measurable functions is measurable, $\mathcal{C}$ is closed under finite disjoint unions. Similarly, since $\Phi_{A\setminus B} = \Phi_A - \Phi_B$ for $A \supseteq B$, $\mathcal{C}$ is closed under proper differences and hence in particular under complementation. Finally, since a monotone sequence of measurable functions has a measurable limit, and $\Phi_{A_n} \uparrow \Phi_A$ whenever $A_n \uparrow A$, it follows that $\mathcal{C}$ is a monotone class.

Let F be any closed set in $\mathcal{X}$ and y any positive number. Choose $\mu \in \mathcal{M}_{\mathcal{X}}$ such that $\mu(F) < y$ and set $\epsilon = y - \mu(F)$. We can then write

$$\{\nu\colon \Phi_F(\nu) < y\} \;=\; \{\nu\colon \nu(F) < y\} \;=\; \{\nu\colon \nu(F) < \mu(F) + \epsilon\},$$

showing that this set of measures is an element of the basis (ii) of Proposition A2.5.I and hence an open set in $\mathcal{M}_{\mathcal{X}}$ and therefore an element of $\mathcal{B}(\mathcal{M}_{\mathcal{X}})$. Thus, $\mathcal{C}$ contains all closed sets, and therefore also $\mathcal{C}$ contains all open sets. From these properties of $\mathcal{C}$, it now follows that $\mathcal{C}$ contains the ring of all finite disjoint unions of differences of open sets in $\mathcal{X}$, and since $\mathcal{C}$ is a monotone class, it must contain all sets in $\mathcal{B}_{\mathcal{X}}$. This shows that Φ_A is $\mathcal{B}(\mathcal{M}_{\mathcal{X}})$-measurable for all Borel sets A and hence *a fortiori* for all sets in any semiring $\mathcal{S}$ generating the Borel sets.

It remains to show that $\mathcal{B}(\mathcal{M}_{\mathcal{X}})$ is the smallest σ-algebra in $\mathcal{M}_{\mathcal{X}}$ with this property. Let $\mathcal{S}$ be given, and let $\mathcal{R}$ be any σ-ring with respect to which Φ_A is measurable for all $A \in \mathcal{S}$. By arguing as above, it follows that Φ_A is also $\mathcal{R}$-measurable for all A in the σ-ring generated by $\mathcal{S}$, which by assumption is $\mathcal{B}_{\mathcal{X}}$. Now suppose we are given $\epsilon > 0$, a measure $\mu \in \mathcal{M}_{\mathcal{X}}$, and a finite family $F_1, \ldots, F_n$ of closed sets. Then, the set

$$\{\nu\colon \nu(F_i) < \mu(F_i) + \epsilon \text{ for } i = 1, \ldots, n \text{ and } |\nu(\mathcal{X}) - \mu(\mathcal{X})| < \epsilon\}$$

is an intersection of sets of $\mathcal{R}$ and hence is an element of $\mathcal{R}$. But this shows that $\mathcal{R}$ contains a basis for the open sets of $\mathcal{M}_{\mathcal{X}}$. Since $\mathcal{M}_{\mathcal{X}}$ is separable, every open set can be represented as a countable union of basic sets, and thus all open sets are in $\mathcal{R}$. Thus, $\mathcal{R}$ contains $\mathcal{B}(\mathcal{M}_{\mathcal{X}})$, completing the proof. □

A2.6. Boundedly Finite Measures and the Space $\mathcal{M}_{\mathcal{X}}^{\#}$

For applications to random measures, we need to consider not only totally finite measures on $\mathcal{B}_{\mathcal{X}}$ but also σ-finite measures with the strong local finiteness condition contained in the following definition.

Definition A2.6.I. *A Borel measure μ on the c.s.m.s. $\mathcal{X}$ is* boundedly finite *if $\mu(A) < \infty$ for every bounded Borel set A.*

We write $\mathcal{M}_{\mathcal{X}}^{\#}$ for the space of boundedly finite Borel measures on $\mathcal{X}$ and generally use the $^{\#}$ notation for concepts taken over from finite to boundedly

finite measures. The object of this section is to extend to $\mathcal{M}_{\mathcal{X}}^{\#}$ the results previously obtained for $\mathcal{M}_{\mathcal{X}}$: while most of these extensions are routine, they are given here for the sake of completeness.

Consider first the extension of the concept of weak convergence. Taking a fixed origin $x_0 \in \mathcal{X}$, let $S_r = S_r(x_0)$ for $0 < r < \infty$ and introduce a distance function $d^{\#}$ on $\mathcal{M}_{\mathcal{X}}^{\#}$ by setting

$$d^{\#}(\mu, \nu) = \int_0^{\infty} \mathrm{e}^{-r} \frac{d_r(\mu^{(r)}, \nu^{(r)})}{1 + d_r(\mu^{(r)}, \nu^{(r)})} \, \mathrm{d}r, \tag{A2.6.1}$$

where $\mu^{(r)}$, $\nu^{(r)}$ are the totally finite restrictions of μ, ν to S_r and d_r is the Prohorov distance between the restrictions. Examining (A2.5.1) where this distance is defined, we see that the infimum cannot decrease as r increases when the number of closed sets to be scrutinized increases, so as a function of r, d_r is monotonic and thus a measurable function. Since the ratio $d_r/(1+d_r) \leq 1$, the integral in (A2.6.1) is defined and finite for all μ, ν. The triangle inequality is preserved under the mapping $x \mapsto x/(1+x)$, while $d^{\#}(\mu, \nu) = 0$ if and only if μ and ν coincide on a sequence of spheres expanding to the whole of $\mathcal{X}$, in which case they are identical.

We call the metric topology generated by $d^{\#}$ the $w^{\#}$-topology ('weak hash' topology) and write $\mu_k \to_{w^{\#}} \mu$ for convergence with respect to this topology. Some equivalent conditions for $w^{\#}$-convergence are as in the next result.

Proposition A2.6.II. *Let $\{\mu_k : k = 1, 2, \ldots\}$ and μ be measures in $\mathcal{M}_{\mathcal{X}}^{\#}$; then the following conditions are equivalent.*

(i) $\mu_k \to_{w^{\#}} \mu$.

(ii) $\int_{\mathcal{X}} f(x)\, \mu_k(\mathrm{d}x) \to \int_{\mathcal{X}} f(x)\, \mu(\mathrm{d}x)$ *for all bounded continuous functions $f(\cdot)$ on $\mathcal{X}$ vanishing outside a bounded set.*

(iii) *There exists a sequence of spheres $S^{(n)} \uparrow \mathcal{X}$ such that if $\mu_k^{(n)}$, $\mu^{(n)}$ denote the restrictions of the measures μ_k, μ to subsets of $S^{(n)}$, then $\mu_k^{(n)} \to \mu^{(n)}$ weakly as $k \to \infty$ for $n = 1, 2, \ldots$.*

(iv) $\mu_k(A) \to \mu(A)$ *for all bounded $A \in \mathcal{B}_{\mathcal{X}}$ for which $\mu(\partial A) = 0$.*

Proof. We show that (i) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (iv) $\Rightarrow$ (i). Write the integral in (A2.6.1) for the measures μ_n and μ as

$$d^{\#}(\mu_k, \mu) = \int_0^{\infty} \mathrm{e}^{-r} g_k(r) \, \mathrm{d}r$$

so that for each k, $g_k(r)$ increases with r and is bounded above by 1. Thus, there exists a subsequence $\{k_n\}$ and a limit function $g(\cdot)$ such that $g_{k_n}(r) \to g(r)$ at all continuity points of g [this is just a version of the compactness criterion for vague convergence on $\mathbb{R}$: r egard each $g_k(r)$ as the distribution function of a probability measure so that there exists a vaguely convergent subsequence; see Corollary A2.6.V or any standard proof of the Helly–Bray results]. By dominated convergence, $\int_0^{\infty} \mathrm{e}^{-r} g(r) \, \mathrm{d}r = 0$ and hence, since $g(\cdot)$

is monotonic, $g(r) = 0$ for all finite $r > 0$. This being true for all convergent subsequences, it follows that $g_k(r) \to 0$ for such r and thus, for these r,

$$d_r(\mu_k^{(r)}, \mu^{(r)}) \to 0 \qquad (k \to \infty).$$

In particular, this is true for an increasing sequence of values r_n, corresponding to spheres $\{S_{r_n}\} \equiv \{S_n\}$, say, on which therefore $\mu_k^{(r_n)} \to \mu^{(r_n)}$ weakly. Thus, (i) implies (iii).

Suppose next that (iii) holds and that f is bounded, continuous, and vanishes outside some bounded set. Then, the support of f is contained in some S_r, and hence $\int f \, \mathrm{d}\mu_k^{(r)} \to \int f \, \mathrm{d}\mu^{(r)}$, which is equivalent to (ii).

When (ii) holds, the argument used to establish (iv) of Theorem A2.3.II shows that $\mu_k(C) \to \mu(C)$ whenever C is a bounded Borel set with $\mu(\partial C) = 0$.

Finally, if (iv) holds and S_r is any sphere that is a continuity set for μ, then by the same theorem $\mu_k^{(r)} \to \mu^{(r)}$ weakly in S_r. But since $\mu(S_r)$ increases monotonically in r, S_r is a continuity set for almost all r, so the convergence to zero of $d^{\#}(\mu_k, \mu)$ follows from the dominated convergence theorem. □

Note that we cannot find a universal sequence of spheres, $\{S^n\}$ say, for which (i) and (ii) are equivalent because the requirement of weak convergence on S^n that $\mu_k(S^n) \to \mu(S^n)$ cannot be guaranteed unless $\mu(\partial S^n) = 0$.

While the distance fiunction $d^{\#}$ of Definition A2.6.I depends on the centre x_0 of the family $\{S_r\}$ of spheres used there, *the $w^{\#}$-topology does not depend on the choice of x_0*. To see this, let $\{S'_n\}$ be any sequence of spheres expanding to $\mathcal{X}$ so that to any S'_n we can first find n' for which $S'_n \subseteq S_{r_{n'}}$ and then find n'' for which $S_{r_{n'}} \subseteq S'_{n''}$. Now weak convergence within a given sphere is subsumed by weak convergence in a larger sphere containing it, from which the asserted equivalence follows.

It should also be noted that for locally compact $\mathcal{X}$, $w^{\#}$-convergence coincides with vague convergence.

The next theorem extends to $w^{\#}$-convergence the results in Propositions A2.5.III and A2.5.IV.

Theorem A2.6.III. (i) *$\mathcal{M}_{\mathcal{X}}^{\#}$ with the $w^{\#}$-topology is a c.s.m.s.*

(ii) *The Borel σ-algebra $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ is the smallest σ-algebra with respect to which the mappings $\Phi_A \colon \mathcal{M}_{\mathcal{X}}^{\#} \mapsto \mathbb{R}$ given by*

$$\Phi_A(\mu) = \mu(A)$$

are measurable for all sets A in a semiring $\mathcal{S}$ of bounded Borel sets generating $\mathcal{B}_{\mathcal{X}}$ and in particular for all bounded Borel sets A.

Proof. To prove separability, recall first that the measures with rational masses on finite support in a separability set $\mathcal{D}$ for $\mathcal{X}$ form a separability set $\mathcal{D}'$ for the totally finite measures on each S_n under the weak topology. Given $\epsilon > 0$, choose R so that $\int_R^{\infty} \mathrm{e}^{-r} \, \mathrm{d}r < \frac{1}{2}\epsilon$. For any $\mu \in \mathcal{M}_{\mathcal{X}}^{\#}$, choose an atomic measure μ_R from the separability set for S_R such that μ_R has support in S_R

and $d_R(\mu_R, \mu^{(R)}) < \frac{1}{2}\epsilon$. Clearly, for $r < R$, we also have

$$d_r(\mu_R^{(r)}, \mu^{(r)}) < \tfrac{1}{2}\epsilon.$$

Substitution in the expression for $d^{\#}$ shows that $d^{\#}(\mu_R, \mu) < \epsilon$, establishing that the union of separability sets is a separability set for measures in $\mathcal{M}_{\mathcal{X}}^{\#}$.

To show completeness, let $\{\mu_k\}$ be a Cauchy sequence for $d^{\#}$. Then, each sequence of restrictions $\{\mu_k^{(r)}\}$ forms a Cauchy sequence for d_r and so has a limit ν_r by Proposition A2.5.III. The sequence $\{\nu_r\}$ of measures so obtained is clearly consistent in the sense that $\nu_r(A) = \nu_s(A)$ for $s \leq r$ and Borel sets A of S_r. Then, the set function

$$\mu(A) = \nu_r(A)$$

is uniquely defined on Borel sets A of S_r and is nonnegative and countably additive on the restriction of $\mathcal{M}_{\mathcal{X}}$ to each S_r. We now extend the definition of μ to all Borel sets by setting

$$\mu(A) = \lim_{r \to \infty} \nu_r(A \cap S_r),$$

the sequence on the right being monotonically increasing and hence having a limit (finite or infinite) for all A. It is then easily checked that $\mu(\cdot)$ is finitely additive and continuous from below and therefore countably additive and so a boundedly finite Borel measure. Finally, it follows from (ii) of Proposition A2.6.II that $\mu_k \to_{w^{\#}} \mu$.

To establish part (ii) of the theorem, examine the proof of Proposition A2.5.IV. Let $\mathcal{C}'$ be the class of sets A for which Φ_A is a $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$-measurable mapping into $[0, \infty)$. Again, $\mathcal{C}'$ is a monotone class containing all bounded open and closed sets on $\mathcal{X}$ and hence $\mathcal{B}_{\mathcal{X}}$ as well as any ring or semiring generating $\mathcal{B}_{\mathcal{X}}$. Also, if $\mathcal{S}$ is a semiring of bounded sets generating $\mathcal{B}_{\mathcal{X}}$ and Φ_A is $\mathcal{R}$-measurable for $A \in \mathcal{S}$ and some σ-ring $\mathcal{R}$ of sets on $\mathcal{M}_{\mathcal{X}}^{\#}$, then Φ_A is $\mathcal{R}$-measurable for $A \in \mathcal{B}_{\mathcal{X}}$. The proposition now implies that $\mathcal{R}^{(r)}$, the σ-algebra formed by projecting the measures in sets of $\mathcal{R}$ onto S_r, contains $\mathcal{B}(\mathcal{M}_{S_r})$. Equivalently, $\mathcal{R}$ contains the inverse image of $\mathcal{B}(\mathcal{M}_{S_r})$ under this projection. The definition of $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ implies that it is the smallest σ-algebra containing each of these inverse images. Hence, $\mathcal{R}$ contains $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$. □

The final extension is of the compactness criterion of Theorem A2.4.I.

Proposition A2.6.IV. *A family of measures $\{\mu_k\}$ in $\mathcal{M}_{\mathcal{X}}^{\#}$ is relatively compact in the $w^{\#}$-topology on $\mathcal{M}_{\mathcal{X}}^{\#}$ if and only if their restrictions $\{\mu_\alpha^{(n)}\}$ to a sequence of closed spheres $\overline{S}_n \uparrow \mathcal{X}$ are relatively compact in the weak topology on $\mathcal{M}_{\overline{S}_n}$, in which case the restrictions $\{\mu_\alpha^F\}$ to any closed bounded F are relatively compact in the weak topology on $\mathcal{M}_F$.*

Proof. Suppose first that $\{\mu_\alpha\}$ is relatively compact in the $w^{\#}$-topology on $\mathcal{M}_{\mathcal{X}}^{\#}$ and that F is a closed bounded subset of $\mathcal{X}$. Given any sequence of the μ_α^F, there exists by assumption a $w^{\#}$-convergent subsequence, $\mu_{\alpha_k} \to_{w^{\#}} \mu$ say. From Proposition A2.6.II, arguing as in the proof of A2.3.II, it follows that for

all bounded closed sets C, $\limsup_{k\to\infty} \mu_{\alpha_k}(C) \le \mu(C)$. Hence, in particular, the values of $\mu_{\alpha_k}(F)$ are bounded above. Moreover, the restrictions $\{\mu^F_{\alpha_k}\}$ are uniformly tight, this property being inherited from their uniform tightness on a closed bounded sphere containing F. Therefore, the restrictions are relatively compact as measures on F, and there exists a further subsequence converging weakly on F to some limit measure, $\mu^{\#}$ say, on F. This is enough to show that the μ^F_α themselves are relatively compact.

Conversely, suppose that there exists a family of spheres S_n, closed or otherwise, such that $\{\mu^{(n)}_\alpha\}$ are relatively compact for each n. By diagonal selection, we may choose a subsequence α_k such that $\mu^{(n)}_{\alpha_k} \to \mu^{(n)}$ weakly for every n and therefore that, if f is any bounded continuous function vanishing outside a bounded set, then $\int f \,\mathrm{d}\mu^{(n)}_{\alpha_k} \to \int f \,\mathrm{d}\mu^{(n)}$. It is then easy to see that the $\mu^{(n)}_\alpha$ form a consistent family (i.e. $\mu^{(n)}_\alpha$ coincides with $\mu^{(m)}_\alpha$ on S_m for $n \ge m$) and so define a unique element μ of $\mathcal{M}^{\#}_{\mathcal{X}}$ such that $\mu_{\alpha_k} \to_{w^{\#}} \mu$. □

The criterion for weak convergence on each S_n can be spelled out in detail from Prohorov's Theorem A2.4.I. A particularly neat result holds in the case where $\mathcal{X}$ is locally (and hence countably) compact when the following terminology is standard. A *Radon measure* in a locally compact space is a measure taking finite values on compact sets. A sequence $\{\mu_k\}$ of such measures converges *vaguely* to μ if $\int f\, d\mu_k \to \int f\, d\mu$ for each continuous f vanishing outside a compact set. Now any locally compact space with a countable base is metrizable, but the space is not necessarily complete in the metric so obtained. If, however, the space is both locally compact and a c.s.m.s., it can be represented as the union of a sequence of compact sets K_n with $K_n \subseteq K^\circ_{n+1}$, and then by changing to an equivalent metric if necessary, we can ensure that the spheres S_n are compact as well as closed (see e.g. Hocking and Young, 1961, Proposition 2.61); we assume this is so. Then, a Borel measure is a Radon measure if and only if it is boundedly finite, and vague convergence coincides with $w^{\#}$-convergence. The discussion around (A2.6.1) shows that the vague topology is metrizable and suggests one form for a suitable metric. Finally, Proposition A2.6.IV takes the following form.

Corollary A2.6.V. *If $\mathcal{X}$ is a locally compact c.s.m.s., then the family $\{\mu_\alpha\}$ of Radon measures on $\mathcal{B}_{\mathcal{X}}$ is relatively compact in the vague topology if and only if the values $\{\mu_\alpha(A)\}$ are bounded for each bounded Borel set A.*

Proof. Assume the metric is so chosen that closed bounded sets are compact. Then, if the $\mu_\alpha(\cdot)$ are relatively compact on each S_n, it follows from condition (i) of Theorem A2.4.I that the $\mu_\alpha(S_n)$ are bounded and hence that the $\mu_\alpha(A)$ are bounded for any bounded Borel set A.

Conversely, suppose the boundedness condition holds. Then, in particular, it holds for S_n, which is compact so the tightness condition (ii) of Theorem A2.4.I is satisfied trivially. Thus, the $\{\mu_\alpha\}$ are relatively compact on each S_n and so by Proposition A2.6.IV are relatively compact in the $w^{\#}$- (i.e. vague) topology. □

A2.7. Measures on Topological Groups

A group $\mathcal{G}$ is a set on which is defined a binary relation $\mathcal{G} \times \mathcal{G} \mapsto \mathcal{G}$ with the following properties.

(i) (*Associative law*) For all g_1, g_2, $g_3 \in \mathcal{G}$, $(g_1 g_2)g_3 = g_1(g_2 g_3)$.

(ii) There exists an *identity element* e (necessarily unique) such that for all $g \in \mathcal{G}$, $ge = eg = g$.

(iii) For every $g \in \mathcal{G}$, there exists a unique *inverse* g^{-1} such that $g^{-1}g = gg^{-1} = e$.

The group is *Abelian* if it also has the property

(iv) (*Commutative law*) For all g_1, $g_2 \in \mathcal{G}$, $g_1 g_2 = g_2 g_1$.

A *homomorphism* between groups is a mapping $\mathcal{T}$ that preserves the group operations in the sense that $(\mathcal{T}g_1)(\mathcal{T}g_2) = \mathcal{T}(g_1 g_2)$ and $(\mathcal{T}g_1)^{-1} = \mathcal{T}g^{-1}$. If the mapping is also one-to-one, it is an *isomorphism*. An *automorphism* is an isomorphism of the group onto itself.

A subgroup $\mathcal{H}$ of $\mathcal{G}$ is a subset of $\mathcal{G}$ that is closed under the group operations and so forms a group in its own right. If $\mathcal{H}$ is *nontrivial* (i.e. neither $\{e\}$ nor the whole of $\mathcal{G}$), its action on $\mathcal{G}$ splits $\mathcal{G}$ into equivalence classes, where $g_1 \equiv g_2$ if there exists $h \in \mathcal{H}$ such that $g_2 = g_1 h$. These classes form the *left cosets* of $\mathcal{G}$ relative to $\mathcal{H}$; they may also be described as the (left) quotient space $\mathcal{G}/\mathcal{H}$ of $\mathcal{G}$ with respect to $\mathcal{H}$. Similarly, $\mathcal{H}$ splits $\mathcal{G}$ into *right cosets*, which in general will not be the same as the left cosets. If $\mathcal{G}$ is Abelian, however, or more generally if $\mathcal{H}$ is a *normal* (or *invariant*) subgroup, which means that for every $g \in \mathcal{G}$, $h \in \mathcal{H}$, $g^{-1}hg \in \mathcal{H}$, the right and left cosets coincide and the products of two elements, one from each of any two given cosets, fall into a uniquely defined third coset. With this definition of multiplication, the cosets then form a group in their own right, namely the *quotient group*. The natural map taking an element from $\mathcal{G}$ into the coset to which it belongs is then a homomorphism of $\mathcal{G}$ into $\mathcal{G}/\mathcal{H}$, of which $\mathcal{H}$ is the *kernel*; that is, the inverse image of the identity in the image space $\mathcal{G}/\mathcal{H}$.

The *direct product* of two groups $\mathcal{G}$ and $\mathcal{H}$, written $\mathcal{G} \times \mathcal{H}$, consists of the Cartesian products of $\mathcal{G}$ and $\mathcal{H}$ with the group operation

$$(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1 h_2),$$

identity $(e_{\mathcal{G}}, e_{\mathcal{H}})$, and inverse $(g, h)^{-1} = (g^{-1}, h^{-1})$. In particular, if $\mathcal{G}$ is a group and $\mathcal{H}$ a normal subgroup, then $\mathcal{G}$ is isomorphic to the direct product $\mathcal{H} \times \mathcal{G}/\mathcal{H}$.

$\mathcal{G}$ is a *topological group* if it has a topology $\mathcal{U}$ with respect to which the mapping $(g_1, g_2) \mapsto g_1 g_2^{-1}$ from $\mathcal{G} \times \mathcal{G}$ (with the product topology) into $\mathcal{G}$ is continuous. This condition makes the operations of left (and right) multiplication by a fixed element of $\mathcal{G}$, and of inversion, continuous. A theory with wide applications results if the topology $\mathcal{U}$ is taken to be locally compact and second countable. It is then metrizable but not necessarily complete in the resulting metric. In keeping with our previous discussion, however, we frequently assume that $\mathcal{G}$ is a *complete separable metric group* (c.s.m.g.) as well

as being locally compact. If, as may always be done by a change of metric, the closed bounded sets of $\mathcal{G}$ are compact, we refer to $\mathcal{G}$ as a σ-group.

Definition A2.7.I. *A σ-group is a locally compact, complete separable metric group with the metric so chosen that closed bounded sets are compact.*

In this context, boundedly finite measures are Radon measures, and the concepts of weak and vague convergence coincide. A boundedly finite measure μ on the σ-group is *left-invariant* if (writing $gA = \{gx : x \in A\}$)

$$\mu(gA) = \mu(A) \qquad (g \in \mathcal{G},\ A \in \mathcal{B}_{\mathcal{G}}), \tag{A2.7.1}$$

or equivalently,

$$\int_{\mathcal{G}} f(g^{-1}x)\,\mu(\mathrm{d}x) = \int_{\mathcal{G}} f(x)\,\mu(\mathrm{d}x) \tag{A2.7.2}$$

for all $f \in \mathrm{BC}(\mathcal{G})$, the class of continuous functions vanishing outside a bounded (in this case compact) set. Right-invariance is defined similarly. A fundamental theorem for locally compact groups asserts that up to scale factors they admit unique left- and right-invariant measures, called *Haar measures*. If the group is Abelian, the left and right Haar measures coincide, as they do also when the group is compact, in which case the Haar measure is totally finite and is uniquely specified when normalized to have total mass unity. On the real line, or more generally on $\mathbb{R}^d$, the Haar measure is just Lebesgue measure $\ell(\cdot)$, and the uniqueness referred to above is effectively a restatement of results on the Cauchy functional equation.

If $\mathcal{G}$ is a topological group and $\mathcal{H}$ a subgroup, the *quotient topology* on $\mathcal{G}/\mathcal{H}$ is the largest topology on $\mathcal{G}/\mathcal{H}$ making the natural map from $\mathcal{G}$ into $\mathcal{G}/\mathcal{H}$ continuous. It is then also an *open map* (i.e. takes open sets into open sets). If it is closed, then the quotient topology for $\mathcal{G}/\mathcal{H}$ inherits properties from the topology for $\mathcal{G}$: it is Hausdorff, or compact, or locally compact if and only if $\mathcal{G}$ has the same property.

These concepts extend to the more general context where $\mathcal{X}$ is a c.s.m.s. and $\mathcal{H}$ defines a group of one-to-one bounded continuous maps T_h of $\mathcal{X}$ into itself such that

$$T_{h_1}(T_{h_2}(x)) = T_{h_1 h_2}(x).$$

Again we assume that $\mathcal{H}$ is a σ-group and that the $\{T_h\}$ act continuously on $\mathcal{X}$, meaning that the mapping $(h, x) \mapsto T_h(x)$ is continuous from $\mathcal{H} \times \mathcal{X}$ into $\mathcal{X}$. The action of $\mathcal{H}$ splits $\mathcal{X}$ into equivalence classes, where $x_1 \equiv x_2$ if there exists $h \in \mathcal{H}$ such that $x_2 = T_h(x_1)$. It acts *transitively* on $\mathcal{X}$ if for every $x_1,\ x_2 \in \mathcal{X}$ there exists an h such that T_h maps x_1 into x_2. In this case, the equivalence relation is trivial: there exists only one equivalence class, the whole space $\mathcal{X}$. In general, the equivalence classes define a quotient space $\mathcal{Q}$, which may be given the quotient topology; with this topology, the natural map taking x into the equivalence class containing it is again both continuous and open. If the original topology on $\mathcal{H}$ is not adjusted to the group action, however, the quotient topology may not be adequate for a detailed discussion of invariant measures.

EXAMPLE A2.7(a). Consider $\mathbb{R}^1$ under the action of scale changes: $x \mapsto \alpha x$ $(0 < \alpha < \infty)$. Here $\mathcal{H}$ may be identified with the positive half-line $(0, \infty)$ with multiplication as the group action. There are three equivalence classes, $(-\infty, 0)$, $\{0\}$, and $(0, \infty)$, which we may identify with the three-point space $\mathcal{Q} = \{-1, 0, 1\}$. The quotient topology is trivial (only $\emptyset$ and the whole of $\mathcal{Q}$), whereas the natural topology for further discussion is the discrete topology on $\mathcal{Q}$, making each of the three points both open and closed in $\mathcal{Q}$. With this topology, the natural map is open but not continuous. It does have, however, a continuous (albeit trivial) restriction to each of the three equivalence classes and therefore defines a Borel mapping of $\mathcal{X}$ into $\mathcal{Q}$. □

An important problem is to determine the structure of boundedly finite measures on $\mathcal{X}$ that are invariant under the group of mappings $\{T_h\}$. In many cases, some or all of the equivalence classes of $\mathcal{X}$ under $\mathcal{H}$ can be identified with replicas of $\mathcal{H}$ so that we may expect the restriction of the invariant measures to such cosets to be proportional to Haar measure. When such an identification is possible, the following simple lemma can be used; it allows us to deal with most of the situations arising from concrete examples of invariant measures [see e.g. Bourbaki (1963) for further background].

Lemma A2.7.II (Factorization Lemma)**.** *Let $\mathcal{X} = \mathcal{H} \times \mathcal{Y}$, where $\mathcal{H}$ is a σ-group and $\mathcal{Y}$ is a c.s.m.s., and suppose that $\mu \in \mathcal{M}_{\mathcal{X}}^{\#}$ is invariant under left multiplication by elements of $\mathcal{H}$ in the sense that, for $A \in \mathcal{B}_{\mathcal{X}}$ and $B \in \mathcal{B}_{\mathcal{Y}}$,*

$$\mu(hA \times B) = \mu(A \times B). \tag{A2.7.3}$$

Then $\mu = \ell \times \kappa$, where ℓ is a multiple of left Haar measure on $\mathcal{H}$ and $\kappa \in \mathcal{M}_{\mathcal{Y}}^{\#}$ is uniquely determined up to a scalar multiple.

PROOF. Consider the set function $\mu_B(\cdot)$ defined on $\mathcal{B}_{\mathcal{H}}$ for fixed $B \in \mathcal{B}_{\mathcal{Y}}$ by

$$\mu_B(A) = \mu(A \times B).$$

Then μ_B inherits from μ the properties of countable additivity and bounded finiteness and so defines an element of $\mathcal{M}_{\mathcal{H}}^{\#}$. But then, from (A2.7.3),

$$\mu_B(hA) = \mu(hA \times B) = \mu(A \times B) = \mu_B(A),$$

implying that μ_B is invariant under left multiplication by elements of $\mathcal{H}$. It therefore reduces to a multiple of left Haar measure on $\mathcal{H}$,

$$\mu_B(A) = \kappa(B) = \ell(A), \qquad \text{say.}$$

Now the family of constants $\kappa(B)$ may be regarded as a set function on $\mathcal{B}_{\mathcal{Y}}$, and, as for μ_B, this function is both countably additive and boundedly finite. Consequently, $\kappa(\cdot) \in \mathcal{M}_{\mathcal{Y}}^{\#}$, and it follows that

$$\mu(A \times B) = \mu_B(A) = \ell(A)\kappa(B).$$

In other words, μ reduces to the required product form on product sets, and since these generate $\mathcal{B}_{\mathcal{X}}$, μ and the product measure $\ell \times \kappa$ coincide. □

To apply this result to specific examples, it is often necessary to find a suitable product representation for the space on which the transformations act. The situation is formalized in the following statement.

Proposition A2.7.III. *Let $\mathcal{X}$ be a c.s.m.s. acted on measurably by a group of transformations $\{T_h: h \in \mathcal{H}\}$, where $\mathcal{H}$ is a σ-group. Suppose, furthermore, that there exists a mapping $\psi: \mathcal{H} \times \mathcal{Y} \mapsto \mathcal{X}$, where $\mathcal{Y}$ is a c.s.m.s. and ψ is one-to-one, both ways measurable, takes bounded sets into bounded sets, and preserves the transformations $\{T_h\}$ in the sense that*

$$T_{h'}\psi(h, y) = \psi(h'h, y) \qquad (h' \in \mathcal{H}). \tag{A2.7.4}$$

Let μ be a measure on $\mathcal{M}_{\mathcal{X}}^{\#}$ that is invariant under the transformation T_h. Then there exists a unique invariant measure $\kappa \in \mathcal{M}_{\mathcal{Y}}^{\#}$ such that, for $\mathcal{B}_{\mathcal{X}}$-measurable nonnegative functions f,

$$\int_{\mathcal{X}} f(x)\,\mu(\mathrm{d}x) = \int_{\mathcal{Y}} \kappa(\mathrm{d}y) \int_{\mathcal{H}} f\big(\psi(h, y)\big)\,\ell(\mathrm{d}h). \tag{A2.7.5}$$

Proof. Let $\tilde{\mu}$ be the image of μ induced on $\mathcal{H} \times \mathcal{Y}$ by the mapping ψ; that is, $\tilde{\mu}(A \times B) = \mu\big(\psi(A \times B)\big)$. Then,

$$\tilde{\mu}(hA \times B) = \mu\big(\psi(hA \times B)\big) = \mu\big(T_h\psi(A \times B)\big) = \mu\big(\psi(A \times B)\big) = \tilde{\mu}(A \times B)$$

so that $\tilde{\mu}$ is invariant under the action of $h \in \mathcal{H}$ on the first argument. Moreover, if A and B are bounded sets in $\mathcal{H}$ and $\mathcal{Y}$, respectively, then by assumption $\psi(A \times B)$ is bounded in $\mathcal{X}$ so that $\tilde{\mu}$ is boundedly finite whenever μ is boundedly finite. Lemma A2.7.II can now be applied and yields the result that

$$\tilde{\mu}(A \times B) = \ell(A)\kappa(B)$$

for some unique boundedly finite measure κ in $\mathcal{M}_{\mathcal{Y}}^{\#}$. This relation establishes the truth of (A2.7.5) for indicator functions $I_{\psi(A\times B)}(x)$ for $A \in \mathcal{B}_{\mathcal{H}}$ and $B \in \mathcal{B}(\mathcal{M}_{\mathcal{Y}}^{\#})$. Using the usual approximation arguments, the result extends to simple functions f and thence to limits of these. It therefore holds for all nonnegative f such that $f \circ \psi$ is measurable on $\mathcal{H} \times \mathcal{Y}$. But this is true for any f that is $\mathcal{B}_{\mathcal{X}}$-measurable and so proves (A2.7.5). □

Example A2.7(b). Let μ be a measure on $\mathbb{R}^2$ that is invariant under rotations about the origin. These may be written T_θ for $\theta \in \mathbb{S}$, $\mathbb{S}$ denoting the circumference of the unit disk with addition modulo 2π. The equivalence classes consist of circles of varying radii centred on the origin, together with the isolated point $\{0\}$. The mapping $(r, \theta) \mapsto (r\cos\theta, r\sin\theta)$ takes the product space $\mathbb{S} \times \mathbb{R}_+$ into $\mathbb{R}^2 \setminus \{0\}$ and is a representation of the required kind for $\mathbb{R}^2 \setminus \{0\}$. We therefore write μ as the sum of a point mass at the origin and a measure on $\mathbb{R}^2 \setminus \{0\}$ that is invariant under rotations and can therefore be represented as the image of the uniform distribution around the circle and a measure κ on the positive half-line. Integration with respect to μ takes the form [see (A2.7.5)]

$$\int_{\mathbb{R}^2} f(x)\,\mu(\mathrm{d}x) = f(0)\mu(\{0\}) + \int_{0+}^{\infty} \kappa(\mathrm{d}r) \int_0^{2\pi} f(r\cos\theta, r\sin\theta)\,\frac{\mathrm{d}\theta}{2\pi}\,. \qquad \square$$

A2.8. Fourier Transforms

In this section, we collect together a few basic facts from classical Fourier transform theory. For brevity, most results are stated for Fourier transforms of functions on $\mathbb{R} \equiv \mathbb{R}^1$; the corresponding results for $\mathbb{R}^d$ can be obtained by no more than changes in the domain of integration and appropriate bookkeeping with multiples of 2π. Both the $\mathbb{R}^d$ theory and the theory of Fourier series, which can be regarded as Fourier transforms of functions defined on the unit circle, are subsumed under the concluding comments concerned with Fourier transforms of functions defined on locally compact Abelian groups. We refer to texts such as Titchmarsh (1937) for more specific material on these topics.

For any real- or complex-valued measurable (Lebesgue) integrable function $f(\cdot)$, its Fourier transform $\tilde{f}(\cdot)$ is defined by

$$\tilde{f}(\omega) = \int_{-\infty}^{\infty} \mathrm{e}^{i\omega x} f(x)\,\mathrm{d}x \qquad (\omega \in \mathbb{R}). \tag{A2.8.1}$$

If f is real and symmetric, then so is $\tilde{f}$. In any case, $\tilde{f}$ is bounded and continuous, while the Riemann–Lebesgue lemma asserts that $f(\omega) \to 0$ as $|\omega| \to \infty$. Furthermore, if $\tilde{f}$ is integrable, then the inverse relation

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{ix\omega} \tilde{f}(\omega)\,\mathrm{d}\omega \tag{A2.8.2}$$

holds. The theory is not symmetric with respect to f and $\tilde{f}$: for a more detailed account of the representation of a function by its inverse Fourier transform, see, for example, Titchmarsh (1937).

A symmetrical theory results if we consider (real- or complex-valued) functions that are square integrable. We have the *Plancherel identities* for square integrable functions f and g,

$$\int_{-\infty}^{\infty} f(x)\overline{g(x)}\,\mathrm{d}x = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{f}(\omega)\overline{\tilde{g}(\omega)}\,\mathrm{d}\omega, \tag{A2.8.3}$$

and, with $g = f$,

$$\int_{-\infty}^{\infty} \left|f(x)\right|^2 \mathrm{d}x = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left|\tilde{f}(\omega)\right|^2 \mathrm{d}\omega. \tag{A2.8.4}$$

Here the Fourier transform cannot be obtained directly from (A2.8.1) but can be represented as a mean square limit

$$\tilde{f}(\omega) = \operatorname*{l.i.m.}_{T\to\infty} \int_{-T}^{T} \mathrm{e}^{i\omega x} f(x)\,\mathrm{d}x, \tag{A2.8.5}$$

the existence of the finite integral following readily from the Schwarz inequality. Since the limit is defined only up to an equivalence, the theory is strictly between equivalence classes of functions—that is, elements of the Hilbert space $L_2(\mathbb{R})$—rather than a theory between individual functions.

An important version for probability theory is concerned with the Fourier transforms of totally finite measures (or signed measures). If G is such a measure, its *Fourier–Stieltjes transform* $\tilde{g}$ is the bounded uniformly continuous function

$$\tilde{g}(\omega) = \int_{-\infty}^{\infty} \mathrm{e}^{i\omega x}\, G(\mathrm{d}x). \tag{A2.8.6}$$

If G is a probability measure, $\tilde{g}(\omega)$ is its characteristic function and $\tilde{g}$ is then a *positive-definite function*: for arbitrary finite families of real numbers $\omega_1, \ldots, \omega_r$ and complex numbers $\alpha_1, \ldots, \alpha_r$,

$$\sum_{i=1}^{r}\sum_{j=1}^{r} \alpha_i \bar{\alpha}_j \tilde{g}(\omega_i - \omega_j) \geq 0. \tag{A2.8.7}$$

Conversely, *Bochner's theorem* asserts that any function continuous at $\omega = 0$ and satisfying (A2.8.7) can be represented as the Fourier transform of a totally finite measure G on $\mathbb{R}$ with $G(\mathbb{R}) = \tilde{g}(0)$. If we take any real or complex integrable function f with any totally finite signed measure G and apply Fubini's theorem to the double integral

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{i\omega x} f(\omega)\, G(\mathrm{d}x)\, \mathrm{d}\omega,$$

which is certainly well defined, we obtain *Parseval's identity*:

$$\int_{-\infty}^{\infty} \tilde{f}(x)\, G(\mathrm{d}x) = \int_{-\infty}^{\infty} f(\omega)\tilde{g}(\omega)\, \mathrm{d}\omega. \tag{A2.8.8}$$

This identity is of basic importance because it shows that G is uniquely determined by $\tilde{g}$. Various more specific inversion theorems can be obtained by taking suitable choices of f followed by a passage to the limit: this approach is outlined in Feller (1966, Section XV.3), for example. In particular, the following two forms are traditional.

(i) For continuity intervals (a, b) of G,

$$G((a,b)) = \lim_{T\to\infty} \int_{-T}^{T} \frac{\mathrm{e}^{-i\omega a} - \mathrm{e}^{-i\omega b}}{i\omega}\, \tilde{g}(\omega)\, \mathrm{d}\omega.$$

(ii) For an atom a of G,

$$G(\{a\}) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \mathrm{e}^{-i\omega a} \tilde{g}(\omega)\, \mathrm{d}\omega.$$

Much of the preceding theory can be extended without difficulty from $\mathbb{R}$ to the case of a locally compact Abelian topological group $\mathcal{G}$. The *characters* of such a group are the continuous homomorphisms of the group into the complex numbers of modulus 1. If χ_1, χ_2 are characters, then so are $\chi_1\chi_2$ and χ_1^{-1}. Thus, the characters form a group in their own right, $\widetilde{\mathcal{G}}$ say, the *dual group* for $\mathcal{G}$. There is a natural topology on $\widetilde{\mathcal{G}}$, namely the smallest making the evaluation mapping $e_g(\chi) \equiv \chi(g)$ continuous for each $g \in \mathcal{G}$, and with this topology $\widetilde{\mathcal{G}}$ also is a locally compact Abelian topological group. If $\mathcal{G} = \mathbb{R}$, the characters are of the form $\mathrm{e}^{i\omega x}$ ($\omega \in \mathbb{R}$), and $\widetilde{\mathcal{G}}$ can be identified with another version of $\mathbb{R}$. If $\mathcal{G} = \mathbb{Z}$, the group of integers, $\widetilde{\mathcal{G}}$ is the circle group, and vice versa. In any case, the original group reappears as the dual of the dual group $\widetilde{\mathcal{G}}$, and if $\mathcal{G}$ is compact, $\widetilde{\mathcal{G}}$ is discrete and conversely.

Now let H and $\widetilde{H}$ denote Haar measure on $\mathcal{G}$ and $\widetilde{\mathcal{G}}$, respectively. If $f\colon \mathcal{G} \mapsto \mathbb{R}$ is measurable and H-integrable, its Fourier transform $\tilde{f}$ is the function defined on $\widetilde{\mathcal{G}}$ by

$$\tilde{f}(\chi) = \int_{\mathcal{G}} \chi(g) f(g)\, H(\mathrm{d}g). \tag{A2.8.9}$$

If also $\tilde{f}$ is $\widetilde{H}$-integrable, then the inverse relation

$$f(g) = \int_{\widetilde{\mathcal{G}}} \overline{\chi(g)} \tilde{f}(\chi)\, \widetilde{H}(\mathrm{d}\chi) \tag{A2.8.10}$$

holds, provided that $\widetilde{H}$ is normed appropriately [otherwise, a normalizing constant such as $1/(2\pi)$ in (A2.8.2) is needed]. Assuming that such a norming has been adopted, the appropriate analogues of (A2.8.4–8) remain true. In particular, we note the generalized Plancherel identity

$$\int_{\mathcal{G}} \left|f(g)\right|^2 H(\mathrm{d}g) = \int_{\widetilde{\mathcal{G}}} \left|\tilde{f}(\chi)\right|^2 \widetilde{H}(\mathrm{d}\chi). \tag{A2.8.11}$$

APPENDIX 3

Conditional Expectations, Stopping Times, and Martingales

This appendix contains mainly background material for Chapter 14. For further discussion and most proofs, we refer the reader to Ash (1972), Chung (1974), Brémaud (1981), and to various references cited in the text.

A3.1. Conditional Expectations

Let $(\Omega, \mathcal{E}, \mathcal{P})$ be a probability space (see Section A1.4), X a random variable (r.v.) with $\mathrm{E}|X| = \int_\Omega |X|\,\mathcal{P}(\mathrm{d}\omega) < \infty$, and $\mathcal{G}$ a sub-σ-algebra of events from $\mathcal{E}$. The *conditional expectation* of X with respect to $\mathcal{G}$, written $\mathrm{E}(X \mid \mathcal{G})$ or $\mathrm{E}_{X|\mathcal{G}}(\omega)$, is the $\mathcal{G}$-measurable function (i.e. a random variable) defined up to values on a set of $\mathcal{G}$ of $\mathcal{P}$-measure zero as the Radon–Nikodym derivative

$$\mathrm{E}(X \mid \mathcal{G}) = \mathrm{E}_{X|\mathcal{G}}(\omega) = \xi_X^{(\mathcal{G})}(\mathrm{d}\omega)/\mathcal{P}^{(\mathcal{G})}(\mathrm{d}\omega),$$

where $\xi_X(A) = \int_A X(\omega)\,\mathcal{P}(\mathrm{d}\omega)$ is the indefinite integral of X and the superscript $(\mathcal{G})$ indicates that the set functions are to be restricted to $\mathcal{G}$.

The $\mathcal{G}$-measurability of $\mathrm{E}(X \mid \mathcal{G})$ implies that

$$\int_U X(\omega)\,\mathcal{P}(\mathrm{d}\omega) = \int_U \mathrm{E}_{X|\mathcal{G}}(\omega)\,\mathcal{P}(\mathrm{d}\omega) \qquad (\text{all } U \in \mathcal{G}), \tag{A3.1.1}$$

an equation, usually taken as the defining relation, that determines the conditional expectation uniquely. Extending (A3.1.1) from $\mathcal{G}$-measurable indicator functions $I_U(\omega)$ to more general $\mathcal{G}$-measurable functions Y, we have, whenever $\mathrm{E}(|X|)$ and $\mathrm{E}(|XY|)$ exist,

$$\mathrm{E}(XY) = \int_\Omega Y(\omega)X(\omega)\,\mathcal{P}(\mathrm{d}\omega) = \int_\Omega Y(\omega)\mathrm{E}_{X|\mathcal{G}}(\omega)\,\mathcal{P}(\mathrm{d}\omega) = \mathrm{E}[Y\,\mathrm{E}(X \mid \mathcal{G})]. \tag{A3.1.2}$$

Now replacing Y by YI_U for $U \in \mathcal{G}$ and using (A3.1.1), there follows the factorization property of conditional expectations that for $\mathcal{G}$-measurable r.v.s Y for which both $\mathrm{E}(|X|)$ and $\mathrm{E}(|XY|)$ exist,

$$\mathrm{E}(XY \mid \mathcal{G}) = Y\,\mathrm{E}(X \mid \mathcal{G}) \quad \text{a.s.} \tag{A3.1.3}$$

Conditional expectations inherit many standard properties of ordinary expectations:

Linearity: $$\mathrm{E}\Big(\sum_{j=1}^{k} \alpha_j X_j \,\Big|\, \mathcal{G}\Big) = \sum_{j=1}^{k} \alpha_j \mathrm{E}(X_j \mid \mathcal{G}); \tag{A3.1.4}$$

Monotonicity: $X \le Y$ a.s. implies $\mathrm{E}(X \mid \mathcal{G}) \le \mathrm{E}(Y \mid \mathcal{G})$ a.s.; (A3.1.5)

Monotone convergence: $X_n \ge 0$ and $X_n \uparrow Y$ a.s. imply that

$$\mathrm{E}(X_n \mid \mathcal{G}) \uparrow \mathrm{E}(Y \mid \mathcal{G}) \quad \text{a.s.}; \tag{A3.1.6}$$

Jensen's inequality: For convex measurable functions $f\colon \mathbb{R} \mapsto \mathbb{R}$ for which $\mathrm{E}[|f(X)|] < \infty$,

$$f(\mathrm{E}[X \mid \mathcal{G}]) \le \mathrm{E}[f(X) \mid \mathcal{G}] \quad \text{a.s.} \tag{A3.1.7}$$

[in (A3.1.7), convexity means that $f\big(\frac{1}{2}(x+y)\big) \le \frac{1}{2}[f(x) + f(y)]$.]

If $\mathcal{G}_1$ and $\mathcal{G}_2$ are two sub-σ-algebras with $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{E}$ and $\mathrm{E}(|X|) < \infty$ as before, the *repeated conditioning* theorem holds:

$$\mathrm{E}[\mathrm{E}(X \mid \mathcal{G}_1) \mid \mathcal{G}_2] = \mathrm{E}[\mathrm{E}(X \mid \mathcal{G}_2) \mid \mathcal{G}_1] = \mathrm{E}(X \mid \mathcal{G}_1), \tag{A3.1.8}$$

yielding as the special case when $\mathcal{G} = \{\emptyset, \Omega\}$

$$\mathrm{E}[\mathrm{E}(X \mid \mathcal{G})] = \mathrm{E}(X). \tag{A3.1.9}$$

Two σ-algebras $\mathcal{G}$ and $\mathcal{H}$ are *independent* if, for all $A \in \mathcal{G}$ and $B \in \mathcal{H}$, $\mathcal{P}(A \cap B) = \mathcal{P}(A)\mathcal{P}(B)$. Given such $\mathcal{G}$ and $\mathcal{H}$, if X is $\mathcal{G}$-measurable and we seek $\mathrm{E}(X \mid \mathcal{H})$, we may expect it to reduce to yield

$$E(X \mid \mathcal{H}) = \mathrm{E}(X). \tag{A3.1.10}$$

This is a special case of the principle of *redundant conditioning*: if the r.v. X is independent of $\mathcal{H}$ [i.e. $\sigma(X)$ and $\mathcal{H}$ are independent σ-algebras] and $\mathcal{G}$ is independent of $\mathcal{H}$, then

$$\mathrm{E}(X \mid \mathcal{G} \vee \mathcal{H}) = \mathrm{E}(X \mid \mathcal{G}), \tag{A3.1.11}$$

reducing to (A3.1.10) for trivial $\mathcal{G}$.

Let $\mathcal{X}$ be a c.s.m.s. and X an $\mathcal{X}$-valued r.v. on $(\Omega, \mathcal{E}, \mathcal{P})$. Given a sub-$\sigma$-algebra $\mathcal{G}$ of $\mathcal{E}$, the *conditional distribution* of X given $\mathcal{G}$ is defined by analogy with (A3.1.1) by

$$\mathcal{P}(X \in A \mid \mathcal{G}) = \mathrm{E}(I_A(X) \mid \mathcal{G}) \qquad (A \in \mathcal{B}_{\mathcal{X}}). \tag{A3.1.12}$$

As in Section A1.5, the question of the existence of *regular* conditional distributions arises. In our present context, we seek a kernel function

$$Q(A,\omega) \qquad (A \in \mathcal{B}(\mathcal{X}),\ \omega \in \Omega)$$

such that for fixed A, $Q(A,\cdot)$ is a $\mathcal{G}$-measurable function of ω [and we identify this with (A3.1.12)], while for fixed ω, we want $Q(\cdot,\omega)$ to be a probability measure on $\mathcal{B}(\mathcal{X})$. Introduce the set function $\pi(\cdot)$ defined initially for product sets $A \times U$ for $A \in \mathcal{B}(\mathcal{X})$ and $U \in \mathcal{G}$ by

$$\pi(A \times U) = \int_U I_A(X(\omega))\,\mathcal{P}(\mathrm{d}\omega). \tag{A3.1.13}$$

Since $\pi(\cdot)$ is countably additive on such sets, it can be extended to a measure, clearly a probability, on $(\mathcal{X} \times \Omega, \mathcal{B}(\mathcal{X}) \otimes \mathcal{G})$. Then Proposition A1.5.III can be applied and yields the following formal statement in which we identify the kernel function $Q(\cdot,\cdot)$ sought above with $\mathcal{P}(X \in A \mid \mathcal{G})$.

Proposition A3.1.I. *Let $\mathcal{X}$ be a c.s.m.s., $(\Omega, \mathcal{E}, \mathcal{P})$ a probability space, and X an $\mathcal{X}$-valued r.v. defined on $(\Omega, \mathcal{E}, \mathcal{P})$. If $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{E}$, then there exists a regular version of the conditional distribution $\mathcal{P}_{X\in\cdot|\mathcal{G}}(\omega)$ such that*

(i) *$\mathcal{P}_{X\in\cdot|\mathcal{G}}(\omega)$ is a probability measure on $\mathcal{B}(\mathcal{X})$ for each fixed ω;*
(ii) *$\mathcal{P}_{X\in A|\mathcal{G}}(\cdot)$ is a $\mathcal{G}$-measurable function of ω for fixed $A \in \mathcal{B}(\mathcal{X})$; and*
(iii) *for each $U \in \mathcal{G}$ and $A \in \mathcal{B}(\mathcal{X})$,*

$$\int_U \mathcal{P}_{X\in A|\mathcal{G}}(\omega)\,\mathcal{P}(\mathrm{d}\omega) = \int_U I_A(X(\omega))\,\mathcal{P}(\mathrm{d}\omega). \tag{A3.1.14}$$

Observe that if $\mathcal{G} = \mathcal{E}$, then the conditional distribution $\mathcal{P}_{X\in\cdot|\mathcal{G}}(\omega)$ is the degenerate distribution concentrated on the point $X(\omega)$. In general, the conditional distribution represents a blurred image of this degenerate distribution, the blurring arising as a result of the incomplete information concerning X carried by the sub-σ-algebra $\mathcal{G}$.

The following question is of the nature of a converse to the proposition. Given $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, $(\Omega, \mathcal{E}, \mathcal{P})$ and a regular kernel $Q(A,\omega)$, can we find a refinement $\mathcal{E}' \supseteq \mathcal{E}$ and an $\mathcal{E}'$-measurable $\mathcal{X}$-valued r.v. X such that $Q(A,\omega)$ coincides with $\mathcal{P}_{X\in A|\mathcal{G}}(\omega)$? If we confine ourselves to the original space, this may not necessarily be possible, but by extending Ω we can accomplish our aim. Take the probability space $(\Omega', \mathcal{E}', \mathcal{P}')$ given by $\Omega' = \mathcal{X} \times \Omega$, $\mathcal{E}' = \mathcal{B}(\mathcal{X}) \otimes \mathcal{E}$ and $\mathcal{P}' = \pi$ as constructed via (A3.1.13) (identifying $\mathcal{G}$ there with $\mathcal{E}$ here), and consider the r.v. $X\colon \mathcal{X} \times \Omega \mapsto \mathcal{X}$ for which $X(\omega') = X(x,\omega) = x$. With the mapping $T\colon \Omega' \mapsto \Omega$ for which $T(\omega') = T(x,\omega) = \omega$, so that $T^{-1}(\mathcal{E})$ is a sub-σ-algebra of $\mathcal{E}'$, we then have

$$\mathcal{P}'_{X\in A|T^{-1}(\mathcal{E})}(\omega') = Q(A, T(\omega')) = Q(A,\omega) \qquad (A \in \mathcal{B}(\mathcal{X})). \tag{A3.1.15}$$

Often the conditioning σ-algebra $\mathcal{G}$ is itself generated by some real- or (more generally) c.s.m.s.-valued r.v. Y. Then $\mathrm{E}(X \mid \mathcal{G})$ is called the conditional expectation of X given Y and $\mathcal{P}(X \in A \mid \mathcal{G})$ the conditional distribution of X given Y, together with the suggestive notation $\mathrm{E}(X \mid Y)$ or $\mathrm{E}_{X|Y}(\omega)$ and $\mathcal{P}(X \in A \mid \mathcal{G})$ or $\mathcal{P}_{X\in A|\mathcal{G}}(\omega)$. Equation (A3.1.13) then implies, for any Borel-measurable function $h(\cdot)$ such that the unconditional expectations exist,

$$\mathrm{E}[Xh(Y) \mid Y] = h(Y)\,\mathrm{E}(X \mid Y). \tag{A3.1.16}$$

The terminology suggests that, although $\mathrm{E}(X \mid Y)$ is defined as an r.v., its value should depend on ω only through $Y(\omega)$. Thus, if Y takes its values in a c.s.m.s. $\mathcal{Y}$, we should look for a real-valued $\mathcal{B}(\mathcal{Y})$-measurable function $h_{X|Y}(y)$ such that

$$E_{X|Y}(\omega) = h_{X|Y}\big(Y(\omega)\big) \quad \text{a.s.} \tag{A3.1.17}$$

That such a function exists is the assertion of the *Doob representation theorem* (e.g. Doob, 1953). It can be established by applying the argument around (A3.1.1) to the measures induced on $\mathcal{B}(\mathcal{Y})$ by the equations

$$\mathcal{P}_Y(B) = \mathcal{P}(Y^{-1}(B)) \qquad (B \in \mathcal{B}(\mathcal{Y})),$$
$$\xi_X(B) = \int_{Y^{-1}(B)} X(\omega)\,\mathcal{P}(\mathrm{d}\omega),$$

and, noting that $\xi_X \ll \mathcal{P}_Y$ on $\mathcal{B}(\mathcal{Y})$, by applying the Radon–Nikodym theorem. Since the product of a finite or denumerably infinite number of c.s.m.s.s can itself be regarded as a c.s.m.s., we state the theorem in the following general form.

Proposition A3.1.II. *Let $(\Omega, \mathcal{E}, \mathcal{P})$ be a probability space, X an integrable real-valued r.v. on Ω, and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{E}$ generated by a countable family of r.v.s $Y = \{Y_1, Y_2, \ldots\}$ taking their values in the c.s.m.s.s $\mathcal{Y}_1, \mathcal{Y}_2, \ldots$ respectively. Then, there exists a Borel-measurable function $h_{X|Y}(\cdot)\colon \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \mapsto \mathbb{R}$ such that*

$$\mathrm{E}_{X|\mathcal{G}}(\omega) = h_{X|Y}(Y_1(\omega), Y_2(\omega), \ldots) \quad \mathcal{P}\text{-a.s.} \tag{A3.1.18}$$

The proposition concerning regular conditional distributions can be transformed in a similar way, yielding a kernel $\mathcal{P}_{X\in A|Y}(y_1, y_2, \ldots)$, which is a probability distribution in A for each vector $(y_1, y_2, \ldots)$, a Borel-measurable function of the family $(y_1, y_2, \ldots)$ for each A, and satisfies

$$\mathcal{P}_{X\in A|\mathcal{G}}(\omega) = \mathcal{P}_{X\in A|Y}(Y_1(\omega), Y_2(\omega), \ldots) \quad \mathcal{P}\text{-a.s.}$$

When densities exist with respect to some underlying measure μ such as Lebesgue measure on $\mathbb{R}^d$, the conditional distributions have the form

$$\mathcal{P}_{X\in A|Y}(y_1, y_2, \ldots) = \frac{\int_A f_{X,Y}(x, y_1, y_2, \ldots)\,\mu(\mathrm{d}x)}{\int_{\mathcal{X}} f_{X,Y}(x, y_1, y_2, \ldots)\,\mu(\mathrm{d}x)},$$

where $f_{X,Y}(\cdot)$ is the joint density for $X, Y_1, Y_2, \ldots$ in the product space $\mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots$, and a similar representation holds for the conditional expectation $h_{X|Y}(\cdot)$.

A3.2. Convergence Concepts

Most of the different notions of convergence and uniform integrability mentioned below are standard. Stable convergence is less familiar and is discussed in more detail.

A sequence of r.v.s $\{X_n: n = 1, 2, \ldots\}$ on a common probability space $(\Omega, \mathcal{E}, \mathcal{P})$ *converges in probability* to a limit r.v. X, also defined on $(\Omega, \mathcal{E}, \mathcal{P})$, if for all $\epsilon > 0$,

$$\mathcal{P}\{|X_n - X| > \epsilon\} \to 0 \qquad (n \to \infty). \tag{A3.2.1}$$

The sequence *converges almost surely* to X if

$$\begin{aligned} 1 &= \mathcal{P}\{\omega: X_n(\omega) \to X(\omega) \ \ (n \to \infty)\} \\ &= \mathcal{P}\Big(\bigcap_{r=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m \geq n} \Big\{\omega: \big|X_m(\omega) - X(\omega)\big| < \frac{1}{r}\Big\}\Big) \\ &= \mathcal{P}\Big(\bigcap_{r=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m \geq n} \Big\{\omega: \big|X_m(\omega) - X_n(\omega)\big| < \frac{1}{r}\Big\}\Big). \end{aligned} \tag{A3.2.2}$$

Both these concepts readily generalize to the case where the r.v.s X and X_n are $\mathcal{X}$-valued for some c.s.m.s. $\mathcal{X}$ by simply replacing the Euclidean distance $|X - Y|$ by the metric $\rho(X, Y)$ for $X, Y \in \mathcal{X}$. The a.s. convergence in (A3.2.2) implies convergence in probability; convergence in probability implies the existence of a subsequence $\{X_{n_k}\}$ that converges a.s. to the same limit.

Returning to the real-valued case, for any given real $p \geq 1$, $\{X_n\}$ *converges in the mean of order* p (or *in pth mean*, or *in* L_p *norm*) if the pth moments exist and

$$\|X_n - X\|_p \equiv [\mathrm{E}(|X_n - X|^p)]^{1/p} \to 0 \qquad (n \to \infty), \tag{A3.2.3}$$

the norm here denoting the norm in the Banach space $L_p(\Omega, \mathcal{E}, \mathcal{P})$ of equivalence classes of r.v.s with finite pth moments. Mean square convergence—i.e. convergence in L_2 norm—has its own notation l.i.m. (Doob, 1953, p. 8) as in Section 8.4. For $p = \infty$, the space $L_\infty(\Omega, \mathcal{E}, \mathcal{P})$ consists of $\mathcal{P}$-*essentially bounded* r.v.s X; that is, r.v.s X for which $|X| \leq M$ a.s. for some $M < \infty$; then

$$\|X\|_\infty = \operatorname{ess\,sup} |X(\omega)| = \inf\{M: |X(\omega)| \leq M \ \text{ a.s.}\}. \tag{A3.2.4}$$

If $X_n \to X$ in pth mean, then $\mathrm{E}(X_n^p) \to \mathrm{E}(X^p)$ $(n \to \infty)$.

Chebyshev's inequality, in the form for an L_p r.v. X,

$$\mathcal{P}\{|X - a| > \epsilon\} \leq \epsilon^{-p}\mathrm{E}(|X - a|^p) \qquad (\epsilon > 0, \text{ real } a), \tag{A3.2.5}$$

shows that convergence in L_p norm implies convergence in probability. The converse requires the additional condition of *uniform integrability*.

Definition A3.2.I. *A family of real-valued r.v.s* $\{X_t: t \in \mathcal{T}\}$ *defined on the common probability space* $(\Omega, \mathcal{E}, \mathcal{P})$ *is* uniformly integrable *if, given* $\epsilon > 0$, *there exists* $M < \infty$ *such that*

$$\int_{|X_t|>M} |X_t(\omega)|\,\mathcal{P}(\mathrm{d}\omega) < \epsilon \qquad (\textit{all } t \in \mathcal{T}). \tag{A3.2.6}$$

Proposition A3.2.II. *Let the r.v.s $\{X_n: n = 1, 2, \ldots\}$ and X be defined on a common probability space $(\Omega, \mathcal{E}, \mathcal{P})$ and be such that $X_n \to X$ in probability. Then, a necessary and sufficient condition for the means to exist and for $X_n \to X$ in L_1 norm is that the sequence $\{X_n\}$ be uniformly integrable.*

Applied to the sequence $\{X_n^p\}$ and noting the inequality $\mathrm{E}(|X_n - X|^p) \le 2^{p-1}[\mathrm{E}(|X_n|^p) + \mathrm{E}(|X|^p)]$ $(1 \le p < \infty)$, the proposition extends in an obvious way to convergence in L_p norm for $1 \le p < \infty$.

A weaker concept than convergence in L_p norm [i.e. strong convergence in the Banach space $L_p(\Omega, \mathcal{E}, \mathcal{P})$] is that of *weak L_p convergence*, namely, that if X_n and $X \in L_p$, then $\mathrm{E}(X_n Y) \to \mathrm{E}(XY)$ $(n \to \infty)$ for all $Y \in L_q$, where $p^{-1} + q^{-1} = 1$.

Let X_n be $\mathcal{X}$-valued for a c.s.m.s. $\mathcal{X}$ with metric ρ. *X_n converges in distribution* if $\mathcal{P}\{X_n \in A\} \to \mathcal{P}\{X \in A\}$ for all $A \in \mathcal{B}(\mathcal{X})$ for which $\mathcal{P}\{X \in \partial A\} = 0$. This type of convergence is not so much a constraint on the r.v.s as a constraint on the distributions they induce on $\mathcal{B}(\mathcal{X})$: indeed, it is precisely weak convergence of their induced distributions. If $X_n \to X$ in probability (or, *a fortiori*, if $X_n \to X$ a.s. or in L_p norm), then from the inequalities

$$\begin{aligned}\mathcal{P}\{X_n \in A\} - \mathcal{P}\{X \in A\} &\le \mathcal{P}\big(\{X_n \in A\} \cap \{X \in A\}\big) \\ &\le \mathcal{P}\big(\{X_n \in A\} \cap \{X \in (A^\epsilon)^c\}\big) + \mathcal{P}\{X \in A^\epsilon\} - \mathcal{P}\{X \in A\} \\ &\le \mathcal{P}\{\rho(X_n, X) > \epsilon\} + \mathcal{P}\{X \in A^\epsilon\} - \mathcal{P}\{X \in A\},\end{aligned}$$

it follows that $X_n \to X$ in distribution, also written $X_n \to_d X$. No general converse statement is possible except when X is *degenerate*; that is, $X = a$ a.s. for some $a \in \mathcal{X}$. For this exceptional case, $X_n \to_d a$ means that for any positive ϵ, $\mathcal{P}\{\rho(X_n, a) < \epsilon\} = \mathcal{P}\{X_n \in S_\epsilon(a)\} \to 1$ $(n \to \infty)$, and this is the same as $X_n \to a$ in probability.

A hybrid concept, in the sense that it depends partly on the r.v.s X_n themselves and partly on their distributions, is that of *stable convergence*.

Definition A3.2.III. *If $\{X_n: n = 1, 2, \ldots\}$ and X are $\mathcal{X}$-valued r.v.s on $(\Omega, \mathcal{E}, \mathcal{P})$ and $\mathcal{F}$ is a sub-σ-algebra of $\mathcal{E}$, then $X_n \to X$* ($\mathcal{F}$-stably) *in distribution if for all $U \in \mathcal{F}$ and all $A \in \mathcal{B}(\mathcal{X})$ with $\mathcal{P}\{X \in \partial A\} = 0$,*

$$\mathcal{P}(\{X_n \in A\} \cap U) \to \mathcal{P}(\{X \in A\} \cap U) \qquad (n \to \infty). \tag{A3.2.7}$$

The hybrid nature of stable convergence is well illustrated by the facts that when $\mathcal{F} = \{\emptyset, \Omega\}$, $\mathcal{F}$-stable convergence is convergence in distribution, whereas when $\mathcal{F} \supseteq \sigma(X)$, we have a.s. convergence in probability because the regular version $\mathcal{P}_{X \in A|\mathcal{F}}(\omega)$ of the conditional distribution appearing in $\mathcal{P}(\{X \in A\} \cap U) = \int_U \mathcal{P}_{X \in A|\mathcal{F}}(\omega)\,\mathcal{P}(\mathrm{d}\omega)$ can be taken as being $\{0, 1\}$-valued, and when such degenerate distributions for the limit r.v. occur, the concepts of convergence in distribution and in probability coincide, as already noted.

In general, stable convergence always implies weak convergence, and it may be regarded as a form of weak convergence of the conditional distributions $\mathcal{P}(X_n \in A \mid \mathcal{F})$. Just as weak convergence can be expressed in equivalent ways, so too can stable convergence as follows (see Aldous and Eagleson, 1978).

Proposition A3.2.IV. *Let* $\{X_n\}$, X *and* $\mathcal{F}$ *be as in Definition* A3.2.III. *Then, the conditions* (i)–(iv) *below are equivalent.*

(i) $X_n \to X$ ($\mathcal{F}$*-stably); that is,* (A3.2.7) *holds.*

(ii) *For all* $\mathcal{F}$*-measurable* $\mathcal{P}$*-essentially bounded r.v.s* Z *and all bounded continuous* $h\colon \mathcal{X} \mapsto \mathbb{R}$,

$$\mathrm{E}[Zh(X_n)] \to \mathrm{E}[Zh(X)] \qquad (n \to \infty). \tag{A3.2.8}$$

(iii) *For all real-valued* $\mathcal{F}$*-measurable r.v.s* Y, *the pairs* (X_n, Y) *converge jointly in distribution to the pair* (X, Y).

(iv) *For all bounded continuous functions* $g\colon \mathcal{X} \times \mathbb{R} \mapsto \mathbb{R}$ *and all real-valued* $\mathcal{F}$*-measurable r.v.s* Y,

$$g(X_n, Y) \to g(X, Y) \qquad (\mathcal{F}\text{-stably}). \tag{A3.2.9}$$

If $\mathcal{X} = \mathbb{R}^d$, *then any of* (i)–(iv) *is equivalent to condition* (v).

(v) *For all real vectors* $t \in \mathbb{R}^d$ *and all* $\mathcal{P}$*-essentially bounded* $\mathcal{F}$*-measurable r.v.s* Z,

$$\mathrm{E}[Z \exp(it'X_n)] \to \mathrm{E}[Z \exp(it'X)] \qquad (n \to \infty). \tag{A3.2.10}$$

Proof. Equation (A3.2.7) is the special case of (A3.2.8) with $Z = I_U(\omega)$ and $h(x) = I_A(x)$ for $U \in \mathcal{F}$ and $A \in \mathcal{B}(\mathcal{X})$, except that such $h(\cdot)$ is not in general continuous: as in the continuity theorem for weak convergence, (A3.2.8) can be extended to the case where h is bounded and Borel measurable and $\mathcal{P}\{X \in \partial h\} = 0$, where ∂h is the set of discontinuities of h. When $\mathcal{X} = \mathbb{R}^d$, (A3.2.10) extends the well-known result that joint convergence of characteristic functions is equivalent to weak convergence of distributions. Note that all of (A3.2.7), (A3.2.8), and (A3.2.10) are contracted versions of the full statement of weak convergence in L_1 of the conditional distributions; namely, that

$$\mathrm{E}(Z\,\mathrm{E}[h(X_n) \mid \mathcal{F}]) \to \mathrm{E}(Z\,\mathrm{E}[h(X) \mid \mathcal{F}]) \qquad (n \to \infty) \tag{A3.2.11}$$

for arbitrary (not necessarily $\mathcal{F}$-measurable) r.v.s Z. However, (A3.2.11) can immediately be reduced to the simpler contracted forms by using the repeated conditioning theorem, which shows first that it is enough to consider the case that Z is $\mathcal{F}$-measurable and second that when Z is $\mathcal{F}$-measurable, the conditioning on $\mathcal{F}$ can be dropped.

If Y is real-valued and $\mathcal{F}$-measurable and in (A3.2.7) we set $U = Y^{-1}(B)$ for $B \in \mathcal{B}(\mathbb{R})$, we obtain

$$\mathcal{P}\{(X_n, Y) \in A \times B\} \to \mathcal{P}\{(X, Y) \in A \times B\},$$

from which (iii) follows. Conversely, taking $Y = I_U$ in (iii) yields (A3.2.7).

Finally, for any two real-valued $\mathcal{F}$-measurable r.v.s Y, Z, repeated application of (iii) shows that (X_n, Y, Z) converges weakly in distribution to the triple (X, Y, Z). Applying the continuous mapping theorem (Proposition A2.2.VII) yields the result that the pairs $(g(X_n, Y), Z)$ converge weakly in distribution to $(g(X, Y), Z)$, which is equivalent to the stable convergence of $g(X_n, Y)$ to $g(X, Y)$ by (iii). Since stable convergence implies weak convergence, (iv) implies (iii). □

When the limit r.v. is independent of the conditioning σ-algebra $\mathcal{F}$, we have a special case of some importance: (A3.2.7) and (A3.2.10) then reduce to the forms

$$\mathcal{P}(X_n \in A \mid U) \to \mathcal{P}\{X \in A\} \qquad (\mathcal{P}(U) > 0) \tag{A3.2.12}$$

and

$$\mathrm{E}[Z \exp(it'X_n)] \to \mathrm{E}(Z)\,\mathrm{E}[\exp(it'X)], \tag{A3.2.13}$$

respectively. In this case, the X_n are said to *converge $\mathcal{F}$-mixing* to X.

In applications, it is often the case that the left-hand sides of relations such as (A3.2.7) converge as $n \to \infty$, but it is not immediately clear that the limit can be associated with the conditional distribution of a well-defined r.v. X. Indeed, in general there is no guarantee that such a limit r.v. will exist, but we can instead extend the probability space in such a way that on the extended space a new sequence of r.v.s can be defined with effectively the same conditional distributions as for the original r.v.s and for which there is $\mathcal{F}$-stable convergence in the limit to a proper conditional distribution.

Lemma A3.2.V. *Suppose that for each $U \in \mathcal{F}$ and for A in some covering ring generating $\mathcal{B}(\mathcal{X})$, the sequences $\{\mathcal{P}(\{X_n \in A\} \cap U)\}$ converge. Then, there exists a probability space $(\Omega', \mathcal{E}', \mathcal{P}')$, a measurable mapping $T\colon (\Omega', \mathcal{E}') \mapsto (\Omega, \mathcal{E})$, and an r.v. X' defined on $(\Omega', \mathcal{E}')$ such that if $\mathcal{F}' = T^{-1}\mathcal{F}$ and $X'_n(\omega') = X_n(T\omega')$, then $X'_n \to X'$ ($\mathcal{F}'$-stably).*

Proof. Set $\Omega' = \mathcal{X} \times \Omega$, and let $\mathcal{E}'$ be the smallest σ-algebra of subsets of Ω' containing both $\mathcal{B}(\mathcal{X}) \otimes \mathcal{F}$ and also $\mathcal{X} \times \mathcal{E}$. Defining T by $T(x, \omega) = \omega$, we see that T is measurable. Also, for each $A \in \mathcal{B}(\mathcal{X})$ and $U \in \mathcal{F}$, the limit $\pi(A \times U) = \lim_{n\to\infty} \mathcal{P}(\{X_n \in A\} \cap U)$ exists by assumption and defines a countably additive set function on such product sets. Similarly, we can set $\pi(\mathcal{X} \times B) = \lim_{n\to\infty} \mathcal{P}(\{X_n \in \mathcal{X}\} \cap B) = \mathcal{P}(B)$ for $B \in \mathcal{E}$. Thus, π can be extended to a countably additive set function, $\mathcal{P}'$ say, on $\mathcal{E}'$. Observe that $\mathcal{F}' = T^{-1}\mathcal{F}$ consists of all sets $\mathcal{X} \times U$ for $U \in \mathcal{F}$. Define also $X'(x, \omega) = x$. Then, for $U' = \mathcal{X} \times U \in \mathcal{F}'$,

$$\mathcal{P}'(\{X'_n \in A\} \cap U') = \mathcal{P}(\{X_n \in A\} \cap U) \to \mathcal{P}'(A \times U) = \mathcal{P}'(\{X' \in A\} \cap U')$$

so that X'_n converges to X' $\mathcal{F}$-stably. □

Each of the conditions (i)–(v) of Proposition A3.2.IV consists of a family of sequences, involving r.v.s X_n converging in some sense, and the family of

the limits is identified with a family involving a limit r.v. X. It is left to the reader to verify via Lemma A3.2.V that if we are given only the convergence parts of any of these conditions, then the conditions are still equivalent, and it is possible to extend the probability space and construct a new sequence of r.v.s X'_n with the same joint probability distributions as the original X_n together with a limit r.v. X' such that $X'_n \to X'$, $\mathcal{F}$-stably, and so on.

In a similar vein, there exists the following selection theorem for stable convergence.

Proposition A3.2.VI. *Let $\{X_n\}$ be a sequence of $\mathcal{X}$-valued r.v.s on $(\Omega, \mathcal{E}, \mathcal{P})$ and $\mathcal{F}$ a sub-σ-algebra of $\mathcal{E}$. If*
(i) *either $\mathcal{F}$ is countably generated or $\mathcal{F} \supseteq \sigma(X_1, X_2, \ldots)$, and*
(ii) *the distributions of the $\{X_n\}$ converge weakly on $\mathcal{B}(\mathcal{X})$,*
then there exists an extended probability space $(\Omega', \mathcal{E}', \mathcal{P}')$, elements T, $\mathcal{F}'$, X'_n defined as in Lemma A3.2.V, a sequence $\{n_k\}$, and a limit r.v. X' such that $\{X'_{n_k}\}$ converges to X', $\mathcal{F}$-stably, as $k \to \infty$.

Proof. Suppose first that $\mathcal{F}$ is countably generated, and denote by $\mathcal{R}$ some countable ring generating $\mathcal{F}$. For each $U \in \mathcal{R}$, the measures on $\mathcal{B}(\mathcal{X})$ defined by

$$Q_n(A; U) = \mathcal{P}(\{X_n \in A\} \cap U)$$

are uniformly tight because they are strictly dominated by the uniformly tight measures $\mathcal{P}(\{X_n \in A\})$. Thus, they contain a weakly convergent subsequence. Using a diagonal selection argument, the subsequence can be so chosen that convergence holds simultaneously for all $U \in \mathcal{R}$. Therefore, we can assume that the sequence $\{Q_{n_k}(A; U)\}$ converges as $k \to \infty$ to some limit $Q(A; U)$ for all A that are continuity sets of this limit measure and for all $U \in \mathcal{R}$.

Given $\epsilon > 0$ and $B \in \mathcal{F}$, there exist U_ϵ, $V_\epsilon \in \mathcal{R}$ such that $U_\epsilon \subseteq B \subseteq V_\epsilon$ and $\mathcal{P}(U_\epsilon) \geq \mathcal{P}(V_\epsilon) - \epsilon$. Then, the two extreme terms in the chain of inequalities

$$\begin{aligned} \lim_{k\to\infty} Q_{n_k}(A; U_\epsilon) &\leq \lim_{k\to\infty} \inf_{j>k} \mathcal{P}(\{X_{n_j} \in A\} \cap B) \\ &\leq \lim_{k\to\infty} \sup_{j>k} \mathcal{P}(\{X_{n_j} \in A\} \cap B) \leq \lim_{k\to\infty} Q_{n_k}(A; V_\epsilon) \end{aligned}$$

differ by at most ϵ, so the sequence $\{\mathcal{P}(\{X_{n_k} \in A\} \cap B)\}$ also converges. The construction of an extended probability space $(\Omega', \mathcal{E}', \mathcal{P}')$ and a limit r.v. X' now follows as in the lemma, establishing the proposition in the case where $\mathcal{F}$ is countably generated.

To treat the case where $\mathcal{F} \supseteq \sigma(X_1, X_2, \ldots)$, consider first the case where $\mathcal{F} = \mathcal{F}_0 \equiv \sigma(X_1, X_2, \ldots)$. This is countably generated because $\mathcal{X}$ is separable and only a countable family of r.v.s is involved. Applying the selection argument and extension of the probability space, we can conclude from (A3.2.10) that

$$\mathrm{E}[Zh(X'_{n_k})] \to \mathrm{E}[Zh(X')] \qquad (\text{any } \mathcal{F}'_0\text{-measurable } Z). \tag{A3.2.14}$$

Now let Z' be any $\mathcal{F}'$-measurable r.v. (where $\mathcal{F} \supset \mathcal{F}_0$). Because $h(X'_{n_k})$ is $\mathcal{F}'_0$-measurable, we can write

$$\mathrm{E}[Z'h(X'_{n_k})] = \mathrm{E}[\,\mathrm{E}(Z' \mid \mathcal{F}'_0)\, h(X'_{n_k})],$$

and the convergence follows from (A3.2.14) by the $\mathcal{F}'_0$-measurability of $\mathrm{E}(Z' \mid \mathcal{F}'_0)$. Thus, for any such Z', $\mathrm{E}[Z'h(X'_{n_k})] \to \mathrm{E}[Z'h(X')]$, implying that $X'_{n_k} \to X'$ ($\mathcal{F}'_0$-stably). □

A systematic account of the topology of stable convergence when $\mathcal{F} = \mathcal{E}$ but no limit r.v. is assumed is given by Jacod and Memin (1984).

A3.3. Processes and Stopping Times

This section is primarily intended as background material for Chapter 14, where the focus is on certain real-valued stochastic processes denoted $\{X_t(\omega)\}$ $= \{X(t,\omega)\} = \{X(t)\}$ on the positive time axis, $t \in (0,\infty) \equiv \mathbb{R}_+$. Other time domains—finite intervals, or $\mathbb{R}$, or (subsets of) the integers $\mathbb{Z} = \{0, \pm 1, \ldots\}$ —can be considered: it is left to the reader to supply appropriate modifications to the theory as needed. Our aim here is to give just so much of the measure-theoretic framework as we hope will make our text intelligible. For a detailed discussion of this framework, texts such as Dellacherie (1972), Dellacherie and Meyer (1978) or Elliott (1982) should be consulted. Condensed accounts of selected results such as given here are also given in Brémaud (1981), Kallianpur (1980), and Liptser and Shiryayev (1977).

While a stochastic process $X(t,\omega)$ may be regarded as an indexed family of random variables on a common probability space $(\Omega, \mathcal{E}, \mathcal{P})$, with index set here taken to be $\mathbb{R}_+$, it is more appropriate for our purposes, as in the general theory, to regard it as a function on the product space $\mathbb{R}_+ \times \Omega$. The stochastic process $X\colon \mathbb{R}_+ \times \Omega \mapsto \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{E}$ is *measurable* when this mapping is measurable; that is, for all $A \in \mathcal{B}(\mathbb{R})$,

$$\{(t,\omega)\colon X(t,\omega) \in A\} \in \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{E}, \tag{A3.3.1}$$

where the right-hand side denotes the product σ-algebra of the two σ-algebras there. As a consequence of this measurability and Fubini's theorem, $X(\cdot,\omega)\colon$ $\mathbb{R}_+ \mapsto \mathbb{R}$ is a.s. measurable, while for measurable functions $h\colon \mathbb{R} \mapsto \mathbb{R}$,

$$Y(\omega) \equiv \int_{\mathbb{R}_+} h(X(t,\omega))\,\mathrm{d}t$$

is a random variable provided the integral exists. A stochastic process on $\mathbb{R}_+$, if defined merely as an indexed family of r.v.s on a common probability space, is necessarily measurable if, for example, the trajectories are either a.s. continuous or a.s. monotonic and right-continuous.

The main topic we treat concerns the evolution of a stochastic process; that is, we observe $\{X(s,\omega)\colon 0 < s \le t\}$ for some (unknown) ω and finite time interval $(0,t]$. It is then natural to consider the σ-algebra

$$\mathcal{F}_t^{(X)} \equiv \sigma\{X(s,\omega)\colon 0 < s \le t\}$$

generated by all possible such evolutions. Clearly,

$$\mathcal{F}_s^{(X)} \subseteq \mathcal{F}_t^{(X)}$$

for $0 < s < t < \infty$. Of course, we may also have some foreknowledge of the process X, and this we represent by a σ-algebra $\mathcal{F}_0$. Quite generally, an expanding family $\boldsymbol{\mathcal{F}} = \{\mathcal{F}_t\colon 0 \le t < \infty\}$ of sub-σ-algebras of $\mathcal{E}$ is called a *filtration* or *history*, and we concentrate on those histories that incorporate information on the process X. For this purpose, we want the r.v. $X(t,\omega)$ to be $\mathcal{F}_t$-measurable (all t); we then say that X is $\boldsymbol{\mathcal{F}}$-*adapted*. We adopt the special notation

$$\boldsymbol{\mathcal{H}} = \{\mathcal{F}_t^{(X)}\colon 0 \le t \le \infty\} \equiv \{\mathcal{H}_t\colon 0 \le t \le \infty\},$$

where $\mathcal{F}_0^{(X)} = \liminf_{t>0} \mathcal{F}_t^{(X)} = \{\emptyset, \Omega\}$ and $\mathcal{F}_\infty^{(X)} = \bigcap_{t>0} \mathcal{F}_t^{(X)}$, and call $\boldsymbol{\mathcal{H}}$ the *internal*, *minimal*, or *natural* history of the process X, both of these last two names reflecting the fact that $\boldsymbol{\mathcal{H}}$ is the smallest family of nested σ-algebras to which X is adapted. Any history of the form $\boldsymbol{\mathcal{F}} = \{\mathcal{F}_0 \vee \mathcal{H}_t\colon 0 \le t \le \infty\}$ is called an *intrinsic* history.

Suppose X is measurable and $\boldsymbol{\mathcal{F}}$-adapted. An apparently stronger condition to impose on X is that of *progressive measurability* with respect to $\boldsymbol{\mathcal{F}}$, meaning that for every $t \in \mathbb{R}_+$ and any $A \in \mathcal{B}(\mathbb{R})$,

$$\{(s,\omega)\colon 0 < s \le t,\ X(s,\omega) \in A\} \in \mathcal{B}((0,t]) \times \mathcal{F}_t. \tag{A3.3.2}$$

Certainly, (A3.3.2) is more restrictive on X than (A3.3.1), and while (A3.3.2) implies (A3.3.1), the converse is not quite true. What can be shown, however, is that given any measurable $\boldsymbol{\mathcal{F}}$-adapted $\mathbb{R}$-valued process X, we can find an $\boldsymbol{\mathcal{F}}$-progressively measurable process Y (that is therefore measurable and $\boldsymbol{\mathcal{F}}$-adapted) that is a *modification* of X in the sense of being defined (like X) on $(\Omega, \mathcal{E}, \mathcal{P})$ and satisfying

$$\mathcal{P}\{\omega\colon X(t,\omega) = Y(t,\omega)\} = 1 \qquad (\text{all } t) \tag{A3.3.3}$$

(see e.g. Dellacherie and Meyer, 1978, Chapter IV, Theorems 29 and 30).

The sets of the form $[s,t] \times U$, $0 \le s < t$, $U \in \mathcal{F}_t$, $t \ge 0$, generate a σ-algebra on $\mathbb{R}_+ \times \Omega$, which may be called the $\boldsymbol{\mathcal{F}}$-*progressive* σ-*algebra*. Then the requirement that the process X be $\boldsymbol{\mathcal{F}}$-progressively measurable may be rephrased as the requirement that $X(t,\omega)$ be measurable with respect to the $\boldsymbol{\mathcal{F}}$-progressive σ-algebra.

A more restrictive condition to impose on X is that it be $\boldsymbol{\mathcal{F}}$-*predictable* (the term $\boldsymbol{\mathcal{F}}$-*previsible* is also used). Call the sub-σ-algebra of $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{E}$ generated by product sets of the form $(s,t] \times U$, where $U \in \mathcal{F}_s$, $t \geq s$, and $0 \leq s < \infty$, the *predictable* σ-*algebra*, denoted $\Psi^{\boldsymbol{\mathcal{F}}}$. (The terminology is well chosen because it reflects what can be predicted at some 'future' time t given the evolution of the process—as revealed by sets $U \in \mathcal{F}_s$—up to the 'present' time s). Then X is $\boldsymbol{\mathcal{F}}$-predictable when it is $\Psi^{\boldsymbol{\mathcal{F}}}$-measurable; that is, for any $A \in \mathcal{B}(\mathbb{R})$,

$$\{(t,\omega)\colon X(t,\omega) \in A\} \in \Psi^{\boldsymbol{\mathcal{F}}}.$$

The archetypal $\boldsymbol{\mathcal{F}}$-predictable process is left-continuous, and this is reflected in Lemma A3.3.I below, in which the left-continuous history $\boldsymbol{\mathcal{F}}_{(-)} \equiv \{\mathcal{F}_{t-}\}$ associated with $\boldsymbol{\mathcal{F}}$ appears: here, $\mathcal{F}_{0-} = \mathcal{F}_0$ and $\mathcal{F}_{t-} = \limsup_{s<t} \mathcal{F}_s = \bigvee_{s<t} \mathcal{F}_s$. Note that if $X(t,\omega)$ is $\mathcal{F}_{t-}$-measurable, its value at t is in fact determined by information at times *prior* to t.

Lemma A3.3.I. *An $\boldsymbol{\mathcal{F}}$-predictable process is $\boldsymbol{\mathcal{F}}_{(-)}$-adapted.*

Proof. Consider first a process of the form

$$X(t,\omega) = I_{(a,b]}(t)\, I_U(\omega) \qquad (0 < a < b < \infty,\ U \in \mathcal{F}_a), \tag{A3.3.4}$$

which is $\boldsymbol{\mathcal{F}}$-predictable by construction of $\Psi^{\boldsymbol{\mathcal{F}}}$. For given t,

$$\{\omega\colon X(t,\omega) = 1\} = \begin{cases} \emptyset & \text{if } a \geq t \text{ or } b < t, \\ U & \text{if } a < t \leq b, \end{cases}$$

so $X(t,\omega)$ is $\mathcal{F}_{t-}$-measurable. Since an arbitrary $\boldsymbol{\mathcal{F}}$-predictable function can be approximated by a linear combination of functions of this type, and since the class of $\boldsymbol{\mathcal{F}}_{(-)}$-adapted processes is closed under linear combinations and monotone limits, standard extension arguments complete the proof. □

Indicator functions as in (A3.3.4), and linear combinations of them, can be used to show that the $\boldsymbol{\mathcal{F}}$-predictable σ-algebra $\Psi^{\boldsymbol{\mathcal{F}}}$ above can be characterized as the σ-algebra generated by the class of bounded left-continuous $\boldsymbol{\mathcal{F}}$-adapted processes (see e.g. Kallianpur, 1980, Lemma 3.1.1).

It is often important to examine the behaviour of a process not at a fixed time t but rather a random time $T = T(\omega)$. Here the definition of *stopping time* is fundamental.

Definition A3.3.II. *Given a history $\boldsymbol{\mathcal{F}}$, a nonnegative r.v. $T\colon \Omega \mapsto [0,\infty]$ is an $\boldsymbol{\mathcal{F}}$-*stopping time *if*

$$\{\omega\colon T(\omega) \leq t\} \in \mathcal{F}_t \qquad (0 \leq t < \infty).$$

If S, T are stopping times, then so are $S \wedge T$ and $S \vee T$. Indeed, given a family $\{T_n\colon n = 1,2,\ldots\}$ of stopping times, $\sup_{n\geq 1} T_n$ is an $\boldsymbol{\mathcal{F}}$-stopping time, while $\inf_{n\geq 1} T_n$ is an $\boldsymbol{\mathcal{F}}_{(+)}$-stopping time.

Since $\{T(\omega) = \infty\} = \bigcap_n \{T(\omega) > n\} \in \mathcal{F}_\infty$, we can also consider extended stopping times as those for which $\mathcal{P}\{T(\omega) < \infty\} < 1$.

While stopping times can be generated in various ways, the most common method is as a first passage time, which for a nondecreasing process usually arises as a level-crossing time.

Lemma A3.3.III. *Let X be an $\boldsymbol{\mathcal{F}}$-adapted monotonically increasing right-continuous process, and let Y be an $\mathcal{F}_0$-measurable r.v. Then $T(\omega) \equiv \inf\{t: X(t,\omega) \geq Y(\omega)\}$ is an $\boldsymbol{\mathcal{F}}$-stopping time, possibly extended, while if X is $\boldsymbol{\mathcal{F}}$-predictable, then T is an (extended) $\boldsymbol{\mathcal{F}}_{(-)}$-stopping time.*

Proof. If Y is constant, $X(t) \geq Y$ if and only if $T \leq t$, and since $\{\omega: X(t,\omega) \geq Y\} \in \mathcal{F}_t$, we also have $\{T(\omega) \leq t\} \in \mathcal{F}_t$. More generally, $X(t,\omega) - Y(\omega)$ is monotonically increasing, right-continuous, and $\boldsymbol{\mathcal{F}}$-adapted (because Y, being $\mathcal{F}_0$-measurable, is necessarily $\mathcal{F}_t$-measurable for every $t > 0$). Then, by the same argument, $\{T(\omega) \leq t\} = \{\omega: X(t,\omega) - Y(\omega) \geq 0\} \in \mathcal{F}_t$. Finally, when X is $\boldsymbol{\mathcal{F}}$-predictable, it is $\boldsymbol{\mathcal{F}}_{(-)}$-adapted, and thus we can replace $\mathcal{F}_t$ by $\mathcal{F}_{t-}$ throughout. □

The next result shows that a process stopped at an $\boldsymbol{\mathcal{F}}$-stopping time T inherits some of the regularity properties of the original process. Here we use the notation

$$X(t \wedge T) = \begin{cases} X(t) & (t \leq T), \\ X(T) & (t > T). \end{cases}$$

Proposition A3.3.IV. *Let $\boldsymbol{\mathcal{F}}$ be a history, T an $\boldsymbol{\mathcal{F}}$-stopping time, and X a process. Then $X(t \wedge T)$ is measurable, $\boldsymbol{\mathcal{F}}$-progressive, or $\boldsymbol{\mathcal{F}}$-predictable, according to whether $X(t)$ itself is measurable, $\boldsymbol{\mathcal{F}}$-progressive, or $\boldsymbol{\mathcal{F}}$-predictable. In all these cases, if $T < \infty$ a.s., then $X(T)$ is an $\mathcal{F}_\infty$-measurable r.v.*

Proof. The product σ-algebra $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{E}$ is generated by sets of the form $(a, \infty) \times B$ for real finite a and $B \in \mathcal{E}$. Since

$$\{(t,\omega): (t \wedge T(\omega), \omega) \in (a,\infty) \times B\} \;=\; (a,\infty) \times (B \cap \{T(\omega) > a\})$$

and $B \cap \{T(\omega) > a\} \in \mathcal{E}$, if X is measurable, so is $Y(t,\omega) \equiv X(t \wedge T(\omega), \omega)$.

The $\boldsymbol{\mathcal{F}}$-predictable σ-algebra $\Psi^{\boldsymbol{\mathcal{F}}}$ is generated by sets of a similar product form but with $B \in \mathcal{F}_a$. Since $\{T(\omega) > a\} \in \mathcal{F}_a$, $(a,\infty) \times (B \cap \{T(\omega) > a\})$ is also a set generating $\Psi^{\boldsymbol{\mathcal{F}}}$, and thus if X is $\boldsymbol{\mathcal{F}}$-predictable, so is Y as before.

Suppose now that X is $\boldsymbol{\mathcal{F}}$-progressive so that for given t in $0 < t < \infty$, $\{X(s,\omega): o < s \leq t\}$ is measurable as a process on $(0,t]$ with probability space $(\Omega, \mathcal{F}_t, \mathcal{P})$. Then, the first argument shows that $Y(s) \equiv X(s \wedge T)$ is a measurable process on this space; that is, $X(t \wedge T)$ is $\boldsymbol{\mathcal{F}}$-progressive.

On the set $\{T < \infty\}$, $X(t \wedge T) \to X(T)$ as $t \to \infty$, so when $\mathcal{P}\{T < \infty\} = 1$, $X(T)$ is an r.v. as asserted. □

As an important corollary to this result, observe that if X is $\boldsymbol{\mathcal{F}}$-progressive and a.s. integrable on finite intervals, then

$$Y(t,\omega) = \int_0^t X(s,\omega)\,\mathrm{d}s$$

is $\boldsymbol{\mathcal{F}}$-progressive, $Y(T)$ is an r.v. if $T < \infty$ a.s., and $Y(t \wedge T)$ is again $\boldsymbol{\mathcal{F}}$-progressive.

We conclude this section with some remarks about the possibility of a converse to Lemma A3.3.I. In the case of a quite general history, no such result of this kind holds, as is shown by the discussion in Dellacherie and Meyer (1978), especially around Chapter IV, Section 97. On the other hand, it is shown in the same reference that when X is defined on the canonical measure space $(\mathcal{M}^{\#}_{[0,\infty)}, \mathcal{B}(\mathcal{M}^{\#}_{[0,\infty)}))$, the two concepts of being $\boldsymbol{\mathcal{F}}_{(-)}$-adapted and $\boldsymbol{\mathcal{F}}$-predictable can be identified, a fact exploited in the treatment by Jacobsen (1982).

The situation can be illustrated further by the two indicator processes

$$V_T^+(t,\omega) \equiv I_{\{T(\omega)\le t\}}(t,\omega), \qquad V_T^-(t,\omega) \equiv I_{\{T(\omega)<t\}}(t,\omega),$$

generated by an $\boldsymbol{\mathcal{F}}$-stopping time T. The trajectories of V_T^+ are right-continuous while those of V_T^- are left-continuous. Since $\mathcal{F}_t \ni \{\omega\colon T(\omega) \le t\} = \{\omega\colon V_T^+(t) = 1\}$, it follows that V_T^+ is $\boldsymbol{\mathcal{F}}$-adapted. So too is V_T^- because

$$\{\omega\colon V_T^-(t) = 1\} = \{\omega\colon T(\omega) < t\} = \bigcup_{n=1}^{\infty} \left\{\omega\colon T(\omega) \le t - \frac{1}{n}\right\} \in \mathcal{F}_t.$$

Hence, both V_T^+ and V_T^- are $\boldsymbol{\mathcal{F}}$-progressively measurable [see the earlier comments or Brémaud (1981, Theorem A1.T33)].

Being left-continuous, V_T^- is $\boldsymbol{\mathcal{F}}$-predictable (e.g. Brémaud, 1981, Theorem 1.T9) and hence also $\boldsymbol{\mathcal{F}}_{(-)}$-adapted. No such statement can be made in general about V_T^+. However, suppose further that T is not only an $\boldsymbol{\mathcal{F}}$-stopping time but also an $\boldsymbol{\mathcal{F}}_{(-)}$-stopping time, so that from the above, V_T^+ is $\boldsymbol{\mathcal{F}}_{(-)}$-adapted. Can we assert that it is $\boldsymbol{\mathcal{F}}$-predictable?

Suppose T is a countably-valued r.v., so for some countable set $\{t_k\} \subset \mathbb{R}_+$,

$$T^{-1}(\{t_k\colon k = 1,2,\ldots\}) = \bigcup_{k=1}^{\infty} T^{-1}(t_k) = \bigcup_{k=1}^{\infty} U_k, \qquad \text{say, } = \Omega.$$

Then

$$\{(t,\omega)\colon V_T^+(t,\omega) = 1\} = \bigcup_{k=1}^{\infty} [t_k, \infty) \times U_k.$$

By assumption, T being an $\boldsymbol{\mathcal{F}}_{(-)}$-stopping time, $U_k \in \mathcal{F}_{t_k-}$, so $U_k \in \sigma\{\bigcup_n \mathcal{F}_{t_k - 1/n}\}$ and hence V_T^+ is $\boldsymbol{\mathcal{F}}$-predictable.

While it can be proved that any $\boldsymbol{\mathcal{F}}$-stopping time can be approximated from above by a sequence of stopping times taking only a countable set of values, this is not enough to treat the general case—indeed, the counterexample considered by Dellacherie and Meyer is just of this indicator function type.

A3.4. Martingales

Definition A3.4.I. *Let $(\Omega, \mathcal{E}, \mathcal{P})$ be a probability space, $\boldsymbol{\mathcal{F}}$ a history on $(\Omega, \mathcal{E})$, and $X(\cdot) \equiv \{X(t): 0 \le t < \infty\}$ a real-valued process adapted to $\boldsymbol{\mathcal{F}}$ and such that $\mathrm{E}(|X(t)|) < \infty$ for $0 \le t < \infty$. Then X is an* $\boldsymbol{\mathcal{F}}$-martingale *if for $0 \le s < t < \infty$,*

$$\mathrm{E}[X(t) \mid \mathcal{F}_s] = X(s) \quad \text{a.s.}, \tag{A3.4.1}$$

an $\boldsymbol{\mathcal{F}}$-submartingale *if*

$$\mathrm{E}[X(t) \mid \mathcal{F}_s] \ge X(s) \quad \text{a.s.}, \tag{A3.4.2}$$

and an $\boldsymbol{\mathcal{F}}$-supermartingale *if the reverse inequality in* (A3.4.2) *holds.*

Strictly, we should speak of X as a $\mathcal{P}$-$\boldsymbol{\mathcal{F}}$-martingale: mostly, it is enough to call it a martingale since both $\mathcal{P}$ and $\boldsymbol{\mathcal{F}}$ are clear from the context.

While the concept of a martingale had its origins in gambling strategies, it has come to play a dominant role in the modern theory of stochastic processes. In our text, we need only a small number of the many striking results concerning martingales and their relatives, principally those connected with stopping times and the Doob–Meyer decomposition.

An important example of a martingale is formed from an $\mathcal{F}_\infty$-measurable r.v. X_∞ with finite mean by taking successive conditional expectations with respect to $\boldsymbol{\mathcal{F}}$: define

$$X(t) = \mathrm{E}(X_\infty \mid \mathcal{F}_t). \tag{A3.4.3}$$

Such a martingale is uniformly integrable. The converse statement is also true (see e.g. Liptser and Shiryayev, 1977, Theorem 3.6).

Proposition A3.4.II. *Let $X(\cdot)$ be a uniformly integrable $\boldsymbol{\mathcal{F}}$-martingale. Then, there exists an $\mathcal{F}_\infty$-measurable r.v. X_∞ such that* (A3.4.3) *holds.*

The following form of the well-known convergence theorem can be found in Liptser and Shiryayev (1977, Theorem 3.3).

Theorem A3.4.III. *Let $X(\cdot)$ be an $\boldsymbol{\mathcal{F}}$-submartingale with a.s. right-continuous trajectories. If $\sup_{0 \le t < \infty} \mathrm{E}[\max(0, X(t))] < \infty$, then there exists an $\mathcal{F}_\infty$-measurable r.v. X_∞ such that*

$$X(t, \omega) \to X_\infty(\omega) \quad (t \to \infty) \quad \text{a.s.}$$

If also $X(\cdot)$ is uniformly integrable, then $\mathrm{E}(|X_\infty|) < \infty$ and $\mathrm{E}(|X(t) - X_\infty|) \to 0$ as $t \to \infty$; that is, $X(t) \to X_\infty$ in L_1 norm.

This theorem can be applied to the example in (A3.4.3) whether the family of σ-algebras $\{\mathcal{F}_t\}$ is increasing (as with a history $\boldsymbol{\mathcal{F}}$) or decreasing. For convenience, we state the result in terms of a two-sided history $\boldsymbol{\mathcal{G}} = \{\mathcal{G}_t: -\infty < t < \infty\}$, defining $\mathcal{G}_\infty$ as usual and $\mathcal{G}_{-\infty} = \bigcap_{-\infty < t < \infty} \mathcal{G}_t = \lim_{t \to -\infty} \mathcal{G}_t$.

Corollary A3.4.IV. *If the r.v. Y is $\mathcal{G}_\infty$-measurable, has finite first moment, and $Y(t) \equiv \mathrm{E}(Y \mid \mathcal{G}_t)$ has a.s. right-continuous trajectories on $-\infty < t < \infty$ for some two-sided history $\boldsymbol{\mathcal{G}}$, then*

$$\mathrm{E}(Y \mid \mathcal{G}_t) \to \begin{cases} Y & (t \to \infty), \\ \mathrm{E}(Y \mid \mathcal{G}_{-\infty}) & (t \to -\infty), \end{cases} \tag{A3.4.4}$$

both a.s. and in L_1 norm.

In most point process applications, the processes concerned are right-continuous by definition, so the sample-path conditions for the convergence results above are automatically satisfied. In the general theory of processes, it is shown that, if the history $\boldsymbol{\mathcal{F}}$ is right-continuous and the σ-algebras are $\mathcal{P}$-complete in the strong sense that $\mathcal{F}_0$ (and hence $\mathcal{F}_t$ for all $t > 0$) contains all $\mathcal{P}$-null sets from $\mathcal{F}_\infty$, there always exists a right-continuous modification of an $\boldsymbol{\mathcal{F}}$-submartingale, with the additional property that this modification also has left limits at each $t > 0$; that is, the (modified) process is càdlàg [see e.g. Liptser and Shiryayev (1977, pp. 55–59) or Dellacherie and Meyer (1980); Elliott (1982) uses corlol, the acronym of the English equivalent, continuous on right, limits on left].

In turning to properties of martingales with fixed times s, t replaced by stopping times S, T, say, we need the notion of σ-algebras consisting of events prior to (and including) the time T and also strictly prior to T.

Definition A3.4.V. *Let $\boldsymbol{\mathcal{F}}$ be a history and T an $\boldsymbol{\mathcal{F}}$-stopping time. The* T-prior σ-algebra *$\mathcal{F}_T$ is the sub-σ-algebra of $\mathcal{F}_\infty$ defined by*

$$\mathcal{F}_T = \{A\colon A \in \mathcal{F}_\infty \text{ and } A \cap \{T \le t\} \in \mathcal{F}_t \text{ for every } t\};$$

the strict T-prior σ-algebra *$\mathcal{F}_{T-}$ is generated by the sets*

$$\{A\colon A \in \mathcal{F}_0\} \cup \big\{A \cap \{T > t\} \text{ for } A \in \mathcal{F}_t \text{ and } t \ge 0\big\}.$$

Clearly, $\mathcal{F}_T$ and $\mathcal{F}_{T-}$ are somewhat different entities (see Dellacherie and Meyer, 1978, p. 117). It can be checked that T is both $\mathcal{F}_T$- and $\mathcal{F}_{T-}$-measurable. A contrast is provided in the next result.

Lemma A3.4.VI. *Let $\boldsymbol{\mathcal{F}}$ be a history, T an $\boldsymbol{\mathcal{F}}$-stopping time, and $X(\cdot)$ an $\boldsymbol{\mathcal{F}}$-progressive process. Then $X(T)$ is $\mathcal{F}_T$-measurable. Further, if $X(\cdot)$ is $\boldsymbol{\mathcal{F}}$-predictable, then $X(T)$ is $\mathcal{F}_{T-}$-measurable.*

Proof. Suppose $X(\cdot)$ is $\boldsymbol{\mathcal{F}}$-progressive. Setting for any $x \in \mathbb{R}$

$$A_x = \{\omega\colon X(T(\omega), \omega) \le x\},$$

$X(T)$ is $\mathcal{F}_T$-measurable if $A_x \cap \{T \le t\} \in \mathcal{F}_t$. But from Proposition A3.4.IV, $X(t \wedge T)$ is $\boldsymbol{\mathcal{F}}$-progressive, and therefore $\boldsymbol{\mathcal{F}}$-adapted, so $\{\omega\colon X(t \wedge T(\omega), \omega) \le x\} \in \mathcal{F}_t$; hence

$$A_x \cap \{T \le t\} = \{\omega\colon X(t \wedge T(\omega), \omega) \le x\} \cap \{T \le t\} \in \mathcal{F}_t.$$

Now suppose that $X(\cdot)$ is $\boldsymbol{\mathcal{F}}$-predictable. To show the $\mathcal{F}_{T-}$-measurability of $X(T)$, look at the inverse image under $X(T)\colon \omega \mapsto X(T(\omega), \omega) \in \mathbb{R}$ of a generating set $(t, \infty) \times A$ $(A \in \mathcal{F}_t)$ of the $\boldsymbol{\mathcal{F}}$-predictable σ-algebra $\Psi^{\boldsymbol{\mathcal{F}}}$, namely

$$\{\omega\colon t < T(\omega) < \infty\} \cap \{\omega\colon \omega \in A\},$$

which is a generating set for $\mathcal{F}_{T-}$. □

The *optional sampling theorem* for martingales follows (see e.g. Liptser and Shiryayev, 1977, pp. 60–61).

Theorem A3.4.VII. *Let $\boldsymbol{\mathcal{F}}$ be a history, S and T the $\boldsymbol{\mathcal{F}}$-stopping times with $S \le T$* a.s.*, and $X(\cdot)$ an $\boldsymbol{\mathcal{F}}$-submartingale that is uniformly integrable and has right-continuous trajectories. Then $\mathcal{F}_S \subseteq \mathcal{F}_T$ and*

$$\mathrm{E}[X(T) \mid \mathcal{F}_S] \ge X(S) \quad \text{a.s.},$$

where equality holds if X is an $\boldsymbol{\mathcal{F}}$-martingale.

Corollary A3.4.VIII. *Let T be an $\boldsymbol{\mathcal{F}}$-stopping time. If $X(\cdot)$ is a uniformly integrable $\boldsymbol{\mathcal{F}}$-martingale (resp. submartingale), then so is $X(t \wedge T)$.*

Proof. For fixed s, t with $s < t$, $s \wedge T$ and $t \wedge T$ are two stopping times satisfying the conditions of the theorem, so

$$\mathrm{E}[X(t \wedge T) \mid \mathcal{F}_{s \wedge T}] \ge X(s \wedge T),$$

and thus $\{X(t \wedge T)\}$ is an $\{\mathcal{F}_{t \wedge T}\}$-martingale. To show the stronger property that it is an $\boldsymbol{\mathcal{F}}$-martingale, note that $\mathcal{F}_{t \wedge T} \subseteq \mathcal{F}_t$ so $\{X(t \wedge T)\}$ is $\boldsymbol{\mathcal{F}}$-adapted, and it remains to show that

$$\int_A X_{t \wedge T}\, \mathcal{P}(\mathrm{d}\omega) \ge \int_A X_{s \wedge T}\, \mathcal{P}(\mathrm{d}\omega) \qquad (\text{all } A \in \mathcal{F}_s), \tag{A3.4.5}$$

knowing that it holds for all $A \in \mathcal{F}_{s \wedge T}$. Express the left-hand side as the sum of integrals over $A_1 = A \cap \{T > s\}$ and $A_2 = A \cap \{T \le s\}$. Certainly, $A_1 \in \mathcal{F}_s$, while

$$A_1 \cap \{s \wedge T \le u\} = A \cap \{T > s\} \cap \{s \wedge T \le u\} = \begin{cases} \emptyset \in \mathcal{F}_u & \text{if } u < s, \\ A_1 \in \mathcal{F}_s & \text{if } u \ge s. \end{cases}$$

Now $\mathcal{F}_s \subseteq \mathcal{F}_u$, so by definition of $\mathcal{F}_{s \wedge T}$, we have $A_1 \in \mathcal{F}_{s \wedge T}$, and (A3.4.5) holds for A_1. On A_2, $t \ge s \ge T$ so $X(t \wedge T) = X(s \wedge T)$ there, and (A3.4.5) holds for A_2. By addition, we have shown (A3.4.5). □

Finally, we quote the form of the Doob–Meyer decomposition theorem used in Chapter 14; see e.g. Liptser and Shiryayev (1977) for proof.

Theorem A3.4.IX (Doob–Meyer)**.** *Let $\boldsymbol{\mathcal{F}}$ be a history and $X(\cdot)$ a bounded $\boldsymbol{\mathcal{F}}$-submartingale with right-continuous trajectories. Then, there exists a unique (up to equivalence) uniformly integrable $\boldsymbol{\mathcal{F}}$-martingale $Y(\cdot)$ and a unique $\boldsymbol{\mathcal{F}}$-predictable cumulative process $A(\cdot)$ such that*

$$X(t) = Y(t) + A(t). \tag{A3.4.6}$$

For nondecreasing processes $A(\cdot)$ with right-continuous trajectories, it can be shown that $\boldsymbol{\mathcal{F}}$-predictability is equivalent to the property that for every bounded $\boldsymbol{\mathcal{F}}$-martingale $Z(\cdot)$ and positive u,

$$\mathrm{E}\bigg[\int_0^u Z(t)\,A(\mathrm{d}t)\bigg] = \mathrm{E}\bigg[\int_0^u Z(t-)\,A(\mathrm{d}t)\bigg].$$

Since for any $\boldsymbol{\mathcal{F}}$-adapted cumulative process ξ and any $\boldsymbol{\mathcal{F}}$-martingale Z, $\mathrm{E}\big[Z(u)\int_0^u \xi(\mathrm{d}t)\big] = \mathrm{E}\big[\int_0^u Z(t)\,\xi(\mathrm{d}t)\big]$, the property above is equivalent to

$$\mathrm{E}[Z(u)A(u)] = \mathrm{E}\bigg[\int_0^u Z(t-)\,A(\mathrm{d}t)\bigg].$$

A cumulative process with this property is referred to in many texts as a *natural increasing process*. The theorem can then be rephrased thus: *every bounded submartingale has a unique decomposition into the sum of a uniformly integrable martingale and a natural increasing function*. The relation between natural increasing and predictable processes is discussed in Dellacherie and Meyer (1980).

The boundedness condition in Theorem A3.4.IX is much stronger than is really necessary, and it is a special case of Liptser and Shiryayev's (1977) 'Class D' condition for supermartingales; namely, that the family $\{X(T)\}$ is uniformly integrable for all $\boldsymbol{\mathcal{F}}$-stopping times. More general results, of which the decomposition for point processes described in Chapter 13 is in fact a special case, relax the boundedness or uniform integrability conditions but weaken the conclusion by requiring $Y(\cdot)$ to be only a local martingale [i.e. the stopped processes $Y(\cdot \wedge T_n)$ are martingales for a suitable increasing sequence $\{T_n\}$ of $\boldsymbol{\mathcal{F}}$-stopping times].

The Doob–Meyer theorem is often stated for supermartingales, in which case the natural increasing function should be subtracted from the martingale term, not added to it.

Given an $\boldsymbol{\mathcal{F}}$-martingale S, it is *square integrable* on $[0,\tau]$ for some $\tau \le \infty$ if $\sup_{0<t\le\tau} \mathrm{E}[X^2(t)] < \infty$. The process $\{X^2(t)\}$ is then an $\boldsymbol{\mathcal{F}}$-submartingale on $[0,\tau]$. When it is a bounded submartingale, the Doob–Meyer theorem as quoted above implies that we have the decomposition

$$X^2(t) = Y_2(t) + A_2(t) \qquad (0 \le t \le \tau) \tag{A3.4.7}$$

for some $\boldsymbol{\mathcal{F}}$-martingale $Y_2(\cdot)$ and $\boldsymbol{\mathcal{F}}$-predictable process $A_2(\cdot)$. It is readily checked that for $0 \le s < t \le \tau$,

$$A_2(t) - A_2(s) = \mathrm{E}[(X_t - X_s)^2 \mid \mathcal{F}_s],$$

hence the name *quadratic variation process* for $A_2(\cdot)$. Equation (A3.4.7) can be established for any square-integrable martingale via the general Doob–Meyer theorem. A significant calculus for such processes, including applications to point processes, can be constructed as in Kunita and Watanabe (1967) and Brémaud (1981).

References with Index

[At the end of each reference entry is the page number or numbers where it is cited. A bibliography of about 600 references up to about 1970, although excluding much of the historical material of Chapter 1 of this book, is given in D.J. Daley and R.K. Milne (1972), The theory of point processes: A bibliography, *Int. Statist. Rev.* **41**, 183–201.]

Aalen, O.O. (1975). *Statistical Inference for a Family of Counting Processes*. Ph.D. thesis, Statistics Dept., University of California, Berkeley. [17, 238]

—— (1978). Non-parametric inference for a family of counting processes. *Ann. Statist.* **6**, 701–726. [238]

Abbé, E. (1879). Über Blutkörper-Zahlung. *Jena Z. Med. Naturwiss.* **13** (New Series 6), 98–105. [8]

Aldous, D. and Eagleson, G.K. (1978). On mixing and stability of limit theorems. *Ann. Probab.* **6**, 325–331. [420]

Ammann, L.P. and Thall, P.F. (1979). Count distributions, orderliness and invariance of Poisson cluster processes. *J. Appl. Probab.* **16**, 261–273. [227]

Andersen, P.K., Borgan, Ø., Gill, R.D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York. [17, 108–109, 238]

Andersson, H. and Britton, T. (2000). *Stochastic Epidemic Models and Their Statistical Analysis*, Lecture Notes in Statistics **151**. Springer-Verlag, New York. [217]

Andrews, G.E. (1976). *The Theory of Partitions*. Addison–Wesley, Reading, MA. [121]

Argabright, L. and de Lamadrid, J.G. (1974). Fourier analysis of unbounded measures on locally compact abelian groups. *Mem. Amer. Math. Soc.* **145**. [303, 358]

Arjas, E., Nummelin, E., and Tweedie, R.L. (1978). Uniform limit theorems for non-singular renewal and Markov renewal processes. *J. Appl. Probab.* **15**, 112–125. [89–90]

Ash, R.B. (1972). *Real Analysis and Probability*. Academic Press, New York. [380, 414]

Athreya, K. and Ney, P.E. (1972). *Branching Processes*. Springer-Verlag, New York. [151]

—— and —— (1978). A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.* **245**, 493–501. [96–97]

——, Tweedie, R.L., and Vere-Jones, D. (1980). Asymptotic behaviour of point-processes with Markov-dependent intervals. *Math. Nachr.* **99**, 301–313. [102]

Baccelli, F. and Brémaud, P. (1994). *Elements of Queueing Theory*. Springer-Verlag, Berlin. [17, 211]

Baddeley, A.J. (1998). A crash course on stochastic geometry. In Barndorff-Nielsen *et al.* (1998), 1–35. [221]

—— (2001). Likelihoods and pseudolikelihoods for Markov spatial processes. In de Gunst, M.C.M., Klaassen, C.A.J., and van der Vaart, W. (Eds.), *State of the Art in Probability and Statistics: Festschrift for Willem R. van Zwet*, Institute of Mathematical Statistics Monograph Series **36**. IMS, Hayward, CA, pp. 21–49. [217]

—— and Møller, J. (1989). Nearest-neighbour Markov point processes and random sets. *Int. Statist. Rev.* **57**, 89–121. [111]

—— and Turner, R. (2000). Practical maximum pseudo-likelihood for spatial point patterns (with Discussion). *Aust. N.Z. J. Statist.* **42**, 283–322. [232]

—— , van Lieshout, M.N.M., and Møller, J. (1996). Markov properties of cluster processes. *Adv. Appl. Probab.* **28**, 346–355. [111]

Barndorff-Nielsen, O.E., Kendall, W.S., and van Lieshout, M.N.M. (Eds.) (1998). *Stochastic Geometry: Likelihood and Computation*. Chapman and Hall, London. [17, 111]

Bartlett, M.S. (1954). Processus stochastiques ponctuels. *Ann. Inst. Henri Poincaré* **14**, 35–60. [14]

—— (1955). *An Introduction to Stochastic Processes*. Cambridge University Press, Cambridge [2nd ed. 1966; 3rd ed. 1978]. [13]

—— (1963). The spectral analysis of point processes. *J. Roy. Statist. Soc. Ser. B* **25**, 264–296. [16, 182, 303, 305]

—— (1964). The spectral analysis of two-dimensional point processes. *Biometrika* **51**, 299–311. [16]

—— (1967). The spectral analysis of line processes. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **3**, 135–153. [329]

—— and Kendall, D.G. (1951). On the use of the characteristic functional in the analysis of some stochastic processes in physics and biology. *Proc. Cambridge Philos. Soc.* **47**, 65–76. [15, 150]

Bateman, H. (1910). Note on the probability distribution of α-particles. *Philos. Mag.* **20** (6), 704–707. [Note to E. Rutherford and H. Geiger, The probability variations in the distribution of α-particles, *Philos. Mag.* **20** (6), 698–704.] [9]

Baudin, M. (1981). Likelihood and nearest-neighbor distance properties of multi-dimensional Poisson cluster processes. *J. Appl. Probab.* **18**, 879–888. [221, 227]

Bebbington, M. and Harte, D. (2001). On the statistics of the linked stress release model. In Daley, D.J. (Ed.), *Probability, Statistics and Seismology, J. Appl. Probab.* **38A**, 176–187. [240, 256]

Benard, C. and Macchi, O. (1973). Detection and emission processes of quantum particles in a chaotic state. *J. Math. Phys.* **14**, 155–167. [140]

Berbée, H. (1983). A bound on the size of point clusters of a random walk with stationary increments. *Ann. Probab.* **11**, 414–418. [301]

Berg, C. and Forst, G. (1975). *Potential Theory on Locally Compact Abelian Groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete **87**. Springer-Verlag, New York. [357]

Berman, M. (1983). Discussion of Ogata's paper. *Bull. Int. Statist. Inst.* **50** (3), 412–422. [244]

Bhabha, H.J. (1950). On the stochastic theory of continuous parametric systems and its application to electron-photon cascades. *Proc. Roy. Soc. London Ser. A* **202**, 301–332. [15, 111, 124, 136]

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York. [391]

Bloom, W.R. (1984). Translation bounded measures and the Orlicz–Paley–Sidon theorem. In *Probability Measures on Groups, VII (Oberwolfach, 1983)*, Lecture Notes in Mathematics **1064**. Springer, Berlin, pp. 1–9. [367]

Bochner, S. (1947). Stochastic processes. *Ann. Math.* **48**, 1014–1061. [15]

—— (1955). *Harmonic Analysis and the Theory of Probability*. University of California Press, Berkeley. [15, 303, 338]

Boel, R., Varaiya, P., and Wong, E. (1975). Martingales on jump processes, I: Representation results, and II: Applications. *SIAM J. Control* **13**, 999–1021 and 1022–1061. [211]

Bogoliubov, N.N. (1946). *Problems of a Dynamical Theory in Statistical Physics* (in Russian). Gostekhizdat, Moscow. [Translated by E.K. Gora in de Boer, J. and Uhlenbeck, G.E. (Eds.), *Studies in Statistical Mechanics*, Vol. 1, North-Holland, Amsterdam, 1962, pp. 5–116.] [15, 111, 124]

Bol'shakov, I.A. (1969). *Statistical Problems in Isolating a Stream of Signals from Noise* (in Russian). Sovyetskoye Radio, Moscow. [146, 185]

Boltzmann, L. (1868). Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten. *Sitzungsber. Math. Naturwiss. Kl. Kais. Akad. Wiss.* **58**, 517–560. [9]

Borovkov, K. and Vere-Jones, D. (2000). Explicit formulae for stationary distributions of stress release models. *J. Appl. Probab.* **37**, 315–321. [245]

Bourbaki, N. (1963). *Éléments de Mathématique, Fasc. XXIX (Livre VI, Intégration, Chaps. 7 et 8)*, Actualités Scientifiques et Industrielles **1306**. Hermann, Paris. [409]

Breiman, L. (1965). Some probabilistic aspects of the renewal theorem. In *Transactions of the Fourth Prague Conference on Information Theory, Statistics and Decision Functions, Prague, 1965.* Academia, Prague, pp. 255–261. [87]

Brémaud, P. (1972). *A Martingale Approach to Point Processes.* Ph.D. thesis, Electrical Engineering Dept., University of California, Berkeley. [211]

—— (1981). *Point Processes and Queues: Martingale Dynamics.* Springer-Verlag, New York. [14, 17, 107–108, 211, 231, 377, 414, 423, 427, 431]

—— and Massoulié, L. (1994). Imbedded construction of stationary point processes and sequences with a random memory. *Queueing Syst.* **17**, 213–234. [252]

—— and —— (1996). Stability of nonlinear Hawkes processes. *Ann. Probab.* **24**, 1563–1588. [202, 252–253, 268, 275]

Brémaud, P. and Massoulié, L. (2001). Hawkes branching point processes without ancestors. *J. Appl. Probab.* **38**, 122–135. [203]

Bretagnolle, J. and Dacunha-Castelle, D. (1967). Sur une classe de marches aléatoires. *Ann. Inst. Henri Poincaré* **3**, 403–431. [90]

Brillinger, D.R. (1972). The spectral analysis of stationary interval functions. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1**, 483–513. [15, 303, 331, 338]

—— (1975a). The identification of point process systems. *Ann. Probab.* **3**, 909–929. [318]

—— (1975b). Stochastic inference for stationary point processes. In Puri, M.L. (Ed.) *Stochastic Processes and Related Topics*. Academic Press, New York, pp. 55–99. [Reprinted in Brillinger (1981).] [17, 318]

—— (1978). Comparative aspects of the study of ordinary time series and point processes. In Krishnaiah, P.R. (Ed.), *Developments in Statistics*, Vol. I. Academic Press, New York, pp. 33–133. [303, 318, 337]

—— (1981). *Time Series: Data Analysis and Theory*, 2nd ed. Holden–Day, San Francisco. [318, 337]

—— (1992). Nerve cell spike train analysis: A progression of technique. *J. Amer. Statist. Assoc.* **87**, 260–271. [318]

Brix, A. and Kendall, W.S. (2002). Simulation of cluster point processes without edge effects. *Adv. Appl. Probab.* **34**, 267–280. [221, 275]

Brown, T.M. and Nair, M. (1988). A simple proof of the multivariate random time change theorem for point processes. *J. Appl. Probab.* **25**, 210–214. [264]

Campbell, N.R. (1909). The study of discontinuous phenomena. *Proc. Cambridge Philos. Soc.* **15**, 117–136. [163]

Cane, V.R. (1974). The concept of accident proneness. *Izv. Mat. Inst. Bulgar. Akad. Sci.* **15**, 183–189. [11–12]

—— (1977). A class of non-identifiable stochastic models. *J. Appl. Probab.* **14**, 475–482. [11]

Carlsson, H. and Nerman, O. (1986). An alternative proof of Lorden's renewal inequality. *Adv. Appl. Probab.* **18**, 1015–1016. [91]

Chernick, M.R., Daley, D.J., and Littlejohn, R.P. (1988). A time-reversibility relationship between two Markov chains with exponential stationary distributions. *J. Appl. Probab.* **25**, 418–422. [105]

Chong, F.S. (1981). A point process with second order Markov dependent intervals. *Math. Nachr.* **103**, 155–163. [105, 298, 303]

Chung, K.L. (1972). Crudely stationary point processes. *Amer. Math. Monthly* **79**, 867–877. [44]

—— (1974). *A Course in Probability Theory*, 2nd ed. Academic Press, New York. [383, 414]

Copson, E.C. (1935). *An Introduction to the Theory of Functions of a Complex Variable*. Oxford University Press, Oxford. [310, 336]

Coram, M. and Diaconis, P. (2002). New tests of the correspondence between unitary eigenvalues and the zeros of Riemann's zeta function. *J. Phys. A* (to appear). [18, 140]

Cox, D.R. (1955). Some statistical methods connected with series of events (with Discussion). *J. Roy. Statist. Soc. Ser. B* **17**, 129–164. [16, 105, 110, 169]

Cox, D.R. (1962). *Renewal Theory*. Methuen, London. [66]

—— (1972a). The statistical analysis of dependencies in point processes. In Lewis (1972), pp. 55–66. [211, 238]

—— (1972b). Regression models and life tables (with Discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187–220. [17, 238]

—— (1975). Partial likelihood. *Biometrika* **62**, 269–276. [236]

—— and Isham, V. (1980). *Point Processes*. Chapman and Hall, London. [66, 105, 295–296, 301–302]

—— and Lewis, P.A.W. (1966). *The Statistical Analysis of Series of Events*. Methuen, London. [16–17, 20, 66, 261–262, 296, 303]

Cramér, H. (1930). *On the Mathematical Theory of Risk*. Skandia Jubilee Volume, Stockholm. [Reprinted in Cramér, H. (1994). *Collected Works* (A. Martin-Löf, Ed.). Springer, Berlin.] [199]

—— and Leadbetter, M.R. (1967). *Stationary and Related Stochastic Processes*. Wiley, New York. [14, 333, 345]

Cressie, N.A.C. (1991). *Statistics for Spatial Data*. Wiley, New York. [Rev. ed. 1993.] [17, 111, 222, 320]

Daley, D.J. (1965). On a class of renewal functions. *Proc. Cambridge Philos. Soc.* **61**, 519–526. [78]

—— (1971). Weakly stationary point processes and random measures. *J. Roy. Statist. Soc. Ser. B* **33**, 406–428. [61, 301, 305, 331]

—— (1972a). A bivariate Poisson queueing process that is not infinitely divisible. *Proc. Cambridge Philos. Soc.* **72**, 449–450. [188]

—— (1972b). Asymptotic properties of stationary point processes with generalized clusters. *Z. Wahrs.* **21**, 65–76. [308]

—— (1973a). Poisson and alternating renewal processes with superposition a renewal process. *Math. Nachr.* **57**, 359–369. [82]

—— (1973b). Markovian processes whose jump epochs constitute a renewal process. *Quart. J. Math. Oxford Ser. (2)* **24**, 97–105. [82]

—— (1974). Various concepts of orderliness for point processes. In Harding and Kendall (1974), pp. 148–161. [47, 52]

—— (1981). The absolute convergence of weighted sums of dependent sequences of random variables. *Z. Wahrs.* **58**, 199–203. [168]

—— (1982). Stationary point processes with Markov-dependent intervals and infinite intensity. In Gani, J. and Hannan, E.J. (Eds.), *Essays in Statistical Science, J. Appl. Probab.* **19A**, 313–320. [94, 105]

—— (1999). The Hurst index of long-range dependent renewal processes. *Ann. Probab.* **27**, 2035–2041. [106]

—— and Milne, R.K. (1975). Orderliness, intensities and Palm–Khinchin equations for multivariate point processes. *J. Appl. Probab.* **12**, 383–389. [331]

—— and Narayan, P. (1980). Series expansions of probability generating functions and bounds for the extinction probability of a branching process. *J. Appl. Probab.* **17**, 939–947. [119, 122]

——, Rolski, T., and Vesilo, R. (2000). Long-range dependent point processes and their Palm–Khinchin distributions. *Adv. Appl. Probab.* **32**, 1051–1063. [106]

Daley, D.J. and Vere-Jones, D. (1972). A summary of the theory of point processes. In Lewis (1972), pp. 299–383. [60]

Daniels, H.E. (1945). The statistical theory of the strength of bundles of threads, I. *Proc. Roy. Soc. London Ser. A* **183**, 405–435. [7]

Darwin, J.H. (1957). The power of the Poisson index of dispersion. *Biometrika* **44**, 286–289. [23]

David, F.N. and Barton, D.E. (1962). *Combinatorial Chance*. Griffin, London. [114–115, 121]

Davidson, R. (1974). Construction of line processes: Second-order properties. In Harding and Kendall (1974), pp. 55–75. [Original publication (1970), *Izv. Akad. Nauk Armen. SSR Ser. Mat.* **5**, 219–234.] [305]

Davies, R.B. (1977). Testing the hypothesis that a point process is Poisson. *Adv. Appl. Probab.* **9**, 724–746. [222, 226, 228–229]

—— (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74**, 33–43. [263]

Dawson, D.A., Fleischmann, K., and Mueller, C. (2000). Finite time extinction of superprocesses with catalysts. *Ann. Probab.* **28**, 603–642. [18]

Dellacherie, C. (1972). *Capacités et Processus Stochastiques*. Springer, Berlin. [423]

—— and Meyer, P.-A. (1978). *Probabilities and Potential*. Hermann, Paris, and North-Holland, Amsterdam. [423–424, 427]

—— and —— (1980). *Probabilités et Potential, Chap. V–VIII, Théorie des Martingales*. Hermann, Paris. [429, 431]

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39**, 1–22. [244]

Diaconis, P. and Evans, S.N. (2000). Immanants and finite point processes. In memory of Gian-Carlo Rota. *J. Combin. Theory Ser. A* **91**, 305–321. [18]

—— and —— (2001). Linear functionals of eigenvalues of random matrices. *Trans. Amer. Math. Soc.* **353**, 2615–2633. [18, 141]

Diggle, P.J. (1983). *Statistical Analysis of Spatial Point Patterns*. Academic Press, London. [17, 111, 300]

——, Fiksel, T., Grabanik, P., Ogata, Y., Stoyan, D., and Tanemura, M. (1994). On parameter estimation for pairwise-interaction point processes. *Internat. Statist. Rev.* **62**, 99–117. [217]

—— and Milne, R.K. (1983). Negative binomial quadrat counts and point processes. *Scand. J. Statist.* **10**, 257–267. [200]

Doob, J.L. (1948). Renewal theory from the point of view of the theory of probability. *Trans. Amer. Math. Soc.* **63**, 422–438. [72]

—— (1949). Time series and harmonic analysis. In Neyman, J. (Ed.), *Berkeley Symposium on Mathematical Statististics and Probabability*. University of California Press, Berkeley, pp. 303–343. [303, 331]

—— (1953). *Stochastic Processes*. Wiley, New York. [61, 303–304, 333, 339, 417]

——, Snell, J.L., and Williamson, R.E. (1960). Application of boundary theory to sums of independent random variables. In *Contributions to Probability and Statistics (Essays in Honor of H. Hotelling)*, Stanford University Press, Stanford, CA, pp. 182–197. [74]

Dwass, M. and Teicher, H. (1957). On infinitely divisible random vectors. *Ann. Math. Statist.* **28**, 461–470. [188]

Eggenberger, F. and Pólya, G. (1923). Über die Statistik verketteter Vorgänge. *Z. Angew. Math. Mech.* **3**, 279–289. [11]

Elliott, R.J. (1982). *Stochastic Calculus and Applications*. Springer-Verlag, New York. [423, 429]

—— , Aggoun, L., and Moore, J.B. (1995). *Hidden Markov Models*. Springer, New York. [244]

Ellis, R.L. (1844). On a question in the theory of probabilities. *Cambridge Math. J.* **4** (21), 127–133. [Reprinted in W. Walton (Ed.) (1863) *The Mathematical and Other Writings of Robert Leslie Ellis*, Deighton Bell, Cambridge, pp. 173–179.] [4]

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events*. Springer, Berlin. [17, 200]

Erlang, A.K. (1909). The theory of probabilities and telephone conversations. *Nyt. Tidsskr. Mat. B* **20**, 33–41. [Reprinted in E. Brockmeyer, H.L. Halstrom and A. Jensen (1948), *The Life and Works of A.K. Erlang*, Copenhagen Telephone Company, Copenhagen, pp. 131–137.] [5, 9]

Feller, W. (1950). *An Introduction to Probability Theory and Its Applications*, Vol. 1. Wiley, New York [2nd ed. 1957; 3rd ed. 1968]. [13]

—— (1966). *An Introduction to Probability Theory and Its Applications*, Vol. 2. Wiley, New York [2nd ed. 1971]. [45, 62, 64, 66, 70, 74, 84, 86–87, 90–91, 412]

—— (1968). [= 3rd ed. of Feller (1950).] [28–30]

Fieger, W. (1971). Die Anzahl der γ-Niveau-Kreuzungspunkte von stochastische Prozessen. *Z. Wahrs.* **18**, 227–260. [59]

Fosam, E.B. and Shanbhag, D.N. (1997). Variants of the Choquet–Deny theorem with applications. *J. Appl. Probab.* **34**, 101–106. [82]

Franken, P., König, D., Arndt, U., and Schmidt, V. (1981). *Queues and Point Processes*. Akademie-Verlag, Berlin. [14, 17]

Fréchet, M. (1940). *Les probabilités associées à un systeme d'événements compatibles et dépendants*, Actualités Scientifiques et Industrielles **859**. Hermann, Paris. [119]

Galambos, J. (1975). Methods for proving Bonferroni inequalities. *J. London Math. Soc.* **9** (2), 561–564. [119]

—— and Kotz, S. (1978). *Characterizations of Probability Distributions*, Lecture Notes in Mathematics **675**. Springer-Verlag, Berlin. [24, 77, 82]

Galton, F. and Watson, H.W. (1874). On the probability of extinction of families. *J. Roy. Anthropol. Inst.* **4**, 138–144. [9]

Gaver, D.P. (1963). Random hazard in reliability problems. *Technometrics* **5**, 211–216. [211]

—— and Lewis, P.A.W. (1980). First-order autoregressive gamma sequences and point processes. *Adv. Appl. Probab.* **12**, 727–745. [92]

Georgii, H.-O. (1988). *Gibbs Measures and Phase Transitions*. W. de Gruyter, Berlin. [18]

Glass, L. and Tobler, W.R. (1971). Uniform distribution of objects in a homogeneous field: Cities on a plain. *Nature* **233**, 67–68. [298]

Goldman, J.R. (1967). Stochastic point processes: Limit theorems. *Ann. Math. Statist.* **8**, 771–779. [31]

Goodman, N.R. and Dubman, M.R. (1969). The theory of time-varying spectral analysis and complex Wishart matrix processes. In Krishnaiah, P.R. (Ed.), *Multivariate Analysis II*. Academic Press, New York, pp. 351–366. [172]

Grandell, J. (1976). *Doubly Stochastic Poisson Processes*, Lecture Notes in Mathematics **529**. Springer-Verlag, New York. [173, 175]

Graunt, J. (1662). *Natural and Political Observations Made Upon the Bills of Mortality*. John Martin, London. [Reprinted in facsimile in *The Earliest Classics: John Graunt and Gregory King* (1973). Gregg International Publishers, Farnborough.] [3]

Greenwood, M. and Yule, G.U. (1920). An enquiry into the nature of frequency distributions of multiple happenings, with particular reference to the occurrence of multiple attacks of disease or repeated accidents. *J. Roy. Statist. Soc.* **83**, 255–279. [10]

Grégoire, G. (1984). Negative binomial distribution for point processes. *Stoch. Proc. Appl.* **16**, 179–188. [200]

Greig-Smith, P. (1964). *Quantitative Plant Ecology*, 2nd ed. Butterworths, London. [296]

Griffiths, R.C., Milne, R.K., and Wood, R. (1979). Aspects of correlation in bivariate Poisson distributions and processes. *Aust. J. Statist.* **21**, 238–255. [188]

Guttorp, P. (1995). *Stochastic Modeling of Scientific Data*. Chapman and Hall, London. [320]

Häberlund, E. (1975). Infinitely divisible stationary recurrent point processes. *Math. Nachr.* **70**, 259–264. [82]

Häggström, O., van Lieshout, M.N.M., and Møller, J. (1999). Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli* **5**, 641–658. [217]

Haight, F.A. (1967). *Handbook of the Poisson Distribution*. Wiley, New York. [9]

Hall, P. (1988). *An Introduction to Coverage Processes*. Wiley, New York. [205]

Hannan, E.J. (1970). *Multiple Time Series*. Wiley, New York. [347]

Harding, E.J. and Kendall, D.G. (Eds.) (1974). *Stochastic Geometry*. Wiley, Chichester. [17]

Harn, K. van (1978). *Classifying Infinitely Divisible Distributions by Functional Equations*, Mathematical Centre Tract **103**. Mathematisch Centrum, Amsterdam. [78]

Harris, T.E. (1956). The existence of stationary measures for certain Markov processes. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **2**, 113–124. [92, 97]

—— (1963). *The Theory of Branching Processes*. Springer-Verlag, Berlin. [16, 151]

—— (1968). Counting measures, monotone random set functions. *Z. Wahrs.* **10**, 102–119. [16, 389]

—— (1971). Random measures and motions of point processes. *Z. Wahrs.* **18**, 85–115. [16]

Hawkes, A.G. (1971a). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58**, 83–90. [183, 211]

—— (1971b). Point spectra of some mutually exciting point processes. *J. Roy. Statist. Soc. Ser. B* **33**, 438–443. [183, 202, 211, 309, 320, 322]

Hawkes, A.G. (1972). Spectra of some mutually exciting point processes with associated variables. In Lewis (1972), pp. 261–271. [183, 202, 320]

—— and Adamopoulos, L. (1973). Cluster models for earthquakes—regional comparisons. *Bull. Int. Statist. Inst.* **45**(3), 454–461. [202, 204, 309]

—— and Oakes, D. (1974). A cluster representation of a self-exciting process. *J. Appl. Probab.* **11**, 493–503. [183]

Hayashi, T. (1986). Laws of large numbers in self-correcting point processes. *Stoch. Proc. Appl.* **23**, 319–326. [240]

Hewitt, E. and Zuckerman, H.S. (1969). Remarks on the functional equation $f(x+y) = f(x) + f(y)$. *Math. Mag.* **42**, 121–123. [64]

Heyde, C.C. and Seneta, E. (1977). *I.J. Bienaymé: Statistical Theory Anticipated*. Springer-Verlag, New York. [9]

Hille, E. and Phillips, R.S. (1957). *Functional Analysis and Semi-Groups*. American Mathematical Society, Providence, RI. [63–64]

Hocking, T.G. and Young, G.S. (1961). *Topology*. Addison–Wesley, Reading, MA, and London. [406]

Holgate, P. (1964). Estimation for the bivariate Poisson distribution. *Biometrika* **51**, 241–245. [188]

Hunter, J.J. (1974a). Renewal theory in two dimensions: Basic results. *Adv. Appl. Probab.* **6**, 376–391. [72]

—— (1974b). Renewal theory in two dimensions: Asymptotic results. *Adv. Appl. Probab.* **6**, 546–562. [72]

Isham, V. (1985). Marked point processes and their correlations. In Droesbecke, F. (Ed.), *Spatial Processes and Spatial Time Series Analysis*. Publications des Facultés Universitaires Saint-Louis, Bruxelles, pp. 63–75. [327]

—— and Westcott, M. (1979). A self-correcting point process. *Stoch. Proc. Appl.* **8**, 335–347. [239]

Ito, Y. (1980). Renewal processes decomposable into i.i.d. components. *Adv. Appl. Probab.* **12**, 672–688. [82]

Jacobsen, M. (1982). *Statistical Analysis of Counting Processes*, Lecture Notes in Statistics **12**. Springer-Verlag, New York. [17, 238, 427]

Jacod, J. (1975). Multivariate point processes: Predictable projections, Radon–Nikodym derivatives, representation of martingales. *Z. Wahrs.* **31**, 235–253. [247]

—— and Memin, J. (1981). Sur un type de convergence intermédiaire entre la convergence en loi et la convergence en probabilité. In *Seminar on Probability, XV (Strasbourg, 1979/1980)*, Lecture Notes in Mathematics **850**, Springer, Berlin, pp. 529–546. [Correction (1983) in *Seminar on Probability, XVII*, Lecture Notes in Mathematics **986**, Springer, Berlin, pp. 509–511.] [423]

Jagers, P. (1974). Aspects of random measures and point processes. In Ney, P. (Ed.), *Advances in Probability and Related Topics*, Vol. 3. Marcel Dekker, New York, pp. 179–239. [15]

—— (1975). *Branching Processes with Biological Applications*. Wiley, London. [151]

Janossy, L. (1948). *Cosmic Rays*. Oxford University Press, Oxford. [15]

—— (1950). On the absorption of a nucleon cascade. *Proc. Roy. Irish Acad. Sci. Sect. A* **53**, 181–188. [111, 124]

Jiřina, M. (1966). Asymptotic behaviour of measure-valued branching processes. *Rozpr. Cesk. Akad. Ved., Rada Mat. Prir. Ved.* **75**(3). [15]

Johnson, N.L and Kotz, S. (1969). *Distributions in Statistics, Vol. I: Discrete Distributions*. Houghton Mifflin, Boston. [2nd ed. 1993. Wiley, New York.] [10, 12]

—— and —— (1970). *Distributions in Statistics, Vol. II: Continuous Univariate Distributions–1*. Houghton Mifflin, Boston. [2nd ed. 1994. Wiley, New York.] [7]

—— and —— (1994). = 2nd ed. of Johnson and Kotz (1970). [82]

Jolivet, E. (1978). Caractérisation et test du caractère agrégatif des processus ponctuels stationnaires sur $\mathbb{R}^2$. In Dacunha-Castelle, D. and Cutsem, B. van (Eds.) *Journées de Statistiques des Processus Stochastiques*, Lecture Notes in Mathematics **636**, Springer-Verlag, Berlin, pp. 1–25. [300]

Jowett, J. and Vere-Jones, D. (1972). The prediction of stationary point processes. In Lewis (1972), pp. 405–435. [331]

Kagan, Y.Y. (1999). Universality of the seismic moment-frequency relation. *Pure Appl. Geophys.* **155**, 537–573. [256]

—— and Schoenberg, F. (2001). Estimation of the upper cutoff parameter for the tapered Pareto distribution. In Daley, D.J. (Ed.), *Probability, Statistics and Seismology, J. Appl. Probab.* **38A**, 158–175. [256]

Kailath, T. and Segall, I. (1975). The modelling of random modulated jump processes. *IEEE Trans. Inf. Theory* **IT-21** (2), 135–142. [211]

Kallenberg, O. (1975). *Random Measures*. Akademie-Verlag, Berlin, and Academic Press, London. [3rd ed. 1983; reprinted with corrections as 4th ed. 1986]. [15, 389]

—— (1983). = 3rd ed. of Kallenberg (1975). [292, 294]

Kallianpur, G. (1980). *Stochastic Filtering Theory*. Springer-Verlag, New York. [423, 425]

Karr, A.F. (1986). *Point Processes and Their Statistical Inference*. Marcel Dekker, New York. [2nd ed. 1991.] [18, 247]

Kathirgamatamby, N. (1953). Note on the Poisson index of dispersion. *Biometrika* **40**, 225–228. [23]

Kelly, F.P. and Ripley, B.D. (1976). A note on Strauss's model for clustering. *Biometrika* **63**, 357–360. [217–218, 228]

Kendall, D.G. (1949). Stochastic processes and population growth. *J. Roy. Statist. Soc. Ser. B* **11**, 230–264. [15]

Kerstan, J. (1964). Teilprozesse Poissonscher Prozesse. In *Transactions of the Third Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, Czech. Academy of Science, Prague, pp. 377–403. [202, 268]

——, Matthes, K., and Mecke, J. (1974). *Unbegrenzt Teilbare Punktprozesse*. Akademie-Verlag, Berlin. [14]

——, ——, and —— (1982). *Infinitely Divisible Point Processes* (in Russian). Nauka, Moscow. [= 3rd ed. of Kerstan *et al.* (1974).] [14]

Khinchin, A.Ya. (1955). *Mathematical Methods in the Theory of Queueing* (in Russian). *Trudy Mat. Inst. Steklov* **49**. [Translated (1960). Griffin, London.] [14, 30, 46–48, 54]

—— (1956). On Poisson sequences of chance events. *Teor. Veroyatnost. i Primenen.* **1**, 320–327. [Translation in *Theory Probab. Appl.* **1**, 291–297.] [52]

Kingman, J.F.C. (1964). On doubly stochastic Poisson processes. *Proc. Cambridge Philos. Soc.* **60**, 923–930. [174]

—— (1972). *Regenerative Phenomena*. John Wiley, London. [82]

—— (1993). *Poisson Processes*. Clarendon Press, Oxford. [18, 33]

—— and Taylor, S.J. (1966). *Introduction to Measure and Probability*. Cambridge University Press, Cambridge. [Chapters 1–9 republished as *Introduction to Measure and Integration* (1973), same publisher.] [332, 368]

Knopoff, L. (1971). A stochastic model for the occurrence of main sequence events. *Rev. Geophys. Space Phys.* **9**, 175–188. [239]

Kolmogorov, A.N. (1935). La transformation de Laplace dans les espaces linéaires. *C. R. Acad. Sci. Paris* **200**, 1717–1718. [14]

Kotz, S. and Shanbhag, D. (1980). Some new approaches to probability distributions. *Adv. Appl. Probab.* **12**, 903–921. [109]

Krickeberg, K. (1974). Moments of point processes. In Harding and Kendall (1974), pp. 89–113. [143]

—— (1980). Statistical problems on point processes. In *Mathematical Statistics*, Banach Centre Publications **6**, PWN, Warsaw, pp. 197–223. [300]

Kunita, H. and Watanabe, S. (1967). On square-integrable martingales. *Nagoya Math. J.* **30**, 209–245. [211, 431]

Kutoyants, Y.A. (1980). *Estimation of Parameters of Stochastic Processes* (in Russian). Armenian Academy of Science, Erevan. [18, 226]

—— (1984). *Parameter Estimation for Stochastic Processes*. Heldermann, Berlin. [Translated by B.L.S. Prakasa Rao and Revised from Kutoyants (1980).] [18, 26, 226, 235]

—— (1998). *Statistical Inference for Spatial Poisson Processes*, Lecture Notes in Statistics **134**. Springer-Verlag, New York. [18]

Lai, C.D. (1978). An example of Wold's point processes with Markov-dependent intervals. *J. Appl. Probab.* **15**, 748–758. [96, 103–104, 243]

Lampard, D.G. (1968). A stochastic process whose successive intervals between events form a first order Markov chain-I. *J. Appl. Probab.* **5**, 648–668. [95, 105]

Lancaster, H.O. (1963). Correlations and canonical forms of bivariate distribution functions. *Ann. Math. Statist.* **34**, 532–538. [95]

Laplace, P.S. (1814). *Essai Philosophique des Probabilités*. Introduction (pp. i–cvi), Théorie Analytique des Probabilités, 2nd ed. [English Translation (1951), *A Philosophical Essay on Probabilities*. Dover, New York.] [3]

Last, G. and Brandt, A. (1995). *Marked Point Processes on the Real Line*. Springer-Verlag, New York. [18]

Lawrance, A.J. (1970). Selective interaction of a stationary point process and a renewal process. *J. Appl. Probab.* **7**, 483–489. [44]

Leadbetter, M.R. (1972). On basic results of point process theory. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **3**, 449–462. [14, 48, 52]

——, Lindgren, G., and Rootzen, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York. [14, 17]

LeCam, L. (1947). Un instrument d'étude des fonctions aléatoires: La fonctionelle caractéristique. *C. R. Acad. Sci. Paris* **224**, 710–711. [15]

LeCam, L. (1961). A stochastic theory of precipitation. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **3**, 165–186. [191]

Lee, P.M. (1968). Some aspects of infinitely divisible point processes. *Stud. Sci. Math. Hungar.* **3**, 219–224. [31, 33]

Lewis, P.A.W. (1964a). A branching Poisson process model for the analysis of computer failure patterns (with Discussion). *J. Roy. Statist. Soc. Ser. B* **26**, 398–456. [16, 182, 192]

—— (1964b). The implications of a failure model for the use and maintenance of computers. *J. Appl. Probab.* **1**, 347–368. [182]

—— (1970). Remarks on the theory, computation and application of the spectral analysis of series of events. *J. Sound Vib.* **12** (3), 353–375. [26]

—— (Ed.) (1972). *Stochastic Point Processes*. Wiley, New York. [211]

—— and Shedler, G.S. (1976). Simulation of nonhomogeneous Poisson processes with log linear rate function. *Biometrika* **63**, 501–506. [24, 269]

Liemant, A., Matthes, K., and Wakolbinger, A. (1988). *Equilibrium Distributions of Branching Processes*, Mathematical Research **42**. Akademie-Verlag, Berlin. [151]

Liggett, T.M. (1999). *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*. Springer-Verlag, Berlin. [18]

Lin, V.Ya. (1965). On equivalent norms in the space of square summable entire functions of exponential type (in Russian). *Mat. Sb. (N.S.)* **67**(109), 586–608. [Translation (1969) *Amer. Math. Soc. Transl.* **79**(2), 53–76.] [358, 364]

Lindvall, T. (1977). A probabilistic proof of Blackwell's renewal theorem. *Ann. Probab.* **5**, 482–485. [83]

—— (1992). *Lectures on the Coupling Method*. Wiley, New York. [83]

Liptser, R.S. and Shiryayev, A.N. (1974). *Statistics of Random Processes* (in Russian). Nauka, Moscow. [Translation (1977, 1978).] [17, 211]

—— and —— (1977). *Statistics of Random Processes, I: General Theory*. Springer-Verlag, New York. [17, 211, 423, 428–431]

—— and —— (1978). *Statistics of Random Processes, II: Applications*. Springer-Verlag, New York. [17, 211]

—— and —— (2000). 2nd ed. of Liptser and Shiryayev (1977, 1978). [17, 211]

Littlewood, D.E. (1950). *The Theory of Group Characters and Matrix Representations of Groups*, 2nd ed. Clarendon Press, Oxford. [140]

Liu, J., Chen, Y., Shi, Y., and Vere-Jones, D. (1999). Coupled stress release model for time dependent seismicity. *Pure Appl. Geophys.* **155**, 649–667. [256]

Loève, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton, NJ. [4th ed. (2 vols.) (1977, 1978). Springer-Verlag, New York.] [29, 33]

Lotka, A.J. (1939). A contribution to the theory of self-renewing aggregates, with especial reference to industrial replacement. *Ann. Math. Statist.* **10**, 1–25. [5]

Lowen, S.B. and Teich, H.C. (1990). Power-law shot noise. *IEEE Trans. Inf. Theory* **IT-36**, 1302–1318. [170]

Lu, C., Harte, D., and Bebbington, M. (1999). A linked stress release model for historical Japanese earthquakes. Coupling among major seismic regions. *Earth Planets Space* **51**, 907–916. [240]

—— and Vere-Jones, D. (2000). Application of linked stress release model to historical earthquake data: Comparison between two kinds of tectonic seismicity. *Pure Appl. Geophys.* **157**, 2351–2364. [256]

Lüders, R. (1934). Die Statistik der seltenen Ereignisse. *Biometrika* **26**, 108–128. [11]

Lukacs, E. (1970). *Characteristic Functions*, 2nd ed. Griffin, London. [79]

Lundberg, F. (1903). *Approximerad framställning av sannolikhetsfunktionen. Återförsäkring av kollektivrisker.* Akad. Afhandling. Almqvist och Wiksell, Uppsala. [199]

Lyon, J.F. and Thoma, R. (1881). Ueber die Methode der Blutkörperzählung. *Virchows Arch. Path. Anat. Physiol.* **84**, 131–154. [9]

Macchi, O. (1971a). Distribution statistique des instants d'émission des photoélectrons d'une lumière thermique. *C. R. Acad. Sci. Paris Ser. A* **272**, 437–440. [139, 172–173, 175]

—— (1971b). Stochastic processes and multicoincidences. *IEEE Trans.* **IT-17** (1), 1–7. [144]

—— (1975). The coincidence approach to stochastic point processes. *Adv. Appl. Probab.* **7**, 83–122. [124, 130, 132, 136–137, 140, 172]

MacDonald, I.L. and Zucchini, W. (1997). *A Hidden Markov and other Models for Discrete-Valued Time Series*. Chapman and Hall, London. [244]

MacMahon, P.A. (1915). *Combinatory Analysis*, Vol. 1. Cambridge University Press, Cambridge. [175]

Main, I.G. (1996). Statistical physics, seismogenesis, and seismic hazard. *Rev. Geophys.* **34**, 433–462. [257]

Maistrov, L.E. (1967). *Probability Theory—An Historical Sketch* (in Russian). Izdat. Nauka, Moscow. [Translated by S. Kotz (Ed.) (1974). Academic Press, New York.] [3]

Massoulié, L. (1998). Stability results for a general class of interacting point process dynamics, and applications. *Stoch. Proc. Appl.* **75**, 1–30. [275]

Matérn, B. (1960). *Spatial Variation.* Meddelanded Stat. Skogsforsk. **49** (5), 1–144. [2nd ed. (1986). Lecture Notes in Statistics **36**, Springer-Verlag, New York.] [16, 298]

Mathéron, G. (1975). *Random Sets and Integral Geometry*. Wiley, New York. [17, 206]

Matthes, K., Kerstan, J., and Mecke, J. (1978). *Infinitely Divisible Point Processes.* Wiley, Chichester [= 2nd ed. of Kerstan, Matthes and Mecke (1974).] [See MKM]

McFadden, J.A. (1956). The axis-crossing intervals of random functions, I. *Trans. Inst. Radio Engnrs.* **IT-2**, 146–150. [14]

—— (1958). The axis-crossing intervals of random functions, II. *Trans. Inst. Radio Engnrs.* **IT-4**, 14–24. [14]

—— (1965). The entropy of a point process. *J. SIAM* **13**, 988–994. [286]

—— and Weissblum, W. (1963). Higher-order properties of a stationary point process. *J. Roy. Statist. Soc. Ser. B* **25**, 413–431. [81]

McKendrick, A.G. (1914). Studies on the theory of continuous probabilities with special reference to its bearing on natural phenomena of a progressive nature. *Proc. London Math. Soc.* **13**(2), 401–416. [9–10]

—— (1926). The application of mathematics to medical problems. *Proc. Edinburgh Math. Soc.* **44**, 98–130. [9]

Mecke, J. (1967). Zum Problem der Zerlegbarkeit stationärer rekurrenter zufälliger Punktfolgen. *Math. Nachr.* **35**, 311–321. [81]

—— (1969). Verschärfung eines Satzes von McFadden. *Wiss. Z. Friedrich-Schiller-Universität Jena* **18**, 387–392. [81]

Meyer, P.A. (1971). Démonstration simplifiée d'un théorème de Knight. In *Séminaire de Probabilités V, Université de Strasbourg, 1969–1970*, Lecture Notes in Mathematics **191**. Springer, Berlin, pp. 191–195. [257]

Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*. Springer-Verlag, London. [93]

Miles, R.E. (1974). On the elimination of edge effects in planar sampling. In Harding and Kendall (1974), pp. 228–247. [303]

Milne, R.K. (1974). Infinitely divisible bivariate Poisson processes (Abstract). *Adv. Appl. Probab.* **6**, 226–227. [188]

—— and Westcott, M. (1972). Further results for Gauss–Poisson processes. *Adv. Appl. Probab.* **4**, 151–176. [185–186, 315]

—— and —— (1993). Generalized multivariate Hermite distributions and related point processes. *Ann. Inst. Statist. Math.* **45**, 367–381. [123, 187]

Minc, H. (1978). *Permanents*. Addison–Wesley, Reading, MA. [173]

[MKM] (1978). [= Matthes, Kerstan and Mecke (1978).] [14, 82, 151, 157, 163, 176]

[MKM] (1982). [See Kerstan, Matthes and Mecke (1982).]

Molchanov, I. (1997). *Statistics of the Boolean Model for Practitioners and Mathematicians*. Wiley, Chichester. [17, 205–206]

Moran, P.A.P. (1967). A non-Markovian quasi-Poisson process. *Stud. Sci. Math. Hungar.* **2**, 425–429. [31, 106]

—— (1968). *An Introduction to Probability Theory*. Clarendon Press, Oxford. [33, 123, 163]

—— (1976a). A quasi-Poisson point process in the plane. *Bull. London Math. Soc.* **8**, 69–70. [31]

—— (1976b). Another quasi-Poisson plane point process. *Z. Wahrs.* **33**, 269–272. [31]

Moyal, J.E. (1962a). The general theory of stochastic population processes. *Acta Math.* **108**, 1–31. [15, 111, 129–130, 150]

—— (1962b). Multiplicative population chains. *Proc. Roy. Soc. London Ser. A* **266**, 518–526. [150–151]

Neuts, M.F. (1979). A versatile Markovian point process. *J. Appl. Probab.* **16**, 764–779. [306]

Newman, D.S. (1970). A new family of point processes characterized by their second moment properties. *J. Appl. Probab.* **7**, 338–358. [174, 185, 220]

Newton, Sir Isaac (1728). *The Chronology of Ancient Kingdoms Amended*. [Published posthumously. See H. Zeitlinger (1927), *A Newton bibliography*. In W.J. Greenstreet (Ed.), *Isaac Newton 1642–1727*. Bell and Sons, London, pp. 148–170.] [5]

Neyman, J. (1939). On a new class of 'contagious' distributions applicable in entomology and bacteriology. *Ann. Math. Statist.* **10**, 35–57. [11]

—— and Scott, E.L. (1958). Statistical approach to problems of cosmology (with Discussion). *J. Roy. Statist. Soc. Ser. B* **20**, 1–43. [12, 181]

Neyman, J. and Scott, E.L. (1972). Processes of clustering and applications. In Lewis (1972), pp. 646–681. [181]

Nummelin, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahrs.* **43**, 309–318. [96–97]

Oakes, D. (1974). A generalization of Moran's quasi-Poisson process. *Stud. Sci. Math. Hungar.* **9**, 433–437. [31]

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimates for stationary point processes. *Ann. Inst. Statist. Math.* **30**, 243–261. [235]]

—— (1981). On Lewis' simulation method for point processes. *IEEE Trans. Inf. Theory* **IT-27**, 23–31. [270–271, 275]

—— (1983). Likelihood analysis of point processes and its applications to seismological data. *Bull. Int. Statist. Inst.* **50** (2), 943–961. [244]

—— (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83**, 9–27. [204, 261, 263]

—— (1992). Detection of precursory relative quiescence before great earthquakes through a statistical model. *J. Geophys. Res.* **97**, 19845–19871. [263]

—— (1998). Space–time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50**, 379–402. [204]

—— (2001). Increased probability of large earthquakes near aftershock regions with relative quiescence. *J. Geophys. Res.* **106**, 8729–8744. [263]

—— and Akaike, H. (1982). On linear intensity models for mixed doubly stochastic Poisson and self-exciting point processes. *J. Roy. Statist. Soc. Ser. B* **44**, 102–107. [234, 309]

——, Akaike, H., and Katsura, K. (1982). The application of linear intensity models to the investigation of causal relations between a point process and another stochastic process. *Ann. Inst. Statist. Math.* **34**, 373–387. [309]

—— and Katsura, K. (1986). Point-process models with linearly parametrized intensity for the application to earthquake catalogue. *J. Appl. Probab.* **23A**, 231–240. [235]

—— and Tanemura, M. (1981). Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure. *Ann. Inst. Statist. Math.* **33B**, 315–338. [217]

—— and —— (1984). Likelihood analysis of spatial point patterns. *J. Roy. Statist. Soc. Ser. B* **46**, 496–518. [128]

—— and Vere-Jones, D. (1984). Inference for earthquake models: A self-correcting model. *Stoch. Proc. Appl.* **17**, 337–347. [239–240, 244]

Ohser, H. and Stoyan, D. (1981). On the second-order and orientation analysis of planar stationary point processes. *Biom. J.* **23**, 523–533. [298]

Orey, S. (1971). *Limit Theorems for Markov Chain Transition Probabilities.* Van Nostrand–Reinhold, London. [92]

Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Ann. Inst. Statist. Math.* **31**, 145–155. [309]

Palm, C. (1943). *Intensitätsschwankungen im Fernsprechverkehr*, Ericsson Technics **44**. [13, 328]

Papangelou, F. (1972a). Summary of some results on point and line processes. In Lewis (1972), pp. 522–532. [23]

Papangelou, F. (1972b). Integrability of expected increments of point processes and a related random change of scale. *Trans. Amer. Math. Soc.* **165**, 483–506. [23]

—— (1974). On the Palm probabilities of processes of points and processes of lines. In Harding and Kendall (1974), pp. 114–147. [257]

Parthasarathy, K.R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York. [381]

Pogorzelski, W.A. (1966). *Integral Equations and Their Applications*. Pergamon Press, Oxford, and PWN, Warsaw. [142]

Poisson, S.D. (1837). *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile, Précédées des Règles Générales du Calcul des Probabilités*. Bachelier, Paris. [8]

Pólya, G. (1931). Sur quelques points de la théorie des probabilités. *Ann. Inst. Henri Poincaré* **1**, 117–162. [11]

Prékopa, A. (1957a). On the compound Poisson distribution. *Acta Sci. Math. Szeged.* **18**, 23–28. [37]

—— (1957b). On Poisson and composed Poisson stochastic set functions. *Stud. Math.* **16**, 142–155. [37]

Preston, C.J. (1976). *Random Fields*, Lecture Notes in Mathematics **534**. Springer-Verlag, New York. [128]

Prohorov, Yu.V. (1956). Convergence of random processes and limit theorems in probability theory (in Russian). *Teor. Veroyatnost. i Primenen.* **1**, 177–238. [Translation in *Theory Probab. Appl.* **1**, 157–214.] [15]

Quine, M.P. and Watson, D.F. (1984). Radial simulation of n-dimensional Poisson processes. *J. Appl. Probab.* **21**, 548–557. [25]

Ramakrishnan, A. (1950). Stochastic processes relating to particles distributed in a continuous infinity of states. *Proc. Cambridge Philos. Soc.* **46**, 595–602. [15, 111, 136]

Rao, C.R. and Shanbhag, D.N. (1986). Recent results on characterizations of probability distributions: A unified approach through extensions of Deny's theorem. *Adv. Appl. Probab.* **18**, 660–678. [74]

Rényi, A. (1967). Remarks on the Poisson process. *Stud. Sci. Math. Hungar.* **5**, 119–123. [31, 33]

Resnick, S.I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag, New York. [17]

Rice, S.O. (1944). Mathematical analysis of random noise. *Bell Syst. Tech. J.* **23**, 282–332 and **24**, 46–156. [Reprinted in N. Wax (Ed.) (1954). *Selected Papers on Noise and Stochastic Processes*. Dover, New York, pp. 133–294.] [14]

Ripley, B. D. (1976). The second-order analysis of spatial point processes. *J. Appl. Probab.* **13**, 255–266. [297, 300]

—— (1977). Modelling spatial patterns (with Discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 172–212. [128, 297]

—— (1981). *Spatial Statistics*. Wiley, New York. [16–17, 111, 222, 297, 300]

—— (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press, Cambridge. [320]

—— and Kelly, F.P. (1977). Markov point processes. *J. London Math. Soc.* **15**, 188–192. [218]

Robertson, A.P. and Thornett, M.L. (1984). On translation bounded measures. *J. Aust. Math. Soc. Ser. A* **37**, 139–142. [358, 367]

Rubin, I. (1972). Regular point processes and their detection. *IEEE Trans. Inf. Theory* **IT-18**, 547–557. [211]

Rudemo, M. (1964). Dimension and entropy for a class of stochastic processes. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **9**, 73–87. [286]

Ruelle, D. (1969). *Statistical Mechanics: Rigorous Results*. Benjamin, New York. [128]

Ryll-Nardzeweki, C. (1961). Remarks on processes of calls. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **2**, 455–465. [14]

Schäl, M. (1971). Über Lösungen einer Erneuerungsgleichung. *Abh. Math. Sem. Univ. Hamburg* **36**, 89–98. [90]

Schlather, M. (2001). Second order characteristics of marked point processes. *Bernoulli* **7**, 99–117. [327]

Schoenberg, F. (1999). Transforming spatial point processes into Poisson processes. *Stoch. Proc. Appl.* **81**, 155–164. [266]

Schoenberg, F.P. (2002). On rescaled Poisson processes and the Brownian bridge. *Ann. Inst. Statist. Math.* **54**, 445–457. [262]

Schwarz, L. (1951). *Théorie des Distributions*, Vol. II. Hermann, Paris. [357]

Seidel, H. (1876). Über die Probabilitäten solcher Ereignisse welche nur seiten vorkommen, obgleich sie unbeschränkt oft möglich sind. *Sitzungsber. Math. Phys. Cl. Akad. Wiss. München* **6**, 44–50. [8]

Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press, London. [17]

Sgibnev, M.S. (1981). On the renewal theorem in the case of infinite variance. *Sibirsk. Mat. Zh.* **22** (5), 178–189. [Translation in *Siberian Math. J.* **22**, 787–796.] [91]

Shi, Y., Liu, J., and Zhang, S. (2001). An evaluation of Chinese annual earthquake predictions, 1990–1998. In Daley, D.J. (Ed.), *Probability, Statistics and Seismology, J. Appl. Probab.* **38A**, 222–231. [276]

Slivnyak, I.M. (1962). Some properties of stationary flows of homogeneous random events. *Teor. Veroyatnost. i Primenen.* **7**, 347–352. [Translation in *Theory Probab. Appl.* **7**, 336–341.] [60]

—— (1966). Stationary streams of homogeneous random events. *Vestn. Harkov. Gos. Univ. Ser. Mech. Math.* **32**, 73–116. [60]

Smith, W.L. (1958). Renewal theory and its ramifications (with Discussion). *J. Roy. Statist. Soc. Ser. B* **20**, 284–302. [72]

—— (1962). On necessary and sufficient conditions for the convergence of the renewal density. *Trans. Amer. Math. Soc.* **104**, 79–100. [91]

Snyder, D.L. (1972). Filtering and detection for doubly stochastic Poisson processes. *IEEE Trans. Inf. Theory* **IT-18**, 97–102. [211]

—— (1975). *Random Point Processes*. Wiley, New York. [16, 211]

—— and Miller, M.I. (1991). *Random Point Processes in Time and Space*. Wiley, New York. [= 2nd ed. of Snyder (1975).] [17, 211]

Solomon, H. and Wang, P.C.C. (1972). Nonhomogeneous Poisson fields of random lines with applications to traffic flow. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **3**, 383–400. [329]

Srinivasan, S.K. (1969). *Stochastic Theory and Cascade Processes*. American Elsevier, New York. [15, 124]

—— (1974). *Stochastic Point Processes and Their Applications*. Griffin, London. [15]

Stone, C. (1966). On absolutely continuous components and renewal theory. *Ann. Math. Statist.* **37**, 271–275. [88, 90]

Stoyan, D. (1983). *Comparison Methods for Queues and Other Stochastic Models*. Wiley, Chichester. [7]

—— (1984). On correlations of marked point processes. *Math. Nachr.* **116**, 197–207. [327]

——, Kendall, W.S., and Mecke, J. (1987). *Stochastic Geometry*. Akademie-Verlag, Berlin, and Wiley, Chichester. [17, 111]

——, Kendall, W.S., and Mecke, J. (1995). *Stochastic Geometry*, 2nd ed. Wiley, Chichester. [1st ed. Stoyan *et al.* (1987).] [17, 111, 205–206, 222]

—— and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields*. Wiley, Chichester. [17, 111, 205, 222]

Strauss, D.J. (1975). A model for clustering. *Biometrika* **62**, 467–475. [217]

'Student' (1907). On the error of counting with a haemacytometer. *Biometrika* **5**, 351–360. [9–10]

Szasz, D.O.H. (1970). Once more on the Poisson process. *Stud. Sci. Math. Hungar.* **5**, 441–444. [31]

Takacs, L. (1965). A moment problem. *J. Aust. Math. Soc.* **5**, 487–490. [122]

—— (1967). On the method of inclusion and exclusion. *J. Amer. Statist. Assoc.* **62**, 102–113. [119]

—— (1976). Some remarks on a counter process. *J. Appl. Probab.* **13**, 623–627. [105]

Teugels, J.L. (1968). Renewal theorems when the first or the second moment is infinite. *Ann. Math. Statist.* **39**, 1210–1219. [106]

Thedeen, T. (1964). A note on the Poisson tendency in traffic distribution. *Ann. Math. Statist.* **35**, 1823–1824. [329]

Thompson, H.R. (1955). Spatial point processes with applications to ecology. *Biometrika* **42**, 102–115. [181]

Thorisson, H. (2000). *Coupling, Stationarity and Regeneration*. Springer, New York. [83]

Thornett, M.L. (1979). A class of second-order stationary random measures. *Stoch. Proc. Appl.* **8**, 323–334. [338–340, 342, 357–358, 367]

Titchmarsh, E.C. (1937). *Introduction to the Theory of Fourier Integrals*. Oxford University Press, Oxford. [411]

Tyan, S. and Thomas, J.B. (1975). Characterization of a class of bivariate distribution functions. *J. Multivariate Anal.* **5**, 227–235. [95]

Utsu, T., Ogata, Y., and Matsu'ura, R.S. (1995). The centenary of the Omori formula for the decay law of aftershock activity. *J. Phys. Earth* **43**, 1–33. [263]

Varnes, D.J. (1989). Predicting earthquakes by analyzing accelerating precursory seismic activity. *Pure Appl. Geophys.* **130**, 661–686. [257]

Vasil'ev, P.I. (1965). On the question of ordinariness of a stationary stream. *Kisinev. Gos. Univ. Ucen. Zap.* **82**, 44–48. [52]

Vere-Jones, D. (1970). Stochastic models for earthquake occurrences (with Discussion). *J. Roy. Statist. Soc. Ser. B* **32**, 1–62. [192, 325, 344]

—— (1974). An elementary approach to the spectral theory of stationary random measures. In Harding and Kendall (1974), pp. 307–321. [331, 334, 341, 345, 357]

—— (1975). A renewal equation for point processes with Markov-dependent intervals. *Math. Nachr.* **68**, 133–139. [95, 103]

—— (1978a). Space time correlations for microearthquakes—a pilot study. Supplement to *Adv. Appl. Probab.* **10**, 73–87. [297–298, 300, 303]

—— (1978b). Earthquake prediction—a statistician's view. *J. Phys. Earth* **26**, 129–146. [239]

—— (1982). On the estimation of frequency in point-process data. In Gani, J. and Hannan, E.J. (Eds.), *Essays in Statistical Science, J. Appl. Probab.* **19A**, 383–394. [226]

—— (1984). An identity involving permanents. *Linear Alg. Appl.* **63**, 267–270. [175]

—— (1988). On the variance properties of stress release models. *Aust. J. Statist.* **30A**, 123–135. [240–241, 245]

—— (1995). Forecasting earthquakes and earthquake risk. *Int. J. Forecasting* **11**, 503–538. [17]

—— (1997). Alpha-permanents and their applications to multivariate gamma, negative binomial and ordinary binomial distributions. *N.Z. J. Math.* **26**, 125–149. [14, 140, 175]

—— (1999). Probabilities and information gain for earthquake forecasting. *Comput. Seismol.* **30** (Geodynamics and Seismology), 248–263. [286]

—— and Davies, R.B. (1966). A statistical survey of earthquakes in the main seismic region of New Zealand. Part II, Time Series Analysis. *N.Z. J. Geol. Geophys.* **9**, 251–284. [163, 344]

—— and Musmeci, F. (1992). A space–time clustering model for historical earthquakes. *Ann. Inst. Statist. Math.* **44**, 1–11. [204]

—— and Ogata, Y. (1984). On the moments of a self-correcting process. *J. Appl. Probab.* **21**, 335–342. [240]

—— and Ozaki, T. (1982). Some examples of statistical inference applied to earthquake data. *Ann. Inst. Statist. Math.* **34**, 189–207. Correction (1987), *Ann. Inst. Statist. Math.* **39**, 243. [203, 234, 309, 337]

——, Robinson, R., and Yang, W. (2001). Remarks on the accelerated moment release model: Problems of model formulation, simulation and estimation. *Geophys. J. Internat.* **144**, 517–531. [256–257]

Von Bortkiewicz, L. (1898). *Das Gesetz der kleinen Zalhlen*. G. Teubner, Leipzig. [See M.P. Quine and E. Seneta (1987). Bortkiewicz's data and the law of small numbers. *Internat. Statist. Rev.* **55**, 173–181.] [9]

Warren, W.G. (1962). *Contributions to the Study of Spatial Point Processes*. Ph.D. thesis, University of North Carolina, Chapel Hill (Statistics Dept. Mimeo Series 337). [181]

—— (1971). The centre-satellite concept as a basis for ecological sampling. In Patil, G.P., Pielou, E.C., and Waters, W.E. (Eds.), *Statistical Ecology*, Vol. 2, Pennsylvania State University Press, University Park, PA, pp. 87–118. [181]

Watanabe, S. (1933). On the theory of durability. *Geophys. Mag. (Tokyo)* **7**, 307–317. [7]

Watanabe, S. (1964). On discontinuous additive functionals and Levy measures of a Markov process. *Japanese J. Math.* **34**, 53–70. [211, 257]

Weibull, W. (1939a). A statistical theory of the strength of materials. *Ing. Vetensk. Akad. Banal. Stockholm*, No. 151. [4]

—— (1939b). The phenomenon of rupture in solids. *Ing. Vetensk. Akad. Banal. Stockholm*, No. 153. [4]

Westcott, M. (1970). Identifiability in linear processes. *Z. Wahrs.* **16**, 39–46. [169]

Whitworth, W.A. (1867). *Choice and Chance, an Elementary Treatise on Permutations, Combinations and Probability, with 300 Exercises.* Cambridge. [2nd ed. (1870); 3rd ed. (1878); 4th ed. . . . *with 640 Exercises* (1886), Deighton Bell, Cambridge; 5th ed. *Choice and Chance, with One Thousand Exercises* (1901), reprinted (1942), Stechert, New York, and (1951), Hafner Publishing, New York.] [9]

—— (1897). *DCC Exercises in Choice and Chance.* Deighton Bell, Cambridge. [Reprinted (1959), Hafner Publishing, New York.] [9]

Wold, H. (1948). On stationary point processes and Markov chains. *Skand. Aktuar.* **31**, 229–240. [14, 92, 105, 110]

—— (1949). Sur les processes stationnaires ponctuels. *Coll. Int. CNRS* **13**, 75–86. [14]

Yaglom, A.Ya. (1961). Second-order homogeneous random fields. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **2**, 593–622. [303]

Yashin, A. (1970). Filtering of jump processes. *Avtomat. i Telemekh.* **1970**(5), 52–58. [Translation in *Automat. Remote Control* **1970**, 725–730.] [211]

Yule, G.U. (1924). A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis. *Philos. Trans. B* **213**, 21–87. [9]

Yvon, J. (1935). *La Théorie Statistique des Fluides et l'Équation d'État*, Actualités Scientifiques et Industrielles **203**. Hermann, Paris. [12, 111, 124]

Zheng, X. (1991). Ergodic theorems for stress release processes. *Stoch. Proc. Appl.* **37**, 239–258. [240–241]

—— and Vere-Jones, D. (1994). Further applications of the stress release model to historical earthquake data. *Tectonophysics* **229**, 101–121. [240]

Zygmund, A. (1968). *Trigonometric Series*, 2nd ed. Cambridge University Press, Cambridge. [341]

Subject Index