

***Probability and Its Applications***

**D.J. Daley  
D. Vere-Jones**

**An Introduction to  
the Theory of  
Point Processes**

**Volume II: General Theory  
and Structure**

**Second Edition**

# Probability and Its Applications

---

*A Series of the Applied Probability Trust*

*Editors: J. Gani, C.C. Heyde, P. Jagers, T.G. Kurtz*

# Probability and Its Applications

---

*Azencott et al.*: Series of Irregular Observations. Forecasting and Model Building  
*Bass*: Diffusions and Elliptic Operators  
*Bass*: Probabilistic Techniques in Analysis  
*Berglund/Gentz*: Noise-Induced Phenomena in Slow-Fast Dynamical Systems: A Sample-Paths Approach  
*Chen*: Eigenvalues, Inequalities and Ergodic Theory  
*Costa/Fragoso/Marques*: Discrete-Time Markov Jump Linear Systems  
*Daley/Vere-Jones*: An Introduction to the Theory of Point Processes I: Elementary Theory and Methods, Second Edition  
*Daley/Vere-Jones*: An Introduction to the Theory of Point Processes II: General Theory and Structure, Second Edition  
*de la Pena/Gine*: Decoupling: From Dependence to Independence, Randomly Stopped Processes U-Statistics and Processes Martingales and Beyond  
*Del Moral*: Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications  
*Durrett*: Probability Models for DNA Sequence Evolution  
*Galambos/Simonelli*: Bonferroni-Type Inequalities with Equations  
*Gani (ed.)*: The Craft of Probabilistic Modelling. A Collection of Personal Accounts  
*Guyon*: Random Fields on a Network. Modeling, Statistics and Applications  
*Kallenberg*: Foundations of Modern Probability, 2nd edition  
*Kallenberg*: Probabilistic Symmetries and Invariance Principles  
*Last/Brandt*: Marked Point Processes on the Real Line  
*Molchanov*: Theory of Random Sets  
*Nualart*: The Malliavin Calculus and Related Topics, 2nd edition  
*Shedler*: Regeneration and Networks of Queues  
*Silvestrov*: Limit Theorems for Randomly Stopped Stochastic Processes  
*Rachev/Rueschendorf*: Mass Transportation Problems. Volume I: Theory  
*Rachev/Rueschendorf*: Mass Transportation Problems. Volume II: Applications  
*Resnick*: Extreme Values, Regular Variation and Point Processes  
*Thorisson*: Coupling, Stationarity and Regeneration

D.J. Daley      D. Vere-Jones

# An Introduction to the Theory of Point Processes

Volume II: General Theory and Structure

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 Mathematics, Statistics  
and Computing Science  
Victoria University of Wellington  
Wellington, New Zealand  
David.Vere-Jones@mcs.vuw.ac.nz

*Series Editors:*

J. Gani  
Stochastic Analysis Group, CMA  
Australian National University  
Canberra. ACT 0200  
Australia

P. Jagers  
Department of Mathematical Sciences  
Chalmers University of Technology  
and Göteborg (Gothenburg)  
SE-412 96 Göteborg  
Sweden

C.C. Heyde  
Stochastic Analysis Group, CMA  
Australian National University  
Canberra, ACT 0200  
Australia

T.G. Kurtz  
Department of Mathematics  
University of Wisconsin  
480 Lincoln Drive  
Madison, WI 53706  
USA

ISBN 978-0-387-21337-8

e-ISBN 978-0-387-49835-5

Library of Congress Control Number: 2007936157

© 2008, 1988 by the Applied Probability Trust

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com



*To Nola,  
and in memory of Mary*



## Preface to Volume II, Second Edition

In this second volume, we set out a general framework for the theory of point processes, starting from their interpretation as random measures. The material represents a reorganized version of those parts of Chapters 6–14 of the first edition not already covered in Volume I, together with a significant amount of new material.

Contrary to our initial expectations, growth in the theoretical aspects of the subject has at least matched the growth in applications. Much of the original text has been substantially revised in order to present a more consistent treatment of marked as well as simple point processes. This applies particularly to the material on stationary processes in Chapter 12, the Palm theory covered in Chapter 13, and the discussion of martingales and conditional intensities in Chapter 14. Chapter 15, on spatial point processes, has also been significantly modified and extended. Essentially new sections include Sections 10.3 and 10.4 on point processes defined by Markov chains and Markov point processes in space; Sections 12.7 on long-range dependence and 12.8 on scale invariance and self-similarity; Sections 13.4 on marked point processes and convergence to equilibrium and 13.6 on fractal dimensions; Sections 14.6 on random time changes and 14.7 on Poisson embedding and convergence to equilibrium; much of the material in Sections 15.1–15.4 on spatial processes is substantially new or revised; and some recent material on point maps and point stationarity has been included in Section 13.3.

As in the first edition, much of the general theory has been developed in the context of a complete separable metric space (c.s.m.s. throughout this volume). Critical to this choice of context is the existence of a well-developed theory of measures on metric spaces, as set out, for example, in Parthasarathy

(1967) or Billingsley (1968). We use this theory at two levels. First, we establish results concerning the space<sup>1</sup>  $\mathcal{M}_{\mathcal{X}}^{\#}$  and  $\mathcal{N}_{\mathcal{X}}^{\#}$  of realizations of random measures and point processes, showing that these spaces themselves can be regarded as c.s.m.s.s, and paying particular attention to sample path properties such as the existence of atoms. Second, leaning on these results, we use the same framework to discuss the convergence of random measures and point processes. The fact that the same theory appears at both levels lends unity and economy to the development, although care needs to be taken in discriminating between the two levels.

The text of this volume necessarily assumes greater familiarity with aspects of measure theory and topology than was the case in Volume I, and the first two appendices at the end of Volume I are aimed at helping the reader in this regard. The third appendix reviews some of the material from martingale theory and the general theory of processes that underlies the discussion of predictability and conditional intensities in Chapter 14.

As was the case in Volume I, we are very much indebted to the friends, critics, reviewers, and readers who have supplied us with comments, suggestions, and corrections at various stages in the preparation of this volume. The list is too long to include in full, but we would like to mention in particular the continuing support and advice we have had from Robin Milne, Val Isham, Rick Schoenberg, Gunther Last, and our long-suffering colleagues in Canberra and Wellington. The patience and expertise of Springer Verlag, as mediated through our long-continued contacts with John Kimmel, are also very much appreciated.

Daryl Daley  
Canberra, Australia

David Vere-Jones  
Wellington, New Zealand

---

<sup>1</sup> 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.



# Contents

Preface to Volume II, Second Edition	vii
Principal Notation	xii
Concordance of Statements from the First Edition	xvi
<b>9 Basic Theory of Random Measures and Point Processes</b>	<b>1</b>
9.1 Definitions and Examples	2
9.2 Finite-Dimensional Distributions and the Existence Theorem	25
9.3 Sample Path Properties: Atoms and Orderliness	38
9.4 Functionals: Definitions and Basic Properties	52
9.5 Moment Measures and Expansions of Functionals	65
<b>10 Special Classes of Processes</b>	<b>76</b>
10.1 Completely Random Measures	77
10.2 Infinitely Divisible Point Processes	87
10.3 Point Processes Defined by Markov Chains	95
10.4 Markov Point Processes	118
<b>11 Convergence Concepts and Limit Theorems</b>	<b>131</b>
11.1 Modes of Convergence for Random Measures and Point Processes	132
11.2 Limit Theorems for Superpositions	146
11.3 Thinned Point Processes	155
11.4 Random Translations	166

<b>12 Stationary Point Processes and Random Measures</b>	<b>176</b>
12.1 Stationarity: Basic Concepts	177
12.2 Ergodic Theorems	194
12.3 Mixing Conditions	206
12.4 Stationary Infinitely Divisible Point Processes	216
12.5 Asymptotic Stationarity and Convergence to Equilibrium	222
12.6 Moment Stationarity and Higher-order Ergodic Theorems	236
12.7 Long-range Dependence	249
12.8 Scale-invariance and Self-similarity	255
<b>13 Palm Theory</b>	<b>268</b>
13.1 Campbell Measures and Palm Distributions	269
13.2 Palm Theory for Stationary Random Measures	284
13.3 Interval- and Point-stationarity	299
13.4 Marked Point Processes, Ergodic Theorems, and Convergence to Equilibrium	317
13.5 Cluster Iterates	334
13.6 Fractal Dimensions	340
<b>14 Evolutionary Processes and Predictability</b>	<b>355</b>
14.1 Compensators and Martingales	356
14.2 Campbell Measure and Predictability	376
14.3 Conditional Intensities	390
14.4 Filters and Likelihood Ratios	400
14.5 A Central Limit Theorem	412
14.6 Random Time Change	418
14.7 Poisson Embedding and Existence Theorems	426
14.8 Point Process Entropy and a Shannon–MacMillan Theorem	440
<b>15 Spatial Point Processes</b>	<b>457</b>
15.1 Descriptive Aspects: Distance Properties	458
15.2 Directional Properties and Isotropy	466
15.3 Stationary Line Processes in the Plane	471
15.4 Space–Time Processes	485
15.5 The Papangelou Intensity and Finite Point Patterns	506
15.6 Modified Campbell Measures and Papangelou Kernels	518
15.7 The Papangelou Intensity Measure and Exvisibility	526
<b>References with Index</b>	<b>537</b>
<b>Subject Index</b>	<b>557</b>

# Chapter Titles for Volume I

1	Early History	1
2	Basic Properties of the Poisson Process	19
3	Simple Results for Stationary Point Processes on the Line	41
4	Renewal Processes	66
5	Finite Point Processes	111
6	Models Constructed via Conditioning: Cox, Cluster, and Marked Point Processes	157
7	Conditional Intensities and Likelihoods	211
8	Second-Order Properties of Stationary Point Processes	288
A1	A Review of Some Basic Concepts of Topology and Measure Theory	368
A2	Measures on Metric Spaces	384
A3	Conditional Expectations, Stopping Times and Martingales	414

# Principal Notation

Very little of the general notation used in Appendices 1–3 in Volume I 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. Generally, the particular interpretation of symbols with more than one use is clear from the context. Where they are given, page numbers indicate the first or significant use of the notation. Page numbers in slant font, such as *158*, refer to Volume I, but such references are not intended to be comprehensive.

Throughout the lists below,  $N$  denotes a point process,  $\xi$  a random measure (or sometimes, as on p. 358, a cumulative process on  $\mathbb{R}_+$ ), and  $\mathcal{X}$  a c.s.m.s.

## Spaces

$\mathbb{C}$	complex numbers	
$\mathbb{R} = \mathbb{R}^1$	real line	
$\mathbb{R}_+, \mathbb{R}_0^+$	nonnegative numbers, positive numbers	358
$\mathbb{R}^d$	$d$ -dimensional Euclidean space	
$\mathbb{S}$	circle group and its representation as $(0, 2\pi]$	
$\mathbb{U}_{2\alpha}^d$	$d$ -dimensional cube of side length $2\alpha$ and vertices $(\pm\alpha, \dots, \pm\alpha)$	
$\mathbb{X}$	countable state space for Markov chain	96
$\mathbb{Z}, \mathbb{Z}_+$	integers of $\mathbb{R}$ , $\mathbb{R}_+$	
$\mathcal{X}$	state space of $N$ or $\xi$ ; often $\mathcal{X} = \mathbb{R}^d$ ; always $\mathcal{X}$ is c.s.m.s. (complete separable metric space)	
$\Omega$	space of probability elements $\omega$	
$\mathcal{E}$	measurable sets in probability space	
$(\Omega, \mathcal{E}, \mathcal{P})$	basic probability space on which $N$ and $\xi$ are defined	<i>158</i> , 7
$\mathcal{X}^{(n)}$	$n$ -fold product space $\mathcal{X} \times \dots \times \mathcal{X}$	<i>129</i>
$\mathcal{X}^\cup$	$= \mathcal{X}^{(0)} \cup \mathcal{X}^{(1)} \cup \dots$	<i>129</i>

$\mathcal{B}(\mathcal{X})$	Borel $\sigma$ -field generated by open spheres of $\mathcal{X}$	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 $\sigma$ -field on product space $\mathcal{X}^{(n)}$	129
$\text{BM}(\mathcal{X})$	bounded measurable functions of bounded support	161, 52
$\text{BM}_+(\mathcal{X})$	nonnegative functions $f \in \text{BM}(\mathcal{X})$	57
$\overline{\text{BM}}_+(\mathcal{X})$	limits of monotone sequences from $\text{BM}_+(\mathcal{X})$	57
$\mathcal{K}$	mark space for marked point process (MPP)	194, 7
$\mathcal{M}_{\mathcal{X}}(\mathcal{N}_{\mathcal{X}})$	totally finite (counting) measures on $\mathcal{X}$	158, 3
$\mathcal{M}_{\mathcal{X}}^{\#}$	boundedly finite measures on $\mathcal{X}$	158, 3
$\mathcal{N}_{\mathcal{X}}^{\#}$	boundedly finite counting measures on $\mathcal{X}$	131, 3
$\mathcal{N}_0^{\#}$	$= \mathcal{N}_{\mathcal{X}}^{\#} \setminus \{\emptyset\}$	90
$\mathcal{N}_{\mathcal{X}}^{\#*}$	simple counting measures in $\mathcal{N}_{\mathcal{X}}^{\#}$	24
$\mathcal{N}_0, \mathcal{N}_{\mathcal{X}}^{\#*}$	subset of $\mathcal{N}_{\mathcal{X}}^{\#*}$ with $N\{0\} > 0$ FIX !!!	24, 290
$\mathcal{S}^+$	doubly infinite sequences of positive numbers $\{t_0, t_{\pm 1}, t_{\pm 2}, \dots\}$ with $\sum_{n=1}^{\infty} t_n = \sum_{n=1}^{\infty} t_{-n} = \infty$	14
$\mathcal{U}$	linear space; complex-valued Borel measurable functions $\zeta$ on $\mathcal{X}$ with $ \zeta  \leq 1$	52; 57
$\mathcal{U} \otimes \mathcal{V}$	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 $h(\cdot)$ with $1 - h(\cdot)$ of bounded support in $\mathcal{X}$	59
$\mathcal{V}_0(\mathcal{X})$	$= \{h \in \mathcal{V}(\mathcal{X}) : \inf_x h(x) > 0\}$ , i.e., $-\log h \in \text{BM}_+(\mathcal{X})$	59
$\overline{\mathcal{V}}(\mathcal{X})$	limits of monotone sequences from $\mathcal{V}(\mathcal{X})$	59
$\mathcal{W} = \mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}$	product space supporting Campbell measure $C_{\mathcal{P}}$	269

## 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  $z \in \mathbb{C}$ .

$\sim$	reduced measure (by factorization)	160, 183
$\#$	extension of concept from totally finite to boundedly finite measure space	158, viii
$F^{n*}$	$n$ -fold convolution power of measure or d.f. $F$	55
$\mathbf{a}, \mathbf{g}$	suffixes for atomic measure, ground process of MPP	4, 3
$\ \mu\ $	variation norm of (signed) measure $\mu$	374
a.e. $\mu$ , $\mu$ -a.e.	almost everywhere with respect to measure $\mu$	376
a.s., $\mathcal{P}$ -a.s.	almost sure, $\mathcal{P}$ -almost surely	376
$A(\cdot), A^{\mathcal{F}}(\cdot)$	$\mathcal{F}$ -compensator for $\xi$ on $\mathbb{R}_+$	358
$A^{(n)}$	$n$ -fold product set $A \times \dots \times A$	130
$\mathcal{A}$	family of sets generating $\mathcal{B}$ ; semiring of bounded Borel sets generating $\mathcal{B}_{\mathcal{X}}$	31, 368
$c_k, c_{[k]}$	$k$ th cumulant, $k$ th 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, 69
$C_k(\cdot), C_{[k]}(\cdot)$	cumulant, factorial cumulant measure	147, 69
$C_2(A \times B) = \text{cov}(\xi(A), \xi(B))$	covariance measure of $\xi$	191, 69
$\check{C}_2(\cdot)$	reduced covariance measure of stationary $N$ or $\xi$	292, 238
$C_{\mathcal{P}}, C_{\mathcal{P}}^!$	Campbell measure, modified Campbell measure	269, 270
$\check{C}_{\mathcal{P}}(\cdot)$	reduced Campbell measure (= Palm measure)	287, 331
$\delta(\cdot)$	Dirac delta function	
$\delta_x(A)$	Dirac measure, $= \int_A \delta(u - x) du = I_A(x)$	382, 3
$\mathcal{D}_\alpha$	Dirichlet process	12
$\Delta F(x) = F(x) - F(x-)$	jump at $x$ in right-continuous function $F$	107
$\Delta^L, \Delta^R$	left- and right-hand discontinuity operators	376
$F(\cdot; \cdot)$	finite-dimensional (fidi) distributions	158, 26
$\mathcal{F}; \mathcal{F}^\dagger$	history on $\mathbb{R}_+; \mathbb{R}$	236, 356; 394
$\Phi(\cdot)$	characteristic functional	15, 54
$G[h] \ (h \in \mathcal{V})$	probability generating functional (p.g.fl.) of $N$	144, 59
$G_c[\cdot]$	p.g.fl. of cluster centre process $N_c$	178
$G_m[\cdot   x]$	p.g.fl. of cluster member process $N_m(\cdot   x)$	178, 192
$G$	expected information gain of stationary $N$ on $\mathbb{R}$	280, 442
$\Gamma(\cdot)$	Bartlett spectrum for stationary $\xi$ on $\mathbb{R}^d$	304, 205
$H(t)$	integrated hazard function (IHF) [ $Q(t)$ in Vol.I]	109, 361
$H(\mathcal{P}; \mu)$	generalized entropy	277, 441
$\mathcal{H}; \mathcal{H}^\dagger$	internal history of $\xi$ on $\mathbb{R}_+; \mathbb{R}$	236, 358; 395
$I_A(x) = \delta_x(A)$	indicator function of element $x$ in set $A$	
$\mathcal{I}$	$\sigma$ -field of events invariant under shift operator $S_u$	194
$J_n(A_1 \times \cdots \times A_n)$	Janossy measure	124
$J_n(\cdot   A)$	local Janossy measure	137, 73
$j_n(x_1, \dots, x_n)$	Janossy density	125, 119, 506
$K$	compact set; generic Borel set in mark space $\mathcal{K}$	371, 8
$\ell(\cdot)$	Lebesgue measure in $\mathcal{B}(\mathbb{R}^d)$	31
$\ell_{\mathcal{K}}(\cdot)$	reference measure on mark space	401
$L[f] \ (f \in BM_+(\mathcal{X}))$	Laplace functional of $\xi$	161, 57
$L_\xi[1 - h]$	p.g.fl. of Cox process directed by $\xi$	170
$\lambda(\cdot), \lambda$	intensity measure of $N$ , intensity of stationary $N$	44, 46
$\lambda^*(t, \omega)$	conditional intensity function	231, 390
$\lambda^\dagger(t, \kappa, \omega)$	complete intensity function for stationary MPP on $\mathbb{R}$	394
$m_k(\cdot) \ (m_{[k]}(\cdot))$	$k$ th (factorial) moment density	136
$\check{m}_2, \check{M}_2$	reduced second-order moment density, measure, of stationary $N$	289
$m_g$	mean density of ground process $N_g$ of MPP $N$	198, 323
$M(A)$	expectation measure $E[\xi(A)]$	65
$M_k(\cdot)$	$k$ th-order moment measure $E[\xi^{(k)}(\cdot)]$	66
$N(A)$	number of points in $A$	42
$N(a, b], = N((a, b])$	number of points in half-open interval $(a, b]$	19, 42

$N(t)$	$= N(0, t] = N((0, t])$	42
$N_c$	cluster centre process	176
$N_m(\cdot \mid x)$	cluster member or component process	176
$N_g$	ground process of MPP	194, 7
$N^*$	support counting measure of $N$	4
$\{(p_n, \Pi_n)\}$	probability measure elements for finite point process	123
$P = (p_{ij})$	matrix of one-step transition probabilities $p_{ij}$ of discrete-time Markov chain on countable state space $\mathbb{X}$	96
$P(z)$	probability generating function (p.g.f.) of distribution $\{p_n\}$	10, 115
$P_0(A)$	avoidance function	31, 33
$\mathcal{P}$	probability measure of $N$ or $\xi$ on c.s.m.s. $\mathcal{X}$	158, 6
$\mathcal{P}_0(\cdot)$	Palm distribution for stationary $N$ or $\xi$ on $\mathbb{R}$	288
$\bar{\mathcal{P}}_0$	averaged (= mean) Palm measure for stationary MPP	319
$\mathcal{P}_{(0, \kappa)}(\cdot)$	Palm measure for $\kappa \in \mathcal{K}$	318
$\mathcal{P}_{x, \kappa}(\cdot)$	local Palm measure for $(x, \kappa) \in \mathcal{X} \times \mathcal{K}$	318
$\{\pi_k\}$	stationary distribution for $(p_{ij})$	96
$\emptyset, \emptyset(\cdot)$	empty set; null measure	17; 88, 292
$Q = (q_{ij})$	$Q$ -matrix of transition rates $q_{ij}$ for continuous-time Markov chain on countable state space $\mathbb{X}$	97
$\rho(x, y)$	metric for $x, y$ in metric space	370
$\rho(y \mid \mathbf{x})$	Papangelou (conditional) intensity	120, 506
$\{S_i\}$	nested bounded sets, $S_i \uparrow \mathcal{X}$ ( $i \rightarrow \infty$ )	16
$\{S_n\}$	random walk, sequence of partial sums	66
$S_r(x)$	sphere of radius $r$ , centre $x$ , in metric space $\mathcal{X}$	371, 5
$S_r$	$= S_r(0)$	459
$\{t_i(N)\}, \{t_i\}$	successive points of $N$ on $\mathbb{R}$ , $t_{-1} < 0 \leq t_0$	15
$\{\tau_i\}$	intervals between points of $N$ on $\mathbb{R}$ , $\tau_i = t_i - t_{i-1}$	15
$\mathcal{T} = \{\mathcal{T}_n\} = \{\{A_{ni}\}\}$	dissecting system of nested partitions	382, 10
	tiling	16
$\mathcal{T}_\infty$	tail $\sigma$ -algebra of process on $\mathbb{R}^d$	208
$U(A) = E[N(A)]$	renewal measure	67

# 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 both volumes of this second edition.

1988 edition	this edition	1988 edition	this edition
2.2.I-III	2.2.I-III	7.1.I-II	9.1.II-III
2.3.III	2.3.I	7.1.III	9.1.VII
2.4.I-II	2.4.I-II	7.1.IV-VI	9.1.IV
2.4.V-VIII	2.4.III-VI	7.1.VII	9.1.V
3.2.I-6.V	3.2.I-6.V	7.1.VIII and 6	9.1.VIII
4.2.I-6.V	4.2.I-6.V	7.1.IX-X	9.1.XV, XII
5.2.I-VII	5.2.I-VII	7.1.XI	9.2.X
5.3.I-III	5.3.I-III	7.1.XII-XIII	6.4.I(a)-(b)
5.4.I-III	5.4.I-III	7.2.I	9.3.VII
5.4.IV-VI	5.4.V-VII	7.2.II	9.3.VIII-IX
5.5.I	5.5.I	7.2.III	9.3.VIII
6.1.I	9.1.I	7.2.IV	9.3.X
6.1.II and 7	9.1.VI	7.2.V-VIII	9.3.XII-XV
6.1.III-IV and 7	9.1.VIII-IX	7.3.I-V	9.2.XI-XV
6.1.V-VII and 7	9.1.XIV-XVI	7.4.I-II	9.4.IV-V
6.2.I-IX	9.2.I-IX	7.4.III	9.5.VI
6.3.I-VI	9.3.I-VI	7.4.IV-V	9.4.VII-VIII
6.3.VII-IX	10.1.II-IV	7.4.VI	9.5.II
6.4.I-II	9.4.I-II	7.4.VII	9.4.IX
6.4.III	9.5.I	8.1.I	(6.1.13)
6.4.IV-V	9.5.IV-V	8.1.II	6.1.II, IV
6.4.VI	9.4.III	8.2.I	6.3.I
6.4.VII-IX	9.4.VI-VIII	8.2.II	6.3.II, (6.3.6)
		8.2.III	Ex.12.1.6



1988 edition	this edition	1988 edition	this edition
8.2.IV	(6.3.6)	12.2.VI–VIII	13.4.I–III
8.3.I–III	6.3.III–V	12.3.I	13.1.IV
8.4.I–VIII	10.2.I–VIII	12.3.II–VI	13.3.I–V
8.5.I–III	6.2.II	12.4.I	13.4.IV
9.1.I–VII	11.1.II–VIII	12.4.II	Ex.13.4.7
9.1.IX	11.1.IX	12.4.III	13.4.V
9.1.X	11.1.XI	12.4.IV–VI	13.4.VII–IX
9.1.XI	12.3.X	13.1.I–III	7.1.I–III
Ex.9.1.3	11.1.X	13.1.IV–VI	7.2.I–III
9.2.I–VI	11.2.I–VI	13.1.VII	7.1.IV
(9.2.12)	11.2.VII	13.2.I–II	14.1.I–II
9.2.VII	10.2.IX	Ex.13.2(b)	14.1.III
9.2.VIII	11.2.VIII	13.2.III–IV	14.1.IV–V
9.3.I–V	11.3.I–V	13.2.V	14.1.VII
9.4.I–V	11.4.I–V	13.2.VI	14.2.I
9.4.VI–IX	13.6.I–IV	13.2.VII–IX	14.2.VII
10.1.I–III	12.1.I–III	13.3.I	14.2.II
10.1.IV	12.1.VI	13.3.II–IV	14.3.I–III
10.1.V–VI	12.4.I–II	13.3.V–VIII	14.4.I–IV
10.2.I–V	12.2.I–V	13.3.IX–XI	14.5.I–III
10.2.VI	12.2.IV	13.4.I–III	14.6.I–III
10.2.VII–VIII	12.2.VI–VII	13.4.III	7.4.I
10.3.I–IX	12.3.I–IX	13.4.IV	14.2.VIII
10.3.X–XII	12.4.III–V	13.4.V	14.2.VII
10.4.I–II	12.6.I–II	13.4.VI	14.2.IX
10.4.III	12.1.IV	13.5.I–II	pp. 394–396
10.4.IV–VII	12.6.III–VI	13.5.III	14.3.IV
10.5.II–III	15.2.I–II	13.5.IV–V	14.3.V–VI
10.6.I–VIII	15.3.I–VIII	13.5.VI	14.8.I
11.1.I–V	8.6.I–V	13.5.VII–IX	14.8.V–VII
11.2.I–II	8.2.I–II	13.5.X	(14.8.9)
11.3.I–VIII	8.4.I–VIII	13.5.XI	14.8.VIII
11.4.I–IV	8.5.I–IV	14.2.I–V	15.6.I–III, V–VI
11.4.V–VI	8.5.VI–VII	14.2.VI–VII	15.7.I–II
12.1.I–III	13.1.I–III	14.3.I–III	15.7.III–V
12.1.IV–VI	13.1.V–VII	Appendix identical except for	
12.2.I	13.2.I	A2.1.IV	A1.6.I
12.2.II–V	13.2.III–VI	A2.1.V–VI	A2.1.IV–V

## CHAPTER 9

# Basic Theory of Random Measures and Point Processes

9.1	Definitions and Examples	2
9.2	Finite-Dimensional Distributions and the Existence Theorem	25
9.3	Sample Path Properties: Atoms and Orderliness	38
9.4	Functionals: Definitions and Basic Properties	52
9.5	Moment Measures and Expansions of Functionals	65

This chapter sets out a framework for developing point process theory as part of a general theory of random measures. This framework was developed during the 1940s and 1950s, and reached a definitive form in the now classic treatments by Moyal (1962) and Harris (1963). It still provides the basic framework for describing point processes both on the line and in higher-dimensional spaces, including especially the treatment of finite-dimensional distributions, moment structure, and generating functionals. In the intervening decades, many important alternative approaches have been developed for more specialized classes of processes, particularly those with an evolutionary structure, and we come to some at least of these in later chapters.

As far as is convenient, we develop the theory in a dual setting, stating results for general random measures alongside the more specific more clearly the features that are peculiar to point processes. Thus, for results that hold in this unified context, proofs are usually given only in the former, more general, setting.

Furthermore, the setting for point processes also handles many of the topics of this chapter for marked point processes (MPPs): an MPP in state space  $\mathcal{X}$  with mark space  $\mathcal{K}$  can be regarded as a point process on the product space  $\mathcal{X} \times \mathcal{K}$  so far as fidi distributions, generating functionals, and moment measures are concerned. It is only when we consider particular cases, such as Poisson and compound Poisson processes or purely atomic random measures, that distinctions begin to emerge, and become more apparent as we move to

discuss stationary processes in Chapter 12 and Palm theory and martingale properties in Chapters 13 and 14.

The other major approach to point process theory is through random sequences of points. We note that this is equivalent to our approach through random measures, at least in our setting that includes point processes in finite-dimensional Euclidean space  $\mathbb{R}^d$ .

Section 9.1 sets out some basic definitions and illustrates them with a variety of examples. The second section introduces the finite-dimensional (fidi) distributions and establishes both basic existence theorems and a version of Rényi's theorem that simple point processes are completely characterized by the behaviour of the *avoidance function* (vacuity function, empty space function), viz. the probability  $P_0(A) \equiv \mathcal{P}\{N(A) = 0\}$  over a suitably rich class of Borel sets  $A$ . Section 9.3 is concerned with the sample path properties of random measures and point processes, and includes a detailed discussion of simplicity (orderliness) for point processes. The final two sections treat generating functionals and moment properties, extending the treatment for finite point processes given in Chapter 5.

## 9.1. Definitions and Examples

Let  $\mathcal{X}$  be an arbitrary complete separable metric space (c.s.m.s.) and  $\mathcal{B}_{\mathcal{X}} = \mathcal{B}(\mathcal{X})$  the  $\sigma$ -field of its Borel sets. Except for case (v) of Definition 9.1.II, all the measures that we consider on  $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$  are required to satisfy the boundedness condition set out in Definition 9.1.I. It extends to general measures the property required of counting measures in Volume I, that bounded sets have finite counting measure and hence, as point sets, they contain only finitely many points and therefore have no finite accumulation points.

**Definition 9.1.I.** *A Borel measure  $\mu$  on the c.s.m.s.  $\mathcal{X}$  is boundedly finite if  $\mu(A) < \infty$  for every bounded Borel set  $A$ .*

This constraint is incorporated into the definitions below of the spaces which form the main arena for the analysis in this volume. They incorporate the basic metric properties of spaces of measures summarized in Appendix A2 of Volume I. In particular we use from that appendix the following.

- (1) The concept of *weak convergence* of totally finite measures on  $\mathcal{X}$ , namely that  $\mu_n \rightarrow \mu$  weakly if and only if  $\int f d\mu_n \rightarrow \int f d\mu$  for all bounded continuous  $f$  on  $\mathcal{X}$  (see Section A2.3).
- (2) The extension of weak convergence of totally finite measures to  $w^\#$  (weak-hash) convergence of boundedly finite measures defined by  $\int f d\mu_n \rightarrow \int f d\mu$  for all bounded continuous  $f$  on  $\mathcal{X}$  vanishing outside a bounded set (Section A2.6).
- (3) The fact that both weak and weak-hash convergence are equivalent to forms of metric convergence, namely convergence in the Prohorov metric

at equation (A2.5.1) and its extension to the boundedly finite case given by equation (A2.6.1), respectively.

Exercise 9.1.1 shows that for sequences of totally finite measures, weak and weak-hash convergence are not equivalent.

Many of our results are concerned with one or other of the first two spaces defined below. Both are closed in the sense of the  $w^\#$ -topology referred to above, and in fact are c.s.m.s.s in their own right (Proposition 9.1.IV). At the same time it is convenient to introduce four further families of measures which play an important role in the sequel.

**Definition 9.1.II.**

- (i)  $\mathcal{M}_\mathcal{X}^\#$  is the space of all boundedly finite measures on  $\mathcal{B}_\mathcal{X}$ .
- (ii)  $\mathcal{N}_\mathcal{X}^\#$  is the space of all boundedly finite integer-valued measures  $N \in \mathcal{M}_\mathcal{X}^\#$ , called counting measures for short.
- (iii)  $\mathcal{N}_\mathcal{X}^{\#*}$  is the family of all simple counting measures, consisting of all those elements of  $\mathcal{N}_\mathcal{X}^\#$  for which

$$N\{x\} \equiv N(\{x\}) = 0 \text{ or } 1 \quad (\text{all } x \in \mathcal{X}). \quad (9.1.1)$$

- (iv)  $\mathcal{N}_{\mathcal{X} \times \mathcal{K}}^{\#g}$  is the family of all boundedly finite counting measures defined on the product space  $\mathcal{B}(\mathcal{X} \times \mathcal{K})$ , where  $\mathcal{K}$  is a c.s.m.s. of marks, subject to the additional requirement that the ground measure  $N_g$  defined by

$$N_g(A) \equiv N(A \times \mathcal{K}) \quad (\text{all } A \in \mathcal{B}_\mathcal{X}) \quad (9.1.2)$$

is a boundedly finite simple counting measure, i.e.  $N_g \in \mathcal{N}_\mathcal{X}^{\#*}$ .

- (v)  $\mathcal{M}_{\mathcal{X},a}^\#$  is the family of boundedly finite purely atomic measures  $\xi \in \mathcal{M}_\mathcal{X}^\#$ .
- (vi)  $\mathcal{M}_\mathcal{X}$  (respectively,  $\mathcal{N}_\mathcal{X}$ ) is the family of all totally finite (integer-valued) measures on  $\mathcal{B}_\mathcal{X}$ .

We introduce the family  $\mathcal{N}_{\mathcal{X} \times \mathcal{K}}^{\#g}$  to accommodate our Definition 9.1.VI(iv) of a marked point process (MPP) (as a process on  $\mathcal{X}$  with marks in  $\mathcal{K}$ ). In it we require the ground process  $N_g$  to be both simple and boundedly finite. Note that in general a simple boundedly finite counting measure on  $\mathcal{B}_{\mathcal{X} \times \mathcal{K}}$  need not be an element of this family  $\mathcal{N}_{\mathcal{X} \times \mathcal{K}}^{\#g}$ . For example, taking  $\mathcal{X} = \mathcal{K} = \mathbb{R}$ , realizations of a homogeneous Poisson process on the plane would have ground process elements failing to be members of  $\mathcal{N}_\mathbb{R}^\#$ . See also Exercises 9.1.3 and 9.1.6.

Note also that although a purely atomic boundedly finite measure can have at most countably many atoms, these atoms may have accumulation points, so representing such measures as a countable set  $\{(x_i, \kappa_i)\}$  of pairs of locations and sizes of the atoms can give a counting measure on  $\mathcal{X} \times \mathbb{R}_+$  that need not be in either  $\mathcal{N}_{\mathcal{X} \times \mathbb{R}_+}^{\#g}$  nor even  $\mathcal{N}_{\mathcal{X} \times \mathbb{R}_+}^\#$  [cf. Proposition 9.1.III(v) below].

In investigating the closure properties of  $\mathcal{M}_\mathcal{X}^\#$  and  $\mathcal{N}_\mathcal{X}^\#$  (Lemma 9.1.V below), we use *Dirac measures* (see Section A1.6) defined for every  $x \in \mathcal{X}$  by

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in \text{Borel set } A, \\ 0 & \text{otherwise.} \end{cases} \quad (9.1.3)$$

**Proposition 9.1.III.** *Let  $\mathcal{X}$  be a c.s.m.s., and  $\mu$  a boundedly finite measure on  $\mathcal{B}_{\mathcal{X}}$  (i.e.,  $\mu \in \mathcal{M}_{\mathcal{X}}^{\#}$ ).*

- (i) *The measure  $\mu$  is uniquely decomposable as*

$$\mu = \mu_a + \mu_d, \quad (9.1.4)$$

where

$$\mu_a = \sum_i \kappa_i \delta_{x_i} \quad (9.1.5)$$

is a purely atomic measure, expressed in terms of the uniquely determined countable set  $\{(x_i, \kappa_i)\} \subset \mathcal{X} \times \mathbb{R}_0^+$ , and  $\mu_d$  is a diffuse measure (i.e., it has no atoms).

- (ii) *A boundedly finite measure  $N$  on  $\mathcal{B}_{\mathcal{X}}$  is a counting measure (i.e.,  $N \in \mathcal{M}_{\mathcal{X}}^{\#}$  belongs to  $\mathcal{N}_{\mathcal{X}}^{\#}$ ), if and only if in part (i) its diffuse component is null, and in (9.1.5) all  $\kappa_i$  are positive integers,  $\kappa_i = k_i$  say, and  $\{x_i\}$  is a countable set with at most finitely many  $x_i$  in any bounded Borel set; that is,*

$$N = \sum_i k_i \delta_{x_i}. \quad (9.1.6)$$

- (iii) *For any  $N \in \mathcal{N}_{\mathcal{X}}^{\#}$ ,*

$$N^* = \sum_i \delta_{x_i} \quad (9.1.7)$$

defines the support counting measure  $N^*$ . Then  $N^* \in \mathcal{N}_{\mathcal{X}}^{\#*}$ ;  $N$  belongs to  $\mathcal{N}_{\mathcal{X}}^{\#*}$  if and only if at (9.1.6)  $k_i = 1$  (all  $i$ ); equivalently,  $N$  coincides with its support counting measure.

- (iv) *Any counting measure  $N \in \mathcal{N}_{\mathcal{X}}^{\#}$  may be represented as a counting measure  $\tilde{N} \in \mathcal{N}_{\mathcal{X} \times \mathbb{Z}_+}^{\#g}$  with representation  $\{(x_i, k_i)\}$ , in which the ground measure  $\tilde{N}_g$  is equal to the support counting measure  $N^*$  and the positive integer-valued marks  $k_i$  represent the multiplicities of the atoms of  $N$ .*
- (v) *There exists a one-to-one, both ways measurable, correspondence between purely atomic boundedly finite measures  $\mu$  on  $\mathcal{B}_{\mathcal{X}}$  and counting measures  $N_{\mu}(A \times K) = \sum_i \delta_{(x_i, \kappa_i)}(A \times K)$  on the Borel sets of  $\mathcal{X} \times \mathbb{R}_0^+$  satisfying the additional requirement that, for all bounded  $A \in \mathcal{B}_{\mathcal{X}}$ ,*

$$\int_{A \times \mathbb{R}_+} \kappa N_{\mu}(dx \times d\kappa) = \sum_{i: x_i \in A} \kappa_i < \infty; \quad (9.1.8)$$

the correspondence is given for bounded  $A \in \mathcal{B}_{\mathcal{X}}$  and any  $K \in \mathcal{B}(\mathbb{R}_0^+)$  that is bounded away from 0 by

$$\mu(A) = \int_{A \times \mathbb{R}_0^+} \kappa N_{\mu}(dx \times d\kappa) \quad (9.1.9a)$$

and

$$N_{\mu}(A \times K) = \lim_{n \rightarrow \infty} \sum_j I_K[\mu(A_{nj})], \quad (9.1.9b)$$

where  $\{\{A_{nj}\}\}$  is a dissecting system of measurable subsets of  $A$ .

*Remarks.* The measure  $N_\mu$  in (v) above, in terms of the representation as  $\sum_i \delta_{(x_i, \kappa_i)}$ , may have an accumulation point  $(x, 0) \in \mathcal{X} \times \mathbb{R}_+$  and therefore fail to be boundedly finite, even though for such a measure  $N_\mu$  we do have  $\sum_i \kappa_i < \infty$ . In this case, we have, for some bounded set  $A$ ,  $N_g(A) = \infty$  whereas for  $\epsilon > 0$ ,  $\#\{(x_i, \kappa_i): x_i \in A, \kappa_i > \epsilon\} < \infty$ . See Definition 9.1.VI(vi).

PROOF. Part (i) is a standard property of  $\sigma$ -finite measures on  $\mathcal{B}_\mathcal{X}$ : see the definition of atomic and diffuse measures in Appendix A1.6.

In part (ii), it is clear from (i) that if the diffuse component is null and the  $\kappa_i$  are positive integers, then the measure is a counting measure. Conversely, if  $N$  is integer-valued, any atom of  $N$  must have positive integral mass; and because  $N$  is boundedly finite there can be at most a finite number of such atoms within any bounded set, and at most countably many in all because we can cover  $\mathcal{X}$  by a countable number of bounded sets. Hence, to complete the proof, it is enough to show that  $N$  has no nonatomic component.

Let  $y$  be an arbitrary point of  $\mathcal{X}$ , and  $\{\epsilon_j: j = 1, 2, \dots\}$  a monotonic sequence of positive reals decreasing to zero, so that the spheres  $S_{\epsilon_j}(y) \downarrow \{y\}$  as  $j \rightarrow \infty$ . Then by the continuity lemma for measures (Proposition A1.3.II),

$$N\{y\} \equiv N(\{y\}) = \lim_{j \rightarrow \infty} N(S_{\epsilon_j}(y)).$$

Each term on the right-hand side is nonnegative integer-valued; the same therefore applies to  $N\{y\}$ . Thus, if  $y$  is not an atom of  $N$ , it must be the limit of a sequence of open spheres for which  $N(S_{\epsilon_j}(y)) = 0$ , hence, in particular, the centre of an open sphere with this property. This shows that the support of  $N$  (the complement of the largest open set with zero measure) consists exclusively of the atoms of  $N$ , or equivalently, that  $N$  is purely atomic. Equation (9.1.6) now follows from (9.1.5).

The properties of  $N^*$  in (iii) follow from the representations in (i) and (ii).

Part (iv) follows from Definition 9.1.II(iv) and part (ii) because  $\delta_{(x_i, \kappa_i)}$  can be identified with an atom in the product space.

For part (v), (9.1.9a) is a restatement of (9.1.5). The condition at (9.1.8) is a restatement of the requirement that the measure  $\mu$  be boundedly finite. The representation in (9.1.9b) mimics the construction using decreasing spheres to prove (ii) but with decreasing sets from a sequence of partitions from the dissecting system. Because  $K$  is bounded away from 0, there are at most a finite number of atoms with locations in  $A$  and values in  $K$ . As the sets in the dissecting system shrink, each of these atoms will ultimately be isolated in one of the subsets, leading to the representation (9.1.9b).  $\square$

Notice that this proof makes essential use of the topological structure of  $\mathcal{X}$ ; Moyal (1962) discusses some of the difficulties that arise in extending it to more general contexts.

Basic properties of  $\mathcal{M}_\mathcal{X}^\#$  are set out in Section A2.6 of Appendix 2, from which the key points for our purposes are set out below, together with their counterparts for  $\mathcal{N}_\mathcal{X}^\#$ .

**Proposition 9.1.IV.**

- (i) Under the  $w^\#$ -topology,  $\mathcal{M}_\mathcal{X}^\#$  is a c.s.m.s. in its own right.
- (ii) The corresponding Borel  $\sigma$ -algebra,  $\mathcal{B}(\mathcal{M}_\mathcal{X}^\#)$  say, is the smallest  $\sigma$ -algebra on  $\mathcal{M}_\mathcal{X}^\#$  with respect to which the mappings  $\mu \mapsto \mu(A)$  are measurable for all  $A \in \mathcal{B}_\mathcal{X}$ .
- (iii) Under the  $w^\#$ -topology,  $\mathcal{N}_\mathcal{X}^\#$  is a c.s.m.s. in its own right, and its Borel sets coincide with the Borel sets of  $\mathcal{N}_\mathcal{X}^\#$  as a subset of  $\mathcal{M}_\mathcal{X}^\#$ .
- (iv)  $\mathcal{B}(\mathcal{N}_\mathcal{X}^\#)$  is the smallest  $\sigma$ -algebra with respect to which the mappings  $N \mapsto N(A)$  are measurable for each  $A \in \mathcal{B}_\mathcal{X}$ .

Statements (i) and (ii) form Theorem A2.6.III. The extensions to counting measures follow from the next lemma.  $\square$

**Lemma 9.1.V.**  $\mathcal{N}_\mathcal{X}^\#$  is a closed subset of  $\mathcal{M}_\mathcal{X}^\#$ .

PROOF. Let  $\{N_k\}$  be a sequence of counting measures converging to some limit measure  $N$  in the  $w^\#$ -topology in  $\mathcal{M}_\mathcal{X}^\#$ . As in the proof of Proposition 9.1.III, let  $y$  be an arbitrary point of  $\mathcal{X}$ , and  $S_{\epsilon_j}(y)$  a sequence of spheres, contracting to  $\{y\}$ , with the additional property that

$$N(\partial S_{\epsilon_j}(y)) = 0 \quad (j = 1, 2, \dots)$$

[this is always possible because  $N(S_\epsilon(y))$ , as a function of  $\epsilon$ , has jumps for at most countably many values of  $\epsilon$ , and thus, the complementary set of values of  $\epsilon$ , being dense, the  $\epsilon_j$  can be chosen in the complementary set]. For each such sphere it follows from the properties of  $w^\#$ -convergence (Proposition A2.6.II) that

$$N_k(S_{\epsilon_j}(y)) \rightarrow N(S_{\epsilon_j}(y)).$$

Once again the terms on the left-hand side are all nonnegative integers, so the same is true for the term on the right-hand side. As in the previous proof, it then follows that  $N$  is purely atomic. This argument shows that  $\mathcal{N}_\mathcal{X}^\#$  is sequentially closed in  $\mathcal{M}_\mathcal{X}^\#$ , and hence closed because  $\mathcal{M}_\mathcal{X}^\#$  is separable (Theorem A2.6.III).  $\square$

Exercise 9.1.2 shows that the spaces in Definitions 9.1.II(iii)–(iv), although measurable subsets of  $\mathcal{M}_\mathcal{X}^\#$ , are not closed in either  $\mathcal{M}_\mathcal{X}^\#$  or  $\mathcal{N}_\mathcal{X}^\#$ .

Similarly, the space  $\mathcal{M}_{\mathcal{X},a}^\#$  is not closed in the weak or weak $^\#$  topologies: the sequence of purely atomic measures with atoms of mass  $1/n$  at  $\{1/n, 2/n, \dots, 1\}$  converges weakly to Lebesgue measure on  $(0, 1)$ .

Properties (ii) and (iv) of Proposition 9.1.IV open the way to defining random measures and point processes as measurable mappings involving the spaces of Definition 9.1.II, and lead to simple characterizations of random measures and point processes.

**Definition 9.1.VI.**

- (i) A random measure  $\xi$  with phase or state space  $\mathcal{X}$ , is a measurable mapping from a probability space  $(\Omega, \mathcal{E}, \mathcal{P})$  into  $(\mathcal{M}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#}))$ .
- (ii) A point process  $N$  on state space  $\mathcal{X}$  is a measurable mapping from a probability space  $(\Omega, \mathcal{E}, \mathcal{P})$  into  $(\mathcal{N}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#}))$ .
- (iii) A point process  $N$  is simple when

$$\mathcal{P}\{N \in \mathcal{N}_{\mathcal{X}}^{\#*}\} = 1. \quad (9.1.10)$$

- (iv) A marked point process on  $\mathcal{X}$  with marks in  $\mathcal{K}$  is a point process  $N$  on  $\mathcal{B}_{\mathcal{X} \times \mathcal{K}}$  for which

$$\mathcal{P}\{N \in \mathcal{N}_{\mathcal{X} \times \mathcal{K}}^{\#g}\} = 1; \quad (9.1.11)$$

its ground process is given by  $N_g(\cdot) \equiv N(\cdot \times \mathcal{K})$ .

- (v) A purely atomic random measure  $\xi$  is a measurable mapping from a probability space  $(\Omega, \mathcal{E}, \mathcal{P})$  into  $(\mathcal{M}_{\mathcal{X},a}^{\#}, \mathcal{B}(\mathcal{M}_{\mathcal{X},a}^{\#}))$ .
- (vi) An extended MPP with positive marks is a point process on  $\mathcal{B}_{\mathcal{X} \times \mathbb{R}_+}$  which is finite-valued on all sets of the form  $A \times K$  for bounded  $A \in \mathcal{B}_{\mathcal{X}}$  and Borel sets  $K \subset (\epsilon, 1/\epsilon)$  for some  $\epsilon > 0$ .

The notation of Definition 9.1.VI(i) is intended to imply that with every sample point  $\omega \in \Omega$ , we associate a particular *realization* that is a boundedly finite Borel measure on  $\mathcal{X}$ ; we denote it by  $\xi(\cdot, \omega)$  or just  $\xi(\cdot)$  (or even  $\xi$ ) when we have no need to draw attention to the underlying spaces. Similar statements can be made for counting measures and point processes  $N(\cdot, \omega)$ ,  $N(\cdot)$ , and so on.

A consequence of Definitions 9.1.VI(i)–(ii) above and Definition 9.1.II(ii) for  $\mathcal{N}_{\mathcal{X}}^{\#}$ , is that a *random measure is a point process if and only if its realizations are a.s. integer-valued*.

Observe that by choosing the state space  $\mathcal{X}$  appropriately, Definitions 9.1.VI(i)–(ii) can be made to include not only a number of important special cases but also a number of apparent generalizations. In the case  $\mathcal{X} = \mathbb{R}$ , discussion of *one-dimensional random measures* is essentially equivalent, as we note in Example 9.1(c), to the discussion of processes with nonnegative increments. The cases  $\mathcal{X} = \mathbb{R}^d$ ,  $d \geq 2$ , correspond to *multidimensional random measures*. If  $\mathcal{X}$  has the product form  $\mathcal{Y} \times \mathcal{K}$ , where  $\mathcal{K}$  is a finite set,  $\{1, \dots, d\}$  say, and we define distance in  $\mathcal{Y} \times \mathcal{K}$  by (for example)

$$d((x, i), (y, j)) = \rho(x, y) + |i - j|,$$

the resulting process is a *multivariate random measure*; each of its  $d$  components is a random measure on  $\mathcal{Y}$ . This itself is a special case of a point process defined on a product space; when both components are metric spaces, any one of a number of combinations of the two individual metrics—the additive form above is one convenient choice—will make the product space into a metric space and so allow the basic machinery to be applied. The assumption that



such a choice can and has been made underlies the introduction of MPPs in Section 6.4 and Definition 9.1.VI(iv) which, as already noted, is equivalent when coupled with Definition 9.1.II(iv) to the requirement that for the ground process,  $N_g(A) \equiv N(A \times \mathcal{K}) < \infty$  a.s. for all bounded  $A \in \mathcal{B}_{\mathcal{X}}$ .

Exercise 9.1.4 indicates that for a marked point process  $N$  and any Borel set  $K \in \mathcal{B}_{\mathcal{K}}$ , the ‘ $K$ -marginal’ process  $N_K(\cdot) = N(\cdot \times K)$  is a well-defined simple point process. That  $N_g$  is boundedly finite and simple follows from Definition 9.1.II(iv).

Exercise 9.1.5 sets out in greater detail the extension described in Definition 9.1.VI(vi). The motivation behind this definition is the construction of the counting measure  $N_\mu$  in Proposition 9.1.III(v) for a purely atomic measure  $\mu$ . Using this definition and the measurability assertion in Proposition 9.1.III(v) leads to the following equivalence for purely atomic random measures.

**Lemma 9.1.VII.** *Equations (9.1.9a,b) establish a one-to-one correspondence between purely atomic random measures  $\xi$  and extended MPPs with positive marks,  $N_\xi$  say, satisfying the condition (9.1.8).*

A realization of a random measure  $\xi$  has the value  $\xi(A, \omega)$  [or we may write just  $\xi(A)$ ] on the Borel set  $A \in \mathcal{B}_{\mathcal{X}}$  [and, similarly,  $N(A)$  for a point process  $N$ ]. For each fixed  $A$ ,  $\xi_A \equiv \xi(A, \cdot)$  is a function mapping  $\Omega$  into  $\mathbb{R}_+$ , and thus it is a candidate for a nonnegative random variable; that it is indeed such is shown in the following proposition.

**Proposition 9.1.VIII.** *Let  $\xi$  (respectively,  $N$ ) be a mapping from a probability space into  $\mathcal{M}_{\mathcal{X}}^\#$  ( $\mathcal{N}_{\mathcal{X}}^\#$ ) and  $\mathcal{A}$  a semiring of bounded Borel sets generating  $\mathcal{B}_{\mathcal{X}}$ . Then  $\xi$  is a random measure ( $N$  is a point process) if and only if  $\xi_A$  ( $N(A)$ ) is a random variable for each  $A \in \mathcal{A}$ .*

PROOF. Let  $\mathcal{U}$  be the  $\sigma$ -algebra of subsets of  $\mathcal{M}_{\mathcal{X}}^\#$  whose inverse images under  $\xi$  are events, and let  $\Phi_A$  denote the mapping taking a measure  $\mu \in \mathcal{M}_{\mathcal{X}}^\#$  into  $\mu(A)$  [hence, in particular,  $\Phi_A: \xi(\cdot, \omega) \mapsto \xi(A, \omega)$ ]. Because  $\xi_A(\omega) = \xi(A, \omega) = \Phi_A(\xi(\cdot, \omega))$  as in Figure 9.1, we have for any  $B \in \mathcal{B}_{\mathbb{R}_+}$

$$\xi^{-1}(\Phi_A^{-1}(B)) = (\xi_A)^{-1}(B).$$

When  $\xi_A$  is a random variable,  $(\xi_A)^{-1}(B) \in \mathcal{E}$ , and then by definition we have  $\Phi_A^{-1}(B) \in \mathcal{U}$ . It now follows from Theorem A2.6.III that  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^\#) \subseteq \mathcal{U}$  and hence that  $\xi$  is a random measure.

Conversely, by definition of  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^\#)$ ,  $\Phi_A^{-1}(B) \in \mathcal{B}(\mathcal{M}_{\mathcal{X}}^\#)$ , and when  $\xi$  is a random measure,  $\xi^{-1}(\Phi_A^{-1}(B)) \in \mathcal{E}$ , so then  $\xi_A$  is a random variable.  $\square$

Taking for  $\mathcal{A}$  the semiring of all bounded sets in  $\mathcal{B}_{\mathcal{X}}$  we obtain the following corollary.

**Corollary 9.1.IX.**  *$\xi: \Omega \mapsto \mathcal{M}_{\mathcal{X}}^\#$  is a random measure (respectively,  $N: \Omega \mapsto \mathcal{N}_{\mathcal{X}}^\#$  is a point process) if and only if  $\xi(A)$  ( $N(A)$ ) is a random variable for each bounded  $A \in \mathcal{B}_{\mathcal{X}}$ .*

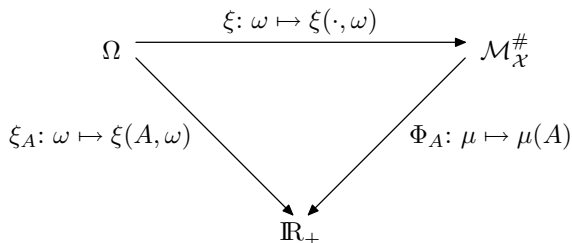


Figure 9.1

One useful consequence of Proposition 9.1.VIII is that we may justifiably use  $\xi(A)$  to denote the random variable  $\xi_A$  as well as the value  $\xi(A, \omega)$  of the realization of the random measure  $\xi$ .

Definitions 9.1.VI on their own do not lend themselves easily to the construction of particular random measures or point processes: for this the most powerful tool is the existence Theorem 9.2.VII below. Nevertheless, using Proposition 9.1.VIII or its corollary, we can handle some simple special cases as below in Examples 9.1(a)–(e) and Exercises 9.1.7–8.

**EXAMPLE 9.1(a)** *Uniform random measure.* Let  $\mathcal{X}$  be the real line, or more generally any Euclidean space  $\mathbb{R}^d$ , and define

$$\xi(A) = \Theta \ell(A),$$

where  $\ell(\cdot)$  denotes Lebesgue measure on  $\mathcal{X}$  and  $\Theta$  is a random multiplier that is nonnegative. To set this up formally, take  $\Omega$  to be a half-line  $[0, \infty)$ ,  $\mathcal{E}$  the Borel  $\sigma$ -algebra on  $\Omega$ , and  $\mathcal{P}$  any probability measure on  $\Omega$ , for example, the measure with gamma density  $x^\alpha e^{-x} / \Gamma(\alpha)$ . This serves as the distribution of  $\Theta$ . For each particular value of  $\Theta$ , the corresponding realization of the random measure is the  $\Theta$  multiple of Lebesgue measure. (Note that this process is random only in a rather artificial sense. Given only one realization of the process, we would have no means of knowing whether it is random. Randomness would only appear if we were to observe many realizations. In the language of Chapter 12, the process is stationary but not ergodic.)

We are left with one task, to verify that the mapping in  $(\mathcal{M}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#}))$  is indeed measurable. By Proposition 9.1.VIII, it is sufficient to verify that the mappings  $\xi(A)$  are random variables for each fixed  $A \in \mathcal{B}_{\mathcal{X}}$ . In our case,  $\xi(A)$  is a multiple of the random variable  $\Theta$ , so the verification is trivial. Thus, we have an example of a random measure.  $\square$

**EXAMPLE 9.1(b)** *Quadratic random measures: measures with  $\chi^2$  density* [see Example 6.1(c) and Exercise 6.1.3]. In Volume I we sketched a class of random measures constructed as follows. Take  $\mathcal{X} = \mathbb{R}$ , and choose any Gaussian process  $Z(t, \omega)$  with a.s. continuous trajectories. [Sufficient conditions for this can be expressed in terms of the covariance function: for example, it is enough for  $Z(\cdot)$  to be stationary with covariance function  $c(u)$  that is continuous at

$u = 0$ ; Cramér and Leadbetter (1967) give further results of this kind.] Then set

$$\xi(A) = \begin{cases} \int_A Z^2(t) dt & \text{for continuous } Z(\cdot, \omega), \\ 0 & \text{otherwise.} \end{cases}$$

Let us prove more formally that this construction defines a random measure. Because  $Z^2(t) \geq 0$ , it is clear that  $\xi(A) \geq 0$ , and countable additivity is a standard property of indefinite integrals. Moreover, because  $Z^2(\cdot)$  is a.s. continuous, it is bounded on bounded sets, and so  $\xi(\cdot)$  is boundedly finite. For almost all  $\omega$ , therefore,  $\xi(\cdot, \omega)$  is a boundedly finite Borel measure.

To complete the proof that  $\xi(\cdot)$  is a random measure we check that the condition of Proposition 9.1.VIII is met. Let  $A$  be any finite half-open interval (left-open right-closed for definiteness), and let

$$\mathcal{T}_n = \{A_{ni} : i = 1, \dots, k_n\}$$

be a sequence of partitions of  $A$  into subintervals with lengths  $1/n$  or less. If  $t_{ni}$  is a representative point from  $A_{ni}$ , it follows from standard properties of the Riemann integral that as  $n \rightarrow \infty$ ,

$$\xi_n(A) \equiv \sum_{i=1}^{k_n} Z^2(t_{ni}) \ell(A_{ni}) \rightarrow \int_A Z^2(t) dt = \xi(A) \quad \text{a.s.}$$

Each  $Z(t)$  is a random variable by assumption, and therefore so too is  $\xi_n(A)$  (as a linear combination of random variables) and  $\xi(A)$  (as the limit of a sequence of random variables). It is then clear that  $\xi(A)$  is a random variable for every set  $A$  in the semiring of finite unions of left-open right-closed intervals. It now follows from Proposition 9.1.VIII that  $\xi(\cdot)$  is a random measure.

The distributions of  $\xi(A)$  are nearly but not quite of gamma form: each particular value  $Z^2(t)$  has a gamma distribution and is proportional to a  $\chi^2$  random variable with one degree of freedom, so that the integral defining  $\xi(A)$  behaves as a linear combination of gamma variables. Its characteristic function can be obtained in the form of an infinite product of rational factors each associated with a characteristic root of the integral operator with kernel  $c^2(u - t)$  on  $A \times A$ .

Exercise 6.1.3 asserts that when  $Z(\cdot)$  is stationary, so too is  $\xi(\cdot)$ , and its moments have been given there [see also Example 9.5(a)], whereas Example 9.3(a) discusses sample-path properties.

Because a quadratic random measure has a gamma process as its density, it is a candidate for the directing measure of the class of Cox processes called *negative binomial processes* in Barndorff-Nielsen and Yeo (1969). More generally, sums of independent quadratic random measures have gamma process densities, and also therefore meet Barndorff-Nielsen and Yeo's definition, but, as they noted, although these point processes have computable moment properties, their distributional properties are not so readily accessible other than in degenerate cases (cf. Exercise 9.1.9). These negative binomial processes differ from those of Example 6.4(b), where the distributions are exactly of

negative binomial form, but, in contrast to this example, the earlier processes are either non-orderly or non-ergodic.  $\square$

EXAMPLE 9.1(c) *Processes with nonnegative increments.* Let  $\mathcal{X} = \mathbb{R}$ . It seems obvious that any stochastic process  $X(t)$ , defined for  $t \in \mathbb{R}$  and possessing a.s. finite-valued monotonic increasing trajectories, should define a random measure through the relation

$$\xi(a, b] = X(b) - X(a). \quad (9.1.12)$$

We show that this is the case at least when  $X(t)$  is also a.s. right-continuous. In any case, (9.1.12) certainly induces, for each realization, a finitely additive set function on the ring of finite unions of half-open intervals. Right-continuity enters as the condition required to secure countable additivity on this ring (compare the conditions in Proposition A2.2.VI and Corollary A2.2.VII). Then the set function defined by (9.1.12) can be extended a.s. to a boundedly finite measure, which we may continue to denote by  $\xi$  on  $\mathcal{B}_{\mathbb{R}}$ .  $\xi$  now represents a mapping from the probability space into  $\mathcal{M}_{\mathbb{R}}^{\#}$ . Because  $X(t)$  is a stochastic process,  $\xi(a, b]$  is a random variable for each half-open interval  $(a, b]$ . Proposition 9.1.VIII now implies that  $\xi$  is a random measure.  $\square$

The condition of right continuity can always be assumed when  $X(t)$  is *stochastically continuous*, that is, whenever for each  $\epsilon > 0$ ,

$$\Pr\{|X(t+h) - X(t)| > \epsilon\} \rightarrow 0 \quad \text{as } h \rightarrow 0, \quad (9.1.13)$$

because we may then define a new process by setting

$$X^*(t) = X(t+0)$$

and it is easy to verify that  $X^*(t)$  is a *version* or *copy* of  $X(t)$ , in the sense that it has the same fidi distributions. This condition is satisfied in particular by processes with stationary independent increments, giving rise to stationary random measures with the completely random property of Section 2.2. The next example illustrates the type of behaviour to be expected; Section 10.1 gives a more complete discussion of completely random measures.

EXAMPLE 9.1(d) *Gamma random measures—stationary case.* We indicated in Example 6.1(b) (see also Exercise 6.1.1) that a stationary random measure is defined by r.v.s  $\xi(A_i)$  that are mutually independent for disjoint Borel sets  $A_i$  in  $\mathbb{R}^d$  and have Laplace–Stieltjes transforms

$$\psi(A_i; s) = (1 + \lambda s)^{-\alpha \ell(A_i)} \quad (\lambda > 0, \alpha > 0, \operatorname{Re}(s) \geq 0);$$

these transforms show that the  $\xi(\cdot)$  are gamma distributed. The convergence  $\psi(A_i) \rightarrow 1$  as  $\ell(A_i) \rightarrow 0$  shows that the process is stochastically continuous, so we can assume right-continuity of the sample paths. The discussion around Proposition 9.1.VIII then implies that the resulting family of random variables can be extended to a random measure.

Despite the condition of stochastic continuity, the measures  $\xi(\cdot)$  here are not absolutely continuous, but on the contrary have a purely atomic character. This follows from the Lévy representation theorem which asserts that a process with independent increments can be represented as the sum of a shift, a Gaussian component, and an integral of Poisson components indexed according to the heights of the jumps with which they are associated [see e.g. Feller (1966, Section XVII.2), Bertoin (1996), or Theorem 10.1.III below]. The existence of a Gaussian component is ruled out by the monotonic character of the realizations, which also implies that the jumps are all positive. Thus, the random measure can be represented as a weighted superposition of Poisson processes, in the same kind of way as the compound Poisson process of Section 2.2. In the present case, however, the random measure has a countable rather than a finite number of atoms in any finite interval, but most atoms are so small that the total mass in such an interval is still a.s. finite. See also Example 9.1(g) and the discussion preceding it.  $\square$

EXAMPLE 9.1(e) *Random probability distributions and Dirichlet processes.* Random probability distributions play an important role in the theory of statistical inference, in particular, as prior distributions in nonparametric inference. Here we outline one method that has been proposed for constructing such distributions. Further constructions are in Exercises 9.1.10 and 9.3.4.

Suppose given a random measure  $\xi$  on the c.s.m.s.  $\mathcal{X}$ , with  $\xi$  a.s. totally finite and nonzero, and define

$$\zeta(A) = \xi(A)/\xi(\mathcal{X}) \quad (A \in \mathcal{B}_{\mathcal{X}}) \quad (9.1.14)$$

[in full,  $\zeta(A, \omega) = \xi(A, \omega)/\xi(\mathcal{X}, \omega)$ ]. Proposition 9.1.VIII shows that  $\zeta$  is a random measure; because  $\zeta(\mathcal{X}) = 1$  a.s., it is a random probability measure.

The *Dirichlet process*  $\mathcal{D}_{\alpha}$  is the random measure  $\zeta$  defined by the ratio at (9.1.14) when  $\xi$  is a gamma random measure as in the previous example and Exercise 6.1.1. Straightforward algebra shows that  $\zeta(A)$  has a beta distribution and, more generally, that the fidi distributions of  $\zeta(A_i)$  over disjoint sets  $A_i$ ,  $i = 1, \dots, r$ , are multivariate beta distributions. Exercise 9.5.1 gives moment measures of the process.

For a Dirichlet process on  $\mathcal{X} = \mathbb{R}$ , write  $F_{\zeta}(\cdot)$  for the random d.f. associated with the random probability distribution  $\zeta$ . Then the random variables  $Z_i = F_{\zeta}(x_i) - F_{\zeta}(x_{i-1})$ ,  $i = 1, \dots, r$ , where  $-\infty = x_0 < x_1 < \dots < x_{r-1} < x_r = \infty$  are such that, if each  $\alpha_i = \alpha((x_{i-1}, x_i]) > 0$ , their joint distribution is singular with respect to  $r$ -dimensional Lebesgue measure but absolutely continuous with respect to  $(r-1)$ -dimensional Lebesgue measure on the simplex  $\{(z_1, \dots, z_{r-1}): z_1 + \dots + z_{r-1} = 1 - z_r \leq 1\}$ , where it has the density function

$$f(z_1, \dots, z_{r-1} \mid \alpha_1, \dots, \alpha_r) = \Gamma\left(\sum_{i=1}^r \alpha_i\right) \prod_{i=1}^r \left(\frac{z_i^{\alpha_i-1}}{\Gamma(\alpha_i)}\right).$$

Ferguson (1973) supposes that an (unobserved) realization  $\zeta$  of  $\mathcal{D}_{\alpha}$  governs independent observations  $X_1, \dots, X_n$  for which the parameter  $\alpha$  specifies the prior distribution of  $\zeta$ . He shows that, conditional on  $(X_1, \dots, X_n) =$

$(x_1, \dots, x_n)$ , the posterior distribution of  $\zeta$  is again that of a Dirichlet process but has parameter  $\alpha + \sum_{i=1}^n \delta_{x_i}$ .

Our later discussion of completely random measures around (10.1.4) implies that both  $\zeta$  and  $\xi$  have realizations that are purely atomic, and hence that the possible d.f.s in such a random distribution are a.s. purely discrete. This is actually an advantage in the above discussion, as it is the feature that allows prior and posterior to have the same distributional form. See also Exercise 9.1.11, where the gamma random measure and the Dirichlet distribution are used to define a prior distribution for an inhomogeneous Poisson process.  $\square$

Concerning the existence of point processes, two basic approaches are widely used in the literature. A point process is defined sometimes as an integer-valued random measure  $N(\cdot, \omega)$ , as above, and sometimes as a sequence of random variables  $\{y_i\}$ . When are these approaches equivalent?

In one direction the argument is straightforward and covered in the next result where we start from certain finite or countably infinite sequences  $\{y_i: i = 1, 2, \dots\}$  of  $\mathcal{X}$ -valued random elements.

**Proposition 9.1.X.** *Let  $\{y_i\}$  be a sequence of  $\mathcal{X}$ -valued random elements defined on a probability space  $(\Omega, \mathcal{E}, \mathcal{P})$ , and suppose that there exists an event  $E_0 \in \mathcal{E}$  such that  $\mathcal{P}(E_0) = 0$  and  $\omega \notin E_0$  implies that for any bounded set  $A \in \mathcal{B}_{\mathcal{X}}$ , only a finite number of the elements of  $\{y_i(\omega)\}$  lie within  $A$ . Define  $N(\cdot)$  to be the zero measure on  $E_0$  and otherwise set*

$$N(A) = \#\{y_i \in A\} = \sum_i \delta_{y_i}(A) \quad (A \in \mathcal{B}_{\mathcal{X}}). \quad (9.1.15)$$

*Then  $N(\cdot)$  is a point process.*

PROOF. The set function  $N(A)$  is clearly a.s. integer-valued and finitely additive on Borel sets. Given a sequence  $\{y_i\}$ , choose  $y \notin \{y_i\}$  and let  $\{A_j\}$  be a sequence of bounded Borel sets decreasing to  $\{y\}$ . Because any  $A_j$  is bounded,  $N(A_j) < \infty$  a.s. Then for each  $y_i \in A_j$ , there exists finite  $j_i$  such that  $y_i \notin A_{j'}$  for  $j' > j_i$ . Therefore, for some finite  $j''$ ,  $N(A_k) = 0$  for all  $k > j''$ , and hence

$$N\left(\bigcap_{j=1}^{\infty} A_j\right) = 0 \quad \text{a.s.};$$

that is,  $N(A_j) \rightarrow 0$  a.s. Thus,  $N(\cdot)$  must be not just finitely but a.s. countably additive. This is enough to show that the sequence  $\{y_i\}$  induces a counting measure on  $\mathcal{X}$ , and so sets up a mapping from its probability space into  $\mathcal{N}_{\mathcal{X}}^{\#}$ .

To show that  $N(\cdot)$  is a point process, the critical step is to show that this mapping is measurable. From Proposition 9.1.VIII it is enough to show that for each Borel set  $A$ ,  $N(A)$  is a random variable. To this end, for each  $k = 1, 2, \dots$  and  $\omega \notin E_0$ , write

$$N_k(A, \omega) = \sum_{i=1}^k \delta_{y_i(\omega)}(A). \quad (9.1.16)$$

Because  $y_i$  is an  $\mathcal{X}$ -valued random variable, and  $A$  is a Borel subset of  $\mathcal{X}$ , each  $\delta_{y_i}(A)$  is a random variable for  $i = 1, \dots, k$  and therefore so too is  $N_k(A)$ . By monotonicity,  $N(A) = \lim_{k \rightarrow \infty} N_k(A)$  is well defined and therefore a random variable as required.  $\square$

The main problem with this approach to point processes arises from the need to express, in terms of the distributions of the  $y_i$ , the condition that with probability 1 the counting measures are boundedly finite. Suppose, for example, that the  $y_i$  are generated as the successive states in a Markov chain with state space  $\mathcal{X}$  and that the transition function for the chain satisfies an irreducibility condition sufficient to imply the usual classification into recurrent and transient chains. Then the local finiteness condition is satisfied if and only if the chain is transient. Exercises 9.1.12–13 illustrate this point.

We now turn to the more difficult question of constructing a sequence of  $\mathcal{X}$ -valued r.v.s from a given point process. Recall from Proposition 9.1.III that the realizations of the random counting measure  $N(\cdot)$  determine a.s. a countable set of atoms without any finite accumulation point, but this ignores the problem of finding a meaningful ordering of the atoms, without which the interpretation of the locations of the atoms as random variables is unresolved. The idea of fixing some enumeration of the points, in a measurable way, prompts the following definition.

**Definition 9.1.XI.** *Let  $N$  be a point process on a c.s.m.s.  $\mathcal{X}$  as in Definition 9.1.VI(ii). A measurable enumeration of  $N$  is a sequence of  $\mathcal{X}$ -valued r.v.s  $\{y_i(N) \equiv y_i(N(\cdot, \omega)) : i = 1, 2, \dots\}$  such that for every bounded Borel set  $A$ ,  $N(A, \omega) = \sum_{i=1}^{\infty} \delta_{y_i(N(\cdot, \omega))}(A)$  a.s.*

As a first illustration of how a sequence of random variables  $\{y_i\}$  can be extracted from the counting measure  $N(\cdot)$ , we formalize the discussion at the end of Section 3.1 concerning the relation between counting and interval properties of a simple point process  $N \in \mathcal{N}_{\mathbb{R}}^{\#*}$ . Retaining the notation of that section, recall the definitions

$$N(t) = \begin{cases} N((0, t]) & (t > 0), \\ 0 & (t = 0), \\ -N((t, 0]) & (t < 0), \end{cases} \quad (9.1.17)$$

and for  $i = 0, \pm 1, \dots$ ,

$$t_i(N) = \inf\{t : N(t) \geq i\}, \quad (9.1.18a)$$

$$\tau_i(N) = t_i(N) - t_{i-1}(N) \quad (9.1.18b)$$

(Figure 9.2 illustrates the relationship between  $\{t_i\}$  and  $\{\tau_i\}$ ). Let  $\mathcal{S}^+$  denote the space of all sequences  $\{\tau_0, \tau_{\pm 1}, \tau_{\pm 2}, \dots; x\}$  of positive numbers  $\tau_i$  satisfying  $0 \leq x < \tau_0$  and

$$\sum_{i=1}^{\infty} \tau_i = \sum_{i=1}^{\infty} \tau_{-i} = +\infty. \quad (9.1.19)$$

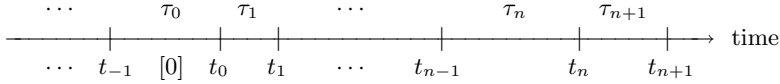


Figure 9.2

Intervals  $\tau_1, \dots, \tau_n, \dots$  between successive points  $t_0, t_1, \dots, t_{n-1}, t_n, \dots$ . When applicable, 0 satisfies  $t_{-1} < 0 \leq t_0$ , so that the interval of length  $\tau_0$  contains 0.

Then adding the relation  $x(N) = -t_0(N)$  to  $\tau_i$  at (9.1.18) defines a mapping  $R: \mathcal{N}_{\mathbb{R}}^{\#*} \rightarrow \mathcal{S}^+$ . The inverse mapping  $R^{-1}$  is defined for  $s^+ \in \mathcal{S}^+$  by

$$t_0(s^+) = -x_0(s^+), \quad t_i(s^+) = \begin{cases} t_{i-1}(s^+) + \tau_i & (i \geq 1), \\ t_{i+1}(s^+) - \tau_{i+1} & (i < 0). \end{cases} \quad (9.1.20)$$

We give  $\mathcal{S}^+$  the Borel  $\sigma$ -algebra  $\mathcal{B}(\mathcal{S}^+)$  obtained in the usual way as the product of  $\sigma$ -algebras on each copy of  $\mathbb{R}_+$ .

**Proposition 9.1.XII.** *The mapping at (9.1.18),  $R$  say, provides a one-to-one both ways measurable mapping of  $\mathcal{N}_{\mathbb{R}}^{\#*}$  into  $\mathcal{S}^+$ . In particular,*

- (i) *the quantities  $\tau_i(N)$  and  $x(N)$  are well-defined random variables when  $N$  is a simple point process; and*
- (ii) *there is a one-to-one correspondence between the probability distributions  $\mathcal{P}^*$  of simple point processes on  $\mathcal{N}_{\mathbb{R}}^{\#*}$  and probability distributions on the space  $\mathcal{S}^+$ .*

PROOF. The relations (9.1.18a) define the  $t_i(N)$  as stopping times for the increasing, right-continuous process  $N(\cdot)$  (see Definition A3.3.II and Lemma A3.3.III). Hence the  $t_i(N)$ , and also therefore the  $\tau_i(N)$ , are random variables whenever  $N$  is a simple point process.

To establish the converse, observe that the  $N$  can be defined in terms of the sequence of random variables  $t_i$ , noting that the requirement (9.1.19) implies that the resulting point process is boundedly finite. Then it follows from Proposition 9.1.X that  $N$  is a well-defined, simple point process.  $\square$

Note, in particular, the intervention of the initial interval  $(0, t_1]$ , the length  $x$  of which must be given separately: there is *not* a one-to-one correspondence between intervals and simple counting measures (Exercise 9.1.14 provides a further example of this type and a counterexample involving the subset  $\mathcal{N}_0 \subset \mathcal{N}_{\mathbb{R}}^{\#*}$  with an atom at 0 and the subset  $\mathcal{S}_0^+$  for which  $t_0 = x = 0$ ). Sigman's (1995) Appendix D discusses these questions also.

There is no analogous simple representation for point processes in  $\mathbb{R}^d$ ,  $d = 2, 3, \dots$ , although Exercise 9.1.15 sketches a possible construction based on distances of points from an origin (see also below Lemma 13.3.III). A general construction that exploits the separability property of the c.s.m.s.  $\mathcal{X}$  was suggested by Nguyen and Zessin (1976) and incorporated into Theorem 1.11.5 of MKM (1982). It is based on an ordered system of *tilings* of the c.s.m.s.  $\mathcal{X}$ , meaning a countable family  $\mathcal{T} = \{\mathcal{T}_n\}$  of 'infinite' dissecting systems of



$$\begin{array}{ccccccc}
& & & x & & y & \\
\mathcal{T}_n : & A_{n1} & \cdots & A_{n,i_n} & \cdots & A_{n,j_n} & \cdots \\
\mathcal{T}_{n+1} : & A_{n+1,1} & \cdots & A_{n+1,i_{n+1}} & \cdots & A_{n+1,j_{n+1}} & \cdots
\end{array}$$

Figure 9.3

Tilings  $\mathcal{T}_n = \{A_{ni}\}$ ,  $\mathcal{T}_{n+1} = \{A_{n+1,i}\}$  with sets containing  $x, y$ .

$\mathcal{X}$ : each tiling  $\mathcal{T}_n = \{A_{ni} : i = 1, 2, \dots\}$  consists of countably many disjoint bounded sets  $A_{ni}$  satisfying

- (i) (Partition and tiling properties)  $A_{ni} \cap A_{nj} = \emptyset$  for  $i \neq j$ , and  $\bigcup_i A_{ni} = \mathcal{X}$ ;
- (ii) (Nesting property)  $A_{n-1,i} \cap A_{nj} = A_{nj}$  or  $\emptyset$ , and any  $A_{n-1,i} \in \mathcal{T}_{n-1}$  is expressible  $A_{n-1,i} = \bigcup_{j \in \mathcal{J}_{n-1,i}} A_{nj}$  for some finite set of indices  $\mathcal{J}_{n-1,i}$ ;
- (iii) (Point-separating property) given distinct  $x, y \in \mathcal{X}$ , there exists an integer  $n(x, y)$  such that for  $n \geq n(x, y)$ ,  $x \in A_{ni}$  implies  $y \notin A_{ni}$ ; and
- (iv) (Enumeration consistency property) given distinct  $x, y \in \mathcal{X}$ , when  $\mathcal{T}_m$  is such that  $x \in A_{mi} \not\supset y$ , for all  $n \geq m$  sets  $A_{n,i_n} \ni x$ ,  $A_{n+1,i_{n+1}} \ni x$  and  $A_{n,j_n} \ni y$ ,  $A_{n+1,j_{n+1}} \ni y$  are such that  $j_n - i_n$  and  $j_{n+1} - i_{n+1}$  have the same sign.

The first three properties above are analogues of the properties of a dissecting system (Definition A1.6.I). We know from Proposition A2.1.IV that on any bounded Borel set in  $\mathcal{B}_{\mathcal{X}}$  there exists a dissecting system. We also know that a c.s.m.s. is covered by the union of a countable increasing family of boundedly finite sets  $S_i$  say; then  $\{A_{1i}\} = \{S_{i+1} \setminus S_i\}$  is a tiling. Now introduce dissecting systems on each  $A_{1i}$ , and enumerate their members in one sequence  $\{A_{2i}\}$  so as to satisfy property (iv); then it is a tiling of  $\mathcal{X}$ .

Given an integer-valued measure  $N$  on  $\mathcal{B}_{\mathcal{X}}$ ,  $N(A_{1i})$  is a finite integer for each  $i$ , and for each  $i = 1, 2, \dots$  we can enumerate the atomic support of  $N$  within those  $A_{1i}$  for which  $N(A_{1i}) \geq 1$  via the tilings. Moreover, if  $x, y \in \mathcal{X}$  are such that  $N(\{x\}) \geq 1$  and  $N(\{y\}) \geq 1$ , then  $x$  and  $y$  will be enumerated (with appropriate multiplicity if either inequality is strict) in a finite number of operations starting from some  $S_i$  that contains both  $x$  and  $y$ . The determination of such  $x$  or  $y$  proceeds via a sequence of nested sets  $A_{n,i_n}$  say for which  $N(A_{n,i_n}) \geq N(A_{n+1,i_{n+1}}) \geq 1$  for all  $n$  with  $\bigcap_{n=1}^{\infty} A_{n,i_n} = \{x\}$  say.

It follows that an enumeration of the points of  $N$  is thereby determined, for a sequence  $y_1, y_2, \dots$ , within a finite number of steps for each  $y_r$  even before its precise location is known.

Now write  $y_r = \lim \bigcap_j A_{j,i_{j,r}}$  for some monotonic decreasing sequence of sets  $A_{j,i_{j,r}}$  for which  $N(A_{j,i_{j,r}}) = 1$  for all sufficiently large  $j$ . For any given finite enumeration of points, the position in the enumeration is found from a finite number of elements of dissecting systems, with all the associated counting measures  $N(\cdot)$  measurable. The limit is therefore measurable; that is, the enumeration is measurable as required.

This outline argument leads to the following assertion.

**Lemma 9.1.XIII.** *Given a point process  $N$  on a c.s.m.s.  $\mathcal{X}$  as in Definition 9.1.VI(ii), there exists a measurable enumeration of  $\mathcal{X}$ -valued random*

elements  $\{y_i\}$  satisfying (9.1.16). This enumeration is uniquely determined by a given family of bounded sets  $\{S_i\}$  with  $S_i \uparrow \mathcal{X}$  and of dissecting systems for each  $S_i \setminus S_{i-1}$ .

A random measure may be regarded as a family of random variables indexed by the Borel sets of  $\mathcal{X}$ , but it is considerably more than this. The additivity and continuity properties of measures require at least the truth of

$$\xi(A \cup B) = \xi(A) + \xi(B) \quad \text{a.s.} \quad (9.1.21)$$

for all pairs of disjoint Borel sets  $A, B$  in  $\mathcal{X}$ , and

$$\xi(A_n) \rightarrow 0 \quad \text{a.s.} \quad (9.1.22)$$

for all sequences of bounded Borel sets  $A_n$  such that  $A_n \downarrow \emptyset$ . It is not quite trivial to prove, but fundamental for the resulting theory, that these conditions are in fact sufficient for the family to form a random measure. The difficulty is associated with the exceptional sets of measure zero: because there is an uncountable family of relations (9.1.21) and (9.1.22), it is not clear that the exceptional sets can be combined to form a single set that is still of measure zero. The next lemma indicates one way around the difficulty.

**Lemma 9.1.XIV.** *Let  $\mathcal{A}$  be a countable ring of bounded Borel sets with the self-approximating property of Definition A2.2.VIII, and  $\xi_A(\omega)$  a family of nonnegative random variables indexed by the sets  $A \in \mathcal{A}$ . In order that, with probability 1, the  $\xi_A(\omega)$  should admit an extension to a measure on  $\sigma(\mathcal{A})$ , it is necessary and sufficient that (9.1.21) hold for all disjoint pairs  $(A, B)$  of sets in  $\mathcal{A}$  and that (9.1.22) hold for all sequences  $\{A_n\}$  of sets in  $\mathcal{A}$  with  $A_n \downarrow \emptyset$ .*

PROOF. The number of sets in  $\mathcal{A}$  being countable, it follows immediately that (9.1.21) implies that the  $\xi_A(\omega)$  are a.s. finitely additive there. To establish countable additivity we use the covering property of Lemma A2.2.IX, from which it follows that it is enough to know that

$$\lim_{n \rightarrow \infty} \xi\left(\bigcup_{i=1}^n F_i(A; 1/k)\right) = \xi(A) \quad (9.1.23)$$

simultaneously for all sets  $A \in \mathcal{A}$  and integers  $k < \infty$ . Because each such relation holds a.s. from (9.1.22), and because the number of sets  $A \in \mathcal{A}$  and integers  $k < \infty$  is countable, this requirement is satisfied almost surely. Then Lemma A2.2.IX implies that the  $\xi_A$  can be a.s. extended to a measure on  $\sigma(\mathcal{A})$ .

The necessity of both conditions follows directly from the additivity and continuity properties of a measure.  $\square$

As an immediate corollary we obtain the following theorem, in which the point process analogues of (9.1.21) and (9.1.22), for  $A, B$ , and  $\{A_n\}$  as above, are

$$N(A \cup B) = N(A) + N(B) \quad \text{a.s.,} \quad (9.1.24a)$$

$$N(A_n) \rightarrow 0 \quad \text{a.s.} \quad (9.1.24b)$$

**Theorem 9.1.XV.** *Let  $\{\xi_A(\omega)\}$  [respectively,  $N_A(\omega)$ ] be a family of non-negative random variables indexed by the sets of  $\mathcal{B}_{\mathcal{X}}$  and a.s. finite-valued (finite integer-valued) on bounded Borel sets. In order that there exist a random measure  $\xi^*(A, \omega)$  (point process  $N$ ) such that, for all  $A \in \mathcal{B}_{\mathcal{X}}$ ,*

$$\xi^*(A, \omega) = \xi_A(\omega) \quad \text{a.s.}, \quad [N(A) = N_A \quad \text{a.s.}], \quad (9.1.25)$$

*it is necessary and sufficient that (9.1.21) [(9.1.24a)] hold for all pairs  $A, B$  of disjoint bounded Borel sets and that (9.1.22) [(9.1.24b)] hold for all sequences  $\{A_n\}$  of bounded Borel sets with  $A_n \downarrow \emptyset$ .*

PROOF. Let  $\mathcal{A}$  be any countable generating ring of bounded Borel sets with the self-approximating property of Definition A2.2.VIII, as, for example, the ring  $\mathcal{C}$  following Lemma A2.2.IX. If (9.1.21) and (9.1.22) hold for Borel sets in general, they certainly hold for sets in  $\mathcal{A}$ . Thus, the conditions of Lemma 9.1.XIV are satisfied, and we can assert that with probability 1 the  $\xi_A(\omega)$ , initially defined for  $A \in \mathcal{A}$ , can be extended to measures  $\xi^*(A, \omega)$  defined for all  $A \in \sigma(\mathcal{A}) = \mathcal{B}_{\mathcal{X}}$ . For  $\omega$  in the  $\mathcal{P}$ -null set,  $U$  say, where the measures cannot be so extended, set  $\xi^*(A, \omega) = 0$ . Then  $\xi^*(A, \omega)$  is a random measure which coincides a.s. with the original random variables  $\xi(A, \omega)$  at least on  $\mathcal{A}$ .

It is not immediately obvious, nor indeed is it necessarily true, that the extensions  $\xi^*(A, \omega)$  coincide with the original random variables  $\xi_A(\omega)$  for  $A \notin \mathcal{A}$ , even outside the exceptional set  $U$  of probability zero where the extension may fail. The best we can do is to show that they are a.s. equal for each particular Borel set  $A$ . The exceptional sets may be different for different  $A$ , and we do not claim that they can be combined into a single exceptional set of measure zero.

Consider the class of sets on which  $\xi^*$  and  $\xi$  coincide a.s. This class includes  $\mathcal{A}$ , and from the relations (9.1.22) it is closed under monotone limits. By the monotone class theorem it therefore includes  $\sigma(\mathcal{A})$ , which by assumption is  $\mathcal{B}_{\mathcal{X}}$ . This proves (9.1.25), and hence also the sufficiency part of the theorem.

Necessity is an easy corollary of the additivity and continuity properties of a measure.

The arguments can be applied equally to the case that the  $\xi$  is a.s. a counting measure, leading to the analogous result for a point process.  $\square$

As a sample application of Theorem 9.1.XV, we outline an approach to the definition of what, loosely speaking, might be termed a conditional random measure. It can be used to provide alternative proofs for the existence of the doubly stochastic and cluster processes introduced in Chapter 6.

**Proposition 9.1.XVI.** *Let  $\xi$  be a random measure defined on the probability space  $(\Omega, \mathcal{E}, \mathcal{P})$  with some c.s.m.s.  $\mathcal{X}$  as state space, and let  $\mathcal{F}$  be a sub- $\sigma$ -algebra of  $\mathcal{E}$ . Then there exists a version of the conditional expectation  $\eta(A, \omega) = \mathbb{E}[\xi(A) \mid \mathcal{F}](\omega)$  such that*

- (i) *for each  $A \in \mathcal{B}_{\mathcal{X}}$ ,  $\eta(A, \cdot)$  is an  $\mathcal{F}$ -measurable r.v.; and*
- (ii)  *$\eta$  is a random measure with state space  $\mathcal{X}$ .*

PROOF. It is easy to see from standard properties of conditional expectations that the additivity and consistency relations (9.1.21) and (9.1.22) are both satisfied for the conditional expectations  $\eta(A, \omega)$ . Furthermore, we may take the probability space here to be  $(\Omega, \mathcal{F}, \mathcal{P}_{\mathcal{F}})$  rather than  $(\Omega, \mathcal{F}, \mathcal{P})$ , where  $\mathcal{P}_{\mathcal{F}}$  denotes the restriction of  $\mathcal{P}$  to sets of  $\mathcal{F}$ , because by definition the conditional expectations are all  $\mathcal{F}$ -measurable. It now follows directly from Theorem 9.1.XV that there exists an  $\mathcal{F}$ -measurable random measure  $\eta^*$  such that  $\eta^*(A) = \eta(A)$  a.s.  $\square$

An almost identical argument leads to the classical result on the existence of regular conditional distributions given a  $\sigma$ -algebra (see Exercise 9.1.16 for variants on this theme).

To conclude this section, we would again emphasize the essential role played by the assumptions in the definitions. For example, the truth of Proposition 9.1.VIII and Theorem 9.1.XV depends in an essential manner on the assumption of nonnegativity. Corresponding statements for random signed measures are false in general: this is shown by the next example which, superficially, would be regarded as a random measure.

EXAMPLE 9.1(f) *Wiener's homogeneous chaos*. For  $A \in \mathcal{B}_{\mathcal{X}}$ , let  $\xi(A)$  have a normal  $N(0, \mu(A))$  distribution, where  $\mu(A)$  is a fixed, boundedly finite, Borel measure on  $\mathcal{X}$ , and suppose that the  $\xi(A)$  are independent for disjoint sets. These two requirements immediately allow the joint distributions of finite families  $\xi(A_1), \dots, \xi(A_k)$  to be written down, and it is easy to check that these joint distributions satisfy the consistency requirements of the Kolmogorov theorem. Thus, there does exist a probability space  $\Omega$  on which the  $\xi(A)$  may be simultaneously defined as random variables.

Now consider the random variable

$$W = \xi(A_1 \cup A_2) - \xi(A_1) - \xi(A_2),$$

where  $A_1, A_2$  are disjoint bounded Borel sets. It is readily checked that  $E(W) = 0$  and

$$\text{var } W = \mu(A_1 \cup A_2) + \mu(A_1) + \mu(A_2) - 2\mu(A_1) - 2\mu(A_2) = 0,$$

so  $W = 0$  a.s. Next consider the sequence  $\{A_n\}$  of disjoint bounded Borel sets with  $A = \bigcup_{j=1}^{\infty} A_j$ , where  $A$  is also bounded, and set

$$W_n \equiv \xi\left(\bigcup_{j=1}^n A_j\right) = \sum_{j=1}^n \xi(A_j) \quad \text{a.s.},$$

where the last equality follows by induction from the previous result. Then

$$\text{var } W_n = \mu\left(\bigcup_{j=1}^n A_j\right) = \sum_{j=1}^n \mu(A_j),$$

and if  $W = \xi(\bigcup_{j=1}^{\infty} A_j)$ , we must have  $\text{var}(W_n - W) \rightarrow 0$ . This shows that

$$\sum_{j=1}^n \xi(A_j) \rightarrow \xi(A)$$

in quadratic mean, and because the  $\xi(A_j)$  are independent, the partial sums converge to  $\xi(A)$  almost surely as well [see, e.g., Moran (1968, Theorem 8.24)]. We have shown that the family  $\{\xi(A_j)\}$  satisfies both (9.1.21) and (9.1.22).

On the other hand it is not true that for almost all  $\omega$  the realizations are signed measures. To see this, let  $\{A_1, \dots, A_n\}$  be a finite partition of  $A$  and set

$$Y_n = \sum_{j=1}^n |\xi(A_j)|.$$

If the realizations  $\xi(\cdot)$  were signed measures, the  $Y_n$  would remain uniformly bounded a.s. over all possible partitions. But

$$E(Y_n) = \sum_{j=1}^n E(|\xi(A_j)|) = \left(\frac{2}{\pi}\right)^{1/2} \sum_{j=1}^n (\mu(A_j))^{1/2},$$

and

$$\sum_{j=1}^n (\mu(A_j))^{1/2} \geq \frac{\sum_{j=1}^n \mu(A_j)}{\max_{1 \leq j \leq n} (\mu(A_j))^{1/2}} = \frac{\mu(A)}{\max_{1 \leq j \leq n} (\mu(A_j))^{1/2}},$$

so  $E(Y_n)$  can be made arbitrarily large by choosing a partition for which  $\max_{1 \leq j \leq n} \mu(A_j)$  is sufficiently small. Because  $\text{var } Y_n \leq \mu(A)$  for every partition, an application of Chebyshev's inequality shows that for any given finite  $y$ , a partition can be found for which  $\Pr\{Y_n \geq y\}$  can be made arbitrarily close to 1. This is impossible if the  $Y_n$  are a.s. bounded.  $\square$

Other examples may fail to be a random measure or point process because they fail to satisfy the bounded finiteness condition, as occurs for example with the jump points of many Lévy processes and for certain point sets for which Mandelbrot (1982, p. 78) proposed the term *dust*. A dust is a point set with infinitely many points in some bounded set and which has topological dimension  $D_T = 0$  (Mandelbrot, 1982, pp. 15, 409–412). [Stoyan and Stoyan (1994, p. 4) describe a dust as an uncountable point set containing no piece of any curve; the uncountability assumption seems unnecessarily restrictive.] For example, the rationals on  $[0, 1]$  constitute an everywhere dense countable dust, whereas the Cantor set (or, *Cantor dust*) on  $[0, 1]$  [see, e.g., Halmos (1950, Exercise 15.5)] is an uncountable nowhere dense dust.

EXAMPLE 9.1(g) *Lévy dust*. Mandelbrot (1982, p. 240) includes the set of zeroes of Brownian motion  $B(\cdot)$  as an example of a *Lévy dust*. Here we use the term here to mean the class of dusts defined via *subordinators* of a Brownian motion process [and, by a *subordinator*  $\eta(\cdot)$  we mean a nonnegative

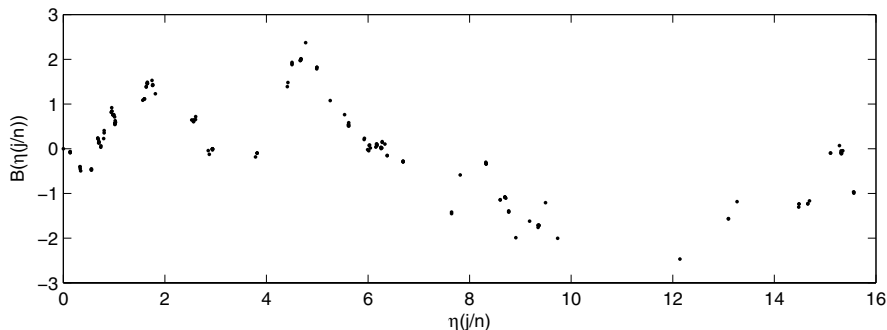


Figure 9.4

A space-time skeleton at  $\{t = j/n: j = 1, \dots, 20n\}$  of points of Lévy dust  $\{(x_i, B(x_i))\}$  for standard Brownian motion  $B(\cdot)$  at jump points  $\{t_i\}$  of gamma random measure subordinator  $\eta(\cdot)$ , with  $x_i = \eta(t_i-)$ . [ $n = 50$ , but  $\eta(j/n) - \eta([j-1]/n) < 10^{-7}$  for over half the skeletal points.]

Lévy process with zero drift coefficient as in, e.g., Bertoin (1996, Chapter III and p. 16)], so that a Lévy dust consists of the set of values of the process

$$y(t) = B[\eta(t)],$$

where  $B(\cdot)$  is a one- or two-dimensional Brownian motion. A subordinator  $\eta(\cdot)$ , being nondecreasing and a pure jump process with independent increments on  $\mathbb{R}_+$ , has a countably infinite number of jump points,  $\{t_i\}$  say, on any bounded interval of positive length. When a Markov process is used as a subordinand, the resultant process is again Markovian, and its range remains a countable set of points which, in the case of Brownian motion  $B(\cdot)$ , is just the countable set  $\{y_i\} = \{B(\eta(t_i))\}$ . Indeed, the process  $B(\eta(\cdot))$  is again a Lévy process (Bertoin's Exercise III.6.1).

Figure 9.4 depicts a space-time skeleton of a sample of these points by plotting  $\{(x_j, y_j)\} = \{(\eta(j/n), B(x_j))\}$ :  $j = 1, \dots, 20n\}$  in the case of a stationary gamma random measure  $\eta(t) = \xi(0, t]$  of Example 9.1(d) with  $\alpha = 1$ .

For Lévy dusts on  $\mathbb{R}^2$ , when  $B(t) = [Z_1(t), Z_2(t)]$  and  $Z_j(t)$  ( $j = 1, 2$ ) are standard independent one-dimensional Brownian motions, a useful approximation for simulation and illustrative purposes is to ignore the ultra-fine structure (infinitely many exceedingly small increments) and treat the process as a random walk  $\{X_n\}$  in  $\mathbb{R}^2$  whose steps  $Y_n = X_{n+1} - X_n$  have an isotropic distribution in  $\mathbb{R}^2$ . For example, when  $\eta(\cdot)$  is a nonnegative stable process, the step lengths follow approximately the Pareto form

$$\Pr\{|Y_n| > r \mid |Y_n| > \delta\} = (\delta/r)^\alpha \quad (r > \delta)$$

[see, e.g., Ogata and Katsura (1991) and the more extended discussion in Martínez and Saar (2002) of astrophysical applications, where the approximation is also known as Rayleigh–Lévy dust or Rayleigh–Lévy flights]. Two illustrations of such approximating point sets are shown in Figure 9.5.  $\square$

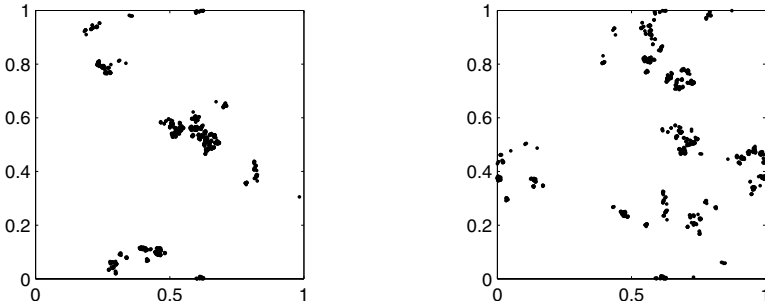


Figure 9.5

Two realizations of 1000 points of ‘Rayleigh–Lévy dust’, that is, a random walk with isotropic steps with d.f. tail  $(r_{\min}/r)^\alpha$  on  $r > r_{\min} = 0.001$ ,  $\alpha = 1.05$  and  $0.9$  (left- and right-hand figures respectively), locations reduced modulo 1 to the unit square.

### Exercises and Complements to Section 9.1

- 9.1.1 By definition, a sequence of totally finite measures  $\{\mu_k\}$  that converges weakly to a totally finite measure  $\mu$  converges in the  $w^\#$  sense to the same limit. Check that the converse need not be true by taking  $\mu_k$  to be Lebesgue measure on  $[0, k]$ . [Hint: As  $k \rightarrow \infty$ ,  $\mu_k$  does not converge weakly but does converge  $w^\#$  to Lebesgue measure on  $[0, \infty)$ .]
- 9.1.2 (a) The sequence of measures  $\{N_k\}$  on  $\mathcal{B}_{\mathbb{R}}$  defined by  $N_k = \delta_0 + \delta_{1/k}$  converges weakly to the measure  $N = 2\delta_0$ ; each  $N_k \in \mathcal{N}_{\mathbb{R}}^{\#*}$  but not the limit measure.
- (b) For  $k = 1, 2, \dots$ , let the measure  $\eta_k$  on  $\mathcal{X} \times \mathcal{K}$ , with  $\mathcal{X} = \mathbb{R}$  and  $\mathcal{K} = \mathbb{Z}_+$ , have unit atoms at all the points  $\{(i + j/2^k, 2^k) : i = 1, 2, \dots; j = 0, 1, \dots, 2^k - 1\}$  so  $\eta_k$  has boundedly finite support in  $\mathbb{R} \times \mathbb{Z}_+$ , and let  $N_r = \sum_{k=1}^r \eta_k$ . Show that each  $N_r$  is an element of  $\mathcal{N}_{\mathcal{X} \times \mathcal{K}}^{\#*}$  but that their limit is not.
- 9.1.3 Show that an MPP can be simple even if its ground process is not simple. [Hint: Simplicity of the MPP implies only that no single location has two identical marks.]
- 9.1.4 Let  $N$  be a simple point process on  $\mathcal{B}_{\mathcal{X} \times \mathcal{K}}$ , and  $K$  a fixed bounded Borel set in  $\mathcal{K}$ . Show that  $N_K(A) = N(A \times K)$  (bounded  $A \in \mathcal{B}_{\mathcal{X}}$ ) defines a simple point process. Deduce that the ground process  $N_g$  of Definition 9.1.V(iv) is well defined.
- 9.1.5 (a) Show that an extended MPP in the sense of Definition 9.1.6(vi) may fail to satisfy the requirements either of an MPP with marks in  $\mathbb{R}_+$  or of a point process on  $\mathbb{R} \times \mathbb{R}_+$ . [Hint: Consider as a counterexample the Poisson process arising in the Lévy representation of the gamma random measure of Example 9.1(d). The problem lies in satisfying the bounded finiteness properties as we have defined them.]
- (b) Show that the mapping  $(x_i, \kappa_i) \mapsto (x_i, \log \kappa_i)$  defines a one-to-one mapping between the realizations of an extended MPP with marks in  $\mathbb{R}_+$  and the

space  $\mathcal{N}_{\mathcal{X} \times \mathbb{R}}^\#$ . Use this mapping to define and explore the properties of a form of weak convergence for sequences of extended MPPs.

- 9.1.6 Show that, except for a set of  $\mathcal{P}$ -measure zero, a realization of a marked point process (Section 6.4) can be regarded as a simple point process on the product space  $\mathcal{X} \times \mathcal{K}^\cup$ , where  $\mathcal{K}^\cup = \mathcal{K}^{(1)} \cup \mathcal{K}^{(2)} \cup \dots$ , each  $\mathcal{K}^{(k)}$  consisting of all ordered sets of  $k$ -tuples of elements of the  $k$ -fold product set of  $\mathcal{K}$  with itself, and the measure on each  $A \times \mathcal{K}^{(k)}$  is symmetric in the subsets of  $\mathcal{K}^{(k)}$ . Conclude that any MPP is equivalent to another MPP whose ground process is simple.
- 9.1.7 Let  $\{X(t): t \in \mathbb{R}\}$  be a measurable nonnegative stochastic process on  $(\Omega, \mathcal{E}, \mathcal{P})$ . Show that, when the integrals concerned are finite, the relation

$$\xi(A, \omega) = \int_A X(t, \omega) dt \quad (\text{bounded } A \in \mathcal{B}_{\mathbb{R}})$$

defines a random measure  $\xi: \Omega \mapsto \mathcal{M}_{\mathbb{R}}^\#$ .

[Hint: Start by considering  $X(t, \omega)$  of the form  $\sum_j c_j I_{A_j}(t) I_{E_j}(\omega)$ .]

- 9.1.8 Let  $N$  be a well-defined point process on  $\mathcal{X} = \mathbb{R}^2$ . With each point  $y_i$  in a realization of  $N$  associate a geometric object in one of the following ways.

- (a) Construct a disk  $S_r(y_i)$  with centre  $y_i$  and radius  $r$ , and let

$$\xi(A) = \sum_i \ell(A \cap S_r(y_i)) \quad (\text{bounded } A \in \mathcal{B}(\mathbb{R}^2))$$

represent the total area of disks intersecting any Borel set  $A$ . Use Proposition 9.1.VIII to verify that  $\xi$  is a well-defined random measure on  $\mathbb{R}^2$ .

- (b) If the radius of each disk is also a random variable, leading to  $S_{R_i}(x_i)$  say, a conditioning argument as in Example 6.4(e), coupled with some condition ensuring the a.s. finiteness of the defining sum  $\sum \ell(A \cap S_{R_i}(x_i))$ , is needed.
- (c) Instead of disks, construct from  $y_i$  as endpoint, a finite line segment  $L_i$  of length  $d$  and random orientation  $\theta_i$  say, for some random variables  $\{\theta_i\}$  that are i.i.d. on  $(0, 2\pi]$ . For any bounded Borel set  $A \subset \mathbb{R}^2$  let  $\ell(A \cap L)$  now denote the Lebesgue measure (in  $\mathbb{R}^1$ ) of the intersect of a line  $L$  with  $A$ . Again use a conditioning argument to show that

$$\xi_L(A) \equiv \sum_i \ell(A \cap L_i) \quad (\text{bounded } A \in \mathcal{B}(\mathbb{R}^2))$$

is a well-defined random measure.

- 9.1.9 Let  $N(A)$  denote the number of points in  $A \in \mathcal{B}_{\mathcal{X}}$  of the negative binomial process of Example 9.1(b), involving a Cox process directed by the random measure  $\xi(A) = \int_A \eta(u) du$  for  $\eta(\cdot)$  a gamma process. Show that  $E(z^{N(A)}) = E(\exp[-(1-z)\xi(A)])$ . Derive a negative binomial approximation for suitably small sets  $A$ , and relate the first two moments of  $N(\cdot)$  to those of  $\xi(\cdot)$ .

9.1.10 *Random probability distributions.*

- (a) Let  $\xi$  be a boundedly finite but not totally finite measure on  $\mathbb{R}_+$ . Use the distribution function

$$F_\eta(x) \equiv 1 - \exp(-\xi[0, x])$$

to define a measure  $\eta$  on  $\mathbb{R}_+$ , with  $\eta(\mathbb{R}_+) = 1$ . Show that when  $\xi$  is a random measure,  $\eta$  is a random probability measure on  $\mathbb{R}_+$ .

- (b) When  $\xi$  is completely random (see Section 10.1),  $\eta$  is a ‘neutral process’ in the terminology of Doksum (1974). Show that when  $\xi$  has no deterministic component (see Theorem 10.1.III), the distribution  $\eta$  is a.s. purely atomic.



9.1.11 *Prior and posterior distributions for an inhomogeneous Poisson process.* Suppose it is desired to fit an inhomogeneous but totally finite Poisson process to one or more sets of observations over  $\mathcal{X}$ . Instead of assuming a specific parametric form for the intensity measure, suppose it is a gamma random measure  $\Lambda$  governed by a constant  $\lambda$  and some totally finite measure  $\alpha(\cdot)$  (see Exercise 6.1.1). Given a realization  $(x_1, \dots, x_n)$ , show that the posterior distribution for  $\Lambda$  is again a gamma random measure, governed by the constant  $\lambda + 1$  and the totally finite measure  $\alpha + \sum_{i=1}^n \delta_{x_i}$ . Equivalently, we may take  $\Lambda = C.F$  where the prior distribution for the constant  $C$  is  $\Gamma(\alpha(\mathcal{X}), \lambda)$ , and the prior distribution for the probability distribution  $F$  has the Dirichlet form  $D_\alpha$ .

9.1.12 Let  $\{X_n\}$  be a stationary ergodic real-valued Markov chain whose first absolute moment is finite, and define  $Y_n = X_1 + \dots + X_n$ . If  $EX_n \neq 0$ , then by the ergodic theorem  $\{Y_n\}$  obeys the strong law of large numbers and therefore satisfies the conditions of Proposition 9.1.X.

9.1.13 Let  $\{Y_n : n = 0, 1, \dots\}$  be a random walk in  $\mathbb{R}^d$ ; that is,  $Y_0 = 0$  and the  $\mathbb{R}^d$ -valued r.v.s  $X_n \equiv Y_n - Y_{n-1}$ ,  $n = 1, 2, \dots$  are i.i.d. Show that the conditions of Proposition 9.1.X are satisfied if either  $d \geq 3$  or else  $d = 1$  or  $2$  and  $E|X_n| < \infty$ ,  $EX_n \neq 0$ .

[Hint: Under the stated conditions, a random walk in  $\mathbb{R}^d$  is transient.]

Note that a *renewal process* is the special case  $d = 1$  and  $X_n \geq 0$  a.s. (and  $X_n \neq 0$  a.s.), so that for some d.f.  $F$  on  $\mathbb{R}_+$ , with  $0 = F(0-) \leq F(0+) < 1 = \lim_{x \rightarrow \infty} F(x)$ , and any positive integer  $r$ ,

$$\mathcal{P}(\{X_i \in (x_i, x_i + dx_i], i = 1, \dots, r\}) = \prod_{i=1}^r [F(x_i + dx_i) - F(x_i)].$$

9.1.14 Given a nonnull counting measure  $N \in \mathcal{N}_{\mathbb{R}}^{\#}$ , define  $\{Y_n : n = 0, \pm 1, \dots\}$  or a subset of this doubly infinite sequence by

$$\begin{aligned} N(0, Y_n) &< n \leq N(0, Y_n] & (n = 1, 2, \dots), \\ N(Y_n, 0] &< -n + 1 \leq N[Y_n, 0] & (n = 0, -1, \dots). \end{aligned}$$

Show that if  $N$  is a point process then  $\{Y_n : -N(-\infty, 0] + 1 \leq n \leq N(0, \infty)\}$  is a set of well-defined r.v.s.

Now let  $\mathcal{N}_0$  be the subspace of  $\mathcal{N}_{\mathbb{R}}^{\#*}$  consisting of simple counting measures on  $\mathbb{R}$  with a point at the origin, so  $N \in \mathcal{N}_0$  is boundedly finite, simple, and  $N\{0\} = 1$ . Show that if the atoms of such  $N$  yield the ordered set  $\{\dots, t_{-1}, t_0 = 0, t_1, \dots\}$  and  $\tau_i = t_i - t_{i-1}$ , then the mapping  $\Theta: \mathcal{N}_0 \mapsto \mathcal{S}_0^+$  which takes the counting measure  $N$  into the space  $\mathcal{S}_0^+$  of doubly infinite positive sequences  $\{\dots, \tau_{-1}, \tau_0, \tau_1, \dots\}$  is one-to-one and both ways measurable with respect to the usual  $\sigma$ -fields in  $\mathcal{N}_0$  and  $\mathcal{S}_0^+$ . Hence, probability measures on  $\mathcal{N}_0$  and  $\mathcal{S}_0^+$  are in one-to-one correspondence.

9.1.15 To establish a measurable enumeration of the points of a point process on  $\mathcal{X} \subseteq \mathbb{R}^d$ , first locate an initial point in the sequence as the point closest to some spatial origin (recall that the point process is on  $\mathcal{X} \subseteq \mathbb{R}^d$ ), with the proviso that in the event of there being two or more points equidistant

from 0 [i.e., for some sphere  $S_r(0)$  and integer  $k \geq 2$ ,  $N(S_r(0)) = k$  and  $N(S_{r-\epsilon}(0)) = 0$  for every  $\epsilon > 0$ ], these  $k$  points are ordered lexicographically in terms of a coordinate system for  $\mathcal{X}$ , yielding  $y_1, \dots, y_k$  say. The remaining points can be found on a sequence of progressively larger spheres centred on 0, using a similar tie-breaking rule as needed. Show that such a sequence of points is a sequence of  $\mathcal{X}$ -valued r.v.s as required in Definition 9.1.XI.

Compare this construction with the discussion of finite point processes summarized in Proposition 5.3.II.

- 9.1.16 (a) Mimic Theorem 9.1.XV to establish the existence of *regular conditional probabilities* on a product space  $\mathcal{X} \times \mathcal{Y}$ , where  $(\mathcal{X}, \mathcal{E})$  is an arbitrary measurable space and  $\mathcal{Y}$  is a c.s.m.s. (cf. Proposition A1.5.III). [Hint: Let  $\pi$  be a probability measure on the product space and  $\pi_{\mathcal{X}}$  the marginal distribution on  $(\mathcal{X}, \mathcal{E})$ . For fixed disjoint  $A, B \in \mathcal{B}_{\mathcal{Y}}$  show the existence of Radon–Nikodym derivatives  $Q(A | x)$ ,  $Q(B | x)$ ,  $Q(A \cup B | x)$  such that

$$Q(A \cup B | x) = Q(A | x) + Q(B | x) \quad (\pi_{\mathcal{X}}\text{-a.e. } x).$$

Now identify  $Q(A | x)$  with  $\xi_A(\omega)$  of the theorem and verify the continuity condition (9.1.22). For alternative approaches see Ash (1972, Section 6.6) and Feller (1966, Section V.10).]

- (b) Extend the above argument to the case of  $\mu$ , a boundedly finite measure on  $\mathcal{X} \times \mathcal{Y}$ , where  $\mathcal{X}, \mathcal{Y}$  are c.s.m.s.s and there exists a boundedly finite measure  $\lambda$  on  $\mathcal{B}_{\mathcal{X}}$  such that  $\mu(\cdot \times B)$  is absolutely continuous with respect to  $\lambda$  for bounded sets  $B \in \mathcal{B}_{\mathcal{Y}}$ ; that is, establish the existence of a family of measures  $\mu(\cdot | x)$  on  $\mathcal{B}_{\mathcal{Y}}$  for all  $x \in \mathcal{X}$  such that  $\mu(B | \cdot)$  is measurable for each bounded  $B \in \mathcal{B}_{\mathcal{X}}$ , and for bounded sets  $A \in \mathcal{B}_{\mathcal{X}}$ ,  $B \in \mathcal{B}_{\mathcal{Y}}$ ,

$$\mu(A \times B) = \int_A \mu(B | x) \lambda(dx).$$

[Hint: Normalize  $\lambda$  so that it is a probability measure on  $A$ .]

## 9.2. Finite-Dimensional Distributions and the Existence Theorem

Only statements about the distributions of a process are amenable, via frequency counts and the like, to direct comparison with observations. This is some justification for the view that the theory of random measures and point processes can be reduced to the study of the measures they induce on  $(\mathcal{M}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#}))$  and  $(\mathcal{N}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#}))$  respectively.

The deeper reason, however, is the unity and clarity that this point of view brings to questions concerning the existence of random measures and point processes. Using the characterization of such distributions through their finite-dimensional (fidi) distributions, as set out in Proposition 9.2.III below, we have an unequivocal answer to the problem of how to establish the existence of a particular class of point processes: can we write down for the class

a unique and consistent family of fidi distributions? This is the underlying reason for the importance of the studies by Moyal (1962) and Harris (1963), who were the first to set up a systematic theory of point processes in these terms. It also opens up the way for extensions to point processes on general types of spaces. On the other hand, the fidi distributions do not always provide the most convenient framework for examining the structure of particular models. For finite processes, the Janossy densities introduced in Chapter 5 are usually the most effective tool; likewise, for evolutionary processes, the conditional intensities introduced in Chapter 7 may prove extremely useful. But in all such cases, basic questions of existence can be referred back to the possibility of constructing a consistent family of fidi distributions.

**Definition 9.2.I.** *The distribution of a random measure or point process is the probability measure it induces on  $(\mathcal{M}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#}))$  or  $(\mathcal{N}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#}))$ , respectively.*

**Definition 9.2.II.** *The finite-dimensional distributions (fidi distributions for short) of a random measure  $\xi$  are the joint distributions, for all finite families of bounded Borel sets  $A_1, \dots, A_k$  of the random variables  $\xi(A_1), \dots, \xi(A_k)$ , that is, the family of proper distribution functions*

$$F_k(A_1, \dots, A_k; x_1, \dots, x_k) = \mathcal{P}\{\xi(A_i) \leq x_i \ (i = 1, \dots, k)\}. \quad (9.2.1)$$

Let us say that the distribution of a random measure is *completely determined* by some quantities  $\psi$  if, whenever two random measures give the same values for  $\psi$ , their distributions coincide. Analogously to Theorem A2.6.III and Proposition 9.1.VIII, we have the following result.

**Proposition 9.2.III.** *The distribution of a random measure is completely determined by the fidi distributions (9.2.1) for all finite families  $(A_1, \dots, A_k)$  of disjoint sets from a semiring  $\mathcal{A}$  of bounded sets generating  $\mathcal{B}_{\mathcal{X}}$ .*

PROOF. Let  $\mathcal{R}$  denote the ring generated by  $\mathcal{A}$ . Then any element  $A$  of  $\mathcal{R}$  can be represented as the finite union of disjoint sets from  $\mathcal{A}$ ,  $A = \bigcup_{i=1}^k A_i$  say, and thus, because

$$\xi(A) = \sum_{i=1}^k \xi(A_i), \quad (9.2.2)$$

the distribution of  $\xi(A)$  can be written down in terms of (9.2.1) for disjoint  $A_i$ . A similar result holds for the joint distributions of the  $\xi(A)$  for any finite family of sets  $A_i$  in  $\mathcal{R}$ .

Now consider the class of subsets of  $\mathcal{M}_{\mathcal{X}}^{\#}$  of the form of *cylinder sets*,

$$\{\xi: \xi(A_i) \in B_i \ (i = 1, \dots, k)\}, \quad (9.2.3)$$

where the  $A_i$  are chosen from  $\mathcal{R}$ , and the  $B_i$  are Borel sets of the real line  $\mathbb{R}$ . These cylinder sets form a ring, and it follows from Theorem A2.5.III

that this ring generates  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ . But the probabilities of all such sets can be determined from the joint distributions (9.2.1). Thus, the distribution of  $\xi$  is known on a ring generating  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$  and it follows from Proposition A1.3.I(b) that it is determined uniquely.  $\square$

In the terminology of Billingsley (1968, p. 15), Proposition 9.2.III asserts that finite families of disjoint sets from a semiring  $\mathcal{A}$  generating  $\mathcal{B}_{\mathcal{X}}$  form a *determining class* for random measures on  $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ . Of course we also have the following corollary.

**Corollary 9.2.IV.** *The distribution of a random measure is completely determined by its fidi distributions.*

For a point process, that is, an integer-valued random measure, it is simplest to specify the fidi distributions in the notation of (5.3.9), namely, for bounded Borel sets  $A_1, A_2, \dots$  and nonnegative integers  $n_1, n_2, \dots$ ,

$$P_k(A_1, \dots, A_k; n_1, \dots, n_k) = \mathcal{P}\{N(A_i) = n_i \ (i = 1, \dots, k)\}. \quad (9.2.4)$$

As in Proposition 9.1.VIII, the distribution of a point process, meaning the measure induced on  $(\mathcal{N}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#}))$ , is completely specified by the fidi distributions of  $N(A)$  for  $A$  in a countable ring generating the Borel sets.

We turn now to the main problem of this section, to find necessary and sufficient conditions on a set of fidi distributions (9.2.1) that will ensure that they are the fidi distributions of a random measure. The conditions fall into two groups: first the consistency requirements of the Kolmogorov existence theorem, and then the supplementary requirements of additivity and continuity needed to ensure that the realizations are measures.

**Conditions 9.2.V** (Kolmogorov Consistency Conditions).

(a) Invariance under index permutations. *For all integers  $k > 0$  and all permutations  $i_1, \dots, i_k$  of the integers  $1, \dots, k$ ,*

$$F_k(A_1, \dots, A_k; x_1, \dots, x_k) = F_k(A_{i_1}, \dots, A_{i_k}; x_{i_1}, \dots, x_{i_k}).$$

(b) Consistency of marginals. *For all  $k \geq 1$ ,*

$$F_{k+1}(A_1, \dots, A_k, A_{k+1}; x_1, \dots, x_k, \infty) = F_k(A_1, \dots, A_k; x_1, \dots, x_k).$$

The first of these conditions is a notational requirement: it reflects the fact that the quantity  $F_k(A_1, \dots, A_k; x_1, \dots, x_k)$  measures the probability of an event

$$\{\omega: \xi(A_i) \leq x_i \ (i = 1, \dots, k)\},$$

that is independent of the order in which the random variables are written down. The second embodies an essential requirement: it must be satisfied if there is to exist a single probability space  $\Omega$  on which the random variables can be jointly defined.

The other group of conditions captures in distribution function terms the conditions (9.1.21) and (9.1.22), which express the fact that the random variables so produced must fit together as measures.

**Conditions 9.2.VI** (Measure Requirements).

- (a) Additivity. For every pair  $A_1, A_2$  of disjoint Borel sets from  $\mathcal{B}_X$ , the distribution  $F_3(A_1, A_2, A_1 \cup A_2; x_1, x_2, x_3)$  is concentrated on the diagonal  $x_1 + x_2 = x_3$ .
- (b) Continuity. For every sequence  $\{A_n: n \geq 1\}$  of bounded Borel sets decreasing to  $\emptyset$ , and all  $\epsilon > 0$ ,

$$1 - F_1(A_n; \epsilon) \rightarrow 0 \quad (n \rightarrow \infty). \quad (9.2.5)$$

Conditions 9.2.V imply the existence of a probability space on which the random variables  $\xi(A)$ ,  $A \in \mathcal{B}_X$ , can be jointly defined. Then Condition 9.2.VI(a) implies

$$\mathcal{P}\{\xi(A_1) + \xi(A_2) = \xi(A_1 \cup A_2)\} = 1. \quad (9.2.6)$$

It follows by induction that a similar relation holds for the members of any finite family of Borel sets. For any given sequence of sets, Condition 9.2.VI(a) implies a.s. finite additivity, and then Condition 9.2.VI(b) allows this finite additivity to be extended to countable additivity.

This leads us to the existence theorem itself: it asserts that, in the case of nonnegative realizations, the Conditions 9.2.V and 9.2.VI are not only necessary but also sufficient to ensure that the fidi distributions can be associated with a random measure. Note that Example 9.1(f) implies that without nonnegativity, the sufficiency argument breaks down. It appears to be an open problem to find necessary and sufficient conditions on the fidi distributions that ensure that they belong to a random signed measure. See Exercise 9.2.4.

**Theorem 9.2.VII.** Let  $F_k(\cdot; \cdot)$  be a family of distributions satisfying the Consistency Conditions 9.2.V. In order that the  $F_k(\cdot)$  be the fidi distributions of a random measure, it is necessary and sufficient that

- (i) the distributions  $F_k(\cdot)$  be supported by the nonnegative half-line; and
- (ii) the  $F_k(\cdot)$  satisfy the Measure Conditions 9.2.VI.

PROOF. Necessity is clear from the necessity part of Theorem 9.1.XIV, so we proceed to sufficiency.

Because the  $F_k(\cdot)$  satisfy the Kolmogorov conditions, there exists a probability space  $(\Omega, \mathcal{E}, \mathcal{P})$  and a family of random variables  $\xi_A$  for bounded  $A \in \mathcal{B}_X$ , related to the given fidi distributions by (9.2.1). Condition (i) above implies  $\xi_A \geq 0$  a.s., and condition (ii) that the random variables  $\xi_A$  satisfy (9.1.21) for each fixed pair of bounded Borel sets. Now the random variables are a.s. monotonic decreasing, so (9.2.5) implies the truth of (9.1.22) for each fixed sequence of bounded Borel sets  $A_n$  with  $A_n \downarrow \emptyset$ . As in earlier discussions, the whole difficulty of the proof revolves around the fact that in general there is

an uncountable number of conditions to be checked, so that even though each individual condition is satisfied with probability 1, it cannot be concluded from this that the set on which they are simultaneously satisfied also has probability 1. To overcome this difficulty, we invoke Theorem 9.1.XIV. It is clear from the earlier discussion that both conditions of Theorem 9.1.XIV are satisfied, so that we can deduce the existence of a random measure  $\xi^*$  such that  $\xi^*(A)$  and  $\xi_A$  coincide a.s. for every Borel set  $A$ . But this implies that  $\xi^*$  and  $\xi$  have the same fidi distributions, and so completes the proof.  $\square$

**Corollary 9.2.VIII.** *There is a one-to-one correspondence between probability measures on  $\mathcal{B}(\mathcal{M}_X^\#)$  and families of fidi distributions satisfying Conditions 9.2.V and 9.2.VI.*

In practice, the fidi distributions are given most often for disjoint sets, so that Condition 9.2.VI(a) cannot be verified directly. In this situation it is important to know what conditions on the joint distributions of the  $\xi(A)$  for disjoint sets will allow such distributions to be extended to a family satisfying Condition 9.2.VI(a).

**Lemma 9.2.IX.** *Let  $F_k$  be the family of fidi distributions defined for finite families of disjoint Borel sets and satisfying for such families the Kolmogorov Conditions 9.2.V. In order for there to exist an extension (necessarily unique) to a full set of fidi distributions satisfying Conditions 9.2.VI(a) as well as 9.2.V, it is necessary and sufficient that for all integers  $k \geq 2$ , and finite families of disjoint Borel sets  $\{A_1, A_2, \dots, A_k\}$ ,*

$$\begin{aligned} \int_0^z F_k(A_1, A_2, A_3, \dots, A_k; dx_1, z - x_1, x_3, \dots, x_k) \\ = F_{k-1}(A_1 \cup A_2, A_3, \dots, A_k; z, x_3, \dots, x_k). \end{aligned} \quad (9.2.7)$$

PROOF. The condition (9.2.7) is clearly a corollary of Conditions 9.2.VI(a) and therefore necessary. We show that it is also sufficient. Let us first point out how the extension from disjoint to arbitrary families of sets can be made. Let  $\{B_1, \dots, B_n\}$  be any such arbitrary family. Then there exists a minimal family  $\{A_1, \dots, A_k\}$  of disjoint sets (formed from the nonempty intersections of the  $B_i$  and  $B_i^c$ ) such that each  $B_i$  can be represented as a finite union of some of the  $A_j$ . The joint distribution  $F_k(A_1, \dots, A_k; x_1, \dots, x_k)$  will be among those originally specified. Using this distribution, together with the representations of each  $\xi(B_i)$  as a sum of the corresponding  $\xi(A_j)$ , we can write down the joint distribution of any combination of the  $\xi(B_i)$  in terms of  $F_k$ . It is clear from the construction that the resultant joint distributions will satisfy Condition 9.2.VI(a) and that only this construction will satisfy this requirement.

To complete the proof it is necessary to check that the extended family of distributions continues to satisfy Condition 9.2.V(b). We establish this by induction on the index  $k$  of the minimal family of disjoint sets generating the given fidi distribution. Suppose first that there are just two sets  $A_1$ ,

$A_2$  in this family. The new distributions defined by our construction are  $F_2(A_1, A_1 \cup A_2)$ ,  $F_2(A_2, A_1 \cup A_2)$ , and  $F_3(A_1, A_2, A_1 \cup A_2)$ . Consistency with the original distributions  $F_2(A_1, A_2)$ ,  $F_1(A_1)$ , and  $F_1(A_2)$  is guaranteed by the construction and by the marginal consistency for distributions of disjoint sets. Only the marginal consistency with  $F_1(A_1 \cup A_2)$  introduces a new element. Noting that by construction we have

$$F_2(A_1, A_1 \cup A_2; x, y) = \int_0^{\min(x, y)} F_2(A_1, A_2; du, y - u),$$

and letting  $x \rightarrow \infty$ , we see that this requirement reduces precisely to (9.2.7) with  $k = 2$ . Similarly, for  $k > 2$ , marginal consistency reduces to checking points covered by the construction, by preceding steps in the induction, by Condition 9.2.V(b) for disjoint sets, or by (9.2.7).  $\square$

EXAMPLE 9.2(a) *Stationary gamma random measure* [see Example 9.1(d)]. The Laplace transform relation

$$\psi_1(A; s) \equiv \psi(A; s) = (1 + \lambda s)^{-\alpha \ell(A)}$$

determines the one-dimensional distributions, and the independent increments property on disjoint sets implies the relation

$$\psi_k(A_1, \dots, A_k; s_1, \dots, s_k) = \prod_{i=1}^k (1 + \lambda s_i)^{-\alpha \ell(A_i)},$$

which determines their joint distributions. Consistency of marginals here reduces to the requirement

$$\psi_{k-1}(A_1, \dots, A_{k-1}; s_1, \dots, s_{k-1}) = \psi_k(A_1, \dots, A_k; s_1, \dots, s_{k-1}, 0),$$

which is trivially satisfied. Also, if  $A_n \downarrow \emptyset$ ,  $\ell(A_n) \rightarrow 0$  by continuity of Lebesgue measure, and thus

$$\psi_1(A; s) = (1 + \lambda s)^{-\alpha \ell(A_n)} \rightarrow 1,$$

which is equivalent to Condition 9.2.VI(b). Finally, to check (9.2.6) we should verify that for disjoint  $A_1$  and  $A_2$ ,

$$\psi_1(A_1 \cup A_2; s) = \psi_2(A_1, A_2; s, s),$$

which is a simple consequence of additivity of Lebesgue measure.

These arguments establish the consistency conditions when the sets occurring in the fidi distributions are disjoint, and it follows from Lemma 9.2.IX that there is a unique consistent extension to arbitrary Borel sets.  $\square$

The basic existence theorem for point processes is somewhat simpler than Theorem 9.2.VII for general random measures, as we now indicate.

**Theorem 9.2.X** (Kolmogorov Existence Theorem for Point Processes).

*In order that a family  $P_k(A_1, \dots, A_k; n_1, \dots, n_k)$  of discrete fidi distributions defined on bounded Borel sets be the fidi distributions of a point process, it is necessary and sufficient that*

(i) for any permutation  $i_1, \dots, i_k$  of the indices  $1, \dots, k$ ,

$$P_k(A_1, \dots, A_k; n_1, \dots, n_k) = P_k(A_{i_1}, \dots, A_{i_k}; n_{i_1}, \dots, n_{i_k});$$

- (ii)  $\sum_{r=0}^{\infty} P_{k+1}(A_1, \dots, A_k, A_{k+1}; n_1, \dots, n_k, r) = P_k(A_1, \dots, A_k; n_1, \dots, n_k)$ ;  
 (iii) for each disjoint pair of bounded Borel sets  $A_1, A_2$ ,  $P_3(A_1, A_2, A_1 \cup A_2; n_1, n_2, n_3)$  has zero mass outside the set where  $n_1 + n_2 = n_3$ ; and  
 (iv) for sequences  $\{A_n\}$  of bounded Borel sets with  $A_n \downarrow \emptyset$ ,  $P_1(A_n; 0) \rightarrow 1$ .

The task of checking the conditions in detail here can be lightened by taking advantage of Lemma 9.2.IX, from which it follows that if the consistency conditions (i) and (ii) are satisfied for disjoint Borel sets, and if for such disjoint sets the equations

$$\begin{aligned} \sum_{r=0}^n P_k(A_1, A_2, A_3, \dots, A_k; r, n-r, n_3, \dots, n_k) \\ = P_{k-1}(A_1 \cup A_2, A_3, \dots, A_k; n, n_3, \dots, n_k) \end{aligned} \quad (9.2.8)$$

hold, then there is a unique consistent extension to a full set of fidi distributions satisfying (iii).

EXAMPLE 9.2(b) *The Poisson process with parameter measure  $\mu$*  [see Section 2.4]. Here the fidi distributions for disjoint Borel sets are readily specified by the generating function relations

$$\Pi_k(A_1, \dots, A_k; z_1, \dots, z_k) = \prod_{i=1}^k \exp[-\mu(A_i)(1 - z_i)], \quad (9.2.9)$$

where  $\Pi_k$  is the generating function associated with the distribution  $P_k$ . Condition (ii) is readily checked by setting  $Z_k = 1$ ; then the term  $1 - z_k$  vanishes and reduces the product to the appropriate form for  $\Pi_{k-1}$ . In generating function terms, equation (9.2.6) becomes, for  $k = 2$ ,

$$\Pi_2(A_1, A_2; z, z) = \Pi_1(A_1 \cup A_2; z),$$

which expresses the additivity of the Poisson distribution. Finally, to check condition (iv) we require  $\Pi_1(A_n; 0) \rightarrow 1$ , that is,

$$\exp[-\mu(A_n)] \rightarrow 1,$$

which is a corollary of the assumption that  $\mu$  is a measure, so  $\mu(A_n) \rightarrow 0$  as  $A_n \downarrow \emptyset$ .

It should be noted that the form (9.2.9) does not hold for arbitrary sets but has to be replaced by such forms as

$$\begin{aligned} \Pi_2(A_1, A_2; z_1, z_2) \\ = \exp[-\mu(A_1)(1 - z_1) - \mu(A_2)(1 - z_2) + \mu(A_1 \cap A_2)(1 - z_1)(1 - z_2)] \end{aligned}$$

when the sets overlap. The extension to arbitrary families of nondisjoint sets is unique, but laborious, and need not be pursued in detail.  $\square$



EXAMPLE 9.2(c) *Finite point processes.* If the distribution of a finite point process is specified in any of the ways described in Proposition 5.3.II, in particular, say, by its Janossy measures [see around (5.3.2)], then the fidi distributions are given by (5.3.13), namely

$$\begin{aligned} n_1! \dots n_k! P_k(A_1, \dots, A_k; n_1, \dots, n_k) \\ = \sum_{r=0}^{\infty} \frac{J_{n+r}(A_1^{(n_1)} \times \dots \times A_k^{(n_k)} \times C^{(r)})}{r!}, \end{aligned} \quad (9.2.10)$$

where  $C$  is the complement of the union of the disjoint sets  $A_1, \dots, A_k$  and  $n = n_1 + \dots + n_k$ .

Although we can infer on other grounds that the point process is well defined, and hence that the fidi distributions must be consistent, it is of interest to check the consistency conditions directly. Because (9.2.10) is restricted to disjoint sets, the appropriate conditions are (i), (ii), and (iv) of Theorem 9.2.X together with (9.2.8).

The permutation condition (i) follows from the symmetry of the Janossy measures. Also, condition (iv) reduces to

$$P_1(A_n; 0) = \sum_{r=0}^{\infty} \frac{J_r((\mathcal{X} \setminus A_n)^{(r)})}{r!} \rightarrow 1 \quad \text{if } A_n \downarrow \emptyset.$$

But then  $\mathcal{X} \setminus A_n \uparrow \mathcal{X}$ , and the result follows from dominated convergence, the fact that the  $J_r(\cdot)$  are themselves measures, and the normalization condition  $\sum_{r=0}^{\infty} [J_r(\mathcal{X}^{(r)})]/r! = 1$  as in (5.3.9).

The additivity requirement (9.2.8) follows from identities of the type

$$\begin{aligned} \sum_{n_1 + \dots + n_k = n} \frac{J_{n+r}(A_1^{(n_1)} \times \dots \times A_k^{(n_k)} \times C^{(r)})}{n_1! \dots n_k!} \\ = \frac{J_{n+r}((A_1 \cup \dots \cup A_k)^{(n)} \times C^{(r)})}{n!}, \end{aligned}$$

which are immediate applications of Lemma 5.3.III. Similarly, the marginal condition (ii) reduces to checking the equations

$$\begin{aligned} \sum_{n_k=0}^{\infty} \sum_{r=0}^{\infty} \frac{J_{\nu+n_k+r}(A_1^{(n_1)} \times \dots \times A_{k-1}^{(n_{k-1})} \times A_k^{(n_k)} \times C^{(r)})}{n_k! r!} \\ = \sum_{s=0}^{\infty} \frac{1}{s!} \sum_{t=0}^s J_{\nu+s}(A_1^{(n_1)} \times \dots \times A_{k-1}^{(n_{k-1})} \times A_k^{(t)} \times C^{(s-t)}) \\ = \sum_{s=0}^{\infty} \frac{J_{\nu+s}(A_1^{(n_1)} \times \dots \times A_{k-1}^{(n_{k-1})} \times (A_k \cup C)^{(s)})}{s!}, \end{aligned}$$

where  $\nu = n_1 + \dots + n_{k-1}$ , the first equation is a regrouping of terms, and the second equation is a further application of Lemma 5.3.III.  $\square$

Underlying Example 9.2(c) is a mapping of  $\mathcal{X}^\cup \mapsto \mathcal{N}_\mathcal{X}$ ; see also Propositions 9.1.XI–XII. The space  $\mathcal{X}^{\cup*} = \bigcup_{n=0}^\infty \mathcal{X}^{(n)*}$ , where  $\mathcal{X}^{(n)*}$  is the  $n$ -fold product of the c.s.m.s.  $\mathcal{X}$  subject to the constraint that any unordered set  $\mathbf{x} = \{x_1, \dots, x_n\} \in \mathcal{X}^{(n)*}$  satisfies  $x_i \neq x_j$  for  $i \neq j$  [cf. (5.3.10)], is a candidate space for describing finite simple point processes [cf. Chapter 5 and Definition 9.1.II(iii)].

Important questions relate to the characterization of subclasses of point processes through their fidi distributions. Characterizing simple point processes leads to the discussion of orderliness taken up in Section 9.3. Marked point processes can be treated as point processes on product spaces, with the defining sets for the fidi distributions restricted to ‘rectangle sets’  $A_i \times K_i$  where  $A_i \in \mathcal{B}_\mathcal{X}$ ,  $K_i \in \mathcal{B}_\mathcal{K}$ . The fidi distributions of a process on the product space correspond to those of a marked point process if the distributions of the ground process, obtained by setting all  $K_i = \mathcal{K}$  in the product sets, as in

$$P_k^g(A_1, \dots, A_k; n_1, \dots, n_k) = P_k(A_1 \times \mathcal{K}, \dots, A_k \times \mathcal{K}; n_1, \dots, n_k),$$

are proper, and satisfy the conditions of Proposition 9.2.X.

To conclude the present section, we outline an extension of Rényi’s (1967) result, quoted at Theorem 2.3.II, that a simple Poisson process in  $\mathbb{R}^d$ , whether homogeneous or not, is determined by the values of its *avoidance function*

$$P_0(A) = \mathcal{P}\{N(A) = 0\} \tag{9.2.11}$$

on a suitably rich class  $\mathcal{A}$  of Borel sets. The essence of this result is that, for a simple point process, the avoidance function alone is enough to determine the full set of fidi distributions. Our aim is to describe an interaction of structural properties of the space  $\mathcal{X}$  and the function  $P_0(\cdot)$  which are enough for  $P_0(\cdot)$  to retain this determining character without the strong probabilistic assumptions of the Poisson process.

Concerning terminology, Kendall (1974) used the term *avoidance function* in a more general (stochastic geometry) context, reflecting the fact that  $P_0(A)$  gives the probability of the support of a random set function avoiding a prescribed set  $A$ ; other possible terms include *zero function*, *avoidance probability function*, and *vacuity function* [McMillan (1953)]. Extensions of Rényi’s result are due to Mönch (1971), who showed that the Poisson assumption is not needed [see also Kallenberg (1973, 1975)], and a characterization of the avoidance function due to Kurtz (1974). Unpublished work of Karbe (1973) is presumably the basis of some discussion in MKM (1978, Section 1.4). Much of the work, largely couched in algebraic language, was developed by McMillan (1953) who used the term *vacuity function* in lectures in Berkeley in 1981.

If only the state space  $\mathcal{X}$  of the simple point process  $N(\cdot)$  were countable, Rényi’s result would be almost trivial, for with  $i, j, \dots$  denoting distinct points

of  $\mathcal{X}$ , we should have for the first few fidi distributions

$$\begin{aligned} P_1(\{i\}; 0) &= 1 - P_1(\{i\}; 1) = P_0(\{i\}), \\ P_2(\{i\}, \{j\}; 0, 0) &= P_0(\{i, j\}), \\ P_2(\{i\}, \{j\}; 0, 1) &= P_0(\{j\}) - P_0(\{i, j\}), \\ P_2(\{i\}, \{j\}; 1, 1) &= 1 - P_0(\{i\}) - P_0(\{j\}) + P_0(\{i, j\}). \end{aligned}$$

Continuing in this way, all the fidi distributions could be built up through a sequence of differencing operations applied to  $P_0(\cdot)$ , and it is clear that the avoidance function would thereby determine the fidi distributions uniquely. Our task here is to extend this argument to a general c.s.m.s.  $\mathcal{X}$  as state space.

Following Kurtz (1974), the equations

$$\Delta(A)\psi(B) = \psi(B) - \psi(A \cup B), \quad (9.2.12a)$$

$$\Delta(A_1, \dots, A_k, A_{k+1})\psi(B) = \Delta(A_{k+1})[\Delta(A_1, \dots, A_k)\psi(B)] \quad (k = 1, 2, \dots), \quad (9.2.12b)$$

define a difference operator  $\Delta(A)$  and its iterates acting on any set function  $\psi(\cdot)$  for  $A, A_1, A_2, \dots, B$  in a ring of sets on which  $\psi(\cdot)$  is defined. This operator is tailored to the needs of (9.2.16) in the lemma below; the sign convention in its definition here is opposite that of Kurtz and Kallenberg.

**Lemma 9.2.XI.** *For every integer  $k \geq 1$  and all Borel sets  $A_1, A_2, \dots, B$ ,*

$$\Delta(A_1, \dots, A_k)P_0(B) = \mathcal{P}\{N(A_i) > 0 \ (i = 1, \dots, k), \ N(B) = 0\}. \quad (9.2.13)$$

PROOF. For  $k = 1$  we have

$$\mathcal{P}\{N(A_1) > 0, \ N(B) = 0\} = P_0(B) - P_0(A_1 \cup B) = \Delta(A_1)P_0(B).$$

The general form follows by an induction argument (see Exercise 9.2.5).  $\square$

As a special case of (9.2.13) with  $B$  the null set,

$$\Delta(A_1, \dots, A_k)P_0(\emptyset) = \mathcal{P}\{N(A_i) > 0 \ (i = 1, \dots, k)\}. \quad (9.2.14)$$

The nonnegativity of (9.2.11) appears later in Theorem 9.2.XV in a characterization of the avoidance function. In the meantime, Lemma 9.2.XI provides a useful notational convention and serves as a reminder that the probability of the complex event on the right-hand side of (9.2.13) can be expressed immediately in terms of the avoidance function.

The basic idea motivating the introduction of the operator  $\Delta$  at (9.2.12) is that it leads to a succinct description of the fidi distributions of a point process when, for suitable sets  $A_i$ ,  $N(A_i)$  is ‘small’ in the sense of having

$$\mathcal{P}\{N(A_i) = 0 \text{ or } 1 \text{ (all } i)\} \approx 1.$$

Such an approximation can be realized only if  $N$  is simple and the class of sets on which the values of the avoidance function are known contains a dissecting system for  $\mathcal{X}$  (see Definition A1.6.I). Fortunately, on a c.s.m.s.  $\mathcal{X}$  the Borel sets  $\mathcal{B}_{\mathcal{X}}$  necessarily contain a dissecting system and hence a dissecting ring (Definition A2.1.V).

**Theorem 9.2.XII** (Rényi, 1967; Mönch, 1971). *The distribution of a simple point process  $N$  on a c.s.m.s.  $\mathcal{X}$  is determined by the values of the avoidance function  $P_0$  on the bounded sets of a dissecting ring  $\mathcal{A}$  for  $\mathcal{X}$ .*

PROOF. It is enough to show that the fdi distributions of a point process as at (9.2.4), involving only bounded subsets of  $\mathcal{X}$ , are determined by the avoidance function. We use an indicator function  $Z(B)$ ,  $B \in \mathcal{A}$ , and a dissecting system  $\mathcal{T}$  as in and below (9.3.12), for which the r.v.s

$$\zeta_n(A) = \sum_{i=1}^{k_n} Z(A_{ni}) \quad (n = 1, 2, \dots), \quad (9.2.15)$$

count the numbers of sets in  $\mathcal{T}_n$  containing points of  $N(\cdot)$ . Because every  $A_{ni}$  is the union of elements in  $\mathcal{T}_{n+1}$ , and the r.v.s  $Z(\cdot)$  are subadditive set functions, it follows that  $\{\zeta_n(A)\}$  is a nondecreasing sequence. Moreover, because  $N$  is simple and  $\{\mathcal{T}_n\}$  is a dissecting system, the limit

$$N(A) \equiv \lim_{n \rightarrow \infty} \zeta_n(A) \quad (9.2.16)$$

exists a.s. Now the joint distribution of the  $Z(A_{ni})$ , and hence of  $\zeta_n(A)$ , and (more generally) of  $\{\zeta_n(A_i) \ (i = 1, \dots, k)\}$ , is expressible directly in terms of the avoidance function: for example,

$$\mathcal{P}\{\zeta_n(A) = r\} = \sum_{\{i_1, \dots, i_r\}} \Delta(A_{ni_1}, \dots, A_{ni_r}) P_0\left(A \setminus \bigcup_{j=1}^r A_{ni_j}\right), \quad (9.2.17)$$

where the sum is taken over all  $\binom{k_n}{r}$  distinct combinations of  $r$  sets from the  $k_n$  ( $\geq r$ ) sets in the partition  $\mathcal{T}_n$  of  $A$ . Rather more cumbersome formulae give the joint distributions of  $\zeta_n(A_i)$ . Because the convergence of  $\{\zeta_n\}$  to its limit is monotonic, the sequence of events  $\{\zeta_n(A_i) \leq n_i \ (i = 1, \dots, k)\}$  is also monotone decreasing in  $n$ , and thus  $\mathcal{P}\{\zeta_n(A_i) \leq n_i \ (i = 1, \dots, k)\} \rightarrow \mathcal{P}\{N(A_i) \leq n_i \ (i = 1, \dots, k)\}$ .

Thus,  $P_0$  determines the fdi distributions as asserted.  $\square$

**Corollary 9.2.XIII.** *Let  $N_1, N_2$  be two point processes on  $\mathcal{X}$  whose avoidance functions coincide on the bounded sets of a dissecting ring for  $\mathcal{X}$ . Then their support point processes  $N_1^*$  and  $N_2^*$  are equivalent.*

Versions of these results that apply to random measures can be given (see Exercise 9.2.7): the avoidance functions are replaced by the Laplace transforms  $E(e^{-s\xi(A)})$  for fixed  $s > 0$ .

We turn finally to a characterization problem.

**Definition 9.2.XIV.** *A set function  $\psi$  defined on a ring  $\mathcal{R}$  of sets is completely monotone on  $\mathcal{R}$  if for every sequence  $\{A, A_1, A_2, \dots\}$  of members of  $\mathcal{R}$ ,*

$$\Delta(A_1, \dots, A_n) \psi(A) \geq 0 \quad (\text{every } n = 1, 2, \dots).$$

Note that this definition of complete monotonicity differs from conventional usage by the omission of a factor  $(-1)^n$  on the left-hand side of the inequality [see also the definition of  $\Delta$  at (9.2.12)].

Using Definition 9.2.XIV, Lemma 9.2.XI asserts that the avoidance function of a point process is completely monotone on  $\mathcal{B}_{\mathcal{X}}$ . Complete monotonicity of a set function is not sufficient on its own to characterize an avoidance function.

**Theorem 9.2.XV** (Kurtz, 1974). *Let  $\psi$  be a set function defined on the members of a dissecting ring  $\mathcal{R}$  covering the c.s.m.s.  $\mathcal{X}$ . In order that there exist a point process on  $\mathcal{X}$  with avoidance function  $\psi$ , it is necessary and sufficient that*

- (i)  $\psi$  be completely monotone;
- (ii)  $\psi(\emptyset) = 1$ ;
- (iii)  $\psi(A_n) \rightarrow 1$  for any bounded sequence  $\{A_n\}$  in  $\mathcal{R}$  for which  $A_n \rightarrow \emptyset$  ( $n \rightarrow \infty$ ); and
- (iv) for every bounded  $A \in \mathcal{R}$ ,

$$\lim_{r \rightarrow \infty} \lim_{n \rightarrow \infty} \left[ \psi(A) + \sum_{k=1}^r \sum_{\{i_1, \dots, i_r\}} \Delta(A_{ni_1}, \dots, A_{ni_r}) \psi\left(A \setminus \bigcup_{j=1}^k A_{ni_j}\right) \right] = 1,$$

where  $\{\mathcal{T}_n\} = \{\{A_{ni} : i = 1, \dots, k_n\}\}$  is a dissecting system for  $A$ ,  $\{\mathcal{T}_n\} \subseteq \mathcal{R}$ , and the inner summation is over all distinct combinations of  $k$  sets from the  $k_n$  sets in the partition  $\mathcal{T}_n$  for  $A$ .

PROOF. The necessity of (i) has been noted in Lemma 9.2.XI, condition (ii) is self-evident, and condition (iii) here is the same as (iv) of Theorem 9.2.X. Condition (iv) here follows most readily from (9.2.17) when written in the form

$$\lim_{r \rightarrow \infty} \lim_{n \rightarrow \infty} \sum_{k=0}^r \mathcal{P}\{\zeta_n(A) = k\} = \lim_{r \rightarrow \infty} \mathcal{P}\{N(A) \leq r\} = 1$$

and expresses the fact that a point process  $N$  is boundedly finite.

For the sufficiency, it is clear from (i) and (ii) that we can construct an indicator process  $Z'$  on bounded  $A \in \mathcal{R}$  with fid distributions (for any finite number  $k$  of disjoint bounded  $A_1, \dots, A_k \in \mathcal{R}$ )

$$\left. \begin{aligned} \Pr\{Z'(A_1) = 0\} &= 1 - \Pr\{Z'(A_1) = 1\} = \psi(A_1) = \Delta(A_1) \psi(\emptyset), \\ \Pr\{Z'(A_i) = 1 \ (i = 1, \dots, k)\} &= \Delta(A_1, \dots, A_k) \psi(\emptyset), \\ \Pr\{Z'(A_j) = 0, Z'(A_i) = 1 \ (\text{all } i \neq j)\} \\ &= \Delta(A_1, \dots, A_{j-1}, A_{j+1}, \dots, A_k) \psi(A_j), \\ \Pr\{Z'(A_i) = 0 \ (\text{all } i)\} &= \psi\left(\bigcup_{i=1}^k A_i\right); \end{aligned} \right\} \quad (9.2.18b)$$

nonnegativity is ensured by (i), summation to unity by (ii), and marginal consistency reduces to

$$\Delta(A_1, \dots, A_{k+1}) \psi(B) + \Delta(A_1, \dots, A_k) \psi(B \cup A_{k+1}) = \Delta(A_1, \dots, A_k) \psi(B).$$

In other words, we have a family of fidi distributions that, being consistent in the sense of the Kolmogorov existence theorem [e.g., Parthasarathy (1967, Chapter V)], enable us to assert the existence of a probability space  $(\mathcal{Z}, \mathcal{E}, \mathcal{P}')$  on which are jointly defined  $\{0, 1\}$ -valued r.v.s  $\{Z'(A): \text{bounded } A \in \mathcal{R}\}$ , and  $\mathcal{P}'$  is related to  $\psi$  via relations such as (9.2.18) (with  $\text{Pr}$  replaced by  $\mathcal{P}'$ ).

We now introduce r.v.s  $\zeta'_n(A)$  (bounded  $A \in \mathcal{R}$ ) much as at (9.2.18) and observe that the subadditivity of  $Z'$  implies that  $\zeta'_n(A)$  are a.s. monotone nondecreasing under refinement as before, so

$$\lim_{n \rightarrow \infty} \zeta'_n(A) \equiv N(A) \quad (9.2.19)$$

exists a.s. and, being the limit of an integer-valued sequence, is itself integer-valued or infinite. From the last relation at (9.2.18b), we have  $\mathcal{P}'\{\zeta'_n(A) = 0\} = \psi(A)$  for all  $n$ , so

$$\mathcal{P}'\{N'(A) = 0\} = \psi(A) \quad (\text{all bounded } A \in \mathcal{R}). \quad (9.2.20)$$

The a.s. finiteness condition that  $N'$  must satisfy on bounded  $A \in \mathcal{R}$  is equivalent to demanding that

$$\lim_{y \rightarrow \infty} \lim_{n \rightarrow \infty} \mathcal{P}'\{\zeta'_n(A) \leq y\} = 1,$$

which, expressed in terms of the functions  $\psi$  via (9.2.21) and relations such as (9.2.18) (with  $\mathcal{P}'$  and  $\psi$  replacing  $\mathcal{P}$  and  $P_0$ ), reduces to condition (iii).

For bounded disjoint  $A, B \in \mathcal{R}$ , we find by using a dissecting system for  $A \cup B$  containing dissecting systems for  $A$  and  $B$  separately that

$$N(A \cup B) = \lim_{n \rightarrow \infty} \zeta'_n(A \cup B) = \lim_{n \rightarrow \infty} [\zeta'_n(A) + \zeta'_n(B)] = N'(A) + N'(B) \quad \text{a.s.},$$

and thus  $N'$  is finitely additive on  $\mathcal{R}$ . Let  $\{A_i\}$  be any disjoint sequence in  $\mathcal{R}$  with bounded union

$$A \equiv \bigcup_{i=1}^{\infty} A_i \in \mathcal{R};$$

we seek to show that  $N'(A) = \sum_{i=1}^{\infty} N'(A_i)$ . Let  $B_r = \bigcup_{i=r+1}^{\infty} A_i = A \setminus \bigcup_{i=1}^r A_i$ , so  $B_r$  is bounded,  $\in \mathcal{R}$ ,  $\downarrow \emptyset$  ( $r \rightarrow \infty$ ), and thus  $\mathcal{P}'\{N'(B_r) = 0\} = \psi(B_r) \uparrow 1$ ; that is,  $N'(B_r) \rightarrow 0$  a.s. Define events  $C_r \in \mathcal{E}$  for  $r = 0, 1, \dots$  by

$$C_0 = \{N': N'(A) = 0\} \quad \text{and} \quad C_r = \{N': N'(B_r) = 0 < N'(B_{r-1})\}.$$

Then  $\mathcal{P}'(C_0 \cup C_1 \cup \dots) = 1$ , and  $N'(A) = \sum_{i=1}^r N'(A_i) + N'(B_r)$  on  $C_r$ . Also, on  $C_r$ , it follows from  $0 = N'(B_r) = \lim_{n \rightarrow \infty} \zeta'_n(B_r)$  that  $N'(A_i) = 0$  for  $i \geq r+1$  and hence  $\sum_{i=r+1}^{\infty} N'(A_i) = 0$  on  $C_r$ . Because  $\mathcal{P}'(\bigcup_{r=0}^{\infty} C_r) = 1$ , it now follows that  $N'$  is countably additive on  $\mathcal{R}$ . Then by the usual extension theorem for measures,  $N'$  can be extended a.s. to a countably additive boundedly finite nonnegative integer-valued measure on  $\mathcal{B}_{\mathcal{X}}$ . This extension, with the appropriate modification on the  $\mathcal{P}'$ -null set where the extension may fail, provides the required example of a point process with avoidance function  $\mathcal{P}'\{N'(A) = 0\} = \psi(A)$  ( $A \in \mathcal{R}$ ) satisfying conditions (i)–(iv).  $\square$

## Exercises and Complements to Section 9.2

- 9.2.1 Give an example of a family of fidi distributions satisfying (9.2.6) for disjoint sets and all other requirements of Theorem 9.2.VII apart from Condition 9.2.VI(a), which is not satisfied. [Hint: Let  $\mathcal{X}$  be a two-point space,  $\{x, y\}$  say, and construct a r.v.  $Z$  and a random set function  $\xi$  for which  $Z$  has the distribution of  $\xi(\{x, y\})$  but  $Z \neq \xi(\{x\}) + \xi(\{y\})$ .]
- 9.2.2 Give an example of a family of fidi distributions that satisfy (9.2.6) for  $k = 2$  but not for some  $k \geq 3$ , and hence do not satisfy the consistency Condition 9.2.V(b). [Hint: Modify the previous example.]
- 9.2.3 Show that the joint distributions of the Dirichlet process of Example 9.1(e) are consistent.
- 9.2.4 Let  $\Psi = \xi_1 - \xi_2$  be the difference of two random measures. For each  $\omega$ , let  $\Psi = \Psi_+ - \Psi_-$  be the Jordan–Hahn decomposition of  $\Psi(\omega)$  (Theorem A1.3.IV). Determine conditions under which the mappings  $\Psi_+$  and  $\Psi_-$  are measurable and hence define random measures. Investigate the extent to which such conditions can be extended to more general settings.
- 9.2.5 Let  $A_1, \dots, A_n$  be disjoint,  $A = \bigcup_{i=1}^n A_i$ , and  $\psi(\emptyset) = 1$ . Verify that the operator  $\Delta$  at (9.2.12) satisfies (a) and (b) below, and complete the induction proof of (9.2.13).

$$(a) \quad \Delta(A)\psi(B) = \sum_{k=1}^n \sum_{1 \leq i_1 < \dots < i_k \leq n} \Delta(A_{i_1}, \dots, A_{i_k}) \psi\left(B \cup \left(A \setminus \bigcup_{j=1}^k A_{i_j}\right)\right),$$

$$(b) \quad \sum_{i=1}^n (1 - \psi(A_i)) = \sum_{k=1}^n \sum_{1 \leq i_1 < \dots < i_k \leq n} k \Delta(A_{i_1}, \dots, A_{i_k}) \psi\left(A \setminus \bigcup_{j=1}^k A_{i_j}\right).$$

- 9.2.6 Let  $A_1, A_2, \dots$  be disjoint,  $A = \bigcup_{i=1}^{\infty} A_i$ , and  $\psi(\emptyset) = 1$ . Show that if

$$1 - \psi(A) = \sum_{k=1}^{\infty} \sum_{1 \leq i_1 < \dots < i_k < \infty} \Delta(A_{i_1}, \dots, A_{i_k}) \psi\left(A \setminus \bigcup_{j=1}^k A_{i_j}\right),$$

then

$$1 - \psi\left(\bigcup_{i=m}^{\infty} A_i\right) = \sum_{k=1}^{\infty} \sum_{m \leq i_1 < \dots < i_k < \infty} \Delta(A_{i_1}, \dots, A_{i_k}) \psi\left(A \setminus \bigcup_{j=1}^k A_{i_j}\right).$$

- 9.2.7 Let  $\xi$  be a random measure. Show that for each fixed  $s > 0$  the transform  $\varphi_s(A) \equiv E(e^{-s\xi(A)})$  is completely monotone in the argument  $A$ . Use this to develop a characterization theorem for random measures parallel to Theorem 9.2.XV. [Hint:  $\varphi_s(A)$  is the avoidance function of a Cox process with intensity measure  $s\xi(\cdot)$ . See Kallenberg (1975, Section 5.3).]

## 9.3. Sample Path Properties: Atoms and Orderliness

The outstanding feature of the *sample paths* or *realizations* of a random measure is their countable additivity. Nevertheless, some additional questions remain: for example, under what conditions on the fidi distributions will

the realizations be almost surely purely atomic? Or purely nonatomic? Or absolutely continuous with respect to some given measure? Similarly, in the point process context, we may ask for analytic conditions equivalent to the sample-path property of simplicity.

This section discusses some basic questions of this kind; Section 10.1 takes the discussion further in the particular context of completely random measures. A technique that runs through our analysis here and later, and has been exploited, for example, by Leadbetter (1968, 1972), Kallenberg (1975), and others, is the use of *dissecting systems*. Recall from Definition A1.6.I that a dissecting system  $\mathcal{T} = \{\mathcal{T}_n: n = 1, 2, \dots\}$  for the space  $\mathcal{X}$  is a nested sequence of finite partitions  $\mathcal{T}_n = \{A_{ni}: i = 1, \dots, k_n\}$  of Borel sets  $A_{ni}$  that ultimately separate points of  $\mathcal{X}$ ; that is, given any two distinct points  $x$  and  $y$  there exists an  $n$  such that  $x$  and  $y$  are contained in distinct members of  $\mathcal{T}_n$  (and hence in distinct members of  $\mathcal{T}_{n'}$  for all  $n' \geq n$ ). We note also that for any  $A \in \mathcal{B}_{\mathcal{X}}$ ,  $\mathcal{T} \cap A$  is a dissecting system for  $A$ .

Proposition A2.1.IV asserts that such systems exist for any c.s.m.s.  $\mathcal{X}$ . If, additionally,  $\mathcal{X}$  is locally compact (such as when  $\mathcal{X}$  is Euclidean), a sufficient condition for a family of nested partitions  $\mathcal{T}$  to be dissecting for a bounded Borel set  $A$  is that

$$\max_{1 \leq i \leq k_n} \text{diam}(A_{ni}) \rightarrow 0 \quad (n \rightarrow \infty). \quad (9.3.1)$$

Returning to sample paths, we start with a consideration of the fixed atoms, abbreviating the notation  $\xi(\{x\})$  to  $\xi\{x\}$ .

**Definition 9.3.I.** *The point  $x_0$  is a fixed atom of the random measure  $\xi$  if  $\mathcal{P}\{\xi\{x_0\} > 0\} > 0$ .*

It must be remarked that the adjective ‘fixed’ here refers to the locations of the atoms of the realizations  $\xi$ ; by contrast, if a realization  $\xi$  has  $\xi\{x\} > 0$  but  $\mathcal{P}\{\xi\{x\} > 0\} = 0$  then such an atom is termed a *random atom*.

**Lemma 9.3.II.** *A random measure  $\xi$  is free of fixed atoms on the bounded Borel set  $A$  if and only if for some (and then every) dissecting system  $\mathcal{T}$  for  $A$ , for every fixed  $\epsilon > 0$ ,*

$$\max_{1 \leq i \leq k_n} \mathcal{P}\{\xi(A_{ni}) > \epsilon\} \rightarrow 0 \quad (n \rightarrow \infty). \quad (9.3.2)$$

PROOF. Assume first that for some dissecting system  $\mathcal{T}$  for  $A$ , some  $\epsilon > 0$ , and some  $\eta > 0$ ,

$$\mathcal{P}\{\xi(A_{nj}) \geq \epsilon\} \geq \eta \quad (9.3.3)$$

for at least one set  $A_{nj}$  in each  $\mathcal{T}_n$ ,  $n = 1, 2, \dots$ . For each  $n$  let  $\mathcal{J}_n$  denote all such index pairs  $(n, j)$ ; set  $\mathcal{J} = \mathcal{J}_1 \cup \mathcal{J}_2 \cup \dots$ . Because the partitions  $\mathcal{T}_n$  are nested, every  $A_{n+1,j} \subseteq A_{n,j'(j)}$  for some  $j'(j)$ , and because  $\xi$  is a measure,  $\xi(A_{n+1,j}) \leq \xi(A_{n,j'(j)})$  a.s. Consequently, the infinite set  $\mathcal{J}$  is such that every  $\mathcal{J}_n$  contains at least one pair  $(n, j)$  for which  $A_{nj} \supseteq A_{n',j'}$  for infinitely many pairs  $(n', j') \in \mathcal{J}$ . Use this property to construct a sequence  $\{(n, j(n)): n = 1, 2, \dots\}$  such that  $A_{n,j(n)} \supseteq A_{n+1,j(n+1)}$  and  $(n, j(n)) \in \mathcal{J}_n$



for each  $n$ . By the separation property of  $\mathcal{T}$ ,  $A_\infty \equiv \lim_{n \rightarrow \infty} A_{n,j(n)}$  is either the empty set  $\emptyset$  or a singleton set,  $\{x_0\}$  say. Because  $A$  is bounded, so is  $A_{1,j(1)}$ , and therefore  $\infty > \xi(A_{n,j(n)})$ . Because  $\xi$  is a random measure, we then have

$$\xi(A_{n,j(n)}) \downarrow \xi(A_\infty) \quad (n \rightarrow \infty),$$

and therefore  $\mathcal{P}\{\xi(A_\infty) \geq \epsilon\} \geq \eta$ . Because  $\xi(\emptyset) = 0$  a.s. and  $\epsilon > 0$ , we must have  $A_\infty = \{x_0\}$ , and thus  $\xi$  has at least one fixed atom.

Conversely, if  $\xi$  has a fixed atom  $x_0 \in A$ , there exists  $\epsilon > 0$  for which  $0 < \mathcal{P}\{\xi\{x_0\} > \epsilon\} \equiv \eta'$ . Then, given any dissecting system  $\mathcal{T}$  for  $A$ , there exists a set  $A_{n,j(n)}$  in each  $\mathcal{T}_n$  such that  $x_0 \in A_{n,j(n)}$ , and

$$\mathcal{P}\{\xi(A_{n,j(n)}) > \epsilon\} \geq \mathcal{P}\{\xi\{x_0\} > \epsilon\} = \eta' \quad (\text{all } n),$$

so (9.3.2) fails for any dissecting system  $\mathcal{T}$  for  $A$ . □

Once the fixed atoms have been identified, one would anticipate representing the random measure as the superposition of two components, the first containing all fixed atoms and the second free from fixed atoms. It is not absolutely clear, however, that this procedure corresponds to a measurable operation on the original process. The following establishes this fact.

**Lemma 9.3.III.** *The set  $D$  of fixed atoms of a random measure is countably infinite at most.*

PROOF. Suppose on the contrary that  $D$  is uncountable. Because  $\mathcal{X}$  can be covered by the union of at most countably many bounded Borel sets, there exists a bounded set,  $A$  say, containing uncountably many fixed atoms.

Define the subset  $D_\epsilon$  of  $D \cap A$  by

$$D_\epsilon = \{x: \mathcal{P}\{\xi\{x\} > \epsilon\} > \epsilon\}, \quad (9.3.4)$$

observing by monotonicity that  $D \cap A = \lim_{\epsilon \downarrow 0} D_\epsilon$ . If  $D_\epsilon$  is finite for every  $\epsilon > 0$ , then by a familiar construction we can deduce that  $D \cap A$  is countable, so for some positive  $\epsilon$  which we fix for the remainder of the proof,  $D_\epsilon$  is infinite.

We can extract from  $D_\epsilon$  an infinite sequence of distinct points  $\{x_1, x_2, \dots\}$  for which the events  $E_n \equiv \{\xi: \xi\{x_n\} > \epsilon\}$  have  $\mathcal{P}(E_n) > \epsilon$ . Because  $\xi$  is boundedly finite,

$$\begin{aligned} 0 &= \mathcal{P}\{\xi(A) = \infty\} \\ &\geq \mathcal{P}\{\xi\{x\} > \epsilon \text{ for infinitely many } x \in D_\epsilon\} \\ &\geq \mathcal{P}\{\text{infinitely many } E_n \text{ occur}\} \\ &= \mathcal{P}\left\{\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k\right\} = \lim_{n \rightarrow \infty} \mathcal{P}\left\{\bigcup_{k=n}^{\infty} E_k\right\} \\ &\geq \epsilon > 0, \end{aligned}$$

thereby yielding a contradiction. □

It is convenient to represent the countable set  $D$  by  $\{x_k\}$  and to write  $U_k \equiv U_k(\omega)$  for the random variable  $\xi\{x_k\}$ . Using Dirac measure  $\delta_x$  as in (9.1.3), the set function  $\xi_c$  defined for bounded Borel sets  $A$  by

$$\xi_c(A, \omega) = \xi(A, \omega) - \sum_{x_k \in D} U_k(\omega) \delta_{x_k}(A)$$

is positive and countably additive in  $A$ , and for every such  $A$  it defines a random variable. Thus, it defines a new random measure that is clearly free from fixed atoms, and we have proved the following extension of Proposition 9.1.III(i) of properties of a fixed measure  $\mu$  to those of a random measure  $\xi$ .

**Proposition 9.3.IV.** *Every random measure  $\xi$  can be written in the form*

$$\xi(\cdot, \omega) = \xi_c(\cdot, \omega) + \sum_{k=1}^{\infty} U_k(\omega) \delta_{x_k}(\cdot),$$

where  $\xi_c$  is a random measure without fixed atoms, the sequence  $\{x_k: k = 1, 2, \dots\}$  constitutes the set  $D$  of all fixed atoms of  $\xi$ , and  $\{U_k: k = 1, 2, \dots\}$  is a sequence of nonnegative random variables.

Consider next the more general question of finding conditions for the trajectories to be a.s. nonatomic. As before, let  $A$  be a bounded Borel set and  $\mathcal{T} = \{\mathcal{T}_n: n = 1, 2, \dots\}$  a dissecting system for  $A$ . For any given  $\epsilon > 0$ , we can ‘trap’ any atoms of  $\xi$  with mass  $\epsilon$  or greater by the following construction. For each  $n$ , set

$$N_\epsilon^{(n)}(A) = \#\{i: A_{ni} \in \mathcal{T}_n, \xi(A_{ni}) \geq \epsilon\}. \quad (9.3.5)$$

Then each  $N_\epsilon^{(n)}(A)$  is a.s. finite, being bounded uniformly in  $n$  by  $\xi(A)/\epsilon$ . Moreover, as  $n \rightarrow \infty$ ,  $N_\epsilon^{(n)}(A)$  converges a.s. to a limit r.v.,  $N_\epsilon(A)$  say, which is independent of the particular dissecting system  $\mathcal{T}$  and which represents the number of atoms in  $A$  with mass  $\epsilon$  or greater (see Exercise 9.3.2 for a more formal treatment of these assertions). Consequently,  $\xi$  is a.s. nonatomic on  $A$  if and only if for each  $\epsilon > 0$ ,  $N_\epsilon(A) = 0$  a.s. Because  $N_\epsilon^{(n)}(A)$  converges a.s. to  $N_\epsilon(A)$  irrespective of the value of the latter, a necessary and sufficient condition for  $N_\epsilon = 0$  a.s. is that  $N_\epsilon^{(n)} \rightarrow 0$  in probability. This leads to the following criterion.

**Lemma 9.3.V.** *The random measure  $\xi$  is a.s. nonatomic on bounded  $A \in \mathcal{B}_X$  if and only if for every  $\epsilon > 0$  and for some (and then every) dissecting system  $\mathcal{T}$  for  $A$ ,*

$$\mathcal{P}\{\#\{i: \xi(A_{ni}) \geq \epsilon\} > 0\} \rightarrow 0 \quad (n \rightarrow \infty). \quad (9.3.6)$$

**Corollary 9.3.VI.** *A sufficient condition for  $\xi$  to be a.s. nonatomic on bounded  $A \in \mathcal{B}_X$  is that for some dissecting system for  $A$  and every  $\epsilon > 0$ ,*

$$\sum_{i=1}^{k_n} \mathcal{P}\{\xi(A_{ni}) \geq \epsilon\} \rightarrow 0 \quad (n \rightarrow \infty). \quad (9.3.7)$$

*If  $\xi$  is a completely random measure then this condition is also necessary.*

PROOF. Equation (9.3.7) is sufficient for (9.3.6) to hold because

$$\mathcal{P}\left(\bigcup_{i=1}^{k_n} \{\xi(A_{ni}) \geq \epsilon\}\right) \leq \sum_{i=1}^{k_n} \mathcal{P}\{\xi(A_{ni}) \geq \epsilon\}.$$

When  $\xi$  is completely random (Definition 10.1.I), the r.v.s  $\xi(A_{ni})$  are mutually independent and hence

$$\begin{aligned} 1 - \mathcal{P}\left(\bigcup_{i=1}^{k_n} \{\xi(A_{ni}) \geq \epsilon\}\right) &= \mathcal{P}\left(\bigcap_{i=1}^{k_n} \{\xi(A_{ni}) < \epsilon\}\right) \\ &= \prod_{i=1}^{k_n} \mathcal{P}\{\xi(A_{ni}) < \epsilon\} = \prod_{i=1}^{k_n} (1 - \mathcal{P}\{\xi(A_{ni}) \geq \epsilon\}). \end{aligned} \quad (9.3.8)$$

If now  $\xi$  is a.s. nonatomic, then by (9.3.6) the left-hand side of (9.3.8) converges to 1 as  $n \rightarrow \infty$ . Finally, the convergence to 1 of the product of the right-hand side of (9.3.8) implies (9.3.7).  $\square$

Exercise 9.3.3 shows that (9.3.7) is not necessary for a random measure to be absolutely continuous. Indeed, it appears to be an open problem to find simple sufficient conditions, analogous to Corollary 9.3.VI, for the realizations of a random measure to be a.s. absolutely continuous with respect to a given measure (see also Exercise 9.1.7).

EXAMPLE 9.3(a) *Quadratic random measures* [see also Example 9.1(b)]. Take  $A$  to be the unit interval  $(0, 1]$ , and for  $n = 1, 2, \dots$  divide this into  $k_n = 2^n$  subintervals each of length  $2^{-n}$  to obtain suitable partitions for a dissecting system  $\mathcal{T}$ . Each  $\xi(A_{ni})$  can be represented in the form

$$\xi_{ni} \equiv \xi(A_{ni}) = \int_{i/k_n}^{(i+1)/k_n} Z^2(t) dt \approx (1/k_n) Z^2(i/k_n).$$

Because  $Z^2(i/k_n)$  has a  $\chi^2$  distribution on one degree of freedom, we deduce (as may be shown by a more careful analysis) that

$$\Pr\{\xi_{ni} > \epsilon\} = \Pr\{Z^2(i/k_n) > k_n \epsilon\} (1 + O(1)) \leq C \exp(-k_n \epsilon)$$

for some finite constant  $C$ . Then

$$\sum_{i=1}^{k_n} \Pr\{\xi_{ni} > \epsilon\} \leq C k_n e^{-k_n \epsilon} \rightarrow 0 \quad (n \rightarrow \infty).$$

Thus,  $\xi$  being an integral with a.s. continuous integrand its trajectories are a.s. nonatomic.  $\square$

We turn now to point processes; ultimately we generalize the results of Section 3.3. Simplicity is again a sample-path property; in terms of Propo-

sition 9.1.III, it occurs when  $N$  coincides with its support counting measure  $N^*$ , although this is not the only way it can be described.

For MPPs, we note as in Exercise 9.1.6 that any MPP that is not simple can be redefined as a simple MPP with marks in the compound space  $\mathcal{K}^\cup$ . In particular, a nonsimple point process on  $\mathcal{X}$  can be regarded as a simple point process on  $\mathcal{X} \times \{1\}^\cup = \mathcal{X} \times \mathbb{Z}_+$  [recall also Proposition 9.1.III(iv)].

In practice, it is arguably more useful to have analytic conditions for a point process to be simple. The main approach in developing such conditions is again via dissecting systems.

We start from the representation (9.1.6) of a counting measure  $N$ , writing it now in the form

$$N = \sum_{k=1}^{\infty} k N_k^*, \quad (9.3.9)$$

where

$$N_k^*(A) = \#\{x_i \in A : N\{x_i\} = k\} \quad (k = 1, 2, \dots). \quad (9.3.10)$$

Then for the support counting measure  $N^*$  of  $N$  we can write (cf. (9.1.7))

$$N^* = \sum_{k=1}^{\infty} N_k^*. \quad (9.3.11)$$

We would like to regard (9.3.9) and (9.3.11) as statements concerning point processes as well as statements about individual realizations. To this end we use Proposition 9.1.VIII. It is clear from the construction that  $N^*$  and each  $N_k^*$  are elements of  $\mathcal{N}_{\mathcal{X}}^{\#*}$ ; the essential point is to show that for any bounded Borel set  $A$ ,  $N^*(A)$  and  $N_k^*(A)$  are r.v.s. We establish this by a construction which plays an important role in later arguments.

Suppose then that  $N$  is a point process, and for bounded  $B \in \mathcal{B}_{\mathcal{X}}$  define the indicator functions

$$Z_k(B) = I_{\{1 \leq N(B) \leq k\}} = \begin{cases} 1 & \text{if } 1 \leq N(B) \leq k, \\ 0 & \text{otherwise,} \end{cases} \quad (9.3.12a)$$

$$Z(B) = I_{\{N(B) \geq 1\}} = \begin{cases} 1 & \text{if } N(B) \geq 1, \\ 0 & \text{otherwise,} \end{cases} \quad (9.3.12b)$$

which are r.v.s because the  $N(B)$  are r.v.s. All of the  $Z_k$ , as well as  $Z$ , are *subadditive set functions*, meaning, for example, that for  $A, B \in \mathcal{B}_{\mathcal{X}}$ ,

$$Z_k(A \cup B) \leq Z_k(A) + Z_k(B).$$

Such subadditive set functions have the important properties that

$$\begin{aligned} \mathcal{P}\{1 \leq N(A \cup B) \leq k\} &= \mathbb{E}[Z_k(A \cup B)] \\ &\leq \mathcal{P}\{1 \leq N(A) \leq k\} + \mathcal{P}\{1 \leq N(B) \leq k\}, \end{aligned} \quad (9.3.13)$$

with similar inequalities for  $\mathcal{P}\{N(B) \geq 1\}$ .

As around (9.3.9), let  $\mathcal{T} = \{\mathcal{T}_n\} = \{\{A_{ni}\}\}$  be a dissecting system for any given bounded  $A \in \mathcal{B}_{\mathcal{X}}$ . Then

$$\zeta_k^{(n)}(A) = \sum_{i: A_{ni} \in \mathcal{T}_n} Z_k(A_{ni}), \quad \zeta^{(n)}(A) = \sum_{i: A_{ni} \in \mathcal{T}_n} Z(A_{ni})$$

are further r.v.s. Here, for example,  $\zeta^{(n)}$  counts the number of subsets in the  $n$ th partition containing at least one point of  $N$ . The  $\zeta_k^{(n)}$  are nondecreasing in  $n$ , as follows from (9.3.13) and the fact that  $\mathcal{T}_{n+1}$  partitions each  $A_{ni} \in \mathcal{T}_n$ . So for fixed  $k$  and  $A$ ,

$$\zeta_k^{(n)}(A) \leq \zeta_k^{(n+1)}(A) \leq N(A) < \infty \quad (n = 1, 2, \dots);$$

hence

$$\zeta_k(A) = \lim_{n \rightarrow \infty} \zeta_k^{(n)}(A) \quad \text{and} \quad \zeta(A) = \lim_{n \rightarrow \infty} \zeta^{(n)}(A), \quad (9.3.14)$$

being monotone limits of bounded sequences, exist a.s. and again define r.v.s.

Although spheres are used in the derivation of (9.1.5), we could equally use elements of  $\mathcal{T}$ , thereby showing explicitly that  $\zeta_1(A) = N_1^*(A)$  and that  $\lim_{k \rightarrow \infty} \zeta_k(A)$ , which exists because  $\zeta_k(A) \leq \zeta_{k+1}(A) \leq N(A)$  ( $k = 1, 2, \dots$ ), equals  $N^*(A)$ . Thus  $N_1^*(A)$ ,  $N^*(A)$ , and

$$N_k^*(A) = \zeta_k(A) - \zeta_{k-1}(A) \quad (k = 2, 3, \dots)$$

are all random variables. It follows from Proposition 9.1.VIII that, whenever  $N$  is a point process, the random set functions  $N_k^*(\cdot)$  and  $N^*(\cdot)$  are also point processes.

We summarize this discussion in the following proposition.

**Proposition 9.3.VII.** *For any point process  $N$ , the constructions at (9.1.7) and (9.3.10) applied to realizations of  $N$  define simple point processes  $N^*$  and  $N_k^*$ , and the relations (9.3.9) and (9.3.11) hold as relations between point processes.*

*In particular,  $N$  is simple if and only if  $N_k^* = 0$  a.s. ( $k = 2, 3, \dots$ ).*

The next result generalizes Proposition 3.3.I and identifies a general form of the intensity with the first moment measure of  $N^*$ .

**Definition 9.3.VIII.** *The intensity measure of a point process  $N$  is the set function*

$$\lambda(A) \equiv \sup_{\mathcal{T}_n \in \mathcal{T}(A)} \sum_{i: A_{ni} \in \mathcal{T}_n} \mathcal{P}\{N(A_{ni}) \geq 1\} \quad (A \in \mathcal{B}_{\mathcal{X}}).$$

**Proposition 9.3.IX** (Khinchin's Existence Theorem). *Whether finite or infinite,*

$$\lambda(A) = \text{EN}^*(A) \equiv M^*(A); \quad (9.3.15)$$

*it is independent of the choice of dissecting system  $\mathcal{T}$ , and defines a measure when it is boundedly finite.*

PROOF. Using the monotonicity of  $\zeta^{(n)}$  and the interchange of limits that monotone convergence permits, we find that for any given dissecting system for  $A$ ,

$$\begin{aligned} M^*(A) &\equiv \text{E}(N^*(A)) = \text{E}\left(\lim_{n \rightarrow \infty} \zeta^{(n)}(A)\right) \\ &= \lim_{n \rightarrow \infty} \text{E}(\zeta^{(n)}(A)) \\ &= \lim_{n \rightarrow \infty} \sum_{i: A_{ni} \in \mathcal{T}_n} \mathcal{P}\{N(A_{ni}) \geq 1\} \\ &= \sup_{\mathcal{T}_n} \sum_{i: A_{ni} \in \mathcal{T}_n} \mathcal{P}\{N(A_{ni}) \geq 1\} = \lambda(A). \end{aligned}$$

The relation  $N^*(A) = \lim_{n \rightarrow \infty} \zeta^{(n)}(A)$ , and hence also  $\lambda(A)$  itself, is independent of the particular choice of dissecting system. The rest of the proposition follows easily from the fact that the first moment measure  $M(\cdot)$  of a point process is indeed a measure [see the discussion around (5.4.1)].  $\square$

A consequence of the equality of  $\lambda(\cdot)$  with  $M^*(\cdot)$  is that it is a well-defined measure, possibly 'extended' in the sense that we may have  $\lambda(A) = \infty$  for some bounded  $A \in \mathcal{B}_{\mathcal{X}}$ .

Notice also that Definition 9.3.VIII and the subsequent proposition form a particular case of a subadditive set function yielding a measure by addition under refinement (see Exercise 9.3.5).

In the next result the direct part is now trivial; it generalizes Proposition 3.3.IV and so may be called *Korolyuk's theorem*. The converse may be referred to as *Dobrushin's lemma* (cf. Proposition 3.3.V), because it extends naturally a result first referenced in Volkonski (1960) for stationary point processes.

**Proposition 9.3.X.** *For a simple point process  $N$ ,*

$$\lambda(A) = M(A) \quad (\text{all } A \in \mathcal{B}_{\mathcal{X}}). \quad (9.3.16)$$

*Conversely, if (9.3.16) holds and  $\lambda(A) < \infty$  for all bounded  $A$ , then  $N$  is simple.*

PROOF. When  $N$  is not simple, there is some bounded  $A \in \mathcal{B}_{\mathcal{X}}$  for which  $\Delta \equiv \mathcal{P}\{N(A) \neq N^*(A)\} > 0$ . Then

$$M(A) = \text{EN}(A) \geq \Delta + \text{EN}^*(A) = \Delta + \lambda(A)$$

and (9.3.16) cannot hold when  $\lambda(A) < \infty$ .  $\square$

Proposition 9.3.VII asserts that each  $N_k^*$  is a simple point process and thus has an intensity measure  $\lambda_k^* = \mathbf{E}N_k^*$ . From (9.3.9) we may therefore deduce the *generalized Korolyuk equation*

$$M(A) = \mathbf{E}N(A) = \sum_{k=1}^{\infty} k\lambda_k^*(A) = \lambda(A) + \sum_{k=1}^{\infty} [\lambda(A) - \lambda_k(A)], \quad (9.3.17)$$

where  $\lambda_k = \lambda_1^* + \cdots + \lambda_k^*$  is the intensity measure of the simple point process  $\zeta_k$ . Exercise 9.3.6 notes a version of (9.3.17) applicable to atomic random measures.

Further analytic conditions for simplicity involve infinitesimals directly, and usually bear the name *orderly* or *ordinary*, the latter deriving from transliteration rather than translation of Khinchin's original terminology.

**Definition 9.3.XI.** A point process  $N$  is ordinary when, given any bounded  $A \in \mathcal{B}_{\mathcal{X}}$ , there is a dissecting system  $\mathcal{T} = \{\mathcal{T}_n\} = \{\{A_{ni} : i = 1, \dots, k_n\}\}$  for  $A$  such that

$$\inf_{\mathcal{T}_n} \sum_{i=1}^{k_n} \mathcal{P}\{N(A_{ni}) \geq 2\} = 0. \quad (9.3.18)$$

Thus, for an ordinary point process, given  $\epsilon > 0$ , there exists a dissecting system for bounded  $A$  and some  $n_\epsilon$  such that for  $n \geq n_\epsilon$ ,

$$\epsilon > \sum_{i=1}^{k_n} \mathcal{P}\{N(A_{ni}) \geq 2\} \geq \mathcal{P}\left(\bigcup_{i=1}^{k_n} \{N(A_{ni}) \geq 2\}\right) \equiv \mathcal{P}(B_n).$$

Now the sequence of sets  $\{B_n\}$  is monotone decreasing, so for any  $\epsilon > 0$ ,

$$\epsilon > \mathcal{P}(\lim_{n \rightarrow \infty} B_n) = \mathcal{P}\{N\{x\} \geq 2 \text{ for some } x \in A\},$$

which establishes the direct part of the following result. The proof of the converse is left as Exercise 9.3.7.

**Proposition 9.3.XII.** An ordinary point process is necessarily simple. Conversely, a simple point process of finite intensity is ordinary, but one of infinite intensity need not be ordinary.

The following two analytic conditions have a pointwise character, and hence may be simpler to check in practice than the property of being ordinary. To describe them, take any bounded set  $A$  containing a given point  $x$ , let  $\mathcal{T}$  be a dissecting system for  $A$ , and for each  $n = 1, 2, \dots$ , let  $A_n(x)$  denote the member of  $\mathcal{T}_n = \{A_{ni}\}$  that contains  $x$ .

**Definition 9.3.XIII.** (a) A point process on the c.s.m.s.  $\mathcal{X}$  is  $\mu$ -orderly at  $x$  when  $\mu$  is a boundedly finite measure and

$$f_n(x) \equiv \frac{\mathcal{P}\{N(A_n(x)) \geq 2\}}{\mu(A_n(x))} \rightarrow 0 \quad (n \rightarrow \infty), \quad (9.3.19)$$

where if  $\mu(A_n(x)) = 0$  then  $f_n(x) = 0$  or  $\infty$  according as  $\mathcal{P}\{N(A_n(x)) \geq 2\} = 0$  or  $> 0$ .

(b) *The process is Khinchin orderly at  $x$  if*

$$g_n(x) \equiv \mathcal{P}\{N(A_n(x)) \geq 2 \mid N(A_n(x)) \geq 1\} \rightarrow 0 \quad (9.3.20)$$

as  $n \rightarrow \infty$ , where  $g_n(x) = 0$  if  $\mathcal{P}\{N(A_n(x)) \geq 1\} = 0$ .

In many situations the state space  $\mathcal{X}$  is a locally compact group with a boundedly finite invariant measure  $\nu$ . If a point process is  $\nu$ -orderly for a dissecting system based on spheres, we speak of the process as being *orderly*. Such usage is consistent with Khinchin's (1955) original use of the term for stationary point processes on  $\mathbb{R}$ ; a point process on  $\mathbb{R}$  uniformly analytically orderly in Daley's (1974) terminology is orderly in the present sense.

**Proposition 9.3.XIV.** *Suppose that for bounded  $A \in \mathcal{B}_{\mathcal{X}}$  the point process  $N$  is  $\mu$ -orderly at  $x$  for  $x \in A$ , and satisfies*

$$\sup_n \sup_{x \in A} f_n(x) < \infty, \quad (9.3.21)$$

where  $f_n(\cdot)$  is defined by (9.3.19). Then  $N$  is simple on  $A$ .

PROOF. From the definition at (9.3.19),

$$\sum_{i=1}^{k_n} \mathcal{P}\{N(A_{ni}) \geq 2\} = \int_A f_n(x) \mu(dx).$$

Here,  $f_n(x) \rightarrow 0$  pointwise, and by using (9.3.21) to justify appealing to the dominated convergence theorem, the integral  $\rightarrow 0$  as  $n \rightarrow \infty$ . The process is thus ordinary, and Proposition 9.3.XII completes the proof.  $\square$

When  $\lambda$  is boundedly finite, the hypotheses here can be weakened by dropping the requirement (9.3.21) and demanding merely that (9.3.19) and (9.3.20) should hold for  $\mu$ -a.e.  $x$  in  $A$ . This observation is the key to obtaining a partial converse to the proposition: we give it in the context of Khinchin orderliness.

**Proposition 9.3.XV.** *A point process  $N$  with boundedly finite intensity measure  $\lambda(\cdot)$  is simple if and only if it is Khinchin orderly for  $\lambda$ -a.e.  $x$  on  $\mathcal{X}$ .*

PROOF. It suffices to restrict attention to a bounded set  $A$ . To prove sufficiency, use the fact that  $\mathcal{P}\{N(A_{ni}) \geq 1\} \leq \lambda(A_{ni})$  to write

$$\sum_{i=1}^{k_n} \mathcal{P}\{N(A_{ni}) \geq 2\} \leq \int_A g_n(x) \lambda(dx), \quad \rightarrow 0 \quad (n \rightarrow \infty)$$

because from (9.3.20),  $1 \geq g_n(x) \rightarrow 0$  for  $\lambda$ -a.e.  $x$  and  $\lambda(A) < \infty$ .

To prove the necessity, suppose that  $N$  is simple: we first show that

$$h_n(x) \equiv \frac{\mathcal{P}\{N(A_n(x)) \geq 1\}}{\lambda(A_n(x))} \rightarrow 1 \quad (n \rightarrow \infty) \quad \text{for } \lambda\text{-a.e. } x. \quad (9.3.22)$$



We establish this convergence property via a martingale argument much as in the treatment of Radon–Nikodym derivatives [see Lemma A1.6.III or, e.g., Chung (1974, Section 9.5(VIII))]. Construct a sequence of r.v.s  $\{X_n\} \equiv \{X_n(\omega)\}$  on a probability space  $(A, \mathcal{B}_A, \mathbb{P})$ , where  $\mathbb{P}(\cdot) = \lambda(\cdot)/\lambda(A)$ , by introducing indicator r.v.s  $I_{ni}(x) = 1$  for  $x \in A_{ni}$ ,  $= 0$  otherwise, and setting

$$X_n(x) = \sum_{i=1}^{k_n} h_n(x) I_{ni}(x). \quad (9.3.23)$$

Let  $\mathcal{F}'_n$  denote the  $\sigma$ -field generated by the sets of  $\mathcal{T}_1 \cup \dots \cup \mathcal{T}_n$ ; because  $\{\mathcal{T}_n\}$  is a nested system of partitions,  $\mathcal{F}'_n$  has quite a simple structure (!). Then,  $\{X_n\}$  is a submartingale because  $\{\mathcal{F}'_n\}$  is an increasing sequence of  $\sigma$ -fields, and on the set  $A_n(x) = A_{ni}$  say,

$$\begin{aligned} \mathbb{E}(X_{n+1} \mid \mathcal{F}'_n) &= \sum_{j: A_{n+1,j} \subseteq A_{ni}} \frac{\mathcal{P}\{N(A_{n+1,j}) \geq 1\}}{\lambda(A_{n+1,j})} I_{n+1,j}(x) \frac{\lambda(A_{n+1,j})}{\lambda(A_{ni})} \\ &\geq \frac{\mathcal{P}\{N(A_{ni}) \geq 1\}}{\lambda(A_{ni})} I_{ni}(x) = X_n \quad \mathbb{P}\text{-a.s.}, \end{aligned}$$

where the inequality comes from the subadditivity of  $\mathcal{P}\{N(\cdot) \geq 1\}$ , so  $\{X_n\}$  is a submartingale. Now  $h_n(x) \leq 1$  (all  $x$ ), so  $X_n(\omega) \leq 1$  (all  $\omega$ ), and we can apply the submartingale convergence theorem [Theorem A3.4.III or, e.g., Chung (1974, Theorem 9.4.4)] and conclude that  $X_n(x)$  converges  $\mathbb{P}$ -a.s. Equivalently,  $\lim_{n \rightarrow \infty} h_n(x)$  exists  $\lambda$ -a.e. on  $A$ , so to complete the proof of (9.3.22) it remains to identify the limit. For this, it is enough to show that  $\limsup_{n \rightarrow \infty} h_n(x) = 1$   $\lambda$ -a.e., and this last fact follows from  $h_n(x) \leq 1$  (all  $x$ ) and the chain of relations

$$\lambda(A) = \lim_{n \rightarrow \infty} \int_A h_n(x) \lambda(dx) \leq \int_A \limsup_{n \rightarrow \infty} h_n(x) \lambda(dx) \leq \lambda(A),$$

in which we have used the lim sup version of Fatou's lemma.

The same martingale argument can be applied to the function

$$h_n^{(1)}(x) \equiv \begin{cases} \mathcal{P}\{N(A_n(x)) = 1\} / \lambda(A_n(x)) & \text{if } \lambda(A_n(x)) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

because the set function  $\mathcal{P}\{N(\cdot) = 1\}$  is again subadditive. Now for a simple point process with boundedly finite  $\lambda$ ,

$$\sum_{i=1}^{k_n} \mathcal{P}\{N(A_{ni}) = 1\} \rightarrow \lambda_1^*(A) = \lambda(A) \quad (n \rightarrow \infty),$$

so it again follows that

$$h_n^{(1)}(x) \rightarrow 1 \quad \lambda\text{-a.e.} \quad (9.3.24)$$

Combining (9.3.22) and (9.3.24),  $g_n(x) = 1 - h_n^{(1)}(x)/h_n(x) \rightarrow 0$  as  $n \rightarrow \infty$  for  $\lambda$ -a.e.  $x$ .  $\square$

Exercises 9.3.9–10 show that the  $\lambda$ -a.e. qualification cannot be relaxed.

Clearly, from (9.3.22) and (9.3.24), the proposition could equally well be phrased in terms of  $\lambda$ -orderliness rather than Khinchin orderliness. In this form the significance of the result is more readily grasped, namely, that for a simple point process with boundedly finite  $\lambda$ , not only are  $M(\cdot)$  and  $\lambda(\cdot)$  interchangeable but also, for suitably ‘small’ sets  $\delta A$  and  $\lambda$ -a.e., we can interpret

$$M(\delta A) = \lambda(\delta A) = \mathcal{P}\{N(\delta A) = 1\}(1 + o(1)). \quad (9.3.25)$$

Note that for any  $x$  with  $\lambda\{x\} > 0$ ,  $\lambda\{x\} = \mathcal{P}\{N\{x\} = 1\} = \mathcal{P}\{N\{x\} \geq 1\}$  because  $N$  is simple. Equation (9.3.25) provides a link between the statements in Chapter 3 of conditional probabilities as elementary limits and those in Chapter 13 derived by direct appeal to the Radon–Nikodym theorem.

The converse parts of the last few propositions in this section have included the proviso that the intensity measure be boundedly finite; without this proviso, the assertions may be false (see Exercise 9.3.11 and the references there).

### Exercises and Complements to Section 9.3

9.3.1 Use Lemma 9.3.II to show that a general gamma random measure process [see Example 9.1(d)] has no fixed atoms if and only if its shape parameter is a nonatomic measure. [See also the remark in Example 9.5(c).]

9.3.2 An elaboration of the argument leading to Lemma 9.3.V is as follows. Suppose given  $\mu \in \mathcal{M}_{\mathcal{X}}^{\#}$ , bounded  $A \in \mathcal{B}_{\mathcal{X}}$ , a dissecting system  $\mathcal{T} = \{\mathcal{T}_n\}$  for  $A$ , and  $\epsilon > 0$ . Define

$$\nu_{\epsilon}^{(n)}(A) = \#\{i: A_{ni} \in \mathcal{T}_n, \mu(A_{ni}) \geq \epsilon\}.$$

Let  $x_1, \dots, x_k$  be atoms in  $A$  whose masses  $u_1, \dots, u_k$  are at least  $\epsilon$  [and, because  $\mu(A) < \infty$ ,  $k \equiv k(\epsilon)$  is certainly finite]. Verify the following.

- (a)  $k = 0$  if and only if  $\nu_{\epsilon}^{(n)}(A) = 0$  for all sufficiently large  $n$  (cf. Lemma 9.3.III).
- (b) More generally, there exists  $n' < \infty$  such that  $\nu_{\epsilon}^{(n)}(A) = k$  for all  $n \geq n'$ .
- (c) Because  $k$  is independent of  $\mathcal{T}$ , so is  $\nu_{\epsilon}(A) \equiv \lim_{n \rightarrow \infty} \nu_{\epsilon}^{(n)}(A)$ .
- (d) Let  $A$  vary over the bounded Borel sets. Then  $\nu_{\epsilon}(\cdot)$  is a measure (in fact,  $\nu_{\epsilon} \in \mathcal{N}_{\mathcal{X}}^{\#*}$  of Definition 9.1.II).
- (e) Given a random measure  $\xi$ , so that  $\xi(\cdot, \omega) \in \mathcal{M}_{\mathcal{X}}^{\#}$ , denote by  $N_{\epsilon}(\cdot, \omega)$  the counting measure that corresponds to  $\nu_{\epsilon}$  defined from  $\mu = \xi(\cdot, \omega)$ . Use Proposition 9.1.VIII to verify that  $N_{\epsilon}$  is a random measure [indeed, by (d) and Definition 9.1.V, it is a simple point process].
- (f) For each  $\epsilon > 0$  the relation  $\mu_{\epsilon}(A) = \sum_{i=1}^{k(\epsilon)} u_i \delta_{x_i}(A)$  defines a measure. Then  $\lim_{\epsilon \downarrow 0} \mu_{\epsilon}(A) = \mu(A)$  if and only if  $\mu$  is purely atomic.

9.3.3 For every  $A \in \mathcal{B}([0,1])$  let  $\xi(A, \omega) = \int_A f(u, \omega) du$ , where for  $0 \leq u \leq 1$ ,  $f(u, \omega) = r(r+1)u(1-u)^{r-1}$  with probability  $1/[r(r+1)]$  for  $r = 1, 2, \dots$ . Note that  $\xi([0,1], \omega) = 1$  a.s., but there is no interval of positive length in the neighbourhood of 0 where  $\xi$  is ‘small’ a.s. Investigate whether (9.3.7) is satisfied.

9.3.4 *Absolutely continuous random distributions* [cf. Example 9.1(e)]. Construct a sequence  $\{F_n(\cdot)\}$  of distribution functions on  $(0, 1]$  as follows:  $F_n(0) = 0$ ,  $F_n(1) = 1$  ( $n = 1, 2, \dots$ ),  $F_1(\frac{1}{2}) = F_n(\frac{1}{2}) = U_{11}$ , and for  $k = 1, 3, \dots, 2^n - 1$ ,

$$\begin{aligned} F_n(k/2^n) &= (1 - U_{nk})F_{n-1}((k-1)/2^n) + U_{nk}F_{n-1}((k+1)/2^n) \\ &= F_{n+r}(k/2^n) \quad (n = 1, 2, \dots; r = 1, 2, \dots), \end{aligned}$$

where  $\mathcal{U}_n \equiv \{U_{nk} : k = 1, \dots, 2^n - 1\}$  is a family of i.i.d.  $(0, 1)$ -valued r.v.s with  $EU_{nk} = \frac{1}{2}$  and  $\sigma_n^2 = \text{var } U_{nk}$ , the families  $\{\mathcal{U}_n\}$  are mutually independent, and  $F_n(x)$  is obtained by linear interpolation between  $F_n(j_n(x)/2^n)$  and  $F_n((j_n(x) + 1)/2^n)$ , where  $j_n(x) = \text{largest integer} \leq 2^n x$ . With  $F_n(x)$  so defined on  $0 \leq x \leq 1$ , the derivative  $f_n(\cdot)$  of  $F_n(\cdot)$  is well defined except at  $x = a_{nj} \equiv j/2^n$  ( $j = 0, 1, \dots, 2^n$ ), where we adopt the convention that  $f(a_{nj}) = 1$  (all  $n$ , all  $a_{nj}$ ).

(a) Show that there is a d.f.  $F$  on  $0 \leq x \leq 1$  such that

$$\Pr\{F_n(x) \rightarrow F(x) \ (0 \leq x \leq 1) \text{ as } n \rightarrow \infty\} = 1.$$

(b) Provided the  $U_{nj}$  are sufficiently likely to be close to  $\frac{1}{2}$ , for which the condition  $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$  is sufficient, show that

$$\Pr\{f_n(x) \rightarrow f(x) \text{ for } 0 < x < 1\} = 1$$

for some density function  $f(\cdot)$  for which  $F(x) = \int_0^x f(u) du$  a.s. for  $0 < x < 1$ . Thus, the random d.f.  $F(\cdot)$  is a.s. absolutely continuous with respect to Lebesgue measure. [Hint: Let the r.v.  $W$  be uniformly distributed on  $[0, 1]$  and independent of  $\{\mathcal{U}_n\}$ . Show that  $\{f_n(W)\}$  is a martingale, and assuming  $\sum \sigma_n^2 < \infty$ , use the mean square martingale convergence theorem to deduce that  $f_n(W)$  converges a.s. and that  $F$  is the integral of its limit.]

(c) Investigate other conditions such as  $\sum_{n=1}^{\infty} \Pr\{|2U_{nk} - 1| < 1 - \epsilon\} = \infty$  for some  $\epsilon > 0$  or  $\liminf_{n \rightarrow \infty} \sigma_n^2 < \frac{1}{4}$  that may be sufficient to imply either that  $F(\cdot)$  is continuous on  $(0, 1)$  or else that  $F(\cdot)$  has jumps.

[Remarks: For constructions related to the above by Kraft (1964), see Dubins and Freedman (1967)—the random d.f.s they construct on  $[0, 1]$  are a.s. singular continuous—and Métivier (1971) where the construction leads to random d.f.s on  $[0, 1]$  that are a.s. absolutely continuous.]

9.3.5 Let the nonnegative set function  $\psi$  defined on the Borel subsets of a space  $\mathcal{X}$  be subadditive under refinement. Define a set function,  $\lambda_\psi$  say, much as in Definition 9.3.VIII. Show that if  $\lambda_\psi(A)$  is finite on bounded  $A \in \mathcal{B}_\mathcal{X}$ , then  $\lambda_\psi$  is a measure. [Hint: Check that  $\lambda_\psi$  is continuous on the empty set.]

9.3.6 *Korolyuk equation for purely atomic random measure*. Let  $N$  be a marked point process on  $\mathcal{X} \times (0, \infty)$  for which the ground process  $N_g$  has boundedly finite intensity measure  $\lambda$ . Denote by  $\{(x_i, \kappa_i)\}$  the points of a realization of  $N$ , and consider the random measure  $\xi_\mathcal{X} = \sum \kappa_i \delta_{x_i}$  [see also (6.4.6) and (9.1.5)]. Show that for each finite  $x > 0$  the set function

$$\lambda_\mathcal{X}(A) \equiv \sup_T \sum_{i=1}^{k_n} \mathcal{P}\{0 < \xi_\mathcal{X}(A_{ni}) \leq x\} \quad (\text{bounded } A \in \mathcal{B}_\mathcal{X}),$$

is an intensity measure for those points with marks  $\leq x$ , and that [cf. (9.3.17)]

$$E[\xi_\mathcal{X}(A)] = \int_0^\infty [\lambda(A) - \lambda_x(A)] dx, \quad \text{finite or infinite.}$$

9.3.7 Use equation (9.3.16) to show that a simple point process with boundedly finite intensity measure is ordinary.

9.3.8 Show that a Poisson process is simple if and only if it is ordinary. [Hint: Show that being simple is equivalent to the parameter measure being nonatomic.]

9.3.9 Let the point process  $N$  on  $\mathbb{R}_+$  have points located at  $\{n^2 U : n = 1, 2, \dots\}$ , where the r.v.  $U$  is uniformly distributed on  $(0, 1)$ . Show that for  $0 < h < 1$ ,

$$\frac{\mathcal{P}\{N(0, h] \geq k\}}{\mathcal{P}\{N(0, h] \geq 1\}} = \frac{1}{k^2}, \quad \frac{\mathcal{P}\{N(0, h] \geq k\}}{\lambda((0, h])} = \frac{6/\pi^2}{k^2}.$$

Conclude that the  $\lambda$ -a.e. constraint in Proposition 9.3.XV cannot be relaxed.

9.3.10 Suppose that the simple point process  $N$  on  $\mathcal{X}$  has a boundedly finite intensity measure  $\lambda(\cdot)$  that is absolutely continuous with respect to a measure  $\mu(\cdot)$  on  $\mathcal{X}$ , and that there is a version of the Radon–Nikodym derivative  $d\lambda/d\mu$  coinciding  $\mu$ -a.e. with a continuous function. Use the techniques of the proof of Proposition 9.3.XV to show that, in the notation of (9.3.22),

$$\frac{\mathcal{P}\{N(A_n(x)) \geq 1\}}{\mu(A_n(x))} \rightarrow \frac{d\lambda}{d\mu} \quad (n \rightarrow \infty) \text{ for } \mu\text{-a.e. } x.$$

Deduce in particular that when  $\mathcal{X} = \mathbb{R}^d$  and  $N$  is stationary,

$$\frac{\mathcal{P}\{N(A_n(x)) \geq 1\}}{\ell(A_n(x))} \rightarrow \text{const.} \quad \ell\text{-a.e. } x,$$

and then use stationarity both to identify the constant and to eliminate the  $\ell$ -a.e. condition (cf. Theorem 13.3.IV).

9.3.11 When  $N$  is a stationary mixed Poisson process on  $\mathbb{R}$  with mixing distribution function  $F(\cdot)$ ,

$$\mathcal{P}\{N(0, x] = k\} = \int_0^\infty e^{-\lambda x} \frac{(\lambda x)^k}{k!} dF(\lambda) \quad (0 < x < \infty, k = 0, 1, \dots).$$

Prove the following.

- (i)  $N$  is simple [because  $F(\infty-) = 1$ ].
- (ii) The intensity of  $N$  is finite or infinite with  $\lim_{y \rightarrow \infty} F_1(y)$ , where

$$F_1(y) = \int_0^y [1 - F(u)] du.$$

(iii) Whether the intensity is finite or not, the conditional probabilities at (9.3.20) converge to zero when  $F_1$  is slowly varying [i.e.,  $F_1(2y)/F_1(y) \rightarrow 1$  as  $y \rightarrow \infty$ ].

(iv)  $N$  is orderly when  $y[1 - F(y)] \rightarrow 0$  as  $y \rightarrow \infty$ .

(v)  $N$  is ordinary when  $\liminf_{y \rightarrow \infty} y[1 - F(y)] = 0$ .

[Hint: It follows from results in Daley (1982b) that these results are in fact necessary and sufficient. It is not difficult to find d.f.s  $F$  showing that none of the implications (iii)  $\Rightarrow$  (iv)  $\Rightarrow$  (v)  $\Rightarrow$  (i) can be reversed; for other examples see Exercise 3.3.2, Daley (1974, 1982a), and MKM (1978, pp. 68, 371), but note that MKM use the term orderly for what we have called ordinary.]

## 9.4. Functionals: Definitions and Basic Properties

In the study of families of independent random measures and point processes, transforms analogous to Laplace–Stieltjes transforms and probability generating functions of nonnegative and integer-valued r.v.s play a central role. In this section we outline a general setting in the context of random measures and give general characterization results, before extending the discussion in Chapter 5 of the properties of probability generating functionals (p.g.fl.s) of point processes. Then in Section 9.5 we discuss moment measures and some of their connections with functionals of random measures and point processes.

Let  $f$  be a Borel measurable function, defined on the same c.s.m.s. space as the random measure  $\xi$ . No special definitions are needed to introduce the integral

$$\xi_f = \int_{\mathcal{X}} f(x) \xi(dx) \equiv \int f d\xi, \quad (9.4.1)$$

for by assumption each realization of  $\xi$  is a boundedly finite Borel measure, and the usual theory of the Lebesgue integral applies on a realization-by-realization basis. In particular, if we introduce the space  $\text{BM}(\mathcal{X})$  of bounded measurable functions which vanish outside a bounded set in  $\mathcal{X}$ , then with probability 1 the integral exists and is finite. It is also a random variable; this can be seen by first taking  $f$  to be the indicator function of a bounded Borel set and then applying the usual approximation arguments using linear combinations and monotone limits. Now the class of r.v.s is closed under both these operations, and  $\xi_f = \xi(A)$  is certainly an r.v. when  $f = I_A$ , so it follows that  $\xi_f$  is a (proper) r.v. for any  $f \in \text{BM}(\mathcal{X})$ .

The study of such random integrals, which are evidently linear in  $f$ , links the theory of random measures with a whole hierarchy of theories of random linear functionals, of which the theory of random distributions is perhaps the most important, and is relevant in discussing second-order properties (see Chapter 8). We pause, therefore, to give a brief introduction to such general theories.

Given any linear space  $\mathcal{U}$  on which the notions of addition and scalar multiplication are defined, the concept of a linear functional, that is, a mapping  $\gamma$  from  $\mathcal{U}$  into the real line satisfying

$$\gamma(\alpha u + \beta v) = \alpha \gamma(u) + \beta \gamma(v) \quad (\alpha, \beta \in \mathbb{R}; u, v \in \mathcal{U}) \quad (9.4.2)$$

makes sense, and we may consider the space of all such linear functionals on a given  $\mathcal{U}$ . Furthermore, if  $\mathcal{U}$  has a topology conformable with the linear operations (i.e., one making these continuous), we may consider the smaller space of continuous linear functionals on  $\mathcal{U}$ . Many different possibilities arise, depending on the choice of  $\mathcal{U}$  and of the topology on  $\mathcal{U}$  with respect to which continuity is defined.

With any such choice there are several ways in which we may associate a random structure with the given space of linear functionals. Of these we

distinguish two general classes, which we call *strict sense* and *broad sense* random linear functionals.

A natural  $\sigma$ -algebra in the space  $\mathcal{C}_{\mathcal{U}}$  of continuous linear functionals on  $\mathcal{U}$  is the smallest  $\sigma$ -algebra with respect to which the mappings  $\gamma: u \mapsto \gamma(u)$  are measurable with respect to each  $u \in \mathcal{U}$ . Endowing  $\mathcal{C}_{\mathcal{U}}$  with this  $\sigma$ -algebra, we may define a *strict sense* random linear functional on  $\mathcal{U}$  as a measurable mapping  $\Gamma(\cdot)$  from a probability space into  $\mathcal{C}_{\mathcal{U}}$ . This ensures, as a minimal property, that  $\Gamma(u)$  is a random variable for each  $u \in \mathcal{U}$ . On the other hand, it is often difficult to determine conditions on the distributions of a family of r.v.s  $\{\Gamma_u\}$ , indexed by the elements of  $\mathcal{U}$ , that will allow us to conclude that the family  $\{\Gamma_u\}$  can be identified a.s. with a random functional  $\Gamma(u)$  in this strict sense. The same difficulty arises if we attempt to define a random linear functional as a probability distribution on  $\mathcal{C}_{\mathcal{U}}$ . How can we tell, from the fid distributions or otherwise, whether such a distribution does indeed correspond to such an object? Even in the random measure case this discussion is not trivial, and in many other situations it remains unresolved.

The alternative, *broad sense*, approach is to accept that a random linear functional cannot be treated as anything more than a family of r.v.s indexed by the elements of  $\mathcal{U}$  and to impose on this family appropriate linearity and continuity requirements. Thus, we might require that

$$\Gamma_{\alpha u + \beta v} = \alpha \Gamma_u + \beta \Gamma_v \quad \text{a.s.} \quad (9.4.2')$$

and, if  $u_n \rightarrow u$  in the given topology on  $\mathcal{U}$ ,

$$\xi_{u_n} \rightarrow \xi_u \quad \text{a.s.} \quad (9.4.3)$$

or, at (9.4.3), we could merely use convergence in probability or in quadratic mean. If  $\Gamma_u = \Gamma(u)$  for all  $u \in \mathcal{U}$ , where  $\Gamma(\cdot)$  is a strict sense random linear functional, then of course both (9.4.2') and (9.4.3) hold a.s. Dudley (1969) reviews some deeper results pertaining to random linear functionals.

EXAMPLE 9.4(a) *Generalized random processes (random Schwartz distributions)*. Take  $\mathcal{X} = \mathbb{R}$ , and let  $\mathcal{U}$  be the space of all infinitely differentiable functions on  $\mathbb{R}$  that vanish outside some finite interval; that is,  $\mathcal{U}$  is a space of *test functions* on  $\mathbb{R}$ . Introduce a topology on  $\mathcal{U}$  by setting  $u_n \rightarrow u$  if and only if the  $\{u_n\}$  vanish outside some common finite interval, and for all  $k \geq 0$ , the  $k$ th derivatives  $\{u_n^{(k)}\}$  converge uniformly to  $u^{(k)}$ . Then  $\mathcal{C}_{\mathcal{U}}$ , the space of all functionals on  $\mathcal{U}$  satisfying (i) the linearity condition (9.4.2), and (ii) the continuity condition  $\gamma(u_n) \rightarrow \gamma(u)$  whenever  $u_n \rightarrow u$  in  $\mathcal{U}$ , is identified with the space of *generalized functions*, or more precisely *Schwartz distributions*. Any ordinary continuous function  $g$  defines such a distribution through the relation

$$\gamma(u) = \int_{-\infty}^{\infty} g(x)u(x) dx,$$

the continuity condition (ii) following from the boundedness of  $g$  on the finite interval outside which the  $u_n$  vanish and the uniform convergence of the  $u_n$  themselves. Similarly, any bounded finite measure  $G$  on  $\mathbb{R}$  defines a distribution by the relation

$$\gamma(u) = \int_{\mathbb{R}} u(x) G(dx).$$

However, many further types of Schwartz distribution are possible, relating, for example, to linear operations on the derivatives of  $u$ .  $\square$

The corresponding strict sense theory has been relatively little used, but the broad sense theory plays a central role in the second-order theory of stationary generalized processes, of which the second-order theory of stationary point processes and random measures forms a special case. A similar theory for random generalized fields can be developed by taking test functions on  $\mathbb{R}^d$  in place of test functions on  $\mathbb{R}$ . Gelfand and Vilenkin (1964) and Yaglom (1961) give systematic treatments of these broad sense theories.

A natural tool for handling any type of random linear functional is the *characteristic functional*, defined by

$$\Phi_{\Gamma}[g] = E[\exp(i\Gamma_g)] \quad (g \in \mathcal{U}), \quad (9.4.4)$$

where  $\Gamma_g$  is a random linear functional (strict or broad sense) on  $\mathcal{U}$ . It can be described as the characteristic function  $E(e^{is\Gamma_g})$  of  $\Gamma_g$ , evaluated at the arbitrary value  $s = 1$  and treated as a function of  $g$  rather than  $s$ .

EXAMPLE 9.4(b) *Gaussian measures on Hilbert space*. Random variables taking their values in a Hilbert space  $\mathcal{H}$  can be placed within the general framework of random linear functionals by taking advantage of the fact that the space of continuous linear functionals on a Hilbert space can be identified with the given Hilbert space itself. In this interpretation  $\Gamma(u)$  is identified with the inner product  $\langle \Gamma, u \rangle$  for  $u \in \mathcal{H}$ .

When  $\mathcal{H}$  is finite-dimensional, the characteristic functional reduces to the multivariate characteristic function

$$E[\exp(i\langle \Gamma, u \rangle)] = E\left[\exp\left(i \sum_{k=1}^n \Gamma_k u_k\right)\right] = \phi(u_1, \dots, u_n) = \phi(u),$$

where  $\Gamma_k$  and  $u_k$  are the coordinates of  $\Gamma$  and  $u$ . In this case a Gaussian measure is just the ordinary multivariate normal distribution: setting the mean terms equal to zero for simplicity, the characteristic function has the form

$$\phi(u) = \exp\left(-\frac{1}{2}u^{\top}Au\right),$$

where  $u^{\top}Au$  is the quadratic form associated with the nonnegative definite (positive semidefinite) symmetric matrix  $A$ . This suggests the generalization to infinite-dimensional Hilbert space of

$$\Phi[u] = E[\exp(i\Gamma_u)] = \exp(-\frac{1}{2}\langle u, Au \rangle), \quad (9.4.5)$$

where  $A$  is now a positive definite self-adjoint linear operator. The finite-dimensional distributions of  $\Gamma_{u_1}, \dots, \Gamma_{u_n}$  for arbitrary  $u_1, \dots, u_n$  in  $\mathcal{H}$  can be determined by setting  $\sum_{k=1}^n s_k u_k$  in place of  $u$  in (9.4.5): they are of multivariate normal form with  $n \times n$  covariance matrix having elements  $\langle u_i, A u_j \rangle$ . From this representation the consistency conditions are readily checked, as well as the linearity requirements (9.4.2). If  $u_n \rightarrow u$  (i.e.,  $\|u_n - u\| \rightarrow 0$ ), it follows from the boundedness of  $A$  that  $\langle (u_n - u), A(u_n - u) \rangle \rightarrow 0$  and hence that  $\Gamma_{u_n} \rightarrow \Gamma_u$  in probability and in quadratic mean.

These arguments suffice to show that (9.4.5) defines a broad sense random linear functional on  $\mathcal{H}$ , but they are not sufficient to imply that (9.4.5) defines a strict sense random linear functional. For this, more stringent requirements are needed; these have their roots in the fact that a probability measure on  $\mathcal{H}$  must be *tight*, and hence in a loose sense approximately concentrated on a finite-dimensional subset of  $\mathcal{H}$ . It is known [see, e.g., Parthasarathy (1967, Chapter 8)] that the necessary and sufficient conditions for (9.4.5) to be the characteristic functional of a strict sense random linear functional on  $\mathcal{H}$  [so that we can write  $\Gamma_u = \Gamma(u) = \langle \Gamma, u \rangle$ ], or, equivalently, of a probability measure on  $\mathcal{H}$  itself, is that the operator  $A$  be of Hilbert–Schmidt type. In this case the characteristic functional has the more special form

$$\Phi[u] = \exp \left( -\frac{1}{2} \sum \lambda_k (\langle h_k, u \rangle)^2 \right),$$

where  $\{h_k\}$  is a complete set of eigenvectors for  $A$ ,  $\{\lambda_k\}$  is the set of corresponding eigenvalues, and  $\sum \lambda_k^2 < \infty$ .  $\square$

Returning to the random measure context, let us first note that random measures can just as easily be characterized by the values of the integrals (9.4.1) as they can by their evaluations on Borel sets; indeed, the latter are just a special case of the former when  $f$  is an indicator function. It follows at once that  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$  is the smallest  $\sigma$ -algebra with respect to which the random integrals  $\int f d\xi$  are measurable for each  $f \in \text{BM}(\mathcal{X})$ , and that a mapping  $\xi$  from a probability space into  $\mathcal{M}_{\mathcal{X}}^{\#}$  is a random measure if and only if  $\int f d\xi$  is a random variable for each  $f \in \text{BM}(\mathcal{X})$  [a smaller class of functions  $f$  suffices: Kallenberg (1983a, Exercise 3.1) indicates a stronger version of this result]. A more useful result is the following analogue of Proposition 9.1.VIII, the proof of which is left to Exercise 9.4.1.

**Proposition 9.4.I.** *Let  $\{\xi_f\}$  be a family of random variables, indexed by the elements  $f$  of  $\text{BM}(\mathcal{X})$ . Then there exists a random measure  $\xi$  such that*

$$\xi_f = \int f d\xi \quad \text{a.s.}$$

*if and only if*

- (i)  $\xi_{\alpha f + \beta g} = \alpha \xi_f + \beta \xi_g$  a.s. for all scalars  $\alpha, \beta$  and  $f, g \in \text{BM}(\mathcal{X})$ ; and
- (ii)  $\xi_{f_n} \rightarrow \xi_f$  a.s. as  $n \rightarrow \infty$  for all monotonically converging nonnegative sequences  $\{f_n\} \subset \text{BM}(\mathcal{X})$  (i.e.  $f_n \geq 0$  and  $f_n \uparrow f$ ).



Conditions (i) and (ii) are, of course, just the conditions (9.4.2) and (9.4.3) in a form suitable for random measures; the importance of the proposition is that it implies that *the broad and strict sense approaches are equivalent for random measures*.

From this point it is easy to move to a characterization of the fidi distributions of the integrals  $\int f d\xi$  for a random measure; we state this in the form of a characterization theorem for *characteristic functionals*, which we define by

$$\Phi_\xi[f] = E[\exp(i \int f d\xi)] \quad (f \in \text{BM}(\mathcal{X})), \quad (9.4.6)$$

as the appropriate special form of (9.4.4).

**Theorem 9.4.II.** *Let the functional  $\Phi[f]$  be real- or complex-valued, defined for all  $f \in \text{BM}(\mathcal{X})$ . Then  $\Phi$  is the characteristic functional of a random measure  $\xi$  on  $\mathcal{X}$  if and only if*

(i) *for every finite family  $f_1, \dots, f_n$  of functions  $f_k \in \text{BM}(\mathcal{X})$ , the function*

$$\phi_n(f_1, \dots, f_n; s_1, \dots, s_n) = \Phi \left[ \sum_{k=1}^n s_k f_k \right] \quad (9.4.7)$$

*is the multivariate characteristic function of proper random variables  $\xi_{f_1}, \dots, \xi_{f_n}$ , which are nonnegative a.s. when the functions  $f_1, \dots, f_n$  are nonnegative;*

(ii) *for every sequence  $\{f_n\} \subset \text{BM}(\mathcal{X})$  with  $f_n \geq 0$  and  $f_n \uparrow f$  pointwise,*

$$\Phi[f_n] \rightarrow \Phi[f]; \text{ and} \quad (9.4.8)$$

(iii)  $\Phi(\mathbf{0}) = 1$  *where  $\mathbf{0}$  here denotes the zero function in  $\text{BM}(\mathcal{X})$ .*

*Moreover when the conditions are satisfied, the functional  $\Phi$  uniquely determines the distribution of  $\xi$ .*

PROOF. If  $\xi$  is a random measure, conditions (i) and (ii) are immediate, and imply that the fidi distributions of  $\xi$ , and hence the distribution of  $\xi$  itself, are uniquely determined by  $\xi$ . If  $f_n \rightarrow f$  pointwise, and the  $f_n$  are either monotonic or bounded by a common element of  $\text{BM}(\mathcal{X})$ , it follows from the Lebesgue convergence theorems that for each realization,

$$\xi_{f_n} = \int f_n d\xi \rightarrow \int f d\xi = \xi_f,$$

so that  $\xi_{f_n} \rightarrow \xi_f$  a.s. Equation (9.4.8) follows from a further application of the dominated convergence theorem, using  $|\exp(i \int f_n d\xi)| \leq 1$ .

Suppose next that conditions (i)–(iii) are satisfied. Condition (i) subsumes both Kolmogorov consistency conditions for the fidi distributions of the r.v.s  $\xi_f$  defined by (9.4.7) for  $f \in \text{BM}(\mathcal{X})$ . For example, in characteristic function

terms, the requirement of marginal consistency reduces to the trivial verification that

$$\Phi\left(\sum_{k=1}^{n-1} s_i f_i + 0 \cdot f_n\right) = \Phi\left(\sum_{k=1}^{n-1} s_i f_i\right).$$

Thus, we may assume the existence of a jointly distributed family of r.v.s  $\{\xi_f, f \in \text{BM}(\mathcal{X})\}$ .

Condition (i) also implies the linearity property (i) of Proposition 9.4.I, for the condition  $\xi_{f_3} = \xi_{f_1} + \xi_{f_2}$  a.s. is equivalent to the identity

$$\Phi[s_1 f_1 + s_2 f_2 + s_3 f_3] = \Phi[(s_1 + s_3) f_1 + (s_2 + s_3) f_2],$$

which will certainly be valid if  $f_3 = f_1 + f_2$ . A similar argument applies when scalar multipliers  $\alpha, \beta$  are included.

Finally, condition (ii) of the theorem implies that the distribution of  $\xi_{f-f_n}$  approaches the distribution degenerate at 0, and hence that  $\xi_{f-f_n}$  converges in probability to zero. From the linearity property of the  $\xi_f$  we deduce that

$$\xi_{f_n} \rightarrow \xi_f \quad \text{in probability.} \quad (9.4.9)$$

However, because we assume in condition (ii) that the sequence  $\{f_n\}$  is monotonic increasing, it follows from condition (i) that the sequence  $\{\xi_{f_n}\}$  is a.s. monotonic increasing. Because  $\xi_{f_n} \leq \xi_f$  a.s. by similar reasoning,  $\xi_{f_n}$  converges a.s. to a proper limit r.v.  $X$  say. But then (9.4.9) implies that  $X = \xi_f$  a.s., so condition (ii) of Proposition 9.4.I is also satisfied. The existence of a random measure with the required properties now follows from Proposition 9.4.I and the part of the theorem already proved.  $\square$

Variants of condition (ii) above are indicated in Exercise 9.4.2.

As described above, the characteristic functional emerges naturally from the context of random linear functionals, but in the study of random measures and point processes, which are nonnegative by definition, it is enough to use real variable counterparts. The *Laplace functional* is defined for  $f \in \text{BM}_+(\mathcal{X})$ , the space of all nonnegative  $f \in \text{BM}(\mathcal{X})$ , by

$$L_\xi[f] = E\left[\exp\left(-\int f d\xi\right)\right] \quad (f \in \text{BM}_+(\mathcal{X})) \quad (9.4.10)$$

[this is the same as (6.1.8)]. An exact counterpart of Theorem 9.4.II holds for Laplace functionals; see Exercise 9.4.3 for some detail. Observe, in particular, that the theorem implies that the distribution of a random measure is completely determined by the Laplace functional.

The class  $\text{BM}_+(\mathcal{X})$  of functions on which the Laplace functional is defined can be restrictive for some applications, as, for example, in discussing the mixing properties of cluster processes in Section 12.3. We therefore define the *extended Laplace functional* for use in such contexts much as for  $L$  but on the space of functions  $\overline{\text{BM}}_+(\mathcal{X})$  consisting of functions  $f$  that are expressible as

the monotone limit of an increasing sequence of functions  $\{f_n\} \subset \text{BM}_+(\mathcal{X})$ . Then by the monotone convergence theorem,

$$\int_{\mathcal{X}} f_n(x) \xi(dx) \uparrow \int_{\mathcal{X}} f(x) \xi(dx) \quad \text{a.s.},$$

whether the limit is finite or infinite, and then by dominated convergence,

$$L_\xi[f_n] \rightarrow L_\xi[f] \equiv \mathbb{E}[\exp(-\int_{\mathcal{X}} f d\xi)] \quad (f \in \overline{\text{BM}}_+(\mathcal{X})), \quad (9.4.11)$$

where we use  $L_\xi$  or, more briefly,  $L$ , both for the functional as originally defined and for its extension to  $\overline{\text{BM}}_+(\mathcal{X})$ .

The extended Laplace functional, as defined over functions  $f \in \overline{\text{BM}}_+(\mathcal{X})$ , has continuity properties as below, but it need not be continuous for monotone sequences  $\{f_n\} \subset \overline{\text{BM}}_+(\mathcal{X})$ : take  $f_\epsilon(x) = \epsilon$  (all  $x \in \mathcal{X}$ ),  $\xi(\mathcal{X}) = \infty$  a.s.; then for all  $\epsilon > 0$ ,  $L[f_\epsilon] = 0 \neq L[\mathbf{0}] = 1$ .

**Proposition 9.4.III.** *The extended Laplace functional  $L[\cdot]$  satisfies*

$$L[f_n] \rightarrow L[f] \quad (f_n, f \in \overline{\text{BM}}_+(\mathcal{X}))$$

whenever either

- (a)  $f_n(x) \uparrow f(x)$ , or
- (b)  $f_n(x) \rightarrow f(x)$  and there exists a nonnegative measurable function  $\Delta(\cdot)$  such that  $\int_{\mathcal{X}} \Delta(x) \xi(dx) < \infty$  a.s. and  $|f_n(x) - f(x)| \leq \Delta(x)$  for all sufficiently large  $n$ .

PROOF. If  $f_n \uparrow f$ , then it is easy to construct a monotone sequence of functions  $\{f'_n\} \subset \text{BM}_+(\mathcal{X})$  with  $f'_n(x) \uparrow f(x)$ , and (9.4.11) holds by definition.

In the other case, we have for all  $n \geq$  some  $n_0$ ,

$$f_n(x) \leq f(x) + \Delta(x) \leq f_{n_0}(x) + 2\Delta(x),$$

so  $\int_{\mathcal{X}} f_n(x) \xi(dx) \leq \int_{\mathcal{X}} [f_{n_0}(x) + 2\Delta(x)] \xi(dx) \leq \infty$  a.s., and by dominated convergence applied to the sequence  $\{f_n(\cdot)\}$ ,

$$\int_{\mathcal{X}} f_n(x) \xi(dx) \rightarrow \int_{\mathcal{X}} f(x) \xi(dx) < \infty \quad \text{a.s.}$$

A second appeal to dominated convergence now implies (9.4.11). □

Under conditions (b) here, it follows that  $L[\epsilon f] \rightarrow 1 = L[\mathbf{0}]$  as  $\epsilon \rightarrow 0$ , so for such  $f_n$  and  $f \in \text{BM}_+(\mathcal{X})$ , the extended Laplace functional has all the properties of the ordinary Laplace functional (9.4.10).

Because random measures are inherently nonnegative, the Laplace functional is often the most appropriate tool to use in handling random measures, just as the Laplace–Stieltjes transform is generally the most useful tool for handling nonnegative random variables. It is only when the r.v. is integer-valued, or, in our context, when the random measure is a point process, that

there are advantages in moving to the probability generating function (p.g.f.) or its counterpart the probability generating functional (p.g.fl). We have already discussed the p.g.fl. of finite point processes in Chapter 5, defined there on the class  $\mathcal{U}$  of complex-valued Borel measurable functions  $\zeta$  satisfying the condition  $|\zeta(x)| \leq 1$  (all  $x$ ). But just as the Laplace functional is discussed more advantageously as a functional over a space of real-valued functions, so is the p.g.fl. discussed better over a narrower class of functions. We largely follow Westcott's (1972) general treatment.

**Definition 9.4.IV.**  $\mathcal{V}(\mathcal{X})$  denotes the class of all real-valued Borel functions  $h$  defined on the c.s.m.s.  $\mathcal{X}$  with  $1 - h$  vanishing outside some bounded set and satisfying

$$0 \leq h(x) \leq 1 \quad (\text{all } x \in \mathcal{X}).$$

$\mathcal{V}_0(\mathcal{X})$  is the subset of  $h \in \mathcal{V}(\mathcal{X})$  satisfying  $\inf_{x \in \mathcal{X}} h(x) > 0$ .

$\bar{\mathcal{V}}(\mathcal{X})$  is the space of functions  $h$  expressible as limits of monotone sequences  $h_n \in \mathcal{V}(\mathcal{X})$ .

Extending Definition 5.5.I, the probability generating functional (p.g.fl.) of a (general) point process  $N$  on the c.s.m.s.  $\mathcal{X}$  is defined by

$$G[h] \equiv G_N[h] = \mathbb{E} \left[ \exp \left( \int_{\mathcal{X}} \log h(x) N(dx) \right) \right] \quad (h \in \mathcal{V}(\mathcal{X})). \quad (9.4.12)$$

Because a point process is a.s. finite on the bounded set where  $1 - h$  does not vanish, the exponential of the integral at (9.4.12) can legitimately be written in the product form

$$G_N[h] = \mathbb{E} \left( \prod_i h(x_i) \right), \quad (9.4.13)$$

where the product is taken over the points of each realization of  $N$  (recall Proposition 9.1.V), with the understanding that it takes the value zero if  $h(x_i) = 0$  for any  $x_i$ , and unity if there are no points of  $N$  within the support of  $1 - h$ . If  $h$  is such that  $-\log h \in \text{BM}_+(\mathcal{X})$ , then the equation

$$G[h] = L_N[-\log h] \quad (9.4.14)$$

relates the p.g.fl. to the Laplace functional of the point process [cf. (9.4.10)]. Indeed,  $-\log h \in \text{BM}_+(\mathcal{X})$  implies that the values of  $h$  lie within a closed subset of  $(0, 1]$ , so because the distribution of a random measure is determined by all  $f \in \text{BM}_+(\mathcal{X})$ , the distribution of a point process is determined by all  $h \in \mathcal{V}_0(\mathcal{X})$ . Although results for point processes need to be proved only with this more restricted class of functions, it is only in our discussion of mixing properties of cluster processes (see Proposition 12.3.IX) that we need the constraint, so mostly we use  $\mathcal{V}(\mathcal{X})$ . Note that if  $\{h_n(x)\}$  is a pointwise convergent sequence of functions  $\in \mathcal{V}_0(\mathcal{X})$  with the support of each  $1 - h_n$  contained by some fixed bounded set, then the pointwise limit,  $h$  say, has  $h \in \mathcal{V}(\mathcal{X})$  but not necessarily  $h \in \mathcal{V}_0(\mathcal{X})$ . To this extent,  $\mathcal{V}(\mathcal{X})$  is a simpler class with which to work.

By putting additional restrictions on the point process, the p.g.fl. can be defined for more general classes of functions  $h$ . For example, if the expectation measure  $M$  of  $N$  exists [cf. around (9.5.1) below], and nonnegative  $h$  is so chosen that

$$\int_{\mathcal{X}} |\log h(x)| M(dx) < \infty, \quad (9.4.15)$$

then the integral in (9.4.12) converges a.s. to a finite quantity, and the expectation exists.

The p.g.fl. of a point process can be characterized as in Theorem 9.4.V. It is an exact analogue of Theorem 9.4.II, so proof is left to the reader.

**Theorem 9.4.V.** *Let the functional  $G[h]$  be real-valued, defined for all  $h \in \mathcal{V}(\mathcal{X})$ . Then  $G$  is the p.g.fl. of a point process  $N$  on  $\mathcal{X}$  if and only if*

(i) *for every  $h$  of the form*

$$1 - h(x) = \sum_{k=1}^n (1 - z_k) I_{A_k}(x),$$

*where the bounded Borel sets  $A_1, \dots, A_n$  are disjoint and  $|z_i| \leq 1$ , the p.g.fl.  $G[h]$  reduces to the joint p.g.f.  $P_n(A_1, \dots, A_n; z_1, \dots, z_n)$  of an  $n$ -dimensional integer-valued random variable;*

(ii) *for every sequence  $\{h_n\} \subset \mathcal{V}(\mathcal{X})$  with  $h_n \downarrow h$  pointwise,  $G[h_n] \rightarrow G[h]$  whenever  $1 - h$  has bounded support; and*

(iii)  *$G[\mathbf{1}] = 1$  where  $\mathbf{1}$  denotes the function identically equal to unity in  $\mathcal{X}$ .*

*Moreover, when these conditions are satisfied, the functional  $G$  uniquely determines the distribution of  $N$ .*

Variants on the continuity condition (ii) are again possible, although more is needed than just pointwise convergence (see Exercise 9.4.5). Indeed, we shall have a need for the *extended p.g.fl.*  $G[\cdot]$  defined by analogy with the extended Laplace functional at (9.4.11) as

$$G[h] \equiv \mathbb{E} \left[ \exp \left( \int_{\mathcal{X}} \log h(x) N(dx) \right) \right] \quad (h \in \bar{\mathcal{V}}(X))$$

[see Definition 9.4.IV for  $\bar{\mathcal{V}}(X)$ ]. Further details are given in Exercise 9.4.6.

**EXAMPLE 9.4(c) Poisson and compound Poisson processes.** The form of the p.g.fl. for the Poisson process is already implicit in the form of the p.g.f. obtained in Chapter 2 and its multivariate extension [see (2.4.5) and (9.2.9)]

$$\Pi_k(A_1, \dots, A_k; z_1, \dots, z_k) = \exp \left( - \sum_{j=1}^k (1 - z_j) \mu(A_j) \right), \quad (9.4.16)$$

where  $A_1, \dots, A_k$  are disjoint and  $\mu(\cdot)$  is the parameter measure of the process. Writing  $h(x) = 1 - \sum_{j=1}^k (1 - z_j) I_{A_j}(x)$ , so that  $h(x) = z_j$  on  $A_j$  and  $= 1$  outside the union of all the  $A_j$ , (9.4.16) is expressible as

$$G[h] = \exp \left( - \int_{\mathcal{X}} [1 - h(x)] \mu(dx) \right), \quad (9.4.17)$$

which is evidently of the required form for the p.g.fl.

The following heuristic derivation of (9.4.17) also throws light on the character of the p.g.fl. Suppose that the support of  $1 - h(x)$  is partitioned into small subsets  $\Delta A_i$  in each of which is a 'representative point'  $x_i$ . Then, approximately,

$$\int_{\mathcal{X}} \log h(x) N(dx) \approx \sum_i \log h(x_i) N(\Delta A_i),$$

where because the r.v.s  $N(\Delta A_i)$  are independent (by assumption),

$$\begin{aligned} \mathbb{E} \left[ \exp \left( \int_{\mathcal{X}} \log h(x) N(dx) \right) \right] &\approx \mathbb{E} \left[ \prod_i \exp \left( \log h(x_i) N(\Delta A_i) \right) \right] \\ &= \prod_i \mathbb{E} [h(x_i)^{N(\Delta A_i)}] = \prod_i \exp \left( -[1 - h(x_i)] \mu(\Delta A_i) \right) \\ &\approx \exp \left( - \int_{\mathcal{X}} [1 - h(x)] \mu(dx) \right). \end{aligned}$$

The corresponding expression for the p.g.fl. of a compound Poisson process, understood in the narrow sense of Section 2.4 as a nonorderly point process, is

$$\begin{aligned} G[h] &= \exp \left( - \int_{\mathcal{X}} [1 - \Pi_{h(x)}(x)] \mu(dx) \right) \\ &= \exp \left( - \int_{\mathcal{X}} \sum_n \pi_n(x) (1 - [h(x)]^n) \mu(dx) \right); \end{aligned} \quad (9.4.18)$$

this reduces to (2.4.5) for the univariate p.g.f. when  $h(x) = 1 - (1 - z)I_A(x)$ . It is not difficult to verify that  $G$  satisfies the conditions of Theorem 9.4.V, and therefore it represents the p.g.fl. of a point process which enjoys the complete independence property discussed further in Section 10.1. Indeed, it is not difficult to check from the representation in (9.4.18) that the compound Poisson process can be characterized as a completely random measure which has no drift component, is free of fixed atoms, and has random atoms of positive integral mass.

The generalized compound Poisson process described in Section 6.4 corresponds to a marked Poisson process with independent nonnegative marks. As a point process on  $\mathcal{X} \times \mathcal{K}$  (with  $\mathcal{K} = \mathbb{R}_+$ ) it has a Laplace functional of the form

$$L[f] = \exp \left( - \int_{\mathcal{X}} \int_{\mathcal{K}} \left( 1 - e^{-f(x, \kappa)} \right) \pi(d\kappa) \mu(dx) \right) \quad (f \in \text{BM}_+(\mathcal{X} \times \mathcal{K})).$$

Here the interpretation as a Poisson process on the product space is immediately evident. Furthermore, when  $\mathcal{K} = \mathbb{R}_+$ , the cumulative process  $\xi(A) = \sum_{i: x_i \in A} \kappa_i$  has the Laplace functional, now defined over  $f \in \text{BM}_+(\mathcal{X})$ ,

$$L[f] = \mathbb{E} \left( e^{-\sum_i \kappa_i f(x_i)} \right) = \exp \left( - \int_{\mathcal{X}} \int_{\mathcal{K}} \left( 1 - e^{-\kappa f(x)} \right) \pi(d\kappa) \mu(dx) \right).$$

This exhibits the process as a completely random measure (Definition 10.1.I) with neither any fixed atoms nor a drift component.  $\square$

EXAMPLE 9.4(d) *Mixed Poisson process.* Referring to (9.4.16), denote the fidi distributions of a Poisson process by  $P_k(\cdot \mid \mu)$  for short. Then  $N$  is a mixed Poisson process when for some r.v.  $\Lambda$  and boundedly finite measure  $\mu$  its fidi distributions  $P_k(\cdot)$  are given by

$$P_k(\cdot) = E[P_k(\cdot \mid \Lambda\mu)], \quad (9.4.19)$$

the expectation being with respect to  $\Lambda$ . Write  $L(s) = E(e^{-s\Lambda})$  [ $\text{Re}(s) \geq 0$ ] for the Laplace–Stieltjes transform of  $\Lambda$ . It then follows from (9.4.19) and (9.4.17) that the p.g.fl. of a mixed Poisson process is given by

$$G[h] = E\left[\exp\left(-\int_{\mathcal{X}} [1-h(x)] \Lambda \mu(dx)\right)\right] = L\left(\int_{\mathcal{X}} [1-h(x)] \mu(dx)\right). \quad \square$$

We have already remarked that one of the most important properties of the transforms of r.v.s is the simplification they afford in handling problems involving sums of independent r.v.s. The summation operator for random measures is defined by

$$(\xi_1 + \xi_2)(A) = \xi_1(A) + \xi_2(A) \quad (\text{all } A \in \mathcal{B}_{\mathcal{X}}), \quad (9.4.20)$$

and it is both obvious and important that it extends the notion of *superposition* of point processes. Note that (9.4.20) has the equivalent form

$$\int f d(\xi_1 + \xi_2) = \int f d\xi_1 + \int f d\xi_2 \quad (\text{all } f \in \text{BM}_+(\mathcal{X})). \quad (9.4.20')$$

Now suppose that  $\{\xi_i: i = 1, 2, \dots\}$  is an infinite sequence of random measures, each defined on  $(\Omega, \mathcal{E}, \mathcal{P})$ , such that

$$\zeta(A) \equiv \sum_{i=1}^{\infty} \xi_i(A) \quad (9.4.21)$$

is a.s. finite on all bounded  $A \in \mathcal{B}_{\mathcal{X}}$ . It is well known and easy to check that a countable sum of measures is again a measure. Thus,  $\zeta(\cdot)$  is a boundedly finite measure, at least on the  $\omega$  set where the  $\xi_i$  are simultaneously measures, which set has probability 1 by assumption and the fact that only a countable family is involved. Redefining  $\zeta$  to be zero on the complementary  $\omega$  set of  $\mathcal{P}$ -measure zero, and observing that a countable sum of r.v.s is again a r.v., we obtain a mapping from  $(\Omega, \mathcal{E}, \mathcal{P})$  into  $\mathcal{M}_{\mathcal{X}}^{\#}$  satisfying the condition of Corollary 9.1.IX and which is therefore a random measure. Thus, we have the following lemma.

**Lemma 9.4.VI.**  $\zeta(\cdot)$  defined at (9.4.21) is a random measure if and only if the infinite sum at (9.4.21) converges for all bounded  $A \in \mathcal{B}_{\mathcal{X}}$ .

No new concepts arise in the following definition of independence of two random measures; it extends to the mutual independence of both finite and infinite families of random measures in the usual way.

**Definition 9.4.VII.** *The random measures  $\xi_1$  and  $\xi_2$  are independent when they are defined on a common space  $(\Omega, \mathcal{E}, \mathcal{P})$  and are such that  $\mathcal{P}(\mathcal{F}_1 \cap \mathcal{F}_2) = \mathcal{P}(\mathcal{F}_1)\mathcal{P}(\mathcal{F}_2)$  for all finite families  $\mathcal{F}_i$  of events defined on  $\xi_i$  ( $i = 1, 2$ ).*

Let  $\xi_i$  have characteristic functional  $\Phi_i$  and Laplace functional  $L_i$ . By writing  $\zeta_n = \xi_1 + \cdots + \xi_n$ , the following assertions are simple consequences of the definitions and Lemma 9.4.VI, and can be proved by methods exploited already (see Exercise 9.4.7).

**Proposition 9.4.VIII.** *When the random measures  $\xi_1, \xi_2, \dots$  are mutually independent, the sum  $\zeta_n = \xi_1 + \cdots + \xi_n$  has the characteristic functional*

$$\Phi_{\zeta_n}[f] = \prod_{i=1}^n \Phi_i[f] \quad (\text{all } f \in \text{BM}(\mathcal{X})) \quad (9.4.22a)$$

and Laplace functional

$$L_{\zeta_n}[f] = \prod_{i=1}^n L_i[f] \quad (\text{all } f \in \text{BM}_+(\mathcal{X})). \quad (9.4.22b)$$

$L_{\zeta_n}[f]$  converges as  $n \rightarrow \infty$  to a nonzero limit  $L_f$  for each  $f \in \text{BM}_+(\mathcal{X})$  if and only if the infinite sum at (9.4.21) is finite on bounded  $A \in \mathcal{B}_{\mathcal{X}}$ , and then  $L_f$  is the Laplace functional of the random measure  $\zeta$  at (9.4.21).

The analogue of this result for the p.g.fl.  $G$  of the superposition of the independent point processes  $N_1, \dots, N_n$  with p.g.fl.s  $G_1, \dots, G_n$  is easily given (see Exercise 9.4.8 for proof).

**Proposition 9.4.IX.** *When the point processes  $N_1, N_2, \dots$  are mutually independent, the superposition  $N_1 + \cdots + N_n$  has p.g.fl.*

$$G[h] = \prod_{i=1}^n G_i[h] \quad (h \in \mathcal{V}(\mathcal{X})).$$

*This sequence of finite products converges if and only if the infinite sum*

$$N(A) = \sum_{i=1}^{\infty} N_i(A) \quad (9.4.23)$$

*is a.s. finite on bounded  $A \in \mathcal{B}_{\mathcal{X}}$ , and the infinite product is then the p.g.fl. of the point process  $N$  at (9.4.23).*

## Exercises and Complements to Section 9.4

9.4.1 Use linear combinations of indicator functions and their limits to prove that conditions (i) and (ii) of Proposition 9.4.I imply the a.s. form of (9.4.1).

9.4.2 Condition (ii) of Theorem 9.4.II requires the continuity of characteristic functionals

$$\Phi[f_n] \rightarrow \Phi[f] \quad (9.4.24)$$

for  $f, f_n \in \text{BM}(\mathcal{X})$  when  $f_n(x) \rightarrow f(x)$  ( $x \in \mathcal{X}$ ) pointwise monotonically from below with  $f, f_n$  in fact  $\in \text{BM}_+(\mathcal{X})$ . Show that (9.4.24) holds without this monotonicity of convergence or nonnegativity of  $f$  if either



- (i)  $\mathcal{P}\{\xi(\mathcal{X}) < \infty\} = 1$ , bounded measurable  $f$  and  $f_n$ , with

$$\sup_{x \in \mathcal{X}} |f(x) - f_n(x)| \rightarrow 0 \quad (n \rightarrow \infty); \text{ or} \quad (9.4.25)$$

- (ii)  $f$  and  $f_n \in \text{BM}(\mathcal{X})$ , the union of their support is a bounded set, and (9.4.25) holds.

Give an example of a random measure  $\xi$  and functions  $f, f_n \in \text{BM}(\mathcal{X})$  satisfying (9.4.25) for which (9.4.24) fails. [Hint: Consider a stationary Poisson process on  $\mathbb{R}_+$  with  $f(x) = 0$  (all  $x \in \mathbb{R}_+$ ),  $f_n(x) = n^{-1}I_{[0,n]}(x)$ .]

9.4.3 *Laplace functional analogues* of various results for characteristic functionals are available, subject to modifications reflecting the different domain of definition; below,  $f \in \text{BM}_+(\mathcal{X})$ .

- (a) [See Theorem 9.4.II.] Show that  $\{L[f]: \text{all } f\}$  uniquely determines the distribution of a random measure  $\xi$ .  
 (b) [See Exercise 9.4.2.] For sequences  $f_n$ , the convergence  $L[f_n] \rightarrow L[f]$  holds as  $\sup_{x \in \mathcal{X}} |f_n(x) - f(x)| \rightarrow 0$  if  
 (i)  $\xi$  is totally bounded; or  
 (ii) the pointwise convergence  $f_n \rightarrow f$  is monotonic; or  
 (iii) there is a bounded Borel set containing the support of every  $f_n$ .

Give examples to show that, if otherwise, the convergence  $L[f_n] \rightarrow L[f]$  may fail.

9.4.4 For a random measure  $\xi$  on the c.s.m.s.  $\mathcal{X}$  with Laplace functional  $L[\cdot]$ , show that for any bounded  $A \in \mathcal{B}_{\mathcal{X}}$ ,  $\mathcal{P}\{\xi(A) = 0\} = \lim_{s \rightarrow \infty} L[sI_A]$ , whereas for any  $A \in \mathcal{B}_{\mathcal{X}}$ ,  $\mathcal{P}\{\xi(A) < \infty\} = \lim_{s \downarrow 0} \lim_{n \rightarrow \infty} L[sI_{A_n}]$ , where  $\{A_n\}$  is an increasing sequence of bounded sets in  $\mathcal{B}_{\mathcal{X}}$  for which  $A = \lim_{n \rightarrow \infty} A_n$  (the case  $A = \mathcal{X}$  is of obvious interest).

9.4.5 Suppose that the functions  $h_n \in \mathcal{V}(\mathcal{X})$  and that  $h_n(x) \rightarrow h(x)$  ( $n \rightarrow \infty$ ) for every  $x \in \mathcal{X}$ . Show that  $G[h_n] \rightarrow G[h]$  ( $n \rightarrow \infty$ ) if, in place of the conditions at (ii) of Theorem 9.4.V, either

- (a)  $N$  is a.s. totally finite, or  
 (b)  $N$  has a finite first moment measure  $M$  and  $\int_{\mathcal{X}} |h_n(x) - h(x)| M(dx) \rightarrow 0$  as  $n \rightarrow \infty$ .

Let  $h_n(x) = 1 - n^{-1}$  for  $|x| < n$ ,  $= 1$  for  $|x| \geq n$ , so that  $h_n(x) \rightarrow 1$  (all  $x$ ) for  $n \rightarrow \infty$ . Show that for a stationary Poisson process at rate  $\lambda$ ,  $G[h_n] = e^{-2\lambda} \neq 1 = G[\lim_{n \rightarrow \infty} h_n(\cdot)]$  (cf. Exercise 9.4.2).

9.4.6 Let  $\{h_n\} \subset \overline{\mathcal{V}}(\mathcal{X})$  have  $h_n(x) \rightarrow h(x)$  pointwise as  $n \rightarrow \infty$ . Show that the extended p.g.fl. convergence result  $G[h_n] \rightarrow G[h]$  holds if any one of the following conditions holds.

- (a)  $h$  is the monotone limit of functions  $h_n \in \mathcal{V}(\mathcal{X})$ .  
 (b)  $|\log[h_n(x)/h(x)]| < \epsilon(x)$  (all  $n$ ) and  $\int_{\mathcal{X}} \epsilon(x) N(dx) < \infty$  a.s.  
 (c)  $\inf_{x \in \mathcal{X}} h_n(x) > c$  for some  $c > 0$  and sufficiently large  $n$ , and

$$\int_{\mathcal{X}} |h_n(x) - h(x)| N(dx) < \infty \quad \text{a.s.}$$

[Hint: This is the p.g.fl. analogue of Proposition 9.4.III. For part (b), use the method of proof of Proposition 9.4.VIII. Part (c) follows from (b). See Daley and Vere-Jones (1987).]

- 9.4.7 For the partial sums  $\zeta_n$  of random measures (cf. Proposition 9.4.VIII), show that  $L_{\zeta_n}[f]$  has a nonzero limit for  $f \in \text{BM}_+(\mathcal{X})$  if and only if the infinite product  $\prod_{i=1}^{\infty} (1 - \exp[-\xi_i(A)]) > 0$  a.s., that is, when the infinite series at (9.4.21) converges. Hence, complete the proof of the proposition.
- 9.4.8 For an infinite sequence  $N_1, N_2, \dots$  of independent point processes, show that the necessary and sufficient condition for the infinite superposition  $\sum_{i=1}^{\infty} N_i$  to be a well-defined point process is the convergence for every bounded  $A \in \mathcal{B}_{\mathcal{X}}$  of the sum  $\sum_{i=1}^{\infty} p_i(A)$ , where  $p_i(A) = \Pr\{N_i(A) > 0\}$ . Hence, establish Proposition 9.4.IX.
- 9.4.9 Let  $N$  be a renewal process with lifetime d.f.  $F$ , and denote by  $G_{b|a}[h]$  the p.g.fl. of the process on the interval  $[a, b]$  conditional on the occurrence of a point of the process at  $a$ . Much as in Bol'shakov (1969), show that this conditional p.g.fl. satisfies the integral equation

$$G_{b|a}[h] = [1 - F(b - a)] + \int_a^b (h(x) + 1) G_{b|x}[h] d_x F(x - a).$$

Similarly, if  $\tilde{G}_{a|b}[h]$  denotes the p.g.fl. conditional on a point at  $b$ ,

$$\tilde{G}_{a|b}[h] = [1 - F(b - a)] + \int_a^b (h(x) + 1) \tilde{G}_{a|x}[h] |d_x F(b - x)|.$$

Find extensions of these equations to the case where the renewal process is replaced by a Wold process.

## 9.5. Moment Measures and Expansions of Functionals

As with ordinary random variables and characteristic functions, the characteristic functional is closely associated with the moment structure of the random measure, which, as in the point process case studied in Chapter 5, is expressed through a family of *moment measures*. In particular, for any random measure  $\xi$  on the c.s.m.s.  $\mathcal{X}$  and any Borel set  $A$ , consider the expectation

$$M(A) = E[\xi(A)] \quad (\text{finite or infinite}). \quad (9.5.1)$$

Clearly,  $M$  inherits the property of finite additivity from the underlying random measure  $\xi$ . Moreover, if the sequence  $\{A_n\}$  of Borel sets is monotonic increasing to  $A$ , then by monotone convergence  $M(A_n) \uparrow M(A)$ . Thus,  $M(\cdot)$  is continuous from below and therefore a measure. In general, it need not take finite values, even on bounded sets, but when it does, we say that the *expectation measure of  $\xi$  exists* and is given by (9.5.1). The expectation measure  $M(\cdot)$  may also be called the *first moment measure* of  $\xi$ .

When  $M$  exists, the above argument can readily be extended to the random integrals  $\int f d\xi$  for  $f \in \text{BM}(\mathcal{X})$ . Thus, if  $f$  is the indicator function of the bounded Borel set  $A$ ,  $E(\int f d\xi) = M(A)$ . Extending in the usual way through linear combinations and monotone limits it follows that

$$E(\int f d\xi) = \int f dM \quad (f \in \text{BM}(\mathcal{X})). \quad (9.5.2)$$

Equations of the form (9.5.2) have been included under the name *Campbell theorem* [see, e.g., Matthes (1972), MKM (1978)] after early work by Campbell (1909) on the shot-noise process in thermionic vacuum tubes [see also Moran (1968, pp. 417–423)]. Campbell measures that we discuss in Chapter 13 constitute a significant extension of this simple concept.

Consider next the  $k$ -fold product of  $\xi$  with itself, that is, the measure defined a.s. for Borel rectangles  $A_1 \times \cdots \times A_k$  by

$$\xi^{(k)}(A_1 \times \cdots \times A_k) = \prod_{i=1}^k \xi(A_i) \quad (9.5.3)$$

and extended to a measure, necessarily symmetric, on the product Borel  $\sigma$ -algebra in  $\mathcal{X}^{(k)}$ . Now the rectangles form a semiring generating this  $\sigma$ -algebra, and (9.5.3) defines a random variable for every set in this semiring, so it follows from Proposition 9.1.VIII that  $\xi^{(k)}$  is a random measure on  $\mathcal{X}^{(k)}$ .

**Definition 9.5.I.** *The  $k$ th order moment measure  $M_k(\cdot)$  of  $\xi$  is the expectation measure of  $\xi^{(k)}$ , whenever this expectation measure exists.*

This identification is illustrated forcefully in the proof below of a result already given at Proposition 5.4.VI and where a key role is played by

$$\text{diag } A^{(k)} \equiv \{(x_1, \dots, x_k) \in \mathcal{X}^{(k)} : x_1 = \cdots = x_k = x \in A\}. \quad (9.5.4)$$

**Proposition 9.5.II.** *A point process  $N$  with boundedly finite second moment measure has  $M_2(\text{diag } A^{(2)}) \geq M(A)$  for all bounded  $A \in \mathcal{B}_{\mathcal{X}}$ ; equality holds if and only if  $N$  is simple.*

PROOF. Let  $\mathcal{T}$  be a dissecting system for  $A$ , so  $\text{diag } A^{(2)}$  equals the monotone limit  $\lim_{n \rightarrow \infty} \bigcup_{i=1}^{k_n} A_{ni} \times A_{ni}$ . Because  $M_2$  is a measure and  $M_2(A^{(2)}) < \infty$ ,

$$\begin{aligned} M_2(\text{diag } A^{(2)}) &= M_2\left(\lim_{n \rightarrow \infty} \bigcup_{i=1}^{k_n} (A_{ni} \times A_{ni})\right) = \lim_{n \rightarrow \infty} M_2\left(\bigcup_{i=1}^{k_n} (A_{ni} \times A_{ni})\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{E}\left[\sum_{i=1}^{k_n} N^2(A_{ni})\right] = M(A) + \lim_{n \rightarrow \infty} \mathbb{E}\left[\sum_{i=1}^{k_n} N(A_{ni})[N(A_{ni}) - 1]\right]. \end{aligned}$$

Write the last term as  $\mathbb{E}(X_n)$ . From the nesting property of the  $\mathcal{T}_n$ , the r.v.s  $X_n$  are a.s. nonincreasing and  $\geq 0$ , so by monotone convergence

$$M_2(\text{diag } A^{(2)}) = M(A) + \mathbb{E}\left(\lim_{n \rightarrow \infty} X_n\right),$$

and  $\lim_{n \rightarrow \infty} X_n = 0$  if and only if  $\lim_{n \rightarrow \infty} \sup_i N(A_{ni}) \leq 1$  a.s.; that is,  $\mathcal{P}\{N(\{x\}) \leq 1 \text{ for all } x \in A\} = 1$ ; equivalently,  $N = N^*$  a.s.  $\square$

The following kindred property of random measures is proved similarly; see Exercise 9.5.5(b), and Section 9.3 and Exercise 9.5.6 for related results.

**Proposition 9.5.III.** *A random measure  $\xi$  with boundedly finite second moment measure  $M_2$  is a.s. nonatomic if and only if  $M_2(\text{diag } \mathcal{X}^{(2)}) = 0$ .*

EXAMPLE 9.5(a) *Mixtures of quadratic random measures.* A Gaussian r.v. has moments of all orders, from which it follows that the same is true for the stationary quadratic random measure in Example 9.1(b). In particular, its first and second moment measures are defined by the equations

$$M(A) = E[\xi(A)] = E\left[\int_A Z^2(x) dx\right] = \sigma^2 \ell(A), \quad (9.5.5a)$$

$$\begin{aligned} M_2(A \times B) &= E[\xi(A)\xi(B)] = E\left[\int_A \int_B Z^2(x)Z^2(y) dx dy\right] \\ &= \int_A \int_B [\sigma^4 + 2c^2(x-y)] dx dy. \end{aligned} \quad (9.5.5b)$$

From these representations it is clear that  $M$  and  $M_2$  are both absolutely continuous with respect to Lebesgue measure on  $\mathbb{R}$  and  $\mathbb{R}^2$ , with derivatives  $\sigma^2$  and  $\sigma^4 + 2c^2(x-y)$ , respectively, where  $c(\cdot)$  is the covariance function for  $Z$ . Similar representations can be obtained for higher moments.  $\square$

EXAMPLE 9.5(b) *Mixed random measure.* Let  $\Lambda$  be a positive r.v., independent of the random measure  $\xi(\cdot)$ , and set

$$\xi_\Lambda(A) = \Lambda \xi(A).$$

Using independence, the  $k$ th order moment measures  $M_\Lambda^{(k)}$  for  $\xi_\Lambda$  are related to those of  $\xi$  by the equations

$$M_\Lambda^{(k)}(\cdot) = E(\Lambda^k) M_k(\cdot).$$

Thus, if  $\Lambda$  has infinite moments of order  $k$  and higher, the same will be true for the moment measures of  $\xi_\Lambda$ , and conversely, if the  $k$ th moment of  $\xi_\Lambda$  is finite, the  $k$ th order moment measure of  $\xi$  exists.

This particular example is nonergodic [meaning, the values of  $M(A)$  cannot be determined from observations on a single realization of the process], but this is not a necessary feature of examples with infinite moment measures: for example, in place of the r.v.  $\Lambda$ , we could multiply  $\xi$  by any continuous ergodic process  $\lambda(t)$  with infinite moments and integrate to obtain a random measure with similar moment properties. This procedure of mixing, or randomizing, with respect to a given parameter of a process is a rich source of examples.  $\square$

EXAMPLE 9.5(c) *Moments of completely random measures* (see Sections 2.2 and 10.1). If  $\xi(\cdot)$  is completely random, the first and second moment measures (assuming these are finite) are given by relations of the type

$$\begin{aligned} M(A) &= E[\xi(A)] = \mu(A), \\ M_2(A \times B) &= E[\xi(A)\xi(B)] = \mu(A)\mu(B) + \text{var } \xi(A \cap B). \end{aligned} \quad (9.5.6)$$

Particular interest here centres on the variance term: it vanishes unless the set  $A \cap B \neq \emptyset$ , so this term represents a measure concentrated along the

diagonal. For the stationary gamma random measure studied in Example 9.1(d),  $\text{var } \xi(A) = \lambda^2 \alpha \ell(A)$  and so

$$M_2(A \times B) = \lambda^2 \alpha \ell(A \cap B) + \lambda^2 \alpha^2 \ell(A) \ell(B). \quad (9.5.7)$$

Thus,  $M_2$  has a constant areal density  $\lambda^2 \alpha^2$  off the diagonal and a concentration with linear density  $\lambda^2 \alpha$  along it. Such concentrations are associated with the a.s. atomic character of the random measure (see Proposition 9.5.III) and should be contrasted with the absolutely continuous moment measures of Example 9.5(a) in which the realizations themselves are a.s. absolutely continuous measures.  $\square$

The next lemma summarizes the relation between moment measures and the moments of random integrals, and is useful in discussing expansions of functionals. Note also the identification property at Exercise 9.5.7.

**Lemma 9.5.IV.** *Let the  $k$ th moment measure  $M_k$  of the random measure  $\xi$  exist. Then for all  $f \in \text{BM}(\mathcal{X})$ , the random integral  $\int f d\xi$  has finite  $k$ th moment satisfying*

$$\mathbb{E} \left[ \left( \int f d\xi \right)^k \right] = \int_{\mathcal{X}^{(k)}} f(x_1) \dots f(x_k) M_k(dx_1 \times \dots \times dx_k). \quad (9.5.8)$$

PROOF. Apply (9.5.2) to the product measure  $\xi^{(k)}$ , for which  $M_k$  is the expectation measure. This gives

$$\begin{aligned} & \int_{\mathcal{X}^{(k)}} h(x_1, \dots, x_k) M_k(dx_1 \times \dots \times dx_k) \\ &= \mathbb{E} \left[ \int_{\mathcal{X}} \dots \int_{\mathcal{X}} h(x_1, \dots, x_k) \xi(dx_1) \dots \xi(dx_k) \right] \end{aligned} \quad (9.5.9)$$

for all  $k$ -dimensional bounded Borel measurable functions  $h(x_1, \dots, x_k): \mathcal{X}^{(k)} \mapsto \mathcal{X}$ . Then (9.5.8) is the special case  $h(x_1, \dots, x_k) = \prod_{i=1}^k f(x_i)$ .  $\square$

We now consider the finite Taylor series expansion of the characteristic functional (see Exercise 9.5.8 for an expansion for the Laplace functional).

**Proposition 9.5.V.** *Let  $\Phi$  be the characteristic functional of the random measure  $\xi$ , and suppose that the  $k$ th moment measure of  $\xi$  exists for some  $k \geq 1$ . Then for each fixed  $f \in \text{BM}(\mathcal{X})$  and real  $s \rightarrow 0$ ,*

$$\Phi[sf] = 1 + \sum_{r=1}^k \frac{(is)^r}{r!} \int_{\mathcal{X}^{(r)}} f(x_1) \dots f(x_r) M_r(dx_1 \times \dots \times dx_r) + o(|s|^k). \quad (9.5.10)$$

Furthermore, if the  $(k+1)$ th moment exists, the remainder term  $o(|s|^k)$  is bounded by

$$\frac{|s|^{k+1}}{(k+1)!} C_f^{k+1} M_{k+1}(A_f^{(k+1)}), \quad (9.5.11)$$

where  $C_f$  is a bound for  $f$ , and  $f$  vanishes outside the bounded Borel set  $A_f$ .

PROOF. Because  $\Phi[sf] = \phi_f(s)$ , where  $\phi_f$  is the ordinary characteristic function for the r.v.  $\int f d\xi$ , both assertions follow from (9.5.8) and the corresponding Taylor series results for ordinary characteristic functions. The bound (9.5.11), with  $C_f$  and  $A_f$  as defined there, is derived from

$$E[|\int f d\xi|^{k+1}] \leq E[(\int |f| d\xi)^{k+1}] \leq E[(C_f \xi(A_f))^{k+1}]. \quad \square$$

The analogy with the finite-dimensional situation may be strengthened by noting that the moment measures can be identified with successive Fréchet derivatives of  $\Phi$ . Specifically, we can write formally

$$M_k(A) = \Phi^{(k)}(I_A).$$

The difficulty with such expressions is that they rarely give much information, either theoretical or computational, concerning the analytic form or other characteristics of the moment measures.

The corresponding expression for the logarithm of the characteristic functional leads to a new family of measures associated with  $\xi$ , the *cumulant measures*. The first cumulant measure coincides with the expectation measure, whereas the second is the *covariance measure* defined by

$$C_2(A \times B) = M_2(A \times B) - M(A)M(B) = \text{cov}(\xi(A), \xi(B)). \quad (9.5.12)$$

In Example 9.5(a), the covariance measure is absolutely continuous with respect to two-dimensional Lebesgue measure, the *covariance density* being given by

$$c_2(x, y) = 2c^2(x - y).$$

The covariance density for this random measure is just the ordinary covariance function of the process forming the density of the random measure. Similar relations between the moment and cumulant densities of the random measure, and the moment and cumulant functions of its density, hold whenever the random measure can be represented as the integral of an underlying process. By contrast, in Example 9.5(c), the covariance measure is singular, consisting entirely of the concentration along the diagonal  $y = x$  with linear density  $\lambda^2 \alpha$ . The relation can be expressed conveniently using the Dirac delta function as

$$c_\xi(x, y) = \lambda^2 \alpha \delta(x - y).$$

The general relation between the moment measures and the cumulant measures is formally identical to the relation between the factorial moment measures and factorial cumulant measures studied in Chapter 5, inasmuch as both are derived by taking logarithms of an expression of the type (9.5.9).

Finally we consider analogues of Proposition 9.5.V for point processes that no longer need be finite as in the discussion in Section 5.4. The advantages of working with factorial moment measures  $M_{[k]}$  remain: the same definition

at (5.4.2) holds when we require the sets  $A_i$  there to be bounded. Then  $M_{[k]}$  exists (i.e., is boundedly finite) if and only if  $M_k$  exists, and the definitions (5.4.3) and (5.4.4) continue, now with bounded sets  $A_i$ .

In the proof below  $N^{[k]}(\cdot)$  denotes the  $k$ -fold factorial product measure as described below (5.4.4) (the notation recalls the factorial power used around Definition 5.2.I). It has expectation measure  $M_{[k]}(\cdot)$ . For simple point processes, integration with respect to  $N^{[k]}(dx_1 \times \cdots \times dx_k)$  is the same as integration with respect to  $N(dx_1) \cdots N(dx_k)$  if we add the restriction that  $x_1, \dots, x_k$  must all be distinct. Thus, the integral at (7.1.13) could be written without the coincidence annihilating function  $I(\cdot)$  there if instead the product measure is replaced by the factorial product measure.

**Proposition 9.5.VI.** *Let  $G$  be the p.g.fl. of a point process whose  $k$ th order moment measure exists for some positive integer  $k$ . Then for  $1 - \eta \in \mathcal{V}(\mathcal{X})$  and  $0 < \rho < 1$ ,*

$$G[1 - \rho\eta] = 1 + \sum_{j=1}^k \frac{(-\rho)^j}{j!} \int_{\mathcal{X}^{(j)}} \eta(x_1) \cdots \eta(x_j) M_{[j]}(dx_1 \times \cdots \times dx_j) + o(\rho^k). \quad (9.5.13)$$

PROOF. Some care is needed in evaluating the difference between  $G[1 - \rho\eta]$  and the finite sum on the right-hand side of (9.5.13). For fixed  $\eta$  and a given realization  $\{y_i\}$  of the point process, consider the expressions

$$S_m(\rho) = 1 + \sum_{j=1}^m \frac{(-\rho)^j}{j!} \int_{\mathcal{X}^{(j)}} \eta(x_1) \cdots \eta(x_j) N^{[j]}(dx_1 \times \cdots \times dx_j),$$

where  $N^{[j]}$  is the modified product counting measure formed by taking all possible ordered  $j$ -tuples of different points of the realizations of  $N$  (with the convention that if  $\{y_i\}$  has multiple points these should be treated as different points with the same state space coordinates, that is, as if they represented distinct particles). Each integral then reduces to a sum

$$Q_j = \sum \eta(y_{i_1}) \cdots \eta(y_{i_j})$$

over all such  $j$ -tuples. Effectively, each sum is a.s. finite, because with probability 1 only a finite number of points of the process will fall within the support of  $1 - \eta$ . Moreover, the sum vanishes whenever  $j$  is larger than the number of points in this support. Because  $N^{[j]}$  includes all possible orderings of a given  $j$ -tuple, each distinct term in the sum occurs  $j!$  times, so that we can write  $Q_j = \sum j! q_j$ , where the sum on  $q_j$  extends over all distinct combinations of  $j$  points from  $\{y_i\}$  (with the same convention as before regarding multiple points). In this notation we have

$$S_m(\rho) = 1 + \sum_{j=1}^m (-\rho)^j q_j, \quad (9.5.14)$$

and it is not difficult to verify (e.g., by induction) that for all  $m$  and  $\eta$ ,

$$S_{2m+1}(\rho) \leq \prod_i [1 - \rho\eta(y_i)] \equiv \Pi(\rho) \leq S_{2m}(\rho), \quad (9.5.15)$$

the product being taken over  $\{y_i\}$ ; Exercise 9.5.9 interprets (9.5.15) in terms of Bonferroni inequalities. Equation (9.5.14) implies that  $|S_k - S_{k-1}| \leq \rho^k q_k$ , which with (9.5.15) and its implication that the sums  $S_k(\rho)$  are alternately above and below  $\Pi(\rho)$  implies both

$$|S_k(\rho) - \Pi(\rho)| \leq \rho^k q_k \quad \text{and} \quad |S_k(\rho) - \Pi(\rho)| \leq \rho^{k+1} q_{k+1}. \quad (9.5.16)$$

Now suppose that  $M_k$ , and hence  $M_{[k]}$ , exist. The first inequality at (9.5.16) implies that  $[S_k(\rho) - \Pi(\rho)]/\rho^k$  is bounded by a random variable with finite expectation

$$E(q_k) = \frac{1}{k!} \int_{\mathcal{X}^{(k)}} \eta(x_1) \dots \eta(x_k) M_{[k]}(dx_1 \times \dots \times dx_k),$$

because  $M_{[k]}$  is just the expectation of  $N^{[k]}$ . The second inequality at (9.5.16) implies that  $[S_k(\rho) - \Pi(\rho)]/\rho^k \rightarrow 0$  a.s. as  $\rho \rightarrow 0$ . The limit behaviour of the remainder term in (9.5.13) now follows by dominated convergence.

Uniqueness of the expansion (9.5.13) follows from the uniqueness of the coefficients in a power series expansion and the fact that, as symmetric measures, the moment measures are uniquely specified by integrals of the type appearing in the expansion (see Exercise 9.5.7).  $\square$

Taking expectations in (9.5.16) yields the following corollary.

**Corollary 9.5.VII.** *When  $M_{k+1}$  exists, the remainder term in (9.5.13) is bounded by*

$$\frac{\rho^{k+1}}{(k+1)!} \int_{\mathcal{X}^{(k+1)}} \eta(x_1) \dots \eta(x_{k+1}) M_{[k+1]}(dx_1 \times \dots \times dx_{k+1}).$$

On taking logarithms of the expression (9.5.13) and using the expansion  $\log(1 - y) = -\sum_{j=1}^k y^j/j + o(|y|^j)$  ( $y \rightarrow 0$ ), equation (9.5.17) below follows.

**Corollary 9.5.VIII.** *Under the conditions of Proposition 9.5.VI, the p.g.fl. can be expressed in terms of the factorial cumulant measures  $C_{[j]}$ , for  $\rho \rightarrow 0$ , as*

$$\log G[1 - \rho\eta] = \sum_{j=1}^k \frac{(-\rho)^j}{j!} \int_{\mathcal{X}^{(j)}} \eta(x_1) \dots \eta(x_j) C_{[j]}(dx_1 \times \dots \times dx_j) + o(\rho^k). \quad (9.5.17)$$

Equation (9.5.17) serves to define the cumulant measures, which can be expressed explicitly in terms of the measures  $M_{[k]}$  as in Chapter 5. Unfortunately it does not seem possible to provide a simple bound for the remainder term in (9.5.17) analogous to that of Corollary 9.5.VII (but, see Exercise 9.5.10).  $\square$



EXAMPLE 9.5(d) *Moment measures of Poisson and compound and mixed Poisson processes* [continued from Examples 9.4(c)–(d)]. Expanding the p.g.fl. at (9.4.17) formally, in terms of  $\eta$  with  $1 - \eta \in \mathcal{V}(\mathcal{X})$ ,

$$\begin{aligned} G[1 + \eta] &= 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \left( \int_{\mathcal{X}} \eta(x) \mu(dx) \right)^k \\ &= 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \eta(x_1) \cdots \eta(x_k) \mu(dx_1) \cdots \mu(dx_k). \end{aligned}$$

Thus, for the Poisson process with parameter measure  $\mu$ , the  $k$ th order factorial moment measure  $M_{[k]}$  is the  $k$ -fold product measure of  $\mu$  with itself. The situation with the cumulant measures is even simpler: here,  $\log G[1 + \eta] = \int_{\mathcal{X}} \eta(x) \mu(dx)$  so that for a Poisson process the second and all higher factorial cumulant measures vanish.

This last result is in marked contrast with the situation for compound Poisson processes for which  $\log G[1 + \eta]$  equals

$$\begin{aligned} \int_{\mathcal{X}} \sum_n ([1 + \eta(x)]^n - 1) \pi_n(x) \mu(dx) &= \sum_{k=1}^{\infty} \int_{\mathcal{X}} \frac{[\eta(x)]^k}{k!} \sum_{n=k}^{\infty} n^{[k]} \pi_n(x) \mu(dx) \\ &= \sum_{k=1}^{\infty} \int_{\mathcal{X}} [\eta(x)]^k m_{[k]}(x) \mu(dx), \end{aligned}$$

where  $m_{[k]}(x)$  is the  $k$ th factorial moment of the batch-size distribution  $\{\pi_n(x)\}$  at the point  $x$ , assuming the moment exists. This representation implies that  $C_{[k]}$  is concentrated on the diagonal elements  $(x, \dots, x)$  where it reduces to a measure with density  $m_{[k]}(x)$  with respect to  $\mu(\cdot)$ .

For the mixed Poisson process, suppose that  $\Lambda$  has a finite  $k$ th moment. Then in a neighbourhood  $\operatorname{Re}(s) \geq 0$  of  $s = 0$ , the Laplace–Stieltjes transform

$$L(s) = 1 + \sum_{j=1}^k \frac{(-s)^j E(\Lambda^j)}{j!} + o(|s|^k).$$

Then for  $\eta$  with  $1 - \eta \in \mathcal{V}(\mathcal{X})$ ,  $G[1 - \rho\eta]$  equals

$$1 + \sum_{j=1}^k \frac{(-\rho)^j E(\Lambda^j)}{j!} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \eta(x_1) \cdots \eta(x_j) \mu(dx_1) \cdots \mu(dx_j) + o(\rho^k),$$

and thus the factorial moment measures of the process are given by

$$M_{[j]}(dx_1 \times \cdots \times dx_j) = E(\Lambda^j) \mu(dx_1) \cdots \mu(dx_j) \quad (j \leq k). \quad (9.5.18)$$

If in particular,  $\mathcal{X} = \mathbb{R}^d$  and  $\mu(\cdot)$  is Lebesgue measure on  $\mathbb{R}^d$ , then  $M_{[j]}$  has a density  $m_{[j]}$  with respect to such Lebesgue measure given by

$$m_{[j]}(x_1, \dots, x_j) = E(\Lambda^j) \quad (j \leq k).$$

Thus, the factorial moment measures for the mixed Poisson process retain the product form of the Poisson case but are multiplied by the scalar factors  $E(\Lambda^j)$ . McFadden (1965a) and Davidson (1974c) established the following converse to this result. Let  $\{M_{[j]}(\cdot)\}$  be a sequence of product measures of the form (9.5.18) with  $\{E(\Lambda^j)\}$  replaced by a sequence  $\{\gamma_j\}$ . Then the  $M_{[j]}(\cdot)$  are the factorial moment measures of a point process if and only if  $\{\gamma_j\}$  is the moment sequence of some nonnegative r.v.  $\Lambda_0$ . A sufficient condition for the resulting process to be uniquely defined is that  $\sum \gamma_j^{-1/j} < \infty$ , in which case it is necessarily the mixed Poisson process with parameter measure  $\Lambda_0\mu$ , where  $\Lambda_0$  has a uniquely defined distribution with moments  $\{\gamma_j\}$ . Proposition 5.4.VII implies a weaker version of this result (see Exercise 9.5.11).

For completeness here recall that in Example 6.4(b) we discussed *negative binomial processes*, meaning, point processes  $N$  for which  $N(A)$  has a negative binomial distribution [but see also Example 9.1(b)]. In particular, we noted two mechanisms leading to such processes, one starting from compound Poisson processes and the other from mixed Poisson processes.  $\square$

Not infrequently our concern with a point process may be with its structure only on some bounded region  $A$  of the state space. Within  $A$  (assumed to be Borel),  $N$  is a.s. finite-valued by assumption, and its probabilistic structure must be expressible in terms of some family of local probability distributions or Janossy measures as in Definition 5.4.IV and, for the p.g.fl., Example 5.5(b). However, because such point processes are in general a.s. infinite on the whole of  $\mathcal{X}$ , no such measures exist for the process as a whole. We illustrate in the next example how local characteristics can be described: we take the case of the negative binomial process in the setting of its local Janossy measure.

**EXAMPLE 9.5(e) Local properties of the negative binomial process.** Recall from Example 5.5(b) that, given the p.g.fl.  $G[\cdot]$  of a point process and a bounded Borel set  $A$ , the p.g.fl.  $G_A[\cdot]$  of the local process on  $A$  is given by

$$G_A[h] = G[1 - I_A + h^*] \quad [h \in \mathcal{V}(A)],$$

where  $h^*(x) = h(x)I_A(x)$  so that  $h^* \in \mathcal{V}(\mathcal{X})$ . Example 6.4(b)(i) gives the p.g.fl.

$$G[h] = \exp \left( \int_{\mathcal{X}} \frac{\log([1 - \rho h(x)]/(1 - \rho))}{\log(1 - \rho)} \mu(dx) \right)$$

for a negative binomial process coming from a Poisson cluster process with clusters degenerate at a point and size following a negative binomial distribution. Then because the integral over  $A^c$  vanishes, we deduce that

$$G_A[1 - I_A + h^*] = \exp \left[ \frac{1}{\log(1 - \rho)} \int_A \log \left( \frac{1 - \rho h}{1 - \rho} \right) \mu(dx) \right].$$

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

$$J_1(dx | A) = \rho p_0(A) \mu(dx),$$

$$J_2(dx_1 \times dx_2 | A) = \rho^2 p_0(A) [\mu(dx_1) \mu(dx_2) + \delta(x_1, x_2) \mu(dx_1)],$$

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  $x_1 = x_2$ .  $\square$

## Exercises and Complements to Section 9.5

9.5.1 *Moment measures of Dirichlet process.* Let  $\xi$  be a random probability measure on  $\mathcal{X}$ . Show that for every  $k$ , the  $k$ th moment measure exists and defines a probability measure on  $\mathcal{X}^{(k)}$ . Find these measures for the Dirichlet process  $\zeta$  of Example 9.1(e), showing in particular that

$$E\zeta(A) = \frac{\alpha(A)}{\alpha(\mathcal{X})}, \quad \text{var } \zeta(A) = \frac{\alpha(A)}{\alpha(\mathcal{X})} \cdot \frac{1 - \alpha(A)/\alpha(\mathcal{X})}{\alpha(\mathcal{X}) + 1}.$$

9.5.2 For the random measure induced by the limit random d.f. of Exercise 9.3.4, show that the first moment measure is Lebesgue measure on  $[0, 1]$ .

9.5.3 Let  $\xi$  be a random measure on  $\mathcal{X} = \mathbb{R}^d$ , and for  $g \in \text{BM}_+(\mathcal{X})$  define  $G(A) = \int_A g(x) \ell(dx)$ , where  $\ell$  denotes Lebesgue measure on  $\mathbb{R}^d$ . Define  $\eta$  on  $\mathcal{B}_{\mathcal{X}}$  by

$$\eta(A) = \int_{\mathcal{X}} G(A - x) \xi(dx).$$

- (a) Show that  $\eta(A)$  is an a.s. finite-valued r.v. for bounded  $A \in \mathcal{B}_{\mathcal{X}}$ , that it is a.s. countably additive on  $\mathcal{B}_{\mathcal{X}}$ , and hence invoke Proposition 9.1.VIII to conclude that  $\eta$  is a well-defined random measure.
- (b) Show that if  $\xi$  has moment measures up to order  $k$ , so does  $\eta$ , and find the relation between them. Verify that the  $k$ th moment measure of  $\eta$  is absolutely continuous with respect to Lebesgue measure on  $(\mathbb{R}^d)^{(k)}$ .
- (c) Denoting the characteristic functionals of  $\xi$  and  $\eta$  by  $\Phi_{\xi}[\cdot]$  and  $\Phi_{\eta}[\cdot]$ , show that for  $f \in \text{BM}_+(\mathcal{X})$ ,

$$h(x) = \int_{\mathcal{X}} f(y) g(y - x) dy$$

is also in  $\text{BM}_+(\mathcal{X})$ , and

$$\Phi_{\eta}[f] = \Phi_{\xi}[h].$$

9.5.4 (*Continuation*). By its very definition,  $\eta$  is a.s. absolutely continuous with respect to Lebesgue measure and its density

$$Y(t) \equiv \int_{\mathcal{X}} g(t - x) \xi(dx),$$

when  $\xi$  is completely random, is called a *linear process*. Find the characteristic functional of  $Y$  when  $\xi$  is a stationary gamma random measure.

[*Remark:* Examples of linear processes are provided by equation (9.5.2) in connection with the original Campbell theorem and by the shot-noise process as in Examples 6.1(d) and 6.2(a). See Exercise 10.1.3(b) for the case that  $\xi$  is completely random; for other references see, e.g., Westcott (1970).]

9.5.5 (a) For the random probability distribution on  $\mathbb{R}_+$  defined as in Exercise 9.1.4 by  $F_\eta(x) = 1 - \exp(-\xi([0, x]))$ , show that  $\xi$  is a.s. nonatomic if and only if the second moment measure  $M_2$  of  $\eta$  has  $M_2(\text{diag}(\mathbb{R}^{(2)})) = 0$ .

(b) Prove the more general Proposition 9.5.III.

9.5.6 (a) Use the intensity measure  $\lambda_k^*$  of the simple point process  $N_k^*$  in the decomposition at (9.3.9) [see also (9.3.17) and Proposition 9.5.II] to show that

$$M_2(\text{diag } A^{(2)}) = M(A) + \sum_{k=2}^{\infty} k(k-1)\lambda_k^*(A) = M(A) + M_{[2]}(\text{diag } A^{(2)}).$$

Conclude more generally that  $M_r(\text{diag } A^{(r)}) = \sum_{k=1}^{\infty} k^r \lambda_k^*(A)$ .

(b) When  $M_r(A^{(r)}) < \infty$ , deduce that  $M_{[r]}(\text{diag } A^{(r)}) = 0$  if and only if  $\mathcal{P}\{N(\{x\}) \leq r-1 \text{ for all } x \in A\} = 1$ .

9.5.7 Let  $M_{(k)}$  be a symmetric measure on  $(\mathcal{X}^{(k)}, \mathcal{B}^{(k)})$ . By starting from indicator functions, show that  $M_{(k)}$  is uniquely determined by integrals of the form

$$\int_{\mathcal{X}^{(k)}} \eta(x_1) \dots \eta(x_k) M_{(k)}(dx_1 \times \dots \times dx_k) \quad (1 - \eta \in \mathcal{V}(\mathcal{X})).$$

9.5.8 Expand  $E(e^{-X-\epsilon Y})$  for nonnegative r.v.s  $X$  and  $Y$  to deduce that if a random measure  $\xi$  has a finite  $k$ th order moment measure, then for  $\epsilon > 0$  and for functions  $f, g \in \text{BM}_+(\mathcal{X})$ , using  $\xi_f$  as at (9.5.1),

$$L[f + \epsilon g] = L[f] - \epsilon E[\xi_g \exp(-\xi_f)] + \dots + \frac{(-\epsilon)^k}{k!} E[(\xi_g)^k \exp(-\xi_f)] + o(\epsilon^k).$$

9.5.9 Let  $Q_K = \prod_{i=1}^K (1 - \alpha_i)$ , where  $0 < \alpha_i < 1$  for  $i = 1, \dots, K$ , and write

$$q_k = \sum_{1 \leq i_1 < \dots < i_k \leq K} \alpha_{i_1} \dots \alpha_{i_k},$$

so that  $Q_K = 1 - q_1 + q_2 - \dots + (-1)^K q_K$ . By interpreting  $\{\alpha_i\}$  as the set of probabilities of some independent events  $A_1, \dots, A_K$ , and using the Bonferroni inequalities (see Exercise 5.2.5), show that for all  $K$  and positive integers  $m \leq \frac{1}{2}K$ ,

$$S_K^{(2m-1)} \leq Q_K \leq S_K^{(2m)},$$

where

$$S_K^{(k)} = 1 - q_1 + q_2 - \dots + (-1)^k q_k, \quad k = 1, \dots, K.$$

Hence, deduce (9.5.15). [*Hint*: See Westcott (1972).]

9.5.10 Give a Laplace functional analogue of Proposition 9.5.V.

[*Hint*: Replace  $(is)^r$  there by  $(-s)^r$ , where now  $\text{Re}(s) \geq 0$ .]

9.5.11 Suppose that for  $j = 1, 2, \dots$ , measures  $M_{[j]}$  on  $\mathcal{B}(\mathcal{X}^{(j)})$  are defined by (9.5.18) with each  $E(\Lambda^j)$  replaced by some  $\gamma_j > 0$  for which  $\sum_{j=1}^{\infty} (1+\epsilon)^j \gamma_j < \infty$  for some  $\epsilon > 0$ . Use (5.4.8) to show that local Janossy measures are defined and that they determine a point process whenever  $\{\gamma_j\}$  is a moment sequence.

## CHAPTER 10

# Special Classes of Processes

10.1	Completely Random Measures	77
10.2	Infinitely Divisible Point Processes	87
10.3	Point Processes Defined by Markov Chains	95
10.4	Markov Point Processes	118

We have already discussed in Volume I a variety of particular models for point processes and random measures, and described many of their properties. With the added benefit of the basic theory in Chapter 9, we return here to the study of four important classes of models: completely random measures; infinitely divisible point processes; point processes generated by Markov chains; and Markov point processes in space. Each class has interest in its own right, and contains models which are widely used in applications. Although it is the intrinsic interest of the models that motivates the discourse, our immediate aims are to use the theory of the last chapter to establish structure theorems for these classes, to show that they are well-defined mathematical objects, and to establish some of their general properties.

A key feature of the first two classes is their close link to the Poisson process. Indeed, they form natural extensions of the compound Poisson processes discussed in Chapters 2 and 9. Because of this feature, many of their properties can be handled compactly by p.g.fl. techniques, and we make extensive use of this approach. It should be borne in mind, however, that the main advantage of this approach lies precisely in its compactness: it quickly summarizes information that can still be derived quite readily without it and that in less tractable examples may not be so easily expressible in p.g.fl. form.

The other two sections illustrate both the power and limitations of the ideas of Markov chains that so pervade applied probability. When the process has a temporal ingredient, it is natural to include this in any probabilistic description so that it is the ‘future’ that is predicted (stochastically) on the basis of sufficient conditions described by the ‘present’ so that any other knowledge from the past is superfluous in terms of making any prediction better informed. In reality, this amounts to a factorization of the stochastic structure

of the evolution of the process as a product of probabilistic terms (densities) 'chained' through adjacent epochs in time.

It is exactly a product probabilistic structure that underlies the class of so-called Markov point processes in space. The class of models that can be described in this way leads to representations of the type now known under the umbrella of Hammersley–Clifford theorems. This Section 10.4 serves as a taste of further results in Chapter 15.

## 10.1. Completely Random Measures

This section represents both an illustration of the ideas of the sample-path properties expounded in Section 9.3 and an extension of the discussion of Section 2.4 on the general Poisson process. The principal result is the Representation Theorem 10.1.III, which comes from Kingman (1967); it is based on a careful study of sample-path structures to which we proceed immediately.

Although completely random measures have been referred to at several points in the text already, we state the following for the record.

**Definition 10.1.I.** *A random measure  $\xi$  on the c.s.m.s.  $\mathcal{X}$  is completely random if for all finite families of disjoint, bounded Borel sets  $\{A_1, \dots, A_k\}$ , the random variables  $\{\xi(A_1), \dots, \xi(A_k)\}$  are mutually independent.*

Of course, the Poisson process discussed extensively in Chapter 2 and elsewhere is the prime example. A compound Poisson process with marks in the c.s.m.s.  $\mathcal{K}$  is a completely random process on  $\mathcal{X} \times \mathcal{K}$  with the additional requirement that the ground process  $N_g$  on  $\mathcal{X}$  be well defined (and then a Poisson process in its own right). In one dimension, a random measure  $\xi$  is completely random if and only if the corresponding cumulative process  $\eta(t) = \int_0^t \xi(dx)$  has independent increments.

Recall from Corollary 9.3.VI that a completely random measure  $\xi$  on  $\mathcal{X}$  is nonatomic if and only if for some dissecting system  $\{\mathcal{T}_n\}$  and every  $\epsilon > 0$ ,  $\sum_{i=1}^{k_n} \mathcal{P}\{\xi(A_{ni}) \geq \epsilon\} \rightarrow 0$  ( $n \rightarrow \infty$ ). A substantial step in the proof of the main representation result, equation (10.1.4) in Theorem 10.1.III, is the following result, which is of interest in its own right.

**Proposition 10.1.II.** *If the completely random measure  $\xi$  is a.s. nonatomic, then there is a fixed nonatomic measure  $\nu$  such that*

$$\xi(\cdot) = \nu(\cdot) \quad \text{a.s.} \quad (10.1.1)$$

PROOF. Let  $\mathcal{T} = \{\mathcal{T}_n\} = \{\{A_{ni}: i = 1, \dots, k_n\}\}$  be a dissecting system for any given bounded Borel set  $A$ ; define the transforms

$$\psi_{ni}(s) = E(\exp[-s\xi(A_{ni})]) \quad (\operatorname{Re}(s) \geq 0).$$

Because  $\xi$  is completely random and  $\xi(A) = \sum_{i=1}^{k_n} \xi(A_{ni})$ , we have

$$\psi_A(s) \equiv E(\exp[-s\xi(A)]) = \prod_{i=1}^{k_n} \psi_{ni}(s) \quad (n = 1, 2, \dots),$$

and

$$1 - \psi_{ni}(s) = E(1 - \exp[-s\xi(A_{ni})]) = \int_0^\infty s e^{-sy} \mathcal{P}\{\xi(A_{ni}) > y\} dy.$$

Appealing to Corollary 9.3.VI and the dominated convergence theorem,  $\xi$  being nonatomic implies that

$$\max_{1 \leq i \leq k_n} [1 - \psi_{ni}(s)] \rightarrow 0 \quad (n \rightarrow \infty)$$

for every fixed real  $s \geq 0$ . Using this result in an expansion of the logarithmic term below, it now follows that

$$\begin{aligned} -\log \psi_A(s) &= -\sum_{i=1}^{k_n} \log \psi_{ni}(s) = -\lim_{n \rightarrow \infty} \sum_{i=1}^{k_n} \log \psi_{ni}(s) \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^{k_n} [1 - \psi_{ni}(s)] \\ &= \lim_{n \rightarrow \infty} \int_0^\infty (1 - e^{-sy}) G_n(dy), \end{aligned} \tag{10.1.2}$$

where  $G_n(\cdot)$  is the sum of the  $k_n$  individual probability measures of  $\xi(A_{ni})$ . Again from (9.3.7),

$$\int_\epsilon^\infty G_n(dy) \rightarrow 0 \quad (n \rightarrow \infty) \tag{10.1.3}$$

for every fixed  $\epsilon > 0$ , and from the limit relation (10.1.2) it follows that  $\int_0^\epsilon y G_n(dy)$  remains bounded as  $n \rightarrow \infty$ . Thus, the sequence of measures

$$H_n(dy) \equiv \min(1, y) G_n(dy)$$

is not merely bounded in total mass but, from (10.1.3), is uniformly tight. Again, using (10.1.3), it follows that the only possible limit for  $\{H_n\}$  is a degenerate measure with its mass concentrated at the origin. Uniform tightness implies the existence of a convergent subsequence, thus there must exist a constant  $\nu \equiv \nu(A)$  and a sequence  $\{n_k\}$  for which  $H_{n_k} \rightarrow \nu(A)\delta_0$  weakly, and therefore

$$\begin{aligned} -\log \psi_A(s) &= \lim_{n_k \rightarrow \infty} \int_0^\infty (1 - e^{-sy}) G_{n_k}(dy) \\ &= \lim_{n_k \rightarrow \infty} \int_0^\infty \frac{1 - e^{-sy}}{\min(y, 1)} H_{n_k}(dy) = s\nu(A). \end{aligned}$$

This result is equivalent to  $\xi(A) = \nu(A)$  a.s. for the given bounded Borel set  $A$ . Because such a relation holds for any bounded Borel set  $A$ , the family  $\nu(\cdot)$  must be a finitely bounded Borel measure and  $\xi = \nu$  must hold for almost all realizations. Finally,  $\xi$  being a.s. free of atoms, the same must be true of  $\nu$  also.  $\square$

The major goal now is the following representation theorem.

**Theorem 10.1.III** (Kingman, 1967). *Any completely random measure  $\xi$  on the c.s.m.s.  $\mathcal{X}$  can be uniquely represented in the form*

$$\xi(A) = \sum_{k=1}^{\infty} U_k \delta_{x_k}(A) + \nu(A) + \int_0^{\infty} y N(A \times dy), \quad (10.1.4)$$

where the sequence  $\{x_k\}$  enumerates a countable set of fixed atoms of  $\xi$ ,  $\{U_k\}$  is a sequence of mutually independent nonnegative random variables determining (when positive) the masses at these atoms,  $\nu(\cdot)$  is a fixed nonatomic boundedly finite measure on  $\mathcal{X}$ , and  $N(\cdot)$  is a Poisson process on  $\mathcal{X} \times (0, \infty)$ , independent of  $\{U_k\}$ , the parameter measure  $\mu$  of which may be unbounded on sets of the form  $A \times (0, \epsilon)$  but satisfies

$$\int_{y > \epsilon} \mu(A \times dy) < \infty, \quad (10.1.5a)$$

$$\int_{0 < y < \epsilon} y \mu(A \times dy) < \infty \quad (10.1.5b)$$

for every bounded Borel set  $A$  and every  $0 < \epsilon < \infty$ , and for all  $x \in \mathcal{X}$ ,

$$\mu(\{x\} \times (0, \infty)) = 0. \quad (10.1.5c)$$

PROOF. The complete independence property shows that the random masses  $U_k$  in the first component in (10.1.4), whose identification is assured by Proposition 9.3.IV, are mutually independent and independent also of the sum of the other two terms. So by considering  $\xi - \sum U_k \delta_{x_k}$ , we may assume that the completely random measure has no fixed atoms. Similar considerations imply (10.1.5c).

From Lemma 9.1.VII we can identify the component with random atoms as an extended MPP on  $\mathcal{X}$  with positive marks [Definition 9.1.VI(vi)],  $\tilde{N}$  say. To show that  $\tilde{N}$  inherits the completely random property, let  $V_j = A_j \times [a_j, b_j)$  ( $j = 1, 2$ ) be any two disjoint product sets of the form described above. If  $A_1 \cap A_2 = \emptyset$ , it is obvious that  $\tilde{N}(V_1)$  and  $\tilde{N}(V_2)$  are independent. Consider the other possibility, that  $A_1 = A_2 = A$  but  $[a_1, b_1)$  and  $[a_2, b_2)$  are disjoint. Let  $\mathcal{T}$  be a dissecting system for  $A$ , and set

$$\begin{aligned} X_{ni} &= \begin{cases} 1 & \text{if } a_1 \leq \xi(A_{ni}) < b_1, \\ 0 & \text{otherwise,} \end{cases} & Y_{ni} &= \begin{cases} 1 & \text{if } a_2 \leq \xi(A_{ni}) < b_2, \\ 0 & \text{otherwise,} \end{cases} \\ p_{ni} &= \mathcal{P}\{X_{ni} = 1\}, & q_{ni} &= \mathcal{P}\{Y_{ni} = 1\}. \end{aligned}$$



Then

$$\tilde{N}(V_1) = \lim_{n \rightarrow \infty} \sum_{i=1}^{k_n} X_{ni} \quad \text{a.s.}, \quad \tilde{N}(V_2) = \lim_{n \rightarrow \infty} \sum_{i=1}^{k_n} Y_{ni} \quad \text{a.s.},$$

and the complete independence property yields for the joint probability generating function (with  $|z_j| \leq 1$  for  $j = 1, 2$ )

$$\begin{aligned} E(z_1^{\tilde{N}(V_1)} z_2^{\tilde{N}(V_2)}) &= \lim_{n \rightarrow \infty} \prod_{i=1}^{k_n} E(z_1^{X_{ni}} z_2^{Y_{ni}}) \\ &= \lim_{n \rightarrow \infty} \prod_{i=1}^{k_n} [1 - p_{ni}(1 - z_1) - q_{ni}(1 - z_2)]. \end{aligned} \quad (10.1.6)$$

Similarly,

$$E(z_1^{\tilde{N}(V_1)}) E(z_2^{\tilde{N}(V_2)}) = \lim_{n \rightarrow \infty} \prod_{i=1}^{k_n} [1 - p_{ni}(1 - z_1)] [1 - q_{ni}(1 - z_2)]. \quad (10.1.7)$$

To establish the independence of  $\tilde{N}(V_1)$  and  $\tilde{N}(V_2)$ , it is enough to show that (10.1.6) and (10.1.7) are equal for all  $0 \leq z_j \leq 1$  ( $j = 1, 2$ ). Take logarithms and use the inequalities

$$x \leq -\log(1 - x) \leq x/(1 - x) \quad (0 \leq x < 1), \quad (10.1.8)$$

and write  $r_{ni} = p_{ni}(1 - z_1) + q_{ni}(1 - z_2)$ , Then for  $0 \leq z_j \leq 1$  in (10.1.6),

$$\begin{aligned} 0 &\leq -\sum_{i=1}^{k_n} \log(1 - r_{ni}) - \sum_{i=1}^{k_n} r_{ni} \\ &\leq R_n \equiv \sum_{i=1}^{k_n} \frac{r_{ni}^2}{1 - r_{ni}} \leq \frac{\max_i r_{ni}}{1 - \max_i r_{ni}} \sum_{i=1}^{k_n} r_{ni}. \end{aligned} \quad (10.1.9)$$

Now from Lemma 9.3.II we must have  $\max_i r_{ni} \rightarrow 0$  ( $n \rightarrow \infty$ ), and by (10.1.6) and the first inequality at (10.1.9),

$$\sum_{i=1}^{k_n} r_{ni} \leq -\sum_{i=1}^{k_n} \log(1 - r_{ni}) \rightarrow -\log E(z_1^{\tilde{N}(V_1)} z_2^{\tilde{N}(V_2)}),$$

which is finite, uniformly in  $n$ . So  $R_n \rightarrow 0$  as  $n \rightarrow \infty$ . Similar estimates apply to the two generating functions at (10.1.7), and because the difference of the leading terms in (10.1.6) and (10.1.7) vanishes, we must have equality of the generating functions as required.

By induction we can demonstrate the independence of any finite family  $\{\tilde{N}(V_j): j = 1, \dots, k\}$  whenever the sets  $V_j$  are rectangular and disjoint as considered. Now by Proposition 9.2.III and its corollary, the distribution of  $\tilde{N}$  is determined by all the joint distributions  $\{\tilde{N}(V_j): j = 1, \dots, k\}$ , and thus  $\tilde{N}(\cdot)$  is completely random. By construction  $\tilde{N}$  is a simple point process,

therefore Theorem 2.4.VII implies that  $\tilde{N}$  is a Poisson process with parameter measure  $\mu$  taking finite values on rectangles of the form already considered.

The tighter boundedness properties in (10.1.5) can be established as follows. First, because  $N_\epsilon(A) = \tilde{N}(A \times (\epsilon, \infty))$  is boundedly finite for fixed  $\epsilon > 0$  and (as just shown) is for fixed  $A$  a Poisson process in the range  $\infty > \epsilon > 0$ , the finiteness of (10.1.5a) is assured. For (10.1.5b), we observe moreover that for every  $\epsilon > 0$ ,

$$\int_0^\epsilon y \tilde{N}(A \times dy) \leq \xi(A) < \infty \quad \text{a.s.}, \quad (10.1.10)$$

because the integral is simply a sum (possibly an infinite series) of the contribution of the random atoms of  $\xi$  to its total mass  $\xi(A)$  on  $A$ . To see that the convergence of the integral at (10.1.10) implies the convergence of its expectation as at (10.1.5b), partition the interval  $(0, \epsilon)$  into the sequence of subintervals  $\{[\epsilon/(r+1), \epsilon/r) : r = 1, 2, \dots]\}$ . Then

$$\int_0^\epsilon y \mu(A \times dy) \leq \sum_{r=1}^\infty y_r \mu(A_r), \quad \text{where } y_r = \frac{\epsilon}{r} \text{ and } A_r = A \times [y_{r+1}, y_r].$$

Let  $\{\zeta_r\}$  be a sequence of independent Poisson r.v.s with parameters  $\mu(A_r)$ . Then for the Laplace transform of  $\sum y_r \zeta_r$  we have

$$\begin{aligned} -\log E \left[ \exp \left( -s \sum_{r=1}^\infty y_r \zeta_r \right) \right] &= -\log \prod_{r=1}^\infty \exp(-\mu(A_r)[1 - \exp(-sy_r)]) \\ &= \sum_{r=1}^\infty \mu(A_r)[1 - \exp(-sy_r)] \geq \frac{1}{2}s \sum_{r=1}^\infty y_r \mu(A_r) \end{aligned}$$

for  $0 < s < \epsilon^{-1}$ , because then  $0 < sy_r \leq 1$ .

Thus, the convergence of  $\sum_{r=1}^\infty y_r \mu(A_r)$  is implied by the a.s. convergence of  $\sum_{r=1}^\infty y_r \zeta_r$ , concerning which we have

$$\sum_{r=1}^\infty y_r \zeta_r = \sum_{r=1}^\infty \frac{\epsilon}{r} \zeta_r \geq \int_0^\epsilon y \tilde{N}(A \times dy) \geq \sum_{r=1}^\infty \frac{\epsilon}{r+1} \zeta_r \geq \frac{1}{2} \sum_{r=1}^\infty y_r \zeta_r.$$

The asserted finiteness at (10.1.5b) is now established.

Finally observe that the measure

$$\tilde{\nu}(A) \equiv \xi(A) - \sum U_k \delta_{x_k}(A) - \int_0^\infty y \tilde{N}(A \times dy)$$

is a.s. nonatomic (by construction) and completely random inasmuch as  $\xi$  is by assumption and the other two terms have been demonstrated to have the property. Then by Proposition 10.1.I,  $\tilde{\nu} = \nu$  a.s. for some fixed nonatomic measure  $\nu$ . The theorem is proved on noting that of the three terms in (10.1.4), the first consists of fixed atoms, the second is a constant measure, and the third is purely atomic, so uniqueness is assured.  $\square$

As a simple special case we obtain the Lévy-type representation for a process with nonnegative independent increments.

EXAMPLE 10.1(a) *Nonnegative Lévy processes.* Here we use ‘Lévy process’ to mean a process  $X(t)$  on the real line with independent increments; non-negativity ensures that the corresponding set process is a measure and not a signed measure. Thus, this example excludes the Brownian motion process and its fractional derivatives. The form of the representation (10.1.4) is unchanged; all that is required is the identification of  $\mathcal{X}$  with  $\mathbb{R}$ .

In applications, it is common to require the process to have increments that are both stationary and independent. Stationarity then rules out the existence of fixed atoms, the fixed measure  $\nu$  reduces to a multiple of Lebesgue measure, and the compound Poisson process in the representation (10.1.4) inherits the stationarity property from  $X(t)$ . Thus the representation takes the simpler form, for any finite interval  $(a, b]$ ,

$$X(b) - X(a) = \nu(b - a) + \int_0^\infty y N((a, b] \times dy),$$

where  $\nu$  is a nonnegative real constant, called the drift coefficient in, for example, Bertoin (1996, p. 16), and  $N$  is an extended stationary compound Poisson process, meaning that the intensity measure of the corresponding Poisson process on  $\mathbb{R} \times \mathbb{R}_+$  has the form  $\mu = \ell \times \Psi$  where  $\Psi$ , although not necessarily totally finite, does have finite total mass beyond any  $\epsilon > 0$  [corresponding to (10.1.5a)], and integrates  $y$  at the origin [corresponding to (10.1.5b)].  $\square$

It is often convenient to describe the representation of Theorem 10.1.III in terms of Laplace functionals, as in the proposition below [cf. Kingman (1967)]. Exercise 10.1.2 summarizes the corresponding representations of the Laplace–Stieltjes transforms for the process increments in the real line case; these are standard representations for the transforms of nonnegative infinitely divisible distributions.

**Proposition 10.1.IV.** *In order that the family  $\{\psi_A(\cdot), A \in \mathcal{B}_\mathcal{X}\}$  denote the Laplace transforms of the one-dimensional distributions of a completely random measure on  $\mathcal{X}$ , it is necessary and sufficient that  $\psi_A(\cdot)$  have a representation of the form, for  $\operatorname{Re}(s) \geq 0$ ,*

$$\log \psi_A(s) = - \sum_{k=1}^{\infty} \theta_k(s) \delta_{x_k}(A) - \int_0^\infty (1 - e^{-sy}) \mu(A \times dy) - s\nu(A), \quad (10.1.11)$$

where  $\{x_k\}$  is a fixed sequence of points, each  $\theta_k(\cdot)$  is the logarithm of the Laplace transform of a positive random variable, and the measures  $\nu, \mu$  have the same properties as in Theorem 10.1.III.

Conversely, given any such family  $\{x_k, \theta_k(\cdot), \nu, \mu\}$ , there exists a completely random measure with one-dimensional Laplace transforms given by (10.1.11).

PROOF. The representation (10.1.11) follows immediately on substituting for  $\xi(A)$  from (10.1.4) in the expectation  $\psi_A(s) = E(e^{-s\xi(A)})$ .

To prove the converse it is sufficient to show that the form (10.1.11), together with the definition of joint distributions through the completely random property, yields a consistent family of finite-dimensional distributions. The details of the verification are left as Exercise 10.1.1.  $\square$

EXAMPLE 10.1(b) *Stable random measures; nonnegative stable processes.* A special case of interest is the class of stable random measures for which the measure  $\mu$  of Theorem 10.1.III takes the form

$$\mu(dx \times dy) = \kappa(dx) y^{-(1+1/\alpha)} dy$$

for  $1 < \alpha < \infty$  and some boundedly finite measure  $\kappa(\cdot)$  on  $\mathcal{B}_{\mathcal{X}}$ . For such random measures the Laplace–Stieltjes transform of the one-dimensional distributions take the form

$$\psi_A(s) = E[e^{-\xi(A)s}] = \exp \left\{ -\kappa(A) \int_0^\infty \frac{1 - e^{-sy}}{y^{1+1/\alpha}} dy \right\} = \exp \{ -C_\alpha \kappa(A) s^{1/\alpha} \}, \quad (10.1.12)$$

where  $C_\alpha = \alpha \Gamma([\alpha - 1]/\alpha)$  [see, e.g., Bertoin (1996, p. 73)].

An alternative representation for the Laplace–Stieltjes transforms of nonnegative stable processes, due to Kendall (1963), is given in Exercises 10.1.2–3.

When  $\mathcal{X} = \mathbb{R}^d$  and  $\kappa$  reduces to Lebesgue measure, the process is both self-similar and stationary, with index of similarity  $\alpha$  (we also discuss self-similar random measures in Section 12.8). For more detail concerning stable random measures and related processes, see Samorodnitsky and Taqqu (1994).  $\square$

The case when the fidi distributions are gamma distributions has already been discussed in Example 9.1(d). A related but more extended family of completely random measures is described below, following Brix (1999) whose exposition covers earlier material.

EXAMPLE 10.1(c) *G-random measures.* Here the one-dimensional distributions have Laplace transforms (10.1.11) of the form

$$\psi_A(s) = \exp \{ -\kappa(A)[(\theta + s)^\rho - \theta^\rho]/\rho \},$$

where  $\rho$  and  $\theta$  are parameters satisfying either  $\rho \leq 0$  and  $\theta > 0$ , or  $0 < \rho \leq 1$  and  $\theta \geq 0$ . The case  $\rho = 0$  can be obtained as a limit for  $\rho \downarrow 0$ , and gives back a gamma distribution for  $\psi_A(s)$ .

When  $\rho < 0$ , the underlying measure  $\mu$  of (10.1.11) is a product of Lebesgue measure and a gamma distribution (i.e., in this case the jumps have gamma distributed heights). In the case  $0 < \rho < 1$  the corresponding density is improper (its integral diverges) but still satisfies the condition  $\int_0^\infty y f(y) dy < \infty$ , implying that conditions (10.1.5) hold. The Lévy representation has density [cf. (10.1.11)]

$$\mu(A \times dy) = \frac{\kappa(A)}{\Gamma(1 - \rho)} y^{-\rho-1} e^{-\theta y} dy.$$

Lee and Whitmore (1993) describe the Lévy processes corresponding to the one-dimensional versions of these processes as *Hougarde processes*.

Brix (1999) also describes the use of these  $G$ -measures, or smoothed versions thereof, as directing measures for a Cox process. In one dimension, that is, when  $\mathcal{X} = \mathbb{R}$ , the smoothed version can then be made to correspond to a type of shot-noise process; Exercise 10.1.7 gives some details.  $\square$

No essentially new ideas arise in extending the complete randomness property to marked point processes. We say that an MPP on  $\mathcal{X}$  with marks in  $\mathcal{K}$  is completely random when the associated point process on  $\mathcal{X} \times \mathcal{K}$  is completely random.

It is somewhat surprising that when the MPP has a simple ground process  $N_g$ , this condition is equivalent to the apparently weaker condition that the random variables  $N(A_i \times K_i)$  should be mutually independent whenever the sets  $A_i$  are disjoint, irrespective of whether the corresponding sets  $K_i$  are disjoint. Because the construction of Exercise 9.1.6 indicates that, by adjusting the mark space if necessary, we can always find an equivalent description of an MPP as an MPP with simple ground process, the lemma below is rather generally applicable.

**Lemma 10.1.V.** *An MPP with simple ground process  $N_g$  is completely random if and only if for every finite  $n$ , bounded  $A_i \in \mathcal{B}_{\mathcal{X}}$  and  $K_i \in \mathcal{B}_{\mathcal{K}}$  ( $i = 1, \dots, n$ ), the random variables  $N(A_i \times K_i)$  are mutually independent whenever the  $A_i$  are mutually disjoint.*

**PROOF.** It is obvious that if the complete randomness property holds, the asserted independence property holds because sets in a product space with disjoint marginals are disjoint.

For the converse, suppose given two product sets in the product space,  $A_j \times K_j$  say, for  $j = 1, 2$ . Consider first the case  $A_1 = A_2 = A$  say but  $K_1 \cap K_2 = \emptyset$ . We want to show that under the condition of  $N_g$  being simple, the  $N(A \times K_j)$  are independent. Let  $\mathcal{T}$  be a dissecting system for  $A$ , and consider for elements  $A_{ni}$  of a partition

$$X_{ni,j} = \min(1, N(A_{ni} \times K_j)).$$

Simplicity of  $N_g$  implies that  $N(A \times K_j) = \lim_{n \rightarrow \infty} \sum_i X_{ni,j}$ . We can now imitate that part of the proof of Theorem 10.1.III around (10.1.6–9) to conclude that the  $N(A \times K_j)$  are independent.

In the general case, with both  $A_{12} = A_1 \cap A_2$  and  $A'_1 \cup A'_2$  nonempty, where  $A'_j = A_j \setminus A_{12}$  for  $j = 1, 2$ , the product sets  $A'_j \times K$  are disjoint in their  $\mathcal{X}$ -components and therefore independent, and  $N(A_{12} \times K_j)$  are independent for disjoint  $K_j$  by the case already considered.

Independence of  $N(V_j)$  ( $j = 1, 2$ ) for arbitrary bounded Borel sets  $V_j$  in the product space follows from their independence when the  $V_j$  are product sets by standard extension arguments.

The argument extends to any finite number of sets by induction.  $\square$

We proceed to examine the structure of a completely random MPP. We know from Chapter 2 (or as a special case of Theorem 10.1.III) that a simple

completely random point process reduces to a Poisson process; in the present case the parameter measure  $\mu_g$  of the ground process must satisfy

$$\mu_g(A) = E[N(A \times \mathcal{K})] < \infty \quad (10.1.13)$$

for all bounded  $A \in \mathcal{B}_{\mathcal{X}}$ , inasmuch as  $N_g$  is boundedly finite by assumption. Introduce a family of probability measures  $P(K | x)$  on the mark space  $(\mathcal{K}, \mathcal{B}(\mathcal{K}))$  by means of the Radon–Nikodym derivatives

$$E[N(A \times K)] \equiv \mu(A \times K) = \int_A P(K | x) \mu_g(dx). \quad (10.1.14)$$

Then the absolute continuity condition  $\mu(\cdot \times K) \ll \mu_g$  is satisfied, and the property  $P(K | x) = 1$  a.s. follows from the definition of  $\mu_g$ . As in the discussion of regular conditional probability (Proposition A1.5.III), we can and do assume that the family  $\{P(B | x): B \in \mathcal{B}(\mathcal{K}), x \in \mathcal{X}\}$  is so chosen that  $P(\cdot | x)$  is a probability measure on  $\mathcal{B}(\mathcal{K})$  for all  $x \in \mathcal{X}$ . With this understanding we arrive at the next proposition which effectively implies that completely independent MPPs reduce to some general type of compound Poisson process.

**Proposition 10.1.VI.** *A completely random MPP with simple ground process is fully specified by the two components:*

- (i) *a Poisson process of locations with parameter measure  $\mu_g$ ; and*
- (ii) *a family of probability distributions  $P(\cdot | x)$  giving the distribution of the mark in  $\mathcal{K}$  with the property that  $P(B | x)$  is measurable in  $x$  for each fixed  $B \in \mathcal{B}(\mathcal{K})$ .*

*Conversely, given such  $\mu_g$  and  $P(\cdot | \cdot)$ , there exists a completely random MPP having these as components.*

PROOF. For the converse, it suffices to construct a Poisson process on  $\mathcal{X} \times \mathcal{K}$  with parameter measure (10.1.14); we leave it to Exercise 10.1.6 to verify that the resultant process is an MPP with the complete randomness property.  $\square$

## Exercises and Complements to Section 10.1

10.1.1 Imitate the discussion of Example 9.2(a) to verify that the fidi distributions of completely random measure, constructed from the one-dimensional distributions as at (10.1.12), satisfy Conditions 9.2.V and 9.2.VI.

10.1.2 Let  $\xi$  be a stationary, completely random measure on  $\mathbb{R}$ , and let  $\psi_t(s)$  denote the Laplace–Stieltjes transform of  $\xi(0, t]$ .

(a) Deduce from Proposition 10.1.IV that

$$\psi_t(s) = \exp \left( -st\nu - t \int_{(0, \infty)} [1 - e^{-sy}] \Psi(dy) \right), \quad (10.1.15)$$

where  $\nu$  is a positive constant, and the  $\sigma$ -finite measure  $\Psi$  on  $(0, \infty)$  satisfies, for some  $\epsilon > 0$ ,  $\int_0^\epsilon y \Psi(dy) < \infty$  and  $\int_\epsilon^\infty \Psi(dy) < \infty$ .

(b) Establish also the equivalent representation

$$\psi_t(s) = \exp \left( -st\nu - t \int_{(0,\infty]} \frac{1 - e^{-sy}}{1 - e^{-y}} G(dy) \right) \quad (10.1.16)$$

for some totally finite measure  $G$  and some nonnegative finite constant  $\nu$  [ $\nu$  here equals  $\nu(0, 1]$  in the notation of (10.1.11)].

[*Remark:* This form parallels that given in Kendall (1963); the measure  $\mu$  of (10.1.5) and (10.1.11) satisfies  $\int_{(0,\infty)} (1 - e^{-y}) \mu(A \times dy) < \infty$ .]

10.1.3 (*Continuation*). Using the representation above and the same notation, show that  $\mathcal{P}\{\xi(0, t] = 0\} > 0$  if and only if both  $\int_{(0,1)} y^{-1} G(dy) < \infty$  and  $\nu = 0$ .

[*Remark:* The condition  $\nu = 0$  precludes any positive linear trend; the other condition precludes the possibility of an everywhere dense set of atoms.]

10.1.4 For a given random measure  $\xi$  let  $\mathcal{Y}$  denote the family of measures  $\eta$  satisfying

$$\mathcal{P}\{\xi(A) \geq \eta(A) \text{ (all } A \in \mathcal{B}_{\mathcal{X}})\} = 1.$$

(a) Define  $\nu_d(A) = \sup_{\eta \in \mathcal{Y}} \eta(A)$ ; check that it is a measure, and confirm that  $\xi - \nu_d$  is a random measure.

(b) Extract from  $\xi - \nu_d$  the random measure  $\zeta_a$  consisting of all the fixed atoms of  $\xi - \nu_d$ , leaving  $\xi_r = \xi - \nu_d - \zeta_a$ .

(c) The result of (a) and (b) is to effect a decomposition  $\xi = \nu_d + \zeta_a + \xi_r$  of  $\xi$  into a deterministic component  $\nu_d$ , a component of fixed atoms  $\zeta_a$ , and a random component  $\xi_r$ . Give an example showing that there may still be bounded  $A \in \mathcal{B}_{\mathcal{X}}$  for which  $\mathcal{P}\{\xi_r(A) \geq \epsilon\} = 1$  for some  $\epsilon > 0$ .

[*Hint:* Let  $\xi_r$  give mass 1 to either  $U$  or  $U + 1$ , where  $U$  is uniformly distributed on  $(0, 1)$ .]

10.1.5 Proposition 10.1.IV coupled with the independence property for disjoint sets  $A$  implies that the Laplace functional (9.4.10) of a completely random measure is expressible for  $f \in \text{BM}_+(\mathcal{X})$  as

$$\begin{aligned} -\log \left\{ \mathbb{E} \left[ \exp \left( - \int_{\mathcal{X}} f(x) \xi(dx) \right) \right] \right\} \\ = \int_{\mathcal{X}} f(x) \alpha(dx) + \int_{\mathcal{X} \times \mathbb{R}_+} (1 - e^{-yf(x)}) \mu(dx \times dy), \end{aligned}$$

where  $\alpha \in \mathcal{M}_{\mathcal{X}}^{\#}$  and  $\gamma \in \mathcal{M}^{\#}(\mathcal{X} \times \mathbb{R}_+)$  satisfies  $\int_{\mathbb{R}_+} (1 - e^{-y}) \gamma(B \times dy) < \infty$  for all bounded  $B \in \mathcal{B}_{\mathcal{X}}$ . [*Hint:* Kallenberg (1983a, Chapter 7) gives an alternative proof. Compare also with Proposition 10.2.IX.]

10.1.6 Verify the assertion that a completely independent MPP has a simple ground process  $N_g$  if and only if (10.1.12) holds for every dissecting system  $\mathcal{T}$  for bounded  $A \in \mathcal{B}_{\mathcal{X}}$ . Without complete independence (10.1.12) need not hold [see (9.3.18) and Proposition 9.3.XII].

10.1.7 *Cox processes directed by stationary  $G$ -processes.* Let  $\xi$  be a stationary  $G$ -random measure on  $\mathbb{R}^d$  as in Example 10.1(c) so that in the notation of the example,  $\kappa(A) = \kappa\ell(A)$  for some finite positive constant  $\kappa$ .

- (a) Show that if  $\xi$  itself is used as the directing measure of a Cox process, then the realizations are stationary but a.s. not simple.
- (b) In the case  $d = 1$ , suppose the directing process is not  $\xi$  but the smoothed version  $X(y) = \int_{-\infty}^y \phi(y-x) \xi(dx)$  for some continuous nonnegative integrable kernel function  $\phi(\cdot)$ ;  $X(\cdot)$  has a density and can be regarded as a type of general shot-noise process [cf. Examples 6.1(d) and 6.2(a)]. Show that the Cox process is well-defined, stationary, a.s. simple, with finite mean rate  $m = \theta^{\alpha-1} \kappa \int_{\mathbb{R}_+} \phi(u) du$  and reduced factorial covariance density

$$c_{[2]}(u) = \kappa \theta^{\alpha-2} (1 - \alpha) \int_{\mathbb{R}_+} \phi(x) \phi(x+u) dx.$$

## 10.2. Infinitely Divisible Point Processes

Our aim in this section is to characterize the class of infinitely divisible point processes and random measures. In the point process case, this question is intimately bound up with characterizations of Poisson cluster processes. The role of infinitely divisible point processes and random measures in limit theorems for superpositions is taken up in the next chapter (Section 11.2).

**Definition 10.2.I.** *A point process or random measure is infinitely divisible if, for every  $k$ , it can be represented as the superposition of  $k$  independent, identically distributed, point process (or random measure) components.*

In symbols, a point process  $N$  is infinitely divisible if, for every  $k$ , we can write

$$N = N_1^{(k)} + \cdots + N_k^{(k)}, \quad (10.2.1)$$

where the  $N_i^{(k)}$  ( $i = 1, \dots, k$ ) are i.i.d. components. Using p.g.fl.s, the condition takes the form (in an obvious notation)

$$G[h] = (G_{1/k}[h])^k \quad (h \in \mathcal{V}(\mathcal{X})). \quad (10.2.2)$$

The p.g.fl. is nonnegative for such  $h$ , so we can restate (10.2.2) as follows. A point process is infinitely divisible if and only if, for every  $k$ , the uniquely defined nonnegative  $k$ th root of its p.g.fl. is again a p.g.fl. Similarly for random measures the defining property can be restated as follows. A random measure is infinitely divisible if and only if, for every integer  $k > 0$ , the uniquely defined  $k$ th root of its Laplace functional is again a Laplace functional.

From these remarks we may immediately verify that, for example, a Poisson process is infinitely divisible (replace the original parameter measure  $\mu$  by the measure  $\mu/k$  for each component), as, more generally, are the Poisson cluster



processes studied in Section 6.3 (replace the parameter measure  $\mu_c$  for the cluster centre process by  $(\mu_c)/k$  and leave the cluster structure unaltered).

In the point process case, any fidi distribution has a joint p.g.f. expressible as  $G[h_A]$ , where  $h_A(\cdot)$  is of the form

$$h_A(x) = 1 - \sum_{i=1}^n (1 - z_i) I_{A_i}(x) \quad (10.2.3)$$

for appropriate subsets  $A_i$  of the set  $A$ . Then from (10.2.2) it follows that the fidi distributions of an infinitely divisible point process are themselves infinitely divisible. Conversely, when a point process has its fidi distributions infinitely divisible, (10.2.2) holds for functions  $h$  of the form (10.2.3). Because such functions are dense in  $\mathcal{V}(\mathcal{X})$ , it follows by continuity as in Theorem 9.4.V that p.g.f.s like  $G[h_A]$  and  $G_{1/k}[h_A]$  define p.g.f.s. Similar arguments apply to the case of a random measure, thereby proving the following lemma.

**Lemma 10.2.II.** *A point process or random measure is infinitely divisible if and only if its fidi distributions are infinitely divisible.*

We now embark on a systematic exploitation of this remark and the results set out in earlier sections concerning the representation of infinitely divisible discrete distributions (see, in particular, Exercises 2.2.2–3). We first consider the case of a finite point process.

**Proposition 10.2.III.** *Suppose that the point process  $N$  with p.g.fl.  $G[\cdot]$  is a.s. finite and infinitely divisible. Then there exists a uniquely defined, a.s. finite point process  $\tilde{N}$ , such that  $\Pr\{\tilde{N} = \emptyset\} = 0$ , and a finite positive number  $\alpha$  such that*

$$G[h] = \exp(\alpha(\tilde{G}[h] - 1)) \quad (h \in \mathcal{V}(\mathcal{X})), \quad (10.2.4)$$

where  $\tilde{G}$  is the p.g.fl. of  $\tilde{N}$  and  $\emptyset$  denotes the null measure.

Conversely, any functional of the form (10.2.4) represents the p.g.fl. of an a.s. finite point process that is infinitely divisible.

PROOF. It is clear that any functional of the form (10.2.4) is a p.g.fl. and that the point process to which it corresponds is infinitely divisible (replace  $\alpha$  by  $\alpha/k$  and take  $k$ th powers). It is also a.s. finite if  $\tilde{N}$  is a.s. finite, because if  $\tilde{G}[\rho I_{\mathcal{X}}] \rightarrow 1$  as  $\rho$  increases to 1, then also  $G[\rho I_{\mathcal{X}}] \rightarrow 1$ , implying  $N$  is a.s. finite (see Exercise 9.4.5).

Suppose conversely that  $N$  is infinitely divisible and a.s. finite, and consider its p.g.fl. When  $h$  has the special form  $\sum_{i=1}^n z_i I_{A_i}(\cdot)$ , where  $A_1, \dots, A_n$  is a measurable partition of  $\mathcal{X}$ , we know from Exercise 2.2.3 that  $G[h]$ , which then reduces to the multivariate p.g.f.  $P(z_1, \dots, z_n)$  of the random variables  $N(A_1), \dots, N(A_n)$ , can be represented in the form

$$P(z_1, \dots, z_n) = \exp(\alpha[Q(z_1, \dots, z_n) - 1]),$$

where  $Q$  is itself a p.g.f. with  $Q(0, \dots, 0) = 0$  and  $\alpha$  is positive, independent of the choice of the partition and equal to  $-\log(\Pr\{N(\mathcal{X}) = 0\})$ . Now consider the function

$$\tilde{G}[h] = 1 + \alpha^{-1} \log G[h].$$

When  $h$  has the above special form,  $\tilde{G}$  reduces to the multivariate p.g.f.  $Q$ . Also,  $\tilde{G}$  inherits continuity from  $G$ . Hence, it is a p.g.fl. by Theorem 9.4.V. To show that the resulting process is a.s. finite consider the behaviour of  $G[\rho I_{\mathcal{X}}]$  as  $\rho$  increases to 1. Because  $N$  itself is a.s. finite,  $G[\rho I_{\mathcal{X}}] \rightarrow 1$  by Exercise 9.4.5. But then  $\log G[\rho I_{\mathcal{X}}] \rightarrow 0$  and so  $\tilde{G}[\rho I_{\mathcal{X}}] \rightarrow 1$ , showing that  $\tilde{G}$  is the p.g.fl. of a point process  $\tilde{N}$  that is a.s. finite.  $\square$

The representation (10.2.4) has a dual interpretation. It shows that any a.s. finite and infinitely divisible process  $N$  can be regarded as the ‘Poisson randomization’ [borrowing a phrase from Milne (1971)] of a certain other point process  $\tilde{N}$ . In this interpretation, the process  $N$  is constructed by first choosing a random integer  $K$  according to the Poisson distribution with probabilities

$$p_n = e^{-\alpha} \alpha^n / n!,$$

and then, given  $K$ , taking the superposition of  $K$  i.i.d. components each having the same distribution as  $\tilde{N}$ .

On the other hand, the process  $N$  can also be related to the cluster processes of Section 6.3. To see this, first represent the p.g.fl.  $\tilde{G}$  in terms of the Janossy measures for  $\tilde{N}$ , so that (10.2.4) becomes

$$\log G(h) = \alpha \left( \sum_{k=1}^{\infty} \frac{1}{k!} \int \cdots \int_{\mathcal{X}^{(k)}} h(x_1) \cdots h(x_k) \tilde{J}_k(dx_1 \times \cdots \times dx_k) - 1 \right).$$

This infinite sum can be rewritten as

$$\log G(h) = \sum_{k=1}^{\infty} \int \cdots \int_{\mathcal{X}^{(k)}} [h(x_1) \cdots h(x_k) - 1] Q_k(dx_1 \times \cdots \times dx_k), \quad (10.2.5)$$

where  $Q_k(\cdot) = (\alpha/k!) \tilde{J}_k(\cdot)$ . Observe finally that this last form is the log p.g.fl. of a Poisson cluster process, as in Proposition 6.3.V.

Both interpretations above coexist for an a.s. finite process: they represent alternative constructions for the same process.

To investigate the behaviour when the a.s. finite condition is relaxed, we first observe that any infinitely divisible process remains infinitely divisible but becomes a.s. finite when we consider its restriction to any bounded Borel set. Its local representation therefore continues to have the form (10.2.4). Rewrite (10.2.4) for the special case that the bounded set is a (large) sphere,  $S_n$  say, and introduce explicitly the distribution,  $\tilde{\mathcal{P}}_n$  say, of the process  $\tilde{N}$

restricted to  $\mathcal{B}(S_n)$ . Writing  $G_n[\cdot]$  for the corresponding p.g.fl. of  $N$ , we have from (10.2.4) that

$$G_n[h] = \exp \left[ \alpha_n \int_{\tilde{N} \in \mathcal{N}^\#(S_n)} \left( \exp \left[ \int_{S_n} \log h(x) \tilde{N}(dx) \right] - 1 \right) \tilde{\mathcal{P}}_n(d\tilde{N}) \right], \quad (10.2.6)$$

where we recall the convention that the inner exponential term is to be counted as unity if  $\tilde{N}$  has no points in the region where  $h$  differs from unity, and as zero if  $\tilde{N}$  has any points in the region where  $h$  vanishes. Bearing this in mind, we have in particular, from (10.2.6),

$$\tilde{Q}_n\{\tilde{N}: \tilde{N}(S_n) > 0\} \equiv \alpha_n \tilde{\mathcal{P}}_n\{\tilde{N}: \tilde{N}(S_n) > 0\} = -\log \mathcal{P}\{N(S_n) = 0\}, \quad (10.2.7)$$

where  $\tilde{Q}_n$  is an abbreviation for  $\alpha_n \tilde{\mathcal{P}}_n$  and we continue the convention that  $\tilde{\mathcal{P}}_n\{\tilde{N}(S_n) = 0\} = 0$ , so that  $e^{-\alpha_n}$  is just the probability that the original process has no points in  $S_n$ .

Each measure  $\tilde{Q}_n$  may be used to induce a similar measure,  $\tilde{Q}_n^*$  say, on the class of cylinder sets in the full space  $\mathcal{N}_\mathcal{X}^\#$  determined by conditions on the behaviour of the counting process on  $S_n$ . Specifically, for  $C$  a set in  $\mathcal{N}_{S_n}^\#$  of the form

$$C = \{\tilde{N} \in \mathcal{N}_{S_n}^\#: \tilde{N}(A_i) = r_i, A_i \subseteq S_n, i = 1, \dots, k\},$$

where the  $r_i$  are nonnegative integers not all zero, we associate the set  $C^*$  in  $\mathcal{N}_\mathcal{X}^\#$  given by  $C^* = \{\tilde{N} \in \mathcal{N}_\mathcal{X}^\#: \tilde{N}(A_i) = r_i, A_i \subseteq S_n, i = 2, \dots, k\}$ , and put

$$\tilde{Q}_n^*(C^*) = \tilde{Q}_n(C).$$

This construction fails for the set in  $\mathcal{N}_\mathcal{X}^\#$  for which  $\tilde{N}(S_n) = 0$ : for this reason we have to define  $\tilde{Q}_n^*$  not on the full sub- $\sigma$ -algebra of cylinder sets with base determined by conditions in  $S_n$ , but on the sub- $\sigma$ -algebra generated by those cylinder sets incorporating the condition  $\tilde{N}(S_n) > 0$ . Let us denote this sub- $\sigma$ -algebra by  $\mathcal{B}_n$ . Then it is clear that the  $\mathcal{B}_n$  are monotonic increasing and that

$$\sigma\left(\bigcup_{n=1}^{\infty} \mathcal{B}_n\right) = \mathcal{B}(\mathcal{N}_0^\#(\mathcal{X})),$$

where  $\mathcal{N}_0^\#(\mathcal{X})$  denotes the space  $\mathcal{N}_\mathcal{X}^\#$  with the null measure  $\emptyset(\cdot)$  omitted. On the union  $\bigcup_{n=1}^{\infty} \mathcal{B}_n$ , we can consistently define a set function  $\tilde{Q}^*$ , the projective limit of  $\{\tilde{Q}_n^*\}$ , by setting

$$\tilde{Q}^*(A) = \tilde{Q}_n^*(A)$$

whenever  $A \in \mathcal{B}_n$ . This is possible because  $\tilde{Q}_m^*$  reduces to  $\tilde{Q}_n^*$  whenever  $m > n$  and we restrict attention to sets in  $\mathcal{B}_n$ . The set function  $\tilde{Q}^*$  is countably additive on each of the  $\mathcal{B}_n$  but not obviously so on their union. The situation, however, is similar to that of the Kolmogorov extension theorem

for stochastic processes, or to the extension theorem considered in Section 9.2, where countable additivity is ultimately a consequence of the metric assumptions imposed on the space  $\mathcal{X}$ . The same argument applies here also; we leave the details to Exercise 10.2.1. It implies that  $\tilde{Q}^*$  has a unique extension to a measure  $\tilde{Q}$  on the  $\sigma$ -algebra  $\mathcal{B}(\mathcal{N}_0^\#(\mathcal{X}))$ .

In addition to the fact that it is defined on the sets of  $\mathcal{N}_0^\#(\mathcal{X})$  rather than  $\mathcal{N}_\mathcal{X}^\#$ ,  $\tilde{Q}$  enjoys one further special property. Equation (10.2.7) implies that for any bounded set  $A$ ,

$$\tilde{Q}\{N: N(A) > 0\} < \infty. \quad (10.2.8)$$

**Definition 10.2.IV.** A boundedly finite measure  $\tilde{Q}$  defined on the Borel sets of  $\mathcal{N}_0^\#(\mathcal{X}) = \mathcal{N}_\mathcal{X}^\# \setminus \{N(\mathcal{X}) = 0\}$ , and satisfying the additional property (10.2.8), is called a KLM measure.

The measure is so denoted for basic contributions to the present theory in Kerstan and Matthes (1964) and Lee (1964, 1967).

**Theorem 10.2.V.** A point process on the c.s.m.s.  $\mathcal{X}$  is infinitely divisible if and only if its p.g.fl. can be represented in the form

$$G[h] = \exp \left[ \int_{\mathcal{N}_0^\#(\mathcal{X})} \left( \exp \left[ \int_{\mathcal{X}} \log h(x) \tilde{N}(dx) \right] - 1 \right) \tilde{Q}(d\tilde{N}) \right] \quad (10.2.9)$$

for some KLM measure  $\tilde{Q}$ . When such a representation exists, it is unique.

PROOF. Suppose that the point process  $N$  is infinitely divisible, and let  $\tilde{Q}$  be the KLM measure constructed as above. When it is equal to unity outside the sphere  $S_n$ , the representation (10.2.9) reduces to (10.2.6) from the construction of  $\tilde{Q}$ , and so the functional  $G$  in (10.2.9) must coincide with the p.g.fl. of the original process.

Conversely, suppose a KLM measure  $\tilde{Q}$  is given, and consider (10.2.9). If we set

$$\alpha_n = \tilde{Q}\{\tilde{N}: \tilde{N}(S_n) > 0\}$$

(finite by assumption in Definition 10.2.IV), (10.2.9) can again be reduced to the form (10.2.6) for functions  $h$  that equal unity outside  $S_n$ , and therefore, by Proposition 10.2.III, it is the p.g.fl. of a local process defined on  $S_n$ . In particular, therefore, (10.2.9) reduces to a joint p.g.f. when  $h$  has the form  $1 - \sum_{i=1}^k (1 - z_i) I_{A_i}$ . The continuity condition follows from the remark already made that (10.2.9) defines a local p.g.fl. when we restrict attention to the behaviour of the process on  $S_n$ . Thus, (10.2.9) is itself a p.g.fl. Infinite divisibility follows from Lemma 10.2.II and the remarks already made concerning the local behaviour of  $G[h]$ . Finally, uniqueness follows from the construction and the uniqueness part of Proposition 10.2.III.  $\square$

EXAMPLE 10.2(a) *Poisson process* [see Example 9.4(c)]. If (10.2.9) is to reduce to the p.g.fl. (9.4.17) of a Poisson process, each  $\tilde{N}$  must be simple and have a

single point as its support; that is, we must have  $\tilde{Q}\{\tilde{N}(\mathcal{X}) \neq 1\} = 0$ , because for given  $\tilde{N}$  the integrand for the integral at (10.2.9) with respect to  $\tilde{Q}$  must reduce to  $h(x) - 1$ , where  $\{x\}$  is the singleton support of  $N$ . In fact, the KLM measure  $\tilde{Q}$  must be related to the parameter measure  $\mu$  by

$$\tilde{Q}\{\tilde{N}(A) = 1\} = \tilde{Q}\{\tilde{N}: \tilde{N}(\mathcal{X}) = 1 = \tilde{N}(A)\} = \mu(A) \quad (\text{bounded } A \in \mathcal{B}_{\mathcal{X}}). \quad \square$$

Further insight into the structure of such processes can be obtained from a classification of the properties of their KLM measures. In particular, we make the following definitions.

**Definition 10.2.VI.** *An infinitely divisible point process is regular if its KLM measure is carried by the set*

$$V_r \equiv \{\tilde{N}: \tilde{N}(\mathcal{X}) < \infty\} \quad (10.2.10a)$$

*and singular if its KLM measure is carried by the complementary set*

$$V_s \equiv \{\tilde{N}: \tilde{N}(\mathcal{X}) = \infty\}. \quad (10.2.10b)$$

We now have the following decomposition result.

**Proposition 10.2.VII.** *Every infinitely divisible point process can be represented as the superposition of a regular infinitely divisible process and a singular infinitely divisible process, the two components being independent.*

PROOF. This follows from the representation (10.2.9) on writing

$$\tilde{Q} = \tilde{Q}_r + \tilde{Q}_s,$$

where for each  $A \in \mathcal{B}(\mathcal{N}_0^\#(\mathcal{X}))$ ,

$$\tilde{Q}_r(A) = \tilde{Q}(A \cap V_r), \quad \tilde{Q}_s(A) = \tilde{Q}(A \cap V_s).$$

Each of  $\tilde{Q}_r(\cdot)$  and  $\tilde{Q}_s(\cdot)$  is again a KLM measure, and because the original p.g.fl. appears as the product of the p.g.fl.s of the two components, the corresponding components themselves must be independent and their superposition must give back the original process.  $\square$

Further characterizations of various classes of infinitely divisible point processes can be given in terms of their KLM measures, as set out below. Some refinements for the stationary case are given in Section 12.4.

**Proposition 10.2.VIII.**

- (i) *An infinitely divisible point process is a.s. finite if and only if it is regular and its KLM measure is totally finite.*

- (ii) *An infinitely divisible point process can be represented as a Poisson cluster process, with a.s. finite clusters, if and only if it is regular.*
- (iii) *An infinitely divisible point process can be represented as a Poisson randomization if and only if its KLM measure is totally finite.*

PROOF. Part (i) is a restatement of Proposition 10.2.III, regularity coming from the assertion that the process  $N$  is a.s. finite, and the total boundedness of the KLM measure from the fact that it can be represented in the form  $\alpha \tilde{\mathcal{P}}$ , where  $0 < \alpha < \infty$  and  $\tilde{\mathcal{P}}$  is a probability measure.

Part (ii) follows from the representation of Poisson cluster processes in Proposition 6.3.V, taking  $\mathcal{X} = \mathcal{Y}$  there, so that the measures  $Q_k(\cdot)$  in that proposition can be combined to give a measure on the space of all finite counting measures, as in Proposition 5.3.II. That this is a KLM measure follows from the absence of any  $Q_0$  term, and the condition (6.3.33), which in the terminology of the present section can be rewritten [see (6.3.35)] as

$$\mu(A) = \tilde{Q}\{\tilde{N}: \tilde{N}(A) > 0\} < \infty \quad (\text{bounded } A \in \mathcal{B}_{\mathcal{X}}).$$

Conversely, we can split the KLM measure of any regular infinitely divisible process into its components on the sets  $V_k = \{\tilde{N}: \tilde{N}(\mathcal{X}) = k\}$ , in each of which it induces a measure  $Q_k$  with the properties described in Proposition 6.3.V.

Finally, part (iii) follows from the observation that here also the KLM measure can be written in the form  $\tilde{Q} = \alpha \tilde{\mathcal{P}}$ , where  $\alpha$  is the parameter of the Poisson randomizing distribution and  $\tilde{\mathcal{P}}$  is the distribution of the point process being randomized, which we assume adjusted if necessary so that  $\tilde{\mathcal{P}}(\{\tilde{N}: \tilde{N}(\mathcal{X}) = 0\}) = 0$ .  $\square$

Extensions to infinitely divisible multivariate and marked point processes are considered in Exercises 10.2.2 and 10.2.5. An analogous representation holds also for infinitely divisible random measures, and is set out below.

**Proposition 10.2.IX.** *A random measure on the c.s.m.s.  $\mathcal{X}$  is infinitely divisible if and only if its Laplace functional can be represented in the form for all  $f \in \text{BM}_+(\mathcal{X})$*

$$-\log L[f] = \int_{\mathcal{X}} f(x) \alpha(dx) + \int_{\mathcal{M}_0^{\#}(\mathcal{X})} \left[ 1 - \exp \left( - \int_{\mathcal{X}} f(x) \eta(dx) \right) \right] \Lambda(d\eta), \quad (10.2.11)$$

where  $\alpha \in \mathcal{M}_{\mathcal{X}}^{\#}$ ,  $\mathcal{M}_0^{\#}(\mathcal{X}) = \mathcal{M}_{\mathcal{X}}^{\#} - \{\emptyset\}$ , and  $\Lambda$  is a  $\sigma$ -finite measure on  $\mathcal{M}_0^{\#}(\mathcal{X})$  satisfying, for every bounded Borel set  $B \in \mathcal{B}_{\mathcal{X}}$  and distribution  $F_1(B; x) \equiv \Lambda\{\eta: \eta(B) \leq x\}$ ,

$$\int_{\mathbb{R}_+} (1 - e^{-x}) F_1(B; dx) < \infty. \quad (10.2.12)$$

A proof involving the inductive limit of the corresponding representations for the Laplace transforms of the fidi distributions can be given along the same

lines as that of Theorem 10.2.V, with (10.2.11) reducing to (10.2.9) when  $\xi$  is a point process. See Exercise 10.2.5.

Results about the convergence of infinitely divisible distributions and their role in limit theorems are reviewed in Chapter 11, notably Section 11.2. Results for the stationary case are outlined in Section 12.4.

## Exercises and Complements to Section 10.2

- 10.2.1 *Kolmogorov extension theorem analogue for  $\tilde{Q}^*$ .* Show that the measure  $\tilde{Q}^*$  defined below (10.2.7) admits consistent fidi distributions in the sense that

$$\tilde{Q}^*\{N: N(A_i) = k_i, i = 1, \dots, n\}$$

satisfy the two consistency Conditions 9.2.V, namely, marginal consistency and symmetry under permutations. Show also that  $\tilde{Q}^*$  is finitely additive and continuous in the sense that for disjoint  $A_i$ ,

$$\tilde{Q}^*\left\{N: N\left(\bigcup_{i=1}^n A_i\right) \neq \sum_{i=1}^n N(A_i)\right\} = 0,$$

and

$$\tilde{Q}^*\{N: N(A_n) > 0\} \rightarrow 0 \quad \text{when } A_n \downarrow \emptyset.$$

[For the latter, write  $V_n = \{N: N(A_n) > 0\}$ . Because  $N(A_n) \downarrow 0$  for all  $N \in \mathcal{N}_{\mathcal{X}}$ ,  $\{V_n\}$  is a monotonic decreasing sequence of sets, say  $V_n \downarrow V$ . Supposing  $N_0 \in V$ , then  $N_0(A_n) > 0$  for all  $n$  giving a contradiction.  $\tilde{Q}^*(V_n) \rightarrow 0$  follows from the countable additivity of  $\tilde{Q}^*(\cdot)$  on  $S_k$ , because we may assume the existence of some  $k$  for which  $A_n \subseteq S_k$  ( $n = 1, 2, \dots$ ).] The same arguments as used in the proof of Lemma 9.2.IX now show that there exists a countably additive set function  $\tilde{Q}$  defined on  $\mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#})$  such that  $\tilde{Q}(C) = \tilde{Q}^*(C)$   $\tilde{Q}^*$ -a.s.; that is,  $\tilde{Q}^*$  admits a countably additive extension  $\tilde{Q}$  as required.

- 10.2.2 For an infinitely divisible multivariate point process [see Definition 6.4.I(a)], show that the KLM measure  $\tilde{Q}$  is defined on  $\mathcal{X} \times \{1, \dots, m\}$  and satisfies for bounded  $A \in \mathcal{B}_{\mathcal{X}}$

$$\tilde{Q}\{\tilde{N} = (\tilde{N}_1, \dots, \tilde{N}_m): \tilde{N}_1(A) + \dots + \tilde{N}_m(A) > 0\} < \infty.$$

- 10.2.3 Show that the KLM measure of the Gauss–Poisson process of Example 6.3(d) is the sum of two components, namely, measures  $\tilde{Q}_j$  concentrated on realizations with  $\tilde{N}(\mathcal{X}) = j$  for  $j = 1, 2$ .

- 10.2.4 Let  $N$  be an infinitely divisible marked point process on the space  $\mathcal{X}$  with marks in  $\mathcal{K}$ , so  $N_g(A) < \infty$  for each bounded  $A \in \mathcal{B}_{\mathcal{X}}$ .

- Write out the representation of  $N$  as an infinitely divisible point process on  $\mathcal{Y} \equiv \mathcal{X} \times \mathcal{K}$ .
- Observe that the ground process  $N_g$  is infinitely divisible, and write down its representation as an infinitely divisible point process on  $\mathcal{X}$ .
- Investigate the relation between the KLM measures  $\tilde{Q}$  on  $\mathcal{B}(\mathcal{Y}_0)$  for  $N$  and  $\tilde{Q}_g$  on  $\mathcal{B}(\mathcal{X}_0)$  for  $N_g$ . In particular, investigate whether the KLM measure  $\tilde{Q}$  has  $\tilde{Q}\{\tilde{N}: \tilde{N}(A \times \mathcal{K}) > 0\} < \infty$  for such  $A$ .
- Investigate the cluster process representation of a regular infinitely divisible marked point process.

10.2.5 *Proof of Proposition 10.2.IX.* To establish (10.2.11), show the following.

- (a)  $\xi$  is infinitely divisible if and only if its fidi distributions are infinitely divisible.
  - (b) Each fidi distribution has a standard representation similar to that of (10.2.11) subject to the constraint (10.2.12).
  - (c) An inductive limit argument as in the proof of Proposition 10.2.III holds.
- [Hint: See Kallenberg (1975, Theorem 6.1).]

## 10.3. Point Processes Defined by Markov Chains

In many applications, point processes arise as an important—and observable—component of the process of primary interest, but not necessarily as the primary process itself. Very commonly, the principal object of study is a Markov process. Our aim in this section is to examine some of the ways Markov processes can give rise to point processes in time, and some of the issues arising in the discussion of such models.

We show first that a Markov or semi-Markov process on finite or countable state space can be regarded equivalently as an MPP  $\{(t_n, \kappa_n)\}$ , where each  $t_n$  is the time of a state transition, and the associated mark  $\kappa_n$  denotes the state entered when the transition occurs.

In the other direction, every point process with a conditional intensity can be represented as a Markov process, via its history, which can be thought of as a Markov process of jump–diffusion type on a general state space, the points of the state space representing past histories as viewed from the present. The process moves continuously between states during the intervals between events, and jumps to a new state whenever an event occurs. It is Markovian because both the timing of the jump (determined by the conditional intensity) and the nature of the jump (determined by the conditional mark distribution) are functions of the current state.

In such great generality, this observation may not have great value, although we shall return to it in Section 12.5 in developing a framework for the discussion of convergence to equilibrium for point processes. However, many important models arise as special cases, when the relevant history can be condensed into a compact and manageable form. A renewal process provides the simplest example, where the relevant history is just the backward recurrence time, which increases at unit rate between events, resets to zero whenever a jump occurs, and constitutes a simple diffusion–jump–type Markov process. For a Wold process, the backward recurrence time and the length of the last complete interval together constitute a Markov process driving the point process. Some other simple examples are described following the discussion of Markov and semi-Markov processes.

We then note an important distinction between models such as the renewal process, in which the underlying Markov process can be directly constructed from observations on the point process, and those for which the point process



forms only part of a more complex Markovian system, which therefore remains at best partially observable if the only available information comes from the point process observations. This latter class includes hidden Markov models (HMMs) for point processes, such as the so-called Markov-modulated Poisson process (MMPP), and the Markovian arrival process (MAP) and batch Markovian arrival processes (BMAPs) developed by Neuts and co-workers, and now widely used in modelling internet protocol (IP) traffic and elsewhere. Difficult issues of parameter estimation arise for such processes: we outline the use of the expectation–minimization (E–M) algorithm for this purpose.

EXAMPLE 10.3(a) *Point process structure of Markov renewal and semi-Markov processes* [e.g. Asmussen (2003, Section VII.4); Çinlar (1975, Chapter 10); Kulkarni (1995, Chapter 9)]. The probability structure of a Markov renewal or semi-Markov process on finite or countable state space  $\mathbb{X} = \{i, j, \dots\}$  is defined by the following ingredients: an initial distribution  $\{p_i^0: i \in \mathbb{X}\}$ ; a matrix of transition probabilities  $(p_{ij})$  with  $\sum_{j \in \mathbb{X}} p_{ij} = 1$  (all  $i \in \mathbb{X}$ ); and two families of distribution functions  $F_{ij}^0(u)$  and  $F_{ij}(u)$   $[(i, j) \in \mathbb{X} \times \mathbb{X} \text{ and } u \in (0, \infty)]$ , defining the initial and subsequent lengths of time the process remains in state  $i$ , given that the next transition takes it to  $j$ . These ingredients can be used to construct a bivariate sequence  $\{(\kappa_n, \tau_n)\}$  satisfying, for  $u \in \mathbb{R}_+$  and sequences  $x_r \in \mathbb{R}_+$  and  $k_r \in \mathbb{X}$  for  $r = 0, 1, \dots$ ,

$$\begin{aligned} \Pr\{\kappa_0 = k_0\} &= p_{k_0}^0, \\ \Pr\{\tau_1 \leq u, \kappa_1 = k_1 \mid \kappa_0 = k_0\} &= F_{k_0 k_1}^0(u) p_{k_0 k_1}, \end{aligned} \quad (10.3.1a)$$

and for  $n = 2, 3, \dots$ ,

$$\begin{aligned} \Pr\{\tau_n \leq u, \kappa_n = k_n \mid \kappa_0 = k_0, (\kappa_r, \tau_r) = (k_r, x_r) \ (r = 1, \dots, n-1)\} \\ = \Pr\{\tau_n \leq u, \kappa_n = k_n \mid \kappa_{n-1} = k_{n-1}\} = F_{k_{n-1} k_n}(u) p_{k_{n-1} k_n}. \end{aligned} \quad (10.3.1b)$$

Letting  $u \rightarrow \infty$  here we see that  $\{\kappa_n: n = 0, 1, \dots\}$  is a discrete-time  $\mathbb{X}$ -valued Markov chain with one-step transition matrix  $(p_{ij})$ , and that for each  $n \geq 1$ ,  $\tau_n$  conditional on  $(\kappa_n, \kappa_{n+1})$  is independent of  $\{(\kappa_r, \tau_r): r = 0, \dots, n-1\}$ . We assume that the process has neither instantaneous states nor explosions (see Exercise 10.3.1). Sufficient conditions, ensuring in particular that  $t_n \equiv t_0 + \tau_1 + \dots + \tau_n \rightarrow \infty$  a.s., are that the matrix  $(p_{ij})$  should be irreducible and have nontrivial invariant measure  $\{\pi_i\}$  satisfying  $\sum_i \pi_i \mu_i < \infty$ , where

$$\mu_i = \sum_{j \in \mathbb{X}} p_{ij} \int_0^\infty u F_{ij}(du) = \sum_{j \in \mathbb{X}} \int_0^\infty u G_{ij}(du) = \int_0^\infty u G_i(du)$$

is the mean sojourn time in state  $i$  and

$$G_i(\cdot) = \sum_{j \in \mathbb{X}} G_{ij}(\cdot) \equiv \sum_{j \in \mathbb{X}} p_{ij} F_{ij}(\cdot).$$

When  $\text{card}(\mathbb{X}) \geq 2$  and  $p_{ii} = 0$  ( $i \in \mathbb{X}$ ), there is a one-to-one measurable mapping between sequences  $\{(t_n, \kappa_n)\}$  and  $\mathbb{X}$ -valued, right-continuous, piecewise-constant functions  $X(t)$  with at most finitely many change-points on any bounded interval, where  $t'$  is a *change-point* of  $X$  if and only if  $X(t'-) \neq X(t') = X(t'+)$ . Clearly, any such  $X(t) \in \mathbb{X}$  for all  $t$  for some countable set of states  $\mathbb{X} = \{i, j, \dots\}$ , and all change-points on  $[0, \infty)$  can be ordered as  $0 = t_0 < t_1 < \dots < t_n \rightarrow \infty$  ( $n \rightarrow \infty$ ). Thus any such  $X(\cdot)$  determines a marked point sequence  $\{(t_n, \kappa_n): n = 0, 1, \dots\}$  for  $\kappa_n \in \mathbb{X}$  for all  $n$ . Conversely, such a marked point sequence determines  $X(\cdot)$  via the relation

$$X(t) = \sum_{n=0}^{\infty} \kappa_n I_{[t_n, t_{n+1})}(t), \quad (10.3.2)$$

that is,  $X(t) = X(t_n + 0) = \kappa_n$  ( $t_n \leq t < t_{n+1}$ ). Write  $\tau_n = t_n - t_{n-1}$  for all  $n$  for which  $t_n < \infty$  (see Figure 9.2).

We call the sequence of marked points  $\{(t_n, \kappa_n)\}$  a *Markov renewal process*, and  $\{X(t)\}$  a *semi-Markov process*. The discussion around (10.3.2) implies that the two are equivalent under the stated conditions. Observe the following.

- (1) When  $\mathbb{X}$  is a one-point set we have a renewal process, delayed if  $F^0 \neq F$ .
- (2) When all  $\tau_n = 1$ ,  $\{X(n): n = 0, 1, \dots\}$  is a discrete-time Markov chain on  $\mathbb{X}$  with one-step transition probability matrix  $(p_{ij})$ .
- (3) When  $F_{ij}^0(u) = F_{ij}(u) = 1 - e^{-q_i u}$  ( $0 < q_i < \infty$ ) for all  $j$ , and  $q_{ij} = q_i p_{ij}$ ,  $X(\cdot)$  is a conservative continuous-time Markov chain with  $Q$ -matrix  $(q_{ij})$ .

A Markov renewal process  $\{(t_n, \kappa_n)\}$  as defined can just as well be interpreted as an MPP with mark space  $\mathcal{K} = \mathbb{X}$  and ground process defined by  $N_g(A) = \#\{n: t_n \in A\}$  for bounded  $A \in \mathcal{B}_{\mathbb{R}_+}$ . By the equivalence just noted, a semi-Markov process can also be treated as an MPP.

Consider first its conditional intensity. For any finite  $t > 0$ , since  $N_g$  is finite on bounded subsets, either  $N_g[0, t) = 0$  or there exists a largest  $t_n \in [0, t)$ , defining  $t_{\text{prev}}$  say with associated mark<sup>1</sup>  $\kappa_{\text{prev}}$ . Suppose that  $t_{\text{prev}}$  is defined, and the transition distribution functions  $F_{ij}(u)$  have densities  $f_{ij}(u)$ , so that there also exist transition-time density functions

$$g_{ij}(t) dt = \Pr\{\tau_n \in (t, t + dt), \kappa_n = j \mid \kappa_{n-1} = i\} = p_{ij} f_{ij}(t) dt. \quad (10.3.3)$$

Then the conditional intensity function  $\lambda_g(t)$  for the ground process depends only on the current state  $X(t-) (= \kappa_{\text{prev}})$  and when it was entered, and we have

$$\lambda_g(t) = g_{\kappa_{\text{prev}}}(t - t_{\text{prev}}) / [1 - G_{\kappa_{\text{prev}}}(t - t_{\text{prev}})],$$

The conditional intensity function  $\lambda^*$  itself is expressible as

$$\lambda^*(t, \kappa \mid \mathcal{H}_t) = \frac{g_{\kappa_{\text{prev}}, \kappa}(t - t_{\text{prev}})}{1 - G_{\kappa_{\text{prev}}}(t - t_{\text{prev}})} = \lambda_g(t) f(\kappa \mid t, \mathcal{H}_t), \quad (10.3.4a)$$

<sup>1</sup> In terms of the pair  $(t_{\text{prev}}, \kappa_{\text{prev}})$ , a Poisson process depends on neither element, a renewal process depends on  $t_{\text{prev}}$  only, a Markov process depends on  $\kappa_{\text{prev}}$  only, and a semi-Markov process depends on the complete pair. See Exercise 10.3.2.

where the conditional mark distribution is given by

$$f(\kappa \mid t, \mathcal{H}_t) = \frac{g_{\kappa_{\text{prev}}, \kappa}(t - t_{\text{prev}})}{\sum_{j \in \mathbb{X}} g_{\kappa_{\text{prev}}, j}(t - t_{\text{prev}})}.$$

For  $k \in \mathbb{X}$ , each  $N_k(A) = \#\{n: t_n \in A \text{ and } \kappa_n = k\}$  is a point process on  $\mathbb{R}$  for which  $N_k(t) \equiv N_k(0, t]$  counts the number of entries into the state  $k$  during  $(0, t]$ .

In the Markov case, the Markov property implies that the conditional intensity  $\lambda^*(t, k) dt = E[dN_k(t) \mid \mathcal{H}_t]$  depends on the past only through the state last entered, so that the ground intensity is given by

$$\lambda_g^*(t) = q_{X(t-)} = q_{\kappa_{\text{prev}}},$$

whereas for the conditional mark distribution

$$f(k \mid t, \mathcal{H}_t) = p_{\kappa_{\text{prev}}, k} = p_{X(t-), k} \quad (10.3.4b)$$

[cf. equation (7.3.3)]. The trajectories of the Markov process are easy to reconstruct from this MPP, because it carries details of which states were entered and for how long. The likelihood for a realization  $\{(t_n, k_n): n = 1, \dots, N(T)\}$ , starting in state  $k_0$  at  $t = 0$  and extending over an observation period  $(0, T)$ , can be written in the form

$$L_T = e^{-q_{k_0} t_1} q_{k_0 k_1} e^{-q_{k_1} (t_2 - t_1)} q_{k_1 k_2} \dots e^{-q_{k_{N(T)}} (T - t_{N(T)})}. \quad (10.3.5)$$

In the semi-Markov case, the MPP again contains complete information about the evolution of the process, and allows the likelihood to be written down explicitly in terms of the  $g_{ij}(\cdot)$ ; details are set out in Exercise 10.3.5.

As in the simple renewal process [see (4.1.5–10)], an important role in describing the properties of the semi-Markov process is played by the Markov renewal operator  $\mathbf{H}(\cdot)$  with elements  $H_{ij}(t) = E[N_j[0, t] \mid (t_0, \kappa_0) = (0, i)]$ .  $\mathbf{H}$  is given by the sum of the series of convolution powers, in which  $\mathbf{G}(t) = (\mathbf{G}_{ij}(t))$ ,

$$\mathbf{H} = \mathbf{I} + \mathbf{G} + \mathbf{G} * \mathbf{G} + \mathbf{G} * \mathbf{G} * \mathbf{G} + \dots,$$

or equivalently from the Markov renewal equation  $\mathbf{H} = \mathbf{I} + \mathbf{G} * \mathbf{H} = \mathbf{I} + \mathbf{H} * \mathbf{G}$ , where convolution of matrices of nondecreasing functions is defined elementwise by

$$((A * B)(t))_{ij} = \sum_{k \in \mathbb{X}} \int_0^t dA_{ik}(u) B_{kj}(t - u)$$

[if instead of nondecreasing functions we have matrices of nonnegative densities then define  $((a * b)(t))_{ij} = \sum_k \int_0^t a_{ik}(u) b_{kj}(t - u) du$ ]. In particular,

assuming densities as above and that  $F_{ij}^0 = F_{ij}$ , the factorial moment densities for the ground process are given by the matrix products

$$m_{[r]}^g(u_1, \dots, u_r) = \mathbf{p}_0^\top \mathbf{H}'(u_1) \mathbf{H}'(u_2 - u_1) \dots \mathbf{H}'(u_r - u_{r-1}) \mathbf{1} \quad (10.3.6)$$

for  $r = 1, 2, \dots$ , where  $\mathbf{p}_0 = (p_i^0)$  is the vector of initial probabilities,  $\mathbf{H}'(t) = (h_{ij}(t)) = (H'_{ij}(t))$ , and  $\mathbf{1}$  is a vector of ones. Ball and Milne (2005) generalize this result to a wide range of point processes defined by transitions between states or groups of states; see also Exercises 10.3.3–4.

Let  $\{\pi_i\}$  be left-invariant for  $(p_{ij})$ . Then  $U_g(x) = \sum_{i \in \mathbb{X}} \pi_i \sum_{j \in \mathbb{X}} H_{ij}(x)$  is the analogue of the renewal function at (4.1.5). See Exercise 13.4.4.  $\square$

The next three examples illustrate further cases where a point process can be expressed in terms of a relatively simple Markov process.

EXAMPLE 10.3(b) *Hawkes process with exponential decay* [see also Exercise 7.2.5 and Example 7.3(c)]. This is the simplest of several such examples considered in Chapter 7. The governing Markov process is of shot-noise type  $Y(t) = \int_0^t e^{-\alpha(t-u)} N(du)$ , and is a linear function of past observations; the conditional intensity function takes the form

$$\lambda^*(t) = \lambda + \nu \int_0^t \alpha e^{-\alpha(t-u)} N(du) = \lambda + \nu \alpha Y(t)$$

when we add in a background rate  $\lambda$ . Exercise 7.2.5 details the forward Kolmogorov equation, the stationary distribution, and the likelihood function, assuming the process starts from the stationary distribution. Examples 7.3(b)–(c) consider various extensions, in particular to situations where the conditional intensity can be a more general function of  $Y(t)$ .  $\square$

EXAMPLE 10.3(c) *Birth-and-death process*. Consider a simple birth-and-death process, with birth rate  $\lambda$  per individual and death rate  $\mu$  per individual. Let  $N^+(t)$  and  $N^-(t)$  denote the numbers of births and deaths recorded up to time  $t$ . Then the population size at time  $t$  is just the difference

$$N(t) = N(0) + N^+(t) - N^-(t).$$

We can treat this as an MPP by supposing that the instants of births and deaths are recorded separately, together forming the ground process

$$N_g(t) = N^+(t) + N^-(t),$$

with the marks ‘+’ denoting a birth and ‘−’ a death. The process  $N(t)$  forms an ergodic Markov process if  $\lambda < \mu$  and there are also new arrivals, or ‘births from external source’ (i.e., immigrants) that occur according to a Poisson process of constant rate  $\nu > 0$ . This is a classic example of a continuous-time Markov process, and it is well known that, for example, a stationary

distribution exists when  $\mu > \lambda$  and is negative binomial [see, e.g., Bartlett (1955, p. 78)] with generating function

$$G(z) = \sum_{j=0}^{\infty} z^j \Pr\{N(t) = j\} = \left( \frac{\mu - \lambda z}{\mu - \lambda} \right)^{-\nu/\mu}.$$

In point process terms, the conditional mark distribution takes the form

$$\begin{aligned} f^*(+ | t) &= [\nu + \lambda N(t-)] / \lambda_g^*(t), \\ f^*(- | t) &= [\mu N(t-)] / \lambda_g^*(t), \end{aligned} \quad \text{where} \quad \lambda_g^*(t) = \nu + (\lambda + \mu)N(t-)$$

[recall (7.3.3)]. The likelihood can be written down from the standard point process formulae (Proposition 7.3.III) provided either  $N(0)$  is known (e.g., the population starts from size zero), or its initial distribution is known [e.g., in the stationary regime, it should be the stationary distribution for  $N(t)$ ]. Exercise 10.3.5 gives some details.  $\square$

EXAMPLE 10.3(d) *Pure death process for software reliability* [Jelinski and Moranda (1972); Singpurwalla and Wilson (1999)]. Consider a new or partially tested piece of software containing a number of errors ('bugs'). Every time the software fails, one of the bugs is discovered and repaired. Suppose that every undiscovered bug contributes a constant component  $\mu$  to the total risk; denote the number of bugs still undiscovered at time  $t$  by  $X(t)$ . Then the conditional intensity at time  $t$  is given by  $\mu X(t)$ . The process  $X(t)$  is clearly Markovian; in fact it constitutes a pure death process with constant death rate  $\mu$ . In principle, it is directly observable only if  $N = X(0)$  is observable, a somewhat unlikely circumstance in the given context. More commonly therefore,  $N$  is treated as an unknown parameter (see Exercise 10.3.5 and, e.g., Singpurwalla and Wilson). But unless  $\mu$  is known, very little information about  $N$  is retrievable from the data on observed failure times, and standard likelihood estimates of  $N$  are either unstable or unobtainable. Hence a Bayesian approach is often preferred, with a prior distribution for the initial state which may be either given subjectively or constructed from experience of previous studies. The posterior distribution of  $N$  can then be used to obtain an estimate of the number of remaining bugs.  $\square$

The last example shows very clearly, as is also true in earlier examples, that *the underlying Markov process cannot be considered as fully observable unless the initial state is known*. If it is not known, the initial state plays a role similar to that of an unknown parameter, which can be either estimated along with the other parameters from the likelihood function (this rarely leads to a satisfactory estimate because the information in the data with any bearing on the initial value is usually limited), or specified in terms of a prior distribution.

Such examples form simple cases of the more general situation where the observed point process carries only partial information about the underlying

Markov process. More commonly, not only the initial state but also the subsequent evolution of the Markov process remains unobserved. The birth-and-death process illustrates also how simple it is for a point process driven by an observable Markov process to transform into an example of this kind. In this example, the situation would be radically altered if the ‘+/-’ marks were no longer observed: for a history with  $n$  observed points, there would then be  $2^n$  different ways in which the marks might be assigned, and obtaining the overall likelihood would entail averaging over  $2^n$  likelihoods conditional on a given ordering.

At this point we enter the territory of hidden Markov models. The observed history, usually the internal history for the point process, is insufficient to allow the reconstruction of the Markov process driving the point process. Consequently the simple likelihoods characterizing the previous examples, which were dependent on knowing the driving process, are no longer available. Put in other terms, the conditional intensity based on the internal history is complex and difficult to handle directly, in contrast to the conditional intensity given the history of the driving process, which commonly has a simple form.

Starting from the 1960s, these ideas led to a filtering theory for point processes, motivated by the analogy with the Kalman filter, and making use of the martingale constructions described in Chapter 14. More recently, a new focus on the problems of parameter and state estimation for such processes has developed through the use of the *E-M (estimation-maximization) algorithm*, an approach we outline shortly. First, however, we explore directly the simplest example of an HMM with point process observations. It played a key role in the early discussion of point process filtering [see, e.g., Yashin (1970); Galchuk and Rosovskii (1971); Snyder (1972); Jowett and Vere-Jones (1972); Rudemo (1972, 1973); Vere-Jones (1975); Brémaud (1981, Chapter 4)], and is also the starting point for the more elaborate HMMs which have come to be used extensively in communication theory and elsewhere.

**EXAMPLE 10.3(e)** *Cox process directed by a simple Markov chain; ‘telegraph signal’ process; Markov modulated Poisson process.* We sketched this model briefly in Exercise 7.2.8. The more extended discussion here generally follows Rudemo (1972).

Suppose given a Markov process  $\{X(t): t \geq 0\}$  on the finite state space  $\{1, \dots, K\}$  with  $Q$ -matrix  $Q = (q_{ij})$  so that  $\sum_{j=1}^K q_{ij} = 0$  ( $i = 1, \dots, K$ ), and  $q_i \equiv -q_{ii}$ , assumed positive for all  $i$ , governs the exponential holding times in state  $i$ , and  $p_{ij} = q_{ij}/q_i$  represents the probability that when a jump occurs from state  $i$  it is into a state  $j \neq i$ . Then the matrix of transition probabilities  $P(t) \equiv (p_{ij}(t))$  satisfies  $P(0) = I$  and the forward and backward equations

$$\frac{dP}{dt} = QP(t) = P(t)Q,$$

with solution  $P(t) = \exp(tQ)$ . Suppose further that while this Markov process is in state  $j$ , points of a Poisson process are generated at rate  $\lambda_j$ , and that

the observational data consist only of these points. The simplest nontrivial case of this set-up occurs with a process  $X(\cdot)$  on two states with  $\lambda_2 \approx 0$  and  $\lambda_1$  somewhat larger: we then have a model of a ‘telegraph signal’ process.

Several estimation problems arise in connection with this process. Let us consider in particular the problem of ‘tracking’ the unobserved Markov process  $X(\cdot)$ , given the observations on the point process. This requires maintaining and updating the family of probabilities

$$\pi_i(t) \equiv \mathcal{P}(\{X(t) = i\} \mid \mathcal{H}_t). \quad (10.3.7)$$

Suppose the points observed on  $(0, T)$  are  $t_1 < \dots < t_{N(T)}$ , with  $t_{N(T)} < T$ . To obtain the  $\pi_i(t)$  above, we consider the evolution of the ‘joint statistics’ defined, on any subinterval of the form  $(0, t)$  for  $0 < t < T$  and with  $t_{N(t)} < t \leq t_{N(t)+1}$ , by

$$\begin{aligned} & p_i(t; t_1, \dots, t_{N(t)}) dt_1 \dots dt_{N(t)} \\ &= \Pr\{X(t) = i \text{ and points occur in } (t_j, t_j + dt_j), j = 1, \dots, N(t)\}. \end{aligned} \quad (10.3.8)$$

Call ‘either  $X(\cdot)$  changes state or there is a point at  $t$ ’ an event at  $t$ . Then, conditional on  $X(t+) = i$ , the time  $\tau$  elapsing until the next event is exponentially distributed with  $\Pr\{\tau > u\} = e^{-(q_i + \lambda_i)u}$ , and, independent of  $\tau$ , the event is either a point with probability  $\lambda_i/(q_i + \lambda_i)$  or a transition of  $X(\cdot)$  from  $i$  to  $j$  with probability  $q_{ij}/(q_i + \lambda_i)$ . Between observed points, therefore, the joint statistics evolve in a similar manner to the basic transition probabilities but with the matrix  $Q - \Lambda$  in place of  $Q$ , where  $\Lambda \equiv \text{diag}(\lambda_1, \dots, \lambda_K)$ . When a jump occurs, the joint statistics are weighted by factors  $\lambda_i$ , corresponding to multiplying the vector of joint statistics by the matrix  $\Lambda$ . It then follows that the vector  $\mathbf{p}(\cdot) \equiv (p_1(\cdot), \dots, p_K(\cdot))^\top$  of joint statistics can be expressed as the matrix product

$$\mathbf{p}(t)^\top \equiv \mathbf{p}(t; t_1, \dots, t_{N(t)})^\top = \mathbf{p}(0)^\top R((0, t]; t_1, \dots, t_{N(t)}), \quad (10.3.9a)$$

where  $R((0, t]) = J(t)$  if  $N(t) = 0$  and otherwise

$$R((0, t]) = J(t_1) \Lambda J(t_2 - t_1) \Lambda \dots J(t_{N(t)} - t_{N(t)-1}) \Lambda J(t - t_{N(t)}), \quad (10.3.9b)$$

$\mathbf{p}(0) = (p_1(0), \dots, p_K(0))^\top$  is the vector of initial probabilities for the process  $X(\cdot)$  and  $J(x) = \exp((Q - \Lambda)x)$ . The probability  $\pi_i(t)$  that at any time  $t \in (0, T]$  the process is in state  $i$ , given the observations  $t_1, \dots, t_{N(t)}$  up to time  $t$  and the initial distribution, is the ratio

$$\pi_i(t) = \frac{p_i(t)}{\mathbf{p}(t)^\top \mathbf{1}}. \quad (10.3.7')$$

Although (10.3.9) gives an explicit representation of the joint statistics, it may be just as convenient, particularly if an updating procedure is envisaged,

to represent their evolution in terms of the differential equations they satisfy between events, and the discrete jumps that occur at the events  $t_n$  on the trajectory. In terms of the  $p_i(t)$  these equations are linear in form as below, with  $D_t \equiv \partial/\partial t$ :

$$D_t p_i(t) = -(\lambda_i + q_i)p_i(t) + \sum_{j \neq i} p_j(t)q_{ji} \quad (t \neq \text{any } t_n), \quad (10.3.10a)$$

$$\Delta p_i(t_n) = p_i(t_n+) - p_i(t_n-) = (\lambda_i - 1)p_i(t_n-). \quad (10.3.10b)$$

Similar equations can be written down for the conditional probabilities  $\pi_i(t)$  but in view of the ratios involved these are nonlinear, having the form

$$D_t \pi_i(t) = -\pi_i(t)[\lambda_i + q_i - \lambda^{\mathcal{H}}(t)] + \sum_{j \neq i} \pi_j(t)q_{ji} \quad (t \neq \text{any } t_n), \quad (10.3.11a)$$

$$\Delta \pi_i(t_n) = \left( \frac{\lambda_i}{\lambda^{\mathcal{H}}(t)} - 1 \right) \pi_i(t_n-), \quad (10.3.11b)$$

where

$$\lambda^{\mathcal{H}}(t) = \sum_{i=1}^K \lambda_i p_i(t) \quad (10.3.12)$$

is the conditional intensity at time  $t$ , given only the internal history  $\mathcal{H}$ .  $\square$

If we consider the same example from the point of view of parameter estimation, (10.3.9) immediately yields the likelihood for observations over the observation period  $(0, T)$  in the form

$$L(t_1, \dots, t_{N(T)}) = \mathbf{p}(T)^\top \mathbf{1} = \mathbf{p}(0)^\top R((0, T]) \mathbf{1}, \quad (10.3.13)$$

where  $\mathbf{1}$  is the column vector of 1s. Because the likelihood is represented here as an explicit function of the  $Q$ -matrix of the hidden Markov process and the rates  $\lambda_i$ , it could in principle be used directly in maximization routines to find likelihood estimates of these quantities. However, this is a cumbersome and unstable process at best, and various techniques have been suggested to help stabilize the estimation procedures. One option, treated in detail in Vere-Jones (1975), is to extend the difference/differential equations from the joint probabilities (and hence the likelihood) to the means and variances of the parameter estimates; the results are still cumbersome and awkward to implement in practice. A more effective approach, indicated already in Exercise 7.2.8, is to bring the E–M algorithm to bear on the problem. This approach also reveals structural features that are important in discussing more general classes of models. The virtues of the E–M algorithm in this context are that it allows us to return to the simpler likelihood structures that occur when the underlying Markov process is known, thus making maximization (the M-step) easy, and replaces the rather unstable direct maximization of the likelihood by a more stable iterative procedure. Nevertheless, numerical issues



remain a major concern, and for the point process models many hundreds or even thousands of observations may be needed to gain stable parameter estimates, and in unfavourable cases (e.g., when some state transitions appear only rarely), much more may be required.

To introduce these ideas, we first outline the basic steps in applying the E–M algorithm to an HMM in discrete time and space. The immediate point process application, treated in Example 10.3(f), assumes Poisson observations such as would result from binning the observed points in Example 10.3(d).

Standard references for the E–M algorithm are Dempster *et al.* (1977), Elliott *et al.* (1995), MacDonald and Zucchini (1997), and particularly relevant examples are discussed, for example, in Qian and Titterton (1990), Deng and Mark (1993), Turner *et al.* (1998). The pioneering work goes back to papers by Baum and Petrie (1966) and Baum and Eagon (1967).

To describe an HMM in discrete time and space we suppose given the matrix  $P = (p_{ij})$ ,  $i, j = 1, \dots, K$ , of one-step transition probabilities of the underlying Markov chain which for simplicity we assume is aperiodic and irreducible. We also suppose given a family of probability distributions: when the chain is in state  $j$  it generates an observation with density  $f_j(z)$  ( $j = 1, \dots, K$  and  $z$  real).

The procedures start by introducing the *forward* and *backward probabilities* defined, respectively, by

$$\alpha_n(j) = \Pr\{X_n = j; Z_1 = z_1, \dots, Z_n = z_n\} \quad (10.3.14a)$$

$$\beta_n(j) = \Pr\{Z_{n+1} = z_{n+1}, \dots, Z_N = z_N \mid X_n = j\} \quad (10.3.14b)$$

[strictly,  $\alpha_n(j) = \alpha(j; z_1, \dots, z_n)$  and  $\beta_n(j) = \beta(j; z_{n+1}, \dots, z_N)$ ]. For  $n = 1, \dots, N - 1$ , these probabilities satisfy the recurrence relations

$$\alpha_{n+1}(j) = \sum_{i=1}^K \alpha_n(i) p_{ij} f_j(z_{n+1}), \quad (10.3.15a)$$

$$\beta_n(j) = \sum_{k=1}^K p_{jk} f_k(z_{n+1}) \beta_{n+1}(k), \quad (10.3.15b)$$

where the  $p_{ij}$  are the one-step transition probabilities of the discrete-time chain (so  $(p_{ij}) = P_{\Delta t}$  in our case), the  $f_j(\cdot)$  are probability densities (Poisson probabilities in our case) for the observations when the chain is in state  $j$ , and  $\alpha_1(\cdot)$  is the vector of initial probabilities and  $\beta_N(\cdot) = \mathbf{1}$ . Matrix versions of these equations are discussed in Exercise 10.3.7. It is assumed that the  $f_j$  are unchanging over the observation period, and that successive observations are conditionally independent, given the corresponding states.

For every  $1 \leq n \leq N$ , the likelihood  $L_N$  of the  $N$  observations, obtained by averaging over the possible state sequences, can be written in the form  $L_N = \sum_{i=1}^K \alpha_n(i) \beta_n(i)$ ; in particular,

$$L_N = \sum_{i=1}^K \alpha_N(i). \quad (10.3.16)$$

In practice it is generally desirable to renormalize the forward and backward probabilities at each step to avoid computations with extremely small numbers (see Exercise 10.3.7). Although the recurrence equations allow the forward and backward probabilities to be computed quickly, their underlying importance is embodied in the following lemma.

**Lemma 10.3.I** (State Estimation Lemma). *For an HMM with  $P$  and  $\{f_j(\cdot)\}$  as above, the conditional probabilities  $d_n(i) = \Pr\{X_n = i \mid Z_1, \dots, Z_N\}$  and  $e_n(i, j) = \Pr\{X_n = i, X_{n+1} = j \mid Z_1, \dots, Z_N\}$  are given, respectively, by*

$$d_n(i) = \frac{\alpha_n(i)\beta_n(i)}{\sum_k \alpha_n(k)\beta_n(k)}, \quad n = 1, \dots, N, \quad (10.3.17a)$$

$$e_n(i, j) = \frac{\alpha_n(i) p_{ij} f_j(Z_{n+1}) \beta_{n+1}(j)}{\sum_{k, \ell} \alpha_n(k) p_{k\ell} f_\ell(Z_{n+1}) \beta_{n+1}(\ell)}, \quad n = 1, \dots, N-1. \quad (10.3.17b)$$

PROOF. The proof of (10.3.17a), as of the likelihood representation (10.3.16), depends on the fact that we can use the Markovian property to break open the joint probability  $\Pr\{X_n = i; Z_1 = z_1, \dots, Z_N = z_N\}$  and write it as the product of the two terms  $\alpha_n(i)$ ,  $\beta_n(i)$  for any  $n$  in  $0 \leq n \leq N$ . Start with

$$\begin{aligned} \Pr\{X_n = i; Z_1 = z_1, \dots, Z_N = z_N\} \\ = \Pr\{Z_1 = z_1, \dots, Z_N = z_N \mid X_n = i\} \Pr\{X_n = i\}. \end{aligned}$$

Now, given the state  $i$  at time  $n$ , the distribution of the observations beyond  $n$  is independent of those before  $n$  and is nothing other than the backward probability  $\beta_n(i)$ . That is because the Markovian character of the transition probabilities means that no relevant information about the future states can be transmitted past the present state  $i$ , once that is given. Likewise the distribution of the observations up to time  $n$  is independent of those which follow, given the state at time  $n$ . The right-hand side of the above equation therefore reduces to

$$[\alpha_n(i)/\Pr\{X_n = i\}] \times \beta_n(i) \times \Pr\{X_n = i\} = \alpha_n(i)\beta_n(i).$$

Rewriting the joint probability in terms of the conditional distribution of the state, given the observations, and normalizing, yields (10.3.17a). Similar reasoning gives (10.3.17b). See also Exercise 10.3.7 for a matrix proof.  $\square$

The E–M algorithm for this setting involves finding those parameter values for the full model [i.e., both  $p_{ij}$  and  $f_j(\cdot)$ ] that maximize the expected value of the likelihood, given the observations and an initial set of parameter values. Finding the expected value of the full likelihood, given the observations, is the E-step; finding the parameter values which maximize this expected likelihood is the M-step. An argument based on Jensen’s inequality shows that each iteration of the algorithm can only increase the likelihood.

To illustrate the algorithm we turn to a discrete-time version of Example 10.3(e) [see also Fischer and Meier-Hellstern (1993), Davison and Ramesh (1993), and Rydén (1994, 1996)].

EXAMPLE 10.3(f) *Discrete-time HMMs with Poisson observations: E–M algorithm analysis.* Because the standard procedures apply to discrete-time processes, a preliminary step in adapting them to the context of Example 10.3(e) is to bin the observations into small time intervals of length  $\Delta t$  (preferably smaller than the mean interval length). The model then reduces to the discrete-time Markov chain with transition matrix

$$P_{\Delta t} = e^{Q\Delta t},$$

whereas the observations consist of the sequence of counts  $Z_1, \dots, Z_N$  in successive bins  $n = 1, \dots, N$ . The counts can be modelled conveniently as Poisson random variables having parameter  $\mu_i = \lambda_i \Delta t$  when the Markov chain is in state  $i$ . It is assumed that changes of state occur only on the boundaries of the  $\Delta t$  intervals.

We take the vector of parameters for this problem to be

$$\theta = \{\pi_1, \dots, \pi_K; p_{11}, \dots, p_{ij}, \dots, p_{KK}; \mu_1, \dots, \mu_K\},$$

where  $\{\mu_i\}$  are the parameters of the distributions  $f_i(\cdot)$ . The complete likelihood  $L_c$ , given successive states  $\{i_1, \dots, i_N\}$  and observations  $\{Z_1, \dots, Z_N\}$ , is given by

$$\log L_c(\theta) = \log \pi_{i_1} + \sum_{n=1}^{N-1} \log p_{i_n, i_{n+1}} + \sum_{n=1}^N \log f_{i_n}(Z_n).$$

In our case, inasmuch as the distributions are Poisson,

$$\log f_{i_n}(Z_n) = Z_n \log \mu_{i_n} - \mu_{i_n} - \log(Z_n!).$$

It is convenient here to rewrite the likelihood by collecting the quantities multiplying a particular term such as  $\log p_{ij}$ ; then  $\log L_c(\theta)$  equals

$$\sum_i \delta_{i_0 i} \log \pi_i + \sum_{i,j} E_{ij} \log p_{ij} + \sum_i (G_i \log \mu_i - D_i \mu_i) + C, \quad (10.3.18)$$

where  $D_i$  counts the total visits to state  $i$ ,  $E_{ij}$  the number of transfers from state  $i$  to state  $j$ , and  $G_i$  the number of points emitted while the chain is in state  $i$ , and  $C$  is a function of the observations  $Z_n$  only. Equation (10.3.18) shows that  $\{\delta_{i_0 i}, D_i, E_{ij}, G_i\}$  is a set of sufficient statistics for the parameters  $\theta$ .

Next, take expectations conditional on the observations  $\mathbf{Z} = \{Z_1, \dots, Z_N\}$  and some initial parameter set  $\theta^*$ , treating the running parameter values in (10.3.18) as fixed numbers. This requires taking an average over state histories of the log-likelihood (10.3.16) (specifically, the various sufficient statistics), conditional on the observations and initial parameters. But these conditional expectations can all be written down in terms of the conditional probabilities of visits to or transfers between successive states, given the observations, and

hence in terms of the quantities appearing in Lemma 10.3.I. Specifically, we find

$$\begin{aligned} \mathbb{E}[\delta_{i_0,i} \mid \mathbf{Z}] &= \pi_i^*, & \mathbb{E}[E_{ij} \mid \mathbf{Z}] &= \sum_{n=1}^{N-1} e_n^*(i, j) \equiv e_{ij}^*, \\ \mathbb{E}[D_i \mid \mathbf{Z}] &= \sum_{n=1}^N d_n^*(i) \equiv d_i^*, & \mathbb{E}[G_i \mid \mathbf{Z}] &= \sum_{n=1}^N Z_n d_n^*(i) \equiv g_i^*, \end{aligned}$$

where \* indicates dependence on the initial parameters. Substitution leads to an expression for the conditioned log-likelihood which exactly replicates the form of (10.3.18), namely,  $\mathbb{E}[\log L_c(\theta) \mid \mathbf{Z}]$  equals

$$\sum_i \pi_i^* \log \pi_i + \sum_{i,j} e_{ij}^* \log p_{ij} + \sum_i (g_i^* \log \mu_i - d_i^* \mu_i) + C. \quad (10.3.19)$$

Equation (10.3.19) constitutes the E-step.

To implement the maximization step we have to find the values of the current parameters that maximize (10.3.19) for given values of the observations and of the starred expressions. Recalling the constraints  $\sum_i \pi_i = 1 = \sum_j p_{ij}$ , we find for the new estimates

$$\hat{\pi}_i = d_1^*(i), \quad \hat{p}_{ij} = \frac{e_{ij}^*}{d_i^*}, \quad \hat{\mu}_i = \frac{g_i^*}{d_i^*}. \quad (10.3.20)$$

These equations, which are commonly referred to as the *Baum–Welch re-estimation equations*, constitute the M-step. Together with the results of Lemma 10.3.I, they summarize the application of the E–M algorithm in the given discrete context. The algorithm proceeds by successive application of the E- and M-steps, using the forward and backward probabilities to evaluate the quantities appearing on the right-hand sides in (10.3.20), and the equations themselves to evaluate the revised parameter estimates.

In practice, a special problem again revolves around estimating the initial distribution, as there is not usually enough information in the data to provide stable estimates for the probabilities  $\pi_i$ . If the  $\pi_i$  are taken to be the stationary probabilities for the chain being estimated, and hence functions of the  $p_{ij}$ , the simplicity of the updating equations is lost, and they become essentially intractable. A reasonable compromise [cf. the discussion in Turner *et al.* (1998)] is to evaluate the stationary distribution for the chain with the initial parameters  $\theta_1$ , and regard this distribution as fixed in the M-step. Then only the transition probabilities and the distribution parameters need to be re-estimated, resulting in modified forms in (10.3.20) with the given initial distribution replacing the estimated initial distribution.

The E–M algorithm may converge rather slowly as it approaches the maximum, in which case it may be more efficient to switch over for the final steps to direct maximization of the likelihood (10.3.16), using the E–M estimates as starting values.  $\square$

The parameters for the continuous-time model which motivated Example 10.3(f) can be related approximately to the parameters of the discrete-time version by the relations

$$\lambda_i \approx \mu_i / \Delta t; \quad q_{ii} \approx -(1 - p_{ii}) / \Delta t; \quad q_{ij} \approx -p_{ij} / \Delta t \quad (i \neq j).$$

It is also possible to apply the E–M algorithm directly to the continuous-time process [see Rydén (1996)], and we now outline this alternative approach.

EXAMPLE 10.3(g) *Cox process directed by a finite Markov chain: Direct E–M analysis* [continued from Example 10.3(e)]. We start from the definition (10.3.9a) for the joint statistic  $p_i(t)$ , which we can identify as the continuous-time analogue of the forward probability  $\alpha_n(i)$ . To emphasize the analogy, we write for the remainder of this discussion, using  $N(t)$  as before,

$$\begin{aligned}\alpha_t(i) &= \Pr\{X(t) = i \text{ and points occur in } (0, t] \text{ at times } t_1, \dots, t_{N(t)}\} \\ &= \mathbf{p}(0)^\top R((0, t]) \boldsymbol{\delta}_i = p_i(t),\end{aligned}$$

where  $\boldsymbol{\delta}_i$  is a  $K$ -vector of 0s except for 1 in the  $i$ th component. Similarly, we can write down an analogue to the backward probability  $\beta_n(i)$ , namely,

$$\begin{aligned}\beta_t(i) &= \Pr\{\text{points occur in } (t, T] \text{ at times } t_{N(t)+1}, \dots, t_{N(T)} \mid X(t) = i\} \\ &= \boldsymbol{\delta}_i^\top R((t, T]) \mathbf{1}.\end{aligned}$$

The differential versions (10.3.10–11) may be regarded as continuous-time analogues of the forward recurrence equations (10.3.15a) in the Baum–Welch formulation of HMMs. The quantities  $\alpha_t(i)$ ,  $\beta_t(i)$  can be computed from the differential equations, or by direct evaluation of (10.3.9a) and its analogue for the backward probabilities. Once again the likelihood can be evaluated as

$$L_c = \sum_i \alpha_t(i) \beta_t(i) = \sum_i \alpha_T(i) \quad (\text{every } t \text{ in } 0 < t < T).$$

The Markov property again allows us to write down state estimation probabilities, analogous to those in Lemma 10.3.I. These take the forms

$$d_t(i) = \Pr\{X(t) = i \mid \text{points at } t_1, \dots, t_{N(T)}\} = \alpha_t(i) \beta_t(i) / L_c, \quad (10.3.21a)$$

$$\begin{aligned}e_t(i, j) dt &= \Pr\{\text{transition } i \text{ to } j \text{ occurs in } (t, t + dt) \mid \text{points at } t_1, \dots, t_{N(T)}\} \\ &= \alpha_t(i) q_{ij} \beta_t(j) / L_c.\end{aligned} \quad (10.3.21b)$$

The complete likelihood can be written down directly in terms of the epochs  $\{s_\ell: \ell = 1, \dots, M\}$  of state transitions,  $s_0 = 0$ , the state transitions themselves  $\{i_{\ell-1} \mapsto i_\ell\}$ , and the event times  $\{t_n\}$ . Recalling the notation  $q_i = -q_{ii}$ ,  $\pi_{ij} = q_{ij}/q_i$ , and taking logarithms, we can write the log-likelihood as

$$\begin{aligned}\log L_c &= \log \pi_{i_0} + \sum_{\ell=0}^{M-1} \left( \log q_{i_\ell, i_{\ell+1}} + [N(s_{\ell+1}) - N(s_\ell)] \log \lambda_{i_\ell} \right. \\ &\quad \left. - (q_{i_\ell} + \lambda_{i_\ell})(s_{\ell+1} - s_\ell) \right) - (q_{i_M} + \lambda_{i_M})(T - t_M) \\ &= \sum_i \left[ \delta_{i_0, i} \log \pi_i - D_i(q_i + \lambda_i) + \sum_{j \neq i} E_{ij} \log q_{ij} - G_i \log \lambda_i \right],\end{aligned}$$

where  $D_i$  accumulates the time spent in state  $i$  during  $(0, T]$ ,  $E_{ij}$  counts the number of transitions from  $i$  to  $j$ , and  $G_i$  counts the number of events which occur while the chain is state  $i$ . Taking expectations after conditioning on the observations yields, for example,

$$E[D_i \mid t_1, \dots, t_{N(T)}] \equiv D_i^* = \int_0^T dt(i) dt,$$

with analogous integral expressions for  $E_{ij}^*$  and  $G_i^*$  but with  $dt$  replaced by  $dN(t)$  (i.e., to give sums rather than integrals). Hence we obtain for the E-step, after substituting for the starred quantities,

$$E(\log L_c \mid \mathcal{H}_{(0,T)}) = \sum_i \left[ d_0(i) \log \pi_i - (q_i + \lambda_i) \int_0^T dt(i) dt \right. \\ \left. + \sum_{j \neq i} \log q_{ij} \int_0^T e_t(i, j) dt - \log \lambda_i \int_0^T dt(i) dN(t) \right],$$

where  $N(\cdot)$  denotes the observed counting process. Then maximizing over the parameters  $\pi_i, q_i, q_{ij}, \lambda_i$ , we obtain for the M-step estimates

$$\begin{aligned} \hat{\pi}_i &= d_0(i), & \hat{q}_{ij} &= \frac{\int_0^T e_t(i, j) dt}{\int_0^T dt(i) dt}, \\ \hat{\lambda}_i &= \frac{\int_0^T dt(i) dN(t)}{\int_0^T dt(i) dt}, & \hat{q}_i &= \frac{\sum_{j \neq i} \int_0^T e_t(i, j) dt}{\int_0^T dt(i) dt}. \end{aligned} \tag{10.3.22}$$

The integrals in these equations can be evaluated with the aid of (10.3.9) by summation of integrals over intervals between the points  $0, t_1, \dots, t_{N(T)}, T$ . From the numerical point of view, a serious problem is the calculation of the matrix exponentials  $J(t)$  in (10.3.9b). If a simple discretization is used for this purpose, one is effectively reverting to the discrete-time model previously discussed, and there are greater advantages in keeping to the discrete set-up all through. One alternative is to use matrix diagonalization methods; a further alternative, suggested by Rydén (1996), is the ‘uniformization’ algorithm outlined in Exercise 10.3.9. A recent discussion of numerical aspects can be found in Roberts, Ephraim, and Dieguez (2006).  $\square$

This model has been the subject of considerable extension and elaboration. One generalization, which combines Example 10.3(g) with elements of the discrete HMM considered above, is to allow the observed process to be an MPP rather than a simple point process; a brief outline is given in Exercise 10.3.9. Example 10.3(f) can also be considered as a simple special case of the more general situation where the counting process  $N(t)$  is not itself Markovian, but forms one component in a more complex process which is

Markovian. Rudemo (1973) appears to be the first to have studied point processes of this kind. He considered a bivariate system  $\{(X(t), N(t))\}$ , where  $N(t)$  is the observed counting process,  $X(t)$  is a  $\mathcal{K}$ -valued unobserved Markov process with  $Q$ -matrix  $Q$ , and counts may be produced either between state transitions, as in the previous example, or at the transitions themselves. The system remains Markovian, and similar issues arise: tracking the current state of the unobserved process  $X(t)$  from observations on the counting process, and estimating the parameters of the bivariate process.

This framework also covers the extensive series of studies by Neuts and co-workers on Markovian and Batch Markovian Arrival Processes (so-called MAPs and BMAPs); see *inter alia* Neuts (1978, 1979, 1989), Ramaswami (1980), Asmussen *et al.* (1996), and Klemm *et al.* (2003). An important feature of these models is that, if they are used as input streams for single-server and other queueing systems, many characteristics such as distributions of queue lengths and waiting times, can be represented as matrix-analytic analogues of the forms which they take in the simpler systems with Poisson or renewal input. Here, we concentrate on the point process properties, referring the reader to accounts of the wider range of applications in the cited references.

The  $Q$ -matrix for a two-component process  $\{(X(t), N(t))\}$  has a block structure of the form

$$\begin{pmatrix} Q_{00} & Q_{01} & Q_{02} & \cdots \\ Q_{10} & Q_{11} & Q_{12} & \cdots \\ Q_{20} & Q_{21} & Q_{22} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix},$$

where each  $Q_{rs}$  is a  $K \times K$  matrix that describes the transition rates of the unobserved process  $X(t)$  which may occur while the counting variable moves from  $r$  to  $s$ . In most cases backward transitions are not possible, so that  $Q_{rs} = 0$  for  $r > s$ ; also, the process  $X(t)$  has stationary transition probabilities and  $N(t)$  has stationary increments. The  $Q$ -matrix then becomes a block-type version of the  $Q$ -matrix of a pure birth process,

$$\begin{pmatrix} Q_0 & Q_1 & Q_2 & \cdots \\ 0 & Q_0 & Q_1 & \cdots \\ 0 & 0 & Q_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}. \quad (10.3.23)$$

Example 10.3(g) is recovered if we take  $Q_0 = Q - \Lambda$ ,  $Q_1 = \Lambda$ ,  $Q_j = 0$  for  $j > 1$ , where  $Q$  is the  $Q$ -matrix of  $X(t)$ . That is to say,  $N(t)$  increases by 1 each time a point occurs, but the value of  $X(t)$  is unaltered, and the transitions of  $X(t)$  do not directly affect the value of  $N(t)$ .

More generally, processes of the type (10.3.23) are characterized by the matrix  $Q_0$ , which describes the transitions of the process  $X(t)$  in the absence of jumps, and the matrices  $Q_1, \dots, Q_L$  describing the transitions of  $X(t)$  which accompany jumps of size  $1, \dots, L$ , respectively.

One of the simplest point processes of this type, often used as a building block in constructing more complex models, is described below.

EXAMPLE 10.3(h) *Renewal process of phase type; PH-distributions* [Neuts (1978)]. This model is motivated by the situation in which points are recorded only when an underlying Markov process enters a particular state, say 0. It then becomes an example of an alternating renewal process, with one type of interval corresponding to sojourns in state 0, and the other to the periods while the process is traversing a path through the remaining states. As such it can be treated as a modified version of Example 10.3(a) as in Exercise 10.3.4.

However, we wish to consider the limiting situation in which sojourns in state 0 are instantaneous, so that the observed process is actually a renewal process. This limiting process is then described by two components: a defective  $K \times K$   $Q$ -matrix with row sums  $\sum_{j=1}^K q_{kj} = -\delta_k < 0$ , and a vector of re-entry probabilities, say  $\pi_k$ , describing the probability of entering state  $k$  at the same time as a renewal occurs. This corresponds to a model governed by a matrix of type (10.3.23) in which  $Q_0 = Q$ ,  $Q_1 = \delta^\top \pi$ ,  $Q_\ell = 0$  for  $\ell > 1$ .

The interval between successive renewals can be represented as the sum of a random number of exponentially distributed waiting times, corresponding to the holding times in each of the states passed through before the next renewal, and mixed over the starting state. Its distribution takes the form

$$1 - F(t) = \pi^\top e^{Qt} \mathbf{1}, \quad (10.3.24)$$

which in Neuts' terminology represents a *PH-distribution with representation*  $(Q, \pi)$ . Special cases, such as mixtures or sums of exponentials, correspond to giving special forms to  $Q$  and  $\pi$ ; see Exercise 10.3.10 for examples. Expressions for the Laplace transform of the above density, as well as for the associated renewal function, are outlined in Exercise 10.3.10.  $\square$

Consider next the BMAP, originally called a versatile Markovian point process by Neuts (1979). This can be considered both as a Markov process with  $Q$ -matrix of the form (10.3.23), and an MPP. Both points of view are helpful in developing properties of the process, as we indicate below. The process accommodates not only batch arrivals, but also some dependence between the lengths of the intervals between arrivals, because the lengths of two consecutive intervals both depend on the state of  $X(t)$  at the arrival time by which they are separated. Variations on this model are extensively used in information-processing networks, where the batch size approximates packet length, and the arrival rate the packet frequency [see Klemm *et al.* (2003)].

EXAMPLE 10.3(i) *BMAP representations: E-M analysis* [Neuts (1979); Rydén (1996); Klemm *et al.* (2003)]. We suppose again that the process is driven by an unobserved Markov process  $X(t)$  with  $K$  states and generator  $Q$ . In this case, however, the observed points correspond to batches of size  $0 < \ell \leq L$ , and may or may not be associated with a change of state. To accommodate the latter possibilities, the matrix  $Q$  is split among the different batch sizes, and written as a sum  $Q = \sum_{\ell=0}^L Q_\ell$ , where the  $Q_\ell$  appear in the representation (10.3.23), so the  $Q_\ell$  govern the transitions associated with arrivals of batch



size  $\ell$  (or no arrivals for  $Q_0$ ). Write  $Q_\ell = (q_{ij}^{(\ell)})$ , so positive elements  $q_{kk}^{(\ell)}$  with  $\ell > 0$  correspond to arrivals for which no change of state occurs.

Although the BMAP is the archetypal process with a representation as at (10.3.23), for estimation purposes it is more fruitful to consider it as an MPP with bivariate marks,  $(k, \ell)$  say, the two components recording the state  $k$  entered and the size  $\ell$  of the associated arrival batch, including the possibility  $\ell = 0$ . Marks of the form  $(k, 0)$  do not occur when  $X(t)$  is in state  $k$ , because they would correspond to transitions from state  $k$  to state  $k$  with no accompanying arrivals, and hence to no effective transition of any kind.

Inasmuch as the structure of (10.3.23) already implies that the transition rates depend only on the state  $k$  and not on the accumulated number of arrivals, the conditional intensity for the associated MPP  $\{t_n, (k_n, \ell_n)\}$  depends on the history only through the current state  $X(t)$ , which we denote by  $\xi$  to ease the notation. In terms of the representation (10.3.23), we have then for the conditional intensity, for  $k = 1, \dots, K$  and  $\ell = 0, \dots, L$ ,

$$\lambda_{k,\ell}^*(t) = \begin{cases} 0 & \text{if } \ell = 0 \text{ and } \xi = k, \\ q_{\xi k}^{(\ell)} & \text{otherwise.} \end{cases}$$

The ground intensity for the overall process (including transitions not associated with arrivals) is given by

$$\lambda_g^*(t) = E[dN_g(t) \mid \xi]/dt = -(Q_0)_{\xi\xi} = q_\xi, \quad (10.3.25a)$$

and the conditional mark distribution takes the form

$$f^*(k, \ell \mid t) = \pi_{k\ell}(\xi) \equiv q_{\xi k}^{(\ell)} / q_\xi, \quad \ell = 0, 1, \dots, L. \quad (10.3.25b)$$

The complete process is stationary if and only if the Markov process  $X(t)$  is stationary, because  $X(t)$  determines the occurrence probabilities for all types of transitions. Because the marginal process  $X(t)$  is Markovian, and governed by the matrix  $Q$ ,  $X(t)$  is stationary if and only if it starts with initial distribution  $\boldsymbol{\pi}$ , satisfying  $\boldsymbol{\pi}^\top Q = \mathbf{0}^\top$  and which we assume to be well-determined and unique. Taking expectations under this distribution, we find for the mean rate of occurrence of all transitions,

$$\bar{\lambda} = E[\lambda_g^*(t)] = \sum_{k=1}^K \pi_k q_{kk}^{(0)} = \sum_{k=1}^K \pi_k q_k \quad (\text{all } t).$$

The mean rate of arrival of batches of size  $\ell > 0$  is  $\lambda_\ell = \sum_{k=1}^K \pi_k \sum_{j=1}^K q_{kj}^{(\ell)}$ , and the overall rate of arrival of batches is

$$\lambda = \sum_{k=1}^K \pi_k \sum_{\ell=1}^L \sum_{j=1}^K q_{kj}^{(\ell)} = \bar{\lambda} - \lambda_0,$$

where  $\lambda_0 = \sum_k \pi_k \sum_{j:j \neq k} q_{kj}^{(0)}$  is the expected rate of transitions not accompanied by arrivals. Other characteristics of the process, including interval distributions and correlations, can be represented in matrix exponential terms along the lines of Example 10.3(h) (see Exercise 10.3.11).

The BMAP models share with the E-M algorithm the feature that elaborations of the forward-backward equations can be used for parameter estimation. Indeed, the steps follow a very similar pattern to those of Example 10.3(f), and we indicate them only in summary form; Klemm *et al.* (2003) can be consulted for further details and numerical aspects.

Suppose that over the observation interval  $(0, T)$ , the process has initial probability distribution  $\{\pi_k^0\}$ , starts from state  $k_0$ , and jumps at times  $\{t_n\}$  with marks  $\{k_n, \ell_n\}$ , corresponding to transitions into states  $k_n$  associated with arrival batches of size  $\ell_n$  (including the possibility  $\ell_n = 0$ ). Then the complete likelihood takes the form

$$L_c = \pi_{k_0}^0 q_{k_0, k_1}^{(\ell_1)} e^{-q_{k_0} t_1} q_{k_1, k_2}^{(\ell_2)} e^{-q_{k_1} (t_2 - t_1)} \dots \quad (10.3.26)$$

Grouping together the terms associated with particular transitions, or sojourns in particular states, this can be rewritten as

$$L_c = \pi_{k_0}^0 \left[ \prod_{j,k,\ell} (q_{jk}^{(\ell)})^{N_{j,k}^{(\ell)}} \right] e^{-\sum q_k D_k}, \quad (10.3.27)$$

where, with reference to the observation period  $(0, T)$ ,  $N_{j,k}^{(\ell)}$  counts the total number of transitions from state  $j$  into state  $k$  associated with an arrival batch of size  $\ell$ , and  $D_k$  is the total length of time that  $X(t)$  is in state  $k$ .

We now turn to extensions of the backwards and forwards probabilities. In the present model, given the initial distribution  $\pi^0$  and the set of observed arrival times and batch sizes  $\{(\tau_n, \ell_n): n = 1, \dots, N(T)\}$  (and note that the  $\{\tau_n\}$  form in general only a subset of the set  $\{t_n\}$  used in describing the complete likelihood), the forward probabilities take the form

$$\alpha_t(k) = (\pi^0)^\top e^{Q_0(\tau_1)} Q_{\ell_1} e^{Q_0(\tau_2 - \tau_1)} Q_{\ell_2} \dots e^{Q_0(t - \tau_{N(t)})} \mathbf{e}_k. \quad (10.3.28)$$

Likewise the backward probabilities  $\beta_t(i)$  can be written

$$\beta_t(k) = \mathbf{e}_k^\top e^{Q_0(\tau_{N(t)+1} - t)} Q_{\ell_{N(t)+1}} e^{Q_0(\tau_{N(t)+2} - \tau_{N(t)+1})} Q_{\ell_{N(t)+2}} \dots e^{Q_0(T - \tau_{N(T)})} \mathbf{1}.$$

It is evident that here also, for every  $t$  in  $(0, T)$ , the incomplete likelihood can be expressed as

$$\mathcal{L}(T) = \sum_k \alpha_t(k) \beta_t(k) = \sum_k \alpha_T(k) = \alpha_T^\top \mathbf{1}. \quad (10.3.29)$$

This expression can be directly maximized to find suitable parameter estimates. Alternatively, as in Example 10.3(h), we can seek more stable procedures based on the E–M algorithm.

Adapting the latter approach, the forward and backward probabilities reappear in the appropriate extension of the state estimation Lemma 10.3.I. Equation (10.3.21a) retains the form

$$d_t(i) \equiv \Pr\{X(t) = i \mid \mathcal{S}\} = \alpha_t(i)\beta_t(i)/\mathcal{L}(T),$$

whereas if  $N_{ij}^{(0)}(t)$  denotes the number of transitions from  $i$  to  $j$  with no associated arrivals in time  $t$ , (10.3.21b) becomes

$$e_t(i, j) dt \equiv E[dN_{ij}^{(0)}(t) \mid \mathcal{S}] = [\alpha_t(i)q_{ij}^{(0)}\beta_t(j)/\mathcal{L}(T)]dt.$$

For the conditional rate at time  $t$  of  $i \mapsto j$  transitions associated with arrivals of batch size  $\ell > 0$  we have

$$e_t^{(\ell)}(j, k) = \frac{\alpha_t(j)q_{jk}^{(\ell)}\beta_t(k)}{\mathcal{L}_T}.$$

Turning finally to the E- and M-steps, we obtain from (10.3.27),

$$\log L_c = \log \pi_{i_0}^0 - \sum_k D_k \log q_k + \sum_{j,k} \sum_{\ell=0}^L N_{ij}^{(\ell)} \log q_{jk}^{(\ell)}. \quad (10.3.30)$$

Taking expectations conditional on the observed sequence  $\mathcal{S} \equiv \{(t_n, \ell_n)\}$  constitutes the E-step, and leads to an expression similar to (10.3.30) but with  $E[S_k \mid \mathcal{S}]$  and  $E[N_{jk}^{(\ell)} \mid \mathcal{S}]$  replacing the corresponding expressions without the expectations. Maximizing with respect to the parameters is again straightforward, and leads to the updated estimates

$$\hat{\pi}_i^0 = \Pr\{X(0) = i \mid \mathcal{S}\}, \quad \hat{q}_{jk}^{(\ell)} = \frac{E(N_{jk}^{(\ell)} \mid \mathcal{S})}{E(S_j \mid \mathcal{S})}, \quad \hat{q}_j = \sum_k \sum_{\ell=0}^L \hat{q}_{jk}^{(\ell)}. \quad (10.3.31)$$

Thus the crucial difficulties are again in evaluating the conditional expectations which appear in these equations, and again these can be represented in terms of the forward and backward probabilities. We find

$$\Pr\{X(0) = i \mid \mathcal{S}\} = d_0(i), \quad (10.3.32a)$$

$$E(S_k \mid \mathcal{S}) = \int_0^T d_t(i) dt, \quad E(N_{jk}^{(0)} \mid \mathcal{S}) = \int_0^T e_t(j, k) dt.$$

The conditional expectations for the cases with  $\ell > 0$  have a slightly different form because they correspond to known times and sizes of arrivals. However, similar reasoning leads to the results

$$E(N_{jk}^{(\ell)} \mid \mathcal{S}) = \int_0^T e_t^{(\ell)}(j, k) dN^\ell(t), \quad (10.3.32b)$$

where  $N^\ell(t)$  counts the number of batches of arrivals of size  $\ell$ .

Evaluation of the matrix exponentials which arise in these formulae can again be tackled by diagonalization or by the uniformization approximation described in Exercise 10.3.8.  $\square$

## Exercises and Complements to Section 10.3

10.3.1 *Existence.* Any state  $i$  of a continuous time Markov chain on countable state space with time-homogeneous transition probabilities (an MC, say) as in Part II of Chung (1967) is either stable or instantaneous according as the corresponding diagonal matrix element  $q_i = -q_{ii}$  is finite or infinite, respectively. An MC with only stable states and all of them nonabsorbing can nevertheless have infinitely many jumps in a finite interval, but for an MC to be regarded as a point process as defined in Chapter 9 we wish to exclude such a possibility. By using the role of the diagonal matrix elements  $q_i = -q_{ii}$  as intensities, argue that a stochastic condition for an MC to have a.s. finitely many jumps  $\{t_n\}$  in bounded subsets of  $\mathbb{R}$  is that

$$\sum_{t_n \in \mathbb{R}_+} 1/q_{X(t_n-)} = \infty \quad \text{a.s.}$$

Verify that the pure birth process with quadratic birth rates fails this condition. [Hint: See, e.g., Feller (1968, Section XVII.4).]

10.3.2 (a) Use  $(t_{\text{prev}}, \kappa_{\text{prev}})$  of Example 10.3(a) with the transition probabilities and distribution functions specifying a semi-Markov process to describe the conditional intensity at time  $t$  of (i) a Poisson process; (ii) a renewal process; (iii) a Markov process (on countable state space); and (iv) a semi-Markov process. Verify the assertion of the footnote in that Example.

(b) Given a Markov renewal process  $\{(t_n, \kappa_n)\}$  around (10.3.1) with  $\text{card}(\mathbb{X}) \geq 2$  but allowing  $\kappa_n = \kappa_{n+1}$  (so  $p_{ii} > 0$  for at least one  $i \in \mathbb{X}$ ),  $X(\cdot)$  at (10.3.2) is still well-defined. Define a subset  $\{(t'_n, \kappa'_n)\}$ , still a Markov renewal process, such that  $\kappa'_{n+1} \neq \kappa'_n$  (all  $n$ ); the relation (10.3.2) still holds (and the sample functions remain the same), and when the original transition matrix  $(p_{ij})$  is irreducible, the transition matrix  $(p'_{ij})$ , where  $p'_{ij} = p_{ij}/(1-p_{ii})$ , now leads to a one-to-one measurable mapping between the latter Markov renewal process and the semi-Markov process  $X(\cdot)$ .

10.3.3 *Markov renewal or semi-Markov process observed on a subset* [cf. Çinlar (1975, 10(1.13))]. Let  $\mathcal{Y} = \{(t_n, \kappa_n)\}$  be a realization of an irreducible Markov renewal process on state space  $\mathbb{X}$ , and let  $\mathbb{X}'$  be a nonempty proper subset of  $\mathbb{X}$ . Construct the subset  $\mathcal{Y}'$  of  $\mathcal{Y}$  via  $\mathcal{Y}' = \{(t_n, \kappa_n): \kappa_n \in \mathbb{X}'\}$  and relabel it sequentially as  $\{(t'_n, \kappa'_n)\}$ .

- (a) Show that  $\mathcal{Y}'$  is an irreducible Markov renewal process on  $\mathbb{X}'$ .
- (b) Let  $\mathcal{Y}$  have matrix renewal function  $\mathbf{H}(t) = (H_{ij}(t))$ . Show that the components of the matrix renewal function of  $\mathcal{Y}'$  are  $\{H_{ij}(t): i, j \in \mathbb{X}'\}$ .
- (c) When  $\text{card}(\mathbb{X}') \equiv R$  say is finite, find an expression for the  $R \times R$  matrix  $\mathbf{G}_{\mathbb{X}'}(t)$  of transition distributions for the process observed only in  $\mathbb{X}'$  in terms of the original matrix  $\mathbf{G}(t) = (G_{ij}(t))$ . In the special case where the original process is Markovian, show that the transition distributions are of the phase-type discussed in Example 10.3(h).

[Hint:  $\{\kappa_n\}$  is a discrete-time Markov chain on  $\mathbb{X}$ , ensuring that  $\{\kappa'_n\}$  is a similar process on  $\mathbb{X}'$ , inheriting irreducibility from  $\mathcal{Y}$ . Its one-step transition probabilities  $(p'_{ij})$  say, for  $i, j \in \mathbb{X}'$ , can be found from the  $p_{ij}$  via taboo probability versions of the Chapman–Kolmogorov equations, or else directly

from a renewal-type equation. The same is true of the components of the matrix of transition distributions  $\mathbf{G}_{\mathbb{X}'}.$

- 10.3.4 (*Continuation*). When  $\mathcal{Y} = \{(t_n, \kappa_n)\}$  is a Markov renewal process, the process  $\mathcal{Y}_2 = \{(t_n, (\kappa_{n-1}, \kappa_n))\}$  is also a Markov renewal process, now with state space  $\mathbb{X}^{(2)}$ . The subset  $\mathcal{Y}'_2$  of observations of transitions of  $\mathcal{Y}$  restricted to a subset  $\mathbb{X}'_2 \subseteq \mathbb{X}^{(2)}$ , is again a Markov renewal process by Exercise 10.3.3(a); for simplicity confine attention to the case  $t_0 = 0$ . Then the first moment  $H_{(i,j),(k,\ell)}(t)$  say, for the number of jumps  $t_n$  in  $(0, t]$  for which  $k \mapsto \ell$  given a jump  $i \mapsto j$  at 0, is independent of  $i$  and satisfies the Markov renewal equation

$$H_{j,(k,\ell)}(t) = \delta_{jk} G_{k\ell}(t) + \sum_{h \in \mathbb{X}} \int_0^t dG_{jh}(u) H_{h,(k,\ell)}(t-u).$$

Then with  $\mathbf{H}_{(k,\ell)} = (H_{1,(k,\ell)} \cdots H_{j,(k,\ell)} \cdots)^\top$  and  $\boldsymbol{\delta}_k = (\delta_{1k} \cdots \delta_{jk} \cdots)^\top$ ,  $\mathbf{H}_{(k,\ell)}(t) = (\mathbf{H} * G_{k\ell})(t) \boldsymbol{\delta}_k$ , provided  $(\mathbf{G}^{(n*)} * \mathbf{H}_{(k,\ell)})(t) \rightarrow \mathbf{0}$  ( $n \rightarrow \infty$ ). The analogue  $m_{[r]}^{\mathbf{g}}$  for  $\mathcal{Y}'_2$  of the factorial moment densities at (10.3.6) when  $\mathbf{G}$  has density  $\mathbf{g} = (g_{ij})$ , in terms now of  $\mathbf{g}_{\mathbb{X}'_2}(\cdot) = (\delta_{(i,j),\mathbb{X}'_2} g_{ij}(\cdot))$  and with  $u_0 = 0 < u_1 < \cdots < u_r$ , is the product of matrices and vectors

$$m_{[r]}^{\mathbf{g}}(u_1, \dots, u_r) = \mathbf{p}_0^\top \prod_{s=1}^r (\mathbf{H}' * \mathbf{g}_{\mathbb{X}'_2}(u_s - u_{s-1})) \mathbf{1}.$$

[*Hint*: Deduce the equation for  $H_{j,(k,\ell)}$  from the Markovian nature of  $\{\kappa_n\}$  and a backwards Chapman–Kolmogorov decomposition. See Ball and Milne (2005); Darroch and Morris (1967) considered the Markovian case earlier.]

- 10.3.5 *Likelihoods for semi-Markov processes*. In the notation of Example 10.3(a), with  $t_0 = 0$  and  $X(0) = \kappa_0$ , show that the likelihood  $L$  of a semi-Markov process  $X(t)$  observed on  $(0, T]$  as having successive jumps at  $t_1, \dots, t_{N(T)}$  into states  $\kappa_1, \dots, \kappa_{N(T)}$  is expressible as

$$L = \left( \prod_{n=1}^{N(T)} g_{\kappa_{n-1}\kappa_n}(t_n - t_{n-1}) \right) [1 - G_{\kappa_{N(T)}}(T - t_{N(T)})],$$

where  $G_k(u) = \sum_j \int_0^u g_{kj}(u) du = \sum_j p_{kj} F_{kj}(u)$ . Write down the conditional intensity for the corresponding MPP, and verify that the above expression coincides with the usual form of the likelihood for an MPP.

- 10.3.6 *Alternative treatments of the Jelinski–Moranda process* [Example 10.3(d)].
- Formulate the likelihood when the initial state is treated as an unknown parameter.
  - Outline a Bayesian approach to the estimation of parameters in the process by finding the form of the posterior distribution for both the initial state and the parameters of the death process.

- 10.3.7 *Forward and backward equations for discrete time HMMs*.

- Use induction to verify formally the forward and backward equations (10.3.17) for estimation in HMM [cf. MacDonald and Zucchini (1997)].

- (b) Alternatively, show that the forward and backward equations reduce to matrix iterations such as

$$\alpha_{n+1}^\top = \alpha_n^\top P D_{n+1},$$

where  $P$  is the transition probability matrix, and  $D_n = \text{diag}\{f_1(z_n), f_2(z_n), \dots\}$ . Hence, for example, we have the explicit form  $\alpha_n^\top = \pi_0^\top P D_1 P D_2 \dots P D_n$ , with a similar expression for the backward probabilities. Use these to give straightforward proofs of the likelihood equation (10.3.16) and (10.3.17) from the state estimation Lemma 10.3.I.

- (c) To introduce normalized forms of the forward and backward probabilities, set  $\alpha_n^\dagger(i) = \alpha_n(i) / (\sum_j \alpha_n(j))$ , and similarly  $\beta_n^\dagger(i) = \beta_n(i) / (\sum_j \beta_n(j))$ . Reformulate equations (10.3.13) for these normalized forms.

[Remark: These quantities are numerically more stable; constants  $\rho_n$  say are needed to recover the original  $\alpha_n(\cdot)$ , and similarly for  $\beta_n(\cdot)$ .]

- 10.3.8 *Uniformization approximation for calculating matrix exponentials* [Gross and Miller (1984)]. Show that if  $m > \max\{-q_{ii}\}$  is chosen just larger than the maximum diagonal element,  $\exp(Qt)$  can be represented in the form  $e^{Qt} = e^{-mt} e^{mtA}$ , where  $A$  denotes the nonnegative matrix  $Q/m + I$ , and hence that

$$\exp(Qt) = e^{-mt} \sum_{n=0}^{\infty} \frac{(mt)^n}{n!} A^n = \sum_{n=0}^{\infty} p_n(t) A^n,$$

where the  $\{p_n(t)\}$  are the probabilities in a  $\text{Poi}(mt)$  distribution. Sufficient iterates of the fixed matrix  $A$ , and a sufficient range of values of the Poisson probabilities, can then be computed to give an effective algorithm for determining values of the matrix exponential to high precision.

- 10.3.9 *MPP extension of Cox process driven by Markov chain* [see Examples 10.3(e) and 10.3(g)]. Let  $Q$  be the matrix of transition rates for a  $K$ -state Markov process  $X(t)$ . Suppose that for  $k = 1, \dots, K$ , when  $X(t) = k$ , points are generated at rate  $\lambda_k$  with conditional mark distribution  $f_k(x)$ . Write down the conditional intensity  $\lambda(t, x)$  of the corresponding MPP in terms of the current state  $\xi$ , and use it to find the complete likelihood for observations over an interval  $(0, T)$ . Verify that the forward probabilities  $\alpha_t(k)$  for the observations up to time  $t$ , given the sequence  $\{(t_n, x_n): n = 1, \dots, N(t)\}$  can be represented as a matrix product  $\alpha_t^\top = \pi_0^\top R(0, t]$  where

$$R(0, t] = J(t_1) \Lambda D(x_1) J(t_2 - t_1) \Lambda D(x_2) \dots J(t - t_{N(t)})$$

and  $J(t)$  has the same interpretation as in (10.3.6b),  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_K)$ , and  $D(x) = \text{diag}(f_1(x), \dots, f_K(x))$ . Use these representations to find a set of sufficient statistics for the complete likelihood, and the appropriate extension of the E- and M-steps of Example 10.3(g).

- 10.3.10 *PH-distributions and their Laplace transforms.*

- (a) Write out the distribution (10.3.24) explicitly in the special case that  $Q$  is a  $2 \times 2$  matrix.
- (b) Show that both sums and mixtures of exponentials can be represented as PH-distributions, and find the elements in their representation.  
[Hint: Restrict  $Q$  to the diagonal and superdiagonal; for a mixture of exponentials  $Q$  has pure diagonal form.]

- (c) Show that the distribution (10.3.24) has Laplace transform  $\ell(s) = \boldsymbol{\pi}^\top [sI - Q]^{-1} \mathbf{1}$ . Use this representation to find the renewal function. Generalize to a matrix renewal function as in Example 10.3(a).

10.3.11 *Interval distributions and correlations for the stationary BMAP process.* Find a matrix exponential (phase-type) representation for the time interval between one batch arrival and the next, first assuming the batches (of sizes  $\ell_1$  and  $\ell_2$  say) are associated with transitions into states  $k_1$  and  $k_2$  respectively, and then without making this assumption. Does this representation imply that the observed process can be regarded as a semi-Markov process with states  $\ell = 1, \dots, L$ ? Find also the correlations between successive intervals.

## 10.4. Markov Point Processes

Markov processes in time heavily influenced the growth of applied probability modelling in the second half of the twentieth century. Dobrushin (1968) successfully described a *spatial* Markov property, and the 1970s saw Hammersley and Clifford (unpublished, 1971), Moran (1973), and Besag (1974), for example, describing processes on two-dimensional lattices with a Markovian property, exploiting adjacency of points on a lattice to limit the range of stochastic dependence [Isham (1981) presents a broad review]. Ripley and Kelly (1977) gave a definitive description of *Markov point processes*, with an important sequel by Baddeley and Møller (1989) and subsequent expansion by Baddeley and co-workers; there is a consolidated account in van Lieshout (2000), and a broad exposition with examples in Møller and Waagepetersen (2004, Chapter 6 and Appendices F and G). Many mathematical properties of Gibbs distributions were anticipated earlier in statistical physics [Ruelle (1969, Chapter 3), Preston (1976)]. Georgii (1988) gives a probabilistic approach to Gibbs measures on multidimensional lattices.

The practical appeal of Markov models lies in the form of the joint probability distribution in many variables: it is expressible as the product of many conditional probabilities each in a small number of variables defined only on ‘adjacent’ time points. This then raises the possibility of specifying the model purely in terms of local conditional probabilities. The Papangelou conditional intensity function in Definition 10.4.I plays this role in point process modelling when coupled with an algebraic relationship property of ‘neighbourliness’ denoted below by  $\sim$ , although the paradigm example of a renewal process fails this relationship in general [see around (10.4.16)].

Consider then simple finite point processes on a c.s.m.s.  $\mathcal{X}$ , often a bounded subset of  $\mathbb{R}^2$  or  $\mathbb{R}^3$ . Proposition 5.3.II gives a canonical space for finite point processes as the union  $\mathcal{X}^\cup$  of all product spaces  $\{\mathcal{X}^{(n)}: n = 0, 1, \dots\}$  with generic element  $\mathbf{x} = \{x_1, \dots, x_n\}$  for which  $n = \text{card}\{\mathbf{x}\} = n(\mathbf{x})$ . In Section 5.3 we used Janossy measures  $\{J_n(\cdot): n = 0, 1, \dots\}$  to describe finite point processes; here we use their density functions  $\{j_n(\mathbf{x}): n = 1, 2, \dots\}$  with respect to  $n$ -fold products of Lebesgue measure  $\ell$  on  $\mathcal{X}$  with  $\ell(\mathcal{X}) < \infty$ , symmetric as

around (5.3.1–2). It is sometimes more convenient to describe the distributions through the density function  $f = \{f_n\}$  with respect to the distribution  $\pi = \{\pi_n\}$  of a totally finite unit-rate Poisson process on the measurable space  $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ ; if more generally this Poisson process has probability measure  $\mathcal{P}_{\mu}$ , then  $f$  is the likelihood of the process relative to  $\mathcal{P}_{\mu}$  as in (7.1.7). When  $\mathcal{X} = \mathbb{R}^d$  and  $\mu = \ell$  these two descriptions are virtually equivalent inasmuch as  $j_n(\mathbf{x})/f_n(\mathbf{x}) = e^{-\ell(\mathcal{X})}$   $\ell$ -a.e. (see Exercise 10.4.1).

For Janossy densities, always symmetric, we have the interpretation as at (5.4.13), rewritten here in notation closer to the present setting, namely

$$j_n(\mathbf{x}) dx_1 \dots dx_n = \Pr \left\{ \begin{array}{l} \text{exactly } n \text{ points in a realization:} \\ \text{one in each subset } (x_i, x_i + dx_i) \\ (i = 1, \dots, n), \text{ and none elsewhere} \end{array} \right\}. \quad (10.4.1)$$

Even more generally,  $f$  (= a collection  $\{f_n\}$  of symmetric functions) may be a density with respect to any totally finite measure on the space  $\mathcal{X}^{\cup*}$ . For example, if  $\mathcal{X}$  is a finite set and we use counting measure on  $\mathcal{X}$  as the reference measure, then for each  $n = 0, 1, \dots, \#(\mathcal{X})$  we should want  $f_n(\mathbf{x})$  equal to  $n!$  times the probability mass associated with the point set  $\mathbf{x}$  for which  $n = n(\mathbf{x})$ .

Realizations of finite point processes with densities as at (10.4.1) are a.s. simple, and for them we have  $\mathbf{x} \in \mathcal{X}^{\cup*}$  a.s., where  $\mathcal{X}^{\cup*}$  denotes the subset of  $\mathcal{X}^{\cup}$  containing in each component  $\mathcal{X}^{(n)}$  only those  $\mathbf{x} = \{x_1, \dots, x_n\}$  for which  $x_i \neq x_j$  ( $i \neq j$ ). This description proves more convenient here than the integer-valued measures  $N$  used in Chapter 9, but they are equivalent because  $N(\cdot) = \sum_{i=1}^{n(\mathbf{x})} \delta_{x_i}(\cdot)$ , and  $N \in \mathcal{N}_{\mathcal{X}}^*$  when  $\mathbf{x} \in \mathcal{X}^{\cup*}$  a.s. (cf. Proposition 9.1.X). We use  $\mathbf{x}$  to denote both an element of  $\mathcal{X}^{\cup*}$  and a subset of  $\mathcal{X}$ . For  $y \in \mathcal{X}$  we usually write  $\mathbf{x} \cup y$  rather than  $\mathbf{x} \cup \{y\}$ . For the difference we write variously

$$\mathbf{x} \setminus y = \mathbf{x} \setminus \{y\} = \mathbf{x}_y.$$

An inclusion written  $\mathbf{y} \subset \mathbf{x}$  is strict. With this understanding, a *nonnegative measurable function*  $f: \mathcal{X}^{\cup*} \rightarrow \mathbb{R}_+$  is a *density function of a simple finite point process* when  $f(\mathbf{x}) = f_{n(\mathbf{x})}(\mathbf{x})$  for  $\mathbf{x} \in \mathcal{X}^{\cup*}$  and  $f$  is integrable as at (10.4.2b). For such a process it is convenient to write (with mixed notation), when the reference measure is  $\pi(\cdot)$ , for a measurable function  $g$ ,

$$\mathbb{E}[g(N)] = \int_{\mathcal{X}^{\cup*}} g(\mathbf{x}) f(\mathbf{x}) \pi(d\mathbf{x}) = \sum_{n=0}^{\infty} \int_{\mathcal{X}^{(n)}} g(\mathbf{x}) f_n(\mathbf{x}) \pi_n(d\mathbf{x}), \quad (10.4.2a)$$

and  $f$  satisfies

$$\int_{\mathcal{X}^{\cup*}} f(\mathbf{x}) \pi(d\mathbf{x}) \equiv \sum_{n=0}^{\infty} \int_{\mathcal{X}^{(n)}} f_n(\mathbf{x}) \pi_n(d\mathbf{x}) = 1. \quad (10.4.2b)$$

Often, the process of interest is on a subset of some Euclidean space  $\mathbb{R}^d$  and it is *regular* in the sense of Definition 7.1.I, whereas the reference probability measure is a unit rate Poisson process, in which case this Poisson process is on a compact subset of  $\mathbb{R}^d$ .



**Definition 10.4.I.** Given a simple finite point process on  $\mathcal{X}$  with a density  $f: \mathcal{X}^{\cup*} \mapsto \mathbb{R}_+ \equiv \{f_n(\mathbf{x}): \mathbf{x} \in \mathcal{X}^{(n)*}, n = 1, 2, \dots\}$ , the function

$$\rho(y | \mathbf{x}) = \begin{cases} f_1(y)/f_0(\emptyset) & (y \in \mathcal{X}, \mathbf{x} = \emptyset), \\ \frac{f_{n+1}(\mathbf{x} \cup y)}{f_n(\mathbf{x})} \equiv \frac{f(\mathbf{x} \cup y)}{f(\mathbf{x})} & (y \in \mathcal{X} \setminus \mathbf{x}, \mathbf{x} \in \mathcal{X}^{(n)*}, n = 1, 2, \dots) \end{cases} \quad (10.4.3)$$

defines its Papangelou conditional intensity [set  $\rho(y | \mathbf{x}) = 0$  if  $f_n(\mathbf{x}) = 0$ ].

We remark that in view of our earlier comments,  $\rho(y | \mathbf{x})$  can just as easily be given in terms of Janossy densities: see Exercise 10.4.2.

Ripley and Kelly's definition of a Markov point process involves a concept of 'adjacency' of pairs of points and an analogue of a 'neighbourhood' based on such a notion of adjacency. This concept for points  $y, z \in \mathcal{X}$  is embodied in some reflexive symmetric relation  $\sim$  (meaning that  $y \sim y$  and for  $z \neq y$ ,  $y \sim z$  if and only if  $z \sim y$ ), as, for example,  $y \sim z$  if and only if  $|y - z| \leq R$  for some finite positive  $R$  (but, see Exercise 10.4.3). Any such relation  $\sim$  defines a  $(\sim)$ -neighbourhood (or just neighbourhood for short)  $b^\sim(y)$  of any  $y \in \mathcal{X}$  by

$$b^\sim(y) = \{z \in \mathcal{X}: z \sim y\}. \quad (10.4.4)$$

This definition is easily extended to  $\mathbf{y} \in \mathcal{X}^{\cup*}$  by setting  $b^\sim(\mathbf{y}) = \{z \in \mathcal{X}: z \sim y \text{ for some } y \in \mathbf{y}\}$ .

**Definition 10.4.II.** A simple finite point process with density function  $f: \mathcal{X}^{\cup*} \mapsto \mathbb{R}_+$  is a Markov point process if for every  $\mathbf{x}$  with  $f(\mathbf{x}) > 0$  its Papangelou conditional intensity  $\rho(y | \mathbf{x}) = f(\mathbf{x} \cup y)/f(\mathbf{x})$  satisfies

$$\rho(y | \mathbf{x}) = g(y, \mathbf{x} \cap b^\sim(y)) \quad (y \in \mathcal{X} \setminus \mathbf{x}), \quad (10.4.5)$$

where  $g: \mathcal{X} \times \mathcal{X}^{\cup*} \rightarrow \mathbb{R}_+$ . Call such  $f$  a Markov density function.

In other words, for  $f$  to be the density function of a Markov point process, we require that, for all  $\mathbf{x} \in \mathcal{X}^{\cup*}$ ,  $y \in \mathcal{X} \setminus \mathbf{x}$  and writing  $n = n(\mathbf{x})$ , the  $(n+1)$ -dimensional joint density function  $f_{n+1}(\mathbf{x} \cup y)$  must be expressible as a product of the  $n$ -dimensional joint density function  $f_n(\mathbf{x})$  and some function  $g(\cdot, \cdot)$  that depends only on  $y$  and those elements of  $\mathbf{x}$  that lie in the  $(\sim)$ -neighbourhood  $b^\sim(y)$  of the 'extra' point  $y$ , i[e.e.,  $g(y, \cdot)$  is independent of all elements of  $\mathbf{x}$  and  $\mathcal{X}$  that are not  $(\sim)$ -neighbours of  $y$ ].

In particular, for  $y, z \in \mathbf{x}$  such that  $z \not\sim y$  (hence,  $z \notin b^\sim(y)$ ), we have  $\mathbf{x}_y \cap b^\sim(y) = (\mathbf{x}_z)_y \cap b^\sim(y)$ . Thus, for such  $y$  and  $z$ , the relations

$$\frac{f(\mathbf{x})}{f(\mathbf{x}_y)} = g(y, \mathbf{x}_y \cap b^\sim(y)) = g(y, (\mathbf{x}_z)_y \cap b^\sim(y)) = \frac{f(\mathbf{x}_z)}{f((\mathbf{x}_z)_y)} \quad \text{if } z \not\sim y \quad (10.4.6)$$

hold for the density function  $f$  of a Markov point process, so that

$$f(\mathbf{x}) = \frac{f(\mathbf{x}_y) f(\mathbf{x}_z)}{f((\mathbf{x}_z)_y)} \quad (y, z \in \mathbf{x}, z \not\sim y). \quad (10.4.6')$$

In (10.4.6),  $f(\mathbf{x})$  is a compact notation for  $f_{n(\mathbf{x})}(\mathbf{x})$ , so  $f(\mathbf{x}_y) = f_{n(\mathbf{x}_y)}(\mathbf{x}_y) = f_{n(\mathbf{x})-1}(\mathbf{x}_y)$ . Theorem 10.4.V shows that there are far-reaching consequences of this condition which states that the conditional intensity of adding an ‘extra’ point  $y$  to a set  $\mathbf{x}_y$  is independent of any point  $z$  that is *not* in the neighbourhood  $b^\sim(y)$ .

EXAMPLE 10.4(a). An inhomogeneous Poisson process with density  $\mu(\cdot)$  has  $j_n(\mathbf{x}) = \prod_{x_i \in \mathbf{x}} \mu(x_i)$ , so

$$\rho(y | \mathbf{x}) = f(\mathbf{x} \cup y) / f(\mathbf{x}) = \mu(y),$$

which, being independent of  $\mathbf{x}$ , is clearly of the required form (10.4.5) for the process to be a Markov point process.  $\square$

A major property of Markov point processes is the Hammersley–Clifford Representation Theorem 10.4.V below. It is important practically because it expresses the joint density function of a point set  $\mathbf{x}$  as a product of (conditional) probability density functions of many smaller subsets  $\mathbf{y} \subset \mathbf{x}$ . When it was first proved, the result also identified classes of Markov random fields on one hand and Gibbs states with nearest neighbour potentials, the latter being already well known in statistical physics [cf. Clifford (1990)].

Specifically, (10.4.7) expresses the density function  $f$  of a Markov point process as products of terms involving another function  $\phi: \mathcal{X}^{\cup*} \mapsto \mathbb{R}_+$  which is simpler than  $f$  in that  $\phi(\mathbf{x}) = 1$  as soon as the set  $\mathbf{x}$  includes a pair of distinct elements,  $y, z$  say, for which  $y \not\sim z$ . In other words,  $\phi(\mathbf{x})$  can differ from 1 only when  $\mathbf{x}$  is a *clique* defined in 10.4.III(a) below.

**Definition 10.4.III** (Cliques). *Let  $\mathbf{x}, \mathbf{y}$  be finite nonempty subsets of  $\mathcal{X}$  (equivalently,  $\mathbf{x}, \mathbf{y} \in \mathcal{X}^{\cup*}$ ), and  $\sim$  a reflexive symmetric relation on elements of  $\mathcal{X}$ .*

- (a)  $\mathbf{y}$  is a clique (write  $(\sim)$ -clique if distinction is needed) if it is the empty set or a singleton set or else  $y \sim z$  for every two-point subset  $\{y, z\} \subseteq \mathbf{y}$ .
- (b) A clique  $\mathbf{y} \subset \mathbf{x}$  is a maximal clique of the set  $\mathbf{x}$  when  $\mathbf{y} \cup \{z\}$  is not a clique for every  $z \in \mathbf{x} \setminus \mathbf{y}$ .
- (c)  $\text{Clq}(\mathbf{x})$  (or  $(\sim)$ - $\text{Clq}(\mathbf{x})$  if needed) is the family of all cliques  $\mathbf{y} \subseteq \mathbf{x}$ .

To understand cliques better, we list below some of their properties, leaving their proof to Exercise 10.4.5. In this list,  $\sim$  is a reflexive symmetric binary relation on elements of the c.s.m.s.  $\mathcal{X}$  containing points  $y, z, \dots$  and finite nonempty subsets  $\mathbf{x}, \mathbf{y}, \dots$ . Note [compare (v) and (vi)] that cliques do not in general yield equivalence relations (this was wrongly claimed on p. 219 of the first printing of Volume I).

- (i) The empty set and all one-point sets  $\{y\}$  are cliques.
- (ii) The two-point set  $\{y, z\}$  with  $y \neq z$  is a clique if and only if  $y \sim z$ .
- (iii) If  $\mathbf{y} \subset \mathbf{x}$  and  $\mathbf{x}$  is a clique, then so also is  $\mathbf{y}$ .
- (iv)  $\sim$  is transitive within a clique.
- (v) Distinct maximal cliques can overlap when  $\sim$  is not transitive.

- (vi) If  $\sim$  is transitive then distinct maximal cliques cannot overlap, and the maximal cliques that are subsets of  $\mathbf{x}$  provide a decomposition of  $\mathbf{x}$  into equivalence classes.

**Definition 10.4.IV.** Let  $h, \phi$  be nonnegative real-valued functions on  $\mathcal{X}^{\cup*}$ .

- (a) A family  $\mathcal{H}$  of subsets of  $\mathcal{X}^{\cup*}$  is hereditary if  $\mathbf{x} \in \mathcal{H}$  implies  $\mathbf{y} \in \mathcal{H}$  whenever  $\mathbf{y} \subset \mathbf{x}$ .  
 (b)  $h$  is an hereditary function if  $h(\mathbf{x}) > 0$  implies  $h(\mathbf{y}) > 0$  for every  $\mathbf{y} \subset \mathbf{x}$ .  
 (c)  $\phi$  is a  $(\sim)$ -interaction function if  $\phi(\mathbf{x}) = 1$  whenever  $\mathbf{x}$  is not a  $(\sim)$ -clique.

Møller and Waagepetersen (2004, Example F.2) describe a process defined on a space with each element of  $\mathcal{H}$  a strict subset of  $\mathcal{X}^{\cup*}$  [i.e.,  $\mathcal{H} \subset \mathcal{X}^{\cup*} \neq \mathcal{H}$ ].

The following theorem is a key result for Markov point processes. Its first version is unpublished [see Besag (1974), discussion by its originators there, and Clifford (1990)]; this version is due largely to Ripley and Kelly (1977).

**Theorem 10.4.V** (Hammersley–Clifford Representation). A probability density function  $f: \mathcal{X}^{\cup*} \mapsto \mathbb{R}_+$  is the density function of a Markov point process if and only if there is a  $(\sim)$ -interaction function  $\phi$  such that for nonempty sets  $\mathbf{x}$ ,

$$f(\mathbf{x}) = \prod_{\mathbf{y} \subseteq \mathbf{x}} \phi(\mathbf{y}) = \prod_{\mathbf{z} \in \text{Clq}(\mathbf{x})} \phi(\mathbf{z}) \quad (\mathbf{x} \in \mathcal{X}^{\cup*}). \quad (10.4.7)$$

*Remark.* Because  $\emptyset \in \text{Clq}(\mathbf{x})$ ,  $\phi(\emptyset)$  appears as a factor of  $f(\mathbf{x})$  for every  $\mathbf{x}$ , and hence acts as a multiplicative normalizing constant in this representation.

**PROOF.** For a  $(\sim)$ -interaction function  $\phi$ ,  $\phi(\mathbf{z}) = 1$  when  $\mathbf{z}$  is not a clique so in (10.4.7) the second equality is trivial and only the first needs proof.

Given  $\phi$ , define  $\tilde{f}(\cdot)$  by either product in (10.4.7), and suppose that  $\tilde{f}$  is integrable and hence can be and is normalized to be a density function on  $\mathcal{X}^{\cup*}$ . Then for  $\mathbf{x}$  with  $\tilde{f}(\mathbf{x}) > 0$  and  $z \in \mathcal{X} \setminus \mathbf{x}$ ,  $\tilde{f}(\mathbf{x} \cup z)/\tilde{f}(\mathbf{x}) = \prod_{\mathbf{y} \subseteq \mathbf{x}} \phi(\mathbf{y} \cup z)$ . Because  $\phi$  is an interaction function, it is possible for  $\phi(\mathbf{y} \cup z) \neq 1$  only when  $\mathbf{y} \cup z$  is a clique, so that  $\mathbf{y} \cup z \subseteq b^\sim(z)$ , hence  $\prod_{\mathbf{y} \subseteq \mathbf{x}, \mathbf{y} \cup z \in \text{Clq}(\mathbf{x})} \phi(\mathbf{y} \cup z) = g(z, \mathbf{x} \cap b^\sim(z))$  for some function  $g$ . We thus have the form at (10.4.5), and  $\tilde{f}$  as defined is the density function of a Markov point process.

Conversely, let  $f$  be the density function of a Markov point process, and define  $\psi$  iteratively by

$$\psi(\mathbf{x}) = \begin{cases} f(\emptyset) & \text{if } \mathbf{x} = \emptyset, \\ 1 & \text{if nonempty } \mathbf{x} \text{ is not a clique,} \\ f(\mathbf{x}) / \prod_{\mathbf{y} \subseteq \mathbf{x}} \psi(\mathbf{y}) & \text{otherwise,} \end{cases} \quad (10.4.8)$$

taking  $0/0 = 1$  if need be. Then  $\psi$  is an interaction function, and it remains to show that the representation (10.4.7) holds with  $\phi = \psi$ .

Suppose  $\mathbf{x}$  is given, with  $n(\mathbf{x}) = r \geq 2$ , and that (10.4.7) has been proved for all  $\mathbf{x}'$  with  $n(\mathbf{x}') \leq r - 1$ ; note that it is true for  $r - 1 = 1$ .

First, if  $f(\mathbf{x}) = 0$  and  $\prod_{\mathbf{y} \subset \mathbf{x}} \psi(\mathbf{y}) = 0$ , then  $\prod_{\mathbf{y} \subseteq \mathbf{x}} \psi(\mathbf{y}) = 0$  also and (10.4.7) holds.

Next, if  $\prod_{\mathbf{y} \subset \mathbf{x}} \psi(\mathbf{y}) > 0$ , then for any  $\mathbf{z} \subset \mathbf{x}$  (hence,  $n(\mathbf{z}) \leq r - 1$ ), we have  $f(\mathbf{z}) = \prod_{\mathbf{y} \subset \mathbf{z}} \psi(\mathbf{y}) > 0$  because  $\mathbf{y} \subset \mathbf{z} \subset \mathbf{x}$ , and the right-hand side of (10.4.6) is positive. If also  $f(\mathbf{x}) = 0$ , then the left-hand side of (10.4.6) is zero so we have reached a contradiction if  $\mathbf{x}$  not a clique, whereas if  $\mathbf{x}$  is a clique then by the last case of (10.4.8), (10.4.7) holds.

Finally, when both  $f(\mathbf{x}) > 0$  and  $\prod_{\mathbf{y} \subset \mathbf{x}} \psi(\mathbf{y}) > 0$ , either  $\mathbf{x}$  is a clique and by the last case of (10.4.8), (10.4.7) holds, or else  $\mathbf{x}$  is not a clique and therefore there exist  $y, z \in \mathbf{x}$  such that  $z \not\sim y$ . Because  $f$  is the density function of a Markov point process, (10.4.6') holds with arguments in the right-hand side there having at most  $r - 1$  points, so that

$$f(\mathbf{x}) = \frac{f(\mathbf{x}_y)f(\mathbf{x}_z)}{f((\mathbf{x}_z)_y)} = \frac{\prod_{\mathbf{w} \subseteq \mathbf{x}_y} \psi(\mathbf{w}) \prod_{\mathbf{w} \subseteq \mathbf{x}_z} \psi(\mathbf{w})}{\prod_{\mathbf{w} \subseteq (\mathbf{x}_z)_y} \psi(\mathbf{w})}. \quad (10.4.9)$$

Now  $\psi(\mathbf{x}) = 1$  because  $\mathbf{x}$  is not a clique, and for any other  $\mathbf{w} \subset \mathbf{x}$ , either (i)  $\mathbf{w}$  contains both  $y$  and  $z$  (and  $\psi(\mathbf{w}) = 1$  because it is not a clique), or (ii)  $\mathbf{w}$  contains neither, in which case  $\mathbf{w} \subseteq (\mathbf{x}_y)_z$ , or (iii)  $\mathbf{w}$  contains exactly one of  $y$  and  $z$  so it is of the form  $\mathbf{w}' \cup y$  or  $\mathbf{w}' \cup z$  for  $\mathbf{w}' \subseteq (\mathbf{x}_y)_z$ . These possibilities and facts imply that the right-hand side of (10.4.9) equals  $\prod_{\mathbf{w} \subseteq \mathbf{x}} \psi(\mathbf{w})$ ; that is, (10.4.6) holds when  $n(\mathbf{x}) = r$ .  $\square$

In typical applications, the Papangelou conditional intensity or the clique density function  $\phi(\cdot)$  may be known, in particular, when  $n(\mathbf{x})$  is 'small' for most, if not all, cliques  $\mathbf{x}$ .

EXAMPLE 10.4(b) *Strauss process* [continued from Example 7.1(c) and Exercise 7.1.8]. In the notation of this chapter, the Janossy density of the Strauss model of Example 7.1(c), for which  $x \sim y$  if and only if  $\|x - y\| \leq R$ , is given for  $0 < \beta < \infty$ ,  $0 < \gamma \leq 1$ , and  $\alpha$  a normalizing constant, by

$$j_n(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \gamma^{m(\mathbf{x}, R)},$$

where  $m(\mathbf{x}, R)$  is the number of distinct elements  $y, z \in \mathbf{x}$  for which  $y \sim z$ . Then

$$\rho(y | \mathbf{x}) = \beta^{n(\mathbf{x} \cup y) - n(\mathbf{x})} \gamma^{m(\mathbf{x} \cup y, R) - m(\mathbf{x}, R)} = \beta \gamma^{n(\mathbf{x} \cap S_R(y))},$$

the exponent of  $\gamma$  being equal to the number of elements of  $\mathbf{x}$  within distance  $R$  of  $y$ . Then  $\rho(y | \mathbf{x})$  is of the required form (10.4.5) for a Markov process, and therefore the representation (7.1.5) for the Janossy density follows from the Hammersley–Clifford theorem.

Kelly and Ripley (1976) showed that the Strauss process is uniquely characterized by two properties: its density function is hereditary, and its Papangelou conditional intensity is of the form, for  $\mathbf{x} \in \mathcal{X}^{\cup*}$  and  $y \in \mathcal{X} \setminus \mathbf{x}$ ,

$$\rho(y | \mathbf{x}) = g(n(\mathbf{x} \cap S_R(y))),$$

where  $g: \mathbb{Z}_+ \mapsto \mathbb{R}_+$  and  $S_R(y)$  is the closed ball with centre  $y$  and radius  $R$ .  $\square$

EXAMPLE 10.4(c) *Area-interaction point process* [Baddeley and van Lieshout (1995), van Lieshout (2000, Section 4.3)]. Suppose a simple finite point process on a compact subset  $\mathcal{X} \subset \mathbb{R}^d$  has density with respect to a Poisson process at unit rate on  $\mathcal{X}$  given by

$$f(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \gamma^{-\ell(\mathcal{X} \cap U_R(\mathbf{x}))} \quad (\mathbf{x} \in \mathcal{X}^{\cup*}; \alpha, \beta, \gamma > 0), \quad (10.4.10)$$

where  $\ell$  denotes Lebesgue measure on  $\mathbb{R}^d$ ,  $U_R(\mathbf{x}) = \bigcup_{i=1}^{n(\mathbf{x})} S_R(x_i)$  is the union of  $n(\mathbf{x})$  spheres with centres  $x_i \in \mathbf{x}$  and common radii  $R > 0$ , and  $\alpha$  is a normalizing constant. In this form it is also called the penetrable spheres model used by Widom and Rowlinson (1970) and others to study liquid–vapour equilibrium questions (note also the next example); the model in  $\mathbb{R}$  is tractable as a Kingman regenerative phenomenon [Hammersley *et al.* (1975)].

Because  $f(\cdot)$  at (10.4.10) is a density function with respect to a unit-rate Poisson distribution,  $\Pr\{N(\mathcal{X}) = n\}$  lies between

$$e^{-\ell(\mathcal{X})} \frac{[\ell(\mathcal{X})]^n}{n!} \cdot \alpha \beta^n \gamma^{-\ell(\mathcal{X})} \quad \text{and} \quad e^{-\ell(\mathcal{X})} \frac{[\ell(\mathcal{X})]^n}{n!} \cdot \alpha \beta^n,$$

hence  $\alpha$  lies between  $(e^{\beta-1}/\gamma)^{-\ell(\mathcal{X})}$  and  $e^{-(\beta-1)\ell(\mathcal{X})}$ , so  $N(\mathcal{X})$  is finite-valued a.s. Thus  $f$  is indeed the density function of a simple finite point process, being Poisson when  $\gamma = 1$ . Its Papangelou conditional intensity is given by

$$\rho(y | \mathbf{x}) = \beta \gamma^{-\ell(\mathcal{X} \cap [S_R(y) \setminus U_R(\mathbf{x})])},$$

and because the set difference here is a function of  $R$ ,  $y$  and those  $x_i \in \mathbf{x}$  for which  $|y - x_i| < 2R$ , a function  $g(y, \mathbf{x} \cap S_{2R}(y))$  can be constructed to satisfy (10.4.5); that is, a process with density (10.4.10) is a Markov point process, with  $x \sim y$  if and only if  $|x - y| \leq 2R$ . It is attractive or repulsive as  $\gamma \geq$  or  $\leq 1$ , respectively, and Poisson for  $\gamma = 1$ , where a simple point process with Papangelou conditional intensity  $\rho$  is called *attractive* (respectively, *repulsive*) whenever, for all  $\mathbf{x}, \mathbf{y}$  with  $\mathbf{x} \subset \mathbf{y}$  and  $z \notin \mathbf{y}$ ,  $\rho(z | \mathbf{x}) \leq \rho(z | \mathbf{y})$  ( $\rho(z | \mathbf{x}) \geq \rho(z | \mathbf{y})$ ).

Baddeley and van Lieshout allow a general version of this model by replacing Lebesgue measure  $\ell$  in the exponent of (10.4.10) by a totally finite Borel regular measure  $\nu$  say and the spheres  $S_R(x_i)$  by compact sets  $\tilde{S}(x) \subset \mathcal{X}$  where  $|\tilde{S}|: \mathcal{X} \mapsto \mathbb{R}_+$  is continuous and bounded. They also show how a birth-and-death process can have the model as a stationary distribution.

Observe that for  $\gamma < 1$ , the process is not as ‘aggressively’ repulsive as the Strauss process in which the exponent of  $\gamma$  increases quadratically in  $n(\mathbf{x})$  whereas in (10.4.10) it changes at most linearly in  $n(\mathbf{x})$ , with such changes becoming closer to zero with greater overlap between different spheres  $S_R(x_i)$  as  $n(\mathbf{x})$  increases.  $\square$

EXAMPLE 10.4(d) *Penetrable spheres mixture model* [Widom and Rowlinson (1970), van Lieshout (2000, Examples 2.8, 2.11)]. Consider a bivariate point process  $(\mathbf{x}_1, \mathbf{x}_2)$  constructed on a bounded set  $\mathcal{X}$  by superposing two independent Poisson processes  $N_j$  at rates  $\beta_j$  ( $j = 1, 2$ ) subject to every  $x \in \mathbf{x}_1$  being

at a minimum distance  $R$  from every  $y \in \mathbf{x}_2$ ; that is,  $d(\mathbf{x}_1, \mathbf{x}_2) > R$ , where for nonempty finite point sets  $\mathbf{x}, \mathbf{y}$ ,  $d(\mathbf{x}, \mathbf{y}) = \min_{x \in \mathbf{x}, y \in \mathbf{y}} d(x, y)$ . Then the density relative to a unit rate Poisson process on  $\mathcal{X}$  equals

$$f(\mathbf{x}_1, \mathbf{x}_2) = \alpha \beta_1^{n(\mathbf{x}_1)} \beta_2^{n(\mathbf{x}_2)} I_{\{d(\mathbf{x}_1, \mathbf{x}_2) > R\}} \quad (10.4.11)$$

for some normalizing constant  $\alpha$ . This density is positive when, given  $\mathbf{x}_1$ , all  $n(\mathbf{x}_2)$  of the points of the second component avoid the union  $U_R(\mathbf{x}_1)$  [cf. (10.4.10)] of circles of radius  $R$  around all the points of the first component. It follows, using Poisson process properties, that the marginal density of the first component equals

$$\alpha \beta_1^{n(\mathbf{x}_1)} \sum_{i=0}^{\infty} \beta_2^i \frac{e^{-\ell(\mathcal{X})}}{i!} [\ell(\mathcal{X} \setminus U_R(\mathbf{x}_1))]^i = \alpha \beta_1^{n(\mathbf{x}_1)} e^{(\beta_2-1)\ell(\mathcal{X})} e^{-\beta_2 \ell(\mathcal{X} \cap U_R(\mathbf{x}_1))}.$$

Thus, the marginal distributions in this mixture model are just the cases  $1 < \gamma = e^{\beta_2}$  or  $e^{\beta_1}$  of the area-interaction model of Example 10.4(c).  $\square$

Given a space  $\mathcal{X}$ , a multitude of possible symmetric reflexive relations can be defined. When two such relations  $\overset{i}{\sim}$  ( $i = 1, 2$ ) are given, the intersection relation  $\overset{\cap}{\sim}$  defined by  $y \overset{\cap}{\sim} z$  if and only if both  $y \overset{1}{\sim} z$  and  $y \overset{2}{\sim} z$ , and the union relation  $\overset{\cup}{\sim}$  defined by  $y \overset{\cup}{\sim} z$  if and only if at least one of  $y \overset{1}{\sim} z$  and  $y \overset{2}{\sim} z$  holds, are both well-defined symmetric reflexive relations. When the relations  $\overset{1}{\sim}$  and  $\overset{2}{\sim}$  are ordered in the sense that (say)  $y \overset{1}{\sim} z$  implies  $y \overset{2}{\sim} z$  for all  $y, z$ , it follows that we can identify  $\overset{\cap}{\sim}$  and  $\overset{\cup}{\sim}$  with  $\overset{1}{\sim}$  and  $\overset{2}{\sim}$ , respectively.

Now suppose that  $\{f'_n\}$  and  $\{f''_n\}$  are Markov density functions for point sets on  $\mathcal{X}$  with respect to  $\overset{1}{\sim}$  and  $\overset{2}{\sim}$ , respectively, and that

$$\frac{1}{c} = \int_{\mathcal{X}^{\cup*}} f'(\mathbf{x}) f''(\mathbf{x}) \pi(d\mathbf{x}) < \infty \quad (10.4.12)$$

for some finite constant  $c > 0$ , so that  $cf = cf'f''$  is a density function, where  $f(\cdot) = \{f_n(\cdot)\} = \{f'_n(\cdot)f''_n(\cdot)\}$ , and  $f(\mathbf{x}) = f'_{n(\mathbf{x})}(\mathbf{x})f''_{n(\mathbf{x})}(\mathbf{x})$ .

**Proposition 10.4.VI.** *Let  $f = f'f''$  be the product of two Markov density functions. When  $cf$  is a density function for some finite positive  $c$ , it is a Markov density function.*

PROOF. The Hammersley–Clifford representation applied to the densities  $f'$  and  $f''$  for  $n = n(\mathbf{x})$  implies that

$$f_n(\mathbf{x}) = \prod_{\mathbf{y} \in (\overset{1}{\sim})\text{-Clq}(\mathbf{x})} \phi'(\mathbf{y}) \prod_{\mathbf{z} \in (\overset{2}{\sim})\text{-Clq}(\mathbf{x})} \phi''(\mathbf{z}),$$

where  $\phi'$  and  $\phi''$  are the interaction functions determined by the densities  $f'$  and  $f''$ . Observe that if we define the function  $\phi$  on  $\mathcal{X}^{\cup*}$  by

$$\phi(\mathbf{x}) = \begin{cases} \phi'(\mathbf{x})\phi''(\mathbf{x}) & \mathbf{x} \text{ is a } (\overset{\cap}{\sim})\text{-clique,} \\ \phi'(\mathbf{x}) & \mathbf{x} \text{ is a } (\overset{1}{\sim})\text{-clique but not a } (\overset{2}{\sim})\text{-clique,} \\ \phi''(\mathbf{x}) & \mathbf{x} \text{ is a } (\overset{2}{\sim})\text{-clique but not a } (\overset{1}{\sim})\text{-clique,} \\ 1 & \text{otherwise (i.e., } \mathbf{x} \text{ is not a } (\overset{\cup}{\sim})\text{-clique),} \end{cases}$$

then  $\phi$  is a  $(\cup)$ -interaction function and the function  $f$  is expressible as in equation (10.4.8). The Hammersley–Clifford theorem now implies that there exists some finite positive  $c$  such that  $cf$  is the density function of a Markov point process on  $\mathcal{X}$ .  $\square$

Proposition 10.4.VI enables us to extend the Strauss model in such a way as to allow degrees of interaction that may depend on the distance between points, as sketched in Exercise 10.4.6. Such a model is still a Gibbs model.

**EXAMPLE 10.4(e) Spatial birth-and-death process** [Preston (1977); see also van Lieshout (2000, pp. 83–87)]. This is a continuous-time space–time Markov process with state space  $\mathcal{X}^{\cup*}$  satisfying the following.

- (a) The only transitions are ‘births’ ( $\mathbf{x} \mapsto \mathbf{x} \cup y$ ) and ‘deaths’ ( $\mathbf{x} \cup y \mapsto \mathbf{x}$ ), where  $\mathbf{x} \in \mathcal{X}^{\cup*}$  and  $y \in \mathcal{X} \setminus \mathbf{x}$ .
- (b) The probability of more than one transition in  $(t, t + h)$  is  $o(h)$ .
- (c) Given the state  $\mathbf{x}$  at  $t$ , the probability of a death  $\mathbf{x} \mapsto \mathbf{x} \setminus y$  ( $y \in \mathbf{x}$ ) during  $(t, t + h)$  equals  $D(\mathbf{x} \setminus y, y)h + o(h)$ , where  $D(\cdot, \cdot): \mathcal{X}^{\cup*} \times \mathcal{X} \mapsto \mathbb{R}_+$  is a  $\mathcal{B}_{\mathcal{X}^{\cup*}} \times \mathcal{B}_{\mathcal{X}}$ -measurable function.
- (d) Given the state  $\mathbf{x}$  at  $t$ , the probability of a birth  $\mathbf{x} \mapsto \mathbf{x} \cup y$  in  $(t, t + h)$ , where  $y \in F \in \mathcal{B}_{\mathcal{X}}$ , equals  $B(\mathbf{x}, F)h + o(h)$ , where  $B(\mathbf{x}, \cdot)$  is a finite measure on  $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ .

Assume that  $B(\mathbf{x}, \cdot)$  has a density  $b(\mathbf{x}, \cdot)$  with respect to the finite measure  $\lambda(\cdot)$  on  $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ , so that intuitively,  $b(\mathbf{x}, y)$  is the transition rate for a birth  $\mathbf{x} \mapsto \mathbf{x} \cup y$ .

Let  $f$  be a Markov function that is the density of a finite point process on  $\mathcal{X}$ . Ripley (1977) observed that if there exists a spatial birth-and-death process such that whenever  $f(\mathbf{x} \cup y) > 0$  it is true that the detailed balance relation

$$b(\mathbf{x}, y)f(\mathbf{x}) = D(\mathbf{x}, y)f(\mathbf{x} \cup y) > 0 \quad (\mathbf{x} \in \mathcal{X}^{\cup*}) \quad (10.4.13)$$

holds, then the birth-and-death process is indecomposable and time-reversible, and its unique equilibrium distribution is the point process with density  $f$ .

Existence and convergence are guaranteed by the following amalgamation of Preston’s (1977) Proposition 5.1 and Theorem 7.1 [see, e.g., Preston or van Lieshout (2000) for proof].

**Proposition 10.4.VII.** *Let  $B(\cdot, \cdot): \mathcal{X}^{\cup*} \times \mathcal{B}_{\mathcal{X}^{\cup*}} \mapsto \mathbb{R}_+$  and  $D(\cdot, \cdot): \mathcal{X}^{\cup*} \times \mathcal{X} \mapsto \mathbb{R}_+$  be such that  $B(\mathbf{x}, \cdot)$  is a finite measure on  $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$  for each  $\mathbf{x} \in \mathcal{X}^{\cup*}$ ,  $B(\cdot, F)$  is  $\mathcal{B}_{\mathcal{X}^{\cup*}}$ -measurable for each  $F \in \mathcal{B}_{\mathcal{X}}$ , and  $D(\cdot, \cdot)$  is  $\mathcal{B}_{\mathcal{X}^{\cup*}} \times \mathcal{B}_{\mathcal{X}}$ -measurable. Define*

$$\beta_n = \sup_{\mathbf{x} \in \mathcal{X}^{(n)*}} B(\mathbf{x}, \mathcal{X}), \quad \delta_n = \inf_{\mathbf{x} \in \mathcal{X}^{(n)*}} \sum_{y \in \mathbf{x}} D(\mathbf{x} \setminus y, y). \quad (10.4.14)$$

Suppose that either

- (a)  $\beta_n = 0$  for all sufficiently large  $n \geq 0$  and  $\delta_n > 0$  for all  $n \geq 1$ ; or

(b)  $\beta_n > 0$  for all  $n > 0$ ,  $\delta_n > 0$  for all  $n \geq 1$ , and

$$\sum_{n=1}^{\infty} \frac{\beta_0 \cdots \beta_{n-1}}{\delta_1 \cdots \delta_n} < \infty, \quad \sum_{n=1}^{\infty} \frac{\delta_1 \cdots \delta_n}{\beta_1 \cdots \beta_n} = \infty. \quad (10.4.15)$$

Then there exists a unique spatial birth-and-death process for which  $B$  and  $D$  are the transition rates (the backwards equations involving  $B$  and  $D$  have a unique solution). The process converges in distribution as  $t \rightarrow \infty$  to its unique equilibrium measure, independent of the initial state.

When either the birth- or death-rate is constant, the equilibrium distribution is an area-interaction process (see Exercise 10.4.8).  $\square$

Powerful as it is, Ripley and Kelly's definition of a Markov point process, Definition 10.4.II, poses problems for a stationary renewal process which, seemingly, should be the simplest nontrivial case of a Markov point process on  $\mathbb{R}$ . To see this, suppose that on the interval  $(0, t)$  there are  $n$  points  $\mathbf{x} = \{x_i: i = 1, \dots, n\}$  with  $0 < x_1 < \cdots < x_n < t$  say, coming from a stationary renewal process whose lifetime d.f.  $F$  has support  $(0, a)$ , density function  $f$ , and finite mean lifetime  $\lambda^{-1} = \int_0^\infty x f(x) dx$  [so,  $\int_0^a f(x) dx = F(a) = 1 > F(a - h)$  for any  $h > 0$ ]. Then it is a standard result (see, e.g., Exercise 7.2.3) that the Janossy density function  $j_n(\mathbf{x})$  is given by

$$j_n(\mathbf{x}) = \lambda[1 - F(x_1)] \left( \prod_{i=1}^{n-1} f(x_{i+1} - x_i) \right) \lambda[1 - F(t - x_n)]. \quad (10.4.16)$$

Consequently, for any  $\mathbf{x}$  such that  $j_n(\mathbf{x}) > 0$  [and then, necessarily,  $\max\{x_1, \max_{2 \leq i \leq n}(x_i - x_{i-1}), t - x_n\} \leq a$ ], we have

$$\frac{j_{n+1}(\mathbf{x} \cup y)}{j_n(\mathbf{x})} = \begin{cases} [1 - F(y)]f(x_1 - y)/[1 - F(x_1)] & \text{if } y < x_1, \\ f(y - x_{i-1})f(x_i - y)/f(x_i - x_{i-1}) & \text{if } x_{i-1} < y < x_i, \\ & i = 2, \dots, n, \\ f(y - x_n)[1 - F(t - y)]/[1 - F(t - x_n)] & \text{if } y > x_n. \end{cases} \quad (10.4.17)$$

It is evident from (10.4.17) that if such a process is to be Markovian in terms of some 'adjacency' relation  $\sim$  and therefore have a Hammersley–Clifford representation as in Theorem 10.4.V, that cliques on which interaction functions are defined can have at most two elements of a set  $\mathbf{x}$ , and that these elements must be nearest neighbours either to the right or left of a given element  $x \in \mathbf{x}$ . Furthermore, even supposing that  $\{x_i, x_{i+1}\}$  is a clique in  $\mathbf{x}$ , adjoining a point  $y \notin \mathbf{x}$  for which  $x_i < y < x_{i+1}$ , would then change the status of  $\{x_i, x_{i+1}\}$  so that it would no longer be a clique. Within the setting of the earlier part of this section, this 'argument' suggests that a stationary renewal process does not generally fit the Ripley–Kelly definition of a Markov point set. However, a word of caution is apposite: a Poisson process in  $\mathbb{R}$  is a renewal process,



and it also satisfies the Ripley–Kelly definition (because of the complete independence property).

We turn our attention therefore to the relation  $y \sim z$  (or its negation) used earlier between elements  $y, z \in \mathbf{x} \subset \mathcal{X}$ . In all the examples noted here and in the literature generally, the relation is defined in fact for any pair  $y, z \in \mathcal{X}$  independent of any set  $\mathbf{x}$  to which either (or both or neither) may belong. The setting of a renewal process suggests we restrict attention to describing the elements of the pair  $\{y, z\}$  as satisfying a reflexive symmetric relation  $y \overset{\sim}{\sim} z$  only when the pair is a subset of  $\mathbf{x}$  and that the relation may depend crucially on  $\mathbf{x}$  in the sense that we may have, for  $y, z \in \mathbf{x}$ ,  $y \overset{\sim}{\sim} z$  but, although for  $w \notin \mathbf{x}$  we necessarily have  $\mathbf{x} \subset (\mathbf{x} \cup \{w\})$ , it need not be the case that  $y \overset{\mathbf{x} \cup w}{\sim} z$ . For example, when  $y, z \in \mathbf{x} \subset \mathbb{R}$  and  $y \overset{\sim}{\sim} z$  means that  $z$  is the nearest right- or left-neighbour of  $y$  from the set  $\mathbf{x}$ , incrementing the set  $\mathbf{x}$  by a point  $w$  lying between  $y$  and  $z$  destroys this nearest right- or left-hand neighbour property.

Suppose then that such a reflexive symmetric relation  $\overset{\sim}{\sim}$  is defined for all two-point subsets of  $\mathbf{x} \in \mathcal{X}^{\cup*}$ , subject to  $\mathbf{x}$  being in an hereditary family  $\mathcal{H}$  (see Definition 10.4.IV). When for  $y, z \in \mathbf{x}$  the property  $y \overset{\sim}{\sim} z$  holds, say that  $y$  and  $z$  are *neighbours within  $\mathbf{x}$* , or  $(\overset{\sim}{\sim})$ -adjacent. For  $\mathbf{y} \subset \mathbf{x}$  the  $(\overset{\sim}{\sim})$ -neighbourhood is

$$b^{\mathbf{x}}(\mathbf{y}) = \{z \in \mathbf{x}: z \overset{\sim}{\sim} y \text{ for some } y \in \mathbf{y}\}. \quad (10.4.18)$$

The set  $\mathbf{y}$  is a  $(\overset{\sim}{\sim})$ -clique if  $y \overset{\sim}{\sim} z$  for all  $y, z \in \mathbf{y}$ , and the  $(\overset{\sim}{\sim})$ -clique indicator function is

$$I^{\mathbf{x}}(\mathbf{y}) = \begin{cases} 1 & \text{if } \mathbf{y} \subset \mathbf{x} \text{ is a } (\overset{\sim}{\sim})\text{-clique,} \\ 0 & \text{otherwise} \end{cases} \quad (10.4.19)$$

[for strict analogy with (10.4.4) the notation  $b^{\mathbf{x}}(\mathbf{y})$  and  $I^{\mathbf{x}}(\mathbf{y})$  would be used here]. Using this notation, we have for example

$$I^{\mathbf{x}}(\{y, z\}) = 1 \quad \text{if and only if } y \overset{\sim}{\sim} z.$$

Observe the status of the sets  $\mathbf{y}$  and  $\mathbf{x}$  in (10.4.18–19):  $\mathbf{y}$  provides the points that are ‘targeted’ for adjacency, and  $\mathbf{x}$  the ‘environment’ within which ‘adjacency’ is defined via the reflexive symmetric relation  $\overset{\sim}{\sim}$ .

Notice that for  $\overset{\sim}{\sim}$  but not  $\sim$ , expanding  $\mathbf{x}$  can destroy the  $(\overset{\sim}{\sim})$ -adjacency of points  $y, z \in \mathbf{x}$  (i.e., there can exist  $u \notin \mathbf{x}$  such that  $y \overset{\sim}{\sim} z$  but  $y \overset{\mathbf{x} \cup u}{\sim} z$ ).

**Definition 10.4.VIII.** A function  $f: \mathcal{H} \mapsto \mathbb{R}_+$  is a  $(\overset{\sim}{\sim})$ -Markov function if for all  $\mathbf{x}$  in the hereditary class  $\mathcal{H}$ , the function  $f$  is hereditary, and for  $y \in \mathcal{X}$  and  $\mathbf{x} \cup y \in \mathcal{H}$  with  $f(\mathbf{x}) > 0$ , the ratio  $\rho(y \mid \mathbf{x})$  depends only on  $y$ ,  $b^{\mathbf{x} \cup y}(y)$ , and the relations  $\overset{\sim}{\sim}$  and  $\overset{\mathbf{x} \cup y}{\sim}$  restricted to the neighbourhood set  $b^{\mathbf{x} \cup y}(y)$ .

Recall that the Hammersley–Clifford Theorem 10.4.V gives a representation of joint densities in terms of simpler ‘interaction’ functions. For this set-dependent adjacency relation  $\overset{\sim}{\sim}$  the analogous function, and the extended theorem, are as follows; its proof is similar to that of Theorem 10.4.V and can be found in Baddeley and Møller (1989).

**Definition 10.4.IX.** Let the function  $\phi: \mathcal{H} \mapsto \mathbb{R}_+$  be hereditary and be such that for  $y \notin \mathbf{x} \in \mathcal{H}$ , if  $\phi(\mathbf{x}) > 0$  and  $\phi(b^{\mathbf{x} \cup y}(y)) > 0$ , then  $\phi(\mathbf{x} \cup y) > 0$ .

A  $(\tilde{\mathbf{x}})$ -interaction function is a function  $\Phi$  defined in terms of such  $\phi$  by a relation of the form

$$\Phi(\mathbf{y} \mid \mathbf{x}) = \phi(\mathbf{y})^{I^{\mathbf{x}}(\mathbf{y})}, \quad (10.4.20)$$

where  $0^0 = 0$ .

**Theorem 10.4.X** (Hammersley–Clifford, extended). Let the relation  $\tilde{\mathbf{x}}$  satisfy the consistency conditions, for finite  $\mathbf{x} \in \mathcal{H}$ ,  $\mathbf{w} \subset \mathbf{z} \in \mathcal{H}$ ,  $y, z \in \mathcal{X}$  but  $y, z \notin \mathbf{z}$ , and  $\mathbf{x} = \mathbf{z} \cup \{y, z\} \in \mathcal{H}$ ,

(C.1)  $I^{\mathbf{z}}(\mathbf{w}) \neq I^{\mathbf{z} \cup y}(\mathbf{w})$  implies  $\mathbf{w} \subset b^{\mathbf{z} \cup y}(y)$ ; and

(C.2) when  $y \tilde{\mathbf{x}} z$ ,  $I^{\mathbf{z} \cup y}(\mathbf{w}) + I^{\mathbf{z} \cup z}(\mathbf{w}) = I^{\mathbf{z}}(\mathbf{w}) + I^{\mathbf{x}}(\mathbf{w})$ .

Then  $f$  is a  $(\tilde{\mathbf{x}})$ -Markov function if and only if

$$f(\mathbf{x}) = \prod_{\mathbf{y} \subseteq \mathbf{x}} \Phi(\mathbf{y} \mid \mathbf{x}) \quad (10.4.21)$$

for all  $\mathbf{x} \in \mathcal{H}$ , where  $\Phi$  is a  $(\tilde{\mathbf{x}})$ -interaction function.

Baddeley, van Lieshout and Møller (1996) considered Poisson cluster processes and showed them to be nearest-neighbour Markov processes as above when the clusters are uniformly bounded, or if the cluster centre process is Markov or nearest-neighbour Markov and the clusters are both uniformly bounded and a.s. non-empty. Thus, the nearest-neighbour Markov property is preserved under random translation but not under random thinning.

Other extensions of  $\sim$  have been suggested: Ord's process in Exercise 10.4.10 gives one, Chin and Baddeley (1999, 2000) looked first at a relation based on components exhibiting pairwise-connectivity and then at interactions *between* components of point configurations, and van Lieshout (2006a) has considered a sequential definition in association with space–time processes.

## Exercises and Complements to Section 10.4

10.4.1 Suppose that a simple finite point process on (a subset of)  $\mathbb{R}^d$  has Janossy density  $\{j_n(\mathbf{x})\}$ , and that its density with respect to an inhomogeneous Poisson process with intensity  $\lambda(x)$  ( $x \in \mathcal{X}$ ) is  $\{f_n(\mathbf{x})\}$ . Show that  $j_n(\mathbf{x}) = e^{-\Lambda(\mathcal{X})} f_n(\mathbf{x}) \prod_{x_i \in \mathbf{x}} \lambda(x_i)$  where  $\Lambda(\mathcal{X}) = \int_{\mathcal{X}} \lambda(u) \ell(du)$ .

10.4.2 Let the finite point process on  $\mathcal{X}$  have Janossy densities  $\{j_n(\cdot)\}$  as in Sections 5.4 and 7.1. Then the Papangelou conditional intensity of Definition 10.4.I is expressible for some finite  $c > 0$

$$\rho(y \mid \mathbf{x}) = \begin{cases} c j_1(y) & (y \in \mathcal{X}, \mathbf{x} = \emptyset), \\ j_{n+1}(\mathbf{x} \cup y) / j_n(\mathbf{x}) & (y \in \mathcal{X} \setminus \mathbf{x}, \mathbf{x} \in \mathcal{X}^{(n)*}, n = 1, 2, \dots). \end{cases}$$

10.4.3 Suppose the reflexive symmetric relation  $\sim$  of a Markov point process is given by  $y \sim z$  if and only if  $|y - z| \leq R$  for  $R = 0$ . Because the only cliques are then singletons, deduce that the point process must be Poisson.

10.4.4 Verify the properties (i)–(vi) of cliques listed after Definition 10.4.III, providing in particular a counterexample to property (v).

10.4.5 Check conditions for Examples 10.4(b) and (c) to be repulsive or attractive.

10.4.6 *Extended Strauss models; multiscale processes.* Suppose given for a finite positive integer  $k$ ,  $0 = R_0 < R_1 < \dots < R_k < R_{k+1} = \infty$ ,  $\beta_j \in (0, \infty)$  and  $\gamma_j \in (0, 1]$  ( $j = 1, \dots, k$ ), and  $\mathcal{X}$  a bounded Borel subset of  $\mathbb{R}^d$ . For  $\mathbf{x} \in \mathcal{X}^{\cup*}$  and  $R \in \mathbb{R}_+$  let  $m(\mathbf{x}, R) = \#\{x_i, x_j \in \mathbf{x} : \|x_i - x_j\| < R\}$  as in Example 10.4(b), and  $\Delta m(\mathbf{x}, R', R'') = m(\mathbf{x}, R'') - m(\mathbf{x}, R')$  for  $0 \leq R' < R'' \leq \infty$ .

- (a) Let  $\alpha_j$  be such that  $f^{(j)}(\mathbf{x}) = \alpha_j \beta_j^{n(\mathbf{x})} \gamma_j^{m(\mathbf{x}, R_j)}$  is a Markov density function for each  $j = 1, \dots, k$ . Use Proposition 10.4.VI to deduce that

$$f(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \prod_{j=1}^k (\gamma_j \dots \gamma_k)^{\Delta m(\mathbf{x}, R_{j-1}, R_j)}$$

is a Markov density function for suitable  $\alpha$  and  $\beta$ .

- (b) More generally,  $f(\mathbf{x}) = \alpha \beta^{n(\mathbf{x})} \prod_{j=1}^k \gamma_j^{\Delta m(\mathbf{x}, R_{j-1}, R_j)}$  is a Markov density function for some finite positive  $\alpha$  and  $\beta$ ; display its Papangelou intensity function. [Hint: Penttinen (1984) or Møller and Waagepetersen (2004, Example 6.2).]

- (c) Define stochastic monotonicity of Markov point sets for such models.

10.4.7 (a) In the setting of Example 10.4(f) verify that a nonnegative function  $f$  defined for all  $\mathbf{x} \in \mathcal{X}^{\cup*}$  by (10.4.13) and integrable as below (10.4.2) is a Markov density function when conditions such as (10.4.15) are satisfied.

- (b) When  $f$  is a Markov function, with  $(\sim)$ -interaction function  $\phi$  say, a representation such as (10.4.7) but with  $\sqrt{\phi}$  holds for  $\sqrt{f} \equiv \{\sqrt{f_n}\}$ . Conclude that when  $\sqrt{f}$  is integrable as at (10.4.2b) it is a Markov function.

10.4.8 Suppose a birth-and-death process as in Example 10.4(e) has constant birth-rate  $B(\mathbf{x}, A) = \ell(\mathcal{X} \cap A)$  for  $A \in \mathcal{B}_{\mathcal{X}}$  and death-rate  $D(\mathbf{x}) = \sum_{y \in \mathbf{x}} D(\mathbf{x} \setminus y, y)$  for a  $\mathcal{B}_{\mathcal{X} \cup*} \times \mathcal{B}_{\mathcal{X}}$ -measurable function  $D(\cdot, \cdot)$  that satisfies condition (a) or (b) of Proposition 10.4.VII. Show that the equilibrium measure is an area-interaction process as in Example 10.4(c).

[Hint: Baddeley and van Lieshout (1995, Section 4) also give a constant death-rate (but variable birth-rate) analogue of this property.]

10.4.9 Let  $\mathcal{X} = \{x_1, y_1, x_2, y_2, z, y_3\}$  and suppose that  $x_1 \sim y_1 \sim x_2 \sim y_2 \sim z \sim y_3 \sim x_1$  but  $u \not\sim v$  for all other pairs  $\{u, v\} \subset \mathcal{X}$ . For any  $\{u, v\} \subset \mathcal{X}$  and  $\mathbf{w} \subset \mathcal{X}$  define  $u \overset{\mathbf{w}}{\sim} v$  if either  $u \sim v$  or else  $u \sim w$  and  $w \sim v$  for some  $w \in \mathbf{w}$ .

- (a) Consider the two sets  $\mathbf{x} = \mathcal{X} \setminus z$  and  $\mathbf{y} = \{y_1, y_2, y_3\}$ , so  $y_2 \not\sim y_3$ . Then  $I^{\mathbf{x}}(\mathbf{y}) = 0$  and  $I^{\mathbf{x} \cup z}(\mathbf{y}) = 1$ , so  $I^{\mathbf{x}}(\mathbf{y}) \neq I^{\mathbf{x} \cup z}(\mathbf{y})$ . But  $y_1 \notin b^{\mathbf{x} \cup z}(z)$ , so condition (C.1) does not hold for such  $\mathcal{X}$  and  $\overset{\mathbf{w}}{\sim}$  as defined.

- (b) Use  $\mathbf{y}$  as above, but now put  $\mathbf{z} = \mathbf{y} \cup z$  and let  $u, v = x_1, x_2$ . Then  $I^{\mathbf{z}}(\mathbf{y}) = I^{\mathbf{z} \cup u}(\mathbf{y}) = I^{\mathbf{z} \cup v}(\mathbf{y}) = 0$  but  $I^{\mathbf{z} \cup \{u, v\}}(\mathbf{y}) = 1$ , so (C.2) also fails.

10.4.10 *Ord's process.* Consider the function  $f(\mathbf{x}) = \alpha \beta^n \prod_{i=1}^n g(\text{area of } C^{\mathbf{x}}(x_i))$ , where  $\mathbf{x} \in$  bounded subregion of  $\mathbb{R}^2$ ,  $n = n(\mathbf{x})$ ,  $C^{\mathbf{x}}(x_i)$  denotes the Voronoi cell associated with  $x_i \in \mathbf{x}$  as in Example 10.4(c), and  $g$  a function described shortly. If  $g$  is not constant then  $f$  is not a Markov function, because  $f(\mathbf{x} \cup y)/f(\mathbf{x})$  depends on neighbours of neighbours of  $y$ , but for positive bounded  $g$ ,  $f$  is a Markov function w.r.t.  $\overset{\sim^2}{\sim}$  defined by  $x_i \overset{\sim^2}{\sim} x_j$  if either  $x_i \overset{\sim}{\sim} x_j$  or else there exists  $x_k \in \mathbf{x}$  such that both  $x_i \overset{\sim}{\sim} x_k$  and  $x_k \overset{\sim}{\sim} x_j$ .

## CHAPTER 11

# Convergence Concepts and Limit Theorems

11.1	Modes of Convergence for Random Measures and Point Processes	132
11.2	Limit Theorems for Superpositions	146
11.3	Thinned Point Processes	155
11.4	Random Translations	166

When random measures and point processes are regarded as probability measures on the appropriate c.s.m.s.  $\mathcal{M}_{\mathcal{X}}^{\#}$  or  $\mathcal{N}_{\mathcal{X}}^{\#}$ , they may be associated with concepts of both weak and strong convergence of measures on a metric space. In this chapter we examine these concepts more closely, finding necessary and sufficient conditions for weak convergence, relating this concept to other possible definitions of convergence, and applying it to some near-classical questions concerning the convergence of superpositions, thinnings, and translations of point processes.

A common theme in the limit theorems described in this chapter is the emergence of the Poisson process as the limit of repeated applications of some stochastic operation on an initial point process. In a loose sense, each of the operations of superposition, thinning, and random translation is entropy increasing; it is not surprising then that among point processes with fixed mean rate, the Poisson process has maximum entropy (see Section 7.6 and the further discussion in Chapter 14). These limit theorems help not only to explain the ubiquitous role of the Poisson process in applications but also to reveal its central place in the structural theory of point processes.

Of course these applications far from exhaust the role of convergence concepts in the general theory of random measures and point processes. Other important applications arise in the discussion of ergodic theorems and convergence to equilibrium in Chapter 12, and in various questions related to Palm theory in Chapter 13 and conditional intensities in Chapter 14. In this chapter we mostly restrict attention to  $\mathcal{X} = \mathbb{R}^d$ , even though extensions to more general locally compact groups are usually possible. Many of these extensions are covered in MKM (1978) and especially MKM (1982), giving systematic extensions of earlier work to the context of a general locally compact group.

## 11.1. Modes of Convergence for Random Measures and Point Processes

In this section we examine different possible modes of convergence for a family of point processes or random measures. We generally suppose that the processes involved are to be thought of as distributions on  $\mathcal{M}_{\mathcal{X}}^{\#}$  or  $\mathcal{N}_{\mathcal{X}}^{\#}$ , where  $\mathcal{X}$  as usual is a general c.s.m.s. Then the question is this: given a sequence of probability measures  $\{\mathcal{P}_n\}$  on  $\mathcal{M}_{\mathcal{X}}^{\#}$ , in what sense should the statement  $\mathcal{P}_n \rightarrow \mathcal{P}$  be understood?

Three types of convergence suggest themselves for immediate consideration: strong convergence of probability distributions on  $\mathcal{M}_{\mathcal{X}}^{\#}$  (i.e.,  $\|\mathcal{P}_n - \mathcal{P}\| \rightarrow 0$ , where  $\|\mathcal{P}\|$  is the variation norm as defined at the end of Section A1.3); weak convergence of probability distributions on  $\mathcal{M}_{\mathcal{X}}^{\#}$  [i.e.,  $\mathcal{P}_n \rightarrow \mathcal{P}$  weakly, meaning  $\mathcal{P}_n \rightarrow \mathcal{P}$  in the sense of weak convergence of measures on the metric space  $\mathcal{M}_{\mathcal{X}}^{\#}$ ; see Definition A2.3.I(i)]; and convergence of the finite distributions [i.e., for all suitable finite families of bounded Borel sets  $A_1, \dots, A_k$ , the joint distributions of the random variables  $\xi(A_1), \dots, \xi(A_k)$  under  $\mathcal{P}_n$  converge weakly to their limit distribution under  $\mathcal{P}$ ]. The consequential matter of convergence of moments is noted before considering in the final part of this section some further questions that arise when we try to relate convergence of measures in  $\mathcal{M}_{\mathcal{X}}^{\#}$  and convergence of the associated cumulative processes (distribution functions) in the function space  $\mathcal{D}(0, \infty)$  or its relatives [see discussion following Example 11.1(c)].

We sometimes adopt a common abuse of terminology by stating that the random measures  $\xi_n$  converge weakly (or strongly) to a limit random variable  $\xi$  when all that is meant is the weak (or strong) convergence of their distributions in  $\mathcal{M}_{\mathcal{X}}^{\#}$ ; in fact there are no requirements for convergence of the random measures themselves (as elements of or mappings into  $\mathcal{M}_{\mathcal{X}}^{\#}$ ). The same abuse applies to “point processes  $N_n$  converge to  $N$ .”

We note first that *strong convergence implies weak convergence*. Indeed, for any set  $U \in \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ , we have in the notation of Section A1.3

$$\|\mathcal{P}_n - \mathcal{P}\| \geq V_{\mathcal{P}_n - \mathcal{P}}(U) \geq |\mathcal{P}_n(U) - \mathcal{P}(U)|.$$

It follows that strong convergence implies  $\mathcal{P}_n(U) \rightarrow \mathcal{P}(U)$  for all  $U \in \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ , which then implies weak convergence by Theorem A2.3.II. The converse is not true; Example 11.1(a) below serves as a counterexample. Indeed, strong convergence implies that any fixed atom for the limit probability must also be a fixed atom for its approximants.

One of the most important applications of convergence in variation norm concerns convergence to equilibrium, or stability properties, of stochastic processes, in which context it is frequently established through the concept of *coupling*. Two jointly defined stochastic processes  $X(t)$  and  $Y(t)$  are said to *couple* if there exists an a.s. finite random variable  $T$  (the *coupling time*) such that  $X(t)$  and  $Y(t)$  are a.s. equal for all  $t \geq T$ . The basic lemma is as below [see, e.g., Lindvall (1992) or Thorisson (2000) for more extended discussion].

**Lemma 11.1.I.** *Let  $X(t, \omega)$ ,  $Y(t, \omega)$ , be two stochastic processes, defined on a common probability space  $(\Omega, \mathcal{E}, \mathcal{P})$ , and taking their values in a common c.s.m.s.  $V$ . Denote by  $\mathcal{P}_t$ ,  $\mathcal{Q}_t$  the distributions of  $X(t)$ ,  $Y(t)$ , respectively, on  $(V, \mathcal{B}_V)$ . Suppose that  $X$  and  $Y$  couple, with coupling time  $T$ . Then*

$$\|\mathcal{P}_t - \mathcal{Q}_t\| \leq 2\mathcal{P}\{T > t\}.$$

PROOF. To establish convergence in variation norm, we first note that

$$\|\mathcal{P}_t - \mathcal{Q}_t\| = \sup_{f: \|f\| \leq 1} \left| \int f(v) \mathcal{P}_t(dv) - \int f(v) \mathcal{Q}_t(dv) \right|,$$

the supremum being taken over all bounded measurable functions  $f$  on  $(V, \mathcal{B}_V)$  with  $\|f\| = \sup_{v \in V} |f(v)| \leq 1$ . (Indeed, the supremum is achieved when  $f = I_{U^+} - I_{U^-}$  in the Jordan–Hahn decomposition of  $\mathcal{P}_t - \mathcal{Q}_t$ : see Theorem A1.3.IV.) We then have

$$\begin{aligned} \left| \int f(v) \mathcal{P}_t(dv) - \int f(v) \mathcal{Q}_t(dv) \right| &\leq \int_{\Omega} |f(X(t, \omega)) - f(Y(t, \omega))| \mathcal{P}(d\omega) \\ &= \int_{T \leq t} |f(X(t, \omega)) - f(Y(t, \omega))| \mathcal{P}(d\omega) \\ &\quad + \int_{T > t} |f(X(t, \omega)) - f(Y(t, \omega))| \mathcal{P}(d\omega) \\ &\leq 2\|f\| \mathcal{P}\{T > t\} \leq 2\mathcal{P}\{T > t\}. \end{aligned} \quad \square$$

This lemma can be applied to point processes by associating  $X(t) \in \mathcal{N}_{\mathcal{X}}^{\#}$  with the shifted version  $S_t N$  of a point process initially defined on  $\mathbb{R}_+$ : see the further discussion on convergence to equilibrium in Section 12.5 where the weaker concept of shift-coupling is also discussed (see around Lemma 12.5.IV).

Although convergence in variation norm is generally the more difficult to establish, once available it is very convenient to use. This is because in addition to its properties as a norm, it also respects convolution in the sense that

$$\|\mu * \nu\| \leq \|\mu\| \|\nu\|$$

(see Exercise 11.1.1). In practice, it is often convenient to work not with the norm on the full space  $\mathcal{M}_{\mathcal{X}}^{\#}$ , but rather with the family of norms on each of the spaces of totally finite measures  $\mathcal{M}_A$  for bounded  $A \in \mathcal{B}_{\mathcal{X}}$ . We have already met examples of this approach in the discussion of convergence to equilibrium of renewal and Wold processes (see in particular Corollary 4.4.VI). Yet another possibility is to look at norm convergence rather than weak convergence for the fidi distributions, an issue that arises in applying the Stein–Chen approach to establishing convergence to Poisson distributions.

Although the main emphasis in our discussions is on weak convergence, an illustration of the sort of analysis required is given in the second half of

Section 11.3, where some of the preliminary inequalities are derived and used to strengthen the convergence statements in the discussion of thinning. Other examples are given from time to time in the exercises and elsewhere.

We turn now to the main topic of this section, namely the weak convergence of random measures and point processes, and its relation to the weak convergence of the finite-dimensional distributions.

In connection with the latter concept, we call the Borel set  $A$  a *stochastic continuity set* for the measure  $\mathcal{P}$  if  $\mathcal{P}\{\xi(\partial A) > 0\} = 0$ , equivalently  $\mathcal{P}\{\xi(\partial A) = 0\} = 1$ . Without a restriction to sets that are continuity sets for the limit measure, convergence of the fidi distributions would be too strong a concept to be generally useful as the following example shows.

**EXAMPLE 11.1(a) Convergence and continuity sets.** Let  $\xi_n$  consist of exactly one point in each interval  $(k, k+1)$ ,  $k = 0, \pm 1, \pm 2, \dots$ , with each such point uniformly distributed over  $(k, k+1/n)$ . Then as  $n \rightarrow \infty$ , we would like to say that the sequence converges to the deterministic point process with one point at each integer. However, if  $A = (0, 1)$ , we have  $\mathcal{P}_n\{\xi(0, 1) > 0\} = 1$  but  $\mathcal{P}\{\xi(0, 1) > 0\} = 0$ . Thus, we can expect difficulties to arise in the definition if the limit random measure has fixed atoms, and these atoms lie on the boundary of the set  $A$  considered in the finite-dimensional distribution.

Similar but more general examples can readily be constructed.  $\square$

Granted the need for the restriction it is important to know that there are ‘sufficiently many’ stochastic continuity sets. Given  $\mathcal{P}$ , let  $\mathcal{S}_{\mathcal{P}}$  denote the class of stochastic continuity sets for  $\mathcal{P}$ . From the elementary properties of set boundaries (see Proposition A1.2.I), it is clear that  $\mathcal{S}_{\mathcal{P}}$  is an algebra (see Exercise 11.1.2). The following lemma is then sufficient for most purposes.

**Lemma 11.1.II.** *Let  $\mathcal{X}$  be a c.s.m.s.,  $\mathcal{P}$  a probability measure on  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ , and  $\mathcal{S}_{\mathcal{P}}$  the class of stochastic continuity sets for  $\mathcal{P}$ . Then for all  $x$  and, given  $x$ , for all but a countable set of values of  $r > 0$ ,  $S_r(x) \in \mathcal{S}_{\mathcal{P}}$ .*

**PROOF.** It is enough to show that for each finite positive  $\varepsilon$ ,  $\delta$ , and  $R$ , the set of  $r$  in  $[0, R]$  satisfying

$$\mathcal{P}\{\xi(\partial S_r(x)) > \delta\} > \varepsilon$$

is finite. Suppose the contrary, and let  $\varepsilon$ ,  $\delta$ , and  $R$  be such that for some countably infinite set  $\{r_1, r_2, \dots\}$  of distinct values of  $r$  in  $0 \leq r \leq R < \infty$ ,  $\mathcal{P}(B_i) > \varepsilon$  for  $i = 1, 2, \dots$ , where  $B_i = \{\xi: \xi(\partial S_{r_i}(x)) > \delta\}$ . Then

$$\varepsilon \leq \limsup_{i \rightarrow \infty} \mathcal{P}(B_i) \leq \mathcal{P}(\limsup_{i \rightarrow \infty} B_i) \leq \mathcal{P}\{\xi(\overline{S_R(x)}) = \infty\}$$

because

$$\xi(\overline{S_R(x)}) \geq \sum_{i=1}^{\infty} \xi(\partial S_{r_i}(x)) = \infty$$

whenever  $\xi(\partial S_{r_i}(x)) \geq \delta > 0$  for an infinite number of values of  $i$ . This contradicts the bounded finiteness of  $\xi$ .  $\square$

**Corollary 11.1.III.** *The stochastic continuity sets of  $\mathcal{P}$  form an algebra that contains both a dissecting ring and a covering ring.*

PROOF. Inspection shows that the constructions of a dissecting system at Proposition A2.1.IV and of a covering ring before Corollary A2.3.III are not affected by replacing any sphere  $S_\alpha(d_i)$  that is not a stochastic continuity set by a marginally smaller sphere  $S_{\alpha'}(d_i)$  that is. Because the remaining stages of the constructions involve only finite unions, intersections, and differences, they do not lead out of the algebra of such continuity sets.  $\square$

We can now state the following more formal definition.

**Definition 11.1.IV.** *The sequence  $\{\xi_n\}$  converges in the sense of convergence of fidi distributions if for every finite family  $\{A_1, \dots, A_k\}$  of bounded continuity sets  $A_i \in \mathcal{B}_\mathcal{X}$  the joint distributions of  $\{\xi_n(A_1), \dots, \xi_n(A_k)\}$  converge weakly in  $\mathcal{B}(\mathbb{R}^k)$  to the joint distribution of  $\xi(A_1), \dots, \xi(A_k)$ .*

The mapping that takes a general element  $\xi$  of  $\mathcal{M}_\mathcal{X}^\#$  into  $\xi(A)$ , where  $A$  is a bounded Borel set, is measurable, essentially by definition of the  $\sigma$ -algebra  $\mathcal{B}(\mathcal{M}_\mathcal{X}^\#)$ , but need not be continuous. To see this last point it is enough to consider variants on Example 11.1(a), where the sequence of measures  $\{\xi_n\}$  converges in the  $w^\#$ -topology in  $\mathcal{M}_\mathcal{X}^\#$  to a limit measure  $\xi$  that has an atom on the boundary  $\partial A$ . However, only those measures  $\xi$  giving nonzero mass to  $\partial A$  can act in this way as discontinuity points of the mapping  $\xi \mapsto \xi(A)$ , for if  $\xi(\partial A) = 0$  and  $\xi_n \rightarrow_{w^\#} \xi$ , then by Theorem A2.3.II (see also Proposition A2.6.II),  $\xi_n(A) \rightarrow \xi(A)$ .

Now let  $\{\mathcal{P}_n\}$  be a sequence of probability distributions, and suppose that  $\mathcal{P}_n \rightarrow \mathcal{P}$  weakly, and that  $A$  is a stochastic continuity set for  $\mathcal{P}$ . This last is just another way of saying that the set  $D$  of discontinuity points for the mapping  $f_A: \xi \mapsto \xi(A)$  satisfies the condition  $\mathcal{P}(D) = 0$ . It then follows from the extended form of the continuous mapping theorem (Proposition A2.3.V) that  $\mathcal{P}_n(f_A^{-1}) \rightarrow \mathcal{P}(f_A^{-1})$ , or in other words that the distribution of  $\xi(A)$  under  $\mathcal{P}_n$  converges to its distribution under  $\mathcal{P}$ . A similar argument applies to any finite family of bounded Borel sets  $\{A_1, \dots, A_k\}$  satisfying  $\mathcal{P}\{\xi(\partial A) = 0\} = 1$  and hence leads to the following lemma.

**Lemma 11.1.V.** *Weak convergence implies weak convergence of the finite-dimensional distributions.*

What is more surprising is that the converse of this statement is also true, so that *for random measures and point processes, the concepts of weak convergence and convergence of fidi distributions are equivalent*. This result, which constitutes the main theorem of this section, is proved at Theorem 11.1.VII.

In preparation for this result, we set out in explicit form the conditions for a family of probability measures on  $\mathcal{B}(\mathcal{M}_\mathcal{X}^\#)$  to be uniformly tight (cf. Appendix A2.4, in particular Theorem A2.4.I). In the proposition below  $T$  refers to an arbitrary index set, not necessarily countable.



**Proposition 11.1.VI.** *For a family of probability measures  $\{\mathcal{P}_t, t \in \mathcal{T}\}$  on  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$  to be uniformly tight, it is necessary and sufficient that, given any closed sphere  $\bar{S} \subset \mathcal{X}$  and any  $\varepsilon, \delta > 0$ , there exists a real number  $M < \infty$  and a compact set  $C \subseteq \bar{S}$  such that, uniformly for  $t \in \mathcal{T}$ ,*

$$\mathcal{P}_t\{\xi(\bar{S}) > M\} < \varepsilon, \quad (11.1.1)$$

$$\mathcal{P}_t\{\xi(\bar{S} \setminus C) > \delta\} < \varepsilon. \quad (11.1.2)$$

*If  $\mathcal{X}$  is locally compact, and in particular if  $\mathcal{X} = \mathbb{R}^d$ , the second condition is redundant.*

PROOF. Uniform tightness means that, for each  $\varepsilon > 0$ , there exists a compact set  $K \in \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$  such that  $\mathcal{P}_t(K) > 1 - \varepsilon$  for all  $t \in \mathcal{T}$ . From Proposition A2.6.IV and Theorem A2.4.I,  $K$  is compact if there exists a sequence of closed spheres  $\bar{S}_n \uparrow \mathcal{X}$  such that for each  $\delta > 0$  and  $n < \infty$  there exist constants  $M_n$  and compact sets  $C_{n,\delta} \subseteq \bar{S}_n$  such that for all  $\xi \in K$ ,

$$(a) \quad \xi(\bar{S}_n) \leq M_n, \text{ and}$$

$$(b) \quad \xi(\bar{S}_n \setminus C_{n,\delta}) \leq \delta.$$

Effectively, (11.1.1) and (11.1.2) are just reformulations of (a) and (b). Indeed, supposing first that (11.1.1) and (11.1.2) are satisfied, choose any sequence of closed spheres  $\bar{S}_n \uparrow \mathcal{X}$  such that each  $\bar{S}_n$  is a stochastic continuity set for  $\mathcal{P}$ . From (11.1.1) we choose  $M'_n$  such that

$$\mathcal{P}_t\{\xi(\bar{S}_n) > M'_n\} < \varepsilon/2^{n+1},$$

and from (11.1.2) we choose the compact set  $C'_{mn} \subseteq \bar{S}_n$  so that

$$\mathcal{P}_t\{\xi(\bar{S}_n \setminus C'_{mn})\} < \varepsilon/2^{m+n+2}.$$

Define the sets, for  $n, m = 1, 2, \dots$ ,

$$Q_n = \{\xi: \xi(\bar{S}_n) \leq M_n\}, \quad Q_{mn} = \{\xi: \xi(\bar{S}_n \setminus C'_{mn}) \leq m^{-1}\},$$

$$K = \bigcap_{n=1}^{\infty} \bigcap_{m=1}^{\infty} (Q_n \cap Q_{mn}).$$

By construction, (a) and (b) are satisfied so  $K$  is compact, and

$$\mathcal{P}_t(K^c) \leq \sum_{n=1}^{\infty} \left[ \mathcal{P}_t(Q_n^c) + \sum_{m=1}^{\infty} \mathcal{P}_t(Q_{mn}^c) \right] \leq \sum_{n=1}^{\infty} \left[ \frac{\varepsilon}{2^{n+1}} + \sum_{m=1}^{\infty} \frac{\varepsilon}{2^{m+n+2}} \right] = \varepsilon.$$

Thus,  $K$  satisfies all the required conditions.

Suppose conversely that the measures  $\mathcal{P}_t$ , are uniformly tight. Given  $\varepsilon$ , choose compact  $K \subset \mathcal{M}_{\mathcal{X}}^{\#}$  and hence deduce the existence of spheres  $\bar{S}_n \uparrow \mathcal{X}$  such that there exist constants  $M_n$  for which (a) holds, and, given  $\delta$ , there exist  $C_{n,\delta}$  such that (b) holds. Given any  $\bar{S}$ , choose  $n$  so that  $\bar{S} \subseteq \bar{S}_n$ , set

$M = M_n$  so that (11.1.1) is true, and, given  $\delta$ , set  $C = C_{n,\delta} \cap \bar{S}$  so that  $\xi(\bar{S} \setminus C) \leq \xi(\bar{S} \setminus C_{n,\delta})$  and hence (11.1.2) holds.  $\square$

EXAMPLE 11.1(b) *Convergence of one-point processes.* Let  $\xi_n$  be the degenerate point process in  $\mathbb{R}$  in which all the mass is concentrated on the counting measure with a single atom at the point  $n$ . Then (11.1.1) holds trivially for all  $S$  and  $\varepsilon$  with  $M = 2$ . In fact  $\varepsilon_n \rightarrow \varepsilon_\infty$  weakly, where  $\varepsilon_\infty$  has all its mass concentrated on the zero random measure 0. Thus, it is important to bear in mind that weak convergence here does not preclude the possibility that the limit point process may be everywhere zero.

Next let  $\eta_n = \sum_{k=1}^n \xi_k$ . Then for all  $n$  we have

$$\eta_n(0, m] \leq m$$

so that (11.1.1) still holds with  $M$  equal to the radius of the sphere  $S$ . In this case  $\eta_n \rightarrow \eta_\infty$  weakly, where  $\eta_\infty$  is the *deterministic point process at unit rate* with an atom at each positive integer.

Finally, let  $\zeta_n = \sum_{k=1}^n k\zeta_{n-k}$ . Here condition (11.1.1) fails and no weak limit exists.  $\square$

The next theorem is the main result of this section. It is a striking consequence of the integer-valued character and locally bounded nature of a point process.

**Theorem 11.1.VII.** *Let  $\mathcal{X}$  be a c.s.m.s. and  $\mathcal{P}$ ,  $\{\mathcal{P}_n: n = 1, 2, \dots\}$  distributions on  $(\mathcal{M}_{\mathcal{X}}^\#, \mathcal{B}(\mathcal{M}_{\mathcal{X}}^\#))$ . Then  $\mathcal{P}_n \rightarrow \mathcal{P}$  weakly if and only if the fidi distributions of  $\mathcal{P}_n$  converge weakly to those of  $\mathcal{P}$ .*

PROOF. The first part of the theorem has already been proved in Lemma 11.1.V. Because the set of all fidi distributions determines a probability measure uniquely, in order to prove the converse it suffices to show that the family  $\{\mathcal{P}_n\}$  is uniformly tight, for then every sequence contains a weakly convergent subsequence, and from the convergence of the fidi distributions this must be the limit measure  $\mathcal{P}$ ; thus, all convergent subsequences have the same limit, and so the whole sequence converges.

To establish tightness, we use the assumption that the fidi distributions converge for stochastic continuity sets of  $\mathcal{P}$  to show that (11.1.1) and (11.1.2) hold for any given  $\bar{S}$ ,  $\varepsilon$ , and  $\delta$ . We start by choosing  $\bar{S}' \supseteq \bar{S}$  to be a stochastic continuity set not only for  $\mathcal{P}$  but also for each of the  $\mathcal{P}_n$ ,  $n = 1, 2, \dots$ . (Because by Lemma 11.1.II only countable sets of exceptional radii are involved, this choice can always be made.) Furthermore, we can choose values  $M$  that are continuity points for the distribution of  $\xi(\bar{S}')$  under  $\mathcal{P}$  and for which  $\mathcal{P}\{\xi(\bar{S}') > M\} < \frac{1}{2}\varepsilon$  and  $\mathcal{P}_n\{\xi(\bar{S}') > M\} \rightarrow \mathcal{P}\{\xi(\bar{S}') > M\}$  as  $n \rightarrow \infty$ . Thus, for  $n > n_0$  say, we have

$$\mathcal{P}_n\{\xi(\bar{S}') > M\} < \varepsilon,$$

and by increasing  $M$  if necessary we can ensure that this inequality holds for all  $n$ . This establishes (11.1.1).

Again working only with spheres that are stochastic continuity sets for  $\mathcal{P}$ , choose spheres  $S_{r_j}(x_i) \equiv S_{ij}$  centred on the points  $x_i$  ( $i = 1, 2, \dots$ ) of a separability set and with radii  $r_j \leq 2^{-j}$ . Define  $C_{ij} = \bar{S}_{ij} \cap \bar{S}'$ . Because

$$\xi\left(\bigcup_{i=1}^K C_{ij}\right) \uparrow \xi(\bar{S}') \quad (K \rightarrow \infty),$$

we can choose  $K_j$  so that, with  $C_j = \bigcup_{i=1}^{K_j} C_{ij}$ ,

$$\mathcal{P}\{\xi(\bar{S}' - C_j) \geq \delta_j\} \leq \varepsilon/2^{j+1},$$

where  $\delta_j \leq \delta/2^j$  is chosen to be a continuity point of the distribution of  $\xi(\bar{S}' - C_j)$  under  $\mathcal{P}$ . Again using the weak convergence of the fidi distributions, and increasing the value of  $K_j$  if necessary, we can ensure as before that the similar inequality

$$\mathcal{P}_n\{\xi(\bar{S}' - C_j) \geq \delta_j\} \leq \varepsilon/2^j \quad (11.1.3)$$

holds for all  $n$ . Now define  $C = \bigcap_{j=1}^{\infty} C_j$ . Then  $C$  is closed, and by construction it can be covered by a finite number of  $\varepsilon$ -spheres for every  $\varepsilon > 0$ , so by Proposition A2.2.II,  $C$  is compact. We have moreover from (11.1.3) that, for every  $n$ ,

$$\begin{aligned} \mathcal{P}_n\{\xi(\bar{S}') - \xi(C) > \delta\} &= \mathcal{P}_n\left\{\xi\left(\bigcup_{j=1}^{\infty} (\bar{S}' - C_j)\right) > \delta\right\} \\ &\leq \sum_{j=1}^{\infty} \mathcal{P}_n\{\xi(\bar{S}' - C_j) > \delta/2^j\} \\ &\leq \sum_{j=1}^{\infty} \mathcal{P}_n\{\xi(\bar{S}' - C_j) > \delta_j\} \leq \sum_{j=1}^{\infty} \frac{\varepsilon}{2^j} = \varepsilon, \end{aligned}$$

thereby establishing (11.1.2). Thus, both conditions of Proposition 11.1.VI are satisfied, and we conclude that the family  $\{\mathcal{P}_n\}$  is tight.  $\square$

Several equivalent conditions for weak convergence can be derived as corollaries or minor extensions to the above theorem. The last condition represents a minor weakening of the full strength of convergence of fidi distributions.

**Proposition 11.1.VIII.** *Each of the following conditions is equivalent to the weak convergence  $\mathcal{P}_n \rightarrow \mathcal{P}$  weakly, where in (i) and (ii),  $f$  ranges over the space of continuous functions vanishing outside a bounded set.*

- (i) *The distribution of  $\int_{\mathcal{X}} f \, d\xi$  under  $\mathcal{P}_n$  converges weakly to its distribution under  $\mathcal{P}$ .*
- (ii) *The Laplace functionals  $L_n[f] \equiv E_{\mathcal{P}_n}[\exp(-\int_{\mathcal{X}} f(x) \xi(dx))]$  converge pointwise to the limit functional  $L[f]$ .*
- (iii) *For point processes, the p.g.f.s  $G_n[h]$  converge to  $G[h]$  for each continuous  $h \in \mathcal{V}_0$ .*

- (iv) For every finite family  $\{A_1, \dots, A_k\}$  from a covering semiring of bounded continuity sets for the limit random measure  $\xi$ , the joint distributions of  $\{\xi_n(A_1), \dots, \xi_n(A_k)\}$  converge weakly in  $\mathcal{B}(\mathbb{R}^k)$  to the joint distribution of  $\{\xi(A_1), \dots, \xi(A_k)\}$ .

PROOF. For any  $f$  as described, the mapping defined by

$$\Phi_f(\xi) = \int_{\mathcal{X}} f(x) \xi(dx);$$

is continuous at  $\xi$  provided  $\xi(Z(f)) = 0$ , where  $Z(f)$  is the set of discontinuities of  $f$ . Hence, in particular,  $\Phi_f$  is continuous for all  $\xi$  whenever  $f$  itself is continuous. Thus, the distributions of  $\Phi_f(\xi)$  under  $\mathcal{P}_n$  converge weakly to its distribution under  $\mathcal{P}$ .

Now suppose that  $f$  is a function of the form  $\sum_i c_i I_{A_i}(x)$ , where  $\sum_i |c_i| < \infty$  and  $\{A_i\}$  is a bounded family of bounded Borel sets that are stochastic continuity sets for  $\mathcal{P}$ . Convergence of the distributions of the integrals  $\int_{\mathcal{X}} f d\xi$  for all such functions  $f$  is equivalent to the joint convergence in distribution of  $\xi(A_1), \dots, \xi(A_k)$  for every finite integer  $k$ , that is, to fidi convergence. Because such functions can be approximated by continuous functions, as, for example, in the proof of Theorem A2.3.II, it follows that (i) implies convergence of the fidi distributions and hence weak convergence.

Condition (ii) is equivalent to (i) by well-known results on Laplace transforms.

Because  $f(x) = -\log h(x)$  is a function as in (i) if and only if  $h$  is continuous and  $h \in \mathcal{V}_0$ , (iii) is equivalent to (ii) when the distributions  $\mathcal{P}_n$  correspond to point processes.

Establishing the sufficiency of the last condition is a matter of verifying that the constructions in the proof of Theorem 11.1.VII can be carried through with sets  $A_i$  drawn from the covering semiring (or, more generally, from the ring it generates, in as much as it is clear that the convergence carries over to sets taken from this generated ring). Because by definition of a covering ring each open sphere can be approximated by sets in the ring, both constructions in the first part of the proof can be so modified, implying that the sequence  $\{\mathcal{P}_n\}$  is uniformly tight. If  $\{\mathcal{P}_{n_k}\}$  is any weakly convergent subsequence from  $\{\mathcal{P}_n\}$ , with limit  $\mathcal{P}'$  say, then  $\mathcal{P}$  and  $\mathcal{P}'$  must have the same fidi distributions for sets drawn from the covering semiring. Then from Proposition 9.2.III it follows that  $\mathcal{P} = \mathcal{P}'$ , and hence as before that  $\mathcal{P}_n \rightarrow \mathcal{P}$  weakly.  $\square$

In the case of point processes, even sharper versions of condition (iv) are possible when the processes are simple, through the use of the avoidance function (see Theorem 9.2.XII). In this case the limit measure must correspond to a simple point process if it is to be uniquely characterized by the avoidance function, and condition (ii) below is such an additional requirement about asymptotic orderliness. Several variants are now possible; the following is perhaps the simplest.

**Proposition 11.1.IX.** Let  $\{\mathcal{P}_n: n = 1, 2, \dots\}$ ,  $\mathcal{P}$  be distributions on  $\mathcal{N}_\chi^\#$  with  $\mathcal{P}$  corresponding to a simple point process, and suppose that  $\mathcal{R} \subseteq \mathcal{S}_\mathcal{P}$  is a covering dissecting ring. In order that  $\mathcal{P}_n \rightarrow \mathcal{P}$  weakly, it is sufficient that

- (i)  $\mathcal{P}_n\{N(A) = 0\} \rightarrow \mathcal{P}\{N(A) = 0\}$  as  $n \rightarrow \infty$  for all bounded  $A \in \mathcal{R}$ ; and
- (ii) for all bounded  $A \in \mathcal{R}$  and partitions  $\mathcal{T}_r = \{A_{ri}: i = 1, \dots, k_r\}$  of  $A$  by sets of  $\mathcal{R}$ ,

$$\limsup_{n \rightarrow \infty} \sup_{\mathcal{T}_r} \sum_{i=1}^{k_r} \mathcal{P}_n\{N(A_{ri}) \geq 2\} = 0. \quad (11.1.4)$$

PROOF. In view of Theorem 9.2.XII it is enough to show that under the stated conditions the family  $\{\mathcal{P}_n\}$  is uniformly tight and that the limit of any weakly convergent subsequence must be a simple point process.

Let  $\bar{S}$  be a closed sphere in  $\mathcal{R}$ , and in (ii) take  $A = \bar{S}$ . Observing that  $\{N(\bar{S}) > k_r\}$  implies  $\{N(A_{ri}) \geq 2 \text{ for at least one } i\}$ ,

$$\sum_{i=1}^{k_r} \mathcal{P}_n\{N(A_{ri}) \geq 2\} \geq \mathcal{P}_n\{N(\bar{S}) > k_r\}.$$

Given  $\varepsilon > 0$ , (11.1.4) implies that the sum on the left-hand side here is bounded by  $\varepsilon$  for  $n \geq n_0$ , hence (by adjusting  $k_r$  if necessary) for all  $n$ , and thus that the first condition for uniform tightness holds.

Condition (11.1.2) here can be stated in the following form. Given  $\varepsilon > 0$ , there exists a compact set  $C$  such that  $\mathcal{P}_n\{N(\bar{S} - C) = 0\} > 1 - \varepsilon$  for  $n = 1, 2, \dots$ . Choose  $C$  so that for the limit distribution we have

$$\mathcal{P}\{N(\bar{S} - C) = 0\} > 1 - \frac{1}{2}\varepsilon.$$

From assumption (i) we have  $\mathcal{P}_n\{N(\bar{S} - C) = 0\} \rightarrow \mathcal{P}\{N(\bar{S} - C) = 0\}$  as  $n \rightarrow \infty$ , from which the required inequality (11.1.2) holds for all sufficiently large  $n$ , and hence (by increasing  $C$  if necessary) for all  $n$ .

Thus, both conditions for uniform tightness of  $\{\mathcal{P}_n\}$  are satisfied. Now let  $\{\mathcal{P}_{n_k}\}$  be any weakly convergent subsequence from the family  $\{\mathcal{P}_n\}$ , with limit  $\mathcal{P}'$  say, so that from Theorem 11.1.VII all fidi distributions converge to those of  $\mathcal{P}'$ ; hence, from (i) of the proposition, for  $A \in \mathcal{R}$ ,

$$\mathcal{P}\{N(A) = 0\} = \mathcal{P}'\{N(A) = 0\}.$$

From this result it follows not necessarily that  $\mathcal{P} = \mathcal{P}'$  but merely that

$$\mathcal{P}\{N(A) = 0\} = \mathcal{P}'\{N^*(A) = 0\} \quad \text{or} \quad \mathcal{P} = (\mathcal{P}')^*,$$

where  $N^*$  is the support point process of  $N$  and  $(\mathcal{P}')^*$  is its distribution (see Corollary 9.2.XIII). However, we have

$$\sum_{i=1}^{k_r} \mathcal{P}'\{N(A_{ri}) \geq 2\} = \sum_{i=1}^{k_i} \lim_{k \rightarrow \infty} \mathcal{P}_{n_k}\{N(A_{ri}) \geq 2\}$$

so that from (11.1.4)

$$\limsup_{\mathcal{T}_r} \sum_{i=1}^{k_r} \mathcal{P}'\{N(A_{ri}) \geq 2\} \leq \lim_{k \rightarrow \infty} \sup_{\mathcal{T}_r} \sum_{i=1}^{k_r} \mathcal{P}_{n_k}\{N(A_{ri}) \geq 2\} = 0,$$

from which it follows (Proposition 9.3.XII) that  $\mathcal{P}'$  is simple, and hence that  $\mathcal{P}' = (\mathcal{P}')^* = \mathcal{P}$ . So all weakly convergent subsequences have limit  $\mathcal{P}'$ , and thus  $\mathcal{P}_n \rightarrow \mathcal{P}$  weakly.  $\square$

Thus far we have considered essentially convergence of probability measures of point processes, but what of their moments? Uniform integrability is an analogous ‘bounding’ condition that ensures that a sequence of moment measures may converge, as in the assertion below (the proof is standard and left to the reader; see also Exercise 11.1.3).

**Proposition 11.1.X.** *Let  $\{N_n\}$  be a weakly convergent sequence of point processes on the c.s.m.s.  $\mathcal{X}$  with limit  $N$  and for which the first moment measures  $\{M_n(A)\} = \{E[N_n(A)]\}$  are finite for bounded  $A \in \mathcal{B}_{\mathcal{X}}$ . These moment measures converge in the  $w^\#$ -topology to the first moment measure  $M$  of  $N$  if and only if for some sequence of spheres  $S_k \uparrow \mathcal{X}$ ,*

$$E[N_n(S_k)I\{N_n(S_k) \geq a\}] \rightarrow 0 \quad (a \rightarrow \infty) \quad \text{uniformly in } n.$$

As an illuminating example of the use of weak convergence arguments, we outline below the proof used by Brémaud and Massoulié (2001) to establish the existence of a type of Hawkes process [Examples 6.3(c), 7.2(b)] with no immigration component: the population is both maintained and balanced purely via its own offspring and their locations.

EXAMPLE 11.1(c) *Existence of Hawkes process without ancestors* [Brémaud and Massoulié (2001)]. We consider a sequence of ordinary Hawkes processes with infectivity functions (intensity measure of offspring process) of the form  $\mu_\epsilon(x) = (1-\epsilon)\mu(x)$ ,  $\int \mu(x) dx = 1$ , with associated immigration rates  $\nu_\epsilon = \epsilon\nu$ . All these processes have the same mean rate

$$\frac{\nu_\epsilon}{1 - \int \mu_\epsilon(x) dx} = \frac{\epsilon\nu}{1 - (1-\epsilon)} = \nu,$$

so it is plausible that as  $\epsilon \rightarrow 0$ , the processes may converge to some limit process with mean rate  $\nu$  and conditional intensity of the form

$$\lambda(t) = \int_{-\infty}^t \mu(t-u) N(du). \quad (11.1.5)$$

Our aim is to establish that such convergence does indeed occur.

In view of Theorem 11.1.VI, it is sufficient to show that the finite dimensional distributions converge to a consistent limit. Fix an interval  $[a, b]$  and consider the total number  $N_\epsilon(a, b)$  of points of the approximating process falling within this interval. From stationarity of the approximating process we have from Markov’s inequality, uniformly for all  $\epsilon > 0$ ,

$$\Pr\{N_\epsilon(a, b) > M\} \leq \frac{E[N_\epsilon(a, b)]}{M} = \frac{\nu(b-a)}{M},$$

so that the left-hand side converges uniformly to 0 as  $M \rightarrow \infty$ . It follows that the distributions of  $N_\epsilon(a, b)$  are uniformly tight. It is easily seen from this that all fidi distributions for the processes  $N_\epsilon$  restricted to  $(a, b)$  are uniformly tight. We can therefore extract a subsequence such that the fidi distributions on  $(a, b)$  converge weakly to some limit process on  $(a, b)$ . By covering the real line with a family of such intervals, we can even find a subsequence along which all the fidi distributions converge weakly. Then it follows from Theorem 11.1.VI that the point processes themselves converge weakly to some limit point process. Exercise 11.1.4 gives a more general version of this argument.

In this model, the weak convergence just established also implies convergence of the expressions for the conditional intensities. To see this, consider expressions of the form which define the conditional intensity, namely

$$E[N(a, b)I_A] = E\left[\int_a^b I_A \lambda(u) du\right], \quad (11.1.6)$$

where  $A \in \mathcal{H}_a$ , the internal (minimal) history for the process (see the discussion in Chapter 7, or later in this volume in Chapter 14). In particular, it is enough to consider  $A$  of the form

$$A = \{N: N(C_1) = n_1, N(C_2) = n_2, \dots, N(C_k) = n_k\}$$

for integers  $k$ ,  $n_1, \dots, n_k$  and sets  $C_i \in (-\infty, a)$ , because sets of this kind generate  $\mathcal{H}_a$ . In this model, the existence of densities for the Poisson processes of new immigrants and offspring implies that all bounded Borel sets  $C \in \mathbb{R}$  are continuity sets. It follows that the function  $N \mapsto N(a, b)I_A$  is continuous in the weak<sup>#</sup> topology in  $\mathcal{B}_N$  (see the discussion following Lemma 11.1.III), and hence, from the continuity theorem (Proposition A2.3.V), that  $E[N_\epsilon(a, b)I_A] \rightarrow E[N(a, b)I_A]$ . Similar arguments apply also to the more complex expressions on the right-hand side of (11.1.6), which for the approximating process  $N_\epsilon$  we can write as

$$E\left[\int_a^b I_A \lambda_\epsilon(x) dx\right] = E\left[\int_a^b I_A \left\{\nu_\epsilon + \int_{-\infty}^u \mu_\epsilon(x-u) N(du)\right\} dx\right].$$

Again the expectations converge to the corresponding form for the limit process, and serve to identify the conditional intensity for the limit process with the form (11.1.5).

Nothing in the argument so far precludes the possibility that the limit point process is a.s. equal to the zero counting measure. Indeed, Brémaud and Massoulié show that if the function  $\mu(x)$  is ‘light-tailed’ (decays at an exponential rate or faster) then the only possible limit processes are degenerate (zero or infinite; see Exercise 11.1.5). Because  $E[N_\epsilon(a, b)] = \nu(b-a)$ , a sufficient condition to ensure that the limit process is nontrivial is that the limit of the first moment measures should be the first moment measure of

the limit process. For this, a uniform integrability condition is needed for the quantities  $N_\epsilon(a, b)\Pr\{N_\epsilon(a, b) > M\}$ , as indicated in Exercise 11.1.3. Because for any random variable  $X \geq 0$ ,  $E[XI_{X>M}] \leq E[X^2]/M$ , a sufficient further condition is boundedness of the variances  $\text{var}[N_\epsilon(a, b)]$ . This requires a careful examination of the spectral properties of  $N_\epsilon$ ; for details see Exercise 11.1.6, where it is shown that the variances remain bounded provided the infectivity function satisfies the ‘heavy-tail’ conditions that for some  $0 < \alpha < \frac{1}{2}$ ,  $t^{1+\alpha}\mu(t)$  is bounded on  $t \geq 0$ , and  $t^{1+\alpha}\mu(t)$  approaches a finite limit as  $t \rightarrow \infty$ .  $\square$

We conclude this section with a few remarks concerning the relation between weak convergence of random measures with state space  $\mathcal{X} = \mathbb{R}_+$  and weak convergence of the associated cumulative processes as elements in  $\mathcal{D}(0, \infty)$ . Here the cumulative function associated with a measure  $\mu$  on  $\mathbb{R}_+$  is defined by

$$F_\mu(x) = \mu\{(0, x]\} \quad 0 < x < \infty.$$

Such functions are monotonic increasing, right-continuous with left limits, and therefore define a subspace of  $\mathcal{D}(0, \infty)$ . Because the metrics in  $\mathcal{M}^\#(\mathbb{R}_+)$  and  $\mathcal{D}(0, \infty)$  are obtained from compounding the analogous metrics over finite intervals, it is sufficient to compare the behaviour of the two metrics over a common finite interval, which for convenience we take as the interval  $(0, 1]$ . In both cases we are effectively concerned with the distance between two cumulative functions  $F, G$  over  $(0, 1]$ . Weak convergence of a family of measures on  $(0, 1]$  is equivalent to convergence of the cumulative functions with respect to the Lévy metric  $\rho_L$ , where  $\rho_L(F, G)$  is defined as the infimum of values  $\varepsilon$  such that for all  $x \in (0, 1]$ ,

$$G(x - \varepsilon) - \varepsilon \leq F(x) \leq G(x + \varepsilon) + \varepsilon$$

[we take  $G(-y) = 0$ ,  $G(1 + y) = G(1)$  ( $y > 0$ ) for the purposes of this definition]. On the other hand convergence of the distribution functions in  $\mathcal{D}(0, 1)$  is equivalent to convergence with respect to the Skorohod metric  $\rho_S$ , where  $\rho_S(F, G)$  is defined as the infimum of values  $\varepsilon$  such that there exists a continuous mapping  $\lambda$  of  $[0, 1]$  onto  $[0, 1]$ , with  $\lambda(0) = 0$ ,  $\lambda(1) = 1$ , for which

$$\sup_{0 \leq x \leq 1} |x - \lambda(x)| < \varepsilon, \quad \sup_{0 \leq x \leq 1} |F(\lambda(x)) - G(x)| < \varepsilon.$$

The statements  $\rho_L(F, G) < \varepsilon$  and  $\rho_S(F, G) < \varepsilon$  both require  $F$  and  $G$  to be close in the sense that uniformly for  $x \in (0, 1]$ , the value of  $F(x)$  differs from a possibly slightly shifted value of  $G(x)$  by less than  $\varepsilon$ , the shift also not being allowed to exceed  $\varepsilon$ . In the case of the Skorohod metric, the degree of shift is controlled by the function  $\lambda(x)$ , whereas in the Lévy case it is constrained not to exceed  $\varepsilon$  but is otherwise not related from one  $x$  to any other.

In both cases the statement  $\rho(F_n, F) \rightarrow 0$  is equivalent to the requirement that  $F_n(x) \rightarrow F(x)$  at all continuity points of  $x$  (see Exercises 11.1.7–8).



Provided therefore that discussion is restricted to the subspace of cumulative processes, the two types of convergence are equivalent. Equivalently, the mapping from  $\mathcal{M}(0, 1)$  into  $\mathcal{D}(0, 1)$ , which takes the measure  $\xi$  into its cumulative function  $F_\xi$ , is both ways continuous. The continuous mapping Theorem A2.3.V therefore yields the following result.

**Lemma 11.1.XI.** *A sequence of random measures  $\{\xi_n\}$  on  $\mathcal{M}^\#(\mathbb{R}^+)$  converges weakly to a random measure  $\xi$  if and only if the corresponding sequence of cumulative processes  $F_{\xi_n}$  converges weakly in  $\mathcal{D}(0, \infty)$  to the cumulative process  $F_\xi$ .*

Extensions to  $\mathbb{R}^d$  can be obtained in terms of the concepts described by Straf (1972).

From the point of view of weak convergence, it is therefore immaterial whether we deal with the random measures directly or the stochastic processes defined by the associated cumulative functions.

When rescaling is involved, as, for example, in the central limit theorem, the limits need no longer correspond to random measures, and functional limit theorems can be obtained. An example is given in Proposition 12.3.X.

## Exercises and Complements to Section 11.1

11.1.1 For totally finite signed measures  $\mu, \nu$  on a c.s.m.s.  $\mathcal{X}$  show that the variation norm  $\|\cdot\|$  (see Section A1.3) has the following properties:

- (a)  $\|\alpha\mu\| = |\alpha| \|\mu\|$  for every scalar  $\alpha$ .
- (b)  $\|\mu + \nu\| \leq \|\mu\| + \|\nu\|$ , with equality if the supports of  $\mu$  and  $\nu$  are disjoint.
- (c)  $\|\mu * \nu\| \leq \|\mu\| \|\nu\|$ , with equality if both measures are of constant sign (and thus  $\|\mu\| = |\mu(\mathcal{X})|$ ).

11.1.2 (a) Let  $N$  be a point process on  $\mathbb{R}$  and  $\mathcal{R}$  the ring generated by half-open intervals such as  $(a, b]$ , so that for  $A \in \mathcal{R}$ ,  $\partial A$  consists of the finite set of endpoints of the constituent intervals of  $A$ . Deduce that  $A$  is a stochastic continuity set unless any of the endpoints of its constituent intervals happen to be fixed atoms of the process.

- (b) Deduce that if  $N$  on  $\mathbb{R}$  is stationary, it has no fixed atoms.
- (c) In general, when  $\mathcal{P}$  is a probability measure on  $\mathcal{M}_\mathcal{X}^\#$ , the stochastic continuity sets for  $\mathcal{P}$  form an algebra.

11.1.3 *Convergence of moment measures.* Let  $N$  be a point process that is the limit of the weakly convergent sequence  $\{N_n\}$  of point processes. Show that  $M(A) = E[N(A)] \leq \liminf_{n \rightarrow \infty} E[N_n(A)] = \liminf_{n \rightarrow \infty} M_n(A)$ .

In order to have equality here, some condition such as the uniform integrability condition of Proposition 11.1.X is needed: weak convergence and existence and boundedness of the moment measures are not enough to ensure their vague convergence.

Suppose that the point processes  $N_n$  are on  $\mathbb{R}$  and defined as follows. Choose with probability  $1 - n^{-1}$  the randomized unit lattice point process with points at  $\{r + U: r = 0, \pm 1, \dots\}$  and  $U$  a random variable uniformly distributed on  $(0, 1)$ , and with probability  $n^{-1}$  the point process with points

on the lattice  $\{r/n: r = 0, \pm 1, \pm 2, \dots\}$  in which the centre of the lattice is located uniformly at random over the unit interval. This sequence  $\{N_n\}$  has the following properties.

- (i) It converges weakly to the randomized unit lattice point process.
- (ii)  $E[N_n(A)] = (2 - n^{-1})\ell(A)$  for  $n = 1, 2, \dots$
- (iii)  $E[N_\infty(A)] = \ell(A)$  for all bounded Borel sets  $A$ .

11.1.4 When  $\mathcal{X} = \mathbb{R}^d$ , every sequence of random measures with locally uniformly bounded first moment measures is relatively weakly compact (i.e., contains a weakly convergent subsequence). In particular, any sequence of random measures with uniformly bounded mean densities is relatively weakly compact. [Hint: Use Markov's inequality to show that for each bounded Borel set  $A$  there exists a constant  $C_A < \infty$  such that  $\Pr\{\xi(A) > M\} \leq C_A/M$  for all  $M > 0$ .]

11.1.5 For a Hawkes process as in Example 11.1(c), show that if  $\int x\mu(x) dx \leq \infty$  then a process with conditional intensity (11.1.5) cannot have a finite mean. [Hint:  $\Pr\{N(\mathbb{R}_+) = 0\} = E[\Pr\{N(\mathbb{R}_+) = 0 \mid \mathcal{H}_0\}] \geq \exp(-\lambda \int t\mu(t) dt)$ ; this yields a contradiction if the process is ergodic and nonzero.]

11.1.6 *Hawkes process without ancestors* [see Example 11.1(c)].

- (i) Suppose the density  $\mu$  satisfies the conditions  $\int_0^\infty \mu(t) dt = 1$  and for finite positive  $R$  and  $r$ ,  $\sup_{t>0} t^{1+\alpha}\mu(t) \leq R$  and  $\lim_{t \rightarrow \infty} t^{1+\alpha}\mu(t) = r$ . Show that

$$\lim_{\omega \rightarrow 0} [\hat{\mu}(\omega) - 1]\omega^{-\alpha} = r \int_0^\infty \frac{e^{iu} - 1}{u^{1+\alpha}} du.$$

- (ii) Let  $V_\epsilon(T)$  denote the variance of the approximating process  $N_\epsilon$  over a fixed interval  $(0, T)$ . Show that as  $\epsilon \rightarrow 0$ ,  $V_\epsilon(T)$  remains bounded, and

$$V_\epsilon(T) \rightarrow \frac{\lambda}{2\pi} \int_{-\infty}^\infty \frac{|e^{i\omega T} - 1|^2}{\omega^2 |1 - \hat{\mu}(\omega)|^2} d\omega.$$

- (iii) Show that the limit variance obtained above is, in fact, the variance of the limiting process described in the example.

[Hint: For (ii), use (i) and equation (8.2.10) for the spectral density of a Hawkes process. For (iii), see Lemma 1 of Brémaud and Massoulié (2001).]

11.1.7 Let  $\rho_L$ ,  $\rho_S$  refer to the Prohorov and Skorohod metrics, respectively, on the space of finite measures on  $(\mathcal{X}, \mathcal{B}_\mathcal{X})$  (the Prohorov metric reduces to the Lévy metric on  $\mathbb{R}$ ). Let  $\{N_n\}$ ,  $N$  be finite counting measures on  $\mathcal{X}$ . Prove that  $\rho_L(N_n, N) \rightarrow 0$  if and only if for all sufficiently large  $n$ ,  $N_n$  has the same number of atoms as  $N$ , and the locations of the atoms in  $N$ , converge to their locations under  $N$ . Deduce that  $\rho_L(N_n, N) \rightarrow 0$  implies  $\rho_S(N_n, N) \rightarrow 0$ . [Hint: See, e.g., Straf (1972).]

11.1.8 Write  $d_0(F, G) = \sup_{x \in \mathbb{R}} |F(x) - G(x)|$  for the sup metric on the space of d.f.s  $F, G$  on  $\mathbb{R}$ , and write  $\mu_F$ ,  $\mu_G$  for the measures generated by such d.f.s. Prove that  $2d_0(F, G) \leq \|\mu_F - \mu_G\|$ .

## 11.2. Limit Theorems for Superpositions

Limit theorems for superpositions of point processes go back at least as far as Palm (1943) and continued in Khinchin (1955) in developing a simple version of Proposition 11.2.VI below for the superposition of a large number of independent identically distributed stationary point processes on  $\mathbb{R}$ . Under suitable conditions, rescaled versions of the resulting processes have a Poisson process limit. Extensions through work of Ososkov (1956), Franken (1963) and Grigelionis (1963) led ultimately to Theorems 11.2.III and 11.2.V in which rescaling is subsumed by convergence of sums in a uniformly asymptotically negligible array.

The formal setting for studying the sum or superposition of a large number of point processes or random measures is a *triangular array*  $\{\xi_{ni}: i = 1, \dots, m_n; n = 1, 2, \dots\}$  and its associated row sums

$$\xi_n = \sum_{i=1}^{m_n} \xi_{ni}, \quad n = 1, 2, \dots$$

If for each  $n$  the processes  $\{\xi_{ni}: i = 1, \dots, m_n\}$  are mutually independent, we speak of an *independent array*, and when they satisfy the condition that for all  $\varepsilon > 0$  and all bounded  $A \in \mathcal{B}_{\mathcal{X}}$

$$\limsup_{n \rightarrow \infty} \mathcal{P}\{\xi_{ni}(A) > \varepsilon\} = 0, \quad (11.2.1)$$

the array is *uniformly asymptotically negligible*, or u.a.n. for short. In the case of a triangular array of point processes, the u.a.n. condition (11.2.1) reduces to the simpler requirement that

$$\lim_{n \rightarrow \infty} \sup_i \mathcal{P}\{N_{ni}(A) > 0\} = 0. \quad (11.2.2)$$

Note that an independent u.a.n. array is called infinitesimal in MKM (1978, Section 3.4), a null-array in Feller (1966) and Kallenberg (1975, Chapter 6), and holospoudic in Chung (1974, Section 7.1). The terminology u.a.n. comes from Loève (1963) (Loève in fact wrote uan).

Although this formal setting can be extended (see Exercise 11.2.3) and the notation simplified by taking  $m_n = \infty$ , from which the finite case is obtained by assuming all but finitely many elements to be zero, we retain the setting with  $m_n < \infty$  for the sake of familiarity.

In the discussion that follows the reader will doubtless observe the very close analogy between the results developed for point processes and the classical theory for sums of i.i.d. random variables in  $\mathbb{R}$ . This is hardly surprising, for a point process is just a particular type of random measure and a random measure is just a random variable taking its values on the metric Abelian group of boundedly finite signed measures on the state space  $\mathcal{X}$ . As such it comes under the extension of the classical theory developed, for example, in Parthasarathy (1967, Chapter 4). We develop results for point processes

directly; however, the reader may find it useful to bear the classical theory in mind as a guide, as for example at the end of this section in reviewing the corresponding results for random measures.

We start with a preliminary result on the convergence of infinitely divisible point processes, continuing with notation from Section 10.2. Let  $\{\tilde{Q}_n: n = 1, 2, \dots\}$  denote a sequence of KLM measures (Definition 10.2.IV) and  $\tilde{Q}$  a limit measure. We cannot immediately speak of the weak convergence of  $\tilde{Q}_n$  to  $\tilde{Q}$ , firstly because the KLM measures are only  $\sigma$ -finite in general, and secondly because they are only defined on the Borel subsets of the space  $\mathcal{N}_0^\#(\mathcal{X}) \equiv \mathcal{N}_\mathcal{X}^\# \setminus \{N(\mathcal{X}) = 0\}$ , which is not complete. To define an appropriate modification of weak convergence for the sequence  $\{\tilde{Q}_n\}$ , recall that for each bounded Borel set  $A$  the KLM measure  $\tilde{Q}$  induces a totally finite measure  $\tilde{Q}^A$  on the space  $\mathcal{N}_0^\#(A)$  of nonzero counting measures on  $A$ . Extend  $\tilde{Q}^A$  to the whole of  $\mathcal{N}_A^\#$  by setting  $\tilde{Q}^{(A)} = \tilde{Q}^A$  on  $\mathcal{N}_0^\#(A)$  and  $\tilde{Q}^{(A)}\{N: N(A) = 0\} = 0$ .

**Definition 11.2.I.** *The sequence of KLM measures  $\{\tilde{Q}_n: n = 1, 2, \dots\}$  converges  $Q$ -weakly to the KLM measure  $\tilde{Q}$  (i.e.,  $\tilde{Q}_n \rightarrow \tilde{Q}$   $Q$ -weakly), if for every bounded Borel set  $A$  that is a stochastic continuity set for  $\tilde{Q}$ , the extended measures  $\tilde{Q}_n^{(A)}$  converge weakly to  $\tilde{Q}^{(A)}$ .*

This requirement can be spelt out more explicitly in terms of fidi distributions or the convergence of functionals. Thus, it is equivalent to requiring that for every finite family of bounded Borel sets  $A_i \in \mathcal{S}_{\tilde{Q}}$  ( $i = 1, \dots, k$ ;  $k \geq 2$ ),

$$\begin{aligned} \tilde{Q}_n \left\{ N: N(A_i) = j_i \ (i = 1, \dots, k), \sum_{i=1}^k j_i > 0 \right\} \\ \rightarrow \tilde{Q} \left\{ N: N(A_i) = j_i \ (i = 1, \dots, k) \right\}. \end{aligned} \quad (11.2.3)$$

A more convenient form for our purposes is the following. For every continuous function  $h \in \mathcal{V}(\mathcal{X})$  equal to one outside some bounded Borel set  $A$ , as  $n \rightarrow \infty$ ,

$$\begin{aligned} \int_{\{N(A) > 0\}} \exp \left( \int_{\mathcal{X}} \log h(x) N(dx) \right) \tilde{Q}_n(dN) \\ \rightarrow \int_{\{N(A) > 0\}} \exp \left( \int_{\mathcal{X}} \log h(x) N(dx) \right) \tilde{Q}(dN). \end{aligned} \quad (11.2.4)$$

These restatements are immediate consequences of the conditions for weak convergence developed in Section 11.1 applied to the measures  $\tilde{Q}_n^{(A)}$ , which, although not probability measures, are totally finite so that the framework for weak convergence remains intact.

**Proposition 11.2.II.** (a) *The set of infinitely divisible distributions is closed in the topology of weak convergence in  $\mathcal{N}_\mathcal{X}^\#$ .*

(b) *If  $\{\mathcal{P}_n: n = 1, 2, \dots\}$  and  $\mathcal{P}$  are infinitely divisible distributions on  $\mathcal{N}_\mathcal{X}^\#$ , and  $\{\tilde{Q}_n\}$ ,  $\tilde{Q}$  the corresponding KLM measures, then  $\mathcal{P}_n \rightarrow \mathcal{P}$  weakly if and only if  $\tilde{Q}_n \rightarrow \tilde{Q}$   $Q$ -weakly.*

PROOF. Suppose that  $\mathcal{P}_n \rightarrow \mathcal{P}$  weakly and that the  $\mathcal{P}_n$  are infinitely divisible. Take any integer  $k$ , and observe that if  $\mathcal{P}_n$  has p.g.fl.  $G_n[\cdot]$ , then  $(G_n[\cdot])^{1/k}$  is also a p.g.fl. and corresponds to the KLM measure  $k^{-1}\tilde{Q}_n(\cdot)$ , where  $\tilde{Q}_n$  is the KLM measure of  $\mathcal{P}_n$ . When  $G_n[h] \rightarrow G[h]$  for all continuous  $h \in \mathcal{V}(\mathcal{X})$ , it follows that  $(G_n[h])^{1/k} \rightarrow (G[h])^{1/k}$ , so that  $(G[h])^{1/k}$  is a p.g.fl. for every integer  $k$ , and hence  $\mathcal{P}$  is infinitely divisible. Using the p.g.fl. representation (10.2.9), we have for all continuous  $h \in \mathcal{V}(\mathcal{X})$ ,

$$\begin{aligned} \int_{\mathcal{N}_0^\#(\mathcal{X})} \left[ \exp \left( \int_{\mathcal{X}} \log h(x) N(dx) \right) - 1 \right] \tilde{Q}_n(dN) \\ \rightarrow \int_{\mathcal{N}_0^\#(\mathcal{X})} \left[ \exp \left( \int_{\mathcal{X}} \log h(x) N(dx) \right) - 1 \right] \tilde{Q}(dN). \end{aligned}$$

Taking  $A$  to be a stochastic continuity set for  $\tilde{Q}$ , we can approximate the step function  $h(x) = 1 - I_A(x)$  arbitrarily closely by  $h \in \mathcal{S}_{\tilde{Q}}$  and thus conclude that

$$\tilde{Q}_n\{N: N(A) > 0\} \rightarrow \tilde{Q}\{N: N(A) > 0\},$$

as the exponential term above vanishes if  $N(A) > 0$ . By subtraction we then have that the integrals

$$\int_{\{N(A) > 0\}} \exp \left( \int_{\mathcal{X}} \log h(x) N(dx) \right) \tilde{Q}_n(dN)$$

converge as required at (11.2.4).

Conversely, when  $\tilde{Q}_n \rightarrow \tilde{Q}$   $Q$ -weakly, the argument can be reversed.  $\square$

Given a point process with distribution  $\mathcal{P}$  and p.g.fl.  $G[\cdot]$ , define its *Poisson approximant* [corresponding to the *accompanying law* in the classical theory—see, e.g., Parthasarathy (1967, VI.6)] to be the Poisson randomization with distribution  $\mathcal{P}^*$  and p.g.fl.  $G^*[\cdot]$  given by

$$G^*[h] = \exp(G[h] - 1) \quad (\text{all } h \in \mathcal{V}(\mathcal{X})). \quad (11.2.5)$$

More generally, given a triangular array  $\{N_{ni}\}$ , the corresponding Poisson approximants are given by  $N_{ni}^*$ , with distributions  $\mathcal{P}_{ni}^*$  and p.g.fl.s  $G_{ni}^*[\cdot]$ ; when the  $N_{ni}$  are independent, take the  $N_{ni}^*$  to be independent also. Because

$$\mathcal{P}^*\{N(A) > 0\} = 1 - \exp[-\mathcal{P}\{N(A) > 0\}] \leq \mathcal{P}\{N(A) > 0\},$$

the triangular array  $\{N_{ni}^*\}$  is u.a.n. whenever  $\{N_{ni}\}$  is a u.a.n. array (see also Exercises 11.2.1–2). The following theorem is basic for point processes.

**Theorem 11.2.III.** *Let  $\{N_{ni}: i = 1, \dots, m_n; n = 1, 2, \dots\}$  be an independent u.a.n. array,  $\{N_{ni}^*\}$  an independent array of corresponding Poisson approximants, and  $N$  an infinitely divisible point process with KLM measure  $\tilde{Q}$ . Then the following assertions are equivalent.*

- (i)  $\sum_{i=1}^{m_n} N_{ni} \rightarrow N$  weakly.
- (ii)  $\sum_{i=1}^{m_n} N_{ni}^* \rightarrow N$  weakly.
- (iii)  $\sum_{i=1}^{m_n} \mathcal{P}_{ni}^{(0)} \rightarrow \tilde{Q}$   $Q$ -weakly.

[In (iii),  $\mathcal{P}_{ni}^{(0)}$  is the restriction of  $\mathcal{P}_{ni}$ , the distribution of  $N_{ni}$ , to  $\mathcal{N}_0^\#(\mathcal{X})$ .]

PROOF. Recall the simple inequalities, valid for  $0 \leq 2\alpha_i \leq 1$  ( $i = 1, \dots, m_n$ )

$$0 \leq -\log \left[ \prod_{i=1}^{m_n} (1 - \alpha_i) \right] - \sum_{i=1}^{m_n} \alpha_i \leq \sum_{i=1}^{m_n} \alpha_i^2 \leq \left( \sum_{i=1}^{m_n} \alpha_i \right) (\max_i \alpha_i). \quad (11.2.6)$$

We apply these inequalities with  $\alpha_i = G_{ni}[h]$ , where  $h(x) = 1$  for  $x$  outside some stochastic continuity set  $A$  for  $\mathcal{P}$ , the distribution of  $N$ . From the u.a.n. condition,  $1 - G_{ni}[h] \leq \Pr\{N_{ni}(A) > 0\} \leq \frac{1}{2}$  ( $i = 1, \dots, m_n$ ) for  $n$  sufficiently large, and thus

$$\begin{aligned} \sum_{i=1}^{m_n} (1 - G_{ni}[h]) &\leq \sum_{i=1}^{m_n} \Pr\{N_{ni}(A) > 0\} \leq -\log \left[ \prod_{i=1}^{m_n} [1 - \Pr\{N_{ni}(A) > 0\}] \right] \\ &= -\log \left( \Pr \left\{ \sum_{i=1}^{m_n} N_{ni}(A) = 0 \right\} \right) \\ &\rightarrow -\log (\Pr\{N(A) = 0\}) < \infty, \end{aligned}$$

so that the left-hand sum here is uniformly bounded (over subsets of  $A$ ) for  $n$  sufficiently large. It follows from (11.2.6) that if one of

$$\prod_{i=1}^{m_n} G_n[h] \quad \text{and} \quad \exp \left( \sum_{i=1}^{m_n} (1 - G_{ni}[h]) \right)$$

converges to a finite nonzero limit, then so does the other, and the limits are equal. This implies the equivalence of (i) and (ii) of the theorem.

The processes  $N_{ni}^*$  are infinitely divisible, with KLM measures  $\mathcal{P}_{ni}^{(0)}$ , so the row sum  $\sum N_{ni}^*$  is infinitely divisible, with KLM measure  $\sum \mathcal{P}_{ni}^{(0)}$ . By appealing to Proposition 11.2.II, the equivalence of (ii) and (iii) follows.  $\square$

The arguments used in the proof lead to an alternative formulation of the result in terms of p.g.f.s (see Exercise 11.2.3).

The following result is an easy corollary.

**Proposition 11.2.IV.** *A point process is infinitely divisible if and only if it can be represented as the limit of the row sums of a u.a.n. array.*

The most important application of Theorem 11.2.III is to finding conditions for convergence to a Poisson process.

**Theorem 11.2.V.** *The triangular u.a.n. array  $\{N_{ni}: i = 1, \dots, m_n; n = 1, 2, \dots\}$  converges weakly to a Poisson process with parameter measure  $\mu$  if and only if for all bounded Borel sets  $A$  with  $\mu(\partial A) = 0$ ,*

$$\sum_{i=1}^{m_n} \Pr\{N_{ni}(A) \geq 2\} \rightarrow 0 \quad (n \rightarrow \infty) \quad (11.2.7)$$

and

$$\sum_{i=1}^{m_n} \Pr\{N_{ni}(A) \geq 1\} \rightarrow \mu(A) \quad (n \rightarrow \infty). \quad (11.2.8)$$

PROOF. Recall from Example 10.2(a) that for a Poisson process the KLM measure  $\tilde{Q}(\cdot)$  is related to the parameter measure  $\mu(\cdot)$  by

$$\mu(A) = \tilde{Q}\{N: N(A) > 0\},$$

and that  $\tilde{Q}$  itself is concentrated on one-point realizations, so that

$$\tilde{Q}\{N: N(\mathcal{X}) > 1\} = 0.$$

It follows from Theorem 11.2.III that if the array converges to a Poisson process, then

$$\sum_{i=1}^{m_n} \Pr\{N_{ni}(A) > 0\} \rightarrow \tilde{Q}\{N: N(A) > 0\} = \mu(A),$$

and

$$\sum_{i=1}^{m_n} \Pr\{N_{ni}(A) \geq 2\} \rightarrow \tilde{Q}\{N: N(A) \geq 2\} = 0,$$

so the conditions (11.2.7) and (11.2.8) are necessary.

Conversely, if (11.2.7) holds for a sequence of sets  $A_n \uparrow \mathcal{X}$ , we must then have  $\tilde{Q}\{N: N(\mathcal{X}) > 1\} = 0$ , so that the limit process must be Poisson, and (11.2.8) identifies the parameter measure as  $\mu$ .  $\square$

We remark that as in other applications of weak convergence, it is sufficient, in checking the conditions of the theorem, to let  $A$  run through the sets of any covering semiring of continuity sets of  $\mu$  [see Proposition 11.1.VIII(iv)].

The following special case was the first to be studied and can be regarded as the prototype limit theorem for point processes.

**Proposition 11.2.VI.** *Let  $N$  be a simple stationary point process on  $\mathcal{X} = \mathbb{R}$  with finite intensity  $\lambda$ , and let  $N_n$  denote the point process obtained by superposing  $n$  independent replicates of  $N$  and dilating the scale of  $\mathcal{X}$  by a factor  $n$ . Then as  $n \rightarrow \infty$ ,  $N_n$  converges weakly to a Poisson process with parameter measure  $\lambda\ell(\cdot)$ , where  $\ell(\cdot)$  denotes Lebesgue measure on  $\mathbb{R}$ .*

PROOF. Here we can envisage a triangular array situation in which each  $N_{ni}$  ( $i = 1, \dots, n$ ) has the same distribution as the original process but on a dilated scale. Hence, using Propositions 3.3.I and 3.3.IV,

$$\Pr\{N_{ni}(0, t] > 0\} = \Pr\{N(0, t/n] > 0\} = (\lambda t/n)(1 + o(1)).$$

Summing on  $i = 1, \dots, n$  leads to (11.2.8) with  $\mu(\cdot) = \lambda\ell(\cdot)$ . Similarly, from Proposition 3.3.V,

$$\Pr\{N_{ni}(0, t] > 1\} = \Pr\{N(0, t/n] > 1\} = o(1/n),$$

and again summing on  $i$  leads to (11.2.7).  $\square$

The statement and proof need change when  $\mathcal{X} = \mathbb{R}^d$ ; see Exercise 11.2.4.

We conclude this section by briefly reviewing some extensions and further developments.

Some of the results that we have handled by generating function arguments can be strengthened to give results concerning bounds in variation norm. In particular, there are elegant bounds that follow via the use of Poisson approximants (see Exercises 11.2.1–2).

Extensions to the multivariate, nonorderly, and marked point process cases can generally be handled by applying the preceding results to the case where  $\mathcal{X}$  has the product form  $\mathcal{X} \times \mathcal{K}$  for an appropriate mark space  $\mathcal{K}$ .

EXAMPLE 11.2(a) *Convergence to a multivariate independent Poisson process.* Suppose there is given a point process in which each point is identifiable as one of a finite set of types  $1, \dots, K$  say. The process can be described by the multivariate processes with component processes  $N_{ni}^{(k)}(\cdot)$  ( $k = 1, \dots, K$ ). We seek conditions for weak convergence of the superpositions to a limit process in which the different types follow independent Poisson processes with parameter measures  $\mu_k$  ( $k = 1, \dots, K$ ). This last process can thus be regarded as a Poisson process on the space  $\mathcal{X} \times \{1, \dots, K\}$  with overall measure  $\mu$  such that  $\mu_k(\cdot) = \mu(\cdot \times \{k\})$ . Similarly, regard the family  $\{N_{ni}^{(k)}: k = 1, \dots, K\}$  as defining a process  $N_{ni}$  on  $\mathcal{X} \times \{1, \dots, K\}$ . To apply Theorem 11.2.V we have to interpret (11.2.7) and (11.2.8), which apply to the overall processes  $N_{ni}$ , in terms of the components. We take  $A$  at (11.2.7) and (11.2.8) to be a product set of the form  $B \times \{1, \dots, K\}$  for some bounded Borel set  $B$  that is a stochastic continuity set for each  $\mu_1, \dots, \mu_k$ ; that is,

$$\mu_k(\partial B) = 0 \quad (k = 1, \dots, K). \quad (11.2.9)$$

Then (11.2.7) becomes

$$\sum_{i=1}^{m_n} \Pr\left\{\sum_{k=1}^K N_{ni}^{(k)}(B) \geq 2\right\} \rightarrow 0 \quad (n \rightarrow \infty), \quad (11.2.10)$$



which incorporates the requirement (crucial if the limit process is to have independent components) that there should be zero limiting probability for two distinct components each to contribute points to the same bounded  $B$ .

Similarly, (11.2.8) takes the form that for all bounded  $B_k$  for which  $\mu_k(\partial B_k) = 0$ ,  $k = 1, \dots, K$ ,

$$\sum_{i=1}^{m_n} \Pr \left\{ \sum_{k=1}^K N_{ni}^{(k)}(B) \geq 1 \right\} \rightarrow \sum_{k=1}^K \mu_k(B_k),$$

which in view of (11.2.10) is satisfied if and only if for each bounded  $B$  for which  $\mu_k(\partial B) = 0$  ( $k = 1, \dots, K$ ),

$$\sum_{i=1}^{m_n} \Pr \{ N_{ni}^{(k)}(B) \geq 1 \} \rightarrow \mu_k(B). \quad (11.2.11)$$

Note also that the u.a.n. condition here becomes

$$\lim_{n \rightarrow \infty} \sup_i \Pr \left\{ \sum_{k=1}^K N_{ni}^{(k)}(B) > 0 \right\} = 0,$$

again apparently incorporating a constraint on the simultaneous occurrence of points of several types. However, because the mark space here consists only of a finite set of types, the u.a.n. condition is equivalent to a componentwise u.a.n. condition, namely,

$$\lim_{n \rightarrow \infty} \sup_i \Pr \{ N_{ni}^{(k)}(B) > 0 \} = 0 \quad (k = 1, \dots, K). \quad (11.2.12)$$

Thus, from Theorem 11.2.V we have the following corollary.

**Corollary 11.2.VII.** *For a  $K$ -variate independent triangular array satisfying the u.a.n. condition (11.2.12), the necessary and sufficient conditions for convergence to a  $K$ -variate Poisson process with independent components are that (11.2.10) and (11.2.11) hold.  $\square$*

Necessary and sufficient conditions for convergence to other types of infinitely divisible point processes, in particular the Poisson cluster process, can be derived by referring back to the general results of Theorem 11.2.III. The procedure is similar to that outlined in Theorem 11.2.V: first identify the KLM measure for the process of interest, and then use (iii) to obtain necessary and sufficient conditions on the component probabilities to ensure convergence to the appropriate limit. The particular case of convergence to a Gauss–Poisson process is outlined in Exercise 11.2.5. There are further complements to Example 11.2(a) in Exercises 11.2.6–7.

Finally, we return to the question of convergence of u.a.n. arrays of random measures broached at the beginning of this section. From the general structural form of an infinitely divisible random measure given in Proposition 10.2.IX, we have the following condition for the convergence of a u.a.n. array of random measures (see also Exercise 11.2.10).

**Proposition 11.2.VIII.** *Let  $\{\xi_{ni}; i = 1, \dots, m_n; n = 1, 2, \dots\}$  be an independent u.a.n. array of random measures on the c.s.m.s.  $\mathcal{X}$ ,  $\xi_n = \sum_{i=1}^{m_n} \xi_{ni}$ , and  $\xi$  an infinitely divisible random measure with Laplace representation (10.2.11). Then necessary and sufficient conditions for  $\xi_n \rightarrow \xi$  weakly, are*

$$(i) \quad \sum_{i=1}^{m_n} \mathcal{P}_{ni}^{(0)} \rightarrow \Lambda \quad Q\text{-weakly on } \mathcal{M}_0^\#(\mathcal{X});$$

(ii) *for all bounded  $A \in \mathcal{B}_{\mathcal{X}}$ ,*

$$\begin{aligned} \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \sum_{i=1}^{m_n} \mathbb{E}[\xi_{ni}(A) I_{[0, \varepsilon]}(\xi_{ni}(A))] \\ = \lim_{\varepsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \sum_{i=1}^{m_n} \mathbb{E}[\xi_{ni}(A) I_{[0, \varepsilon]}(\xi_{ni}(A))] = \alpha(A); \text{ and} \end{aligned}$$

$$(iii) \quad \lim_{r \rightarrow \infty} \limsup_{n \rightarrow \infty} \sum_{i=1}^{m_n} \mathcal{P}\{\xi_{ni}(A) > r\} = 0.$$

Just as for point processes, the conditions (i)–(iii) of this proposition can be summarized more succinctly into the single requirement that the Laplace functionals should satisfy

$$\sum_{i=1}^{m_n} (1 - L_{ni}[f]) \rightarrow \int_{\mathcal{X}} f(x) \alpha(dx) - \int_{\mathcal{M}_{\mathcal{X}}^\#} \left[ \exp \left( - \int_{\mathcal{X}} f(x) \xi(dx) \right) - 1 \right] \Lambda(d\xi). \quad (11.2.13)$$

A sketch proof, completely analogous to that of Theorem 11.2.III, is set out in Exercises 11.2.8–9. For a more detailed treatment see Kallenberg (1975, Chapter 6).

## Exercises and Complements to Section 11.2

11.2.1 (a) Let  $\mathcal{P}$ ,  $\mathcal{P}^*$  be the distributions of a totally finite point process on  $\mathcal{X}$  and its Poisson approximant, respectively. Show that

$$\|\mathcal{P} - \mathcal{P}^*\| \leq 2(\mathcal{P}\{N(\mathcal{X}) > 0\})^2.$$

(b) Denoting the convolution of measures  $\mathcal{P}_{ni}$  by  $*\mathcal{P}_{ni}$ , show that

$$\|*\mathcal{P}_{ni} - *\mathcal{P}_{ni}^*\| \leq 2 \sum_i (\mathcal{P}_{ni}\{N(\mathcal{X}) > 0\})^2.$$

(c) Conclude that for independent u.a.n. arrays, the weak convergence of the probability measures in parts (i) and (ii) of Theorem 11.2.III can be replaced by their strong convergence on bounded Borel sets  $A$ .

11.2.2 For finite point processes  $\mathcal{P}_j$  with Poisson approximants  $\mathcal{P}_j^*$  ( $j = 1, 2$ ), show that  $\|\mathcal{P}_1^* - \mathcal{P}_2^*\| \leq 2\|\mathcal{P}_1 - \mathcal{P}_2\|$ .

11.2.3 Restate Theorem 11.2.III as follows: the necessary and sufficient condition for convergence of an independent u.a.n. array is that

$$\sum_{i=1}^{m_n} (1 - G_{ni}[h]) \rightarrow \int_{\mathcal{N}_0^\#(\mathcal{X})} \left( \exp \left[ \int_{\mathcal{X}} \log h(x) N(dx) \right] - 1 \right) Q(dN)$$

for all  $h \in \mathcal{V}(\mathcal{X})$ , and that the right-hand side then equals  $\log G[h]$ , where  $G$  is the p.g.fl. of the limit point process.

Hence or otherwise deduce that Theorem 11.2.III remains valid in the case  $m_n = \infty$  provided only that the resultant superpositions are well defined.

11.2.4 *Extensions of the prototype limit result for superpositions.*

- (a) Formulate statements analogous to Proposition 11.2.VI when  $N$  is stationary but instead of simple it is (i) nonorderly; or (ii) a marked point process (does the limiting process have independent marks?).
- (b) Consider the case  $\mathcal{X} = \mathbb{R}^d$  for some  $d \geq 2$  in which  $N_n$  is as stated in Proposition 11.2.VI except that any  $A \in \mathcal{B}_{\mathcal{X}}$  is now rescaled by a factor  $n^{1/d}$ . What conditions on  $N$  suffice for the weak convergence of  $N_n$ ? What is needed to obtain nontrivial conclusions in the scenarios of (a)?

11.2.5 For a Gauss–Poisson process [see Example 6.3(d)] the KLM measure  $\tilde{Q}$  is described in Exercise 10.2.3. Use this to deduce that if an independent u.a.n. array converges to a Gauss–Poisson process, then the following hold for bounded  $A, A_1, A_2$  in  $\mathcal{B}_{\mathcal{X}}$ :

- (a)  $\sum_i \Pr\{N_{ni}(A) \geq 3\} \rightarrow 0$ ;
- (b)  $\sum_i \Pr\{N_{ni}(A) = 1\} \rightarrow \tilde{Q}_1\{N(A) = 1\} + 2\tilde{Q}_2\{N(A) = 1\}$ ;
- (c)  $\sum_i \Pr\{N_{ni}(A_1) = 1, N_{ni}(A_2) = 1\} \rightarrow \tilde{Q}_2\{N(A_1) = 1, N(A_2) = 1\}$  for disjoint  $A_1, A_2$ .

11.2.6 Express the result of Example 11.2(a) in terms of multivariate p.g.fl.s.

11.2.7 Formulate a limit theorem for superpositions of independent marked point processes on the space  $\mathcal{X} \times \mathcal{K}$  by regarding the components and the limit as point processes on  $\mathcal{X} \times \mathcal{K}$ . [Hint: Compare with Example 11.2(a), and assume first that the marks are independent as in Proposition 6.4.IV.]

11.2.8 Compare the statements of Theorem 11.2.III and Proposition 11.2.VIII. To establish the latter, proceed as below.

- (a) Introduce Poisson approximants and show that for a u.a.n. array their sum converges if the sum of original summands converges.
- (b) Prove an analogue of Proposition 11.2.II for the convergence of infinitely divisible random measures in terms of the convergence of their components  $\alpha_n$  and  $\Lambda_n$ .
- (c) Finally, apply part (b) to the Poisson approximants.

11.2.9 (*Continuation*). Show that the equivalence of (11.2.13) and Proposition 11.2.VIII can be regarded as a continuity theorem for Laplace functionals complicated by the detail of the behaviour near the zero measure.

11.2.10 Interpret the conditions in Proposition 11.2.VIII in terms of the setting of Chapter 4 of Parthasarathy (1967).

## 11.3. Thinned Point Processes

The notion of thinning a point process to construct another point process has been described in Example 4.3(a) for renewal processes in the simplest case where the thinning occurs independently for each point. The idea underlying the operation is that, in principle, points of a process may occur at any of a very large number of locations, but it is only at a relatively small proportion of such locations that points are observed. The limit theorems described below formulate sufficient conditions for this process of *rarefaction* (or *thinning* or *deletion*) to lead in the limit to a Poisson process.

Let  $N$  be a point process on  $\mathcal{X} = \mathbb{R}^d$  and  $p(\cdot)$  a measurable function on  $\mathcal{X}$  with  $0 \leq p(x) \leq 1$  (all  $x$ ).  $N_{p(\cdot)}$  is obtained from  $N$  by *independent thinning according to  $p(\cdot)$*  when the following holds. Let the realization  $N(\cdot, \omega)$  consist of the countable set of points  $\{x_i\}$  (cf. Proposition 9.1.V); construct a subset of these points by taking each  $x_i$  in turn, deleting it with probability  $1 - p(x_i)$  and retaining it with probability  $p(x_i)$ , independently for each point; and regard the set of points so retained as defining a realization of the thinned point process  $N_{p(\cdot)}(\cdot, \omega)$ .

Some form of rescaling is needed before a limit theorem can emerge: the simplest set-up is the following. Take  $\mathcal{X} = \mathbb{R}$  and  $p(x) = p$  (all  $x$ ), and after thinning contract the scale in  $\mathcal{X}$  by an amount  $p$ , so that the point  $x \in \mathbb{R}$  is mapped into  $px$ ; equivalently, the point process, say  $\tilde{N}_p(\cdot)$ , resulting from both thinning and scale-contraction, has  $\tilde{N}_p(A) = k$  only if from the original process exactly  $k$  of the  $N(p^{-1}A)$  points in the set  $p^{-1}A$  are retained in the thinning process.

**Proposition 11.3.I.** *Let  $\tilde{N}_p(\cdot)$  denote the sequence of point processes obtained by independent thinning and contraction at rate  $p = 1/T$  from a point process  $N(\cdot)$  on  $\mathcal{X} = \mathbb{R}$ , and let  $N_\infty$  denote a stationary Poisson process at rate  $\lambda$ . Then*

$$\tilde{N}_p(\cdot) \rightarrow N_\infty \quad \text{weakly}$$

*if and only if as  $p \rightarrow 0$ , for every bounded  $A \in \mathcal{B}_\mathcal{X}$ ,*

$$pN(p^{-1}A) \equiv (1/T)N(TA) \rightarrow \lambda \ell(A) \quad \text{in probability.} \quad (11.3.1)$$

PROOF. For independent thinnings, the mechanics are most easily described in terms of p.g.fl.s. Indeed, the thinned process can be regarded as an especially simple form of cluster process, in which each of the points of the original process may be regarded as the centre of a cluster, and the cluster itself is either empty (if the point is deleted) or has just one point at the site of the cluster centre. Suppose first that we are given a general thinning function  $p(x)$ . Then for  $h \in \mathcal{V}(\mathbb{R})$ , the p.g.fl. of the cluster member process, given a centre at  $y$ , and in the notation of equations (6.3.6) and (6.3.7), is  $G_m[h \mid y] = p(y)h(y) + 1 - p(y)$ . Thus, the p.g.fl.  $G_{p(\cdot)}[h]$  can be written in the abbreviated notation

$$G_{p(\cdot)}[h] = G[1 - p + ph], \quad (11.3.2)$$

where  $G$ , the p.g.fl. of the original process, here plays the role of the p.g.fl. of the cluster centre process. In particular, it follows easily from this representation, that a Poisson process remains Poisson after independent deletions [see Exercise 11.3.1(a)].

Now suppose that the deletion function  $p(x) \equiv p$  (all  $x$ ), and denote by  $\tilde{G}_p$  the p.g.fl. of the point process  $N_p$  after deletion and rescaling. From equation (11.3.2) we obtain

$$\begin{aligned}\tilde{G}_p[h] &= E\left(\exp \int_{\mathcal{X}} \log h(x) N_p(dx)\right) \\ &= E\left(\exp \int_{\mathcal{X}} \log (1 - p[1 - h(x/p)]) N(dx)\right) \\ &= E\left(\exp \int_{\mathcal{X}} \log (1 - p[1 - h(x)]) N(dx/p)\right).\end{aligned}$$

The logarithmic term here equals  $-p[1 - h(x)](1 + O(p))$  for  $p \downarrow 0$ , so, using continuity of the generating function with respect to convergence in probability,

$$\begin{aligned}\tilde{G}_p[h] &= E\left[\exp\left(-\int_{\mathcal{X}} [1 - h(x)]p(1 + O(p)) N(dx/p)\right)\right] \\ &\rightarrow \exp\left(-\int_{\mathcal{X}} \lambda[1 - h(x)] \ell(dx)\right)\end{aligned}$$

if and only if (11.3.1) holds.  $\square$

An equivalent proof in terms of the convergence of one-dimensional distributions and using the characterization result of Theorem 9.2.XII is given in Westcott (1976) [see Exercise 11.3.2(a) and Proposition 11.1.IX]. Proof in the case of a renewal process is simpler [see Example 4.3(a) and Exercise 11.3.1(b)].

Equation (11.3.1) requires the individual realizations to satisfy an almost sure averaging property with a deterministic limit; if instead of (11.3.1) we have

$$pN(p^{-1}A, \omega) \rightarrow \lambda(\omega)\ell(A) \quad \text{in probability} \quad (11.3.3)$$

for some r.v.  $\lambda(\cdot)$  defined on the space  $(\Omega, \mathcal{F}, \mathcal{P})$  on which  $N$  is defined [and, implicitly,  $(\Omega, \mathcal{F}, \mathcal{P})$  is assumed to be large enough to embrace the independent thinning process], the conclusion of Proposition 11.3.I is modified as below. This in turn is a special case of Theorem 11.3.III, so the proof is omitted.

**Proposition 11.3.II.**  $\tilde{N}_p(\cdot)$  converges weakly to a mixed Poisson process, with mixing random variable  $\lambda$ , if and only if (11.3.3) holds.

The formulation at (11.3.1) or (11.3.3) specifies a particular form of the measure approximated by  $\lambda\ell(A)$ , namely,  $pN(p^{-1}A)$ . An alternative approach is simply to postulate the existence of a sequence of point processes  $\{N_n(\cdot)\}$

such that, given a sequence of thinning probability functions  $\{p_n(x)\}$  satisfying

$$0 \leq p_n(x) \leq 1 \quad (x \in \mathcal{X}, n = 1, 2, \dots) \quad (11.3.4a)$$

and

$$\sup_{x \in \mathcal{X}} p_n(x) \rightarrow 0 \quad (n \rightarrow \infty), \quad (11.3.4b)$$

the sequence of random measures  $\Lambda_n$ , where  $\Lambda_n(A) = \int_A p_n(x) N_n(dx)$ , then satisfies

$$\Lambda_n \rightarrow \Lambda \quad \text{weakly} \quad (11.3.5)$$

for some limit random measure  $\Lambda(\cdot)$ . Here, we may also allow the functions  $p_n(\cdot)$  to be stochastic, subject to the constraints at (11.3.4) and (11.3.5).

Note that the operation of ‘scale-contraction’ needs care when the space  $\mathcal{X} = \mathbb{R}^d$  say: the independent thinning operation implies that the expectation measure  $M(\cdot)$  of  $N$  becomes  $pM(\cdot)$  after thinning, so we should look for convergence of  $pM(A/p^{1/d})$  in order to obtain a nontrivial  $d$ -dimensional analogue of the following basic result (see Exercise 11.3.4).

**Theorem 11.3.III.** *Let  $\{p_n(x): x \in \mathcal{X}, n = 1, 2, \dots\}$  be a sequence of measurable stochastic processes satisfying (11.3.4),  $\{N_n: n = 1, 2, \dots\}$  a sequence of point processes, and  $\tilde{N}_n$  the process obtained from  $N_n$  by independent thinning according to  $p_n$ . Then there exists a point process  $N$  for which*

$$\tilde{N}_n \rightarrow N \quad \text{weakly}$$

*if and only if (11.3.5) holds for some random measure  $\Lambda$ , in which case  $N$  is the Cox process directed by  $\Lambda$ .*

The statement in this theorem allows for more general limits than Proposition 11.3.II precisely because no construction of the ‘increasingly dense’ processes  $N_n(\cdot)$  is specified. In the context of Proposition 11.3.II, it is not possible to have any Cox process other than the mixed Poisson process because when  $pN(p^{-1}A, \omega) \rightarrow \Lambda(A, \omega)$ , say, as  $p \rightarrow 0$ , then also  $p_1 p_2 N((p_1 p_2)^{-1}A, \omega) \rightarrow p_2 \Lambda(p_2^{-1}A, \omega)$  as  $p_1 \rightarrow 0$ , and taking (for example)  $\mathcal{X} = \mathbb{R}$  and  $A = (0, 1]$ , we then have  $\Lambda((0, p_2^{-1}], \omega) = p_2^{-1} \Lambda((0, 1], \omega)$  for all  $0 < p_2 < 1$ , from which it follows that  $\Lambda(\cdot, \omega)$  coincides with  $\lambda(\omega)\ell(\cdot)$  for some r.v.  $\lambda(\cdot)$ .

PROOF. We have for the p.g.fl.  $\tilde{G}_n$  of  $\tilde{N}_n$ , with  $h \in \mathcal{V}$ ,

$$\tilde{G}_n[h] = E \left( \exp \int_{\mathcal{X}} \log(1 - p_n(x)[1 - h(x)]) N_n(dx) \right).$$

When equations (11.3.4) are satisfied we can write

$$-\log(1 - p_n(x)[1 - h(x)]) = p_n(x)[1 - h(x)][1 + R_n(x)],$$

where  $|R_n(x)| \leq \frac{1}{2}p_n(x)$  and  $\theta_n = \sup_{x \in \mathcal{X}} |R_n(x)| \rightarrow 0$  as  $n \rightarrow \infty$ . We can therefore write

$$-\int_{\mathcal{X}} \log(1 - p_n(x)[1 - h(x)]) N_n(dx) = \int_{\mathcal{X}} [1 - h(x)](1 + R_n(x)) \Lambda_n(dx).$$

If now (11.3.5) holds, then the random variables  $\int_{\mathcal{X}} [1 - h(x)] \Lambda_n(dx)$  converge in distribution to  $\int_{\mathcal{X}} [1 - h(x)] \Lambda(dx)$ , so that their Laplace transforms, and hence also the p.g.f.s  $\tilde{G}_n[h]$ , converge to the Laplace transform of their limit, namely,

$$\tilde{G}_n[h] \rightarrow E\left(\exp\left[-\int_{\mathcal{X}} [1 - h(x)] \Lambda(dx)\right]\right). \quad (11.3.6)$$

The right-hand side here is just the p.g.f. of the Cox process directed by  $\Lambda$ , which completes the proof that (11.3.5) is sufficient.

Suppose conversely that the point processes  $\tilde{N}_n$  converge. We first establish that the random measures  $\Lambda_n$  are weakly compact. Referring to Proposition 11.1.V, let  $\bar{S}$  be a closed sphere in  $\mathcal{X}$ , and consider the random variables  $\Lambda_n(\bar{S})$ . Weak convergence of the sequence  $\{\tilde{N}_n\}$  implies  $\tilde{G}_n[h] \rightarrow G_\infty[h]$ , say, for  $h \in \mathcal{V}$ , and in particular for  $h = h_z$ , where  $0 \leq z \leq 1$  and

$$h_z(x) = \begin{cases} z & (x \in \bar{S}), \\ 1 & (x \notin \bar{S}). \end{cases}$$

But, assuming (11.3.4), this is equivalent to convergence of the Laplace transforms

$$E(\exp[-(1 - z)\Lambda_n(\bar{S})])$$

to the limit  $G_\infty[h_z]$ , which is continuous in  $z$  as  $z \rightarrow 1$ . Then the continuity theorem for Laplace transforms implies that the limit is the Laplace transform of a proper distribution, and hence that the distributions of the random variables  $\Lambda_n(\bar{S})$  are uniformly tight. Thus, given  $\epsilon > 0$  we can find  $M < \infty$  such that for all  $n$ ,

$$\mathcal{P}\{\Lambda_n(\bar{S}) > M\} < \epsilon;$$

that is, (11.1.1) holds.

As for (11.1.2), if the  $\tilde{N}_n$  converge weakly then, given  $\eta > 0$ , there exists a compact  $C$  such that for  $n = 1, 2, \dots$ ,

$$\mathcal{P}\{\tilde{N}_n(\bar{S} - C) > 0\} < \eta$$

[i.e., we use the necessity of (11.1.2) for the point processes]. Now set  $h(x) = 0$  or 1 as  $x \in$  or  $\notin \bar{S} - C$ , and deduce as above that from (11.3.4),

$$1 - \mathcal{P}\{\tilde{N}_n(\bar{S} - C) > 0\} - E[\exp(-\Lambda_n(\bar{S} - C))] \rightarrow 0 \quad (n \rightarrow \infty).$$

Thus, for sufficiently large  $n$ ,

$$\mathbb{E}(1 - \exp[-\Lambda_n(\bar{S} - C)]) \leq \mathcal{P}\{\tilde{N}_n(\bar{S} - C) > 0\} + \eta \leq 2\eta. \quad (11.3.7)$$

But by a basic inequality, because  $1 - e^{-x}$  is nonnegative and monotonic,

$$\mathcal{P}\{\Lambda_n(\bar{S} - C) > \delta\} \leq \frac{\mathbb{E}[1 - e^{-\Lambda_n(\bar{S} - C)}]}{1 - e^{-\delta}} \leq \frac{2\eta}{1 - e^{-\delta}}.$$

Thus, no matter how small  $\delta$  and  $\epsilon$ , we can find  $C$  such that  $\mathcal{P}\{\Lambda_n(\bar{S} - C) > \delta\} < \epsilon$  for all sufficiently large  $n$ , and hence (by modifying  $C$  if necessary) for all  $n > 0$ .

Thus, both conditions (11.1.1) and (11.1.2) hold for the sequence  $\Lambda_n$ . It is now a simple matter to deduce from the convergence of the  $\tilde{G}_n$  that any limit random measure  $\Lambda$  must satisfy

$$\mathbb{E}\left[\exp - \int_{\mathcal{X}} (1 - h(x)) \Lambda(dx)\right] = G_{\infty}[h].$$

It follows that the limit  $\Lambda$  must be unique and that the Laplace functionals of the  $\Lambda_n$  converge to that of  $\Lambda$ , so that (11.3.5) holds.  $\square$

One of the important applications of point process methods, and of the concept of thinning in particular, is to the study of high-level crossings of a continuous stochastic process. We do not treat this topic in detail, for which see Leadbetter, Lindgren, and Rootzen (1983) and the earlier text by Cramér and Leadbetter (1967), apart from briefly indicating one possible approach as follows.

Consider a nonnegative, discrete time process  $\{X_n\}$  and associate with each  $n$  the point  $(n, X_n)$  of a marked point process in  $\mathbb{R} \times \mathbb{R}_+$ , where  $\mathbb{R}_+$  plays the role of the mark space  $\mathcal{K}$  in Definition 9.1.V. Let the underlying process of time points in  $\mathbb{R}$  be thinned by rejecting all pairs  $(n, X_n)$ , for which (say)  $X_n \leq M$ , and let the time axis be rescaled suitably. We may now seek conditions under which the rescaled process converges to a limit as  $M \rightarrow \infty$ . In general, the thinnings here are not independent and the resulting process may exhibit substantial clustering properties. With suitable precautions, however, and assuming some asymptotic independence or mixing conditions, we may anticipate convergence to a Poisson limit.

A richer theory might be expected to result if one could retain the values of the marks accepted, albeit themselves rescaled in an appropriate manner. We now give an example of this kind, in the especially simple case where the marks  $\{X_n\}$  are i.i.d.; yet another approach to dependent thinnings, where the probability of thinning is allowed to depend on the previous history, is outlined in Proposition 14.2.XI.



EXAMPLE 11.3(a). *Thinning by the tails of a regularly varying distribution.* With the set-up just described, suppose that the initial points  $\{t_i\}$  form a stationary, ergodic process  $N_0$  with finite mean rate  $m$  and that the marks  $X_i$  are i.i.d. with regularly varying tails [see, e.g., Feller (1971, Section VIII.8) or Bingham, Goldie, and Teugels (1987)], so that for some  $\alpha > 0$ ,

$$1 - F(x) = L(x)x^{-\alpha} \quad (x \rightarrow \infty),$$

where  $L(x)$  is slowly varying at infinity; that is,  $L(cx)/L(x) \rightarrow 1$  for all  $c > 0$ .

Consider now a sequence of point processes on the space  $\mathbb{R} \times \mathbb{R}_+$  obtained in the following manner. For each  $n = 1, 2, \dots$ , set

$$N_n((t_1, t_2] \times (u, v]) = \#\{(t_i, x_i): nt_1 < t_i \leq nt_2 \text{ and } a_n u < x_i \leq a_n v\},$$

where the sequence of constants  $\{a_n: n = 1, 2, \dots\}$  is defined by

$$1 - F(a_n) = 1/n, \quad \text{equivalently, } a_n = F^{-1}(1 - 1/n),$$

and we assume for convenience that the distribution of  $F$  is continuous so the inverse  $F^{-1}$  is well defined.

Because the marks are independent, the p.g.fl. of the marked process on  $\mathbb{R} \times \mathbb{R}_+ = \mathcal{X} \times \mathcal{K}$  can be written, for suitable functions  $h(\cdot)$ , in the form

$$G[h] = E\left(\exp \int_{\mathbb{R}} \left\{ \log \left[ \int_{\mathbb{R}_+} h(t, y) dF(y) \right] \right\} N_0(dt)\right).$$

The function  $h$  here must of course lie in  $\mathcal{V}(\mathcal{X} \times \mathcal{K})$  (see Proposition 6.4.IV), but, because the limit point process is boundedly finite only in subsets of the mark space bounded away from the origin,  $h$  should be equal to unity in a neighbourhood of the origin on the mark space  $\mathbb{R}_+$ . In effect, the metric in the mark space should be modified so that the origin on the  $y$ -axis becomes a point at  $\infty$ . With this modification the p.g.fl. theory carries through without change.

For the rescaled process, we have to consider

$$\begin{aligned} \int_{\mathbb{R}_+} h\left(\frac{t}{n}, \frac{y}{a_n}\right) dF(y) &= \int_{\mathbb{R}_+} h\left(\frac{t}{n}, y\right) dF(a_n y) \\ &= 1 - \int_0^\infty \left[1 - h\left(\frac{t}{n}, y\right)\right] dF(a_n y). \end{aligned}$$

The assumption of regular variation of  $F$  is equivalent to the weak convergence of the measures  $nF(a_n \cdot)$  to the measure  $\nu$  defined by  $\nu(y, \infty) = y^{-\alpha}$ , because for any interval  $(u, v]$  with  $0 < u < v < \infty$ ,

$$\begin{aligned} n \int_u^v dF(a_n y) &= n[(1 - F(a_n u)) - (1 - F(a_n v))] \\ &= \frac{1 - F(a_n u)}{1 - F(a_n)} - \frac{1 - F(a_n v)}{1 - F(a_n)} = \frac{L(a_n u)}{L(a_n)} u^{-\alpha} - \frac{L(a_n v)}{L(a_n)} v^{-\alpha} \\ &\rightarrow u^{-\alpha} - v^{-\alpha} \quad (n \rightarrow \infty). \end{aligned}$$

Consequently, the innermost integral in the expression for  $G[h]$  is expressible as

$$\int_0^\infty h\left(\frac{t}{n}, \frac{y}{a_n}\right) dF(y) = 1 - n^{-1}(1 + o(1)) \int_0^\infty \left[1 - h\left(\frac{t}{n}, y\right)\right] d\nu(y);$$

thus, the p.g.fl. of the rescaled process,  $G_n$  say, is given by

$$\begin{aligned} G_n[h] &= E \exp \left( \int_{\mathbb{R}} \left\{ \log \left[ \int_{\mathbb{R}_+} h\left(\frac{t}{n}, \frac{y}{a_n}\right) F(dy) \right] \right\} N_0(dt) \right) \\ &= E \exp \left( \int_{\mathbb{R}} n^{-1}(1 + o(1)) \int_{\mathbb{R}_+} \left[1 - h\left(\frac{t}{n}, y\right)\right] \nu(dy) N_0(dt) \right) \\ &\rightarrow E \exp \left( \int_{\mathbb{R}} \int_{\mathbb{R}_+} (1 - h(u, y)) \nu(dy) m du \right) \quad \text{as } n \rightarrow \infty, \end{aligned}$$

using the ergodicity of  $N_0$ . The limit process is thus a Poisson process on  $\mathbb{R} \times \mathbb{R}_+$  with intensity measure  $m\ell(\cdot) \times \nu$ . For any  $c > 0$ , the limit process restricted to points with marks above  $c$  is a compound Poisson process where the marks are distributed on  $(c, \infty)$  according to the distribution with d.f.  $F_c(x) = 1 - (x/c)^{-\alpha}$ .

Strictly speaking, the overall process is not a compound Poisson process as defined in Section 6.4, because the ground process is not boundedly finite. Such extended compound Poisson processes appear also in the discussion of stable random measures (see Section 10.2) and self-similar MPPs in Section 12.8. For further examples and applications see Resnick (1986, 1987).  $\square$

We now indicate how the convergence in Theorem 11.3.III can be strengthened. The theorem is the weak convergence of the fidi distributions: we prove the stronger property that the fidi distributions converge in variation norm, and at the same time provide a bound on the rate of convergence.

For probability measures  $\mathcal{P}_1, \mathcal{P}_2$  on the measurable space  $(\Omega, \mathcal{E})$  the *variation metric*  $d(\mathcal{P}_1, \mathcal{P}_2)$  can be defined by

$$d(\mathcal{P}_1, \mathcal{P}_2) = \sup_{B \in \mathcal{E}} |\mathcal{P}_1(B) - \mathcal{P}_2(B)|.$$

This metric has the probabilistic interpretation that

$$d(\mathcal{P}_1, \mathcal{P}_2) = \inf \Pr\{\omega: X(\omega) \neq Y(\omega)\},$$

where the infimum is taken over all pairs of measurable functions  $X, Y$  on  $(\Omega, \mathcal{E}, \Pr)$  inducing the measures  $\mathcal{P}_1, \mathcal{P}_2$ , respectively. A pair  $(X, Y)$  for which equality holds constitutes a maximal coupling for the probability measures  $\mathcal{P}_1, \mathcal{P}_2$ . The metric  $d(\cdot, \cdot)$  differs by a factor of 2 from the variation distance because, in the notation of Appendix A1.3 where it is defined,

$$\begin{aligned} V_{\mathcal{P}_1 - \mathcal{P}_2} &= \|\mathcal{P}_1 - \mathcal{P}_2\| = \int_{\Omega} |\mathcal{P}_1(d\omega) - \mathcal{P}_2(d\omega)| \\ &= \sup_{\mathcal{T}(\Omega)} \sum_1^{n(\mathcal{T})} |\mathcal{P}_1(A_i) - \mathcal{P}_2(A_i)| = 2d(\mathcal{P}_1, \mathcal{P}_2). \end{aligned}$$

This notation  $\|\cdot\|$  is as in MKM (1978) where  $\text{Var}(\cdot) = \|\cdot\|$  is also used.

Because the limit r.v. in Theorem 11.3.III is Poisson, our concern is with  $d(\mathcal{P}_n, \mathcal{P}_\infty)$ , where the limit probability measure  $\mathcal{P}_\infty$  is Poisson and nonatomic. The Renyi–Mönch Theorem 9.2.XII asserts that such measures are characterized by their one-dimensional distributions, and by Proposition 11.1.IX it is then enough here to consider the quantity

$$d(N_n(A), N(A)) \equiv d(\Pr\{N_n(A) \in \cdot\}, \Pr\{N(A) \in \cdot\})$$

for any bounded Borel set  $A$  [we abuse the notation in replacing the distributions of r.v.s in  $d(\cdot, \cdot)$  by the r.v.s themselves]. Furthermore, for nonnegative integer-valued r.v.s,  $X, Y$  say, with distributions  $\{p_k\}, \{q_k\}$  say, we have

$$\begin{aligned} d(X, Y) = d(\{p_k\}, \{q_k\}) &= \sup_{A \subset \mathbb{Z}_+} \left| \sum_{k \in A} (p_k - q_k) \right| = \frac{1}{2} \sum_{k=0}^{\infty} |p_k - q_k| \\ &= \sum_{k=0}^{\infty} (p_k - q_k)_+. \end{aligned}$$

**Proposition 11.3.IV.** *In the setting of Theorem 11.3.III, for bounded  $A \in \mathcal{B}_X$ ,*

$$d(\tilde{N}_n(A), N(A)) \leq E \left( \sup_{x \in A} |p_n(x)| + [1 - \exp(-|\Lambda_n(A) - \Lambda(A)|)] \right).$$

PROOF. Observe that it is enough to prove the result in the context that the functions  $p_n$  and measures  $\Lambda_n, \Lambda$  are deterministic, for if otherwise, describe these entities as functions of  $\omega'$  with distribution  $\mu(\cdot)$ . Then

$$\begin{aligned} d(N_{p_n}(A), N(A)) &= \sup_B |\mathcal{P}_1(B) - \mathcal{P}_2(B)| \\ &= \sup_B \left| \int_{\Omega} (\mathcal{P}_1(B; \omega') - \mathcal{P}_2(B; \omega')) \mu(d\omega') \right| \\ &\leq \int_{\Omega} \sup_B |\mathcal{P}_1(B; \omega') - \mathcal{P}_2(B; \omega')| \mu(d\omega') \\ &= \int_{\Omega} d(N_{p_n}(A), N(A) \mid \omega') \mu(d\omega'). \end{aligned}$$

The first term in the bound comes from Lemma 11.3.V below, and the second comes from the fact that for Poisson r.v.s  $X, Y$  with means  $\lambda, \mu$ ,  $d(X, Y) \leq d(0, Z) = 1 - e^{-|\lambda - \mu|}$ , where  $Z$  is Poisson with mean  $|\lambda - \mu|$ .  $\square$

The second term in the bound can be tightened: see Exercise 11.3.5.

**Lemma 11.3.V.** *Let  $X_1, \dots, X_n$  be independent Bernoulli r.v.s with  $p_j = \Pr\{X_j = 1\} = 1 - \Pr\{X_j = 0\}$ , and  $Y$  a Poisson r.v. with mean  $\lambda = \sum_{j=1}^n p_j$ .*

Then the distributions of  $Y$  and  $S = \sum_{j=1}^n X_j$  have variation distance  $d(S, Y)$  bounded as in

$$d(S, Y) \leq \frac{C \sum_{j=1}^n p_j^2}{\sum_{j=1}^n p_j} \leq C \max_j \{p_j\}, \quad (11.3.8)$$

where  $C \leq 0.71$  when  $\max_j \{p_j\} \leq 0.25$ . Always,  $C \leq 1$ .

*Remarks.* Although Stein–Chen methods have been much used to prove results like this lemma, we give a direct approach based on Fourier transforms. The method of proof below draws in part on work of Samuels (1965) and Kerstan (1964a). It shows that the value of the constant  $C$  depends on the value of  $\max\{p_j\}$  and can be reduced further by supposing that the maximum is smaller than 0.25. Its smallest value, when the maximum  $\rightarrow 0$ , is equal (by the method below) to 0.409, which is tighter than that quoted by Romanowska (1978) for the more restricted case of a simple binomial approximated by a Poisson. Denoting the middle term in (11.3.7) by  $C\varpi$ , the computations in the later part of the proof below can be tightened further by retaining  $\xi$  in (11.3.8) as a function of  $\theta$  and integrating numerically; then 0.61 can replace  $0.70789 \approx 0.71$  in the theorem. The general result that  $C \leq 1$  follows from work of Barbour and Hall (1984) and is not discussed here.

PROOF. The sum  $S$  has a Poisson binomial distribution,  $\{b_k\}$  say, for which the generating function is

$$\sum_{k=0}^n b_k z^k = \prod_{j=1}^n (1 - p_j + p_j z) \quad (|z| \leq 1).$$

An inequality of Newton used in Samuels (1965) implies that  $c_k = b_k / \binom{n}{k}$  is a log concave sequence; that is,  $c_k^2 \geq c_{k-1} c_{k+1}$  for  $k = 0, 1, \dots, n$ ; equivalently,

$$b_k^2 \geq (1 + (n - k)^{-1})(1 + k^{-1})b_{k-1}b_{k+1}$$

for  $k = 1, \dots, n - 1$ . Write  $\pi_k = \pi_k(\lambda) \equiv e^{-\lambda} \lambda^k / k!$ . Then

$$d(S, Y) = \sum_{k=0}^n \{b_k - \pi_k\}_+ = \sum_{k=0}^n \pi_k (b_k / \pi_k - 1)_+,$$

and this summation will involve nonzero terms on a single interval of integers,  $\{k_0 + 1, \dots, k_1\}$  say, if the sequence of ratios  $\{b_k / \pi_k\}$  is unimodal. For this it suffices that the ratio of ratios  $(b_k / \pi_k) / (b_{k-1} / \pi_{k-1})$ , which equals  $k b_k / \lambda b_{k-1}$ , be monotonic in  $k$ , and this is implied by the corollary to the Newton inequality (see Exercise 11.3.6). Consequently, the sup distance

$$\begin{aligned} d_0(S, Y) &\equiv \sup_k |\Pr\{S \leq k\} - \Pr\{Y \leq k\}| \\ &= \max \left( \sum_{k=0}^{k_0} (\pi_k - b_k), \sum_{k=k_1+1}^n (\pi_k - b_k) \right). \end{aligned}$$

By addition, we thus have  $2d_0(S, Y) \geq d(S, Y) \geq d_0(S, Y)$ , so to bound the variation distance  $d$ , it suffices to bound the sup distance  $d_0$ .

The Fourier inversion relation gives

$$\begin{aligned} d_0(S, Y) &= \sup_k \left| \frac{1}{2\pi i} \int_{-\pi}^{\pi} e^{-i\theta k} \frac{E(e^{i\theta S}) - E(e^{i\theta Y})}{1 - e^{i\theta}} d\theta \right| \\ &\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|E(e^{i\theta S}) - E(e^{i\theta Y})|}{|2 \sin \frac{1}{2}\theta|} d\theta. \end{aligned}$$

The characteristic functions here are  $E(e^{i\theta Y}) = \exp[-\lambda(1 - e^{i\theta})]$  and

$$\begin{aligned} E(e^{i\theta S}) &= \prod_{j=1}^n (1 - p_j(1 - e^{i\theta})) \\ &= E(e^{i\theta Y}) \prod_{j=1}^n (\exp[p_j(1 - e^{i\theta})])[1 - p_j(1 - e^{i\theta})]. \end{aligned}$$

Each term in the product here is of the form

$$\begin{aligned} [1 - p(1 - e^{i\theta})] \exp[p(1 - e^{i\theta})] &= 1 - \sum_{k=2}^{\infty} \frac{[p(1 - e^{i\theta})]^k (k-1)}{k!} \\ &\equiv 1 - p^2(1 - e^{i\theta})^2 f(p; \theta) \quad \text{say,} \end{aligned}$$

so

$$\begin{aligned} |E(e^{i\theta S}) - E(e^{i\theta Y})| &= |E(e^{i\theta Y})| \left| \prod_{j=1}^n [1 - p_j^2(1 - e^{i\theta})^2 f(p_j; \theta)] - 1 \right| \\ &= e^{-\lambda(1 - \cos \theta)} \left| \sum_{r=1}^n [-(1 - e^{i\theta})^2]^r \left[ \sum_{j_1 < \dots < j_r} (p_{j_k}^2 f(p_{j_k}; \theta)) \right] \right| \\ &\leq e^{-\lambda(1 - \cos \theta)} \left( \prod_{j=1}^n (1 + 2(1 - \cos \theta) p_j^2 |f(p_j; \theta)|) - 1 \right) \\ &\leq e^{-\lambda(1 - \cos \theta)} (\exp[\lambda \xi(\theta)(1 - \cos \theta)] - 1), \end{aligned}$$

where

$$\xi(\theta) = \frac{2 \sum_{j=1}^n p_j^2 |f(p_j; \theta)|}{\sum_{j=1}^n p_j}$$

and

$$|f(p; \theta)| \leq \sum_{k=2}^{\infty} \frac{|2p \sin \frac{1}{2}\theta|^{k-2} (k-1)}{k!} = \frac{1 - (1 - \psi)e^{\psi}}{\psi^2},$$

with  $\psi = |2p \sin \frac{1}{2}\theta|$ . This bound is a maximum for  $|\theta| \leq \pi$  at  $\theta = \pi$  where it equals  $(1 - (1 - 2p)e^{2p})/4p^2$ , which increases monotonically with  $p$  in  $p > 0$ .

Write  $\xi = \xi(\pi)$ . Then using  $d \leq 2d_0$  and the bounds above gives

$$\begin{aligned}
 d(S, Y) &\leq \frac{1}{\pi} \int_0^\pi \frac{e^{-\lambda(1-\cos\theta)} [e^{\lambda\xi(1-\cos\theta)} - 1]}{\sin \frac{1}{2}\theta} d\theta \\
 &= \frac{2\lambda}{\pi} \int_0^\pi \sin \frac{1}{2}\theta d\theta \int_0^\xi \exp[-2\lambda(1-y) \sin^2 \frac{1}{2}\theta] dy \\
 &= \frac{4\lambda}{\pi} \int_0^1 du \int_0^\xi \exp[-2\lambda(1-y)(1-u^2)] dy \\
 &\leq \frac{4\lambda}{\pi} \int_0^\xi dy \int_0^1 \exp[-2\lambda(1-y)(1-u)] du \\
 &\leq \frac{2}{\pi} \int_0^\xi \frac{dy}{1-y} = -\frac{2\log(1-\xi)}{\pi}, \tag{11.3.9}
 \end{aligned}$$

uniformly in  $\lambda$ . For  $\max_j \{p_j\} \leq 0.25$ ,  $\xi \leq 0.3513$ , and thus  $d(S, Y) \leq [(2\pi^{-1} \log(1/0.6487))/0.25]\varpi = 1.102\varpi$ , where we follow Le Cam (1960) in writing  $\varpi = \sum_{j=1}^n p_j^2 / \sum_{j=1}^n p_j$ . Alternatively, the inequality (11.3.8) can be replaced by the tighter bound based on

$$\sup_{\alpha > 0} \int_0^1 \alpha \exp[-\alpha(1-u^2)] du = c = 0.642374$$

(this supremum is attained at  $\alpha = 2.255$ ), leading to

$$d(S, Y) \leq \frac{2c}{\pi} \int_0^\xi \frac{dy}{1-y} = -\frac{2c\log(1-\xi)}{\pi} < \frac{2c\xi}{\pi(1-\xi)}.$$

For  $\max_j \{p_j\} \leq 0.25$ , this yields  $d(S, Y) \leq 0.70789\varpi$ , and for fixed  $n$ ,

$$\lim_{\max_j \{p_j\} \downarrow 0} d(S, Y) \leq \frac{2c\xi}{\pi} = 0.40895\xi. \quad \square$$

## Exercises and Complements to Section 11.3

11.3.1 (a) Verify that if a Poisson process in  $\mathcal{X} = \mathbb{R}^d$  with intensity measure  $\Lambda(\cdot)$  is subjected to independent thinning with retention function  $p(x)$  for measurable  $p(\cdot)$ , then the thinned process is Poisson with intensity measure  $\Lambda_p$ , where

$$\Lambda_p(A) = \int_A p(x) \Lambda(dx).$$

(b) When  $N$  in Proposition 11.3.I is a renewal process, use *inter alia* transform techniques to show the following [cf. Rényi (1956)].

- (i) The rarefaction of any renewal process is a renewal process.
- (ii) The only renewal process invariant under the operations is the Poisson.
- (iii) Starting from any renewal process with finite mean lifetime, the limit under the operations is a Poisson process.

- 11.3.2 (a) Apply Proposition 11.1.IX to furnish an alternative proof of Proposition 11.3.I via avoidance functions.
- (b) Give a direct proof of Proposition 11.3.II either via p.g.f.s or by extension of part (a).
- 11.3.3 Use Theorem 11.3.III to show that a distribution  $\mathcal{P}$  on  $\mathcal{N}_{\mathcal{X}}^{\#}$  is a Cox distribution if and only if for each  $c$  in  $0 < c < 1$  there exists a distribution  $\mathcal{P}_c$ , which under independent random thinning with constant probability  $c$  yields  $\mathcal{P}$  [see Mecke (1968)].
- 11.3.4 *Extensions of the prototype limit result for thinnings.*
- (a) Formulate statements analogous to Proposition 11.3.I when  $N$  is stationary but instead of simple it is (i) nonorderly and either (i') batches of points are treated *in toto*, with all points retained with probability  $p$  and otherwise all deleted; (i'') each point of a batch is thinned independently (with retention probability  $p$  per point); or (ii) a marked point process. In case (ii) does the limit process have independent marks?
- (b) Consider the case  $\mathcal{X} = \mathbb{R}^d$  for some  $d \geq 2$  in which  $N_n$  is as stated in Proposition 11.3.I except that any  $A \in \mathcal{B}_{\mathcal{X}}$  is now rescaled by a factor  $n^{1/d}$ . What conditions on  $N$  suffice for the weak convergence of  $N_n$ ? How are the results of (a) affected by having  $\mathcal{X} = \mathbb{R}^d$ ?
- 11.3.5 (a) Show that  $a_k \equiv \sup_{\lambda > 0} \pi_k(\lambda) \sqrt{\lambda}$  occurs for  $\lambda = k + \frac{1}{2}$  and that the sequence of ratios  $\{a_{k+1}/a_k\}$  is monotonic in  $k$ . Show that  $a_0 = 1/\sqrt{2e\lambda} \geq a_k > a_{k+1}$  for  $k = 0, 1, \dots$ , and that  $\pi_k(\lambda) \leq 1/\sqrt{2e\lambda}$  (every  $\lambda > 0$ ).
- (b) Express  $d(\{\pi_k(\lambda)\}, \{\pi_k(\mu)\})$  as an integral with density  $\pi_j(\cdot)$  for some  $j$ , and hence deduce that this variation metric is bounded above by  $\sqrt{2/e} |\sqrt{\lambda} - \sqrt{\mu}|$ . Thus, except for small  $\Lambda_n, \Lambda$ , the last term in Proposition 11.3.IV can be tightened [see Daley (1987)].
- 11.3.6 (a) Let  $b_k(n; p) = \binom{n}{k} p^k (1-p)^{n-k}$ ,  $k = 0, 1, \dots, n$ ,  $0 < p < 1$ , denote binomial probabilities, and write  $\pi_k(\lambda)$  for Poisson probabilities as in the proof of Lemma 11.3.V. Show directly that  $\{b_k(n; p)/\pi_k(\lambda)\}$  is a unimodal sequence in  $k$  by virtue of the monotonicity of  $(k+1)b_{k+1}/b_k$ .
- (b) Now let  $\{b_k\} = \{b_k(n; p_1, \dots, p_n)\}$  denote the Poisson binomial distribution of  $S$  of the lemma. Show by induction on  $n$  that  $\{(k+1)b_{k+1}/b_k\}$  is monotonic in  $k$ .
- (c) Deduce that  $d(S, Y) \leq \max_k \{b_k/\pi_k\} - 1$ .

## 11.4. Random Translations

The remaining class of stochastic operations that we consider has as its prototype *random translations*: each point  $x_i$  in the realization of some initial point process  $N_0$  is shifted independently of its neighbours through a random vector  $Y_i$ , the  $Y_i$  forming a sequence of i.i.d. random variables. Exercises 2.3.4(b) and 8.2.7 give simple cases. Generalizations occur if the translations are replaced by more general clustering mechanisms in which the mean number of points per cluster is held equal to one: we defer discussion to Section 13.5 in order to take advantage of some properties of Palm distributions.

One possible approach to such problems is to view the resultant process as the superposition of its individual clusters, one from each point of the initial realization, and to seek to apply the results of Section 11.2 on triangular arrays. Here the  $n$ th row in the array relates to the process derived from  $n$  stages of clustering (translation); although the number of terms  $N_{ni}$  in the  $n$ th row is infinite, this does not affect the validity of the criteria provided their superpositions are well defined (see Exercise 11.2.3), as they must be in this case for the resultant processes themselves to be well defined.

In the case of random translations in  $\mathbb{R}^d$  this leads us to a result of the following kind. Let  $\nu(\cdot)$  denote the common distribution on  $\mathcal{B}(\mathbb{R}^d)$  for the translations  $Y_i$ , and  $\nu_n(\cdot)$  the  $n$ -fold convolution of  $\nu$ , corresponding therefore to the effect of  $n$  successive random translations. For  $h \in \mathcal{V}$  and  $\mathcal{X} = \mathbb{R}^d$ , the p.g.fl. after  $n$  translations takes the form

$$\begin{aligned} G_n[h] &= G_0 \left[ \int_{\mathcal{X}} h(\cdot + y) \nu_n(dy) \right] \\ &= E \left( \exp \left[ \int_{\mathcal{X}} \log \left( \int_{\mathcal{X}} h(x + y) \nu_n(dy) \right) N_0(dx) \right] \right), \end{aligned} \quad (11.4.1)$$

where  $G_0[\cdot]$  is the p.g.fl. of the initial process  $N_0$ .

Each process  $N_{ni}$  contains just one point, and the u.a.n. condition (11.2.2) reduces to the requirement

$$\sup_i \nu_n(A + x_i) = \sup_i \Pr\{Y_1 + \cdots + Y_n \in x_i + A\} \rightarrow 0 \quad (n \rightarrow \infty). \quad (11.4.2)$$

To use Theorem 11.2.V to prove convergence to a Poisson limit, observe that condition (11.2.7) is trivial here because  $\mathcal{P}\{N_{ni}(A) \geq 2\} \equiv 0$ , and (11.2.8) translates into

$$\int_{\mathcal{X}} \nu_n(A + x) N_0(dx) = \sum_i \nu_n(A + x_i) \rightarrow \mu(A) \quad (n \rightarrow \infty). \quad (11.4.3)$$

The problem then is to find conditions on the initial process  $N_0$  and the distribution  $\nu$  that will ensure the truth of (11.4.2) and (11.4.3). The former of these is closely related to the concept of the *concentration function*  $Q_A(F)$  of a distribution  $F$  on  $\mathcal{B}(\mathbb{R}^d)$  defined for bounded Borel sets  $A$  by

$$Q_A(F) = \sup_{x \in \mathbb{R}^d} F(x + A).$$

Clearly the expression in (11.4.2) is bounded above by  $Q_A(\nu_n)$ , so the u.a.n. statement there is a direct consequence of the lemma below [for a non-Fourier analytic proof see Ibragimov and Linnik (1971, Chapter 15, Section 2)].

**Lemma 11.4.I.** *Let  $F$  be a distribution on  $\mathcal{B}(\mathbb{R}^d)$  and  $F^{n*}$  its  $n$ th convolution power. For bounded  $A$ ,  $Q_A(F^{n*}) \rightarrow 0$  as  $n \rightarrow \infty$  if and only if the support of  $F$  contains at least two distinct points.*



PROOF. If  $F$  is degenerate, then so is  $F^{n*}$  and  $Q_A(F) = Q_A(F^{n*}) = 1$  for every  $n$  for every nonempty  $A$ . Otherwise, we show in fact that

$$Q_A(F^{n*}) \leq c(F, A)/n^{1/2} \quad (11.4.4)$$

for some constant  $c(F, A)$ , observing that it is enough to prove this in the case  $d = 1$  because for  $d \geq 2$ , noting that  $Q_A(F)$  increases for monotonic increasing  $A$ , we can embed  $A$  in a set corresponding to a marginal distribution. Exercise 11.4.2 shows that the order  $n^{-1/2}$  of the bound in (11.4.4) is tight.

Write  $\tilde{H}(y) = (\sin \frac{1}{2}y/\frac{1}{2}y)^2$  for the c.f. of the probability measure  $H(\cdot)$  with triangular density function  $H'(x) = (1 - |x|)_+$ . Then the Parseval relation (A2.8.8) yields for any d.f.  $G$ , positive  $a$ , and real  $\gamma$ ,

$$\int_{-\infty}^{\infty} \tilde{H}(a(x - \gamma)) G(dx) = \frac{1}{a} \int_{-\infty}^{\infty} \tilde{G}(\omega) e^{-i\omega\gamma} \left(1 - \frac{|\omega|}{a}\right)_+ d\omega. \quad (11.4.5)$$

Substitute  $G = F^{n*}$ , and recognize that the integral on the right-hand side here is over the interval  $(-a, a)$ , and the left-hand side is real, so that an upper bound that is uniform in  $\gamma$  is given by

$$\frac{1}{a} \int_{-a}^a |\tilde{F}(\omega)|^n d\omega \leq \frac{1}{a} \int_{-a}^a \exp\left(-\frac{1}{2}n(1 - |\tilde{F}(\omega)|^2)\right) d\omega,$$

where we have used the inequality  $x = 1 + (x - 1) < e^{x-1}$  with  $x = |\tilde{F}(\omega)|^2$ . Now  $|\tilde{F}|^2$  is the characteristic function of the d.f.  $F_s$  of the symmetrized r.v.  $X' - X''$ , where  $X'$  and  $X''$  are i.i.d. like  $X$  with d.f.  $F$ . Thus,

$$\begin{aligned} 1 - |\tilde{F}(\omega)|^2 &= 1 - E(\exp[i\omega(X' - X'')]) = 1 - E(\cos \omega(X' - X'')) \\ &= 2E(\sin^2 \frac{1}{2}\omega(X' - X'')) \geq 2 \int_{|y|>b} \sin^2 \frac{1}{2}\omega y F_s(dy) \end{aligned}$$

for some positive  $b$ , which, because  $F$  is nondegenerate, can be and is so chosen that

$$\int_{|y|>b} F_s(dy) = \Pr\{|X' - X''| > b\} \equiv \eta > 0.$$

Use Jensen's inequality in the form  $\exp(E[f(Y)]) \leq E(\exp[f(Y)])$  to write

$$\begin{aligned} \frac{1}{a} \int_{-a}^a |\tilde{F}(\omega)|^n d\omega &\leq \frac{1}{a} \int_{-a}^a d\omega \int_{|y|>b} \frac{\exp[-n\eta \sin^2 \frac{1}{2}\omega y]}{\eta} F_s(dy) \\ &= \frac{1}{a} \int_{|y|>b} \frac{F_s(dy)}{\eta|y|} \int_{|z|<a|y|} \exp[-n\eta \sin^2 \frac{1}{2}z] dz \\ &\leq (\text{const.}) \int_{|z|<ab} \exp[-n\eta \sin^2(z/2)] dz \leq \frac{(\text{const.})}{(n\eta)^{1/2}}, \end{aligned}$$

where the constant is independent of  $n$  but may depend on  $a$  and  $b$  (and hence on  $A$  and  $F$ ).

Refer back to (11.4.5) and observe that  $\tilde{H}(y) \geq 4/\pi^2$  for  $|y| < \pi$ , so the left-hand side there is bounded below by  $(4/\pi^2)[G(\gamma + \pi/a) - G(\gamma - \pi/a - 0)]$ .

Then

$$Q_A(F^{n*}) \leq \frac{\frac{1}{4}\pi^2(\text{const.})}{\sqrt{n\eta}} \quad \text{for any } A \subseteq \left[-\frac{\pi}{a}, \frac{\pi}{a}\right].$$

Fix  $a = \pi/\rho b$  for any  $\rho > 1$ , so that then for any  $A \subseteq [-\rho b, \rho b]$ , (11.4.4) holds. But  $b$  is positive, so we can choose  $\rho$  arbitrarily large, proving the lemma.  $\square$

Condition (11.4.3) is more difficult to realize, although it is certainly satisfied when the initial realization is sufficiently regular. Suppose, as an extreme example, that  $d = 1$  and that  $N_0$  consists of the lattice process with a point at every integer. We can then proceed via an application of the Poisson summation formula (see Exercise 8.6.4), which for sufficiently smooth functions  $g$  yields

$$\sum_{k=-\infty}^{\infty} (\nu_n * g)(k) = \sum_{j=-\infty}^{\infty} \tilde{\nu}_n(2\pi j) \tilde{g}(2\pi j),$$

where  $\tilde{\nu}_n(\omega) = \tilde{\nu}(\omega)^n$  is the Fourier–Stieltjes transform of  $\nu^{n*}$ ,  $\tilde{g}$  is the ordinary Fourier transform of the continuous integrable function  $g$ , and we assume  $g$  is so chosen that  $\tilde{g}$  is integrable and  $\sum_{j=-\infty}^{\infty} |\tilde{g}(2\pi j)|$  is convergent.

Recall that a distribution in  $\mathbb{R}^1$  is *nonlattice* if its distribution is not concentrated on a lattice (shifted multiple of the set of integers) in  $\mathbb{R}^1$ , and that its characteristic function then satisfies

$$|\tilde{\nu}(\omega)| < 1 \quad (\omega \neq 0). \quad (11.4.6)$$

Letting  $n \rightarrow \infty$  in (11.4.4) we obtain by dominated convergence

$$\sum_{k=-\infty}^{\infty} (\nu_n * g)(k) \rightarrow \tilde{\nu}_n(0) \tilde{g}(0) = \int_{\mathbb{R}} g(x) dx.$$

To show that this implies (11.4.3) for intervals it is enough to sandwich the indicator function of an interval  $A$  between two functions  $g_1, g_2$  with the properties required above, that is, in such a way that for given  $\epsilon > 0$ ,

$$\sup_x |g_2(x) - g_1(x)| < \epsilon \quad \text{and} \quad g_1(x) \leq I_A(x) \leq g_2(x)$$

[this can be achieved, e.g., by taking for  $g_2$  the function  $I_{A\epsilon/2} * t_{\epsilon/2}$  and for  $g_1$  the function  $I_{A-\epsilon/2} * t_{\epsilon/2}$ , where  $t_\alpha$  is the triangular distribution on base  $(-\alpha, \alpha)$ ]. Thus, (11.4.3) holds whenever  $\nu$  is nonlattice and  $A$  is an interval. Because it is enough in Theorem 11.2.V to let  $A$  run through intervals, we can conclude from that theorem that *if the initial process is lattice with a point on every integer but the distribution  $\nu$  is nonlattice, then the processes  $N_n$ , obtained by successive random translations according to  $\nu$ , converge weakly to the stationary Poisson process with unit rate.*

Obviously, the condition that the initial points lie on a lattice can be relaxed, but, unfortunately, it cannot be relaxed far enough to apply almost surely to the realizations of a typical stationary point process. Indeed, it follows from a theorem of Stone (1968) that the essential requirement on the initial process, if this is regarded as fixed, is that

$$N_0(x + A_n)/\ell(A_n) \rightarrow \text{const.} \quad \text{uniformly in } x, \quad (11.4.7)$$

where  $A_n$  is the sequence of hypercubes  $\mathbb{U}_n^d$  of side  $n$  in  $\mathbb{R}^d$  or, more generally, a convex averaging sequence in the sense of Definition 12.2.I. Such a condition is not satisfied almost surely even by the realizations of a stationary Poisson process. On the other hand, averaged forms of (11.4.5), that is, with convergence in  $L_1$  or  $L_2$ , follow directly from the ergodic theorems of Section 12.2, and in these the uniformity is trivial when the initial process is stationary. We therefore seek an alternative approach that will bypass the probability one requirements on the initial configuration.

For this purpose we return to a direct study of the p.g.fl. at (11.4.1) of the processes  $N_n$ . To establish weak convergence of the translated versions  $N_n$  to a Poisson limit with rate  $m$ , it is enough to show that for  $h \in \mathcal{V}$ ,

$$-\int_{\mathcal{X}} \log \left[ \int_{\mathcal{X}} h(x+y) \nu_n(dy) \right] N_0(dx) \xrightarrow{p} m \int_{\mathcal{X}} [1 - h(x)] dx,$$

because this implies convergence of the Laplace transforms of the random variables on the left-hand side, and hence of the p.g.fl.s.

To ease the notation write  $u(x) = 1 - h(x)$ , so that  $u$  vanishes outside a bounded set and satisfies  $0 \leq u \leq 1$ . Then the above requirement becomes

$$-\int_{\mathcal{X}} \log \left[ 1 - \int_{\mathcal{X}} u(x+y) \nu_n(dy) \right] N_0(dx) \xrightarrow{p} m \int_{\mathcal{X}} u(x) dx. \quad (11.4.8)$$

From Lemma 11.4.I, if  $\nu$  has at least two points in its support, we can easily deduce that

$$\theta_n \equiv \sup_x \int_{\mathcal{X}} u(x+y) \nu_n(dy) \rightarrow 0 \quad (n \rightarrow \infty).$$

We can therefore approximate the logarithm by its leading term, with remainder, for sufficiently large  $n$ , bounded by

$$\left| \int_{\mathcal{X}} u(x+y) \nu_n(dy) + \log \left( 1 - \int_{\mathcal{X}} u(x+y) \nu_n(dy) \right) \right| \leq \theta_n \int_{\mathcal{X}} u(x+y) \nu_n(dy).$$

Suppose now that  $N_0$  is stationary with finite mean rate  $m$ . Then (11.4.8) is implied by the corresponding  $L_1$  convergence, which leads us to estimate the expected difference by

$$\begin{aligned} & \mathbb{E} \left| m \int_{\mathcal{X}} u(x) dx + \int_{\mathcal{X}} \log \left( 1 - \int_{\mathcal{X}} u(x+y) \nu_n(dy) \right) N_0(dx) \right| \\ & \leq \mathbb{E} \left| m \int_{\mathcal{X}} u(x) dx - \int_{\mathcal{X}} \int_{\mathcal{X}} u(x+y) \nu_n(dy) N_0(dx) \right| \\ & \quad + \mathbb{E} \left| \int_{\mathcal{X}} \left\{ \int_{\mathcal{X}} u(x+y) \nu_n(dy) + \int_{\mathcal{X}} \log[1 - u(x+y)] \nu_n(dy) \right\} N_0(dx) \right|, \end{aligned}$$

where the second expectation is bounded by

$$\theta_n \mathbb{E} \left[ \int_{\mathcal{X}} \int_{\mathcal{X}} u(x+y) \nu_n(dy) N_0(dx) \right] = m \theta_n \int_{\mathcal{X}} u(x) dx,$$

which tends to zero by Lemma 11.4.I. Thus, for weak convergence to the Poisson limit it is enough to show that for measurable  $u$  with bounded support and  $0 \leq u \leq 1$ ,

$$\mathbb{E} \left| m \int_{\mathcal{X}} u(x) dx - \int_{\mathcal{X}} \int_{\mathcal{X}} u(x+y) \nu_n(dy) N_0(dx) \right| \rightarrow 0. \quad (11.4.9)$$

A complication arises here if the stationary distribution  $N_0$  is nonergodic, for the ergodic theorems then assert convergence not to the constant  $m$  but to the random variable (asymptotic density of  $N_0$ )

$$Y = \mathbb{E}[N_0(\mathbb{U}^d) \mid \mathcal{I}], \quad (11.4.10)$$

where  $\mathbb{U}^d$  is the unit cube in  $\mathbb{R}^d$ ,  $\mathcal{I}$  is the invariant  $\sigma$ -algebra, and  $\mathbb{E}(Y) = m$  (see Theorem 12.2.IV). In this case (11.4.7) should be replaced by the more general requirement, justified by a completely analogous argument, that

$$\mathbb{E} \left| Y \int_{\mathcal{X}} u(x) dx - \int_{\mathcal{X}} \int_{\mathcal{X}} u(x+y) \nu_n(dy) N_0(dx) \right| \rightarrow 0. \quad (11.4.11)$$

Note that (11.4.9) and (11.4.11) may be regarded as  $L_1$  versions of (11.4.3).

A full discussion of (11.4.11) involves further delicate analysis of the convolution powers of  $\nu$ ; we content ourselves here with the much easier  $L_2$  version, assuming the initial point process has bounded finite second moment measure. This leads us to the following theorem.

**Theorem 11.4.II.** *Let  $N_0$  be a second-order stationary point process on  $\mathcal{X} = \mathbb{R}^d$  and  $\nu$  a distribution on  $\mathbb{R}^d$  that is nonlattice. Then the sequence of point processes  $\{N_n\}$ , derived from  $N_0$  by successive random translations according to  $\nu$ , converges weakly to the stationary mixed Poisson process with p.g.fl.*

$$G[h] = \mathbb{E} \left[ \exp \left\{ -Y \int_{\mathcal{X}} [1 - h(x)] dx \right\} \right], \quad (11.4.12)$$

where  $Y$  is given by (11.4.10).

PROOF. We again use a Fourier argument, observing that in the ergodic case

$$\begin{aligned} & \mathbb{E} \left| m \int_{\mathcal{X}} u(x) dx - \int_{\mathcal{X}} \int_{\mathcal{X}} u(x+y) \nu_n(dy) N_0(dx) \right|^2 \\ &= \text{var} \left( \int_{\mathcal{X}} \int_{\mathcal{X}} u(x+y) \nu_n(dy) N_0(dx) \right) \\ &= \int_{\mathcal{X}} |\tilde{u}(\omega)|^2 |\tilde{\nu}(\omega)|^{2n} \Gamma(d\omega), \end{aligned} \quad (11.4.13)$$

where  $\Gamma(\cdot)$  is the Bartlett spectrum introduced in Definition 8.2.II. The validity of the above relation follows from the Parseval relation (8.6.10) and Lemma 8.6.V, ensuring the  $\Gamma$ -integrability of  $|\tilde{u}(\omega)|^2$  and hence of  $|\tilde{\nu}(\omega)|^{2n} |\tilde{u}(\omega)|^2$ .

Because  $|\tilde{\nu}(\omega)| < 1$  for  $\omega \neq 0$  [this holds for nonlattice distributions in  $\mathbb{R}^d$  for arbitrary  $d \geq 1$ : see (11.4.6) for the case  $d = 1$ ], the right-hand side of the identity above converges to  $\Gamma\{0\}$ , which being equal to  $\text{var } Y$  where  $Y$  is given by (11.4.10) (see Exercise 12.2.9), vanishes for an ergodic process.

In the nonergodic case we have to replace  $\Gamma$  by a modified measure  $\Gamma^*$  introduced as the Fourier transform of the modified covariance measure

$$C^*(A \times B) = E[(N_0(A) - Y\ell(A))(N_0(B) - Y\ell(B))].$$

Then  $\Gamma^*$  differs from  $\Gamma$  precisely by the absence of the atom at zero. We can now argue as in the ergodic case and deduce that

$$E \left| \int_{\mathcal{X}} \int_{\mathcal{X}} u(x+y) \nu_n(dy) N_0(dx) - Y \int_{\mathcal{X}} u(x) dx \right|^2 \rightarrow \Gamma^*\{0\} = 0.$$

This result implies (11.4.11) and so completes the proof.  $\square$

It is easily seen that both ordinary and mixed stationary Poisson processes are invariant under the operation of random translation (see Exercise 11.4.1). As a corollary to Theorem 11.4.II we now have the following converse.

**Corollary 11.4.III.** *Suppose that  $N$  is a stationary, second-order point process that is invariant under the operation of random translation according to a nonlattice distribution  $\nu$ . Then  $N$  is a stationary mixed Poisson process.*

PROOF. Take  $N$  as the initial distribution in the theorem, and observe that if  $N$  is invariant the weak limit of the  $N_n$  must coincide with  $N$ .  $\square$

A second case of interest arises when the random translations  $\nu_n$  are derived from the movements over time  $n$  of particles with fixed but random and independently chosen velocities as in Example 8.3(g). If these velocities have a common distribution  $\nu$ , we can then write

$$\nu_n(dx) = \nu(n^{-1}dx)$$

and observe that, from Exercise 11.4.5(d),  $Q_A(\nu_n) = Q_{A/n}(\nu) \rightarrow 0$ . Moreover, the integral at (11.4.13) becomes

$$\int_{\mathcal{X}} |\tilde{u}(\omega)|^2 |\nu(n\omega)|^2 \Gamma(d\omega).$$

Now if  $\nu$  is absolutely continuous,  $|\tilde{\nu}(n\omega)| \rightarrow 0$  as  $n \rightarrow \infty$  for every  $\omega \neq 0$  by the Riemann–Lebesgue lemma, so that the proof of (11.4.9) and its extension (11.4.11) to the ergodic case can be completed as in the previous discussion. The restriction to integer values  $n$  is immaterial here, and we therefore obtain the following further result.

**Theorem 11.4.IV.** *Let  $N_0$  be as in Theorem 11.4.II, and for all  $t \geq 0$  let the point processes  $N_t$  be derived from  $N_0$  by random translations through time  $t$  by fixed but random velocities with common distribution  $\nu$ . If  $\nu$  is absolutely continuous with respect to Lebesgue measure on  $\mathbb{R}^d$ , the processes  $N_t$  converge weakly to a mixed Poisson process as in (11.4.12).*

**Corollary 11.4.V.** *Let the point process  $N_t$  represent the position at time  $t$  of a system of particles moving in  $\mathbb{R}^d$  with fixed velocities chosen independently and randomly according to a distribution  $\nu$  that is absolutely continuous with respect to Lebesgue measure in  $\mathbb{R}^d$ . If the distribution of  $N_t$  is independent of  $t$ , spatially homogeneous, and of second order, then  $N_t$  is a mixed Poisson process as in (11.4.12).*

A more general type of location-dependent random translation is illustrated in the following example.

**EXAMPLE 11.4(a)** *Markov shifts (random translations).* Suppose given a point process on  $\mathcal{X}$  with p.g.fl.  $G[h]$  ( $h \in \mathcal{V}(\mathcal{X})$ ) and that any particle of this process initially at  $x$  is shifted into any  $A \in \mathcal{B}_{\mathcal{X}}$  with probability  $p(A | x)$ , where

$$p(\mathcal{X} | x) = \int_{\mathcal{X}} p(dy | x) \leq 1 \quad (\text{all } x),$$

the shortfall  $q(x) = 1 - p(\mathcal{X} | x)$  being the probability of deletion of the particle. Arguing as for Exercise 11.3.1 yields

$$G_m[h | x] = q(x) + \int_{\mathcal{X}} h(y) p(dy | x) = 1 - \int_{\mathcal{X}} [1 - h(y)] p(dy | x)$$

for the p.g.fl. of the (zero- or one-point) cluster associated with  $x$ , from which the resultant p.g.fl. for the translated process equals  $G[G_m[h | \cdot]]$ . The  $k$ th factorial moment  $M_{[k]}^{\text{tr}}$  for the shifted process is given in terms of the corresponding moment of the initial process by

$$\int_{\mathcal{X}} \cdots \int_{\mathcal{X}} p(dy_1 | x_1) \cdots p(dy_k | x_k) M_{[k]}(dx_1 \times \cdots \times dx_k).$$

When the initial process is Poisson with parameter measure  $\mu(\cdot)$  so that  $\log G[h] = - \int_{\mathcal{X}} [1 - h(x)] \mu(dx)$ , the p.g.fl. of the shifted process equals

$$\exp \left( - \int_{\mathcal{X}} \int_{\mathcal{X}} [1 - h(y)] p(dy | x) \mu(dx) \right),$$

so the shifted process is Poisson also, with parameter measure

$$\mu^{\text{tr}}(A) = \int_{\mathcal{X}} p(A | x) \mu(dx) \quad (\text{bounded } A \in \mathcal{B}_{\mathcal{X}}).$$

A situation of particular interest arises if  $\mu^{\text{tr}} = \mu$ ; that is,  $\mu$  is an invariant measure (not necessarily totally finite) for the Markov transition kernel  $p(\cdot | \cdot)$ . It follows from the last relation that a Poisson process with this parameter measure is invariant under the Markov shift operation, a result due to Derman (1955).

Consider finally the case of a pure shift (so that  $q(x) = 0$  for all  $x \in \mathcal{X}$ ). Suppose that  $\mathcal{X} = \mathbb{R}^d$  and  $\mu(dx) = \mu\ell(dx)$  where on the right-hand side,  $\mu$  is a constant and  $\ell$  denotes Lebesgue measure on  $\mathcal{B}_{\mathbb{R}^d}$ . Then the initial process is stationary and  $p(dy | x) = F(d(y - x))$ , meaning that the shifts are identically distributed about the positions of the initial points; that is, we have random translations of the points. Then  $\mu^{\text{tr}} = \mu\ell$  and consequently a stationary Poisson process is invariant under a process of i.i.d. shifts.  $\square$

Before leaving this topic we make a few remarks concerning the  $L_1$  theory referred to briefly before Theorem 11.4.II. A key step here is to show that in both situations considered, the distributions  $\nu_n$  satisfy the condition, for all bounded Borel sets  $A \in \mathbb{R}^d$ ,

$$\int_{\mathbb{R}^d} |\nu_n(y + A) - \nu_n(x + y + A)| dy \rightarrow 0 \quad \text{uniformly in } x. \quad (11.4.14)$$

This condition, or the apparently stronger but in fact equivalent condition

$$\|\nu_n * \gamma_1 - \nu_n * \gamma_2\| \rightarrow 0 \quad (11.4.15)$$

for all pairs  $\gamma_1, \gamma_2$  of distributions absolutely continuous with respect to Lebesgue measure in  $\mathbb{R}^d$ , is referred to in MKM (1978) as *weak asymptotic uniformity* of the sequence  $\{\nu_n\}$ . A particular example of such a sequence is the sequence of uniform distributions on the sets  $\{A_n\}$  of a convex averaging sequence. The major technical difficulty is then to show that the standard form of conclusion of the mean ergodic theorem, which can be written as

$$\mathbb{E} \left| \int_{\mathbb{R}^d} H_n(x + A) N_0(dx) - Y\ell(A) \right| \rightarrow 0, \quad (11.4.16)$$

where  $H_n$  is this special case of a weakly asymptotically uniform sequence, can be extended to the general case and therefore implies (11.4.11) in each of the two situations under consideration.

The definitive treatment of the  $L_1$  case was given by Stone (1968), after earlier work by Dobrushin (1956) and Maruyama (1955) in the context of iterated random translation, and by Breiman (1963) and Thedéen (1964) for the random velocities scheme. Further extensions and generalizations occur in a series of papers by Matthes and co-workers; for details we refer to MKM (1978) especially Chapter 11 and the further references there. An algebraic treatment of (11.4.14) and related properties, when  $\nu_n$  are convolution powers, is contained in the papers by Stam (1967a, b). The second-order treatment used to prove Theorem 11.4.II is an extension of the discussion in Vere-Jones (1968). Some partial results concerning (11.4.14) and related topics are covered in Exercises 11.4.4–5.

## Exercises and Complements to Section 11.4

11.4.1 Show that in the stationary case, both ordinary and mixed Poisson processes are invariant under the operation of random translation.

[Hint: Use the p.g.f.l. representations at (11.4.1) and (11.4.12).]

- 11.4.2 The binomial distribution  $\{b_k(n; p)\} = \left\{\binom{n}{k} p^k (1-p)^{n-k}\right\}$  with  $0 < p < 1$  is the  $n$ -fold convolution of the simplest nondegenerate d.f.  $F$  that can arise with Lemma 11.4.I. The order  $1/\sqrt{n}$  of the bound at (11.4.4) is tight because

$$\frac{1}{\sqrt{2\pi(n+1)p(1-p)}} \leq Q_{\{0\}}(\{b_k(n; p)\}) \leq \frac{1}{\sqrt{4(n+1)p(1-p)}}$$

[see, e.g., MKM (1978, pp. 476–477) or else Daley (1987)].

- 11.4.3 Let  $P(x, A)$  denote a stochastic or substochastic kernel defined for all  $x \in \mathbb{R}^d$  and  $A \in \mathcal{B}(\mathbb{R}^d)$  such that it has an infinite invariant measure  $\nu$ . Consider the operation of random translation according to the kernel  $P$  [i.e., a point initially at  $x$  is translated to a new point  $y$  according to the distribution  $P(x, \cdot)$ ].

- The Poisson process with intensity measure  $\nu$  is invariant under this operation.
- If  $P$  is continuous, then the initial process  $N_0$  is invariant under this operation if and only if it is a Cox process directed by  $Y\nu$ , where  $Y$  is a nonnegative random variable.
- Investigate conditions under which the sequence of point processes  $\{N_n\}$  obtained from an initial process  $N_0$  by successive iteration of this operation will converge to a limit of the form described in (b).

[Hint: See Kerstan and Debes (1969) and Debes *et al.* (1971). Part (a) goes back to Derman (1955). The case where  $\nu$  is totally finite is discussed in MKM (1978, Section 4.8).]

- 11.4.4 For a given distribution  $F$  on  $\mathbb{R}^d$ , let  $\mathcal{S}$  denote the set of points  $a$  in  $\mathbb{R}^d$  such that for all intervals  $A$ ,

$$\sup_x |F^{n*}(x + a + A) - F^{n*}(x + A)| \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (11.4.17)$$

Prove the following.

- $\mathcal{S}$  is an algebra.
  - If  $a \in \text{supp}(F)$ , then  $a \in \mathcal{S}$ .
  - If  $\text{supp}(F)$  is contained in no proper subalgebra of  $\mathbb{R}^d$ , then  $\mathcal{S} = \mathbb{R}^d$ .
- 11.4.5 (Continuation). A sequence of measures  $\{\nu_n\}$  is *weakly asymptotically uniformly distributed* if for all absolutely continuous distributions  $\sigma$  on  $\mathcal{B}(\mathbb{R}^d)$  and all  $x \in \mathbb{R}^d$ ,

$$\|\sigma * \nu_n * \delta_x - \sigma * \nu_n\| \rightarrow 0 \quad (n \rightarrow \infty). \quad (11.4.18)$$

- Show that (11.4.17) implies (11.4.18) in the special case that  $\sigma$  is the uniform distribution on the interval  $A$  and  $\{\nu_n\} = \{F^{n*}\}$ .
- Extend this result and deduce that if  $F$  is nonlattice, the sequence of convolution powers of  $F$  is weakly asymptotically uniformly distributed.
- Prove that (11.4.18) is equivalent to

$$\|\nu_n * \sigma_1 - \nu_n * \sigma_2\| \rightarrow 0 \quad (n \rightarrow \infty)$$

for all pairs of absolutely continuous distributions  $\sigma_1$  and  $\sigma_2$ .

- If (11.4.18) holds then  $Q_A(\nu_n) \rightarrow 0$  (cf. Lemma 11.4.I).

[Hint: For further details and applications see MKM (1978, Chapter 11).]



## CHAPTER 12

# Stationary Point Processes and Random Measures

12.1	Stationarity: Basic Concepts	177
12.2	Ergodic Theorems	194
12.3	Mixing Conditions	206
12.4	Stationary Infinitely Divisible Point Processes	216
12.5	Asymptotic Stationarity and Convergence to Equilibrium	222
12.6	Moment Stationarity and Higher-order Ergodic Theorems	236
12.7	Long-range Dependence	249
12.8	Scale-invariance and Self-similarity	255

Stationary point processes play an exceptionally important role in applications and lead also to a rich theory. They have appeared already in Chapters 3, 6 and 8, where some basic properties and applications were outlined, and they are central to the discussion not only in the present chapter, but also in Chapters 13 and 15, and to a lesser extent Chapter 14 also.

Our main purpose in this chapter is to develop a systematic study of the theory of point processes and random measures that are invariant under shifts in  $d$ -dimensional Euclidean space  $\mathbb{R}^d$ . Much of the theory is likely to appear familiar: some parts are merely variants of the corresponding theory of stationary continuous processes; the greater part can be deduced from the theory of stationary random distributions, or, in  $\mathbb{R}^1$ , from the theory of processes with stationary increments, but some parts, especially applications, are peculiar to point processes and random measures.

Although we have chosen to develop the basic theory for shifts  $S_u$  acting on the canonical space  $\mathcal{M}_{\mathcal{X}}^{\#}$ , with  $\mathcal{X} = \mathbb{R}^d$ , the underlying ideas are capable of extension in several directions. In the first instance this refers to point processes and random measures invariant under more general forms of group action, for example, to processes in two- or three-dimensional Euclidean space that are invariant under rotations (i.e., isotropy) as well as shifts (i.e., homogeneity), and to processes on other types of manifold, such as the surface of a sphere or cylinder. In fact, many of the topics included in this chapter can

be developed with almost equal facility (i.e., requiring nothing or little more than changes of wording or interpretation) for point processes on a locally compact metric group, and are so developed in the Russian edition of MKM (1982). We examine some such extensions in the present chapter, especially in connection with scale-invariance and self-similarity, whereas others, including isotropy, are taken up in Chapter 15 on spatial point processes.

At the same time, the shifts studied in this chapter are examples of a *flow* on a probability space  $(\Omega, \mathcal{E}, \mathcal{P})$ , meaning in general a group  $\{\theta_g\}$  of measurable one-to-one transformations of  $(\Omega, \mathcal{F})$  onto itself. The probability measure  $\mathcal{P}$  is invariant under the flow if for all  $E \in \mathcal{E}$  and all  $\theta_g$ ,  $\mathcal{P}(\theta_g E) = \mathcal{P}(E)$ . This more general concept [see, e.g., Baccelli and Brémaud (2003)] is useful in unifying the treatment of processes, including marked point processes (MPPs) and Cox processes, where the canonical space  $\mathcal{M}_{\mathcal{X}}^{\#}$  needs extending to accommodate information about the outcomes of auxiliary random variables or processes.

An important technical role in establishing the form of both probability and moment structures for stationary processes is played by the factorization theorems summarized in Appendix A2 as Lemma A2.7.II and Theorem A2.7.III. In their basic form they assert that if a measure  $\mu$  on a product space  $\mathbb{R}^d \times \mathcal{K}$  is invariant under shifts in the first component, then  $\mu$  reduces to a product of Lebesgue measure in  $\mathbb{R}^d$  and a fixed measure  $\kappa$  on  $\mathcal{K}$ . These factorization results extend to more general contexts with  $\mathbb{R}^d$  replaced by a  $\sigma$ -group  $\mathcal{H}$ , and Lebesgue measure replaced by Haar measure on  $\mathcal{H}$ . They underlie the structure not only of stationary MPPs, where they apply most obviously, but also of stationary Poisson and Poisson cluster processes, of the moment measures of stationary processes, and of the Palm theory which is the subject of Chapter 13.

After a first section on basic concepts and examples, the chapter covers ergodic theorems, moment and mixing properties (Sections 12.2–4), stationary infinitely divisible point processes (Section 12.5), convergence to equilibrium (Section 12.6), long-range dependence (Section 12.7), and scale-invariance and self-similarity (Section 12.8).

## 12.1. Stationarity: Basic Concepts

We consider first  $\mathcal{X} = \mathbb{R}^d$  and invariance properties with respect to translations (or shifts). For arbitrary  $u$ ,  $x \in \mathcal{X}$ , and  $A \in \mathcal{B}_{\mathcal{X}}$ , write

$$T_u x = x + u, \quad T_u A = A + u = \{x + u : x \in A\}. \quad (12.1.1)$$

Then  $T_u$  induces a transformation  $S_u$  of  $\mathcal{M}_{\mathcal{X}}^{\#}$  (and also of  $\mathcal{N}_{\mathcal{X}}^{\#}$ ) through the equation<sup>1</sup>

$$(S_u \mu)(A) = \mu(T_u A) \quad (\mu \in \mathcal{M}_{\mathcal{X}}^{\#}, A \in \mathcal{B}_{\mathcal{X}}). \quad (12.1.2)$$

It is clear that  $S_u \mu \in \mathcal{M}_{\mathcal{X}}^{\#}$  whenever  $\mu \in \mathcal{M}_{\mathcal{X}}^{\#}$ ; that is,  $S_u$  maps  $\mathcal{M}_{\mathcal{X}}^{\#}$  into (indeed, onto) itself. Moreover,  $S_u$  is continuous: to see this, let  $\{\mu_n\}$  be

---

<sup>1</sup> With the operators  $T$  and  $S$ , as defined, the Dirac measure  $\delta(\cdot)$  has the property that

a sequence of measures on  $\mathcal{B}_{\mathcal{X}}$  converging in the  $w^\#$ -topology to a limit  $\mu$ , and let  $f$  be a bounded continuous function vanishing outside a bounded set. Then its translate  $f(x - u)$  has similar properties, so from properties of  $w^\#$ -convergence (Proposition A2.6.II),

$$\begin{aligned} \int_{\mathcal{X}} f(x) (S_u \mu_n)(dx) &= \int_{\mathcal{X}} f(x - u) \mu_n(dx) \\ &\rightarrow_{w^\#} \int_{\mathcal{X}} f(x - u) \mu(dx) = \int_{\mathcal{X}} f(x) (S_u \mu)(dx). \end{aligned}$$

An application of the sufficiency half of Proposition A2.6.II shows that  $S_u \mu_n \rightarrow_{w^\#} S_u \mu$ , and hence that  $S_u$  is continuous. Because a shifted counting measure is again a counting measure, and  $\mathcal{N}_{\mathcal{X}}^\#$  is closed in  $\mathcal{M}_{\mathcal{X}}^\#$ , the same conclusion holds for the effects of shifts  $T_u$  on counting measures. This establishes the following simple but important result.

**Lemma 12.1.I.** *For  $\mathcal{X} = \mathbb{R}^d$  and  $u \in \mathbb{R}^d$ , both the mappings  $S_u: \mathcal{M}_{\mathcal{X}}^\# \mapsto \mathcal{M}_{\mathcal{X}}^\#$  and  $S_u: \mathcal{N}_{\mathcal{X}}^\# \mapsto \mathcal{N}_{\mathcal{X}}^\#$  defined at (12.1.2) via the shift operator  $T_u$  are continuous (and hence measurable) and one-to-one.*

It now follows that if  $\xi$  is a random measure or point process, then so is  $S_u \xi$  for every  $u \in \mathbb{R}^d$  because  $S_u \xi$  is then the composition of two measurable mappings. This remark enables us to make the following definition.

**Definition 12.1.II.** *A random measure or point process  $\xi$  with state space  $\mathcal{X} = \mathbb{R}^d$  is stationary if, for all  $u \in \mathbb{R}^d$ , the fidi distributions of the random measures  $\xi$  and  $S_u \xi$  coincide.*

If extra emphasis is needed, we call such random measures *strictly stationary* or *stationary as a whole* to distinguish them from random measures that are stationary in weaker senses such as second-order stationarity following Proposition 8.1.I. Note also that this definition is the natural extension to  $\mathbb{R}^d$  and  $\mathcal{M}_{\mathcal{X}}^\#$  of Definition 3.2.I, and lies behind the summary of stationarity properties in Proposition 6.1.I.

Definition 12.1.II can be stated in a compact form by defining a ‘lifted’ operator or transformation  $\widehat{S}_u$  that functions at a third level of abstraction, on measures  $\mathcal{P}$  on the Borel sets of  $\mathcal{M}_{\mathcal{X}}^\#$ . For  $B \in \mathcal{B}(\mathcal{M}_{\mathcal{X}}^\#)$  set

$$\widehat{S}_u \mathcal{P}(B) = \mathcal{P}(S_u B), \quad (12.1.3)$$

where  $S_u B = \{S_u \mu: \mu \in B\}$ . The remark following Lemma 12.1.I implies that  $\widehat{S}_u$  maps  $\mathcal{M}_{\mathcal{X}}^\#$  (or  $\mathcal{N}_{\mathcal{X}}^\#$ ) into itself, and an argument similar to the proof of

---


$$(S_u \delta_x)(\cdot) = \delta_{x-u}(\cdot) \text{ and}$$

$$\int f(x) (S_u \mu)(dx) = \int f(x) \mu(dx + u) = \int f(x - u) \mu(dx).$$

Sometimes, an operator  $S'_u$  for which  $S'_u = S_{-u}$  is used instead. Then the  $+$  and  $-$  signs in the above equations are interchanged. The operator  $S_u$  we use is the same as  $T_u$  in MKM (1978, p. 258) and as  $T_{-u}$  in Kallenberg (1975 or 1983a, Exercise 10.10).

that lemma shows that the mapping is even continuous (see Exercise 12.1.1). Then Definition 12.1.II is equivalent to stating that a random measure on  $\mathbb{R}^d$  is stationary if its distribution on  $\mathcal{M}_{\mathbb{R}^d}^\#$  is invariant under shifts  $\widehat{S}_u$ .

These concepts and results can be extended to the more general context of a flow referred to in the introduction. In this case the probability space  $(\Omega, \mathcal{E}, \mathcal{P})$  is not given an explicit structure, but it is supposed to be capable of supporting a family of one-to-one measurable transformations of  $\Omega$  onto itself,  $\{\theta_u: u \in \mathcal{G}\}$  say, where  $\mathcal{G}$  has a group structure,  $\theta_{u+v} = \theta_u \circ \theta_v$ ,  $\theta_0$  is the identity, and  $(\theta_u)^{-1} = \theta_{-u}$ . The link to a transformation acting directly on the random measure  $\xi: (\Omega, \mathcal{E}) \mapsto (\mathcal{M}_{\mathcal{X}}^\#, \mathcal{B}_{\mathcal{M}_{\mathcal{X}}^\#})$  is then provided by taking  $\mathcal{G} = \mathcal{X} = \mathbb{R}^d$  and requiring that

$$\xi(A, \theta_u \omega) = \xi(A + u, \omega)$$

or, more briefly,  $\xi(\theta_u \omega) = S_u \xi(\omega)$ .

Most of the examples below illustrate specific cases where the flow is defined for  $u \in \mathcal{G} = \mathbb{R}^d$ , but of course shifts in  $\mathbb{R}^d$  are not the only actions which can be defined by flows.

In the case of a marked point process, it is enough to take for  $\Omega$  the space of counting measures on the product space  $\mathbb{R}^d \times \mathcal{K}$  (i.e.,  $\mathcal{N}_{\mathbb{R}^d \times \mathcal{K}}^\#$ ), and to consider for the flow the family of translations  $S'_u$  acting on the first component only, so that  $S'_u N(A \times K) = N(T_u A \times K)$ . The MPP is stationary if its fidi distributions are invariant under the translations  $\{S'_u\}$  (equivalently, its probability distribution on  $\mathcal{M}_{\mathbb{R}^d \times \mathcal{K}}^\#$  is invariant under the lifted operators  $\{\widehat{S}'_u\}$ ). In discussing a stationary MPP  $N$  say, we mostly write  $S_u N$  rather than  $S'_u N$ , the restriction of the shift to  $\mathbb{R}^d$  being understood.

We proceed to a detailed study of stationarity of random measures in  $\mathbb{R}^d$ , illustrating the constructions in terms of shifts  $\widehat{S}_u$  on the canonical probability space.

The results of Chapter 9 imply that the distribution of a random measure is completely determined either by its fidi distributions or by its Laplace functional (see Propositions 9.2.III and 9.4.II). Similarly, the distribution of a point process is completely determined by its fidi distributions or by its p.g.fl. (Theorem 9.4.V) or, if the point process is simple, by its avoidance function (Theorem 9.2.XII). Applying these criteria to the definition of stationarity, we deduce that a random measure is stationary if and only if its fidi distributions are stationary, or, equivalently, if and only if its Laplace functional is stationary. Similarly, a point process is stationary if and only if its p.g.fl. is stationary, or, if it is simple, if and only if its avoidance function is stationary. Spelling out the details of these remarks yields the following theorem.

**Theorem 12.1.III.** *Let  $\xi$  be a random measure on state space  $\mathcal{X} = \mathbb{R}^d$ . Each of the following conditions is necessary and sufficient for  $\xi$  to be stationary.*

- (i) *For each  $u \in \mathbb{R}^d$  and  $k = 1, 2, \dots$ , the fidi distributions satisfy*

$$F_k(A_1 \dots, A_k; x_1, \dots, x_k) = F_k(A_1 + u, \dots, A_k + u; x_1, \dots, x_k). \quad (12.1.4)$$

(ii) For each  $u \in \mathbb{R}^d$  and  $f \in \text{BM}(\mathbb{R}^d)$  the characteristic functional satisfies

$$\Phi[f(\cdot)] = \Phi[f(\cdot - u)]. \quad (12.1.5)$$

When  $\xi$  is a point process  $N$  the conditions are equivalent to the following.

(iii) For each  $u \in \mathbb{R}^d$  and  $h \in \mathcal{V}(\mathbb{R}^d)$ , the p.g.fl.  $G$  satisfies

$$G[h(\cdot)] = G[h(\cdot - u)]. \quad (12.1.6)$$

If also  $N$  is simple, the conditions are equivalent to the following.

(iv) For each  $u \in \mathbb{R}^d$  and all bounded Borel sets  $A \in \mathcal{B}(\mathbb{R}^d)$ , the avoidance function  $P_0(\cdot)$  of  $N$  satisfies

$$P_0(A) = P_0(A + u). \quad (12.1.7)$$

Furthermore, in (i) and (iv) it is sufficient for the results to hold for disjoint sets  $A_i$  and  $A$  from a dissecting semiring generating  $\mathcal{B}(\mathbb{R}^d)$ .

The final statement of the theorem implies that it is enough in (i) and (iv) to have the statements holding for disjoint sets that can be represented as finite unions of half-open rectangles. It is not possible to relax this condition significantly: in  $\mathbb{R}^1$  Lee's counterexample quoted in Exercise 2.3.1 exhibits two processes with the same distributions for  $N(I)$  whenever  $I$  is an interval, one of these processes being stationary (indeed, a stationary Poisson process) and the other not. See also Exercise 12.1.2.

Analogues of this proposition hold for other examples of flows. The details for MPPs are spelled out in Exercise 12.1.3. The next few examples illustrate some applications of Theorem 12.1.III and its extensions.

**EXAMPLE 12.1(a)** *Stationarity of Poisson and compound Poisson processes* [continued from Example 9.4(c); see also Lemma 6.4.VI]. From the representation of the Poisson process p.g.fl. at (9.4.17) we have, for  $\mathcal{X} = \mathbb{R}^d$ ,

$$\log G[h(\cdot - u)] = \int_{\mathcal{X}} [h(x - u) - 1] \mu(dx) = \int_{\mathcal{X}} [h(y) - 1] (S_u \mu)(dy), \quad (12.1.8)$$

which under the assumption of stationarity at (12.1.6) is to be equal to

$$\int_{\mathcal{X}} [h(y) - 1] \mu(dy).$$

Because the measure  $\mu$  is completely determined by its integrals of functions of the form  $h(y) - 1$  for  $h \in \mathcal{V}(\mathbb{R}^d)$ , it follows that a Poisson process is stationary if and only if its parameter measure is invariant under translation. Now the only measure on  $\mathbb{R}^d$  invariant under translations is Lebesgue measure, so  $\mu(\cdot)$  must be a multiple of Lebesgue measure on  $\mathbb{R}^d$ ; that is, for some  $\mu \geq 0$ ,

$$\mu(\cdot) = \mu \ell(\cdot).$$

Thus, a Poisson process on  $\mathbb{R}^d$  is stationary if and only if it has a constant intensity with respect to Lebesgue measure on  $\mathbb{R}^d$ .

Alternatively, we may first observe that a stationary random measure can have no fixed atoms, then use the fact (Theorem 2.4.II) that a Poisson process has no fixed atoms if and only if its parameter measure is nonatomic, in which case the process is simple, and finally appeal to (12.1.7), which yields

$$e^{-\mu(A)} = e^{-\mu(A+u)},$$

implying the same result even more directly.

The compound Poisson process, in the general sense of Section 6.4, is an example of a marked point process, but essentially similar techniques can be applied. Using the p.g.fl. approach from Exercise 12.1.3(iii), we have to check, for a constant rate Poisson ground process and i.i.d. marks, that (12.1.8) holds for a function  $h(u, \kappa)$  of the two variables. In fact we have

$$\begin{aligned} \log G[h(\cdot - u, \cdot)] &= \int_{\mathcal{X}} \int_{\mathcal{K}} [h(x - u, \kappa) - 1] \mu \, dx \, \pi(d\kappa) \\ &= \int_{\mathcal{X}} \int_{\mathcal{K}} [h(y, \kappa) - 1] \mu \, dy \, \pi(d\kappa) = \log G[h(\cdot, \cdot)]. \quad \square \end{aligned}$$

EXAMPLE 12.1(b) *Stationarity is preserved by simple random thinnings and translations* [continued from Sections 11.3 and 11.4]. Let  $N$  be a stationary point process and assume that each point  $x_i$  of a realization of  $N$  is independently and randomly shifted through a random distance  $X_i$ , where the  $\{X_i\}$  are identically distributed with common d.f.  $F(\cdot)$ ; to accommodate deletions, we allow the distribution to be defective, and set  $q = 1 - F(\mathbb{R}^d)$ . Then from equation (11.4.1) the respective p.g.fl.s  $G$  and  $G_0$  of the shifted process and  $N$  are related by

$$G[h(\cdot)] = G_0[q + \int_{\mathcal{X}} h(y) F(dy - \cdot)].$$

Much as in the previous example, when  $G_0$  is itself stationary,  $G_0[h(\cdot)] = G_0[h(\cdot - u)]$  for all  $u \in \mathbb{R}^d$  and  $h \in \mathcal{V}(\mathbb{R}^d)$ . The right-hand side of the expression for  $G[\cdot]$  then equals

$$G_0[q + \int_{\mathcal{X}} h(y - u) F(dy - \cdot)] = G[h(\cdot - u)],$$

so by (iii) the transformed process is again stationary. Pure translations occur when  $q = 0$ , else random deletions when  $F$  is concentrated at 0.  $\square$

The stationarity of Cox processes and some cluster processes can be verified by similar techniques. Cox processes are important as examples where the flow needs to be defined initially on an extension of the canonical space  $\mathcal{M}_{\mathcal{X}}^{\#}$ .

EXAMPLE 12.1(c) *Mixed Poisson and Cox processes*. For the case of a mixed Poisson process, we may take  $\Omega = \mathcal{N}_{\mathbb{R}^d \times \mathbb{R}_+}^{\#}$ . Then the pair  $(N, \lambda)$  corresponds to the choice of a counting measure  $N \in \mathcal{N}_{\mathbb{R}^d}^{\#}$  and a rate  $\lambda \in \mathbb{R}_+$ . The distribution  $\mathcal{P}$  can be generated by conditioning as outlined in Section 6.1:

$$\mathcal{P}(V \times A) = \int_A \text{Poi}(V \mid \lambda) \Pi(d\lambda) \quad (A \in \mathcal{B}_{\mathbb{R}_+}),$$

where  $V$  is a set of realizations from  $\mathcal{N}_{\mathbb{R}^d}^\#$ ,  $\text{Poi}(\cdot \mid \lambda)$  is the probability distribution on  $\mathcal{N}_{\mathbb{R}^d}^\#$  of a Poisson process at rate  $\lambda$ , and  $\Pi$  is the distribution of  $\lambda$  on the Borel sets of  $\mathbb{R}_+$ . As in Example 12.1(a), the flow is the family of shifts  $S_u$  on the first component. Stationarity is guaranteed, because invariance of the mixture  $\mathcal{P}(\cdot)$  is implied by invariance of each of the conditional distributions  $\text{Poi}(\cdot \mid \lambda)$ .

The case of a Cox process is only a little more complicated. Here we may take  $\Omega = \mathcal{N}_{\mathcal{X}}^\# \times \mathcal{M}_{\mathcal{X}}^\#$ , where the first component refers to the realizations of the point process and the second to the realizations of the directing random measure. The flow must now act simultaneously on both components, so that

$$(N(A, \theta_u \omega), \xi(B, \theta_u \omega)) = (N(A + u, \omega), \xi(B + u, \omega)) \quad (12.1.9)$$

or in more compact notation

$$\theta_u(N, \xi) = (S_u N, S_u \xi). \quad (12.1.9')$$

We have then for the distribution  $\mathcal{P}$  of the Cox process on  $\Omega$  as above,

$$\mathcal{P}(V \times W) = \int_W \text{Poi}(V \mid \xi) \mathcal{Q}(d\xi) \quad (W \in \mathcal{B}(\mathcal{M}_{\mathbb{R}^d}^\#)), \quad (12.1.10)$$

where  $V$  a set of counting measures,  $W$  is a set of directing measures  $\xi$ ,  $\text{Poi}(\cdot \mid \xi)$  is now the distribution of the inhomogeneous Poisson process with parameter measure  $\xi$ , and  $\mathcal{Q}$  is the distribution of the directing random measure  $\xi$ . Exercise 12.1.4 indicates how to show that the resultant process is stationary if and only if  $\mathcal{Q}$  is stationary. In this case the bivariate process  $(N(\cdot), \xi(\cdot))$  is also stationary, its distribution being invariant under the same shift acting on both components.  $\square$

Similar constructions are possible in other cases where the evolution of the random measure under study is associated with the evolution of some auxiliary process. For stationarity of a general cluster process see Exercise 12.1.6; the important example of a Poisson cluster process is summarized shortly in Proposition 12.1.V where for the first time we meet a measure in  $\mathbb{R}^d$  invariant under the group of *diagonal shifts*  $D_x^{(k)}$  defined for  $k \in \mathbb{Z}_+$  and  $x \in \mathbb{R}^d$  by

$$D_x^{(k)}(y_1, \dots, y_k) = (x + y_1, \dots, x + y_k), \quad (12.1.11)$$

where  $y = (y_1, \dots, y_k)$  and  $y_i \in \mathbb{R}^d$  for  $i = 1, \dots, k$ , so first we examine the structure of such measures. As in Appendix A2.7, the cosets under this group of transformations are images of the main diagonal  $y_1 = \dots = y_k$ . The action of  $D_x^{(k)}$  along any such coset is just a shift through the vector  $x$ . Thus, we should anticipate that any measure invariant under the diagonal shifts should reduce to a multiple of Lebesgue measure in each such coset. The next lemma makes this idea precise: by the *diagonal subspace* we mean the space  $\{(y_1, \dots, y_k) : y_1 = \dots = y_k \in \mathbb{R}^d\}$ .

**Lemma 12.1.IV** (Diagonal Shifts Lemma). *Let  $\mu$  be a boundedly finite Borel measure on  $\mathcal{X}^{(k)}$  with  $\mathcal{X} = \mathbb{R}^d$ . Then  $\mu$  is invariant under the diagonal shifts  $D_x^{(k)}$  of (12.1.12) if and only if it can be represented as a product of Lebesgue measure on the diagonal subspace and a reduced measure  $\check{\mu}$  on  $\mathcal{X}^{(k-1)}$  such that, for any function  $f \in \text{BM}(\mathcal{X}^{(k)})$ , and  $k > 0$ ,*

$$\begin{aligned} & \int_{\mathcal{X}^{(k)}} f(x_1, \dots, x_k) \mu(dx_1 \times \dots \times dx_k) \\ &= \int_{\mathcal{X}} dx \int_{\mathcal{X}^{(k-1)}} f(x, x + y_1, \dots, x + y_{k-1}) \check{\mu}(dy_1 \times \dots \times dy_{k-1}), \end{aligned} \quad (12.1.12)$$

where in the case  $k = 1$ ,  $\check{\mu}(\cdot) = m\delta_0(\cdot)$  in which  $\delta_0$  denotes Dirac measure,  $m = \mu(\mathbb{U}^d)$ , and  $\mathbb{U}^d$  is the unit  $d$ -dimensional hypercube.

PROOF. Consider the mapping from  $\mathcal{X} \times \mathcal{X}^{(k-1)}$  into  $\mathcal{X}^{(k)}$  defined by

$$(x, (y_1, \dots, y_{k-1})) \mapsto (x, x + y_1, \dots, x + y_{k-1}). \quad (12.1.13)$$

Given any  $(x_1, \dots, x_k) \in \mathcal{X}^{(k)}$ , we have uniquely  $x = x_1$  and  $y_i = x_{i+1} - x_1$  ( $i = 1, \dots, k-1$ ), so the mapping is one-to-one and onto; it is clearly continuous and hence measurable. Under the mapping, the action of the diagonal shifts  $D_x^{(k)}$  on  $\mathcal{X}^{(k)}$  is reduced to the ordinary shift  $T_x$  on the  $\mathcal{X}$  component of the product  $\mathcal{X} \times \mathcal{X}^{(k-1)}$ . We therefore have a representation of the original space  $\mathcal{X}^{(k)}$  to which we can apply Lemma A2.7.II and assert that the image,  $\mu^*$  say, of  $\mu$  induced by the mapping (12.1.13) reduces to a product of  $d$ -dimensional Lebesgue measure along  $\mathcal{X}$  and some measure  $\check{\mu}$  on the other factor space  $\mathcal{X}^{(k-1)}$ ; that is,  $\mu^* = \ell \times \check{\mu}$ . Then  $\check{\mu}$  and  $\mu$  are related as at (12.1.12).  $\square$

For an alternative approach see Exercises 12.1.8–9 and 12.6.1–2.

**Proposition 12.1.V.** *A Poisson cluster process with a.s. finite clusters, and both cluster centres and cluster members in  $\mathcal{X} = \mathbb{R}^d$ , is stationary if and only if it can be represented in such a way that*

- (i) *the cluster centres form a stationary Poisson process in  $\mathbb{R}^d$ ; and*
- (ii) *the cluster members depend only on their positions relative to the cluster centre, and not on the location of the cluster centre itself.*

*In particular, the p.g.fl. of a stationary Poisson cluster process with a.s. finite clusters has a unique representation (its regular representation) of the form*

$$\log G[h] = \mu_c \sum_{k=1}^{\infty} \frac{\pi_k}{k!} \int_{\mathcal{X}} dx \int_{\mathcal{X}^{(k-1)}} [h(x)h(x+y_1)\dots h(x+y_{k-1}) - 1] P_{k-1}(dy_1 \times \dots \times dy_{k-1}), \quad (12.1.14)$$

where  $\mu_c$  is the intensity of the cluster centre process,  $\{\pi_k: k \geq 1\}$  is a proper probability distribution of cluster sizes,  $P_0(\cdot) = \delta_0(\cdot)$ , and for  $k \geq 2$ ,  $P_{k-1}(\cdot)$  is a symmetric probability distribution describing the locations of the remaining  $k-1$  cluster members relative to an arbitrary cluster member chosen as origin.



*Remark.* As already noted around Proposition 6.3.V, the representation of a cluster process in terms of cluster centre and cluster member processes is not unique. In the present context, it is even possible to construct a stationary Poisson cluster process from nonstationary components; that is, they do not satisfy conditions (i) and (ii) (see Exercise 12.1.5). What the proposition asserts is that, even in such cases, the process will have an alternative representation where the above conditions do hold, and that when there is more than one representation, the regular representation is always available as an option.

PROOF. Recall from Proposition 6.3.V that a Poisson cluster process with a.s. finite clusters has a unique representation with p.g.fl. of the form

$$\log G[h] = \sum_{k=1}^{\infty} \frac{1}{k!} \int_{\mathcal{X}^{(k)}} (h(x_1) \dots h(x_k) - 1) K_k(dx_1 \times \dots \times dx_k), \quad (12.1.15)$$

where the Khinchin measure  $K_k$  has the representation

$$K_k(B) = \int_{\mathcal{X}} J_k(B \mid y) \mu_c(dy) \quad (12.1.16)$$

in terms of the Janossy density  $J_k(\cdot \mid y)$  of the cluster member process and the intensity measure  $\mu_c$  of the Poisson cluster centre process. (Note that we here assume that both cluster centres and cluster members have points in the same space  $\mathcal{Y} = \mathcal{X} = \mathbb{R}^d$ .) If the process is stationary, uniqueness of the representation (12.1.14) implies that each of the measures  $K_k$  must be invariant under diagonal shifts  $(x_1, \dots, x_k) \mapsto (u + x_1, \dots, u + x_k)$ , or equivalently

$$K_k(T_u A_1 \times \dots \times T_u A_k) = K_k(A_1 \times \dots \times A_k).$$

The diagonal shifts Lemma 12.1.IV now implies that  $K_k(\cdot)$  must reduce to a product of Lebesgue measure along the diagonal, and a boundedly finite measure  $\check{K}_{k-1}(\cdot)$  on  $\mathcal{B}(\mathcal{X}^{(k-1)})$  such that (12.1.15) holds.

These ingredients can be used to construct a candidate process for the regular representation. We introduce Janossy measures  $J_0 = 0$ ,  $J_1(dx \mid y) = \delta_y(dx)$ , and, for  $k > 1$ ,

$$J_k(dx_1 \times dx_2 \times \dots \times dx_k \mid x) = (1/\mu_c) \delta_0(dy_1) \check{K}_{k-1}(dy_2 \times \dots \times dy_k),$$

where  $y_i = x_i - x$  for  $i = 1, \dots, k$  and  $\mu_c = \sum_{k=1}^{\infty} \check{K}_{k-1}(\mathcal{X}^{(k-1)})/k!$  is the candidate intensity of a stationary Poisson cluster centre process. We interpret

$$\pi_k = J_k(\mathcal{X}^{(k)})/k!$$

as the probability that a cluster has  $k$  members, and

$$\tilde{P}_{k-1}(\cdot) = (k! \mu_c \pi_k)^{-1} \check{K}_{k-1}(\cdot)$$

as the symmetric probability distribution describing the location of the remaining cluster members relative to a given cluster member as centre; from symmetry the cluster member selected as centre may be regarded as being chosen uniformly at random. We leave the reader to verify that back-substitution of these candidate elements results both in a Poisson cluster process with the required properties and in (12.1.14) being satisfied.

This establishes the necessity of a representation satisfying the conditions (i) and (ii) of the proposition, as well as the form of the regular representation. That the two conditions are sufficient to guarantee stationarity is a matter of verification. In p.g.fl. terms, condition (ii) of the proposition implies that for every  $x$ ,  $G_m[h(\cdot) \mid x] = G_m[h(\cdot + x) \mid 0]$ . It is then straightforward to check that condition (12.1.6) in Theorem 12.1.III is met. Exercise 12.1.6 extends this argument to more general cluster processes; an alternative approach using Radon–Nikodym derivatives and the disintegration of measures is sketched in Exercises 12.1.8–9.  $\square$

The next example examines a particular case of this representation in detail, and shows that it is not always the most natural or convenient for further manipulations.

EXAMPLE 12.1(d) *The regular representation of a stationary Neyman–Scott process* [continued from Example 6.3(a)]. In the Neyman–Scott model the cluster members have a common distribution  $F(\cdot)$  about the cluster centre. To obtain the regular representation we should refer the distribution of the cluster members to an arbitrarily chosen member of the cluster itself as origin. For clusters with just one element we have evidently

$$\tilde{\mathcal{P}}_0(A) = \delta_0(A);$$

that is, the cluster is necessarily centred at its sole representative. For  $k = 2$  we obtain

$$\tilde{\mathcal{P}}_1(A) = \Pr\{Y - X \in A\} = \int_{\mathcal{X}} F(x + A) F(dx),$$

where  $X$  and  $Y$  are independent r.v.s with the distribution  $F$ .

Similarly for general  $k \geq 2$ ,

$$\tilde{\mathcal{P}}_{k-1}(A_2 \times \cdots \times A_k) = \int_{\mathcal{X}} F(dx) F(x + A_2) \cdots F(x + A_k).$$

In the branching process interpretation of the cluster members,  $\tilde{\mathcal{P}}_{k-1}$  gives the distribution of the locations of the other siblings given that the arbitrarily chosen member comes from a family of size  $k$ .  $\square$

The use of (i) or (iv) rather than (ii) or (iii) of Theorem 12.1.III is indicated in the next example. It shows that a stationary measure of renewal type can be defined on  $\mathcal{N}_{\mathbb{R}}^{\#}$  irrespective of whether the interval distribution has a finite or infinite first moment. Without a finite mean, this measure is not totally finite and so cannot be used directly to define a stationary point process; it is used in Exercise 12.4.6 to exhibit an example of a weakly singular infinitely divisible point process.

EXAMPLE 12.1(e) *Stationary regenerative measure and renewal process* [see also Exercises 9.1.13–14]. Let  $\mu$  be a measure, not necessarily totally finite, defined on the space of sequences  $\{Y_n: n \in \mathbb{Z}_+\}$  satisfying  $Y_0 = 0 \leq Y_1 \leq \dots \leq Y_n \rightarrow \infty$  ( $n \rightarrow \infty$ ), and put  $\tau_n = Y_n - Y_{n-1}$  ( $n = 1, 2, \dots$ ). We call  $\mu$  *regenerative* when for any positive integer  $r$  and  $\tau_i \in \mathbb{R}_+$  ( $i = 1, \dots, r$ ),

$$\mu(\{\tau_i \in (x_i, x_i + dx_i], i = 1, \dots, r\}) = \mu_1(dx_1) \prod_{i=2}^r dF(x_i),$$

where  $\mu_1$  is a boundedly finite measure on  $\mathbb{R}_+$  and  $F$  is a d.f. on  $\mathbb{R}_+$ . Then it follows as in Exercise 9.1.14 that  $\mu$  defines a measure on  $\mathcal{B}(\mathcal{N}_{\mathbb{R}_+}^\#)$ .

Our first aim is to show that when

$$\mu_1(dx_1) = [1 - F(x_1)] dx_1,$$

the measure  $\mu$  is invariant under shifts  $S_u$  for  $u > 0$ , even if it is not a probability measure as in Definition 12.1.II. When the counting measure  $N$  on  $\mathbb{R}_+$  consists of unit atoms at  $Y_1, Y_2, \dots$ , with  $N(0, Y_r] = r$  for  $r = 1, 2, \dots$  as in Exercise 9.1.12, the successive atoms  $\{Y'_n\}$  for the counting measure  $S_u N$  for  $u > 0$  are given by

$$Y'_n = Y_{n+\nu} - u,$$

where the index  $\nu = 0$  if  $Y_1 > u$ ,  $= \sup\{n: Y_n \leq u\}$  otherwise. Consequently, writing  $\tau'_n = Y'_n - Y'_{n-1}$  ( $Y'_0 \equiv 0$ ), the measure  $S_u \mu$  on  $\{Y'_n\}$  is related to  $\mu$  by

$$\begin{aligned} (S_u \mu)(\{\tau'_i \in (x_i, x_i + dx_i], i = 1, \dots, r\}) \\ &= \mu(\{\tau_1 \in (u + x_1, u + x_1 + dx_1], \tau_i \in (x_i, x_i + dx_i], i = 2, \dots, r\}) \\ &\quad + \sum_{j=1}^{\infty} \mu(\{\tau_1 + \dots + \tau_j \leq u, \tau_1 + \dots + \tau_{j+1} \in (u + x_1, u + x_1 + dx_1], \\ &\quad \tau_{j+i} \in (x_i, x_i + dx_i], i = 2, \dots, r\}) \\ &= \left[ \mu(\{\tau_1 \in (u + x_1, u + x_1 + dx_1]\}) + \sum_{j=1}^{\infty} \mu(\{\tau_1 + \dots + \tau_j \leq u, \right. \\ &\quad \left. \tau_1 + \dots + \tau_{j+1} \in (u + x_1, u + x_1 + dx_1]\}) \right] \times \prod_{i=2}^r dF(x_i). \end{aligned}$$

For convenience, integrate  $x_1$  over  $(0, y]$  say, so that on the right-hand side, when  $\mu_1(0, y] = \int_0^y [1 - F(x_1)] dx_1$ , the  $(r-1)$ -fold product of terms  $dF(\cdot)$  has coefficient

$$\int_u^{u+y} [1 - F(x_1)] dx_1 + \sum_{j=1}^{\infty} \int \dots \int [1 - F(t_1)] dt_1 dF(t_2) \dots dF(t_{j+1}),$$

where the multiple integral is over the set  $\{t_1 + \cdots + t_j \leq u < t_1 + \cdots + t_{j+1} \leq u+y\}$ . Writing  $U_0(x) = \sum_{j=1}^{\infty} F^{j*}(x)$ , so that  $U_0$  satisfies the renewal equation

$$U_0(x) = F(x) + \int_0^x F(x-y) dU_0(y),$$

the multiple integral can be expressed as

$$\begin{aligned} & \int_0^y [1 - F(x_1)] dx_1 \int_{u-x_1}^{u-x_1+y} dF(x_2) \\ & \quad + \int_0^u U_0(dv) \int_0^{u-v} [1 - F(x_1)] dx_1 \int_{u-x_1-v}^{u-x_1-v+y} dF(x_2) \\ & = \int_0^u [1 - F(u-x)] dx \int_x^{x+y} dF(z) \\ & \quad + \int_0^u U_0(dv) \int_0^{u-v} [1 - F(u-v-x)] dx \int_x^{x+y} dF(z). \end{aligned}$$

Here, the second term equals

$$\begin{aligned} & \int_0^u dx \int_0^{u-x} [1 - F(u-x-v)] dU_0(v) \int_x^{x+y} dF(z) \\ & = \int_0^u F(u-x) dx \int_x^{x+y} dF(z), \end{aligned}$$

so the coefficient of the  $(r-1)$ -fold product of terms  $dF(\cdot)$  equals

$$\int_u^{u+y} [1 - F(x)] dx + \int_0^u [F(y+x) - F(x)] dx = \int_0^y [1 - F(x)] dx,$$

showing that  $\mu$  is invariant as required.

When  $\int_0^\infty [1 - F(x)] dx \equiv \lambda^{-1} < \infty$ ,  $\lambda\mu(\cdot)$  is a probability measure, and so also is the measure it induces on  $\mathcal{B}(\mathcal{N}_{\mathbb{R}_+}^\#)$ . We can then identify the counting measure  $N(\cdot)$  with such a stationary distribution as a *stationary renewal process*.  $\square$

We use the following proposition in discussing stationary infinitely divisible point processes; the result is of wider importance (see, e.g., Section 3.4 and the discussion of parallel lines in a stationary line process in Section 15.4). An analogue for MPPs is at Exercise 12.2.10.

**Proposition 12.1.VI** (Zero-infinity Dichotomy). *For a stationary random measure  $\xi$  on  $\mathcal{X} = \mathbb{R}^d$ ,*

$$\mathcal{P}\{\xi(\mathcal{X}) = 0 \text{ or } \infty\} = 1. \quad (12.1.17)$$

**PROOF.** The assertion is equivalent to showing that  $\mathcal{P}\{0 < \xi(\mathcal{X}) < \infty\} = 0$ . Supposing the contrary, it necessarily follows that there exist some positive

constants  $a$  and  $\gamma$  such that for the hypercube  $\mathbb{U}_\gamma^d$  with vertices  $\{\pm\frac{1}{2}\gamma, \dots, \pm\frac{1}{2}\gamma\}$  and its complement  $(\mathbb{U}_\gamma^d)^c$ ,

$$\mathcal{P}\{\xi: \xi(\mathbb{U}_\gamma^d) > a, \xi((\mathbb{U}_\gamma^d)^c) < a\} = \alpha > 0.$$

Write  $T_{\gamma r}\mathbb{U}_\gamma^d$  for the shift of  $\mathbb{U}_\gamma^d$  through the vector  $\gamma r = (\gamma r_1, \dots, \gamma r_d)$ , where  $r$  has integer-valued components so  $r \in \mathbb{Z}^d$ , and consider the events

$$V_r = \{\xi: \xi(T_{\gamma r}\mathbb{U}_\gamma^d) > a, \xi(T_{\gamma r}(\mathbb{U}_\gamma^d)^c) < a\}.$$

By stationarity,  $\mathcal{P}(V_r) = \mathcal{P}(V_0) = \alpha$  for all such  $r$ , and because the events  $V_r$  are disjoint for distinct  $r$ ,

$$\mathcal{P}\left(\bigcup_{r \in \mathbb{Z}^d} V_r\right) = \sum_{r \in \mathbb{Z}^d} \mathcal{P}(V_r) = \infty \cdot \alpha,$$

which is impossible when  $\mathcal{P}$  is a probability measure unless  $\alpha = 0$ . □

Equation (12.1.17) prompts the following definition.

**Definition 12.1.VII.** A random measure  $\xi$  is nonnull when  $\mathcal{P}\{\xi = \emptyset\} = 0$ .

It follows from (12.1.17) that a nonnull stationary random measure on  $\mathcal{X} = \mathbb{R}^d$  satisfies  $\mathcal{P}\{\xi(\mathcal{X}) = \infty\} = 1$ .

The discussion so far has centred on invariance with respect to shifts in  $\mathbb{R}^d$ , but, as mentioned earlier, the ideas can be carried over with only nominal changes to processes invariant under other types of transformation, such as rotations, permutations of coordinates, or changes of scale. To conclude this section we examine one such example where, as in  $\mathbb{R}^d$ , the state space itself is the group.

In such cases we should anticipate that a basic role will be played by Haar measure which, analogous to the uniform distribution on the circle, or Lebesgue measure on the line, is the unique measure on the group invariant under the group actions.

In the case of a Poisson process, for example, the properties of the process are determined by the parameter measure, which inherits the property of being invariant under the group actions from invariance under the corresponding flow. But the only measures invariant under the group actions are multiples of the Haar measure, and so the parameter measure itself must be a multiple of Haar measure [recall Examples 12.1(a) and (c)].

Even when there is no obvious governing measure, Haar measure will reappear in the moment measures, and lurks in the background behind the finite-dimensional distributions. Its role in the latter context can be seen most clearly when the state space is compact as in the next example.

**EXAMPLE 12.1(f)** *Stationary point process on the circle*  $\mathbb{S}$ . For a point process with state space the circle  $\mathbb{S}$ , which we identify with angles  $\theta$  modulo  $2\pi$ , the compactness of  $\mathbb{S}$  implies that the process necessarily has a.s. finite

realizations, so explicit constructions in terms of Janossy measures are possible. Thus, supposing that the realization consists of exactly  $n$  points, defined by angles  $\{\theta_1, \dots, \theta_n\}$ , its distribution can be described by the symmetrized probability measure (conditional on  $n$ )

$$\Pi_n(d\theta_1 \times \dots \times d\theta_n) = \frac{J_n(d\theta_1 \times \dots \times d\theta_n)}{J_n(\mathbb{S}^{(n)})}.$$

Stationarity (invariance under rotations) implies that for all  $\theta \in \mathbb{S}$  and  $A_1, \dots, A_n \in \mathcal{B}(\mathbb{S})$ ,

$$\Pi_n(T_\theta A_1 \times \dots \times T_\theta A_n) = \Pi_n(A_1 \times \dots \times A_n)$$

so that we again have invariance under diagonal shifts. Here Lemma 12.1.IV implies that  $\Pi_n$  can be written in terms of a product of the uniform measure on  $\mathbb{S}$  and a reduced probability measure  $\check{\Pi}_n$  on a space of  $n - 1$  arguments  $\phi_1, \dots, \phi_{n-1}$ : for  $g \in \text{BM}(\mathbb{S}^{(n)})$  we have

$$\begin{aligned} \int_{\mathbb{S}^{(n)}} g(\theta_1, \dots, \theta_n) \Pi_n(d\theta_1 \times \dots \times d\theta_n) \\ = \int_{\mathbb{S}} \frac{d\theta}{2\pi} \int_{\mathbb{S}^{(n-1)}} g(\theta, \theta + \phi_1, \dots, \theta + \phi_{n-1}) \check{\Pi}_n(d\phi_1 \times \dots \times d\phi_{n-1}). \end{aligned} \quad (12.1.18)$$

The interpretation of this result is quite simple. If the distribution  $\Pi_n$  of  $n$  points is rotationally invariant, it can be described by locating one point uniformly around the circle and the other  $n - 1$  points relative to it according to the reduced distribution  $\check{\Pi}_n$ . The symmetry properties of  $\Pi_n$  imply that it is immaterial which point is designated as the one to be uniformly distributed, and stationarity (i.e., rotational invariance) implies that it is immaterial which point of the circle is chosen to play the role of origin. For example, if  $n = 2$  and densities exist, the distribution of the two points is completely described by a symmetrical density function  $f(\cdot)$  such that

$$\Pi_2(d\theta_1 \times d\theta_2) = (2\pi)^{-1} f(\theta_2 - \theta_1) d\theta_1 d\theta_2.$$

Note that here, as in general, it is a necessary consequence of stationarity that any one-dimensional marginal distribution such as  $\Pi_2(\cdot \times \mathbb{S})$  must be uniform.

As a more specific example of such a process, consider first any symmetric distribution  $g(\theta)$  about the origin (pole)  $\theta = 0$  [see, e.g., Mardia and Jupp (2000) for examples]. Take any fixed or random number of points independently distributed about the origin, to form the circular analogue of a Neyman–Scott cluster,  $N(\cdot \mid 0)$  say. Then shift the origin to an angle uniformly distributed over  $\mathbb{S}$ . This is already a single-cluster, stationary process on  $\mathbb{S}$ . Finally, consider the superposition of  $N'$  such processes, where  $N'$  is Poisson distributed with mean  $\nu$ . The result is a cluster process on  $\mathbb{S}$  analogous to a Neyman–Scott process in time or space.

In this case, unfortunately, the reduction of  $\Pi_n(\cdot)$  to  $\check{\Pi}_n(\cdot)$  defined by (12.1.18) is of little direct value in computing its properties. Even if the realization consists of only two points, the density function  $f$  for the angular separation of the two points will be an awkward mixture of densities that arise from pairs of points coming from either single or different clusters.  $\square$

For stationary point processes in general, the locations of points relative to a given point of the process as origin are independent of where that point itself is located. This is the theme of the Palm theory for stationary point processes discussed in Section 13.3.

Often the main difficulty in applying the group concepts relates to the fact that the group  $\mathcal{G}$  of transformations may split the space into equivalence classes in quite a complex manner. By contrast, the shifts act transitively on the whole space (any point can be transformed by a member of  $\mathcal{G}$  into any other point) so that the equivalence classes are trivial, the whole space forming the unique equivalence class.

Marked point processes on  $\mathbb{R}^d$  form the canonical example of the sort of structure to be expected in more general cases. Here the state space has the representation  $\mathcal{X} = \mathbb{R}^d \times \mathcal{K}$ , in which the first factor is the group and the second can be regarded as a representation of the space of equivalence classes. This product form is the desired endpoint of analyses based on Lemma A2.7.II and Proposition A2.7.III. Any measure on  $\mathcal{X}$  that is invariant under the group actions can then be expressed as the product of Haar measure on the group and a measure on the other component  $\mathcal{K}$ . Of course the probability measure defining the point process does not live on  $\mathcal{X}$  itself, but on  $\mathcal{N}_{\mathcal{X}}^{\#}$ , but once again the underlying factorization of measures on  $\mathcal{X}$  generally carries with it some corresponding simplifications of the probability distributions and the moment measures. To illustrate, we consider an extension of the previous example to the marked case.

**EXAMPLE 12.1(g)** *A stationary MPP on  $\mathbb{S}$ .* To extend Example 12.1(f) to an MPP, start by supposing that the number  $n$  of points in the ground process is fixed, where the ground process is again specified by locating an initial point uniformly at random around the circle, and then locating the other points relative to it according to a reduced  $(n - 1)$ -dimensional symmetric distribution  $\check{\Pi}_n(\phi_1, \dots, \phi_{n-1})$  say. To take the specific case when  $n = 2$  as an example, the process can be specified by two components:

- (i) a distribution  $F(\phi)$  for the angular separation  $\phi$  (in a given direction, clockwise say); and
- (ii) a family of symmetric bivariate distributions,  $G_2(K_1, K_2 \mid \phi)$  say, for the marks (with  $K_i \in \mathcal{B}_{\mathcal{K}}$  for a mark space  $\mathcal{K}$  that is a c.s.m.s.), given the angular separation  $\phi$ .

More generally, the associated distribution of marks can be specified by a family of  $n$ -dimensional symmetric distributions on  $\mathcal{K}$ ,  $G_n(K_1, \dots, K_n \mid \theta, \theta + \phi_1, \dots, \theta + \phi_{n-1})$  say, indexed by the angular locations; in the stationary case each  $G_n$  is independent of  $\theta \in \mathbb{S}$ . The simplest case is that of independent

marks, but in general the joint distribution of the set of marks may depend on both the number and relative angles of the points in the ground process. Symmetry implies that the marginal distributions of the multivariate mark distribution are equal, and for a fixed number of points this can be taken as defining what is meant by the stationary mark distribution. However, in the general case of a random number of points, the marginal distributions may also depend on the value of  $n$ , so that the stationary mark distribution appears as a weighted average of the stationary mark distributions for realizations with different numbers of points (see Exercise 12.1.10).  $\square$

## Exercises and Complements to Section 12.1

12.1.1 (a) Modify the argument leading to Lemma 12.1.I to show that when  $\mathcal{X} = \mathbb{R}^d$ ,  $S_u \mu$  is jointly continuous as a mapping from  $\mathcal{M}_{\mathcal{X}}^{\#} \times \mathcal{X}$  into  $\mathcal{M}_{\mathcal{X}}^{\#}$ .

(b) Show that  $\hat{S}_u$  defined at (12.1.3) is a continuous and hence measurable mapping of the space  $\mathcal{M}^{\#}(\mathcal{M}_{\mathcal{X}}^{\#})$  of boundedly finite measures on  $\mathcal{M}_{\mathcal{X}}^{\#}$  into itself, and that  $\hat{S}_u$  preserves measure and hence maps the set of all probability measures on  $\mathcal{M}_{\mathcal{X}}^{\#}$  into itself. Verify that  $\hat{S}_u$  acts on the space  $\mathcal{M}^{\#}(\mathcal{N}_{\mathcal{X}}^{\#})$  of all boundedly finite measures on  $\mathcal{N}_{\mathcal{X}}^{\#}$  in a similar way.

[Hint:  $\hat{S}_u$  inherits the properties of  $S_u$  in much the same way as  $S_u$  inherits the properties of  $T_u$ . Continuity depends ultimately on the upper continuity of a measure at the empty set (Proposition A1.3.II).]

12.1.2 Give examples of nonstationary point processes for which (a) the avoidance function is stationary; (b) the one-dimensional distributions are stationary.

[Hint: For integer-valued r.v.s  $X$ ,  $Y$ , and  $X + Y$ , with  $X, Y \geq 0$  a.s., find a bivariate distribution for dependent  $X$  and  $Y$  with the same marginal distribution for  $X + Y$  as though  $X$ ,  $Y$  are independent. Take  $\mathcal{X} = \mathbb{Z}$  and define the fidi distributions of a point process by using the dependent and independent bivariate distributions for alternate pairs of integers (see Ripley, 1976).]

12.1.3 *Stationarity conditions for marked point processes.* Verify that the following conditions for MPPs on  $\mathbb{R}^d$  (i.e., for point processes on state space  $\mathbb{R}^d \times \mathcal{K}$ ), are equivalent, each corresponding to stationarity.

(i) For each  $u \in \mathbb{R}^d$ ,  $k = 1, 2, \dots$ , and families  $A_1, \dots, A_k \in \mathcal{B}_{\mathbb{R}^d}$  and  $K_1, \dots, K_k \in \mathcal{B}_{\mathcal{K}}$ , the fidi distributions satisfy

$$\begin{aligned} \mathcal{P}(\{N(A_i \times K_i) = n_i \ (i = 1, \dots, k)\}) \\ &= P_k(A_1 \times K_1, \dots, A_k \times K_k; n_1, \dots, n_k) \\ &= P_k((A_1 + u) \times K_1, \dots, (A_k + u) \times K_k; n_1, \dots, n_k). \end{aligned}$$

(ii) For each  $u \in \mathbb{R}^d$  and  $f \in \text{BM}(\mathbb{R}^d \times \mathcal{K})$ , with  $S_u f(x, \kappa) = f(x - u, \kappa)$ , the characteristic functional satisfies

$$\Phi[S_u f] = \Phi[f].$$

(iii) For each  $u \in \mathbb{R}^d$  and  $h \in \mathcal{V}(\mathbb{R}^d \times \mathcal{K})$ , the p.g.fl.  $G$  satisfies

$$G[S_u h] = G[h].$$



12.1.4 Starting from a Cox process  $N$  which with its directing measure  $\xi$  satisfies (12.1.9), use the p.g.fl. of  $N$  and the Laplace functional  $L_\xi$  of  $\xi$ , which are related by (6.2.3) of Proposition 6.2.II, to verify that a Cox process  $N$  on  $\mathbb{R}^d$  directed by  $\xi$  is stationary if and only if  $\xi$  is stationary. Using a similar approach show that the joint process  $(N, \xi)$  is stationary.

[Hint: Stationarity means invariance as indicated at (12.1.4). Check that this holds if and only if  $\xi$  is invariant.]

12.1.5 The following two examples show that a stationary cluster process can be realized from nonstationary components.

- (a) Random thinning with deletion probability  $\mu(x)/[1 + \mu(x)]$  at  $x \in \mathbb{R}^1$  of an inhomogeneous Poisson process at rate  $[1 + \mu(x)] dx$ , where  $\mu(x) \geq 0$  (all  $x$ ), yields a stationary Poisson process (cf. Exercise 11.3.1).
- (b) Take a simple point process on  $\mathbb{R}$  with points at  $\{2n + U: n = 0, \pm 1, \dots\}$ , where the r.v.  $U$  is uniformly distributed on  $(0, 1)$ , to be a (nonstationary) cluster centre process. Let clusters be independent and let them consist of precisely two points at distances  $X_1$  and  $1 + X_2$  from the cluster centre, where for each cluster  $X_1$  and  $X_2$  are i.i.d. r.v.s. Then the cluster process so constructed is the same as the random translation of a stationary deterministic process at unit rate.

12.1.6 Let  $N$  be a cluster process (see Definition 6.3.I) with stationary centre process  $N_c$  on  $\mathbb{R}^d$  and independent component processes  $N_m(\cdot | y)$  ( $y \in \mathbb{R}^d$ ) for which the fidi distributions of  $N_m(\cdot | y)$ , relative to  $y$ , are independent of  $y$ . Denote the p.g.fl. of  $N_m$  by  $G_m[\cdot | y]$ . Referring to Lemma 6.3.II and Exercise 6.3.2, show that a stationary cluster process  $N$  is well defined if and only if

$$\int_{\mathbb{R}^d} (1 - G_m[h | y]) N_c(dy) < \infty \quad \text{a.s.} \quad (h \in \mathcal{V}(\mathbb{R}^d)). \quad (12.1.19)$$

[Hint: Homogeneity of the  $N_m$  means that  $G_m[h(\cdot) | y] = G_m[h(\cdot + y) | 0]$ .]

12.1.7 *Stationary deterministic lattice processes* [see Example 8.2(e) for the case  $d = 1$ ]. Let the r.v.  $Y$  be uniformly distributed over the unit cube  $\mathbb{U}^d$  in  $\mathbb{R}^d$ , and let  $\mathbb{Z}^d$  denote the set of all integer-valued lattice points in  $\mathbb{R}^d$ . Show that the point process  $N$  with sample realizations  $\{n + Y: n \in \mathbb{Z}^d\}$  is stationary. (Call  $N$  the *stationary cubic lattice process* at unit rate in  $\mathbb{R}^d$ .) If the span of the lattice in the direction of the  $x_i$ -axis,  $i = 1, \dots, d$ , is changed from 1 to  $a_i$ , where the positive reals  $a_i$  satisfy  $\prod_{i=1}^d a_i = 1$ , verify that stationarity at unit rate is retained.

12.1.8 (a) Let  $f(\cdot)$  be a nonnegative measurable function on  $\mathbb{R}$  satisfying for each fixed  $u \in \mathbb{R}$

$$f(x + u) = f(x) \quad (\text{a.e. } x). \quad (12.1.20)$$

Show that there exists a finite constant  $\alpha$  such that  $f(x) = \alpha$  a.e. [Hint:  $F(y) = \int_0^y f(x) dx$  satisfies the Hamel equation  $F(x + y) = F(x) + F(y)$ .]

- (b) Extend the result of (a) to  $\mathbb{R}^d$ . [Hint: Apply (a) in a coordinatewise manner, deducing at the first step, for example, that in place of the constant  $\alpha$  is a measurable function  $\alpha(x_{d-1})$  ( $x_{d-1} \in \mathbb{R}^{d-1}$ ) satisfying (12.1.20).]

12.1.9 *Radon–Nikodym approach to construction of stationary cluster elements.*

- (a) Check that each of the measures  $K_k^1(\cdot) \equiv K_k(\cdot \times \mathcal{X}^{(k-1)})$  in Proposition 12.1.V reduces to a multiple of Lebesgue measure.
- (b) Define the Radon–Nikodym derivatives  $P_{k-1}(\cdot | x)$  as in the discussion under (6.3.34) by

$$\int_A P_{k-1}(B | x) K_k(dx \times \mathcal{X}^{(k-1)}) = K_k(A \times B) \quad (A \in \mathcal{B}_{\mathcal{X}}, B \in \mathcal{B}(\mathcal{X}^{(k-1)}))$$

and observe that, for each fixed  $u$ ,  $P_{k-1}(T_u A_2 \times \cdots \times T_u A_k | x + u)$  and  $P_{k-1}(A_2 \times \cdots \times A_k | x)$  are versions of the same density and hence equal a.e.

- (c) For fixed  $A_2, \dots, A_k$ , show that the function  $P_{k-1}(T_u A_2 \times \cdots \times T_u A_k | u)$  in part (b) is a measurable function of  $u$ , implying by Exercise 12.1.8(b) that it reduces a.e. to a constant  $P_{k-1}(A_2 \times \cdots \times A_k | 0)$ .  
[Hint: For fixed  $u$ , the Radon–Nikodym theorem shows that  $P_{k-1}(T_u A_2 \times \cdots \times T_u A_k | x + u) = P_{k-1}(A_2 \times \cdots \times A_k | x) K_k^1$ -a.e.  $x$ . Integrate  $K_k(T_u A_1 \times \cdots \times T_u A_k)$  over  $u$  and use Fubini's theorem to express the result as an integral whose density with respect to the product measure  $du \times dx$  is  $P_{k-1}(T_u A_2 \times \cdots \times T_u A_k | x + u)$ , thereby showing via the Radon–Nikodym theorem its joint measurability in  $x$  and  $u$ . Hence, by putting  $x = 0$ , deduce that  $P_{k-1}(T_u A_2 \times \cdots \times T_u A_k | u)$  is a measurable function of  $u$  that is a.e. equal to  $P_{k-1}(A_2 \times \cdots \times A_k | 0)$ .]
- (d) Take a countable semiring  $\mathcal{A}$  generating  $\mathcal{B}(\mathbb{R}^d)$  and show that  $P_{k-1}$  is countably additive on product sets of the form  $A_2 \times \cdots \times A_k$  for  $A_i \in \mathcal{A}$  and so can be extended uniquely to a measure  $\tilde{P}_{k-1}$  on  $\mathcal{B}(\mathbb{R}^{(k-1)d})$  such that for all product sets with  $A_i \in \mathcal{B}(\mathbb{R}^d)$ ,

$$P_{k-1}(T_u A_2 \times \cdots \times T_u A_k | u) = \tilde{P}_{k-1}(A_2 \times \cdots \times A_k) \quad \text{a.e.}$$

- 12.1.10 In the setting and notation of Examples 12.1(f)–(g), put  $\pi_n = \Pr\{N(\mathbb{S} \times \mathcal{K})\}$ . Verify that the stationary mark distribution, for  $K \in \mathcal{B}_{\mathcal{K}}$ , equals

$$\frac{1}{\sum_{n=1}^{\infty} n \pi_n} \sum_{n=1}^{\infty} n \pi_n \int_{\mathbb{S}} \frac{d\theta}{2\pi} \int_{\mathbb{S}^{(n-1)}} G_n(K, \mathcal{K}, \dots, \mathcal{K} | \theta, \theta + \phi_1, \dots, \theta + \phi_{n-1}) \check{\Pi}_n(d\phi_1 \times \cdots \times d\phi_{n-1}).$$

- 12.1.11 *Renewal process and random walk on  $\mathbb{S}$ .* Suppose given a probability distribution  $G(d\theta)$  on  $(0, 2\pi]$ , interpreted as the length of a step in the clockwise direction around the circumference of a circle, and for  $n = 1, 2, \dots$  let  $U_n(A)$  denote the expected number of visits within the first  $n$  steps to the set  $A \subseteq (0, 2\pi]$ . Find conditions on  $G$  such that

$$U_n(A)/n \rightarrow 2\pi \ell(A)/\bar{\theta},$$

where  $\bar{\theta} = \int_0^{2\pi} \theta G(d\theta)$  is necessarily finite and bounded by  $2\pi$ . Investigate the behaviour when the conditions fail. [Hint: Formulate versions of the direct Riemann integrability and spread-out conditions of Section 4.4 for  $G$ , and apply the results for the real line, then wrap around the circle.]

## 12.2. Ergodic Theorems

In this section we review some basic ergodic theorems and develop them for random measures and point processes. There are diverse examples of their application through the rest of this chapter and the next, and significant extensions of the theory in Sections 13.4–5.

Let  $(\Omega, \mathcal{E}, \mu)$  be a measure space and  $S$  a measure-preserving operator on this space; that is,  $\mu(S^{-1}E) = \mu(E)$  for  $E \in \mathcal{E}$ . The classical ergodic theorems assert the convergence, in some sense and under appropriate conditions, of the averages  $n^{-1} \sum_{r=1}^n f(S^r \omega)$  to a limit function  $\bar{f}(\omega)$ , which is invariant under the action of  $S$  [i.e.,  $\bar{f}(S\omega) = \bar{f}(\omega)$ ] for a measurable function  $f$ . When  $f$  is  $\mu$ -integrable, the limit function  $\bar{f}$  is also  $\mu$ -integrable and the *individual ergodic theorem* asserts convergence  $\mu$ -a.e. When  $f \in L_p(\mu)$  for some  $1 \leq p < \infty$ ,  $\bar{f} \in L_p(\mu)$  also and the *statistical ergodic theorem* asserts convergence in the  $L_p$  norm. When  $\mu$  is a probability measure, the limit function  $\bar{f}(\omega)$  is a random variable that can be identified with the conditional expectation of  $f$  with respect to the  $\sigma$ -algebra  $\mathcal{I}$  of *invariant events* under  $S$ , that is, of those sets  $E \in \mathcal{E}$  for which  $\mu(S^{-1}E \triangle E) = 0$ . Writing  $X$  for  $f(\omega)$ ,  $X_n$  for  $f(S^n \omega)$ , and  $Y_X = E(X | \mathcal{I})$  for  $\bar{f}$ , the individual ergodic theorem in the probability case can be written more graphically in the form

$$\frac{1}{n} \sum_{r=1}^n X_r \xrightarrow{\text{a.s.}} E(X | \mathcal{I}) \equiv Y_X. \quad (12.2.1)$$

An important special case arises when the probability measure is such that the events in  $\mathcal{I}$  all have probability measure either 0 or 1. In this case the transformation  $S$  is said to be *metrically transitive* with respect to the measure  $\mu$ , and the process  $\{X_n\}$ , or its distribution, is said to be *ergodic*. In such circumstances the only invariant functions are constants, the conditional expectation in (12.2.1) reduces to the ordinary expectation, and (12.2.1) takes the familiar form

$$\frac{1}{n} \sum_{r=1}^n X_r \xrightarrow{\text{a.s.}} m \equiv EX.$$

For a fuller discussion of these results with proofs and references, see, for example, Billingsley (1965).

One other prefatory remark is in order. Given a stationary process  $\{X(t): t \in \mathbb{R}\}$ , define the two  $\sigma$ -fields  $\mathcal{I}_1$  and  $\mathcal{I}$  of sets  $E \in \mathcal{E}$  that are invariant under the shift transformations  $\{S_n: n = 0, \pm 1, \dots\}$  and  $\{S_t: t \in \mathbb{R}\}$ , respectively. In general  $\mathcal{I} \neq \mathcal{I}_1$ , with  $\mathcal{I} \subseteq \mathcal{I}_1$ ; of course, if  $\mathcal{I}_1$  is trivial, then so is  $\mathcal{I}$ . This and the consequences of the sandwich relation below cover our main concerns.

We consider first the implications of these results for stationary random measures on  $\mathbb{R}$ . Here we take  $\Omega = \mathcal{M}_{\mathbb{R}}^{\#}$  and  $S$  as the shift through the unit distance. The measure-preserving character of  $S$  is then a corollary of

stationarity. The simplest choice for  $X$  is the random variable  $X = \int_0^1 \xi(dx)$ , which has finite expectation whenever  $\xi$  has finite mean intensity. Then

$$X_n = \int_0^1 \xi(n+dx) = \int_n^{n+1} \xi(dx)$$

and the assertion in (12.2.1) becomes

$$\frac{\xi(0, n]}{n} \xrightarrow{\text{a.s.}} \mathbb{E} \left( \int_0^1 \xi(dx) \mid \mathcal{I} \right). \quad (12.2.2)$$

If, in particular,  $\xi$  is ergodic then

$$\frac{\xi(0, n]}{n} \xrightarrow{\text{a.s.}} m. \quad (12.2.3)$$

The results (12.2.2) and (12.2.3) seem simple, but they can be applied to many more general situations of which the simplest is to a continuous-time process. Observe first that from the simple sandwich relation

$$\frac{[T]}{T} \cdot \frac{\xi(0, [T])}{[T]} \leq \frac{\xi(0, T]}{T} \leq \frac{[T+1]}{T} \cdot \frac{\xi(0, [T+1])}{[T+1]}$$

we easily extend (12.2.2) to arbitrary intervals as for Proposition 3.5.I, so that

$$\frac{\xi(0, T]}{T} \xrightarrow{\text{a.s.}} \mathbb{E} \left( \int_0^1 \xi(dx) \mid \mathcal{I} \right). \quad (12.2.4)$$

Because the limit is invariant under all shifts  $\{S_t: t \in \mathbb{R}\}$ , it is  $\mathcal{I}$ -measurable rather than just  $\mathcal{I}_1$ -measurable, so that the conditional expectation can and will be taken with respect to  $\mathcal{I}$ .

As a corollary, consider the behaviour of a nonnegative measurable function  $f(\cdot)$  on  $\mathbb{R}$ , applied to a stationary measurable stochastic process  $X(\cdot)$  on  $\mathbb{R}$ . If  $\mathbb{E}[f(X(t))] < \infty$ , we can define a random measure  $\xi$  with finite mean intensity by setting

$$\xi(A) = \int_A f(X(t)) dt.$$

Applying (12.2.4) to such  $\xi$  yields the result

$$\frac{1}{T} \int_0^T f(X(t)) dt \xrightarrow{\text{a.s.}} \mathbb{E}[f(X(t)) \mid \mathcal{I}].$$

The only restrictive feature is the limitation to nonnegative functions  $f$ : this is not inherent in the ergodic problem but arises from our concern with random measures rather than random signed measures.

Similar results hold in higher-dimensional spaces and in the more general context of metric groups considered at the end of Section 12.1. The main point of difficulty concerns the choice of averaging sets to replace the intervals  $(0, n]$  in (12.2.2). Even in the plane it is not difficult to find sequences  $\{A_n\}$  with  $A_n \subset A_{n+1}$  and  $\ell(A_n) \rightarrow \infty$  such that the analogue of (12.2.2) fails in some cases (see Exercise 12.2.1). To consider this question further, let  $(\Omega, \mathcal{E}, \mu)$  be a measure space acted on measurably by the group of measurable transformations  $\{S_g: g \in \mathcal{G}\}$ , meaning that  $(g, \omega) \mapsto S_g \omega$  is jointly measurable, where  $\mathcal{G}$  is a  $\sigma$ -group with unique right-invariant Haar measure  $\chi$ . Note the most important fact that the averaging in ergodic theorems takes place over sets in  $\mathcal{G}$  and not the state space  $\mathcal{X}$ . For example, the individual ergodic theorem takes the form that, for suitable sequences  $\{A_n\}$ ,

$$\frac{\int_{A_n} f(S_g \omega) \chi(dg)}{\chi(A_n)} \rightarrow \bar{f}(\omega) \quad \mu\text{-a.e.}, \quad (12.2.5)$$

where, in the probability case,  $\bar{f}(\omega)$  is the conditional expectation  $E(f | \mathcal{I})$  with respect to the  $\sigma$ -algebra of events invariant under the whole family  $\{S_g: g \in \mathcal{G}\}$ .

A thorough discussion of extensions of the classical ergodic theorems in this context is given by Tempel'man (1972) [see also Tempel'man (1986) and Sinai (2000, Chapter 4, Section 3.3)] who sets out a range of conditions on the sequence  $\{A_n\}$ —some necessary, others sufficient—for the validity both of (12.2.5) and of corresponding statistical ergodic theorems. For the present discussion we adopt only the simplest of the conditions he describes.

**Definition 12.2.I.** Let  $\mathcal{X} = \mathbb{R}^d$ . The sequence  $\{A_n\}$  of bounded Borel sets in  $\mathbb{R}^d$  is a convex averaging sequence if

- (i) each  $A_n$  is convex;
- (ii)  $A_n \subseteq A_{n+1}$  for  $n = 1, 2, \dots$ ; and
- (iii)  $r(A_n) \rightarrow \infty$  ( $n \rightarrow \infty$ ), where  $r(A) = \sup\{r: A \text{ contains a ball of radius } r\}$ .

Using this terminology, we set out versions of the individual and statistical ergodic theorems, referring to Tempel'man (1972) for proofs and further extensions.

**Proposition 12.2.II.** (a) (Individual Ergodic Theorem for  $d$ -dimensional Shifts). Let  $(\Omega, \mathcal{E}, \mathcal{P})$  be a probability space,  $\{S_x: x \in \mathbb{R}^d\}$  a group of measure-preserving transformations acting measurably on  $(\Omega, \mathcal{E}, \mathcal{P})$  and indexed by the points of  $\mathbb{R}^d$ ,  $\{A_n: n = 1, 2, \dots\}$  a convex averaging sequence in  $\mathbb{R}^d$ , and  $\mathcal{I}$  the  $\sigma$ -algebra of events in  $\mathcal{E}$  that are invariant under the transformations  $\{S_x\}$ . Then for all measurable functions (random variables)  $f$  on  $(\Omega, \mathcal{E}, \mathcal{P})$  with  $E(|f|) < \infty$ ,

$$\frac{\int_{A_n} f(S_x \omega) dx}{\ell(A_n)} \xrightarrow{\text{a.s.}} E(f | \mathcal{I}). \quad (12.2.6)$$

(b) (Statistical Ergodic Theorem for  $d$ -dimensional Shifts). *Under the same conditions as in (a) and for  $p \geq 1$ ,*

$$\mathbb{E} \left| \frac{\int_{A_n} f(S_x \omega) dx}{\ell(A_n)} - \mathbb{E}(f \mid \mathcal{I}) \right|^p \rightarrow 0 \quad (\text{all } f \in L_p(\mathcal{P})). \quad (12.2.7)$$

*Remark.* In general the statistical ergodic theorem holds under weaker conditions on the sequence  $\{A_n\}$  than the individual ergodic theorem. Versions of the theorem remain true when the probability measure  $\mathcal{P}$  is replaced by a  $\sigma$ -finite measure  $\mu$ , subject of course to the condition that  $\int |f(\omega)| \mu(d\omega) < \infty$ ; see Proposition 12.4.V for an application.

Our task is to apply these theorems to stationary random measures on the c.s.m.s.  $\mathcal{X}$ : we consider two cases,  $\mathcal{X} = \mathbb{R}^d$  and  $\mathcal{X} = \mathbb{R}^d \times \mathcal{K}$ , for unmarked and marked point processes, respectively, where the c.s.m.s.  $\mathcal{K}$  is a space of marks.

When  $\mathcal{X} = \mathbb{R}^d$ , we identify  $(\Omega, \mathcal{E})$  with the space  $(\mathcal{M}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#}))$  of boundedly finite measures  $\xi$  defined on  $\mathcal{B}_{\mathcal{X}}$ , and  $S_x$  with the shift taking  $\xi(\cdot)$  into  $\xi(\cdot + x)$ . If  $\xi$  has finite first moment measure, stationarity requires that this should reduce to a constant multiple  $m\ell(\cdot)$  of Lebesgue measure on  $\mathbb{R}^d$ .

More generally, if  $\mathcal{X} = \mathbb{R}^d \times \mathcal{K}$ , we still take  $(\Omega, \mathcal{E}) = (\mathcal{M}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#}))$  but identify  $\{S_x\}$  with shifts in the first coordinate only. Under stationarity, the first moment measure becomes a measure on the product space and it is invariant under shifts in the first component. The factorization Lemma A2.7.II then implies that the first moment measure has the product form  $\ell \times \nu$ , where  $\nu$  is a boundedly finite measure on  $\mathcal{K}$ . If the ground process has finite first moment measure, then it must be a multiple  $m_g \ell(\cdot)$  of Lebesgue measure on  $\mathbb{R}^d$ . In this case,  $\nu(\mathcal{K}) < \infty$  and  $\nu$  can be normalized to a probability measure  $\pi(\cdot)$  on  $(\mathcal{K}, \mathcal{B}_{\mathcal{K}})$ , the *stationary mark distribution*. The first moment measure is then of the form  $m_g \ell \times \pi$ .

We proceed to an extension of these remarks to the conditional expectation of  $\xi$  with respect to the appropriate invariant  $\sigma$ -algebra  $\mathcal{I}$ .

**Lemma 12.2.III.** *Let  $\xi$  be a random measure on the product space  $\mathcal{X} = \mathbb{R}^d \times \mathcal{K}$  and  $\mathcal{I}$  the  $\sigma$ -algebra of invariant events with respect to the shifts  $S_x$  in  $\mathbb{R}^d$ . When  $\xi$  is stationary with respect to these shifts and such that its expectation measure exists, there exists an  $\mathcal{I}$ -measurable random measure  $\psi(\cdot)$  on  $\mathcal{K}$  such that for all nonnegative measurable functions  $f$  on  $\mathcal{X}$ ,*

$$\mathbb{E} \left( \int_{\mathbb{R}^d \times \mathcal{K}} f(x, \kappa) \xi(dx \times d\kappa) \mid \mathcal{I} \right) = \int_{\mathcal{K}} \psi(d\kappa) \int_{\mathbb{R}^d} f(x, \kappa) \ell(dx) \quad \mathcal{P}\text{-a.s.} \quad (12.2.8)$$

*In particular, for bounded  $B \in \mathcal{B}(\mathbb{R}^d)$  and  $K \in \mathcal{B}_{\mathcal{K}}$ ,*

$$\mathbb{E}(\xi(B \times K) \mid \mathcal{I}) = \ell(B) \psi(K) \quad \mathcal{P}\text{-a.s.} \quad (12.2.9)$$

PROOF. Let  $X$  be any r.v. on  $(\Omega, \mathcal{E}, \mathcal{P})$  with finite expectation and  $G \in \mathcal{I}$  be an invariant set, so that  $\mathcal{P}(G \triangle S_x G) = 0$ . Then

$$\begin{aligned} \int_G X(\omega) \mathcal{P}(d\omega) &= \int_{S_x G} X(\omega) \mathcal{P}(d\omega) = \int_G X(S_{-x}\omega) \mathcal{P}(d(S_{-x}\omega)) \\ &= \int_G X(S_{-x}\omega) \mathcal{P}(d\omega), \end{aligned}$$

so for all  $x \in \mathbb{R}^d$ ,

$$\mathbb{E}(X \mid \mathcal{I}) = \mathbb{E}(S_x X \mid \mathcal{I}) \quad \mathcal{P}\text{-a.s.} \quad (12.2.10)$$

Take  $X = \xi(A)$ , and recall from Proposition 9.1.XV that there is a version of the conditional expectation,  $\mathbb{E}(\xi(\cdot) \mid \mathcal{I}) \equiv \eta(\cdot)$  say, which is again a random measure. Then (12.2.10) asserts that

$$\eta(S_x A) = \eta(A) \quad \mathcal{P}\text{-a.s.}$$

Take  $A$  of the form  $B \times K$  as at (12.2.9), and let  $B$  and  $K$  run through the members of countable rings generating  $\mathcal{B}(\mathbb{R}^d)$  and  $\mathcal{B}_{\mathcal{K}}$ , respectively, and  $x$  through a countable dense set in  $\mathbb{R}^d$ . Because only a countable family of null sets is involved, we can assume that (12.2.10) holds simultaneously for all such  $B$ ,  $K$ ,  $x$ , and for  $\omega$  outside a single set  $V$  with  $\mathcal{P}(V) = 0$ . For  $\omega \notin V$  it now follows from Lemma A2.7.II that

$$\eta(B \times K, \omega) = \ell(B) \psi(K, \omega)$$

for some kernel  $\psi$  on  $\mathcal{K} \times \Omega$ . But  $\eta(\cdot)$  was chosen to be an  $\mathcal{I}$ -measurable random measure, so for each  $K$  the left-hand side is an  $\mathcal{I}$ -measurable r.v. (more precisely, it can be extended to all  $\omega \in \Omega$ , in such a way as to form such a r.v.). Also, for a fixed  $\omega \notin V$ ,  $\psi(K, \omega)$  is countably additive and its extension to  $V$  can be constructed so as to retain this property. Thus,  $\psi(\cdot)$  is a random measure on  $\mathcal{K}$ , from Proposition 9.1.VIII. This establishes (12.2.9), and (12.2.8) follows by standard extension arguments.  $\square$

Applying the definition in this lemma to a stationary MPP leads to an analogue of Proposition 12.1.VI (see Exercise 12.2.10).

When  $\mathcal{X} = \mathbb{R}^d$  in Lemma 12.2.III,  $\mathcal{K}$  reduces to a single point, and thus the random measure  $\psi$  is then an  $\mathcal{I}$ -measurable random variable

$$Y = \mathbb{E}(\xi(\mathbb{U}^d) \mid \mathcal{I}),$$

where  $\mathbb{U}^d$  is the unit cube in  $\mathbb{R}^d$ . Then (12.2.10) becomes the more familiar assertion that

$$\mathbb{E}(\xi(A) \mid \mathcal{I}) = Y \ell(A) \quad \mathcal{P}\text{-a.s.} \quad (12.2.10')$$

We can now state the main theorem of this section. It treats both marked and unmarked processes, and combines simple versions of both the individual and the statistical ergodic theorems. For more extensive results see MKM (1978, Section 6.2), Nguyen and Zessin (1979a), and the further discussion in Sections 13.4–5.

**Theorem 12.2.IV.** *Let the random measure  $\xi$  on  $\mathcal{X} = \mathbb{R}^d \times \mathcal{K}$  for some c.s.m.s.  $\mathcal{K}$  be stationary with respect to shifts on  $\mathbb{R}^d$  and have boundedly finite expectation measure  $\ell \times \nu$ . Let  $\psi$  be the invariant random measure defined as in Lemma 12.2.III. Then for any convex averaging sequence  $\{A_n\}$  on  $\mathbb{R}^d$  and any  $\nu$ -integrable function  $h$  on  $\mathcal{K}$ ,*

$$\frac{1}{\ell(A_n)} \int_{\mathcal{K}} h(\kappa) \xi(A_n \times d\kappa) \rightarrow \int_{\mathcal{K}} h(\kappa) \psi(d\kappa) \quad (n \rightarrow \infty) \quad (12.2.11)$$

a.s. and in  $L_1$  norm. If the second moment measure exists and

$$\mathbb{E} \left[ \left( \int_{\mathcal{K}} h(\kappa) \xi(\mathbb{U}^d \times d\kappa) \right)^2 \right] < \infty,$$

then convergence at (12.2.11) also holds in mean square.

For an unmarked process ( $\mathcal{K}$  reduces to a single point) the statements are equivalent to

$$\frac{\xi(A_n)}{\ell(A_n)} \rightarrow Y = \mathbb{E}[\xi(\mathbb{U}^d) \mid \mathcal{I}] \quad (n \rightarrow \infty) \quad (12.2.12)$$

a.s. and in  $L_1$  mean, and also in mean square if the second moment measure of  $\xi$  exists.

PROOF. We give details mainly for the unmarked case, and consider the proof of (12.2.12).

For some fixed  $\varepsilon > 0$ , let  $g_\varepsilon$  be a continuous function in  $\mathbb{R}^d$  such that

- (i)  $g_\varepsilon(x) \geq 0$ ,  $\int_{\mathbb{R}^d} g_\varepsilon(x) dx = 1$ ; and
  - (ii) the support of  $g_\varepsilon(\cdot) \subseteq S_\varepsilon(0)$ , the ball in  $\mathbb{R}^d$  with centre at 0 and radius  $\varepsilon$ .
- Now define a function  $f$  on  $\mathcal{M}_{\mathcal{X}}^\#$  by

$$f(\xi) = \int_{\mathbb{R}^d} g_\varepsilon(y) \xi(dy).$$

It is clear that  $f$  is measurable and, because  $\xi$  has finite expectation measure,  $f$  is a  $\mathcal{P}$ -integrable function with

$$\mathbb{E}(f) = m \int_{\mathbb{R}^d} g_\varepsilon(x) dx = m.$$

Observe that when  $\{A_n\}$  is a convex averaging sequence, so are the related sequences  $\{A_n^\varepsilon\}$  and  $\{A_n^{-\varepsilon}\}$  with elements defined by

$$A_n^{-\varepsilon} = \{x: S_\varepsilon(x) \subseteq A_n\} \quad \text{and} \quad A_n^\varepsilon = \bigcup_{x \in A_n} S_\varepsilon(x).$$

Also,

$$f(S_x \xi) = \int_{\mathbb{R}^d} g_\varepsilon(y) \xi(x + dy) = \int_{\mathbb{R}^d} g_\varepsilon(u - x) \xi(du).$$



This leads to the sandwich relation

$$\begin{aligned} \int_{A_n^{-\varepsilon}} f(S_x \xi) dx &= \int_{\mathbb{R}^d} \xi(du) \int_{A_n^{-\varepsilon}} g_\varepsilon(u-x) dx \\ &\leq \xi(A_n) \leq \int_{\mathbb{R}^d} \xi(du) \int_{A_n^\varepsilon} g_\varepsilon(u-x) dx = \int_{A_n^\varepsilon} f(S_x \xi) dx, \end{aligned}$$

where the inequalities are consequences of properties (i) and (ii) of  $g_\varepsilon(\cdot)$ , which further imply that

$$\int_{A_n^{-\varepsilon}} g_\varepsilon(u-x) dx = 0 \quad (u \notin A_n) \quad \text{and} \quad \int_{A_n^\varepsilon} g_\varepsilon(u-x) dx = 1 \quad (u \in A_n).$$

Invoking Proposition 12.2.II(a), we obtain

$$Y \liminf \frac{\ell(A_n^{-\varepsilon})}{\ell(A_n)} \leq \liminf \frac{\xi(A_n)}{\ell(A_n)} \leq \limsup \frac{\xi(A_n)}{\ell(A_n)} \leq Y \limsup \frac{\ell(A_n^\varepsilon)}{\ell(A_n)} \quad \mathcal{P}\text{-a.s.}$$

Because  $r(A_n) \rightarrow \infty$  and  $A_n$  is convex,  $\ell(A_n^\varepsilon)/\ell(A_n) \rightarrow 1$  and  $\ell(A_n^{-\varepsilon})/\ell(A_n) \rightarrow 1$  as  $n \rightarrow \infty$ . This establishes the a.s. assertion in (12.2.12).

Also from Proposition 12.2.II(b), with  $p = 1$ , we have, with  $f$  as defined above and for  $n \rightarrow \infty$ ,

$$\mathbb{E} \left| \frac{\int_{A_n^{-\varepsilon}} f(S_x \xi) dx}{\ell(A_n^{-\varepsilon})} - Y \right| \rightarrow 0 \quad \text{and} \quad \mathbb{E} \left| \frac{\int_{A_n^\varepsilon} f(S_x \xi) dx}{\ell(A_n^\varepsilon)} - Y \right| \rightarrow 0.$$

Denoting the first terms in these differences by  $L_n$  and  $U_n$ , respectively, these equations imply that  $\mathbb{E}|U_n - L_n| \rightarrow 0$  as  $n \rightarrow \infty$ . Furthermore,

$$\frac{\ell(A_n^{-\varepsilon})}{\ell(A_n)} L_n \leq \frac{\xi(A_n)}{\ell(A_n)} \leq \frac{\ell(A_n^\varepsilon)}{\ell(A_n)} U_n,$$

where the coefficients of  $L_n$  and  $U_n$  converge to 1, so  $\mathbb{E}|\xi(A_n)/\ell(A_n) - Y| \rightarrow 0$  as  $n \rightarrow \infty$ . This establishes the  $L_1$  convergence in (12.2.12). When the second moment measure exists, a similar argument with the  $L_2$  norm replacing the  $L_1$  norm establishes the  $L_2$  convergence.

Turning now to the general marked case, let  $h(\cdot): \mathcal{K} \mapsto \mathbb{R}$  be measurable and  $\nu$ -integrable. Define a function  $f$  on  $\mathcal{M}_\mathcal{X}^\#$  by

$$f(\xi) = \int_{\mathbb{R}^d \times \mathcal{K}} g_\varepsilon(y) h(\kappa) \xi(dy \times d\kappa),$$

and observe that  $f(S_x \xi)$  is the same integral with  $g_\varepsilon(y)$  replaced by  $g_\varepsilon(y-x)$ . Then form the integrals

$$\int_{A_n^{-\varepsilon}} f(S_x \xi) dx \quad \text{and} \quad \int_{A_n^\varepsilon} f(S_x \xi) dx,$$

and invoke the general forms of Proposition 12.2.II and Lemma 12.2.III to assert that

$$\frac{\int_{A_n^\varepsilon} f(S_x \xi) dx}{\ell(A_n^\varepsilon)} \rightarrow \int_{\mathcal{K}} h(\kappa) \psi(d\kappa),$$

with a similar statement holding with  $A_n^\varepsilon$  replaced by  $A_n^{-\varepsilon}$ ; here  $\psi(\cdot)$  is the invariant random measure defined by Lemma 12.2.III. Similar inequalities and arguments now apply as in the unmarked case, and yield (12.2.11) in its a.s.,  $L_1$  and  $L_2$  forms.  $\square$

As simple special cases of (12.2.12) and (12.2.11), the theorem yields the following corollaries.

**Corollary 12.2.V.** (a) *When  $\xi$  is stationary and metrically transitive with finite mean density  $m$ ,*

$$\frac{\xi(A_n)}{\ell(A_n)} \rightarrow m \quad \text{a.s. and in } L_1 \text{ norm.} \quad (12.2.13)$$

(b) *If  $\{(x_i, \kappa_i)\}$  is the realization of a stationary ergodic MPP on  $\mathbb{R}^d \times \mathcal{K}$ , then with  $\nu$  as in Theorem 12.2.IV,*

$$\frac{1}{\ell(A_n)} \sum_{i: x_i \in A_n} h(\kappa_i) \xrightarrow{\text{a.s.}} \int_{\mathcal{K}} h(\kappa) \nu(d\kappa). \quad (12.2.14)$$

For versions of the  $L_2$  norm results, see Exercises 12.2.7–8.

Numerous other special cases and corollaries follow from Theorem 12.2.IV such as the following (see also the exercises to this section).

**Proposition 12.2.VI.** *Under the conditions of Theorem 12.2.IV, for any measurable integrable function  $h(\cdot)$  on  $\mathbb{R}^d$  and with  $Y$  as at (12.2.12),*

$$\frac{\int_{\mathbb{R}^d} h(y) \xi(A_n + y) dy}{\ell(A_n)} = \frac{\int_{A_n} \xi(dx) \int_{\mathbb{R}^d} h(u - x) dx}{\ell(A_n)} \rightarrow Y \int_{\mathbb{R}^d} h(y) dy$$

a.s. and in  $L_1$  norm.  $(12.2.15)$

In all of these results, the convex averaging sequence  $\{A_n\}$  can of course be specialized to sequences of balls about the origin or nested hyper-rectangles whose smallest dimension  $\rightarrow \infty$ .

Higher-order ergodic theorems, requiring the existence of higher-order moment measures, are discussed in Section 12.6 and reappear in Chapter 13 in connection with higher-order Palm distributions. A different type of extension is outlined briefly in the proposition below.

**Proposition 12.2.VII** (Weighted Averages). *Let  $\{a_n(\cdot)\}$  be a monotonic increasing sequence of nonnegative functions, convex upward,  $\{A_n\}$  a convex*

averaging sequence in  $\mathbb{R}^d$ , and  $\xi$  a stationary random measure on  $\mathbb{R}^d$  with finite intensity  $m$ . Then as  $n \rightarrow \infty$ ,

$$\frac{\int_{A_n} a_n(x) \xi(dx)}{\int_{A_n} a_n(x) dx} \xrightarrow{\text{a.s.}} Y \equiv E(\xi(\mathbb{U}^d) | \mathcal{I}). \quad (12.2.16)$$

PROOF. Define an associated random measure  $\xi'$  on  $\mathbb{R}^d \times \mathbb{R}$  by  $\xi'(A \times B) = \xi(A)\ell(B)$ . Then  $\xi'$  is a stationary random measure in  $\mathbb{R}^{d+1}$ , and the sets

$$A'_n = \{(x, u) : x \in A_n, 0 \leq u \leq a_n(x)\}$$

are those of a convex averaging sequence in  $\mathbb{R}^{d+1}$ . Equation (12.2.16) follows by applying Theorem 12.2.IV to  $\xi'$ .  $\square$

Some other classes of weighting functions can be handled by using the more general averaging sequences considered by Tempel'man (1972). In particular, this includes the class  $a_n(x) = a(x/t_n)$ , where the nonnegative measurable function  $a(\cdot)$  has bounded support (e.g., the unit cube) and  $\{t_n\}$  is a sequence of nonnegative reals  $\rightarrow \infty$ ;  $a(\cdot)$  need not be convex upward. Also, the assumptions of Proposition 12.2.VII can be trivially extended to the case where there exist positive constants  $b_n$  such that  $a_n(x) \leq b_n a_{n+1}(x)$ .

Ergodic theorems are important in nearly all branches of point process theory, whether in establishing properties of point process models, or in developing estimation and testing procedures in statistics, or in analyzing the behaviour of simulation routines for point process models. We conclude this section with an application to the frequency of occurrence of special configurations of points in a Poisson process.

Discussions of particular configurations of points are closely related to one method of introducing Palm probabilities as ergodic limits (see in particular Theorem 13.2.VI). Such results can often be reduced to a direct application of Theorem 12.2.IV itself by introducing a suitable auxiliary random measure. We illustrate the procedure in a case where the expectation in the limit can be evaluated explicitly.

EXAMPLE 12.2(a) *Configurations in a Poisson process.* Let  $N$  be a stationary Poisson process in  $\mathbb{R}^d$  at rate  $\mu$ . Consider first the configuration consisting of a single point of the process with no neighbours within a distance  $a$ . The general (estimation) procedure is to take a convex region  $A$ , which we suppose to be a member of a convex averaging sequence, and count the number of points in  $A$  satisfying the required condition. Write this as the sum

$$Y(A) = \sum_{i: x_i \in A} I_B(S_{x_i} N),$$

where  $B = \{N : N(\{0\}) = 1 = N(S_a(0))\}$ . Evidently, the sum can also be written as the counting process integral

$$Y(A) = \int_A I_B(S_x N) N(dx),$$

and can be regarded as the value of a further point process  $Y(\cdot)$  if  $A$  is allowed to range more generally over bounded sets of  $\mathcal{B}(\mathbb{R}^d)$ . In fact,  $Y(\cdot)$  here is just a dependent thinning of the original process (see Section 11.3).

Applying Theorem 12.2.IV to  $Y(\cdot)$  yields the result that for increasing  $A$ ,

$$\frac{Y(A)}{\ell(A)} \rightarrow \mathbb{E} \left[ \int_{\mathbb{U}^d} I_B(S_x N) N(dx) \right] = \mu p_B, \quad (12.2.17)$$

where  $p_B$  may be regarded as the probability that a given point will be retained in the thinning process: later, we show that  $p_B$  can be interpreted as the Palm probability of the event  $B$ . In the special case considered here, we can evaluate the expectation by a simple approximation argument using the independence properties of the Poisson process as follows. The probability that there is a point in the small region  $(x, x + \delta x)$  and none in the remainder of a ball  $S_a(x)$  centred at  $x$  is  $\mu \ell(\delta x) \exp \{ -\mu [\ell(S_a(x)) - \ell(\delta x)] \}$ , and because the process is simple this is also the expected number of such configurations associated with the element  $(x, x + \delta x)$ . Integration over  $\mathbb{U}^d$  gives the limit as  $\mu e^{-\mu V(a)}$ , where  $V(a) = \ell(S_a(x))$  is the volume of a sphere of radius  $a$ .

Theorem 12.2.IV here asserts that the average density of points in  $A$  that have no neighbours closer than  $a$ , approaches this value as a limit when  $\ell(A) \rightarrow \infty$  through a convex averaging sequence.

Similarly, the average density of points in  $A$  that have at least  $k$  neighbours within a distance  $a$  approaches the limit

$$\mu [1 - e^{-\mu V(a)} (1 + \mu V(a) + \cdots + [\mu V(a)]^{k-1} / (k-1)!)].$$

Finally, consider the numbers of pairs of points that lie within a distance  $a$  of one another. Taking one point of any such pair as a reference origin, at  $x$  say, the number of such pairs to which it belongs is just  $N(S_a(x)) - 1$ . Summing over all points in  $A$  leads to the integral

$$Y_2(A) = \frac{1}{2} \int_A [N(S_a(x)) - 1] N(dx),$$

the factor  $\frac{1}{2}$  arising (asymptotically when the edge effects from points near the boundary of  $A$  become negligible) from the fact that each point of each pair is counted twice. Dividing by  $\ell(A)$  and noting that  $Y_2(\cdot)$  defines a random measure to which the theorems can be applied, we obtain the quantity  $\mathbb{E}[Y_2(\mathbb{U}^d)]$  as the limiting value of such an average density of pairs; again this can be evaluated by an approximation argument that uses the independence properties of the Poisson process as

$$\begin{aligned} \mathbb{E}[Y_2(\mathbb{U}^d)] &= \lim_{\delta x \rightarrow 0} \frac{1}{2} \int_{\mathbb{U}^d} \mathbb{E}[N(S_a(x) \setminus S_{\delta x}(x)) N(dx)] \\ &= \lim_{\delta x \rightarrow 0} \frac{1}{2} \int_{\mathbb{U}^d} \mu (V(a) - V(\delta x)) \mu dx = \frac{1}{2} \mu^2 V(a). \end{aligned}$$

Of course, the expected numbers of pairs is related to the second factorial moment measure of the process, and the general version of the above argument leads to a higher-order ergodic theorem in which the reduced factorial moment measures appear as ergodic limits.  $\square$

## Exercises and Complements to Section 12.2

12.2.1 Suppose the stationary point process  $N(\cdot)$  in  $\mathbb{R}^2$  has sample realizations  $\{Y + (n_1, 2n_2): n_1, n_2 = 0, \pm 1, \dots\}$ , where  $Y$  is uniformly distributed on the rectangle  $(0, 1] \times (0, 2]$ . Contrast the a.s. limits of  $N(A_n)/\ell(A_n)$  ( $n \rightarrow \infty$ ) when

- (1°)  $A_n = (0, n] \times (0, 1]$ ;      (2°)  $A_n = \bigcup_{j=1}^n (j-1, j] \times (j-1, j]$ ;  
 (3°)  $A_n = (0, 1] \times (0, n]$ ;      (4°)  $A_n = (0, n] \times (0, n]$ ;  
 (5°)  $A_n = \{(x, y): x \in (0, n], 0 \leq y \leq 2x/n\}$ ; and  
 (6°)  $A_n = \{(x, y): x \in (0, n], x \leq y \leq x+1\}$ .

12.2.2 Show that Theorem 12.2.IV, which in the text is deduced from Proposition 12.2.II, implies the latter in the sense that, if  $\xi$  is a stationary random measure on  $\mathbb{R}^d$  satisfying the assumptions of Theorem 12.2.IV and (12.2.12) holds, then (12.2.6) holds also. [Hint: For such  $\xi$  and  $f(\xi(\cdot))$  a functional with finite expectation, define a new random measure  $\xi_f$  by

$$\xi_f(A) = \int_A f(S_x \xi(\cdot)) dx \quad (\text{bounded } A \in \mathcal{B}(\mathbb{R}^d)).$$

Apply Theorem 12.2.IV to  $\xi_f$  to deduce Proposition 12.2.II(a) as applied to the original random measure  $\xi$ .]

12.2.3 Show formally that (a) if  $\xi$  has a trivial invariant  $\sigma$ -field, then the only invariant functions are the constant functions; and (b) metric transitivity implies ergodicity [i.e., (12.2.13) holds].

12.2.4 Let  $\mathcal{F}$  be a  $\sigma$ -algebra in  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$  and  $\xi(\cdot)$  a random measure on  $\mathcal{X}$  with finite expectation measure.

(a) Use Theorem 9.1.XIV to show that there exists a random measure  $\psi$  such that

$$\psi(A) = E[\xi(A) \mid \mathcal{F}] \quad \text{a.s.} \quad (\text{bounded } A \in \mathcal{B}_{\mathcal{X}}).$$

(b) Show that if  $\mathcal{F}$  is chosen to be the  $\sigma$ -algebra of events invariant under shifts  $S_x$ , then  $\psi(\cdot)$  is invariant in the sense that

$$S_x \psi(A) = \psi(A) \quad \text{a.s.} \quad (\text{bounded } A \in \mathcal{B}_{\mathcal{X}}).$$

[Hint: Show first that the indicator function of any event in  $\mathcal{F}$  is invariant, and ultimately that any  $\mathcal{F}$ -measurable r.v. is invariant.]

12.2.5 Interpret the strong law result at Exercise 4.1.1 in the setting of Theorem 12.2.IV.

12.2.6 Establish statistical ergodic theorem versions of the individual ergodic theorems at Theorem 12.2.IV, Corollaries 12.2.V–VI and Proposition 12.2.VII. [Hint: Use Proposition 12.2.II(b).]

12.2.7  $L_2$  convergence of a stationary MPP in  $\mathbb{R}^d$ . Show that if the stationary ergodic MPP  $\xi$  in  $\mathbb{R}^d$  satisfies the conditions of Theorem 12.2.IV, including the existence of finite second moments, then for any convex averaging sequence  $\{A_n\}$ ,

$$\lim_{n \rightarrow \infty} E \left[ \left( \frac{\xi(A_n \times K)}{\ell(A_n)} - \psi(K) \right)^2 \right] = 0,$$

where  $\psi(\cdot)$  is as in Lemma 12.2.III.

12.2.8 Let the simple point process  $N$  on  $\mathbb{R}$  have stationary second-order distributions and finite second-order moment. Then the expectation function  $U(\cdot)$  of (3.5.2) satisfies  $U(x)/x \rightarrow \lambda'$  ( $x \rightarrow \infty$ ) for some  $\lambda' \geq \lambda$  [Exercise 8.1.3(e) or Lemma 9 in Daley (1971)]. Examine the implications of Proposition 12.2.II for an  $L_2$  norm result.

12.2.9 (a) Use Theorem 12.2.IV, the inversion result at (8.6.8), and the identification of the Bartlett spectrum  $\Gamma(\cdot)$  as in Definition 8.2.II to show that for a nonergodic second-order stationary random measure  $\xi$  on  $\mathbb{R}^d$ ,

$$\Gamma(\{0\}) = \text{var } Y,$$

where  $Y$  is the  $\mathcal{I}$ -measurable r.v. as in (12.2.10') and (12.2.12).

(b) For a second-order stationary random measure  $\xi$  on  $\mathbb{R}$  and with  $V(x) = \text{var } \xi(0, x]$ , recall from Exercise 8.1.3(b) that  $\lim_{x \rightarrow \infty} x^{-2}V(x)$  exists and is finite. Defining  $v(s) = \int_0^\infty e^{-sx} dV(x)$ , use an Abelian theorem for Laplace–Stieltjes transforms [e.g., Widder (1941, p. 181)] to show that  $\lim_{s \rightarrow 0} s^2 v(s) = \lim_{x \rightarrow \infty} 2x^{-2}V(x)$ . Then show from (8.2.3) that

$$\frac{1}{2}s^2 v(s) = \Gamma(\{0\}) + \int_{\mathbb{R} \setminus \{0\}} \frac{\Gamma(d\omega)}{1 + \omega^2/s^2},$$

and conclude that  $\text{var } \xi(0, x] \sim x^2 \Gamma(\{0\})$  ( $x \rightarrow \infty$ ).

12.2.10 Let  $N$  be a marked point process on  $\mathbb{R}^d \times \mathcal{K}$ , stationary as in Lemma 12.2.III. Extend Proposition 12.1.VI to the statement that for each  $K \in \mathcal{B}(\mathcal{K})$ ,

$$\mathcal{P}\{N(\mathbb{R}^d \times K) = \infty\} + \mathcal{P}\{N(\mathbb{R}^d \times K) = 0\} = 1.$$

12.2.11 A *stationary nonisotropic point process* in  $\mathbb{R}^2$ . For  $i \in \mathbb{Z}$  let  $\{N_i(\cdot)\}$  be independent copies of a simple stationary point process on  $\mathbb{R}$  for which  $\text{var } N_i(0, 1] < \infty$  and with generic realization  $\{x_{ij}: j \in \mathbb{Z}\}$  at rate  $\lambda_1$ ; let  $\{y_i: i \in \mathbb{Z}\}$  be a realization of some other simple stationary point process  $N^0(\cdot)$  on  $\mathbb{R}$  at rate  $\lambda_0$  and with finite second moment measure, independent of the  $N_i$ . Write  $N(\cdot)$  for the counting measure of the point process in  $\mathbb{R}^2$  with realizations  $\{(x_{ij}, y_i): i, j \in \mathbb{Z}\}$ .

- Show that  $N$  is a stationary point process in  $\mathbb{R}^2$ .
- Using the convex averaging sets  $\{A_{1n}\}$  and  $\{A_{2n}\}$  specified by  $A_{1n} = (0, n] \times (0, n^2]$  and  $A_{2n} = (0, n^2] \times (0, n]$ , show that both  $N(A_{1n})/\ell(A_{1n})$  and  $N(A_{2n})/\ell(A_{2n}) \xrightarrow{\text{a.s.}} \mathbb{E}[N(A_{11})] = \lambda_0 \lambda_1$ .
- Use the representation  $N((0, x] \times (0, y]) = \sum_{i=1}^{N^0(0, y]} N_i(0, x]$  for the rectangle  $(0, x] \times (0, y]$  to show that  $\text{var } N((0, x] \times (0, y])$  equals

$$\lambda_0 y \lambda_1 \int_0^x [1 + 2U(u)] du + \lambda_0 \lambda_1^2 x^2 \int_0^y 2[U^0(v) - \lambda_0 v] dv,$$

where  $U(\cdot)$  and  $U^0(\cdot)$  are the expectation functions for the  $N_i$  and  $N^0$ , respectively. Deduce that, when  $\text{var } N_1(0, x] = O(x)$  and  $\text{var } N^0(0, y] = O(y)$  for  $x, y \rightarrow \infty$ , and for the same convex averaging sets as in (b),  $\text{var } N(A_{1n}) = O(n^4)$  but  $\text{var } N(A_{2n}) = O(n^5)$  and  $\ell(A_{1n}) = n^3 = \ell(A_{2n})$ .

- (d) Suppose that both  $N_i(\cdot)$  and  $N(\cdot)$  have covariance density functions,  $c(\cdot)$  and  $c_0(\cdot)$  say. Show that

$$\begin{aligned} \Pr\{N((x, x+dx) \times (y, y+dy)) > 0 \mid N(\{0,0\}) = 1\} \\ = \begin{cases} \lambda[c(x) + \lambda] dx & \text{if } y = 0, \\ \lambda\lambda_0[c_0(y) + \lambda_0] dx dy & \text{if } y \neq 0. \end{cases} \end{aligned}$$

Does  $N$  have a reduced covariance density,  $c((x, y))$  say?

## 12.3. Mixing Conditions

In practice, the useful applications of the ergodic theorem are to those situations where the ergodic limit is constant or, in other words, where the process is metrically transitive (the invariant  $\sigma$ -algebra is trivial). It is therefore important to characterize as fully as possible the various classes of processes that have this property. Now the absence of nontrivial invariant events is closely related to the absence of long-term dependence, and thus, checking for metric transitivity is generally accomplished by verifying that some kind of asymptotic independence or *mixing* condition is satisfied. This section contains a review of such conditions and outlines some of their applications.

As in the previous section we suppose that either  $\mathcal{X} = \mathbb{R}^d$  or  $\mathcal{X} = \mathbb{R}^d \times \mathcal{K}$ , and write  $S_x$  for the operator on  $\mathcal{M}_{\mathcal{X}}^{\#}$  defined by shifts as in (12.1.2) or, for  $\mathcal{X} = \mathbb{R}^d \times \mathcal{K}$ , by shifts in the first coordinate  $S_x\xi(\cdot, K) = \xi(\cdot + x, K)$ . We also write  $\mathbb{U}_{2a}^d$  for the hypercube in  $\mathbb{R}^d$  with sides of length  $2a$  and vertices  $(\pm a, \dots, \pm a)$ , and  $\mathcal{P}$  for the probability measure of a random measure on  $\mathcal{M}_{\mathcal{X}}^{\#}$  or a point process on  $\mathcal{N}_{\mathcal{X}}^{\#}$ .

**Definition 12.3.I.** A stationary random measure (respectively, point process) on state space  $\mathcal{X} = \mathbb{R}^d$  or  $\mathbb{R}^d \times \mathcal{K}$  is

- (i) ergodic if, for all  $V, W$  in  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$  [respectively,  $\mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#})$ ]

$$\frac{1}{\ell(\mathbb{U}_a^d)} \int_{\mathbb{U}_a^d} (\mathcal{P}(S_x V \cap W) - \mathcal{P}(V)\mathcal{P}(W)) dx \rightarrow 0 \quad (a \rightarrow \infty); \quad (12.3.1)$$

- (ii) weakly mixing if for all such  $V, W$ ,

$$\frac{1}{\ell(\mathbb{U}_a^d)} \int_{\mathbb{U}_a^d} |\mathcal{P}(S_x V \cap W) - \mathcal{P}(V)\mathcal{P}(W)| dx \rightarrow 0 \quad (a \rightarrow \infty); \quad (12.3.2)$$

- (iii) mixing if for all such  $V, W$ ,

$$\mathcal{P}(S_x V \cap W) - \mathcal{P}(V)\mathcal{P}(W) \rightarrow 0 \quad (\|x\| \rightarrow \infty); \quad (12.3.3)$$

- (iv)  $\psi$ -mixing (on  $\mathbb{R}^1$ ) if for  $u > 0$ ,  $t \in \mathbb{R}$ , and a function  $\psi(u)$  with  $\psi(u) \downarrow 0$  as  $u \rightarrow \infty$ ,

$$|\mathcal{P}(V \cap W) - \mathcal{P}(V)\mathcal{P}(W)| \leq \psi(u) \quad (12.3.4)$$

whenever  $V \in \sigma\{\xi(A): A \subseteq (-\infty, t]\}$  and  $W \in \sigma\{\xi(B): B \subseteq (t+u, \infty)\}$ .

The conditions are written in order of increasing strength: it is clear that mixing implies weak mixing, which in turn implies ergodicity. Furthermore, any completely random measure, such as the Poisson process, clearly satisfies all four conditions. The first three conditions apply to point processes and random measures generally; the fourth is introduced specifically to illustrate the central limit theorem at Proposition 12.3.X. Before examining the conditions in more detail, we show that in general it is enough to check the properties on any semiring of events generating the Borel sets in  $\mathcal{M}_X^\#$ : replacing  $\mathcal{M}_X^\#$  by  $\mathcal{N}_X^\#$  throughout leads to the same statement for point processes.

**Lemma 12.3.II.** *For a stationary random measure the limits in (12.3.1–4) hold for all  $V, W \in \mathcal{B}(\mathcal{M}_X^\#)$ , if and only if they hold for  $V, W$  in a semiring  $\mathcal{S}$  generating  $\mathcal{B}(\mathcal{M}_X^\#)$ .*

PROOF. We establish the truth of the assertion for (12.3.3) (mixing); the other cases are proved similarly.

Let  $\mathcal{F} \subseteq \mathcal{B}(\mathcal{M}_X^\#)$  denote the class of sets for which (12.3.3) holds. It is clear that if (12.3.3) holds for finite families of disjoint sets  $V_1, \dots, V_j$  and  $W_1, \dots, W_k$ , then it holds also for  $V = \bigcup_{i=1}^j V_i$  and  $W = \bigcup_{i=1}^k W_i$ . So, if (12.3.3) holds for sets in a semiring  $\mathcal{S}$ , it holds for sets in the ring  $\mathcal{R}$  generated by  $\mathcal{S}$ .

Suppose that  $W \in \mathcal{F}$  and  $V_n \in \mathcal{F}$  for  $n = 1, 2, \dots$  with  $V_n \uparrow V$ . Now

$$\begin{aligned} & |\mathcal{P}(S_x V \cap W) - \mathcal{P}(V)\mathcal{P}(W)| \\ & \leq |\mathcal{P}(S_x V \cap W) - \mathcal{P}(S_x V_n \cap W)| + |\mathcal{P}(S_x V_n \cap W) - \mathcal{P}(V_n)\mathcal{P}(W)| \\ & \quad + |\mathcal{P}(V_n)\mathcal{P}(W) - \mathcal{P}(V)\mathcal{P}(W)|, \end{aligned}$$

in which the first term on the right-hand side is bounded above by

$$\mathcal{P}(S_x(V \triangle V_n) \cap W) < \mathcal{P}(S_x(V \triangle V_n)) = \mathcal{P}(V \triangle V_n),$$

and the last term equals  $|\mathcal{P}(V_n) - \mathcal{P}(V)|\mathcal{P}(W)$ . By the continuity of  $\mathcal{P}(\cdot)$ , both these terms  $\rightarrow 0$  as  $n \rightarrow \infty$ , so, given  $\varepsilon > 0$ , we can fix  $n$  large enough such that each term  $< \varepsilon$ , uniformly in  $x$ . For the middle term, having fixed  $n$ , (12.3.3) holds for the pair  $V_n, W$ , so for  $\|x\| > \text{some } x_0$ , this term  $< \varepsilon$  also. Thus,

$$|\mathcal{P}(S_x V \cap W) - \mathcal{P}(V)\mathcal{P}(W)| < 3\varepsilon \quad (\text{all } \|x\| > x_0).$$

Similarly, we may also replace  $W$  by a sequence  $\{W_n\} \subseteq \mathcal{R}$  with  $W_n \rightarrow W$ , showing that  $\mathcal{F}$  is closed under monotone limits. Thus,  $\mathcal{F}$  is a monotone class which, because it includes  $\mathcal{R}$ , includes  $\sigma\{\mathcal{R}\} = \mathcal{B}(\mathcal{M}_X^\#)$ .  $\square$

Our aim now is to establish links with the theorems of the previous section. The next proposition establishes the equivalence of metric transitivity (trivial invariant  $\sigma$ -algebra) with the ergodicity condition at (i) of Definition 12.3.I above. It implies that in talking of an ergodic point process we may use the two criteria indifferently.



**Proposition 12.3.III.** *A stationary random measure or point process is ergodic if and only if it is metrically transitive; that is, the invariant  $\sigma$ -algebra  $\mathcal{I}$  is trivial.*

PROOF. Let  $\xi$  be ergodic as at (12.3.1) above and let  $A$  be an invariant event. Putting  $V = W = A$  in (12.3.1), observe from invariance that  $\mathcal{P}(S_x A \cap A) = \mathcal{P}(A)$  and hence, using ergodicity, that

$$\frac{1}{\ell(\mathbb{U}_a^d)} \int_{\mathbb{U}_a^d} (\mathcal{P}(A) - [\mathcal{P}(A)]^2) dx \rightarrow 0 \quad (a \rightarrow \infty),$$

which is possible only if  $\mathcal{P}(A) = 0$  or  $1$ .

Conversely, suppose that  $\mathcal{I}$  is trivial, so that (12.2.6) takes the form, in the notation as there,

$$\frac{1}{\ell(A_n)} \int_{A_n} f(S_x \xi) dx \rightarrow E[f(\xi)] \quad \text{a.s.}$$

Let  $V, W$  be as in Definition 12.3.I and take  $f(\xi) = I_V(\xi)$ , so that  $Ef(\xi) = \mathcal{P}(V)$  and (12.2.6) yields

$$\frac{1}{\ell(A_n)} \int_{A_n} I_V(S_x \xi) dx \rightarrow \mathcal{P}(V) \quad \text{a.s.}$$

Writing  $g_n(\xi)$  for the left-hand side of this equation, observe that  $0 \leq g_n(\xi) \leq 1$  and that  $g_n(\xi)$  is a measurable function of  $\xi$  in  $(\mathcal{M}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#}))$ . Integrating over  $W$  and using dominated convergence and Fubini's theorem, we obtain

$$\frac{1}{\ell(A_n)} \int_{A_n} \int_W I_V(S_x \xi) dx \mathcal{P}(d\xi) \rightarrow \mathcal{P}(V)\mathcal{P}(W),$$

which reduces to (12.3.1) in the special case  $A_n = \mathbb{U}_n^d$ . □

Just as ergodicity is related to the invariant  $\sigma$ -algebra  $\mathcal{I}$  being trivial, so mixing is related to the  $\sigma$ -algebra of *tail events* defined on the process  $\xi$  being trivial (this  $\sigma$ -algebra being, in general, larger than  $\mathcal{I}$ ). To define these events, denote by  $\mathcal{T}_a$ , for each  $a > 0$ , the  $\sigma$ -algebra of events defined by the behaviour of  $\xi$  outside  $\mathbb{U}_a^d$ ; that is,  $\mathcal{T}_a$  is the smallest  $\sigma$ -algebra in  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$  with respect to which the  $\xi(A)$  are measurable for  $A \in \mathcal{B}(\mathcal{X} \setminus \mathbb{U}_a^d)$ .

**Definition 12.3.IV.** *The tail  $\sigma$ -algebra of the process  $\xi$  is the intersection  $\mathcal{T}_{\infty} \equiv \bigcap_{a>0} \mathcal{T}_a = \bigcap_{n=1}^{\infty} \mathcal{T}_n$ . An element of  $\mathcal{T}_{\infty}$  is a tail event.*

Thus,  $\mathcal{T}_{\infty}$  defines the class of events that are determined by the behaviour of  $\xi$  outside any bounded subset of  $\mathcal{X}$ . It is not difficult to show that, modulo sets of  $\mathcal{P}$ -measure zero, any invariant event is in the tail  $\sigma$ -algebra (see Exercise 12.3.2). The converse is not true, however: periodic processes provide typical examples of processes that are ergodic but for which the tail  $\sigma$ -algebra is non-trivial (see Exercise 12.3.1).

The triviality result referred to above is set out below. Triviality of the tail  $\sigma$ -algebra is also closely related to the concept of *short-range correlation*; see Exercise 12.3.4 for a definition and details of the relationship. Note that the term short-range correlation is well established in the physics literature, although it relates to stochastic dependence rather than any second-order product-moment property; paradoxically, the terms long- and short-range dependence (see Section 12.7 for the point process setting) are established in the statistical literature as pertaining to a second-order or correlational (!) property.

**Proposition 12.3.V.** *If the tail  $\sigma$ -algebra is trivial, then the random measure  $\xi$  is mixing.*

PROOF. Let  $V$  be any set in  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ . Because  $\mathcal{T}_n \downarrow \mathcal{T}_{\infty}$ , we have for any random variable with expectation, and hence in particular for the indicator function  $I_V(\cdot)$ ,

$$\mathbb{E}(I_V \mid \mathcal{T}_n) \rightarrow \mathbb{E}(I_V \mid \mathcal{T}_{\infty}) \quad \text{a.s.}$$

[this is a standard result for backward martingales; see, e.g., Chung (1974, Theorem 9.4.7)].

When  $\mathcal{T}_{\infty}$  is trivial, the right-hand side here reduces to  $\mathbb{E}(I_V) = \mathcal{P}(V)$  a.s. Consequently, given  $\varepsilon > 0$ , we can choose  $n_0$  such that for  $n \geq n_0$ ,

$$|\mathbb{E}(I_V \mid \mathcal{T}_n) - \mathcal{P}(V)| < \varepsilon \quad \text{a.s.} \quad (12.3.5)$$

Now let  $W$  be a cylinder set belonging to the  $\sigma$ -algebra generated by the family  $\{\xi(A) : A \in \mathcal{B}(\mathbb{U}_a^d)\}$ . For  $\|x\|$  sufficiently large, namely,  $\|x\| > d^{1/2}n$ ,  $T_x \mathbb{U}_a^d$  lies in the complement of  $\mathbb{U}_n^d$  and hence  $S_x W \in \mathcal{T}_n$ . For fixed  $n$  and  $\|x\|$  large enough, the indicator function  $I_{S_x W}$  is therefore  $\mathcal{T}_n$ -measurable and so satisfies

$$\mathbb{E}(I_{S_x W} I_V \mid \mathcal{T}_n) = I_{S_x W} \mathbb{E}(I_V \mid \mathcal{T}_n) \quad \text{a.s.}$$

Taking expectations and using (12.3.5), we obtain

$$\mathcal{P}(S_x W \cap V) = \mathcal{P}(S_x W) \mathcal{P}(V) + \mathbb{E}(I_{S_x W} Y),$$

where the r.v.  $Y$  has  $|Y| < \varepsilon$  a.s. This establishes the mixing property for arbitrary  $V$  and any cylinder set  $W$ . Because the cylinder sets generate  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ , the proposition follows from Lemma 12.3.II.  $\square$

Mixing is defined above in terms of probabilities of events; the conditions can equally be stated in terms of expectations of random variables defined on the process: see Exercises 12.3.5–6 for details in the context of convergence to equilibrium, which topic is discussed more fully in Section 12.5.

We already observed in the proof of the last proposition that it is enough to verify the mixing or weak mixing or ergodicity conditions for cylinder sets, that is, for sets of the type that occur in the definition of the fidi distributions. Although this may sometimes be convenient, it is generally easier to check

the conditions in a form that relates to the generating functionals rather than directly to the fidi distributions. The next proposition provides such conditions: the discussion here, as in the applications that follow, is based on Westcott (1972). In the proposition below we use the shift operator  $S_x$  defined on functions  $h(\cdot)$  by  $(S_x h)(y) = h(y + x)$  [compare with (12.1.1)].

**Proposition 12.3.VI.** (a) Let  $\xi$  be a random measure,  $L[\cdot]$  its Laplace functional, and  $h_1, h_2$  functions in  $\text{BM}_+(\mathcal{X})$ .

(i)  $\xi$  is ergodic if and only if for all such  $h_1, h_2$ ,

$$\frac{1}{\ell(\mathbb{U}_n^d)} \int_{\mathbb{U}_n^d} (L[h_1 + S_x h_2] - L[h_1]L[h_2]) dx \rightarrow 0 \quad (n \rightarrow \infty). \quad (12.3.6)$$

(ii)  $\xi$  is weakly mixing if and only if for all such  $h_1, h_2$ ,

$$\frac{1}{\ell(\mathbb{U}_n^d)} \int_{\mathbb{U}_n^d} |L[h_1 + S_x h_2] - L[h_1]L[h_2]| dx \rightarrow 0 \quad (n \rightarrow \infty). \quad (12.3.7)$$

(iii)  $\xi$  is mixing if and only if for all such  $h_1, h_2$ ,

$$L[h_1 + S_x h_2] \rightarrow L[h_1]L[h_2] \quad (\|x\| \rightarrow \infty). \quad (12.3.8)$$

(b) Let  $N$  be a point process,  $G[\cdot]$  its p.g.fl., and  $h_1, h_2$  functions in  $\mathcal{V}(\mathcal{X})$ .

(i)  $N$  is ergodic if and only if for all such  $h_1, h_2$ ,

$$\frac{1}{\ell(\mathbb{U}_n^d)} \int_{\mathbb{U}_n^d} (G[h_1 S_x h_2] - G[h_1]G[h_2]) dx \rightarrow 0 \quad (n \rightarrow \infty). \quad (12.3.9)$$

(ii)  $N$  is weakly mixing if and only if for all such  $h_1, h_2$ ,

$$\frac{1}{\ell(\mathbb{U}_n^d)} \int_{\mathbb{U}_n^d} |G[h_1 S_x h_2] - G[h_1]G[h_2]| dx \rightarrow 0 \quad (n \rightarrow \infty). \quad (12.3.10)$$

(iii)  $N$  is mixing if and only if for all such  $h_1, h_2$ ,

$$G[h_1 S_x h_2] \rightarrow G[h_1]G[h_2] \quad (\|x\| \rightarrow \infty). \quad (12.3.11)$$

PROOF. (a) Let  $\{A_j: j = 1, \dots, J\}$  and  $\{B_k: k = 1, \dots, K\}$  be bounded on  $\mathcal{X}$ , and consider the family of random variables  $\{\xi(A_j)\} \cup \{\xi(T_x B_k)\}$ . If  $\xi$  is mixing then (12.3.3) implies that the joint distribution of this family converges to the product of the two joint distributions of the families  $\{\xi(A_j)\}$  and  $\{\xi(B_k)\}$ . It follows that the multivariate Laplace transforms of these joint distributions satisfy for real nonnegative  $\{\alpha_j\}$  and  $\{\beta_k\}$

$$\begin{aligned} & \mathbb{E} \left[ \exp \left( - \sum_{j=1}^J \alpha_j \xi(A_j) - \sum_{k=1}^K \beta_k \xi(T_x B_k) \right) \right] \\ & \rightarrow \mathbb{E} \left[ \exp \left( - \sum_{j=1}^J \alpha_j \xi(A_j) \right) \right] \mathbb{E} \left[ \exp \left( - \sum_{k=1}^K \beta_k \xi(B_k) \right) \right]. \end{aligned}$$

But this is just the statement (12.3.8) for the special case that the  $h_i$  are the simple functions

$$h_1(x) = \sum_{j=1}^J \alpha_j I_{A_j}(x), \quad h_2(x) = \sum_{k=1}^K \beta_k I_{B_k}(x). \quad (12.3.12)$$

Now any  $h \in \text{BM}_+(\mathcal{X})$  can be monotonically and uniformly approximated by simple functions of this form, so an argument similar to that of Lemma 12.3.II shows that (12.3.8) holds as stated.

Conversely, when (12.3.8) holds, take  $h_1, h_2$  to be simple functions as at (12.3.12). Then it follows from the continuity theorem for Laplace transforms that the joint distributions of  $\{\xi(A_j)\}$  and  $\{\xi(T_x B_k)\}$  converge for all families  $\{A_j\}$  and  $\{B_k\}$ , and hence that (12.3.3) holds for the corresponding cylinder sets, that is, a semiring generating  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ . Then by Lemma 12.3.II, (12.3.3) holds generally and so  $\xi$  is mixing.

Analogous statements hold in the other cases; we omit the details.  $\square$

It is important for our proof below of Proposition 12.3.IX to observe that the convergence properties in equations (12.3.6–11) hold for wider classes of functions than those with bounded support. For example, when each  $h_i$  is the monotone limit of a sequence  $\{h_{in}\} \subset \text{BM}_+(\mathcal{X})$ , (12.3.8) holds provided we interpret the functionals as extended Laplace functionals [see (9.4.11)]. To see this, recall that any function in  $\text{BM}_+(\mathcal{X})$  is the monotone limit of simple functions, and first assume that the functions  $h_1$  and  $\{h_{2n}\}$  are simple functions. By the argument leading to (12.3.12) we then have, when  $\xi$  is stationary and mixing,  $L[h_1 + S_x h_{2n}] \rightarrow L[h_1]L[h_{2n}]$  as  $\|x\| \rightarrow \infty$ . Thus,

$$\begin{aligned} 0 &\leq |L[h_1]L[h_2] - L[h_1 + S_x h_2]| \\ &\leq |L[h_1](L[h_2] - L[h_{2n}])| + |L[h_1]L[h_{2n}] - L[h_1 + S_x h_{2n}]| \\ &\quad + |L[h_1 + S_x h_{2n}] - L[h_1 + S_x h_2]| \\ &\equiv \delta_{1n} + \delta_{2n}(x) + \delta_{3n}(x), \quad \text{say.} \end{aligned}$$

Now

$$\begin{aligned} \delta_{3n}(x) &= |\mathbb{E}\{\exp(-\int h_1 d\xi) [\exp(-\int S_x h_{2n} d\xi) - \exp(-\int S_x h_2 d\xi)]\}| \\ &\leq |L[S_x h_{2n}] - L[S_x h_2]| \quad \text{by nonnegativity and monotonicity,} \\ &= |L[h_{2n}] - L[h_2]| \quad \text{by stationarity,} \end{aligned}$$

and by monotone convergence,  $L[h_{2n}] \downarrow L[h_2]$  as  $n \rightarrow \infty$ . Similarly,  $\delta_{1n} \leq |L[h_{2n}] - L[h_2]|$  and so, given  $\varepsilon > 0$ , we can make both  $\delta_{1n} < \varepsilon$  and  $\delta_{3n}(x) < \varepsilon$ , uniformly in  $x$ , by choosing  $n$  sufficiently large. Fixing such  $n$ , (12.3.12) now implies that we can make  $\delta_{2n}(x) < \varepsilon$  by taking  $\|x\|$  sufficiently large. Thus, (12.3.8) holds for simple  $h_1 \in \text{BM}_+(\mathcal{X})$  and  $h_2 \in \overline{\text{BM}_+}(\mathcal{X})$ , and a similar argument establishes it for  $h_1 \in \overline{\text{BM}_+}(\mathcal{X})$  as well.

Extensions of the rest of (12.3.6–11) are established in a similar manner.

It is also pertinent to note that in part (b), it is enough to restrict the functions  $h_i$  to the subspace  $\mathcal{V}_0(\mathcal{X}) \subset \mathcal{V}(\mathcal{X})$ . This follows directly from part (a) and the remark following (9.4.14).

Using these results, we can investigate the mixing and ergodicity properties of some classes of point processes, namely Cox and cluster processes in Propositions 12.3.VII–IX, and interval properties (e.g., renewal processes) in Exercise 12.4.1 and Section 13.4.

**Proposition 12.3.VII.** *A stationary Cox process  $N$  on  $\mathcal{X} = \mathbb{R}^d$  is mixing, weakly mixing, or ergodic if and only if the random measure  $\Lambda$  directing  $N$  has the same property.*

PROOF. Recall from Proposition 6.2.II that the p.g.fl.  $G[\cdot]$  of a Cox process  $N$  is related to the Laplace functional  $L[\cdot]$  of the random measure  $\Lambda$  directing  $N$  by  $G[h] = L[1 - h]$  for  $h \in \mathcal{V}(\mathcal{X})$ . To verify the mixing property we start from the relation

$$\begin{aligned} G[h_1 S_x h_2] &= L[1 - h_1 S_x h_2] \\ &= L[1 - h_1 + S_x(1 - h_2) - (1 - h_1)S_x(1 - h_2)]. \end{aligned}$$

Because each  $1 - h_i$  ( $i = 1, 2$ ) has bounded support, the last term vanishes for sufficiently large  $\|x\|$ , and for such  $x$ , appealing to (12.3.8) for  $\|x\| \rightarrow \infty$ ,

$$\begin{aligned} G[h_1 S_x h_2] &= L[(1 - h_1) + S_x(1 - h_2)] \\ &\rightarrow L[1 - h_1]L[1 - h_2] = G[h_1]G[h_2]. \end{aligned}$$

This argument is reversible and proves the result concerning the mixing property. Proofs for the weakly mixing and ergodicity properties are similar.  $\square$

**Corollary 12.3.VIII.** *A stationary mixed Poisson process  $N$  is mixing if and only if it is a simple Poisson process.*

PROOF. From Example 9.4(d) and Exercise 12.1.4, the directing measure  $\xi$  must be a random multiple of Lebesgue measure,  $\Lambda\ell(\cdot)$  say, so for  $h \in \mathcal{V}(\mathbb{R}^d)$  the p.g.fl.  $G[h]$  of  $N$  equals

$$\phi_\Lambda \left( \int_{\mathbb{R}^d} [1 - h(x)] dx \right),$$

where  $\phi_\Lambda(\cdot)$  denotes the Laplace–Stieltjes transform of  $\Lambda$ . Now as  $\|x\| \rightarrow \infty$ ,

$$L[(1 - h_1) + S_x(1 - h_2)] \rightarrow \phi_\Lambda \left[ \int_{\mathbb{R}^d} [1 - h_1(x)] dx + \int_{\mathbb{R}^d} [1 - h_2(x)] dx \right],$$

which can equal the product of the  $\phi_\Lambda \left( \int_{\mathbb{R}^d} [1 - h_i(x)] dx \right)$  for all  $h_i \in \mathcal{V}(\mathbb{R}^d)$  if and only if  $\phi_\Lambda$  is an exponential function, and hence the distribution of  $\Lambda$  is concentrated at a single point.  $\square$

By a *stationary cluster process*  $N$  on  $\mathbb{R}^d$ , we mean a process as in Exercise 12.1.6 (or Proposition 12.1.V for a stationary Poisson cluster process).

**Proposition 12.3.IX.** *A stationary cluster process is mixing, weakly mixing, or ergodic, whenever the cluster centre process has the same property.*

PROOF. We give details for the mixing case only; the other cases can be treated in a similar fashion. Also, although we generally follow the p.g.fl. proof of Westcott (1971) [for an alternative proof see MKM (1978, Proposition 11.1.4)], some further argument is needed as in Daley and Vere-Jones (1987). In particular, we use the idea of extended p.g.fl.s and use Proposition 12.3.VI(b) with functions  $h_i \in \mathcal{V}_0(\mathcal{X})$  (see the last remark following that proposition). Then in view of (12.1.18) and (12.3.11) with  $h_i \in \mathcal{V}_0(\mathcal{X})$ , it is enough to deduce from  $G_c[h_1 S_x h_2] \rightarrow G_c[h_1] G_c[h_2]$  as  $\|x\| \rightarrow \infty$  that

$$G_c[G_m[h_1 S_x h_2 \mid \cdot]] \rightarrow G_c[G_m[h_1 \mid \cdot]] G_c[G_m[h_2 \mid \cdot]]. \quad (12.3.13)$$

Formally, the mixing property implies that the right-hand side here is the limit at  $\|x\| \rightarrow \infty$  of

$$G_c[G_m[h \mid \cdot] S_x G_m[h_2 \mid \cdot]] \equiv G_c[\tilde{h}_1 S_x \tilde{h}_2], \quad (12.3.14)$$

where  $\tilde{h}_i(y) = G_m[h_i \mid y]$ ; so, formally, it is enough to show that as  $\|x\| \rightarrow \infty$ ,

$$G_c[\tilde{h}_1 S_x \tilde{h}_2] - G_c[G_m[h_1 S_x h_2 \mid \cdot]] \rightarrow 0. \quad (12.3.15)$$

We have said ‘formally’ because, although  $h_i \in \mathcal{V}_0(\mathbb{R}^d)$ , the same need not necessarily be true of  $\tilde{h}_i$ . However, by replacing the generic cluster  $N_m(\cdot \mid 0)$  by  $N_{mn}(\cdot \mid 0) = N_m(\cdot \cap \mathbb{U}_n^d \mid 0)$  and letting  $n \rightarrow \infty$ , each  $\tilde{h}_i$  is expressed as the limit of the monotonic sequence  $\{\tilde{h}_{in}\}$  for which  $h_{in} \in \mathcal{V}_0(\mathbb{R}^d)$ . Consequently, appealing to the convergence properties of extended p.g.fl.s established in Exercise 9.4.6(b) and the extended form of the mixing property (12.3.11) noted below Proposition 12.3.VI, it follows that when  $N_c(\cdot)$  is mixing,  $G_c[\tilde{h}_1 S_x \tilde{h}_2] \rightarrow G_c[\tilde{h}_1] G_c[\tilde{h}_2]$  as  $\|x\| \rightarrow \infty$ .

To complete the proof, define

$$\Delta_x(u) = G_m[h_1 S_x h_2 \mid u] - G_m[h_1 \mid u] S_x G_m[h_2 \mid u].$$

As upper bounds on  $\Delta_x(u)$  we have

$$\begin{aligned} \Delta_x(u) &\leq G_m[S_x h_2 \mid u](1 - G_m[h_1 \mid u]) \leq 1 - G_m[h_1 \mid u], \\ \Delta_x(u) &\leq G_m[h_1 \mid u](1 - G_m[S_x h_2 \mid u]) \leq 1 - G_m[S_x h_2 \mid u]; \end{aligned}$$

as lower bounds we have

$$\begin{aligned} \Delta_x(u) &\geq G_m[h_1 S_x h_2 \mid u] - G_m[S_x h_2 \mid u] \\ &= -E\left\{\exp\left[\int_{\mathcal{X}} \log[S_x h_2(y)] N_m(dy \mid u)\right]\right. \\ &\quad \left.\times (1 - \exp[\int_{\mathcal{X}} \log h_1(y) N_m(dy \mid u)])\right\} \\ &\geq -(1 - G_m[h_1 \mid u]) \end{aligned}$$

and, similarly,  $\Delta_x(u) \geq -(1 - G_m[S_x h_2 | u])$ . Because  $h_2 \in \mathcal{V}_0(\mathbb{R}^d)$  and  $N_m(\mathbb{R}^d | u) < \infty$  a.s.,  $G_m[S_x h_2 | u] \rightarrow 1$  as  $\|x\| \rightarrow \infty$  and thus  $\Delta_x(u) \rightarrow 0$  as  $\|x\| \rightarrow \infty$ . Also, because the cluster process exists, (12.1.18) holds, and therefore  $|\Delta_x(u)|$  is bounded above by  $1 - G_m[h_1 | u]$ , which is  $N_c$ -integrable a.s. Indeed, again because  $h_1 \in \mathcal{V}_0(\mathbb{R}^d)$  and  $N_m(\mathbb{R}^d | u) < \infty$  a.s., it also holds that  $1 \geq G_m[h_1 | u] \geq c_1$ , uniformly in  $u$ , for some positive constant  $c_1$ , and similarly for  $G_m[S_x h_2 | u]$ . Thus,

$$\chi_x(u) \equiv G_m[h_1 | u] S_x G_m[h_2 | u] \geq c > 0$$

for some constant  $c$ , uniformly in  $x$  and  $u$ . Now

$$\begin{aligned} & |G_c[G_m[h_1 S_x h_2 | \cdot]] - G_c[G_m[h_1 | \cdot] S_x G_m[h_2 | \cdot]]| \\ &= |G_c[\chi_x + \Delta_x] - G_c[\chi_x]| \\ &= |\mathbb{E}[\exp(\int_{\mathbb{R}^d} \log[\chi_x(u) + \Delta_x(u)] N_c(du)) - \exp(\int_{\mathbb{R}^d} \log \chi_x(u) N_c(du))]| \\ &\leq 1 - \mathbb{E} \exp\left(-\int_{\mathbb{R}^d} \log\left(1 + \frac{|\Delta_x(u)|}{\chi_x(u)}\right) N_c(du)\right). \end{aligned}$$

This expression  $\rightarrow 0$  as  $\|x\| \rightarrow \infty$  because  $1 \geq \chi_x(u) \geq c > 0$  (all  $x$  and  $u$ ),  $\Delta_x(u) \rightarrow 0$  pointwise, and  $\Delta_x(u)$ , and hence  $c^{-1}\Delta_x(u)$  also, is  $N_c$ -integrable a.s. uniformly in  $x$  (see Exercise 9.4.6). Thus, (12.3.13) is proved.  $\square$

One of the classical applications of mixing conditions is in establishing conditions for a central limit theorem. Here, the point process or random measure character of the realizations plays an entirely minor role; it is not the local behaviour but the behaviour over large time spans that is important. Results for point processes and random measures can, indeed, be written down directly from the results in texts such as Billingsley (1968) for stochastic processes in general. Because rescaling is involved, the limits need no longer correspond to random measures, and convergence needs to be expressed in terms of convergence in a function space such as  $\mathcal{D}(0, 1)$ . For example, adapting Billingsley's Theorem 20.1 to random measures yields the following.

**Proposition 12.3.X.** *Let the stationary random measure  $\xi$  on  $\mathcal{X} = \mathbb{R}$  be  $\psi$ -mixing for some continuous, monotonic, nonnegative function  $\psi(\cdot)$  on  $\mathbb{R}_+$  for which  $\int_0^\infty [\psi(t)]^{1/2} dt < \infty$ , and have boundedly finite first and second moment measures, with mean rate  $m$  and reduced covariance measure  $C(\cdot)$  satisfying  $0 < \sigma^2 = C(\mathbb{R}) < \infty$ . Then the sequence of random processes  $\{Y_n\}$  defined by*

$$Y_n(t) = \{\xi(0, nt] - mnt\} / \sigma n^{1/2} \quad (0 \leq t \leq 1)$$

*converges weakly in  $\mathcal{D}(0, 1)$  to the Wiener process on  $(0, 1)$ .*

## Exercises and Complements to Section 12.3

12.3.1 Prove that a stationary renewal process can exist if and only if the lifetime distribution has a finite mean and that such a process is ergodic but need

not be mixing. Show that if the lifetime distribution is nonlattice then the process is mixing. [Hint: Use the renewal theorem in different forms. A periodic renewal process can be made into a stationary process by suitably distributing the initial point (see e.g. the stationary deterministic process at Exercise 12.1.7), but such a process is not mixing because, for example, the events  $V = \{N(0, \frac{1}{2}] > 0\}$  and  $W = \{N(\frac{1}{2}, 1] > 0\}$  do not satisfy the mixing property (12.3.1) when  $x$  is an integer and the process has period 1.]

12.3.2 Show that any event in  $\mathcal{I}$  is equal (modulo sets of measure zero) to an event in  $\mathcal{T}_\infty$ , but not conversely. [Hint: See Exercise 12.3.1 for the converse. For  $W \in \mathcal{I}$ , consider  $W = \bigcap_{n=1}^\infty S_{x_n} V$ , where  $\|x_n\| \rightarrow \infty$  as  $n \rightarrow \infty$ .]

12.3.3 As an example of a cluster process that is mixing but for which the cluster centre process is not mixing, take the cluster centre process to be a mixture of two Neyman–Scott cluster processes with member distributions  $F_1$  and  $F_2$  about the centre, and the cluster member process again of Neyman–Scott type with distribution  $F_3$  such that  $F_1 * F_3 = F_2 * F_3$ .

12.3.4 Say that a process  $\xi$  has *short-range correlation* if for  $W \in \mathcal{B}(\mathcal{M}_\mathcal{X}^\#)$  with  $\mathcal{P}(W) > 0$ , and arbitrary  $\varepsilon > 0$ , there exists a bounded set  $A \in \mathcal{B}_\mathcal{X}$  such that on the sub- $\sigma$ -algebra of  $\mathcal{B}(\mathcal{M}_\mathcal{X}^\#)$  determined by  $\{\xi(B): B \in \mathcal{X} \setminus A\}$ , the variation norm of the difference as below satisfies

$$\|\mathcal{P}_{\mathcal{X} \setminus A}(\cdot \mid W) - \mathcal{P}_{\mathcal{X} \setminus A}(\cdot)\| < \varepsilon.$$

Show that  $\xi$  has short-range correlation if and only if the tail  $\sigma$ -algebra  $\mathcal{T}_\infty$  is trivial. [Hint: Consider the sequence of Radon–Nikodym derivatives  $p_n(\cdot) \equiv d\mathcal{P}_{\mathcal{X} \setminus A_n}(\cdot \mid W)/d\mathcal{P}_{\mathcal{X} \setminus A_n}(\cdot)$  for  $A_n = \mathbb{U}_n^d$  (for example). Show that these functions  $\{p_n(\cdot)\}$  constitute a martingale that converges to a limit  $p_\infty(\cdot)$ , which is  $\mathcal{T}_\infty$ -measurable, and that  $h_\infty \neq \text{constant a.s.}$  if  $\|\mathcal{P}_{\mathcal{X} \setminus A}(\cdot \mid W) - \mathcal{P}_{\mathcal{X} \setminus A}(\cdot)\| > c > 0$  for some real  $c$  and every bounded  $A \in \mathcal{B}_\mathcal{X}$ . The result is attributed to Lanford and Ruelle (1969) in MKM (1982, Theorem 1.10.1).]

12.3.5 Let r.v.s  $X(\xi), Y(\xi)$  be defined on the stationary random measure  $\xi$  and have finite expectations. Show that  $\xi$  is mixing if and only if for all such r.v.s  $X, Y$ ,

$$\int_{\mathcal{M}_\mathcal{X}^\#} X(S_x \xi) Y(\xi) \mathcal{P}(d\xi) \rightarrow E[X(\xi)] E[Y(\xi)] \quad (\|x\| \rightarrow \infty).$$

12.3.6 *Stability condition for mixing process.* Let  $\mathcal{P}, \mathcal{P}_0$  be measures on  $\mathcal{M}_\mathcal{X}^\#$  with  $\mathcal{P}_0 \ll \mathcal{P}$ . Show that if  $\mathcal{P}$  is mixing then for  $\mathcal{P}_x$  defined by

$$\mathcal{P}_x(V) = \int_{\mathcal{M}_\mathcal{X}^\#} I_V(S_x \xi) \mathcal{P}_0(d\xi) \quad (V \in \mathcal{B}(\mathcal{M}_\mathcal{X}^\#)),$$

$\mathcal{P}_x \rightarrow \mathcal{P}$  weakly as  $\|x\| \rightarrow \infty$ . [Hint: Let  $p(\xi)$  be a measurable version of the Radon–Nikodym derivative; apply Exercise 12.3.5.]

12.3.7 A process is mixing of order  $k$  if for all  $V_0, \dots, V_{k-1} \in \mathcal{B}(\mathcal{M}_\mathcal{X}^\#)$ ,

$$\mathcal{P}(V_0 \cap S_{x_1} V_1 \cap \dots \cap S_{x_{k-1}} V_{k-1}) \rightarrow \mathcal{P}(V_0) \mathcal{P}(V_1) \dots \mathcal{P}(V_{k-1})$$

as  $\|x_i\| \rightarrow \infty$  ( $i = 1, \dots, k-1$ ) in such a way that  $\|x_i - x_j\| \rightarrow \infty$  also for all  $i \neq j$ . Show that when  $\mathcal{T}_\infty$  is trivial the process is mixing of all orders. [Hint: Use the method of proof of Proposition 12.3.V. See also MKM (1978, Theorem 6.3.6) and MKM (1982, Theorem 6.2.9).]



12.3.8 Verify the assertions of Proposition 12.3.IX concerning conditions for a stationary cluster process to be weakly mixing or ergodic.

## 12.4. Stationary Infinitely Divisible Point Processes

This shorter section discusses stationarity and mixing conditions for infinitely divisible point processes, resuming the notation and terminology of Section 10.2. It can be viewed as an extended example concerning mixing conditions, at the same time yielding a more refined classification of such point processes.

Consider first stationarity. Let  $\tilde{Q}$  denote the KLM measure of a stationary infinitely divisible process so that from (10.2.9),

$$\begin{aligned} \log G[h(\cdot - u)] &= \int_{\mathcal{N}_0^\#(\mathcal{X})} \left( \exp \left[ \int_{\mathcal{X}} \log h(x - u) \tilde{N}(\mathrm{d}x) \right] - 1 \right) \tilde{Q}(\mathrm{d}\tilde{N}) \\ &= \int_{\mathcal{N}_0^\#(\mathcal{X})} \left( \exp \left[ \int_{\mathcal{X}} \log h(y) \tilde{N}(\mathrm{d}y) \right] - 1 \right) \tilde{Q}(S_u \mathrm{d}\tilde{N}). \end{aligned} \quad (12.4.1)$$

When  $N$  is stationary, this must coincide with the original form of (10.2.9),

$$\log G[h] = \int_{\mathcal{N}_0^\#(\mathcal{X})} \left( \exp \left[ \int_{\mathcal{X}} \log h(y) \tilde{N}(\mathrm{d}y) \right] - 1 \right) \tilde{Q}(\mathrm{d}\tilde{N}).$$

As in (12.1.3), we can define a new measure  $\hat{S}_u \tilde{Q}$  by

$$\hat{S}_u \tilde{Q}(B) = \tilde{Q}\{\tilde{N}: S_u \tilde{N} \in B\}$$

so that (12.4.1) could equally well be written in the form (10.2.9) with  $\hat{S}_u \tilde{Q}$  in place of  $\tilde{Q}$ . But by Theorem 10.2.V the KLM measure is unique, so  $\tilde{Q}$  and  $\hat{S}_u \tilde{Q}$  must coincide, and we have established the following result (see Exercise 12.4.1 for a generalization).

**Proposition 12.4.I.** *An infinitely divisible point process on  $\mathbb{R}^d$  is stationary if and only if its KLM measure is stationary (i.e., invariant under the shifts  $\hat{S}_u$ ).*

In Section 10.2 we established some relationships between total finiteness of the KLM measure and the representation of an infinitely divisible point process as a Poisson cluster process. This relationship can be sharpened when the point process is stationary.

**Proposition 12.4.II.** *Let  $N$  be a stationary infinitely divisible point process on  $\mathbb{R}^d$ .*

- (a) *If  $N$  has a representation as a Poisson randomization then it is singular.*
- (b)  *$N$  is regular if and only if it can be represented as a Poisson cluster process with a stationary Poisson process of cluster centres and a cluster structure that depends only on the relative locations of the points in a cluster and not on the location of the cluster itself.*

PROOF. Suppose first that the KLM measure  $\tilde{Q}$  is totally finite, so that  $\tilde{\mathcal{P}}(\cdot) \equiv \tilde{Q}(\cdot)/\tilde{Q}(\mathcal{N}_0^\#(\mathcal{X}))$  is the probability measure of a stationary point process  $\tilde{N}$ . The special property (10.2.8) of the KLM measure now implies that  $\tilde{\mathcal{P}}\{\tilde{N}(\mathcal{X}) = 0\} = 0$ , so coupled with Proposition 12.1.VI we must have  $\tilde{\mathcal{P}}\{\tilde{N}(\mathcal{X}) = \infty\} = 1$ , which with (10.2.10b) proves (a).

From the decomposition at Proposition 10.2.VII of an infinitely divisible point process into its regular and singular components, and from the fact that  $\tilde{S}_u$  maps  $\tilde{\mathcal{P}}\{\tilde{N}(\mathcal{X}) = \infty\}$  into itself, it follows that we may discuss the effects of stationarity separately for each type of process. From the discussion around Proposition 10.2.VIII, we know already that a regular infinitely divisible point process has a representation as a Poisson cluster process. Statement (b) then follows from Proposition 12.1.V.  $\square$

Observe that, from (a) of this proposition, the stationary singular infinitely divisible distributions can be classified into those with totally finite KLM measures, namely the Poisson randomizations, and those with unbounded KLM measure. An alternative and more interesting classification can be based on the  $\tilde{Q}$ -measures of the sets of trajectories with zero or positive asymptotic densities, as indicated below and in Example 13.2(c). First, however, note that Proposition 12.3.IX implies that a stationary Poisson cluster process, and hence any regular stationary infinitely divisible process, is necessarily mixing and hence ergodic. Interest centres therefore around the mixing properties of the singular stationary infinitely divisible processes. We follow essentially Kerstan and Matthes (1967) and MKM (1978, Chapter 6), starting from a simple general property.

**Proposition 12.4.III.** *If the stationary random measure  $\xi$  on  $\mathcal{X} = \mathbb{R}^d$  is ergodic, it cannot be represented as a mixture of two distinct stationary processes.*

PROOF. Suppose the contrary, so that  $\mathcal{P} = \alpha\mathcal{P}_1 + (1 - \alpha)\mathcal{P}_2$  say, where  $0 < \alpha < 1$  and the other three terms are stationary probability measures on  $\mathcal{M}_\mathcal{X}^\#$ . Evidently,  $\mathcal{P}_1 \ll \mathcal{P}$  and the Radon–Nikodym derivative  $d\mathcal{P}_1/d\mathcal{P}$  is invariant under shifts  $S_x$ . When  $\mathcal{P}$  is ergodic the only invariant functions are constants a.s., so  $\mathcal{P}_1 = c\mathcal{P}$ ; hence,  $c = 1$  because both  $\mathcal{P}$  and  $\mathcal{P}_1$  are probability measures. Thus,  $\mathcal{P}_1 = \mathcal{P}$ , and similarly  $\mathcal{P}_2 = \mathcal{P}$ , showing that the decomposition is trivial.  $\square$

As a converse to this result, by noting that a Poisson randomization is by definition a nontrivial discrete mixture of distinct components, we conclude as follows.

**Corollary 12.4.IV.** *No stationary Poisson randomization, nor more generally any process that can be represented as the superposition of a stationary process and a Poisson randomization, is ergodic.*

From this result it may seem plausible that no singular stationary infinitely divisible process could be ergodic. Such is not the case: the next result, due

to Kerstan and Matthes (1967), spells out in terms of the KLM measure those properties that lead to ergodicity or mixing, and we show that such properties include examples of singular infinitely divisible processes.

**Proposition 12.4.V.** *Let  $N$  be a stationary infinitely divisible point process on  $\mathcal{X} = \mathbb{R}^d$  with KLM measure  $\tilde{Q}$ .*

(a)  *$N$  is ergodic if and only if for all bounded  $A, B \in \mathcal{B}_{\mathcal{X}}$ ,*

$$\frac{1}{\ell(\mathbb{U}_n^d)} \int_{\mathbb{U}_n^d} \tilde{Q}\{\tilde{N}: N(T_x A) > 0 \text{ and } \tilde{N}(B) > 0\} dx \rightarrow 0 \quad (n \rightarrow \infty), \quad (12.4.2)$$

*in which case  $N$  is also weakly mixing.*

(b)  *$N$  is mixing if and only if for all such  $A, B$ ,*

$$\tilde{Q}\{\tilde{N}: \tilde{N}(T_x A) > 0 \text{ and } \tilde{N}(B) > 0\} \rightarrow 0 \quad (\|x\| \rightarrow \infty). \quad (12.4.3)$$

PROOF. We consider first part (b). We use the p.g.fl. formulation of the mixing condition, writing now

$$\tilde{G}[h] = \log G[h] = \int_{\mathcal{N}_{\mathcal{X}}^{\#}} \left[ \exp \left( \int_{\mathcal{X}} \log h(x) N(dx) \right) - 1 \right] \tilde{Q}(d\tilde{N}) \quad (h \in \mathcal{V}(\mathbb{R}^d)).$$

Then the mixing condition (12.3.11) is expressible as

$$\tilde{G}[h_1 S_x h_2] \rightarrow \tilde{G}[h_1] + \tilde{G}[h_2] \quad (\|x\| \rightarrow \infty). \quad (12.4.4)$$

To show that this condition is the same as (12.4.3), take  $h_1 = 1 - I_B$ ,  $h_2 = 1 - I_A$ , and consider the difference

$$\tilde{G}[h_1 S_x h_2] - \tilde{G}[h_1] - \tilde{G}[h_2] = \int_{\mathcal{N}_{\mathcal{X}}^{\#}} \left( \prod_i a_i b_i - \prod_i a_i - \prod_i b_i + 1 \right) \tilde{Q}(dN), \quad (12.4.5)$$

where  $a_i = h_1(x_i)$ ,  $b_i = h_2(x_i - x)$ , and the  $x_i$  are the points of the particular realization  $\tilde{N}$ . For the particular  $h_1, h_2$ , the integrand on the right vanishes except for realizations of  $\tilde{N}$  for which both  $\tilde{N}(B) > 0$  and  $\tilde{N}(T_x A) > 0$ , when it reduces to unity. Thus, the left-hand side of (12.4.5) coincides with the expression at (12.4.1) for such  $h_1, h_2$ , and thus (12.4.4) implies (12.4.3).

Conversely, when (12.4.3) holds, (12.4.4) holds whenever  $1 - h_i$  are indicator functions as above. For more general  $h_i \in \mathcal{V}(\mathbb{R}^d)$ , the difference at (12.4.5) is dominated by the corresponding difference when  $h_1$  and  $h_2$  are replaced by  $1 - I_B$  and  $1 - I_A$ , where  $B$  and  $A$  are the supports of  $1 - h_1$  and  $1 - h_2$ . Thus, (12.4.3) also implies (12.4.4) for these more general  $h_i$ , and part (b) is established.

To prove part (a) we need to develop some auxiliary results. Because  $\tilde{Q}$  is stationary, although perhaps only  $\sigma$ -finite, an extension of the ergodic

theorem (see comment following Proposition 12.2.II) can be applied to deduce the limit, for any convex averaging sequence  $\{A_n\}$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{\ell(A_n)} \int_{A_n} f(S_x \tilde{N}) dx = \bar{f}(\tilde{N}) \quad \tilde{Q}\text{-a.e.} \quad (12.4.6)$$

whenever the  $\mathcal{B}(\mathcal{N}_\chi^\#)$ -measurable function  $f$  satisfies  $\int_{\mathcal{N}_\chi^\#} f(\tilde{N}) \tilde{Q}(d\tilde{N}) < \infty$ .

Let  $I_V$  be the indicator function of the set  $V = \{\tilde{N}: \tilde{N}(A) > 0\}$  for bounded  $A \in \mathcal{B}(\mathbb{R}^d)$ . Then  $\tilde{Q}(V) < \infty$  by (10.2.8), so the limit in (12.4.6),  $\bar{I}_V(\tilde{N})$  say, exists  $\tilde{Q}$ -a.e. We assert that if the infinitely divisible process is ergodic, then this limit must be zero  $\tilde{Q}$ -a.e. To see this, consider for any fixed positive  $c$  the set  $J_c = \{\tilde{N}: \bar{I}_V(\tilde{N}) > c > 0\}$ , which is measurable and invariant because  $\bar{I}_V(\cdot)$  is. Furthermore,

$$\begin{aligned} c\tilde{Q}(J_c) &\leq \int_{J_c} \bar{I}_V(\tilde{N}) \tilde{Q}(d\tilde{N}) \leq \int_{\mathcal{N}_\chi^\#} \bar{I}_V(\tilde{N}) \tilde{Q}(d\tilde{N}) \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{\ell(A_n)} \int_{A_n} \int_{\mathcal{N}_\chi^\#} I_V(S_x \tilde{N}) \tilde{Q}(d\tilde{N}) dx = \tilde{Q}(V) < \infty. \end{aligned}$$

Consequently,  $J_c$  has finite  $\tilde{Q}$  measure and is invariant. If in fact  $\tilde{Q}(J_c) > 0$ , we can construct a stationary probability measure  $\mathcal{P}_V$  on  $\mathcal{N}_\chi^\#$  by setting

$$\mathcal{P}_V(\cdot) = \tilde{Q}(J_c \cap \cdot) / \tilde{Q}(J_c),$$

and it then follows that the original process has the Poisson randomization of  $\mathcal{P}_V$  as a convolution factor. But for an ergodic process this is impossible by Corollary 12.4.IV, so  $\tilde{Q}(J_c) = 0$  for every  $c > 0$ , and  $\bar{I}_V(\tilde{N}) = 0$   $\tilde{Q}$ -a.e. as asserted.

Now let  $B$  be any bounded set in  $\mathcal{B}(\mathbb{R}^d)$ , write  $W = \{\tilde{N}: \tilde{N}(B) > 0\}$ , and consider the relations

$$\begin{aligned} \frac{1}{\ell(A_n)} \int_{A_n} \tilde{Q}\{\tilde{N}: \tilde{N}(A) > 0 \text{ and } \tilde{N}(B+x) > 0\} dx \\ &= \frac{1}{\ell(A_n)} \int_{A_n} \int_{\mathcal{N}_\chi^\#} I_V(\tilde{N}) I_{S_x W}(\tilde{N}) \tilde{Q}(d\tilde{N}) dx \\ &= \int_{\mathcal{N}_\chi^\#} I_V(\tilde{N}) \left( \frac{1}{\ell(A_n)} \int_{A_n} I_{S_x W}(\tilde{N}) dx \right) \tilde{Q}(d\tilde{N}). \end{aligned}$$

We have just shown that the inner integral here  $\rightarrow 0$  as  $n \rightarrow \infty$   $\tilde{Q}$ -a.e., and because this integral  $\leq 1$  and  $\tilde{Q}(V) < \infty$ , we can apply the dominated convergence theorem and conclude that the entire expression  $\rightarrow 0$  as  $n \rightarrow \infty$ ; that is, (12.4.2) holds.

The converse implication depends on the inequality [see under (12.4.5)]

$$\begin{aligned} 0 &\leq \tilde{G}[h_1 S_x h_2] - \tilde{G}[h_1] - \tilde{G}[h_2] \\ &\leq \tilde{Q}\{\tilde{N}: \tilde{N}(A) > 0 \text{ and } \tilde{N}(T_x B) > 0\} = q(x) \quad \text{say,} \end{aligned}$$

where  $A, B$  are the supports of  $1 - h_1$  and  $1 - h_2$ . Using the elementary inequality  $e^\alpha - 1 \leq \alpha e^\alpha$  ( $\alpha > 0$ ) and taking exponentials, we obtain

$$0 \leq G[h_1 S_x h_2] - G[h_1]G[h_2] \leq G[h_1]G[h_2]q(x)e^{q(x)}. \quad (12.4.7)$$

Because  $q(x) < \tilde{Q}\{\tilde{N}: \tilde{N}(A) > 0\} < \infty$  uniformly in  $x$ , it follows that the difference in (12.4.7) is bounded by  $Kq(x)$  for some finite positive constant  $K$ . Then the integral at (12.3.9) is bounded by

$$\frac{K}{\ell(\mathbb{U}_n^d)} \int_{\mathbb{U}_n^d} q(x) \, dx,$$

which  $\rightarrow 0$  as  $n \rightarrow \infty$  when (12.4.3) holds. Proposition 12.3.VI(b) part (i) now shows that the process must be ergodic, and because the difference at (12.4.7) is already nonnegative, the stronger convergence statement at (12.3.10) must hold, and the process is in fact weakly mixing as asserted.  $\square$

The arguments used in the preceding proof can be taken further. Consider in particular the invariant function  $\bar{I}_V(\cdot)$  introduced below (12.4.6), putting  $A = \mathbb{U}^d$  say for definiteness. The trajectories  $\tilde{N}$  can be classified  $\tilde{Q}$ -a.e. by defining

$$\mathcal{S}_w = \{\tilde{N}: \bar{I}_V(\tilde{N}) = 0\}, \quad \mathcal{S}_s = \{\tilde{N}: \bar{I}_V(\tilde{N}) > 0\}.$$

Both these subsets of  $\mathcal{N}_\lambda^\#$  are measurable, and their complement has zero  $\tilde{Q}$  measure, so  $\tilde{Q}$  can be decomposed as  $\tilde{Q} = \tilde{Q}_w + \tilde{Q}_s$ , where

$$\tilde{Q}_w(\cdot) = \tilde{Q}(\mathcal{S}_w \cap \cdot), \quad \tilde{Q}_s = \tilde{Q}(\mathcal{S}_s \cap \cdot).$$

The notation here comes from the definition of a singular stationary infinitely divisible point process as *weakly singular* if  $\tilde{Q}_s$  vanishes and *strongly singular* if  $\tilde{Q}_w$  vanishes. Just as a general infinitely divisible point process can be represented as the superposition of regular and singular components, so in the stationary case the singular component can be further represented as the superposition of weakly singular and strongly singular components. Evidently, a Poisson randomization is strongly singular, and any stationary singular process that is ergodic must be weakly singular (in fact, the condition is necessary and sufficient: see Exercises 12.4.2–4). Examples of weakly singular processes are not easy to construct: one such construction that uses a modified randomization procedure starting from a renewal process with infinite mean interval length is indicated in Exercise 12.4.6. Other examples arise from the so-called stable cluster processes described in Section 13.5. See also Example 13.3(b). For additional results see MKM (1978, Sections 6.3 and 9.6), and MKM (1982, Section 6.2).

## Exercises and Complements to Section 12.4

12.4.1 Show that if an infinitely divisible point process on the c.s.m.s.  $\mathcal{X}$  is invariant under a  $\sigma$ -group  $\{T_g: g \in \mathcal{G}\}$  of transformations (Definition A2.7.I), then its KLM measure is also invariant under the transformations  $\hat{S}_g$  induced by  $\{T_g\}$ . [Hint: See Proposition 12.4.I.]

12.4.2 Show that a singular infinitely divisible process is strongly singular if and only if it can be represented as a countable superposition of Poisson randomizations. [Hint: Consider the restriction  $\tilde{Q}_n$  of  $\tilde{Q}$  to the set

$$J_n = \{\tilde{N}: \tilde{I}_V(\tilde{N}) \in ((n+1)^{-1}, n^{-1}]\},$$

where  $V = \{\tilde{N}: \tilde{N}(A) > 0\}$  for some bounded  $A \in \mathcal{B}(\mathbb{R}^d)$  (cf. the proof of Proposition 12.4.II). Then  $\tilde{Q}_n$  is totally finite, so it can be the KLM measure of a Poisson randomization. The original process is equivalent to the convolution of these randomizations.]

12.4.3 Show that for a singular infinitely divisible point process the conditions below are equivalent.

- (a) The process is weakly singular;
- (b) The process is ergodic;
- (c) If  $V \in \mathcal{I}$  then  $\tilde{Q}(V) = 0$  or  $\infty$ .

[Hint: To show (a)  $\Leftrightarrow$  (b), use the characterization of weak singularity in terms of  $\tilde{I}_V$  together with relevant parts of the proof of Proposition 12.4.II. Modify the argument of Exercise 12.4.2 to show that (a)  $\Leftrightarrow$  (c).]

12.4.4 Show that  $\mathcal{T}_\infty$  is trivial for every regular infinitely divisible point process. [Hint: Use the short-range correlation inequality of Exercise 12.3.4 and equation (12.4.7) to show that for some positive  $c$ ,

$$|\mathcal{P}(V \cap W) - \mathcal{P}(V)\mathcal{P}(W)| \leq c \tilde{Q}\{\tilde{N}(B) > 0 \text{ and } \tilde{N}(\mathcal{X} \setminus A) > 0\},$$

where for bounded  $A, B \in \mathcal{B}_\mathcal{X}$ ,  $V$  and  $W$  are in the sub- $\sigma$ -algebras determined by

$$\{\xi(B'): B' \in \mathcal{B}_\mathcal{X}, B' \subseteq B\} \quad \text{and} \quad \{\xi(A'): A' \in \mathcal{B}_\mathcal{X}, A' \in \mathcal{X} \setminus A\}.$$

Now use the regularity of  $\xi$  to show that the right-hand side  $\rightarrow 0$  as  $A \uparrow \mathcal{X}$ , and hence that events in  $\mathcal{T}_\infty$  have probability 0 or 1. MKM (1982, p. 97) attributes the result to K. Hermann.]

12.4.5 For  $n = 1, 2, \dots$  let  $N_n$  be independent stationary infinitely divisible point processes on  $\mathbb{R}$  that are Poisson randomizations (with mean number = 1) of Poisson processes at rate  $\lambda_n$ , where  $\sum_{n=1}^\infty \lambda_n < \infty$ . Verify that  $N^* = \sum_{n=1}^\infty N_n$  is a well-defined stationary point process that is singular infinitely divisible with infinite KLM measure.

12.4.6 Let a stationary infinitely divisible point process on  $\mathbb{R}_+$  have as its (stationary) KLM measure a finite positive multiple  $\alpha\mu$  of the shift-invariant regenerative measure of Example 12.1(e), so that (using the notation from there)

$$\tilde{Q}\{\tilde{N}(0, y] > 0\} = \alpha \int_0^y [1 - F(u)] du,$$

which has a finite or infinite limit according as the d.f.  $F(\cdot)$  has finite or infinite mean. Show that for  $x > y > 0$  and  $z > 0$ ,

$$\begin{aligned} \tilde{Q}(\{\tilde{N}(0, y] > 0\}, \{\tilde{N}(x, x+z] > 0\}) \\ = \alpha \int_0^y [1 - F(u)] du \int_{x-u}^{x-u+z} [1 - F(x-u+z-v)] U_0(dv), \end{aligned}$$

and that this quantity has limit zero when  $F$  has infinite mean. Hence, conclude that the point process is then weakly singular. [See MKM (1978, Section 9.6) for some other details.]

12.4.7 Show that if an infinitely divisible process is mixing then it is mixing of all orders as in Exercise 12.3.7.

## 12.5. Asymptotic Stationarity and Convergence to Equilibrium

The issue of convergence to equilibrium, or stability, for point process models has already surfaced in Chapter 4, where we illustrated the use of coupling arguments to obtain classical results for renewal and Wold processes, and in Exercise 12.3.6, where convergence was established assuming both mixing and an absolute continuity condition. This section treats convergence to equilibrium in the more general context of simple and marked point processes on  $\mathbb{R}$ . Similar results can be developed for random measures with essentially only changes in terminology, but are left to the exercises. The special problem of convergence to equilibrium from the Palm distribution is taken up in Section 13.4.

The tools available to tackle these problems have been extended greatly by the development of coupling and shift-coupling methods, which we illustrate in the present section, and the work of Brémaud and Massoulié on Poisson embedding, which we outline in Section 14.7 and use there to develop conditions for convergence to equilibrium in terms of conditional intensities. Lindvall (1992) is a basic reference on coupling ideas; for shift-coupling see Aldous and Thorisson (1993) and Thorisson (1994, 2000). In the point process context, systematic treatments concentrating on applications to queues appear in Franken *et al.* (1981), Sigman (1995), and Baccelli and Brémaud (1994); convergence to equilibrium has been discussed, for example, by Lindvall (1988) for finite-memory processes, Sigman (1995) for shift-coupling and queues, Brémaud and Massoulié (1996) and Massoulié (1998) for generalizations of the Hawkes process, and Last (2004) for the stress-release model and its analogues. Additional references are given in these papers and below.

Apart from the basic theoretical interest of these issues, their importance has increased in recent years because of the greatly increased role of simulation methods, especially Markov chain Monte Carlo methods, and the associated

need to estimate the ‘burn-in times’ required for initial system values to approximate the equilibrium values that are usually the endpoint in view.

When such problems arise in applications, we are commonly given a ‘law of evolution’ of the process and seek to establish, first, whether such a law is compatible with a stationary form for the process, and second, whether the process will converge to that stationary form when started at  $t = 0$  from some ‘initial distribution’.

To try to capture these ideas more precisely, consider first the action of shifts on a point process (simple or marked) defined on the half-line  $[0, \infty) = \mathbb{R}_+$ . If the process has distribution  $\mathcal{P}$  say, on  $\mathcal{B}(\mathcal{N}_{\mathbb{R}_+ \times \mathcal{K}}^\#)$ , then for  $u > 0$ , the action of the general shift operator  $\hat{S}_u$  on  $\mathcal{P}$  is well defined: it corresponds to shifting the time origin to  $u$ . Thus  $\hat{S}_u \mathcal{P} \equiv \mathcal{P}_u$  is a probability measure defined on the Borel sets of  $\mathcal{N}^\#([-u, \infty) \times \mathcal{K})$ . Any such measure can be projected onto the smaller sub- $\sigma$ -algebra  $\mathcal{B}(\mathcal{N}_{\mathbb{R}_+ \times \mathcal{K}}^\#)$ , on which there is then a measure that corresponds to a new (simple or marked) point process on  $\mathbb{R}_+ \times \mathcal{K}$ .

For the rest of this section the notation  $\mathcal{P}_u = \hat{S}_u \mathcal{P}$  denotes the probability measures of the shifted point processes induced in this way on  $\mathbb{R}_+ \times \mathcal{K}$ . We call an MPP on the half-line *stationary* if its distribution is invariant under these positive shifts. Trivially, the restriction to the half-line of an MPP stationary on the whole line is stationary on the half-line. Conversely, any MPP stationary on the half-line can be extended to a process stationary on the full line by shifting the origin forward (which does not change the distributions) and taking the limit of these shifted processes (Exercise 12.5.1).

In the discussion below we use the notation  $(C, 1)$ , as for Cesàro summability in analysis, to denote the behaviour of integral averages.

**Definition 12.5.I.** *Let  $N$  be an MPP on  $\mathbb{R}_+ \times \mathcal{K}$ , and let  $\mathcal{P}, \mathcal{P}_u$  be the probability measures on  $\mathcal{B}(\mathcal{N}_{\mathbb{R}_+ \times \mathcal{K}}^\#)$  associated as above with  $N$  and its shifted versions  $S_u N$ . Then  $N$  is asymptotically stationary (resp.,  $(C, 1)$ -asymptotically stationary) if as  $u \rightarrow \infty$ ,  $\mathcal{P}_u$  (resp.,  $\bar{\mathcal{P}}_u = u^{-1} \int_0^u \mathcal{P}_v dv$ ) converges weakly to a limit  $\mathcal{P}^*$  corresponding to a limit MPP  $N^*$ . It is strongly asymptotically stationary (resp., strongly  $(C, 1)$ -asymptotically stationary) if the convergence holds in variation norm.*

Brémaud and Massoulié and others refer to asymptotic stationarity as a *stability* property, having in mind the analogy with the stability of differential equations.

**Lemma 12.5.II.** *Any limit measure  $\mathcal{P}^*$  arising in Definition 12.5.I necessarily corresponds to a stationary point process on  $\mathbb{R}_+$ .*

PROOF. When

$$\hat{S}_u \mathcal{P} \rightarrow \mathcal{P}^* \quad \text{weakly} \quad (u \rightarrow \infty), \quad (12.5.1)$$

then for all finite  $u$ , the sequences  $\hat{S}_{x+u} \mathcal{P}$  and  $\hat{S}_x \mathcal{P}$  must have the same weak limit  $\mathcal{P}^*$  as  $x \rightarrow \infty$ , implying that  $\mathcal{P}^* = \hat{S}_u \mathcal{P}^*$ .



A similar argument holds in the  $(C, 1)$  case if we apply the shift  $\widehat{S}_x$  to the integrals and note that the contributions from the integrals  $\int_0^x$  and  $\int_u^{u+x}$  on the end-intervals are asymptotically negligible.  $\square$

The convergence at (12.5.1) is equivalent to requiring that for  $u > 0$  the fidi distributions should satisfy

$$F_k(A_1 + u, \dots, A_k + u; x_1, \dots, x_k) \rightarrow F_k^*(A_1, \dots, A_k; x_1, \dots, x_k), \quad (12.5.2)$$

either directly or in the  $(C, 1)$  sense, where on the right-hand side  $F^*$  refers to the fidi distributions of  $\mathcal{P}^*$ . Thus, if the shifted distributions of the initial point process converge to limits which constitute a family of fidi distributions, then these limit distributions must be those of a stationary point process. The underlying problem can be phrased alternatively as follows. Given a stationary MPP with associated probability measure  $\mathcal{P}^*$ , find the class of MPP distributions  $\mathcal{P}$  for which (12.5.1) or (12.5.2) holds. In this sense, the problem of convergence to equilibrium can also be interpreted as a *domain of attraction* problem.

Consider then what more stringent requirements may be needed for convergence to equilibrium without the  $(C, 1)$ -averaging. It is convenient here to specify the distribution  $\mathcal{P}$  of the initial point process on the half-line by treating it as the conjunction of two components: a distribution  $\Pi(\cdot)$  of initial conditions  $Z$  defined on some appropriate c.s.m.s.  $\mathcal{Z}$ , and a kernel  $\mathcal{P}(\cdot | z)$  that governs the evolution of the process  $N$  given the initial condition  $Z = z$ , so that

$$\mathcal{P}(V) = \Pr\{N \in V\} = \int_{\mathcal{Z}} \mathcal{P}(V | z) \Pi(dz) \quad (V \in \mathcal{B}(\mathcal{N}_{\mathbb{R}_+}^{\#})). \quad (12.5.3)$$

This formulation follows Brémaud and Massoulié (1994, 1996), which in turn has its origins in the fundamental paper of Kerstan (1964b).

For  $u \geq 0$  let  $\mathcal{P}_u(\cdot)$  denote the distribution of  $S_u N$  (so that in our earlier notation,  $\mathcal{P}_u = \widehat{S}_u \mathcal{P}$  and  $\mathcal{P}_0 = \mathcal{P}$ ), and suppose that there exists a family of distributions  $\Pi_u$  on the space  $\mathcal{Z}$  of initial conditions such that for these  $u$ ,

$$\mathcal{P}_u(V) = \int_{\mathcal{Z}} \mathcal{P}(V | z) \Pi_u(dz) \quad (V \in \mathcal{B}(\mathcal{N}_{\mathbb{R}_+}^{\#})) \quad (12.5.4)$$

with the same kernel  $\mathcal{P}(\cdot | z)$  as in (12.5.3). We can interpret  $\Pi_u$  as the initial conditions that apply when the time origin is shifted to  $u$ . When such a representation is available, it is not difficult to show that convergence of the distributions  $\Pi_u$  of the initial conditions implies convergence of the shifted distributions  $\mathcal{P}_u$ .

**Proposition 12.5.III.** *Let  $\mathcal{Z}$  be a c.s.m.s. of initial conditions  $Z$  for some point process and  $\{\Pi_u\}$  a family of probability measures on  $\mathcal{B}_{\mathcal{Z}}$ . Let  $\mathcal{P}(V | z)$*

be a measurable family of point processes on  $\mathbb{R}_+$ , and suppose that for all  $u \geq 0$  and  $V \in \mathcal{B}(\mathcal{N}_{\mathbb{R}_+}^\#)$ ,

$$\mathcal{P}_u(V) \equiv \widehat{S}_u \mathcal{P}_0(V) \equiv \widehat{S}_u \int_{\mathcal{Z}} \mathcal{P}(V \mid z) \Pi_0(dz) = \int_{\mathcal{Z}} \mathcal{P}(V \mid z) \Pi_u(dz).$$

- (a) If  $\Pi_u \rightarrow \Pi^*$  weakly, and the kernel  $\mathcal{P}(\cdot \mid z)$  takes continuous functions on  $\mathcal{N}_{\mathbb{R}_+}^\#$  into continuous functions on  $\mathcal{Z}$ , then  $\mathcal{P}_u \rightarrow \mathcal{P}^*$  weakly and  $\mathcal{P}^*(V) = \int_{\mathcal{Z}} \mathcal{P}(V \mid z) \Pi^*(dz)$  (i.e. the point process  $N$  with distribution  $\mathcal{P} = \mathcal{P}_0$  is asymptotically stationary with limit distribution  $\mathcal{P}^*$ ).
- (b) If  $\Pi_u \rightarrow \Pi^*$  in variation norm, then the point process  $N$  of (a) is strongly asymptotically stationary with the same limit distribution  $\mathcal{P}^*$ , and

$$\|\mathcal{P}_u - \mathcal{P}^*\| \leq \|\Pi_u - \Pi^*\|. \quad (12.5.5)$$

PROOF. To establish asymptotic stationarity we have to prove weak convergence of the  $\mathcal{P}_u$ . This means showing that  $\int_{\mathcal{N}^\#(\mathbb{R}_+)} h(N) \mathcal{P}_u(dN)$  converges for any bounded, continuous function  $h(N)$  on  $\mathcal{N}_{\mathbb{R}_+}^\#$ . But under assumption (a), the integral  $\int_{\mathcal{N}^\#(\mathbb{R}_+)} h(N) \mathcal{P}(dN \mid z)$  defines a bounded continuous function of  $z \in \mathcal{Z}$ , and then the weak convergence of  $\Pi_u$  to  $\Pi^*$  implies

$$\begin{aligned} \int_{\mathcal{N}^\#(\mathbb{R}_+)} h(N) \mathcal{P}_u(dN) &= \int_{\mathcal{Z}} \left( \int_{\mathcal{N}^\#(\mathbb{R}_+)} h(N) \mathcal{P}(dN \mid z) \right) \Pi_u(dz) \\ &\rightarrow \int_{\mathcal{Z}} \left( \int_{\mathcal{N}^\#(\mathbb{R}_+)} h(N) \mathcal{P}(dN \mid z) \right) \Pi^*(dz) = \int_{\mathcal{N}^\#(\mathbb{R}_+)} h(N) \mathcal{P}^*(dN). \end{aligned}$$

Under assumption (b), the same relations hold under the weaker requirement that the integral  $\int_{\mathcal{N}^\#(\mathbb{R}_+)} h(N) \mathcal{P}(dN \mid z)$  is merely bounded and measurable, which follows from the boundedness of  $h$  and the measurability condition on the family  $\mathcal{P}(\cdot \mid z)$ .

The bound in (12.5.5) follows from the fact that  $\mathcal{P}(\cdot \mid z)$  is a stochastic kernel, so that, for any totally finite signed measure  $\Psi$  on  $\mathcal{B}_{\mathcal{Z}}$ ,

$$\begin{aligned} &\left\| \int_{\mathcal{N}^\#(\mathbb{R}_+) \times \mathbb{R}_+} \mathcal{P}(dN \mid z) \Psi(dz) \right\| \\ &\leq \sup_{h: |h(N)| \leq 1} \int_{\mathcal{N}^\#(\mathbb{R}_+)} |h(N)| \int_{\mathbb{R}_+} \mathcal{P}(dN \mid z) [\Psi^+(dz) + \Psi^-(dz)] \leq \|\Psi\|, \end{aligned}$$

where  $\Psi^+, \Psi^-$  are the positive and negative parts of  $\Psi$  from its Jordan–Hahn decomposition, and the supremum is taken over all measurable functions bounded by 1.  $\square$

The major advantage of this form of representation is that the process of ‘initial conditions’ can usually be represented as a Markov process  $Z(t)$  say,

thus reducing the problem of convergence to equilibrium for point processes to the well-developed theory of convergence to equilibrium for Markov processes.

Such a representation is always possible in principle because the initial condition can be taken as the history of the process up to time 0, and such a history can itself be regarded as a point in a c.s.m.s. This still leaves us within the framework of Markov processes on a c.s.m.s. (or more generally a Polish space). Consequently, problems for point processes concerning conditions for the existence of and convergence to equilibrium distributions, are reduced to corresponding problems for general Markov chains [see, e.g., Meyn and Tweedie (1993a, b)]. Similarly, the notions of coupling and shift-coupling, which again have their origins in Markov chain theory, can be invoked and applied within the point process context.

The treatment becomes greatly simplified if a condensed representation can be found, for example, one in which the space of initial conditions can be reduced to a low-dimensional Euclidean space. The goal in practice, then, is to identify a simple representation for the space  $\mathcal{Z}$  of initial conditions, and to examine the conditions for convergence of the induced Markov process  $Z(\cdot)$ .

To establish a representation (12.5.4) for a given point process model, the first step is to identify a convenient state space  $\mathcal{Z}$  for the Markov process of initial conditions. The essential feature that such a representation should possess is that knowledge of the current value  $Z(u)$  should be sufficient to fully determine the fidi distributions for the evolution of the process beyond time  $u$ , any further information from the past being redundant.

The other components needed to complete the representation are the form of the state transition probabilities for the Markov process  $Z(\cdot)$  and the mappings  $\mathcal{P}(\cdot | z)$  determining the evolution of the process from a given value of  $z$ . Most commonly, this is done by identifying a mapping,  $\Phi: \mathcal{N}_{\mathbb{R}_-}^\# \mapsto \mathcal{Z}$  say, that condenses the information from the past trajectory  $\{N(t): t < 0\}$  of the point process onto a particular value  $z = Z(0) \in \mathcal{Z}$ . When this transformation is applied to  $S_u N$ , its value  $\Phi(S_u N) = Z(u)$  represents the updated value of the initial condition, so that from there on its evolution is governed by  $\mathcal{P}(\cdot | Z(u))$ . Under these conditions the transition probabilities for the process  $Z(\cdot)$   $B \in \mathcal{B}_{\mathcal{Z}}$  are given by

$$P_u(B | z) = \mathcal{P}(S_u \Phi^{-1}(B) | z) = \hat{S}_u \mathcal{P}(\Phi^{-1}(B) | z) \quad (B \in \mathcal{B}_{\mathcal{Z}}).$$

Integrating over the further evolution from  $u$  to  $u + v$  yields the Chapman–Kolmogorov equation

$$\begin{aligned} P_{u+v}(B | z) &= \int_{\mathcal{Z}} \mathcal{P}(S_v \Phi^{-1}(B) | z') P_u(dz' | z) \\ &= \int_{\mathcal{Z}} P_v(B | z') P_u(dz' | z). \end{aligned}$$

Thus, once the mapping  $\Phi$  has been identified, the provisions of Proposition 12.5.III, especially the bound (12.5.5), can be invoked to establish asymptotic stationarity for the associated point process.

We now illustrate the proposition with some well-known examples.

EXAMPLE 12.5(a) *Convergence to equilibrium of renewal and Wold processes.* Consider first the simpler case of the renewal process, supposing the lifetime distribution to be nonlattice, with finite mean  $\mu = \lambda^{-1}$ . A convenient choice for  $Z$  here is the forward recurrence time defined as in (3.4.15). The space  $\mathcal{Z}$  becomes the half-line  $\mathbb{R}_+$  and the conditional point process corresponding to  $\mathcal{P}(\cdot | z)$  is the ordinary renewal process (with initial point at 0) shifted through  $z$ . It remains to investigate the conditions in parts (a) and (b) of the proposition.

Continuity of the integral  $\int_{\mathcal{N}^\#(\mathbb{R}_+)} h(N) \mathcal{P}(dN | z)$  as a function of  $Z$ , here the value  $z$  say of the forward recurrence time, follows from the observation (Lemma 12.1.I) that the shifted realization  $S_u N$  of an element  $N \in \mathcal{N}^\#_{\mathbb{R}_+}$  depends continuously on  $u$ , so if  $h(\cdot)$  is a bounded and continuous function of  $N(\cdot)$ ,  $h(S_u N)$  is a bounded continuous function of  $u$ . Continuity of  $\int_{\mathcal{N}^\#(\mathbb{R}_+)} h(N) \mathcal{P}(dN | Z = z)$  then follows by dominated convergence.

The other requirement of condition (a) amounts here to the weak convergence of the distribution  $\Pi_u$  of the forward recurrence time at time  $u$  to its equilibrium form  $\Pi^*$ . This was established for nonlattice lifetime distributions with finite mean in Example 4.4(a). In fact, the stronger result required for condition (b) holds when the lifetime distribution is spread out, as noted in Exercise 4.4.4. Thus a renewal process having a spread-out lifetime distribution with finite mean is both weakly and strongly asymptotically stationary, from an arbitrary initial condition.

Similar results for the Wold process can be obtained from the discussion in Section 4.5. Here it is convenient to describe the initial conditions in terms of the pair  $(X, Y) = (t_0(N), t_0(N) - t_{-1}(N))$  representing the time to the first event and the length  $L_0(N)$  of the (unobserved) initial interval. Let  $(X_u, Y_u)$  denote their values after a shift of the time origin through  $u$ . Conditions for the convergence of the pair  $(X_u, Y_u)$  are obtained in Proposition 4.5.VI. For example, if the recurrence time distribution  $G$  in that proposition is nonlattice but spread-out, and the stationary interval distribution has finite mean, then convergence holds in variation norm and the associated Wold process is strongly asymptotically stationary. In particular this is the case when the transition kernel satisfies the simpler Condition 4.5.I'.  $\square$

The next example introduces two extensions of principle: the point process under consideration is now an MPP, and the underlying Markov process may not be directly observable. The extension to a marked process causes no essential difficulties; the function  $h$  in the proposition now needs to be taken as a function on the space  $\mathcal{N}^\#_{\mathbb{R}_+ \times \mathcal{K}}$  rather than on  $\mathcal{N}^\#_{\mathbb{R}_+}$ , with the shifts defined on the first component only, but no essentially new points arise. To accommodate the non-observable parts of the process, or more generally any random variables in addition to those described by the observed history of the point process, the probability space needs to be taken in the general form  $(\Omega, \mathcal{E})$  and the explicitly defined shifts  $S_u$  on  $\mathcal{N}^\#_{\mathbb{R}_+}$  replaced by a general flow

$\theta_u$  on  $(\Omega, \mathcal{E})$  which is compatible with the shifts [see the discussion following (12.1.3) and elsewhere in Section 12.1]. The initial conditions must then carry sufficient information to define the initial state of all components of the process, observable or otherwise, and it is again their evolution with  $u$  which governs the convergence of the point process.

EXAMPLE 12.5(b) *Convergence to equilibrium of BMAPs and MMPPs* [see Examples 10.3(e), (h), (i)]. The BMAPs of Example 10.3(i) are treated there as examples of MPPs. But, any BMAP shares with the simpler MMPP model of Examples 10.3(e),(g) the feature that its evolution is fully determined by the evolution of a hidden or at best partly observed Markov process,  $X(t)$  say. Indeed, equations (10.3.23) spell out the form of the conditional intensity as a function of the current state of the process. In such a situation, the space of initial conditions  $\mathcal{Z}$  reduces to the state space  $\mathbb{X}$  of the process  $X(t)$  (and in these examples  $\mathbb{X}$  is just a finite set of points,  $\{1, \dots, K\}$  say). The family of distributions  $\mathcal{P}(\cdot | z)$  of Proposition 12.5.III becomes the family of distributions on  $(\mathcal{N}_{\mathbb{R}}^{\#}, \mathcal{B}_{\mathcal{N}_{\mathbb{R}}^{\#}})$  of the BMAP started from the initial condition  $X(0) = z$ . The distribution  $\Pi_0$  of the proposition is the distribution of  $X(0)$  on the finite set  $\mathbb{X}$ ,  $\pi_0 = \{\pi_k(0)\}$  say, and  $\Pi_u$ , corresponding to the distribution of  $X(u)$ , is given by  $\pi_u^{\top} = \pi_0^{\top} P^u$ , where  $P^u = e^{uQ}$  is the matrix of transition probabilities  $p_{ij}(u)$ . Because  $\mathbb{X}$  is finite, the transition probabilities  $P^u$  converge in variation norm to the limit matrix  $\pi_{\infty} \mathbf{1}^{\top}$ , where  $\pi_{\infty}$  is the stationary distribution, as soon as the process  $X(t)$  is irreducible. It then follows from Proposition 12.5.III(b), and specifically from (12.5.5), that a BMAP with irreducible governing process  $X(t)$  is strongly asymptotically stationary, no matter what the initial condition.  $\square$

In tackling more complex examples, several different approaches have been suggested. In most cases, as illustrated already for the renewal process, the process  $Z_u$  takes the form of a continuous-time, mixed jump–diffusion process on a continuous state space such as  $\mathbb{R}^d$ . Conditions for the convergence to equilibrium for such processes, based on extensions of Foster–Lyapunov conditions for stability, have been set out with a considerable degree of generality in Meyn and Tweedie (1993b); their book, Meyn and Tweedie (1993a), contains a compendium of related material on general state-space Markov chains.

Application of the Markov process conditions to any particular point process example requires a careful analysis of the properties of the induced process  $\{Z_u\}$ , and may be far from a trivial exercise (cf. Exercise 12.5.2). For example, in a comprehensive study, Last (2004) has used this approach to derive stability properties for a class of models that are related to the stress-release model of Example 7.2(g). Further examples include models in reliability for repairable systems [Kijima (1989), Last and Szekli (1998)] and work-load processes [Browne and Sigman (1992)]; see also Exercise 12.5.3. The difficulty revolves around the subtle behaviour of the Markov process itself. Zheng (1991) derived general conditions for irreducibility and positive recurrence of the simple stress-release model, and Last extended the discussion to more

general processes, and gave estimates for the rate of convergence and convergence of moments. For the more complex linked stress-release model and its variants, as in Example 7.3(d), many basic questions remain open. Even sufficient conditions for ergodicity have been established only in very special cases [Bebbington and Borovkov (2003)], and no necessary and sufficient conditions are known.

Another general approach is to invoke coupling or shift-coupling results. These have the advantage that they directly yield convergence in variation norm. They may be applied to the process of initial conditions, to the point process itself, or to the process of intervals appearing in the Palm distributions. Early applications of coupling arguments to point processes appear in Hawkes and Oakes (1974), Berbée (1979), and Lindvall (1988). Recent texts such as Lindvall (1992), Sigman (1995), and Thorisson (2000) cover much wider territory.

We gave both the underlying definition of coupling and the basic coupling inequality in Lemma 11.1.I. To apply the inequality directly to point processes, we take the stochastic processes  $X(t)$ ,  $Y(t)$  in Lemma 11.1.I to be shifted versions  $S_t N$ ,  $S_t N'$ , of two simple or marked point processes  $N$ ,  $N'$ , so that the parameter  $t$  refers to the extent of the time shift. It is convenient to treat the processes as defined on a common probability space  $(\Omega, \mathcal{E}, \Pr)$ , and to denote by  $\theta_t$  the flows associated with the shifts  $S_t$ .

We say that  $N$  and  $N'$  *couple* if there exist versions  $\tilde{N}$ ,  $\tilde{N}'$  of  $N$ ,  $N'$ , and a finite stopping time  $T$ , such that  $\theta_T \tilde{N} = \theta_T \tilde{N}'$  a.s. The last equation means that the trajectories on  $[0, \infty)$  of the shifted versions are a.s. equal, so that  $\theta_{T+t} \tilde{N} = \theta_{T+t} \tilde{N}'$  a.s. for  $t > 0$ . Similarly we say that  $N$  and  $N'$  *shift-couple* if there exist versions  $\tilde{N}$ ,  $\tilde{N}'$  of  $N$ ,  $N'$  respectively, and finite stopping times  $T$ ,  $T'$  such that  $\theta_T \tilde{N} = \theta_{T'} \tilde{N}'$ ; that is,  $\theta_{t+T} \tilde{N} = \theta_{t+T'} \tilde{N}'$  a.s. for  $t \geq 0$ . In this context the basic coupling inequality of Lemma 11.1.I extends as follows. Note that because coupling equalities hold between shifts of versions of the original MPPs, rather than between the original MPPs themselves, the inequalities involve only the distributions of the original processes.

**Lemma 12.5.IV** (Coupling Inequalities). *Let  $N$ ,  $N'$  be jointly distributed MPPs on  $\mathbb{R}_+ \times \mathcal{K}$ , and denote by  $\mathcal{P}$ ,  $\mathcal{P}'$  the probability measures induced on  $\mathcal{B}(\mathcal{N}^\#(\mathbb{R}_+ \times \mathcal{K}))$  by  $N$ ,  $N'$ , respectively, and  $\|\cdot\|$  the total variation norm.*

(a) *Suppose  $N$  and  $N'$  couple, with coupling time  $T$ . Then*

$$\|\hat{S}_t \mathcal{P} - \hat{S}_t \mathcal{P}'\| \leq 2 \Pr\{T > t\}. \quad (12.5.6a)$$

(b) *Suppose  $N$  and  $N'$  shift-couple, with coupling times  $T$ ,  $T'$ . Then*

$$\left\| \frac{1}{t} \int_0^t \hat{S}_u \mathcal{P} \, du - \frac{1}{t} \int_0^t \hat{S}_u \mathcal{P}' \, du \right\| \leq 2 \Pr\{\max(T, T') > Ut\}, \quad (12.5.6b)$$

where  $U$  is uniformly distributed on  $(0, 1)$ , independent of  $T$ ,  $T'$ ,  $N$ ,  $N'$ .

PROOF. The first inequality is proved in Lemma 11.1.I. For the other we follow Thorisson (1994; 2000, Theorem 5.3.1). By definition of shift-coupling, there exist a.s. finite stopping times  $T, T'$  such that  $\theta_T N = \theta_{T'} N'$  is a successful coupling of  $N$  and  $N'$ .

Let  $U$  be a r.v. uniformly distributed on  $(0, 1)$  independent of  $(T, T', N, N')$ . Then so is  $1 - U$ , and hence for any finite  $t > 0$ ,  $\theta_{Ut} N$  and  $\theta_{(1-U)t} N$  are copies of the same process. For such  $U$  and  $t$ ,  $(T + Ut) \bmod t = t(T/t + U - \lfloor T/t + U \rfloor)$  is uniformly distributed on  $(0, t)$ , independent of  $(T, T', N, N')$  ( $\lfloor x \rfloor$  denotes the largest integer  $\leq x$ ). Therefore  $\theta_{(T+Ut) \bmod t} N$  is also a copy of  $\theta_{Ut} N$ . Similarly,  $\theta_{(T'+Ut) \bmod t} N'$  is a copy of the process  $\theta_{Ut} N'$ .

On the set  $C \equiv \{Ut + \max(T, T') \leq t\}$ ,  $(T + Ut) \bmod t = T + Ut$ , and thus,

$$\theta_{(T+Ut) \bmod t} N = \theta_{Ut} \theta_T N = \theta_{Ut} \theta_{T'} N' = \theta_{(T'+Ut) \bmod t} N' \quad \text{on } C,$$

the central equality by the assumed shift-coupling. Thus on  $C$  we have a shift-coupling of  $\theta_{Ut} N$  and  $\theta_{Ut} N'$ , and therefore

$$\|\hat{S}_{Ut} \mathcal{P} - \hat{S}_{Ut} \mathcal{P}'\| \leq 2[1 - \Pr(C)].$$

But  $1 - \Pr(C) = \Pr\{Ut + \max(T, T') > t\} = \Pr\{\max(T, T') > (1 - U)t\}$ , which is the same as  $\Pr\{\max(T, T') > Ut\}$ . Taking expectations over  $U$  in the displayed inequality yields the left-hand side of (12.5.6b).  $\square$

As a corollary, we obtain the following conditions for asymptotic stationarity.

**Proposition 12.5.V.** *Let  $N, N'$  be jointly distributed MPPs on  $\mathbb{R}_+ \times \mathcal{K}$ , and suppose that  $N'$  is stationary.*

- (a) *If  $N$  couples with  $N'$ , then  $N$  is strongly asymptotically stationary, with limit process  $N'$ .*
- (b) *If  $N$  shift-couples with  $N'$ , then  $N$  is strongly  $(C, 1)$ -asymptotically stationary, with limit process  $N'$ .*

PROOF. The proof follows from Lemma 12.5.IV and the observation that when  $N'$  is stationary, its distribution is invariant under the shifts  $\hat{S}_u$ .  $\square$

As a further corollary of these results it follows that the ergodic theorems from Section 12.2 extend to processes which are asymptotically stationary. Suppose that  $N$  is asymptotically stationary and that both  $N$  and the limit process  $N'$  have boundedly finite first moment measures. The limit process  $N'$  is stationary by assumption, so that the ergodic Theorem 12.2.IV applies to  $N'$ , as also to any of its a.s. versions. If  $N$  shift couples to  $N'$  there are versions of both processes, the realizations of which coincide a.s. after certain realization-dependent but finite time-shifts. The existence a.s. of the  $(C, 1)$  limits  $t^{-1} \int_0^t f(S_u N) du$  for the realizations of the version of the stationary process therefore implies the existence a.s. of the same  $(C, 1)$  limits for the version of the approximating process. But the a.s. existence of the  $(C, 1)$  limits does not depend on which version is used (the versions are equal outside a set of realizations with zero probability measure), and so the ergodic behaviour applies also to the original version of the approximating process.

**Corollary 12.5.VI.** *The a.s. ergodic results of Theorem 12.2.IV apply to asymptotically stationary MPPs whenever the original process and its stationary limit have boundedly finite first moment measures.*

In fact, a great deal more follows from the work of Aldous and Thorisson, who showed in particular that coupling is strongly linked to mixing properties and the behaviour of the tail  $\sigma$ -field  $\mathcal{T}_\infty$ , whereas shift-coupling is linked in a parallel way to ergodicity properties and the behaviour of the invariant  $\sigma$ -field  $\mathcal{I}$ . The results depend on a deep analysis of the ergodic properties of Markov chains, which we do not develop in detail here, referring to the papers cited for proofs, and to Sigman (1995) for an elaboration of their applications to point processes. The most important results for our purposes are summarized in the proposition below [see in particular Thorisson (1994, Section 4)], where for the sake of completeness Proposition 12.5.V is included.

**Proposition 12.5.VII.** *Suppose that  $N$  and  $N'$  are two MPPs on  $\mathcal{X} = \mathbb{R}_+ \times \mathcal{K}$ , and that  $N'$  is stationary.*

(A) (Coupling Equivalences) *The following statements are equivalent.*

- (i)  *$N$  and  $N'$  couple.*
- (ii)  *$N$  is strongly asymptotically stationary with limit process  $N'$ .*
- (iii)  *$N$  and  $N'$  induce the same probability distribution on the tail  $\sigma$ -algebra  $\mathcal{T}_\infty$ .*

(B) (Shift-Coupling Equivalences) *The following statements are equivalent.*

- (i)  *$N$  and  $N'$  shift-couple.*
- (ii)  *$N$  is strongly  $(C, 1)$ -asymptotically stationary with limit process  $N'$ .*
- (iii)  *$N$  and  $N'$  induce the same probability distribution on the invariant  $\sigma$ -algebra  $\mathcal{I}$ .*

Let us note at least one of the remarkable consequences of these results.

**Corollary 12.5.VIII.** *Weak and strong  $(C, 1)$ -asymptotic stationarity coincide.*

PROOF. Suppose  $N$  is weakly  $(C, 1)$ -asymptotically stationary with limit process  $N'$ , and denote by  $\mathcal{P}, \mathcal{P}'$  the probability measures they induce on  $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ . Then for any invariant set  $E \in \mathcal{N}_\mathcal{X}^\#$ ,

$$\mathcal{P}(E) = \widehat{S}_x \mathcal{P}(E) \rightarrow \mathcal{P}'(E),$$

so  $\mathcal{P}$  and  $\mathcal{P}'$  coincide on  $\mathcal{I}$ . Thus condition B(iii) holds and hence  $N$  and  $N'$  shift-couple. The rest of the corollary follows from Proposition 12.5.VII.  $\square$

In applications we are still left with the problem of identifying situations where coupling holds. For point processes, identifying shift-coupling is generally easier for the associated sequence of intervals than for the continuous-time process, because regeneration points, which often lie at the heart of coupling arguments, are more commonly linked to the intervals. Thus a common approach is to establish shift-coupling for the intervals, and then to refer to the



Palm equivalence between counting and interval properties to establish shift-coupling and associated ergodic results for the continuous-time process. See the further discussion in Section 13.5 and Sigman (1995).

As an example of coupling (rather than shift-coupling) arguments, we outline in the next example the basic argument used by Hawkes and Oakes (1974) to establish convergence to equilibrium of the simple Hawkes process. The extensive work on linear and nonlinear Hawkes processes by Brémaud, Massoulié, and colleagues is outlined in Section 14.7 in conjunction with their ‘Poisson embedding’ technique.

EXAMPLE 12.5(c) *Convergence to equilibrium of the Hawkes process* [continued from Examples 6.3(c), 7.2(b)]. We introduced the Hawkes process at Example 6.3(c) as an example of a Poisson cluster process with a special kind of branching process cluster. Then Example 7.2(b) shows that, when started with the zero initial condition at  $t = 0$ , the process is characterized by a conditional intensity function of the form

$$\lambda_0(t) = \nu + \int_0^t \mu(t-u) N_0(du), \quad (12.5.7)$$

where  $\nu$  is an immigration rate and  $\mu(u)$  is a density for the intensity measure, assumed to satisfy the condition

$$\int_0^\infty \mu(x) dx \equiv \rho < 1. \quad (12.5.8)$$

The existence of a stationary version of the process follows from the Poisson clustering representation, or can be established by letting the origin retreat to  $-\infty$  in (12.5.7) (see Exercise 12.5.4).

We now compare two versions of the process:  $N_0$  starts from 0 at time 0 and follows (12.5.7), whereas  $N^\dagger$  is a stationary version with the complete conditional intensity

$$\lambda^\dagger(t) = \nu + \int_{-\infty}^t \mu(t-u) N^\dagger(du) \quad (12.5.9)$$

and mean rate  $m = \nu/(1 - \rho)$ .

For both versions, we consider the effect of shifting the origin forward to  $s$ ; equivalently, we consider the shifted versions  $S_s N_0$  and  $S_s N^\dagger$  that bring the origin back to 0.  $S_s N^\dagger$  can be split into two components: one component has the same structure as  $S_s N_0$ , being built up from clusters initiated by immigrants arriving after time  $-s$ , and the component  $N_{-s}^\dagger$  consists of the ‘offspring’ of points that occurred before time  $-s$ . On  $\mathbb{R}_+$  the contributions from the latter form a Poisson process whose intensity, conditional on  $\mathcal{H}_{-s}$ , is given by

$$\lambda_{-s}^\dagger(t) = \int_{-\infty}^{-s} \mu(t-u) N_{-s}^\dagger(du). \quad (12.5.10)$$

Note that from the integrability of  $\mu$  as at (12.5.8),

$$\mathbb{E}[\lambda_{-s}^\dagger(t)] = m \int_{t+s}^{\infty} \mu(x) dx \rightarrow 0 \quad (s \rightarrow \infty),$$

and more generally, for any  $T < \infty$ ,

$$\begin{aligned} \Pr\{N_s^\dagger(0, T) > 0\} &= \mathbb{E}[1 - \exp(-\int_0^T \lambda_{-s}^\dagger(t) dt)] \\ &\leq \mathbb{E}[\int_0^T \lambda_{-s}^\dagger(t) dt] \\ &\leq mT \int_s^{\infty} \mu(x) dx \rightarrow 0 \quad (s \rightarrow \infty). \end{aligned} \quad (12.5.11)$$

Consider now the corresponding probability measures on  $\mathcal{N}^\#(\mathbb{R}_+)$ ,  $\mathcal{P}_0$  and  $\mathcal{P}^\dagger$  say, but restrict attention to their projections onto  $\mathcal{N}^\#([0, S])$ . Using the property  $\|\mathcal{P} * \mathcal{Q}\| = \|\mathcal{P}\| \cdot \|\mathcal{Q}\|$  of the variation norm, which implies that the contribution from the common distribution of  $S_s N_0$  disappears from the difference below, we have

$$\|\hat{S}_s \mathcal{P}_0 - \mathcal{P}^\dagger\|_{[0, S]} = \|\hat{S}_s \mathcal{P}_0 - \hat{S}_s \mathcal{P}^\dagger\|_{[0, S]} \leq \Pr\{N_{-s}^\dagger[0, S] > 0\} \rightarrow 0 \quad (s \rightarrow \infty).$$

This implies convergence of the fidi distributions of the two processes on  $[0, S]$ , for any  $S > 0$ , hence their weak convergence, and hence the weak asymptotic stationarity of  $N_0$ .

We can extend this result to initial conditions specified by a distribution  $\mathcal{P}^J$  of a point process  $N^J$  on  $\mathbb{R}_-$ , assuming that for  $t > 0$  the process evolves according to the same dynamics as before. As in the previous discussion, for  $t > 0$  the conditional intensity can then be written as the sum of two components, one corresponding to  $N_0$  and deriving from immigrants arriving after  $t = 0$ , and the other the residual risk remaining from events occurring before  $t = 0$ . Denoting the shifted version of the latter process by  $N_{-s}^J$ , we can write its contribution to the conditional intensity as in (12.5.10) but with  $N_{-s}^J$  replacing  $N_{-s}^\dagger$ . Similar arguments to those used before show that, for a given realization of  $N^J$  on  $\mathbb{R}_-$ ,

$$\Pr\{N_{-s}^J[0, S] > 0\} \leq \int_{t=0}^S \int_{-\infty}^{-S} S_s N^J(du) \mu(t-u) dt = \sum_{i=1}^{\infty} \Delta(s, i),$$

where  $\Delta(s, i) = \int_{-s-u_i}^{-s-u_i+S} \mu(x) dx$ , and the  $u_i$  are the points of  $N^J$  in  $\mathbb{R}_-$ .

If the initial distribution  $\mathcal{P}^J$  is concentrated on an individual realization, it follows that the resultant process will be weakly asymptotically stationary provided  $\Sigma(s) \equiv \sum_i \Delta(s, i) < \infty$ , and  $\Sigma(s) \rightarrow 0$  as  $s \rightarrow \infty$ . For under these conditions, comparing the two projections on  $[0, S]$ , we have

$$\begin{aligned} \|\hat{S}_s \mathcal{P}^J - \mathcal{P}^\dagger\|_{[0, S]} &\leq \|\hat{S}_s \mathcal{P}_0 - \hat{S}_s \mathcal{P}^J\|_{[0, S]} + \|\hat{S}_s \mathcal{P}_0 - \mathcal{P}^\dagger\|_{[0, T]} \\ &\leq \Pr\{N_{-s}^J(0, S) > 0\} + \Pr\{N_{-s}^\dagger[0, S] > 0\} \\ &\rightarrow 0 \quad (s \rightarrow \infty). \end{aligned}$$

If the initial distribution  $\mathcal{P}^J$  is such that  $N^J$  has mean density  $m(u)$  on  $\mathbb{R}_-$ , then a similar conclusion holds provided

$$\int_{t=-\infty}^{-s} m(u) \left( \int_0^s \mu(t-u) dt \right) du \rightarrow 0 \quad (s \rightarrow \infty),$$

which is certainly satisfied if, for example,  $m(u)$  is bounded.

Under slightly stronger conditions we can establish strong asymptotic stationarity as well. Suppose that, in addition to (12.5.8),  $\mu(x)$  also satisfies the condition

$$\int_0^\infty x \mu(x) dx < \infty. \quad (12.5.12)$$

This condition implies that, if a parent event has a direct offspring, the mean gestation period, that is to say, the mean time to the appearance of the offspring, is finite, and hence, because (12.5.8) implies that the mean number of offspring is also finite, that the random time  $T$  from the appearance of an ancestor (cluster centre) to the last of his descendants (last cluster member) is also finite; that is,  $E(T) < \infty$ .

The probability that a cluster is initiated at some time  $-u \leq -s$ , and still produces some members in  $\mathbb{R}_+$ , is given by  $1 - F_T(u) = \Pr\{T > u\}$ . Treating this as a thinning probability, and using the Poisson character of the arrival of ancestors (cluster centres), the probability that (in the stationary process) none of the ancestors arriving before  $-s$  produces offspring in  $\mathbb{R}_+$  is equal to

$$\Pr\{N_{-s}^\dagger[0, \infty) > 0\} = 1 - \exp\left(-\nu \int_s^\infty [1 - F_T(u)] du\right).$$

Because  $\int_0^\infty [1 - F_T(u)] du = E(T) < \infty$ , we have  $\Pr\{N_{-s}^\dagger[0, \infty) > 0\} \rightarrow 0$ .

The occurrence time  $T_s$  for the last point in  $\mathbb{R}_+$  of a cluster initiated before time  $-s$  is therefore a.s. finite, and acts as a coupling time between the processes  $S_s N_0$  and  $N^\dagger$  on  $\mathbb{R}_+$ . The coupling time inequality (12.5.6a) then yields

$$\|\hat{S}_s \mathcal{P}_0 - \mathcal{P}^\dagger\|_{[0, \infty]} \leq 2\Pr\{T_s > 0\} = 2\Pr\{N_{-s}^\dagger[0, \infty) > 0\} \rightarrow 0 \quad (s \rightarrow \infty).$$

Thus, when the density  $\mu$  satisfies both (12.5.8) and (12.5.13), the process starting from zero is strongly asymptotically stationary.

Again the result can be extended to more general initial conditions  $N^J$ , for example, when

$$\int_{t=0}^\infty \int_{-\infty}^0 \mu(t-u) N^J(du) dt = \sum \sigma(|u_j|) < \infty, \quad (12.5.14)$$

where  $\sigma(u) = \int_u^\infty \mu(x) dx$  and  $\{u_j\}$  enumerates the points of  $N^J$  over  $(-\infty, 0)$ . The arguments are similar to those used before, and are outlined in Exercise 12.5.6.  $\square$

Note that the argument used above to derive strong asymptotic stationarity is not peculiar to Hawkes processes, but holds for a wide range of Poisson cluster processes. Exercise 12.5.7 gives some details and examples.

## Exercises and Complements to Section 12.5

- 12.5.1 Prove the statement above Definition 12.5.I, that a process stationary under positive shifts on the half-line can be extended to a process stationary on the whole line. [Hint: Consider  $\{\hat{S}_u\mathcal{P}: u > 0\}$ .]
- 12.5.2 *Generalized stress release model.* Following Last (2004), consider a piecewise linear Markov process  $X(t)$  determined by the following components: a positive constant  $\nu$  representing the linear increase of  $X(t)$  between jumps; a locally bounded risk function  $\Psi(x): \mathbb{R} \mapsto \mathbb{R}_+$  representing the instantaneous rate of occurrence of jumps given that  $X(t-) = x$ ; and a stochastic kernel  $J(x, B)$  representing the probability of a jump into  $B \in \mathcal{B}_{\mathbb{R}_+}$  given  $X(t) = x \in \mathbb{R}$ .
- Show that, as for the BMAP processes, the process  $X(t)$  uniquely determines the associated MPP, say  $N_X \equiv \{(t_n, \kappa_n)\}$  of jump times and jump sizes, and that conversely knowledge of the MPP on  $\mathbb{R}_+$  determines  $X(t)$  uniquely up to the initial value  $X(0)$ .
  - Show also that if  $X(t)$  is positive recurrent, in the sense that, if  $X(t)$  has distribution  $\Pi_t$  on  $\mathbb{R}_+$ , there exists a stationary distribution  $\Pi^*$  such that  $\|\Pi_t - \Pi^*\| \rightarrow 0$  as  $t \rightarrow \infty$ , then  $N_X$  is strongly asymptotically stationary.
  - Show that if  $X(t)$  is ‘geometrically ergodic’ in the strong sense that  $\|\Pi_t - \Pi^*\| \leq C \exp(-\beta t)$  for positive constants  $C, \beta$ , then also  $N_X$  is geometrically asymptotically stationary in the sense that in Definition 12.5.I,  $\|\mathcal{P}_u - \mathcal{P}^*\| \leq C' \exp(-\beta' t)$  for some constants  $C', \beta'$ .
- [Hint: Use the norm-preserving property of the stochastic kernel, as in establishing (12.5.5).]
- 12.5.3 (*Continuation*). *Application to reliability models.* For each of the examples listed below, identify the form of the components, and find sufficient conditions to ensure that the resulting process  $X(t)$  is (i) well-defined, (ii) positive recurrent, and (iii) geometrically ergodic.
- The *simple stress release model* of Example 7.2(g) and Exercises 7.2.9–10.
  - The *repairable system model* [e.g., Block *et al.* (1985), Kijima (1989), and Last and Szekli (1998)], in which  $X(t)$  denotes the ‘virtual age’ of a system subject to failure, and after every failure an instantaneous repair takes place, restoring the system to some fraction  $0 \leq \theta < 1$  of its ‘age’ before failure.
  - The *workload-dependent queueing process* [e.g., Browne and Sigman (1992), Meyn and Tweedie (1993b)], in which  $W(t) = \max\{-X(t), 0\}$  decreases linearly between jumps until either another jump (upwards) occurs, or  $W(t) = 0$  in which case it remains zero until the next jump occurs.  $W(t)$  here can be interpreted as the workload in a queueing system in which both the arrival and the service rates may depend on the current value of  $W(t)$ .
- [Hint: The basic aim is to find variations on condition (7.1.3) that will allow Foster–Lyapunov drift conditions, as developed in Meyn and Tweedie, to be applied to the process in question. Once convergence of the Markov process has been established, Proposition 12.5.III can be invoked to transfer the results to the associated point process. Last (2004) gives a very general discussion, as well as applications to the specific examples mentioned.]
- 12.5.4 Investigate conditions under which asymptotic stationarity of the point process implies, in an appropriate sense, asymptotic convergence of its first,

second, and higher moment measures to the moment measures of the stationary version of the process.

- 12.5.5 Consider the Hawkes process of Example 12.5(c). Show that, for the process started from zero as in (12.5.7), there is an instantaneous mean rate  $m(t) = E[\lambda_0(t)]$  which is always finite and satisfies  $m(t) \leq m = \nu/(1 - \rho)$ . Imitate the arguments leading to (12.5.11) to show that the sequence  $\hat{S}_s \mathcal{P}_0$  defines a strong Cauchy sequence of point processes on  $[0, S]$ , and so implies the existence of a limit point process on  $[0, S]$ , which in turn, because  $S$  is arbitrary, implies the existence of the distribution  $\mathcal{P}^\dagger$  of an equilibrium process  $N^\dagger$  on  $\mathbb{R}_+$  such that  $\hat{S}_s \mathcal{P}_0 \rightarrow \mathcal{P}^\dagger$  weakly as  $s \rightarrow \infty$ .
- 12.5.6 Let  $N^J$  be an initial realization on  $\mathbb{R}_-$  satisfying (12.5.14), and suppose that (12.5.12) holds. Show that the process started from  $N^J$  is strongly asymptotically stationary. [Hint: Extend the arguments below (12.5.12) to show that for both  $\hat{S}_s N^J$  and  $\hat{S}_s N^\dagger$ , the probability that the residual components on  $\mathbb{R}_+$  are nonempty converge to zero, so that both couple to  $\hat{S}_s N_0$ .]
- 12.5.7 *Convergence to equilibrium of Poisson cluster processes.*
- Consider a Poisson cluster process as in Proposition 6.3.III with constant intensity  $\mu_c$  for the cluster centres, and stationary cluster structure such that the time  $T$  from the first to the last member of the cluster satisfies  $E(T) < \infty$ . Imitate the last part of the discussion in Example 12.5(c) to show that the process is strongly asymptotically stationary.
  - Show that the conditions are satisfied for the Neyman–Scott process whenever both the number  $N$  of points and the distance  $X$  of a satellite point from the cluster centre have finite (absolute) first moments. Similarly for the Bartlett–Lewis model show that the conditions are satisfied whenever the number of points  $N$  and the distance  $X$  between successive cluster points both have finite means.
- 12.5.8 (*Continuation*). Investigate conditions under which the strong asymptotic stationarity for Poisson cluster processes can be strengthened to geometric asymptotic stationarity as in Exercise 12.5.2(c).

## 12.6. Moment Stationarity and Higher-order Ergodic Theorems

The essential simplification of the moment structure implied by stationarity derives from the application of Lemma A2.7.II as in the diagonal shifts Lemma 12.1.IV. It amounts to a diagonal factorization: each moment measure is represented as a product of Lebesgue measure along the main diagonal and a *reduced measure* in a complementary subspace. These reduced measures determine the moment structure of the process and have long been studied, usually as densities, in applications. They appear in many different guises, notably as the moment measures of the Palm distributions in Chapter 13, and in the higher-order ergodic theorems discussed at the end of this section and

again in Section 13.4. Their Fourier transforms provide the various spectra of the random measure, the second-order version (Bartlett spectrum) being discussed in detail in Chapter 8. The role of the factorization theorem in the present context emerged in the work of Brillinger (1972) and Vere-Jones (1971), being subsumed by the more general results on the disintegration of moment measures developed in Krickeberg (1974b).

We start with some definitions, supposing again that  $\mathcal{X} = \mathbb{R}^d$ , and writing  $T_x$  for the shift operator as in (12.1.1). Although we include MPPs in the exposition below, for which stationarity was noted below Definition 12.1.II, the resulting expressions are more involved and are left largely as exercises.

**Definition 12.6.I.** (a) *A random measure or point process on  $\mathcal{X} = \mathbb{R}^d$  is  $k$ th-order stationary if its  $k$ th moment measure exists, and for each  $j = 1, \dots, k$ , bounded Borel sets  $A_1, \dots, A_j$ , and  $x \in \mathbb{R}^d$ ,*

$$M_j(T_x A_1 \times \cdots \times T_x A_j) = M_j(A_1 \times \cdots \times A_j). \quad (12.6.1)$$

(b) *An MPP on  $\mathbb{R}^d \times \mathcal{K}$  is  $k$ th-order stationary if its  $k$ th moment measure exists, and for each  $j = 1, \dots, k$ , bounded Borel sets  $A_1, \dots, A_k$  and  $K_1, \dots, K_k \in \mathcal{B}_{\mathcal{K}}$ ,  $M_j(T_x A_1 \times K_1 \times \cdots \times T_x A_j \times K_j)$  is independent of  $x \in \mathbb{R}^d$ .*

If  $\xi$  is a stationary random measure, the joint distributions of  $\{\xi(T_x A_1), \dots, \xi(T_x A_j)\}$  coincide with those of  $\{\xi(A_1), \dots, \xi(A_j)\}$  (see around Definition 12.1.II), so that a stationary random measure for which the  $k$ th-order moment measure exists is  $k$ th-order stationary. The converse implication is not true in general (see Exercise 8.1.1), but in particular parametric models, moment stationarity, even of relatively low order, generally requires stationarity of the process as a whole. For example, a Poisson process is stationary if and only if it is first-order stationary.

The imposition of conditions on  $M_j$  for  $j < k$  in Definition 12.6.I is certainly redundant in the case of a simple point process, for the lower-order moment measures appear as diagonal concentrations (see Proposition 9.5.II) and are thereby identified uniquely (see Exercise 9.5.7). It may be redundant more generally, but the question appears to be open. It is relatively easy, however, to find a process for which the second cumulant measure is stationary but the expectation measure is not (see Exercise 8.1.2).

The case  $k = 1$  of the condition (12.6.1) simply asserts that the expectation measure  $M_1(\cdot)$  is invariant under shifts. It must therefore reduce to a multiple of the unique measure on  $\mathbb{R}^d$  with this property, namely, Lebesgue measure. We thus have the following proposition that incorporates parts of Propositions 6.1.I, 8.1.I and 8.3.II.

**Proposition 12.6.II.** (a) *A random measure on  $\mathbb{R}^d$  is first-order stationary if and only if its expectation measure is a finite positive multiple  $m$  (the mean density) of Lebesgue measure  $\ell$  on  $\mathbb{R}^d$ .*

(b) *A marked random measure or MPP on  $\mathbb{R}^d \times \mathcal{K}$  is first-order stationary if and only if its expectation measure is a product  $\ell \times F$  of Lebesgue measure*

on  $\mathbb{R}^d$  and a boundedly finite measure  $F$  on  $\mathcal{B}_{\mathcal{K}}$ ;  $F$  is totally finite if and only if the expectation measure of the ground process has finite mean density  $m_g$ , and  $m_g = F(\mathcal{K})$ .

The proportionality constant referred to in this proposition is usually denoted  $m$  and called either the *mean density* or the *rate* of the process.

For  $k > 1$ , the conditions at (12.6.1) imply, via the generating properties of rectangle sets, that the whole measure  $M_k$  is invariant under the group of diagonal shifts  $D_x^{(k)}$  defined by (12.1.11). The diagonal shifts Lemma 12.1.IV now implies the following proposition for the unmarked case. The corresponding results for the marked case are outlined in Exercise 12.6.9.

**Proposition 12.6.III.** *For any  $k$ th-order stationary random measure or point process on  $\mathbb{R}^d$ , there exist reduced measures  $\check{M}_k, \check{M}_{[k]}, \check{C}_k$ , and  $\check{C}_{[k]}$  related to the corresponding  $k$ th-order measures  $M_k, M_{[k]}, C_k$ , and  $C_{[k]}$  through equations, valid for any function  $f \in \text{BM}(\mathcal{X}^{(k)})$ , of the type*

$$\begin{aligned} & \int_{\mathcal{X}^{(k)}} f(x_1, \dots, x_k) M_k(dx_1 \times \dots \times dx_k) \\ &= \int_{\mathcal{X}} dx \int_{\mathcal{X}^{(k-1)}} f(x, x + y_1, \dots, x + y_{k-1}) \check{M}_k(dy_1 \times \dots \times dy_{k-1}). \end{aligned} \quad (12.6.2)$$

We call  $\check{M}_k, \check{M}_{[k]}, \check{C}_k$ , and  $\check{C}_{[k]}$  the *reduced  $k$ th-order moment measure*, the *reduced  $k$ th-order factorial moment measure*, the *reduced  $k$ th-order cumulant measure*, and the *reduced  $k$ th-order factorial cumulant measure*, respectively; see Proposition 13.2.VI for their interpretation as moment measures of the Palm distribution. For  $k = 1$  these reduced measures all coincide and equal the mean density  $m = M_1(\mathbb{U}^d)$ . For  $k = 2$  we mostly use  $\check{C}_2$ , which we also call the *reduced covariance measure*. It is defined on  $\mathcal{B}_{\mathcal{X}}$ , and its properties and applications form the main content of Chapter 8.

Note that the disintegration furnished by (12.6.2) is of the form

$$M_k = \ell \times \check{M}_k, \quad (12.6.3)$$

where  $\ell$  denotes standard Lebesgue measure on  $\mathbb{R}^d$  (so,  $\ell(\mathbb{U}^d) = 1$ ) and thus *any scale factors remain in the reduced measure*. The same disintegration result can also be obtained via an argument involving Radon–Nikodym derivatives with respect to the first-moment measure, as in Exercises 12.1.8–9. This alternative approach is outlined in Exercises 12.6.1–2 and leads to a decomposition of the form

$$M_k = M_1 \times (m^{-1} \check{M}_k)$$

with  $M_1 = m\ell$ . This and its role in Palm theory has led some authors to adopt  $m^{-1} \check{M}_k$  as the definition of the reduced measure; we have preferred not to adopt this convention, mainly because of its incompatibility with the usual definition of the stationary form of the covariance function when the measure is absolutely continuous.

Some of the more accessible properties of the reduced moment measures are given in the next proposition; analogous statements for factorial moment and cumulant measures can be given (see Exercise 12.6.3). A more extended list of properties for the case  $k = 2$  is given in Proposition 8.1.II for the unmarked point processes and in Proposition 8.3.II for MPPs.

**Proposition 12.6.IV.** *Let  $\check{M}_k$  be the  $k$ th-order reduced moment measure for the  $k$ th-order stationary random measure  $\xi$  on  $\mathcal{X} = \mathbb{R}^d$ .*

- (i)  $\check{M}_k(\cdot)$  is a symmetric measure on  $\mathcal{X}^{(k-1)}$  (invariant under permutations of the arguments in the product space) and is invariant also under the ‘shift reflection’ transformation mapping  $(u_1, u_2, \dots, u_{k-1})$  into  $(-u_1, u_2 - u_1, \dots, u_{k-1} - u_1)$ .
- (ii) When  $M_k$  is absolutely continuous with density  $m_k$ ,  $\check{M}_k$  is also absolutely continuous, and its density  $\check{m}_k$  is related to  $m_k$  by

$$m_k(x_1, x_2, \dots, x_k) = \check{m}_k(x_2 - x_1, \dots, x_k - x_1). \quad (12.6.4)$$

- (iii) For all bounded Borel sets  $A_1, \dots, A_{k-1} \in \mathcal{X}$ ,

$$\check{M}_k(A_1 \times \dots \times A_{k-1}) = \mathbb{E} \left( \int_{\mathbb{U}^d} \xi(x + A_1) \dots \xi(x + A_{k-1}) \xi(dx) \right). \quad (12.6.5)$$

PROOF. In (12.6.2) set

$$f(x_1, \dots, x_k) = g(x_1)h(x_2 - x_1, \dots, x_k - x_1),$$

where  $g(\cdot)$  and  $h(\cdot)$  are bounded Borel functions of bounded support on  $\mathcal{X}$  and  $\mathcal{X}^{(k-1)}$ , respectively, so that

$$\begin{aligned} & \int_{\mathcal{X}^{(k)}} g(x_1)h(x_2 - x_1, \dots, x_k - x_1) M_k(dx_1 \times \dots \times dx_k) \\ &= \int_{\mathcal{X}} g(x) dx \int_{\mathcal{X}^{(k-1)}} h(y_1, \dots, y_{k-1}) \check{M}_k(dy_1 \times \dots \times dy_{k-1}). \end{aligned} \quad (12.6.6)$$

Now let the variables  $x_2, \dots, x_k$  in the argument of  $h(\cdot)$  on the left-hand side be permuted. Because of the symmetry properties of  $M_k(\cdot)$ , this leaves the integral unaltered. Observe also that it corresponds to permuting the variables  $y_1, \dots, y_{k-1}$  in the argument of  $h(\cdot)$  on the right-hand side of (12.6.6). Equivalently, it corresponds to leaving the variables in  $h(\cdot)$  unaltered and permuting the variables in  $\check{M}_k$ . Because a measure on  $\mathcal{X}^{(k-1)}$  is uniquely determined by the integrals of all such functions  $h(\cdot)$ , it follows that  $\check{M}_k$  must be invariant under permutations of its arguments; that is, it is symmetric.

Alternatively, if we interchange  $x_1$  and  $x_2$  on the left-hand side of (12.6.6), the integral is unaltered, and from (12.6.3) the right-hand side becomes

$$\int_{\mathcal{X}} dx \int_{\mathcal{X}^{(k-1)}} g(x + y_1)h(-y_1, y_2 - y_1, \dots, y_{k-1} - y_1) \check{M}_k(dy_1 \times \dots \times dy_{k-1}).$$



But  $\int_{\mathcal{X}} g(x + y_1) dx = \int_{\mathcal{X}} g(x) dx$ , so we conclude that

$$\begin{aligned} & \int_{\mathcal{X}^{(k-1)}} h(y_1, \dots, y_{k-1}) \check{M}_k(dy_1 \times \dots \times dy_{k-1}) \\ &= \int_{\mathcal{X}^{(k-1)}} h(-y_1, y_2 - y_1, \dots, y_{k-1} - y_1) \check{M}_k(dy_1 \times \dots \times dy_{k-1}), \end{aligned}$$

from which there follows the shift reflection invariance assertion in (i).

If  $M_k$  has density  $m_k$  with respect to Lebesgue measure on  $\mathbb{R}^{dk}$ , invariance of  $M_k$  implies that  $m_k(x, x + y_1, \dots, x + y_{k-1})$  is independent of  $x$ , so on cancelling the factor  $\int_{\mathcal{X}} g(x) dx$  in (12.6.6) we obtain

$$\begin{aligned} & \int_{\mathcal{X}^{(k-1)}} h(y_1, \dots, y_{k-1}) \check{M}_k(dy_1 \times \dots \times dy_{k-1}) \\ &= \int_{\mathcal{X}^{(k-1)}} h(y_1, \dots, y_{k-1}) m_k(x, x + y_1, \dots, x + y_{k-1}) dy_1 \dots dy_{k-1}. \end{aligned}$$

Thus,  $\check{M}_k$  is absolutely continuous with density  $m_k(x, x + y_1, \dots, x + y_{k-1})$ , which is equivalent to the assertion (ii).

Finally, in (12.6.6) set  $g(x) = I_{\mathbb{U}^d}(x)$  and  $h(y_1, \dots, y_{k-1}) = \prod_{j=1}^{k-1} I_{A_j}(y_j)$ . Then because  $\int_{\mathcal{X}} g(x) dx = \ell(\mathbb{U}^d) = 1$ , (12.6.5) follows directly.  $\square$

Although the reduced measures  $\check{M}_k$  are necessarily nonnegative, the same is not true of reduced cumulant measures  $\check{C}_k$ . In the simplest nontrivial case,  $\check{C}_2(A) = \check{M}_2(A) - m^2 \ell(A)$  [see (8.1.6)], so that for its Jordan–Hahn decomposition  $\check{C}_2 = \check{C}_2^+ - \check{C}_2^-$  into positive and negative parts, we have

$$\check{C}_2^-(A) \leq m^2 \ell(A) < \infty \quad (\text{bounded } A \in \mathcal{B}_{\mathcal{X}}).$$

For the simple (stationary) Poisson process,  $\check{M}_2(A) = (m^2 + m)\ell(A)$  so  $\check{C}_2^+(A) = m\ell(A)$ ,  $\check{C}_2^-(A) = 0$ , but for the stationary deterministic process on  $\mathbb{R}$  with span  $a$  as in Example 8.3(e),  $\check{C}_2^+$  consists of atoms of mass  $1/a$  at the points  $ka$  ( $k \in \mathbb{Z}$ ) and  $\check{C}_2^-(A) = \ell(A)/a^2$  for  $A \in \mathcal{B}_{\mathbb{R}}$ . Thus, although this process has  $0 \leq \text{var } N(0, x] \leq \frac{1}{4}$ , neither of  $\check{C}_2^+$  and  $\check{C}_2^-$  is totally finite.

The identification at (12.6.5) of  $\check{M}_k(\cdot)$  as an expectation suggests the existence of higher-order ergodic theorems in which the reduced moment measures appear as the ergodic limits. To identify the limits in the nonergodic situation, we use the following application of Lemma 12.2.III, where  $\mathcal{I}$  again denotes the  $\sigma$ -algebra of invariant events.

**Lemma 12.6.V.** *Let  $\xi$  be a strictly stationary random measure with finite  $k$ th moment measure. Then there exists a symmetric  $\mathcal{I}$ -measurable random measure  $\check{\psi}_k$  on  $\mathcal{X}^{(k-1)}$ , invariant also under the shift reflections of Proposition 12.6.IV, such that for bounded Borel functions  $f$  of bounded support on  $\mathcal{X}^{(k)}$ ,*

$$\begin{aligned} & \mathbb{E} \left( \int_{\mathcal{X}^{(k)}} f(x_1, \dots, x_k) \xi(dx_1) \dots \xi(dx_k) \mid \mathcal{I} \right) \\ &= \int_{\mathcal{X}} dx \int_{\mathcal{X}^{(k-1)}} f(x, x + y_1, \dots, x + y_{k-1}) \check{\psi}_k(dy_1 \times \dots \times dy_{k-1}). \end{aligned} \quad (12.6.7)$$

In particular, for bounded  $A_1, \dots, A_{k-1} \in \mathcal{B}_{\mathcal{X}}$ ,

$$\check{\psi}_k(A_1 \times \dots \times A_{k-1}) = \mathbb{E} \left( \int_{\mathbb{U}^d} \xi(x + A_1) \dots \xi(x + A_{k-1}) \xi(dx) \mid \mathcal{I} \right). \quad (12.6.8)$$

PROOF. Represent  $\mathcal{X}^{(k)}$  in the product form  $\mathcal{X} \times \mathcal{X}^{(k-1)}$  via the mapping (12.1.13). On the space  $\mathcal{X}^{(k)}$ ,  $\xi$  induces a new random measure, namely, the  $k$ -fold product  $\xi^{(k)}$  of  $\xi$  with itself, and  $\xi^{(k)}$  is stationary with respect to diagonal shifts. Its image under (12.1.13) is therefore stationary with respect to shifts in the first component. We now have a situation to which the general result in Lemma 12.2.III applies, with  $\mathcal{X}^{(k-1)}$  playing the role of the mark space  $\mathcal{K}$ . On the product space  $\mathcal{X} \times \mathcal{X}^{(k-1)}$  there exists a  $\sigma$ -algebra of sets invariant under shifts in the first component, and the image of  $\xi^{(k)}$  under (12.1.13) has a conditional expectation with respect to this  $\sigma$ -algebra, which factorizes into a product of Lebesgue measure on  $\mathcal{X}$  and an  $\mathcal{I}$ -measurable random measure  $\check{\psi}_k$  on  $\mathcal{X}^{(k-1)}$ , which is readily checked as having the properties described in the lemma.  $\square$

Before proceeding to the ergodic theorems, we give a further example, albeit somewhat artificial in character, to illustrate some of the types of behaviour that can occur in the nonergodic case.

EXAMPLE 12.6(a) *Poisson cluster process with dependent clusters.* Suppose that  $\mathcal{X} = \mathbb{R}$  and that cluster centres occur at rate  $\lambda$ . Set up a common pattern for the clusters from a fixed realization  $\{y_1, \dots, y_Z\}$  of a finite Poisson process on  $\mathbb{R}$  with a nonatomic parameter measure  $\mu(\cdot)$ , so that  $Z$  is a Poisson r.v. with mean  $\mathbb{E}(Z) = \mu(\mathbb{R}) = \nu$  and, conditional on  $Z$ , the r.v.s  $y_1, \dots, y_Z$  are i.i.d. r.v.s with distribution  $\mu(\cdot)/\nu$ . Then, given a realization  $\{x_i\}$  of the cluster centre process, we associate with the cluster centre  $x_i$  the cluster  $(x_i + y_j; j = 1, \dots, Z)$ , so that the whole process has as its realization the points  $\{x_i + y_j; i = 0, \pm 1, \dots; j = 1, \dots, Z\}$ . The r.v.s  $\{Z, y_1, \dots, y_Z\}$  define a  $\sigma$ -algebra of events, which, in fact, coincides with the invariant  $\sigma$ -algebra  $\mathcal{I}$  for the whole process. We can then compute moment characteristics of the process as follows.

(1°) For  $k = 1$ , the mean density given the invariant  $\sigma$ -algebra  $\mathcal{I}$ , that is, the r.v.  $Y$  of (12.2.12), here equals  $\lambda Z$ ; the mean density of the whole process equals

$$m = \mathbb{E}(\lambda Z) = \lambda \mu(\mathbb{R}) = \lambda \nu.$$

(2°) For  $k = 2$  and before reduction, the second-order moment measure given  $\mathcal{I}$  has three components: a multiple  $\lambda^2 Z^2$  of Lebesgue measure in the plane; a line concentration with density  $\lambda Z$  along the main diagonal  $x = y$ ; and line concentrations of density  $\lambda$  along each of the  $Z(Z - 1)$  lines  $y = x + y_i - y_j$ , where  $i \neq j$  but both orderings are permitted. Then the reduced moment measure  $\check{\psi}_2(\cdot)$  on  $\mathcal{B}_{\mathbb{R}}$  can be written

$$\check{\psi}_2(du) = \lambda^2 Z^2 du + \lambda Z \delta(u) du + \lambda \sum_{i \neq j} \delta(u - y_i + y_j) du,$$

where the  $\delta$ -function terms represent atoms at 0 and the points  $\pm|y_i - y_j|$  ( $i \neq j$ ). Taking expectations leads to the reduced second moment measure:

$$\check{M}_2(du) = \lambda^2 \nu(\nu + 1) du + \lambda \nu \delta(u) du + \lambda \int_{\mathbb{R}} \mu(x + du) \mu(dx).$$

(3°) Third- and higher-order moments can be built up in a similar way by considering all possible locations of triples of points  $\{y_i\}$  and so on.

Observe that  $\check{\psi}_2$  is just the form that the reduced moment measure would take if the cluster structure were fixed for all realizations: the process would then be infinitely divisible and ergodic.

A variant of this Poisson cluster process, but having conditionally independent clusters, can be obtained by treating the clusters as a Cox process directed by some a.s. finite random measure  $\xi$  replacing the fixed measure  $\mu(\cdot)$  above: regard  $\xi$  as fixed for any given realization with the points in each cluster now being determined mutually independently according to a Poisson process with parameter measure  $\xi(\cdot)$ . Then  $\check{M}_2$  as above equals the random measure  $\check{\psi}_2$  of this process and the new  $\check{M}_2$  is obtained by a further averaging over the realizations of  $\xi$ . Further variants of the model are possible.  $\square$

We are now in a position to state the higher-order version of the ergodic Theorem 12.2.IV [see also Nguyen and Zessin (1976)]. Extensions to the marked case of Lemma 12.6.V and the result below are outlined in Exercise 12.6.10. For further extensions see Sections 13.4–5, especially Propositions 13.4.I and 13.4.III.

**Theorem 12.6.VI.** *Let  $\xi$  be a strictly stationary random measure for which the  $k$ th moment measure exists,  $\check{\psi}_k$  the invariant random measure defined by (12.6.8), and  $B_1, \dots, B_{k-1}$  a family of bounded Borel sets in  $\mathbb{R}^d$ . Then, for any convex averaging sequence  $\{A_n\}$  in  $\mathbb{R}^d$ , as  $n \rightarrow \infty$ ,*

$$\frac{1}{\ell(A_n)} \int_{A_n} \xi(x + B_1) \dots \xi(x + B_{k-1}) \xi(dx) \xrightarrow{\text{a.s.}} \check{\psi}_k(B_1 \times \dots \times B_{k-1}). \quad (12.6.9)$$

*In particular, if  $\xi$  is ergodic,*

$$\frac{1}{\ell(A_n)} \int_{A_n} \xi(x + B_1) \dots \xi(x + B_{k-1}) \xi(dx) \xrightarrow{\text{a.s.}} \check{M}_k(B_1 \times \dots \times B_{k-1}). \quad (12.6.10)$$

**PROOF.** Given a bounded Borel function  $g$  of bounded support on  $\mathcal{X}^{(k)}$ , consider the random function

$$f(\xi) \equiv \int_{\mathcal{X}^{(k)}} g(x_1, \dots, x_k) \xi(dx_1) \dots \xi(dx_k),$$

noting that, by assumption, we have  $E[f(\xi)] < \infty$ .

Appealing to Proposition 12.2.II(a) and evaluating the limit in (12.2.6) from (12.6.7), we obtain, for  $n \rightarrow \infty$ ,

$$\begin{aligned} & \frac{1}{\ell(A_n)} \int_{A_n} f(S_x \xi) \, dx \\ &= \frac{1}{\ell(A_n)} \int_{A_n} dx \int_{\mathcal{X}^{(k)}} g(u_1 - x, \dots, u_k - x) \xi(du_1) \dots \xi(du_k) \\ &\xrightarrow{\text{a.s.}} \int_{\mathcal{X}} dx \int_{\mathcal{X}^{(k-1)}} g(x, x + y_1, \dots, x + y_{k-1}) \check{\psi}_k(dy_1 \times \dots \times dy_{k-1}). \end{aligned} \quad (12.6.11)$$

In particular, by taking  $g(x_1, \dots, x_k) = g_\varepsilon(x_1)h(x_2 - x_1, \dots, x_k - x_1)$ , where  $g_\varepsilon(\cdot)$  has the same properties as in the proof of Theorem 12.2.IV, it follows, for example, that  $\int_{A_n^\varepsilon} f(S_x \xi) \, dx$  equals

$$\begin{aligned} & \int_{A_n^\varepsilon} g_\varepsilon(u_1 - x) \, dx \int_{\mathcal{X}^{(k)}} h(u_2 - u_1, \dots, u_k - u_1) \xi(du_1) \dots \xi(du_k) \\ &= \int_{A_n^\varepsilon} g_\varepsilon(u_1 - x) \, dx \int_{\mathcal{X}^{(k)}} h(v_1, \dots, v_{k-1}) \xi(du_1) \xi(u_1 + dv_1) \dots \xi(u_1 + dv_{k-1}) \\ &\geq \int_{A_n} \xi(du_1) \int_{\mathcal{X}^{(k-1)}} h(v_1, \dots, v_{k-1}) \xi(u_1 + dv_1) \dots \xi(u_1 + dv_{k-1}). \end{aligned}$$

Thus, we can use the approximation argument exploited in Theorem 12.2.IV to deduce that, for nonnegative bounded functions  $h$  of bounded support in  $\mathcal{X}^{(k-1)}$ , as  $n \rightarrow \infty$ ,

$$\begin{aligned} & \frac{1}{\ell(A_n)} \int_{A_n} \xi(du) \int_{\mathcal{X}^{(k-1)}} h(v_1, \dots, v_{k-1}) \xi(u + dv_1) \dots \xi(u + dv_{k-1}) \\ &\xrightarrow{\text{a.s.}} \int_{\mathcal{X}^{(k-1)}} h(v_1, \dots, v_{k-1}) \check{\psi}_k(dv_1 \times \dots \times dv_{k-1}). \end{aligned} \quad (12.6.12)$$

Equations (12.6.9) and (12.6.10) are now easily derived as special cases of (12.6.12).  $\square$

It is, of course, a corollary of (12.6.8) that

$$\mathbb{E}(\check{\psi}_k(B_1 \times \dots \times B_{k-1})) = \check{M}_k(B_1 \times \dots \times B_{k-1}).$$

An  $L_1$  version of Theorem 12.6.VI is given in Exercise 12.6.8.

For point processes, the left-hand side of (12.6.10) suggests a natural class of nonparametric estimates for the reduced moment measures, as for example the estimate

$$\widehat{M}_{[2]}(B; A) = \frac{1}{\ell(A)} \sum_{i: x_i \in A} N^*(x_i + B), \quad (12.6.13)$$

where  $N^*(B) = N(B) - \delta_0(B)$ , introduced at (8.1.25) in discussing second-order factorial moment measures. In practice, estimates of this kind are

subject to serious biases arising from edge effects, as discussed briefly around (8.1.26) and in greater detail in texts such as SKM (1995). In such contexts, Theorem 12.6.VI provides a starting point for proving the consistency of the estimates or of variants in which  $\ell(A)$  is replaced by  $N(A)$ , itself representing an estimate of  $M_1(A) = m\ell(A)$ . The resulting quantity can be written

$$\widehat{M}_{[2]}(B; A) \approx \frac{m}{N(A)} \sum_{i=1}^{N(A)} N^*(x_i + B) \quad (12.6.14)$$

and represents the average, over the points of  $A$ , of the counts of points in sets  $B$  relative to the points of  $A$  as origin. As such, it is a point estimate of the first-order moment measure of the Palm process associated with  $N$ , as discussed further in Section 13.4. Again, Theorem 12.6.VI provides the basis for proving the consistency of such estimates, and is so used in the discussion of fractal dimension in Section 13.6.

Often, the natural interpretation of the estimates such as in (12.6.13) is in terms of point configurations. For example, in the case  $k = 3$ , the third-order factorial moment measure gives information about the occurrence of triplets of points of the realization, taking one point of the triplet as origin. In the discussion of Section 13.6, for example, use is made of sets

$$B_{k-1,r} = \{(u_1, \dots, u_{k-1}) : \max \|u_j\| < r\},$$

which for  $k = 3$  gives information about the proportion of triplets in the realization with the property that all three points of the triplet lie within a maximum distance  $r$  of one of the three points. Kagan (see Exercise 12.6.11) has used estimates of two-, three- and four-point configurations at ‘small’ scale in examining possible relations between shocks and aftershocks in earthquake studies.

As for finite processes in Chapter 5, moment densities are often used as an aid to understanding both qualitative and quantitative behaviour of models, as we illustrate in our concluding example.

**EXAMPLE 12.6(b) *Interacting point processes.*** Suppose given two stationary simple point processes  $N_j$  ( $j = 0, 1$ ) with mean densities  $m_j$ . They evolve independently except that each successive point  $t_i$  of the process  $N_1$  is followed by a *dead time*  $Z_i$ , during which any point of the process  $N_0$  is deleted. Suppose that  $Z_i = \min(Y_i, t_{i+1} - t_i)$ , where  $\{Y_i\}$  is a sequence of i.i.d. nonnegative r.v.s independent of both  $N_0$  and  $N_1$ . We observe  $N_1$  and the thinned process  $N_2$  consisting of those points of  $N_0$  that are not deleted. Our aim is to describe the first and second moment measures of the output  $(N_1, N_2)$ , particularly as they relate to the same measures for  $N_0$ .

To this end it is convenient to use the  $\{0, 1\}$ -valued process  $J(t)$  for which

$$J(t) = \begin{cases} 0 & \text{if } 0 < t - t_i \leq Z_i \text{ for some } i, \\ 1 & \text{otherwise (so, } t \in \bigcup_i (t_i + Z_i, t_{i+1}]). \end{cases}$$

Because the marked point process  $\{t_i, (Y_i, t_{i+1} - t_i)\}$  on  $\mathbb{R} \times \mathbb{R}_+^2$  is stationary whenever  $N_1$  (with realizations the points  $\{t_i\}$ ) is stationary, it follows that then  $J(\cdot)$  is itself a stationary process, with first and second moments

$$\alpha = \mathbb{E}J(t) \quad (\text{all } t), \quad \beta(u) = \mathbb{E}(J(t)J(t+u)) \quad (\text{all } t, u).$$

Furthermore,  $J$  is determined by  $N_1$  and  $\{Y_i\}$ , and is thus independent of  $N_0$ . Consequently from  $N_2(dx) = J(x) N_0(dx)$  it follows that  $N_2$  has mean density  $m_2$  given by

$$m_2 = \mathbb{E}N_2(0, 1] = \mathbb{E}\left[\int_0^1 J(x) N_2(dx)\right] = \mathbb{E}\left[\int_0^1 \mathbb{E}[J(x)] N_2(dx)\right] = \alpha m_0.$$

In addition, for bounded measurable  $h$  of bounded support in  $\mathbb{R}^2$ , and writing  $D = \{(x, y): x = y\}$  for the diagonal of  $\mathbb{R}^2$ , the second factorial measure  $M_{[2]}^{(2)}(\cdot)$  of  $N_2$  satisfies

$$\begin{aligned} \int_{\mathbb{R}^2 \setminus D} h(x, y) M_{[2]}^{(2)}(dx \times dy) &= \mathbb{E}\left[\int_{\mathbb{R}^2 \setminus D} h(x, y) N_2(dx) N_2(dy)\right] \\ &= \mathbb{E}\left[\int_{\mathbb{R}^2 \setminus D} h(x, y) J(x) J(y) N_0(dx) N_0(dy)\right]. \end{aligned}$$

Thus, when  $N_0$  has a reduced factorial moment density  $\check{m}_{[2]}(\cdot)$ , we have

$$\begin{aligned} \mathbb{E}\left[\int_{\mathbb{R}^2 \setminus D} h(x, y) J(x) J(y) \check{m}_{[2]}(x - y) dx dy\right] \\ = \int_{\mathbb{R}^2 \setminus D} h(x, y) \beta(x - y) \check{m}_{[2]}(x - y) dx dy, \end{aligned}$$

and  $N_2$  has a reduced factorial moment measure which likewise has a density  $\check{m}_{[2]}^{(2)}(\cdot)$ ; it is given by

$$\check{m}_{[2]}^{(2)}(u) = \beta(u) \check{m}_{[2]}(u).$$

Finally, for the cross-intensity we find similarly (using differential notation for brevity) that

$$\mathbb{E}(N_1(dx) N_2(dy)) = m_0 \gamma(y - x) dx dy,$$

where  $\gamma(u) dx = \mathbb{E}[J(x+u) N_1(dx)]$ . Here,  $\gamma(u)$  can be interpreted as the rate of occurrence of points  $t_i$  of  $N_1$  such that  $t_i + u$  lies outside any dead-time interval.

Any more detailed evaluation of the quantities  $\alpha$ ,  $\beta(\cdot)$ , and  $\gamma(\cdot)$  of the process  $N_1$  requires in turn more specific detail about its structure. Ergodicity of  $N_1$  is enough to show via the ergodic theorem that

$$\alpha = \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n (X_i - Z_i)}{\sum_{i=1}^n X_i} = m_1 \mathbb{E}[(X_i - Y_i)_+] = m_1 \int_0^\infty G(v) (1 - F(v)) dv,$$

where  $G$  is the common d.f. of the  $Y_i$ , and  $X_i = t_{i+1} - t_i$  with common d.f.  $F$  given by

$$F(x) = \lim_{h \downarrow 0} \Pr\{N_1(0, x] \geq 1 \mid N_1(-h, 0] \geq 1\}$$

(see Chapters 3 and 13). The functions  $\beta$  and  $\gamma$  are necessarily more complex, involving joint distributions, as we now illustrate in the case that  $N_1$  is a stationary renewal process with lifetime d.f.  $F$  as just given. Writing

$$\pi(t) = \Pr\{J(t) = 0 \mid N_1 \text{ has a point at the origin}\}$$

for the conditional dead-time probability function, the regenerative properties of  $N_1$  imply that  $\pi(\cdot)$  satisfies the renewal equation

$$\begin{aligned} \pi(t) &= (1 - G(t))(1 - F(t)) + \int_0^t \pi(t - v) dF(v) \\ &= \int_0^t (1 - G(t - v))(1 - F(t - v)) dU(v), \end{aligned}$$

where  $U(\cdot)$  is the renewal function generated by the d.f.  $F$  [see (4.1.7)]. When  $F$  is such that the nonlattice form of the renewal Theorem 4.4.I holds, (4.4.2) yields

$$\pi(t) \rightarrow \int_0^\infty (1 - G(v))(1 - F(v)) m_1 dv \quad (t \rightarrow \infty).$$

Thus, writing  $B$  for a stationary backward recurrence time r.v. (see Section 4.2), we also have

$$\lim_{t \rightarrow \infty} \pi(t) = \Pr\{Z > B\} = 1 - \alpha.$$

For general  $t$ , we have  $\gamma(t) = m_1[1 - \pi(t)]$ , so it remains only to identify  $\beta(\cdot)$ . When  $N_1$  is stationary, write  $B_t$  and  $T_t$  for the backward and forward recurrence time r.v.s at time  $t$ , noting that stationarity implies from (4.2.7) and (4.4.2) that their joint distribution has

$$\Pr\{B_t > x, T_t > y\} = m_1 \int_{x+y}^\infty (1 - F(v)) dv.$$

For  $u > 0$  we have

$$\begin{aligned} \beta(u) &= E(J(t)J(t+u)) \\ &= \Pr\{B_t > Z, T_t > u\} + \int_0^u (1 - \pi(u - v)) \Pr\{B_t > Z, T_t \in (v, v + dv)\} \\ &= \alpha - m_1 \int_0^\infty dG(z) \int_0^u \pi(u - v)(1 - F(z + v)) dv, \\ &\rightarrow \alpha^2 \quad (u \rightarrow \infty) \quad \text{when } \pi(t) \rightarrow 1 - \alpha \quad (t \rightarrow \infty). \end{aligned}$$

The properties of  $\beta(u)$  and  $\gamma(u)$  noted for large  $u$  reflect asymptotic independence, but it is primarily the local properties, for  $u$  closer to zero, that are of interest in practice, because it is these that reflect any causal relation between the two processes  $N_0$  and  $N_1$ . Parametric models are then called for, but take us further from the theme of this section so we refer, for example, to Lawrance (1971) or Sampath and Srinivasan (1977) for specific details (see also Exercises 12.6.12–13).  $\square$

## Exercises and Complements to Section 12.6

12.6.1 For a random measure (not necessarily stationary) on the c.s.m.s.  $\mathcal{X}$  for which the  $k$ th moment measure  $M_k(\cdot)$  ( $k \geq 2$ ) is boundedly finite, so  $M(\cdot) \equiv M_1(\cdot)$  exists, establish the existence of a Radon–Nikodym family  $\check{M}_k(B \mid x)$  with the properties below, including the ‘disintegration’ property in (c).

- (a) For each bounded  $B \in \mathcal{B}(\mathcal{X}^{(k-1)})$ ,  $\check{M}_k(B \mid x)$  is a measurable function of  $x$  that is integrable with respect to  $M(\cdot)$  on bounded sets.
- (b) For  $M$ -almost all  $x$ ,  $\check{M}_k(\cdot \mid x)$  is a boundedly finite measure on  $\mathcal{B}(\mathcal{X}^{(k-1)})$ .
- (c) For  $B \in \mathcal{B}(\mathcal{X}^{(k-1)})$ , use the fact that  $M_k(\cdot \times B) \ll M(\cdot)$  to conclude that

$$M_k(A \times B) = \int_A \check{M}_k(B \mid x) M(dx) \quad [A \in \mathcal{B}(\mathcal{X})].$$

12.6.2 (*Continuation*). Arguing as in Exercises (12.1.8–9), deduce that when  $\mathcal{X} = \mathbb{R}^d$  and the process is  $k$ th-order stationary, there exists a version of  $\check{M}_k(\cdot \mid x)$  that is invariant under simultaneous translations; that is,  $\check{M}_k(D_y^{(k-1)}B \mid x+y) = \check{M}_k(B \mid x)$ . Hence, give an alternative proof of Proposition 12.6.III.

12.6.3 Give analogous statements to those of Proposition 12.6.IV for the reduced factorial moment measure and for the reduced ordinary and factorial cumulant measures. Investigate the analogue of (12.6.6) when  $M_k$  is replaced by  $M_{[k]}$  and  $g$  and  $h$  are indicator functions as in the proof of (12.6.5). Relate the case  $k = 2$  to the ergodicity result underlying (12.6.13).

12.6.4 Find the reduced moment and cumulant measures for a stationary renewal process. In particular, show that if the renewal function has a density  $h(\cdot)$ , then reduced  $k$ th factorial moment measures exist for all  $k = 2, 3, \dots$  and have densities

$$\check{m}_{[k]}(x_1, \dots, x_{k-1}) = \lambda h(x'_1) h(x'_2 - x'_1) \cdots h(x'_{k-1} - x'_{k-2}),$$

where  $\{x'_1, \dots, x'_{k-1}\}$  is the set  $\{x_1, \dots, x_{k-1}\}$  arranged in ascending order. [Hint: See Example 5.4(b) and Exercise 7.2.3.]

12.6.5 (a) Show that when the reduced  $k$ th factorial cumulant measure of a  $k$ th-order stationary point process is totally finite, the  $k$ th cumulant of  $N(A)$  is asymptotically proportional to  $\ell(A)$  as  $A \uparrow \mathcal{X}$  through a convex averaging sequence.

- (b) Show that a stationary Poisson cluster process for which the cluster size distribution has finite  $k$ th moment, satisfies the conditions of (a).
- (c) Show that the conditions of (a) are not satisfied for  $k \geq 2$  by either of (i) any nontrivial mixture of Poisson processes; and (ii) a stationary renewal process whose lifetime distribution has finite first moment but infinite second moment. [Hint: Compare Exercises 4.1.1–2 and 4.4.5(c).]



12.6.6 Let the stationary random measure  $\xi$  on  $\mathbb{R}^d$  have finite  $k$ th moment measure  $M_k$ . Show that for bounded Borel functions  $f$  of bounded support in  $(\mathbb{R}^d)^{(k-1)}$ , as  $n \rightarrow \infty$ ,

$$\mathbb{E} \left| \frac{1}{\ell(A_n)} \int_{A_n} f(x_1, \dots, x_k) \xi(dx_1) \dots \xi(dx_k) - \int_{\mathcal{X}^{(k-1)}} f(x, x + y_1, \dots, x + y_{k-1}) \check{\psi}_k(dy_1 \times \dots \times dy_{k-1}) \right| \rightarrow 0.$$

12.6.7 Suppose the stationary random measure  $\xi(\cdot)$  has the mean square continuous nonnegative random function  $\eta(\cdot)$  as density; that is,  $\xi(A) = \int_A \eta(u) du$  a.s.

(a) Prove that the covariance measure  $C_2$  of  $\xi(\cdot)$  has no atom at the origin.

(b) When  $\eta(\cdot)$  is stationary and  $\text{cov}(\eta(x), \eta(y)) = \sigma^2 \rho(|x - y|)$ , show that

$$\text{var } \xi(0, x] = 2\sigma^2 \int_0^x (x - y) \rho(y) dy \quad (x > 0).$$

(c) Interpret the results of (a) and (b) in the (degenerate) case that  $\xi(A) = Y \ell(A)$  for a r.v.  $Y$  with  $\mathbb{E}(Y^2) < \infty$ .

12.6.8  $L_1$  version of Theorem 12.6.VI. Let the strictly stationary random measure  $\xi$  of Theorem 12.6.VI be ergodic and have finite  $k$ th moment measure. Show that under the conditions of the theorem, as  $n \rightarrow \infty$ ,

$$\mathbb{E} \left| \frac{1}{\ell(A_n)} \int_{A_n} \xi(x + B_1) \dots \xi(x + B_{k-1}) \xi(dx) - \check{M}_k(B_1 \times \dots \times B_{k-1}) \right| \rightarrow 0.$$

12.6.9 *Higher-order moments for marked random measures.* Show that for any  $k$ th-order stationary marked random measure or MPP on  $\mathbb{R}^d$  with marks in  $\mathcal{K}$  there exists a reduced measure  $\check{M}_k$  such that for any function  $f \in \text{BM}(\mathcal{X}^{(k)} \times \mathcal{K}^{(k)})$ , and writing  $\mathbf{x}_k, \boldsymbol{\kappa}_k$  for  $(x_1, \dots, x_k), (\kappa_1, \dots, \kappa_k)$  and so on,

$$\begin{aligned} & \int_{\mathcal{X}^{(k)} \times \mathcal{K}^{(k)}} f(\mathbf{x}_k, \boldsymbol{\kappa}_k) M_k(d(\mathbf{x}_k \times \boldsymbol{\kappa}_k)) \\ &= \int_{\mathcal{X}} dx \int_{\mathcal{X}^{(k-1)} \times \mathcal{K}^{(k)}} f(x, x + \mathbf{y}_{k-1}, \boldsymbol{\kappa}_k) \check{M}_k(d(\mathbf{y}_{k-1} \times \boldsymbol{\kappa}_k)). \end{aligned}$$

In particular, use this reduced measure to imitate (12.6.5) for the marked case.

12.6.10 *Higher-order ergodic theorem for the marked case.*

(a) *Extension of Lemma 12.6.VI.* As in Lemma 12.2.III, let  $\xi$  be a random measure on the product space  $\mathcal{X} \times \mathcal{K}$ , and let  $\mathcal{I}$  be the associated  $\sigma$ -algebra of events invariant under shifts  $S_x$ ,  $x \in \mathcal{X} = \mathbb{R}^d$ . Establish the existence of an  $\mathcal{I}$ -measurable random measure  $\check{\Psi}_k(\mathbf{x}_{k-1}, \boldsymbol{\kappa}_k)$  on  $\mathcal{X}^{(k-1)} \times \mathcal{K}^{(k)}$  such that for bounded Borel functions  $f$  on  $\mathcal{X}^{(k-1)} \times \mathcal{K}^{(k)}$  with bounded support

$$\begin{aligned} & \mathbb{E} \left[ \int_{\mathcal{X}^{(k)} \times \mathcal{K}^{(k)}} f(\mathbf{x}_k, \boldsymbol{\kappa}_k) \prod_{i=1}^k \xi(dx_i \times d\kappa_i) \mid \mathcal{I} \right] \\ &= \int_{\mathcal{X}} dx \int_{\mathcal{X}^{(k-1)} \times \mathcal{K}^{(k)}} f(x, x + \mathbf{y}_{k-1}, \boldsymbol{\kappa}_k) \check{\Psi}_k(d(\mathbf{y}_{k-1} \times \boldsymbol{\kappa}_k)). \end{aligned}$$

(b) Establish a corresponding version of Proposition 12.6.VII for MPPs, with  $\check{\Psi}_k$  replacing  $\check{\psi}_k$ .

- 12.6.11 *Kagan's conjectures.* On the basis of empirical evidence from current earthquake catalogues [see also Kagan and Knopoff (1980)], Kagan (1981a, b) conjectured that earthquakes in the crust have a type of self-similar distribution in which the second-, third-, and fourth-order factorial moment densities have the respective forms:

$$\begin{aligned} m_{[2]}(x, y) &\sim 1/D(x, y); & m_{[3]}(x, y, z) &\sim 1/A(x, y, z); \\ m_{[4]}(w, x, y, z) &\sim 1/V(w, x, y, z), \end{aligned}$$

where  $D$  is the distance,  $A$  the area, and  $V$  the volume described by the respective arguments of the densities.

- (a) Investigate conditions under which the above conjectures are compatible with the existence of a stationary process with the prescribed densities. [Hint: Consider integrability conditions at the origin.]  
 (b) Investigate conditions under which they might be approximately true for processes of fractal type, meaning, that they are concentrated in 'random faults' on lower-dimensional elements such as lines or surfaces.

[Remark: These conjectures are still unresolved; for further discussion see Kagan and Vere-Jones (1995).]

- 12.6.12 Suppose that the process  $N_0$  in Example 12.6(b) is Poisson with rate parameter  $\lambda_0$ . Then the output process  $N_2$  has covariance density

$$\lambda_0^2 c(u) = \lambda_0^2 \operatorname{cov}(J(0), J(u)) = \lambda_0^2 [\beta(u) - \alpha^2].$$

Show that, when  $N_1$  is a Poisson process at rate  $\lambda_1$  and the  $Y_i$  are exponentially distributed with mean  $1/\mu$ ,

$$\begin{aligned} \alpha &= \frac{\mu}{\lambda_1 + \mu}, & \beta(u) &= \frac{\mu(\lambda_1 e^{-(\lambda_1 + \mu)u} + \mu)}{(\lambda_1 + \mu)^2}, \\ \pi(t) &= \frac{\lambda_1 + \mu e^{-(\lambda_1 + \mu)t}}{\lambda_1 + \mu}, & \gamma(u) &= \lambda_1(1 - \pi(u)) = \frac{\lambda_1 \mu(1 - e^{-(\lambda_1 + \mu)t})}{\lambda_1 + \mu}. \end{aligned}$$

- 12.6.13 Replace the inhibitory mechanism of Example 12.6(b) defined via the i.i.d. sequence  $\{Y_i\}$  by  $Z_i = \min(T_{t_i}, t_{i+1} - t_i)$ , where  $T_t$  is the forward recurrence time r.v. of the process  $N_0$ . Show that when  $N_0$  and  $N_1$  are independent stationary simple point processes with intensity  $\lambda_0$ , the output process has intensity

$$\lambda_2 = \lambda_0 \int_0^\infty G(t) dF(t),$$

where  $G$  is now the d.f. of a lifetime r.v.  $Y$  for the process  $N_0$  and  $F$  is the d.f. of the backward recurrence time r.v. of the process  $N_1$ .

## 12.7. Long-range Dependence

Long-range dependence (LRD) of a stochastic process could in principle relate to any of several characteristics of the process, but it has now generally come to be associated with second moments. For stationary point processes and random measures on  $\mathbb{R}$ , we define it via the following variance property given in Daley and Vesilo (1997) and already referred to in Exercise 4.5.13.

**Definition 12.7.I.** A second-order stationary point process or random measure  $\xi$  on  $\mathbb{R}$  is long-range dependent if

$$\limsup_{t \rightarrow \infty} \frac{\text{var } \xi(0, t]}{t} = \infty. \quad (12.7.1)$$

When the  $\limsup$  here is finite, the random measure is short-range dependent.

Definition 12.7.I is less easy to apply in higher dimensions, although essentially the same ideas can be used. One approach, outlined in Exercise 12.7.3, is to consider the growth of  $\text{var } \xi(A_n)$  relative to  $\ell(A_n)$  as  $A_n$  increases through a convex averaging sequence of sets. In the treatment and examples below, we mostly restrict attention to point processes in  $\mathbb{R}$ .

Recall that the second moment measure of a stationary random measure on an interval  $(0, t)$  in  $\mathbb{R}$  cannot grow faster than  $O(t^2)$  (see Exercise 12.7.1), so that

$$\limsup_{t \rightarrow \infty} t^{-2} \text{var } \xi(0, t] < \infty. \quad (12.7.2)$$

This property implies that if in the denominator in (12.7.1) we replace  $t$  by  $t^\alpha$ , the ratio can be infinite in the limit only if  $\alpha < 2$ . It is therefore convenient to delineate this range by an index as in the next definition [the name recalls early work on long-range dependence in flow records of the river Nile by the British engineer Hurst; see Beran (1994)].

**Definition 12.7.II.** A long-range dependent stationary random measure  $\xi$  has Hurst index

$$H = \sup \left\{ h: \limsup_{t \rightarrow \infty} \frac{\text{var } \xi(0, t]}{t^{2h}} = \infty \right\}. \quad (12.7.3)$$

Then, the Hurst index must lie in the interval  $[0, 1]$ , although for long-range dependence it is more narrowly confined to  $\frac{1}{2} \leq H \leq 1$ . Furthermore, to have  $H < 1$ , we can immediately rule out the nonergodic case, for unless the invariant  $\sigma$ -field  $\mathcal{I}$  is trivial, it follows from Exercise 12.2.9 that

$$\text{var } \xi(0, x] \sim x^2 \Gamma(\{0\}) \quad (x \rightarrow \infty),$$

where  $\Gamma(\{0\}) = \text{var } Y$  and  $Y = E(\xi(\mathbb{U}) \mid \mathcal{I})$ , so  $H = 1$  when  $\text{var } Y > 0$ .

For a stationary Poisson process at rate  $\lambda$  the ratio at (12.7.1) equals  $\lambda$  for all  $t$  so it cannot be long-range dependent. The next few examples indicate further possibilities of short- and long-range dependence.

**EXAMPLE 12.7(a).** For a stationary renewal process  $N(\cdot)$  with renewal function  $U(\cdot)$  as in Section 4.1, it follows from (3.5.7) that

$$\text{var } N(0, t] = \lambda \int_0^t (2[U(u) - \lambda u] - 1) du \quad (12.7.4)$$

and [cf. Exercise 4.4.5(d)] that  $\text{var } N(0, t] \leq (\text{const.})t$  for some finite constant if and only if  $\sup_{t \geq 0} [U(t) - \lambda t] < \infty$ , which is the case if and only if the lifetime

distribution underlying  $U(\cdot)$  has its second moment finite. In other words, *the independence between intervals (of a renewal process) is not sufficient to eliminate the possibility of long-range dependence.*

The Hurst index for such  $N$  is identified in Exercise 12.7.4.  $\square$

EXAMPLE 12.7(b) *Cluster process.* Let  $V^c(x) = \text{var } N_c(0, x]$  denote the variance function [see (8.1.13b)] of the stationary cluster centre process  $N_c(\cdot)$  of a cluster process  $N(\cdot)$  with independent identically distributed component processes  $N_m(\cdot \mid \cdot)$  (see Definition 6.3.I). Equation (53) of Daley (1972) shows that when  $N$  is stationary and  $N_c$  is orderly at rate  $\lambda_c$ , and the second moment of the component process  $N_m$  is finite, the variance function of  $N$  is given by

$$\begin{aligned} \text{var } N(0, x] &= \lambda_c \int_{-\infty}^{\infty} \mathbb{E}([N_m((-u, x - u] \mid 0)]^2) du \\ &\quad + \int_{-\infty}^{\infty} m_1(dy) \int_{-\infty}^{\infty} m_1(dz) \left[ -\lambda_c(x - |y - z|)_+ \right. \\ &\quad \left. + V^c(y - z + x) - 2V^c(y - z) + V^c(y - z - x) \right], \end{aligned}$$

where  $m_1(y) = \mathbb{E}N_m((-\infty, y] \mid 0)$  and  $V^c(x) = V^c(-x)$  for  $x < 0$ . The three variance terms here equal  $\text{cov}(N_c(0, x], N_c(y - z, y - z + x])$ ; this is dominated by  $V^c(x)$ , which implies that when  $\lim_{x \rightarrow \infty} V^c(x)/x^\alpha$  exists finite and equals  $\lambda_{2,\alpha}$  say, the dominated convergence theorem can be applied when  $\alpha > 1$  to conclude that  $\lim_{x \rightarrow \infty} x^{-\alpha} \text{var } N(0, x] = m_1^2 \lambda_{2,\alpha}$ , where  $m_1 = m_1(\infty)$ .

In this example, long-range dependent behaviour of the cluster centre process [as shown by  $V^c(x) \sim \text{const. } x^\alpha$ ] is carried over into the cluster process itself, magnified by the square of the mean cluster size.

When  $V^c(x)/x \rightarrow \lambda_2$  say (so  $N_c$  is not long-range dependent) the same argument [e.g., Daley (1972)] gives instead

$$V(x) \sim m_1^2 \lambda_2 x + (\text{var } N_m(\mathbb{R} \mid 0)) \lambda_c x \quad (x \rightarrow \infty);$$

that is, variability in the component processes is no longer swamped. Consistent with this fact, having heavy-tailed distributions for the distances from the cluster centre to the cluster members is not sufficient to cause long-range dependence in the process as a whole. For example, the variance in a Neyman–Scott process with Poisson centre process [see, e.g., (6.3.19) and (12.7.5) below] is given by

$$\frac{\text{var } N(0, x]}{x} = \int_{-x}^x \left(1 - \frac{|u|}{x}\right) \check{C}_{[2]}(du) = \lambda_c [m_1^2 + m_{[2]} F * F^-(x)],$$

where  $m_{[2]}$  is the second factorial moment of the cluster size distribution, and  $F * F^-(\cdot)$  is the distribution function of the distance  $X_1 - X_2$  between two members of a cluster. When  $m_{[2]}$  is finite, this converges to a finite limit irrespective of the character of the distribution of  $F$ .  $\square$

EXAMPLE 12.7(c) *Superpositions*. Suppose the stationary random measure  $\xi$  is expressible as the sum of the two stationary random measures  $\xi_1$  and  $\xi_2$ , and that all processes have finite second moments; we write  $\xi(t) = \xi(0, t]$  and so on. Then for  $t > 0$  and from the Cauchy–Schwarz inequality,

$$\begin{aligned} \text{var } \xi(t) &= \left( \sqrt{\text{var } \xi(t)} \right)^2 = \text{var } \xi_1(t) + \text{var } \xi_2(t) + 2 \text{cov}(\xi_1(t), \xi_2(t)) \\ &\leq (\geq) \left( \sqrt{\text{var } \xi_1(t)} \pm \sqrt{\text{var } \xi_2(t)} \right)^2. \end{aligned} \quad (12.7.5)$$

Let  $H, H_1, H_2$  denote the Hurst indexes of  $\xi, \xi_1, \xi_2$ . Using (12.7.5), deduce that either  $H < H_1 = H_2$  or  $H = \max(H_1, H_2)$ , and the latter holds when  $\xi_1$  and  $\xi_2$  are independent. Daley and Vesilo (2000) give details and applications to some queueing examples.  $\square$

Because the variance properties are controlled by the reduced second moment measure  $\check{C}_2$ , we should expect the distinction between long- and short-range dependence to be expressible in terms of this measure. The next lemma gives a partial resolution of this question; Example 12.7(e) indicates that the complement of these sufficient conditions does not yield a set of necessary conditions.

**Lemma 12.7.III.** *Let  $\xi$  be a second-order stationary random measure in  $\mathbb{R}$ , and write  $\check{C}_2 = \check{C}_2^+ - \check{C}_2^-$  for the Jordan–Hahn decomposition of its reduced covariance measure into its positive and negative parts.*

- (a)  *$\xi$  is short-range dependent if  $\check{C}_2$  is totally finite. When  $\check{C}_2^-$  is totally finite,  $\xi$  is long-range dependent if and only if the positive part  $\check{C}_2^+$  is not totally finite.*
- (b)  *$\xi$  is short-range dependent if its Bartlett spectrum  $\Gamma(\cdot)$  has a bounded density in a neighbourhood of the origin.*

PROOF. The results are proved from the relations [cf. equations (8.1.13) and (8.2.3)]

$$\begin{aligned} \frac{\text{var } \xi(0, t]}{t} &= \int_{-t}^t (1 - |u|/t) \check{C}_2(du) = \int_{-\infty}^{\infty} \left(1 - \frac{|u|}{t}\right)_+ \check{C}_2(du) \\ &= \int_{-\infty}^{\infty} \left(\frac{\sin \frac{1}{2}\theta}{\frac{1}{2}\theta}\right)^2 \Gamma(t d\theta), \end{aligned} \quad (12.7.6)$$

expressing the middle integral as a difference of two integrals (involving  $\check{C}_2^+$  and  $\check{C}_2^-$ , respectively) to which the monotone convergence theorem can be applied. Similarly, if  $\Gamma(\cdot)$  has a density  $\gamma(\theta)$  which is bounded in a neighbourhood of  $\theta = 0$ , it follows from dominated convergence that the final integral remains bounded as  $t \rightarrow \infty$ .  $\square$

EXAMPLE 12.7(d) *LRD Cox process*. The Cox process  $N(\cdot)$  directed by the stationary random measure  $\xi(\cdot)$  on  $\mathcal{B}_{\mathbb{R}}$  is stationary and their variance functions are related by  $\text{var } N(0, t] = E(\xi(0, 1])t + \text{var } \xi(0, t]$  (Proposition 6.2.II). Thus,  $N(\cdot)$  has exactly the same long-range dependence behaviour as  $\xi(\cdot)$ .

For example, when  $\xi(\cdot)$  accumulates mass at unit rate during just one of the phases of an alternating renewal process with generic lifetimes  $X_j$  ( $j = 1, 2$ ) say, each with finite first moments, whenever one of the  $X_j$  has infinite second moment the random measure  $\xi$  (and hence  $N$  also) is long-range dependent. Details are left to Exercise 12.7.5.  $\square$

EXAMPLE 12.7(e) *Deterministic point process.* Example 8.2(e) shows that the reduced covariance measure  $\check{C}_2$  of a stationary deterministic point process  $N$  with span  $a$  has positive atoms  $\check{C}_2^+(\{ka\}) = 1/a$  for  $k = \pm 1, \pm 2, \dots$ ,  $\check{C}_2^+(A) = 0$  whenever  $A \cap \{ka\} = \emptyset$  for all such  $k$ , and  $\check{C}_2^-(du) = \ell(du)/a^2$ . This process has a periodic variance function, with

$$0 \leq \text{var } N(0, t] \leq \frac{1}{4},$$

so it is not long-range dependent. On the other hand the positive and negative parts of  $\check{C}_2$  are mutually singular [ $\check{C}_2^+$  is purely atomic and  $\check{C}_2^-$  is a multiple of Lebesgue measure], and both  $\check{C}_2^+(0, u]$  and  $\check{C}_2^-(0, u]$  increase indefinitely with the length  $u$  of the interval, whereas  $\check{C}_2(0, ka] = 0$  for  $k = 1, 2, \dots$ .

This example shows that we cannot exclude the possibility of the process being short-range dependent even when  $\check{C}_2$  fails to be totally finite.  $\square$

Long-range dependence is frequently understood in terms of power-law decay of the correlation function. In the point process context, this means looking at the covariance density function rather than the variance function of which it is the second derivative [Exercises 8.1.3(g), 12.7.2]. For example, Ogata and Katsura (1991) discuss situations in which the covariance density shows a power-law decay but describe it in terms of ‘fractal behaviour’.

We note here the difficulty of maintaining a coherent and consistent naming system in situations where interest from the media and the general public preponderates. Fractals, with all the striking images associated with them in Mandelbrot’s writings form a case in point. Anything exhibiting some form of scaling behaviour or self-similarity, particularly when it is linked to power-law decay, is almost automatically labelled a ‘fractal’ in a nonmathematical context. By contrast, in this book we have tried to establish and maintain distinctions between the concepts of long-range dependence, representing (as in Exercise 4.5.13 and this section) a variance property related to power-law decay in a covariance or correlation function; heavy-tailed behaviour in a probability distribution; scale-invariance and self-similarity (discussed in the next section); and fractal behaviour which we take up in Section 13.6 and relate to the behaviour of moment densities at very small distances or time intervals. In a similar vein, the Hurst index is indeed an index and not dependent on any parametric setting.

## Exercises and Complements to Section 12.7

12.7.1 When the set  $A \subset \mathbb{R}^d$  is a hyper-rectangle of side-lengths  $n_1, \dots, n_d$  and  $\xi$  is a second-order stationary random measure, express  $\xi(A)$  as the sum of

the measures on  $\nu_A = \prod_{i=1}^d n_i$  unit cubes, and use the Cauchy–Schwartz inequality to conclude that the second moment measure  $M_2$  of  $\xi$  satisfies  $M_2(A) \leq (\nu_A^2) M_2(\mathbb{U}^d)$ . Compare with Exercise 8.1.3(b).

12.7.2 Show that a stationary point process on  $\mathbb{R}$  with reduced covariance density  $\check{c}(x) \sim a/x^\gamma$  ( $x \rightarrow \infty$ ) for  $a > 0$  and  $0 < \gamma < 1$  has Hurst index  $1 - \frac{1}{2}\gamma$ .

12.7.3 Call a stationary random measure  $\xi: \mathcal{B}(\mathbb{R}^d) \mapsto \mathbb{R}_+$  *long-range dependent* whenever  $\limsup_{n \rightarrow \infty} (\text{var}[\xi(A_n)]/\ell(A_n)) = \infty$  for some convex averaging sequence  $\{A_n; n = 1, 2, \dots\}$ .

(a) Show that this definition is independent of the convex averaging sequence.

(b) Show, analogously to the first part of (12.7.6), that

$$\frac{\text{var} \xi(A_n)}{\ell(A_n)} = \int_{\mathbb{R}^d} \frac{\ell(A_n \cap T_z A_n)}{\ell(A_n)} \check{C}_2(dz).$$

Show that the ratio in the integrand is always bounded by 1, and is close to 1 for small  $z$ . Hence provide an extension of Lemma 12.7.III to  $\mathbb{R}^d$ .

12.7.4 *Hurst index of a LRD renewal process.* Let  $N$  be a stationary renewal process as in Example 12.7(a) whose generic lifetime r.v.  $X$  with d.f.  $F$  and tail  $\bar{F}(x) = 1 - F(x)$  has moment index  $\kappa$  defined by

$$\kappa \equiv \inf\{k: E(X^k) = \infty\} = \liminf_{x \rightarrow \infty} [-\log \bar{F}(x)]/\log x$$

[see, e.g., Daley and Goldie (2005) for the equality] for which  $1 < \kappa < 2$ . Use (12.7.4) and the asymptotic behaviour  $U(t) - \lambda t \sim \lambda^2 \int_0^\infty \min(u, t) \bar{F}(u) du$  [Sgibnev (1981)] to deduce that its Hurst index  $H = \frac{1}{2}(3 - \kappa)$ .

[Hint: Daley (1999) gives a proof by contradiction; the proof via (12.7.4) and Exercise 4.4.5(c) as just indicated is direct.]

12.7.5 *LRD in ON/OFF processes.* Let  $I(t)$  be the stationary  $\{0, 1\}$ -valued process generated by the ON phases of a stationary alternating renewal process  $N_{\text{alt}}$  with lifetime d.f.s  $F_1$  for such ON-phases and  $F_0$  for the OFF phases (so  $I(t) = 1$  when the alternating renewal process is in an ON phase,  $= 0$  otherwise). Let  $T_1(t)$  denote the accumulated duration during  $(0, t)$  for which  $I(u) = 1$  for  $0 < u < t$ , so that  $T_1(t) = \int_0^t I(u) du$ . Let  $T_0(t) = t - T_1(t)$  denote the corresponding duration for which  $I(\cdot)$  is in the OFF phase. Observe that  $\text{var} T_j < \infty$  for  $j = 0, 1$  because these  $T_j$  are bounded r.v.s, and in fact

$$\text{cov}(T_1(t), T_0(t)) = \text{cov}(T_1(t), t - T_1(t)) = -\text{var} T_1(t) = -\text{var} T_0(t).$$

Consequently, if the Cox process  $N_{\text{Cox}}$  directed by the ON phases of  $N_{\text{alt}}$  is LRD (i.e.,  $\limsup_{t \rightarrow \infty} t^{-1}[\text{var} T_1(t)] = \infty$ ), then so too is the Cox process directed by the OFF phases of  $N_{\text{alt}}$ . Show that  $N_{\text{Cox}}$ ,  $N_{\text{alt}}$ , and the stationary random measure  $\xi$  with  $I(\cdot)$  as its density, all have the same Hurst index. [Hint: Daley (2007) studies this example further, showing that for  $t \rightarrow \infty$ ,  $(\text{var} N_{\text{Cox}}(0, t))/(\text{var} N_{\text{alt}}(0, t))$  has a limit if one of the lifetime distributions  $F_j$  ( $j = 0, 1$ ) has finite second moment but if both have infinite second moment the ratio can oscillate indefinitely. Daley, Rolski and Vesilo (2007) extend this work to a Cox process driven by a LRD semi-Markov process.]

## 12.8. Scale-invariance and Self-similarity

In this last section we look at processes where invariance under multiplicative actions (scale changes) rather than additive actions (translations) plays the key role. Such processes include the important class of *self-similar* processes (also called *auto-modelling* in the older Russian literature). In the context of the present book, the most relevant concept is that of a self-similar random measure, where nonnegativity plays a dominant role, and the theory is rather different from that of the fractional Brownian motions and related two-sided processes which have become familiar in finance and other application areas. The self-similar random measures we consider are purely atomic, and can be described by the marked point process of the locations and sizes of the atoms. Moreover, self-similarity of the random measure can be restated in terms of a corresponding invariance property of this marked point process; we call this property *biscale invariance*.

We first consider the simpler case of point processes invariant under scale changes about a fixed origin. A process on  $\mathcal{X} = \mathbb{R}^d$  is called *scale-invariant* if it is invariant under the group of scale changes

$$\{T_\alpha: 0 < \alpha < \infty\}, \quad \text{where for } x \in \mathbb{R}^d, \quad T_\alpha x = \alpha x. \quad (12.8.1)$$

This group splits  $\mathbb{R}^d$  into equivalence classes, one of which is the origin, and the others can be identified with rays originating from the origin. Now  $\mathbb{R}^d \setminus \{0\}$  can be written as the product  $\mathbb{S}_d \times \mathbb{R}_0^+$ , where  $\mathbb{S}_d$ , the group of  $d$ -dimensional rotations, can in turn be identified with the surface of the  $d$ -dimensional unit sphere, and  $\mathbb{R}_0^+$  denotes the open half-line  $(0, \infty) = \mathbb{R}_+ \setminus \{0\}$ . Note that  $\mathbb{S}_1$  is just the two-point group  $\mathcal{T}_2 = \{-1, 1\}$  under multiplication.

$\mathbb{R}_0^+$  is a group under multiplication, and it has the unique invariant measure  $h(dx) = dx/x$ . It is now obvious, but also follows formally from Lemma A2.7.II, that any measure on  $\mathbb{R}^1$  that is invariant under scale changes, can be represented as the sum of a point mass at the origin, and the direct product of a two-point mass on  $\mathcal{T}_2$  and the measure  $h(\cdot)$  on  $\mathbb{R}_0^+$ . Similarly, a scale-invariant measure on  $\mathbb{R}^d$  can be represented as the sum of a point mass at the origin and a measure  $\kappa(d\theta) dr/r$  on  $\mathbb{R}^d \setminus \{0\} = \mathbb{S}_d \times \mathbb{R}_0^+$ , where  $\kappa(\cdot)$  is an arbitrary totally finite measure on  $\mathbb{S}_d$ .

The position of the origin is unimportant here, but it is clear that the above structure is incompatible with translations in  $\mathbb{R}^d$ , so that a measure on  $\mathbb{R}^d$  cannot be invariant under both translations and scale changes, a result with important consequences for both scale-invariant and self-similar random measures.

**EXAMPLE 12.8(a)** *Scale-invariant Poisson processes on  $\mathbb{R}^d$ .* As in Example 12.1(a) we deduce that if a Poisson process on  $\mathbb{R}^d$  is invariant under scale changes, its parameter measure must have the same property, and must therefore have the structure described above, namely, the sum of a point mass at the origin and a measure  $\kappa(d\theta) dr/r$  on  $\mathbb{S}_d \times \mathbb{R}_0^+$ .  $\square$



Part (a) of the proposition below is an immediate consequence of this discussion. A similar argument applied to the first moment measure, whenever it exists, yields part (b).

**Proposition 12.8.I.** (a) *A Poisson process on  $\mathbb{R}^d$  cannot be simultaneously invariant under both scale changes and translations.*

(b) *No stationary random measure on  $\mathbb{R}^d$  with finite expectation measure can be scale-invariant.*

We turn next to a discussion of self-similarity for random measures. Here, a change in scale is balanced by a compensating change in mass.

**Definition 12.8.II.** *Let  $D$  be a finite positive constant. A random measure is self-similar with similarity index  $D$ , (self-similar or  $D$ -self-similar for short), if its distribution is invariant under the group of transformations  $\{R_\alpha^{(D)}: \alpha \in \mathbb{R}_0^+\}$  defined on boundedly finite measures  $\mathcal{M}_\mathcal{X}^\#$  by*

$$R_\alpha^{(D)}\mu(A) = \alpha^{-D}\mu(\alpha A) \quad (A \in \mathcal{B}_\mathcal{X}). \quad (12.8.2)$$

Zähle (1988) uses the shorter terminology above, and includes discussion of nonprobability measures on  $\mathcal{M}_\mathcal{X}^\#$ . Even more extensive work is covered in Zähle (1990a, b, 1991).

Note that self-similarity, like scale-invariance, refers in the first instance to invariance relative to a fixed origin; it is only under stationarity, or some explicit rule describing how the invariance properties alter as we shift the origin, that the concept extends beyond this case.

The transformations  $R_\alpha^{(D)}$  do not result directly from transformations of the phase space  $\mathcal{X}$  into itself, but do still induce a group, the *renormalization group*, of bounded continuous transformations of  $\mathcal{M}^\#(\mathbb{R}^d)$  into itself (see Exercise 12.8.1).

We start with two negative results. Because of the change in mass, the renormalization group does not map  $\mathcal{N}^\#(\mathbb{R}^d)$  into itself. This justifies the following.

**Proposition 12.8.III.** *A point process cannot be self-similar.*

The class of (deterministic) measures invariant under  $R_\alpha^{(D)}$  is not a rich family: on  $\mathbb{R}_+$  it is confined to measures with power-law densities [*hyperbolic densities* in the usage of Mandelbrot (1982, p. 204)]

$$f_D(x) = Cx^{1-D} \quad (C > 0, x > 0).$$

Only in the trivial case  $D = 1$  is the measure invariant under both the similarity transformations  $R_\alpha^{(D)}$  and translations; it reduces then to a multiple of Lebesgue measure.

The situation for general random measures is more rewarding. To give a preview of the issues which arise, we examine first, without assuming stationarity, the consequences of self-similarity on the representation for completely random measures given in Theorem 10.1.III.

Suppose that  $\xi$  is a completely random measure defined on  $\mathcal{X} = \mathbb{R}^d$ , so that, as in (10.1.4), it can be represented in terms of a drift, a set of fixed atoms, and a Poisson process  $N$  in  $\mathbb{R}^d \times \mathbb{R}_0^+$ . Operating on this representation by  $R_\alpha^{(D)}$ , and using differential notation for brevity, we find

$$\begin{aligned} R_\alpha^{(D)}\xi(dx) &= \alpha^{-D}\nu(\alpha dx) + \int_0^\infty y \alpha^{-D} N(\alpha dx \times dy) + \sum_k U_k \alpha^{-D} \delta_{x_k/\alpha}(dx) \\ &= \alpha^{-D}\nu(\alpha dx) + \int_0^\infty y N(\alpha dx \times \alpha^D dy) + \sum_k U_k \alpha^{-D} \delta_{x_k/\alpha}(dx). \end{aligned}$$

This last expression again corresponds to a completely random measure: the measure  $\nu$  has been transformed by  $R_\alpha^{(D)}$  as at (12.8.2), the Poisson process  $N$  has been subjected to the *biscale transformation*  $S_\alpha^{(D)}$  on  $\mathcal{X} \times \mathbb{R}_0^+$  given by

$$S_\alpha^{(D)}(N(A \times K)) = N(\alpha A \times \alpha^D K), \quad (12.8.3)$$

and the fixed atoms have been transformed both in mass and in location.

If the distribution of the completely random measure is to remain invariant under all transformations  $R_\alpha^{(D)}$  ( $\alpha > 0$ ), then it is clear that there can be no fixed atoms, that the (deterministic) measure  $\nu$  must be invariant under the transformations  $R_\alpha^{(D)}$ , and that the parameter measure  $\mu$  of the bivariate Poisson process  $N$  must be invariant under the transformations  $S_\alpha^{(D)}$ . Thus, we have reduced the problem of characterizing the class of self-similar completely random measures to the problem of characterizing the classes of measures invariant under these two groups of transformations.

For simplicity we consider only the case  $\mathcal{X} = \mathbb{R}^1$ ; the details for  $\mathcal{X} = \mathbb{R}^d$  are similar (see Exercise 12.8.2 for the case  $d = 2$ ).

As in Example 12.8(a), it is necessary to consider separately the action of the transformations on  $\mathbb{R}_0^+$ ,  $\{0\}$  and  $\mathbb{R}_0^- = \{x: -x \in \mathbb{R}_0^+\}$ . Because the processes have no fixed atoms,  $\nu\{0\} = 0$  and  $\mu$ , the parameter measure of  $N$ , has  $\mu(\{0\} \times \mathbb{R}_0^+) = 0$ . Thus, we may consider the effect of  $R_\alpha^{(D)}$  on measures  $\nu$  acting on  $\mathcal{B}_{\mathbb{R}_+}$ , and of  $S_\alpha^{(D)}$  on measures  $\mu$  on  $\mathcal{B}_{\mathbb{R}_+ \times \mathbb{R}_+}$ , with similar results following for the components on  $\mathbb{R}_-$  and  $\mathbb{R}_- \times \mathbb{R}_+$ .

From (12.8.2), invariance of  $\nu$  under  $R_\alpha^{(D)}$  implies that  $\nu$  is absolutely continuous on  $\mathbb{R}_0^+$ ; its density with respect to Lebesgue measure  $\ell$  is given by

$$\frac{d\nu}{d\ell}(x) = c_1 x^{D-1} \quad (x \in \mathbb{R}_0^+).$$

Similarly, on  $\mathbb{R}_0^-$ ,

$$\frac{d\nu}{d\ell}(x) = c_2 |x|^{D-1} \quad (x \in \mathbb{R}_0^-).$$

Next, consider the representation of the parameter measure  $\mu$  on the quadrant  $\mathbb{R}_0^+ \times \mathbb{R}_0^+$ . Invariance of the distribution of  $N$  under (12.8.3) implies that

$\mu$  itself is invariant under the biscal shifts  $S_\alpha^{(D)}$ , and hence is invariant under shifts along the curves  $x^D/y = \text{constant}$ . This suggests writing

$$u = \log x, \quad v = D \log x - \log y,$$

so that in the  $(u, v)$ -plane the transformation (12.8.3) becomes

$$(u, v) \mapsto (u + \log \alpha, v). \quad (12.8.4)$$

We now deduce from Lemma A2.7.II that  $\tilde{\mu}$ , the image of  $\mu$  under this mapping, reduces to a product of Lebesgue measure along the  $u$ -axis and an arbitrary  $\sigma$ -finite measure  $\tilde{\rho}_1$  along the  $v$ -axis. Thus, integration with respect to  $\mu$  in this quadrant can be represented in the form

$$\int_{\mathbb{R}_0^+ \times \mathbb{R}_0^+} f(x, y) \mu(dx \times dy) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(e^u, e^{Du-v}) du \tilde{\rho}_1(dv). \quad (12.8.5)$$

Similar considerations may be applied on  $\mathbb{R}_0^- \times \mathbb{R}_0^+$ , with a possibly different measure  $\tilde{\rho}_2$  replacing  $\tilde{\rho}_1$ . The finiteness constraints (10.1.5), when expressed in the form

$$\int_0^\infty (1 - e^{-y}) \mu(A \times dy) < \infty \quad (\text{bounded } A \in \mathcal{B}(\mathbb{R})),$$

lead to the requirements that for  $i = 1, 2$ ,

$$\int_{-\infty}^0 (1 + |v|) \tilde{\rho}_i(dv) < \infty \quad \text{and} \quad \int_0^\infty e^{-v} \tilde{\rho}_i(dv) < \infty. \quad (12.8.6a)$$

If, in particular,  $\tilde{\rho}_1$  is absolutely continuous with respect to Lebesgue measure, then it is more convenient to write, for  $(x, y) \in \mathbb{R}_0^+ \times \mathbb{R}_0^+$  and with  $\tilde{\rho}_1(dv) = \rho_1(v) dv$ ,

$$\mu(dx \times dy) = \frac{\rho_1(\log(x^D/y))}{xy} dx dy \equiv \frac{\eta_1(x^D/y)}{xy} dx dy$$

for some nonnegative, locally integrable function  $\eta_1$  on  $\mathbb{R}_0^+$ . If also a similar representation holds for  $(x, y) \in \mathbb{R}_0^- \times \mathbb{R}_0^+$ , then an analogous representation holds for some similar function  $\eta_2$  on  $\mathbb{R}_0^+$  and  $|x|$  in place of  $x$ .

When such absolute continuity conditions hold, the Laplace functional  $L[f]$  for the random measure  $\xi$  can be written

$$L[f] = \exp(L_1[f] + L_2[f]) \quad (f \in \text{BM}_+(\mathbb{R})),$$

where  $L_1[f]$  equals

$$-c_1 \int_0^\infty x^{D-1} f(x) dx + \int_0^\infty \int_0^\infty (e^{yf(x)} - 1) \frac{\eta_1(x^D/y)}{xy} dx dy, \quad (12.8.6b)$$

and a similar expression holds for  $L_2$  with  $c_1$  replaced by  $c_2$ ,  $\eta_1(x^D/y)$  by  $\eta_2(|x|^D/y)$ , and integration of  $x$  is over  $(-\infty, 0)$  in place of  $(0, \infty)$ . Conditions (12.8.6a) transform to

$$\int_0^1 \frac{1 + |\log z|}{z} \eta_i(z) dz < \infty \quad \text{and} \quad \int_1^\infty \frac{\eta_i(z)}{z^2} dz < \infty. \quad (12.8.7)$$

We thus have a complete answer to the representation problem in the one-dimensional case.

**Proposition 12.8.IV.** *A completely random measure on  $\mathbb{R}$  is self-similar if and only if, in terms of the representation in Theorem 10.1.III, there are no fixed atoms, and the measures  $\nu, \mu$  can be written, as in (12.8.5), in terms of positive constants  $c_1, c_2$  and measures  $\tilde{\rho}_1, \tilde{\rho}_2$  satisfying (12.8.6a). In particular, when these measures have densities, the Laplace functional of the random measure can be written in the form (12.8.6b) for nonnegative, locally integrable functions  $\eta_1, \eta_2$  satisfying (12.8.7).*

For applications we generally require the random measure to be stationary as well as self-similar, in which case the representation must also be invariant under shifts along the  $x$ -axis. Then, the first term must vanish unless  $D = 1$ , when it reduces to a constant multiple of Lebesgue measure along the whole real axis. For the measures in the second term, the additional condition is easily seen to be satisfied if and only if  $\eta_1(v) = \rho v^{1/D} = \eta_2(v)$ , corresponding to

$$\mu(dx \times dy) = \rho y^{-(1+1/D)} dx dy. \quad (12.8.8)$$

Then the constraints at (12.8.7) require  $D < \infty$  and  $D > 1$  respectively. Hence, it follows that the class of completely random measures that are both stationary and self-similar reduces, for  $D = 1$ , to the trivial example of a constant multiple of Lebesgue measure, and for  $1 < D < \infty$ , to the stable processes with index  $\alpha = 1/D$  and Laplace functional of the form [cf. Example XIII.7(c) of Feller (1966)]

$$\begin{aligned} -\log L[f] &= \rho \int_{-\infty}^{\infty} dx \int_0^{\infty} \frac{1 - e^{-yf(x)}}{y^{1+1/D}} dy \\ &= \rho D \Gamma(1 - D^{-1}) \int_{-\infty}^{\infty} [f(x)]^{1/D} dx. \end{aligned} \quad (12.8.9)$$

**Corollary 12.8.V.** *A completely random measure on  $\mathbb{R}$  is both stationary and self-similar if and only if there is no drift or atomic component, and the Poisson process in representation (10.1.4) has density (12.8.8), with  $1 < D < \infty$ , or equivalently if and only if it reduces to a stable process whose Laplace functional has the form (12.8.9).*

The representation of the stable processes arising in the preceding analysis suggests a more general structural form for self-similar random measures, namely, a representation of the type

$$\xi(A) = \int_0^{\infty} y N(A \times dy), \quad (12.8.10)$$

for some class of extended MPPs  $N$  more general than the Poisson process.

The use of an extended MPP here can be avoided by altering the metric on  $\mathbb{R}_0^+$  to (say)  $d(x, y) = |\log(x/y)|$ , thus effectively removing the origin to an inaccessible boundary point. With such a change of metric, the process remains boundedly finite, although its ground process remains undefined as a

point process. A change of metric of this kind corresponds to representing the mark on a logarithmic or decibel scale, as indeed is commonly done in physical applications, where, for example, the energy release might be the physically meaningful mark, for which some form of self-similarity could be claimed, although the most convenient measure of size is more often a measurement on a decibel scale. The same sort of transformation lies behind (12.8.4). Exercise 12.8.3 restates the requirements of self-similarity in terms of measurements on an associated logarithmic or decibel scale.

We proceed to a more systematic examination of the properties of random measures defined by (12.8.10). As in the independent case, a key role is played by the biscale transformations, and we adopt the following definition.

**Definition 12.8.VI.** *An extended MPP on  $\mathcal{X} = \mathbb{R}^d \times \mathbb{R}_0^+$  is biscale invariant, with index  $D$  ( $D$ -biscale invariant for short), if, for every real constant  $\alpha > 0$ , its fidi distributions are invariant under the biscale transformations  $S_\alpha^{(D)}(x, y)$  of (12.8.3).*

**Proposition 12.8.VII.** *Let  $N$  be an extended MPP on  $\mathbb{R}^d$  with marks in  $\mathbb{R}_0^+$ .*

- (a) *In order that (12.8.10) should define a valid (boundedly finite) random measure  $\xi$  on  $\mathbb{R}^d$ , it is necessary and sufficient that for all bounded  $A \in \mathcal{B}_{\mathbb{R}^d}$ , the integral (12.8.8) should converge a.s. If  $N$  has boundedly finite expectation measure  $M(\cdot \times \cdot)$ , then it is sufficient that  $M$  should satisfy conditions (10.1.5a) and (10.1.5b).*
- (b) *If  $N(\cdot)$  is stationary (i.e., invariant under shifts in its first argument), then  $\xi$  is stationary.*
- (c) *If  $N$  is  $D$ -biscale invariant, then  $\xi$  is  $D$ -self-similar.*
- (d) *If the expectation measure  $M$  exists and (b) is satisfied, then there exists a  $\sigma$ -finite measure  $\phi$  on  $\mathbb{R}_0^+$  such that  $M = \ell \times \phi$ , where  $\ell$  is Lebesgue measure on  $\mathbb{R}^d$ . If also (c) is satisfied, with index  $D$ , then  $\phi$  reduces to the power-law with density*

$$\phi(\kappa) = c\kappa^{-(1+1/D)} \quad (c > 0, 1 < D < \infty). \quad (12.8.11)$$

PROOF. Provided it is understood that  $K \in \mathcal{B}_{\mathbb{R}_0^+}$  is bounded away from both 0 and  $\infty$ , it follows from Proposition 9.1.VIII that

$$\zeta(A \times K) = \int_{A \times K} \kappa N(dt \times d\kappa)$$

defines a boundedly finite random measure  $\zeta$  on  $\mathbb{R} \times \mathcal{K}$ . Then a.s. convergence of the integral (12.8.10) to a finite value is enough to ensure that  $\xi$ , the ground measure for  $\zeta$  (set  $K = \mathcal{K}$ ), is also a boundedly finite random measure.

If the expectation measure  $M$  for  $N$  exists, and  $A$  is bounded, then condition (10.1.5a) implies that the integral  $\int_0^\epsilon \kappa N(A \times d\kappa)$  converges a.s., and condition (10.1.5b) implies that  $N(A \times (\epsilon, \infty))$  is a.s. finite, the two together being sufficient to imply the convergence a.s. of (12.8.10).

Assertion (b) is an easy consequence of the definitions of stationarity, and (c) follows from the following equations, where we first consider the distribution of  $\mu(A)$  for some bounded set  $A \in \mathcal{B}_{\mathbb{R}^d}$ :

$$\begin{aligned} R_\alpha^{(D)}\xi(A) &= \int_0^\infty \kappa \alpha^{-D} N(\alpha A \times dy) \\ &= \int_0^\infty \kappa N(\alpha A \times \alpha^D d\kappa) \\ &= \int_0^\infty \kappa N(A \times d\kappa) = \xi(A), \end{aligned}$$

in which the last line follows from the assumed biscale invariance of  $N$ . Thus the one-dimensional distributions of  $\xi$  are invariant under  $R_\alpha^{(D)}$ . For all  $r > 1$ , similar arguments can be applied to the  $r$ -dimensional distributions of  $\xi(A_1), \dots, \xi(A_r)$  for bounded Borel sets  $A_1, \dots, A_r$ . Such fidi distributions are sufficient to determine the distributions of  $\xi$  and  $R_\alpha^{(D)}\xi$  completely, so it follows that the two random measures must be equal in distribution.

Finally, (d) follows on taking expectations through the equations expressing invariance of  $N$  under shifts in its first component and under biscale invariance, and repeating the argument leading to (12.8.8).  $\square$

The representation implies that, as in the Poisson case of Proposition 12.8.I, the power-law form (12.8.11) is the only possible form for the stationary mark distribution, even though it is unbounded and cannot be normalized to form a probability distribution. Of course, over any lower threshold  $\kappa_0 > 0$ , it can be normalized to form a Pareto distribution with distribution function

$$F(\kappa \mid \kappa_0) = 1 - (\kappa/\kappa_0)^{-(1/D)} \quad (\kappa \geq \kappa_0).$$

Our main interest now is in exhibiting random measures of the above type that are both self-similar and stationary. The stable processes form one such example: constructing additional examples is not a trivial exercise. We attempt this only for processes in one dimension ( $\mathcal{X} = \mathbb{R}$ ), where we can specify the model via its conditional intensity function (see Section 7.3 and the more extended discussion in Chapter 14). Because we are concerned with processes with an infinite past, the appropriate version of the conditional intensity is the complete conditional intensity,  $\lambda^*(t, \kappa)$ , representing the current risk (of a point in  $[t, t + dt)$  with mark in  $[\kappa, \kappa + d\kappa)$ ) given the whole past back to  $-\infty$ . The next result gives necessary conditions which must be satisfied by the complete conditional intensity if the underlying point process is to be stationary and  $D$ -biscale invariant. In the present context it is desirable to reflect the dependence on the past explicitly in the notation, so anticipating what we use in Section 14.7, we therefore write

$$\lambda^*(t, \kappa) = \psi_t(N_{t-}, \kappa),$$

where  $\psi_t$  is a functional of the point process realization  $N_-t-$  on  $(-\infty, t)$  and the mark  $\kappa$ . We also use  $S_\alpha^{(D)}\mathcal{H}_t$  to denote the past at time  $t$  of a transformed process where, taking  $t$  as the time origin, the times of past events in the original process are inflated by a factor  $\alpha$  and their marks by a factor  $\alpha^D$ . Similarly, we use  $T_\tau\mathcal{H}_t$  to denote the past at time  $t$  of a version of the original process shifted through  $\tau$ .

**Lemma 12.8.VIII.** *Let  $N$  be an extended MPP with state space  $\mathbb{R} \times \mathbb{R}_0^+$ , and complete conditional intensity function*

$$\lambda^*(t, \kappa) dt d\kappa = E[N(dt \times d\kappa) \mid \mathcal{H}_{t-}] = \psi_t(N_{t-}, \kappa) dt d\kappa.$$

(a) *If  $N$  is stationary then for all real  $t$ ,*

$$\psi_t(N_{t-}, \kappa) = \psi_0(S_t N_{t-}, \kappa) \equiv \psi(S_t N_{t-}, \kappa)$$

*is independent of  $t$  so that  $\lambda^*(t, \kappa) = \psi(S_t N_{t-}, \kappa)$  for all  $t > 0$ .*

(b) *If  $N$  is also  $D$ -biscale invariant, then for all real  $\alpha > 0$ ,*

$$\psi(S_\alpha^{(D)} N_{0-}, \alpha^D \kappa) = \alpha^{-(1+D)} \psi(N_{0-}, \kappa),$$

*and for all  $t$ ,  $\alpha > 0$ ,*

$$E[N(d(\alpha t) \times d(\alpha^D \kappa)) \mid R_\alpha^{(D)}\mathcal{H}_t] = \psi(S_t N_{t-}, \kappa) dt d\kappa = \lambda^*(t, \kappa) dt d\kappa.$$

PROOF. Conditions (a) and (b) are to be understood as equalities of functionals of the infinite past, suitably adjusted where appropriate. Thus, condition (a) means that if there are previous occurrences at  $\{t_i: t_i < t\}$ , and the times are shifted so that these become  $\{t_i + \tau: t_i < t\}$ , then the value of the conditional intensity for the shifted process at time  $t + \tau$  coincides with the value of the conditional intensity for the original process at time  $t$ . Now, under the assumptions, the conditional intensities can be expressed in terms of the fidi distributions and vice versa, so the statement is equivalent to equality of the fidi distributions under shifts and hence is a consequence of stationarity. It can be satisfied only if the conditional intensity depends on the past occurrence times through the differences  $t - t_i$  and not on the absolute values  $t_i$ . It is a necessary condition (but not sufficient) for the conditional intensity itself to be a stationary process in time.

Similarly, to justify condition (b), consider a simultaneous inflation of the time scale (from origin  $t = 0$  back into the past) by a factor  $\alpha$ , and of the mark scale by a factor  $\alpha^D$ . If the underlying process is  $D$ -biscale invariant, the conditional intensity at  $t = 0$  for the inflated process must have the same value as the conditional intensity for the original process at  $t = 0$ , yielding the condition (b) for  $t = 0$ . It implies that, as a function of past events, the conditional intensity at time  $t = 0$  must be a function of the ratios  $t_i/\kappa_i^{1/D}$ . Because by assumption the underlying process is also stationary, a similar

condition must hold for all  $t$ . Note that although the left-hand side of the last equation in condition (b) refers to a conditional intensity at time  $\alpha t$  it cannot be equated with  $\lambda^*(\alpha t)$  [as wrongly asserted in Vere-Jones (2005)] because the conditioning histories are different.  $\square$

Conditions (a) and (b) are both satisfied when the conditional intensity can be expressed in the form

$$\lambda^*(t, \kappa) = \kappa^{-(1+1/D)} h \left( \left\{ \frac{\kappa}{\kappa_i}, \frac{(t - t_i)^D}{\kappa_i} \right\} \right), \quad (12.8.12)$$

where  $h$  is a function of the infinite set of pairs of arguments  $(\kappa/\kappa_i, (t-t_i)^D/\kappa_i)$  involving the times and marks  $t_i, \kappa_i$  of past events (i.e., with  $t_i < t$ ). To understand the form of the arguments of  $h$ , note that, in general, the complete conditional intensity should be a function of  $(t, \kappa)$  and the infinite set of pairs  $(t_i, \kappa_i)$ . Setting  $\tau = -t$  in condition (a) shows that the arguments of  $h$  can be reduced to  $\kappa$  and the pairs  $(t - t_i, \kappa_i)$ . Then setting  $\alpha = 1/\kappa$  and using (b) the arguments reduce to the pairs  $((t - t_i)/\kappa^{1/D}, \kappa_i/\kappa)$ , which is equivalent to the form in (12.8.12). Note that the initial term  $\kappa^{-(1+1/D)}$  arises from the inflation of the infinitesimal elements  $dt d\kappa$ .

Unfortunately, the conditions of the lemma provide no guarantee that in any particular case a process with the proposed form of conditional intensity function exists, or, if it does exist, that it is uniquely specified and inherits the invariance properties of the conditional intensities. To illustrate the latter point, consider a Hawkes process, with conditional intensity as set out in Example 7.2(b), but with criticality constant  $\nu \geq 1$ . The proposed conditional intensity satisfies condition (a), but the only corresponding point process is explosive and does not admit any stationary version.

In developing a potential model, therefore, it is necessary to check two points: that the proposed conditional intensity satisfies the conditions of the lemma, and that a process with this conditional intensity exists and is both stationary and self-similar. The stable processes correspond to the choice  $h \equiv \text{const}$ . The next example demonstrates that the proposed class of processes is not limited to the stable processes.

**EXAMPLE 12.8(b).** *Self-similar ETAS model.* Recall from Examples 6.4(d) and 7.3(b) that the standard ETAS model has conditional intensity of the form, for  $M > M_0$ ,  $p > 0$ ,

$$\lambda^*(t, M) = \beta e^{-\beta(M_i - M_0)} \left\{ \mu_c + A \sum_{i: t_i < t} e^{\alpha(M - M_0)} \frac{p c^p}{(c + t - t_i)^{1+p}} \right\}, \quad (12.8.13)$$

where the condition for stability (existence of a stationary version) is that

$$\rho = A \int_{M_0}^{\infty} e^{\alpha(M - M_0)} \beta e^{-\beta(M_i - M_0)} dM = \frac{A\beta}{\beta - \alpha} < 1. \quad (12.8.14)$$



The mark here is in the logarithmic (magnitude) scale. To find a self-similar variant, we should convert back to the original (energy) scale  $\kappa = e^M$ , and investigate the effect of lowering the magnitude cut-off  $M_0 = \log \kappa_0$ . In this formulation, the conditional intensity for the ETAS model takes the form

$$\lambda^*(t, \kappa) = \frac{\beta}{\kappa_0} \left( \frac{\kappa}{\kappa_0} \right)^{-(1+\beta)} \left\{ \mu + \sum_{i:t_i < t} A \left( \frac{\kappa_i}{\kappa_0} \right)^\alpha \frac{p c^p}{(c + t - t_i)^{1+p}} \right\}. \quad (12.8.15)$$

This is reminiscent of the structural form (12.8.12), but several modifications are needed before it can represent the complete intensity of a stationary, biscale invariant process. The essential steps are the following.

- 1°. The threshold  $\kappa_0$  should disappear. To make this possible we need at least to set  $\alpha = \beta$ .
- 2°. This now violates the stability condition (12.8.14); to ensure the existence of a suitable point process we replace the factor  $\kappa_i^\alpha$  inside the summation by the *stabilizing factor*

$$S \left( \frac{\kappa}{\kappa_i} \right) = \left[ \min \left( \frac{\kappa}{\kappa_i}, \frac{\kappa_i}{\kappa} \right) \right]^{\delta/2} \quad (\delta > 0). \quad (12.8.16)$$

- 3°. To ensure that all terms have a self-similar form, we replace the constant  $c$  in (12.8.13) by the function  $c(\kappa_i) = C \kappa_i^{1/D}$  of the size of the generating event at  $t_i$ .

The resulting complete intensity can be written in the form

$$\lambda^*(t, \kappa) = \kappa^{-(1+1/D)} \left\{ \mu + \eta \sum_{i:t_i < t} \left[ 1 + \frac{t - t_i}{C \kappa_i^{1/D}} \right]^{-(1+p)} S \left( \frac{\kappa}{\kappa_i} \right) \right\}. \quad (12.8.17)$$

The quantities  $\mu$ ,  $\eta$ ,  $C$ ,  $p$ ,  $\delta$ , and  $D$  are positive constants and constitute the parameters of the model. The parameters  $\mu$  and  $\eta$  are rates per unit time and energy,  $C$  has units time/energy $^{1/D}$ , and the remaining parameters are dimensionless. The compatibility with (12.8.12) is now evident, but it is not yet clear whether there exists a well-defined process with this form of complete intensity.

To clarify this point, we revert to the cluster process interpretation of the Hawkes process (see Section 6.3). To accommodate the mark structure, the state space should be changed from  $\mathbb{R}$  to  $\mathbb{R} \times \mathbb{R}_0^+$ . Our candidate model then has cluster centres forming a Poisson process on  $\mathbb{R} \times \mathbb{R}_0^+$ , with mark-time intensity  $\mu \kappa^{-(1+1/D)}$ , and the cluster members from a parent at  $(t_i, \kappa_i)$  are the total offspring, from all generations, of a branching process in which first generation offspring form an independent Poisson process with intensity

$$\theta(t, \kappa \mid t_i, \kappa_i) = \frac{\eta}{\kappa^{1+1/D}} \left[ 1 + \frac{t - t_i}{C \kappa_i^{1/D}} \right]^{-(1+p)} S \left( \frac{\kappa}{\kappa_i} \right) \quad (t > t_i, \kappa > 0),$$

and the offspring from this and all later generations independently follow the same Poisson process relative to their own parent. Using this formulation, we can make use of the general criterion for the existence of a Poisson cluster process at Proposition 6.3.III, namely, the convergence, for each bounded Borel set  $B \in \mathcal{X} \equiv \mathbb{R} \times \mathbb{R}_0^+$  (the set  $B$  should also be bounded away from 0 in  $\mathcal{K} = \mathbb{R}_0^+$ ), of the integral

$$\int_{\mathcal{X}} \Pr\{N(B \mid x) > 0\} \mu_c(dx), \quad (12.8.18)$$

where  $N(\cdot \mid x)$  is the cluster member process from a cluster centre at  $x$  and  $\mu_c(\cdot)$  the expectation measure for the process  $N_c(\cdot)$  of cluster centres.

In addition to being nonnegative, the essential characteristic of the kernel  $\theta$  is that it should admit the function

$$\psi(t', \kappa') = (\kappa')^{-(1+1/D)}$$

as the eigenfunction corresponding to a positive (and hence maximum) eigenvalue  $\rho < 1$ . To clarify its behaviour in this regard, it is convenient to rewrite  $\theta$  in the more general form

$$\theta(t, \kappa \mid t', \kappa') = \rho f(t - t', \kappa') P(\kappa, \kappa') \left(\frac{\kappa}{\kappa'}\right)^{-(1+1/D)}, \quad (12.8.19)$$

where  $f$  is normalized to be a probability density function in  $u = t - t'$  and  $P$  is normalized to be a Markov transition kernel in  $\kappa'$ . Straightforward computations show that this is achieved in the present instance by setting

$$\begin{aligned} f(t - t', \kappa') &= \frac{p}{C(\kappa')^{1/D}} \left[ 1 + \frac{t - t'}{C(\kappa')^{1/D}} \right]^{-(1+p)}, \\ P(\kappa, \kappa') &= \frac{\delta}{\kappa'} \left[ \min \left( \frac{\kappa}{\kappa'}, \frac{\kappa'}{\kappa} \right) \right]^{\delta/2}, \\ \rho &= \eta C / \delta p. \end{aligned}$$

For a branching process with Poisson located offspring, the first generation of offspring from an ancestor at  $(t_c, \kappa_c)$  follows a Poisson process with intensity  $\theta(t, \kappa \mid t_c, \kappa_c)$ , the second generation follows a Poisson process with intensity

$$\theta^{(2)}(t, \kappa \mid t_c, \kappa_c) = \int_{\mathcal{X}} \theta(t, \kappa \mid t', \kappa') \theta(t', \kappa' \mid t_c, \kappa_c) dt' d\kappa',$$

and in general the  $k$ th generation follows a Poisson process with intensity given by the  $k$ th iterate of  $\theta$ , say  $\theta^{(k)}(\cdot \mid \cdot)$ . Then, treating  $t_c$  and  $\kappa_c$  as the coordinates of a cluster centre (independent arrival) and writing  $B$  for a Borel

subset of  $\mathbb{R} \times \mathbb{R}_0^+$  that is both bounded and bounded away from 0 in its second argument, we can estimate the integral by

$$\begin{aligned} \int_{\mathcal{X}} \Pr\{N(B \mid x) > 0\} \mu_c(dx) &\leq \int_{\mathcal{X}} \sum_k \Pr\{N_k(B \mid x) > 0\} \mu_c(dx) \\ &\leq \int_{\mathcal{X}} \sum_k \mathbb{E}[N_k(B \mid x)] \mu_c(dx) \\ &= \mu \sum_k \int_{\mathbb{R}} \int_{\mathbb{R}_0^+} \theta^{(k)}(B \mid t_c, \kappa_c) \psi(t_c, \kappa_c) dt_c d\kappa_c, \end{aligned}$$

where  $N_k(\cdot \mid x)$  is the process of  $k$ th generation offspring,  $\theta^{(k)}(B \mid \cdot) = \int_B \theta^{(k)}(t, \kappa \mid \cdot) dt d\kappa$ , and the intensity  $\mu_c(dx)$  of cluster centres has been replaced by the specific form  $\mu \psi(t_c, \kappa_c) dt_c d\kappa_c$ . But because  $\psi(t, \kappa)$  is an eigenfunction, this last sum reduces to

$$M(B) = \mu \sum_{k=0}^{\infty} \rho^k \left( \int_B \kappa^{-(1+1/D)} dt d\kappa \right).$$

If  $\rho < 1$  and  $B$  is bounded away from 0 on the  $\kappa$  axis, this sum is certainly finite, and then represents the expected number of points falling into  $B$ , namely,

$$M(B) = \frac{\mu}{1-\rho} \int_B \kappa^{-(1+1/D)} dt d\kappa. \quad (12.8.20)$$

The argument shows that, although the number of cluster members is infinite in total, most are very small, and only a finite number fall into a bounded set  $B$  bounded away from 0 on the energy axis. A similar statement holds for the overall process.

Finally, given that such a cluster process exists, and has the above mean rate, it follows that the series defining  $\lambda^*(t, \kappa)$  converges almost surely, and then represents the total risk of an event in  $dt \times d\kappa$ , given the locations of all points with  $t_i < t$ , that is, given the complete history of the process up to time  $t$ .  $\square$

The model can be modified and extended in various ways. In particular, the power-law form for the density function  $f$  of time-delays is not an inevitable feature of self-similarity. It can just as well be replaced by an exponential (short-tailed) form without affecting either stationarity or self-similarity. An outline is given in Exercise 12.8.4. Thus there is no necessary connection between self-similarity and long-range dependence, as there is in the case of the fractional Brownian motions. On the other hand, self-similarity does imply power-law growth in some sense, as is apparent from the very definition. However, because the moments are infinite, this sense cannot be expressed in terms of the rate of growth of the moment functions.

A related and important approach to self-similarity for random measures is developed in the papers by Zähle already quoted. Zähle suggests basing the property, not on absolute locations relative to the state space  $\mathcal{X}$ , but on locations relative to a given point of the realization, that is, on the Palm distributions of the process. This allows the treatment of some examples, such as Lévy dust [Example 9.1(g)] which lie outside the more restricted development in the text above. Zähle gives a general algebraic treatment, and develops many important properties of self-similar random measures, such as the dimensionality of the set of atoms, but he does not examine the probabilistic structure of the atoms from the point of view considered in this section.

### Exercises and Complements to Section 12.8

- 12.8.1 Show that the renormalization group  $R_\alpha^{(D)}$  acts boundedly and continuously on  $\mathcal{M}_\mathcal{X}^\#$ . [Hint: For suitable  $f$ ,  $\int f(x) R_\alpha^{(D)} \mu(dx) = \alpha^{-D} \int f(y/\alpha) \mu(dy)$ . Now imitate the proof of continuity of the shifts  $S_x$  in Lemma 12.1.I.]
- 12.8.2 Develop a representation for self-similar completely random measures in  $\mathbb{R}^2$  analogous to that set out in Proposition 12.8.IV for such measures in  $\mathbb{R}$ . Then consider the simplifications which occur on assuming that the random measure is in addition (a) homogeneous (i.e., stationary with respect to shifts in  $\mathbb{R}^2$ ), (b) isotropic, or (c) both. [Hint: Consider the effects on the intensity of the Poisson process in the representation (10.1.4).]
- 12.8.3 Let  $N$  be a self-similar extended MPP on  $\mathbb{R}^d \times \mathbb{R}_+$ ; transform the mark-scale by setting  $q = \log \kappa$ . Show that  $D$ -self-similarity is equivalent to requiring the transformed process  $N^*$  on  $\mathbb{R}^d \times \mathbb{R}$  to have fidi distributions that are invariant under the transformations

$$E_\alpha^{(D)}(A \times Q) = \alpha A \times (S_{D \log \alpha} Q).$$

Restate Proposition 12.8.VII and Lemma 12.8.VIII in terms of the transformed process  $N^*$ .

- 12.8.4 Investigate in detail the properties of a version of the self-similar ETAS model where the normalized density for the time delay in (12.8.19) has the exponential form  $f(t - t_i) = c_\kappa e^{c_\kappa(t - t_i)}$ . In particular check that the form (12.8.12) can be sustained, and that the existence of a stationary process can be established as in Example 12.8(b).
- 12.8.5 *Long-range dependence of self-similar ETAS model.*
- (a) In the model of Example 12.8(b), show that the first- and higher-order moment measures of self-similar random measures do not exist, so that long-range dependence in the sense of Section 12.7 cannot be defined.
  - (b) Investigate conditions for long-range dependence of the process  $N_K$  restricted to any mark set  $K$  bounded away from 0 and  $\infty$ .

## CHAPTER 13

# Palm Theory

13.1	Campbell Measures and Palm Distributions	269
13.2	Palm Theory for Stationary Random Measures	284
13.3	Interval- and Point-stationarity	299
13.4	Marked Point Processes, Ergodic Theorems, and Convergence to Equilibrium	317
13.5	Cluster Iterates	334
13.6	Fractal Dimensions	340

In Section 3.4 we gave a brief introduction to Palm–Khinchin equations and noted that, for a stationary point process on the line, they provide a link between counting and interval properties. In this chapter we study this link both in more detail and in a more general setting. It is a topic that continues to find new applications, both within point process theory itself, and in the applications of that theory to ergodic theory, queueing theory, stochastic geometry and many other fields. Its continuing relevance is linked to the shift of viewpoint that it entails: from an absolute frame of reference outside the process under study, to a frame of reference inside the process (meaning, for a point process, relative to a point of the process). Such a change of viewpoint is usually insightful, and sometimes essential, in seeking an understanding of point process properties.

Early contributions by Palm (1943) and Khinchin (1955) have already been noted in Chapter 3. Subsequently, the general theme was taken up by Kaplan (1955), who was influenced by Doob’s (1948) work on renewal processes, and Slivnyak (1962, 1966). This work examined point processes in  $\mathbb{R}$  with the property of interval-stationarity; successful extension of this idea to point processes in  $\mathbb{R}^d$  has been much more recent.

A critical development in the study of stationary random measures and point processes was the formulation by Kummer and Matthes (1970) of what they called Campbell measure, in essence a refinement of the Radon–Nikodym approach that Ryll–Nardzewski (1961) and Papangelou (1970, 1974a) used earlier. The later evolution of their work can be traced through the three editions—in German, English, and Russian—of Matthes, Kerstan, and Mecke,

referred to as MKM (1974, 1978, 1982). The relation with ergodic theory (the theory of flows and of flows under a function) was studied by Neveu (1968, 1976), Papangelou (1970), and Delasnerie (1977). Baccelli and Brémaud (1994) exploit the links between material in this chapter and the martingale approach outlined in Chapter 14, while Sigman (1995) and Thorisson (2000) use shift-coupling arguments in an alternative approach to the main limit theorems and their applications in queueing theory and elsewhere. More recently, Thorisson (2000), Timár (2004), Heveling and Last (2005), and others, have shown how to extend the concept of interval-stationarity for a point process in  $\mathbb{R}$  to a more general concept of point-stationarity in  $\mathbb{R}^d$  for  $d \geq 2$ , although it dates back at least to Mecke (1975).

The theory has many applications, notably in queueing theory, where work was initiated by König and Matthes (1963); see also König, Matthes, and Nawrotski (1967) and Franken (1975), Franken *et al.* (1981), and Brandt, Franken, and Lisek (1990), among many others. Related applications in stochastic geometry are presented in Stoyan and Mecke (1983) and, more profusely, in Stoyan, Kendall, and Mecke (1987, 1995) [SKM (1987, 1995) below].

In our discussion, which has been strongly influenced by MKM (1978), the major emphasis concerns the stationary case. The main results are derived by a factorization of the Campbell measure, which parallels the factorization of the moment measures given in Section 12.6. Indeed, the reduced moment measures reappear in this chapter as multiples of the moment measures of the Palm distribution.

The definition of Campbell measure and a brief account of the Radon–Nikodym approach is given in Section 13.1. The main results for stationary random measures are set out in Section 13.2, and Section 13.3 develops the basic relationships between stationarity of the measure, and stationarity relative to points of the process. This includes the interpretation of the Palm distribution as the distribution ‘conditional on a point at the origin,’ the equivalence between stationarity of the point process and stationarity of the intervals for a one-dimensional point process, and its recent extensions to point-stationarity in higher dimensions. Ergodicity and convergence to equilibrium from the Palm distribution are discussed in Section 13.4, which also outlines extensions to MPPs. Section 13.5 gives the discussion of cluster iterates deferred from Chapter 11, and Section 13.6 looks at an interpretation of fractal dimensions in terms of moments of the Palm distribution.

## 13.1. Campbell Measures and Palm Distributions

For any random measure  $\xi$ , including possibly a point process, on the c.s.m.s.  $\mathcal{X}$ , we introduce a measure  $C_{\mathcal{P}}(\cdot \times \cdot)$  on the product space  $\mathcal{W} \equiv \mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}$  by setting, for  $A \in \mathcal{B}_{\mathcal{X}}$ ,  $U \in \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ , and  $\mathcal{P}$  the distribution of  $\xi$  on  $\mathcal{M}_{\mathcal{X}}^{\#}$ ,

$$C_{\mathcal{P}}(A \times U) = \mathbb{E}[\xi(A)I_U(\xi)] = \int_U \int_A \xi(dx) \mathcal{P}(d\xi). \quad (13.1.1a)$$

It represents a refinement of the first moment measure  $M(A) = C_{\mathcal{P}}(A \times \mathcal{M}_{\mathcal{X}}^{\#})$ , which results when  $U$  is expanded to cover the full space  $\mathcal{M}_{\mathcal{X}}^{\#}$ .

Write  $\mathcal{B}_{\mathcal{W}}$  for the product Borel  $\sigma$ -field  $\mathcal{B}_{\mathcal{X}} \otimes \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ ; that is,  $\mathcal{B}_{\mathcal{W}}$  is generated by all  $A \times U \in \mathcal{W}$  with  $A \in \mathcal{B}_{\mathcal{X}}$  and  $U \in \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ . The set function  $C_{\mathcal{P}}(\cdot)$  is clearly countably additive on such product sets but is totally finite if and only if the first moment measure exists and is totally finite. To see that  $C_{\mathcal{P}}(\cdot)$  is always at least  $\sigma$ -finite, let  $\{A_m\}$  ( $m = 1, 2, \dots$ ) be a sequence of bounded Borel sets covering  $\mathcal{X}$ , and define

$$U_{mn} = \{\xi: \xi(A_m) \leq n\} \quad (n = 1, 2, \dots).$$

Then the inequalities

$$C_{\mathcal{P}}(A_m \times U_{mn}) = \int_{A_m} \int_{U_{mn}} \xi(dx) \mathcal{P}(d\xi) \leq n \mathcal{P}(U_{mn}) \leq n$$

imply that  $C_{\mathcal{P}}$  is certainly finite on each set  $A_m \times U_{mn}$ . These sets cover  $\mathcal{W}$ , because for any given  $(x, \xi) \in \mathcal{W}$  we can select  $A_m \ni x$  and then, because any  $\xi \in \mathcal{M}_{\mathcal{X}}^{\#}$  is a.s. boundedly finite, given  $A_m$  we can find  $n$  such that  $\xi(A_m) \leq n$ , so  $(x, \xi) \in A_m \times U_{mn}$ . It then follows that the set function  $C_{\mathcal{P}}$  extends uniquely to a  $\sigma$ -finite measure on  $\mathcal{B}_{\mathcal{W}}$ . We continue to use  $C_{\mathcal{P}}$  for this extension.

It is also convenient to introduce here the *modified Campbell measure*<sup>1</sup>  $C_{\mathcal{P}}^!(\cdot \times \cdot)$  which plays an important role in the analysis of spatial point processes in Chapter 15. It is defined much as in (13.1.1a) but specifically for simple point processes  $N$ , and with the special feature of excluding the point at the origin which is characteristic of the ordinary Palm distributions. For  $A \in \mathcal{B}_{\mathcal{X}}$  and  $U \in \mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#*})$  and  $\mathcal{P}$  the distribution of  $N$  on  $\mathcal{N}_{\mathcal{X}}^{\#*}$ , we write

$$C_{\mathcal{P}}^!(A \times U) = \mathbb{E} \left[ \int_A I_U(N \setminus x) N(dx) \right], \quad (13.1.1b)$$

where  $N \setminus x$  denotes the realization of  $N$  modified by the removal of any point that there may be at location  $x$  (sometimes  $N \setminus x$  is written loosely as  $N - \delta_x$ ).

**Definition 13.1.I.** (a) *The Campbell measure  $C_{\mathcal{P}}$  associated with the random measure  $\xi$  on the c.s.m.s.  $\mathcal{X}$ , having distribution  $\mathcal{P}$  on  $\mathcal{M}_{\mathcal{X}}^{\#}$ , is the unique extension of the set function defined at (13.1.1a) to a  $\sigma$ -finite measure on  $\mathcal{B}_{\mathcal{W}}$ .*

<sup>1</sup> Kallenberg (1983a, §12.3) uses the term ‘reduced Campbell measure,’ which we avoid here to eliminate any confusion with the term ‘reduced moment measure’ used, for example, onwards from Proposition 12.6.III.

(b) The modified Campbell measure  $C_{\mathcal{P}}^!$  associated with the simple point process  $N$  on the c.s.m.s.  $\mathcal{X}$  with distribution  $\mathcal{P}$  on  $\mathcal{N}_{\mathcal{X}}^{\#*}$ , is the unique extension of the set function defined at (13.1.1b) to a  $\sigma$ -finite measure on  $\mathcal{B}_{\mathcal{W}}$ .

For the remainder of this chapter we deal only with the ordinary Campbell measure and associated Palm measures, leaving until Chapter 15 any results we need from the analogous development of modified Campbell and Palm measures (but see Exercise 13.2.7).

By following the usual route from indicator functions to simple functions and limits of simple functions, the quantity defined initially at (13.1.1a) extends to the following integral form.

**Lemma 13.1.II.** *For  $\mathcal{B}_{\mathcal{W}}$ -measurable functions  $g(x, \xi)$  that are either non-negative or  $C_{\mathcal{P}}$ -integrable,*

$$\begin{aligned} \int_{\mathcal{W}} g(x, \xi) C_{\mathcal{P}}(dx \times d\xi) &= \mathbb{E} \left( \int_{\mathcal{X}} g(x, \xi) \xi(dx) \right) \\ &= \int_{\mathcal{M}_{\mathcal{X}}^{\#}} \int_{\mathcal{X}} g(x, \xi) \xi(dx) \mathcal{P}(d\xi). \end{aligned} \quad (13.1.2)$$

We have already noted the connection between Campbell measure and the first moment measure which results on setting  $U = \mathcal{M}_{\mathcal{X}}^{\#}$  in (13.1.1a); it yields

$$C_{\mathcal{P}}(A \times \mathcal{M}_{\mathcal{X}}^{\#}) = \mathbb{E}[\xi(A)] = M(A)$$

whenever the first moment measure  $M(\cdot)$  exists. The link with Campbell's theorem noted around (9.5.2) follows most easily from (13.1.2). When  $M(\cdot)$  exists, and  $g$  is a function of  $x$  only, (13.1.2) reduces to

$$\mathbb{E} \left( \int_{\mathcal{X}} g(x) \xi(dx) \right) = \int_{\mathcal{X}} g(x) \int_{\mathcal{M}_{\mathcal{X}}^{\#}} \xi(dx) \mathcal{P}(d\xi) = \int_{\mathcal{X}} g(x) M(dx),$$

that is, precisely (9.5.2), of which Campbell's (1909) original result is the special case for a stationary Poisson process. No doubt it was this link with (9.5.2) that Kummer and Matthes (1970) had in mind in coining the term *Campbell measure* for the measure  $C_{\mathcal{P}}$ .

Several further comments should be made concerning this definition. As in Chapter 9, the role of the canonical probability space  $(\mathcal{M}_{\mathcal{X}}^{\#}, \mathcal{B}_{\mathcal{M}_{\mathcal{X}}^{\#}}, \mathcal{P})$  can be replaced by a more general probability space  $(\Omega, \mathcal{E}, \mathcal{P})$  without altering the basic character of the definition, provided only that the probability space is rich enough to support the random measure. In this case the product measurable space  $\mathcal{W}$  above is replaced by the product measurable space

$$(\mathcal{W}^*, \mathcal{B}_{\mathcal{W}^*}) \equiv (\Omega \times \mathcal{X}, \mathcal{E} \otimes \mathcal{B}_{\mathcal{X}}),$$

and the defining property (13.1.1a) of the Campbell measure becomes

$$C_{\mathcal{P}}(A \times U) = \mathbb{E}[\xi(A)I_U] = \int_U \int_A \xi(dx, \omega) \mathcal{P}(d\omega) \quad (U \in \mathcal{E}).$$



In this chapter, as in the last, we develop the basic theory for the canonical probability space. In Chapter 14, however, the more general definition as above is needed when we consider conditioning on random variables external to the point process itself.

A second point to note is that refinements of higher-order moment measures can be defined in a similar way to the Campbell measure itself [see, e.g., Kallenberg (1975, p. 69; 1983a, p. 103)]. For example, a second-order Campbell measure  $C_{\mathcal{P}}^{(2)}$  can be defined on  $\mathcal{X}^{(2)} \times \mathcal{M}_{\mathcal{X}}^{\#}$  by setting

$$C_{\mathcal{P}}^{(2)}(A \times B \times U) = E[\xi(A)\xi(B)I_U(\xi)] \quad (13.1.3)$$

for  $A, B \in \mathcal{B}_{\mathcal{X}}, U \in \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ . Clearly, the second moment measure  $M_2(A \times B)$ , when it exists, appears as the marginal distribution on integrating out  $\mathcal{P}$  (see Exercise 13.1.1).

Finally, observe that the construction is not restricted to measures  $\mathcal{P}$  on  $\mathcal{M}_{\mathcal{X}}^{\#}$  which have total mass one, but can be carried through for any measure  $Q$  on  $\mathcal{M}_{\mathcal{X}}^{\#}$  for which

- (i)  $Q$  is  $\sigma$ -finite, and
- (ii) there exists a suitable family  $\{A_m\}$  covering  $\mathcal{X}$  such that, for all  $(m, n)$ ,  $Q(U_{mn}) = Q(\{\xi: \xi(A_m) \leq n\}) < \infty$  [see below (13.1.1a)].

Such a construction, of  $C_Q$  say, starting from  $C_Q(A \times U) = \int_U \int_A \xi(dx) Q(d\xi)$  with  $A$  and  $U$  as in (13.1.1a), is important (and always possible) for the KLM measures  $\tilde{Q}$  associated with an infinitely divisible random measure (see Exercise 13.1.2). We briefly digress to examine its definition in this more general case, for while it is clear from the construction that  $Q$  determines  $C_Q$  uniquely, the converse is true only if some additional information is given. The situation is summarized in Lemma 13.1.III. A characterization of measures that can appear as Campbell measures, based on Wegmann (1977), is outlined in Exercise 13.1.3.

**Lemma 13.1.III.** *When the measure  $Q$  on  $\mathcal{M}_{\mathcal{X}}^{\#}$  satisfies (i) and (ii) above, the corresponding Campbell measure  $C_Q$  determines  $Q$  uniquely on  $\mathcal{M}_{\mathcal{X}}^{\#} \setminus \{\emptyset\}$ ; it determines  $Q$  uniquely on  $\mathcal{M}_{\mathcal{X}}^{\#}$  if, in particular, either  $Q$  is a probability measure, or  $Q(\{\emptyset\}) = 0$ .*

PROOF. Using the assumptions, choose a bounded set  $A$  within a finite union of the  $A_m$ . Then  $\xi(A)$  is  $Q$ -a.e. finite, and setting  $V_x = \{\xi: \xi(A) \leq x\}$  for arbitrary  $x > 0$ , we can define

$$F_A(x) = Q(V_x) < \infty.$$

Clearly  $F_A(\cdot)$  is the d.f. of a probability distribution when  $Q$  is a probability measure, but will not be so in general. Then

$$C_Q(A \times V_x) = \int_0^x y dF_A(y) = G_A(x) \quad \text{say;}$$

that is, the Campbell measure determines  $G_A(\cdot)$ , a weighted version of  $F_A(\cdot)$ . Furthermore, apart from the value of  $F_A(0+)$ ,  $F_A(\cdot)$  can be recovered from  $C_Q$  via the relation

$$dF_A(x) = x^{-1} dG_A(x).$$

By varying the choice of  $A$  we see that  $C_Q$  determines all the one-dimensional ‘distributions’ of  $Q$ . Analogous arguments apply to the multivariate fidi distributions [recall that the basic sample functions  $\xi(\cdot)$  are  $Q$ -random measures] and show that the Campbell measure determines all the fidi distributions of  $Q$ , hence  $Q$  itself, up to the value of  $Q(\{\emptyset\})$ . This last is clearly determined if  $Q(\mathcal{M}_\mathcal{X}^\#) = 1$ , or if its value is explicitly prescribed.  $\square$

The key to the introduction of Palm distributions in general is the relation between the Campbell measure  $C_\mathcal{P}$  and the first moment measure  $M(\cdot)$ . Whenever  $M(\cdot)$  exists as a boundedly finite measure, for each fixed  $U \in \mathcal{B}(\mathcal{M}_\mathcal{X}^\#)$ ,  $C_\mathcal{P}(\cdot \times U)$  is then absolutely continuous with respect to  $M(\cdot)$ . We can thus introduce the Radon–Nikodym derivative as a  $\mathcal{B}_\mathcal{X}$ -measurable function  $\mathcal{P}_x(U)$  satisfying, for each  $A \in \mathcal{B}_\mathcal{X}$ ,

$$\int_A \mathcal{P}_x(U) M(dx) = C_\mathcal{P}(A \times U), \quad (13.1.4)$$

and  $\mathcal{P}_x(U)$  is defined uniquely up to values on sets of  $M$ -measure zero. Moreover, for each fixed, bounded Borel set  $A$ ,

$$C_\mathcal{P}(A \times U)/M(A) = C_\mathcal{P}(A \times U)/C_\mathcal{P}(A \times \mathcal{M}_\mathcal{X}^\#)$$

is a probability measure on  $\mathcal{M}_\mathcal{X}^\#$ . Just as in the discussion of regular conditional probabilities (see Proposition A1.5.III), it follows that the family  $\{\mathcal{P}_x(U)\}$  can be chosen so that

(A) for each fixed  $U \in \mathcal{B}(\mathcal{M}_\mathcal{X}^\#)$ ,  $\mathcal{P}_x(U)$  is a measurable function of  $x$  that is  $M$ -integrable on bounded subsets of  $\mathcal{X}$ ; and

(B) for each fixed  $x \in \mathcal{X}$ ,  $\mathcal{P}_x(U)$  is a probability measure on  $U \in \mathcal{B}(\mathcal{M}_\mathcal{X}^\#)$ .

We call each such measure  $\mathcal{P}_x(\cdot)$  a *local Palm distribution* for  $\xi$ , and the family of such measures satisfying (A) and (B) the *Palm kernel* associated with  $\xi$ . Then the discussion above implies the following result, in which (13.1.5) follows from (13.1.4) by the usual extension arguments.

**Proposition 13.1.IV.** *Let  $\xi$  be a random measure whose first moment measure  $M$  exists. Then  $\xi$  admits a Palm kernel, that is, a regular family of local Palm distributions  $\{\mathcal{P}_x(\cdot): x \in \mathcal{X}\}$  which are defined uniquely up to values on  $M$ -null sets, and for all  $\mathcal{B}_\mathcal{W}$ -measurable functions  $g$  they are either nonnegative or  $C_\mathcal{P}$ -integrable, and satisfy*

$$\mathbb{E} \left( \int_{\mathcal{X}} g(x, \xi) \xi(dx) \right) = \int_{\mathcal{X} \times \mathcal{M}_\mathcal{X}^\#} g(x, \xi) C_\mathcal{P}(dx \times d\xi) = \int_{\mathcal{X}} \mathbb{E}_x[g(x, \xi)] M(dx), \quad (13.1.5)$$

where

$$\mathbb{E}_x[g(x, \xi)] = \int_{\mathcal{M}_x^\#} g(x, \xi) \mathcal{P}_x(d\xi) \quad (x \in \mathcal{X}).$$

Note that this proposition holds equally for random measures and point processes, nor does it require  $\xi$  to be stationary. When  $\xi$  is stationary, the local Palm distributions become translated versions of a single basic distribution, so that

$$\mathcal{P}_x(S_x U) = \mathcal{P}_0(U) \quad (\ell\text{-a.e. } x). \quad (13.1.6)$$

A more general version of (13.1.6) is given in Section 13.2 where a factorization argument is used and there is no requirement for the existence of first moments. An outline proof in the present setting, using arguments similar to those of Exercise 12.1.9, is sketched in Exercise 13.1.4(a).

We turn to illustrate the nature of local Palm distributions, first for a random measure with density.

**EXAMPLE 13.1(a)** *Palm distributions for a random measure with density.* Suppose that the random measure  $\xi$  on  $\mathbb{R}$  has trajectories with a.s. continuous locally bounded derivatives  $d\xi(x, \omega)/dx = X(x, \omega)$  and that the first moment measure  $M(\cdot)$  has a continuous locally bounded density  $dM(x)/dx = m(x) = \mathbb{E}[X(x)]$ . Here we use  $\omega \in \Omega$  for the probability space, rather than  $\xi \in \mathcal{M}_x^\#$ , merely to avoid ambiguity of notation. In (13.1.5) let  $g(\cdot)$  run through a sequence of functions of the form  $g_n(x, \omega) = h_n(x)I_U(\omega)$ , where  $U \in \mathcal{B}(\Omega)$  is fixed and  $\{h_n(x)\}$  is a sequence of functions converging to  $\delta_{x_0}$  for some  $x_0 \in \mathbb{R}$ , with  $m$  and  $X$  continuous at  $x_0$ . From (13.1.5) we obtain

$$\int_U \mathcal{P}(d\omega) \int_{\mathcal{X}} h_n(x) X(x, \omega) dx = \int_{\mathcal{X}} m(x) h_n(x) \mathcal{P}_x(U) dx.$$

Using the a.s. continuity and local boundedness, the left-hand side converges as  $n \rightarrow \infty$  to

$$\int_U X(x_0, \omega) \mathcal{P}(d\omega),$$

and the right-hand side converges to  $m(x_0) \mathcal{P}_{x_0}(U)$ , where these functions are continuous in  $x$  at  $x_0$  by assumption, so

$$\mathcal{P}_{x_0}(U) = \frac{1}{m(x_0)} \int_U X(x_0, \omega) \mathcal{P}(d\omega). \quad (13.1.7)$$

Thus, the measure  $\mathcal{P}_{x_0}(\cdot)$  appears as a reweighted version of the probability measure  $\mathcal{P}$ , the weight for the particular realization  $\xi(\omega)$  being taken as proportional to the value of the density  $X(x_0, \omega)$  at the chosen point  $x_0$ . Alternatively, if  $Y$  is any random variable defined on the process, and  $\mathbb{E}_{x_0}(\cdot)$  denotes expectations with respect to the Palm distribution at  $x_0$ , (13.1.7) is equivalent to

$$\mathbb{E}_{x_0}(Y) = \frac{\mathbb{E}[X(x_0)Y]}{\mathbb{E}[X(x_0)]}.$$

Some condition such as the assumed continuity in  $x$  at  $x_0$  of  $\mathcal{P}_x(U)$  is essential, as can be shown by counterexamples where the right- and left-limits at  $x_0$  of  $\mathcal{P}_x(U)$  exist but are different [see Leadbetter (1972) for the point process context].  $\square$

Greater interest attaches to random measures that are a.s. purely atomic. In this case the Campbell measure inherits a singular structure from  $\xi$ , in that its support is restricted to the subset of  $\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}$  defined by

$$\mathcal{U} = \{(x, \xi) \in \mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#} : \xi \in \mathcal{M}_{\mathcal{X}}^{\#}, \xi(\{x\}) > 0\}. \quad (13.1.8)$$

For point processes the support is further restricted to the subset of  $\mathcal{X} \times \mathcal{N}_{\mathcal{X}}^{\#}$  defined by

$$\mathcal{V} = \{(x, N) \in \mathcal{X} \times \mathcal{N}_{\mathcal{X}}^{\#} : N \in \mathcal{N}_{\mathcal{X}}^{\#}, N(\{x\}) \geq 1\}. \quad (13.1.9)$$

The relevant properties of  $\mathcal{U}$  and  $\mathcal{V}$  are summarized in the next proposition.

**Proposition 13.1.V.** *Let  $\mathcal{U}$  and  $\mathcal{V}$  be defined by (13.1.8–9).*

- (i)  $\mathcal{U}$  is a Borel subset of  $\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}$  and  $\mathcal{V}$  is a Borel subset of  $\mathcal{X} \times \mathcal{N}_{\mathcal{X}}^{\#}$ .
- (ii) A random measure  $\xi$  with distribution  $\mathcal{P}$  on  $\mathcal{M}_{\mathcal{X}}^{\#}$  is purely atomic if and only its Campbell measure  $C_{\mathcal{P}}$  satisfies  $C_{\mathcal{P}}(\mathcal{U}^c) = 0$ .
- (iii)  $\xi$  is a point process if and only if  $C_{\mathcal{P}}(\mathcal{V}^c) = 0$ .

PROOF. Consider for each  $n > 0$  a partition of  $\mathcal{X}$  into a countable family  $\mathcal{T}_n$  of Borel subsets  $\{A_{nm} : m = 1, 2, \dots\}$  of diameter  $\leq n^{-1}$ , and set

$$V_{nmk} = \{(x, \xi) : x \in A_{nm}, \xi(A_{nm}) \geq 1/k\}.$$

Clearly,  $V_{nmk} \in \mathcal{B}_{\mathcal{X}} \otimes \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ . We assert that

$$\mathcal{U} = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=1}^{\infty} V_{nmk}, \quad (13.1.10)$$

implying *inter alia* that  $\mathcal{U}$  is a measurable subset of  $\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}$ .

To justify (13.1.10), consider any  $(x, \xi) \in \mathcal{U}$ . Because  $\xi(\{x\}) > 0$  there exists  $k'$  such that  $\xi(\{x\}) \geq 1/k'$ . For each  $n$ ,  $x$  belongs to just one element of  $\mathcal{T}_n$ ,  $A_{nm'}$  say, for which  $\xi(A_{nm'}) \geq \xi(\{x\}) \geq 1/k'$ . Hence  $(x, \xi) \in V_{nm'k'}$ . Thus every element in  $\mathcal{U}$  is an element in the right-hand side of (13.1.10).

Conversely, for fixed  $k$ , suppose  $(x, \xi) \in \bigcap_n \bigcup_m V_{nmk}$ , and for each  $n$  let  $A_n(x)$  denote the unique  $A_{nm}$  containing  $x$ . Then

$$\xi(\{x\}) = \lim_{n \rightarrow \infty} \xi(A_n(x)) \geq 1/k,$$

implying that  $(x, \xi) \in \mathcal{U}$ , so (13.1.10) holds as asserted. Part (i) is shown.

To check (ii), let  $\mathcal{T}_n = \{A_{nm} : m = 1, \dots, r_n\}$  now be a dissecting system for an arbitrary bounded  $A \in \mathcal{B}_{\mathcal{X}}$ , and let  $\mathcal{U}_A$  denote the analogue of  $\mathcal{U}$  at

(13.1.10) with these redefined  $\{A_{nm}\}$ . Use the representation of  $\xi$  in terms of the MPP  $N_\xi$  to write, for such bounded  $A$ ,

$$M(A) = E_{\mathcal{P}}[\xi(A)] = E_{\mathcal{P}^*} \left[ \int_{A \times (0, \infty)} \kappa N(dx \times d\kappa) \right] < \infty, \quad (13.1.11)$$

where expectations are written with respect to the measures  $\mathcal{P}$  for  $\xi$  and  $\mathcal{P}^*$  for  $N$  under the one-to-one measurable mapping  $\xi \leftrightarrow N_\xi$  [cf. Proposition 9.1.V(v)]. Set

$$V_{nmk}^* = \left\{ (x, N): \int_{A_{nm} \times (1/k, \infty)} \kappa N(dx \times d\kappa) \geq \frac{1}{k} \right\}.$$

We assert that for each  $n$  and  $k$ ,  $C_{\mathcal{P}^*}(\bigcap_{m=1}^{\infty} (V_{nmk}^*)^c)$  vanishes. Indeed,

$$C_{\mathcal{P}^*} \left( \bigcap_{m=1}^{\infty} (V_{nmk}^*)^c \right) = E_{\mathcal{P}^*} \left[ \int_{\mathcal{X} \times (0, \infty)} \prod_{m=1}^{\infty} [1 - I_{V_{nmk}^*}(x, N)] \kappa N(dx \times d\kappa) \right],$$

and the right-hand side equals

$$E_{\mathcal{P}^*} \left[ \sum_{m=1}^{\infty} \int_{A_{nm} \times (0, \infty)} [1 - I_{V_{nmk}^*}(x, N)] \kappa N(dx \times d\kappa) \right],$$

because for  $x \in A_{nm}$ , only the term involving  $A_{nm}$  can contribute a term different from unity to the infinite product, so that the integral of the product reduces to the sum of integrals as shown. But if  $(x, N) \in V_{nmk}^*$ ,  $I_{V_{nmk}^*}(x, N) = 1$ , so the integrand in that term, and hence its integral, vanishes. If, alternatively,  $(x, N) \notin V_{nmk}^*$ , then there are no atoms in  $A_{nm}$  of mass  $> 1/k$  [i.e.,  $N(A_{nm} \times (1/k, \infty)) = 0$ ], so again the integral vanishes. Thus all terms in the sum vanish, and our assertion is justified.

Let

$$\mathcal{U}_{A,k}^* = \{(x, N): x \in A, N(\{x\} \times (1/k, \infty)) = 1\}.$$

Then, as for (13.1.6), we can show that

$$\mathcal{U}_{A,k}^* = \bigcap_n \bigcup_{m=1}^{r_n} V_{nmk}^*,$$

and it follows from the assertion just proved that

$$C_{\mathcal{P}^*}((\mathcal{U}_{A,k}^*)^c) = C_{\mathcal{P}^*} \left( \bigcup_n \bigcap_{m=1}^{r_n} (V_{nmk}^*)^c \right) = 0.$$

Define  $\mathcal{U}_A^*$  analogously to  $\mathcal{U}_A$  but in terms of the counting measures, that is,  $\mathcal{U}_A^* = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=1}^{r_n} V_{nmk}^*$ , and consider the difference

$$C_{\mathcal{P}^*}(\mathcal{U}_A^*) - C_{\mathcal{P}^*}(\mathcal{U}_{A,k}^*) = E_{\mathcal{P}^*} \left[ \int_{A \times (0, 1/k]} \kappa I_{\mathcal{U}_A^*}(x, N) N(dx \times d\kappa) \right].$$

Letting  $k \rightarrow \infty$ , the difference converges to zero by (13.1.7) and dominated convergence, so that  $C_{\mathcal{P}^*}((\mathcal{U}_A^*)^c) = 0$ , equivalently,  $C_{\mathcal{P}}((\mathcal{U}_A)^c) = 0$ .

Because the space  $\mathcal{X}$  is separable, we can cover it by a countable family of bounded sets  $A_i$  on each of which  $C_{\mathcal{P}}((\mathcal{U}_{A_i})^c) = 0$ . Also, because

$$\mathcal{U}^c = \{(x, \xi) : \xi(\{x\}) = 0\} = \bigcup_i \{(x, \xi) : x \in A_i, \xi(\{x\}) = 0\} = \bigcup_i \mathcal{U}_{A_i}^c,$$

assertion (ii) in the Proposition now follows.

Assertion (iii) follows by a similar (and simpler (!)) argument applied to the space of counting measures.  $\square$

The proposition above shows that for random measures that are a.s. purely atomic, each local Palm measure  $\mathcal{P}_x$  inherits the singular structure of the Campbell measure, in the form of an atom at the point  $x$  selected as a local origin. In the point process case, this leads directly to the interpretation of the Palm distribution  $\mathcal{P}_x$  as a distribution conditional on the occurrence of a point at  $x$ . For a direct approach to this definition, under some assumptions, see Exercise 13.1.9. The stationary case is discussed in detail in Section 13.3.

More generally, for a purely atomic random measure, it follows from the proposition that the only contributions to the local Palm distribution  $\mathcal{P}_x$  come from realizations of  $\xi$  which have an atom at  $x$ . The relationships are most easily explored in the case of a random measure with only finitely many atoms, or equivalently, a finite point process with positive marks, as in the next example. The special case of a nonsimple point process is illustrated in Exercise 13.1.6; the general relationship is sketched in Exercise 13.1.7.

**EXAMPLE 13.1(b) *Purely atomic random measure with finitely many atoms.*** On a state space  $\mathcal{X}$  we suppose given a random measure  $\xi$  that is purely atomic and which a.s. has only finitely many atoms, so  $\xi(\mathcal{X}) < \infty$  and we also assume that  $M(\mathcal{X}) = E[\xi(\mathcal{X})] < \infty$ . Denote  $\mathcal{P}\{\xi \text{ has } n \text{ atoms}\} = p_n$  for some probability distribution  $\{p_n\}$  with  $\sum n p_n = \mu$  denoting the mean number of atoms. Let  $\{x_i : i = 1, \dots, n\}$  denote the locations of the atoms, and  $\{\kappa_i\}$  their masses, supposing that the locations are i.i.d. with distribution  $F(\cdot)$ , and that the masses are independently distributed with distributions  $\Pi(\cdot | x)$  conditional on the locations. We can thus describe a realization  $\xi$  by means of the subset  $\{y_1, \dots, y_n\}$  for some finite integer  $n$  and the pairs  $y_i = (x_i, \kappa_i)$  that are i.i.d. on  $\mathcal{X} \times \mathbb{R}_0^+$  with distribution  $\Psi(dy) = \Psi(d(x, \kappa)) = \Pi(d\kappa | x) F(dx)$ . With this notation and recalling (9.1.4) and Proposition 9.1.III(v), the process can also be identified as an MPP  $N$  on  $\mathcal{X}$  with positive marks, and for  $A \in \mathcal{B}_{\mathcal{X}}$ ,  $\xi(A) = \int_A \xi(dx) = \sum_{x_i \in A} \kappa_i$ .

To identify the Palm kernel for  $\xi$ , consider first the left-hand side of the defining equation (13.1.5), taking the function  $g(x, \xi)$  there to be of product form  $\alpha(x)h(\xi)$ , where  $h$  on  $\mathcal{Y}^\cup$  is defined piecewise as  $h_n(y_1, \dots, y_n)$  on  $\mathcal{Y}^{(n)}$  where it is symmetric in the indices  $(1, \dots, n)$  for positive  $n$ ;  $h$  can be arbitrary

for  $n = 0$  because the integral vanishes when  $\xi(\mathcal{X}) = 0$ . Then

$$\mathbb{E} \left[ \int_{\mathcal{X}} g(x, \xi) \xi(dx) \right] = \sum_{n=1}^{\infty} p_n \int_{\mathcal{Y}^{(n)}} \left[ \sum_{j=1}^n \kappa_j \alpha(x_j) h_n(y_1, \dots, y_n) \right] \prod_{i=1}^n \Psi(dy_i).$$

Because  $h_n$  and the joint distribution are symmetric, this can be rewritten as

$$\mu \int_{\mathcal{Y}} \kappa \alpha(x) \Pi(d\kappa \mid x) F(dx) \int_{\mathcal{Y}^{(n-1)}} \left[ \sum_{j=1}^{n-1} \kappa_j h_{n-1}(y_1, \dots, y_{n-1}) \right] \prod_{i=1}^{n-1} \Psi(dy_i),$$

where  $\mu = \sum_{n=1}^{\infty} n p_n$  denotes the mean number of atoms. Introducing the mean atomic mass  $m(x) = \int_{\mathbb{R}_0^+} \kappa \Pi(d\kappa \mid x)$ , the first moment measure is then given by

$$M(dx) = \mu m(x) F(dx).$$

Inspecting the right-hand side of (13.1.5), define for  $x \in \mathcal{X}, \kappa \in \mathbb{R}_0^+, n \in \mathbb{Z}_+$ ,

$$p_n^* = n p_n / \mu \quad \text{and} \quad \Pi^*(d\kappa \mid x) = \kappa \Pi(d\kappa \mid x) / m(x),$$

and identify  $\mathbb{E}_x[g(x, \xi)]$  as

$$\alpha(x) \sum_{n=1}^{\infty} p_n^* \int_{\mathbb{R}_0^+} \Pi^*(d\kappa \mid x) \int_{\mathcal{Y}^{(n-1)}} \left( \sum_{j=1}^{n-1} \kappa_j \right) h_{n-1}(y_1, \dots, y_{n-1}) \prod_{i=1}^{n-1} \Psi(dy_i).$$

We can now see that (13.1.5) is indeed satisfied if we take for  $\mathcal{P}_x(\cdot)$  on  $\mathcal{Y}^{(n)}$  for  $n = 1, 2, \dots$  the symmetrized version of the measure

$$P_n^*(x; dy_1 \times \dots \times dy_n) = p_n^* \delta_x(dx_1) \Pi^*(d\kappa_1 \mid x_1) \prod_{i=2}^n \Psi(dy_i). \quad (13.1.12)$$

In comparison with the corresponding component of the original measure, the above component is ‘tilted’ in two respects: the distributions of the realizations are weighted by the number of points they contain (a realization with  $n$  points has  $n$  different possibilities of locating a point at a given origin), and the distribution of the mass of the atom at  $x$  is weighted by the mass of original atom at  $x$ .

Exercise 13.1.7 extends this example to a more general setting.  $\square$

Palm distributions for an MPP are best introduced by treating the MPP as a point process on  $\mathcal{X} \times \mathcal{K}$ . Each local Palm distribution  $\mathcal{P}_{(x, \kappa)}$  then represents the behaviour of the process given the occurrence of a point at  $x$  with mark  $\kappa$ . The relation of the local Palm distributions to the Palm distribution of the ground process which we assume has finite first moment measure, can be deduced from the following representation, where the function  $g(\cdot, \cdot)$  of (13.1.5) is taken to have the special form  $g(x, N)$  for  $x \in \mathcal{X}$  and  $N \in \mathcal{N}_{\mathcal{X} \times \mathcal{K}}^{\#}$ :

$$\begin{aligned}
\mathbb{E} \left[ \int_{\mathcal{X} \times \mathcal{K}} g(x, N) N(d(x, \kappa)) \right] &= \int_{\mathcal{X} \times \mathcal{K}} M(d(x, \kappa)) \int_{\mathcal{N}_{\mathcal{X} \times \mathcal{K}}^{\#}} g(x, N) \mathcal{P}_{(x, \kappa)}(dN) \\
&= \int_{\mathcal{X} \times \mathcal{K}} M^g(dx) \Pi(d\kappa | x) \int_{\mathcal{N}_{\mathcal{X} \times \mathcal{K}}^{\#}} g(x, N) \mathcal{P}_{(x, \kappa)}(dN) \\
&= \int_{\mathcal{X}} M^g(dx) \int_{\mathcal{N}_{\mathcal{X} \times \mathcal{K}}^{\#}} g(x, N) \int_{\mathcal{K}} \Pi(d\kappa | x) \mathcal{P}_{(x, \kappa)}(dN) \\
&= \int_{\mathcal{X}} M^g(dx) \int_{\mathcal{N}_{\mathcal{X} \times \mathcal{K}}^{\#}} g(x, N) \bar{\mathcal{P}}_x(dN). \tag{13.1.13}
\end{aligned}$$

In this chain of relations,  $M^g(\cdot) = M(\cdot \times \mathcal{K})$  is the first-moment measure for the ground process, and we have used the disintegration  $M(d(x, \kappa)) = M^g(dx) \Pi(d\kappa | x)$ . The measure  $\bar{\mathcal{P}}_x(\cdot)$  defined at the last step can be interpreted as an ‘average’ local Palm distribution representing the behaviour of the process given the occurrence of a point at  $x$  with unspecified mark. The ground measure of the Palm distribution can now be obtained as the projection of  $\bar{\mathcal{P}}_x(\cdot)$ , which lives on the Borel sets of  $\mathcal{N}_{\mathcal{X} \times \mathcal{K}}^{\#}$ , onto the Borel sets of  $\mathcal{N}_{\mathcal{X}}^{\#}$ . Indeed, denoting this projection by  $\bar{\mathcal{P}}_x^g$  and taking  $g(x, N)$  in (13.1.13) to be a function  $g^*(x, N_g)$  of  $x$  and  $N_g$  only, (13.1.13) reduces to

$$\mathbb{E} \left[ \int_{\mathcal{X}} g^*(x, N_g) N_g(dx) \right] = \int_{\mathcal{X}} M^g(dx) \int_{\mathcal{N}_{\mathcal{X}}^{\#}} g^*(x, N_g) \bar{\mathcal{P}}_x^g.$$

Exercise 13.1.10 gives some further details.

To illustrate this situation, suppose in Example 13.1(b) above, we treat the process not as a random measure  $\xi$  but as an MPP. Then the representation (13.1.12) for the local Palm distribution of  $\xi$  should be replaced by

$$P_n^*(y; dy_1 \times \cdots \times dy_n) = p_n^* \delta_y(y_1) \prod_{i=2}^n \Psi(dy_i) \tag{13.1.14a}$$

for the local Palm distribution for the MPP, and for its ground process,

$$\bar{P}_n^g(x; dx_1 \times \cdots \times dx_n) = p_n^* \delta_x(x_1) \prod_{i=2}^n F(dx_i). \tag{13.1.14b}$$

We conclude this section with a characterization of Palm distributions via Laplace functionals, and give some results that can be deduced via this characterization. Much as at (9.4.18), write  $L[f]$ , where  $f \in \text{BM}_+(\mathcal{X})$ , for the Laplace functional of a random measure with distribution  $\mathcal{P}$ , and  $\{L_x[f]\}$  for the family of Laplace functionals derived from the associated Palm kernel at (13.1.5), so that for  $x \in \mathcal{X}$  and  $f \in \text{BM}_+(\mathcal{X})$ ,

$$L_x[f] = \int_{\mathcal{M}_{\mathcal{X}}^{\#}} \exp \left[ - \int_{\mathcal{X}} f(y) \xi(dy) \right] \mathcal{P}_x(d\xi). \tag{13.1.15}$$



**Proposition 13.1.VI.** *Let  $\xi$  be a random measure with finite first moment measure  $M$  and  $L[f]$ ,  $L_x[f]$  the Laplace functionals associated with the original random measure and its Palm kernel, respectively. Then the functionals  $L[f]$  and  $L_x[f]$  satisfy the relation, for  $f, g \in \text{BM}_+(\mathcal{X})$ ,*

$$\lim_{\varepsilon \downarrow 0} \frac{L[f] - L[f + \varepsilon g]}{\varepsilon} = \int_{\mathcal{X}} g(x) L_x[f] M(dx). \quad (13.1.16)$$

*Conversely, if a family  $\{L_x[f]\}$  satisfies (13.1.16) for all  $f, g \in \text{BM}_+(\mathcal{X})$  and some random measure  $\xi$  with Laplace functional  $L[\cdot]$  and first moment measure  $M(\cdot)$ , then the functionals  $\{L_x[f]\}$  coincide  $M$ -a.e. with the Laplace functionals of the Palm kernel associated with  $\xi$ .*

PROOF. Because the first moment measure exists, a finite Taylor expansion (see Exercise 9.5.8) for  $\varepsilon > 0$  and  $f, g \in \text{BM}_+(\mathcal{X})$  yields

$$L[f + \varepsilon g] = L[f] - \varepsilon E \left( \int_{\mathcal{X}} g(x) \exp \left[ - \int_{\mathcal{X}} f(y) \xi(dy) \right] \xi(dx) \right) + o(\varepsilon). \quad (13.1.17)$$

From (13.1.6),

$$\begin{aligned} E \left( \int_{\mathcal{X}} g(x) \exp \left[ - \int_{\mathcal{X}} f(y) \xi(dy) \right] \xi(dx) \right) \\ = \int_{\mathcal{X}} g(x) M(dx) \int_{\mathcal{M}_{\mathcal{X}}^{\#}} \exp \left[ - \int_{\mathcal{X}} f(y) \xi(dy) \right] \mathcal{P}_x(d\xi) \\ = \int_{\mathcal{X}} g(x) L_x[f] M(dx); \end{aligned}$$

substitution into (13.1.17) followed by rearrangement leads to (13.1.16).

To prove the converse, suppose that (13.1.16) holds for the family of functions  $\{\tilde{L}_x[f]\}$ ; because (13.1.16) holds for all  $g \in \text{BM}_+(\mathcal{X})$ , the measures  $L_x[f]M(dx)$  and  $\tilde{L}_x[f]M(dx)$  coincide, so  $L_x$  and  $\tilde{L}_x$  agree for  $M$ -a.e.  $x$ .  $\square$

The relation at (13.1.16) is useful in identifying the form of the Palm kernel in some simple cases.

EXAMPLE 13.1(c) *The Palm kernel for a Poisson process.* For a Poisson process with parameter measure  $\mu(\cdot)$ , we have from below (9.4.18) that

$$\log L[f] = - \int_{\mathcal{X}} (1 - e^{-f(x)}) \mu(dx).$$

Then

$$\begin{aligned} \frac{dL[f + \varepsilon g]}{d\varepsilon} &= -L[f + \varepsilon g] \frac{d}{d\varepsilon} \left( \int_{\mathcal{X}} (1 - e^{-f(x) - \varepsilon g(x)}) \mu(dx) \right) \\ &= -L[f + \varepsilon g] \left( \int_{\mathcal{X}} g(x) e^{-f(x) - \varepsilon g(x)} \mu(dx) \right) \\ &\rightarrow -L[f] \int_{\mathcal{X}} g(x) e^{-f(x)} \mu(dx) \quad (\varepsilon \rightarrow 0). \end{aligned}$$

This can be put in the form of (13.1.16) via the identification  $M(\cdot) = \mu(\cdot)$ , and

$$L_x[f] = e^{-f(x)} L[f] = L_{\delta_x}[f] L[f], \quad (13.1.18)$$

where on the right-hand side,  $L_{\delta_x}[\cdot]$  denotes the Laplace functional of the degenerate random measure with an atom of unit mass at  $x$  and no other mass.

The interpretation of (13.1.18) is that the local Palm distribution  $\mathcal{P}_x(\cdot)$  coincides with the distribution of the original process except for the addition of an extra point at  $x$  itself to each trajectory. By regarding the local Palm distribution as being conditional on a point at  $x$ , the independence properties of the Poisson process then imply that, apart from a given point at  $x$ , the probability structure of the conditional process is identical to that of the original process.

The relation embodied in (13.1.18) can also be written in the form

$$\mathcal{P}_x = \mathcal{P} * \delta_x, \quad (13.1.19)$$

where  $\delta_x$  denotes a degenerate random measure as in (13.1.18). Equation (13.1.19) is the focus of a characterization of the Poisson process.  $\square$

**Proposition 13.1.VII** [Slivnyak (1962); Mecke (1967)]. *The distribution of a random measure with finite first moment measure satisfies the functional relation (13.1.19) if and only if the random measure is a Poisson process.*

PROOF. The necessity of (13.1.19) has been shown above. For the converse, suppose that  $\mathcal{P}_x$  satisfies (13.1.19). Then from (13.1.16) we obtain

$$\frac{dL[\varepsilon f]}{d\varepsilon} = -L[\varepsilon f] \int_{\mathcal{X}} f(x) e^{-\varepsilon f(x)} M(dx),$$

where  $M$  is the first moment measure, assumed to exist. Using  $\log L[0] = \log 1 = 0$ ,

$$-\log L[f] = \int_{\mathcal{X}} \int_0^1 f(x) e^{-\varepsilon f(x)} d\varepsilon M(dx) = \int_{\mathcal{X}} (1 - e^{-f(x)}) M(dx),$$

so that  $L[f]$  is the Laplace functional of a Poisson process with parameter measure equal to  $M(\cdot)$ .  $\square$

## Exercises and Complements to Section 13.1

13.1.1 (a) Check that modified Campbell measure  $C_{\mathcal{P}}^1(\cdot)$  has a unique extension from (13.1.1b) to sets in  $\mathcal{B}_{\mathcal{X}} \otimes \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ .

(b) Show that the set function  $C_{\mathcal{P}}^{(2)}(\cdot)$  defined at (13.1.3) has a unique extension to a  $\sigma$ -finite measure on  $\mathcal{X} \times \mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}$ . When the second moment measure exists, define a second-order family of local Palm distributions  $\mathcal{P}_{x,y}^{(2)}(U)$  satisfying

$$\int_{A \times B} \mathcal{P}_{x,y}^{(2)}(U) M_2(dx \times dy) = C_{\mathcal{P}}^{(2)}(A \times B \times U).$$

13.1.2 Use the analogue of (13.1.1) to define the Campbell measure  $C_{\tilde{Q}}(\cdot)$  of the KLM measure  $\tilde{Q}$ , using (10.2.8) to establish the  $\sigma$ -finiteness property of  $C_{\tilde{Q}}(\cdot)$  on appropriate subsets of  $\mathcal{X} \times \mathcal{N}_{\mathcal{X}}^{\#}$ . Appeal to Lemma 13.1.III to establish that  $C_{\tilde{Q}}(\cdot)$  determines  $\tilde{Q}$  uniquely.

13.1.3 A measure  $C(\cdot)$  on  $\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}$  is the Campbell measure  $C_{\mathcal{P}}$  of some random measure  $\xi$  with  $\sigma$ -finite first moment measure if and only if the following three conditions hold:

- (i)  $C(A \times \mathcal{M}_{\mathcal{X}}^{\#}) < \infty$  for bounded  $A \in \mathcal{B}_{\mathcal{X}}$ ;
- (ii)  $\int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} g(x, \eta) C(dx \times d\eta) = 0$  whenever  $\int_{\mathcal{X}} g(x, \eta) \eta(dx) = 0$  for each  $\eta \in \mathcal{M}_{\mathcal{X}}^{\#}$ ; and
- (iii)  $1 - \phi_A \equiv \int_{A \times \{\eta: \eta(A) > 0\}} [\eta(A)]^{-1} C(dx \times d\eta) \leq 1$  for bounded  $A \in \mathcal{B}_{\mathcal{X}}$ .

When these conditions hold,  $\inf_A \phi_A = \mathcal{P}\{\xi = \emptyset\}$ .

[Hint: For the converse, define a measure  $\mathcal{P}$  on  $\mathcal{M}_{\mathcal{X}}^{\#}$  by

$$\int_{\mathcal{M}_{\mathcal{X}}^{\#}} f(\eta) \mathcal{P}(d\eta) = \int_{\mathcal{X}} \int_{\mathcal{M}_{\mathcal{X}}^{\#}} k(x, S_{-x}\eta) f(S_{-x}\eta) C(dx \times d\eta),$$

where  $k(\cdot)$  satisfies (13.2.9); then verify that  $C$  and  $C_{\mathcal{P}}$  coincide. See Wegmann (1977) for details.]

13.1.4 (a) Use arguments analogous to those of Exercise 12.1.9 to show that (13.1.6) holds when the process  $N$  is stationary and has boundedly finite first moment measure.

- (b) What can be said about the local second-order Palm measure  $\mathcal{P}_{x,y}^{(2)}(\cdot)$  when the process is stationary?

[Hint: Use Exercise 12.1.8 much as in Exercise 12.1.9.]

13.1.5 Let  $\xi$  be a random measure supported by  $\mathbb{Z} = \{0, \pm 1, \dots\}$  and with finite first moment measure. Describe its Palm and Campbell measures [see (13.1.4)]. When  $\xi$  is a simple point process, reinterpret the Palm measure as a conditional distribution.

13.1.6 Discuss two possible interpretations for the local Palm distributions for a nonsimple point process, the first based on the occurrence of a point at  $x$ , and the second based on the occurrence of a point of given multiplicity at  $x$ . Show that the first can be represented as an average of the second.

13.1.7 Let  $\Psi: \mathcal{M}_{\mathcal{X}}^{\#} \rightarrow \mathcal{N}^{\#}(\mathcal{X} \times \mathbb{R}_0^+)$  denote the mapping of a purely atomic boundedly finite random measure  $\xi$  on  $\mathcal{X}$  into an integer-valued extended random measure  $N_{\xi}$  on  $\mathcal{X} \times \mathbb{R}_0^+$  as in Proposition 9.1.III(v).

- (a) Show that a measure  $\mathcal{P}$  on  $\mathcal{M}_{\mathcal{X}}^{\#}$  induces a measure  $\mathcal{P}^*$  on  $\mathcal{N}^{\#}(\mathcal{X} \times \mathbb{R}_0^+)$  and conversely.
- (b) Let  $C_{\mathcal{P}}$  and  $C_{\mathcal{P}^*}$  denote the corresponding Campbell measures. Show that

$$\begin{aligned} \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} g(x, \xi) C_{\mathcal{P}}(dx \times d\xi) &= E_{\mathcal{P}} \left[ \int_{\mathcal{X}} g(x, \xi) \xi(dx) \right] \\ &= E_{\mathcal{P}^*} \left[ \int_{\mathcal{X} \times \mathbb{R}_0^+} \kappa g(x, \Psi(N)) N(dx \times d\kappa) \right]. \end{aligned}$$

- (c) Writing  $h(x, \kappa, N) = \kappa g(x, \Psi(N))$  as in part (b), and  $\mathcal{Y} = \mathcal{X} \times \mathbb{R}_0^+$ , show that

$$\begin{aligned} \mathbb{E}_{\mathcal{P}^*} \left( \int_{\mathcal{X} \times \mathbb{R}_0^+} h(x, \kappa, N) N(dx \times d\kappa) \right) \\ = \int_{\mathcal{Y} \times \mathcal{N}_{\mathcal{Y}}^{\#}} h(x, \kappa, N) C_{\mathcal{P}^*}(dx \times d\kappa \times dN) \\ = \int_{\mathcal{Y}} M^*(dx \times d\kappa) \mathbb{E}_{(x, \kappa)}[h(x, \kappa, N)], \end{aligned}$$

where  $M^*$  is the first moment measure of the extended MPP on  $\mathcal{Y}$  and

$$\mathbb{E}_{(x, \kappa)}[h(x, N)] = \int_{\mathcal{N}_{\mathcal{Y}}^{\#}} h(x, N) \mathcal{P}_{(x, \kappa)}(dN),$$

where the Palm kernel  $\mathcal{P}_{(x, \kappa)}(\cdot)$  is conditioned both on the location and the mass of a given atom. Simplifications of the relations above depend on particular features in the model as in Example 13.1(b) and (13.1.15).

- 13.1.8 Show that, although the arguments leading to Proposition 13.1.V can fail when  $C_{\mathcal{P}}(A \times \mathcal{M}_{\mathcal{X}}^{\#}) = M(A) = \infty$  for some bounded  $A \in \mathcal{B}_{\mathcal{X}}$ , they can be recovered by introducing a nonnegative function  $h(\cdot)$  such that

$$\int_{A \times \mathcal{M}_{\mathcal{X}}^{\#}} h(x, \xi) C_{\mathcal{P}}(dx \times d\xi) < \infty \quad (\text{bounded } A \in \mathcal{B}_{\mathcal{X}}),$$

and defining a modified Campbell measure  $H(dx \times d\xi) = h(x, \xi) C_{\mathcal{P}}(dx \times d\xi)$ . [Hint: Apply the arguments leading to Proposition 13.1.V to  $H$ , and use the results for  $H$  to derive corresponding results for  $C_{\mathcal{P}}$  itself. A possible function for  $h$  is  $I_B(x)/[1 + \xi(B)]$  for some fixed bounded  $B \in \mathcal{B}_{\mathcal{X}}$ . Similar arguments occur in the discussion following (13.2.4).]

- 13.1.9 Establish conditions for a limit interpretation of the local Palm distributions  $\mathcal{P}_x$  as conditional distributions, given the occurrence of a point in a neighbourhood of  $x$ . [Hint: As in Example 13.1(a), consider the behaviour of functions of the form  $g_n(x, \omega) = h_n(x)I_U(\omega)$ , where  $h_n(x)$  is a  $\delta$ -sequence. First take  $U$  to be the event that a point occurs within a neighbourhood of  $x$ , then a compound event incorporating the occurrence of a point near  $x$  with additional conditions. Find appropriate continuity conditions for the ratio to converge. The stationary case is discussed in detail in Section 13.3; see also Leadbetter (1972).]

- 13.1.10 *Local Palm measures for an MPP and its ground process.* Let  $N \equiv \{(x_i, \kappa_i)\}$  be an element of  $\mathcal{N}_{\mathcal{X} \times \mathcal{K}}^{\#}$ . Denote by  $\psi_g$  the projection of  $N$  onto the corresponding realization  $N_g \equiv \{x_i\}$  of its ground process which has measure  $\mathcal{P}_g$  [for this projection, regard  $N$  as the pair  $(N_g, \mathcal{S})$  where  $\mathcal{S}$  is the ordered sequence of marks  $\{\kappa_i\}$  from the space  $\mathcal{K}^{\infty}$  of such sequences].

- (a) Verify that every probability measure  $\mathcal{P}$  on  $\mathcal{B}(\mathcal{N}_{\mathcal{X} \times \mathcal{K}}^{\#})$  admits the disintegration  $\mathcal{P}(dN) = \mathcal{P}_g(dN_g) \nu(d\mathcal{S} \mid N_g)$ , where  $\nu$  is a regular probability kernel on the product space  $\mathcal{N}_{\mathcal{X}}^{\#} \times \mathcal{K}^{\infty}$ .

- (b) Use this kernel to express formally the link between the Palm measures  $\mathcal{P}_0$  and  $\mathcal{P}_0^g$  of an MPP and its ground process respectively, namely,

$$\int_{\mathcal{N}_{\mathcal{X} \times \mathcal{K}}^\#} g(x, \kappa, N) \mathcal{P}_0(dN) = \int_{\mathcal{N}_{\mathcal{X}}^\#} \mathcal{P}_0^g(dN_g) \int_{\psi_g^{-1}(N_g)} g(x, \kappa, N_g, \mathcal{S}) \nu(d\mathcal{S} \mid N_g).$$

- (c) Check that the measure  $\bar{\mathcal{P}}_x^g$  introduced below (13.1.13) is the marginal distribution, in the above sense, corresponding to the measure  $\bar{\mathcal{P}}_x$ .

### 13.1.11 Higher-order versions of local Palm measure.

- (a) Let the random measure  $\xi$  have boundedly finite second-order moment measure  $M_2$ . Show that for any bounded measurable function  $g(x, y, \xi)$ , the second-order kernel  $\mathcal{P}_{x,y}^{(2)}(\cdot)$  on  $\mathcal{M}_{\mathcal{X}}^\#$ , which exists by Exercise 13.1.1, satisfies

$$\begin{aligned} \int_{\mathcal{M}_{\mathcal{X}}^\#} \mathcal{P}(d\xi) \int_{\mathcal{X}^{(2)}} g(x, y, \xi) \xi(dx) \xi(dy) \\ = \int_{\mathcal{X}^{(2)}} M_2(dx \times dy) \int_{\mathcal{M}_{\mathcal{X}}^\#} g(x, y, \xi) \mathcal{P}_{x,y}^{(2)}(d\xi). \end{aligned}$$

- (b) Show that when  $\xi$  is stationary, there exists a one-parameter kernel  $\check{\mathcal{P}}_x^{(2)}(\cdot)$  such that

$$\mathcal{P}_{x,y}^{(2)}(\cdot) = S_{-x} \check{\mathcal{P}}_{y-x}^{(2)}.$$

- (c) Evaluate explicitly the forms of the kernels  $\mathcal{P}_{x,y}^{(2)}(\cdot)$  and  $\check{\mathcal{P}}_u^{(2)}(\cdot)$  for the simple i.i.d. model of Example 13.1(b).

## 13.2. Palm Theory for Stationary Random Measures

Throughout this section we suppose that  $\mathcal{X} = \mathbb{R}^d$  and that the random measure  $\xi$  is stationary (Definition 12.1.II), or, equivalently, that its distribution  $\mathcal{P}$  is invariant under the transformation  $\xi \mapsto S_u \xi$  (all  $u \in \mathcal{X}$ ). Recall also the notation  $\hat{S}_u \mathcal{P}$  in (12.1.3) for the transformation on distributions induced by  $S_u$ :

$$(\hat{S}_u \mathcal{P})(B) = \mathcal{P}(S_u B) \quad (B \in \mathcal{B}(\mathcal{M}_{\mathcal{X}}^\#)).$$

In fact, the only properties of  $\mathbb{R}^d$  that are critical for the discussion are the existence and uniqueness of the invariant measure, various standard results such as that any measure  $\mu \in \mathcal{M}_{\mathcal{X}}^\#$  is determined by its integrals  $\int_{\mathcal{X}} f(x) \mu(dx)$  for bounded nonnegative measurable functions  $f$ , and the factorization Lemma A2.7.II. Thus, the results hold equally when  $\mathcal{X}$  is the circle group  $\mathbb{S}$ , or more generally whenever  $\mathcal{X}$  is an Abelian  $\sigma$ -group as defined in Section A2.7. Explicit treatments from this more general point of view are given by Mecke (1967) and in MKM (1982).

The results below are stated for general random measures, but most of the illustrations concern simple point processes. Extensions to stationary MPPs are given at the beginning of Section 13.4.

We start by investigating the effect of stationarity on the Campbell measure, again using  $\mathcal{W}$  to denote the product space  $\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^\#$ .

**Proposition 13.2.I.** *Let  $\xi$  be a random measure on  $\mathcal{X} = \mathbb{R}^d$  with distribution  $\mathcal{P}$  and  $C_{\mathcal{P}}$  the associated Campbell measure on  $\mathcal{W}$ . Then  $\xi$  is stationary if and only if  $C_{\mathcal{P}}$  is invariant under the group of transformations  $\Theta_u: \mathcal{W} \mapsto \mathcal{W}$  defined for each  $u \in \mathcal{X}$  by*

$$\Theta_u(x, \xi) = (x - u, S_u \xi). \quad (13.2.1a)$$

The interpretation of (13.2.1) is that, if we shift the origin to  $u$ , the Campbell measure of the shifted process relative to  $u$  is the same as the Campbell measure of the original process relative to 0. Note also that by writing  $\hat{\Theta}_u$  for the mapping on measures on  $\mathcal{W}$  induced by  $\Theta_u$ , the assertion of the proposition is expressed more succinctly as

$$\mathcal{P} = \hat{S}_u \mathcal{P} \quad (\text{all } u) \quad \text{if and only if} \quad C_{\mathcal{P}} = \hat{\Theta}_u C_{\mathcal{P}} \quad (\text{all } u). \quad (13.2.1b)$$

PROOF. When  $C_{\mathcal{P}}$  is the Campbell measure associated with  $\mathcal{P}$ , the Campbell measure associated with  $\hat{S}_u \mathcal{P}$  is to be found from [cf. (13.1.2)]

$$\begin{aligned} \int_{\mathcal{X}} \int_{\mathcal{M}_{\mathcal{X}}^{\#}} g(x, \xi) \xi(dx) \mathcal{P}(S_u d\xi) &= \int_{\mathcal{X}} \int_{\mathcal{M}_{\mathcal{X}}^{\#}} g(x, \xi) S_u \xi(dx - u) \mathcal{P}(S_u d\xi) \\ &= \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} g(x, \xi) C_{\mathcal{P}}(\Theta_u(dx \times d\xi)) \\ &= \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} g(x, \xi) \hat{\Theta}_u C_{\mathcal{P}}(dx \times d\xi) \end{aligned}$$

and is thus equal to  $\hat{\Theta}_u C_{\mathcal{P}}$ . Consequently, if  $\mathcal{P}$  and  $\hat{S}_u \mathcal{P}$  coincide, so do  $C_{\mathcal{P}}$  and  $\hat{\Theta}_u C_{\mathcal{P}}$ .

Conversely, if  $C_{\mathcal{P}}$  and  $\hat{\Theta}_u C_{\mathcal{P}}$  coincide, then it follows from Lemma 13.1.III that the probabilities  $\mathcal{P}$  and  $\hat{S}_u \mathcal{P}$  from which they are derived coincide.  $\square$

An alternative criterion, due to Mecke (1975), can be deduced as a consequence of the above result.

**Proposition 13.2.II.**  *$\mathcal{P}$  as in Proposition 13.2.I is stationary if and only if its associated Campbell measure  $C_{\mathcal{P}}$  satisfies*

$$\begin{aligned} \int_{\mathcal{X}} \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} g(x, y, S_y \xi) C_{\mathcal{P}}(dy \times d\xi) dx \\ = \int_{\mathcal{X}} \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} g(y, x, S_y \xi) C_{\mathcal{P}}(dy \times d\xi) dx \end{aligned} \quad (13.2.2)$$

for every nonnegative  $\mathcal{B}_{\mathcal{X}} \otimes \mathcal{B}_{\mathcal{X}} \otimes \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ -measurable function  $g(\cdot)$ .

PROOF. The action of the mappings  $\hat{\Theta}_u$  on a measure  $\Psi$  on  $\mathcal{B}_{\mathcal{X}} \otimes \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$  can be represented by requiring the equations

$$\begin{aligned} \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} h(y, \xi) \hat{\Theta}_u \Psi(dy \times d\xi) &= \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} h(\Theta_{-u}(y, \xi)) \Psi(dy \times d\xi) \\ &= \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} h(y + u, S_{-u} \xi) \Psi(dy \times d\xi) \end{aligned} \quad (13.2.3)$$

to hold for any measurable nonnegative  $h$ .  $\mathcal{P}$  being stationary implies by Proposition 13.2.I that  $C_{\mathcal{P}}$  on  $\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}$  is stationary. Put  $\Psi = C_{\mathcal{P}}$  in (13.2.3), set  $h(y, \xi) = g(x, y, S_y \xi)$  and  $u = -x$ , and then integrate with respect to  $x$  over  $\mathcal{X}$ ; this yields

$$\begin{aligned} \int_{\mathcal{X}} \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} g(x, y, S_y \xi) C_{\mathcal{P}}(dy \times d\xi) dx \\ = \int_{\mathcal{X}} \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} g(x, y - x, S_y \xi) C_{\mathcal{P}}(dy \times d\xi) dx. \end{aligned}$$

But because  $dx$  is the invariant measure on  $\mathcal{X}$ , integration over the whole of  $\mathcal{X}$  for any nonnegative measurable function  $m(\cdot, \cdot)$  satisfies

$$\int_{\mathcal{X}} m(x, y - x) dx = \int_{\mathcal{X}} m(y - x, x) dx.$$

Applying this and Fubini's theorem reduces the right-hand side of the previous equation to

$$\int_{\mathcal{X}} \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} g(y - x, x, S_y \xi) C_{\mathcal{P}}(dy \times d\xi) dx.$$

Then reversing the operation with  $\Theta_u$  gives the right-hand side of (13.2.2).

Conversely, suppose that (13.2.2) holds; let  $j(x)$  be a measurable nonnegative function satisfying  $\int_{\mathcal{X}} j(x) dx = 1$ . Apply (13.2.3) with  $\Psi = C_{\mathcal{P}}$  to general nonnegative  $h$ , multiply by  $j$ , and integrate over  $x$ ; this yields

$$\begin{aligned} \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} h(y, \xi) \widehat{\Theta}_u C_{\mathcal{P}}(dy \times d\xi) \\ = \int_{\mathcal{X}} \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} j(x) h(y + u, S_{-(u+y)} S_y \xi) C_{\mathcal{P}}(dy \times d\xi) dx \\ = \int_{\mathcal{X}} \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} j(y) h(x + u, S_{-(u+x)} S_y \xi) C_{\mathcal{P}}(dy \times d\xi) dx \quad \text{by (13.2.2),} \\ = \int_{\mathcal{X}} \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} j(y) h(x, S_{-x} S_y \xi) C_{\mathcal{P}}(dy \times d\xi) dx \quad \text{by invariance of } x, \\ = \int_{\mathcal{X}} \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} j(x) h(y, \xi) C_{\mathcal{P}}(dy \times d\xi) dx \quad \text{by (13.2.2),} \\ = \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} h(y, \xi) C_{\mathcal{P}}(dy \times d\xi) \quad \text{on integrating out } x. \quad \square \end{aligned}$$

We now look for a product representation of  $\mathcal{W}$  to which Lemma A2.7.II can be applied. Consider the transformation  $D: \mathcal{W} \mapsto \mathcal{W}$  for which  $D(x, \psi) = (x, S_{-x}\psi)$  for  $(x, \psi) \in \mathcal{W}$ . Much as in Exercise 12.1.1, this transformation is continuous and hence measurable. It is also one-to-one and onto, because the inverse mapping  $D^{-1}$  has  $D^{-1}(x, \xi) = (x, S_x \xi)$ . Observe that

$$\Theta_u(D(x, \psi)) = \Theta_u(x, S_{-x}\psi) = (x - u, S_{u-x}\psi) = D(x - u, \psi),$$

so that  $D$  provides a representation of  $\mathcal{W}$  under the actions of the group  $\Theta_u$ . Thus, with  $\tilde{C}_{\mathcal{P}}$  denoting the image of  $C_{\mathcal{P}}$  under  $D$ ,  $\tilde{C}_{\mathcal{P}}$  is invariant under shifts in its first argument.

It may appear now that Lemma A2.7.II should be applicable and thereby yield the required decomposition of  $\tilde{C}_{\mathcal{P}}$ . However, there is a technical difficulty: it is not obvious in general (and may not be true if the first moments are infinite) that  $\tilde{C}_{\mathcal{P}}$  takes finite values on products of bounded subsets of  $\mathcal{W}$ .

To overcome this difficulty, we use the  $\sigma$ -finiteness of  $\tilde{C}_{\mathcal{P}}$  to construct a modified measure, with the same invariance properties, to which we can apply Proposition 13.2.I. Indeed, the  $\sigma$ -finiteness of  $\tilde{C}_{\mathcal{P}}$  implies the existence of a strictly positive function  $h(x, \psi)$  such that, provided  $\mathcal{P}(\{\emptyset\}) < 1$ ,

$$0 < \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} h(x, \psi) \tilde{C}_{\mathcal{P}}(dx \times d\psi) < \infty \quad (13.2.4)$$

(see, e.g., the constructions in Exercises 13.1.8 and 13.2.3). Define  $\alpha(\psi) = \int_{\mathcal{X}} h(x, \psi) dx = \int_{\mathcal{X}} h(x+y, \psi) dy$  (all  $x \in \mathcal{X}$ ), so that  $\alpha(\psi) > 0$ , and let  $g(\cdot)$  be any nonnegative Lebesgue integrable function on  $\mathcal{X}$ . By using the invariance properties of  $\tilde{C}_{\mathcal{P}}$  we obtain

$$\begin{aligned} \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} g(x) \alpha(\psi) \tilde{C}_{\mathcal{P}}(dx \times d\psi) &= \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} \int_{\mathcal{X}} g(x) h(x+y, \psi) dy \tilde{C}_{\mathcal{P}}(dx \times d\psi) \\ &= \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} \int_{\mathcal{X}} g(u-y) h(u, \psi) dy \tilde{C}_{\mathcal{P}}(du \times d\psi) \\ &= \int_{\mathcal{X}} g(u-y) dy \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} h(u, \psi) \tilde{C}_{\mathcal{P}}(du \times d\psi) \\ &< \infty. \end{aligned}$$

The finiteness of the integral on the left-hand side for all such integrable  $g$  shows that the modified measure  $\alpha(\psi) \tilde{C}_{\mathcal{P}}(dx \times d\psi)$  takes finite values on products of bounded sets, and indeed on sets  $A \times \mathcal{M}_{\mathcal{X}}^{\#}$  for bounded  $A$ . Inasmuch as the presence of the multiplier  $\alpha(\psi)$  does not affect invariance of the measure with respect to shifts in  $x$ , the factorization Lemma A2.7.II can still be applied and yields

$$\alpha(\psi) \tilde{C}_{\mathcal{P}}(dx \times d\psi) = {}_{\alpha} \check{C}_{\mathcal{P}}(d\psi) \ell(dx)$$

for some uniquely defined finite measure  ${}_{\alpha} \check{C}_{\mathcal{P}}$  on  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ . We now define the Palm measure being sought by setting

$$\check{C}_{\mathcal{P}}(d\psi) = {}_{\alpha} \check{C}_{\mathcal{P}}(d\psi) / \alpha(\psi).$$

This measure is still  $\sigma$ -finite and satisfies the defining relation

$$\tilde{C}_{\mathcal{P}}(dx \times d\psi) = \check{C}_{\mathcal{P}}(d\psi) \ell(dx).$$

Its properties are spelled out in more detail in Theorem 13.2.III where the relations (13.2.5–6) are referred to as the *refined Campbell theorem*.



**Theorem 13.2.III.** (a) Let  $\xi$  be a stationary random measure on  $\mathcal{X} = \mathbb{R}^d$ . Then there exists a unique  $\sigma$ -finite measure  $\check{C}_{\mathcal{P}}$  on  $\mathcal{M}_{\mathcal{X}}^{\#}$  such that for any  $\mathcal{B}_{\mathcal{X}} \otimes \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ -measurable nonnegative, or  $C_{\mathcal{P}}$ -integrable, function  $g(\cdot)$ ,

$$\begin{aligned} \mathbb{E} \left( \int_{\mathcal{X}} g(x, \xi) \xi(dx) \right) &= \int_{\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}} g(x, \xi) C_{\mathcal{P}}(dx \times d\xi) \\ &= \int_{\mathcal{X}} dx \int_{\mathcal{M}_{\mathcal{X}}^{\#}} g(x, S_{-x}\psi) \check{C}_{\mathcal{P}}(d\psi), \end{aligned} \quad (13.2.5)$$

equivalently,

$$\mathbb{E} \left( \int_{\mathcal{X}} g(x, S_x \xi) \xi(dx) \right) = \int_{\mathcal{X}} dx \int_{\mathcal{M}_{\mathcal{X}}^{\#}} g(x, \psi) \check{C}_{\mathcal{P}}(d\psi). \quad (13.2.6)$$

(b) The measure  $\check{C}_{\mathcal{P}}$  is totally finite if and only if  $\xi$  has finite mean density  $m$ , in which case  $m = \check{C}_{\mathcal{P}}(\mathcal{M}_{\mathcal{X}}^{\#})$  and the measure  $m^{-1}\check{C}_{\mathcal{P}}(\cdot)$  coincides with the regular local Palm distribution  $\mathcal{P}_x(S_{-x}(\cdot))$  for  $\ell$ -a.e.  $x$ .

PROOF. Equations (13.2.5–6) are applications to the present setting of the factorization theorem, specifically, of (A2.7.5). The  $\sigma$ -finiteness of  $\check{C}_{\mathcal{P}}$  follows from the construction preceding the theorem.

To prove part (b), set  $g(x, \xi) = I_A(x)$  for some bounded  $A \in \mathcal{B}_{\mathcal{X}}$ . Then (13.2.5) yields

$$M(A) = \mathbb{E}[\xi(A)] = \ell(A) \int_{\mathcal{M}_{\mathcal{X}}^{\#}} \check{C}_{\mathcal{P}}(d\psi),$$

so that if  $M(A) < \infty$ ,  $\check{C}_{\mathcal{P}}$  must be totally finite with  $\check{C}_{\mathcal{P}}(\mathcal{M}_{\mathcal{X}}^{\#}) = M(A)/\ell(A) = m$ . Conversely, if  $\check{C}_{\mathcal{P}}(\mathcal{M}_{\mathcal{X}}^{\#}) < \infty$ , and equal to  $m'$  say, then  $M(A) = m'\ell(A) < \infty$ .

Furthermore, setting  $g(x, \xi) = f(x)h(S_x \xi)$  in (13.1.5) and (13.2.5) above yields

$$m \int_{\mathcal{X}} \int_{\mathcal{M}_{\mathcal{X}}^{\#}} f(x)h(S_x \xi) \mathcal{P}_x(d\xi) dx = \int_{\mathcal{X}} f(x) dx \int_{\mathcal{M}_{\mathcal{X}}^{\#}} h(\psi) \check{C}_{\mathcal{P}}(d\psi)$$

for all measurable Lebesgue-integrable  $f$ , so for all  $\check{C}_{\mathcal{P}}$ -integrable  $h$  it follows that

$$m \int_{\mathcal{M}_{\mathcal{X}}^{\#}} h(S_x \xi) \mathcal{P}_x(d\xi) = \int_{\mathcal{M}_{\mathcal{X}}^{\#}} h(\psi) \check{C}_{\mathcal{P}}(d\psi) \quad (\ell\text{-a.e. } x),$$

which in turn implies that  $m\mathcal{P}_x(S_{-x}d\xi)$  and  $\check{C}_{\mathcal{P}}(d\psi)$  must coincide for  $\ell$ -a.e.  $x$  as measures.  $\square$

**Definition 13.2.IV.** The Palm measure associated with the stationary random measure  $\xi$  on  $\mathbb{R}^d$  or its distribution  $\mathcal{P}$  is the measure  $\check{C}_{\mathcal{P}}(\cdot)$  on  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$  defined by (13.2.5); when the mean density  $m$  is finite, the probability measure  $\mathcal{P}_0(\cdot) = m^{-1}\check{C}_{\mathcal{P}}(\cdot)$  is the Palm distribution for  $\xi$ .

*Remark.* In the point process case, it follows from Proposition 13.1.V that the measure  $\check{C}_{\mathcal{P}}$  is supported by the subspace of counting measures  $N$  on  $\mathbb{R}^d$

for which  $N\{0\} \geq 1$ ; we denote this space  $\mathcal{N}_0^\#(\mathbb{R}^d)$  in general, and  $\mathcal{N}_0^{\#*}$  when the counting measures are also simple with the space  $\mathbb{R}^d$  omitted when the dimension  $d$  is clear from the context. Thus, a generic element  $N_0 \in \mathcal{N}_0^{\#*}$  is boundedly finite, simple, and has  $N\{0\} = 1$ .

If there is danger of ambiguity, we may use the phrase ‘stationary Palm measure’ (or ‘distribution’) to distinguish  $\mathcal{P}_0(\cdot)$  from the local versions described onwards from (13.1.4) in the last section. As there, we can gain insight into the character and interpretation of the Palm measure by looking first at a finite point process.

EXAMPLE 13.2(a). *Stationary point process on  $\mathbb{S}$*  [see Example 12.1(f)]. A stationary point process on  $\mathbb{S}$  has probability distributions that are invariant under rotation, and we showed in Example 12.1(f) that this implies that the symmetrized probability measures  $\Pi_n$  describing such a process have reduced forms  $\check{\Pi}_n$  satisfying (12.1.17); they describe the positions, relative to an arbitrarily selected initial point of the realization, of the other  $n-1$  points.

It is obvious that the Palm measure for the process should be closely related to the reduced distributions  $\check{\Pi}_n$ . The diagonal decompositions summarized by (12.1.17) for a given value of  $n$  can be expressed in portmanteau form via a mapping of  $\mathcal{X}^\cup \setminus \{\emptyset\} \mapsto \mathcal{X} \times \mathcal{X}^\cup$ , similar to the mapping  $D$  preceding (13.2.4), where on the component  $\mathcal{X}^{(n)}$

$$(\theta_1, \dots, \theta_n) \mapsto (\theta_1; \theta_2 - \theta_1, \dots, \theta_n - \theta_1).$$

Any measure  $\mathcal{P}$  on  $\mathcal{X}^\cup$  satisfying  $\mathcal{P}(\{\emptyset\}) = 0$  is thereby mapped into a product of uniform measure on  $\mathcal{X}$  and a reduced measure on  $\mathcal{X}^\cup$ ,  $\check{\mathcal{P}}$  say, which consists of the measures  $\check{\Pi}_n$  with weightings  $p_n = \mathcal{P}\{\xi(\mathcal{X}) = n\}$ . However, this is not quite the Palm distribution, as we can see by reference to any of the previous formulae, such as (13.2.6), from which follows the relation

$$\check{\mathcal{C}}_{\mathcal{P}}(\Gamma) = \frac{1}{2\pi} \mathbb{E} \left( \int_0^{2\pi} I_\Gamma(S_\theta N) N(d\theta) \right).$$

Then for a set  $\Gamma$  determined by a realization with  $n$  points of the form  $(0, \phi_1, \dots, \phi_{n-1})$ ,

$$\begin{aligned} \check{\mathcal{C}}_{\mathcal{P}}(\Gamma) &= \frac{p_n}{2\pi} \mathbb{E} \left( \sum_{i=1}^n I_\Gamma(S_{\theta_i} N) \mid n \right) \\ &= \frac{p_n}{2\pi} \int_{\mathcal{X}^{(n)}} \sum_{i=1}^n I_\Gamma(S_{\theta_i} N) \Pi_n(d\theta_1 \times \dots \times d\theta_n) \\ &= \frac{np_n}{2\pi} \int_{\mathcal{X}} d\theta \int_{\mathcal{X}^{(n-1)}} I_\Gamma(N_0) \check{\Pi}_n(d\phi_1 \times \dots \times d\phi_{n-1}), \end{aligned}$$

where now  $N_0$  is a generic counting measure with points at  $0, \phi_1, \dots, \phi_{n-1}$ . The factor  $n$  arises because the  $n$  terms in the sum give identical integrals

on account of the symmetry properties of  $\Pi_n$ . For such  $\Gamma$  we therefore have  $\check{C}_{\mathcal{P}}(\Gamma) = np_n \check{\Pi}_n(\Gamma)$ . Thus, just as in the more general situation of Exercise 13.1(b), the Palm measure requires a weighting by the factor  $n$ . The Palm distribution is then obtained by normalizing this weighted form, which requires  $\sum_{n=1}^{\infty} np_n = E[N(\mathcal{X})] < \infty$ . The intuitive explanation is as before: a realization with  $n$  points is  $n$  times more likely to locate a point at the origin than a realization with just one point, and must be weighted accordingly in taking the expectation.  $\square$

Equations (13.2.5) and (13.2.6) yield a range of striking formulae as special cases and corollaries. One of the simplest is the following interpretation of the Palm probabilities for point processes.

**EXAMPLE 13.2(b) Palm probabilities as rates.** An important interpretation of the Palm measure for a point process comes from (13.2.6), which yields for  $\Gamma \in \mathcal{N}_0^{\#\#}$ ,

$$\check{C}_{\mathcal{P}}(\Gamma) = E(\#\{i: x_i \in \mathbb{U}^d \text{ and } S_{x_i}N \in \Gamma\}), \quad (13.2.7)$$

where on the right-hand side it is to be understood that each  $x_i$  is a point of the realization  $N$ . Thus,  $\check{C}_{\mathcal{P}}(\Gamma)$  is the expected number of points of the process in the unit cube (or, because the process is stationary, their expected rate), which, when the origin is transferred to the point in question, are associated with the occurrence of  $\Gamma$ . As Matthes and others have suggested, we can regard this as the rate of occurrence of marked points where a point is marked if and only if  $\Gamma$  occurs when the origin is shifted to the point in question. The Palm distribution then appears as a ratio of rates

$$\mathcal{P}_0(\Gamma) = m(\Gamma)/m, \quad (13.2.8)$$

where  $m(\Gamma)$  is the rate of the marked process and  $m$  is the rate of the original process.  $\square$

Next we use (13.2.5–6) to give more explicit formulae expressing the Palm measure  $\check{C}_{\mathcal{P}}$  in terms of  $\mathcal{P}$  and vice versa. Setting  $g(x, \xi) = j(x)h(\xi)$  in (13.2.5), where  $\int_{\mathcal{X}} j(x) dx = 1$ , we obtain at (13.2.10) below the expression for  $\check{C}_{\mathcal{P}}$  in terms of  $\mathcal{P}$ . For an inverse relation, more subtlety is needed: suppose that  $k: \mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#} \mapsto \mathbb{R}_+$  is a normalizing kernel for the realizations  $\xi$ , that is, a nonnegative measurable function satisfying

$$k(x, \emptyset) = 0 \quad \text{and} \quad \int_{\mathcal{X}} k(x, \xi) \xi(dx) = 1 \quad \text{for each } \xi \neq \emptyset. \quad (13.2.9)$$

Substituting  $g(x, \xi) = k(x, \xi)f(\xi)$  for some nonnegative  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ -measurable  $f(\cdot)$  in (13.2.5), equation (13.2.11) is obtained. We summarize as below.

**Proposition 13.2.V.** *Let  $\mathcal{P}$  be stationary on  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$  for  $\mathcal{X} = \mathbb{R}^d$  and  $\check{C}_{\mathcal{P}}$  its Palm measure.*

(a) For any nonnegative  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ -measurable function  $h(\cdot)$ , and nonnegative  $\mathcal{B}_{\mathcal{X}}$ -measurable  $j$  with  $\int_{\mathcal{X}} j(x) dx = 1$ ,

$$\int_{\mathcal{M}_{\mathcal{X}}^{\#}} h(\psi) \check{C}_{\mathcal{P}}(d\psi) = \int_{\mathcal{M}_{\mathcal{X}}^{\#}} \mathcal{P}(d\xi) \int_{\mathcal{X}} j(x) h(S_x \xi) \xi(dx) \quad (13.2.10a)$$

$$= E_{\mathcal{P}} \left( \int_{\mathcal{X}} j(x) h(S_x \xi) \xi(dx) \right). \quad (13.2.10b)$$

(b) For any nonnegative  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ -measurable function  $f(\cdot)$  and for  $k(\cdot)$  satisfying (13.2.9),

$$E_{\mathcal{P}}[f(\xi)] = \int_{\mathcal{M}_{\mathcal{X}}^{\#} \setminus \{\emptyset\}} \int_{\mathcal{X}} k(x, S_{-x}\psi) f(S_{-x}\psi) \check{C}_{\mathcal{P}}(d\psi) dx + f(\emptyset) \mathcal{P}\{\xi = \emptyset\}, \quad (13.2.11)$$

where  $\emptyset$  denotes the null measure.

As a particular corollary of the proposition set  $h(\psi) = I_{\{\emptyset\}}(\psi)$  in (13.2.10). Then

$$\check{C}_{\mathcal{P}}(\{\psi = \emptyset\}) = 0. \quad (13.2.12)$$

From (13.2.11) it can be seen that  $\check{C}_{\mathcal{P}}$  determines  $\mathcal{P}$  up to  $\mathcal{P}\{\xi = \emptyset\}$ . For the rest of this section we assume also that  $\mathcal{P}\{\xi = \emptyset\} = 0$ , in which case it follows from Proposition 12.1.VI that  $\mathcal{P}\{\xi(\mathcal{X}) = \infty\} = 1$ .

Equation (13.2.10) is often presented in the special case  $j(x) = I_{\mathbb{U}^d}(x)$ , when it takes the form

$$\int_{\mathcal{M}_{\mathcal{X}}^{\#}} h(\psi) \check{C}_{\mathcal{P}}(d\psi) = E_{\mathcal{P}} \left( \int_{\mathbb{U}^d} h(S_x \xi) \xi(dx) \right).$$

In the point process case this can be interpreted as follows: shift the origin successively to each point of the process in  $\mathbb{U}^d$  and sum the values of  $h$  obtained from these shifted versions of  $\xi$ , thus obtaining an expected rate of contributions to the sum. When  $h$  is specialized to the indicator of some event  $U$  of the process, the value of the integral on the right-hand side is just the rate of occurrence of points for which  $U$  occurs when the origin is shifted to the point in question.

The two equations (13.2.10–11) can be specialized in various obvious ways. If we take  $h$  or  $f$  to be the indicator function of a set  $U \in \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$ , we obtain direct expressions for  $\check{C}_{\mathcal{P}}(U)$  in terms of  $\mathcal{P}$  and for  $\mathcal{P}(U)$  in terms of  $\check{C}_{\mathcal{P}}$ .

When the mean density  $m$  exists, the equations can be put in the more symmetrical forms using  $\mathcal{P}_0$  from Definition 13.2.IV,

$$E_{\mathcal{P}_0}(h(\psi)) = m^{-1} E_{\mathcal{P}} \left( \int_{\mathbb{U}^d} h(S_x \xi) \xi(dx) \right), \quad (13.2.13)$$

$$E_{\mathcal{P}}(f(\xi)) = m E_{\mathcal{P}_0} \left( \int_{\mathcal{X}} k(x, S_{-x}\psi) f(S_{-x}\psi) dx \right). \quad (13.2.14)$$

Specific examples of functions  $k(\cdot)$  satisfying (13.2.9) are given in Section 13.3; Exercise 13.2.3 outlines Mecke's (1967) general construction.

We now consider the moment measures of the Palm distribution. There is no loss of generality here in speaking of moments of the Palm distribution, rather than of the Palm measure, because the existence of higher moments implies the existence of the first moment, which ensures that the Palm measure is totally finite and so can be normalized to yield the Palm distribution.

Following the notation of the earlier sections, let  $\psi$  denote a stationary random measure, and write

$$\mathring{M}_k(A_1 \times \cdots \times A_k) = \mathbb{E}_{\mathcal{P}_0}(\psi(A_1) \cdots \psi(A_k)) \quad (A_1, \dots, A_k \in \mathcal{B}_{\mathcal{X}})$$

for the  $k$ th moment measure for  $\mathcal{P}_0$ .

**Proposition 13.2.VI.** *For  $k = 1, 2, \dots$ , the  $k$ th moment measure  $\mathring{M}_k$  of the Palm distribution exists if and only if the  $(k+1)$ th moment measure of the original random measure exists, in which case it is related to the reduced  $(k+1)$ th moment measure  $\check{M}_{k+1}$  by*

$$\mathring{M}_k(\cdot) = m^{-1} \check{M}_{k+1}(\cdot). \quad (13.2.15)$$

PROOF. The result is a further application of (13.2.5) and Fubini's theorem. For nonnegative measurable  $g$  and  $h$  with  $h(\cdot)$  on  $\mathcal{X}^{(k)}$  and  $g(\cdot)$  integrable on  $\mathcal{X}$ , set

$$g(x, \xi) = g(x) \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} h(y_1 - x, \dots, y_k - x) \xi(dy_1) \cdots \xi(dy_k).$$

Then the left-hand side of (13.2.5), using also (12.6.6), becomes

$$\begin{aligned} & \int_{\mathcal{X}^{(k+1)}} g(x) h(y_1 - x, \dots, y_k - x) M_{k+1}(dx \times dy_1 \times \cdots \times dy_k) \\ &= \int_{\mathcal{X}} g(x) dx \int_{\mathcal{X}^{(k)}} h(u_1, \dots, u_k) \check{M}_{k+1}(du_1 \times \cdots \times du_k), \end{aligned}$$

whereas using Fubini's theorem with the right-hand side yields

$$\begin{aligned} & m \mathbb{E}_{\mathcal{P}_0} \left( \int_{\mathcal{X}} g(x) dx \int_{\mathcal{X}^{(k)}} h(u_1, \dots, u_k) \psi(du_1) \cdots \psi(du_k) \right) \\ &= m \int_{\mathcal{X}} g(x) dx \int_{\mathcal{X}^{(k)}} h(u_1, \dots, u_k) \mathring{M}_k(du_1 \times \cdots \times du_k). \end{aligned}$$

But  $h \geq 0$  is arbitrary, so (13.2.15) follows, together with finiteness (because  $m$  is finite).  $\square$

The results of Section 12.4 and as above can be summed up in the following diagram:

$$\begin{array}{ccc} \mathcal{P}(d\xi) \xi(dx) & \xrightarrow{\text{moments}} & M_{k+1} \\ \downarrow \text{reduction} & & \downarrow \text{reduction} \\ m \mathcal{P}_0(d\xi) & \xrightarrow{\text{moments}} & m \mathring{M}_k \equiv \check{M}_{k+1} \end{array}$$

Again the results are conveniently illustrated with respect to the finite case of a point process on a circle, as set out in Exercise 13.2.6.

Moment measures can also be defined for higher-order and modified Palm distributions derived from the corresponding Campbell measures introduced around (13.1.3). For example, the first-order moment measure for the second-order Palm distribution  $\mathring{M}_1^{(2)}(dz \mid x, y)$  can be defined as the Radon–Nikodym derivative of the third-order moment measure with respect to the second-order moment measure:

$$\mathring{M}_1^{(2)}(dz \mid x, y) = \frac{M_3(dz \times dx \times dy)}{M_2(dx \times dy)}. \quad (13.2.16)$$

When the process is stationary this reduces to a function of  $u = y - x$  and  $v = z - x$ . Further discussion of these quantities and of the corresponding moment measures for the modified Palm distributions are briefly set out in Exercise 13.2.9.

We turn finally to a characterization of Palm measures due to Mecke (1975). Keeping to Mecke's context, we first observe that the relation between a stationary measure  $\mathcal{P}$  on  $\mathcal{B}(\mathcal{M}_X^\#)$  and its Palm measure  $\check{C}_{\mathcal{P}}$  can be extended to general  $\sigma$ -finite measures  $\mathcal{P}$ . The definition of stationarity carries over to this context, and equations (13.2.5–6) continue to define a  $\sigma$ -finite measure  $\check{C}_{\mathcal{P}}$  in terms of an initial  $\sigma$ -finite measure  $\mathcal{P}$ , whether or not the initial measure is totally finite. The only step that needs checking is that the argument establishing the  $\sigma$ -finiteness of  $\check{C}_{\mathcal{P}}$  remains valid (see Exercise 13.2.8). In this context, neither  $\mathcal{P}$  nor  $C_{\mathcal{P}}$  need be totally finite but both may be.

The previous results can be succinctly summarized for this extended context as follows.

**Proposition 13.2.VII.** *Let  $\mathcal{Q}$  be a  $\sigma$ -finite measure on  $\mathcal{B}(\mathcal{M}_X^\#)$  and  $C_{\mathcal{Q}}$  its associated Campbell measure. Then the following assertions are equivalent.*

- (i)  $\mathcal{Q}$  is stationary (invariant under shifts  $\hat{S}_u$ ,  $u \in \mathcal{X}$ ).
- (ii)  $C_{\mathcal{Q}}$  is invariant under the transformations  $\hat{\Theta}_u$ ,  $u \in \mathcal{X}$  defined at (13.2.1b).
- (iii)  $C_{\mathcal{Q}}$  factorizes as in (13.2.6) into a product of Lebesgue measure and an associated  $\sigma$ -finite measure  $\mathcal{R}$  on  $\mathcal{B}(\mathcal{M}_X^\#)$ , and  $\mathcal{R}$  coincides with  $\check{C}_{\mathcal{Q}}$ .

PROOF. Only the last assertion, equivalent to asserting the uniqueness of the Palm factorization, requires further comment. Suppose in fact that two distinct measures  $\mathcal{R}$  and  $\mathcal{R}'$  both satisfy (13.2.6) for the same  $\mathcal{Q}$ . Then there must be some measurable, nonnegative  $f$  with  $\int_{\mathcal{M}_X^\#} f d\mathcal{R} \neq \int_{\mathcal{M}_X^\#} f d\mathcal{R}'$ . But each of  $\mathcal{R}$  and  $\mathcal{R}'$  is associated with an inversion formula (13.2.11), leading to a contradiction if both  $\mathcal{R}$  and  $\mathcal{R}'$  are associated with the same  $\mathcal{Q}$ .  $\square$

The third assertion of the proposition invites the further question: which measures  $\mathcal{R}$  can appear in a Palm factorization. Although the factorization lemma implies that any  $\mathcal{R}$  inserted in the right-hand side of (13.2.6) yields a measure on  $\mathcal{W}$  that is invariant under  $\hat{\Theta}_u$ , in general this measure will not

be a Campbell measure, that is, it will not be generated by some underlying stationary  $\mathcal{Q}$  on  $\mathcal{M}_{\mathcal{X}}^{\#}$ . The constraint which  $\mathcal{R}$  must satisfy in order to be a stationary Palm measure is given in the following result of Mecke (1975, Theorem 1.7).

**Theorem 13.2.VIII.** *A measure  $\mathcal{R}$  on  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#})$  with  $\mathcal{X} = \mathbb{R}^d$ , is the Palm measure of some stationary,  $\sigma$ -finite (but not necessarily finite) measure  $\mathcal{Q}$  on  $(\mathcal{M}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#}))$  if and only if the following three conditions hold:*

- (i)  $\mathcal{R}$  is  $\sigma$ -finite;
- (ii)  $\mathcal{R}(\{\emptyset\}) = 0$ ; and
- (iii) for all measurable nonnegative  $h(\cdot, \cdot)$  on  $\mathcal{X} \times \mathcal{M}_{\mathcal{X}}^{\#}$ ,

$$\int_{\mathcal{M}_{\mathcal{X}}^{\#}} \int_{\mathcal{X}} h(-y, S_y \psi) \psi(dy) \mathcal{R}(d\psi) = \int_{\mathcal{M}_{\mathcal{X}}^{\#}} \int_{\mathcal{X}} h(y, \psi) \psi(dy) \mathcal{R}(d\psi). \quad (13.2.17)$$

If, in addition,  $\eta = \int_{\mathcal{M}_{\mathcal{X}}^{\#}} \int_{\mathcal{X}} k(y, S_{-y} \psi) dy \mathcal{R}(d\psi) \leq 1$  for some  $k(\cdot)$  satisfying (13.2.9), then  $\mathcal{R}$  is the Palm measure of a probability measure  $\mathcal{P}$  on  $\mathcal{M}_{\mathcal{X}}^{\#}$  satisfying  $\mathcal{P}\{\xi = \emptyset\} = 1 - \eta$ .

PROOF. Suppose first that  $\mathcal{R} = \check{\mathcal{C}}_{\mathcal{Q}}$  as in Proposition 13.2.VII for some stationary,  $\sigma$ -finite measure  $\mathcal{Q}$ . Then (i) follows from Theorem 13.2.III, and (ii) is (13.2.12). To establish (iii), let  $j(x)$  be nonnegative, measurable, with  $\int_{\mathcal{X}} j(x) dx = 1$ , and apply (13.2.10) to the left-hand side of (13.2.17), regarded as a function of  $\psi$ . This yields

$$\int_{\mathcal{M}_{\mathcal{X}}^{\#}} \int_{\mathcal{X}} \int_{\mathcal{X}} j(x) h(-y, S_y S_x \xi) \xi(dx) S_x \xi(dy) \mathcal{Q}(d\xi).$$

Because  $\int_{\mathcal{X}} f(y) S_x \xi(dy) = \int_{\mathcal{X}} f(y - x) \xi(dy)$  for any nonnegative measurable function  $f$ , the displayed expression can be rewritten as

$$\int_{\mathcal{M}_{\mathcal{X}}^{\#}} \int_{\mathcal{X}} \int_{\mathcal{X}} j(x) h(x - y, S_y \xi) \xi(dx) \xi(dy) \mathcal{Q}(d\xi),$$

and using now the reverse of this transformation to the integration over  $x$ , gives

$$\int_{\mathcal{M}_{\mathcal{X}}^{\#}} \int_{\mathcal{X}} \int_{\mathcal{X}} j(x + y) h(x, S_y \xi) S_y \xi(dx) \xi(dy) \mathcal{Q}(d\xi).$$

Rewriting the integration over  $\psi$  and  $y$  in terms of the factorization as in the right-hand side of (13.2.6) (with  $y$  and  $\xi$  in place of  $x$  and  $\psi$  there), gives

$$\int_{\mathcal{X}} \int_{\mathcal{M}_{\mathcal{X}}^{\#}} \int_{\mathcal{X}} j(x + y) h(x, \xi) \xi(dx) \mathcal{R}(d\xi) dy.$$

Finally, using  $\int_{\mathcal{X}} j(x + y) dy = 1$  and integrating out  $y$ , yields the right-hand side of (13.2.17).

For the reverse argument, suppose that  $\mathcal{R}$  satisfies conditions (i)–(iii) and that  $k$  is a normalizing kernel satisfying (13.2.9). Use  $\mathcal{R}$  to define a measure  $\mathcal{Q}$  on  $\mathcal{B}(\mathcal{M}_\mathcal{X}^\#)$  by  $\mathcal{Q}(\{\emptyset\}) = 0$  and the first term in (13.2.11), so that

$$\int_{\mathcal{M}_\mathcal{X}^\#} f(\xi) \mathcal{Q}(d\xi) = \int_{\mathcal{M}_\mathcal{X}^\# \setminus \{\emptyset\}} \int_{\mathcal{X}} k(y, S_{-y}\psi) f(S_{-y}\psi) \mathcal{R}(d\psi) dy.$$

We insert this measure into the left-hand side of the fundamental equation (13.2.6) and obtain

$$\begin{aligned} & \int_{\mathcal{M}_\mathcal{X}^\#} \int_{\mathcal{X}} g(x, S_x\xi) \xi(dx) \mathcal{Q}(d\xi) \\ &= \int_{\mathcal{M}_\mathcal{X}^\#} \int_{\mathcal{X}} \int_{\mathcal{X}} g(x, S_{x-y}\psi) k(y, S_{-y}\psi) S_{-y}\psi(dx) dy \mathcal{Q}(d\psi) \\ &= \int_{\mathcal{M}_\mathcal{X}^\#} \int_{\mathcal{X}} \int_{\mathcal{X}} g(u+y, S_u\psi) k(y, S_{-y}\psi) \psi(du) dy \mathcal{Q}(d\psi), \quad \text{putting } x = u+y, \\ &= \int_{\mathcal{M}_\mathcal{X}^\#} \int_{\mathcal{X}} \int_{\mathcal{X}} g(v, S_u\psi) k(v-u, S_{u-v}\psi) \psi(du) dv \mathcal{Q}(d\psi), \quad \text{putting } y = v-u, \end{aligned}$$

where in the last equation we used the invariance of Lebesgue measure under shifts. In general this expression does not simplify, but when (iii) holds, then with  $h(y, \psi) = \int_{\mathcal{X}} g(v, S_u\psi) k(v-u, S_{u-v}\psi) dv$ , we can write for the right-hand side

$$\int \int \int g(v, \psi) k(v+u, S_{-v}\psi) \psi(du) dv \mathcal{R}(d\psi) = \int \int g(v, \psi) dv \mathcal{R}(d\psi),$$

because

$$\int k(v+u, S_{-v}\psi) \psi(du) = \int k(x, S_{-v}\psi) S_{-v}\psi(dx) = 1.$$

The last equation but one shows that the measure  $\mathcal{Q}$  defined from  $\mathcal{R}$  factorizes as in Proposition 13.2.VII(iii), and hence first is stationary, and second, identifies  $\mathcal{R}$  as the reduced Campbell measure of  $\mathcal{Q}$ . By extension of Definition 13.2.IV we call such an  $\mathcal{R}$  a Palm measure also.

Finally, the normalizing condition entailing  $\eta \leq 1$  comes from Proposition 13.2.V(b).  $\square$

**EXAMPLE 13.2(c)** *Palm factorization of the KLM measure for a stationary infinitely divisible point process.* From Proposition 12.4.I we know that an infinitely divisible point process is stationary if and only if its KLM measure  $\tilde{Q}$  is stationary. In that case, a Palm factorization can be applied to the measure

$$C_{\tilde{Q}}(A \times U) = \int_U \tilde{N}(A) \tilde{Q}(d\tilde{N}) \quad (U \in \mathcal{B}(\mathcal{N}_\mathcal{X}^\# \setminus \{\emptyset\})),$$



giving

$$\int_{\mathcal{X}} \int_{\mathcal{N}_{\mathcal{X} \setminus \{\emptyset\}}^{\#}} g(x, S_x \tilde{N}) \tilde{N}(\mathrm{d}x) \tilde{Q}(\mathrm{d}\tilde{N}) = \int_{\mathcal{X}} \mathrm{d}x \int_{\mathcal{N}_0^{\#}} g(x, \tilde{N}_0) \check{C}_{\tilde{Q}}(\mathrm{d}\tilde{N}_0),$$

where  $\tilde{N}_0$  here denotes a generic element of  $\mathcal{N}_0^{\#}$ . Let us write for brevity

$$\check{C}_{\tilde{Q}} = \tilde{Q}_0,$$

and note that for a point process, as in the probability case,  $\tilde{Q}_0$  is defined on  $\mathcal{B}(\mathcal{N}_0^{\#})$ ;  $\tilde{Q}_0$  may or may not be totally finite.

We can now examine the properties of  $\tilde{Q}_0$  for the various types of stationary, infinitely divisible point processes.

(1°) Suppose that  $\mathcal{P}$  is regular and therefore has a representation as the regular version of a stationary Poisson cluster process (Proposition 12.4.II). Here  $\tilde{Q}_0$  is closely related to the symmetrized measures  $\tilde{P}_{k-1}$  used in defining the regular representation. Regard  $\tilde{P}_{k-1}$  not as a measure on  $\mathcal{X}^{(k-1)}$  but as a measure on the set  $\mathcal{D}_k$  of counting measures in  $\mathcal{N}_0^{\#}$  containing just  $k-1$  points in addition to the point at the origin. Then we have

$$\int_{\mathcal{D}_k} h(\tilde{N}_0) \tilde{Q}_0(\mathrm{d}\tilde{N}_0) = k \int_{\mathcal{X}^{(k-1)}} h\left(\delta_0 + \sum_{i=1}^{k-1} \delta_{x_i}\right) \tilde{P}_{k-1}(\mathrm{d}x_1 \times \cdots \times \mathrm{d}x_{k-1}),$$

the factor  $k$  arising here, as in Example 13.2(a), from the  $\tilde{N}(\mathrm{d}u)$  integration in the Campbell measure. The normalized measures

$$kp_k \tilde{P}_{k-1}(\cdot)/m,$$

where  $m = \sum kp_k$  and  $\{p_k\}$  describes the distribution of the cluster size, can be interpreted loosely as providing the conditional distribution of the other cluster members, given that the point at the origin is arbitrarily chosen from a randomly selected population of i.i.d. families of  $k$  members.

We see that in the regular case, the measure  $\tilde{Q}_0$  is supported by the set of counting measures in  $\mathcal{N}_0^{\#}$  with finite support.

(2°) If the process is strongly singular [see the definitions of  $\mathcal{S}_s$  at the end of Section 12.4 and of  $\bar{I}_V(\tilde{N})$  below (12.4.6)], the KLM measure itself is supported on the set of counting measures with positive ergodic limits [i.e.,  $\bar{I}_V(\tilde{N}) > 0$ ], and it follows, as in the discussion of Theorem 13.3.II below, that  $\tilde{Q}_0$  is concentrated on the space of sequences that also have positive ergodic limits. This is the situation if, in particular, the process is a Poisson randomization, in which case  $\tilde{Q}_0 = \mu \check{C}_{\mathcal{P}}$ , where  $\mu$  is the parameter of the Poisson distribution and  $\mathcal{P}$  is the point process that is being ‘randomized’.

(3°) Finally, the weakly singular processes have  $\tilde{Q}_0$  measures concentrated on the subset of  $\mathcal{N}_0^{\#}$  of counting measures that have infinite support but are asymptotically sparse in the sense that their ergodic limits are zero.

In summary we have proved the following statement.

**Proposition 13.2.IX.** *A stationary infinitely divisible point process is regular, strongly singular, or weakly singular, according to whether the Palm measure  $\hat{Q}_0$  is supported, respectively, by the finite counting measures, the counting measures with infinite support and positive ergodic limits, or the counting measures with infinite support but zero ergodic limits.*

Exercise 13.2.10 exhibits the Palm measure  $\mathcal{P}$  of a stationary infinitely divisible random measure with finite first moment measure as a convolution of  $\mathcal{P}$  with the reduced Campbell measure of its KLM measure.  $\square$

### Exercises and Complements to Section 13.2

13.2.1 Let the random measure  $\xi$  on the integers as in Exercise 13.1.5 be stationary. Describe  $\xi$  as a stationary sequence  $\{X_n\} \equiv \{\xi\{n\}\}$  of nonnegative r.v.s. What aspects of this sequence are described by  $\mathcal{P}$  and  $\mathcal{P}_0$ , respectively? When  $\xi$  is a simple point process, give analogues and simple cases of equations (13.2.2–3) and (13.2.6–7).

13.2.2 Let  $\{\mathcal{P}_n\}$  be a sequence of probability measures on  $\mathcal{M}_X^\#$  and  $\mathcal{P}$  a limit measure. If  $\{\mathcal{P}_n\}$  and  $\mathcal{P}$  are stationary and  $\mathcal{P}_n \rightarrow \mathcal{P}$  weakly, investigate under what conditions and in what sense there is convergence of the Campbell measures  $C_{\mathcal{P}_n}$  to  $C_{\mathcal{P}}$ . Consider also conditions under which, for a stationary process, weak convergence of the underlying probability measures implies weak convergence of the associated Palm distributions, and *vice versa*.

[Hint: Consider first the convergence of the first moment measures, which is necessary for any meaningful sense of convergence for the Campbell measures. For the second part, use the representation theorems first to establish convergence of the fidi distributions, and then refer to Section 11.1.]

13.2.3 Let  $\{A_n\}$  be a covering of  $\mathcal{X}$  by disjoint bounded Borel subsets of  $\mathcal{X}$ . Define the functions  $a(\cdot)$  on  $\mathcal{W} \setminus (\mathcal{X} \times \{\emptyset\})$  and  $k(\cdot)$  for  $\xi \neq \emptyset$  by

$$a(x, \xi) = \begin{cases} 1 & \text{if } x \in A_n \text{ and } \xi(A_n) = 0, \\ \sum_{n=1}^{\infty} 2^{-n} \frac{I_{A_n}(x)}{\xi(A_n)} & \text{otherwise,} \end{cases}$$

$$k(x, \xi) = \frac{a(x, \xi)}{\int_{\mathcal{X}} a(y, \xi) \xi(dy)}.$$

Verify that the function  $k(\cdot)$  satisfies (13.2.9).

13.2.4 Let the stationary random measure  $\xi$  have finite first moment measure. Show that the Laplace functional result of Proposition 13.1.VI simplifies on writing  $L_x[f] = L_0[S_{-x}f]$ .

13.2.5 (a) Let  $\xi', \xi''$  be stationary random measures with probability measures  $\mathcal{P}', \mathcal{P}''$ . For nonnegative  $p, q$  with  $p + q = 1$ , the random measure  $\xi$ , which equals either  $\xi'$  or  $\xi''$  with probabilities  $p, q$ , respectively, has probability measure  $\mathcal{P} = p\mathcal{P}' + q\mathcal{P}''$ . Find its Campbell and Palm measures.

(b) Let  $\xi', \xi''$  as in (a) be independent random measures, and let  $\xi = \xi' + \xi''$ . Using  $*$  to denote convolution, show that the Palm measure  $\mathcal{P}_0$  of  $\xi$  is given by

$$\mathcal{P}_0 = \mathcal{P}'_0 * \mathcal{P}'' + \mathcal{P}' * \mathcal{P}''_0.$$

- 13.2.6 Find explicit expressions for the reduced moment measures for a stationary point process on the circle in terms of the distributions  $\{p_n\}$ ,  $\check{\Pi}_n(\cdot)$  in the notation of Example 13.2(a). In particular, expanding the second moment measure by (5.4.7) and with  $p_n^* = np_n/m$ ,  $m = \sum_n np_n$ ,

$$\begin{aligned} \check{M}_{[2]}(B) d\theta &= \sum_{n=0}^{\infty} (n+1)(n+2)p_{n+2} \Pi_{n+2}(d\theta \times S_{-\theta}B \times \mathcal{X}^{(n)}) \\ &= \frac{m d\theta}{2\pi} \sum_{n=0}^{\infty} (n+1)p_{n+2}^* \check{\Pi}_{n+2}(B \times \mathcal{X}^{(n)}) \\ &= (m d\theta/2\pi) E_{\mathcal{P}_0}[N(B)]. \end{aligned}$$

- 13.2.7 Investigate properties of the modified Campbell measure  $C_{\mathcal{P}}^{\dagger}$  introduced in (13.1.1b) and the corresponding modifications of local Palm distributions. In particular, find the form of these distributions for the process considered in Example 13.1(b). [Hint: See Kallenberg (1983a, §12.3), (MKM, 1978, §5.4).]
- 13.2.8 Extend the argument of Exercise 13.1.2 from the special case of KLM measures to show that if a stationary,  $\sigma$ -finite measure  $\mathcal{R}$  on  $(\mathcal{M}_{\mathcal{X}}^{\#}, \mathcal{B}(\mathcal{M}_{\mathcal{X}}^{\#}))$  is associated with a Campbell measure  $\mathcal{C}_{\mathcal{R}}$  by (13.1.1), then  $\mathcal{C}_{\mathcal{R}}$  is again  $\sigma$ -finite. Show that in this case also, a function  $h$  can be found satisfying (13.2.4), and used to establish the existence of a  $\sigma$ -finite reduced version  $\check{\mathcal{C}}_{\mathcal{R}}$  satisfying (13.2.5) and (13.2.6) of Theorem 13.2.III.
- 13.2.9 (a) Define a second-order Campbell measure as in (13.1.3), and use it to define a family of second-order local Palm distributions  $\mathcal{P}_{x,y}^{(2)}(\cdot)$ . What simplification can be expected if the underlying random measure is stationary?
- (b) Suppose that first-, second-, and third-order moment measures  $M_1$ ,  $M_2$ ,  $M_3$  exist for the random measure  $\xi$ . Define the reduced moment measure  $\check{M}_1^{(2)}(dz | x, y)$  as the Radon-Nikodym derivative in

$$M_3(C, A, B) = \int_{A \times B} \check{M}_1^{(2)}(C | x, y) M_2(dx \times dy),$$

and show that it can be interpreted as the first moment measure of the second-order local Palm distribution as in (a). What simplifications can be expected if the underlying random measure is stationary? [Hint: Both sides equal  $E[\xi(A)\xi(B)\xi(C)]$ , assumed finite for bounded  $A, B, C \in \mathcal{B}_{\mathcal{X}}$ .]

- (c) Investigate the forms of these measures for (i) Poisson, (ii) renewal, and (iii) two-point cluster Poisson processes.
- 13.2.10 Let  $\mathcal{P}$  be the probability measure of a stationary infinitely divisible random measure with finite first moment measure, and let  $\tilde{Q}$  be its KLM measure. Prove that  $\mathcal{P}_0 = \tilde{Q}_0 * \mathcal{P}$ , where  $\tilde{Q}_0$  is the reduced Campbell measure  $\check{C}_{\tilde{Q}}$ . [Hint: Let  $\mathcal{P}$  and  $\tilde{Q}$  have Laplace functionals  $L_{\mathcal{P}}$  and  $L_{\tilde{Q}}$ . Relate  $\log L_{\mathcal{P}}[f]$  to  $L_{\tilde{Q}}[f]$  and deduce that for  $f, g \in BM_+(\mathbb{R}^d)$  [see (13.1.15)],

$$\int_{\mathbb{R}^d} g(x) L_{\mathcal{P}}[S_{-x}f; 0] m_{\mathcal{P}} dx = L_{\mathcal{P}}[f] \int_{\mathbb{R}^d} g(x) L_{\tilde{Q}}[S_{-x}f; 0] m_{\tilde{Q}} dx.]$$

### 13.3. Interval- and Point-stationarity

When the ideas of Section 13.2 are specialized to point processes, a number of new features arise; we review them in this section. In particular, we consider a number of results for the important special case where  $\mathcal{X}$  is the real line, and the point process is simple with finite mean rate  $m$ . The central result here is the correspondence, foreshadowed already in Chapter 3, between such point processes and stationary sequences of intervals. Finding a counterpart for this result in spaces of higher dimension seemed an impossible task at the time of the first edition of this book, but the recent attack on the problem by Thorisson, Last, Heveling, and others has shown this not to be the case, and is introduced at the end of the section.

As noted below Definition 13.2.IV, it follows from Proposition 13.1.V that, for a simple stationary point process, the support for its Palm measure can be taken to be the space  $\mathcal{N}_0^{\#*}$  of simple counting measures with a point at the origin. From Proposition 12.1.VI it follows, because  $N\{0\} = 1$ , that the Palm measure is in fact supported by elements  $N \in \mathcal{N}_0^{\#*}$  which cannot be empty and must therefore satisfy  $N(\mathbb{R}^d) = \infty$ . When the state space is the line, this means that  $N(-\infty, 0] = N(0, \infty) = \infty$ .

For this special case, it is clear also that there is a one-to-one both ways measurable mapping  $\Phi$  between the space  $\mathcal{N}_0^{\#*}$  [with the  $\sigma$ -algebra of Borel sets inherited from  $\mathcal{B}(\mathcal{N}_{\mathbb{R}}^{\#*})$ ] and the space  $\mathcal{T}^+$  of doubly infinite sequences of positive numbers. Denoting the points of a generic element  $N_0 \in \mathcal{N}_0^{\#*}$  by  $\{\dots, t_{-1}(N_0), t_0(N_0) = 0, t_1(N_0), \dots\}$  with  $t_i(N_0) < t_{i+1}(N_0)$  ( $i = 0, \pm 1, \dots$ ), the mapping  $\Phi$  associates the points of  $N_0$  with the sequence of intervals  $\tau_i \equiv t_i(N_0) - t_{i-1}(N_0)$ ; that is,

$$\Phi N_0 = \{\tau_i\} \equiv \{t_i(N_0) - t_{i-1}(N_0) : i = 0, \pm 1, \dots\}.$$

For measurability see Exercise 9.1.14. Every measure on  $(\mathcal{N}_0^{\#*}, \mathcal{B}(\mathcal{N}_0^{\#*}))$ , and in particular every Palm measure  $\check{C}_{\mathcal{P}}(\cdot)$ , then induces a measure on  $(\mathcal{T}^+, \mathcal{B}(\mathcal{T}^+))$ , say  $(\check{C}_{\mathcal{P}}\Phi^{-1})(\cdot)$ .

These remarks pave the way for the results setting up a correspondence between counting properties and interval properties. The correspondence is essentially a restatement of the equations representing the Palm measure in terms of the Campbell measure and hence of its underlying probability measure, and *vice versa*. We state the theorem first in its most striking form, for the case of finite mean density  $m$ .

**Theorem 13.3.I** [Ryll-Nardzewski (1961), Slivnyak (1962, 1966), Kaplan (1955)]. *There is a one-to-one correspondence between the distributions  $\mathcal{P}$  on  $\mathcal{B}(\mathcal{N}_{\mathbb{R}}^{\#*})$  of simple nonnull stationary point processes on the line with finite mean density  $m$ , and the distributions  $\Pi$  on  $\mathcal{B}(\mathcal{T}^+)$  of stationary sequences of positive random variables with finite mean  $m^{-1}$ . If  $\mathcal{P}_0$  is the image in*

$\mathcal{B}(\mathcal{N}_0^{\#*})$  under  $\Phi^{-1}$  of the measure  $\Pi$  on  $\mathcal{B}(\mathcal{T}^+)$ , this relation is effected by the equations

$$\mathbb{E}_{\mathcal{P}_0}(h(N_0)) = m^{-1} \mathbb{E}_{\mathcal{P}} \left( \sum_{i=1}^{N(0,1]} h(S_{t_i} N) \right) \quad (13.3.1)$$

for nonnegative  $\mathcal{B}(\mathcal{N}_0^{\#*})$ -measurable  $h(\cdot)$ , and

$$\begin{aligned} \mathbb{E}_{\mathcal{P}}(g(N)) &= m \mathbb{E}_{\mathcal{P}_0} \left( \int_0^{t_1(N_0)} g(S_t N_0) dt \right) \\ &= m \int_0^\infty \mathbb{E}_{\mathcal{P}_0} [g(S_t N_0) I_{\{t_1(N_0) > t\}}(N_0)] dt \end{aligned} \quad (13.3.2)$$

for nonnegative  $\mathcal{B}(\mathcal{N}_{\mathbb{R}}^{\#*})$ -measurable  $g(\cdot)$ .

PROOF. Equations (13.3.1–2) are adaptations to the present context of equations (13.2.10–11), but further comments are required. In (13.3.1), the points  $t_i$  refer to points of  $N$  lying in the unit interval  $(0, 1]$ ; each  $S_{t_i}$  satisfies  $(S_{t_i} N)(\{0\}) = N(\{t_i\}) = 1$ , and so with probability 1 it can be identified with an element of  $\mathcal{U}_0$  whenever  $h(S_{t_i} N)$  is well defined. In the exceptional  $\mathcal{P}$ -null set, where  $(t_i, S_{t_i} N) \notin \mathcal{U}_0$ , the value of  $h(S_{t_i} N)$  can be represented arbitrarily (and set equal to zero, say).

To derive (13.3.2) from (13.2.11) set there

$$k(x, N) = \begin{cases} 1 & \text{if } x = t_0(N), \\ 0 & \text{otherwise.} \end{cases} \quad (13.3.3a)$$

Observe that for a simple nonnull counting measure (see Definition 12.1.VII),

$$\int_{\mathbb{R}} k(x, N) N(dx) = N(\{t_0(N)\}) = 1, \quad (13.3.3b)$$

so that  $k(\cdot)$  satisfies (13.2.9) when also  $\mathcal{P}$  corresponds to a simple point process, that is,  $\mathcal{P}(\mathcal{N}_{\mathbb{R}}^{\#*}) = 1$ . Now substituting in (13.2.11), the term  $k(x, S_{-x} N_0)$  in the integral on the right-hand side of (13.2.11) equals unity provided  $x = t_0(S_{-x} N_0)$ , which, because the counting measure  $S_{-x} N_0$  for  $N_0 \in \mathcal{N}_0^{\#*}$  consists of atoms of unit mass at  $\{x + t_i(N_0) : i = 0, \pm 1, \dots\}$ , is true for  $x \leq 0 < x + t_1(N_0)$ , that is, for  $-\tau_1 < x \leq 0$ . Changing the variable of integration from  $x$  to  $-x$  leads to (13.3.2).

We now show that if  $\mathcal{P}$  is stationary on  $\mathcal{B}(\mathcal{N}_{\mathbb{R}}^{\#*})$  then  $\Pi$  is stationary on  $\mathcal{B}(\mathcal{T}^+)$  and conversely. Following Franken *et al.* (1981) we argue as follows. First, using the stationarity of  $\mathcal{P}$  we can extend (13.3.1) to give

$$\mathbb{E}_{\mathcal{P}_0}(h(N_0)) = \frac{m}{T} \mathbb{E}_{\mathcal{P}} \left\{ \sum_{i=1}^{N(0,T]} h(S_{t_i} N) \right\} \quad (0 < T < \infty).$$

Define the shift operator  $\vartheta: \mathcal{T}^+ \mapsto \mathcal{T}^+$  by  $\{\vartheta\tau_i\} = \{\tau_{i-1}\}$ . Then its image  $\Theta: \mathcal{N}_0^{\#*} \mapsto \mathcal{N}_0^{\#*}$ , where  $\Theta N_0 = \Phi^{-1}\{\vartheta\tau_i\} = \Phi^{-1}(\vartheta(\Phi N_0))$ , satisfies

$$\mathbb{E}_{\mathcal{P}_0}(h(\Theta N_0)) = \frac{m}{T} \mathbb{E}_{\mathcal{P}} \left( \sum_{i=1}^{N(0,T]} h(S_{t_{i-1}N}) \right),$$

from which we have

$$|\mathbb{E}_{\mathcal{P}_0}(h(N_0)) - \mathbb{E}_{\mathcal{P}_0}h(\Phi N_0)| \leq (m/T) \mathbb{E}_{\mathcal{P}}(|h(S_{t_0}N)| + |h(S_{t_{N'+1}}N)|),$$

where  $N' = N(0, T]$ . Then for all bounded  $h$ , the right-hand side  $\rightarrow 0$  as  $T \rightarrow \infty$ . Consequently, the expectations on the left-hand side coincide for all bounded measurable  $h$ , so that the measures  $\Pi$  and  $\Pi \circ \vartheta$ , equivalent to  $\mathcal{P}_0$  and  $\mathcal{P}_0 \circ \Theta$ , are therefore equal; that is,  $\Pi$  is invariant under  $\vartheta$ , and thus its iterates are also invariant.

Similarly, when  $\Pi$  is stationary, (13.3.2) can be extended by iteration to

$$\mathbb{E}_{\mathcal{P}}[g(N)] = \frac{1}{mr} \mathbb{E}_{\mathcal{P}_0} \left( \int_0^{t_r(N_0)} g(S_t N_0) dt \right) \quad (r = 1, 2, \dots).$$

Replacing  $N$  by  $S_t N$ , subtracting, and letting  $r \rightarrow \infty$ , we find in an analogous fashion that  $\mathcal{P}$  is stationary under shifts  $S_t$ .

We come finally to the question of uniqueness. Suppose we are given a stationary measure  $\Pi$  on  $\mathcal{T}^+$  and that  $\mathcal{P}$  is constructed from  $\Pi$  via  $\mathcal{P}_0$  and (13.3.2). Then  $\mathcal{P}$ , which is clearly a probability measure, has an associated Palm measure  $\check{\mathcal{C}}_{\mathcal{P}}$  that satisfies the equation analogous to (13.3.2); namely,

$$\mathbb{E}_{\mathcal{P}}[g(N)] = \int_{\mathcal{N}_0^{\#*}} \check{\mathcal{C}}_{\mathcal{P}}(dN_0) \int_0^{t_1(N_0)} g(S_t N_0) dt. \quad (13.3.4)$$

Substituting  $g(N) = h(S_{t_0(N)}N)$ , the inner integral in (13.3.4) becomes

$$\int_0^{t_1(N_0)} h(S_{t_0(S_t N_0)} S_t N_0) dt.$$

Now for  $0 < t < t_1(N_0)$ ,  $S_t N_0$  has points at  $t_i(N_0) - t$ , so the point of  $S_t N_0$  lying in  $(-\infty, 0)$  and nearest to the origin is at  $-t$ . In this range of  $t$ , therefore, the argument of  $h$  reduces to  $N_0$ , and (13.3.2) yields for this  $g$

$$\mathbb{E}_{\mathcal{P}}[g(N)] = m \mathbb{E}_{\mathcal{P}_0}[t_1(N_0)h(N_0)].$$

Similarly, (13.3.3) yields

$$\mathbb{E}_{\mathcal{P}}[g(N)] = \int_{\mathcal{N}_0^*} t_1(N_0)h(N_0) \check{\mathcal{C}}_{\mathcal{P}}(dN_0).$$

Both these equations hold for nonnegative  $\mathcal{B}(\mathcal{N}_0^{\#*})$ -measurable  $h$ , so it follows that the measures  $mt_1(N_0)\Pi(dN_0)$  and  $t_1(N_0)\check{\mathcal{C}}_{\mathcal{P}}(dN_0)$  coincide, thereby

identifying  $m\Pi$  as the Palm measure for  $\mathcal{P}$ . Thus,  $\Pi$  is determined uniquely by  $\mathcal{P}$ .

Again, if  $\mathcal{P}$  is given and  $\Pi$  is determined (through  $\mathcal{P}_0$ ) by (13.3.1), then we know already that  $\Pi$  is the Palm distribution of  $\mathcal{P}$ , and hence from (13.3.2) that  $\mathcal{P}$  is uniquely determined by  $\Pi$ . Either equation on its own is enough to imply a one-to-one correspondence.  $\square$

Theorem 13.3.I is a substantial generalization of the Palm–Khinchin equations of Section 3.4, and it provides the most satisfactory approach to the determination of the point process associated with a given process of intervals. The intuitive content of (13.3.2) can be expressed as follows. To embed a stationary sequence of intervals  $\{\tau_n: n = 0, \pm 1, \dots\}$  with distribution  $\Pi$  as in Theorem 13.3.I into a stationary point process on  $\mathbb{R}$ , first select a realization  $\{\tau_n\}$ , and choose a number  $X$  uniformly at random on a suitably large interval  $(0, T)$  say, with  $T \gg \text{any } \tau_n$  [roughly speaking, this is like taking  $r$  large in the display before (13.3.4)]. Then define a realization  $\{t_n\}$  of the point process on  $\mathbb{R}$  by relabelling the sequence

$$t'_r = \begin{cases} -X + \tau_1 + \dots + \tau_r & (r = 0, 1, \dots), \\ -X - \tau_0 - \dots - \tau_{r+1} & (r = -1, -2, \dots), \end{cases}$$

as  $t_n = t'_{r'+n}$  ( $n = 0, \pm 1, \dots$ ), where we identify  $r'$  from  $t'_{r'} = \inf\{t'_r: t'_r > 0\}$ . Then

$$t_n - t_{n-1} = \tau_{r'+n} \quad (\text{all } n).$$

From the choice of  $X$ ,  $\Pr\{\tau_{r'} > x\} = \int_x^\infty u \Pi(du)$ , that is, the length-biased distribution of the common distribution of the stationary sequence  $\{\tau_n\}$ .

The next example utilizes a more direct construction that follows Palm's original suggestion. It incorporates the idea of a point chosen uniformly at random from within an initial interval selected by the length-biased form of the stationary interval distribution spelled out in detail for the Wold process around (4.5.3a). Equation (13.3.2) is a more formal and general way of expressing the same ideas.

**EXAMPLE 13.3(a) Renewal and Wold processes.** Suppose first that  $\{\dots, L_{-1}, L_0, L_1, \dots\}$  is a sequence of i.i.d. positive r.v.s, which is therefore stationary and so describes a distribution  $\Pi$  on  $\mathcal{T}^+$ . Indeed,  $\Pi$  is just the product measure on  $(\mathbb{R}_+)^{(\infty)}$  derived from multiple copies of the measure  $F(dx)$  associated with each of the  $L_i$ . To fit into the framework of the theorem we must have

$$F(0+) = 0, \quad \int_0^\infty x F(dx) = m^{-1} < \infty.$$

In (13.3.2) take  $g(N) = I_\Gamma(N)$ , where  $\Gamma = \Gamma_1 \equiv \{N: t_1(N) > x\}$ . Then the term  $g(S_t N_0)$  on the right-hand side of (13.3.2) equals unity for  $0 < t < t_1(N_0) - x$  and  $t_1(N_0) > x$ , and zero otherwise, so that

$$\begin{aligned} E_{\mathcal{P}}(g(N)) &= \mathcal{P}(\Gamma_1) = \mathcal{P}\{t_1(N) > x\} = m E_\Pi[(L_1 - x)I_{\{L_1 > x\}}] \\ &= m \int_x^\infty (y - x) F(dy) = m \int_x^\infty [1 - F(y)] dy, \end{aligned}$$

which is the first of the Palm–Khinchin equations (3.4.9) and shows in the renewal case that the first interval after a fixed origin (and in view of stationarity the choice of origin is immaterial) has the distribution of the forward recurrence time [see Example 4.1(c)].

Next, take

$$\Gamma = \Gamma_2 \equiv \{N: t_1(N) > x, t_2(N) - t_1(N) > y\}.$$

We obtain similarly

$$\begin{aligned} \mathcal{P}(\Gamma_2) &\equiv \mathcal{P}\{t_1(N) > x, t_2(N) - t_1(N) > y\} = m \mathbb{E}_{\Pi}[(L_1 - x)I_{\{L_1 > x, L_2 > y\}}] \\ &= m \int_x^{\infty} [1 - F(u)] du (1 - F(y)), \end{aligned}$$

on account of the assumed independence of the  $\{L_i\}$ . The first equality here is the second of the Palm–Khinchin equations. The other equality shows that for a stationary renewal process, the length of the second interval after the origin is independent of the first.

In the case of a Wold process (see Section 4.5), the intervals  $\{L_i\}$  form a stationary Markov chain with stationary distribution  $\pi(\cdot)$  say and transition kernel  $P(x, B) = \Pr\{L_{i+1} \in B \mid L_i = x\}$ . Again we must assume that  $\pi(\{0\}) = 0 = \pi((-\infty, 0])$  and  $\int_0^{\infty} x \pi(dx) = m^{-1} < \infty$ . For  $\Gamma_1$  and  $\Gamma_2$  as above we find that  $t_1(N)$  has the same kind of forward recurrence time distribution with  $\pi(\cdot)$  in place of  $F(\cdot)$ , and that  $t_1(N)$  and  $t_2(N) - t_1(N)$  have the joint distribution

$$F_2(dx \times dy) = m dx \int_x^{\infty} \pi(du) P(u, dy).$$

Thus, the marginal distribution of  $t_2(N) - t_1(N)$  is now given by

$$F_2(\mathbb{R}_+ \times dy) = m \int_0^{\infty} dx \int_x^{\infty} \pi(du) P(u, dy),$$

and in general neither this interval nor any of the later intervals has exactly the stationary interval distribution.  $\square$

The analysis of Section 13.2 allows us to construct a Palm measure even for a process with infinite intensity. It is therefore natural to seek a version of Theorem 13.3.I valid even for processes with infinite mean rate; this is possible if  $\Pi$  is allowed to have infinite total mass. In fact, the proof of Theorem 13.3.I carries over with only notational changes as soon as we replace  $m\Pi$  by the measure induced on  $\mathcal{T}^+$  by the Palm measure  $\check{C}_{\mathcal{P}}$ , which remains  $\sigma$ -finite but not necessarily totally finite. For brevity we state the theorem below in terms of a measure  $\mathcal{R}$  on the space  $\mathcal{N}_0^{\#*}$  rather than a measure on the space  $\mathcal{T}^+$  of interval sequences.



**Theorem 13.3.II.** *There is a one-to-one correspondence between distributions  $\mathcal{P}$  on  $\mathcal{B}(\mathcal{N}_{\mathbb{R}}^{\#*})$  of simple nonnull stationary point processes on  $\mathbb{R}$ , and stationary  $\sigma$ -finite (but not necessarily totally finite) measures  $\mathcal{R}$  on  $\mathcal{B}(\mathcal{N}_0^{\#*})$  satisfying*

$$\int_{\mathcal{N}_0^{\#*}} t_1(N_0) \mathcal{R}(dN_0) = \int_{\mathcal{N}_{\mathbb{R}}^{\#*}} \mathcal{P}(dN) = 1. \quad (13.3.5)$$

*The correspondence is effected via nonnegative  $\mathcal{B}(\mathcal{N}_0^{\#*})$ -measurable  $h$  and nonnegative  $\mathcal{B}(\mathcal{N}_{\mathbb{R}}^{\#*})$ -measurable  $g$  in the equations*

$$\int_{\mathcal{N}_0^{\#*}} h(N_0) \mathcal{R}(dN_0) = \int_{\mathcal{N}_{\mathbb{R}}^{\#*}} \left( \sum_{i=1}^{N(0,1]} h(S_{t_i} N) \right) \mathcal{P}(dN) \quad (13.3.6a)$$

and

$$\int_{\mathcal{N}_{\mathbb{R}}^{\#*}} g(N) \mathcal{P}(dN) = \int_{\mathcal{N}_0^{\#*}} \mathcal{R}(dN_0) \int_0^{t_1(N_0)} g(S_t N_0) dt. \quad (13.3.6b)$$

PROOF. The normalization condition (13.3.5) follows from setting  $f(\xi) \equiv 1$  in (13.2.11). The remaining results paraphrase those of Theorem 13.3.I; details of the proofs are left to the reader.  $\square$

We return now to the more general context of point processes on  $\mathbb{R}^d$ . In the absence of a total ordering on  $\mathbb{R}^d$ , it is not immediately apparent what should be the exact counterparts of the preceding results. Some initial progress can be made by replacing the role of  $\tau_1$  above by the point of the realization,  $x^*(N)$  say, that is closest to the origin. We first check that this concept is well defined.

**Lemma 13.3.III.** *Let  $N$  be a simple nonnull stationary point process on  $\mathcal{X} = \mathbb{R}^d$ . Then the set*

$$\{N: \text{there exist } x', x'' \text{ with } \|x'\| = \|x''\| \text{ and } N(\{x'\}) \geq 1, N(\{x''\}) \geq 1\}$$

*is  $\mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#*})$ -measurable and has  $\mathcal{P}$ -measure zero.*

PROOF. The set  $J \subset \mathcal{X}^{(2)}$  defined by

$$J = \{(x, y): \|x\| = \|y\|, x \neq y\}$$

is a measurable set in  $\mathcal{B}(\mathcal{X}^{(2)})$  by inspection, and we can write

$$\int_{\mathcal{X}^{(2)}} I_J(x, y) N(dx) N(dy) = \int_{\mathcal{X}} h(x, N) N(dx),$$

where

$$h(x, N) = \begin{cases} 1 & \text{if } N(\{y\}) > 0 \text{ for some } y \text{ with } \|y\| = \|x\|, \\ 0 & \text{otherwise,} \end{cases}$$

is measurable. Applying (13.2.5), we obtain

$$\mathbb{E} \left( \int_{\mathcal{X}} h(x, N) N(dx) \right) = \int_{N_0^{\#*}} \check{C}_{\mathcal{P}}(dN_0) \int_{\mathcal{X}} h(x, S_{-x}N_0) dx.$$

The function  $h(x, S_{-x}N_0)$  equals 1 only on the at most countable set of surfaces

$$\{y: \|y + x_i\| = \|x_i\|, y \neq x_i\}$$

obtained by letting  $x_i$  run through the points of  $N_0$ . For  $d = 1$ , the surface consists of the single point  $y = -2x_i$ ; for  $d > 1$ , it consists of a surface in  $\mathbb{R}^d$  of dimension  $d - 1$ , and so is of zero  $\mathbb{R}^d$ -Lebesgue measure. In either case, the inner integral vanishes for each  $N_0$ , and so the expectation is zero.  $\square$

It follows from Lemma 13.3.III that with probability 1 the distances from the origin to the points of a realization of a nonnull stationary simple point process in  $\mathbb{R}^d$  can be set out in a strictly increasing sequence

$$0 < r_1(N) < r_2(N) < \dots$$

In this case the quantities  $r_i(N)$  are well-defined random variables because

$$\{r_i(N) < a\} \quad \text{if and only if} \quad \{N(S_a(0)) \geq i\},$$

and for given  $i$  there is a.s. a unique point of the process,  $x_i^*(N)$  say, associated with a given distance  $r_i$ . In the exceptional set (of probability 0) where there is no such unique point, we can put all the  $x_i^*(N)$  equal to zero.

It follows that the  $x_i^*(N)$  form a measurable enumeration of the points of the realization (Definition 9.1.XI), for the measurability of sets such as

$$\{N: x_i^*(N) \in A\} = \bigcap_n \bigcup_k \left\{ \frac{k}{n} \leq r_i(N) < \frac{k+1}{n}; N(A) > 0 \right\} \quad (A \in \mathcal{B}(\mathbb{R}^d))$$

implies that each  $x_i^*(N)$  is a well-defined random element of  $\mathbb{R}^d$ . In the sequel we mostly use the point of a realization  $N$  that is closest to the origin, which we denote for brevity by

$$x^*(N) = x_1^*(N). \quad (13.3.7)$$

One immediate use for  $x^*(N)$  is to develop an inversion formula extending (13.3.2) to the case  $\mathcal{X} = \mathbb{R}^d$ . For this we need the concept of a *Voronoi polygon*: for any given realization  $N$  of a nonnull simple point process, and any point  $u \in \mathbb{R}^d$ , the Voronoi polygon  $V_u(N)$  with ‘centre’  $u$  is the subset

$$V_u(N) = \{x: \|x - u\| < \|x - x_j\|, x_j \in N \text{ and } x_j \neq u\}$$

of points  $x \in \mathbb{R}^d$  that lie closer to  $u$  than to any point  $x_j$  of  $N$ . Consequently,

$$\int_{V_u(N)} N(dx) = \begin{cases} 0 & \text{if } u \notin N, \\ 1 & \text{when } u \in N \text{ and } N \text{ is simple.} \end{cases} \quad (13.3.8)$$

In particular, if  $N_0 \in \mathcal{N}_0^{\#*}$ , we write  $V_0(N_0)$  for the Voronoi polygon about the origin (which is a point of  $N_0$ ), and note that

$$\int_{V_0(N_0)} N_0(dx) = N_0(V_0(N_0)) = 1,$$

so that  $I_{V_0(N_0)}(x) = k(x, N_0)$  where  $k(\cdot)$  is a function as in (13.2.9) and used in the inversion formulae (13.2.11) and its variants. The inversion formula itself takes the form, for nonnegative  $\mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#*})$ -measurable  $g(\cdot)$ ,

$$\mathbb{E}_{\mathcal{P}}(g(N)) = \int_{\mathcal{N}_0^{\#*}} \int_{V_0(N_0)} g(S_x N_0) dx \mathcal{R}(dN_0), \quad (13.3.9a)$$

where the measure  $\mathcal{P}$  on  $\mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#*})$  is a probability measure provided  $\mathcal{R}$  satisfies the normalizing condition

$$\int_{\mathcal{N}_0^{\#*}} \int_{V_0(N_0)} dx \mathcal{R}(dN_0) = 1. \quad (13.3.9b)$$

If  $\mathcal{R}$  is totally finite with mass  $m$ , say  $\mathcal{R} = m\mathcal{P}_0$ , the right-hand side can be written as

$$m \mathbb{E}_{\mathcal{P}_0} \left( \int_{V_0(N(S_x N_0))} g(S_x N_0) dx \right), \quad (13.3.9c)$$

in which case the left-hand side of (13.3.9a) corresponds to the probability distribution of a stationary simple point process with finite mean density  $m$  and Palm distribution  $\mathcal{P}_0$  [the proof of this fact is similar to that of (13.3.2) and left to Exercise 13.3.4(a)].

The inversion formula (13.3.9a) is not as useful in  $\mathbb{R}^d$  ( $d \geq 2$ ) as is (13.3.2) in  $\mathbb{R}$ . This is chiefly a reflection of the increased structural complexity of the higher-dimensional Euclidean spaces, but a few simple results can be deduced from it, such as the intuitively obvious fact that the expected hypervolume (i.e., Lebesgue measure) of the Voronoi polygon about the origin equals  $m^{-1}$  [see Exercise 13.3.4(b)]. It does not supply a full counterpart in  $\mathbb{R}^d$  of Theorems 13.3.I and 13.3.II, because it does not address the issue of defining a notion of stationarity for measures in  $\mathcal{N}_0^{\#*}$  extending that of interval-stationarity for point processes on the line; this extension must await the discussion of point-stationarity and Theorem 13.3.IX.

A further use of  $x^*(N)$  is in the next theorem, which establishes the conditional probability interpretation of the Palm distribution for point processes in general Euclidean spaces  $\mathbb{R}^d$ .

**Theorem 13.3.IV.** *Let  $N$  be a simple stationary point process in  $\mathbb{R}^d$  with finite mean rate  $m$ , distribution  $\mathcal{P}$ , and Palm distribution  $\mathcal{P}_0$ , and let  $\{A_n: n = 1, 2, \dots\}$  be a nested sequence of bounded Borel sets with nonempty interiors satisfying*

$$\text{diam}(A_n) \rightarrow 0 \quad (n \rightarrow \infty). \quad (13.3.10)$$

Then as  $n \rightarrow \infty$ ,

$$\ell(A_n)^{-1} \mathcal{P}\{N(A_n) > 0\} \rightarrow m. \quad (13.3.11)$$

If, furthermore, the sets  $\{A_n\}$  are spheres in  $\mathbb{R}^d$  centred at the origin, then for bounded continuous nonnegative Borel functions  $f$  on  $\mathcal{N}_{\mathcal{X}}^{\#*}$ ,

$$\mathbb{E}_{\mathcal{P}}(f(N) \mid N(A_n) > 0) \rightarrow \mathbb{E}_{\mathcal{P}_0}(f(N_0)). \quad (13.3.12)$$

PROOF. The first assertion (13.3.11) is a corollary to the discussion on intensities in Chapter 9 [see, in particular, (9.3.22) and Exercise 9.3.11]. A further corollary of the same discussion, which we need in the sequel, is that

$$\mathcal{P}\{N(A_n) \geq 2\} / \ell(A_n) \rightarrow 0 \quad (n \rightarrow \infty) \quad (13.3.13)$$

(see Proposition 9.3.XV). We approach the assertion at (13.3.12) via the following result.

**Proposition 13.3.V.** *Let  $N$ ,  $\mathcal{P}$ ,  $\mathcal{P}_0$ ,  $m$ , and  $\{A_n\}$  be as in Theorem 13.3.IV and  $x^*(N)$  as at (13.3.7). Then for bounded nonnegative  $\mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#*})$ -measurable  $f(\cdot)$ ,*

$$\mathbb{E}_{\mathcal{P}}(f(S_{x^*(N)}N) \mid N(A_n) > 0) \rightarrow \mathbb{E}_{\mathcal{P}_0}[f(N_0)]. \quad (13.3.14)$$

PROOF. Note first that

$$\begin{aligned} & \left| \mathbb{E}_{\mathcal{P}}(f(S_{x^*(N)}N) \mid N(A_n) > 0) - \mathbb{E}_{\mathcal{P}_0}(f(N_0)) \right| \\ & \leq (m\ell(A_n))^{-1} \left| \mathbb{E}_{\mathcal{P}}[f(S_{x^*(N)}N) I_{\{N(A_n) > 0\}}] - m\ell(A_n) \mathbb{E}_{\mathcal{P}_0}(f(N_0)) \right| \\ & \quad + (m\ell(A_n))^{-1} \mathbb{E}_{\mathcal{P}}[f(S_{x^*(N)}N) I_{\{N(A_n) > 0\}}] |1 - m\ell(A_n) / \mathcal{P}\{N(A_n) > 0\}| \\ & \equiv J_1 + J_2 \quad \text{say.} \end{aligned}$$

In  $J_2$ , the modulus of the difference converges to zero as  $n \rightarrow \infty$  by (13.3.11), and the multiplier remains finite because

$$\begin{aligned} \frac{\mathbb{E}_{\mathcal{P}}[f(S_{x^*(N)}N) I_{\{N(A_n) > 0\}}]}{m\ell(A_n)} & \leq \left( \sup_{N \in \mathcal{N}_{\mathcal{X}}^{\#*}} f(N) \right) \frac{\mathcal{P}\{N(A_n) > 0\}}{m\ell(A_n)} \\ & \rightarrow \sup_{N \in \mathcal{N}_{\mathcal{X}}^{\#*}} f(N) \quad \text{by (13.3.11),} \end{aligned}$$

and this supremum is finite by the boundedness assumption on  $f(\cdot)$ . Thus  $J_2 \rightarrow 0$  as  $n \rightarrow \infty$ . For  $J_1$ , we note first from the proof of Theorem 13.3.I and (13.2.8) that

$$m\ell(A_n) \mathbb{E}_{\mathcal{P}_0}(f(N_0)) = \mathbb{E}_{\mathcal{P}} \left( \int_{A_n} f(S_x N) N(dx) \right) = \mathbb{E}_{\mathcal{P}} \left( \sum_{x_i \in A_n} f(S_{x_i} N) \right).$$

It is thus enough to consider the difference

$$\frac{1}{\ell(A_n)} \left| \mathbb{E}_{\mathcal{P}} \left( f(S_{x^*(N)}N) I_{\{N(A_n) > 0\}} - \sum_{x_i \in A_n} f(S_{x_i} N) \right) \right|,$$

which certainly vanishes when  $N(A_n) = 0$ . When  $N(A_n) > 0$ , then we have  $x^*(N) \in A_n$ , implying that it can be identified with one of the  $x_i \in A_n$ , so the first term cancels with one of the elements of the sum. Consequently, the difference is dominated by  $(\sup_{N \in \mathcal{N}_{\mathcal{X}}^{\#*}} f(N)) \mathcal{P}\{N(A_n) \geq 2\} / \ell(A_n)$ , which tends to zero by (13.3.13).  $\square$

Resuming the proof of Theorem 13.3.IV, Proposition 13.3.V implies that it is enough to establish the convergence to zero of the difference

$$\begin{aligned} & |\mathbb{E}_{\mathcal{P}}[f(N) \mid N(A_n) > 0] - \mathbb{E}_{\mathcal{P}}[f(S_{x^*(N)}N) \mid N(A_n) > 0]| \\ & \leq \mathbb{E}_{\mathcal{P}}[|f(N) - f(S_{x^*(N)}N)| \mid N(A_n) > 0] \\ & \leq \mathbb{E}_{\mathcal{P}}[\sup_{x \in A_n} |f(S_{x^*(N)}N) - f(S_{x+x^*(N)}N)| \mid N(A_n) > 0], \end{aligned} \quad (13.3.15)$$

because  $x^*(N) \in A_n$  under the condition  $N(A_n) > 0$ . Fixing the set  $A_n$  for the supremum as  $A_{n_0}$  say, and letting  $n \rightarrow \infty$  for the conditioning, this last expression converges by Proposition 13.3.V to

$$\mathbb{E}_{\mathcal{P}_0}(\sup_{x \in A_n} |f(N_0) - f(S_x N_0)|).$$

Inasmuch as  $f$  is uniformly bounded and continuous, and the shift operation is continuous also, the argument of the supremum converges to zero pointwise as  $n_0 \rightarrow \infty$ , and then by dominated convergence the expectation must converge to zero.  $\square$

We note that in both Theorem 13.3.IV and Proposition 13.3.V the convergence results are sufficient to imply that both  $\mathcal{P}\{\cdot \mid N(A_n) > 0\}$  and  $\mathcal{P}\{S_{x^*}(\cdot) \mid N(A_n) > 0\}$  converge weakly to  $\mathcal{P}_0\{\cdot\}$  as  $n \rightarrow \infty$ . In fact the convergence in Proposition 13.3.V can be strengthened: see Exercise 13.3.6.

The results just proved also provide some kind of analogue to the differential form of the Palm–Khinchin equations given at (3.4.11) of Chapter 3. Even in the one-dimensional case, however, it is not easy to provide a completely satisfactory account by this differential approach [see Slivnyak (1962, 1966), Leadbetter (1972), and Exercise 3.4.4].

It is a far more difficult exercise to extend to  $\mathbb{R}^d$  the interval-stationarity interpretation of a stationary point process on  $\mathbb{R}$ . Results in this direction centre around the concepts of *point maps* and *point-stationarity*, initially considered by Mecke (1975) (the term point-stationarity is generally used in  $\mathbb{R}^d$  for what in  $\mathbb{R}^1$  has often been called *cycle stationarity*). One way of looking at the intervals in a one-dimensional point process is as mappings, each of which links two points of a realization. The underlying idea is that, even in higher dimensions, we may be able to define a family of mappings that link points of the process in such a way that invariance under these mappings provides a characterization of the Palm measure. What distinguishes the two cases  $\mathbb{R}^1$  and  $\mathbb{R}^d$  is that the mappings in  $\mathbb{R}^1$  can be restricted to points that are left or right neighbours, whereas in  $\mathbb{R}^d$  they need not have any particular proximity relation.

Extending the notation from the beginning of the section, with  $\mathcal{X} = \mathbb{R}^d$ , let  $\{x_i(N)\}$  be a measurable enumeration of the points (of support) of  $N \in \mathcal{N}_{\mathcal{X}}^{\#*}$ . We use  $x \in N$  as shorthand for  $x \in \text{supp}(N)$ , so  $x \in N \in \mathcal{N}_{\mathcal{X}}^{\#*}$  implies that  $N(\{x\}) = 1$ . In view of Proposition 12.1.VI and the fact that the point process is stationary, we assume that  $\mathcal{P}\{N(\mathcal{X}) = \infty\} = 1$ . Then we are in fact interested in that part of the subspace  $\mathcal{N}_0^{\#*}$  for which  $N(\mathcal{X}) = \infty$ , and the enumeration  $\{x_i(N)\}$  is countably infinite.

Let  $\Psi(N, x)$  be a measurable mapping from  $\mathcal{N}_{\mathcal{X}}^{\#*} \times \mathcal{X}$  to  $\mathcal{X}$ . Call  $\Psi$  *covariant* when for all  $N \in \mathcal{N}_{\mathcal{X}}^{\#*}$  and  $x, y \in \mathcal{X}$ ,

$$\Psi(S_y N, S_y x) \equiv \Psi(N - y, x - y) = \Psi(N, x) - y \equiv S_y \Psi(N, x), \quad (13.3.16)$$

so for covariant  $\Psi$ ,  $S_x \Psi(N, x) = \Psi(S_x N, 0)$ .

**Definition 13.3.VI.** A point map is a covariant mapping  $\Psi: \mathcal{N}_{\mathcal{X}}^{\#*} \times \mathcal{X} \mapsto \mathcal{X}$  such that

$$\Psi(N, x) \begin{cases} \in N & \text{if } x \in N, \\ = x & \text{if } x \notin N. \end{cases} \quad (13.3.17)$$

For covariant  $\Psi(N, \cdot)$ ,  $\Psi(N, x) = S_{-x} \Psi(S_x N, S_x x) = S_{-x} \Psi(S_x N, 0)$ , so  $\Psi$  being a point map and  $x \in N$  is the same as  $S_x N \in \mathcal{N}_0^{\#*}$ , and the first (and critical) case of (13.3.17) is equivalent to requiring that

$$\text{for } N \in \mathcal{N}_0^{\#*}, \Psi(N, 0) \text{ is again a point of } N. \quad (13.3.18)$$

[Earlier work in Thorisson (1995, 2000) and Heveling and Last (2005) introduced a point map as a mapping from  $\mathcal{N}_{\mathcal{X}}^{\#*} \mapsto \mathcal{X}$  satisfying (13.3.18). Their subsequent work starts from Definition 13.3.VI.]

We define the composition of two point maps  $\Psi_1$  and  $\Psi_2$  as follows. Take  $x \in N \in \mathcal{N}_0^{\#*}$  and consider  $(\Psi_2 \circ \Psi_1)(N, x) \equiv \Psi_2(N, \Psi_1(N, x))$ . Because  $y = \Psi_1(N, x) \in N$  by (13.3.18), it follows that  $z \in N$  also, that is,  $\Psi_2 \circ \Psi_1$  is well defined as a point map and, with  $N$  specified, it is indeed the usual composition of the two point maps.

Consequently, starting from  $x_0 \equiv x \in N$ , defining  $x_n = \Psi(N_0, x_{n-1})$  for  $n = 1, 2, \dots$  yields a sequence of points in  $N$ , but they need not be distinct, nor need they enumerate the elements of  $N$  when  $N$  is countable. Now our interest is in countably infinite point sets  $N \in \mathcal{N}_0^{\#*}$ , so the sequence  $\{x_n\}$  just defined by the  $n$ th point map iterates of  $x_0$  is well defined, and either constitutes an infinite chain or reduces to recurring cycles.

Call a point map *bijective* if for any  $N \in \mathcal{N}_{\mathcal{X}}^{\#*}$ , the function  $\Psi(N, \cdot)$  is a one-to-one mapping on  $\mathcal{X}$ . It is evident that a bijective point map  $\Psi$  as at (13.3.17) maps  $N$  in a one-to-one manner onto itself, and for  $x \notin N$  it is just the identity. We can then define the inverse point map  $\Psi^{-1}$  for  $z \in N$  as the solution  $x$  of  $z = \Psi(N, x)$ , and write  $x = \Psi^{-1}(N, z)$ .

We associate with every point map  $\Psi$  the *point shift*  $S^{\Psi}$  on  $\mathcal{N}_{\mathcal{X}}^{\#*}$  defined by

$$S^{\Psi} N = S_{\Psi(N, 0)} N. \quad (13.3.19)$$

Because  $\Psi(N, 0) = 0$  for  $N \notin \mathcal{N}_0^{\#\#}$ ,  $S^\Psi N = N$  unless  $N \in \mathcal{N}_0^{\#\#}$ , in which case the effect of  $S^\Psi$  on  $N$  is to shift the points  $\{x_i(N)\}$  of  $N$  to  $\{x_i(N) - \Psi(N, 0)\}$ . Now by virtue of  $\Psi$  as a point map and  $0 \in N \in \mathcal{N}_0^{\#\#}$ ,  $\Psi(N, 0) \in N$ , that is, it is one of the points enumerated as  $\{x_i(N)\}$ , and this point therefore gives  $x_i(N) - \Psi(N, 0) = 0$ ; that is,  $0 \in S_{\Psi(N, 0)} N$  when  $N \in \mathcal{N}_0^{\#\#}$ . Thus,  $S^\Psi N$  shifts the origin to  $\Psi(N, 0)$ , and  $S^\Psi$  defines an operator on  $\mathcal{N}_\mathcal{X}^{\#\#}$  that maps the space  $\mathcal{N}_0^{\#\#}$  into itself (although not necessarily onto itself).

For any  $N \in \mathcal{N}_\mathcal{X}^{\#\#}$ , the point sets  $N$  and its shifted version  $S_y N$  (any  $y \in \mathcal{X}$ ) consist of points whose relative positions in  $\mathcal{X}$  are not changed by the shift. However, the point shift  $S^\Psi$  when operating on  $N \in \mathcal{N}_0^{\#\#}$  yields a point set  $S^\Psi N \in \mathcal{N}_0^{\#\#}$  that in general differs from the original point set.

The most familiar examples are for right- and left-shifts for sequences in  $\mathbb{R}^1$  as in the next example.

**EXAMPLE 13.3(b)** *Right- and left-shifts as bijective point maps in  $\mathbb{R}^1$ .* For  $\mathcal{X} = \mathbb{R}$ , enumerate the points of  $N \in \mathcal{N}_\mathcal{X}^{\#\#}$  as  $\{x_i\} \equiv \{x_i(N) : i = 0, \pm 1, \dots\}$  where  $x_i < x_{i+1}$  with  $x_0 \leq 0 < x_1$ . For  $N \in \mathcal{N}_0^{\#\#}$  with  $N(\mathbb{R}_-) = N(\mathbb{R}_+) = \infty$ , define  $\Psi(N, 0) = x_1(N)$  [i.e., for such  $N$ ,  $\Psi(N, 0)$  is the first point of  $N$  to the right of the origin,  $x_1(N)$ ], and for other  $N \in \mathcal{N}_0^{\#\#}$  define  $\Psi(N, x) = x$ . Because  $\Psi$  is a point map and therefore covariant,

$$\Psi(N, x_i(N)) = \Psi(N - x_i(N), 0) + x_i(N) = x_{i+1}(N).$$

Again because  $\Psi$  is a point map,  $\Psi(N, x) = x$  whenever  $x \notin N$ , and therefore we have shown that this point map is indeed simply a right-shift.

When  $N \ni 0$ , the corresponding point shift  $S^\Psi$  shifts the origin for  $S^\Psi N$  to  $x_1(N)$ , thus subtracting  $x_1(N)$  from each of the points of  $N$ , yielding  $x_i(S^\Psi N) = x_{i+1}(N) - x_1(N)$ . For example, we then have  $\Psi(S^\Psi N, 0) = x_2(N) - x_1(N)$ .

For any  $x \in N$ ,  $x_\ell(N)$  say, we have  $\Psi(N, x_\ell) = \Psi(N - x_\ell, 0) + x_\ell = x_{\ell+1}(N)$ , so that  $\Psi^{-1}(N, x_\ell) = x_{\ell-1}(N)$ . It then follows that if we define  $\Psi^{n+1} = \Psi \circ \Psi^n$  with  $\Psi^1 = \Psi$  (and thus  $\Psi^{-n} = (\Psi^{-1})^n$ ), that  $\{\Psi^n(N_0, x_0)\} = \{x_n(N_0)\}$ , i.e. this sequence enumerates the points of  $N_0$ .

Observe that the point map  $\Psi_k \equiv \Psi^k$  above has the property that  $\Psi_k(N, x)$  maps  $x$  to the  $k$ th point of  $N$  to the right of  $x$ , and is again a bijective point map. □

The earlier theorems of this chapter can be reinterpreted as assertions that the Palm relations define a one-to-one correspondence between stationary point processes on  $\mathcal{N}_\mathbb{R}^{\#\#}$  and point processes on  $\mathcal{N}_0^{\#\#}$  which are invariant under the bijective point maps taking one point of the realization to the next.

When  $\mathcal{X} = \mathbb{R}^d$  with  $d > 1$ , however, the problem of finding a suitable family of bijective point maps becomes nontrivial. The next example, due to Häggström and quoted in Thorisson (2000, §9.2.8), underlies the extension of Theorem 13.3.II that we describe shortly.

EXAMPLE 13.3(c) *Mutual nearest neighbour matching.* Given distinct points  $x', x''$  in  $N \in \mathcal{N}_0^{\#*}$ ,  $x'$  and  $x''$  are mutual nearest neighbours in  $N$  when the element of  $N$  closest to  $x'$  is  $x''$  and the element of  $N$  closest to  $x''$  is  $x'$ . Define a point map for  $N \in \mathcal{N}_0^{\#*}$  by

$$\Psi(N, 0) = \begin{cases} x & \text{if } x \text{ and } 0 \text{ are mutual nearest neighbours in } N, \\ 0 & \text{otherwise.} \end{cases}$$

Then as  $x$  runs through all the elements of  $N$ ,

$$\Psi(N, x) = \begin{cases} y_x & \text{if } y_x \in N \text{ is mutual nearest neighbour to } x, \\ x & \text{otherwise,} \end{cases} \quad (13.3.20)$$

so  $\Psi$  is a bijective point map. For  $\Psi$  defined in this way, we see that for  $x \in N$ , either  $x$  has a mutual nearest neighbour  $y_x \in N$  or not, and

$$\Psi(N, \Psi(N, x)) = \begin{cases} \Psi(N, y_x) & \text{if the first case,} \\ \Psi(N, x) & \text{otherwise,} \end{cases} = x \text{ in either case,}$$

so that  $\Psi^{-1} = \Psi$ ; that is, this point map is self-inverse.  $\square$

A simple case of a systematic approach to the construction of bijective maps in  $\mathbb{R}^d$  due to Heveling and Last starts from the following definition.

**Definition 13.3.VII.** For any bounded Borel set  $B$ , a  $B$ -selective point map  $\Psi_B(N, x)$  is a mapping  $\mathcal{N}_0^{\#*} \times \mathcal{X} \mapsto \mathcal{X}$  such that

- (1) if  $N \cap [B \cup (-B)]$  consists of a single point  $x_B$  say, and if also 0 is the only point of  $N$  in  $(B + x_B) \cup (-B + x_B)$ , then  $\Psi_B(N, 0) = x_B$ ;
- (2) for  $x \in N$ ,  $\Psi_B(N, x)$  is defined from (1) and the covariant property; and
- (3) in all other cases,  $\Psi_B(N, x) = x$ .

Property (2) here ensures that  $\Phi_B(N, x)$  is in fact a point map: we call it the  $B$ -selective point map. Observe that the conditions ensure that a unique point of  $\mathbb{R}^d$  is defined for every  $N$  and  $x$ . Note, in particular, that  $\Psi_B(N, 0) = 0$  whenever there are two or more points of  $N$  in  $B \cup (-B)$ .

Because  $B \cup (-B)$  is symmetric and  $x_B = \Psi_B(N, 0) = x_{-B}$ , the map  $\Psi_B$  is in fact bijective. To see this, if a point  $x_B \neq 0$  exists as in (1) for given  $N \in \mathcal{N}_0^{\#*}$ , the associated point shift  $S^{\Psi_B}$  interchanges 0 and  $x_B$ , leaving the other points of  $N$  unchanged.  $\Psi_B$  is also self-inverse in the sense that  $\Psi_B \circ \Psi_B = I$  (identity).

The aim now is to build up a comprehensive set of mappings via an exhaustive area search by letting  $B$  vary over a sufficiently wide class of testing sets. To this end, let  $\{\mathcal{T}_n\} = \{\{B_{ni}\}\}$  be a nested system of tilings of  $\mathbb{R}^d$ , namely, a system of partitions of  $\mathcal{X}$  like a dissecting system, each  $\mathcal{T}_n$  consisting of a denumerable number of disjoint sets for which the norm  $\|\mathcal{T}_n\| = \sup_i \{\text{diam}(B_{ni})\} \rightarrow 0$  as  $n \rightarrow \infty$ . As the sets get smaller, their ability to distinguish between points of the realization increases, until finally every



point in the realization can be interchanged with the origin by one of the point shifts  $S^{\Psi_{B_{ni}}}$ . It is now plausible that requiring a measure on  $\mathcal{N}_0^{\#*}$  to be invariant under the denumerable family of point shifts

$$S^{ni} \equiv S^{\Psi_{B_{ni}}}, \quad (13.3.21)$$

may be a condition comparable to interval stationarity in the one-dimensional case.

This is a little too simple in general, because of the complexities of the possible configurations of points in a general element of  $\mathcal{N}_\mathcal{X}^{\#*}$ . To avoid this difficulty, we impose on the point process a property similar to but stronger than simplicity. We say that the element  $N$  of  $\mathcal{N}_0^{\#*}$  in  $\mathbb{R}^d$  is *unaligned* if it contains no equidistant pairs of points that are collinear. Exercise 13.3.11 shows that the set  $\mathcal{U}_\mathcal{X}$  of all unaligned elements in  $\mathcal{X} = \mathbb{R}^d$  is measurable. Then the point process as a whole is *unaligned* if  $\mathcal{P}(\mathcal{U}_\mathcal{X}) = 1$ , and we show that under this condition, invariance under the point shifts  $S^{ni}$  is enough to imply stationarity of the original point process.

In the general case treated by Heveling and Last (2005), this collection of point shifts has to be augmented by additional point shifts that deal with situations where the configurations may contain finite chains of equidistant collinear points. In any case, the aim is to establish the existence of a sufficiently comprehensive set of bijective point maps to justify the far-reaching generalization of the one-dimensional results outlined in the definition and theorem which follow.

**Definition 13.3.VIII.** A  $\sigma$ -finite measure  $\mathcal{R}$  on  $\mathcal{B}(\mathcal{N}_0^{\#*})$  is point-stationary when it is invariant under all bijective point maps.

**Theorem 13.3.IX.** If the measure  $\mathcal{P}$  on  $\mathcal{B}(\mathcal{N}_\mathcal{X}^{\#*})$  corresponds to a stationary point process on  $\mathcal{X} = \mathbb{R}^d$ , then its associated Palm measure  $\mathcal{R}$  on  $\mathcal{B}(\mathcal{N}_0^{\#*})$  is point-stationary.

Conversely, if the  $\sigma$ -finite measure  $\mathcal{R}$  on  $\mathcal{B}(\mathcal{N}_0^{\#*})$  is unaligned, satisfies the normalizing condition (13.3.5), and is invariant under the family of point shifts  $\{S^{ni}\}$  associated with the sets  $B_{ni}$  from a nested sequence of symmetric tilings  $\{\mathcal{T}_n\}$  satisfying  $\|\mathcal{T}_n\| \rightarrow 0$ , then it is the Palm measure associated with an unaligned stationary point process.

PROOF. Start by appealing to the general relations between stationary random measures and their associated Palm measures that are embodied in equation (13.2.6) and its specializations (13.2.10) and (13.2.11). We have to verify that if  $\mathcal{R}$  is the Palm measure for a stationary point process on  $\mathcal{X} = \mathbb{R}^d$  and  $h(\cdot)$  a bounded, measurable, nonnegative function on  $\mathcal{N}_\mathcal{X}^{\#*}$ , the quasi Palm expectation operator  $E_\mathcal{R}$  defined by  $E_\mathcal{R}[h(N)] \equiv \int_{\mathcal{N}_0^{\#*}} h(N) \mathcal{R}(dN)$  satisfies

$$E_\mathcal{R}[h(S^\Psi N)] = E_\mathcal{R}[h(N)] \quad (13.3.22)$$

for any bijective point map  $\Psi$ .

Writing  $j(x) = I_{\mathbb{U}^d}(x)$  for the indicator function of the unit cube in  $\mathbb{R}^d$ , (13.2.10b) yields

$$\begin{aligned}
 \mathbb{E}_{\mathcal{R}}[h(S^{\Psi}N)] &= \int_{\mathcal{N}_0^{\#*}} \int_{\mathcal{X}} j(x) h\{S^{\Psi}(S_x N)\} N(dx) \mathcal{P}(dN) \\
 &= \int_{\mathcal{N}_0^{\#*}} \int_{\mathcal{X}} j(x) h\{S_{\Psi(S_x N, 0)}(S_x N)\} N(dx) \mathcal{P}(dN) \quad \text{by (13.3.19),} \\
 &= \int_{\mathcal{N}_0^{\#*}} \int_{\mathcal{X}} j(x) h\{S_{\Psi(N, x) - x}(S_x N)\} N(dx) \mathcal{P}(dN) \quad \text{by (13.3.16),} \\
 &= \int_{\mathcal{N}_0^{\#*}} \int_{\mathcal{X}} j(x) h\{S_{\Psi(N, x)} N\} N(dx) \mathcal{P}(dN) \quad \text{because } S_{-x} S_x N = N, \\
 &= \int_{\mathcal{N}_0^{\#*}} \int_{\mathcal{X}} j(\Psi^{-1}(N, y)) h(S_y N) N(dy) \mathcal{P}(dN) \quad \text{setting } y = \Psi(N, x).
 \end{aligned}$$

Because  $\Psi$  is bijective, so is  $\Psi^{-1}$ . Returning to the basic relation (13.2.6), and using Fubini's theorem, the last term in the above chain can be evaluated as

$$\int_{\mathcal{N}_0^{\#*}} h(N) \left( \int_{\mathcal{X}} j(\Psi^{-1}(N, y) - y) dy \right) \mathcal{R}(dN) = \mathbb{E}_{\mathcal{R}}[h(N)].$$

Thus, (13.3.22) holds and the direct part is proved.

To prove the converse part, call a family  $\{\Psi_{ni}\}$  of bijective point maps *distinctive* if, for fixed  $n$ ,  $N \in \mathcal{N}_0^{\#*}$ , and  $x \in N$ , the unions  $\bigcup_r \Psi_{ni}^r(N, x)$  are disjoint for distinct  $i = 1, 2, \dots$ , and *exhaustive* if

$$N(\cdot) = \lim_{n \rightarrow \infty} \sum_{i,r} \delta_{\Psi_{ni}^r(N, 0)}(\cdot) \quad (\text{every } N \in \mathcal{N}_0^{\#*}) \quad (13.3.23)$$

[here,  $\Psi^r$  denotes the  $r$ -fold product as below (13.3.18), else see Exercise 13.3.10]. The last condition implies that there are sufficiently many point maps to distinguish the points in any realization of a simple point process. The rest of the proof consists of two stages.

(1°) *If  $\mathcal{R}$  is  $\sigma$ -finite, and invariant under a family of bijective point shifts which is both distinctive and exhaustive, then it is the Palm measure of some stationary measure.* Let  $\Psi$  be a bijective point map from the specified family, and define its cycle length relative to the realization  $N \in \mathcal{N}_0^{\#*}$  by

$$C_{\Psi}(N) = \inf\{r: \Psi^r(N, 0) = 0\}.$$

For any given value  $k$  of the cycle length, for the successive points within that cycle, invariance under  $\Psi$  implies that for any nonnegative measurable function  $f(N, x)$ ,

$$\int_{\{C_{\Psi}=k\}} f(S^{\Psi}N, -\Psi(N, 0)) \mathcal{R}(dN) = \int_{\{C_{\Psi}=k\}} f(N, \Psi^{-1}(N, 0)) \mathcal{R}(dN). \quad (13.3.24)$$

For it is easy to check that the length of the cycle is invariant under  $S^\Psi$ , and applying  $S^{\Psi^{-1}}$  to the terms in the left-hand side leads directly to the form in the right-hand side [see Exercise 13.3.10(d)]. The same equation holds also for each of the iterates of  $\Psi$ , because the length of the cycle is not affected, and the remaining features again follow from properties of  $\Psi^{-1}$ .

Still continuing with  $\Psi$  fixed, let  $N_\Psi$  denote the (reduced) realization, derived from  $N \in \mathcal{N}_0^{\#*}$ , and comprising the points  $\{\Psi^r(N, 0): r = 1, \dots, C_\Psi\}$ . Applying (13.3.24) to each of the iterates in turn, yields for each  $k$ ,

$$\begin{aligned} \int_{\{C_\Psi=k\}} \left( \int_{\mathcal{X}} f(S_x N, -x) N_\Psi(dx) \right) \mathcal{R}(dN) \\ = \int_{\{C_\Psi=k\}} \left( \int_{\mathcal{X}} f(N, x) N_\Psi(dx) \right) \mathcal{R}(dN). \end{aligned}$$

This result holds for all values of the cycle length, including the case  $k = \infty$ , so we can amalgamate the above equations for all  $k > 0$  (omitting the zero iterate to avoid duplication) and obtain

$$E_0 \left( \int_{\mathcal{X}} f(S_x N, -x) (N_\Psi - \delta_0)(dx) \right) = E_0 \left( \int_{\mathcal{X}} f(N, x) (N_\Psi - \delta_0)(dx) \right). \quad (13.3.25)$$

We next amalgamate these equations also over the mappings  $\Psi_{ni}$ , holding  $n$  fixed, and summing over  $i$ . Because the family is distinctive, and the origin is excluded, there are no overlaps. Let  $N_n$  denote the point process obtained by amalgamating all the points in all the  $(N_{\Psi_{ni}} - \delta_0)$  for  $i = 1, 2, \dots$ , and including the origin just once. Then (13.3.25) holds in the form

$$E_0 \left( \int_{\mathcal{X}} f(S_x N, -x) N_n(dx) \right) = E_0 \left( \int_{\mathcal{X}} f(N, x) N_n(dx) \right).$$

Finally, because the family is exhaustive, we can let  $n \rightarrow \infty$ , so that  $N_n \rightarrow N_0$  for each  $N_0$ , leading to

$$E_0 \left( \int_{\mathcal{X}} f(S_x N, -x) N(dx) \right) = E_0 \left( \int_{\mathcal{X}} f(N, x) N(dx) \right).$$

It now follows from Theorem 13.2.VIII that  $\mathcal{R}$  is the Palm measure on  $\mathcal{N}_0^{\#*}$  for some stationary measure;  $\mathcal{R}$  is a probability measure when the normalization condition at (13.3.6) holds.

(2°) *The family of  $B$ -selective bijective point maps  $\Psi_{ni}$  derived from the sets  $\{B_{ni}\}$  of the symmetric tilings  $\mathcal{T}_n$  is both distinctive and exhaustive over the set of unaligned elements of  $\mathcal{N}_0^{\#*}$ .* Because the  $\Psi_{ni}$  are self-inverse, they are also idempotent, so that in considering the powers  $\Psi_{ni}^r$  required by the distinctiveness condition it is sufficient to take  $r = 1$ . In this case the condition merely requires the points  $x_B$  and  $x_{B'}$  identified by distinct  $B$  and  $B'$  to be

distinct; this follows trivially from the fact the subsets in a given level of the tiling are by definition disjoint.

In considering exhaustiveness, the crucial point here is that when the point process is unaligned,  $S^{ni} \equiv S^{\Psi_{ni}}$  either leaves the point at the origin invariant, or interchanges the point at the origin with the unique point in  $B_{ni}$ . Under these circumstances, every point in the realization  $N_0$  will ultimately be exchanged with the origin by one of the  $S^{ni}$ , without altering other points in the configuration. Thus it will ultimately appear in the sum at (13.3.23). Note that this will not be the case if the realization contains equi-spaced sequences of more than two collinear points, for then at least one point pair  $x_1, -x_1$  will be repeated in every set  $B_{ni}$  containing at least one of the two points, sending the value of  $\Psi^{ni}$  to 0, and hence ensuring that  $x_1$  never appears in the union. The converse is proved.  $\square$

The complete version of this theorem in Heveling and Last (i.e., without the restriction to unaligned processes) allows an equivalence statement for point processes in  $\mathbb{R}^d$  analogous to that for point processes on the line given by Proposition 13.3.II as follows (see their paper for proof).

**Theorem 13.3.X.** *There is a one-to-one correspondence between the distributions of simple stationary point processes in  $\mathbb{R}^d$  (i.e., on  $\mathcal{N}_{\mathcal{X}}^{\#*}$  with  $\mathcal{X} = \mathbb{R}^d$ ) and point-stationary measures on  $\mathcal{N}_0^{\#*}$  satisfying the normalization conditions  $\mathcal{P}\{N = \emptyset\} = 0$  and (13.3.9b).*

In fact even the restriction to probability measures can be dropped, as in the discussion around Theorem 13.2.VIII, and the statement presented in symmetric form between  $\sigma$ -finite stationary measures on  $\mathcal{N}_{\mathcal{X}}^{\#*}$  and  $\sigma$ -finite point-stationary measures on  $\mathcal{N}_0^{\#*}$ .

## Exercises and Complements to Section 13.3

*Note:* Assume below that  $N$  is nonnull, that is,  $\mathcal{P}\{N = \emptyset\} = 0$ .

### 13.3.1 Variants of Theorems 13.3.I–II.

- (a) [cf. (13.3.2) and (13.2.6)]. Substitute  $g(x, N) = h(x, N)k(x, N)$  with  $k(\cdot)$  as in (13.3.3) to show that

$$\mathbb{E}_{\mathcal{P}}(h(t_1(N), S_{t_1(N)}N)) = m\mathbb{E}_{\mathcal{P}_0}\left(\int_0^{t_1(N_0)} h(x, N_0) dx\right).$$

- (b) Use (13.3.2) to show that  $m = \mathbb{E}_{\mathcal{P}_0}(t_1(N_0))$ , and hence write (13.3.2) as

$$\mathbb{E}_{\mathcal{P}}(g(N)) = \mathbb{E}_{\mathcal{P}_0}\left(\int_0^{t_1(N_0)} g(S_t N_0) dt\right) / \mathbb{E}_{\mathcal{P}_0}(t_1(N_0)).$$

- 13.3.2 For simple stationary  $N$  use (13.3.2) to find the joint distribution of  $X = t_1(N)$  and  $\tau_1 \equiv t_1(N) - t_0(N)$  in terms of the d.f.  $F(\cdot)$  of  $\tau_1$ . In particular, show the following.

- (a) The joint distribution for  $(X, \tau_1)$  has a density function representation

$$f(u, v) \, du \, dv = \begin{cases} mv^{-1}(1 - F(v)) \, du \, dv & (0 \leq u \leq v), \\ 0 & (u > v). \end{cases}$$

- (b) The conditional distribution of  $X$  given  $\tau_1$  is uniform on  $(0, \tau_1)$ .  
 (c) The conditional distribution of  $X$  given the whole sequence  $\{\tau_n: n = 0, \pm 1, \dots\}$  is uniform on  $(0, \tau_1)$ .

[Hint: Let  $A \in \mathcal{B}(\mathcal{N}_{\mathbb{R}}^{\#*})$  belong to the sub- $\sigma$ -field generated by the  $\tau_i$  and start from the definition of conditional expectation, so that for measurable  $h$ ,

$$\mathbb{E}(h(X)I_A \mid \sigma\{\tau_i\}) = I_A(\{\tau_i\}) \int_0^{\tau_1} h(x) \, dx.]$$

- 13.3.3 Use (13.3.2) and Exercises 13.3.1–2 to provide an alternative derivation of the formulae in Exercise 3.4.4 for  $Q(B_k)$  and  $\Pr(B_k)$ .

- 13.3.4 (a) Define  $k(x, N) = 1$  if  $x = x^*(N)$ ,  $= 0$  otherwise. Verify that  $k(\cdot)$  so defined satisfies (13.2.9) and establish (13.3.9) from (13.2.14) as in the derivation of (13.3.2).

- (b) For a simple stationary point process in  $\mathbb{R}^d$  with mean rate  $m$ , use (13.3.9) to show that  $\mathbb{E}[\ell(V_0(N))] = \mathbb{E}[\int_{V_0(N)} dx] = 1/m$ .

- 13.3.5 Supply an alternative derivation of (13.3.11) from (13.3.3) and (13.3.5).

- 13.3.6 The conclusion of Proposition 13.3.V can be strengthened to  $\|\tilde{\mathcal{P}}_n - \mathcal{P}_0\| \rightarrow 0$ , where  $\tilde{\mathcal{P}}_n$  is the measure on  $\mathcal{N}_0^{\#*}$  induced by the conditional probabilities  $\mathbb{E}_{\mathcal{P}}(f(S_{x^*(N)}N) \mid N(A_n) > 0)$  and convergence is with respect to the variation norm. [Hint: The basic inequalities depend on  $f$  only through  $\|f\|$  and hence hold uniformly in  $f$  for  $\|f\| \leq 1$  say.]

- 13.3.7 Use the inversion equation (13.3.9) to deduce that under the conditions of Theorem 13.3.IV,  $\mathcal{P}\{N(A_n) > 1\} = o(\ell(A_n)) \equiv \ell(A_n)o(1)$  and  $\mathcal{P}\{N(A_n) = 1\} = m\ell(A_n)(1 + o(1))$  [cf. Theorem 1.2.12 of Franken *et al.* (1981)]. For simple stationary point processes on  $\mathbb{R}$  as in Theorem 13.3.I, the analogue of (13.3.9) is

$$\mathbb{E}_{\mathcal{P}}(g(N)) = m\mathbb{E}_{\mathcal{P}_0}\left(\int_{t_{-1}(N_0)/2}^{t_1(N_0)/2} g(S_t N_0) \, dt\right).$$

- 13.3.8 Under the assumptions that  $N$  is stationary, simple, and has finite mean rate  $m$ , show that in Theorem 13.3.IV there is a nested sequence  $\{A_n\}$  with  $\ell(A_n) \rightarrow 0$ , not satisfying (13.3.10), for which (13.3.11) fails. Investigate whether (13.3.11) holds with (13.3.10) but without  $\{A_n\}$  being nested, and whether (13.3.12) holds for more general sets  $A_n$  than spheres.

[Hint: Let the realizations span a lattice, and let  $A_n$  be the union of two small spheres with centres at two lattice points.]

- 13.3.9 Consider a Poisson process in  $\mathbb{R}^1$  with a point at 0. Suppose that the origin is shifted to the point of the process closest to the origin. Show that the shifted point process cannot be Poisson.

[Hint: The interval between the new origin and the old is shorter than the other interval with an endpoint at the old origin, so these intervals cannot be independently distributed (Thorisson, 2000, Example 9.2.1).]

13.3.10 *Properties of point maps.*

- (a) *Products.* Given two point maps  $\Psi_1$  and  $\Psi_2$ , define the product map  $\Psi_2 \circ \Psi_1$  as below (13.3.18).

Give a proof or counterexample to decide whether the product is (i) associative, and (ii) commutative. [*Hint:* Mecke (1975, §2) asserts that it is associative and distributive but not commutative.]

If  $\Psi_1$  and  $\Psi_2$  are bijective, check that  $\Psi_2 \circ \Psi_1$  is bijective. When  $\Psi_1$  is bijective, can  $\Psi_2 \circ \Psi_1$  be bijective without  $\Psi_2$  being bijective?

- (b) *Inverse.* The inverse  $\Psi^{-1}$  of a bijective point map is well defined [see further below (13.3.18)].

More generally, given a point map  $\Psi$ , define  $\Phi: \mathcal{N}_{\mathcal{X}}^{\#*} \times \mathcal{X} \mapsto \mathcal{X}$ , written  $\Phi(N, z) = x$ , to be any solution  $x$  of  $z = \Psi(N, x)$ . If there is a unique solution, check whether  $\Phi$  is (i) covariant, (ii) a point map, (iii) bijective.

- (c) When  $\Psi^{-1}$  is well defined, show that  $S^{\Psi^{-1}}(S^{\Psi}N) = S_0N = N$ .  
 (d) For bijective  $\Psi$ , show that  $\Psi(S^{\Psi}N, 0) = -\Psi^{-1}(N, 0)$  [cf. equation (4.12) of Heveling and Last (2005)].  
 (e) When the cycle length  $C_{\Psi}(N)$  as above (13.3.24) is finite and equal to  $k$  say, verify that  $\Psi^r(N, x) = (\Psi^{-1})^{(k-r)}(N, x)$ , where  $\Psi^r = \Psi \circ \dots \circ \Psi$  denotes the  $r$ -fold product map of  $\Psi$ .

## 13.4. Marked Point Processes, Ergodic Theorems, and Convergence to Equilibrium

In this section we further examine the role of the stationary Palm distribution  $\mathcal{P}_0$ , especially in questions related to ergodicity and convergence to equilibrium. Before doing so we outline briefly the extensions of the previous theory to marked random measures and MPPs. Initially we state the results for general random measures; in the examples and development subsequent to Example 13.4(b) we focus on MPPs. For convenience we generally assume that the ground process of any MPP has finite mean rate, and is simple.

We first examine the extension of the Campbell theory results to the marked case. As already mentioned in Section 13.1, application of the Radon–Nikodym derivative approach leads to families of local Palm distributions  $\mathcal{P}_{(x, \kappa)}$  indexed by an element  $(x, \kappa)$  in the product space  $\mathcal{X} \times \mathcal{K}$ . The Campbell measure itself becomes a *marked Campbell measure* on Borel sets of the space  $\mathcal{X} \times \mathcal{K} \times \Omega$ , where  $\Omega = \mathcal{M}_{\mathcal{X} \times \mathcal{K}}^{\#}$  in the canonical framework. When the MPP is stationary, the arguments of Section 12.2 show that the local families of Palm measures are invariant under shifts in the location, but the dependence on the mark remains. Thus we obtain a family of stationary Palm distributions  $\mathcal{P}_{(0, \kappa)}$  on  $\mathcal{B}(\mathcal{M}_{\mathcal{X} \times \mathcal{K}}^{\#})$  which in the point process case can be interpreted as the behaviour of the process conditional on the occurrence of a point at the origin with mark  $\kappa$ .

The other important ingredient in the marked case is the stationary mark distribution. This appeared first as the nonnormalized measure  $\nu(\cdot)$  on  $\mathcal{K}$  introduced in the discussion of marked random measures above Lemma 12.2.III.

When the ground process has finite mean rate  $m_g$ ,  $\nu$  is a totally finite measure, and we can write

$$\mathbb{E}[\xi(A \times K)] = \ell(A) \nu(K) = m_g \ell(A) \pi(K),$$

where  $\pi$  is the stationary mark distribution. In this situation the reduced Campbell measure itself takes the form

$$\check{C}_P(d\xi \times d\kappa) = \nu(d\kappa) \mathcal{P}_{(0,\kappa)}(d\xi) = m_g \pi(d\kappa) \mathcal{P}_{(0,\kappa)}(d\xi). \quad (13.4.1)$$

The arguments are outlined in Exercise 13.4.1.

In the ergodic theorems of Sections 12.2 and 12.6, in particular Theorem 12.6.VI, the limits of products of random measures averaged over increasing sets are related to the reduced moment measures. Proposition 13.4.I below develops similar results for more general functionals of the random measure, with the limit identified as an integral against  $\check{C}_P$ . The proposition is stated for marked random measures; the special case of MPPs is examined in detail in the discussion around Proposition 13.4.IV. Results for the unmarked case follow by letting the functional  $g$  be independent of  $\kappa$  and integrating over  $\kappa$ .

**Proposition 13.4.I.** *Let  $\xi$  be a strictly stationary, ergodic, marked random measure on  $\mathbb{R}^d \times \mathcal{K}$ , with probability measure  $\mathcal{P}$  on  $\mathcal{B}(\mathcal{M}_{\mathbb{R}^d \times \mathcal{K}}^\#)$ , finite ground rate  $m_g$ , stationary mark distribution  $\pi$ , and reduced Campbell measure (i.e., Palm measure)  $\check{C}_P$ . Let  $g(\xi, \kappa)$  be a  $\mathcal{B}(\mathcal{M}_{\mathcal{X}}^\# \times \mathcal{K})$ -measurable nonnegative function on  $\mathcal{M}_{\mathcal{X}}^\# \times \mathcal{K}$ . Then for any convex averaging sequence  $\{A_n\}$ , as  $n \rightarrow \infty$ ,*

$$\begin{aligned} \frac{1}{\ell(A_n)} \int_{A_n} g(S_x \xi, \kappa) \xi(dx \times d\kappa) &\rightarrow \int_{\mathcal{K}} \int_{\mathcal{M}_{\mathcal{X}}^\#} g(\psi, \kappa) \check{C}_P(d\psi \times d\kappa) \quad \mathcal{P}\text{-a.s.} \\ &= m_g \int_{\mathcal{K}} \int_{\mathcal{M}_{\mathcal{X}}^\#} g(\psi, \kappa) \mathcal{P}_{(0,\kappa)}(d\psi) \pi(d\kappa). \end{aligned} \quad (13.4.2a)$$

PROOF. The result is an extension of the individual ergodic Theorem 12.2.II and the approximation arguments used in deriving Theorem 12.2.IV. As in the latter theorem, we give details mainly for the unmarked case. Suppose first that  $g$  is a nonnegative measurable function on  $\mathcal{M}_{\mathcal{X}}^\#$  satisfying  $\int_{\mathcal{M}_{\mathcal{X}}^\#} g(\xi) \check{C}_P(d\xi) < \infty$   $\mathcal{P}$ -a.s., and introduce the function

$$f(\xi) = \int_{\mathcal{X}} g(S_u \xi) g_\epsilon(u) \xi(du),$$

where  $g_\epsilon(\cdot)$  is a continuous function ‘close’ to a  $\delta$ -function and integrating to 1. As in the proof of Theorem 12.2.IV, we find

$$\begin{aligned} \int_{A_n^\epsilon} f(S_x \xi) dx &= \int_{\mathcal{X}} g_\epsilon(y - x) \int_{A_n^\epsilon} g(S_y \xi) \xi(dy) dx \\ &\geq \int_{\mathcal{X}} g_\epsilon(y - x) \int_{A_n^{-\epsilon}} g(S_y \xi) \xi(dy) dx = \int_{A_n^{-\epsilon}} f(S_x \xi) dx. \end{aligned}$$

From Theorem 12.2.II, using also (13.2.5), we have, as  $n \rightarrow \infty$ ,

$$\begin{aligned} \frac{1}{\ell(A_n)} \int_{A_n} f(S_x \xi) dx &\rightarrow E(f(\xi)) = \int_{\mathcal{X}} g_\epsilon(u) du \int_{\mathcal{M}_x^\#} g(\psi) \check{C}_{\mathcal{P}}(d\psi) \\ &= \int_{\mathcal{M}_x^\#} g(\psi) \check{C}_{\mathcal{P}}(d\psi), \end{aligned}$$

with similar results if  $A_n$  is replaced by  $A_n^\epsilon$  or  $A_n^{-\epsilon}$ . Because  $\epsilon$  is arbitrary, (13.4.2) follows when the limit is finite  $\mathcal{P}$ -a.s. If not, then replace  $g(\cdot)$  by an increasing sequence of functions  $\{g_r(\cdot)\}$  for which  $g_r(\xi) \uparrow g(\xi)$  ( $r \rightarrow \infty$ ), as, for example,  $g_r(\xi) = \min(g(\xi), r\alpha(\xi))$ , where  $\alpha(\cdot)$  is as in the discussion below (13.2.4). Then for every  $r$ ,

$$\liminf_{n \rightarrow \infty} \frac{1}{\ell(A_n)} \int_{A_n} g(S_x \xi) \xi(dx) \geq \int_{\mathcal{M}_x^\#} g_r(\xi) \check{C}_{\mathcal{P}}(d\xi),$$

and the right-hand side  $\rightarrow \infty$  as  $r \rightarrow \infty$ .

For the general marked case we start from a nonnegative function  $g(\xi, \kappa)$  satisfying  $\int_{\mathcal{M}_x^\# \times \mathcal{K}} g(\xi, \kappa) \check{C}_{\mathcal{P}}(d\xi \times d\kappa) < \infty$   $\mathcal{P}$ -a.s., and introduce the function

$$f(\xi) = \int_{\mathcal{X} \times \mathcal{K}} g(S_u \xi, \kappa) g_\epsilon(u) \xi(du \times d\kappa).$$

The rest of the proof then proceeds much as for the unmarked case. □

Note that for  $g(\xi)$  a function of the realization  $\xi$  only, (13.4.1) yields

$$\begin{aligned} \frac{1}{\ell(A_n)} \int_{A_n} g(S_x \xi) \xi(dx) &\rightarrow \int_{\mathcal{K}} \int_{\mathcal{M}_x^\#} g(\psi) \check{C}_{\mathcal{P}}(d\psi \times d\kappa) \quad \mathcal{P}\text{-a.s.} \\ &= m_g \int_{\mathcal{M}_x^\#} g(\psi) \bar{\mathcal{P}}_0(d\psi), \end{aligned} \tag{13.4.2b}$$

where  $\bar{\mathcal{P}}_0$  is not the Palm distribution of the ground process, but rather the averaged form  $\int_{\mathcal{K}} \mathcal{P}_{(0, \kappa)}(\cdot) \pi(d\kappa) \equiv \bar{\mathcal{P}}_0(\cdot)$ . The difference arises because the realizations  $\xi$  here are for a marked process, whereas the realizations of the ground process are unmarked. In the point process case the measure  $\bar{\mathcal{P}}_0$  can be interpreted as the Palm distribution dependent on the occurrence of an arbitrary point (i.e., a point with an arbitrary mark) at the origin.

The example below illustrates how marks can affect ergodic limits in some simple cases.

**EXAMPLE 13.4(a) Ergodic limits for processes with independent and unpredictable marks.** Consider first the case of a stationary marked Poisson process with nonnegative marks. Suppose the underlying Poisson process has intensity  $\mu$  and the marks have distribution function  $F(\cdot)$ . Because of the total lack of memory, the Palm distributions  $\mathcal{P}_{(0, \kappa)}$  are independent of  $\kappa$  and all reduce to the distribution of the original marked Poisson process, with  $m_g = \mu$  and



$\nu(dx) = \mu \pi(dx) = \mu dF(x)$ . Let the function  $g$  be the indicator function of the event  $\Gamma$  that  $T_M > \tau$ , where  $T_M$  is the time from the origin to the first point of the process with mark greater than  $M$ . Then the ergodic limit on the right-hand side of (13.4.2) becomes

$$\mu \int_{\mathcal{K}} \mathcal{P}_{(0,\kappa)}(\Gamma) \pi(d\kappa) = \mu \exp\{-\mu[1 - F(M)]\tau\}. \quad (13.4.3)$$

Note that the function  $g$  here does not depend on the mark  $\kappa$  at the origin. Also, because, in approaching the ergodic limit, averages are taken over all points of the process, irrespective of the mark, the constant preceding the exponential in (13.4.3) is  $\mu$ , and not  $\mu[1 - F(M)]$ . The latter rate could arise in situations that allow  $g$  to depend explicitly on  $\kappa$ . Suppose, for example, we required the mark at the origin to be greater than  $M$  before a contribution to  $\Gamma$  could be counted, so that  $g(\xi, \kappa) = I_{\Gamma}(\xi) I_{(M,\infty)}(\kappa)$ ; the limit on the right-hand side of (13.4.3) then becomes  $\mu[1 - F(M)] \exp\{-\mu[1 - F(M)]\tau\}$ . In this case we are just looking at the time intervals between points where the mark exceeds  $M$ , and the limit has the form we would expect from a Poisson process with rate  $\mu[1 - F(M)]$ .

The situation changes significantly if we allow extensions to processes with unpredictable marks. For example, consider a process of independent exponential intervals, in which the length of the interval following a point with mark  $\kappa$  is exponentially distributed with mean  $\kappa$ . In this case the ground process is a renewal process in which successive intervals are i.i.d. with common d.f. given by the mixture distribution

$$G(x) = 1 - \int_0^\infty e^{-x/\kappa} dF(\kappa)$$

with mean  $\int_0^\infty \kappa F(d\kappa)$ , which we again denote by  $1/\mu$ . The Palm distribution  $\mathcal{P}_{(0,\kappa)}$  now depends crucially on the mark  $\kappa$  at the origin, inasmuch as the time to the next point of the process (i.e., the first interval  $X_1$ ), is exponential with mean  $\kappa$ , and the remaining intervals  $X_2, X_3, \dots$  are i.i.d. with d.f.  $G$ . With no other constraints, and the same event  $\Gamma$  as before,  $\mathcal{P}_{(0,\kappa)}(\Gamma)$  can be evaluated as the sum of the probabilities

$$\Pr\{X_1 > \tau, \kappa_1 > M\} + \Pr\{X_1 < \tau, X_1 + X_2 > \tau, \kappa_1 < M, \kappa_2 > M\} + \dots,$$

where  $\kappa_1, \kappa_2, \dots$  are the marks of the successive points following the origin, and the special distribution of  $X_1$  must be observed. If we look at the ergodic limit, still with  $g = I_{\Gamma}$ , the dependence on the value of the initial mark is lost. If the further constraint is added that the initial mark  $\kappa$  must exceed  $M$  before a contribution to  $\Gamma$  is counted, the initial factor  $\mu$  on the right-hand side of (13.4.3) must first be multiplied by the rate factor  $1 - F(M)$ , as in the previous case, but the calculation of  $\mathcal{P}_{(0,\kappa)}(\Gamma)$  must be modified to allow for the fact that the distribution of  $X_1$  is now constrained by the requirement  $\kappa > M$ . Some further details and examples are given in Exercise 13.4.2.  $\square$

Higher moment results analogous to Proposition 13.2.VI can be developed for the marked case as special cases of the above proposition, but are relatively more complex. Reducing the second moment measure leads to a measure on  $\mathcal{X} \times \mathcal{K} \times \mathcal{K}$  which can also be described as a first-order moment measure of the Palm distribution, taking the form  $\check{M}_1(du \times d\kappa_1 \times d\kappa_2) = m_g^{-1} \check{M}_2(du \times d\kappa_1 \times d\kappa_2)$ . The relation here is between an initial point of the process taken as origin and a second point of the process,  $(0, \kappa_1)$  and  $(u, \kappa_2)$  say. If we standardize the second-order reduced measure by dividing by the reduced second moment measure for the ground process, we obtain a bivariate distribution  $\pi_2(d\kappa_1 \times d\kappa_2 \mid u)$  for the marks at 0,  $u$  respectively, conditioned by the occurrence of points at 0 and at  $u$  (see Lemma 8.3.III). Then we can write

$$\check{M}_1(du \times d\kappa_1 \times d\kappa_2) = \check{M}_1^g(du) \pi_2(d\kappa_1 \times d\kappa_2 \mid u), \quad (13.4.4)$$

where  $\check{M}_1^g$  is the first moment measure for the Palm distribution of the ground process. As noted in the discussion of Lemma 8.3.III and Example 8.3(e), the bivariate distribution  $\pi_2$  need not have marginals which reduce to the stationary distribution  $\pi$ ; indeed, the distribution need not even be symmetric (see Exercise 13.4.3). The reason for the discrepancy is that we are not merely conditioning on a point with a given mark at the origin, but on the occurrence of two points, one at a specified distance from the other, and this additional conditioning alters the distributions in general.

Next we formulate extensions of the last proposition to the nonergodic case; these are summarized in the next two results, for which proofs are outlined in Exercise 13.4.5. To avoid notational confusion, we write  $\omega$  for the element of the probability space even if this is the canonical space, and  $E(\cdot \mid \mathcal{I})(\omega)$  for conditional expectations with respect to the  $\sigma$ -algebra  $\mathcal{I}$  of invariant events.

**Lemma 13.4.II.** *Let  $\mathcal{P}$  be the distribution of a stationary marked random measure on  $\mathbb{R}^d \times \mathcal{K}$ ,  $\check{C}_{\mathcal{P}}$  its reduced (marked) Campbell measure, and  $\mathcal{I}$  the  $\sigma$ -algebra of invariant events under shifts in  $\mathcal{X} = \mathbb{R}^d$ . Then there exists an invariant random measure  $\zeta(\cdot, \cdot \mid \omega)$ , defined on sets in  $\mathcal{B}(\mathcal{M}_{\mathcal{X} \times \mathcal{K}}^{\#})$ , such that for nonnegative and  $\check{C}_{\mathcal{P}}$ -integrable functions  $g(x, \kappa, \xi)$ ,*

$$E\left(\int_{\mathcal{X}} g(x, \kappa, S_x \xi) \xi(dx \times d\kappa) \mid \mathcal{I}\right)(\omega) = \int_{\mathcal{X}} dx \int_{\mathcal{M}_{\mathcal{X}}^{\#}} g(x, \kappa, \psi) \zeta(d\psi \times d\kappa \mid \omega) \quad (13.4.5a)$$

with

$$E(\zeta(d\kappa \times d\psi)) = \check{C}_{\mathcal{P}}(d\kappa \times d\psi) = \nu(d\kappa) \mathcal{P}_{(0, \kappa)}(d\psi). \quad (13.4.5b)$$

In particular, (13.4.5a) with  $g = I_{A \times K \times \Gamma}$ ,  $A \in \mathcal{B}_{\mathcal{X}}$ ,  $K \in \mathcal{K}$ ,  $\Gamma \in \mathcal{B}(\mathcal{M}_{\mathcal{X} \times \mathcal{K}}^{\#})$  yields

$$\zeta(K \times \Gamma \mid \omega) = \frac{1}{\ell(A)} E\left(\int_A I_{\Gamma}(S_x \xi) \xi(dx, K) \mid \mathcal{I}\right)(\omega) \quad (\ell(A) > 0).$$

The next result establishes convergence of the sample path averages to the invariant random measure  $\zeta$ . Note that  $\zeta$  here is a refinement of the random measure  $\psi$  of Lemma 12.2.III. Both  $\zeta$  and  $\psi$  are a.s. totally finite with random total mass  $Y$  as defined around (12.2.10'). The measure  $\psi$  is the marginal measure of  $\zeta$  on  $\mathcal{K}$  after integrating out  $\xi$ , as in

$$\psi(K) = \int_{\mathcal{M}_{\mathcal{X}}^{\#}} \zeta(K \times d\xi) \quad (K \in \mathcal{B}_{\mathcal{K}}; Y = \psi(\mathcal{K})).$$

**Theorem 13.4.III.** *Let  $\xi$  be a strictly stationary marked random measure on  $\mathbb{R}^d \times \mathcal{K}$ ,  $\zeta$  the invariant random measure defined in Lemma 13.4.II, and with  $\mathcal{X} = \mathbb{R}^d$ , let  $h(\cdot, \cdot)$  be a  $\mathcal{B}(\mathcal{M}_{\mathcal{X} \times \mathcal{K}}^{\#})$ -measurable nonnegative or  $\check{C}_{\mathcal{P}}$ -integrable function on  $\mathcal{M}_{\mathcal{X} \times \mathcal{K}}^{\#}$ . Then for any convex averaging sequence  $\{A_n\}$  and  $n \rightarrow \infty$ ,*

$$\frac{1}{\ell(A_n)} \int_{A_n} h(S_x \xi, \kappa) \xi(dx \times d\kappa, \omega) \rightarrow \int_{\mathcal{M}_{\mathcal{X} \times \mathcal{K}}^{\#}} h(\psi, \kappa) \zeta(d\psi \times d\kappa \mid \omega) \quad \mathcal{P}\text{-a.s.} \quad (13.4.6a)$$

In particular, for  $h(\psi, \kappa) = I_{\Gamma \times K}(\psi, \kappa)$  with  $\Gamma \in \mathcal{B}(\mathcal{M}_{\mathcal{X} \times \mathcal{K}}^{\#})$  and  $K \in \mathcal{B}_{\mathcal{K}}$ ,

$$\frac{1}{\ell(A_n)} \int_{A_n} I_{\Gamma \times K}(S_x \xi, \kappa) \xi(dx \times d\kappa, \omega) \rightarrow \zeta(\Gamma \times K \mid \omega) \quad \mathcal{P}\text{-a.s.} \quad (13.4.6b)$$

Notice that the random measure  $\zeta(\cdot)$  is associated with the reduced Campbell measure (i.e., the Palm measure) rather than the Palm distribution. Thus, in combining limits over different ergodic components, it is the conditional Palm measures, rather than the normalized Palm distributions, that combine linearly according to (13.4.5b). In considering ergodic results, it may seem more natural to combine the normalized limits in a linear manner, as suggested by Sigman (1995); this leads to the ‘alternative Palm distribution’ described in Exercise 13.4.6, where some consequences and elementary results are given. The next example illustrates the point.

**EXAMPLE 13.4(b) A mixed Poisson process.** Consider two stationary Poisson processes on  $\mathbb{R}$  with parameters  $\lambda, \mu$  ( $\lambda \neq \mu$ ), selected with probabilities  $p$  and  $q = 1 - p$ , respectively. The invariant  $\sigma$ -field contains two nontrivial events  $J_1, J_2$  say, corresponding to the choice of the  $\lambda$  and  $\mu$  processes, respectively. Let  $\Gamma = \{N(0, a] = 0\}$  for some fixed  $a > 0$ . Then  $S_x \Gamma = \{N(x, x + a] = 0\}$ , and

$$E(I_{\Gamma}(\xi)) = p e^{-\lambda a} + q e^{-\mu a},$$

whereas

$$\zeta(\Gamma, \omega) = \lambda e^{-\lambda a} I_{J_1}(\omega) + \mu e^{-\mu a} I_{J_2}(\omega),$$

in which the indicator variables have multipliers representing the average rate of occurrence of points followed by an empty interval of length  $a$ . The measure  $\check{C}_{\mathcal{P}}(\Gamma)$  represents the overall average rate of occurrence of such points, weighted according to the probabilities of the two components, and so it is

given by

$$\check{C}_{\mathcal{P}}(\Gamma) = p\lambda e^{-\lambda a} + q\mu e^{-\mu a}, \quad (13.4.7a)$$

corresponding to (13.4.5a). The overall Palm probability of  $\Gamma$  equals

$$(p\lambda e^{-\lambda a} + q\mu e^{-\mu a}) / (p\lambda + q\mu), \quad (13.4.7b)$$

the denominator representing the overall rate of occurrence of points. It can be interpreted as the limit of the conditional probability that the next point will occur after lapse of time  $a$ , given that a point occurs in a small interval about the origin. This expression is to be contrasted with the situation for the Palm distributions: the Palm probability of  $\Gamma$  on  $J_1$  is  $e^{-\lambda a}$  and on  $J_2$  it is  $e^{-\mu a}$ , but the overall Palm probability of  $\Gamma$  is not  $p e^{-\lambda a} + q e^{-\mu a}$  as might have been expected. See Exercise 13.4.6 for the latter probability.  $\square$

For simple ergodic point processes, the previous theorems, coupled with the interpretation of the Palm distribution as the distribution of the process at an ‘arbitrary point’, lead to a circle of important results on convergence to equilibrium. Specifically, starting from an ‘arbitrary point’ (i.e., with the Palm distribution for a simple point process) and translating through some  $x \in \mathbb{R}^d$ , we seek conditions under which the translated distributions converge to the stationary distribution as  $\|x\| \rightarrow \infty$ . Dually, starting from an ‘arbitrary location’ (i.e., the stationary distribution) and observing the process at the  $n$ th point nearest to that location, we seek conditions under which the distribution of the process, relative to that  $n$ th point as origin, converges to the Palm distribution as  $n \rightarrow \infty$ .

To approach these ideas for MPPs, we first establish some relevant notation. For MPPs, the Palm measure  $\mathcal{P}_{(0,\kappa)}(\cdot)$  is supported by the subspace,  $\mathcal{N}_{(0,\kappa)}^\#$  say, of marked counting measures having a point with mark  $\kappa$  at the origin. The averaged form,

$$\bar{\mathcal{P}}_0(\cdot) = \int_{\mathcal{K}} \mathcal{P}_{(0,\kappa)}(\cdot) \pi(d\kappa) \quad (13.4.8)$$

which we call the *mean Palm distribution*, occurs frequently in ergodic limits. It can be interpreted as a measure on the subspace  $\mathcal{N}_0^{\mathcal{K}\#} = \bigcup_{\kappa \in \mathcal{K}} \mathcal{N}_{(0,\kappa)}^\#$  of marked counting measures on  $\mathbb{R}^d$  with a point at the origin whose mark  $\kappa \in \mathcal{K}$  there is unspecified. We use  $N^{\mathcal{K}}$  to denote a generic marked counting measure with marks in  $\mathcal{K}$ ,  $N_g^{\mathcal{K}}$  to denote its associated ground counting measure,  $N_0^{\mathcal{K}}$  to denote a marked counting measure having a point with unspecified mark from  $\mathcal{K}$  at the origin, and  $N_{(0,\kappa)}$  to denote a marked counting measure with mark  $\kappa$  at the origin.

Using this notation, inversion theorems for MPPs on  $\mathbb{R}^d$  corresponding to the results in Section 13.3 can be developed. Thus (13.3.1) can be extended to MPPs on  $\mathbb{R}^d$ : for bounded, nonnegative functions  $h(\cdot)$  of  $N_0^{\mathcal{K}}$ ,

$$\mathbb{E}_{\bar{\mathcal{P}}_0} [h(N_0^{\mathcal{K}})] = \int_{\mathcal{K}} \mathbb{E}_{\mathcal{P}_{(0,\kappa)}} [h(N_{(0,\kappa)})] \pi(d\kappa) = \frac{1}{m_g} \mathbb{E}_{\mathcal{P}} \left( \sum_{i: x_i \in \mathbb{U}^d} h(s_{x_i} N^{\mathcal{K}}) \right), \quad (13.4.9)$$

where  $\{x_i \equiv x_i(N^\mathcal{K})\}$  denotes some measurable enumeration of the points of the realization  $N^\mathcal{K}$ .

The analogue of (13.3.2) for  $\mathbb{R}^d$  is (13.3.9b), which for MPPs becomes, for bounded nonnegative functions  $g(\cdot)$  of  $N^\mathcal{K}$ ,

$$\begin{aligned} \mathbb{E}_{\mathcal{P}}[g(N^\mathcal{K})] &= m_g \int_{\mathcal{K}} \mathbb{E}_{\mathcal{P}_{(0,\kappa)}} \left[ \int_{V_0(N_{(0,\kappa)})} g[S_x N_{(0,\kappa)}] dx \right] \pi(d\kappa) \\ &= m_g \mathbb{E}_{\bar{\mathcal{P}}_0} \left[ \int_{V_0(N_0^\mathcal{K})} g[S_x N_0^\mathcal{K}] dx \right], \end{aligned} \quad (13.4.10)$$

where, for example,  $V_0(N_0^\mathcal{K})$  denotes the Voronoi polygon about the origin in  $\mathbb{R}^d$  formed by the locations (points of the ground process) of the realization  $N_0^\mathcal{K}$ . Exercise 13.4.1 sketches arguments that justify these inversion formulae.

Using these representations and notations, we obtain the following combination of results from Section 12.2 and Proposition 13.4.I.

**Proposition 13.4.IV.** *Let  $\mathcal{P}$  be the distribution of a stationary ergodic MPP on  $\mathbb{R}^d$  with marks in  $\mathcal{K}$ , finite ground rate  $m_g$ , and stationary mark distribution  $\pi(\cdot)$ ; let  $\mathcal{P}_{(0,\kappa)}$  be the associated family of stationary Palm distributions, and  $\bar{\mathcal{P}}_0$  the mean Palm distribution defined at (13.4.8). If  $\{A_n\}$  is a convex averaging sequence in  $\mathbb{R}^d$  satisfying  $\ell(A_n)/\ell(A_{n+1}) \rightarrow 1$  ( $n \rightarrow \infty$ ), and  $h(\cdot)$  a bounded, measurable, nonnegative function of  $N^\mathcal{K}$ , then*

$$\frac{1}{\ell(A_n)} \int_{A_n} h(S_x N_0^\mathcal{K}) dx \rightarrow \mathbb{E}_{\mathcal{P}}[h(N^\mathcal{K})] \quad (n \rightarrow \infty) \quad \bar{\mathcal{P}}_0\text{-a.s.} \quad (13.4.11)$$

Furthermore, if  $x_j^* \equiv x_j^*(N^\mathcal{K})$  as above (13.3.7) and  $g(\cdot)$  is a bounded, measurable, nonnegative function of  $N_0^\mathcal{K}$ , then

$$\frac{1}{n'} \sum_{j=1}^{n'} g[S_{x_j^*}(N^\mathcal{K})] \rightarrow \mathbb{E}_{\bar{\mathcal{P}}_0}[g(N_0^\mathcal{K})] \quad (n' \rightarrow \infty) \quad \mathcal{P}\text{-a.s.} \quad (13.4.12)$$

PROOF. We start by considering (13.4.12). In the present context of MPPs and with  $\{A_n\}$  a sequence of spheres, (13.4.2) of Proposition 13.4.I can be put in the form

$$\frac{1}{\ell(A_n)} \sum_{j=1}^{N_g^\mathcal{K}(A_n)} g[S_{x_j^*}(N^\mathcal{K})] \rightarrow m_g \mathbb{E}_{\bar{\mathcal{P}}_0}[g(N_0^\mathcal{K})] \quad \mathcal{P}\text{-a.s.}, \quad (13.4.13)$$

because the locations of points in  $A_n$  are precisely those of  $N_g^\mathcal{K}$  with modulus  $r_j$  satisfying  $r_j = \|x_j^*\| \leq \text{radius}(A_n)$ . Taking  $g \equiv 1$ , it follows that  $N_g^\mathcal{K}(A_n)/\ell(A_n) \rightarrow m_g$ , so in place of (13.4.13) we can write

$$\frac{1}{N_g^\mathcal{K}(A_n)} \sum_{j=1}^{N_g^\mathcal{K}(A_n)} g(S_{x_j^*} N^\mathcal{K}) \rightarrow \mathbb{E}_{\bar{\mathcal{P}}_0}(g(N_0)) \quad \mathcal{P}\text{-a.s.} \quad (13.4.14)$$

We can suppose that the  $\{A_n\}$  have been so chosen that  $\ell(A_n)/\ell(A_{n+1}) \rightarrow 1$  as  $n \rightarrow \infty$ , in which case we have

$$N_g^K(A_n)/N_g^K(A_{n+1}) \rightarrow 1 \quad \mathcal{P}\text{-a.s.}$$

For brevity, write  $g_j = g(S_{x_j^*}N)$ , and  $N_k = N_g^K(A_{n(k)})$ , where  $n(k)$  is chosen to satisfy  $x_k^* \in (A_{n(k)+1} \setminus A_{n(k)})$ . Then the inequalities

$$\frac{1}{N_{k+1}} \sum_{j=1}^{N_k} g_j \leq \frac{1}{N_k} \sum_{j=1}^{N_k} g_j \leq \frac{1}{N_k} \sum_{j=1}^{N_{k+1}} g_j$$

show that the limit relation (13.4.14) implies (13.4.12).

Equation (13.4.11) has the same form as (12.2.6) in the ergodic case, with one notable exception: (12.2.6) holds outside a set of counting measures with  $\mathcal{P}$ -measure zero, whereas in (13.4.11) the corresponding result is required to hold outside a subset of  $\mathcal{N}_0^K$ , which itself has  $\mathcal{P}$ -measure zero. Thus we cannot immediately draw conclusions from (12.2.6). To get over this difficulty, let  $\Gamma$  be the subset of  $\mathcal{N}_{\mathcal{X}}^{\#}$  ( $\mathcal{X} = \mathbb{R}^d \times \mathcal{K}$ ) on which (13.4.11) holds and  $\Gamma_0$  its restriction to  $\mathcal{N}_0^K$ . From (13.4.9), taking  $h$  as the indicator of  $\Gamma_0$ ,

$$\begin{aligned} \bar{\mathcal{P}}_0(\Gamma_0) &= \frac{1}{m_g} \mathbb{E}_{\mathcal{P}} \left( \sum_{i=1}^{N_g^K(\mathbb{U}^d)} I_{\Gamma_0}(S_{x_i^*}(N^K)) \right) \\ &= \frac{1}{m_g} \mathbb{E}_{\mathcal{P}} \left( \int_{\mathbb{U}^d} I_{\Gamma_0}(S_x(N)) N(dx) \right) \\ &= \frac{1}{m_g} \int_{\Gamma} \int_{\mathbb{U}^d} I_{\Gamma_0}(S_x N) N(dx) \mathcal{P}(dN) \end{aligned} \quad (13.4.15)$$

because  $\Gamma$  has  $\mathcal{P}$ -measure one. Now  $\Gamma$  is invariant, so  $S_x N \in \Gamma$  whenever  $N \in \Gamma$ . But if  $S_x N \in \Gamma_0$ , as required by the indicator function  $I_{\Gamma_0}$  in the integrand, we must have  $N(\{x\}) = 1$ , and so the last line in (13.4.15) can be rewritten

$$\bar{\mathcal{P}}_0(\Gamma_0) = \frac{1}{m} \int_{\Gamma} N(\mathbb{U}^d) \mathcal{P}(dN) = \frac{\mathbb{E}_{\mathcal{P}}(N(\mathbb{U}^d))}{m} = 1.$$

This establishes (13.4.11), the set of  $\bar{\mathcal{P}}_0$ -measure zero being taken as the relative complement of  $\Gamma_0$  in  $\bar{\mathcal{P}}_0$ .  $\square$

In the nonergodic case, the argument leading to (13.4.15) can be extended to show that  $\mathcal{P}$  and  $\bar{\mathcal{P}}_0$  induce the same measures on the invariant  $\sigma$ -algebra  $\mathcal{I}$ ; see Exercise 13.4.7.

Equation (13.4.11) of Proposition 13.4.IV can be interpreted as an a.s. statement about convergence to equilibrium. To put the result in the context of Section 12.5, take expectations with respect to  $\bar{\mathcal{P}}_0$  and  $\mathcal{P}$  in equations (13.4.11) and (13.4.12), respectively, and apply the dominated convergence theorem. Then conclude as follows.

**Corollary 13.4.V.** *Suppose that the conditions of Proposition 13.4.IV hold, with  $g$  and  $h$  bounded. Then*

$$\frac{1}{\ell(A_n)} \int_{A_n} E_{\bar{\mathcal{P}}_0}(h(S_x N_0)) dx \rightarrow E_{\mathcal{P}}(h(N)) \quad (n \rightarrow \infty), \quad (13.4.16)$$

and

$$\frac{1}{k} \sum_{j=1}^k E_{\mathcal{P}}(g(S_{x_j^*} N)) \rightarrow E_{\bar{\mathcal{P}}_0}(g(N_0)) \quad (k \rightarrow \infty). \quad (13.4.17)$$

Notice that although the left-hand side of (13.4.17) converges to the Palm distribution in the ergodic case, this is not so in the nonergodic case, when it converges rather to the ‘alternative Palm distribution’ of Exercise 13.4.6.

Equation (13.4.16) can be interpreted as asserting weak convergence of the measures  $(1/\ell(A_n)) \int_{A_n} \hat{S}_x \bar{\mathcal{P}}_0 dx$  to the limit measure  $\mathcal{P}$ . In the one-dimensional context, if we consider intervals  $A_n = [0, n]$  and project onto the half-line  $\mathbb{R}_+$  (see Exercise 13.4.8 for details), this is nothing other than the weak  $(C, 1)$ -asymptotic stationarity of the measure  $\mathcal{P}_0$ . Corollary 12.6.VIII then implies that  $\mathcal{P}_0$  is also strongly  $(C, 1)$ -asymptotically stationary, with the same limit measure  $\mathcal{P}$ . Because the weak and strong versions coincide, we call such a process simply  $(C, 1)$ -asymptotically stationary.

To develop a similar interpretation of equation (13.4.17), first recall that, in the one-dimensional case, every stationary MPP can be associated with a stationary *marked interval process*, and vice versa, where by a marked interval process we mean a sequence of pairs  $\{(\tau_i, \kappa_{i-1})\}$ , with  $\tau_i$  the length of the  $i$ th interval and  $\kappa_{i-1}$  the mark associated with its left-hand endpoint. As in the discussion in Section 13.3, the probability distribution of a marked interval process can be treated either as a distribution on the space of sequences  $\{(\tau_i, \kappa_{i-1})\}$ , or as a distribution on the subspace  $\mathcal{N}_0^\#$  of MPPs with a point at the origin. Note that if the sequence of pairs is stationary, then its distribution on  $\mathcal{N}_0^\#$  is the mean Palm distribution  $\bar{\mathcal{P}}_0$  for some stationary MPP with measure  $\mathcal{P}$  related to  $\bar{\mathcal{P}}_0$  by (13.4.9–10). Note also that for a stationary MPP, the initial interval  $(0, t_1(N))$  has the stationary interval distribution for the ground process, and the initial mark  $\kappa_1$  has a form of stationary mark distribution which is different in general from the stationary mark distribution  $\pi$ . There is also a stationary joint distribution for the initial interval and the initial mark; see Exercise 13.4.9, which also sketches out a proof of the stationarity results summarized above.

The concepts of coupling and shift-coupling, as of weak and strong  $(C, 1)$ -asymptotic stationarity, can be developed for an interval process, whether simple or marked, just as easily as, and in a parallel fashion to, those for a point process. As for point processes, the concepts of weak and strong  $(C, 1)$ -asymptotic stationarity coincide, and can be referred to simply as  $(C, 1)$ -asymptotic stationarity. Now (13.4.17) can be interpreted as asserting that, if we start an interval process from the first (or indeed the  $k$ th) point following

the origin in a stationary MPP, that interval process is  $(C, 1)$ -asymptotically stationary with limit process equal to the stationary interval process associated with the original MPP.

Sigman (1995) calls a point process  $N$  ‘event-stationary’ when the sequence of intervals initiated by  $t_1(N)$  is interval-stationary. In this terminology, the MPP corresponding to  $\bar{\mathcal{P}}_0$  is event-stationary. In general, a stationary point process will not be event-stationary in this sense, but Corollary 13.4.V implies that on  $\mathbb{R}_+$ , a stationary MPP is  $(C, 1)$ -asymptotically event-stationary, and conversely that an event-stationary MPP is  $(C, 1)$ -asymptotically stationary.

Finally in this circle of ideas, these last conclusions can be extended to processes that are not themselves stationary, but only asymptotically stationary, as summarized below.

**Proposition 13.4.VI.** *Suppose that the MPP  $N^K$  is either stationary or  $(C, 1)$ -asymptotically stationary with limit measure  $\mathcal{P}$ . Then the interval process, starting from  $t_1(N)$  as origin, is  $(C, 1)$ -asymptotically stationary with limit measure  $\bar{\mathcal{P}}_0$  associated with  $\mathcal{P}$  through (13.4.9–10).*

*Conversely, suppose that an MPP  $N_0^K$  defined on  $\mathcal{N}_0^\#$  represents an interval process which is either stationary or  $(C, 1)$ -asymptotically stationary, with limit measure  $\bar{\mathcal{P}}_0$ . Then  $N_0^K$  is  $(C, 1)$ -asymptotically stationary with limit measure  $\mathcal{P}$  associated with  $\bar{\mathcal{P}}_0$  through (13.4.9–10).*

PROOF. We start from the observation that if two MPPs  $N$  and  $N'$  shift-couple, then the same is true of the associated interval processes  $N_0$  and  $N'_0$ , started at  $t_1(N)$  and  $t_1(N')$ , respectively, and vice versa. Indeed, if there are stopping times  $T, T'$  and versions  $\tilde{N}, \tilde{N}'$  of the MPPs such that  $\tilde{N}(t+T), \tilde{N}'(t+T')$  are a.s. equal for  $t \geq 0$ , then the corresponding interval processes, say  $\tilde{N}_0$  and  $\tilde{N}'_0$  are equal after discrete times  $J = \tilde{N}(T) + 1$ ,  $J' = \tilde{N}'(T) + 1$  and so the interval processes shift-couple. Conversely, if the two interval processes, corresponding to MPPs  $N_0, N'_0$ , shift-couple, with coupling times  $J, J'$  respectively, then  $N_0$  and  $N'_0$  shift-couple as point processes with coupling times  $T = t_J(N)$ ,  $T' = t_{J'}(N')$ , respectively.

Suppose now that  $N$  is  $(C, 1)$ -asymptotically stationary with limit process  $N'$  having distribution  $\mathcal{P}$ , so that  $N$  and  $N'$  shift-couple. Then the corresponding interval processes, which we may associate with point processes  $N_0$  and  $N'_0$ , also shift-couple, so that  $N_0$  is  $(C, 1)$ -asymptotically stationary with limit process  $N'_0$ . This last process is not itself stationary, but corresponds to the process on the left-hand side of (13.4.16), which by Corollary 13.4.V is  $(C, 1)$ -asymptotically interval stationary with limit the stationary interval process,  $N''_0$  say, associated with the stationary process  $N'$ ; thus  $N'_0$  shift-couples to  $N''_0$ . The transitivity of shift-coupling now implies that, as interval processes,  $N_0$  shift-couples to  $N''_0$  and is therefore  $(C, 1)$ -asymptotically stationary with limit process  $N''_0$ .

But we already know that the distribution of a stationary point process is associated with the distribution of the corresponding stationary interval process through equations (13.4.9–10), and so the first statement of the



proposition follows. The second statement follows by an analogous argument with the roles of the point and interval processes reversed.  $\square$

The last proposition finds applications in queueing theory and related fields, where it provides a starting point for a systematic approach to results on convergence to equilibrium, and the forms of the resulting stationary distributions. Accounts are given in Franken *et al.* (1981), Baccelli and Brémaud (1994), and Sigman (1995), and we do not attempt to repeat the material here. However, to give the flavour of the applications, we indicate how they apply to regenerative processes.

**EXAMPLE 13.4(c) *Regenerative processes via embeddings.*** Call an MPP  $N^{\mathcal{K}}$  on  $\mathbb{R}_+$ , with internal history  $\mathcal{H}$ , *regenerative* if the sequence of event times  $\{t_i: i \geq 1\}$  includes an embedded renewal process. More precisely, we require the existence of a subsequence  $\{\tau_j = t_{i_j}: j = 1, 2, \dots\}$  (the *regeneration points*) such that

- (1) for  $j \geq 1$ , the intervals  $\tau_{j+1} - \tau_j$  form an i.i.d. family with proper d.f.  $F(x)$  (and in the sequel, we suppose that this distribution has a finite mean, so that the corresponding renewal process has a stationary version); and
- (2) between regeneration points, the successive families of marked points

$$N^{(j)} = \{(0, \kappa_{i_j}), ((t_{1+i_j} - t_{i_j}), \kappa_{1+i_j}), ((t_{2+i_j} - t_{i_j}), \kappa_{2+i_j}), \dots, ((t_{i_{j+1}} - t_{i_j}), \kappa_{i_{j+1}})\}$$

are i.i.d. versions of a finite MPP on  $\mathcal{X} = [0, \infty) \times \mathcal{K}$ .

To bring such processes within the purview of the previous theory, we first regard the sequence  $\{(\tau_j, N^{(j)})\}$  as a marked renewal process with i.i.d. marks  $N^{(j)}$ , treating the latter as random variables on the portmanteau space  $\mathcal{X}^{\cup}$  of Chapter 5 [equation (5.3.10)]. If the intervals in the renewal process have finite mean length, the resultant renewal process is stationary, and (having i.i.d. marks) so is the associated marked renewal process. The first interval may be exceptional (i.e., the renewal process may be a delayed renewal process) but in any case the process is asymptotically stationary, hence  $(C, 1)$ -asymptotically stationary. The stationary mark distribution is nothing other than the common distribution, say  $\mathcal{P}_f$ , of the finite MPPs  $N^{(j)}$ .

These remarks imply not only that the embedded marked renewal process is asymptotically stationary, but also that the original MPP is asymptotically stationary. The underlying reason is that any bounded functional of the MPP can be represented, in terms of the sequence of renewal times and the i.i.d. sequence  $N^{(j)}$ , as a bounded functional of the marked renewal process. Hence the original MPP is stationary (expectations of bounded functionals invariant under shifts) if and only if the latter is stationary. It follows also that if the original MPP is started from a regeneration point, it is asymptotically stationary; that is, there is a shift-coupling with the asymptotic form of the original MPP. Moreover, if we take any stopping time  $T$ , defined in terms of the initial finite process  $N^{(0)}$ , the process started at  $T$  also shift-couples to

the limit process (modify the shift). Thus the original MPP is asymptotically stationary if started from any such  $T$ , in particular, from any of the points of the MPP in the initial cycle.

A more explicit argument, exhibiting the form of the ergodic limit in terms of the  $N^{(j)}$  is outlined by Sigman (1995, Section 2.6). For another approach use the ideas of Lemma 12.5.III, taking the initial condition to be the history of the current process  $N^{(j)}$  up to the time-point chosen as origin.  $\square$

As in other situations of this kind, although these general results are useful for discussing the convergence to stationarity and existence of stationary versions of the particular process in view, the hard step is the evaluation of the form of the stationary distribution, which in the present case means evaluating the explicit form of the distribution of the i.i.d. components  $N^{(j)}$ . If, for example, the regenerative process is taken to be the arrival of a customer at an empty queue in an  $M/G/1$  queueing system, and the original MPP is the number of customers waiting in the queue at the arrival of a new customer, explicit arguments are still needed to evaluate the form of the stationary distribution of the queue size.

Although the results have been stated for regenerative processes, the independence of the cycles has been used only tangentially. The essential points of the argument are the existence of an embedded sequence of points which are  $(C, 1)$ -asymptotically stationary, and a law of large numbers for the behaviour within cycles. For extensions in this direction see Sigman for references.

A stronger form of convergence holds when the process is mixing, and we can drop the  $(C, 1)$ -averaging that stems from the ergodic theorem. For a brief statement of such results, we return to unmarked processes, leaving the reader to formulate extensions.

**Proposition 13.4.VII.** *Let  $\mathcal{P}$  be the distribution of a simple stationary mixing point process on  $\mathbb{R}^d$ , with finite density, and let  $\mathcal{P}_0$  be the corresponding Palm distribution. Then*

$$\hat{S}_x \mathcal{P}_0 \rightarrow \mathcal{P} \quad \text{weakly} \quad (\|x\| \rightarrow \infty). \quad (13.4.18)$$

PROOF. To establish weak convergence we need to show that for all bounded continuous  $f$  on  $\mathcal{M}_{\mathcal{X}}^{\#}$ ,

$$E_{\mathcal{P}_0}(f(S_x N_0)) \rightarrow E_{\mathcal{P}}(f(N)) \quad \text{as} \quad \|x\| \rightarrow \infty.$$

Proposition 13.3.V is a convenient starting point for the proof. From that result, for any given  $\epsilon > 0$  and sufficiently small sphere  $A_n$ ,

$$|E_{\mathcal{P}_0}(f(S_x N_0)) - E_{\mathcal{P}}(f(S_x S_{x^*(N)} N) \mid N(A_n) > 0)| < \frac{1}{2}\epsilon \quad (13.4.19)$$

for each fixed  $x \in \mathcal{X}$  [replace  $f(\cdot)$  by  $f(S_x \cdot)$  in (13.3.14)]. Inspection of the proof of (13.3.14) shows that the inequality is uniform in  $x$  (see Exercise 13.3.6) because the two critical inequalities used in its proof depend only on

(13.3.11) and  $\sup |f(\cdot)|$ , both of which are independent of  $x$ . Consequently, we can fix  $n$  from (13.4.19) and proceed by simply evaluating the difference

$$\begin{aligned} & |\mathbb{E}_{\mathcal{P}}(f(S_x S_{x^*(N)} N) \mid N(A_n) > 0) - \mathbb{E}_{\mathcal{P}}(f(N))| \\ &= \frac{|\mathbb{E}_{\mathcal{P}}(f(S_x S_{x^*(N)} N) I_{\{N(A_n) > 0\}}(N)) - \mathbb{E}_{\mathcal{P}}(f(N)) \mathbb{E}_{\mathcal{P}}(I_{\{N(A_n) > 0\}}(N))|}{\mathcal{P}\{N(A_n) > 0\}}. \end{aligned}$$

Now it is enough to apply the result of Exercise 12.3.5 with

$$X(N) = f(S_x S_{x^*(N)} N), \quad Y(N) = I_{\{N(A_n) > 0\}}(N)$$

to deduce that this difference can also be made less than  $\frac{1}{2}\epsilon$  by taking  $\|x\|$  sufficiently large.  $\square$

To illustrate the close connection between these results and the classical renewal theorems, we prove a result for mixing second-order processes that Delasnerie (1977) attributes to Neveu. It asserts that for large  $\|x\|$ , the reduced second moment measure approximates its form under a Poisson process.

**Theorem 13.4.VIII.** *Let  $N$  be a stationary second-order point process in  $\mathbb{R}^d$  with density  $m$  and reduced second moment measure  $\check{M}_2(\cdot)$ . If  $N$  is mixing then as  $\|x\| \rightarrow \infty$ ,*

$$S_x \check{M}_2(\cdot) \rightarrow_w m^2 \ell(\cdot). \quad (13.4.20)$$

PROOF. The formal connection here lies with the representation of the reduced moment measures as moments of the Palm distribution. However, we do not need to call on the Palm theory as such; it is enough to use the definition of the reduced moment measure, which yields, with  $b^*(x) = b(-x)$  as in Chapter 12 and functions  $a(\cdot)$ ,  $b(\cdot)$  that vanish outside a bounded set,

$$\begin{aligned} \mathbb{E} \left( \int_{\mathcal{X}} \int_{\mathcal{X}} a(u-x) b(v) N(du) N(dv) \right) &= \int_{\mathcal{X}} (a * b^*)(v-x) \check{M}_2(dv) \\ &= \int_{\mathcal{X}} (a * b^*)(v) S_x \check{M}_2(dv). \end{aligned}$$

Now when  $N$  is mixing, the first expectation converges as  $\|x\| \rightarrow \infty$  to

$$m^2 \int_{\mathcal{X}} a(u) du \int_{\mathcal{X}} b(v) dv = m^2 \int_{\mathcal{X}} (a * b^*)(v) dv,$$

which by letting  $a(\cdot)$  run through bounded continuous functions ensures the weak convergence of  $b * (S_x \check{M}_2)(\cdot)$  to  $b * (m^2 \ell(\cdot))$ , from which a standard sandwich argument yields (13.4.20).  $\square$

This result assumes a more familiar form in the case that  $d = 1$  (i.e.,  $\mathcal{X} = \mathbb{R}$ ) when expressed in terms of the expectation function  $U(\cdot)$  introduced in Theorem 3.5.III. Then, for a second-order stationary simple point process on  $\mathbb{R}$  we have

$$\begin{aligned} U(x) &= 1 + \lim_{h \downarrow 0} \mathbb{E}_{\mathcal{P}}(N(0, x] \mid N(-h, 0] > 0) \\ &= 1 + \mathbb{E}_{\mathcal{P}_0}(N(0, x]) = 1 + \check{M}_2(0, x]/m, \end{aligned}$$

leading to the following corollary to Theorem 13.4.VIII.

**Corollary 13.4.IX** (Generalized Blackwell Theorem). *Let  $N$  be a simple stationary mixing point process on  $\mathbb{R}$  with finite second moment measure. For all  $h > 0$ ,*

$$U(x+h) - U(x) \rightarrow mh \quad (x \rightarrow \infty).$$

Because a renewal process with  $m < \infty$  has finite second moment measure and is mixing if its lifetime distribution  $F$  is nonlattice, this corollary includes the standard version of Blackwell's theorem (Theorem 4.4.1) as a special case. However, the simplicity of the above argument as compared with the intricacies of Chapter 4 is misleading, because what is obscured here is the fact that to prove that a renewal process is mixing, a result close to Blackwell's theorem must be assumed. For this case, therefore, the corollary would then become the conclusion of a somewhat circular argument. More generally, there is no very simple relation between mixing of the basic process and mixing of the sequence of intervals, in contrast to the case of ergodicity for which the concepts coincide (see Exercise 13.4.9). For the Wold process, similar questions regarding the lattice structure have to be overcome as in Chapter 4, and, additionally, the function  $U(\cdot)$  refers to expectations when the initial interval has the stationary distribution. Consequently, further extensions are needed to cover the case of a process starting with an arbitrary distribution for the initial interval.

## Exercises and Complements to Section 13.4

### 13.4.1 Reduced Campbell measure for marked random measures and MPPs.

- (a) The marked Campbell measure on  $\mathcal{X} \times \mathcal{K} \times \mathcal{M}_{\mathcal{X} \times \mathcal{K}}^\#$  is defined for  $A \in \mathcal{B}_{\mathcal{X}}$ ,  $K \in \mathcal{B}_{\mathcal{K}}$ , and  $U \in \mathcal{B}(\mathcal{M}_{\mathcal{X} \times \mathcal{K}}^\#)$  by

$$C_{\mathcal{P}}(A \times K \times U) = \int_U \int_A \int_{\mathcal{K}} \xi(dx \times d\kappa) \mathcal{P}(d\xi).$$

Show that if  $\mathcal{P}$  is stationary under shifts on  $\mathcal{X} = \mathbb{R}^d$ ,  $C_{\mathcal{P}}$  is invariant under the actions of the transformations  $\Theta_u(x, \kappa, \xi) = (x - u, \kappa, S_u \xi)$ .

- (b) Find a mapping on the triple product space analogous to  $D$  in (13.2.4), and show that when  $\mathcal{P}$  corresponds to a stationary random marked measure  $\xi$ , the marked Campbell measure factorizes into a product of Lebesgue measure on  $\mathcal{X}$  and a reduced Campbell measure  $\check{C}_{\mathcal{P}}(d\kappa \times d\psi)$  defined by the following extension of (13.2.6),

$$\mathbb{E} \left( \int_{\mathcal{X} \times \mathcal{K}} g(x, \kappa, S_x \xi) \xi(dx \times d\kappa) \right) = \int_{\mathcal{X}} dx \int_{\mathcal{K} \times \mathcal{M}_{\mathcal{X} \times \mathcal{K}}^\#} g(x, \kappa, \psi) \check{C}_{\mathcal{P}}(d\kappa \times d\psi).$$

- (c) If the ground process has finite mean rate  $m_g$ , with associated stationary mark distribution  $\pi$ , show (using a disintegration argument) that the right-hand side of the above equation can be written

$$m_g \int_{\mathcal{X}} dx \int_{\mathcal{K}} \pi(d\kappa) \int_{\mathcal{M}_{\mathcal{X} \times \mathcal{K}}^\#} g(x, \kappa, \psi) \mathcal{P}_{(0, \kappa)}(d\psi),$$

where  $\{\mathcal{P}_{(0, \kappa)}\}$  is a family of stationary Palm distributions conditional on the mark  $\kappa$ .

- 13.4.2 *MPP with exponentially distributed intervals.* Consider the second situation in Example 13.4(a), where successive intervals are exponentially distributed with mean length equal to the mark  $\kappa$  of the point initiating the interval. Find a series expansion for  $\mathcal{P}_{(0,\kappa)}(\Gamma)$  as a function of  $M$  and  $\kappa$ . Is the averaged form  $\mathcal{P}_0^*(\Gamma)$  smaller or larger than its value in the first situation, where the intervals are i.i.d. exponential with the same mean  $\mu$  as in the mixed case. Does a version of the renewal theorem hold for the mixed process?
- 13.4.3 *A marked Gauss–Poisson process.* Consider a stationary Gauss–Poisson process [see, e.g., Exercise 7.1.9(a)], in which the parent point has mark  $A$  and the offspring point, following it after a distance  $u$  with d.f.  $dF(u) = f(u) du$  ( $u \geq 0$ ), has mark  $a$ . If parent points occur as a stationary Poisson process with rate  $\lambda$ , find the reduced second-order moment densities  $m_{A,a}^{(2)}(u)$ ,  $m_{a,A}^{(2)}(u)$ ,  $m_{A,A}^{(2)}(u)$ ,  $m_{a,a}^{(2)}(u)$ , and the bivariate  $(2 \times 2)$  distribution  $\pi_2(\kappa_1, \kappa_2)$ . Show in particular that  $m_{A,a}^{(2)}(u) = m_{a,A}^{(2)}(-u) \neq m_{a,A}^{(2)}(u)$ .
- 13.4.4 *Markov renewal process.* Let  $X(t)$  be a stationary semi-Markov process on countable state space  $\mathbb{X}$ , with counting function  $N_g(\cdot)$ , as in Example 10.3(a), and  $H_g(x) = E[N_g[0, x] \mid \text{transition in } X(t) \text{ at } t = 0]$ ,  $= E_{\bar{\mathcal{P}}_0}(N_g[0, x]) = \sum_i \pi_i \sum_j H_{ij}(x)$  [cf. (13.4.8–9) and Example 10.3(a) for notation]. Use  $\pi_k = \sum_i \pi_i \sum_j \int_0^x H_{ij}(du) \int_{x-u}^\infty G_{jk}(dz)$  (all  $x > 0$ ) to show that  $H_g(\cdot)$  is subadditive; that is,  $H_g(x + y) \leq H_g(x) + H_g(y)$  for  $x, y > 0$ . [Hint: Daley, Rolski, and Vesilo (2007) gives a proof.]
- 13.4.5 *Ergodic properties of nonergodic MPPs.* Mimic the steps in the proofs of Lemma 12.2.VI and Theorem 13.2.III to establish Lemma 13.4.II and Theorem 13.4.III.
- 13.4.6 *Alternative version of Palm distribution* [cf. Sigman (1995, Section 4.4)].
- Establish a form of Corollary 13.4.V for the nonergodic case, and hence verify the comment below the corollary.
  - Apply this form to Example 13.4(b) and verify the comment at the end of the example.
  - Find a Radon–Nikodym interpretation of this distribution [Sigman p. 65 quotes Nieuwenhuis (1994) and Thorisson (1995)].
- [Hint: The basic link comes from the r.v.  $Y$  controlling the invariant distribution. For (i), use Exercise 13.4.4 to express the result of the corollary in terms of the invariant random measure. In particular, argue from (13.4.11) (in the unmarked case) that

$$\frac{1}{k} \sum_{j=1}^k g(S_{t_j} N) \rightarrow \frac{1}{Y} \int \zeta(dN \mid \mathcal{I}),$$

where we have written  $\zeta(\cdot \mid \mathcal{I})$  to emphasize the dependence on the invariant  $\sigma$ -algebra. Now, following Sigman, write

$$\frac{1}{k} \sum_{j=1}^k g(S_{t_j} N) \rightarrow E[g(N_0) \mid \mathcal{I}] = \int g(N_0) \mathcal{P}_0(dN \mid \mathcal{I})$$

for an ergodic measure relative to discrete shifts. Deduce that  $E[\zeta(dN \mid \mathcal{I})] = (1/m)E[Y \mathcal{P}_0(dN \mid \mathcal{I})]$ .

- 13.4.7 *Equivalence of invariant  $\sigma$ -algebras for  $\mathcal{P}$  and  $\mathcal{P}_0$ .* If  $\mathcal{P}$  is stationary and  $\Gamma$  is an invariant set in  $\mathcal{I}$  with  $\mathcal{P}(\Gamma) = p$ , then  $\Gamma_0 = \Gamma \cap \mathcal{N}^\#(X_0)$  is invariant in  $\mathcal{I}_0$  with  $\mathcal{P}_0(\Gamma_0) = p$ . Deduce that a stationary point process  $N$  is ergodic if and only if the associated Palm distribution  $\mathcal{P}_0$  is ergodic (e.g., when  $\mathcal{X} = \mathbb{R}$ ,  $N$  is ergodic if and only if the corresponding interval process is ergodic).

[*Hint:* Replace the expectation in (13.4.15) by the conditional expectation with respect to  $\mathcal{I}$ ; then the integral on the right-hand side of (13.4.15) is zero outside  $\Gamma$ . A similar argument holds inside  $\Gamma$ . To prove the converse, if  $\Gamma_0 \in \mathcal{I}_0$ , first put  $\Gamma = \{N: S_{x_1(N)}N \in \Gamma_0\}$ , then use (13.2.9) with  $h$  the indicator of  $\Gamma$ . Otherwise put, ergodicity is equivalent to the requirement that all invariant sets have probability 0 or 1; now use Lemma 13.4.II, except that a converse to the lemma is needed:  $\mathcal{P}_0(\Gamma_0) = 1$  implies the existence of  $\Gamma$  in  $\mathcal{B}(\mathcal{N}_{\mathcal{X}}^\#)$ , such that  $\Gamma$  is invariant with  $\mathcal{P}(\Gamma) = 1$  and  $\Gamma_0 = \Gamma \cap \mathcal{N}_0^*$ .]

- 13.4.8 Let  $\mathcal{P}$  be the distribution of a stationary point process on  $\mathbb{R}_+$  for which  $\mathcal{P}\{N(\mathbb{R}) = \infty\} = 1$ . Restate (13.4.16–17) in terms of integrals on  $(0, t_n]$  with  $t_n \rightarrow \infty$ , where  $t_n$  is the  $n$ th point of the process in  $\mathbb{R}_+$ .

- 13.4.9 *Stationary marked interval process.* Consider a stationary sequence of pairs  $\{(\tau_i, \kappa_{i-1})\}$ , as below (13.4.17), with stationary mean ground rate  $m_g$ . Identify its distribution with the distribution of an MPP with a point at the origin with mark distributed according to the stationary distribution of marks in the original sequence. Then find a one-dimensional version of (13.4.10) to define the distribution of an MPP  $\mathcal{P}$  and verify that it is stationary. Find the distribution of the time to the first point occurring after the origin, and of the mark distribution of that point. Construct an example to illustrate that this may differ from the stationary mark distribution. [*Hint:* Condition on the value of the mark of the point at the origin, which by assumption has the stationary mark distribution. For a counterexample, consider, e.g., an alternating renewal process. See also Sigman (1995, Section 3.4).]

- 13.4.10 Let the multivariate point process  $N(\cdot) = (N_0(t), N_1(t), \dots, N_k(t))$  in  $t > 0$  be such that the successive points  $t_{00} \leq 0 < t_{01} < t_{02} < \dots$  of the component process  $N_0$  are regeneration epochs for  $N(\cdot)$  in the sense that, for  $u_i \geq 0$ , the conditional distributions of  $N(t + u_i)$  given  $t_{0j} = t$  for some  $j \geq 1$  are independent of  $t$ , and the interval lengths  $\{t_{0,j+1} - t_{0j}\}$  are i.i.d. r.v.s with density function  $f(\cdot)$ . Define

$$H(t) = E(N_1(t) \mid t_{00} = 0, t_{01} = t), \quad V(t) = \text{var}(N_1(t) \mid t_{00} = 0, t_{01} = t).$$

Assuming the moments are finite, show that

$$\lim_{t \rightarrow \infty} \frac{\text{var}(N_1(t))}{E(N_1(t))} \geq \beta \equiv \frac{\int_0^\infty V(t)f(t)dt}{\int_0^\infty H(t)f(t)dt},$$

with equality holding if and only if  $H(t)/t$  is constant a.e. on the support of  $f(\cdot)$ . [See Berman (1978) for this and other results concerning such multivariate point processes that are regenerative in the sense of Smith (1955). Results concerning ergodicity and Palm distributions have been extended from regenerative processes to the context of marked point processes; there are details and several examples in Franken *et al.* (1981).]

### 13.5. Cluster Iterates

We turn next to the operation of iterated cluster formation, already mooted in Section 11.4, but postponed then because it makes essential use of Palm theory concepts. We consider clusters with mean size unity, but exclude the case  $\mathcal{P}\{N(\mathcal{X} \mid x) = 1\} = 1$  (all  $x$ ) inasmuch as this is the case of random translations already considered in Section 11.4.

Standard results on branching processes [e.g., Harris (1963, Chapter 1)] imply that for critical branching processes (mean cluster size  $m = 1$ ) the offspring from a given ancestor eventually become extinct, or in other words the iterated clusters eventually collapse to the zero measures. In such circumstances it may seem surprising that stable limit behaviour can occur. The explanation is to be found in the infinite character of the initial distribution, which allows local depletion to be perpetually replenished by immigration from more successful clusters in distant parts of the state space. The higher the dimension of the space, the greater the opportunities for such replenishment become, so that stable behaviour is the norm for  $d \geq 3$ , whereas it is the exception for  $d = 1$  or even  $d = 2$ . An earlier account may be found in Liemant, Matthes, and Wakolbinger (1988); Wakolbinger has several subsequent publications with various co-workers.

The nature of the limiting behaviour is most easily understood by studying first the situation where the initial process is Poisson, with mean density equal to unity say. Any limit, although no longer Poisson, is still infinitely divisible, so the discussion can be phrased in terms of the convergence of the associated KLM measures  $\tilde{Q}_n$  to their limit (Proposition 11.2.II). Because the successive Poisson cluster processes formed by iterating the clustering operation are also stationary, the discussion can be further reduced to the study of the Palm measures  $\tilde{Q}_0^{(n)}$  associated with these KLM measures [see Example 13.2(c), especially the discussion leading to Proposition 13.2.IX, and Exercise 13.2.10]. The essential question can now be phrased in terms of the Palm measures  $\{\tilde{Q}_0^{(n)}\}$ , namely, find conditions on the cluster mechanism such that  $\tilde{Q}_0^{(n)}$  converges to some boundedly finite limit measure  $\tilde{Q}_0^{(\infty)}$  say.

The cluster mechanism itself can be conveniently specified in terms of

- (A) a distribution  $\{\pi_k: k \geq 0\}$  for the size of the cluster; and
- (B) a family of symmetric d.f.s  $P_k(dx_1 \times \cdots \times dx_k)$  ( $k \geq 1$ ) specifying the locations of the cluster members relative to the cluster centre at the origin.

The assumptions  $m \equiv E\{N(\mathcal{X} \mid x)\} = 1$  and  $\mathcal{P}\{N(\mathcal{X} \mid x) = 1\} < 1$  imply both

$$1 = \sum_{k=0}^{\infty} \pi_k = \sum_{k=0}^{\infty} k\pi_k \quad \text{and} \quad \pi_0 > 0.$$

Also, because by assumption the cluster mechanisms are homogeneous in space, the locations relative to a cluster centre not at 0 are specified by the appropriately shifted versions of the  $P_k(\cdot)$ .

The KLM measure corresponding to the Poisson cluster process formed at the first stage of clustering is concentrated on the set of totally finite counting measures and allocates mass  $\pi_k$  to those trajectories containing just  $k$  points. Because  $m = 1$ , it may be considered as a probability distribution on  $\mathcal{N}_0(\mathcal{X})$  in its own right. As in Example 13.2(c), the associated Palm measure is then defined in terms of a modified cluster structure, in which the cluster size is distributed according to

$$\{\pi'_k\} = \{k\pi_k: k \geq 0\}$$

(note that  $\pi'_0 = 0$ ), and the locations of the cluster members are specified by placing one cluster member at the origin and distributing the remaining  $k - 1$  points about it according to the symmetrized measures

$$\tilde{\mathcal{P}}_{k-1}(A_2 \times \cdots \times A_k) = \int_{\mathcal{X}} P_k(dy \times (y + A_2) \times \cdots \times (y + A_k)).$$

Note that the Palm clusters considered here differ from the clusters arising in the regular representation (Propositions 12.1.V and 12.4.II) only through the relative weightings given to the different cluster sizes. Note also that the intensity measure for the underlying cluster process, given by

$$\rho(dx) = \sum_{k=1}^{\infty} k\pi_k \int_{\mathcal{X}^{(k-1)}} P_k(dx \times dy_2 \times \cdots \times dy_k), \quad (13.5.1)$$

is here a probability measure on  $\mathcal{X}$ , whereas the intensity measure for the Palm cluster process is given by

$$\tilde{\rho}(dx) = \delta_0(dx) + \sum_{k=2}^{\infty} k(k-1)\pi_k \int_{\mathcal{X}^{(k-1)}} P_k((y_2 + dx) \times dy_2 \times \cdots \times dy_k).$$

Now consider the Palm cluster resulting from two stages of clustering. To ease the notation only, we use here density notation, with corresponding lower case symbols. First note that the quantity

$$k\pi_k p_k(y, x_2 + y, \dots, x_k + y) \quad (13.5.2)$$

can be interpreted as the joint density of locating the parent (cluster centre) at  $-y$  and  $k - 1$  siblings at  $x_2, \dots, x_k$ , given one point of the cluster at the origin (cf. Exercise 1.2.5). The marginal density for the parent, given a point at the origin, is thus

$$g(y) = \sum_{k=1}^{\infty} k\pi_k \int \cdots \int p_k(y, x_2 + y, \dots, x_k + y) dx_2 \cdots dx_k = \rho(y),$$

where we here write  $\rho(y)$  for the density of the intensity measure (13.5.1). The members of the two-stage Palm cluster can now be classified into three groups: first, the point located at the origin; second, its immediate siblings,



jointly located with the cluster centre according to (13.5.2); and third, its ‘cousins’, found by locating a grandparent and a set of ‘uncles’ by (13.5.2), but given the parent at  $-y$  rather than at the origin, and then superposing the clusters generated by each of the uncles. In symbols, we may write briefly

$$\tilde{N}_2 = \delta_0 + N_1 + N_2.$$

Evidently, this process, introduced by Kallenberg (1977a) and called by him the ‘method of backward trees’, can be continued. At each stage we move one generation further back, taking the location of what was previously the oldest ancestor as origin, locating the ancestor of next order and the siblings of the previously oldest ancestor by (13.5.2), then moving forward to add in to the current generation the superposition of clusters of appropriate order deriving from the siblings of the previously oldest ancestor.

The processes  $\tilde{N}_n$  developed in this way have a monotonic character, because we can imagine them as defined on a common probability space and embedded into an indefinitely continued process of superposition of this kind. Whether the Palm measures  $Q_0^{(n)}$  converge to some limit  $Q_0^{(\infty)}$  thus reduces to the question of whether this process of superposition produces in the limit an a.s. boundedly finite limit measure. Because each stage is formed from its predecessor by an independent operation representing a shift of locations and corresponding augmentation of the number of branches by the distribution (13.5.2), it follows from the Hewitt–Savage zero–one law that the limit is boundedly finite either with probability 1 or with probability 0.

This dichotomy allows us to make the following definition.

**Definition 13.5.I.** *The cluster mechanism described by (A) and (B) is stable or unstable according as the sequence of processes  $\tilde{N}_n$  described above converges a.s. to a boundedly finite limit or diverges a.s.*

In complete generality, the problem of determining conditions that are necessary and sufficient for the stability of a given cluster mechanism appears to be still open. What is known is that the conditions are closely linked to the behaviour of the random walk with step-length distribution governed by the symmetrized form

$$\sigma = \rho * \rho_-, \quad (13.5.3)$$

where  $\rho_-(A) = \rho(-A)$ , of the intensity measure for the clusters.

To see how this measure arises, suppose that the mean square cluster size  $\sum_{k=1}^{\infty} k^2 \pi_k$  is finite; this ensures that the Palm clusters also have finite intensity, which we write in the form  $\hat{\rho} = \delta_0 + \hat{\rho}$ , where

$$\hat{\rho}(A) = \sum_{k=2}^{\infty} (k-1) \tilde{\mathcal{P}}_{k-1}(A \times \mathcal{X} \times \cdots \times \mathcal{X}),$$

and we note

$$\hat{\rho}(\mathcal{X}) = \sum_{k=2}^{\infty} k(k-1) \pi_k.$$

Consider now the differences  $\tilde{N}_{n+1} - \tilde{N}_n$  between the Palm clusters at the  $(n+1)$ th and  $n$ th stages of clustering. To obtain the intensity measure of this increment, we should start from  $\rho_-^{n*}$ , representing the  $n$  steps taken to the left (i.e., according to  $\rho_-$ ) to locate the position of the  $n$ th generation ancestor, convolve this with  $\hat{\rho}$  to obtain the locations of that ancestor's siblings, and finally convolve again with  $\rho_+^{n*}$  to obtain the intensity measure of the superposition of the  $n$ th stage clusters generated by those siblings. Thus, if  $\tilde{\rho}_n$  denotes the intensity measure for  $\tilde{N}_n$ , we have

$$\tilde{\rho}_{n+1} = \tilde{\rho}_n + \rho_-^{n*} * \hat{\rho} * \rho_+^{n*} = \tilde{\rho}_n + \sigma^{n*} * \hat{\rho};$$

hence,

$$\tilde{\rho}_n = \delta_0 + \hat{\rho} * (\delta_0 + \sigma + \sigma^{2*} + \cdots + \sigma^{(n-1)*}).$$

The series on the right-hand side converges toward the renewal measure (boundedly finite or infinite) of the random walk with step-length distributions  $\sigma$ . It is boundedly finite if and only if the random walk is transient (see Exercise 9.1.11). On the other hand, it follows from the monotonic character of the  $\tilde{N}_n$  that they converge to a boundedly finite limit with boundedly finite first moment measure if and only if the sequence  $\tilde{\rho}_n$  so converges. We are thus led to the following result [see Liemant (1969, 1975)].

**Proposition 13.5.II.** *A critical cluster member process with finite mean square cluster size is stable if and only if the random walk generated by the symmetrized intensity measure (13.5.3) is transient.*

Recall that random walks in three or more dimensions are necessarily transient, so that a properly three-dimensional cluster process with finite mean square cluster size is always stable. In one or two dimensions, however, a random walk is not necessarily transient: it is if the step distribution has nonzero mean ( $d = 1$  or  $2$ ) or infinite variance ( $d = 2$ ), so it is only under particular conditions that the associated cluster process can be stable.

Kallenberg (1977a) gives further results concerning stability. In particular, for a process of Neyman–Scott type, transience of the random walk alone is necessary and sufficient for stability, but necessary and sufficient conditions for stability cannot be formulated in full generality solely in terms of conditions on the cluster size distribution and the cluster intensity measure.

Granted that the cluster mechanism is stable, convergence to a limiting process can be established by arguments similar to those used in discussing random translations. As above, write  $Q_0^{(n)}$  for the Palm measure corresponding to the Poisson cluster process formed after  $n$  stages of clustering from a Poisson process with unit rate, and suppose the clusters are stable, so that by hypothesis

$$Q_0^{(n)} \rightarrow Q_0^{(\infty)} \quad (n \rightarrow \infty) \quad (\text{weakly in } \mathcal{N}_0^\#)$$

for some limit distribution  $Q_0^{(\infty)}$ . This convergence implies the corresponding assertion for the associated KLM measures (see Exercise 13.5.3), namely,

$$\tilde{Q}_n \rightarrow \tilde{Q}_\infty \quad (n \rightarrow \infty) \quad (\text{weakly}). \quad (13.5.4)$$

Unlike  $Q_0^{(n)}$ ,  $Q_0^{(\infty)}$  (and hence also  $\tilde{Q}_\infty$ ) allots zero mass to the totally finite counting measures. To see this, observe that the successive increments  $\tilde{N}_{n+1} - \tilde{N}_n$  are independent and nonnegative, and that for positive constants  $c, c'$ ,

$$\mathcal{P}\{\tilde{N}_{n+1} - \tilde{N}_n \geq 1\} \geq c\mathcal{P}\{Z_n > 0\} \geq c'/n,$$

where  $\{Z_n\}$  is a Galton–Watson branching process governed by the cluster size distribution  $\{\pi_k\}$  (see Harris, 1963, Chapter 1). Because the sum of these terms diverges, it follows from the Borel–Cantelli lemmas that with probability 1, an infinite number of the events on the left-hand side occur. Thus,  $\lim_{n \rightarrow \infty} \tilde{N}_n(\mathcal{X})$  is infinite a.s., which is equivalent to the assertion that  $Q_0^{(\infty)}$  allots zero mass to the counting measures with finite total mass.

Now the various Poisson cluster processes formed from the initial Poisson process all have unit rate, so their distributions on  $\mathcal{N}_\mathcal{X}^\#$  are weakly relatively compact, and from (13.5.4) above and Proposition 11.2.II it follows that the limit of any weakly converging subsequence must be infinitely divisible with KLM measure  $\tilde{Q}_\infty$ . This limit process must therefore be the overall weak limit. Recalling that an infinitely divisible point process is singular if its KLM measure is supported by the counting measures with infinite total mass, we can assert the following result.

**Lemma 13.5.III.** *In the stable case, the Poisson cluster processes derived from an initial Poisson process of unit rate converge weakly to a limit point process that is stationary, singular infinitely divisible, and has KLM measure  $\tilde{Q}_\infty$ .*

It can be shown further that the limit process is actually mixing and therefore weakly singular (Fleischmann, 1978). Also, if we start from an initial Poisson process of rate  $\lambda$ , the Poisson cluster processes converge to a limit point process that is infinitely divisible with KLM measure  $\lambda\tilde{Q}_\infty$ .

Granted the Palm versions of the cluster iterates converge to the limit  $\tilde{Q}_\infty$ , we may raise more generally the question of convergence of the processes formed by successive clusterings from a general initial distribution. Because stability implies that the intensity measure  $\rho(\cdot)$  of the cluster member process has at least two points in its support, we obtain an estimate for the p.g.fl.  $G_n[h]$  for the  $n$ th cluster iterate with cluster centre at the origin, namely,

$$\sup_x |G_n[T_x h] - 1| \leq \theta_n \equiv \sup_x \int_{\mathcal{X}} \rho^{n*}(x + dy)(1 - h(y)) \quad (h \in \mathcal{V}),$$

and the last quantity  $\rightarrow 0$  as  $n \rightarrow \infty$  by Lemma 11.4.I.

Using this estimate, we can start the discussion along much the same lines as that of Theorem 11.4.II. As in that proof, the above estimate implies

$$\mathbb{E} \left( \left| - \int_{\mathcal{X}} \log G_n[T_x h] N_0(dx) - \int_{\mathcal{X}} (1 - G_n[T_x h]) N_0(dx) \right| \right) \rightarrow 0.$$

If now  $Y$  is given by (11.4.10) and we write  $G_\infty[h]$  for the putative limit of the terms in the exponent, namely,

$$G_\infty[h] = \int_{\mathcal{N}_0^\#(\mathcal{X})} \left\{ \exp \left[ \int_{\mathcal{X}} \log h(x) \tilde{N}(\mathrm{d}x) \right] - 1 \right\} \tilde{Q}_\infty(\mathrm{d}\mathcal{N}^\#),$$

then to complete the proof along the lines of Theorem 11.4.II we need to show that

$$\mathbb{E} \left\{ \left| \int_{\mathcal{X}} (1 - G_n[T_x h]) N_0(\mathrm{d}x) - Y G_\infty[h] \right|^2 \right\} \rightarrow 0.$$

Now we have already established that the KLM measures of the Poisson cluster process converge to  $\tilde{Q}_\infty$ , so we have further that

$$\int_{\mathcal{X}} (1 - G_n[T_x h]) \mathrm{d}x \rightarrow G_\infty[h],$$

and hence it is in fact sufficient to show that

$$\mathbb{E} \left\{ \left| \int_{\mathcal{X}} (1 - G_n[T_x h]) N_0(\mathrm{d}x) - Y \int_{\mathcal{X}} (1 - G_n[T_x h]) \mathrm{d}x \right|^2 \right\} \rightarrow 0. \quad (13.5.5)$$

At this point we meet a difficulty, because we do not have the detailed information concerning the asymptotic behaviour of the cluster p.g.f.s  $1 - G_n[t_x h]$ , in  $x$  and  $n$ , which would correspond to the information concerning the convolution powers  $1 - \int_{\mathcal{X}} h(y) \rho^{n*}(x + \mathrm{d}y)$  used to complete the proof of Theorem 11.4.II.

Such information can in fact be obtained by the ‘method of reduced trees’ introduced by Fleischmann and Prehn (1974, 1975). The underlying idea here is that if the clustering survives a large number of generations, the offspring in the current generation come with high probability from a single line in the family tree, other lines having become extinct. In other words, the current offspring have a single common ancestor a few generations back, so that for all generations preceding that it is enough to track the positions of this single ancestor and its forebears. Each backward step in this reduced part of the tree corresponds to a step in a random walk governed by the distribution  $\rho(\cdot)$ , as discussed below (13.5.2). Hence, we may approximate the p.g.f.s  $G_n[T_x h]$  in (13.5.5) by the corresponding p.g.f.s  $\int_{\mathcal{X}} h(y) \rho^{n*}(x + \mathrm{d}y)$  for the random translations process governed by the distribution  $\rho$ . Then we may refer to the proof of Theorem 11.4.II again to deduce that for this process, assuming  $\rho$  is nonlattice, the terms corresponding to those in (13.5.5) are asymptotically equal.

These considerations lead to the limit result set out in Theorem 13.5.IV below, representing an extension of Theorem 11.4.II. For details of the argument, as well as extensions and strengthenings of the theorem and analogous results for subcritical branching mechanisms, see MKM (1978, Chapter 12) and the extended and updated version in MKM (1982, Chapter 10).

**Theorem 13.5.IV.** *Let  $N_0$  be a second-order stationary point process on  $\mathcal{X}=\mathbb{R}^d$ , and let  $\{\pi_k, P_k\}$  be a stable cluster mechanism in the sense of Definition 13.5.I. Furthermore, let  $\tilde{Q}_\infty$  denote the limiting KLM measure (13.5.4) associated with the iterates of the cluster mechanism and  $\{N_n\}$  the sequence of point processes derived from  $N_0$  by successive independent clusterings according to  $\{\pi_k, P_k\}$ . If the intensity measure  $\rho$  for  $\{\pi_k, P_k\}$  is nonlattice, then the sequence  $\{N_n\}$  converges weakly to the limit point process with p.g.fl.*

$$G[h] \equiv \mathbb{E} \left[ \exp \left( Y \int_{\mathcal{N}_0^\#(\mathcal{X})} [\exp(\int_{\mathcal{X}} \log h(x) \tilde{N}(dx)) - 1] \tilde{Q}_\infty(d\tilde{N}) \right) \right],$$

where  $Y = \mathbb{E}[N_0(\mathbb{U}^d) \mid \mathcal{I}]$ .

### Exercises and Complements to Section 13.5

- 13.5.1 Find the class of point processes invariant under the cluster operation of Theorem 13.5.IV (i.e., find an analogue of Corollary 11.4.III to Theorem 11.4.II for Theorem 13.5.IV).
- 13.5.2 Let the critical branching mechanism  $\{\pi_k, P_k\}$  be neither stable nor a random translation. Show that the cluster iterates  $\{N_n\}$  starting from a stationary second-order point process  $N_0$  converge weakly to the zero point process. [Hint: Consider first the case where  $N_0$  is a stationary Poisson process.]
- 13.5.3 Complete the convergence results relating KLM measures and the Palm measures of  $n$ -stage Poisson cluster processes asserted in the text following Proposition 13.5.III. [Hint: See the discussion of weak convergence of KLM measures preceding Proposition 11.2.I, and link to Exercise 13.5.2.]

## 13.6. Fractal Dimensions

Although the concept of a fractal set, and many related concepts associated with the names fractals and multifractals, were introduced principally in the work of Mandelbrot (1982), the fractal dimensions we consider here have their origins in earlier work of Rényi (1959), who at that time was studying generalizations of the concept of entropy. They are characteristics of a measure rather than a set. Their link to point processes might seem obscure, because a point process, as a measure, has the dimension of a single point, namely, zero. The connection arises through estimation procedures for fractal dimensions, for example, box-counting and Grassberger–Procaccia estimates, which can be and often are applied to a wide range of point process data. In this case it turns out, as we describe below, that the estimates can be related to the moment measures of the Palm distribution of the underlying point process. Indeed, in such contexts, the estimation procedures might well be better directed toward a careful study of these moment measures than to limit properties which have relatively limited application and are fraught with practical difficulties.

The material for this section is based on Vere-Jones (1999). Falconer (1990) and Cutler (1991) provide basic references for the Rényi dimensions and the multifractal formalism, and Harte (2001) gives a broad overview of both concepts and statistical issues. The topic now comprises a major field in its own right, and has generated a huge literature; we consider here only the specific issues that arise in interpreting estimates of multifractal dimensions when the estimates are derived from point process data.

The Rényi or multifractal dimensions  $D_q$  are defined for a measure  $\mu$  on  $\mathbb{R}^d$  as the limits

$$D_q(\mu) = \lim_{\delta \rightarrow 0} \frac{\log \left\{ \int_{\mathcal{X}} [\mu(S_\delta(x))]^{q-1} \mu(dx) \right\}}{(q-1) \log \delta} \quad (13.6.1a)$$

for  $q \neq 1$ , where  $S_\delta(x)$  is a sphere radius  $\delta$ , centre  $x$ . In the special case  $q = 1$ ,

$$D_1(\mu) = \lim_{\delta \rightarrow 0} \frac{\int_{\mathcal{X}} \log [\mu(S_\delta(x))] \mu(dx)}{\log \delta}, \quad (13.6.1b)$$

a quantity sometimes called the *entropy dimension*.

For measures that have a bounded density with respect to Lebesgue measure over some bounded set  $A$ , and vanish outside that set, the Rényi dimensions are all equal to the dimension  $d$  of the space on which the measure is defined. For singular measures, the situation can be more complicated. In the case of the Cantor measure, for example, the Rényi dimensions are still equal, but to a value less than the dimension of the space. Such a measure is *unifractal*. In still other examples, the growth rates of the measure may vary from point to point, resulting in different weights being given to the increment  $\mu(dx)$  for different  $x$ ; in such examples, the Rényi dimensions will vary with  $q$ , and the measure is described as *multifractal*. Simple variants on the Cantor measure which possess this property are outlined in Exercise 13.6.1.

When  $\mu$  is a probability measure over a bounded observation region  $A$ , one might attempt to ascertain the values of the Rényi dimensions empirically by generating i.i.d. observations according to  $\mu$ , and replacing  $\mu$  in (13.6.1) by the corresponding empirical distribution  $\hat{\mu}(B) = N(B)/N(A)$ . In this context, integrals of the type appearing in the numerators of (13.6.1) are referred to as *correlation integrals* [cf. Harte (2001)], although strictly speaking the term refers to the particular case  $q = 2$ .

To keep to quantities that can be related to point process moment measures, we restrict our discussion to multifractal dimensions of positive integral order  $q = k \geq 2$ . In such cases, a correlation integral which we denote by  $C_k(\cdot)$  has a particularly simple interpretation, namely,

$$C_k(\delta, A, \mu) = \int_A [\mu(S_\delta(x))]^{k-1} \mu(dx) = \Pr\{M_k \leq \delta\}, \quad (13.6.2)$$

where

$$M_k = \max \{ \|X_1 - X_k\|, \|X_2 - X_k\|, \dots, \|X_{k-1} - X_k\| \}$$

and the  $X_j$  are i.i.d. with common distribution  $\mu$  which vanishes outside  $A$ . It then follows (see Exercise 13.6.2) that the relation  $D_k(\mu) = \eta$  is equivalent to the statement that, at the origin, the distribution of  $M_k$  has power-law growth of order  $\nu = (k-1)\eta$ , so that  $\Pr\{M_k \leq \delta\} = \phi(y) y^\nu$  for some function  $\phi(y)$  with  $|\log \phi(y)| = o(|\log y|)$ .

In applications, empirical correlation integrals of the above kind may be calculated for many different types of data, and used to obtain estimates of some quantity purporting to be a fractal dimension. In circumstances when there is no obvious generating measure  $\mu$ , however, it is not clear whether the quantities obtained empirically have any meaningful interpretation. The discussion which follows has the aim of elucidating this point in situations where the data are generated by a point process.

We start by writing the correlation integral in (13.6.1a) in a form more convenient for calculations with general point processes. Let  $\mathcal{I}_{k,\delta}(\cdot)$  denote the indicator function of the set

$$U_{k,\delta} = \{\mathbf{x}: \max_i |x_i - x_k| \leq \delta\},$$

so that

$$\int_A [\mu(S_\delta(x))]^{k-1} \mu(dx) = \int_{A^{(k)}} \mathcal{I}_{k,\delta}(\mathbf{x}) \mu^{(k)}(d\mathbf{x}), \quad (13.6.3)$$

where  $\mathbf{x}$  denotes a  $k$ -vector  $(x_1, \dots, x_k)$  with each component  $x_i \in \mathbb{R}^d$ . Replacing  $\mu$  by the empirical measure  $\hat{\mu}$  formed from the counting measure  $N(\cdot)$  from a finite set of observations  $\{x_1, \dots, x_{N(A)}\}$  over some bounded region  $A$ , we obtain  $C_k(\delta, A, \hat{\mu})$  which we write as

$$\begin{aligned} \hat{C}_k(\delta, A) &= \int_A [\hat{\mu}(S_\delta(u))]^{k-1} \hat{\mu}(du) \\ &= \frac{1}{[N(A)]^k} \int_{A^{(k)}} \mathcal{I}_{k,\delta}(u_1, \dots, u_k) N^{(k)}(du_1 \times \dots \times du_k) \\ &= \frac{1}{[N(A)]^k} \sum_{\text{perm}} \mathcal{I}_{k,\delta}(x_1^*, \dots, x_k^*), \end{aligned} \quad (13.6.4)$$

the last sum being taken over all permutations of the points  $x_1, \dots, x_{N(A)}$  of the realization in  $A$  taken  $k$  at a time.

Points lying outside  $A$  are ignored, a convention which can lead to bias if the measure of interest has continued support outside  $A$ . Because we are concerned with the limits for small testing sets, such edge effects are generally of smaller order than the main terms, but their possible existence needs to be borne in mind. In practice, they can be important even when the size of the testing set is as small as 10% of the observation region; see Harte (2001), especially Chapters 9 and 10, for discussion and illustration of these and related aspects.

For computations, it is convenient to consider contributions only from distinct pairs, triplets, and so on, taking advantage of the symmetries in the

counting measure to reduce the number of terms to be considered. In the present case, the function  $\mathcal{I}_{k,\delta}(x_1, \dots, x_k)$  is symmetric in the first  $k - 1$  arguments, so that each combination of one point of the realization for  $x_k$ , and a set of  $k - 1$  points for  $x_1, \dots, x_{k-1}$ , will be repeated  $(k - 1)!$  times, leading to the alternative estimate, a *quasi* factorial moment estimator,

$$\widehat{C}_{[k]}(\delta, A) = \sum_{j=1}^{N(A)} \sum_{\text{comb}'} \mathcal{I}_{k,\delta}(x_1^*, \dots, x_k^*) / \left[ N(A) \binom{N(A) - 1}{k - 1} \right], \quad (13.6.5)$$

where the inner sum is taken over distinct combinations of one term  $x_j^*$  and  $k - 1$  different terms  $\{x_1^*, \dots, x_k^*\} \setminus \{x_j^*\}$  from  $x_1, \dots, x_{N(A)}$ . No combinations with repeated points from the realization appear in this representation, so that in taking expectations it should be written as an integral against the modified product counting measure  $N^{[k]}$  used in defining the factorial moments (see Section 5.2). Another way of writing this last formula, which may help to show up its link to power-law growth, is outlined in Exercise 13.6.3.

We describe two general situations where the estimates based on the correlation dimension estimates (13.6.4) and (13.6.5) do lead to consistent estimates of a multifractal dimension. Both situations relate to a space–time point process, both use the same estimates, but the estimates are embedded in different limit processes, and the quantities that they estimate are likewise different. In one case the process is stationary in time and observations accumulate over a fixed spatial region, whereas in the other, the process is homogeneous in space, and observations are considered over an expanding sequence of spatial sets, the time span being held fixed. In the first situation, observations accumulate ever more densely over a bounded spatial region, until in the limit their spatial distribution approximates that of the first spatial moment measure over that region. In this situation the estimates can be related to the fractal dimensions of the first spatial moment measure. In the second situation, although we do not accumulate information about density variations in any particular spatial subregion, we do start to collect information about the behaviour of groups of points at various relative distances from each other, leading to the possibility of estimating the power-law growth of the reduced moment measures when spatial homogeneity is assumed.

The proposition below establishes consistency of the correlation integral estimates in these two cases. It provides a starting point for considering the limit behaviour of the multifractal dimension estimates themselves. We describe the process as space–time (cf. Section 15.4 below), this being a convenient nomenclature for the two components of the state space in which the point process exists. In part (b) the ‘time’ variable plays no role so mention of it has been omitted.

**Proposition 13.6.I.** *Let  $N$  be a simple space–time point process, with state space  $\mathcal{X} = \mathbb{R} \times \mathbb{R}^d$ , and  $k > 1$  a positive integer.*

- (a) *Suppose that  $N$  is stationary and ergodic in time, and that its first spatial moment measure exists and has stationary spatial distribution  $\mu$  over a*



given spatial set  $A$ . Then, for  $A$  fixed and time interval  $T \rightarrow \infty$ , both

$$\widehat{C}_k(\delta, A) \rightarrow C_k(\delta, A, \mu) \quad \text{and} \quad \widehat{C}_{[k]}(\delta, A) \rightarrow C_k(\delta, A, \mu)$$

a.s. and in  $L_1$ -norm.

- (b) Suppose that  $N$  is stationary on  $\mathcal{X} = \mathbb{R}^d$ , and that its moment measures exist up to order  $k$ . Let  $\check{M}_k(\cdot)$ ,  $\check{M}_{[k]}(\cdot)$  denote its reduced ordinary and factorial measures of order  $k$ ,  $\check{M}_{k-1}(\cdot)$  and  $\check{M}_{[k-1]}(\cdot)$  the corresponding Palm moment measures,  $m$  the mean spatial density,  $\{A_n; n = 1, 2, \dots\}$  a convex averaging sequence of sets in  $\mathbb{R}^d$ ,  $S_\delta^{(k)}$  the set  $\{x_1, \dots, x_k: \max |x_i| \leq \delta\}$ , and with  $N_n = N(A_n)$ ,  $N_n^{[r]} = N_n! / (N_n - r)!$ . Then as  $n \rightarrow \infty$ ,

$$\begin{aligned} \widehat{C}_k^*(\delta, A_n) &\equiv N_n^{k-1} \widehat{C}_k(\delta, A_n) \rightarrow m^{-1} \check{M}_k(S_\delta^{(k-1)}) = \check{M}_{k-1}(S_\delta^{(k-1)}), \\ \widehat{C}_{[k]}^*(\delta, A_n) &\equiv N_n^{[k-1]} \widehat{C}_{[k]}(\delta, A_n) \rightarrow m^{-1} \check{M}_{[k]}(S_\delta^{(k-1)}) \end{aligned}$$

a.s. and in  $L_1$  norm.

PROOF. The proofs are exercises in using the ergodic theorems of Sections 12.2 and 12.6, and the link between reduced moment measures and the moment measures of the Palm distribution established in Section 13.4 and quoted in the proposition.

For case (a) we need the extension of Proposition 12.2.IV to product integrals [see below (12.2.15)]. Write  $\mathbf{x}$  for the  $k$ -vector  $(x_1, \dots, x_k)$  as earlier,  $N_T(\cdot)$  for the projection of  $N(\cdot \times (0, T))$  onto the spatial component  $\mathbb{R}^d$  of  $\mathcal{X}$ , and  $N_T(A)$  for the total number of points in the observation region  $A$ . The proof of the following lemma is sketched in Exercise 13.6.4.

**Lemma 13.6.II.** *With the notation just given, suppose that the assumptions of Proposition 13.6.I(a) are satisfied and that  $h(\cdot)$  is a  $\mu^{(k)}$ -integrable function on  $(\mathbb{R}^d)^{(k)}$ . Then as  $T \rightarrow \infty$  with region  $A$  fixed,*

$$\frac{1}{T^k} \int_{A^{(k)}} h(\mathbf{x}) N_T^{(k)}(d\mathbf{x}) \rightarrow m^k \int_{A^{(k)}} h(\mathbf{x}) \mu^{(k)}(d\mathbf{x}). \quad (13.6.6)$$

If in addition  $h$  is symmetric, then

$$\left( \sum_{\text{comb}} h(\mathbf{x}_i) \right) / \binom{N_T(A)}{k} \rightarrow \int_{A^{(k)}} h(\mathbf{x}) \mu^{(k)}(d\mathbf{x}), \quad (13.6.7)$$

where the sum is taken over all combinations  $\mathbf{x}_i$  of  $k$  distinct elements from the realization  $(x_1, \dots, x_{N_T(A)})$ .

To establish part (a) of the proposition, apply (13.6.6) twice, the first time with  $h = \mathcal{I}_{k,\delta}$  and the second time with  $h \equiv 1$ , and take the ratio. Comparing the resulting equations with the expressions (13.6.3) and (13.6.4) yields the first statement from part (a) of the proposition. The second statement from part (a) follows similarly from (13.6.5) by applying the variant of (13.6.2) which holds for functions with symmetry only in the first  $k - 1$  arguments.

It is noteworthy here that the same limit is obtained whether or not we allow repeated indices in the sums over permutations of the sample elements. This is because the contributions from terms with multiple points, corresponding to the diagonal concentrations in the moment measures, are of lower order in  $N$  (or  $T$ ) than the terms from sets of distinct points.

The argument for (b) rests on the higher-order ergodic Theorem 12.6.VI. Replacing the sets  $B_i$  in (12.6.10) by spheres  $S_\delta \equiv S_\delta(0) \subset \mathbb{R}^d$  we obtain

$$\begin{aligned} \frac{1}{\ell(A_n)} \int_{A_n} \left( \prod_{j=1}^{k-1} N(x + S_\delta) \right) N(dx) &= \frac{1}{\ell(A_n)} \sum_{i=1}^{N_n} N^{(k-1)}((x_i + S_\delta)^{(k-1)}) \\ &\rightarrow \check{M}_k(S_\delta^{(k-1)}). \end{aligned} \quad (13.6.8)$$

This form also neglects edge effects, for it assumes that the process is observed not only within  $A_n$  but also within any parts of the translated spheres  $T_{x_i}S_\delta$  which happen to fall outside  $A_n$  even for  $x_i \in A_n$ . A sandwiching argument shows that such edge effects are asymptotically negligible provided  $\ell(A_n^{(\epsilon)})/\ell(A_n) \rightarrow 1$ , which we assume.

Now the sum in the middle term of (13.6.8) is just another way of expressing  $\sum_{\text{perm}} \mathcal{I}_{k,\delta}(x_1^*, \dots, x_k^*)$  from (13.6.4). Rewriting this in terms of  $C_k(\delta, A_n)$ , adjusting the scaling factor [only a single integral over  $A_n$  is involved in (13.6.8), whereas a multiple integral over  $(0, T)^{(k)}$  is implicit in (13.6.4) and Lemma 13.6.II], and recalling that  $N(A_n)/\ell(A_n) \rightarrow m$ , we obtain the first statement in part (b) of the proposition.

For the second statement, we omit repeated points, thereby obtaining a representation in terms of the factorial product counting measures  $N^{*[r]}$  whose expectation defines the factorial moments. In place of (13.6.8) above we start from

$$\frac{1}{\ell(A_n)} \sum_{i=1}^{N_n} N^{*[k-1]}[(x_i + S_\delta)^{(k-1)}] \rightarrow \check{M}_{[k]}(S_\delta^{(k-1)}). \quad (13.6.9)$$

The left-hand side of (13.6.9) can be rewritten as

$$\frac{(k-1)!}{\ell(A_n)} \sum_{i=1}^{N_n} \sum_{\text{comb}'} \mathcal{I}_{[k,\delta]}(x_1^*, \dots, x_k^*).$$

Replacing  $\ell(A_n)$  by  $N_n$ , the result is of the same form as (13.6.5) except for the factor  $N_n^{[k-1]}$ . Incorporating this factor and rewriting the expression in terms of the sum in (13.6.5) completes the proof<sup>2</sup> of (b).

As already noted, the limits in (b) can equally well be written in terms of moments of the Palm probabilities  $\check{M}_{k-1}(S_\delta^{(k-1)}) = m^{-1} \check{M}_k(S_\delta^{(k-1)})$ , representing the moment measures (of one order lower) for the process conditioned

<sup>2</sup> Note two errata in Vere-Jones (1999): (a) the last term in equation (35), should start with  $1/(q-1)!$  not  $(q-1)!$ , and (b) the scaling factor in equations (37) and (41) should read  $(N_n - 1) \dots (N_n - k + 1)$  and not  $(N_n - 1)!$  or  $[N(\mathcal{X}) - 1]!$ .

on a point at the origin. This interpretation is a natural one in the present context because the construction is based on the maxima  $M_k$  [see below (13.6.3)] which already singles out one point of the group as a local origin.  $\square$

A range of different estimates based on the correlation integrals have been proposed and their behaviour analyzed in different contexts. For example, the simplest, naïve, estimate is based directly on the definition (13.6.1a) and takes the form, for  $k = 2, 3, \dots$ ,

$$\hat{D}_{[k]}(\delta, A, T) = \frac{\log \hat{C}_{[k]}(\delta, A)}{(k-1) \log \delta}. \quad (13.6.10)$$

Its major drawback in practice is that the behaviour is unreliable when  $\delta$  is small because of measurement error and lack of data within very small spheres; on the other hand for large  $\delta$  it is likely to be significantly biased by edge effects, which we have ignored in the discussion above, and by any departures from simple power-law growth as the test region expands. To avoid at least some of these difficulties, Grassberger and Procaccia (1983) suggested replacing (13.6.10) by

$$\hat{D}_{[k]}^*(\delta, A, T) = \frac{\log \hat{C}_{[k]}(\delta_1, A) - \log \hat{C}_{[k]}(\delta_2, A)}{(k-1)[\log \delta_1 - \log \delta_2]}, \quad (13.6.11)$$

where the interval  $(\delta_1, \delta_2)$  is chosen essentially by inspection, to obtain a portion of the graph of (13.6.10) linear in  $\delta$ , and taking  $\delta_1$  as small as seems reasonable.

Let us note in passing, however, that although the Grassberger–Procaccia and similar procedures may establish the existence of power-law growth over a certain distance range, something which may well be of importance in its own right, it is another matter to assert that the power-law index for this range necessarily coincides with the limiting value at vanishingly small distances. In applications, the difficulty of distinguishing the two situations has been one factor leading to confusion over whether fractal behaviour refers generally to power-law growth, or specifically to the limiting behaviour near zero.

Another approach, developed by Mikosch and Wang (1995), is based on the Hill (1975) estimate; it uses extreme quantiles to estimate the limiting power-law growth. Mikosch and Wang also advocate use of a bootstrap method to estimate the dimension estimates and their confidence bounds. The Hill method is broadly similar to the approach of Takens (1985) which also bases the estimate on the behaviour in the extreme tail. The methods are reviewed and compared in Harte (1998, 2001) where the use of the Hill method is illustrated in practical situations in which bias from measurement error and finite boundary effects cannot be ignored.

Whatever form of estimate is adopted, it is a further nontrivial exercise to establish conditions for consistency of the estimates. Because the fractal dimension is itself defined as a limit, a double limit problem is involved: as

either  $T$  or  $N(T) \rightarrow \infty$ , and as  $\delta \rightarrow 0$ . In general, the maximum rate at which  $\delta \rightarrow 0$  will be constrained by the rate at which the study region expands in time or space. A slightly unusual form of limit process is required, in which one dimension shrinks to zero and the other expands to infinity.

Results for the naïve estimate (13.6.10) are summarized in the theorem below, and in Exercises 13.6.5–6. Consistency of a Grassberger–Procaccia type of estimate, for the index of power-law growth over a predetermined range, is more easily established on the basis of Proposition 13.6.I; an outline is given in Exercise 13.6.7.

**Theorem 13.6.III.**

- (a) Suppose that the conditions of Proposition 13.6.I(a) hold, that the moment measures for the space–time process  $N$  exist up to order  $2k$ , and that the stationary space distribution  $\mu$  has  $k$ th fractal dimension  $D_k(\mu)$ . Defining

$$\widehat{D}_{[k]}(\delta, A, T)^+ = \min\{\widehat{D}_{[k]}(\delta, A, T), d\},$$

if the controlled diagonal growth conditions of Exercise 13.6.5 hold, then  $\widehat{D}_{[k]}(\delta, A, T)^+$  is a mean-square consistent estimate of  $D_k(\mu)$ .

- (b) Suppose that the conditions of Proposition 13.6.I(b) hold, that the moment measures of  $N$  exist up to order  $2k$ , and that the reduced factorial moment measure  $\check{M}_{[k]}$  is such that the limit

$$D_{[k]}^* \equiv \lim_{\delta \rightarrow 0} \frac{\log \check{M}_{[k]}(S_\delta^{(k-1)})}{(k-1) \log \delta} \quad (13.6.12)$$

exists and is finite. Setting

$$\widehat{D}_{[k]}^*(\delta, A, T)^+ = \min \left\{ \frac{\log \widehat{C}_{[k]}^*(\delta, A)}{(k-1) \log \delta}, d \right\},$$

if the bounded growth conditions of Exercise 13.6.6 hold, then  $\widehat{D}_{[k]}^*(\delta, A, T)^+$  is a mean-square consistent estimate of  $D_{[k]}^*$ .

PROOF. We make a few comments only, referring to Vere-Jones (1999) and the Exercises 13.6.5–6 for details.

The nature of the problems which arise is most easily illustrated by considering the expected value of the quantity which appears in the numerator of (13.6.5). Written out directly in terms of the modified product counting measure  $N^{[k]}(\cdot)$ , which omits repeated points, it takes the form

$$\int_{A^{(k)}} \mathcal{I}_{k,\delta}(x_1^*, \dots, x_k^*) N^{(k)}(dx_1^*, \dots, dx_k^*),$$

where  $A$  is the region over which points are observed.

In part (a), the spatial process in view is the ground process of the space–time process over the region  $A \times (0, T)$ , time being treated as a mark and then

ignored. Its expectation therefore reduces to the integral over  $[A \times (0, T)]^{(k)}$  of the  $k$ th order factorial moment  $M_{[k]}(\cdot)$  of the space-time process:

$$\begin{aligned} \mathbb{E} \left[ \sum_{j=1}^{N(A)} \sum_{\text{comb}'} \mathcal{I}_{k,\delta}(x_1^*, \dots, x_k^*) \right] \\ = \int_{[A \times (0, T)]^{(k)}} \mathcal{I}_{k,\delta}(x_1, \dots, x_k) M_{[k]}(dx_1 \times dt_1 \times \dots \times dx_k \times dt_k). \end{aligned}$$

In case (a), we write

$$\begin{aligned} M_{[k]}(dx_1 \times dt_1 \times \dots \times dx_k \times dt_k) \\ = \prod_{i=1}^k M_1(dx_i \times dt_i) + \Delta(dx_1 \times dt_1 \times \dots \times dx_k \times dt_k), \end{aligned} \quad (13.6.13)$$

where in view of stationarity,  $M_1(dx \times dt) = m \mu(dx) dt$ ,  $m$  being the mean rate of occurrence of points over the whole region  $A$  and  $\mu$  being normalized to a probability measure. The product term is what we would expect if we were looking for dimension estimates of  $\mu$ , whereas the term  $\Delta(\cdot)$ , which in general consists of an amalgam of lower order factorial moment and cumulant measures, defines the error which must be controlled if consistent estimates are to be obtained. Indeed, the ‘controlled diagonal growth’ condition of Exercise 13.6.5 puts a bound on the growth of the expected value of the integral against  $\Delta$ , and the ‘controlled bi-diagonal growth’ condition puts a bound on the growth of its variance.

In part (b), the time variable plays no significant role, but the process is homogeneous in space. In this case, omitting the  $dt$  terms, we can write

$$\begin{aligned} \int_{A^{(k)}} \mathcal{I}_{k,\delta}(x_1, \dots, x_k) M_{[k]}(dx_1 \times \dots \times dx_k) \\ = \int_A dx_k \int_{A^{(k-1)} - x_k \mathbf{1}} I_{(\max(|u_1|, \dots, |u_{k-1}|) < \delta)} \check{M}_{[k]}(du_1 \times \dots \times du_{k-1}) \\ = \ell(A)(1 - \epsilon^*) \check{M}_{[k]}(S_\delta^{(k-1)}), \end{aligned}$$

where  $\epsilon^*$  is a measure of the edge effects (which become negligible when  $\delta$  is very small) but there is no other bias term. Here, only a constraint on the growth of the variance is needed to guarantee consistency, as indicated in Exercise 13.6.6.  $\square$

**EXAMPLE 13.6(c) Space-time Poisson processes.** Consider first the setting in part (a) of both Proposition 13.6.IV and Theorem 13.6.VI. Here it is natural to consider a stationary but spatially inhomogeneous Poisson process with intensity measure of the form  $\Lambda(dt \times dx) = m \mu(dx) dt$ , where  $m$  is the mean rate per unit time and space, and the stationary distribution  $\mu$  over space is normalized to form a probability distribution. This is essentially the classical case of finding the fractal dimension from i.i.d. observations with distribution

$\mu$ . The conditions for part (a) hold for general  $\mu$ . Because all factorial cumulants of order 2 and greater vanish for a Poisson process, it is most convenient to deal with the combinatorial form (13.6.5). Then the only constraints on the rate at which  $\delta \rightarrow 0$  are imposed by the growth behaviour of  $\mu$ . In particular, the rate of growth of the variance of the estimate depends on the rate of growth of the distributions of the random variables  $M_2$  and  $M_3$  in (13.6.2). Checking the conditions in Exercise 13.6.4, we find that a sufficient condition for the naïve estimate to be consistent is that  $\delta \rightarrow 0$  no faster than  $T^{-1/D_2}$ , where  $D_2$  is the Rényi dimension of order 2 of  $\mu$ . It will differ from the dimension of the space only in some extreme cases, as when  $\mu$  is concentrated along a set of lines, or has some other singularity which affects the dimension.

If we take the other interpretation, of parts (b), then we must assume that the process is homogeneous in space. The factorial cumulants vanish, and the reduced factorial moment measure  $\tilde{M}_{[r]}(\cdot)$  is proportional to Lebesgue measure on  $(\mathbb{R}^d)^r$ . Hence the dimension is the dimension of the space.  $\square$

The next example illustrates the type of situation more commonly met with in geophysical or similar applications, where dimension estimates are used to gain an impression of the character of small-scale clustering behaviour. We use it also to illustrate the possibility that different rates of power-law growth may occur over different ranges in time or space, and that the dimension estimates may pick up any one or a combination of these, according to the range selected.

**EXAMPLE 13.6(d)** *A space-time Neyman-Scott process with singular components* [cf. Example 6.3(a) and Exercise 6.3.10]. We consider a two-dimensional spatial region, and suppose that cluster centres follow a Poisson process with space-time intensity  $\Lambda(dx \times dt)$ , and that cluster members are i.i.d. about the cluster centre, with a spatial distribution described by the probability measure  $B(dy)$ ,  $y$  being the position of the cluster member relative to the cluster centre as origin, and after a time lag governed by the probability measure  $C(dt)$ . Suppose also that the number of points in a cluster has a discrete distribution  $\{p_n: n > 0\}$  with finite factorial moments  $\nu_{[j]}$  for  $j = 1, \dots, k$ . We consider the behaviour of the correlation integral ( $k = 2$ ) and its estimates under various assumptions concerning the components.

Consider first the behaviour under assumptions of Theorem 13.6.III(a), so that  $\Lambda(dx \times dt) = \lambda dt \theta(dx)$ , where  $\theta$  is a probability measure. After integrating out the time factor, we find for the ground (spatial) process

$$M_1^g(dx) = \lambda T \nu_1 \int B(dx - y) \theta(dy). \quad (13.6.14)$$

Also for  $k = 2$ , the expansion (13.6.13) takes the particularly simple form [cf. equations (6.3.5), (6.3.17)],

$$\begin{aligned} M_{[2]}^g(dx_1 \times dx_2) &= M_1(dx_1) M_2(dx_2) + C_{[2]}^g(dx_1 \times dx_2) \\ &= (\lambda T \nu_1)^2 \theta(dx_1) \theta(dx_2) + \lambda T \nu_2 \int B(dx_1 - y) B(dx_2 - y) \theta(dy). \end{aligned} \quad (13.6.15)$$

Integrating against  $I_{2,\delta}$  we obtain, denoting the convolution of  $\theta$  and  $b$  by  $\theta * B$ ,

$$\begin{aligned} \int_{A \times A} I_{2,\delta}(x_1, x_2) M_{[2]}^g(dx_1 \times dx_2) &= (\lambda T \nu_1)^2 \int_A (\theta * B)[S_\delta(x)] (\theta * B)(dx) \\ &\quad + \lambda T \nu_2 \int_A \left[ \int B(dx_1 - y) B(dx_2 - y) \right] \theta(dy). \end{aligned} \quad (13.6.16)$$

Bearing in mind that  $\theta$  is a probability distribution, the second term on the right-hand side can be written in the form

$$\lambda T \nu_2 [1 - \epsilon(\delta, A)] \int_A B[S_\delta(x)] B(dx), \quad 0 < \epsilon(\delta, A) < 1,$$

where the correction term  $\epsilon$  becomes negligible as  $\delta \rightarrow 0$ .

The order of growth of the terms in the right-hand side of (13.6.16) is  $T^2 \delta^{D_2(\theta * B)}$  for the first, and  $T \delta^{D_2(B)}$  for the second. Their relative behaviour can be very different under different assumptions. If both cluster centre and cluster member processes have bounded densities, then all terms grow like small areas (i.e., proportional to  $\delta^2$ ), so that the correlation dimension is 2. Moreover the second term in (13.6.16) is then  $O(T)$  and the first term is  $O(T^2)$  so the first term dominates and the controlled growth conditions are satisfied even for  $\delta = O(1/T)$ .

If the clusters are dispersed, so that  $B$  has a bounded density  $b$  say, but the cluster centres are concentrated along a set of lines in two-dimensional space, then the correlation dimension of  $\theta$  is 1, but the first moment measure of the point process still has a smooth density  $m_1(x) = \lambda T \nu_1 \int b(x - y) \theta(dy)$  corresponding to a correlation dimension of 2. This is typically the situation when observations are contaminated by spatial measurement errors, so the correlation integral grows as  $O(\delta^2)$  until  $\delta$  reaches the same order of magnitude as the limits of measurement error, or in our case the effective range of the distribution  $B$ , say  $\delta_0$ . When  $\delta$  is increased beyond  $\delta_0$ , the linear concentration of  $\theta$  starts to tell, and the correlation integral starts to grow as  $O(\delta \times \delta_0)$  [i.e., as  $O(\delta)$ ], corresponding to a fractal dimension estimate of 1 rather than 2.

If the converse situation obtains, so that cluster centres are smoothly (say uniformly) distributed over the observation area, but the cluster members are distributed along a line, then the first moment measure is now uniform, corresponding to growth rate  $O(\delta^2)$ , but the first term may no longer dominate the expression (13.6.16). Its overall contribution is  $O(T^2 \delta^2)$ , whereas that of the second term is  $O(T\delta)$ . For fixed  $T$ , the correlation integral grows initially as  $O(\delta)$  and then as  $O(\delta^2)$ , so it shows two regions of power-law growth. If we look for a consistent estimate, and take  $\delta = O(T^{-(1+\eta)})$ , then the first term will be  $O(T^{-2\eta})$  and the second term  $O(T^{-\eta})$ . In this case the bounded growth conditions break down, and the dimension estimate is dominated by the behaviour of the local clustering. If we take  $\delta = O(T^{-(1-\eta)})$ , then the first

term is  $O(T^{2n})$  and the second is  $O(T^n)$ , so the first-order term dominates; in this case the clusters accumulate sufficiently thickly over the observation region that their local linear structure is no longer the dominating feature, and the dimension estimates revert to estimates of the first moment measure. In general, the more extreme the concentrations of the local clusters are, the more slowly  $\delta$  will need to approach zero before the conditions of the theorem are satisfied. In such situations, the correlation integral may show a sequence of ranges, all with different power-law growths: which of these is picked out as the dimension estimate will depend on the rate at which  $\delta$  approaches 0 as  $T \rightarrow \infty$  (see Exercise 13.6.8 for a more elaborate example).

Finally consider the behaviour under the assumptions of part (b). The same basic equations hold, and the estimates are still controlled by the growth behaviour of the second factorial moment in (13.6.16). Ignoring the time coordinate, and recalling that the process is assumed spatially homogeneous, (13.6.16) directly gives the reduced moment measure  $\bar{M}_{[2]}(S_\delta(0))$ , with the additional simplification that the first moment measure here averages out to a multiple of Lebesgue measure;  $\theta$  in the second term is again proportional to Lebesgue measure. Then the same issues arise, but in a slightly different form, because the question now is to determine the correlation dimension of (13.6.16) as a whole, meaning therefore the initial growth rate, from whichever term that rate happens to derive. In the examples we have just been considering, when  $B$  has a bounded density both terms are  $O(\delta^2)$  so the correlation dimension is again 2. With locally linear clusters, it is the second term which will dominate for  $\delta \rightarrow 0$  and so the dimension here will be 1. We see that for this particular example, the effect of changing from the assumptions in (a) to those in (b) is to switch attention from the growth rate of the first moment measure to the growth rate of whatever feature of the cluster structure dominates the behaviour in the initial range of power-law growth.  $\square$

Another commonly occurring and widely studied situation relates to processes generated by a dynamical system, where the measure  $\mu$  of interest is an invariant measure for the process. The common examples are deterministic in character, but can be randomized by introducing a random starting point. We do not look at this example in great detail, referring the reader rather to Cutler (1991), Serinko (1994), and the review in Harte (2001) for further discussion and references.

**EXAMPLE 13.6(e)** *Point processes generated by a dynamical system* [Serinko (1994)]. Let  $\Theta$  be a measurable mapping taking the closed, bounded set  $A \in \mathbb{R}^d$  into itself, and let  $\mu$  be a totally finite invariant measure for  $\Theta$ ; we suppose  $\mu$  normalized to form a probability measure. We consider the application of Theorem 13.6.III(a) to the point process formed by the sequence  $\{x_n\} = \{\Theta^n x_0: n = 0, 1, \dots\}$ . If the initial value  $x_0$  itself has distribution  $\mu$ , then the sequence  $\{x_n\}$  is stationary; we suppose that it is also ergodic. The process may also be regarded as a space-time point process with discrete time variable. The results of Proposition 13.6.I and Theorem 13.6.III(a) are not



affected by the character of the time variable, because they all concern limits for large values of  $T$ . The task is then to find the moment measures of the resulting point process, and to verify the conditions of Theorem 13.6.III(a).

For bounded measurable functions  $h(x)$  on  $A$  and  $h_2(x, y)$  on  $A^{(2)}$  we can characterize the actions of the first and second moment measures of the process on  $A$ , up to time  $T$ , by means of the equations

$$\int_A h(x) M_1(dx) = \sum_{k=1}^T \int_A h(\Theta^k x) \mu(dx) = T \int_A h(x) \mu(dx)$$

and

$$\begin{aligned} \int_{A^{(2)}} h(x, y) M_2(dx \times dy) &= \sum_{k=1}^T \sum_{\ell=1}^T \int_A h(\Theta^k x, \Theta^\ell x) \mu(dx) \\ &= T \int_A h_2(x, x) \mu(dx) + \sum_{r=1}^{T-1} (T-r) \int_A [h_2(x, \Theta^r x) + h_2(\Theta^r x, x)] \mu(dx). \end{aligned}$$

To verify the bounded diagonal growth condition for  $M_2$  we should set  $h_2$  in the second equation to be the indicator  $I_{\|x-y\| \leq \delta}$  and examine the behaviour of the integral as  $T \rightarrow \infty$ ,  $\delta \rightarrow 0$ . Evidently, the critical feature will be the rate at which  $\|\Theta^r x - x\|$  increases with  $r$ . If the rate is fast enough, the bounded growth condition will hold. For large  $r$ , we expect the distribution of  $\Theta^r x$  to approximate  $\mu$ , for  $\mu$ -almost all  $x$ . The rate at which this occurs is governed by mixing conditions on  $\Theta$ . When appropriate mixing conditions are satisfied, therefore, we expect the estimates to be consistent. Serinko (1994) gives details from a somewhat different point of view.  $\square$

## Exercises and Complements to Section 13.6

**13.6.1 Multinomial measures.** Let  $b$  be a positive integer, and consider a division of the unit interval into  $b$  successive subintervals of length  $1/b$ , then a further subdivision of each such subinterval into  $b$  equal sub-subintervals, and so on. Starting with unit mass for the whole interval, at each stage of this process, divide the mass of any given interval among its component subintervals in proportions  $\{p_1, p_2, \dots, p_b\}$ , with  $\sum_{r=1}^b p_r = 1$ . At the  $n$ th stage of this process, the subinterval corresponding to the  $b$ -adic expansion  $0.\omega_1\omega_2 \dots \omega_n$  will have mass  $\prod_{j=1}^n p_{\omega_j}$ . Denote the corresponding probability distribution by  $\mu_n$ . By considering the values of the distribution function at points with finite  $b$ -adic expansions, or otherwise, show that the Rényi dimension of order  $q$  of  $\mu_n$  is approximately equal to  $-\log_b[\sum_{r=1}^b p_r^q]/(q-1)$ , and converges to this value as  $n \rightarrow \infty$ . Show also that as  $n \rightarrow \infty$ , the measures  $\mu_n$  converge weakly to a limit  $\mu$ , and that the fractal dimensions converge to the fractal dimensions of the limit measure. Investigate conditions under which the  $D_q$  are equal. [See, e.g., Harte (2001, Chapter 3).]

13.6.2 Prove the equivalence of the relations [see below (13.6.2)]  $D_k(\mu) = \eta$  and  $\Pr\{M_k < y\} = \phi(y) y^{(k-1)\eta}$  for some function  $\phi(y)$  such that  $\log \phi(y)/\log y \rightarrow 0$  as  $y \rightarrow 0$ . [Hint: To establish the basic link between  $C_k(\delta, A, \mu)$  and the distribution of  $M_k$ , condition on  $X_k$  and then take expectations.]

13.6.3 For a given set  $\{x_1, \dots, x_k\}$  of  $k$  distinct points, let the function

$$N_\delta(x_1, \dots, x_k) = \sum_{i=1}^k \left( \prod_{j \neq i} I_{S_\delta}(x_i - x_j) \right)$$

count the number of points of the set with the property that the remaining  $k-1$  points of the set lie within distance  $\delta$  of the selected point. Show that in this notation, (13.6.5) can be written as

$$C_{[k]}(\delta, A) = \frac{\sum_{\text{comb}} N_\delta(x_1^*, \dots, x_k^*)}{k} \bigg/ \binom{N(A)}{k},$$

the sum being taken over all distinct combinations  $(x_1^*, \dots, x_k^*)$  of observation points.

13.6.4 Prove Lemma 13.6.II, assuming first that  $h$  is of product form  $\prod_1^k h_r(x_r)$  with each  $h_r$  measurable and bounded on  $A$ , and extending via linear combinations to general  $h$ . [Hint: When  $h$  is a product, (13.6.6) is just the product of limits from the one-dimensional case. To derive (13.6.7) take ratios of (13.6.5) first for the given  $h$  and then for  $h \equiv 1$ . Because of the symmetry of  $h$ , each combination of distinct terms appears  $k!$  times, and contributions from repeated arguments are of lower order and can be neglected in the limit.]

13.6.5 Suppose  $N(\cdot \times \cdot)$  is an orderly space-time point process, stationary and ergodic with respect to time, observed over the bounded spatial region  $A$ , and with moment measures existing up to order  $2k$ . In Theorem 13.6.III(a) let  $\Delta_k$  denote the signed measure  $\Delta_k = M_k - m^k(\mu \times \ell)^{(k)}$ ,  $\Delta_k^+$  its total variation, and let  $V_{k,\delta}$  denote the restriction of the set  $U_{k,\delta}$  defined in (13.6.3) to  $A^{(k)}$ . Say that  $N$  has *controlled diagonal growth of order  $k$* , if  $\Delta_k$  satisfies the condition (controlling the bias in replacing the  $k$ th moment measure by its first-order approximation)

$$\Delta_k^+[V_{k,\delta} \times (0, T)^k] \leq CT^{(k-\alpha)}\delta^\beta$$

for some positive constants  $C, \alpha, \beta$  and sufficiently small  $\delta, 1/T$ . Furthermore, say that  $N$  has *controlled bi-diagonal growth of order  $k$*  if  $\Delta_{2k}$  satisfies the condition (controlling the growth of the variance)

$$\Delta_{2k}^+[V_{k,\delta}^{(2)} \times (0, T)^{2k}] \leq C' T^{(2k-2\eta)}\delta^{2\nu}$$

for positive constants  $C', \eta, \nu$  and sufficiently small  $\delta, 1/T$ . Show that if

- (i) the stationary distribution  $\mu$  over  $A$  has  $k$ th order Rényi dimension  $D_k(\mu)$ ,
- (ii)  $N$  has controlled diagonal and bi-diagonal growth of order  $k$ , with constants as above, and

- (iii)  $\delta$  is chosen to vary with  $T$  in such a way that  $\delta \rightarrow 0$  but  $T^r \delta \rightarrow \infty$ , where

$$r > \max \left\{ \frac{\alpha}{(k-1)D_k(\mu) - \beta}, \frac{\eta}{(k-1)D_k(\mu) - \nu} \right\},$$

then the consistency result of Theorem 13.6.III(a) holds.

[Hint: See Vere-Jones (1999, Proposition 2) for details.]

- 13.6.6 Let  $N(\cdot \times \cdot)$  be orderly, stationary (homogeneous), and ergodic with respect to space; let  $N$  be observed over a convex averaging sequence  $\{A_n\}$  of sets in  $\mathcal{X} = \mathbb{R}^d$ , and suppose its moment measures up to order  $2k$  exist. In Theorem 13.6.III(b), let  $\Delta_{2k}^*$  denote the signed measure  $\Delta_{2k}^* = M_{[2k]} - M_{[k]} \times M_{[k]}$ , and  $V_{k,\delta}^n$  the restriction of  $U_{k,n}$  defined in (13.6.18) to the set  $A_n$ . Say that  $N$  has *controlled Palm growth of order  $k$*  if  $\Delta_{2k}^*$  satisfies the condition (controlling the growth of the variance)

$$(\Delta_{2k}^*)^+(V_{k,\delta}^{(2)}) \leq K \ell(A_n)^{2k-2\eta} \delta^{2\nu}$$

for positive constants  $(K, \eta, \nu)$  and sufficiently small  $\delta, 1/n$ . Show that if

- (i) the limit  $D_{[k]}^*$  of (13.6.12) exists,
- (ii)  $N$  has controlled Palm growth of order  $k$ , with constants as above, and
- (iii)  $\delta$  is chosen to vary with  $T$  in such a way that  $\delta \rightarrow 0$  but  $T^r \delta \rightarrow \infty$ , where

$$r > \frac{\eta}{(k-1)D_{[k]}^* - \nu},$$

then the consistency result of Theorem 13.6.III(b) holds.

[Hint: See Vere-Jones (1999, Proposition 4) for details.]

13.6.7 *Grassberger–Procaccia estimates.*

- (a) Suppose the correlation integral  $C(k, \delta, \mu)$  of (13.6.2), where  $\mu$  is the stationary spatial distribution under assumptions (a) of Proposition 13.6.I, shows power-law growth over a given interval  $(a, b)$ , so that

$$\frac{\log C(k, b, \mu) - \log C(k, a, \mu)}{\log b - \log a} = \eta,$$

say. Show that under assumptions (a), replacing  $C$  in the above expression by its sample counterpart (13.6.5) yields a consistent estimate of  $\eta$ .

- (b) Formulate and prove a similar result under assumptions (b).

13.6.8 *Further Neyman–Scott examples.*

- (a) Consider Example 13.6(d) but with homogeneous cluster centres and cluster structure determined by a spatial component  $B$  that has gamma distribution on a line with shape parameter  $\alpha < \frac{1}{2}$ , so that the density has a singularity at 0. Show that if  $X, Y$  both have such a distribution, then  $X - Y$  also has a singularity at 0, with initial power-law growth  $2\alpha$ . Show that the correlation integral (13.6.31) has three ranges of power law growth, initially with  $\delta = 2\alpha$ , then with  $\delta = 1$ , and finally with  $\delta = 2$ . Investigate the consequences for the correlation dimension estimates.
- (b) Extend to  $\mathbb{R}^3$ , and investigate also the behaviour when both cluster centre and cluster structure components have linear or planar concentrations.

## CHAPTER 14

# Evolutionary Processes and Predictability

14.1	Compensators and Martingales	356
14.2	Campbell Measure and Predictability	376
14.3	Conditional Intensities	390
14.4	Filters and Likelihood Ratios	400
14.5	A Central Limit Theorem	412
14.6	Random Time Change	418
14.7	Poisson Embedding and Existence Theorems	426
14.8	Point Process Entropy and a Shannon–MacMillan Theorem	440

The ideas discussed in this chapter have transformed the study of point processes over the last few decades. They provide the background not only for the results on conditional intensities and likelihoods summarized in Chapter 7, but also for general theories of estimation, prediction and control that have been influential as much in the engineering as in the statistical communities. The introduction to Chapter 7 provides some references to the early literature. Last and Brandt (1995) give a thorough study of the theory for marked point processes; other recent texts covering related material include Asmussen (1987, 2003), Baccelli and Brémaud (1994, 2003) and Andersen *et al.* (1993).

Broadly speaking, the chapter provides a setting for the functionals that arise in describing the evolution, or ‘dynamics’, of a simple or marked point process evolving in time. This setting embraces certain broad structural features of a point process embodied in the Doob–Meyer decomposition, Theorem A3.4.IX, which is a basic tool. A point process  $N(\cdot)$  on  $\mathbb{R}_+$  is equivalent to the nondecreasing function  $N(t) = N((0, t])$  for which it is plausible that we should be able to separate  $N(\cdot)$  into a ‘generally’ increasing part  $A(\cdot)$ , its compensator (here,  $A$  comes from *accroissement* or ‘growing’ part), and the ‘unpredictable’ variable part  $M(\cdot)$ ; that is,  $N(t) = A(t) + M(t)$ . The essence of the Doob–Meyer result is that such a decomposition is possible with  $A(\cdot)$  ‘predictable’ and  $M(\cdot)$  a martingale.

The notion of *predictability* which arises here is crucial for the development of a rigorous theory; it forms one part of the so-called ‘general theory of

processes', as set out in Dellacherie and Meyer (1978), for example. A brief outline of some key points is provided in Appendix 3.

Section 14.1 gives a general introduction to the notions of compensator and point-process martingale, and their links to the Doob–Meyer theorem. In Section 14.2 we extend these concepts to random measures and marked point processes. Section 14.3 re-introduces the conditional intensity, which appears here as a Radon–Nikodym derivative of Campbell measure on  $\Omega \times \mathbb{R}$ , when a certain absolute continuity condition is satisfied. Later sections discuss various applications, including the likelihood and time-change theorems treated more informally in Sections 7.2 and 7.4, a martingale-type central limit theorem, and the notion of entropy rate which lies behind the discussion in Section 7.6.

## 14.1. Compensators and Martingales

The aim of this section is to introduce the basic ideas of the martingale approach to the study of point processes on the open half-line<sup>1</sup>  $\mathbb{R}_0^+ \equiv (0, \infty)$ . Many of the technicalities are summarized in Appendix 3, so as to allow scope here to stress the connections with other aspects of point process theory. In much of the discussion we are concerned with a random measure on  $\mathbb{R}_0^+$ , for although the case of point processes is of paramount importance, the more general theory can be covered with little extra effort. Thus, general results are stated in terms of the *cumulative process*

$$\xi(t) \equiv \xi(t, \omega) \equiv \xi((0, t], \omega)$$

for some random measure  $\xi(\cdot, \omega)$  (we use  $\xi$  for both, but the abuse of notation should not lead to difficulties). Observe that such processes have trajectories that are a.s. monotonic increasing and right-continuous, as is true in particular for the *counting processes*

$$N(t) \equiv N(t, \omega) \equiv N((0, t], \omega)$$

of a point process  $N(\cdot, \omega)$  on  $\mathbb{R}_+$ . A more important extension, taken up in Section 14.2, is to multivariate and marked point processes. Because the mark may include a spatial location as well as a size or indicator variable, this extension also includes space–time processes. In such cases, it is necessary to consider, not just a single cumulative process, but a family of cumulative processes indexed by the bounded Borel sets of the mark space.

To facilitate the discussion of martingale properties, we suppose throughout the chapter that, unless otherwise stated, the point processes and random measures in view have boundedly finite first moment measures, or finite mean rates in the case of a stationary process.

The ‘information available at time  $t$ ’ is represented mathematically by a  $\sigma$ -algebra  $\mathcal{F}_t$  of sets from the underlying probability space  $(\Omega, \mathcal{E}, \mathcal{P})$ . The

---

<sup>1</sup> We use  $\mathbb{R}_+ = [0, \infty)$  and  $\mathbb{R}_0^+ = (0, \infty)$  to distinguish the closed and open half-lines.

accumulation of information with time is reflected in  $\mathcal{F}_t$  being a member of an increasing family  $\mathcal{F} = \{\mathcal{F}_s: 0 \leq s < \infty\}$  of  $\sigma$ -algebras; that is,  $\mathcal{F}_s \subseteq \mathcal{F}_t$  for  $0 \leq s \leq t < \infty$ .  $\mathcal{F}$  is called a *history* for the process  $\xi$  provided  $\xi(t)$  is  $\mathcal{F}_t$ -measurable for all  $0 \leq t < \infty$  (i.e.,  $\xi$  is  $\mathcal{F}$ -adapted).  $\mathcal{F}_0$  plays the special role of subsuming all information available before observations commence at  $0+$ , and the  $\sigma$ -algebra  $\mathcal{F}_\infty \equiv \bigvee_{t \geq 0} \mathcal{F}_t$  subsumes all information in the history  $\mathcal{F}$ . For the rest of this section ‘the pair  $(\xi, \mathcal{F})$ ’ or ‘the process  $(\xi, \mathcal{F})$ ’ always means a *cumulative process*  $\xi$  and a *history*  $\mathcal{F}$  such that  $\xi$  is  $\mathcal{F}$ -adapted.

The history  $\mathcal{H}$  consisting of the  $\sigma$ -algebras  $\mathcal{H}_t$  generated for each  $t$  by  $\{\xi(s): 0 < s \leq t\}$  plays a special role: we call it the *internal history* (it is also called the *natural* or *minimal* history, reflecting the fact that  $\mathcal{H}$  is the smallest family of nested  $\sigma$ -algebras to which the observed values of  $\xi$  are adapted). Note that  $\mathcal{H}_0 = \{\emptyset, \Omega\}$ . Histories with the particular structure  $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{H}_t$ , with  $\mathcal{F}_0$  in general nontrivial, are called by Brémaud *intrinsic histories*; among other uses, they are important in the analysis of doubly stochastic processes. The special considerations which are associated with a stationary process observed over a finite or infinite past are examined as part of the discussion of complete conditional intensities in Sections 14.3 and 14.7.

A history  $\mathcal{F}$  is called *right-continuous* if  $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s \equiv \mathcal{F}_{t+}$ . Counting processes and other cumulative processes, being right-continuous and boundedly finite by assumption, necessarily yield internal histories that are right-continuous. In general, right-continuity represents a mild constraint on the admissible forms of conditioning information; in any case, whenever the process is adapted to the history  $\mathcal{F}$ , it is adapted also to the right-continuous history  $\mathcal{F}_{(+)} \equiv \{\mathcal{F}_{t+}: 0 \leq t < \infty\}$  (see Exercise 14.1.3).

It is part of our basic framework that the realizations of the cumulative process are a.s. finite for finite  $t$ , for this reflects our assumption that, as random measures, the trajectories are a.s. boundedly finite (elements of  $\mathcal{M}_{\mathbb{R}_+}^\#$ ). This assumption rules out the possibility of explosions, and imposes a certain requirement on the sequence of  $\mathcal{F}$ -stopping times defined for  $n = 0, 1, \dots$  (see Definition A3.3.II and Lemma A3.3.III for background) by

$$\begin{aligned} T_n &\equiv T_n(\omega) = \sup\{t: \xi(t, \omega) < n\} \\ &= \begin{cases} \infty & \text{if } \xi(t, \omega) < n \text{ for all } 0 < t < \infty, \\ \inf\{t: \xi(t, \omega) \geq n\} & \text{otherwise,} \end{cases} \end{aligned} \quad (14.1.1)$$

namely, that  $T_n \rightarrow \infty$  a.s. as  $n \rightarrow \infty$ . Exercise 14.1.5 addresses the case that the sequence is finite; otherwise we suppose the sequence continues indefinitely.

A feature of the general theory of processes is that the family  $\xi(t, \omega)$  is regarded as a single real-valued mapping  $\xi: \mathbb{R}_+ \times \Omega \mapsto \mathbb{R}_+$  rather than as an indexed family of r.v.s. The product  $\sigma$ -algebra  $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{E}$  of sets from the product space  $\mathbb{R}_+ \times \Omega$  contains a hierarchy of important sub- $\sigma$ -algebras, each of which is associated with a corresponding class of processes, namely, the measurable, progressively measurable, and predictable processes; these are defined and discussed briefly in Section A3.3.

In this section we need especially the concept of an  $\mathcal{F}$ -predictable process  $X$ , which is a process measurable with respect to the  $\mathcal{F}$ -predictable  $\sigma$ -algebra, which in turn is the sub- $\sigma$ -algebra of  $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{E}$  generated by all product sets of the form  $(s, t] \times U$  for  $U \in \mathcal{F}_s$  and  $0 \leq s < t < \infty$  (see above Lemma A3.3.I).

The main result of this section is a theorem which asserts that, for every pair  $(\xi, \mathcal{F})$ , where  $\xi$  is  $\mathcal{F}$ -adapted, there exists an integrated form of the conditional intensity function described in Section 7.2 that is predictable, and is the key to a martingale property extending the result of Lemma 7.2.V. In general, what must be subtracted from an increasing process  $\xi$  to yield a martingale is called a *compensator*; it is formally defined as follows.

**Definition 14.1.I.** Let  $\xi(t)$  be an  $\mathcal{F}$ -adapted cumulative process on  $\mathbb{R}_+$ . An  $\mathcal{F}$ -compensator for  $\xi$  is a monotonic nondecreasing right-continuous predictable process  $A(\cdot)$  such that for each  $n$  and  $\mathcal{F}$ -stopping time  $T_n$  at (14.1.1), the stopped process  $\{\xi(t \wedge T_n) - A(t \wedge T_n); 0 \leq t < \infty\}$  is an  $\mathcal{F}$ -martingale.

In passing, note that a process  $\{D(t); 0 \leq t < \infty\}$  such that  $\{D(t \wedge T_n); 0 \leq t < \infty\}$  is a martingale for some sequence of stopping times  $\{T_n\}$  for which  $E|D(t \wedge T_n)| < \infty$  (all  $t \geq 0, n = 1, 2, \dots$ ), is often called a *local martingale*. The notion occurs repeatedly in more general treatments of point processes [e.g., Liptser and Shiryaev (1974, 1977, 1978, 2000)]. See also Exercise 14.1.7.

Example 14.1(a) below, although trivial, illustrates the fact that the compensator is effectively of interest only for processes with jumps: indeed, as subsequent examples illustrate, the compensator may be regarded as a device for smoothing out jumps and producing an a.s. diffuse random measure from a random measure that may have atoms, but has no fixed atoms.

**EXAMPLE 14.1(a) Cumulative process with density.** Suppose  $\xi(\cdot)$  is the cumulative process of an absolutely continuous random measure, with density  $x(t, \omega)$  some  $\mathcal{F}$ -progressively measurable nonnegative process. Then  $\xi(t) = \int_0^t x(u) du$  is its own compensator, for because  $x(\cdot, \omega) \geq 0$ ,  $\xi$  is monotonic nondecreasing and continuous, and for this reason predictable, inasmuch as it is both  $\mathcal{F}$ -adapted and left-continuous.  $\square$

Given a pair  $(\xi, \mathcal{F})$  as in Definition 14.1.I, the first problem is to give conditions that ensure that a compensator for  $\xi$  exists. We start with the simplest example, a one-point process consisting of a single point whose location is defined by a positive r.v.  $X$  with d.f.  $F$  [see Example 7.4(b)]. The associated counting process is defined by

$$N(t, \omega) = I_{(0, t)}(X(\omega)) \quad (0 < t < \infty, \omega \in \Omega). \quad (14.1.2)$$

If we let  $\mathcal{F}$  coincide with the internal history of the process, that is,  $\mathcal{H}_t \in \mathcal{H}$  is the  $\sigma$ -algebra generated by the sets  $\{\{\omega: X(\omega) \leq s\}; 0 < s \leq t\}$ , we can give a direct construction of the  $\mathcal{H}$ -compensator without any need to appeal to the deeper theorems of the general theory of processes.

Observe first that  $N$  is monotonic nondecreasing, right-continuous, and even uniformly bounded so there is no problem about the existence of moments. Next, because  $N(t, \omega) = 1$  implies  $N(t', \omega) = 1$  for all  $t' \geq t$ , the compensator, for the same  $\omega$ , must also be constant for such  $t' \geq t$ . On the other hand, if  $N(t, \omega) = 0$ , then we know that  $X(\omega) > t$ , and thus, in a small interval  $(t, t + dt)$ , we can expect

$$E[dN(t, \omega)] \approx \frac{dF(t)}{1 - F(t)}, \quad (14.1.3)$$

which equals  $h(t) dt$  if the d.f. has a density and hence a hazard function  $h$ .

These heuristics are approximately correct; the key to obtaining a precise statement is the integrated hazard function (IHF) of Definition 4.6.IV.

**Lemma 14.1.II.** *The one-point process  $N$  at (14.1.2) generated by the positive r.v.  $X$  has  $\mathcal{H}$ -compensator*

$$A(t, \omega) = H(t \wedge X(\omega)) \quad (0 < t < \infty, \omega \in \Omega), \quad (14.1.4a)$$

where  $H$  is the IHF of  $X$ ,

$$H(t) = \int_0^t \frac{dF(x)}{1 - F(x-)}. \quad (14.1.4b)$$

The compensator  $A(t)$  so defined is continuous except at jumps  $u_i$  of  $F$ , where in terms of  $\Delta F(u) = F(u) - F(u - 0)$ ,  $A(\cdot)$  has jumps of height

$$a_i = \Delta F(u_i) / [1 - F(u_i -)] \leq 1, \quad (14.1.5)$$

with equality if and only if  $\Delta F(u_i) = 1 - F(u_i -)$ .

PROOF. Note that for each  $\mathcal{H}_t$  the set  $\{\omega: X(\omega) > t\}$  constitutes a large ‘atom’ (i.e., a subset of  $\Omega$  that cannot be decomposed by the  $\sigma$ -algebra). Of course,  $\mathcal{H}_0 = \{\emptyset, \Omega\}$ , whereas  $\mathcal{H}_\infty$  is the  $\sigma$ -algebra generated by the r.v.  $X$ .

Like  $H$  itself (see Definition 4.6.IV),  $H(t \wedge X(\omega))$  is monotonic increasing and right-continuous in  $t$ . To verify that it is predictable, we first check that  $X(t, \omega) \equiv t \wedge X(\omega)$  is predictable, so we study

$$\{(t, \omega): X(t, \omega) > x\} = \{t > x\} \times \{\omega: X(\omega) > x\}.$$

Now  $\{\omega: X(\omega) > x\} \in \mathcal{H}_x$ , so the set in  $(t, \omega)$  has the form of a generating set for the predictable  $\sigma$ -algebra. Thus,  $X(t, \omega)$  is predictable.

The IHF  $H(x)$  is monotonic increasing and right-continuous in  $x$  and thus has a uniquely defined inverse  $H^{-1}$  for which  $H(x) \geq y$  if and only if  $x \geq H^{-1}(y)$ . In particular,  $\{X(t, \omega) \geq H^{-1}(y)\}$  is a predictable set, so  $H(X(t, \omega))$  is a predictable process.

It remains to verify the martingale property that for fixed  $s$  and  $t$  with  $0 \leq s < t$ ,

$$E[N(t \wedge X) - H(t \wedge X) \mid \mathcal{H}_s] = N(s \wedge X) - H(s \wedge X) \quad \text{a.s.} \quad (14.1.6)$$



Note first that, because of the special structure of  $\mathcal{H}_t$  here, we have for any bounded function  $g(\cdot)$ ,

$$\mathbb{E}[g(X) \mid \mathcal{H}_t] = \begin{cases} g(X) & \text{on } \{X(\omega) \leq t\}, \\ \frac{\mathbb{E}[g(X)I_{\{X>t\}}]}{\mathbb{E}(I_{\{X>t\}})} & \text{on } \{X(\omega) > t\}, \end{cases} \quad (14.1.7)$$

because when  $X > t$ ,  $\mathbb{E}[g(X) \mid \mathcal{H}_t] = \mathbb{E}[g(X) \mid X > t]$ , so the second case of (14.1.7) can be written in terms of the d.f.  $F(\cdot)$  of  $X$  as

$$\mathbb{E}[g(X) \mid \mathcal{H}_t] = \frac{1}{1 - F(t)} \int_t^\infty g(u) F(du) \quad \text{on } \{t < X(\omega)\}.$$

On  $\{s \geq X(\omega)\}$ ,  $N(t \wedge X) = N(s \wedge X) = 1$  and  $H(t \wedge X) = H(s \wedge X) = H(X)$ , which for  $X \leq s$  is  $\mathcal{H}_s$ -measurable, so (14.1.6) holds in this case. On the complement where  $\{s < X(\omega)\}$ , using (14.1.7),

$$\mathbb{E}[N(t \wedge X) \mid \mathcal{H}_s] = \frac{1}{1 - F(s)} \int_s^t F(du) = \frac{F(t) - F(s)}{1 - F(s)},$$

so that from Lemma 4.6.I we obtain

$$\begin{aligned} & [1 - F(s)][\mathbb{E}(H(t \wedge X) \mid \mathcal{H}_s) - H(s)] \\ &= \int_s^t [H(u) - H(s)] F(du) + [H(t) - H(s)][1 - F(t)] = F(t) - F(s). \end{aligned}$$

Thus,  $\mathbb{E}[N(t \wedge X) - H(t \wedge X) \mid \mathcal{H}_s] = -H(s)$  on  $\{s < X(\omega)\}$ , so (14.1.6) holds generally as asserted.

It is a standard property of the distribution function  $F$  of an honest positive r.v. that  $F(x) = F_a(x) + F_c(x)$  for a purely atomic function  $F_a$  and a continuous function  $F_c$  (see above Lemma A1.6.II). Similarly, the IHF  $H$  of  $F$ , being a monotonic nondecreasing function, can be decomposed as  $H(x) = H_a(x) + H_c(x)$ , with

$$H_a(x) + H_c(x) = \int_0^x \frac{dF_a(u) + dF_c(u)}{1 - F(u-)} = \sum_{i: u_i \leq x} a_i + \int_0^x \frac{dF_c(u)}{1 - F(u-)},$$

where on the right-hand side the sum is atomic and the integral is a continuous function of  $x$ . The asserted nature of  $A(\cdot)$  follows.  $\square$

**EXAMPLE 14.1(b)** *One-point process with absolutely continuous or discontinuous  $\mathcal{H}$ -compensator.* Suppose first  $X$  above has an exponential distribution, say  $F(x) = 1 - e^{-\lambda x}$ , so that its IHF  $H(t) = \lambda t$ . Then the corresponding one-point process has  $A(t) = \lambda \min(t, X)$  which is differentiable except at  $X$ , and in any case is absolutely continuous with density  $\lambda^*(t) = \lambda$  ( $t \leq X$ ),  $= 0$  ( $t > X$ ).

Now suppose that  $X = 1$  with probability  $p$ , and with probability  $1 - p$  is exponential as above, so that  $X$  has survivor function

$$S(t) = \begin{cases} p + (1 - p)e^{-\lambda t} & (t < 1), \\ (1 - p)e^{-\lambda t} & (t \geq 1). \end{cases}$$

Then from (4.6.4)  $X$  has IHF

$$H(t) = \begin{cases} \lambda t - \log [1 + p(e^{\lambda t} - 1)] & (t < 1), \\ \lambda t & (t \geq 1), \end{cases}$$

with a jump of size  $\Delta H = \log [1 + p(e^\lambda - 1)]$  at  $t = 1$ . The compensator is given by  $A(t) = H(t \wedge X)$ , and thus has a discontinuity at  $x = 1$  if  $X \geq 1$ . The risk is reduced in the interval  $(0, 1)$  because there is a positive probability that the event will occur at the end of the interval rather than randomly during the interval.

Typically, discontinuities in the compensator are associated with the occurrence of deterministic elements such as fixed atoms as around (14.1.5). See Exercise 14.1.10 for further examples.  $\square$

An important extension of Example 14.1(b) can be given when  $\mathcal{F}$  is an intrinsic history for the one-point process and so consists of  $\sigma$ -algebras of the form  $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{H}_t$ . At least in the case that  $X$  has a regular conditional probability distribution given  $\mathcal{F}_0$ , a version of which we denote  $F(\cdot | \mathcal{F}_0)$ , the influence of  $\mathcal{F}_0$  can be described very simply: all we need to do in Lemma 14.1.II is to replace the distribution of  $X$  and its IHF by this conditional distribution  $F(\cdot | \mathcal{F}_0)$  and its associated IHF,

$$H(t | \mathcal{F}_0) = \int_0^t \frac{dF(u | \mathcal{F}_0)}{1 - F(u- | \mathcal{F}_0)}.$$

**Lemma 14.1.III.** *A one-point process with prior  $\sigma$ -algebra  $\mathcal{F}_0$  and regular conditional distribution  $F(\cdot | \mathcal{F}_0)$  for  $X$  has compensator  $H(t | \mathcal{F}_0)$  relative to the intrinsic history  $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{H}_t$ .*

PROOF. Note first that because  $\mathcal{F}_t \supseteq \mathcal{F}_0$ ,  $E(\cdot | \mathcal{F}_0) = E[E(\cdot | \mathcal{F}_t) | \mathcal{F}_0]$ . We now claim that for nonnegative measurable functions  $g: \mathbb{R} \mapsto \mathbb{R}$ ,

$$E[g(X) | \mathcal{F}_t] = \begin{cases} g(X) & \text{on } \{X(\omega) \leq t\}, \\ \frac{\int_t^\infty g(u) F(du | \mathcal{F}_0)}{1 - F(t | \mathcal{F}_0)} & \text{on } \{X(\omega) > t\}. \end{cases} \quad (14.1.8)$$

The first part of (14.1.8) is obvious, while on  $\{X(\omega) > t\}$ ,  $\mathcal{F}_t$  consists entirely of sets of the form  $U \cap \{X(\omega) > t\}$  for some  $U \in \mathcal{F}_0$ . In this case we can write

$$\int_{U \cap \{X > t\}} g(X(\omega)) \mathcal{P}(d\omega) = \int_U I_{\{X > t\}} g(X(\omega)) \mathcal{P}(d\omega)$$

$$\begin{aligned}
&= \int_U \mathbb{E}[g(X)I_{\{X>t\}} \mid \mathcal{F}_0] \mathcal{P}(d\omega) \\
&= \int_U \frac{I_{\{X>t\}}}{\mathbb{E}(I_{\{X>t\}})} \mathbb{E}[g(X)I_{\{X>t\}} \mid \mathcal{F}_0] \mathcal{P}(d\omega) \\
&= \int_U \frac{\mathbb{E}[g(X)I_{\{X>t\}} \mid \mathcal{F}_0]}{\mathbb{E}(I_{\{X>t\}} \mid \mathcal{F}_0)} \mathcal{P}(d\omega). \tag{14.1.9}
\end{aligned}$$

The first expression in this chain reduces to the left-hand side of (14.1.8) and, from the assumption that  $F(\cdot \mid \mathcal{F}_0)$  is a version of the regular conditional distribution, the last expression reduces to the right-hand side of (14.1.8) on  $\{X(\omega) > t\}$  as asserted.

This result can now be used in place of (14.1.6) to establish the compensator property of the conditional IHF, provided at least that we can manipulate the conditional distributions in the same way as unconditional distributions: this is certainly the case when we can choose a regular version of the conditional distribution.  $\square$

EXAMPLE 14.1(c) *One-point process with randomized hazard function.* To take a specific example, suppose that  $X$  has a negative exponential distribution with parameter  $\lambda$ , where  $\lambda$  itself is a positive r.v. determined by  $\mathcal{F}_0$  (i.e.,  $\lambda$  is  $\mathcal{F}_0$ -measurable). Then the  $\mathcal{F}$ -compensator,  $A^{\mathcal{F}}(t, \omega)$  say, can be represented in terms of the IHF of the exponential ( $\lambda$ ) distribution, namely,

$$A^{\mathcal{F}}(t, \omega) = \lambda(t \wedge X(\omega)).$$

On the other hand, to find the  $\mathcal{H}$ -compensator we must first evaluate the survivor function for the resultant mixed exponential distribution. If, for example,  $\lambda$  itself has a unit exponential distribution with density  $e^{-\lambda} d\lambda$ , then the unconditional survivor function is

$$H(t) = \mathbb{E}[H(t) \mid \mathcal{F}_0] = \int_0^\infty e^{-\lambda t} e^{-\lambda} d\lambda = \frac{1}{1+t}.$$

The IHF is therefore equal to  $\log(1+t)$ , and for the  $\mathcal{H}$ -compensator we obtain

$$A^{\mathcal{H}}(t, \omega) = \log(1 + t \wedge X(\omega)).$$

Such examples show that the choice of prior  $\sigma$ -algebra can drastically affect the form of the compensator.  $\square$

We can now construct the compensator for a simple point process with respect to the *intrinsic history*  $\mathcal{F} = \{\mathcal{F}_0 \vee \mathcal{H}_t: 0 < t < \infty\}$ ; that is, we allow some initial conditioning as in the last example. Such a history  $\mathcal{F}$  is completely described by the initial  $\sigma$ -algebra  $\mathcal{F}_0$  and the family of stopping times  $\{T_n\}$  as at (14.1.1): in view of the assumed simplicity,  $\{T_n\}$  is a.s. a strictly increasing sequence. Given

$$\mathcal{F}_{(n-1)} \equiv \mathcal{F}_{T_{n-1}},$$

which means we are given  $\mathcal{F}_0$  and  $T_1, \dots, T_{n-1}$ , choose a family of regular conditional distributions  $G_n(x \mid \mathcal{F}_{(n-1)})$  for the distributions of the successive differences

$$\tau_n = T_n - T_{n-1} \quad (n = 1, 2, \dots, \quad T_0 \equiv 0).$$

Writing

$$N(t) = \sum_{n=1}^{\infty} [N(t \wedge T_n) - N(t \wedge T_{n-1})] = \sum_{n=1}^{\infty} N^{(n)}(t) \quad \text{say,}$$

each  $N^{(n)}(\cdot)$  is a one-point process with a single point of increase at  $T_n$ . Defining now the IHFs  $H_n(\cdot) \equiv H_n(\cdot \mid \mathcal{F}_{(n-1)})$  from the conditional d.f.s  $G_n(\cdot \mid \mathcal{F}_{(n-1)})$  by

$$H_n(x \mid \mathcal{F}_{(n-1)}) = \int_0^x \frac{G_n(du \mid \mathcal{F}_{(n-1)})}{1 - G_n(u- \mid \mathcal{F}_{(n-1)})},$$

we assert that the  $\mathcal{F}$ -compensator for  $N^{(n)}(\cdot)$  has the form

$$A^{(n)}(t, \omega) = \begin{cases} 0 & \text{on } t < T_{n-1}(\omega), \\ H_n(t - T_{n-1}) & \text{on } T_{n-1}(\omega) \leq t < T_n(\omega), \\ H_n(T_n - T_{n-1}) & \text{on } T_n(\omega) \leq t. \end{cases} \quad (14.1.10)$$

Then by additivity,  $N(\cdot)$  has the  $\mathcal{F}$ -compensator

$$A(t, \omega) = \sum_{n=1}^{\infty} A^{(n)}(t, \omega).$$

To establish (14.1.10), note that predictability of  $A^{(n)}(\cdot)$  is established as in Lemma 14.1.II, so it remains to show that each difference  $Z^{(n)}(t, \omega) \equiv N^{(n)}(t, \omega) - A^{(n)}(t, \omega)$  is an  $\mathcal{F}$ -martingale. We establish the requisite equality

$$\mathbb{E}[Z^{(n)}(t) \mid \mathcal{F}_s] = Z^{(n)}(s) \quad (14.1.11)$$

for  $0 < s \leq t$  separately on the sets  $B_n = \{\omega: T_{n-1} \leq s\}$  and  $B_n^c$ , observing that  $B_n$  and  $B_n^c \in \mathcal{F}_{(n-1)}$ .

Considering first the subsets of  $B_n$ , we have

$$\mathcal{F}_s \cap \{T_{n-1}(\omega) \leq s < T_n(\omega)\} = \mathcal{F}_{(n-1)} \cap \{T_{n-1}(\omega) \leq s < T_n(\omega)\}, \quad (14.1.12)$$

which means that, given any  $C \in \mathcal{F}_s$ , there exists  $C' \in \mathcal{F}_{(n-1)}$  such that  $C \cap B_n = C' \cap B_n$  and conversely: that this is so is clear from the structure of the  $\sigma$ -algebra  $\mathcal{F}_s$  (because  $\mathcal{F}_s \supset \mathcal{H}_s$ ) and a consideration of the basic sets such as  $\{\omega: N(s, \omega) = k\}$ . Now on  $B_n$ , the stopping time  $\tau_n$  plays the same role for  $N^{(n)}(\cdot)$  as  $X$  plays for the one-point process of Example 14.1(b), with

$\mathcal{F}_{(n-1)}$  here playing the role of  $\mathcal{F}_0$  there. In particular, on  $B_n$  we have for any bounded measurable function  $f(\cdot)$

$$\mathbb{E}[f(\tau_n) \mid \mathcal{F}_s] = \frac{\mathbb{E}[f(\tau_n)I_{\{\tau_n > x_n\}} \mid \mathcal{F}_{(n-1)}]}{\mathbb{E}[I_{\{\tau_n > x_n\}} \mid \mathcal{F}_{(n-1)}]} = \int_{x_n}^{\infty} \frac{f(u) G_n(du \mid \mathcal{F}_{(n-1)})}{1 - G_n(x_n \mid \mathcal{F}_{(n-1)})},$$

where  $x_n = s - T_{n-1}$ , necessarily  $\geq 0$  on  $B_n$ . In principle, this evaluation of the conditional expectation involves the extension of (14.1.8) to the case where  $t$  there (equals  $x_n$  here) is a r.v. measurable with respect to the prior  $\sigma$ -algebra  $\mathcal{F}_0$  there (which is  $\mathcal{F}_{(n-1)}$  here). However, scrutiny of (14.1.9) and the surrounding argument shows that nothing need be altered there, with (14.1.8) remaining  $\mathcal{F}_0$ -measurable, so (14.1.9) is still valid. Thus, on  $B_n$ , proof of the martingale equality (14.1.11) follows as in Lemma 14.1.III.

On the sets  $\{s < t < T_{n-1}(\omega)\}$  and  $\{s \geq T_n(\omega)\}$  the equality is trivial because all terms are zero. There remains the case  $\{s < T_{n-1}(\omega) \leq t\}$ . Here we proceed by conditioning first on  $\mathcal{F}_{(n-1)}$ , when equality follows as a special case of the above, because this equality is not affected by further conditioning on  $\mathcal{F}_s$ . We summarize this discussion as follows.

**Theorem 14.1.IV.** *Let  $N$  be the counting process of a simple point process on  $(0, \infty)$ ,  $\mathcal{F}$  a history for  $N$  of the form  $\{\mathcal{F}_0 \vee \mathcal{H}_t\}$ , and  $\{T_n\}$  the sequence of stopping times at (14.1.1). Suppose there exist regular versions  $G_n(\cdot \mid \mathcal{F}_{(n-1)})$  of the conditional d.f.s of the intervals  $\tau_n = T_n - T_{n-1}$ , given  $\mathcal{F}_{(n-1)}$ , such that  $1 - G_n(x-) > 0$  for  $x > 0$ . Then a version of the  $\mathcal{F}$ -compensator for  $N$  is given by*

$$A(t, \omega) = \sum_{n=1}^{\infty} A^{(n)}(t, \omega), \quad (14.1.13a)$$

where

$$A^{(n)}(t, \omega) = \begin{cases} 0 & (t \leq T_{n-1}(\omega)), \\ \int_0^{\tau_n \vee (t - T_{n-1})+} \frac{G_n(du \mid \mathcal{F}_{(n-1)})}{1 - G_n(u- \mid \mathcal{F}_{(n-1)})} & (t > T_{n-1}(\omega)). \end{cases} \quad (14.1.13b)$$

The following special case ties in the result above with the earlier discussion of Section 7.2, in particular with Proposition 7.2.I and Definition 7.2.II.

**Corollary 14.1.V.** *The  $\mathcal{F}$ -compensator  $A(\cdot)$  at (14.1.13) is absolutely continuous a.s. if and only if the conditional d.f.s  $G_n(\cdot \mid \mathcal{F}_{(n-1)})$  have absolutely continuous versions, with densities  $g_n(\cdot \mid \mathcal{F}_{(n-1)})$  say, in which case one version of the  $\mathcal{F}$ -compensator is given by*

$$A(t, \omega) = \int_0^t \lambda^*(u, \omega) du,$$

where

$$\lambda^*(t, \omega) = \sum_{n=1}^{\infty} \lambda_n^*(t, \omega) \equiv \sum_{n=1}^{\infty} \frac{g_n(t \wedge T_n - T_{n-1} \mid \mathcal{F}_{(n-1)}) I_{\{T_{n-1} < t \leq T_n\}}}{1 - G_n(\tau_n \vee (t - T_{n-1})_+ - \mid \mathcal{F}_{(n-1)})}.$$

In particular, the compensator has this form when  $N$  is regular; in this case  $\lambda^*(t, \omega)$  defined above coincides a.s. and  $t$ -a.e. with the function denoted  $\lambda^*(\cdot)$  in Definition 7.2.II.

The construction also yields the following important result.

**Proposition 14.1.VI.** *Under the conditions of Theorem 14.1.IV, the point process defines the compensator uniquely, and, conversely, the compensator uniquely defines the process.*

PROOF. It is clear from the construction that the compensator is uniquely defined by the process. Within the class of processes referred to in Theorem 14.1.IV, the converse is also true, for, step by step, the compensator determines the functions  $G_n$  and hence ultimately the full set of fidi distributions for the process. Some further details are given in Exercises 14.1.13–14.  $\square$

As noted in Proposition 7.2.IV, this result has the corollary that if the compensator is absolutely continuous, then also the conditional intensity function characterizes the point process uniquely.

As an example to illustrate how the reconstruction of the point process from its compensator proceeds, consider Watanabe's characterization of the Poisson process, the first and still one of the most striking applications of martingale ideas to point processes. Watanabe's proof used martingale calculus ideas of the type that we develop further in ensuing sections. Here we treat it as a simple special case of Proposition 14.1.VI and its corollary noted above.

EXAMPLE 14.1(d) *Watanabe's theorem* (Watanabe, 1964). Suppose that the conditional intensity is a constant,  $\lambda^*(t) \equiv \lambda$ , so  $A(t, \omega) = \lambda t$ . Then there is no dependence on the past, and an induction argument shows that all the conditional d.f.s  $G_n(u \mid \mathcal{F}_{(n-1)})$  of (14.1.13) have the same form  $1 - e^{-\lambda u}$ , and that the random variables to which they relate are mutually independent. Thus successive intervals are independently and exponentially distributed, and the process must be Poisson.  $\square$

Much of the preceding discussion can be extended in a straightforward manner to MPPs, but is deferred to the more extended discussion in the next section. For completely general histories it is not possible to provide a representation for the  $\mathcal{F}$ -compensator analogous to that of Theorem 14.1.IV, even in the point process case: there are too many different ways in which the conditioning information can affect the duration of the interval lengths. In specific situations, however, when the history can be identified with the internal history for some larger process of which the point process forms part, it may not be too difficult to identify appropriate distributional forms.

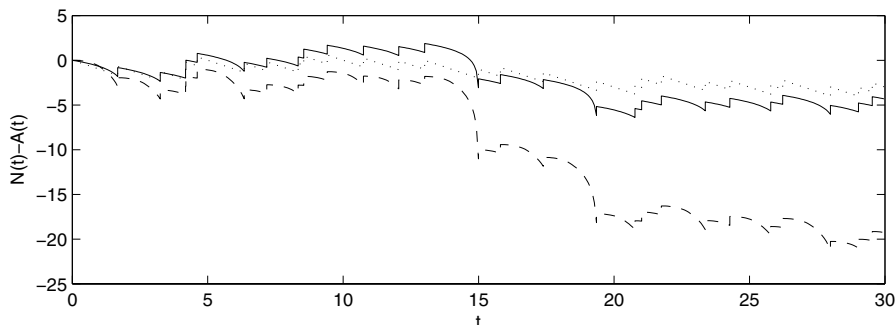


Figure 14.1. Martingale  $N(t) - A(t)$  for compensator (—) of renewal process with uniform  $(0,2)$  lifetimes, with incorrect ‘martingales’ from a unit rate Poisson process (·····) and triangular  $(0,2)$  lifetimes (---).

EXAMPLE 14.1(e) *Simple and modulated renewal process* [continued from Example 7.2(e) and Exercise 7.2.11(c)]. For a simple renewal process with absolutely continuous lifetime distribution and hence a hazard function  $h(\cdot)$  say, we know already that the conditional intensity has the form  $h(t - T_{N(t)})$ . Evidently, the corresponding compensator has the form

$$A(t) = \sum_{n=1}^{N(t)} H(T_n - T_{n-1}) + H(t - T_{N(t)}), \quad (14.1.14)$$

where  $H(\cdot)$  is the IHF corresponding to  $h$  (for  $t < T_1$ ,  $N(t) = 0$  and the summation term vanishes). From Theorem 14.1.IV it is easy to see that this holds for a general renewal process whose lifetime r.v.s are positive a.s.

In Figure 14.1 we have plotted (as the continuous curve) the martingale  $M(t) = N(t) - A(t)$  that results from a realization of a renewal process on  $(0, 30)$  in which the lifetimes are uniformly distributed on  $(0, 2)$  so they have mean 1 and the IHF  $H(u) = -\log(1 - \frac{1}{2}u)$  ( $0 < u < 2$ ).

Two further ‘martingales’ are plotted in Figure 14.1 to show the effect of using (14.1.14) with incorrect compensators as a result of making an incorrect assumption as to the underlying lifetime distribution. The dashed curve results from using the IHF  $H(u) = u$  as would hold for a unit rate Poisson process, and the dashed curve from using  $H(u) = -\log(1 - \frac{1}{2}t^2)$  for  $0 < t \leq 1$ ,  $= \log[2/(2 - t)^2]$  for  $1 \leq t < 2$ , as holds for a renewal process whose lifetimes have the triangular density  $1 - |t - 1|$  on  $(0, 2)$ . See Exercise 14.1.11.

Suppose next that, as in Example 7.2(e), we also observe a family (vector) of stochastic processes

$$\{X(t): 0 < t < \infty\} \equiv \{X_1(t), \dots, X_k(t): 0 < t < \infty\},$$

and identify the history  $\mathcal{F}$  with the joint history generated by the  $\sigma$ -algebras

$$\mathcal{F}_t = \mathcal{H}_t^N \vee \mathcal{H}_t^X,$$

where on the right-hand side there are denoted  $\sigma$ -algebras of the internal histories of  $\{N(t): 0 < t < \infty\}$  and  $\{X(t): 0 < t < \infty\}$ . Returning to the

absolutely continuous case for ease of exposition, suppose that the hazard function in successive intervals is modified in a multiplicative fashion by some nonnegative function  $\psi(X_1(\cdot), \dots, X_k(\cdot))$ ; that is, we take

$$\lambda^*(t) = h(t - T_{N(t)}) \psi(X_1(t), \dots, X_k(t)).$$

In this set-up, the  $\mathcal{F}$ -compensator would be found by integrating  $\lambda^*(t)$  over successive intervals. As a very particular illustration, suppose  $k = 1$ , that  $\psi(\cdot)$  is the step function  $\psi(x) = 2$  if  $x > 0$  and  $\psi(x) = 1$  otherwise, and that  $h(x) = \lambda$ , corresponding to exponential interarrival times in the absence of the modulating factor. Then for any measurable  $X(\cdot)$  we would have

$$A(t) = \lambda(t + Y_t),$$

where the random variable  $Y_t$  is the length of time for which  $X(s) > 0$  during the interval  $(0 < s < t)$ . This assumes that the process  $X(t)$  is observable; when this is not the case a filtering problem is involved, requiring averaging of the  $\mathcal{F}$ -intensity over the coarser  $\sigma$ -algebras of the internal history, as discussed further in Sections 14.3 and 14.4.  $\square$

For more general processes  $\xi(\cdot)$  and histories  $\mathcal{F}$ , even if we cannot establish explicit representations, many important results can still be derived from the Doob–Meyer decomposition, as, for example, below where we show both the existence and uniqueness of the compensator for a cumulative process  $\xi$  with general history  $\mathcal{F}$  and in discussing quadratic variation..

**Theorem 14.1.VII.** *Let  $\{\xi(t): t > 0\}$  be a cumulative process adapted to the right-continuous history  $\mathcal{F}$ . Then  $\xi(\cdot)$  admits an  $\mathcal{F}$ -compensator  $A(t, \omega)$ , which is uniquely defined  $\mathcal{P}$ -a.e. in the sense that for any other compensator  $\tilde{A}(t, \omega)$ ,  $\mathcal{P}\{\xi(t, \omega) = \tilde{A}(t, \omega) \text{ (all } t)\} = 1$ .*

PROOF. We again use the stopped process

$$\xi_n(t) = \xi(t \wedge T_n),$$

where the stopping times  $\{T_n\}$  are as at (14.1.1). Because  $\xi_n(t, \omega) \leq n$ , each  $\xi_n(\cdot)$  is uniformly bounded in  $(t, \omega)$ , and also has bounded first moment, so that in addition it is uniformly integrable in  $t$ . Also, each  $\xi_n(t)$  has its trajectory a.s. nondecreasing in  $t$ , so for  $0 < s < t$ ,

$$\mathbb{E}[\xi_n(t) \mid \mathcal{F}_s] = \mathbb{E}[\xi_n(t) - \xi_n(s) \mid \mathcal{F}_s] + \mathbb{E}[\xi_n(s) \mid \mathcal{F}_s] \geq \xi_n(s) \quad \text{a.s. on } \mathcal{F}_s.$$

Thus,  $\{\xi_n(t): 0 < t < \infty\}$  is a right-continuous, bounded submartingale with respect to the history  $\mathcal{F}$ , and the Doob–Meyer decomposition (Theorem A3.4.IX) can be applied. It implies that there exists a right-continuous non-decreasing  $\mathcal{F}_{(n)}$ -predictable process  $A_n(\cdot)$  and an  $\mathcal{F}$ -martingale  $M_n(\cdot)$  such that

$$\xi_n(t) = A_n(t) + M_n(t).$$



Moreover, the processes  $A_n$ ,  $M_n$  are uniquely defined  $\mathcal{P}$ -a.s. as functions on  $(0, \infty)$ , which implies that the functions  $\{A_n(t)\}$  are a.s. nested in the sense that for  $m > n$ ,

$$A_n(t, \omega) \leq A_m(t, \omega) \quad \text{a.s. for } t \leq T_n(\omega).$$

Now the definition of a cumulative process in terms of a boundedly finite random measure requires  $T_n \rightarrow \infty$  a.s.; letting  $n \rightarrow \infty$  it follows that, a.s. for all  $t$ , a well-defined limit  $A(t, \omega)$  exists and defines an  $\mathcal{F}$ -adapted process such that for every  $n > 0$ ,

$$A(t, \omega) = A_n(t, \omega) \quad (t \leq T_n(\omega)).$$

Clearly,  $A(t, \omega)$  inherits the monotonicity and right-continuity properties from each member of the sequence  $\{A_n(t, \omega)\}$ .

For predictability, observe that because  $T_n \rightarrow \infty$ , the left-continuous  $\mathcal{F}_{(n)}$ -adapted processes generate left-continuous  $\mathcal{F}$ -adapted processes, so  $A(t, \omega)$ , which is  $\mathcal{F}_{(n)}$ -predictable for each  $n$ , is also  $\mathcal{F}$ -predictable. Finally, uniqueness of the overall decomposition  $\xi(t) = A(t) + M(t)$  follows from the uniqueness of the Doob–Meyer decomposition on each of the sets  $t \leq T_n$ .  $\square$

The results do not address directly the question of which predictable cumulative processes can be compensators, nor which nonnegative predictable processes can be conditional intensities. To make sense of this question we suppose first that the processes are defined on the canonical space  $\mathcal{X} \times \mathcal{N}_{\mathcal{X}}^{\#}$ , with  $\mathcal{X} = \mathbb{R}_+$ , where a predictable cumulative process  $A(t, N)$  takes the form of a function  $\Psi\{N(t), T_1(N), \dots, T_{N(t)}(N)\}$  of the points of the realization occurring before time  $t$ . It is then a matter of finding conditions on the function  $\Psi$  that allow a consistent set of integrated hazard functions to be defined from it. These conditions must ensure that the hazard functions satisfy the requirements set out in Exercise 14.1.8. A brief outline of the argument is sketched in Exercises 14.1.13–14; a thorough discussion, incorporating also the marked case, is given by Last and Brandt (1995, Chapter 8).

Another important application of the Doob–Meyer decomposition is in proving the existence of the *quadratic variation* of a martingale when the martingale itself is square integrable. Let  $\xi(t)$  be an  $\mathcal{F}$ -adapted cumulative process on  $\mathbb{R}_+$  with finite second moments

$$\mathbb{E}([\xi(t)]^2) < \infty \quad (0 < t < \infty),$$

write  $A(t)$  for its  $\mathcal{F}$ -compensator (which then necessarily exists), and  $M(t)$  for the  $\mathcal{F}$ -martingale  $\xi(t) - A(t)$ . Then  $[M(t)]^2$  is again an  $\mathcal{F}$ -adapted process whose expected increments are nonnegative because

$$\begin{aligned} \mathbb{E}[(M(t))^2 - (M(s))^2 \mid \mathcal{F}_s] \\ &= \mathbb{E}[(M(t) - M(s))^2 \mid \mathcal{F}_s] + 2\mathbb{E}[(M(t) - M(s))M(s) \mid \mathcal{F}_s] \quad (14.1.15) \\ &= \mathbb{E}[(M(t) - M(s))^2 \mid \mathcal{F}_s], \end{aligned}$$

using the fact that  $M(\cdot)$  is an  $\mathcal{F}$ -martingale. Thus  $[M(t)]^2$  is an  $\mathcal{F}$ -sub-martingale, which therefore has a Doob–Meyer decomposition, say

$$[M(t)]^2 = Q(t) + M_2(t), \quad (14.1.16)$$

where  $M_2(t)$  is the  $\mathcal{F}$ -martingale component, and the  $\mathcal{F}$ -compensator  $Q(t)$  is called the *quadratic variation process*. The name stems from the fact, as follows from (14.1.16) on taking expectations and using the martingale property of  $M_2$ , that

$$\begin{aligned} \mathbb{E}(Q(t) - Q(s) \mid \mathcal{F}_s) &= \mathbb{E}([M(t)]^2 - [M(s)]^2 \mid \mathcal{F}_s) - \mathbb{E}[M_2(t) - M_2(s) \mid \mathcal{F}_s] \\ &= \mathbb{E}([M(t) - M(s)]^2 \mid \mathcal{F}_s) \end{aligned} \quad (14.1.17)$$

(cf. the last equation of A3.4, where, however, the conditional expectation on the left-hand side has been omitted). But  $\mathbb{E}(M(t) - M(s) \mid \mathcal{F}_s) = 0$ , so (14.1.17) shows that the increments in  $Q$  are the conditional variances of the increments in the martingale  $M$ , and hence the terminology.

The right-hand side of (14.1.17) also provides one approach to evaluating  $Q$ : write the argument as

$$[M(t) - M(s)]^2 = \int_s^t \int_s^t M(du) M(dv) = \int_{(s,t] \times (s,t]} (M \times M)(du \times dv),$$

and consider the integral on the three regions  $D_1 = \{s < u < v \leq t\}$ ,  $D_2 = \{s < v < u \leq t\}$ , and  $D_3 = \{s < u = v \leq t\}$ . On  $D_1$  the martingale property implies that

$$\begin{aligned} \mathbb{E}[M(du) M(dv) \mid \mathcal{F}_s] &= \mathbb{E}(\mathbb{E}[M(du) M(dv) \mid \mathcal{F}_u] \mid \mathcal{F}_s) \\ &= \mathbb{E}(M(du) \mathbb{E}[M(dv) \mid \mathcal{F}_u] \mid \mathcal{F}_s) = 0, \end{aligned}$$

and similarly for the integral over  $D_2$ . This leaves only the conditional expectation of the integral over  $D_3$ , hence

$$\mathbb{E}(Q(t) - Q(s) \mid \mathcal{F}_s) = \mathbb{E} \left[ \int_{s < u=v \leq t} (M \times M)(du \times dv) \mid \mathcal{F}_s \right]. \quad (14.1.18)$$

The diagonal component of the product measure  $M \times M$  is zero except where  $M$  itself has atoms, namely, at jumps of either  $N$  or its compensator  $A$  (or both),  $\Delta N$  and  $\Delta A$  say. The case of a simple point process with continuous compensator is particularly simple, for here the diagonal component is zero except at the jumps of  $N$ , for which  $(\Delta N)^2 = \Delta N$ . In this case we obtain

$$\mathbb{E}(Q(t) - Q(s) \mid \mathcal{F}_s) = \mathbb{E} \left[ \int_s^t dN(u) \mid \mathcal{F}_s \right] = \mathbb{E}(A(t) - A(s) \mid \mathcal{F}_s),$$

this last step following from the martingale property of  $N - A$ . It now follows from the uniqueness of the Doob–Meyer decomposition, and the fact that both  $Q$  and  $A$  are monotonic increasing and predictable, that  $Q = A$  a.s.

**Proposition 14.1.VIII.** *Let  $N$  be a simple point process with internal history  $\mathcal{H}$ , continuous  $\mathcal{H}$ -compensator  $A(\cdot)$ , and  $\mathcal{H}$ -quadratic variation  $Q(\cdot)$ . Then under the conditions of Theorem 14.1.IV the processes  $Q(t)$  and  $A(t)$  coincide a.s.; that is,  $\mathcal{P}\{Q(t) = A(t) \text{ (all } t)\} = 1$ .*

This rather surprising result may be regarded as a further illustration of the locally Poisson character of a process with a conditional intensity, because it reflects the property that the mean and the variance of a Poisson distribution are equal. Then  $\sqrt{\lambda^*(t)}$  plays the role of a local standard deviation. Exercise 14.1.17 gives a useful application.

The proposition extends easily to multivariate point processes, as follows.

**EXAMPLE 14.1(f)** *Quadratic variation for simple and multivariate point processes.* Let  $N$  be a simple point process with jump-points  $\{t_i\}$  and  $\mathcal{F}$ -intensity  $\lambda^*(t)$  on  $0 < t < \infty$ . Then the martingale  $M(t) = N(t) - A(t)$  is the difference of a pure jump process  $N$  and a continuous process  $A$ , and from Proposition 14.1.VIII,

$$Q(t) = A(t) \quad \text{a.s.}$$

Thus both the compensator and the quadratic variation of the point process are continuous, and they reduce to the same process.

For an  $\mathcal{F}$ -adapted multivariate point process  $N_j(t)$  ( $j = 1, \dots, J$ ), the discussion can be extended to the cross terms  $Q_{jk}(dt) = E[M_j(dt) M_k(dt)]$  ( $j \neq k$ ). Provided the ground process is simple and  $N_j$  has conditional intensity  $\lambda_j^*$  say, the probability that both components have a point in the same interval of length  $dt$  is  $O[(dt)^2]$ , so that, writing the conditional expectation  $E([M_j(t) - M_j(s)][M_k(t) - M_k(s)] \mid \mathcal{F}_s)$  as the sum of integrals over the regions  $D_1$ ,  $D_2$ , and  $D_3$  as below (14.1.17), all three terms now vanish. Thus the quadratic variation matrix equals  $\text{diag}(A_1(\cdot), \dots, A_k(\cdot))$  in which the  $j$ th term is the compensator for  $N_j$  and has conditional  $\mathcal{F}$ -intensity  $\lambda_j^*(\cdot)$ .  $\square$

When the compensator for  $N$  is not continuous, it is more difficult to evaluate the expectation in (14.1.18) (see Exercise 14.1.16 for a simple example). An alternative, to which we now turn, is to imitate the direct analysis that led to the form of the one-point process established in Lemma 14.1.II [Elliott's (1976) derivation of Lemma 14.1.IX uses properties of orthogonal square integrable martingales, as also in the book form of Elliott (1982, Section 15.1)]. The quadratic variation for general point processes then follows from a sequence of extensions analogous to those used in establishing the general form of the point process compensator in Theorem 14.1.IV from Lemma 14.1.III. For further discussions see Elliott or Karatzas and Shreve (1988).

For a one-point process, as in general, the quadratic variation  $Q$  can be written as the sum  $Q_c + Q_a$  of continuous and atomic components, where  $Q_c = A_c$  and  $Q_a \ll A_a$ ; specifically, in the notation of Lemma 14.1.II,

$$Q_a(t) = \sum_{u_i \leq \min(t, X)} (1 - a_i) a_i < \sum_{u_i \leq \min(t, X)} a_i = A_a(t). \quad (14.1.19)$$

**Lemma 14.1.IX.** *Let the  $\mathcal{H}$ -martingale  $M$  for the one-point process  $N$  at (14.1.2) be square integrable. Then  $N$  has  $\mathcal{H}$ -predictable quadratic variation*

$$Q(t) = H(t \wedge X) - H_2(t \wedge X), \quad (14.1.20)$$

where  $H_2(t) = \sum_{u_i \leq t} a_i^2$  and  $H(\cdot)$ ,  $a_i$ , and  $u_i$  are as in (14.1.4b) and (14.1.5).

PROOF. Using the notation of Lemma 14.1.II we have

$$N(t, \omega) = H(\min(t, X(\omega))) + M(t, \omega),$$

where  $H$  is the IHF given by (14.1.4b) and  $M$  is the  $\mathcal{H}$ -martingale

$$M(t, \omega) = \begin{cases} -H(t) & \text{if } X(\omega) > t, \\ 1 - H(X) & \text{if } X(\omega) \leq t, \end{cases}$$

with  $\mathcal{H}$  the intrinsic history determined by the positive r.v.  $X$ . Recall also from around (4.6.4) that the IHF  $H$  (denoted  $Q$  there) satisfies the equation

$$F(t) = \int_0^t [1 - F(u-)] dH(u) = [1 - F(t-)]H(t) + \int_0^t H(u) dF(u), \quad (14.1.21a)$$

so that

$$[1 - F(t-)]H(t) = \int_0^t [1 - H(u)] dF(u). \quad (14.1.21b)$$

The  $\mathcal{H}$ -submartingale  $\{[M(t, \omega)]^2\}$  has the Doob–Meyer decomposition

$$[M(t, \omega)]^2 = Q(t, \omega) + M_2(t, \omega)$$

for some quadratic variation process  $Q$  and  $\mathcal{H}$ -martingale  $M_2$ . Equation (14.1.20) asserts that  $Q(t, \omega) = H_c(t, \omega) + \sum_{u_i \leq \min(t, X)} a_i(1 - a_i)$ , which is nondecreasing and  $\mathcal{H}$ -predictable as in the proof of Lemma 14.1.II. So to prove the lemma it suffices to show that the  $\mathcal{H}$ -adapted process

$$M_2(t, \omega) = \begin{cases} [H(t)]^2 - H(t) + H_2(t) & \text{if } X(\omega) > t, \\ [1 - H(X)]^2 - H(X) + H_2(X) & \text{if } X(\omega) \leq t, \end{cases} \quad (14.1.22)$$

is an  $\mathcal{H}$ -martingale, for which it is enough to check that  $E(M_2(t, \omega) \mid \mathcal{H}_s) = M_2(s, \omega)$  for  $t > s$ . From (14.1.22) this is trivially true on  $X(\omega) \leq s$ . On  $X(\omega) > s$  we use (14.1.7) to evaluate  $J(s, t) \equiv E(M_2(t, \omega) - M_2(s, \omega) \mid X > s)$ . Then  $[1 - F(s)]J(s, t)$  equals

$$\begin{aligned} & \int_s^t ([H(u) - 1]^2 - H(u) + H_2(u)) dF(u) \\ & + [1 - F(t)](H(t)[H(t) - 1] + H_2(t)) - [1 - F(s)](H(s)[H(s) - 1] + H_2(s)), \\ & = J_0 + J_1 + J_2 + J_3 \quad \text{say,} \end{aligned} \quad (14.1.23)$$

where  $J_0$  involves all the terms with  $H_2$  and is given by

$$\begin{aligned} J_0 &= \int_s^t H_2(u) dF(u) + [1 - F(t)]H_2(t) - [1 - F(s)]H_2(s) \\ &= \int_s^t [1 - F(u-)] dH_2(u) = \sum_{s < u_i \leq t} a_i^2 [1 - F(u_i-)], \end{aligned}$$

$J_1$  is the rest of the integral, and  $J_2$  and  $J_3$  are the boundary terms.

The term  $J_1$  can be written as

$$J_1 = \int_s^t [1 - 2H(u)] dF(u) + \int_s^t [H(u) - 1]H(u) dF(u),$$

and using Lemma 4.6.I, the second integral equals

$$F(t)[H(t) - 1]H(t) - F(s)[H(s) - 1]H(s) - \int_s^t F(u-) d([H(u) - 1]H(u)),$$

where writing  $H_c$  for the continuous component of  $H$  and  $a_i = \Delta H(u_i)$ , the last term equals

$$\begin{aligned} & - \int_s^t F(u-) [2H(u) - 1] dH_c(u) - \sum_{s < u_i \leq t} F(u_i-) (\Delta H(u_i) [2H(u_i) - 1] - a_i^2) \\ &= \sum_{s < u_i \leq t} a_i^2 F(u_i-) - \int_s^t F(u-) [2H(u) - 1] dH(u) \\ &= \sum_{s < u_i \leq t} a_i^2 F(u_i-) + \int_s^t [2H(u) - 1] dF(u) - \int_s^t [2H(u) - 1] dH(u), \end{aligned}$$

in which at the last step we have used  $dH(u) = dF(u)/[1 - F(u-)]$ . Now

$$\int_s^t 2H(u) dH(u) = [H(t)]^2 - [H(s)]^2 + \sum_{s < u_i \leq t} a_i^2$$

(use Lemma 4.6.I or Exercise 14.1.19). Putting these results together we find that  $[1 - F(s)]J(s, t)$  vanishes identically for  $s < t$  as required.  $\square$

## Exercises and Complements to Section 14.1

14.1.1 Consider a simple point process  $N$  on  $\mathcal{X} = \mathbb{Z}_+ \equiv \{0, 1, \dots\}$  adapted to the history  $\mathcal{F} = \{\mathcal{F}_n: n \in \mathbb{Z}_+\}$ . Show the following:

- (a) A process  $\{X_n: n \in \mathbb{Z}_+\}$  is  $\mathcal{F}$ -adapted if  $X_n$  is  $\mathcal{F}_n$ -measurable for each  $n = 0, 1, \dots$ .
- (b)  $\{X_n\}$  is  $\mathcal{F}$ -predictable if each  $X_n$  is  $\mathcal{F}_{n-1}$ -measurable for  $n = 1, 2, \dots$ .
- (c) An  $\mathcal{F}$ -adapted simple point process  $N$  on  $\mathbb{Z}_+$  has  $\mathcal{F}$ -compensator  $A_n$  given by

$$A_n = \sum_{k=1}^n E[N(\{k\}) \mid \mathcal{F}_{k-1}].$$

14.1.2 (*Continuation*). Suppose that  $N$  is a discrete time renewal process with lifetime distribution  $\{f_r\} = \{\Pr\{X = r\}: r = 1, 2, \dots\}$  (so  $f_0 = 0$ ). Then if

$$N(n) = \#\{j = 1, \dots, n: \text{renewal at } j\} = \#\{1 \leq j \leq n: N(\{j\}) = 1\},$$

and  $T(k) = \inf\{j: N(k) = j\}$ ,

$$\Delta A_n \equiv A_n - A_{n-1} = \begin{cases} \frac{f_{n-T(N(n-1))}}{q_{n-T(N(n-1))}} & \text{on } \{N(n-1) \geq 1\}, \\ f_{n-1}^{(0)} & \text{on } \{N(n-1) = 0\}, \end{cases}$$

where  $q_r = 1 - f_1 - \dots - f_{r-1}$  and  $\{f_r^{(0)}\}$  is the distribution of the initial length. Deduce that if  $\{f_r\}$  is the geometric distribution  $\{(1 - \rho)\rho^{r-1}: r = 1, 2, \dots\}$ , then  $A_n = n(1 - \rho)$ , representing a discrete analogue of the Poisson process.

14.1.3 Consider the history  $\mathcal{F}_{(+)} \equiv \{\mathcal{F}_{t+}: 0 \leq t < \infty\}$ , where  $\mathcal{F}_{t+} \equiv \bigcap_{s>t} \mathcal{F}_s$ . Prove that  $\mathcal{F}_{(+)}$  is right-continuous, and that if  $\xi(t)$  is  $\mathcal{F}$ -adapted and right-continuous, then it is also  $\mathcal{F}_{(+)}$ -adapted.

14.1.4 Show that the history generated by a cumulative process or random measure on  $\mathbb{R}_+$  need not necessarily be right-continuous. Investigate whether the origin is the only possible exceptional time.

[Hint: Consider Lebesgue measure multiplied by a nondegenerate r.v.]

14.1.5 Let the sequence of stopping times at (14.1.1) satisfy  $T_{n+1} = T_n$  for  $n \geq n_0$  for some finite integer  $n_0$ , so that  $T_\infty \equiv \lim_{n \rightarrow \infty} T_n < \infty$ .

(a) Prove that  $T_\infty$  is a stopping time.

(b) Show that the conclusions of Lemma 14.1.II and Theorem 14.1.IV continue to hold for  $t$  satisfying  $0 \leq t < T_\infty$ .

What can be said when  $T_n < T_{n+1} \uparrow T_\infty < \infty$ ?

14.1.6 (a) Let  $\{X(t): t \geq 0\}$  be an  $\mathcal{F}$ -submartingale. Show that if  $EX(t) = EX(0)$  for some  $t > 0$ , then  $Y(s, \omega) = X(\min(s, t), \omega)$  is an  $\mathcal{F}$ -martingale.

(b) Let  $S$  and  $T$  be two  $\mathcal{F}$ -stopping times, and  $\mathcal{F}_S$  the  $S$ -prior  $\sigma$ -algebra (see Definition A3.4.V). For  $U \in \mathcal{F}_S$  and  $S \leq T$  a.s., show that  $I_U(\omega)$  is  $\mathcal{F}$ -predictable. [Hint: Suppose first that  $S$  and  $T$  are both countably-valued r.v.s, and then use the fact that general stopping times can be approximated from above by such countably-valued stopping times.]

14.1.7 Let  $\xi(t)$  be an  $\mathcal{F}$ -adapted cumulative process with compensator  $A(t)$ ; define  $Z(t) = \xi(t) - A(t)$ . Show that for any bounded  $\mathcal{F}$ -predictable process  $Y(t)$ ,

$$\eta(t) \equiv \int_0^t Y(u) Z(du)$$

is an  $\mathcal{F}$ -local martingale; that is,  $\eta(t \wedge T_n)$  is an  $\mathcal{F}$ -martingale for each stopping time  $T_n$  as at (14.1.1). [Hint: The case  $Y(t, \omega) = I_{\{(s, t] \times B\}}$  for  $B \in \mathcal{F}_s$  and  $s < t$ , reduces to the defining requirement of the compensator. An extension argument completes the proof.]

- 14.1.8 Let  $N$  be a point process with internal history  $\mathcal{H}$ . A process  $X(t, \omega)$  on  $\mathbb{R}_+$  is  $\mathcal{H}$ -predictable if and only if for  $n = 0, 1, \dots$  there exist  $(\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{H}_{T_n})$ -measurable functions  $f_n(t, \omega)$  such that

$$X(t, \omega) = f_n(t, \omega) \quad \text{on } T_n < t \leq T_{n+1}.$$

[Hint: First write  $X(t, \omega) = \sum_{n=0}^{\infty} X(t, \omega) I_{(T_n, T_{n+1}]}(t, \omega)$ . Then argue as around (14.1.8) that, on each such set, the predictability requirement implies that  $X(t, \omega)$  must be  $\mathcal{H}_{T_n}$ -measurable.]

- 14.1.9 (a) Verify that the quantity  $N(t) - A(t)$  in the one-point process of Lemma 14.1.II has a stopping time representation as in Exercise 14.1.7.
- (b) Extend the argument above to the one-point processes with prior  $\sigma$ -algebra treated in Lemma 14.1.III.
- 14.1.10 (a) For Example 14.1(c), compare the two compensators with the internal history, and the intrinsic history where  $\mathcal{F}_0$  contains the event  $\{X = 1\}$ .
- (b) For Example 8.5(b), write down the compensator for the bivariate Poisson process in which each parent point is followed after a fixed delay  $h$  by a single offspring point. Find the compensators for the joint process, for the parent points only, and for the offspring points only. For the last case, compare the compensators when the history does or does not contain information about the past of the parent process.
- (c) Construct examples of discontinuous compensators first for the one-point process when the d.f. of  $X$  is a mixture of discrete and continuous components, and then [cf. (b)] extend to situations where the conditional hazard functions are themselves discontinuous.
- 14.1.11 In the setting of Example 14.1(e), the sum in (14.1.14)  $\sim E[H(T_1)]N(t)$  for large  $t$ . When the IHF  $H = H_a$  comes from a lifetime d.f.  $F_a$  different from the lifetime d.f.  $F$  underlying  $N(\cdot)$ ,  $E[H_a(T_1)] = \int_0^\infty [\bar{F}(u)/\bar{F}_a(u)] F_a(du)$ . This quantity equals 1 for the Poisson process, but as soon as it differs from 1,  $|N(t) - A_a(t)| = O(t) \gg |N(t) - A(t)| = o(t)$  for large  $t$ . In Figure 14.1, the upper tail behaviour of  $\bar{F}_a$  for the triangular lifetime distribution differs markedly from  $\bar{F}$ .
- 14.1.12 Show that the compensator of a sum of independent cumulative processes is the sum of the compensators of the components. Also consider more general linear combinations. [Hint: Take care with the histories.]
- 14.1.13 Let  $\{Q_0(x), Q_k(x; t_1, \dots, t_k): k = 1, 2, \dots\}$  be a family of nonnegative functions defined on  $0 \leq x < \infty$ ,  $0 < t_1 < \dots < t_k < \dots$ , satisfying the following conditions.
- $Q_k(x; t_1, \dots, t_k)$  is measurable in  $t_1, \dots, t_k$  for fixed  $x$ , and monotonic nondecreasing and right-continuous in  $x$  for fixed  $t_1, \dots, t_k$ .
  - $Q_0(0) = Q_k(0; t_1, \dots, t_k) = Q_k(x; t_1, \dots, t_k) = 0$  for  $x \leq t_k$ .
  - At any discontinuity  $x_i$  say, of  $Q_k$ ,  $\Delta Q_k(x_i) \equiv Q_k(x_i) - Q_k(x_i-) \leq 1$ , with  $\Delta Q_k(x_i) = 1$  only if  $Q_k(x; t_1, \dots, t_k) = Q_k(x_i; t_1, \dots, t_k)$  for all  $x > x_i$ .

Interpret the  $\{Q_k(\cdot)\}$  as the IHFs for the successive intervals of a simple point process  $N$  on  $(\mathcal{N}^\#(\mathbb{R}_+), \mathcal{B}(\mathcal{N}^\#(\mathbb{R}_+)))$  by showing that for given  $Q_k$  the corresponding d.f.s are determined uniquely from the results in Section 4.6.

Deduce that  $N$  has  $\mathcal{H}$ -compensator  $A(\cdot)$  given by  $A(t) = Q_k(t; t_1, \dots, t_k)$  for  $t_k < t \leq t_{k+1}$ , and that only finitely many points can occur in any finite interval, provided also that

(iv) the Janossy measures corresponding to the IHFs  $Q_k$  are proper.

Thus, the  $\mathcal{H}$ -compensator  $A$  determines the distribution of  $N$  [cf. Boel *et al.* (1975), Davis (1976), Brémaud (1981)].

14.1.14 (*Continuation*). Replace the functions  $Q_k(\cdot)$  by a family of regular conditional IHFs given a prior  $\sigma$ -algebra  $\mathcal{F}_0$ . Extend the uniqueness statement to a point process  $N$  with history  $\{\mathcal{F}_0 \vee \mathcal{H}_t\}$ .

Find an extension to the case where the given process  $N$  is conditioned by a point process  $N'$  evolving simultaneously with  $N$ .

[*Hint*: Use a history with  $\mathcal{F}_t = \mathcal{F}_t^{(N)} \vee \mathcal{F}_t^{(N')}$ ; no essential differences arise, but the functions  $Q_k$  are now functions of the points  $t'_i$  say, of the process  $N'$  as well as the points  $t_i$  of  $N$ , and can therefore change at both  $t_i$  and  $t'_i$ .]

14.1.15 *Quadratic variation (i): discrete-time point process*. Let  $\{N_n: n = 1, 2, \dots\}$  denote a discrete-time simple point process on  $\mathbb{Z}_0^+$  so  $N_n = \sum_{i=1}^n I_i$  for  $\{0, 1\}$ -valued r.v.s  $I_i$ . Then [cf. Exercise 14.1.1(c)] its Doob–Meyer decomposition  $N_n = A_n + M_n$  for an  $\mathcal{F}$ -martingale  $\{M_n\}$  and  $\mathcal{F}$ -compensator  $\{A_n\}$  is determined by  $A_n = \sum_{i=1}^n E(I_i | \mathcal{F}_{i-1})$ .

Check that  $[M_n]^2$  is an  $\mathcal{F}$ -submartingale. The Doob–Meyer decomposition justifies the representation  $[M_n]^2 = Q_n + M_{2,n}$  where  $\{M_{2,n}\}$  is an  $\mathcal{F}$ -martingale and

$$Q_n = \sum_{i=1}^n E(I_i | \mathcal{F}_{i-1})[1 - E(I_i | \mathcal{F}_{i-1})].$$

When  $N_n$  is a discrete-time renewal process with lifetime distribution  $\{f_j\}$  (so  $f_j = \Pr\{I_1 = \dots = I_{j-1} = 0, I_j = 1\}$ ), and  $T_n = \sup\{i \leq n: I_i = 1\}$ ,  $E(I_n | \mathcal{F}_{n-1}) = a_n - T_{n-1}$ , where  $a_j = f_j/(1 - F_{j-1})$ ,  $F_j = f_1 + \dots + f_j$ . For such a renewal process, with successive lifetimes  $\{\tau_r\}$  say, the quadratic variation is expressible in terms of  $H_{(2)}(k) = \sum_{j=1}^k a_j(1 - a_j)$  [cf. (14.1.14)] as

$$Q_n = \sum_{i=1}^n a_{i-T_{i-1}}(1 - a_{i-T_{i-1}}) = \sum_{r=1}^{N_n} H_{(2)}(\tau_r) + H_{(2)}(n - T_n).$$

14.1.16 *Quadratic variation (ii): continuous time examples*.

(a) Use (14.1.18) to show generally, if informally, that for a simple point process, the quadratic variation  $Q(u)$  increases by  $dA_c(u)$  as  $u$  passes through a point of continuity of the compensator  $A$ , and by the sum  $a(1-a)^2 + (1-a)a^2 = a(1-a)$  when  $u$  passes through a point where  $A$  has a jump of height  $a$ , the two terms in the sum representing a weighted average of the jumps associated with the occurrence and nonoccurrence of a point at  $u$ . Verify that the result is compatible with Lemma 14.1.IX for the case of a one-point process.

(b) Consider the one-point process when  $X$  has the d.f.

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{if } x < 1, \\ 1 - (1-a)e^{-\lambda x} & \text{if } 1 \leq x, \end{cases}$$



for some  $a \in (0, 1)$ , so the corresponding IHF equals  $\lambda x + aI_{\{x \geq 1\}}$ . Show that the quadratic variation  $Q(x, \omega) = \lambda \min(x, X) + a(1 - a)I_{\{X \geq x\}}$ .

- (c) Investigate the forms of the compensator and quadratic variation for a one-point multivariate point process.

14.1.17 Let  $N$  be a simple stationary point process on  $\mathbb{R}$  with continuous compensator  $A$ , mean rate  $m$ , and finite second moment. Show that for real or complex  $f \in L_1 \cap L_2$ ,

$$\text{var} \left( \int_{\mathbb{R}} f(x) [dN(t) - dA(t)] \right) = m \int_{\mathbb{R}} |f(t)|^2 dt.$$

Investigate what changes are needed if the compensator is not continuous. [Hint: Use results on the quadratic variation process in Proposition 14.1.IX. This idea is used in Brémaud, Massoulié and Ridolfi (2005) to establish results for the Bartlett spectrum of a process with a conditional intensity.]

14.1.18 *Unbounded compensators.*

- (a) Suppose the d.f.  $F$  of a nonnegative r.v. is purely atomic with atoms of mass  $(1 - p)p^{n-2}$  at points  $u_n = 1 - 1/n$  for  $n = 2, 3, \dots$ . Show that a one-point process as in Lemma 14.1.II whose determining r.v.  $X$  has such  $F$  as its d.f. has an explosive compensator.
- (b) Show a similar property for the d.f.  $F(x) = 1 - \sqrt{1 - x}$ .

14.1.19 For real-valued functions  $F$  and  $G$  of locally bounded variation on  $\mathbb{R}$ , show that for any finite open interval  $(a, b)$ ,

$$\begin{aligned} \int_{(a,b)} F(x) dG(x) + \int_{(a,b)} G(x) dF(x) &= F(b-)G(b-) - F(a+)G(a+) \\ &\quad + \sum_{a < x_i < b} [\Delta^L F(x_i) \Delta^L G(x_i) - \Delta^R F(x_i) \Delta^R G(x_i)], \end{aligned}$$

where the sum extends over all discontinuities of such functions for which  $\Delta^L F(x) = F(x) - F(x-)$ ,  $\Delta^R F(x) = F(x+) - F(x)$ . Recover Lemma 4.6.I. [Hint: See, e.g., Asplund and Bungart (1966, Proposition 8.5.5).]

## 14.2. Campbell Measure and Predictability

The properties of compensators are closely linked with concepts from the general theory of processes, especially the Campbell measure on  $\mathbb{R}_+ \times \Omega$  defined on product sets  $B \times U$  for  $B \in \mathcal{B}(\mathbb{R}_+)$  and  $U \in \mathcal{E}$  by

$$C_\xi(B \times U) = \int_U \xi(B, \omega) \mathcal{P}(d\omega). \quad (14.2.1)$$

In Chapter 13 we mostly took  $\Omega$  to be the canonical space  $\mathcal{M}_X^\#$  or  $\mathcal{N}_X^\#$ ; here we prefer a general  $\Omega$ , because of the greater flexibility this allows in the choice of the history  $\mathcal{F}$ . We explore this connection first for simple point processes,

and then extend the discussion to MPPs. As an application of these ideas we examine an extension of the limit theorems of Section 11.3 for thinned processes to situations where the thinning probability may depend on the past history of the process.

As in Section 13.1, the set function  $C_\xi$  can be extended to a  $\sigma$ -finite measure on the product  $\sigma$ -algebra  $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{E}$ . This done, we focus attention on its restriction to the sets of the  $\mathcal{F}$ -predictable  $\sigma$ -algebra  $\Psi^{\mathcal{F}}$  of Section A3.3. In particular, for  $U \in \mathcal{F}_s$  and  $B = (s, t]$  for  $0 \leq s < t < \infty$ , we have from the defining relation

$$C_\xi((s, t] \times U) = \int_U \mathcal{P}(d\omega) \int_{(s, t]} \xi(dx, \omega).$$

On the other hand, the martingale relation of Theorem 14.1.VII implies that for  $U \in \mathcal{F}_s$  and  $n = 1, 2, \dots$ ,

$$\begin{aligned} \int_U [\xi(t \wedge T_n, \omega) - A(t \wedge T_n, \omega)] \mathcal{P}(d\omega) \\ = \int_U [\xi(s \wedge T_n, \omega) - A(s \wedge T_n, \omega)] \mathcal{P}(d\omega), \end{aligned}$$

which on rearrangement gives

$$\int_U \mathcal{P}(d\omega) \int_{s \wedge T_n}^{t \wedge T_n} \xi(dx, \omega) = \int_U \mathcal{P}(d\omega) \int_{s \wedge T_n}^{t \wedge T_n} A(dx, \omega).$$

Then monotone convergence can be used to let  $n \rightarrow \infty$  and we conclude that

$$\int_U \mathcal{P}(d\omega) \int_s^t \xi(dx, \omega) = \int_U \mathcal{P}(d\omega) \int_s^t A(dx, \omega),$$

or, equivalently,

$$\mathbb{E} \int_{\mathbb{R}_+} I_{(s, t] \times U}(x, \omega) \xi(dx, \omega) = \mathbb{E} \int_{\mathbb{R}_+} I_{(s, t] \times U}(x, \omega) A(dx, \omega). \quad (14.2.2)$$

Comparison with (13.1.2) shows that this is just the assertion that on  $\Psi^{\mathcal{F}}$  the Campbell measures induced by  $\xi$  and its compensator  $A$  coincide. More generally, the indicator function in (14.2.2) can be replaced by any  $\Psi^{\mathcal{F}}$ -measurable function  $Y(\cdot)$ . But such a function is just an  $\mathcal{F}$ -predictable process. This argument can be reversed, and leads to the following alternative characterization of the compensator, equivalent to the martingale characterization, but couched here in terms of the general theory of processes.

**Proposition 14.2.I.** *Given a cumulative process  $\xi$  adapted to the history  $\mathcal{F}$ , its compensator is the unique  $\mathcal{F}$ -predictable cumulative process  $A$  satisfying*

$$\mathbb{E} \int_{\mathbb{R}_+} Y(t, \omega) \xi(dt, \omega) = \mathbb{E} \int_{\mathbb{R}_+} Y(t, \omega) A(dt, \omega) \quad (14.2.3)$$

for every nonnegative  $\mathcal{F}$ -predictable process  $Y(\cdot)$ .

It is on account of this relation that the compensator is referred to as the *dual predictable projection* in literature that adheres to the terminology of the general theory of processes. The use of the word ‘projection’ relates to the fact that any conditional expectation can be viewed as a projection from one space of functions to a subspace of functions having a coarser structure: see Krickeberg (1965, Chapter IV) for an introduction to this circle of ideas. The analogue is clearer in the next proposition, which introduces the predictable projection itself.

**Proposition 14.2.II.** *Let  $X(t, \omega)$  be a measurable stochastic process on the probability space  $(\Omega, \mathcal{E}, \mathcal{P})$  satisfying, for all  $0 < t < \infty$ ,*

$$\mathbb{E} \int_0^t |X(s, \omega)| \, ds < \infty,$$

*and let  $\mathcal{F}$  be a history for  $X$ . Then there exists an  $\mathcal{F}$ -predictable process  $X^{\mathcal{F}}(t, \omega)$  such that for all bounded  $\mathcal{F}$ -predictable processes  $Y(t, \omega)$ ,*

$$\mathbb{E} \int_0^t Y(u, \omega) X(u, \omega) \, du = \mathbb{E} \int_0^t Y(u, \omega) X^{\mathcal{F}}(u, \omega) \, du \quad (\text{all } 0 < t < \infty). \quad (14.2.4)$$

*Moreover,  $X^{\mathcal{F}}(\cdot)$  is uniquely defined up to its values on a predictable set of  $(\ell \times \mathcal{P})$ -measure zero.*

PROOF. First let  $Y(u, \omega)$  be the indicator function of the generating set  $(s, t] \times U$  ( $0 \leq s < t < \infty$ ,  $U \in \mathcal{F}_s$ ) of  $\Psi^{\mathcal{F}}$ , for which (14.2.4) takes the form of the defining relation for the Radon–Nikodym derivative

$$\int_{(s, t] \times U} X(u, \omega) \, du \, \mathcal{P}(d\omega) = \int_{(s, t] \times U} \tilde{X}(u, \omega) \, du \, \mathcal{P}(d\omega).$$

Because the Radon–Nikodym derivative is unique up to values on null sets of the defining  $\sigma$ -algebra, this both identifies  $X^{\mathcal{F}}$  as the Radon–Nikodym derivative, and justifies the uniqueness statement in the proposition.  $\square$

The process  $X^{\mathcal{F}}$  can be called the  *$\mathcal{F}$ -predictable projection of  $X$* , or predictable projection for short. Note that the technique of proof is exactly that used in the study of conditional expectation, except that here the measure on  $\mathbb{R}_+ \times \Omega$  is no longer finite nor is  $X$  necessarily integrable on  $\mathbb{R}_+ \times \Omega$ .

We turn now to the extension of these ideas to MPPs. A *marked cumulative process* is a random kernel  $\xi(t, K, \omega)$ ,  $\mathbb{R}_+ \times \mathcal{B}_{\mathcal{K}} \times \Omega \mapsto [0, \infty]$  such that

- (i) for each fixed  $K \in \mathcal{B}_{\mathcal{K}}$ , the process  $\xi(\cdot, K)$  is monotonic increasing and right-continuous in its first argument  $t$ , and
- (ii) for each fixed  $t \in \mathbb{R}_+$ ,  $\xi(t, \cdot)$  is a boundedly finite random measure on  $\mathcal{B}_{\mathcal{K}}$ .

Such marked cumulative processes arise, in particular, from counting the number of points with marks in  $K$  from an MPP observed over the time interval  $(0, t]$ . This is the only case we consider in detail.

Results for marked cumulative processes, analogous to the Doob–Meyer decomposition or the projection theorems just described, can be obtained by considering decompositions of the three-component product space  $\mathbb{R}_+ \times \mathcal{K} \times \Omega$ . The *marked Campbell measure* on this product space is defined by the relations

$$C_\xi(B \times K \times U) = \int_U \xi(B \times K, \omega) \mathcal{P}(d\omega), \quad (14.2.5)$$

where  $B$  and  $K$  are bounded Borel sets from  $\mathbb{R}_+$  and  $\mathcal{K}$ , respectively. For a given history  $\mathcal{F}$ , the *mark-predictable  $\sigma$ -algebra*,  $\Psi_{\mathcal{K}}^{\mathcal{F}}$  is generated by product sets of the form  $(s, t] \times K \times U$  with  $0 < s < t$ ,  $K \in \mathcal{B}_{\mathcal{K}}$  and  $U \in \mathcal{F}_s$ . Equivalently, it is the product of the predictable  $\sigma$ -algebra  $\Psi^{\mathcal{F}}$  with  $\mathcal{B}_{\mathcal{K}}$ . A marked cumulative process is *mark-predictable* if, as a function of three arguments, it is measurable with respect to the mark-predictable  $\sigma$ -algebra.

**Definition 14.2.III.** The  $\mathcal{F}$ -compensator of an MPP  $N$  on  $\mathcal{X}$  with marks in  $\mathcal{K}$  is any mark-predictable, cumulative process  $A(t, K, \omega)$  such that, for each  $K \in \mathcal{B}_{\mathcal{K}}$ ,  $A(t, K, \omega)$  is the  $\mathcal{F}$ -compensator for the simple point process  $N_K(t) \equiv N((0, t] \times K)$ .

The existence and structure of compensators in the marked case are clarified by introducing the ground process  $N_g$  of Section 6.4. By assumption  $N_g$  is a well-defined point process in its own right. For the present context, we also require  $N_g$  to have finite first moment measure, and hence its own compensator  $A_g$  relative to the history  $\mathcal{F}$ . For a given mark set  $K$  and history  $\mathcal{F}$ , we can also define the compensator  $A(t, K)$  for  $N_K(t)$ . For each such  $K$ , the associated Campbell measure satisfies

$$C_{N_K}(dt \times d\omega) \ll C_{N_g}(dt \times d\omega).$$

If we think of this result as a relation between measures on the product  $\sigma$ -algebra  $\Psi^{\mathcal{F}} \times \mathcal{E}$ , it follows as in Proposition A1.5.III for the existence of regular conditional probabilities, that we can write

$$C_\xi((s, t] \times K \times U) = C_{N_K}((s, t] \times U) = \int_U \int_s^t F(K | u, \omega) C_{N_g}(du \times d\omega) \quad (14.2.6)$$

for  $s < t$ ,  $U \in \mathcal{F}_s$ , and some *predictable kernel*  $F(K | t, \omega)$ , which is

- (i)  $\Psi^{\mathcal{F}}$ -measurable, and hence predictable, for each fixed  $K$ ; and
- (ii) a probability distribution on  $\mathcal{B}_{\mathcal{K}}$  for  $(\ell \times \mathcal{P})$ -almost all pairs  $(t, \omega)$ .

We now show that if  $A_g(t, \omega)$  is the  $\mathcal{F}$ -compensator for the ground process, then with  $F$  determined by (14.2.6), the process  $A(t, K, \omega)$  defined by

$$A(t, K, \omega) = \int_0^t F(K | u, \omega) A_g(du, \omega) \quad (14.2.7)$$

is an  $\mathcal{F}$ -compensator for the marked process. Using the Doob–Meyer decomposition for  $N_g$ , we have, for any nonnegative predictable process  $X(t, \omega)$ ,

$$\mathbb{E} \left[ \int_{\mathbb{R}_+} X(t, \omega) N_g(dt, \omega) \right] = \mathbb{E} \left[ \int_{\mathbb{R}_+} X(t, \omega) A_g(dt, \omega) \right].$$

Take  $X(t, \omega) = \int_{\mathcal{K}} Y(t, \kappa, \omega) F(d\kappa \mid t, \omega)$  for any given mark-predictable process  $Y(t, \kappa, \omega)$ , and observe that the properties of  $Y$  and the kernel  $F$  imply that  $X$  is indeed predictable. The equations in the chain below then hold (Fubini's theorem implies finiteness of one and every element of the chain):

$$\begin{aligned} \mathbb{E} \left[ \int_{\mathbb{R}_+ \times \mathcal{K}} Y(t, \kappa, \omega) \xi(dt \times d\kappa, \omega) \right] &= \int_{\mathbb{R}_+ \times \mathcal{K} \times \Omega} Y(t, \kappa, \omega) C_{\xi}(dt \times d\kappa \times d\omega) \\ &= \int_{\mathbb{R}_+ \times \Omega} \int_{\mathcal{K}} Y(t, \kappa, \omega) F(d\kappa \mid t, \omega) C_{N_g}(dt \times d\omega) \\ &= \mathbb{E} \left[ \int_{\mathbb{R}_+} X(t, \omega) N_g(dt, \omega) \right] \\ &= \mathbb{E} \left[ \int_{\mathbb{R}_+} X(t, \omega) A_g(dt, \omega) \right] \\ &= \int_{\Omega} \int_{\mathbb{R}_+} \int_{\mathcal{K}} Y(t, \kappa, \omega) F(d\kappa \mid t, \omega) A_g(dt, \omega) \mathcal{P}(d\omega), \end{aligned}$$

which both establishes the decomposition of the marked Campbell measure into the product of the predictable kernel  $F(d\kappa \mid t, \omega)$  and the Campbell measure for the ground process, and the martingale property for the marked process. In particular, setting  $Y(t, \kappa, \omega) = I_K(\kappa)X(t, \omega)$  yields the equations

$$\int_{\mathbb{R}_+ \times \Omega} X(t, \omega) C_{N_K}(dt \times d\omega) = \int_{\Omega} \int_{\mathbb{R}_+} X(t, \omega) F(K \mid t, \omega) A_g(dt, \omega) \mathcal{P}(d\omega), \quad (14.2.8)$$

implying that the compensator for  $N_K$  can indeed be represented in the form (14.2.7). Uniqueness follows from the uniqueness of the Radon–Nikodym derivatives in (14.2.6).

In summary, we have established the first part of the next theorem. For the rest, in the special case that the history  $\mathcal{F}$  is either the internal history  $\mathcal{H}$  of the MPP or an intrinsic history  $\mathcal{F} = \mathcal{F}_0 \vee \mathcal{H}$ , we can develop a more explicit representation for the compensator extending that of Theorem 14.1.IV. Again we start by considering the canonical example of a one-point MPP defined by a bivariate distribution  $\Gamma$  on  $\mathcal{B}(\mathbb{R}_+ \times \mathcal{K})$  for the time  $X$  and mark  $Y$  of the unique point and develop an analogue of Lemma 14.1.II [cf., e.g., Last and Brandt (1995, Proposition 1.7.1)]. The family of conditional distributions  $\mu(K \mid t) = \Pr\{Y \in K \mid X = t\}$  for  $(t, K) \in \mathbb{R}_+ \times \mathcal{K}$  is trivially predictable because for every  $K$  it is a function of  $t$  that is either deterministic or determined entirely by the prior  $\sigma$ -algebra  $\mathcal{F}_0$ .

From this point onwards the development largely duplicates the discussion leading to Theorem 14.1.IV and the associated assertion in Proposition 14.1.VI that for internal or intrinsic histories, the compensator and the point process determine each other uniquely. The arguments are briefly outlined in Exercises 14.2.1–3 and the results summarized in part (b) of the theorem.

**Theorem 14.2.IV.** (a) Let  $N$  be an  $\mathcal{F}$ -adapted marked point process with ground process  $N_g$  having finite first moment measure. Then an  $\mathcal{F}$ -compensator for  $N$  exists, is  $(\ell \times \mathcal{P})$ -a.e. unique, and can be represented as in (14.2.7) in terms of a mark-predictable kernel  $F$ .

(b) Suppose, in particular, that  $\mathcal{F}$  is a history for  $N$  of the form  $\{\mathcal{F}_0 \vee \mathcal{H}_t\}$ , that  $\{(T_n, \kappa_n)\}$  is the sequence of stopping times and marks as at (14.1.1), and that there exist regular versions  $\Gamma_n(\cdot, \cdot \mid \mathcal{F}_{(n-1)})$  of the bivariate conditional distributions of the pairs  $(\tau_n, \kappa_n)$ , where  $\tau_n = T_n - T_{n-1}$ , given the  $\sigma$ -algebras  $\mathcal{F}_{(n-1)}$  generated by the history up to time  $T_{n-1}$ . Also let  $G_n(\cdot \mid \mathcal{F}_{(n-1)})$  denote the marginal conditional distribution function for  $\tau_n$ , and  $F_n(\cdot \mid \tau_n, \mathcal{F}_{(n-1)})$  the conditional distribution of  $\kappa_n$  given  $\tau_n$  and  $\mathcal{F}_{(n-1)}$ , and suppose  $G_n$  satisfies  $1 - G_n(x-) > 0$  for  $x > 0$ . Then a version of the  $\mathcal{F}$ -compensator  $A(t, K, \omega)$  at (14.2.7) is given by

$$A_g(t, \omega) = \sum_{n=1}^{\infty} A_g^{(n)}(t, \omega), \quad (14.2.9)$$

$$F(K \mid t, \omega) = \sum_{n=1}^{\infty} I_{(T_{n-1}(\omega), T_n(\omega)]}(t) F_n(K \mid t - T_{n-1}(\omega), \mathcal{F}_{(n-1)}),$$

where

$$A_g^{(n)}(t, \omega) = \int_0^{\tau_n \vee (t - T_{n-1}(\omega))} \frac{G_n(du \mid \mathcal{F}_{(n-1)})}{1 - G_n(u- \mid \mathcal{F}_{(n-1)})}.$$

(c) Under the conditions in part (b), the form of the compensator and the fidi distributions for the process determine each other uniquely.

EXAMPLE 14.2(a) *Semi-Markov processes* [see Example 10.3(a)]. In viewing a semi-Markov process  $X(\cdot)$  as an MPP in Section 10.3 we detailed its conditional intensity, its ground process, and its conditional mark distribution, in terms of the structural elements of the process given there, including the transition-time density functions  $g_{jk}(t)$  ( $j, k \in \mathbb{X}$ ,  $t > 0$ ). In terms of a realization  $\{(t_i, \kappa_i)\}$  as before,  $\sigma$ -fields  $\mathcal{F}_{(n)}$  are determined by  $\{(t_i, \kappa_i) : i \leq n\}$ , consistent with the setting also of Theorem 14.2.IV, and then with  $t_{n-1} < t \leq t_n$  and  $\tau_n = t_n - t_{n-1}$ ,

$$G_n(t \mid \mathcal{F}_{(n-1)}) = \sum_{j \in \mathbb{X}} \int_0^{t - t_{n-1}} g_{\kappa_{n-1}, j}(u) du = G_{\kappa_{n-1}}(t - t_{n-1}),$$

$$F_n(\{\kappa\} \mid \tau_n, \mathcal{F}_{(n-1)}) = \frac{g_{\kappa_{n-1}, \kappa}(\tau_n)}{\sum_{j \in \mathbb{X}} g_{\kappa_{n-1}, j}(\tau_n)},$$

$$A_g^{(n)}(t, \omega) = \int_0^{t \wedge t_n(\omega) - t_{n-1}(\omega)} \frac{\sum_{j \in \mathbb{X}} g_{\kappa_{n-1}, j}(u) du}{1 - G_{\kappa_{n-1}}(u-)},$$

in which we observe that  $A_g^{(n)}$  is an IHF.

In the simpler case that  $X$  is a countable state space continuous-time Markov process with transition matrix  $(q_{ij})$  and no instantaneous states, these functions simplify (with  $t$  and  $\tau_n$  as above) to

$$\begin{aligned} G_n(t \mid \mathcal{F}_{(n-1)}) &= 1 - \exp(-q_{\kappa_{n-1}}(t - t_{n-1})), \\ F_n(\{\kappa\} \mid \tau_n, \mathcal{F}_{(n-1)}) &= p_{\kappa_{n-1}, \kappa} = q_{\kappa_{n-1}, \kappa} / q_{\kappa_{n-1}}, \\ \Gamma_n(t, \{\kappa\} \mid \mathcal{F}_{(n-1)}) &= \Pr\{\tau_n > u, \kappa_n = \kappa \mid \mathcal{F}_{(n-1)}\} = p_{\kappa_{n-1}, \kappa} e^{-q_{\kappa_{n-1}} u} \\ &= [1 - G_n(u \mid \mathcal{F}_{(n-1)})] F_n(\{\kappa\} \mid u, \mathcal{F}_{(n-1)}), \end{aligned}$$

and  $F(\{\kappa\} \mid t, \omega) = p_{X(t-), \kappa}$ . Further discussion is given in Last and Brandt (1995, Exercise 4.3.4).  $\square$

The projection operator onto the mark-predictable functions, and its dual, give rise to projection theorems analogous to Theorems 14.2.I and 14.2.II, which we state below for completeness; proofs are sketched in Exercise 14.2.4. See also Exercise 14.2.5.

**Proposition 14.2.V.** *Given a marked cumulative process  $\xi$  adapted to the history  $\mathcal{F}$ , its compensator  $A$  is the unique  $\mathcal{F}$ -predictable cumulative marked process satisfying, for every nonnegative mark-predictable process  $Y(\cdot, \cdot)$ ,*

$$\begin{aligned} \mathbb{E} \int_{\mathbb{R}_+ \times \mathcal{K}} Y(t, \kappa, \omega) \xi(dt \times d\kappa, \omega) &= \mathbb{E} \int_{\mathbb{R}_+ \times \mathcal{K}} Y(t, \kappa, \omega) A(dt \times d\kappa, \omega) \\ &= \mathbb{E} \int_{\mathbb{R}_+} \left[ \int_{\mathcal{K}} Y(t, \kappa, \omega) F(d\kappa \mid t, \omega) \right] A_g(dt, \omega). \end{aligned} \quad (14.2.10)$$

In the next theorem we assume the existence of a boundedly finite (hence  $\sigma$ -finite) reference measure  $\ell_{\mathcal{K}}$  on the Borel sets  $\mathcal{B}_{\mathcal{K}}$  of the mark space.

**Proposition 14.2.VI.** *Let  $X(t, \kappa, \omega)$  be a measurable marked stochastic process on the probability space  $(\Omega, \mathcal{E}, \mathcal{P})$  satisfying*

$$\mathbb{E} \int_0^t \int_{\mathcal{K}} |X(s, \kappa, \omega)| ds \ell_{\mathcal{K}}(d\kappa) < \infty \quad (\text{all } 0 < t < \infty, \text{ bounded } K \in \mathcal{B}_{\mathcal{K}}),$$

*and let  $\mathcal{F}$  be a history for  $X$ . Then there exists an  $\mathcal{F}$ -predictable marked process  $X^{\mathcal{F}}(t, \kappa, \omega)$  such that for all bounded  $\mathcal{F}$ -mark-predictable processes  $Y(t, \kappa, \omega)$ , all  $0 < t < \infty$ , and all  $K \in \mathcal{B}_{\mathcal{K}}$ ,*

$$\begin{aligned} \mathbb{E} \int_0^t \int_{\mathcal{K}} Y(u, \kappa, \omega) X(u, \kappa, \omega) du \ell_{\mathcal{K}}(d\kappa) \\ = \mathbb{E} \int_0^t \int_{\mathcal{K}} Y(u, \kappa, \omega) X^{\mathcal{F}}(u, \kappa, \omega) du \ell_{\mathcal{K}}(d\kappa). \end{aligned} \quad (14.2.11)$$

Moreover,  $X^{\mathcal{F}}(\cdot)$  is uniquely defined up to its values on a predictable set of  $(\ell \times \ell_{\mathcal{K}} \times \mathcal{P})$ -measure zero.

We conclude this section with some results illustrating applications of the preceding ideas to limit theorems stated in terms of the convergence of compensators. The following technical lemma is a useful aid in these proofs.

**Lemma 14.2.VII.** *Let  $\mathcal{F}$  be a history,  $N(\cdot)$  a simple point process that is  $\mathcal{F}$ -predictable, and  $X(\cdot)$  a nonnegative  $\mathcal{F}$ -adapted process with finite first moment. Then the cumulative process*

$$\eta(t) = \sum_{i: t_i \leq t} X(t_i) = \int_0^t X(s+) dN(s), \quad (14.2.12)$$

where  $N(\cdot)$  has jump points  $\{t_i\}$ , is  $\mathcal{F}$ -adapted with  $\mathcal{F}$ -compensator

$$\alpha(t) = \sum_{i: t_i \leq t} \zeta(t_i) \equiv \sum_{i: t_i \leq t} E[X(t_i) | \mathcal{F}_{t_i-}]. \quad (14.2.13)$$

PROOF. All the processes  $N$ ,  $\eta$ , and  $\alpha$  are clearly right-continuous and monotonic increasing, with their points of increase confined to  $\{t_i\}$  (or a subset of these jump points of  $N$ ), where the sizes of their jumps are 1,  $X(t_i)$ , and  $\zeta(t_i)$ , respectively.  $N$  is  $\mathcal{F}$ -adapted by assumption, and it is easily checked that so too are  $\eta$  and  $\alpha$ .

Two more steps are needed to prove the lemma: we must show that  $\alpha(\cdot)$  is  $\mathcal{F}$ -predictable, and that  $\eta(\cdot) - \alpha(\cdot)$  is an  $\mathcal{F}$ -local martingale. Because  $\alpha(\cdot)$  can be written in the form

$$\alpha(t, \omega) = \sum_{i=1}^{\infty} \zeta(t_i) V_{t_i}^+(t, \omega),$$

where  $V_T^+(t, \omega) \equiv I_{\{\omega: T(\omega) \leq t\}}(t, \omega)$  is the right-continuous indicator process of the stopping time  $T$  (see the end of Appendix A3.3 for discussion), we limit ourselves here to establishing these steps in the special case of a one-point process, that is, where

$$\eta(t, \omega) = X(T) V_T^+(t, \omega) \quad (14.2.14)$$

and

$$\alpha(t, \omega) = \zeta(T) V_T^+(t, \omega), \quad (14.2.15)$$

with  $\zeta(T) = E[X(T) | \mathcal{F}_{T-}]$  for some  $\mathcal{F}$ -stopping time  $T$  such that  $V_T^+(\cdot)$  is predictable. The extension to the general case is left to Exercise 14.2.6.

Let  $Y$  be any  $\mathcal{F}_{T-}$ -measurable r.v., and consider the process  $Y V_T^+(t)$ . Indeed, taking  $Y$  first to be an indicator r.v. of some basic  $\mathcal{F}_{T-}$ -measurable set of the form  $B_s \cap \{T(\omega) > s\}$  for some  $B_s \in \mathcal{F}_s$  and  $s \in \mathbb{R}_+$ ,  $\alpha(t, \omega) = 1$  on the intersection of the sets

$$\{(t, \omega): T(\omega) \leq t\} \quad \text{and} \quad \{(t, \omega): t > s, \omega \in B_s \cap \{\omega: T(\omega) > s\}\}.$$

The first set here is in the predictable  $\sigma$ -algebra  $\Psi^{\mathcal{F}}$  by assumption, and the second is of the form of a generating set for  $\Psi^{\mathcal{F}}$  and is therefore predictable also. A similar argument holds when  $Y$  is the indicator r.v. of the other type of generating set for  $\Psi^{\mathcal{F}}$ , namely, an element of  $\mathcal{F}_0$ . Now the class of predictable processes is closed under the formation of nonnegative linear combinations and monotone limits, so it follows that  $Y V_T^+(\cdot)$  is a predictable process whenever



$Y$  is  $\mathcal{F}_{T-}$ -measurable and nonnegative, and hence in particular when it has the form  $E[X(T) \mid \mathcal{F}_{T-}]$ .

For the second part of the proof we have to show that for every  $A_s \in \mathcal{F}_s$  and  $s < t$ ,

$$\int_{A_s} [\eta(t, \omega) - \alpha(t, \omega)] \mathcal{P}(d\omega) = \int_{A_s} [\eta(s, \omega) - \alpha(s, \omega)] \mathcal{P}(d\omega),$$

with  $\eta$  and  $\alpha$  as at (14.2.14) and (14.2.15), so that

$$\eta(t, \omega) - \alpha(t, \omega) = (X(T) - E[X(T) \mid \mathcal{F}_{T-}]) I_{\{T(\omega) \leq t\}}(t, \omega).$$

The indicator function vanishes on  $\{\omega: T(\omega) > t\}$ , and on  $\{T(\omega) \leq s\}$  for  $s < t$ , so it is enough to consider  $\{\omega: s < T(\omega) \leq t\}$ , and hence show that, for  $T > s$ ,

$$\int_{A_s \cap \{T > s\}} (X(T) - E[X(T) \mid \mathcal{F}_{T-}]) \mathcal{P}(d\omega) = 0. \quad (14.2.16)$$

Because the set  $A_s \cap \{\omega: T(\omega) > s\}$  is an element of both  $\mathcal{F}_s$  and  $\mathcal{F}_{T-}$ , the integrand can be replaced by its conditional expectation with respect to  $\mathcal{F}_{T-}$ , and thus (14.2.16) holds.  $\square$

The following result of Brown (1978, 1982, 1983) establishes convergence to Poisson and Cox processes via conditions on compensators.

**Theorem 14.2.VIII.** *Let  $\{\mathcal{F}^{(n)}: n = 1, 2, \dots\}$  be a sequence of histories defined on a common probability space  $(\Omega, \mathcal{E}, \mathcal{P})$ ,  $\{N_n\}$  a sequence of simple point processes for which for each  $n$ ,  $N_n$  is  $\mathcal{F}^{(n)}$ -adapted and has  $\mathcal{F}^{(n)}$ -compensator  $A_n$ . Suppose  $A(\cdot)$  is a cumulative process defined on  $(\Omega, \mathcal{E}, \mathcal{P})$  with continuous trajectories and such that for each  $t > 0$ ,*

(i)  $A(t)$  is  $\mathcal{F}_0^{(n)}$ -measurable for every  $n = 1, 2, \dots$ ; and

(ii)  $A_n(t) \rightarrow A(t)$  ( $n \rightarrow \infty$ ) in probability.

*Then  $N_n$  converges weakly to a Cox process directed by  $A$ .*

PROOF. We first show that conditions (i) and (ii) imply the tightness of the distributions on  $\mathcal{N}_{\mathbb{R}_+}^\#$  induced by  $\{N_n\}$ , so that subsequences converge weakly to the distribution of some point process  $N$ , and then stable convergence is used to help in identifying the limit distribution.

Tightness in the present context amounts to showing that for each  $t \in \mathbb{R}_+$ , given  $\epsilon > 0$ , we can find some  $a_0$  and some subsequence  $\{n_k\}$  such that for all  $n \in \{n_k\}$ ,

$$\mathcal{P}\{N_n(t) > a\} < \epsilon \quad (\text{all } a > a_0); \quad (14.2.17)$$

mostly, we take ‘ $n$  in a subsequence  $\{n_k\}$ ’ as understood. Now for any event  $U \in \mathcal{E}$ ,

$$|\mathcal{P}(U \mid \{A(t) \leq M\}) - \mathcal{P}(U)| \leq \frac{\mathcal{P}\{A(t) > M\}}{\mathcal{P}\{A(t) \leq M\}} \leq \frac{1}{12} \epsilon$$

for  $M$  sufficiently large. Consequently, we can establish (14.2.17) by showing that for all  $n$ ,

$$\mathcal{P}_0(\{N_n(t) > a\} \mid \{A(t) \leq M\}) < \frac{2}{3}\epsilon \quad (\text{all } a > a_0). \quad (14.2.18)$$

Write  $N'_n(\cdot) = I_B N_n(\cdot)$ ,  $A'_n(\cdot) = I_B A_n(\cdot)$ ,  $A'(\cdot) = I_B A(\cdot)$ , where  $B$  is the indicator function of the set  $\{A(t) \leq M\}$ . Equation (14.2.18) will be established if we can show that  $\mathcal{P}(\{N'_n(t) > a\}) < \frac{2}{3}\epsilon$  (all  $a > a_0$ ).

From condition (ii) of the theorem we have immediately the property

(ii)'  $A'_n(\cdot) \rightarrow A(\cdot)$  ( $n \rightarrow \infty$ ) in probability,

so by a diagonal selection argument we can find a subsequence  $\{n'_k\}$  such that  $A'_{n'_k}(s) \rightarrow A'(s)$  a.s. for a countable set of values of  $s$ . But  $A'_{n'_k}(s)$  and  $A'(s)$  are both monotone in  $s$ , finite for finite  $s$ , so

$$\sup_{0 \leq s \leq t} |A'_{n'_k}(s) - A'(s)| \rightarrow 0 \quad (n \rightarrow \infty) \quad \text{a.s.}$$

Then identifying  $\{n_k\} = \{n'_k\}$ ,

$$\sup_{0 \leq s \leq t} |A'_n(s) - A'(s)| \rightarrow 0 \quad \text{in probability.} \quad (14.2.19)$$

By the continuity and  $\mathcal{F}_0^{(n)}$ -measurability of  $A'(\cdot)$ ,

$$T_n \equiv \inf\{s: A'_n(s) > A'(t) + 1\}$$

is an extended  $\mathcal{F}_{(-)}^{(n)}$ -stopping time, with the property implied by (14.2.19) that  $\mathcal{P}\{T_n \geq t\} \rightarrow 1$  as  $n \rightarrow \infty$ . Consequently, there exists  $\Delta$  such that, for all  $n$ ,  $\mathcal{P}\{|N'_n(t \wedge T_n) - N'_n(t)| > \Delta\} < \frac{1}{3}\epsilon$ . Now

$$\mathcal{P}\{N'_n(t \wedge T_n) > a + \Delta\} \leq \frac{\mathbb{E}[N'_n(t \wedge T_n)]}{a + \Delta} = \frac{\mathbb{E}[A'_n(t \wedge T_n)]}{a + \Delta} \leq \frac{M + 1}{a + \Delta} < \frac{\epsilon}{3}$$

for  $a$  sufficiently large and all  $n$ . Equation (14.2.18), and hence (14.2.17), now follow.

Assume there is given a subsequence  $\{n_k\}$  for which the distributions induced by  $N_n(\cdot)$  converge weakly to those of some limit point process  $N(\cdot)$ : our aim is to show that all such limit distributions coincide with those of a Cox process directed by  $A(\cdot)$ . This will follow from the characterization result in Theorem 14.6.I if we can show that there exists a version of  $N(\cdot)$  having  $A(\cdot)$  as its  $\mathcal{F}$ -compensator, where  $\mathcal{F} = \{\mathcal{F}_t\}$ ,  $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{F}_t^{(N)}$ , and  $\mathcal{F}_0 = \sigma\{A(s): 0 < s < \infty\}$ .

As in the first part of the proof, it is enough to fix some  $t < \infty$  and to assume that the subsequence is so chosen and the point process so modified that  $A_n(s) \rightarrow A(s)$  uniformly a.s. on  $(0, t]$  and that the compensators  $A_n(t)$  are a.s. uniformly bounded by finite  $M$  say.

Referring to the properties of  $\mathcal{F}$ -stable convergence of distributions, with the space  $\mathcal{X}$  of Definition A3.2.III taken to be  $\mathcal{N}_{(0,t]}^\#$  and  $\mathcal{F} = \{\mathcal{F}_t\}$  as above, it follows from Proposition A3.2.VI that a version of  $N$  can be defined on the (possibly enlarged) space  $(\Omega, \mathcal{E}, \mathcal{P})$  in such a way that if  $U \in \mathcal{F}_s$  for some  $s$  in  $(0, t]$ ,

$$\int_U N_n(s) \mathcal{P}(d\omega) \rightarrow \int_U N(s) \mathcal{P}(d\omega) \quad (14.2.20)$$

[see condition (iv) of Proposition A3.2.VI]. But the convergence of the compensators  $\{A_n(\cdot)\}$  implies a result corresponding to (14.2.20) with  $A_n$  and  $A$  replacing  $N_n$  and  $N$ , so we also have

$$\int_U [N_n(s) - A_n(s)] \mathcal{P}(d\omega) \rightarrow \int_U [N(s) - A(s)] \mathcal{P}(d\omega). \quad (14.2.21)$$

Before we can invoke the martingale property for the left-hand side here, we have to show that (14.2.21) continues to hold when  $U$  on the left-hand side is replaced by some approximating set  $U_n$  from  $\mathcal{F}_s^{(n)}$ . It is enough to take  $U$  to be a generating set  $V \cap W$ , where  $V \in \mathcal{F}_0$ ,  $W = \{\omega: N(\cdot, \omega) \in B, B \in \mathcal{B}(\mathcal{N}_{(0,s]}^\#)\}$ . Define  $U_n = V \cap W_n$ , where  $W_n$  is defined as  $W$  with  $N$  replaced by  $N_n$ . The properties of  $\mathcal{F}$ -stable convergence imply that  $\mathcal{P}(U \Delta U_n) \rightarrow 0$  (see Proposition A3.2.IV). Furthermore, the integrands at (14.2.21) are uniformly integrable because of the boundedness of their second moments as follows from

$$\begin{aligned} \mathbb{E}([N_n(s)]^2) &\leq \mathbb{E}([N_n(t)]^2) = \mathbb{E}\left[\left(\sum_i \Delta N_n(t_i)\right)^2\right] \\ &= \mathbb{E}\left[\sum_i [\Delta N_n(t_i)]^2 + 2 \sum_i \Delta N_n(t_i) \sum_{t_j < t_i} \Delta N_n(t_j)\right] \\ &= \mathbb{E}\left[N_n(t) + 2 \int_0^t N_n(s-) dN_n(s)\right], \quad \text{because } N_n \text{ is simple,} \\ &= \mathbb{E}\left[A_n(t) + 2 \int_0^t N_n(s-) dA_n(s)\right] \\ &\leq \mathbb{E}[A_n(t) + 2N_n(t) A_n(t)] \leq M + 2M^2. \end{aligned}$$

Using these two facts, we may replace (14.2.21) by

$$\int_{U_n} [N_n(s) - A_n(s)] \mathcal{P}(d\omega) \rightarrow \int_U [N(s) - A(s)] \mathcal{P}(d\omega). \quad (14.2.22)$$

Take  $s = t$  in this equation, but in the definition of  $W_n$ , and hence of  $U_n$ , take  $B \in \mathcal{B}(\mathcal{N}_{(0,s]}^\#)$  with  $s < t$ . Then the martingale property of  $N_n$  implies

$$\int_{U_n} [N_n(t) - A_n(t)] \mathcal{P}(d\omega) = \int_{U_n} [N_n(s) - A_n(s)] \mathcal{P}(d\omega).$$

Using (14.2.22), the same equality holds with  $N_n$  and  $A_n$  replaced by  $N$  and  $A$ , which is just the conditional expectation property identifying  $N(\cdot) - A(\cdot)$  as a martingale, and hence  $A(\cdot)$  as the  $\mathcal{F}$ -compensator for  $N(\cdot)$ .  $\square$

This result is applicable, for example, to ‘dependent thinnings’ where the probability of a point being deleted is allowed to depend on either or both of the past of the original (unthinned) point process and of the history of previous deletions. To see this, consider the special case of Lemma 14.2.VII in which  $X$  is a  $\{0, 1\}$ -valued process so that  $\eta$  at (14.2.12) is a thinning of the process  $N$ , with

$$\zeta(t_i) = \Pr\{t_i \text{ from } N \text{ retained in } \eta\}.$$

Applying the theorem above leads to the corollary below.

**Corollary 14.2.IX.** *Let  $\{N_n, \mathcal{F}^{(n)}: n = 1, 2, \dots\}$  be a sequence of simple point processes and associated histories, all defined on a common probability space  $(\Omega, \mathcal{E}, \mathcal{P})$  and such that  $N_n$  is  $\mathcal{F}^{(n)}$ -predictable for each  $n$ . Let  $\{X_n(\cdot)\}$  be a family of  $\mathcal{F}^{(n)}$ -adapted  $\{0, 1\}$ -valued processes, and let  $\eta_n(\cdot)$ ,  $\zeta_n(\cdot)$ , and  $\alpha_n(\cdot)$  be defined as at (14.2.12) and (14.2.13). If for all  $t > 0$*

$$\alpha_n(t) \rightarrow A(t) \quad (n \rightarrow \infty) \quad \text{in probability,} \quad (14.2.23)$$

*where  $A(\cdot)$  has continuous paths and is  $\mathcal{F}_0^{(n)}$ -measurable for each  $n$ , then the thinned processes  $\eta_n(\cdot)$  converge in distribution to a Cox process directed by  $A(\cdot)$ .*

**EXAMPLE 14.2(b) A point process with controlled thinning.** In quality control and similar contexts, the detection and elimination of errors may be regarded as a thinning operation on an original stream of errors. The error rate in production may vary between batches, so that in order to achieve a uniform low level of errors in the output, screening may need to be intensified according to current estimates of the error rate.

As a crude model, suppose the actual error rate equals some quantity  $\lambda \equiv \lambda(\omega)$  per item produced; here,  $\lambda$  is constant within a batch but may vary between batches, and  $\omega \in \Omega$  for some probability space  $(\Omega, \mathcal{E}, \mathcal{P})$ . Suppose that probabilistic screening of items occurs, the  $\{0, 1\}$ -valued r.v.  $X(t_i)$  indicating elimination of ( $X(t_i) = 0$ ) or failure to detect ( $X(t_i) = 1$ ) an error at  $t_i$ , and  $E(X(t_i) \mid \mathcal{F}_{t_i-}) = \zeta(t_i)$  for some process  $\zeta(\cdot)$ . Let  $Y(t)$  and  $Z(t)$  denote, respectively, the numbers of errors detected and undetected in time  $(0, t)$ , so that assuming the time-scale unit is chosen according to the rate of production of items (whether faulty or not),  $[Y(t) + Z(t)]/t \approx \lambda$  for large  $t$  assuming ergodicity. The aim is to choose  $\zeta(\cdot)$  (and this choice is assumed to be available to the producer) in such a way that  $Z(t)/t \lesssim \gamma$ , asymptotically in  $t$  ( $=$  long batch run), for some small residual error rate  $\gamma$  per item produced. One such possibility is to have

$$1 - \zeta(t, \omega) = \frac{\gamma t}{\gamma t + Y^*(t-, \omega)}, \quad Y^*(t, \omega) \equiv \max(1, Y(t, \omega)), \quad (14.2.24)$$

noting that this implies that  $\zeta(t, \omega) \rightarrow 1$  as  $t \rightarrow 0$  and that  $\zeta(t, \omega)$  remains closer to 1 until (if ever)  $Y^*(t-, \omega)/(\gamma t)$  is no longer significantly larger than 1; if  $\lambda \gg \gamma$  we expect intuitively that this would not occur, because the aim is to make  $Z(t, \omega)/(\gamma t)$  asymptotically about 1 and so  $Y^*(t, \omega)/(\gamma t)$  is like  $(\lambda - \gamma)/\gamma$ .

We therefore study the asymptotic behaviour in  $Z(\cdot)$  for  $\gamma \rightarrow 0$  by considering a sequence of schemes, indexed by  $\gamma$ , and show how Corollary 14.2.IX may be applicable. Suppose that the realizations  $N(\cdot)$  are kept fixed and that each of a sequence of thinning operations is applied to the same realizations, so that  $N_\gamma(t) = N(t)$  and  $Y_\gamma(t)$  and  $Z_\gamma(t)$  are the result of a dependent thinning via a  $\{0, 1\}$ -valued process  $X_\gamma(t)$  for which  $E[X_\gamma(t_i) \mid \mathcal{F}_{t_i-}] = \zeta_\gamma(t_i)$  and  $\zeta_\gamma(\cdot)$  is related to  $Y_\gamma(\cdot)$  as at (14.2.24). Finally, change the time scale by setting  $\tau = \gamma t$  and defining the processes

$$\begin{aligned}\tilde{Y}_\gamma(\tau) &= Y_\gamma(\tau/\gamma), & \tilde{Z}_\gamma(\tau) &= Z_\gamma(\tau/\gamma), \\ \tilde{\zeta}_\gamma(\tau) &= \zeta_\gamma(\tau/\gamma) = 1 - \frac{\tau}{\tau + \tilde{Y}_\gamma(\tau/\gamma-, \omega)}.\end{aligned}$$

Set up a scheme of histories  $\mathcal{F}^{(\gamma)}$  as follows. Take  $\mathcal{F}_0^{(\gamma)}$  to include the information on  $\tilde{N}_\gamma(t) = N_\gamma(\tau/\gamma)$  for every  $\gamma > 0$ , and define  $\mathcal{F}_\tau^{(\gamma)} = \mathcal{F}_0^{(\gamma)} \vee \sigma\{\tilde{X}_\gamma(s): 0 < s \leq \tau\}$ . This choice ensures that for each  $\gamma$  we have  $\mathcal{F}^{(\gamma)}$ -predictability of  $\tilde{N}_\gamma$  and that  $\tilde{X}_\gamma$  is  $\mathcal{F}^{(\gamma)}$ -adapted, leading to the expression for the  $\mathcal{F}^{(\gamma)}$ -compensator for the thinned (output) process  $\tilde{Z}_\gamma$ ,

$$\tilde{A}_\gamma(\tau) = \int_0^\tau [1 - \tilde{\zeta}_\gamma(u)] d\tilde{N}_\gamma(u) = \int_0^{\tau/\gamma} [1 - \zeta_\gamma(u)] dN(u).$$

We must now determine whether the processes  $\tilde{A}_\gamma(\cdot)$  converge in probability as  $\gamma \rightarrow 0$ , and whether the limit function is the compensator of a Poisson process (because this is the plausible limit process under thinning). The triangle inequality gives

$$\begin{aligned}|\tilde{A}_\gamma(\tau) - \tau| &\leq \left| \frac{\gamma}{\lambda} \int_0^\tau d\tilde{N}_\gamma(v) - \tau \right| + \left| \frac{\gamma}{\lambda} \int_0^{\tau'} d\tilde{N}_\gamma(v) \right| \\ &\quad + \left| \int_0^{\tau'} [1 - \tilde{\zeta}_\gamma(v)] d\tilde{N}_\gamma(v) \right| + \left| \int_{\tau'}^\tau [1 - \tilde{\zeta}_\gamma(v) - \frac{\gamma}{\lambda}] d\tilde{N}_\gamma(v) \right|, \quad (14.2.25)\end{aligned}$$

and as indicated in Exercise 14.2.7, provided  $N(t, \omega)/t \rightarrow \lambda$  a.s. ( $t \rightarrow \infty$ ),  $\tau'$  can be chosen here to make each of these four terms converge to zero in probability as  $\gamma \rightarrow 0$ .

We therefore conclude that, in the asymptotic sense indicated by change of time scale, the thinning procedure will be effective in reducing the output error rate to the required low level  $\gamma$ , irrespective of  $\lambda$ , and that the resultant process will be asymptotically Poisson in character.

Observe that although conceptually we regard  $N$  as being known *a priori* [i.e.,  $\sigma(N) \subseteq \mathcal{F}_0^{(\gamma)}$ ],  $\zeta(\cdot)$  is in fact  $\mathcal{H}_{(-)}^{(Y)}$ -adapted, where  $\mathcal{H}^{(Y)}$  is the internal history of the observed process of errors.  $\square$

## Exercises and Complements to Section 14.2

14.2.1 *One-point MPP compensator.* Consider a one-point MPP determined by the bivariate distribution  $\Gamma(\cdot, \cdot)$  on  $\mathcal{B}_{\mathbb{R}_+} \times \mathcal{B}_{\mathcal{K}}$  with marginal distribution function  $F_g(t) = \Gamma([0, t], \mathcal{K})$ , and conditional distribution  $F(K | t) = \Pr\{Y \in K | X = t\}$  as in Theorem 14.2.IV(b). Let  $H_g(t) = \int_0^t dF_g(u)/[1 - F_g(u-)]$ . Show that  $A(t, K, \omega) = \int_0^{t \wedge X} F(K | u) dH_g(u)$  is an MPP compensator for the one-point process as in Definition 14.2.III.

Extend to the case that  $\Gamma$  is itself determined by a prior  $\sigma$ -algebra  $\mathcal{F}_0$ .

14.2.2 (*Continuation*). Use the previous exercise to justify the representations in (14.2.9) in Theorem 14.2.IV(b).

[*Hint:* The representation of  $A_g(t)$  repeats the argument leading to Theorem 14.1.IV. The essential remaining step is to establish the predictability of  $F(K | t, \omega)$  for fixed  $K \in \mathcal{B}_{\mathcal{K}}$ : this is done sequentially by successive conditioning on  $\sigma$ -algebras  $\mathcal{F}_{(n-1)}$ , much as in the discussion around (14.1.9).]

14.2.3 (*Continuation*). Check the assertion in Theorem 14.2.IV(c) by constructing the fidi distributions for the MPP in terms of the components  $\{G_n\}$  and  $\{F_n\}$  of the compensator, and *vice versa*.

14.2.4 Starting from the definition of the mark-predictable  $\sigma$ -algebra, argue as for Proposition 14.2.I to prove Proposition 14.2.V. Similarly, extend the Radon–Nikodym argument for Proposition 14.2.II to prove Proposition 14.2.VI.

14.2.5 Regard a marked cumulative process  $\xi$  as being measure-valued in the sense that for each  $t > 0$ ,  $\xi(t, \cdot)$  is a measure on  $\mathcal{B}_{\mathcal{K}}$ . Consider the latter as an element in the ordered space of measures in which the sequence of measures is monotonic increasing in  $t$ . Investigate whether the projection Propositions 14.2.V–VI remain valid in this measure-valued context.

14.2.6 (a) To complete the proof of Lemma 14.2.VII, extend the discussion from the one-point case given after (14.2.14–15) to any simple point process  $N$ .

(b) Let  $\mathcal{F}$  be a history,  $\xi$  an  $\mathcal{F}$ -predictable cumulative process, and  $X$  a nonnegative  $\mathcal{F}$ -adapted process with finite first moment. Show that if there exists an  $\mathcal{F}$ -predictable version of the conditional expectation  $\zeta(t) \equiv E[X(t) | \mathcal{F}_{t-}]$ , then the process  $\eta(t) \equiv \int_0^t X(u) d\xi(u)$  has  $\mathcal{F}$ -compensator  $\alpha(t) \equiv \int_0^t \zeta(u) d\xi(u)$ .

[*Hint:* The required measurability and predictability properties are preserved in (a) because only linear combinations and a.s. limits are involved. To establish the compensator property in (b), use Fubini's theorem and the argument in the text to show that for every  $A_s \in \mathcal{F}_s$  and  $s < t$ ,

$$\int_{A_s} \mathcal{P}(d\omega) \int_s^t (X(u) - E[X(u) | \mathcal{F}_{u-}]) du = 0.]$$

14.2.7 In Example 14.2(b), show that each of the four terms on the right-hand side of (14.2.25) converges in probability to zero as  $\gamma \rightarrow 0$  on the assumption that  $N(t)/t$  converges a.s. to some limit  $\lambda \equiv \lambda(\omega) \in \mathcal{F}_0^{(\gamma)}$  for all  $\gamma$ .

[*Hint:* The convergence is shown directly for the first three terms; for the last, investigate  $\sup_{t' < v < t} |1 - \tilde{\zeta}_\gamma(v) - \gamma/\lambda|$  via  $Y_\gamma(t)/t \leq N(t)/t$  a.s. in one direction and  $\liminf_{t \rightarrow \infty} Y_\gamma(t)/t$  in the other direction.]

14.2.8 Extend Exercises 14.1.12–13 to the marked case by exhibiting the form of the  $\mathcal{H}$ -marked compensator in terms of suitably defined conditional distributions and intensity functions.

14.2.9 Prove Theorem 14.2.VIII by imitating the construction and proof of Watanabe's theorem at Proposition 14.6.I.

### 14.3. Conditional Intensities

In this section we examine in more detail the properties of both simple and marked point processes with absolutely continuous compensator, extending results for conditional intensities constructed via hazard functions as in Sections 7.2–3 and Corollary 14.1.V. Note that for MPPs we prefer to define the conditional intensity as a density not only with respect to time but also with respect to a reference measure in the mark space, rather than working with families of conditional intensities indexed by Borel sets in the mark space. The main reason for this choice is that in applications, it is usually the density form that suggests itself most naturally. The main topics we consider are conditions for the existence of conditional intensities and, when they exist, the selection of predictable versions. We consider first processes defined on a half-line  $\mathbb{R}_+$  as in Sections 14.1–2, and then look at *complete intensities* for stationary processes on the whole line  $\mathbb{R}$ .

We start from the following definitions, covering successively the simple and marked cases. We again assume the existence of first moment measures, including in the marked case the first moment measure of the ground process.

**Definition 14.3.I.** (a) Let  $\xi$  be a cumulative process with history  $\mathcal{F}$  and  $\mathcal{F}$ -compensator  $A$ . An  $\mathcal{F}$ -intensity for  $\xi$  is any  $\mathcal{F}$ -adapted process  $\lambda^*(u, \omega)$ , measurable with respect to the product  $\sigma$ -algebra  $\mathcal{B}_{\mathbb{R}_0^+} \times \mathcal{E}$ , and such that a.s. for all  $t$ ,

$$A(t, \omega) = \int_0^t \lambda^*(u, \omega) du. \quad (14.3.1a)$$

(b) Let  $\xi$  be a marked cumulative process, with mark space  $(\mathcal{K}, \mathcal{B}_{\mathcal{K}}, \ell_{\mathcal{K}})$  and  $\mathcal{F}$ -compensator  $A$ . An  $\mathcal{F}$ -intensity for  $\xi$  is any  $\mathcal{F}$ -adapted process  $\lambda^*(u, \kappa, \omega)$ , measurable with respect to the three-fold product  $\sigma$ -algebra  $\mathcal{B}_{\mathbb{R}_0^+} \times \mathcal{B}_{\mathcal{K}} \times \mathcal{E}$ , and such that, a.s. for all  $0 < t < \infty$  and  $K \in \mathcal{B}_{\mathcal{K}}$ ,

$$A(t, K, \omega) = \int_{(0, t] \times K} \lambda^*(u, \kappa, \omega) du \ell_{\mathcal{K}}(d\kappa). \quad (14.3.1b)$$

In other words, we require  $\int_0^t \lambda^*(u, \omega) du$  to be  $\mathcal{P}$ -indistinguishable from  $A(t, \omega)$ , with an equivalent statement in the marked case. The existence of the integrals on an a.s. basis is not automatic for a general measurable process  $\lambda^*$ , but is so if  $\lambda^*$  is  $\mathcal{F}$ -progressively measurable (see Section A3.3 for

the definition of progressively measurable unmarked processes; the extension to MPPs follows as in the definition of predictable MPPs). In fact, it is clearly desirable, whenever possible, to choose a version of the intensity that is predictable, because the intensity is used to calculate risks before rather than after the occurrence of points. The projection Propositions 14.2.I–II and 14.2.V–VI help to clarify the circumstances under which this is possible. Regard the absolute continuity as referring not to the a.s. properties of the compensator for fixed  $\omega$  but rather to a property of the Campbell measure on the appropriate product space. The Radon–Nikodym derivative on the restriction of the product  $\sigma$ -algebra to the predictable  $\sigma$ -algebra then provides us with the predictable version that we want.

These remarks are formalized below, where we state the results for the general marked case; those for simple point processes follow on collapsing the mark space to a single point (see Exercise 14.3.1 for an explicit statement). Unmarked point processes which are not simple can be treated as MPPs with integer marks.

**Proposition 14.3.II.** *Let  $\xi$  be an  $\mathcal{F}$ -adapted marked cumulative process with  $\mathcal{F}$ -compensator  $A$ .*

(a) *A necessary and sufficient condition for the existence of an  $\mathcal{F}$ -mark-predictable intensity  $\lambda^*$  for  $A$  is that the marked Campbell measure at (14.2.5) be absolutely continuous with respect to  $\ell \times \ell_{\mathcal{K}} \times \mathcal{P}$  on  $\Psi_{\mathcal{K}}^{\mathcal{F}}$ .*

(b) *When the condition in (a) is satisfied, the Radon–Nikodym derivative  $dC_{\xi}(t, K, \omega)/d(\ell \times \ell_{\mathcal{K}} \times \mathcal{P}) = \lambda^*(t, \kappa, \omega)$  is  $\mathcal{F}$ -mark-predictable, and provides an  $\mathcal{F}$ -mark-predictable version of the conditional intensity. This version then coincides, except possibly on a set of  $(\ell \times \ell_{\mathcal{K}} \times \mathcal{P})$ -measure zero, with any  $\mathcal{F}$ -intensity for  $\xi$  satisfying Definition 14.3.I(b).*

(c) *When an  $\mathcal{F}$ -mark-predictable intensity exists, the ground process has an  $\mathcal{F}$ -predictable intensity*

$$\lambda_g(t, \omega) = \int_{\mathcal{K}} \lambda^*(t, \kappa, \omega) \ell_{\mathcal{K}}(d\kappa), \quad (14.3.2a)$$

and the  $\mathcal{F}$ -predictable version of the conditional mark distribution  $F(\cdot | t, \omega)$ , introduced in (14.2.7), is  $(\ell \times \mathcal{P})$ -a.e. absolutely continuous with respect to  $\ell_{\mathcal{K}}$ , with density  $f(\kappa | t, \omega) = \lambda^*(t, \kappa, \omega)/\lambda_g(t, \omega)$   $\ell_{\mathcal{K}}$ -a.e. so that for  $K \in \mathcal{B}_{\mathcal{K}}$ ,

$$F(K | t, \omega) = \int_{\mathcal{K}} f(\kappa | t, \omega) \ell_{\mathcal{K}}(d\kappa) = \frac{\int_{\mathcal{K}} \lambda^*(t, \kappa, \omega) \ell_{\mathcal{K}}(d\kappa)}{\int_{\mathcal{K}} \lambda^*(t, \kappa, \omega) \ell_{\mathcal{K}}(d\kappa)} \quad (\ell \times \mathcal{P})\text{-a.e.} \quad (14.3.2b)$$

PROOF. Suppose first that an  $\mathcal{F}$ -intensity exists. Consider any basic set  $(s, t] \times K \times U$ , with  $s < t$ ,  $K \in \mathcal{B}_{\mathcal{K}}$  and  $U \in \mathcal{F}_s$ , for the mark-predictable  $\sigma$ -algebra. Now  $(s, t] \times K \times U$  is an element of the threefold product  $\sigma$ -algebra, so it follows from Definition 14.3.I(b) that

$$\int_{(s, t] \times K \times U} \lambda^*(u, \kappa, \omega) du \ell_{\mathcal{K}}(d\kappa) \mathcal{P}(d\omega) = \mathbb{E}[I_U(A(t, K) - A(s, K))].$$



By the martingale property embodied in (14.2.3), the right-hand side here equals

$$\mathbb{E}[I_U(N(t, K) - N(s, K))] = C_{\mathcal{P}}((s, t] \times K \times U).$$

It follows that the Campbell measure  $C_{\mathcal{P}}$  is absolutely continuous on  $\Psi^{\mathcal{F}}$ .

Now suppose conversely that absolute continuity as above holds. Then the Radon–Nikodym theorem implies the existence of at least one version of the density that is itself  $\Psi_{\mathcal{K}}^{\mathcal{F}}$ -measurable. For this version, the above equations continue to hold and show that the density is in fact an  $\mathcal{F}$ -predictable intensity. This version is *a fortiori* measurable with respect to the full product  $\sigma$ -algebra, so the Campbell measure is absolutely continuous with respect to the product measure on this larger  $\sigma$ -algebra also. The uniqueness results associated with the Radon–Nikodym theorem then imply that this predictable version should differ from any other  $\mathcal{F}$ -intensity at most on a subset of  $(\ell \times \ell_{\mathcal{K}} \times \mathcal{P})$ -measure zero.

When the marked Campbell measure is absolutely continuous, the Campbell measure for the ground process, obtained by setting  $K = \mathcal{K}$  in the definition, is likewise absolutely continuous with respect to  $\ell \times \mathcal{P}$  on the twofold product space  $\mathcal{X} \times \Omega$ , and so a predictable version of the intensity exists also for the ground process. Indeed, the integral  $\lambda_g(t, \omega) = \int_{\mathcal{K}} \lambda(t, \kappa, \omega) \ell_{\mathcal{K}}(d\kappa)$  is still predictable, satisfies the requirements of a density for the Campbell measure of the ground process, and thus coincides with any other version of its predictable intensity  $(\ell \times \mathcal{P})$ -a.e. Finally, the ratio  $\int_{\mathcal{K}} \lambda(t, \kappa, \omega) \ell_{\mathcal{K}}(d\kappa) / \lambda_g(t, \omega)$  is predictable and satisfies the defining equation (14.2.6) for the predictable version of the conditional mark distribution.  $\square$

Note that in the marked case it follows from standard results on product spaces (see the discussion around Fubini's Theorem A1.5.I) that for each fixed  $\kappa$ , the function  $\lambda^*(t, \kappa, \omega)$  is  $\mathcal{F}$ -predictable, and that for fixed  $K$ ,  $\lambda_K^*(t, \omega) = \int_K \lambda^*(t, \kappa, \omega) \ell_{\mathcal{K}}(d\kappa)$  is an  $\mathcal{F}$ -conditional intensity for the point process  $N_K(\cdot) = N(\cdot \times K)$ . Note also that throughout the discussion we require the marked conditional intensity to be a density not only in time but also with respect to the reference measure  $\ell_{\mathcal{K}}$  on the mark space; for an alternative approach, aimed at omitting this last requirement, see Exercise 14.3.2.

The proposition implies that the existence of any version of the  $\mathcal{F}$ -conditional intensity, being enough to ensure absolute continuity of the Campbell measure, is enough to ensure also the existence of an  $\mathcal{F}$ -predictable version of the intensity. This raises the question of the relation between the two intensities. In this situation the construction in Proposition 14.3.II reduces to the construction of the predictable projection (as in Propositions 14.2.V–VI) of the initial version of the conditional intensity.

As a specific example, let  $\mathcal{F}$  and  $\mathcal{G}$  be two histories for a process  $(\xi, \mathcal{F})$  with  $\mathcal{G} \subseteq \mathcal{F}$  in the sense that  $\mathcal{G}_t \subseteq \mathcal{F}_t$  for every  $t \geq 0$ . A typical *filtering problem* is to find the  $\mathcal{G}$ -intensity for  $\xi$  given its  $\mathcal{F}$ -intensity. Again, this problem is readily solved by appeal to Propositions 14.3.II and 14.2.V: the  $\mathcal{G}$ -predictable projection of the  $\mathcal{F}$ -intensity is a  $\mathcal{G}$ -intensity. Inasmuch as this resolves the

question of existence of a  $\mathcal{G}$ -intensity, actually computing the projection from its definition as a Radon–Nikodym derivative in the product space  $\mathbb{R}_+ \times \Omega$  may be difficult. A simpler relation such as

$$\tilde{\lambda}^{\mathcal{G}}(t, \kappa) = \mathbb{E}[\lambda^{\mathcal{F}}(t, \kappa) \mid \mathcal{G}_{t-}] \quad (14.3.3)$$

is intuitively plausible, but the difficulty is that there is no guarantee in general that the right-hand side has a version that is even measurable, let alone predictable. Fortunately, this is usually more of a theoretical difficulty than a practical problem, because in most practical applications a version of (14.3.3) can be found with continuity properties which allow it to be identified as a predictable version of the  $\mathcal{G}$ -intensity. The lemma below covers the general situation.

**Lemma 14.3.III.** *Suppose that the intensity  $\tilde{\lambda}^{\mathcal{G}}(t, \kappa, \omega)$  at (14.3.3), with  $\mathcal{G} \subseteq \mathcal{F}$ , admits a version whose trajectories are  $(\ell_{\mathcal{K}} \times \mathcal{P})$ -a.e. left-continuous in  $t$ , or, more generally, which is  $\mathcal{G}$ -mark-predictable. Then this version is also a version of the  $\mathcal{G}$ -mark-predictable intensity.*

PROOF. Suppose such a version as described exists,  $\mu(t, \kappa, \omega)$  say, so that it necessarily satisfies

$$\int_B \int_K \mu(u, \kappa, \omega) \ell_{\mathcal{K}}(d\kappa) \mathcal{P}(d\omega) = \int_B \int_K \lambda^{\mathcal{F}}(u, \kappa, \omega) \ell_{\mathcal{K}}(d\kappa) \mathcal{P}(d\omega)$$

for all  $B \in \mathcal{G}_s \subseteq \mathcal{F}_s$ ,  $K \in \mathcal{K}$ , and  $u > s$ , by definition of the  $\mathcal{F}$ -intensity and the inclusion  $\mathcal{G}_u \supseteq \mathcal{G}_s$ . From the definition of  $\mu$  the left-hand side is measurable in  $u$  and can be integrated over  $(s, t]$  to give

$$\int_{(s,t] \times B \times K} \mu(u, \kappa, \omega) du \ell_{\mathcal{K}}(d\kappa) \mathcal{P}(d\omega) = \int_{(s,t] \times B \times K} \lambda^{\mathcal{F}}(u, \omega) du \ell_{\mathcal{K}}(d\kappa) \mathcal{P}(d\omega).$$

These sets  $(s, t] \times B \times K$  with  $B \in \mathcal{G}_s$  generate  $\Psi^{\mathcal{G}}$ , so this is just the assertion that  $\mu$  is a version of the  $\mathcal{G}$ -predictable projection of  $\lambda^{\mathcal{F}}$ .  $\square$

Most commonly, the coarser  $\sigma$ -algebra  $\mathcal{G}$  is the internal history  $\mathcal{H}$ , and the problem is to find the intensity for the internal history in terms of a larger history, such as some intrinsic history, for which the intensity is more easily calculated. The computation is effectively a form of Bayes' theorem; the following example is a convenient illustration.

**EXAMPLE 14.3(a) Mixed Poisson process intensities.** Suppose there is given a realization  $t_1, \dots, t_N$  on  $(0, t]$  of a mixed Poisson process with rate parameter  $\mu$  that is treated as a random variable with d.f.  $F$  on  $(0, \infty)$ . Take  $\mathcal{F}$  as the intrinsic history, so  $\mathcal{F}_t = \sigma\{\mu\} \vee \mathcal{H}_t$ . Then the  $\mathcal{F}$ -intensity is just  $\mu$  itself.

To find the  $\mathcal{H}$ -intensity we investigate the form of  $\mathbb{E}(\mu \mid \mathcal{H}_{t-})$ , and assume for simplicity that  $F$  has a density  $f$ . The point about the internal history is that it has a simple structure and we can appeal to the existence of regular

conditional distributions expressed in the form of densities. Bayes' theorem here implies for the conditional density of  $\mu$  given  $\mathcal{H}_t$

$$p(\mu \mid \mathcal{H}_{t-}) = \frac{p(\mathcal{H}_{t-} \mid \mu) f(\mu)}{\int_0^\infty p(\mathcal{H}_{t-} \mid \mu) f(\mu) d\mu} = \frac{(\mu t)^{N'} e^{-\mu t} f(\mu)}{\int_0^\infty (\mu t)^{N'} e^{-\mu t} f(\mu) d\mu},$$

where  $N' = N(t-, \omega)$ , and

$$E(\mu \mid \mathcal{H}_{t-}) = \frac{\int_0^\infty \mu^{N'+1} e^{-\mu t} f(\mu) d\mu}{\int_0^\infty \mu^{N'} e^{-\mu t} f(\mu) d\mu}.$$

For example, if  $\mu$  has the exponential density  $\alpha e^{-\alpha\mu}$ , then

$$E(\mu \mid \mathcal{H}_{t-}) = \frac{N' + 1}{t + \alpha} \equiv \frac{N(t-, \omega) + 1}{t + \alpha}, \quad (14.3.4)$$

and because this function is a.s. left-continuous in  $t$ , it can be taken as a version of the  $\mathcal{H}$ -intensity.  $\square$

For stationary point processes on the whole line  $\mathbb{R}$  it is natural to use a form of conditional intensity that depends on the entire past rather than on the past since a fixed origin. Strictly speaking, such processes lie outside the framework we have considered hitherto, because our definition of compensator was restricted to processes on the half-line  $\mathbb{R}_+$ . However, the definition is readily extended to processes on  $\mathbb{R}$ . Suppose  $\xi(t)$  is an  $\mathcal{F}^\dagger$ -adapted process on  $(-\infty, \infty)$ , where  $\mathcal{F}^\dagger$  is a history on  $\mathbb{R}$  (earlier, our histories have been defined as families  $\{\mathcal{F}_t: t \in \mathbb{R}_+\}$ ). Then for any real  $t_1$  and  $t \geq t_1$  we can consider the process  $\xi_1(t)$  on  $[t_1, \infty)$  with respect to the intrinsic history  $\mathcal{G}_t^1 = \mathcal{F}_{t_1}^\dagger \vee \mathcal{F}_{(t_1, t]}^\dagger$ , and define for it a compensator  $A^1(t)$  using the definition for a process on  $\mathbb{R}_+$ . Similarly we can define a compensator  $A^2(t)$  for  $t \geq t_2 > t_1$  with respect to the intrinsic history  $\mathcal{G}_t^2 = \mathcal{F}_{t_2}^\dagger \vee \mathcal{F}_{(t_2, t]}^\dagger$ . Now for  $i = 1, 2$ ,  $\mathcal{G}_t^i = \mathcal{F}_t^\dagger$  for  $t \geq t_i$ , from which it follows that the compensators coincide a.s. on the segment  $[t_2, \infty)$  where they overlap, and hence, by taking a sequence of values  $t_i$  extending back to  $-\infty$ , that there exists a unique process  $A(t)$  on  $(-\infty, \infty)$  such that  $A(t)$  coincides a.s. with  $A^1(t)$  for any particular choice of  $t_1$ .

From this point we can proceed as in the case of a half-line, defining an  $\mathcal{F}^\dagger$ -conditional intensity as any  $\mathcal{F}^\dagger$ -measurable process acting as a density for the compensator  $A(t)$  defined above. To preserve a distinctive notation, we use  $\lambda^\dagger(t)$  to denote such a conditional intensity, and call it a *complete conditional intensity* (continuing if need be to use  $\lambda^\mathcal{F}$  to notate the specific history). Similarly in the marked case, we can write  $\lambda^\dagger(t, \kappa) = \lambda_g^\dagger(t) f(\kappa \mid t)$  for the ground intensity and conditional mark distribution relative to a complete history. The necessary and sufficient condition for the existence of a complete intensity is the absolute continuity of the Campbell measure relative to the product measure  $\ell \times \mathcal{P}$  on the predictable  $\sigma$ -algebra on the full space  $\mathbb{R} \times \Omega$ , with

a corresponding extension in the marked case. When this condition is satisfied the Radon–Nikodym derivative provides a predictable version of the complete intensity. As in the earlier discussion, we can always obtain a predictable version from any other complete intensity by projection onto the predictable  $\sigma$ -algebra, and in the examples we assume that such a choice has been made.

EXAMPLE 14.3(b) *Renewal process with density* [continued from Example 14.1(e)]. Here and for the Wold process, the complete intensity function has the form of the hazard function given in Example 14.1(e) but without the correction term due to the choice of a fixed origin; that is, for the renewal process,  $\lambda^\dagger(t) = h(t - t_{N(t-)})$ , where  $t_{N(t-)} = \sup\{t_i: t_i(\omega) < t\}$ . This illustrates the fact that the complete intensity function is frequently simpler analytically, as well as being amenable to probabilistic study. See Exercise 14.3.4 for the Wold process.  $\square$

EXAMPLE 14.3(c) *Hawkes process* [continued from Examples 6.3(b), 7.2(d) and 8.5(d)]. This process is a stationary Poisson cluster process with clusters that are finite branching processes described by an offspring intensity measure that has support in  $\mathbb{R}_+$  and density function  $\mu(\cdot)$  of total mass  $\nu < 1$ . From this Poisson character and because the support of  $\mu$  is contained by the half-line, the conditional behaviour of the process is very simple: given the history up to time  $t$ , the process of births in  $[t, \infty)$  to individuals themselves born before  $t$  is conditionally a Poisson process with intensity at time  $t + u$  given by  $\lambda + \sum_{t_i < t} \mu(t - t_i + u)$ . Of course, any individuals born in  $[t, t + u)$  will also contribute to the total birth rate. However, knowing from Example 6.3(c) that the expected number of births on a finite interval is finite, it follows from the above remark that

$$\lim_{u \rightarrow 0} \frac{E[N(t+u) - N(t) \mid \mathcal{H}_t]}{u} = \lambda + \sum_{t_i \leq t} \mu(t - t_i), \quad (14.3.5)$$

which suffices to establish the right-hand side as a version of the conditional intensity with respect to the internal history  $\mathcal{H}^\dagger = \{\mathcal{H}_t: t \in \mathbb{R}\}$ . In practice,  $\mu$  is usually continuous apart from a possible jump at zero, so by either defining  $\mu(0) = 0$  or restricting the summation at (14.3.5) to birth epochs  $t_i < t$ , we obtain a left-continuous version that can therefore be taken as the required predictable version of the complete intensity function.

As already remarked in Example 8.5(d), the linear representation here has exactly the same form as that derived from the second-order theory of Chapter 8 when the Hawkes process has rational spectral density. In this case at least, as for Gaussian processes, the ‘best’ predictors are linear, at least where prediction of the intensity is in view.  $\square$

We proceed to a more detailed study of the complete intensity function for stationary processes. For simplicity we consider unmarked processes with just their internal histories, so that we can take for  $(\Omega, \mathcal{E})$  the canonical

space  $(\mathcal{N}_{\mathbb{R}}^{\#}, \mathcal{B}(\mathcal{N}_{\mathbb{R}}^{\#}))$ , with the history  $\mathcal{H}^{\dagger}$  consisting of the  $\sigma$ -algebras  $\mathcal{H}_t^{\dagger} = \sigma\{N(u) - N(s) : -\infty < s < u \leq t\}$  for  $-\infty < t < \infty$ .

Let us first consider the effect of stationarity on the conditional intensity. As in Proposition 13.2.I, stationarity of the underlying measure  $\mathcal{P}$  implies invariance of the Campbell measure under the transformations  $\Theta_u: (x, \xi) \mapsto (x - u, S_u \xi)$ . Observing that these transformations commute with the operation of extracting the Radon–Nikodym derivative, we see that when a predictable complete intensity exists, it should satisfy

$$\frac{d(C_{\mathcal{P}})}{d(\ell \times \mathcal{P})} = \frac{d(\Theta_u(C_{\mathcal{P}}))}{d(\ell \times \mathcal{P})} = \Theta_u \left( \frac{d(C_{\mathcal{P}})}{d(\ell \times \mathcal{P})} \right),$$

so that

$$\lambda^{\dagger}(t, \xi) = \Theta_u \lambda^{\dagger}(t, \xi) = \lambda^{\dagger}(t - u, S_u \xi) = \lambda^{\dagger}(0, S_t \xi). \quad (14.3.6)$$

Observe that in our notation  $\lambda^{\dagger}(t, \xi)$  we have chosen to emphasize the dependence of the intensity on the sample realization  $\xi$  (in the canonical space as we are assuming). Because of predictability, the conditional intensity  $\lambda^{\dagger}(t, \xi)$  depends on  $\xi$  only through its past (i.e., its restriction to  $\mathcal{F}_{t-}^{\dagger}$ ). Specifically, appealing to Lemma A3.3.I and the converse statement for the canonical set-up referred to in the discussion after Proposition A3.3.IV [see also Jacobsen (1982)],  $\lambda^{\dagger}(0, \xi)$  is  $\mathcal{H}_{0-}^{\dagger}$ -measurable, and hence can be represented as a measurable function of the restriction of  $\xi$  to  $(-\infty, 0)$ . Then what (14.3.6) implies is that the form of the dependence of  $\lambda^{\dagger}(t)$  on the past up to  $t$  is the same as the form of dependence of  $\lambda^{\dagger}(0)$  on the past up to 0. For a stationary point process, this means that  $\lambda^{\dagger}(t)$  can be represented as a function of the sequence  $\{t - t_i\}$  of intervals from  $t$  back to events  $t_i$  with  $t_i < t$ . A similar statement can be made when the process is stationary and the history includes information on the past of an ancillary stationary process, but general statements in the noncanonical framework are less easy to formulate. Questions relating to the stationarity and uniqueness of a process whose conditional intensity can be represented as in (14.3.6) are taken up in Section 14.7.

Reasoning similar to that leading to (14.3.6) can be used to clarify the form of the absolute continuity condition on the Campbell measure when the process is stationary, under which condition the Campbell measure can be represented in terms of the associated Palm distribution  $\mathcal{P}_0$  (see Definition 13.2.III and recall that we assume the existence of first moments). We may then anticipate that the absolute continuity condition should therefore be expressible in terms of  $\mathcal{P}_0$ . To see what form this condition might take, start from (13.2.6) of the refined Campbell theorem relating  $\mathcal{P}_0$  and the underlying stationary measure  $\mathcal{P}$ , which here can be written for a general random measure on  $\mathcal{B}_{\mathbb{R}}$  as

$$m \int_{\mathbb{R}} \int_{\mathcal{M}_{\mathbb{R}}^{\#}} h(t, \xi) \mathcal{P}_0(d\xi) \ell(dt) = \int_{\mathbb{R} \times \mathcal{M}_{\mathbb{R}}^{\#}} h(t, S_{-t} \xi) C_{\mathcal{P}}(dt \times d\xi)$$

for nonnegative measurable functions  $h$  on  $\mathcal{B}(\mathbb{R} \times \mathcal{M}_{\mathbb{R}}^{\#})$ . Now suppose that the absolute continuity condition holds, so that for nonnegative predictable

functions  $h(x, \xi)$  we can also write

$$\begin{aligned} \int_{\mathbb{R} \times \mathcal{M}_{\mathbb{R}}^{\#}} h(t, S_{-t}\xi) C_{\mathcal{P}}(dt \times d\xi) &= \int_{\mathbb{R}} \int_{\mathcal{M}_{\mathbb{R}}^{\#}} h(t, S_{-t}\xi) \lambda^{\dagger}(t, \xi) dt \mathcal{P}(d\xi) \\ &= \int_{\mathbb{R}} \int_{\mathcal{M}_{\mathbb{R}}^{\#}} h(t, \xi) \lambda^{\dagger}(0, \xi) dt \mathcal{P}(dS_t \xi) \\ &= \int_{\mathbb{R}} \int_{\mathcal{M}_{\mathbb{R}}^{\#}} h(t, \xi) \lambda^{\dagger}(0, \xi) \widehat{S}_t \mathcal{P}(d\xi) dt. \end{aligned}$$

Because  $\mathcal{P}$  is stationary, it is invariant under  $\widehat{S}_t$ , and the above equations yield, for predictable  $h(t, \xi)$ ,

$$m \int_{\mathbb{R}} \int_{\mathcal{M}_{\mathbb{R}}^{\#}} h(t, \xi) \mathcal{P}_0(d\xi) \ell(dt) = \int_{\mathbb{R} \times \mathcal{M}_{\mathbb{R}}^{\#}} h(t, \xi) \lambda^{\dagger}(0, \xi) \mathcal{P}(d\xi) \ell(dt). \quad (14.3.7)$$

The last equation may seem somewhat paradoxical, at least in the point process context, because  $\mathcal{P}_0$  is then defined on a subspace of  $\mathcal{P}$ -measure zero, namely, the realizations with a point at the origin. The explanation of this paradox lies in the predictability requirement, for on both sides of the equation the integration in fact is over the restrictions of the measures  $\xi$  to  $\mathbb{R}_- = (-\infty, 0)$ . Thus (14.3.7) asserts that the projection of  $m\mathcal{P}_0$  onto  $\mathcal{M}^{\#}(\mathbb{R}_-)$  is absolutely continuous with respect to the projection of  $\mathcal{P}$  onto  $\mathcal{M}^{\#}(\mathbb{R}_-)$ , with  $\lambda^{\dagger}(0, \xi)$  acting as the Radon–Nikodym derivative.

We summarize the preceding discussion as follows.

**Proposition 14.3.IV.** *Let  $\xi$  be a stationary random measure or point process in  $\mathbb{R}$  with finite mean rate  $m$ , distribution  $\mathcal{P}$  on  $\mathcal{B}(\mathcal{M}_{\mathbb{R}}^{\#})$ , and stationary Palm distribution  $\mathcal{P}_0$ . If a complete intensity function  $\lambda^{\dagger}$  exists, then it is stationary in the sense of (14.3.6), and  $m\mathcal{P}_0 \ll \mathcal{P}$  on  $\mathcal{H}_{0-}^{\dagger}$  with Radon–Nikodym derivative equal to  $\lambda^{\dagger}(0, \xi)$   $\mathcal{P}$ -a.e.*

*Conversely, if  $m\mathcal{P}_0 \ll \mathcal{P}$  on  $\mathcal{H}_{0-}^{\dagger}$ , the Radon–Nikodym derivative can be taken as one version of  $\lambda^{\dagger}(0, \xi)$ , in which case (14.3.6) is a stationary  $\mathcal{H}^{\dagger}$ -predictable version of the complete intensity.*

**PROOF.** The last assertion follows from the fact that if the Radon–Nikodym derivative is taken as defining  $\lambda^{\dagger}(0, \xi)$ , then it may be supposed  $\mathcal{H}_{0-}^{\dagger}$ -measurable, in which case  $\lambda^{\dagger}(t, \xi)$  defined by (14.3.6) is  $\mathcal{H}_{t-}^{\dagger}$ -measurable for all  $t$  and is thus  $\mathcal{H}_{(-)}^{\dagger}$ -adapted. But in the canonical framework, such a process is also  $\mathcal{H}^{\dagger}$ -predictable (cf. the remarks in Appendix A3.3 already referenced).  $\square$

Returning specifically to the point process case, we next establish a hazard function representation for  $\lambda^{\dagger}$  analogous to that for  $\lambda^*$  (on  $\mathbb{R}_+$ ) given in Corollary 14.1.V. Note that because the past of the process can be represented as a point in a c.s.m.s., we can and do assume the existence of regular conditional probabilities, given the past. In fact it is convenient to adopt the point of view of Theorem 13.3.I and treat the distributions as distributions on sequences of

intervals. This leads to the following formulation, where we use the notation  $T_u$  for the backward recurrence time at time  $u$ ,  $\boldsymbol{\tau}(u) = \{\tau_{-1}(u), \tau_{-2}(u), \dots\}$  for the vector of intervals between consecutive points prior to  $u$ , and  $F(\cdot \mid \boldsymbol{\tau})$  for the conditional distribution function of the interval-length  $\tau_0$  given the vector  $\boldsymbol{\tau} = (\tau_{-1}, \tau_{-2}, \dots)$  of the preceding intervals. Because we are dealing here with stationary processes, the distributions involved can all be derived from the stationary Palm distribution, as indicated in the proof.

**Proposition 14.3.V.** *For a simple stationary point process on  $\mathbb{R}$ , and in the notation above, a version of the  $\mathcal{H}^\dagger$ -compensator is given by*

$$dA(u) = \frac{F(T_u + du \mid \boldsymbol{\tau}(u)) - F(T_u \mid \boldsymbol{\tau}(u))}{1 - F(T_u - \mid \boldsymbol{\tau}(u))}.$$

PROOF. Because of stationarity, we can take  $u = 0$  without loss of generality. Consider the three variables  $t_1$  (the first point in  $\mathbb{R}_+$ ), the backward recurrence time  $T_0$ , and the vector of intervals  $\boldsymbol{\tau}(0)$ , all to be considered as functions of the realization  $N$ . Let  $h(t_1, T_0, \boldsymbol{\tau}(0))$  be any jointly measurable function of these variables on the product space  $\mathbb{R}_+ \times \mathbb{R}_+ \times (\mathbb{R}_+)^{(\infty)}$ , which is given the usual Borel  $\sigma$ -algebra. From (13.3.2) and standard Palm theory we have

$$\mathbb{E}_{\mathcal{P}}(h(t_1, T_0, \boldsymbol{\tau})) = m \mathbb{E}_{\mathcal{P}_0} \left( \int_0^{\tau_0} h(x, \tau_0 - x, \boldsymbol{\tau}) dx \right),$$

where  $\tau_0 = t_1 + T_0$  is the length of the current interval. Recall that under  $\mathcal{P}_0$ ,  $(\tau_0, \boldsymbol{\tau})$  and  $\boldsymbol{\tau}$  have the same distribution, and that we can write, symbolically,  $\mathbb{E}_{\mathcal{P}_0}(\cdot) = \mathbb{E}_{\mathcal{P}_0}[\mathbb{E}_{\tau_0}(\cdot \mid \boldsymbol{\tau})]$ . We evaluate the inner conditional expectation via the conditional distribution  $F(\cdot \mid \boldsymbol{\tau})$ , and thus obtain

$$\mathbb{E}_{\mathcal{P}}(h(t_1, T_0, \boldsymbol{\tau})) = m \mathbb{E}_{\mathcal{P}_0} \left( \int_0^\infty dx \int_0^\infty h(x, y, \boldsymbol{\tau}) dy F(x + y \mid \boldsymbol{\tau}) \right).$$

This relation shows that the joint distribution of  $(t_1, T_0, \boldsymbol{\tau})$  has the form (in infinitesimal notation)

$$m dx d_t F(x + t \mid \boldsymbol{\tau}) M_0(d\boldsymbol{\tau}),$$

where  $M_0$  is the measure induced by  $\mathcal{P}_0$  on  $(\mathbb{R}_+)^{(\infty)}$ , so for the distribution of  $t_1$  conditional on  $T_0$  and  $\boldsymbol{\tau}$ , we have

$$\Pr\{t_1 \leq t \mid T_0, \boldsymbol{\tau}\} = \frac{F(T_0 + t \mid \boldsymbol{\tau}) - F(T_0 - \mid \boldsymbol{\tau})}{1 - F(T_0 - \mid \boldsymbol{\tau})}.$$

Appealing to the properties of regular conditional probabilities, it now follows, as in the proof of Lemma 14.1.III, that the  $\mathcal{H}^\dagger$ -compensator here has the form asserted.  $\square$

**Corollary 14.3.VI.** *A complete intensity for a simple stationary point process exists if and only if the conditional distribution  $F(\cdot \mid \boldsymbol{\tau}) \ll \ell(\cdot)$ , in which case, using  $f(\cdot \mid \boldsymbol{\tau})$  for a density for  $F(\cdot \mid \boldsymbol{\tau})$ , a version of the  $\mathcal{H}^\dagger$ -conditional*

intensity is given by

$$\lambda^\dagger(u) = \frac{f(T_u \mid \tau(u))}{1 - F(T_u \mid \tau(u))}. \quad (14.3.8)$$

In examples, it is usually the case that either  $f(\cdot \mid \tau)$  is continuous or it can be chosen to be left-continuous, and in either circumstance (14.3.8) then gives a predictable version of the  $\mathcal{H}^\dagger$ -intensity provided that at a point of the process,  $T_u$  is interpreted as the length of the preceding interval and not as zero. A corresponding statement for MPPs is outlined in Exercise 14.3.5(b).

### Exercises and Complements to Section 14.3

- 14.3.1 Let  $\xi$  be an  $\mathcal{F}$ -adapted cumulative process with  $\mathcal{F}$ -compensator  $A$ .
- (a) A necessary and sufficient condition for the existence of an  $\mathcal{F}$ -predictable intensity  $\lambda^*$  for  $A$  is that the Campbell measure at (14.2.1) be absolutely continuous with respect to  $\ell \times \mathcal{P}$  on  $\Psi^{\mathcal{F}}$ .
  - (b) When the condition in (a) is satisfied, the Radon–Nikodym derivative  $dC_\xi(t, \omega)/d(\ell \times \mathcal{P}) = \lambda^*(t, \omega)$  is  $\mathcal{F}$ -predictable, and provides an  $\mathcal{F}$ -predictable version of the conditional intensity. This version coincides, except possibly on a set of  $(\ell \times \mathcal{P})$ -measure zero, with any  $\mathcal{F}$ -intensity for  $\xi$  satisfying Definition 14.3.I(a).
- 14.3.2 Develop an alternative approach to the existence of mark-predictable intensities by starting from the definition of a *conditional intensity measure*: a kernel  $\lambda(t, \omega, K)$  from the product space  $\mathbb{R}_+ \times \Omega$  onto Borel sets of  $\mathcal{K}$ , which for each  $K$  is measurable with respect to the predictable  $\sigma$ -algebra. Then  $\lambda$  can be represented as  $\lambda_g(t, \omega) F(K \mid t, \omega)$  where  $F(K \mid t, \omega)$  is a probability kernel, and  $\lambda_g(t, \omega)$  is a predictable intensity for the ground process. Establish projection theorems, definitions via Radon–Nikodym derivatives, and so on, much as in the text. See Brémaud (1981) and Jacod (1975) for details.
- 14.3.3 As an extension of Lemma 14.3.III, show that if the process  $X(t, \omega)$  is left-continuous and integrable and the history  $\mathcal{F}_t$  is also left-continuous, then there exists a predictable version of the conditional expectation  $E[X(t) \mid \mathcal{F}_t]$  [see Mertens (1972)].
- 14.3.4 Follow reasoning similar to that used for the renewal process in Example 14.3(b) to find explicitly the form of the complete intensity function for a Wold process with transition kernel  $P(x, A)$ .
- 14.3.5 (a) Formulate and prove an extension of Proposition 14.3.IV for MPPs, retaining the canonical framework. [Hint: In the marked case, the role of  $\mathcal{P}_0$  is played by the bivariate measure  $\ell_{\mathcal{K}}(d\kappa) \mathcal{P}_{(0, \kappa)}$ .  $\lambda^\dagger(0, \kappa, \xi)$  can be identified as the Radon–Nikodym derivative of this bivariate measure with respect to the product measure  $\ell \times \ell_{\mathcal{K}} \times \mathcal{P}$  on  $\mathcal{H}_{(-)}^\dagger$ .]
- (b) Show that for a stationary MPP with finite mean ground rate, the statement equivalent to (14.3.8) takes the form

$$\lambda^\dagger(u, \kappa) = \frac{f(T_u, \kappa \mid \sigma(u))}{1 - F(T_u, \mathcal{K} \mid \sigma(u))}, \quad (14.3.9)$$

where  $\sigma(u)$  is the family of pairs  $((\tau_{-1}(u), \kappa_{-1}(u)), (\tau_{-2}(u), \kappa_{-2}(u)), \dots)$ , with  $\kappa_i(u)$  the mark at the endpoint of the interval  $\tau_i(u)$ .



## 14.4. Filters and Likelihood Ratios

Much of the early work on the martingale approach to point processes was motivated by the communications engineering context, where the transfer of information by a pulsed signal rather than a continuously modulated signal, had become a major consideration. The emphasis in the engineering literature was on extending the Kalman–Bucy updating algorithms, which allowed real-time estimation and control for processes with a linear Gaussian structure, to point processes (‘jump processes’) and other more general contexts. Noise pulses arising at various stages of the transmission process contaminate the received signal, which is typically a Poisson process with two components, namely, the original signal and the noise. The practical questions to be resolved concern the estimation of the original signal on the basis of the point process observed at the receiving end of the transmission line, and its use in control procedures.

The last two decades have seen the development of a much wider interest in these models and the estimation procedures that go with them, often in the terminology of the hidden Markov models discussed in Section 10.3. In most of these applications, a key problem is estimating the state of some unobserved system at a particular time, either from the information available up to that time (*filtering problem*), or from information available over a longer time period (*smoothing problem*). But whereas the discussion in Section 10.3 used intuitive arguments based on standard properties of discrete state Markov processes, here we provide an introduction to the more general martingale approach, making use of the concepts developed earlier in this chapter.

Brémaud and Jacod (1977) provide a helpful informal guide to the relationships between the point process procedures and earlier work on Gaussian filters. Fuller discussions of the filtering problems, with further references, can be found in Brémaud (1981) and Snyder and Miller (2000), and a very general treatment of the marked case is in Last and Brandt (1995, Chapter 11). We start with a recapitulation of the simplest version of the hidden Markov models discussed in Section 10.3.

EXAMPLE 14.4(a) *Cox process directed by a finite state Markov process* [continued from Example 10.3(d), Exercise 7.2.8]. In Example 10.3(d) we derived explicit estimates of the current state of a Markov chain  $X(t)$  which, when it is in state  $j$ , generates points of a Poisson process at rate  $\lambda_j$ . Such an example contains in embryonic form most of the features of the general problem of filtering for point processes. In particular, we adduced earlier an estimate  $\pi_i(t)$  of the probability of the current state, given the observations on the point process, as a ratio involving the ‘joint statistics’ or ‘forward probabilities’  $p_i(t)$  [cf. (10.3.4)], and the likelihood, which is just their sum  $\sum p_i(t)$ :

$$\pi_i(t) = \Pr\{X(t) = i \mid \text{observations}\} = p_i(t) / \mathbf{p}(t)^\top \mathbf{1}. \quad (14.4.1a)$$

Let  $\mathcal{F}$  denote the joint history of the point process and the Markov chain, and  $\mathcal{H}$  the internal history of the point process. Then  $\pi_i(t)$  can be identified

as the conditional expectation of the indicator function  $E[I\{X(t) = i\} \mid \mathcal{H}_t]$ . Similarly, we can identify the  $\mathcal{F}$ -conditional intensity as  $\lambda_{\mathcal{F}}^*(t) = \lambda_{X(t-)}$ , and from Lemma 14.3.III we obtain the  $\mathcal{H}$ -intensity as the projection

$$\lambda_{\mathcal{H}}^*(t) = E[\lambda_{X(t-)} \mid \mathcal{H}_t] = \sum_{i=1}^K \pi_i(t) \lambda_i = \frac{\sum_{i=1}^K p_i(t) \lambda_i}{\sum_{i=1}^K p_i(t)}. \quad (14.4.1b)$$

Equations (14.4.1) are typical filtering equations, and one of the main goals of this section is to find their extensions to more general models. Note also the possibility of developing a succinct system of equations for updating such estimates as in (10.3.6–9). Because of the importance of obtaining real-time estimates of the state, it is the latter topic that is dominant in much of the earlier literature. Both the differential equations that hold between observed points, and the difference equations that hold when such points are traversed, can be incorporated into a single set of integral equations, and such integral equations then form the main object of study. They play a role here somewhat analogous to that of stochastic differential equations in models exploiting diffusion concepts.  $\square$

Martingale representations fit naturally into this discussion in which two main approaches can be identified, the ‘innovations’ approach, in which the martingale representations are studied via filtering results such as those described in Proposition 14.3.II and Lemma 14.3.III, and the method of ‘reference probabilities,’ in which they enter via likelihoods. We outline an introduction to the second approach, based on Brémaud (1981) which may be consulted for a more extended account.

We start with a re-examination and extension of Propositions 7.2.III and 7.3.III concerning the structure of likelihood ratios for point processes which are regular (i.e., for which the Janossy measures have densities), and the conditional intensities are taken with respect to the internal histories. We state the result for MPPs; the simple point process appears as the special case that the mark space  $\mathcal{K}$  reduces to a single point.

**Proposition 14.4.I.** *Consider an MPP  $N$ , with ground process  $N_g$ , state space  $\mathbb{R}_+ \times \mathcal{K}$ , adapted to the internal history  $\mathcal{H}$  on the probability space  $(\Omega, \mathcal{E})$ . For  $j = 1, 2$  let  $\mathcal{P}_j$  be a probability distribution on  $(\Omega, \mathcal{E})$  such that for each  $\mathcal{P}_j$  the ground process is boundedly finite (hence nonexplosive) with boundedly finite first moment measure, and for each  $t > 0$ ,  $\mathcal{P}_j^t$  denotes the restriction of  $\mathcal{P}_j$  to  $\mathcal{H}_t$ . Suppose that  $\mathcal{P}_1$  has a strictly positive mark-predictable  $\mathcal{H}$ -intensity  $\lambda_1^{\mathcal{H}}(t, \kappa, \omega)$  relative to the reference measure  $\ell_{\mathcal{K}}$  on  $\mathcal{B}_{\mathcal{K}}$ . Then the necessary and sufficient condition for  $\mathcal{P}_2^t \ll \mathcal{P}_1^t$  for  $0 < t \leq \infty$  is the existence of a nonnegative, mark-predictable process  $\mu(t, \kappa, \omega)$ , such that under  $\mathcal{P}_2$ ,  $N$  admits the mark-predictable intensity  $\lambda_2^{\mathcal{H}}(t, \kappa, \omega) = \mu(t, \kappa, \omega) \lambda_1^{\mathcal{H}}(t, \kappa, \omega)$ . When this condition is satisfied, the likelihood ratio*

$$L_t(\omega) = \frac{d\mathcal{P}_2^t(N(\omega))}{d\mathcal{P}_1^t(N(\omega))}$$

has a right-continuous version with left limits given  $\mathcal{P}_1$ -a.s. by  $L_t(\omega) = 1$  for  $t < t_1(\omega)$ , and for  $t \geq t_1$  it equals

$$\left( \prod_{0 < t_i \leq t} \mu(t_i, \kappa_i, \omega) \right) \exp \left( - \int_{(0, t] \times \mathcal{K}} [\mu(s, \kappa, \omega) - 1] \lambda_1^{\mathcal{H}}(s, \kappa, \omega) \, ds \, \ell_{\mathcal{K}}(d\kappa) \right), \quad (14.4.2)$$

where  $\{(t_i, \kappa_i): i = 1, 2, \dots\}$  is an enumeration of the points of  $N$ . Moreover, if for  $j = 1, 2$ ,  $\lambda_g^{(j)}(t)$  is the ground intensity and  $f^{(j)}(\kappa | t)$  is the density of the mark-predictable kernel, and  $\mu_g(t, \omega) = \lambda_g^{(1)}(t, \omega) / \lambda_g^{(2)}(t, \omega)$ , then for each  $T > 0$ ,  $L_t(\omega)$  is the unique such solution of the integral equation

$$L_t(\omega) = 1 + \int_0^t L_{s-}(\omega) [\mu_g(s, \omega) - 1] Z^*(ds, \omega) \quad (0 \leq t \leq T), \quad (14.4.3)$$

where  $Z^*(t)$  is the  $(\mathcal{H}, \mathcal{P}_1)$ -martingale defined by

$$dZ^*(t) = \int_{\mathcal{K}} \frac{f^{(2)}(\kappa | t)}{f^{(1)}(\kappa | t)} N(dt \times d\kappa) - \lambda_g^{(1)}(t) dt.$$

PROOF. Because we are concerned here only with the internal history intensities, a straightforward proof of (14.4.2) can be written down by starting from the representation of the likelihoods in terms of Janossy measures, as in the discussion leading to Proposition 7.2.III. By assumption, the Janossy measures under  $\mathcal{P}_1$  are absolutely continuous (with respect to Lebesgue measure) with densities determined by the conditional intensity  $\lambda_1^{\mathcal{H}}(t, \kappa, \omega)$ . Then the existence of the likelihood ratio implies absolute continuity for the Janossy measures under  $\mathcal{P}_2$  also. In turn, the densities under  $\mathcal{P}_2$  can be used to define the conditional intensity under  $\mathcal{P}_2$ , namely,  $\lambda_2^{\mathcal{H}}(t, \kappa, \omega)$ . Finally, restating the Janossy densities in terms of the conditional intensities leads to the form of the likelihood ratio given in (14.4.2).

The fact that the conditional hazard functions are measurable functions of  $(t_i, \kappa_i)$  for  $t_i \leq t$  implies that, as a function on  $(0, t] \times \mathcal{K} \times \Omega$ ,  $\mu(s, \kappa, \omega)$  is  $(\mathcal{B}(0, t) \otimes \mathcal{B}_{\mathcal{K}} \otimes \mathcal{H}_t)$ -measurable, and therefore, from the projection theorem, possesses an  $\mathcal{H}$ -mark-predictable version. Use of this version in (14.4.2) leaves the likelihood unaltered except possibly on a set of  $(\ell \times \ell_{\mathcal{K}} \times \mathcal{P}_1)$ -measure zero.

Conversely, suppose that  $\mathcal{P}_2 \ll \mathcal{P}_1$  on  $\mathcal{H}_T = \mathcal{H}_{[0, T]}$ , and that the conditional intensity  $\lambda_1^{\mathcal{H}}$  exists. We consider the likelihood ratio in the successive intervals  $(0, t_1)$ ,  $[t_1, t_2)$ ,  $\dots$ . In the interval  $(0, t_1)$ , absolute continuity of the likelihoods means that the distribution function of the length of the first interval under  $\mathcal{P}_2$  is absolutely continuous with respect to its distribution under  $\mathcal{P}_1$ . The latter has a positive density, so this is equivalent to the assertion that its distribution under  $\mathcal{P}_2$  is absolutely continuous with respect to Lebesgue measure. The ratio of the two densities on the event  $t_1 > t$  is also the ratio of the two conditional intensities, and so defines an appropriate form of  $\mu$  on this

event. When  $t_1 \leq t < t_2$ , a similar argument applied to the conditional distribution for the length of the second interval, given  $t_1$ , establishes the existence of an appropriate  $\mu$  on this event also. Proceeding in this way, and observing that the process is free from explosions under both probabilities, shows that the  $\mathcal{P}_2^{\mathcal{H}}$  conditional intensity exists in general. Both conditional intensities, and hence also their ratio, can be given predictable versions without altering the ratio except on sets of  $(\ell \times \ell_{\mathcal{K}} \times \mathcal{P}_1)$ -measure zero.

Finally, the equivalence of (14.4.2) and (14.4.3) is a consequence of the exponential formula (Lemma 4.6.II) if in that formula we take  $Z^*(t)$  in place of the monotonic function  $F(t)$ , and the ratio  $[\mu_g(t) - 1]$  in place of the function  $u(t)$ . The integrability requirement on  $u$  here follows from the equations

$$\begin{aligned} \mathbb{E}_1 \left( \int_0^t \mu_g(s) \frac{f^{(2)}(\kappa | s)}{f^{(1)}(\kappa | s)} dN_g(s) \right) &= \mathbb{E}_1 \left( \int_{(0,t] \times \mathcal{K}} \mu(s, \kappa) N(ds \times d\kappa) \right) \\ &= \mathbb{E}_1 \left( \int_{(0,t] \times \mathcal{K}} \mu(s, \kappa) \lambda^{(1)}(s, \kappa) ds \ell_{\mathcal{K}}(d\kappa) \right) \end{aligned}$$

and the last expression equals  $\mathbb{E}_2(N_g(0, t]) < \infty$ , where for  $j = 1, 2$ ,  $\mathbb{E}_j$  denotes expectation with respect to  $\mathcal{P}_j$ . Equality of the right-hand sides of the first and second of these equations is also the essential calculation needed to establish the martingale property of  $Z^*$ , because it shows that  $\mathbb{E}[dZ^*(t) | \mathcal{F}_t] = 0$ .  $\square$

It is often convenient to write (14.4.2) in terms of the ground processes  $N_g(t)$  and conditional mark distributions  $f^{(j)}(\kappa | t)$  (cf. Proposition 7.3.III) as

$$\begin{aligned} L_t(\omega) &= \left[ \prod_{i=1}^{N_g(t)} \frac{\lambda_g^{(2)}(t_i, \omega)}{\lambda_g^{(1)}(t_i, \omega)} \right] \\ &\quad \times \exp \left( - \int_{(0,t] \times \mathcal{K}} [\lambda_g^{(2)}(t, \omega) - \lambda_g^{(1)}(t, \omega)] dt \ell_{\mathcal{K}}(d\kappa) \right) \prod_{i=1}^{N_g(t)} \frac{f^{(2)}(\kappa_i | t_i)}{f^{(1)}(\kappa_i | t_i)}. \end{aligned}$$

When  $N$  is a simple point process, both expressions take the familiar form

$$\begin{aligned} L_t(\omega) &= \left( \prod_{0 < t_i \leq t} \mu(t_i, \omega) \right) \exp \left( - \int_0^t [\mu(s, \omega) - 1] \lambda^{(1)}(s, \omega) ds \right) \\ &= \left( \prod_{0 < t_i \leq t} \frac{\lambda^{(2)}(t_i, \omega)}{\lambda^{(1)}(t_i, \omega)} \right) \exp \left( - \int_0^t [\lambda^{(2)}(s, \omega) - \lambda^{(1)}(s, \omega)] ds \right), \end{aligned}$$

and the martingale  $dZ^*(t)$  reduces to  $dZ(t) = dN(t) - \lambda^{(1)}(t) dt$ .

**Corollary 14.4.II.** *For an MPP satisfying the conditions of Proposition 14.4.I, there exists a sequence of stopping times  $S_n \rightarrow \infty$  a.s. as  $n \rightarrow \infty$  such that for each  $n$ ,  $L_{t \wedge S_n}$  is an  $\mathcal{H}$ -martingale under  $\mathcal{P}_1$ .*

PROOF. Because  $L(t) \equiv L_t$  is left-continuous and nonnegative in each finite interval  $(0, t]$ , it is also bounded a.s. on such intervals. Furthermore, if  $T_n$  is defined for the ground process  $N_g$  as at (14.1.1),

$$E_1 \int_0^{t \wedge T_n} \mu_g(s) \lambda_g^{(1)}(s) ds = E_2[N_g(t \wedge T_n)] \leq n < \infty,$$

and

$$E_1 \int_0^{t \wedge T_n} \lambda_g^{(1)}(s) ds = E_1[N_g(t \wedge T_n)] \leq n < \infty.$$

Defining

$$S_n = T_n \wedge \inf\{t: L(t) \geq n\},$$

we necessarily have  $S_n \rightarrow \infty$  and thus

$$E_1 \int_0^{t \wedge S_n} |L(s-) [\mu_g(s) - 1]| \lambda_g^{(1)}(s) ds \leq n^2.$$

Thus, the quantity  $L(s-)[\mu_g(s) - 1]$  on the right-hand side of (14.4.3) is predictable, so because  $Z^*(\cdot)$  is an  $\mathcal{H}$ -martingale under  $\mathcal{P}_1$ , the likelihoods  $L(t \wedge S_n)$  must also form a martingale (see Exercise 14.1.7).  $\square$

This corollary can also be rephrased as stating that *under  $\mathcal{P}_1$ , the likelihood ratios  $L_t$  form an  $\mathcal{H}$ -local martingale.*

It should be emphasized that the treatment in the last few results involves some substantial simplifications. First, by restricting the discussion to boundedly finite processes, we rule out ‘explosive’ situations, where the sequence of time points  $T_n$  may approach a finite limit point. In general, if this requirement is dropped, similar results hold on the interval up to the time of the first ‘explosion’. See, for example, Liptser and Shiryaev (1978) or Last and Brandt (1995) for treatments when this constraint is relaxed.

Second, the absolute continuity condition imposed upon the compensator under  $\mathcal{P}_1$  rules out situations where the compensators may have discontinuities. Such discontinuities correspond to atoms in the conditional distributions defining the compensator, as illustrated in Exercise 14.1.10(c). Likelihood ratios can certainly be considered for such processes, with the requirement that the compensator for the derived process has discontinuities only where the base process compensator has discontinuities. Then the general form of the IHF comes into play; a simple example, involving a Poisson process with fixed atoms, is given in Exercise 14.4.4, leading to additional factors of the form

$$\prod \frac{1 - \Delta A_i^{(2)}}{1 - \Delta A_i^{(1)}}, \quad (14.4.4)$$

where for  $j = 1, 2$ , the  $\Delta A_i^{(j)}$  are the jumps in the compensator under  $\mathcal{P}_j$ ,  $\Delta A_i^{(2)} = \mu(\tau_i) \Delta A_i^{(1)}$  where the  $\tau_i$  are the corresponding time points, and, for

a simple point process, both sets of jumps are required to satisfy  $\Delta A_i^{(j)} \leq 1$ . Again see, for example, Boel, Varaiya and Wong (1975); Jacod (1975); Liptser and Shirayev (1978); Last and Brandt (1995, (10.1.14)) for further details.

The third simplification is a consequence of considering only internal histories. Situations where one conditional intensity depends on information additional to the record of the occurrence times and the other does not, as happens in the hidden Markov schemes described earlier, require some filtering procedure to reduce the more complex intensity to a form of internal intensity, as we consider below Proposition 14.4.IV. In yet other situations, the point process likelihood may be only a partial likelihood, depending on the values of random variables treated as constants for the purposes of likelihood estimation.

We proceed to a consideration of additional issues from the filtering perspective, limiting the discussion to simple point processes to avoid too cumbersome a treatment.

The form of the likelihood ratio prompts the following more general question. Given a point process on  $\mathbb{R}_+$  with measure  $\mathcal{P}$  and  $\mathcal{F}$ -intensity  $\lambda^{\mathcal{F}}$  under  $\mathcal{P}$ , and some further nonnegative  $\mathcal{F}$ -predictable function  $\mu(\cdot)$ , does the product  $\mu(t, \omega) \lambda^{\mathcal{F}}(t, \omega)$  represent the  $\mathcal{F}$ -intensity of the point process under the new measure  $\mathcal{P}_1(d\omega) = L_t(\omega) \mathcal{P}(d\omega)$ ? The answer in general is no, because under the new measure there is no guarantee, without some further constraints on  $\mu(\cdot)$ , that the new trajectories will be a.s. boundedly finite. In other words, the new measure may not necessarily be a probability measure, the possible mass deficiency corresponding to the probability that the realizations no longer lie within  $\mathcal{N}_{(0,t]}^\#$ . It seems rather difficult to find conditions directly on  $\mu$  that will avert this possibility; it is obvious that boundedness of  $\mu(t, \omega) \lambda^{\mathcal{H}}(t, \omega)$  on  $(0, t]$  is sufficient, but this is too restrictive to be generally useful. An alternative stratagem is simply to require  $\mathcal{P}_1$  to form a probability measure, and this leads to the following result.

**Proposition 14.4.III.** *Let  $N$  be a point process on  $\mathbb{R}_+$  defined on the probability space  $(\Omega, \mathcal{E}, \mathcal{P})$ ,  $\mathcal{F}$  a right-continuous history for  $N$  such that  $N$  admits the intensity  $\lambda^{\mathcal{F}}(t, \omega)$ , and  $\mu(t, \omega)$  a nonnegative  $\mathcal{F}$ -predictable process satisfying for some  $t > 0$*

$$\int_0^t \mu(s, \omega) \lambda^{\mathcal{F}}(s, \omega) ds < \infty \quad \mathcal{P}\text{-a.s.}$$

*Then for  $L_t(\omega)$  defined by (14.4.2), the necessary and sufficient condition for  $\mathcal{P}_1(d\omega) = L_t(\omega) \mathcal{P}(d\omega)$  to be a probability measure on  $(\Omega, \mathcal{E})$  is that*

$$\mathbb{E}[L_t(\omega)] = \int_{\Omega} L_t(\omega) \mathcal{P}(d\omega) = 1; \quad (14.4.5)$$

*then  $N$  has an  $\mathcal{F}$ -intensity under  $\mathcal{P}_1$  equal to  $\mu(s, \omega) \lambda^{\mathcal{F}}(s, \omega)$  for  $0 \leq s \leq t$ .*

PROOF. The relation at (14.4.5) is clearly necessary and sufficient for  $\mathcal{P}_1$  to be a probability measure, because it is simply the statement  $\int_{\Omega} \mathcal{P}_1(d\omega) = 1$ .

The substantial part is to show that when it is satisfied,  $\mu(s, \omega) \lambda^{\mathcal{F}}(s, \omega)$  is the  $\mathcal{F}$ -intensity on  $(0, t]$  for  $N$  under  $\mathcal{P}_1$ , for which it is sufficient to show that for every nonnegative  $\mathcal{F}$ -predictable process  $Y(s, \omega)$  on  $(0, t]$ ,

$$\mathbb{E}_1 \left( \int_0^t Y(s) dN(s) \right) = \mathbb{E}_1 \left( \int_0^t Y(s) \mu(s) \lambda^{\mathcal{F}}(s) ds \right).$$

By definition of  $\mathcal{P}_1$ , these expectations  $\mathbb{E}_1(\cdots)$  under  $\mathcal{P}_1$  can be rewritten as expectations  $\mathbb{E}(\cdots)$  with respect to  $\mathcal{P}$ , writing  $L(t) = L_t(\omega)$ ,

$$\mathbb{E} \left( L(t) \int_0^t Y(s) dN(s) \right) \quad \text{and} \quad \mathbb{E} \left( L(t) \int_0^t Y(s) \mu(s) \lambda^{\mathcal{F}}(s) ds \right),$$

respectively. Now under condition (14.4.5),  $L(s, \omega)$  is not merely a local martingale but is in fact a martingale [this follows from the defining relation (14.4.2), where it is necessarily a submartingale, and the relation  $\mathbb{E}_1(L_t) = \mathbb{E}_1(L_0) = 1$  then shows it to be a martingale]. From the results in Section A3.3 concerning the Doob–Meyer decomposition and natural increasing functions, these expectations  $\mathbb{E}(\cdots)$  can now be rewritten as

$$\mathbb{E} \left( \int_0^t L(s) Y(s) dN(s) \right) \quad \text{and} \quad \mathbb{E} \left( \int_0^t L(s) Y(s) \mu(s) \lambda^{\mathcal{F}}(s) ds \right),$$

respectively. But then from the definition of  $L(s)$  and the fact that  $\lambda^{\mathcal{F}}(\cdot)$  is the  $\mathcal{F}$ -intensity of  $N$  under  $\mathcal{P}$ , the first of these integrals equals

$$\begin{aligned} \mathbb{E} \left( \int_0^t L(s-) Y(s) dN(s) \right) &= \mathbb{E} \left( \int_0^t L(s-) Y(s) \mu(s) \lambda^{\mathcal{F}}(s) ds \right) \\ &= \mathbb{E} \left( \int_0^t L(s) Y(s) \mu(s) \lambda^{\mathcal{F}}(s) ds \right), \end{aligned}$$

where the natural increasing property of  $\int_0^s Y(u) \mu(u) \lambda^{\mathcal{F}}(u) du$  justifies the last equality.  $\square$

Next, we formulate a general version of the Bayesian type of formulae that appear in Examples 14.3(a) and 14.4(a). In this version, the role of the likelihood itself (i.e., the density with respect to Lebesgue measure) is taken by the likelihood ratio with respect to a ‘reference probability’  $\mathcal{P}_1$ , which we discuss shortly.

**Proposition 14.4.IV.** *Let  $\mathcal{G} \subseteq \mathcal{F}$  be two histories for the point process  $N$  on  $\mathbb{R}_+$  and  $\mathcal{P}_1, \mathcal{P}_2$  two probability measures, all defined on the measurable space  $(\Omega, \mathcal{E})$ . Suppose that for  $j = 1, 2$  and  $t$  in some interval  $(0, T]$ , the restrictions  $\mathcal{P}_j^t$  of  $\mathcal{P}_j$  to  $\mathcal{F}_t$  satisfy  $\mathcal{P}_2^t \ll \mathcal{P}_1^t$ , and let  $L_t = d\mathcal{P}_2^t/d\mathcal{P}_1^t$  be the likelihood ratio. Then for any real-valued bounded  $\mathcal{F}$ -adapted process  $Z_t$ ,*

$$\mathbb{E}_2(Z_t \mid \mathcal{G}_t) = \frac{\mathbb{E}_1(Z_t L_t \mid \mathcal{G}_t)}{\mathbb{E}_1(L_t \mid \mathcal{G}_t)} \quad \mathcal{P}_1\text{-a.s.} \quad (14.4.6)$$

PROOF. First we show that the product  $E_1(L_t | \mathcal{G}_t) E_2(Z_t | \mathcal{G}_t)$  can be taken as a version of the conditional expectation  $E_1(Z_t L_t | \mathcal{G}_t)$ . By definition, for  $B \in \mathcal{G}_t$ ,

$$\int_B E_1(Z_t L_t | \mathcal{G}_t) \mathcal{P}_1(d\omega) = \int_B Z_t L_t \mathcal{P}_1(d\omega) = \int_B Z_t \mathcal{P}_2(d\omega),$$

and, similarly, because  $E_2(Z_t | \mathcal{G}_t)$  is  $\mathcal{G}_t$ -measurable,

$$\int_B E_2(Z_t | \mathcal{G}_t) E_1(L_t | \mathcal{G}_t) \mathcal{P}_1(d\omega) = \int_B E_2(Z_t | \mathcal{G}_t) \mathcal{P}_2(d\omega) = \int_B Z_t \mathcal{P}_2(d\omega),$$

implying the desired  $\mathcal{P}_1$ -a.s. equality.

Now the set  $D$ , where  $E_1(L_t | \mathcal{G}_t) = 0$ , is  $\mathcal{G}_t$ -measurable and satisfies

$$\mathcal{P}_2(D) = \int_D L_t \mathcal{P}_1(d\omega) = \int_D E_1(L_t | \mathcal{G}_t) \mathcal{P}_1(d\omega) = 0,$$

so the  $\mathcal{P}_1$ -a.s. equality remains true when put in the ratio form (14.4.6) whatever particular definition is used for the conditional expectations when the denominator of (14.4.6) is zero.  $\square$

These results form the starting point of a general attack on the problem of filtering for point processes. The reference probability  $\mathcal{P}_1$  is generally the distribution of a Poisson process at unit rate, this process being assumed to be independent of the other random variables of interest, and in particular of those governing the signal process. The smaller history  $\mathcal{G}$  is commonly the internal history  $\mathcal{H}$ , in which case the ratios at (14.4.6) are quite analogous to the ratios occurring earlier in the examples. Thus, in the mixed Poisson process of Example 14.3(a), the  $\mathcal{F}$ -likelihood ratio equals  $(\mu t)^{N'} e^{-(\mu-1)t}$ , and the quantity under the expectation sign in the numerator at (14.4.6) equals  $\mu(\mu t)^{N'} e^{-(\mu-1)t}$  which, on taking conditional expectations, cancelling the factor  $t^{N'} e^t$ , and noting that  $\mu$  is independent of  $\mathcal{H}_t$  under  $\mathcal{P}_1$ , leads back to the expression for the  $\mathcal{H}$ -intensity obtained above (14.3.3). Similarly the updating formulae in Example 14.4(a) follow from (14.4.6) by setting  $Z_t = I\{X(t) = i\}$ .

We now quote some more general results which can be obtained by this method. Essentially, we follow the development in Brémaud (1981) which gives more details.

Updating formulae can be developed for both numerator and denominator of (14.4.6), essentially as corollaries to the integral equation (14.4.3) already obtained for the likelihood. The denominator requires the conditional expectation with respect to  $\mathcal{G}_t$  of the  $\mathcal{F}$ -likelihood ratio defined in terms of the  $\mathcal{F}$ -intensity. At first sight it may seem surprising that the result of this is simply to replace the  $\mathcal{F}$ -intensity by the  $\mathcal{G}$ -intensity in the expression for the likelihood ratio, which in the language of engineers implies the ‘separation of estimation and detection’ [recall that ‘detection’ is implemented through tests based on the likelihood ratio, and ‘estimation’ comes from the updating formulae based on (14.4.6)]. The result is most easily proved by first reverting to the integral equation form for the likelihood, taking conditional



expectations, replacing  $\lambda^{\mathcal{F}}$  by  $\lambda^{\mathcal{G}}$ , and then returning to the basic form via a second application of the exponential formula (see Exercise 14.4.5 for details).

Updating the numerator at (14.4.6) when  $\mathcal{G} = \mathcal{H}$  is also effected by writing down the integral equation for the likelihood ratio, multiplying through by  $Z_t$ , and taking conditional expectations with respect to  $\mathcal{H}$ .

Many of the most important examples in signal processing, and in many other fields also, relate to extensions of the Markov-directed Cox processes discussed earlier in this chapter and Section 10.3. In this case there occur a number of simplifications which allow the updating formulae for the numerator to be reduced to relatively manageable form, as we indicate below.

Suppose first that  $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{H}_t$ , where  $\mathcal{F}_0$  incorporates full information about a prior or directing (signal) process  $X_t$ , so that  $\mathcal{F}_0 \supseteq \sigma\{X_t: 0 \leq t < \infty\}$ . Then under  $\mathcal{P}_1$ ,  $N$  is independent of  $\sigma\{X_t\}$ , and still using the  $\mathcal{F}$ -predictable version of the intensity, we obtain

$$\mathbb{E}_1(Z_t L_t \mid \mathcal{H}_t) = \mathbb{E}_1(Z_t) + \int_0^t \mathbb{E}_1(Z_t L_{s-} [\lambda^{\mathcal{F}}(s) - 1] \mid \mathcal{H}_s) (dN(s) - ds). \quad (14.4.7)$$

Suppose, in particular, that  $N$  is a Cox process directed by some nonnegative measurable function  $\mu$  of  $X_t$  so that  $\lambda(t) = \mu(X_t) \geq 0$  a.s., and that  $X_t$  is a stationary Markov process on  $\mathbb{R}$  with transition semigroup  $P^t$ . The main interest in the filtering context centres on estimating features of the directing process  $X_t$ , so we take  $Z_t = f(X_t)$  for some bounded continuous function  $f: \mathbb{R} \mapsto \mathbb{R}$ . The Markov property for  $X_t$  then implies

$$\begin{aligned} \mathbb{E}_1[L_t f(X_t)] \\ = \mathbb{E}_1[f(X_t)] + \int_0^t \mathbb{E}_1(L(s-)[\lambda(s) - 1]P^{t-s}f(X_s) \mid \mathcal{H}_s)(dN(s) - ds). \end{aligned}$$

Moreover, if the transition probabilities  $P^t(x, B)$  ( $x \in \mathbb{R}$ ,  $B \in \mathcal{B}(\mathbb{R})$ ) are continuous functions of  $x$  for  $t > 0$ , this expression can be written explicitly in terms of these probabilities and the associated distributions  $P_t(B) = \mathcal{P}\{X(t) \in B\}$  and  $\Pi_t(B) = \Pi_t(B \mid N) \equiv \mathbb{E}[L_t I_B(X_t) \mid \mathcal{H}_t]$ , the latter being a regular version of the conditional probability expressed as a function of the realization  $N$ . The previous equation then takes the form

$$\Pi_t(B) = P_t(B) + \int_{\mathbb{R}} \int_0^t \Pi_s(dx) [\mu(x) - 1] P^{t-s}(x, B) (dN(s) - ds), \quad (14.4.8)$$

from which the updating character is more readily apparent.

Explicit expressions depend on the nature of the governing Markov process. The simplest case is that of a pure jump process with a denumerable set of states. The transition probabilities  $P^{t-s}$  in (14.4.8) then take the matrix exponential form  $e^{(t-s)Q}$  and the equations reduce to the updating formulae already noted in Example 14.4(a) and Section 10.3. The one point of difference is that, because we are here considering the likelihood ratio, it is the difference of the conditional intensities (the second being just unity as it corresponds

to a unit rate Poisson process) which appears in the integrand. This leads to a multiplicative factor which cancels in the ratio at (14.4.6). Even in the more difficult cases of diffusion and mixed jump-diffusion processes, a similar formal representation holds, but with the role of  $Q$  taken by the infinitesimal generator. This is illustrated in the final example, which was a further starting point for point-process filtering theory in Snyder (1972, 1975).

EXAMPLE 14.4(b) *Cox process directed by a Markov diffusion process.* Suppose there is given a diffusion process on  $\mathbb{R}_+$ , whose densities  $p_t(y)$  satisfy the forward equation (with  $D_t = \partial/\partial t$ ,  $D_y = \partial/\partial y$ ,  $D_y^2 = \partial^2/\partial y^2$ ),

$$D_t p_t(y) = -D_y[\beta(y)p_t(y)] + \frac{1}{2}D_y^2[\alpha(y)p_t(y)] \equiv \mathcal{L}(p_t(y)),$$

for drift and diffusion terms  $\alpha(\cdot)$  and  $\beta(\cdot)$  and diffusion operator  $\mathcal{L}$  [see, e.g., Feller (1966, Chapter XIV)]. Denoting the density of  $\Pi_t$  at (14.4.8) by  $\pi_t$  and differentiating (14.4.8) between jumps, we obtain [recalling  $p^0(x, y) = \delta(x - y)$  and using the linearity of  $\mathcal{L}$ ]

$$\begin{aligned} D_t \pi_t(y) &= \mathcal{L}(p_t(y)) + \int_{\mathbb{R}} \int_0^t \pi_s(x) [\mu(x) - 1] \mathcal{L}(p^{t-s}(x, y)) \, dx \, [dN(s) - ds] \\ &\quad - \pi_t(y) [\mu(y) - 1] \\ &= \mathcal{L}(\pi_t(y)) - \pi_t(y) [\mu(y) - 1], \end{aligned} \tag{14.4.9}$$

and at any jump  $t$  of  $N$  we have  $\pi_{t+}(y) = \pi_{t-}(y)\mu(y)$ .

The differential equation (14.4.9) may be compared with the differential equation for the joint probabilities in Example 14.4(a). As in that example, it is just the forward differential equation for the transition probabilities modified by the extra term needed to preclude the occurrence of additional points during the interval under examination. Similarly, the expression for the jump term has the same general character as in that example. One point to re-emphasize is that the updating formulae for the joint probabilities, derived from the numerator of (14.4.6), are much easier to handle than the nonlinear equations which arise in trying to update the state probabilities directly.  $\square$

The main point about (14.4.8) and the various special cases that can be deduced from it is that it provides a very general formulation of the updating equations for the ‘joint statistics’ (forward probabilities) which first appeared in (10.3.6) and (10.3.13a). The existence of such updating formulae depends on the Markov structure of the unobserved process and the assumed independence of the observations (the points of the point process) given the current state of the Markov process. The resulting filtering theory has been mainly applied to the problem of state estimation.

Whether there exist corresponding extensions to the E–M algorithm, used in Section 10.3 to provide a tractable approach to the problem of parameter estimation, is still a subject of current research. A variety of approaches has been suggested, from analytic results in special cases to various forms of Monte Carlo estimation, including particle filters, in more complex situations.

An alternative in such more complex situations is to revert to a discrete context, for example, by approximating a diffusion process by a random walk, and taking advantage of the extensive literature and software for discrete hidden Markov models. There are important potential applications in many fields, including finance, neurophysiology, meteorology, and geophysics, as well as in the original context of signal processing. See references under hidden Markov models in Section 10.3 for some more recent work.

## Exercises and Complements to Section 14.4

14.4.1 Extend Example 14.4(a) to the context where the underlying Markov process is bivariate,  $(X(t), Y(t))$  say, with  $X(t)$   $\{1, \dots, K\}$ -valued as before and unobserved, and  $Y(t)$  observed with values  $\{\alpha, \beta, \dots\}$ , so that  $\{(X(t), Y(t))\}$  has  $Q$ -matrix  $(q_{(i,\alpha),(j,\beta)})$  say. Denote the internal history of the observed process  $Y(\cdot)$  by  $\mathcal{H} = \{\mathcal{H}_t\} \equiv \{\sigma(\{Y(s): 0 \leq s \leq t\})\}$ .

By analogy with (14.4.1) set  $\hat{p}_X(t) \equiv (\hat{p}_1(t), \dots, \hat{p}_K(t))^\top$  with  $\hat{p}_i(t) = \mathcal{P}\{X(t) = i, \text{ observed } Y(s) (0 \leq s \leq t)\}$ , so that [cf. (10.3.6–7)] on intervals of constancy for  $Y(\cdot)$ ,  $Y(t) = \alpha$  say,

$$D_t \hat{p}_i(t) = -q_{(i,\alpha)(j,\alpha)} \hat{p}_i(t) + \sum_{k \neq i} q_{(k,\alpha),(i,\alpha)} \hat{p}_k(t),$$

and where  $\alpha = Y(t-) \neq Y(t) = \beta$ ,  $\hat{p}_i(t) = \hat{p}_i(t-) q_{(i,\alpha)(i,\beta)} / |q_{(i,\alpha)(i,\alpha)}|$ .

- Recover Example 14.4(a) from Example 10.3(d) by taking  $X(t)$  as earlier and  $Y(t) = N(t)$ .
- As another special case, derive the corresponding equations for the process in which  $Y(t)$  simply counts the jumps in the process  $X(\cdot)$ , which is otherwise unobserved.

[Hint: See Rudemo (1972, 1973) for further details and special cases.]

14.4.2 Consider a Neyman–Scott cluster process with Poisson cluster centre process at rate  $\mu$  and clusters of random size  $\nu$ ,  $\Pr\{\nu = j\} = q_j$  for  $j = 0, 1, \dots$ , located at independent exponentially distributed distances, mean  $1/\lambda$ , from the cluster centre. The observed process  $N(t)$  consists of both the cluster centres and the points of the cluster, without distinction, lying in the interval  $(0, t]$ .

Write  $X(t)$  for the numbers of points generated from centres  $t'_j < t$  at locations  $t_i \geq t$ . Then  $X(\cdot)$  is an unobserved Markov process on  $\{0, 1, \dots\}$  governing the observed points, and  $\lambda^*(t) = \mu + \lambda X(t-)$ . Show that the process fits the context of Exercise 14.4.1, albeit a countable state space for  $X(\cdot)$ , when  $Y(t) = N(t)$  and the nonzero off-diagonal transition rates are given for all  $i, j, r = 0, 1, \dots$  by  $q_{(i,r)(i+j,r+1)} = \mu q_j$  and  $q_{(i,r)(i-1,r+1)} = \lambda i$ . Deduce that the joint p.g.f.  $P_t(w, z) = E(w^{X(t)} z^{Y(t)} | X(0))$  is given by

$$P_t(w, z) = [z + (w - z)e^{-\lambda t}]^{X(0)} \exp \left( -\mu \int_0^t [1 - zQ(z + (w - z)e^{-\lambda u})] du \right),$$

and that

$$\lim_{t \rightarrow \infty} E(w^{X(t)}) = \exp \left( -\frac{\mu}{\lambda} \int_0^1 \frac{1 - Q(1 - (1 - w)v)}{v} dv \right).$$

[Hint: See also Jowett and Vere-Jones (1972).]

14.4.3 To complete the proof of Proposition 14.4.I, suppose that  $\mathcal{P}_2 \ll \mathcal{P}_1$ , where  $\mathcal{P}_1$  has  $\mathcal{H}$ -intensity  $\lambda_1(\cdot)$ . Arguing as in Proposition 7.1.III, this implies that  $\mathcal{P}_2$  is absolutely continuous with respect to a Poisson process at unit rate, and thus has  $\mathcal{H}$ -intensity  $\lambda_2(\cdot)$  say. Set  $\mu(t, \omega) = \lambda_2(t, \omega)/\lambda_1(t, \omega)$  if  $\lambda_1(t, \omega) \neq 0$ ,  $= 1$  otherwise, and let  $\tilde{\mu}$  be the predictable projection of  $\mu$ . Show that  $\mu$  defined by (14.4.2) has the properties of the Radon–Nikodym derivative  $d\mathcal{P}_2/d\mathcal{P}_1$ .

14.4.4 Let  $N = N_1 + N_2$  be the sum of two Poisson processes, for which the parameter measures are  $\mu$  times Lebesgue measure, and a purely atomic measure with atoms of fixed mass  $a$  at each integer. Write down the likelihood ratio for  $N$  against a reference measure of the same type, with  $\mu = a = 1$ , and verify that it can be written in the form (14.4.4). [Hint: The term due to the atomic component has the same structure as a set of i.i.d. Poisson variables.]

14.4.5 (a) For a process with  $\mathcal{F}$ -intensity  $\lambda^{\mathcal{F}}$  let  $L_t$  be its likelihood ratio relative to a unit rate Poisson process (see around (14.4.7)). If  $\mathcal{G}_t \subseteq \mathcal{F}_t$  (all  $t \geq 0$ ) for a history  $\mathcal{G}$ ,  $E(L_t | \mathcal{G}_t)$  has the same form as  $L_t$  but with  $\mathcal{G}$ -intensity  $E(\lambda_t^{\mathcal{F}} | \mathcal{G}_{t-})$  in place of  $\lambda^{\mathcal{F}}$ . [Hint: From the integral equation for  $L_t$  construct  $E(L_t | \mathcal{G}_t)$ . Deduce that, because  $N_s$  is  $\mathcal{G}_t$ -measurable for  $t > s$ ,

$$\begin{aligned} E_1 \left[ \int_0^t L_{s-} [1 - \lambda^{\mathcal{F}}(s)] (dN(s) - ds) \mid \mathcal{G}_t \right] \\ = \int_0^t E_1 [L_{s-} (1 - \lambda^{\mathcal{F}}(s)) \mid \mathcal{G}_t] (dN(s) - ds). \end{aligned}$$

Now  $L_{s-}$  and  $1 - \lambda^{\mathcal{F}}(s)$  are  $\mathcal{G}_{s-}$ -measurable, so the integrand on the right-hand side equals  $E_1(L_{s-} | \mathcal{G}_{s-}) E_2(1 - \lambda^{\mathcal{F}}(s) | \mathcal{G}_{s-})$   $\mathcal{P}_2$ -a.s.]

(b) Show similarly that (14.4.7) holds [cf. Brémaud (1981, Chapter VI.3)].

14.4.6 Derive the updating equations (10.3.6–7) of Example 10.3(d) from (14.4.8) by using standard Chapman–Kolmogorov equations for the derivatives of transition probabilities and considering (14.4.8) at and between jumps of  $N(\cdot)$ .

Convert these equations into updating equations for the conditional probabilities  $\Pi_t(\cdot)/L_t$ , so that with  $\hat{\pi}_i(t) = \mathcal{P}\{X(t) = i | \mathcal{H}_t\}$ , either  $t$  is a point of the realization in  $\mathcal{H}_t$  and  $\hat{\pi}_i(t) = (\lambda_i/\hat{\lambda})\hat{\pi}_i(t-)$ , or else

$$\frac{\partial \hat{\pi}_i(t)}{\partial t} = \sum_{j \neq i} q_{ji} \hat{\pi}_j(t) - (\lambda_i - \hat{\lambda}) \hat{\pi}_i(t).$$

14.4.7 Suppose in Example 14.4(a) that the transition rates ( $q_{ij}$ ) and the rates  $\lambda_i$  are functions of some parameter  $\alpha$ , and give  $\alpha$  some prior distribution. Extend the updating equations [cf. also Example 14.3(a)] to obtain the integral equation for the joint statistic  $h_i(\alpha, t)$  for  $X(t), \alpha, N(\cdot)$ , corresponding to (14.4.8),

$$h_i(\alpha, t) = \sum_j \int_0^t h_j(\alpha, s) (\mu_j(\alpha) - 1) p_{ij}(t - s; \alpha) dZ(s).$$

Consider the special case where  $X(\cdot)$  is  $\{1, 2\}$ -valued, with unknown emission rates  $\lambda_1, \lambda_2$  but known exponential holding times in each state. Investigate the consequences of assuming prior distributions for  $\lambda_1, \lambda_2$  that are independent gamma distributions.

## 14.5. A Central Limit Theorem

If the Bayesian approach, with its close links with updating formulae, has been the main focus of attention in the engineering literature, there has also been a substantial development in the application of ‘classical’ statistical procedures to inference problems for point processes. In particular, the monograph of Kutoyants (1980), especially in its revised edition in English [Kutoyants (1984b)], develops the asymptotic results for maximum likelihood estimates based on the representations given earlier in this section. An important part of this development is establishing asymptotic normality for the likelihood derivatives  $D_\theta \log L_t$  (as usual,  $D_\theta \equiv \partial/\partial\theta$ ), where  $\theta$  is the parameter under study. From the representation at (7.2.4) or (14.4.2) it is readily seen that under suitable regularity conditions these take the form

$$\int_0^t \frac{D_\theta \lambda^{\mathcal{H}}(s)}{\lambda^{\mathcal{H}}(s)} [dN(s) - ds].$$

Evaluated at the true parameters, such integrals reduce to integrals with respect to the point process martingale  $Z(s) = N(s) - A(s)$ , and hence they are themselves martingales. It is then possible to apply to them general central limit theorems for martingales, as, for example, did Rebolledo (1980), or else to develop versions of such theorems specially tailored for the point process context, as did Kutoyants (1979, 1984b). We follow the latter approach and give a slight extension to Kutoyants’ work to allow for the possibility of a random variance term that leads to a mixed normal distribution. A convenient framework for this extension is the concept of stable convergence in distribution, as described in Section A3.2. However, we do not use the full strength of stable convergence with respect to the  $\sigma$ -algebra  $\mathcal{F}_\infty$  but only stable convergence with respect to the  $\sigma$ -algebra generated by the limit r.v. itself, as this is all that is needed here to discuss the convergence to mixtures of normals. Jarupskin (1984) obtained stronger forms, requiring further conditions.

**Theorem 14.5.I.** *Let  $N$  be a simple point process on  $\mathbb{R}_+$ ,  $\mathcal{F}$ -adapted and with continuous  $\mathcal{F}$ -compensator  $A$ . Suppose that for each  $T > 0$  an  $\mathcal{F}$ -predictable process  $f_T(t)$  is given and that there exists a positive  $\mathcal{F}_0$ -measurable random variable  $\eta$  such that*

- (i)  $E(\int_0^T [f_T(u)]^2 dA(u)) < \infty$ ;
- (ii) as  $T \rightarrow \infty$ ,  $E(\int_0^T [f_T(u)]^2 dA(u)) \rightarrow \eta^2$  in probability; and
- (iii) there exists  $\delta > 0$  such that as  $T \rightarrow \infty$ ,

$$E\left[\int_0^T |f_T(u)|^{2+\delta} dA(u)\right] \rightarrow 0.$$

*Then the random integral*

$$X_T = \int_0^T f_T(x) [dN(x) - dA(x)]$$

converges  $\mathcal{F}_0$ -stably in distribution to a limit random variable  $U\eta$ , where  $U$  is independent of  $\mathcal{F}_0$  and has a normal  $N(0, 1)$  distribution.

PROOF. It is most convenient to make use of the form (A3.2.10) for stable convergence, because the exponential formula can again be used to good effect to simplify the limiting form of expectations  $E[Z \exp(iyX_T)]$ .

To this end, consider the process, for fixed real  $y$ ,

$$\zeta_T(t, y) = \exp \left( iy \int_0^t f_T(u) [dN(u) - dA(u)] + \frac{1}{2} y^2 \int_0^t [f_T(u)]^2 dA(u) \right).$$

Here,  $A(t)$  and  $N(t)$  are, respectively, continuous and pure jump processes, so  $\zeta_T(t, y)$  can be written in terms of its continuous and jump components as

$$\begin{aligned} \zeta_T(t, y) = \exp & \left( \int_0^t \left( \frac{1}{2} y^2 [f_T(u)]^2 - iy f_T(u) \right) dA(u) \right) \\ & \times \prod_i \left[ 1 + (\exp[iy f_T(t_i)] - 1) \Delta N(t_i) \right], \end{aligned} \quad (14.5.1)$$

where the product is taken over the jump points  $t_i$  of the realization of  $N$  over the interval  $(0, t]$ , and because  $N$  is assumed simple,  $\Delta N(t_i) = 1$  a.s. for all  $i$ . Comparing this expression for  $\zeta_T(t, y)$  with the exponential formula at (4.6.2), it can be deduced that  $\zeta_T(t, y)$  is the unique solution, bounded in  $[0, T]$ , of the integral equation

$$\begin{aligned} & \zeta_T(t, y) - 1 \\ &= \int_0^t \zeta_T(u-, y) \left[ \left[ \frac{1}{2} y^2 [f_T(u)]^2 - iy f_T(u) \right] dA(u) + [\exp[iy f_T(u)] - 1] dN(u) \right] \\ &= \int_0^t \zeta_T(u-, y) (\exp[iy f_T(u)] - 1) [dN(u) - dA(u)] \\ & \quad + \int_0^t \zeta_T(u-, y) (\exp[iy f_T(u)] - 1 - iy f_T(u) + \frac{1}{2} y^2 [f_T(u)]^2) dA(u). \end{aligned}$$

Now let  $\tau$  denote the stopping time (recall that  $\eta$  is  $\mathcal{F}_0$ -measurable)

$$\tau = \inf \left\{ t: \int_0^t [f_T(u)]^2 dA(u) \geq \eta^2 \right\},$$

and let  $Z$  be any  $\mathcal{F}_0$ -measurable, essentially bounded random variable. Setting  $t = T \wedge \tau$  in the integral equation, multiplying by  $Z$ , and taking conditional expectations with respect to  $\mathcal{F}_0$ , the optional sampling theorem implies that

$$E \left[ Z \int_0^{T \wedge \tau} \zeta_T(u-, y) (\exp[iy f_T(u)] - 1) [dN(u) - dA(u)] \mid \mathcal{F}_0 \right] = 0;$$

because both of  $Z\zeta_T(u-, s)$  (which is left-continuous) and  $Z(\exp[iyf_T(u)] - 1)$  are  $\mathcal{F}$ -predictable, the latter function is bounded, and the integral on  $(0, t)$  is an  $\mathcal{F}$ -martingale. Thus, we obtain the estimate

$$|\mathbb{E}([Z\zeta_T(T \wedge \tau) | \mathcal{F}_0] - Z)| \leq \mathbb{E} \left[ |Z| \int_0^{T \wedge \tau} |\zeta_T(u-, y)| |R(f_T(u), y)| dA(u) \mid \mathcal{F}_0 \right],$$

where

$$R(f_T(u), y) = \exp[iyf_T(u)] - 1 - iyf_T(u) + \frac{1}{2}y^2[f_T(u)]^2.$$

If  $\delta > 0$  is chosen as in condition (iii) of the theorem, then some finite  $C(\delta)$  exists such that

$$|R(f_T(u), y)| \leq C(\delta) |y|^{2+\delta} |f_T(u)|^{2+\delta}.$$

Furthermore, on  $0 < u < T \wedge \tau$ ,

$$|\zeta_T(u-, y)| \leq \exp \left( \frac{1}{2}y^2 \int_0^{T \wedge \tau} [f_T(u)]^2 dA(u) \right) \leq \exp(\frac{1}{2}y^2\eta^2).$$

Making use of these inequalities, we obtain the further estimate, writing  $\|Z\| = \text{ess sup } |Z(\omega)|$ ,

$$\begin{aligned} & |\mathbb{E}[Z\zeta_T(T \wedge \tau) | \mathcal{F}_0] - Z| \\ & \leq \|Z\| C(\delta) |y|^{2+\delta} \mathbb{E} \left[ \int_0^{T \wedge \tau} |f_T(u)|^{2+\delta} dA(u) \mid \mathcal{F}_0 \right]. \end{aligned} \quad (14.5.2)$$

Multiplying through by the  $\mathcal{F}_0$ -measurable function  $\exp(-\frac{1}{2}y^2\eta^2)$  shows that the right-hand side of (14.5.2) is an upper bound on

$$|\mathbb{E}(Z[\rho_T e^{iyX_T} - e^{-y^2\eta^2/2}] | \mathcal{F}_0)|, \quad (14.5.3)$$

where  $\rho_T$  equals

$$\exp \left[ iy \int_{T \wedge \tau}^T f_T(u) [dN(u) - dA(u)] - \frac{1}{2}y^2 \left( \eta^2 - \int_0^T [f_T(u)]^2 dA(u) \right)_+ \right].$$

Taking the expectation of (14.5.3) and its bound from (14.5.2), the bound converges to zero from assumption (iii). Also, because  $|\rho_T| \leq 1$  and  $\rho_T \rightarrow 1$  in probability from assumption (ii),  $\mathbb{E}[Z(\rho_T - 1)e^{iyX_T}] \rightarrow 0$ , so that for (14.5.3) we obtain the limit relation

$$\mathbb{E}(Ze^{iyX_T}) \rightarrow \mathbb{E}(Ze^{-y^2\eta^2/2}).$$

It now follows from Proposition A3.2.IV that there exists a random variable  $X$  such that  $X_T \rightarrow X$  ( $\mathcal{F}_0$ -stably) and for each bounded,  $\mathcal{F}_0$ -measurable  $Z$ ,

$$\mathbb{E}(Ze^{iyX}) = \mathbb{E}(Ze^{-y^2\eta^2/2}).$$

This equality is equivalent to  $E(e^{-sX} \mid \mathcal{F}_0) = e^{-s^2\eta^2/2}$ , and hence to

$$E(e^{iyX/\eta} \mid \mathcal{F}_0) = e^{-y^2/2}.$$

We deduce that  $X/\eta = U$  is independent of  $\mathcal{F}_0$  and has a unit normal distribution.  $\square$

**Corollary 14.5.II.** *Under the conditions of the theorem, the distributions of the random integrals  $X_T$  converge weakly to the mixed normal distribution with characteristic function  $\phi(y) = E(e^{-y^2\eta^2/2})$ . Thus, when  $\eta$  is a.s. constant, the  $X_T$  converge weakly in distribution to the normal distribution  $N(0, \eta^2)$ .*

This corollary merely restates the fact that  $\mathcal{F}_0$ -stable convergence implies weak convergence.

**Corollary 14.5.III.** *If  $B_T^2 = \int_0^T [f_T(u)]^2 dA(u) > 0$ , then the randomly normed integrals  $X_T/B_T$  converge  $\mathcal{F}_0$ -stably in distribution to the unit normal random variable  $U$ .*

PROOF. We use the result from Proposition A3.2.IV that if  $X_n \rightarrow X$  ( $\mathcal{F}$ -stably) then  $g(X_n, Y) \rightarrow g(X, Y)$  ( $\mathcal{F}$ -stably) for bounded continuous functions  $g(\cdot)$ . Supposing first that  $Y$  is bounded away from zero and  $X$  is essentially bounded, we can take  $g(x, y) = x/y$  so that  $X_n/Y \rightarrow X/Y$  ( $\mathcal{F}$ -stably). The constraint on  $X$  is immaterial in that it is given that  $\mathcal{P}\{|X| < \infty\} = 1$  (because  $X$  is a well-defined r.v.). Now suppose also that  $Y_n \rightarrow Y$  in probability, where each  $Y_n$  is a.s. positive and  $\mathcal{F}$ -measurable. Then  $(X_n, Y_n) \rightarrow (X, Y)$  in distribution and thus, again,  $X_n/Y_n \rightarrow X/Y$  ( $\mathcal{F}$ -stably). Finally, by approximating  $Y$  by a sequence of r.v.s bounded away from zero, the result extends to the case  $Y > 0$  a.s. Taking  $Y_n = B_{T_n}$  and  $X_n = X_{T_n}$  for some sequence  $T_n \rightarrow \infty$  and  $\mathcal{F} = \mathcal{F}_0$ , the result follows.  $\square$

The form of condition (iii) is not the most general possible. Kutoyants noted that it may be replaced by a Lindeberg type of condition, although the Liapounov type of condition suffices for most applications. Versions of the theorem for multivariate and MPPs can be given [see Kutoyants (1984a, b)].

The major application of the theorem is to the proof of the asymptotic normality of parameter estimates. This application is discussed and illustrated at length in Kutoyants (1980) for the case of inhomogeneous Poisson processes, and in Kutoyants (1984b) for more general processes. The next two examples may also help illustrate the range of applications for the theorem.

EXAMPLE 14.5(a) *Poisson and mixed Poisson processes.* As the simplest possible example, let  $N$  be a simple Poisson process with rate  $\mu$ . Successful application of the theorem relies on identifying the appropriate norming function  $f_T(\cdot)$  for the quantity of interest. Here, to study  $N(t)$ , recall first that its  $\mathcal{H}$ -compensator is  $\mu t$ . Thus, we need  $f_T(\cdot)$  to satisfy

$$\int_0^T [f_T(u)]^2 \mu dt \rightarrow \text{const.} \quad (T \rightarrow \infty),$$



so  $f_T(u) = T^{-1/2}$  is the simplest choice here, with the constant  $= \mu$  and non-random. Then the left-hand side of (iii) reduces to  $\mu T^{-\delta/2} \rightarrow 0$  ( $T \rightarrow \infty$ ) as required. Thus,

$$\frac{N(T) - \mu T}{T^{1/2}} \rightarrow \mu^{1/2} U \quad \text{in distribution,}$$

with  $U$  a standard normal r.v. as in the theorem.

If in fact the process is mixed Poisson with  $\mu$  a r.v. as in Example 14.3(a), the same conclusion holds provided we use the  $\mathcal{F}$ -compensator with  $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{H}_t$ . Indeed, from Corollary 14.5.III we should have

$$\frac{N(T) - \mu T}{(\mu T)^{1/2}} \rightarrow U \quad \text{in distribution.}$$

If we want to devise a result concerning an estimate of  $\mu$ , it is preferable to express the left-hand side here as  $[N(T)/T - \mu]/(\mu/T)^{1/2}$  and then observe that, as  $T \rightarrow \infty$ , we have  $N(T)/T \rightarrow \mu$  a.s. As a result of this, we can replace  $\mu$  in the denominator and deduce further that

$$\frac{N(T)/T - \mu}{[N(T)]^{1/2}/T} \rightarrow U \quad \text{in distribution,}$$

with  $\mu$  the only quantity on the left-hand side that is unknown at  $T$ .

A final possibility would be to use the  $\mathcal{H}$ -compensator, which in the special case given in Example 14.3(a) has the form

$$dA(t) = \frac{N(t-) + 1}{t + \alpha} dt,$$

and leads to virtually the same conclusions. □

In examples of this kind, where  $\mathcal{F}_0$  is either trivial or very simple, there is little advantage in using the extensions to random norming. Only weak convergence is asserted and the theorem sheds no light on whether the estimates converge  $\mathcal{H}_\infty$ -stably, for example. For detail on this question see Jarupskin (1984). It also underlies the next example.

**EXAMPLE 14.5(b) Simple birth process.** This is a standard example [see, e.g., Keiding (1975); Basawa and Scott (1983)] for showing ‘nonergodic’ behaviour in the sense that the asymptotic distribution of the maximum likelihood estimate is not normal but a mixture of normals. If the probability of an individual producing offspring in time  $(t, t + dt)$  is  $\lambda dt$ , and all individuals reproduce independently, it is known that with  $N(t)$  denoting the sum of the initial number  $n_0$  and the number of individuals born in  $(0, t]$  and  $q_t = 1 - p_t = e^{-\lambda t}$ ,

$$\mathcal{P}\{N(t) = n\} = q_t^{n_0} p_t^{n-n_0} \binom{n-1}{n-n_0}, \quad (14.5.4)$$

that

$$N(t)e^{-\lambda t} \rightarrow W \quad \text{a.s.}, \quad (14.5.5)$$

where  $W$  is a r.v. which, if  $n_0 = 1$ , has the unit exponential distribution, and that

$$\hat{\lambda}_t = \frac{N(t) - n_0}{\int_0^t N(u) du}$$

is the maximum likelihood estimate of  $\lambda$ . Clearly, the process may be treated as a point process, and it is then of interest to see what light the present methods shed on the behaviour of the likelihood estimate.

The conditional intensity of the process with respect to the internal history  $\mathcal{H}$  generated by the  $N(t)$  themselves is just equal to  $\lambda N(t-)$ . If we use this history, the first derivative of the likelihood of the process on  $(0, T)$  is proportional to

$$N(T) - \lambda \int_0^T N(t-) dt,$$

which has variance function

$$\lambda \int_0^T N(t) dt \sim E(e^{\lambda T}).$$

This suggests that the norming factor  $k(T) = e^{-\lambda T/2}$  is appropriate, but because  $W$  is not  $\mathcal{F}_0$ -measurable with this choice of history, further discussion is required. In fact, what is needed is the  $\mathcal{F}$ -intensity when  $\mathcal{F}_0 = \sigma\{W\}$  and  $\mathcal{F}_t = \mathcal{F}_0 \vee \mathcal{H}_t^{(N)}$ . The history  $\mathcal{F}$  is a refinement of the internal history, and to find the  $\mathcal{F}$ -compensator we have to discuss the behaviour of the process conditional on the value of  $W$ . This can be computed by writing down from (14.5.4) the joint distribution of  $N(s)$  and  $N(t)$  for  $s > t$ , conditioning on  $N(s)$ , and letting  $s \rightarrow \infty$ , taking into account (14.5.5) and using Stirling's formula [cf. Keiding (1975)]. The result can be stated as follows. Given  $N(t)$  and  $W$ , the conditional distribution of  $N(s) - N(t)$  is Poisson with parameter

$$\lambda(s | t, W) = W e^{\lambda t} (e^{\lambda(s-t)} - 1).$$

Hence the  $\mathcal{F}$ -intensity is

$$\lambda^{\mathcal{F}}(t) = \lambda W e^{\lambda t}.$$

Note that  $E[\lambda^{\mathcal{F}}(t) | \mathcal{H}_t] = \lambda e^{\lambda t} E[W | N(t-)] = \lambda e^{\lambda t} N(t-) e^{-\lambda t} = \lambda N(t-)$ , which is just the  $\mathcal{H}$ -compensator if a predictable version of  $N(t)$  is taken.

We now consider the asymptotic behaviour of the scaled difference

$$\Delta(T) = e^{-\lambda T/2} (\hat{\lambda}_T - \lambda) = \frac{e^{-\lambda T/2} [N(T) - \lambda \int_0^T N(u) du]}{\lambda e^{-\lambda T} \int_0^T N(u) du}. \quad (14.5.6)$$

Applying the theorem, we find after simple computations that the pair

$$e^{-\lambda T/2} \int_0^T [dN(u) - \lambda W e^{\lambda u} du] = e^{-\lambda T/2} [N(T) - n_0 - W(e^{\lambda T} - 1)]$$

and

$$\begin{aligned} e^{-\lambda T/2} \int_0^T \lambda(T-u) [dN(u) - \lambda W e^{\lambda u} du] \\ = e^{-\lambda T/2} \left[ \int_0^T N(u) du - n_0 T - W(e^{\lambda T} - 1 - \lambda T) \right] \end{aligned}$$

converges  $\mathcal{F}_0$ -stably to the pair  $(Z_1 W^{1/2}, Z_2 W^{1/2})$ , where  $Z_1, Z_2$  are independent of  $\mathcal{F}_0$  and jointly normally distributed with covariance matrix  $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ , so that in fact we can write  $Z_2 = Z_1 - Z'_1$  where  $Z_1, Z'_1$  are independent unit normal r.v.s. Thus, the numerator in the term on the right-hand side of (14.5.6) converges  $\mathcal{F}_0$ -stably to  $Z'_1 W^{1/2}$  and from (14.5.5) the denominator converges a.s. to  $W$ . Hence,

$$\Delta(t) \sim Z' W^{-1/2} \quad (\mathcal{F}_0\text{-stably}).$$

but an exponential r.v.  $W$  has just the form  $\frac{1}{2}\chi_{(2)}^2$ , where  $\chi_{(2)}^2$  denotes a chi-square r.v. on two degrees of freedom, and so the ratio has a  $t$ -distribution on two degrees of freedom [again, see Keiding (1975)].  $\square$

## Exercises and Complements to Section 14.5

14.5.1 For the self-correcting or stress-release model of Isham and Westcott (1979) for which

$$\lambda(t) = \exp(\alpha + \beta[t - \rho N(t)]),$$

show that conditions for estimators of  $\beta$  and  $\rho$  to have central limit theorem properties hold for  $\beta > 0$  and  $\rho > 0$  but fail when  $\beta = 0$ .

[Hint: When  $\beta > 0, \rho > 0$ , the process  $X(t) = t - \rho N(t)$  is Markovian and the law of large numbers implies that condition (ii) of Theorem 14.5.I holds, but this fails when  $\beta = 0$ . See also Vere-Jones and Ogata (1984).]

## 14.6. Random Time Change

The topics of both this section and the next describe methods for reducing more general point processes to Poisson processes, emphasizing yet again the fundamental role played by Poisson processes in point process theory.

We start by recapitulating some introductory material, including Watanabe's (1964) characterization of the Poisson process as a process with deterministic compensator. The time-change theorems in this section were introduced in Section 7.4 and linked there to Ogata's (1988) residual analysis for

checking the goodness-of-fit for a point process model [and now extended to the wider range of residual methods introduced for spatial point processes by Baddeley and co-workers, reviewed in Baddeley *et al.* (2005) and discussed briefly in Sections 15.4–5 below]. Our main goal here is the extension of the time-change theorem to MPPs, first proved for multivariate point processes by Meyer (1971) [but see also Doléans-Dade (1970)] using orthogonal martingale arguments, and later extended and simplified by Brown and Nair (1988). The proof we give is for general MPPs, and appears to be new, combining generating functional arguments with the use of the exponential formula much as in Brémaud (1981) and Brown and Nair, but avoiding the second-order theory used in the orthogonal martingale arguments. To illustrate the technique we first use it to give an extension of Watanabe's theorem to Cox processes; in essence the proof is a minor variation of those in Brown (1978) and Brémaud.

Random time-change results for spatial processes are much more problematic: see Nair (1990) and Schoenberg (1999).

**Theorem 14.6.I.** *Let  $N$  be a simple point process on  $\mathbb{R}_+$  adapted to the history  $\mathcal{F}$ . If the  $\mathcal{F}$ -compensator  $A$  of  $N$  is continuous and  $\mathcal{F}_0$ -measurable, then  $N$  is a Cox process directed by  $A$ .*

PROOF. With a view to characterizing the process via its p.g.fl., take a fixed continuous nonnegative  $h \in \mathcal{V}(\mathbb{R}_+)$  (cf. Definition 9.4.IV), with  $h(0) = 1$  and  $h(u) = 1$  outside a finite interval  $[0, T)$ , and for a given realization  $\{t_i\}$  of the process consider the expression

$$\Phi(h) = \left[ \prod_{0 \leq t_i < \infty} h(t_i) \right] e^{-\int_0^\infty [h(u)-1] dA(u)}. \quad (14.6.1)$$

In fact both the product and the integral can be restricted to any finite interval  $(0, T)$  outside which  $h(u) = 1$ , without altering their values. Now identify the quantity

$$H(t) = \left[ \prod_{0 \leq t_i < t} h(t_i) \right] e^{-\int_0^t [h(u)-1] dA(u)}. \quad (14.6.2)$$

with the solution (4.6.2) of the integral equation (4.6.3), namely,

$$H(t) = H(0) + \int_0^t H(x-) u(x) dF(x), \quad (4.6.3)$$

by setting  $F(x) = N(x) - A(x)$  ( $x > 0$ ) and  $u(x) = h(x) - 1$ . Thus,  $H(t)$  satisfies

$$H(t) = 1 + \int_0^t H(s-) [h(s) - 1] [dN(s) - dA(s)]. \quad (14.6.3)$$

The integrand on the right-hand side is left-continuous and  $\mathcal{F}$ -adapted, so it is  $\mathcal{F}$ -predictable (see the discussion following Lemma A3.3.I). It now follows

from Exercise 14.1.7 that the integral is an  $\mathcal{F}$ -martingale, so on taking conditional expectations with respect to  $\mathcal{F}_0$ , we obtain  $E[H(t) \mid \mathcal{F}_0] = 1$ . Letting  $t \rightarrow \infty$  and appealing to monotone convergence yields  $E[\Phi(h) \mid \mathcal{F}_0] = 1$ .

By assumption the compensator  $A$  is  $\mathcal{F}_0$ -measurable, so the term with the exponential [see (14.6.2)] can be taken outside the expectation, and this last equation can be rewritten as

$$E\left[\prod_{\{t_i\}} h(t_i) \mid \mathcal{F}_0\right] = \exp\left[\int_0^\infty (h(u) - 1) dA(u)\right], \quad (14.6.4)$$

which we recognize as identifying the p.g.fl. of the given point process, conditioned by  $\mathcal{F}_0$ , with that of a Poisson process with (given random) mean  $A(t)$ , so unconditionally it is a Cox process directed by  $A(\cdot)$ . The uniqueness theorem for p.g.fl.s completes the proof.  $\square$

**Corollary 14.6.II.** *Let  $N$  be a simple point process with internal history  $\mathcal{H}$  and let its  $\mathcal{H}$ -intensity be the deterministic function  $\mu(\cdot)$ . Then  $N$  is a Poisson process with density function  $\mu(\cdot)$ .*

Watanabe discussed the special case  $\mu(t) = \mu$  (all  $t > 0$ ). Both his result and its generalization are closely related to a theorem in Papangelou (1972), already stated in Theorem 7.4.I, that any simple point process with continuous compensator is locally Poisson in character, in the sense that there exists a local transformation of the time axis that converts the process into a Poisson process. We develop a proof based on a use of the exponential formula similar to that of Theorem 14.6.I.

Consider then a simple point process that is  $\mathcal{F}$ -adapted for some general history  $\mathcal{F}$  for which  $A$  is the  $\mathcal{F}$ -compensator, and consider the time-change defined by  $\tau = A(t)$  ( $t \in \mathbb{R}_+$ ), equivalently,  $t = A^{-1}(\tau) = \inf\{t: A(t) \geq \tau\}$ . Note that if the compensator  $A$  is continuous, then  $A^{-1}$  is right-continuous (and monotonic, like  $A$ ), with jumps at the at most countable set of values of  $\tau$  of constancy of  $A$ , and  $A(A^{-1}(\tau)) = \tau$  for all  $\tau > 0$ . It follows (see Lemma A3.3.III) that for every  $\tau$ ,  $A^{-1}(\tau)$  is an  $\mathcal{F}$ -stopping time. Moreover, the  $\sigma$ -algebra

$$\tilde{\mathcal{F}}_\tau \equiv \mathcal{F}_{A^{-1}(\tau)}$$

is well defined (Definition A3.4.V), with  $\tilde{\mathcal{F}}_\tau \subseteq \tilde{\mathcal{F}}_v$  for  $\tau \leq v$  (Theorem A3.4.VII), and  $\tilde{\mathcal{F}} \equiv \{\tilde{\mathcal{F}}_\tau: 0 < \tau < \infty\}$  constitutes a history for the process  $\tilde{N}(\tau) = N(A^{-1}(\tau))$ .

We now imitate the proof of Proposition 14.6.I, but take the p.g.fl. in the transformed space rather than the original space. Write  $\tau_i = A(t_i)$ , and again suppose  $h(\cdot) \in \mathcal{V}(\mathbb{R}_+)$  is continuous and equal to unity outside some finite interval  $(0, T)$ . In place of (14.6.1), consider

$$\tilde{\Phi}(h) = \left[ \prod_{0 \leq \tau_i < \infty} h(\tau_i) \right] e^{-\int_0^\infty [h(\tau) - 1] d\tau}, \quad (14.6.5)$$

and identify it with the integral equation version

$$\tilde{\Phi}(h) = 1 + \int_0^\infty \tilde{H}(\tau-)[h(\tau) - 1][d\tilde{N}(\tau) - d\tau],$$

where  $\tilde{H}(t)$  is defined as in (14.6.2) by the integral up to  $t$ . Now substitute  $\tau = A(t)$ , write  $H(t) = \tilde{H}(A(t))$ , and take expectations. This yields

$$\mathbb{E}[\tilde{\Phi}(h)] = 1 + \mathbb{E}\left[\int_0^\infty H(t-)[h(A(t)) - 1][dN(t) - dA(t)]\right].$$

Because  $h(A(t)) - 1$  and  $H(t-)$  are still both  $\mathcal{F}$ -predictable, the expectation of the integral above vanishes because of the martingale property of  $N - A$ , so that  $\mathbb{E}[\tilde{\Phi}(h)] = 1$ . Now if we take expectations in (14.6.5) we see that

$$\mathbb{E}\left[\prod_{0 \leq \tau_i < \infty} h(\tau_i)\right] = \exp\left(-\int_0^\infty [1 - h(\tau)] d\tau\right),$$

which shows that the p.g.fl. of the transformed time process is that of a unit-rate Poisson process. Hence, because the p.g.fl. determines the process uniquely, we have proved the following result.

**Proposition 14.6.III.** *Let  $N$  be an  $\mathcal{F}$ -adapted nonterminating simple point process on  $\mathbb{R}_+$  with continuous  $\mathcal{F}$ -compensator  $A$  for some history  $\mathcal{F}$ . Then the randomly rescaled process  $\tilde{N}(\tau) = N(A^{-1}(\tau))$  is a unit-rate Poisson process on  $\mathbb{R}_+$ .*

When  $N$  is an  $\mathcal{F}$ -adapted simple MPP on  $\mathbb{R}_+ \times \mathcal{K}$ , we can use a similar argument to establish a comparable time-change transformation of  $N$  into a marked compound Poisson process  $\tilde{N}$ . In place of a continuous  $\mathcal{F}$ -compensator we assume that  $N$  has a simple ground process with finite first moment measure and an  $\mathcal{F}$ -intensity  $\lambda^{\mathcal{F}}(t, \kappa, \omega)$  underlying the  $\mathcal{F}$ -compensator  $A(t, K, \omega)$  whose existence is guaranteed by (14.2.7). We also assume that the MPP is  $\ell_{\mathcal{K}}$ - $\mathcal{F}$  nonterminating with respect to the reference measure  $\ell_{\mathcal{K}}$  on the mark space  $(\mathcal{K}, \mathcal{B}_{\mathcal{K}})$ , meaning that

$$a(t, \kappa, \omega) \equiv \int_0^t \lambda^{\mathcal{F}}(u, \kappa, \omega) du \rightarrow \infty \quad (t \rightarrow \infty) \quad \ell_{\mathcal{K}} \times \mathcal{P}\text{-a.e.}, \quad (14.6.6)$$

and from Proposition 14.3.II(b),

$$\lambda^{\mathcal{F}}(t, \kappa, \omega) = \lambda_g^{\mathcal{F}}(t, \omega) f(\kappa | t, \omega) \quad (\ell \times \ell_{\mathcal{K}} \times \mathcal{P})\text{-a.e.}$$

Recall from Lemma 6.4.VI that a compound Poisson process with mark kernel  $F(\cdot | \cdot)$  and ground Poisson process with measure  $\mu(\cdot)$  has p.g.fl.  $G[h] = \mathbb{E}[\exp(\int_{\mathcal{X} \times \mathcal{K}} \log h(x, \kappa) N(dx, d\kappa))]$  for nonnegative measurable functions  $h(x, \kappa)$  in  $\mathcal{V}(\mathcal{X} \times \mathcal{K})$ . We evaluate  $G[h]$  as

$$\exp\left(\int_{\mathcal{X}} \int_{\mathcal{K}} [h(x, \kappa) - 1] F(d\kappa | x) \mu(dx)\right) = \exp\left(\int_{\mathcal{X}} [h_F(x) - 1] \mu(dx)\right),$$

where  $h_F(x) = \int_{\mathcal{K}} h(x, \kappa) F(d\kappa | x)$ , and  $h_F \in \mathcal{V}(\mathcal{X})$ .

**Theorem 14.6.IV** (MPP Random Time-change). *Let  $N$  be an  $\mathcal{F}$ -adapted MPP on  $\mathbb{R}_+$  with simple ground process having finite first moment measure and whose  $\mathcal{F}$ -compensator  $A$  admits an  $\mathcal{F}$ -intensity  $\lambda^{\mathcal{F}}(u, \kappa, \omega)$  such that  $N$  is  $\ell_{\mathcal{K}}$ - $\mathcal{F}$  nonterminating. Let  $\tilde{N}$  denote the rescaled MPP defined to have a point at  $(a(t, \kappa, \omega), \kappa)$  if and only if  $N$  has a point at  $(t, \kappa)$ , where  $a(t, \kappa, \omega)$  is defined by (14.6.6). Then  $\tilde{N}$  is a stationary compound Poisson process with unit ground intensity and stationary mark distribution  $\ell_{\mathcal{K}}$ .*

PROOF. In place of the function  $h$  of (14.6.1), consider a nonnegative function  $h(t, \kappa)$  which is jointly continuous in the two variables  $(t, \kappa)$  and satisfies  $h(t, \kappa) = 1$  whenever  $t = 0$ , or  $t > T$  for some finite  $T$ . Mimicking more (14.6.5) rather than (14.6.1), consider

$$H^{\mathcal{F}}(t) = \left[ \prod_{0 \leq t_i \leq t} h(a(t_i, \kappa_i, \omega), \kappa_i) \right] e^{-\int_0^t \int_{\mathcal{K}} [h(a(s, \kappa, \omega), \kappa) - 1] \lambda^{\mathcal{F}}(s, \kappa) \ell_{\mathcal{K}}(d\kappa) ds}. \quad (14.6.7)$$

In the exponential term here, (14.6.6) implies that  $h(a(s, \kappa)) = 1$  for sufficiently large  $s$  for  $\ell_{\mathcal{K}}$ -almost all  $\kappa$ , so an integral there taken instead over  $(0, \infty)$  remains finite. First, however, use the factorization  $\lambda^{\mathcal{F}} = \lambda_g f$  already noted and consider the integral of  $\kappa$ -dependent terms over  $\mathcal{K}$ : define both

$$u^{\mathcal{F}}(s) = \int_{\mathcal{K}} [h(a(s, \kappa), \kappa) - 1] f(\kappa | s) \ell_{\mathcal{K}}(d\kappa) = \bar{h}(s) - 1, \quad (14.6.8a)$$

where  $\bar{h}(s) = \int_{\mathcal{K}} h(a(s, \kappa), \kappa) f(\kappa | s) \ell_{\mathcal{K}}(d\kappa)$ , and a process  $F^{\mathcal{F}}$  on  $\mathbb{R}_+$  by

$$\Delta F^{\mathcal{F}}(t_i) = h(a(t_i, \kappa_i), \kappa_i) / \bar{h}(t_i) \quad \text{and} \quad dF_c^{\mathcal{F}}(s) = -\lambda_g(s) ds, \quad (14.6.8b)$$

where the  $(t_i, \kappa_i)$  are the points of the given realization  $N$  and  $s \notin \{t_i\}$ . The product terms in the exponential formula (4.6.2) are then of the type

$$(1 + u^{\mathcal{F}}(t_i)) \Delta F^{\mathcal{F}}(t_i) = \bar{h}(t_i) h(a(t_i, \kappa_i), \kappa_i) / \bar{h}(t_i) = h(a(t_i, \kappa_i), \kappa_i),$$

and the integral term in (4.6.2) is of the form

$$\int_0^t u^{\mathcal{F}}(s) dF_c^{\mathcal{F}}(s) = - \int_0^t \lambda_g(s) \int_{\mathcal{K}} [h(a(s, \kappa), \kappa) - 1] f(\kappa | s) \ell_{\mathcal{K}}(d\kappa) ds.$$

Hence both terms in (14.6.7) are recovered.

Because  $h$ , and hence  $\bar{h}$ , are bounded functions, and the expected number of points in any finite interval is finite (assumed throughout the chapter), the boundedness requirements of Lemma 4.6.II are satisfied, and we can identify  $H^{\mathcal{F}}(t)$  of (14.6.7) with the solution of the integral equation (4.6.3), written here as

$$H^{\mathcal{F}}(t) = 1 + \int_0^t H^{\mathcal{F}}(s-) [\bar{h}(s) - 1] dF^{\mathcal{F}}(s). \quad (14.6.9)$$

We assert that the process  $F^{\mathcal{F}}(s)$  is an  $\mathcal{F}$ -martingale; to see this, write it in the form

$$dF^{\mathcal{F}}(s) = \int_{\mathcal{K}} \frac{h(a(s, \kappa), \kappa)}{\bar{h}(s)} N(ds \times d\kappa) - \lambda_g(s) ds.$$

For fixed  $\kappa$ , the function  $h(a(s, \kappa), \kappa)$  is  $\mathcal{F}$ -predictable, as therefore is the integral  $\bar{h}(s) = \int_{\mathcal{K}} h(a(s, \kappa), \kappa) \ell_{\mathcal{K}}(d\kappa)$ . Taking expectations over a range  $(a, b)$  therefore yields

$$\begin{aligned} & \mathbb{E} \left[ \int_a^b dF^{\mathcal{F}}(s) \mid \mathcal{F}_a \right] \\ &= \mathbb{E} \left[ \int_a^b \left\{ \int_{\mathcal{K}} \frac{h(a(s, \kappa), \kappa)}{\bar{h}(s)} f(\kappa \mid s) \ell_{\mathcal{K}}(d\kappa) - 1 \right\} \lambda_g(s) ds \mid \mathcal{F}_a \right] = 0, \end{aligned}$$

where we see from the definition of  $\bar{h}$  that the integrand vanishes for each  $s$ .

Next, let  $t \rightarrow \infty$  in (14.6.9), denoting the limit by  $\Phi(h)$ , and take expectations. Expanding the integral on the right-hand side of (14.6.9) and appealing to Fubini's theorem we obtain

$$\begin{aligned} & \mathbb{E} \left[ \int_0^\infty H^{\mathcal{F}}(x-) [\bar{h}(x) - 1] dF^{\mathcal{F}}(x) \right] \\ &= \mathbb{E} \left[ \int_0^\infty H^{\mathcal{F}}(x-) \int_{\mathcal{K}} [h(a(x, \kappa), \kappa) - 1] f(\kappa \mid x) \ell_{\mathcal{K}}(d\kappa) dF^{\mathcal{F}}(x) \right] \\ &= \int_{\mathcal{K}} \mathbb{E} \left[ \int_0^\infty H^{\mathcal{F}}(x-) [h(a(x, \kappa), \kappa) - 1] f(\kappa \mid x) dF^{\mathcal{F}}(x) \right] \ell_{\mathcal{K}}(d\kappa). \end{aligned}$$

Again, for fixed  $\kappa$ , the functions in the integrand are  $\mathcal{F}$ -predictable, and  $F^{\mathcal{F}}(\cdot)$  is a martingale, so that the expectation vanishes. This leaves us with the result

$$\begin{aligned} 1 &= \mathbb{E}[\Phi(h)] \\ &= \mathbb{E} \left[ \left( \prod h(a(t_i, \kappa_i), \kappa_i) \right) e^{-\int_0^\infty \int_{\mathcal{K}} [h(a(t, \kappa), \kappa) - 1] f(\kappa \mid t) \ell_{\mathcal{K}}(d\kappa) \lambda_g(t) dt} \right] \\ &= \mathbb{E} \left[ \prod h(\tau_i, \kappa_i) \right] e^{-\int_0^\infty \int_{\mathcal{K}} [h(\tau, \kappa) - 1] \ell_{\mathcal{K}}(d\kappa) d\tau} \equiv \mathbb{E} \left[ \prod h(\tau_i, \kappa_i) \right] / G[h], \end{aligned}$$

in which  $G[h]$  is the p.g.fl. of a compound Poisson process with unit-rate ground process and location-independent mark-kernel density  $\ell_{\mathcal{K}}$ . Here we have used the nonterminating property of the compensator to justify taking the integral over  $(0, \infty)$ , and by (14.6.6) and (14.3.2b) the change of variable of integration has density  $f(\kappa \mid t) \lambda_g(t)$ . This integral is nonrandom, so it can be taken outside the expectation, and then identified with the reciprocal of the p.g.fl.  $G[h]$  as just described. The theorem follows.  $\square$

For an earlier attempted proof of this result, with further references, see Vere-Jones and Schoenberg (2004). When the mark space is finite, we obtain the following specialization to multivariate processes.



**Corollary 14.6.V.** *Let  $N$  denote a nonterminating multivariate point process with components  $N_k$ ,  $k = 1, \dots, K$ , reference measure  $\pi = \{\pi_1, \dots, \pi_K\}$  satisfying  $\pi_k > 0$ ,  $k = 1, \dots, K$ ,  $\sum_{k=1}^K \pi_k = 1$ , simple ground process, and  $\mathcal{F}$ -conditional intensity  $\lambda_k(t)$ . Let  $a(t, k) = \int_0^t \lambda_k(s) ds$  and denote by  $\tilde{N}$  the rescaled MPP defined to have a point at  $\{a(t, k), k\}$  if and only if the  $k$ th component of  $N$  contains a point at  $t$ . Then  $\tilde{N}$  is a stationary compound Poisson process with unit intensity and stationary mark distribution  $\pi$ .*

*Equivalently, if the rescaling is performed so that  $\tilde{N}$  has a point at  $(a(t, k)\pi_k, k)$  whenever the original process has a point at  $(t, k)$  (this is the same as the reference measure  $\pi$  assigning unit mass to each component), then the resultant process consists of  $K$  independent, unit-rate Poisson processes, one for each mark.*

Suppose that the mark space is the real line; let  $F_{\ell_K}(m) \equiv \int_{\kappa < m} \ell_K(d\kappa)$  be the cumulative measure corresponding to the probability measure  $\ell_K$ , and suppose that  $F_{\ell_K}(m)$  is continuous as a function of  $m$ . Then a simple rescaling of the mark space, taking  $m^* = F_{\ell_K}(m)$ , converts the stationary mark distribution for the transformed process into the uniform distribution on  $[0, 1]$ . Now a compound Poisson process with constant rate and uniform mark distribution can equally be interpreted as a two-dimensional Poisson process on a strip; hence we obtain the following result.

**Corollary 14.6.VI.** *Suppose that the MPP  $N$  has real marks, that the conditions of Theorem 14.6.IV hold, and that the reference probability measure  $\ell_K$  admits a continuous cumulative version  $F_{\ell_K}$ . Then the doubly transformed process  $\tilde{N}^*$ , defined to have a point at  $(a(t_i, m_i), F_{\ell_K}(m_i))$  when the original process has a point at  $(t_i, m_i)$ , is a two-dimensional Poisson process with unit rate over the half-strip  $\mathbb{R}_+ \times [0, 1]$ .*

The role of absolute continuity in Proposition 14.6.IV and its corollaries, in particular the requirement that the compensator  $A(t, K)$  have a density with respect to the mark  $\kappa$  as well as time  $t$ , is illustrated in the next example.

**EXAMPLE 14.6(a) Poisson process with alternating marks.** Let  $\{t_i\}$  be a realization of a Poisson process at unit rate on  $\mathbb{R}_+$ , and attach to points  $\{t_{2i-1}\}$  the mark 1, and to the points  $\{t_{2i}\}$  the mark 2. Then, writing  $t_0 = 0$  and using the notation of Corollary 14.6.V, for  $i = 1, 2, \dots$ ,

$$(\lambda_1(u), \lambda_2(u)) = \begin{cases} (1, 0) & \text{for } t_{2i-2} \leq u < t_{2i-1}, \\ (0, 1) & \text{for } t_{2i-1} \leq u < t_{2i}. \end{cases}$$

Let  $U_i = t_i - t_{i-1}$  ( $i = 1, 2, \dots$ ), so that for  $t_{2i} \leq t < t_{2i+2}$ ,

$$a(t, k) = \int_0^t \lambda_k(u) du = \begin{cases} U_1 + \dots + U_{2i-1} + \min((t - t_{2i})_+, U_{2i+1}) & (k = 1), \\ U_2 + \dots + U_{2i} + \min((t - t_{2i+1})_+, U_{2i+2}) & (k = 2). \end{cases}$$

It is immediately evident that each case on the right-hand side here is a sum of independent unit exponential random variables, and so corresponds to a

Poisson process with unit rate; furthermore, the two cases are independent. It follows that the conditional intensity of the initial MPP can be written in the form  $\lambda^*(t, \kappa) = \lambda_g^*(t)f(\kappa | t)$ , where  $\lambda_g^*(t) = 1$  and

$$f(0 | t) = i(t) = 1 - f(1 | t),$$

where  $i(t)$  is the parity of  $N(t)$ , so that  $i(t) = 1$  if  $N(t)$  is odd,  $= 0$  otherwise. The mark space here has just two points, and the natural reference measure has unit atoms at each point, so that the function  $f$  described above can be regarded as a density with respect to this measure.

So far, this example has illustrated the behaviour to be expected when the conditions of Corollary 14.6.V are satisfied. Now suppose, as a variant on this example where the conditions are not satisfied, that the sequence  $\{t_i\}$  is as before, but that  $t_{2i-1}$  has mark  $\kappa_{2i-1} = X_i$  that is uniformly distributed in  $(0, 1)$ , independently of the other random variables, and  $\kappa_{2i} = 2 - X_i$ . With  $U_i = t_i - t_{i-1}$  as before, we can define the marked point process compensator for  $K \in \mathcal{B}((0, 2))$  in the form

$$A(t, K) = \begin{cases} (U_1 + \cdots + U_{2i-1} + (t - t_{2i}))I_K(X_i), & t_{2i} \leq t < t_{2i+1}, \\ (U_2 + \cdots + U_{2i-2} + (t - t_{2i-1}))I_K(2 - X_i), & t_{2i-1} \leq t < t_{2i}. \end{cases}$$

Here, in the first case,  $F(\cdot | t)$  is just the uniform distribution on  $(0, 1)$ , but in the second case,  $F(\cdot | t)$  has an atom in  $2 - X_i$ , because the value of  $X_i$  is now known at time  $t$ . For this model, therefore, the absolute continuity requirement is not satisfied, and, as the reader may check, the conclusion of Theorem 14.6.IV fails.  $\square$

## Exercises and Complements to Section 14.6

14.6.1 In Example 14.6(a), show that each of the four terms on the right-hand side of (14.6.15) converges in probability to zero as  $\gamma \rightarrow 0$  on the assumption that  $N(t)/t$  converges a.s. to some limit  $\lambda \equiv \lambda(\omega) \in \mathcal{F}_0^{(\gamma)}$  for all  $\gamma$ .

[Hint: The convergence is shown directly for the first three terms; for the last, investigate  $\sup_{t' < v < t} |1 - \tilde{\zeta}_\gamma(v) - \gamma/\lambda|$  via  $Y_\gamma(t)/t \leq N(t)/t$  a.s. in one direction and  $\liminf_{t \rightarrow \infty} Y_\gamma(t)/t$  in the other direction.]

14.6.2 Apply the random time-change to data such as those shown in Figure 14.1 in Section 14.1.

14.6.3 *Alternative proof of multivariate time change Theorem 14.6.V* [Brown and Nair (1988)]. Consider a  $K$ -variate point process  $\{N_k(t): k = 1, \dots, K\}$  on  $\mathbb{R}_+$ , with occurrence times  $\{t_i^{(k)}: i = 1, 2, \dots; k = 1, \dots, K\}$ , set  $t_0^{(0)} = 0$  (all  $k$ ), and let  $\mathcal{F}$  denote the history  $\{\mathcal{F}_s: s \geq 0\}$  where the  $\sigma$ -field  $\mathcal{F}_s$  is determined by all those  $t_i^{(k)} < s$ . Let  $A_k(\cdot)$  denote the  $\mathcal{F}$ -compensator for the  $k$ th component process  $N_k(\cdot)$ ,  $M_k(t) = N_k(t) - A_k(t)$  its  $\mathcal{F}$ -martingale. Assume that  $t_i^{(k)} \neq t_{i'}^{(j)}$  a.s. for  $j \neq k$  and any  $i, i'$ , and that  $A_k(t)$  is a.s. continuous. For finite positive  $\theta_{ki}$  let  $H_{ki}(t)$  denote the stochastic process

$$H_{ki}(t) = \begin{cases} \exp(-\theta_{ki}[A_k(t) - A_k(t_{i-1}^{(k)})]) & t_{i-1}^{(k)} < t \leq t_i^{(k)}, \\ 0 & \text{otherwise,} \end{cases}$$

and define  $\tau_{ki} = A_k(t_i^{(k)}) - A_k(t_{i-1}^{(k)})$ , the duration of the  $i$ th interval of the  $k$ th component on transformed time-scale (note that each component has a different transformation). Then by the monotonicity of  $A_k(\cdot)$ ,  $0 < H_{ki}(t) \leq 1$ , and for  $i \neq i'$ ,  $H_{ki}(t)H_{ki'}(t) = 0$ .

- (a) Show that  $H_{ki}$ , being left-continuous and  $\mathcal{F}$ -adapted, is predictable, and for all  $k, i$  satisfies

$$\begin{aligned} \left| \int_{\mathbb{R}_+} H_{ki}(t) M_k(dt) \right| &\leq \int_{\mathbb{R}_+} H_{ki}(t) [N_k(dt) + A_k(dt)] \\ &\leq \theta_{ki}^{-1} [1 - e^{-\theta_{ki}\tau_{ki}}] < \theta_{ki}^{-1} < \infty. \end{aligned}$$

- (b) Conclude that, because  $M_k$  is a martingale,  $E(\int_{\mathbb{R}_+} H_{ki}(t) M_k(dt)) = 0$ , with  $E[\int_{\mathbb{R}_+} H_{ki}(t) N_k(dt)] = E(e^{-\theta_{ki}\tau_{ki}})$  and

$$E \int_{\mathbb{R}_+} H_{ki}(t) dA_k(t) = E \int_{A_k(t_{i-1}^{(k)})}^{A_k(t_i^{(k)})} e^{-\theta_{ki}[u - A_k(t_{i-1}^{(k)})]} du = \frac{E(1 - e^{-\theta_{ki}\tau_{ki}})}{\theta_{ki}},$$

so  $E(e^{-\theta_{ki}\tau_{ki}}) = 1/(1 + \theta_{ki})$ , and hence each  $\tau_{ki}$  has a unit exponential distribution.

- (c) Use the martingale property and integration over the diagonal  $D_3$  and the two open half-planes  $D_1$  and  $D_2$  much as in the argument following (14.1.16) to show that

$$\begin{aligned} E\left[\int_{\mathbb{R}_+} H_{ki}(u) M_k(du) \times \int_{\mathbb{R}_+} H_{ki'}(v) M_k(dv)\right] &= 0 \quad (i \neq i'), \\ E\left[\int_{\mathbb{R}_+} H_{ji}(u) M_j(du) \times \int_{\mathbb{R}_+} H_{ki'}(v) M_k(dv)\right] &= 0 \quad (j \neq k). \end{aligned}$$

- (d) Use a double induction on  $i$  and  $k$  to show that for all positive integers  $m \leq K$  and  $n$ ,

$$E\left[\prod_{i=1}^n \prod_{k=1}^m \int_{\mathbb{R}_+} H_{ki}(t) M_k(dt)\right] = 0,$$

which by parts (c) and (b) is equivalent to

$$E\left[\prod_{i=1}^n \prod_{k=1}^m (e^{-\theta_{ki}\tau_{ki}} - 1/(1 + \theta_{ki}))\right] = 0.$$

- (e) Part (d) implies that  $E[\prod_{i=1}^n \prod_{k=1}^m e^{-\theta_{ki}\tau_{ki}}] = \prod_{i=1}^n \prod_{k=1}^m 1/(1 + \theta_{ki})$  for all  $\theta_{ki} \geq 0$ , so the random variables  $\tau_{ki}$  are mutually independent with unit exponential distributions. The time-changed processes are therefore independent Poisson processes.

## 14.7. Poisson Embedding and Existence Theorems

This section centres on the theme, already developed in a simulation context in Section 7.5, that a wide range of processes can be derived by thinning a Poisson process in a suitably history-dependent manner. One reason for returning to this topic here is to outline the *Poisson embedding* technique developed in a cycle of papers by Brémaud, Massoulié, and others. This technique

provides a powerful framework for discussing the existence, uniqueness, and convergence to equilibrium, of stationary versions of a point process specified via its conditional intensity function. The discussion in this respect supplements the earlier discussion of convergence to equilibrium in Section 12.5.

We begin here by extending the thinning theorem of Proposition 7.5.I to MPPs, at the same time giving an explicit construction for the Poisson process from which a given point process can be derived by thinning. Like Watanabe's theorem, this proposition has a long history. The original results stem from Kerstan (1964b) and Grigelionis (1971); both preceded the independent development of thinning algorithms in Lewis and Shedler (1976). The extension to MPPs follows Brémaud and Massoulié (1994, 1996). We restrict attention to conditional intensities derived from internal histories, but Massoulié (1998) extends the results to situations where information from an external process can be incorporated into the history (see also Exercises 14.7.1–2).

**Proposition 14.7.I.** (a) *Let  $\tilde{N}$  be a Poisson process defined on the probability space  $(\Omega, \mathcal{E}, \mathcal{P})$ , having state space  $\mathbb{R} \times \mathcal{K} \times \mathbb{R}_+$ , with intensity measure*

$$\tilde{\mu}(dt \times d\kappa \times dx) = dt \times \ell_\kappa(d\kappa) \times dx \quad (14.7.1)$$

*equal to the reference measure, and let  $\tilde{\mathcal{H}}$  denote the internal history of  $\tilde{N}$ . Suppose that  $\lambda^*(t, \kappa)$  is a nonnegative  $\tilde{\mathcal{H}}$ -adapted and  $\tilde{\mathcal{H}}$ -mark-predictable process defined on the same probability space, and that the integrated process  $\lambda_g^*(t) = \int_{\mathcal{K}} \lambda^*(t, \kappa) \pi(d\kappa)$  is a.s. finite and locally integrable on  $\mathbb{R}$ . Define the point process  $N$  by*

$$N(dt \times d\kappa) = \tilde{N}(dt \times d\kappa \times (0, \lambda^*(t, \kappa)]). \quad (14.7.2)$$

*Then  $N$  is an  $\tilde{\mathcal{H}}$ -adapted MPP with  $\tilde{\mathcal{H}}$ -conditional intensity  $\lambda^*(t, \kappa)$  and ground intensity  $\lambda_g^*(t)$ .*

(b) *Let  $N(\cdot, \cdot)$  be an MPP on  $\mathbb{R} \times \mathcal{K}$ , with internal history  $\mathcal{H}$  and  $\mathcal{H}$ -predictable intensity  $\lambda^*(t, \kappa)$ . Then, extending the probability space if necessary, there exists a Poisson process  $\tilde{N}$  on  $\mathbb{R} \times \mathcal{K} \times \mathbb{R}_+$ , with intensity measure (14.7.1), such that  $\tilde{N}$  is  $\tilde{\mathcal{H}}$ -adapted and (14.7.2) holds.*

PROOF. (a) That  $N$  is  $\tilde{\mathcal{H}}$ -adapted follows from its definition in terms of  $\tilde{N}$ . Given any  $\tilde{\mathcal{H}}$ -mark-predictable, nonnegative process  $Y(t, \kappa)$ , and given also the predictability assumptions in the proposition, we then have

$$\begin{aligned} \mathbb{E} \left[ \int_{\mathbb{R} \times \mathcal{K}} Y(t, \kappa) N(dt \times d\kappa) \right] &= \mathbb{E} \left[ \int_{\mathbb{R} \times \mathcal{K} \times (0, \lambda_g^*(t)]} Y(t, \kappa) \tilde{N}(dt \times d\kappa \times dx) \right] \\ &= \mathbb{E} \left[ \int_{\mathbb{R}} \int_{\mathcal{K}} \int_0^{\lambda^*(t, \kappa)} Y(t, \kappa) dt \pi(d\kappa) dx \right] \\ &= \mathbb{E} \left[ \int_{\mathbb{R}} \int_{\mathcal{K}} Y(t, \kappa) \lambda^*(t, \kappa) dt \pi(d\kappa) \right], \end{aligned}$$

which is enough to establish  $\lambda^*$  as the conditional intensity for  $N$ . The assertions concerning the ground process follow simply.

(b) The construction depends on first defining a Poisson process  $N^*(dt \times d\kappa \times du)$  on  $\mathbb{R} \times \mathcal{K} \times \mathbb{R}_+$  with intensity (14.7.1), and a sequence of i.i.d. uniform (0,1) r.v.s  $U_k$ , such that both  $N^*$  and the given process  $N$  as well as  $\{U_k\}$  are all mutually independent. This can always be done at the expense, if needed, of expanding the probability space on which the original point process is defined. With each point  $(t_i, \kappa_i)$  from the original process, associate the point  $(t_i, \kappa_i, U_i \lambda^*(t_i, \kappa_i))$  in the ‘random strip’  $0 \leq x \leq \lambda^*(t, \kappa)$  in the third component of the extended space  $\mathbb{R} \times \mathcal{K} \times \mathbb{R}_+$ . Build the required process  $\tilde{N}$  by amalgamating these points from within the random strip with points of  $N^*$  from outside it. The probability of finding a point of the resultant process in  $(t, t + dt) \times (\kappa, \kappa + d\kappa) \times (x, x + dx)$ , given the history  $\tilde{\mathcal{H}}_{t-}$ , is then  $\lambda^*(t, \kappa) dt \pi(d\kappa) \lambda^*(t, \kappa)^{-1} dx = dt \pi(d\kappa) dx$  for points inside the random strip, and the same for points outside it. The resultant process is therefore a Poisson process with the required intensity measure.  $\square$

Grigelionis (1971) describes representations of the type (14.7.2) as ‘stochastic integrals with respect to a Poisson process;’ the construction above shows that the general class of MPPs defined by conditional intensities can be represented as stochastic integrals in this way. In the French literature and elsewhere, a point process  $N$  satisfying (14.7.2) is said to satisfy a stochastic differential equation driven by the Poisson process  $\tilde{N}$ .

In Proposition 14.7.I we do not require the function  $\lambda^*(t)$  to depend only on the past events of the newly defined process  $N$ ; it could depend on aspects of the behaviour of  $\tilde{N}$  outside the strip; with minor changes it could be made to depend even on the past of ancillary processes. In applications, however, we generally want  $\lambda^*$  to depend only on the past of  $N$  itself, that is, to be not  $\tilde{H}$ -adapted but  $\mathcal{H}$ -adapted where  $\mathcal{H}$  is the internal history of  $N$ . This can be achieved by requiring  $\lambda^*(t)$  to be a functional  $\psi(\cdot, \cdot)$  from the space  $\mathcal{N}_{\mathbb{R}_+ \times \mathcal{K}}^{\#*}$  into  $\mathbb{R}_+$ , and then setting  $\lambda^*(t, \kappa) = \psi(N(t_-), \kappa)$ , where  $N(t_-)$  is the part of the realization of the process  $N$  itself preceding time  $t$ .

The change of point of view here is that the conditional expectation in the definition of  $\lambda^*$  is regarded not directly as a function of  $\omega$  but as a function of  $\omega$  through the realization corresponding to  $\omega$  in the conditioning space  $\mathcal{N}_{\mathbb{R}_+}^{\#*}$ .

Here the c.s.m.s. property of  $\mathcal{N}_{\mathbb{R}}^{\#*}$  is of critical importance in ensuring that the resulting function  $\psi$  is measurable.

Brémaud, Massoulié, and coworkers use these ideas as the starting point for an extended discussion of existence theorems and convergence to equilibrium properties for point processes with a specified conditional intensity function. Brémaud and Massoulié (1996) and Massoulié (1998) are two key references; more recent papers by Brémaud, Nappo, and Torrisi (2002) and Torrisi (2002) extend the discussion to rates of convergence to equilibrium. In the present section we illustrate their approach in a simplified form which seems to us

more intuitive for applications to examples such as the ETAS model which have recurred in this book.

We first suppose that for  $t \geq 0$  the process is specified by a conditional intensity of the form

$$\begin{aligned}\lambda^*(t, \kappa) dt d\kappa &= \psi(S_t N_-, \kappa) = E[N(dt \times d\kappa) \mid \mathcal{F}_{t-}] \\ &= \psi(\{(t - t_i, \kappa_i): t_i < t\}, \kappa) dt \times d\kappa,\end{aligned}\tag{14.7.3}$$

where  $\psi(N, \kappa)$  is a nonnegative function of  $\kappa$  and the restriction of the realization  $N$  of a simple point process to the half-line  $\mathbb{R}_-$ ; that is, it is a nonnegative functional on  $\mathcal{N}_{\mathbb{R}_- \times \mathcal{K}}^{\#*}$ .

In terms of the discussion around Lemma 12.5.III, the representation in (14.7.3) allows us to interpret the MPP in terms of a Markov process on the space  $\mathcal{Z} = \mathcal{N}_{\mathbb{R}_- \times \mathcal{K}}^{\#*}$  and is equivalent to requiring the resulting Markov process to have stationary transition probabilities (see Exercise 14.7.3). Thus the techniques developed in this section are an alternative to the direct analysis of such a Markov process, and are particularly effective when the conditional intensity retains something of the linear structure characteristic of Hawkes processes. Last (1996, 2004) discusses and illustrates the Markov process approach.

Note also that (14.7.3) is the form derived in (14.3.5) for the complete intensity when the process is stationary. However, unlike the situation for point processes defined by their internal conditional intensities on the half-line  $\mathbb{R}_+$ , specifying a conditional intensity on  $\mathbb{R}$  by (14.7.3) does not in general specify a unique point process; there may be a variety of transient solutions, and there may or may not exist a stationary solution.

Our first objective is to seek conditions for the existence of a unique MPP that (i) satisfies given initial conditions which we suppose are specified in terms of a distribution of the point process over  $\mathbb{R}_-$ , and (ii) evolves for  $t \geq 0$  according to the representation (14.7.3). We then look for the existence of stationary solutions and for conditions governing convergence to equilibrium.

We start by constructing a sequence of approximations to the final process, imitating the classical Picard proof for the existence of solutions to a differential equation, and making use of the Poisson embedding results to construct the successive approximations.

The first of the approximations,  $N^0$  say, is specified by the initial condition for  $t < 0$  and for  $t \geq 0$  has  $\lambda^0(t, \kappa) \equiv 0$ , so  $N^0(\mathbb{R}_+) = 0$  a.s. For  $r > 0$ , the subsequent approximations  $N^r$  have the same realization  $N_-$  on  $\mathbb{R}_-$  but for  $t \geq 0$  each  $N^r$  is defined on a pathwise basis via the recursions

$$\lambda^{r+1}(t, \kappa) = \psi(S_t N_-^r, \kappa),\tag{14.7.4a}$$

$$N^{r+1}(dt \times d\kappa) = \tilde{N}(dt \times d\kappa \times (0, \lambda^{r+1}(t, \kappa)]),\tag{14.7.4b}$$

where  $\tilde{N}$  is a realization of a Poisson process constructed as in Proposition 14.7.I(a).

Following Brémaud and Massoulié, we now impose a condition on  $\psi$  which plays a role in the subsequent discussion similar to that of a Lipschitz condition in the classical existence and stability proofs for differential equations; for convenience we therefore refer to it as a *Lipschitz condition*. It takes the general form

$$|\psi(N, \kappa) - \psi(N', \kappa)| \leq \int_{\mathbb{R}_- \times \mathcal{K}} h(-s, \eta, \kappa) (N \Delta N') (ds \times d\eta), \quad (14.7.5)$$

where the nonnegative mapping  $h: \mathbb{R}_+ \times \mathcal{K} \times \mathcal{K} \mapsto \mathbb{R}_+$  is contractive in a sense defined more precisely below, and the MPP  $N \Delta N'$  is defined from MPPs with conditional intensities  $\lambda_N$  and  $\lambda_{N'}$  as the MPP with conditional intensity

$$\lambda_{N \Delta N'}(t, \kappa) = \begin{cases} 0 & \text{for } t < 0, \\ |\lambda_N(t, \kappa) - \lambda_{N'}(t, \kappa)| & \text{for } t \geq 0. \end{cases}$$

Poisson embedding provides a natural approach to defining such a process. Assuming both  $N$  and  $N'$  can be and are defined on the same probability space,  $N \Delta N'$  is formed from the points in the Poisson process  $\tilde{N}$  of (14.7.2) whose third component lies between the two curves  $x = \lambda_N(t, \kappa)$  and  $x = \lambda_{N'}(t, \kappa)$ . We use this construction repeatedly below.

We now seek conditions, on the kernel  $h$  of (14.7.5) and the initial process  $N_-$ , that are sufficient to ensure the existence of a unique MPP satisfying the prescribed initial distribution and with conditional intensity described by the given functional  $\psi$  for  $t > 0$ . Two broad classes of conditions can be identified in the literature. One class, first studied in the seminal paper by Kerstan (1964b), imposes an overall bound on the functional  $\psi$  and as a consequence can relax the conditions on  $h$ ; Massoulié (1998, Theorem 1) gives a general version from this class and is outlined in simplified form in Exercise 14.7.5. The second class, which we now follow, imposes no overall bound on  $\psi$  but requires the kernel  $h$  to define a contraction operator on a suitable space of functions. The specific conditions we use are adapted from those in Theorem 2 of Massoulié (1998); in particular, we omit the dependence on an external process allowed in Massoulié's treatment. Further variants are summarized in Exercises 14.7.5–6 and references noted there.

The following are the conditions we impose on  $h$ .

**Conditions 14.7.II.** *The kernel  $h$  in the Lipschitz condition (14.7.5) satisfies*

(a) *for all pairs  $(\eta, \kappa) \in \mathcal{K} \times \mathcal{K}$ ,*

$$0 < \int_0^\infty h(t, \eta, \kappa) dt \equiv H(\eta, \kappa) < \infty; \quad \text{and} \quad (14.7.6a)$$

(b) *there exists a positive,  $\mathcal{B}_{\mathcal{K}}$ -measurable, and  $\ell_{\mathcal{K}}$ -integrable function  $r(\kappa)$  satisfying, for some positive  $\rho < 1$ , the subinvariance condition*

$$\int_{\mathcal{K}} H(\eta, \kappa) r(\eta) \ell_{\mathcal{K}}(d\eta) \leq \rho r(\kappa). \quad (14.7.6b)$$

The existence, for some  $\rho > 0$ , of a  $\rho$ -subinvariant function for a positive kernel such as  $H(\eta, \kappa)$  is a mild constraint only on the regularity of the kernel. When the mark space is finite, it is a direct consequence of the Perron–Frobenius theorem, when  $\rho$  can be taken to be the maximum eigenvalue of the matrix, and equality rather than inequality holds in (14.7.6b). A similar result holds whenever  $H$  defines a compact operator (Lerch’s theorem). More general conditions can be established using generating function arguments, and are outlined, for example, in Vere-Jones (1968) for the denumerable case and more generally in Liggett (1985). The requirement  $H(\eta, \kappa) > 0$  for all  $\eta, \kappa$ , which implies also  $r(\kappa) > 0$  (all  $\kappa$ ), is a strict irreducibility condition, imposed for convenience in order to avoid having to detail the possibilities when the kernel is reducible.

Note that the kernel  $h$  can be interpreted as the analogue of the matrix density function for a Markov renewal process [see Example 10.3(a)] for which the renewal distributions are defective (hence, the processes are transient).

The major constraint is the scaling requirement  $\rho < 1$ , which is analogous to the subcriticality requirement in the branching process interpretation of a Hawkes process. The main difference between our approach and that in Massoulié (1998) is in the controlling role we give to the subinvariant function  $r(\cdot)$ , including the assumption that it is  $\ell_{\mathcal{K}}$ -integrable. This last assumption is associated with the requirement for the process to have finite ground intensity and could be relaxed if this requirement were dropped. These assumptions increase the transparency of the arguments at the loss of some generality.

The existence of the subinvariant function  $r(\cdot)$  in (14.7.6b) implies that the kernel  $H(\eta, \kappa)$  defines a bounded linear operator  $H$  on the space  $\mathcal{K}_1$ , say, of measurable functions  $f(\kappa)$  satisfying  $\int_{\mathcal{K}} f(\kappa) r(\kappa) \ell_{\mathcal{K}}(d\kappa) < \infty$ , and that its transpose  $H^*$  defines a bounded linear operator on the space  $\mathcal{K}_{\infty}$ , say, of measurable functions  $g(\eta)$  satisfying  $\text{ess sup}[g(\eta)/r(\eta)] < \infty$ , through the respective actions

$$(Hf)(\eta) = \int_{\mathcal{K}} H(\eta, \kappa) f(\kappa) \ell_{\mathcal{K}}(d\kappa), \quad (14.7.7a)$$

$$(H^*g)(\kappa) = \int_{\mathcal{K}} H(\eta, \kappa) g(\eta) \ell_{\mathcal{K}}(d\eta). \quad (14.7.7b)$$

Moreover, under the given conditions, each of the operators  $H$  and  $H^*$  has its norm bounded by  $\rho$ , as indicated in Exercise 14.7.4.

Among other consequences, the contractive condition implies that the sum  $\sum_{n=0}^{\infty} H^n$  converges geometrically fast and defines a bounded limit operator,  $R$  say, so that for functions  $f \in \mathcal{K}_1$  and  $g \in \mathcal{K}_{\infty}$ ,

$$\begin{aligned} \sum_{n=0}^{\infty} \int_{\mathcal{K}} \int_{\mathcal{K}} g(\eta) H^n(\eta, \kappa) f(\kappa) \ell_{\mathcal{K}}(d\eta) \ell_{\mathcal{K}}(d\kappa) \\ = \int_{\mathcal{K}} \int_{\mathcal{K}} g(\eta) R(\eta, \kappa) f(\kappa) \ell_{\mathcal{K}}(d\eta) \ell_{\mathcal{K}}(d\kappa) < \infty. \end{aligned} \quad (14.7.8)$$

Our basic result can now be stated as follows.



**Proposition 14.7.III.** *Suppose that the functional  $\psi$  and kernel  $h$  satisfy Conditions 14.7.II for some  $\ell_{\mathcal{K}}$ -integrable function  $r(\kappa)$ , and that for some  $C < \infty$ ,*

$$\psi(\emptyset, \kappa) \leq Cr(\kappa). \quad (14.7.9)$$

*Suppose also that an initial MPP  $N_-$  is given on  $\mathbb{R}_-$  such that for some  $D < \infty$ ,*

$$\epsilon(t, \kappa) \equiv \mathbb{E} \left[ \int_{\mathbb{R}_- \times \mathcal{K}} h(t-s, \eta, \kappa) N_-(ds \times d\eta) \right] \leq Dr(\kappa). \quad (14.7.10)$$

*Then there exists a unique MPP  $N$  on  $\mathbb{R}$ , with  $N = N_-$  on  $\mathbb{R}_-$  and conditional intensity (14.7.3) on  $\mathbb{R}_+$ , and with finite, bounded mean ground rate on  $\mathbb{R}_+$ .*

PROOF. Without loss of generality we suppose here and throughout that

$$\int_{\mathcal{K}} r(\kappa) \ell_{\mathcal{K}}(d\kappa) = 1.$$

Then the subinvariant function  $r(\cdot)$  plays the role of the density of a dominating stationary mark distribution.

We return to the sequence of approximations (14.7.4). Writing  $\lambda_{\Delta}^n(t, \kappa) = |\lambda^{n+1}(t, \kappa) - \lambda^n(t, \kappa)|$ , we have  $\lambda_{\Delta}^n(t, \kappa) = 0$  for  $t < 0$  because all approximations coincide with the initial condition  $N_-$  on  $\mathbb{R}_-$ , and for  $t \geq 0$  (14.7.5) implies that for  $n = 1, 2, \dots$ ,

$$\begin{aligned} \mathbb{E}(\lambda_{\Delta}^n(t, \kappa)) &= \mathbb{E}(|\psi(S_t N^n, \kappa) - \psi(S_t N^{n-1}, \kappa)|) \\ &\leq \mathbb{E} \left[ \int_{(-\infty, t) \times \mathcal{K}} h(t-s, \eta, \kappa) N^n \Delta N^{n-1}(ds \times d\eta) \right] \\ &= \mathbb{E} \left[ \int_{-\infty}^t \int_{\mathcal{K}} h(t-s, \eta, \kappa) \lambda_{\Delta}^{n-1}(s, \eta) ds \ell_{\mathcal{K}}(d\eta) \right], \end{aligned} \quad (14.7.11)$$

the last equality following from the  $\tilde{\mathcal{H}}$ -martingale property applied to the point process  $N^n \Delta N^{n-1}$ , which has conditional intensity  $|\lambda^n - \lambda^{n-1}|$ .

Writing  $\phi^n(\kappa) = \sup_{t \in \mathbb{R}_+} \mathbb{E}[\lambda_{\Delta}^n(t, \kappa)]$ , for  $n = 0$  we have from (14.7.9) and (14.7.10a),

$$\phi^0(\kappa) = \psi(S_t \emptyset) + [\psi(S_t N_0) - \psi(S_t \emptyset)] \leq Cr(\kappa) + \sup_t \epsilon(t, \kappa) \leq (C + D)r(\kappa).$$

For  $n = 1, 2, \dots$ , (14.7.11) now implies

$$\phi^n(\kappa) \leq \int_{\mathcal{K}} \phi^{n-1}(\eta) H(\eta, \kappa) \ell_{\mathcal{K}}(d\eta)$$

so that on using the inequalities for  $H^*$  following from Conditions 14.7.II,

$$M_n \equiv \sup_{\kappa} [\phi^n(\kappa)/r(\kappa)] \leq \rho M_{n-1} \leq \rho^n M_0,$$

where  $M_0 \leq C + D < \infty$ . The series  $\sum_{n=0}^{\infty} \mathbb{E}[\lambda_{\Delta}^n(t, \kappa)]$  therefore converges uniformly and absolutely in  $(t, \kappa)$ . Considering the ground processes  $N_g^n$ , it then follows that for any finite  $T > 0$ ,

$$\sum_{n=0}^{\infty} \mathbb{E}(|N_g^n(0, T) - N_g^{n-1}(0, T)|) = \sum_{n=0}^{\infty} \mathbb{E} \left[ \int_0^T \lambda_{\Delta}^n(t, \kappa) dt d\kappa \right] \leq \frac{(C + D)T}{1 - \rho}. \quad (14.7.12)$$

Now the ground processes  $N_g^n$  can differ by at most positive integer values, so at most a finite number of the differences  $N_g^n(0, T) - N_g^{n-1}(0, T)$  can be nonzero, a.s. It follows that the ground processes  $N_g^n$ , and hence also the full processes  $N^n$ , must be a.s. all equal after a finite number of terms, and so must converge almost surely and in expectation to some limit process  $N$ .

Such convergence implies in turn that the fidi distributions of the processes  $N^n(\cdot)$  converge for all finite intervals, implying finally that the processes  $N^n$  converge weakly to the limit process  $N$  (Theorem 11.1.VI).

Moreover, from the convergence of the series  $\sum_{n=0}^{\infty} \phi^n(\kappa)$ , we have also that

$$\sup_{t>0} \mathbb{E}[\lambda(t, \kappa)] \leq (C + D)r(\kappa)/(1 - \rho), \quad (14.7.13)$$

implying that for  $t > 0$  the limit process has finite mean ground rate and mark distribution dominated by a multiple of  $r(\kappa)$ .

The limit point process  $N$  starts from the same distribution of initial conditions on  $\mathcal{N}_{\mathbb{R}_-}^{\#*}$  as the approximants  $N^n$ , and on  $\mathbb{R}_+$ ,  $N$  has conditional intensity  $\lambda$  which, using Fatou's lemma and condition (14.7.6b), satisfies

$$\begin{aligned} \mathbb{E}(|\lambda(t, \kappa) - \psi(S_t N, \kappa)|) &\leq \lim_{n \rightarrow \infty} \mathbb{E}(|\lambda^n(t, \kappa) - \psi(S_t N, \kappa)|) \\ &= \lim_{n \rightarrow \infty} \mathbb{E}(|\psi(S_t N^{n-1}) - \psi(S_t N)|) \\ &\leq \lim_{n \rightarrow \infty} \mathbb{E} \left[ \int_{-\infty}^t \int_{\mathcal{K}} h(t - s, \eta, \kappa) (N^n \Delta N^{n-1})(dt \times d\eta) \right]. \end{aligned}$$

But  $h(\cdot, \cdot)$  is  $(\ell \times \ell_{\kappa})$ -integrable and  $N^n \Delta N^{n-1}$  converges weakly to the zero process, so the last limit is zero, showing that  $\lambda(t, \kappa) = \psi(S_t N, \kappa)$  a.s.

Finally, uniqueness follows from Propositions 14.1.VI and 14.2.IV, because we may regard the realization on  $\mathbb{R}_-$  as defining a prior  $\sigma$ -algebra  $\mathcal{F}_0$ , and the compensator defines the process uniquely when the history is an intrinsic history  $\mathcal{H}_t \vee \mathcal{F}_0$ .  $\square$

We turn now to stationary processes and convergence to equilibrium, recalling from Section 12.5 the terms weak and strong asymptotic stationarity to describe MPPs converging weakly or strongly (i.e. in variation norm) from given initial conditions to a stationary process. The next result is again a variant of the corresponding results in Brémaud and Massoulié (1996) and Massoulié (1998), to which we refer for further discussion and extensions.

**Theorem 14.7.IV.** *Suppose that the functional  $\psi$  and kernel  $h$  of (14.7.5) satisfy Conditions 14.7.II for some  $\ell_{\mathcal{K}}$ -integrable function  $r(\kappa)$ , and that  $0 < \psi(\emptyset, \kappa) < Cr(\kappa)$  for some  $C > 0$ . Then there exists a unique stationary MPP  $N^\dagger$  whose complete intensity is given by (14.7.3), which has finite mean ground rate*

$$m_g^\dagger = \mathbb{E} \left[ \int_{\mathcal{K}} \psi(S_0 N_-, \kappa) \ell_{\mathcal{K}}(d\kappa) \right], \quad (14.7.14)$$

*and whose stationary mark distribution is dominated by  $r(\kappa)$ . Furthermore, given initial conditions satisfying (14.7.10), the unique MPP  $N$  specified by Proposition 14.7.III is weakly asymptotically stationary with limit process  $N^\dagger$ . If in addition the function  $h$  in (14.7.5) satisfies, for some  $D' < \infty$ , the condition*

$$\int_0^\infty \int_{\mathcal{K}} t h(t, \eta, \kappa) r(\eta) dt \ell_{\mathcal{K}}(d\eta) < D' r(\kappa), \quad (14.7.15)$$

*then  $N$  is strongly asymptotically stationary with the limit  $N^\dagger$ .*

PROOF. Again we start the existence proof by constructing a sequence of approximating processes, but in the present situation the processes are defined over the whole line  $\mathbb{R}$ , starting from  $N^0$  which is taken to be the empty process over the whole line. It means that if  $\psi(\emptyset, \kappa) \equiv 0$ , then all subsequent approximations are also empty, and the construction fails to lead to any non-trivial stationary process. Otherwise,  $N^1$  is a stationary compound Poisson process with intensity  $\lambda^1(t, \kappa) = \lambda_g f(\kappa)$  for some finite, nonzero constant  $\lambda_g$  and stationary probability density  $f$  dominated by some multiple of  $r$ .

The arguments leading to (14.7.12) now carry over with only minor changes, the supremum in the definition of  $\phi^n(\kappa)$  now being taken over the whole real line. They lead as before to the existence of a well-defined limit MPP  $N^\dagger$ . In this case, however, it follows from the stationarity of  $N^0$  and the time invariant character of  $\psi$ , that  $N^1$ , and by induction all subsequent approximants, are also stationary, and hence that the limit process  $N^\dagger$  is stationary. Moreover, the boundedness conditions on  $\mathbb{E}[\lambda(t, \kappa)]$ , which follow as in (14.7.13), imply here that  $N^\dagger$  has finite, constant ground rate  $m_g$ , and stationary mark distribution dominated by some multiple of  $r(\kappa)$ .

Uniqueness of the stationary solution will follow if we can establish the asymptotic stationarity results, for any stationary solution  $N^*$  of (14.7.3), satisfying the conditions of the theorem, also satisfies the conditions of Proposition 14.7.III, (14.7.10) following here from the assumption that the stationary mark distribution is bounded by a multiple of  $r(\cdot)$ . If we assume the weak asymptotic stationarity results, therefore, this second solution should be weakly asymptotically stationary with limit  $N^\dagger$ , which is possible only if  $N^*$  and  $N^\dagger$  coincide.

It remains to prove the assertions concerning asymptotic stationarity. We suppose  $N$  and  $N^\dagger$  are defined as in the theorem. From the Lipschitz condition we obtain for the difference between the corresponding intensities,  $\lambda_\Delta(t, \kappa) \equiv$

$|\lambda(t, \kappa) - \lambda^\dagger(t, \kappa)|$  say,

$$\begin{aligned}
 E(\lambda_\Delta(t, \kappa)) &= E(|\psi(S_t N, \kappa) - \psi(S_t N^\dagger, \kappa)|) \\
 &\leq E \left[ \int_{(-\infty, t) \times \mathcal{K}} h(t-s, \eta, \kappa) (N \Delta N^\dagger)(ds \times d\eta) \right] \\
 &= E \left[ \int_{-\infty}^t \int_{\mathcal{K}} h(t-s, \eta, \kappa) \lambda_\Delta(s, \eta) ds \ell_{\mathcal{K}}(d\eta) \right] \\
 &= a(t, \kappa) + E \left[ \int_0^t \int_{\mathcal{K}} h(t-s, \eta, \kappa) \lambda_\Delta(s, \eta) ds \ell_{\mathcal{K}}(d\eta) \right], \quad (14.7.16)
 \end{aligned}$$

where

$$a(t, \kappa) = E \left[ \int_{-\infty}^0 \int_{\mathcal{K}} h(t-s, \eta, \kappa) \lambda_\Delta(s, \eta) ds \ell_{\mathcal{K}}(d\eta) \right]. \quad (14.7.17)$$

If we set  $g(t, \kappa) = 0$  for  $t < 0$ ,  $g(t, \kappa) = E[\lambda_\Delta(t, \kappa)]$  for  $t \geq 0$ , (14.7.16) has the form of a Markov renewal equation for the joint density with respect to  $ds \ell_{\mathcal{K}}(d\kappa)$ , namely,

$$g(t, \kappa) = a(t, \kappa) + \int_0^t \int_{\mathcal{K}} h(t-s, \eta, \kappa) g(s, \eta) ds \ell_{\mathcal{K}}(d\eta); \quad (14.7.18)$$

formally this has the solution

$$g(t, \kappa) = \int_0^t \int_{\mathcal{K}} a(t-s, \eta) R^*(s, \eta, \kappa) ds \ell_{\mathcal{K}}(d\kappa), \quad (14.7.19)$$

in which  $R^*(t, \eta, \kappa) = \sum_{n=0}^{\infty} h^{n*}(t, \eta, \kappa)$  is the sum of the iterates of  $h$  under the Markov convolution operation

$$(h * g)(t, \eta, \kappa) = \int_0^t \int_{\mathcal{K}} h(t-s, \eta, \nu) g(s, \nu, \kappa) ds \ell_{\mathcal{K}}(d\nu).$$

Note that, in the notation of (14.7.8),  $\int_0^\infty R^*(t, \eta, \kappa) dt = R(\eta, \kappa) < \infty$  for all  $t \geq 0$ , and moreover that as a sum of iterates of the contraction operator  $H$ ,  $R$  also satisfies

$$\int_{\mathcal{K}} R(\eta, \kappa) r(\eta) \ell_{\mathcal{K}}(d\eta) = \int_{\mathbb{R}_+ \times \mathcal{K}} R^*(t, \eta, \kappa) r(\eta) dt \ell_{\mathcal{K}}(d\kappa) \leq (1 - \rho)^{-1} r(\kappa). \quad (14.7.20)$$

Because both intensities  $\lambda(t, \kappa)$  and  $\lambda^\dagger(t, \kappa)$  are dominated uniformly in  $t$  by multiples of  $r(\kappa)$ , it follows easily from (14.7.17) that  $a(t, \kappa)$  is dominated by a multiple of  $r(\cdot)$ . Hence the integral in (14.7.19) converges, and from (14.7.18) and (14.7.20) we see that  $g(t, \kappa)$  is also dominated by a multiple of  $r(\cdot)$ .

We now consider the ground process of  $N_\Delta$ ,  $N_{\Delta,g}$  say, over the finite interval  $(t, t + T)$ , with the aim of proving that it converges weakly to the empty measure. We have

$$\Pr\{N_{\Delta,g}(t, t + T) > 0\} \leq E[N_{\Delta,g}(t, t + T)] = \int_t^{t+T} \int_{\mathcal{K}} g(u, \kappa) du \ell_{\mathcal{K}}(d\kappa).$$

Substituting for  $g(\cdot, \cdot)$  from (14.7.19) and using Fubini's theorem, leads to the estimates

$$\begin{aligned} & \int_t^{t+T} \int_{\mathcal{K}} g(u, \kappa) du \ell_{\mathcal{K}}(d\kappa) \\ &= \int_t^{t+T} \int_{\mathcal{K}} \left[ \int_0^u \int_{\mathcal{K}} a(u-s, \eta) R^*(s, \eta, \kappa) ds \ell_{\mathcal{K}}(d\eta) \right] du \ell_{\mathcal{K}}(d\kappa) \\ &\leq T \int_0^{t+T} \int_{\mathcal{K} \times \mathcal{K}} a(u-s, \eta) R^*(s, \eta, \kappa) ds \ell_{\mathcal{K}}(d\eta) \ell_{\mathcal{K}}(d\kappa) \\ &\leq T \int_0^{t+T} \int_{\mathcal{K} \times \mathcal{K}} \frac{a(u-s, \eta)}{r(\eta)} R^*(s, \eta, \kappa) r(\eta) ds \ell_{\mathcal{K}}(d\eta) \ell_{\mathcal{K}}(d\kappa). \quad (14.7.21) \end{aligned}$$

Now  $a(t, \kappa)/r(\kappa)$  is bounded, and as  $t \rightarrow \infty$ , using (14.7.17) and the inequality  $f(t, \kappa) \leq C'r(\kappa)$ , we see that

$$\begin{aligned} a(t, \kappa) &= \int_{-\infty}^0 \int_{\mathcal{K}} h(t-s, \eta, \kappa) f(s, \eta) ds \ell_{\mathcal{K}}(d\eta) \\ &\leq C' \int_t^\infty \int_{\mathcal{K}} h(u, \eta, \kappa) r(\eta) du \ell_{\mathcal{K}}(d\eta) \rightarrow 0 \end{aligned}$$

from the integrability of  $h$ . Also, as a function of  $(s, \eta)$  after integrating out  $\kappa$ , the second integral in (14.7.21) converges from (14.7.20), and so (14.7.21) as a whole represents a moving average of the function  $a(t, \kappa)/r(\kappa)$  which is bounded and converges to zero as  $t \rightarrow \infty$ . Thus  $\Pr\{N_{\Delta,g}(t, t + T) > 0\} \rightarrow 0$ , from which we deduce that the fidi distributions of  $S_t N_{\Delta,g}$ , and hence those of  $S_t(N \Delta N^\dagger)$  itself, converge to the fidi distributions of the empty process, and hence that the fidi distributions of  $S_t N$  converge weakly to those  $S_t N^\dagger$ . But this implies (by Theorem 11.1.VII again) the weak convergence of  $S_t N$  to  $S_t N^\dagger$ , and hence the weak asymptotic stationarity of  $N$ .

This argument fails if  $T = \infty$ , but under the additional condition (14.7.15), the expected value of the total number of points in  $N_\Delta$  is finite. Indeed, from the convolution representation (14.7.18), we have

$$\begin{aligned} E[N_{\Delta,g}(0, \infty)] &= \int_0^\infty \int_{\mathcal{K}} f(t, \kappa) dt \ell_{\mathcal{K}}(d\kappa) \\ &= \int_0^\infty \int_{\mathcal{K}} \left[ \int_0^\infty \int_{\mathcal{K}} a(u, \eta) R^*(s, \eta, \kappa) ds \ell_{\mathcal{K}}(d\eta) \right] du \ell_{\mathcal{K}}(d\kappa) \\ &= \int_0^\infty \int_{\mathcal{K} \times \mathcal{K}} a(u, \eta) R(\eta, \kappa) \ell_{\mathcal{K}}(d\eta) \ell_{\mathcal{K}}(d\kappa) du. \end{aligned}$$

Substituting for  $a(u, \eta)$  from (14.7.17), using the inequality  $f(t, \kappa) \leq C'r(\kappa)$ , and changing the order of integration, the last expression becomes

$$\begin{aligned} & \int_0^\infty \int_{\mathcal{K} \times \mathcal{K}} \left[ \int_u^\infty \int_{\mathcal{K}} h(v, \nu, \eta) f(u - v, \nu) \, d\nu \, \ell_{\mathcal{K}}(d\nu) \right] R(\eta, \kappa) \, d\eta \, \ell_{\mathcal{K}}(d\eta) \, \ell_{\mathcal{K}}(d\kappa) \\ & \leq C' \int_0^\infty \int_u^\infty \int_{\mathcal{K} \times \mathcal{K} \times \mathcal{K}} h(v, \nu, \eta) r(\nu) R(\eta, \kappa) \, d\eta \, d\nu \, \ell_{\mathcal{K}}(d\nu) \, \ell_{\mathcal{K}}(d\eta) \, \ell_{\mathcal{K}}(d\kappa) \\ & = C' \int_0^\infty \int_{\mathcal{K} \times \mathcal{K} \times \mathcal{K}} v h(v, \nu, \eta) r(\nu) \, d\nu \, \ell_{\mathcal{K}}(d\nu) R(\eta, \kappa) \, \ell_{\mathcal{K}}(d\eta) \, \ell_{\mathcal{K}}(d\kappa). \end{aligned}$$

Incorporating the condition (14.7.15) we obtain for the last integral,  $J$  say, the telescoping sequence of reductions

$$J \leq D' \int_{\mathcal{K} \times \mathcal{K}} r(\eta) R(\eta, \kappa) \, \ell_{\mathcal{K}}(d\eta) \, \ell_{\mathcal{K}}(d\kappa) \leq \rho D' \int_{\mathcal{K}} r(\kappa) \, \ell_{\mathcal{K}}(d\kappa) = \rho D' < \infty.$$

But if the expected number of points of the difference process on  $\mathbb{R}_+$  is finite, there must be (with probability 1) a finite last occurrence time, say  $L$ , of points in  $N_{\Delta, g}$ . This  $L$  acts as a coupling time for the two processes  $N$  and  $N^\dagger$ , and strong asymptotic stationarity then follows from the basic coupling inequality of Lemma 11.1.I.  $\square$

The arguments are considerably simplified if the process is unmarked, and are outlined in Exercises 14.7.5–7.

**EXAMPLE 14.7(a) Nonlinear Hawkes and ETAS models** [see Example 7.3(b)]. Recall that the simple Hawkes model has conditional intensity of the form  $\lambda^*(t) = \lambda + \int_0^t \mu(t-s) N(ds)$ , with nonlinear version

$$\lambda^*(t) = \Phi \left( \lambda + \int_0^t \mu(t-s) N(ds) \right). \quad (14.7.22)$$

This falls within the ambit of Proposition 14.7.III provided  $\Phi(\cdot)$  satisfies a standard Lipschitz condition of the form

$$|\Phi(x) - \Phi(y)| \leq \alpha |x - y| \quad (x, y \geq 0).$$

In this case conditions (14.7.6a–b) reduce, respectively, to  $\mu(t) < \infty$  and  $\rho = \alpha \int_0^\infty \mu(u) \, du < 1$ . When  $\Phi(x) \equiv x$  this reduces to the usual stability requirement for the Hawkes process.

The standard initial condition is to suppose the process empty for  $t < 0$ , which certainly satisfies condition (14.7.10), whereas (14.7.9) reduces to  $\Phi(\lambda) < \infty$ . Thus a version of the nonlinear process, starting from the empty initial condition, exists provided that both  $\Phi(\lambda) < \infty$  and  $\alpha \int_0^\infty \mu(u) \, du < 1$ .

From Theorem 14.7.IV, the same conditions also imply the existence of a stationary version of the process, with complete conditional intensity  $\lambda^\dagger(\cdot)$

as in (14.7.21) but with the integral taken from  $-\infty$ . The same conditions imply weak asymptotic stationarity of the process started from the empty initial condition. The extra condition (14.7.15) required for strong asymptotic stationarity here reduces to

$$\int_0^\infty t\mu(t) dt < \infty.$$

Consider next a nonlinear version of the ETAS model as a simple example of a nonlinear marked Hawkes process. As in (7.3.10) we write the conditional intensity for the nonlinear version in the form

$$\lambda^*(t, \kappa) = \Phi\left(\lambda + \int_{(0,t) \times \mathcal{K}} \mu(t-s)A(\eta) N(ds \times d\eta)\right)f(\kappa).$$

where  $f(\kappa)$  is the density of the distribution of the ‘unpredictable marks’ and  $A(\eta)$  [alias  $\psi(\eta)$  in (7.3.10)] measures the increase in ‘productivity’ with the increase in  $\eta$ . Assuming a simple Lipschitz condition on  $\Phi$  as above, we can identify the kernel  $h$  of (14.7.5) with

$$h(t, \eta, \kappa) = \alpha \mu(t)A(\eta)f(\kappa).$$

Then condition (14.7.6a) reduces to  $\int_0^\infty \mu(t) dt < \infty$  and the crucial condition (14.7.6b), if we identify the subinvariant vector  $r$  with  $f$ , to the requirement that

$$\rho = \alpha \int_{\mathcal{K}} A(\eta)f(\eta) \ell_{\mathcal{K}}(d\eta) \int_0^\infty \mu(t) dt < 1.$$

Note that this requires the convergence of the integral  $\int_{\mathcal{K}} A(\eta)f(\eta) \ell_{\mathcal{K}}(d\eta)$ , and that in this case the subinvariant vector is strictly invariant. This condition, together with the condition  $\Phi(\lambda) < \infty$ , imply both the existence of a stationary version, and the existence and weak asymptotic stationarity of a version started from the empty initial condition. The extra condition for strong asymptotic stationarity is the same as in the unmarked case.

In the more explicit spatial version of Example 6.4(b), without the nonlinear generalization, we leave the reader to check that the conditions reduce to those quoted in the example.  $\square$

## Exercises and Complements to Section 14.7

**14.7.1 Poisson embedding results for general histories.** Extend Proposition 14.7.I(a) and (b) to the situation where the history  $\mathcal{F}_t$  can be represented in the form  $\mathcal{F}_t = \mathcal{H}_t \vee \mathcal{G}_t$ , where  $\mathcal{H}_t$  refers to the internal history of the point process, and  $\mathcal{G}_t$  to the history of a process evolving contemporaneously in parallel with the point process.

[Hint: The crucial point for part (a) is to define the underlying Poisson process on a history sufficiently rich for both the point process and the auxiliary process to be  $\mathcal{H}$ -adapted and  $\tilde{\mathcal{H}}$ -predictable. See Massoulié (1998).]

14.7.2 (*Continuation*). Specialize Proposition 14.7.I to the case that the conditional intensity of the MPP can be represented in the form (14.7.3).

[*Hint*: The main changes relate to the requirement of  $\tilde{\mathcal{H}}$ -predictability, and the appropriate introduction of the functional  $\psi$ .]

14.7.3 Denote by  $\mathcal{Z}_N(t)$  the process whose value at time  $t$  is the realization  $S_t N_-$  on  $\mathbb{R}_-$  of an MPP with conditional intensity function of the form (14.7.3). Verify the Markov property for  $\mathcal{Z}_N(t)$  and examine the form of its infinitesimal generator.

14.7.4 Verify that the operators  $H$  and  $H^*$  of (14.7.7) do indeed define bounded linear operators on the spaces of functions  $\mathcal{K}_1$  and  $\mathcal{K}_\infty$  defined in the text, and that their norms are bounded by  $\rho$ . [*Hint*: Use (14.7.6b) and a Fubini theorem argument applied to

$$\|Hf\| = \int_{\mathcal{K}} |(Hf)(\eta)| r(\eta) \ell_{\mathcal{K}}(d\eta) \leq \int_{\mathcal{K} \times \mathcal{K}} H(\eta, \kappa) |f(\kappa)| r(\eta) \ell_{\mathcal{K}}(d\eta) \ell_{\mathcal{K}}(d\kappa),$$

$$\|H^*g\| = \text{ess sup}_{\kappa \in \mathcal{K}_\infty} \left\{ \frac{1}{r(\kappa)} \left| \int H(\eta, \kappa) g(\eta) \ell_{\mathcal{K}}(d\eta) \right| \right\},$$

assuming for  $H$  that  $\int_{\mathcal{K}} |f(\kappa)| \ell_{\mathcal{K}}(d\kappa) \equiv \|f\| < \infty$ , and for  $H^*$ , that  $\|g\| \equiv \text{ess sup}[g(\kappa)] < \infty$ .

14.7.5 Check and prove directly the following restriction of Proposition 14.7.III to simple (unmarked) point processes. Let  $\Psi(N)$  be a mapping  $\Psi: \mathcal{N}_{\mathbb{R}_-}^\# \mapsto \mathbb{R}_+$ , satisfying  $\Psi(\emptyset) = C < \infty$  and the Lipschitz condition  $|\psi(N) - \psi(N')| \leq \int_{\mathbb{R}_-} h(-s) (N \Delta N')(ds)$ , where  $h: \mathbb{R}_+ \mapsto \mathbb{R}_+$  satisfies  $0 < \int_0^\infty h(t) dt < \rho < 1$ . Suppose also that the initial condition satisfies  $\epsilon(t) = E[\int_{\mathbb{R}_-} h(t-s) N_-(ds)] \leq D < \infty$ . Then there exists a unique point process  $N$  with finite mean ground rate, initial condition  $N_-$ , and conditional intensity  $\lambda^*(t) = \Psi(S_t N)$  for  $t \geq 0$ . State and prove a corresponding extension of Theorem 14.7.IV. [Brémaud and Massoulié (1996) give a version without the restriction to processes with finite mean rate.]

14.7.6 (*Continuation*). As a variant, prove that similar uniqueness theorems hold under the conditions that  $\Psi$  is bounded overall, but the requirement that  $\rho < 1$  is weakened to  $\rho < \infty$ . Investigate an analogous theorem for MPPs under the assumption that  $\Psi(N, \kappa) \leq Mr(\kappa)$  for some finite  $M$  and function  $r$  that acts as a  $\rho$ -subinvariant function for the kernel  $h(t, \eta, \kappa)$ .

[For other variants, see also Kerstan (1964b), Brémaud and Massoulié (1996), and, for the marked case, Massoulié (1998).]

14.7.7 (*Continuation*). Say that a point process or MPP has *bounded memory* if the functional  $\Psi$  depends on  $N$  only on either (a) its past in the finite interval  $(-a, 0)$ , or (b) a finite number of occurrence times  $t_{-k}$  with  $t_{-k} < 0$ . Investigate existence and stability theorems for point processes and MPPs with finite memory. [See Lindvall (1988) for a treatment of case (b) based on a regeneration point argument. See also Brémaud and Massoulié (1996).]



## 14.8. Point Process Entropy and a Shannon–MacMillan Theorem

In this section we return to the discussion in Section 7.6 of point process entropy and information gain. Our aim is to extend and consolidate the theoretical background to the results presented there concerning entropy rates, likelihoods, and information gain. In particular we complete the discussion of Proposition 7.6.II and establish an ergodic result for point process entropy analogous to the Shannon–MacMillan theorem for independent sequences.

In introducing any general concept of entropy it is important to bear in mind that entropy, like likelihood, is best regarded as defined relative to a reference measure which can be a probability measure, but need not be totally finite. The entropy of a discrete (i.e., purely atomic) distribution  $\{p_k: k = 0, 1, \dots\}$  can be defined directly as the expectation

$$H_a = E(-\log p_k) = -\sum p_k \log p_k, \quad (14.8.1a)$$

but the natural analogue for a continuous distribution, with density  $f(x)$  say on  $\mathcal{X} \subseteq \mathbb{R}$ , namely,

$$H_c = -\int_{\mathcal{X}} f(x) \log f(x) dx, \quad (14.8.1b)$$

is scale-dependent, and cannot be reached as the limit of approximating discrete distributions. For example, in approximating the continuous uniform distribution on  $(0, 1)$  by a sequence of discrete uniform distributions with mass  $1/n$  at each of the points  $(k - \frac{1}{2})/n$ , the entropy of this discrete approximation equals  $\log n$  and consequently diverges as  $n \rightarrow \infty$ . Similarly, for the uniform distribution on the unit hypercube in  $\mathbb{R}^d$ , the discrete approximation to the entropy equals  $d \log n = \log(n^d) = \log(\# \text{ points used in discrete approximation})$ , and again diverges as  $n \rightarrow \infty$ . Indeed, it was just the differences in the rates of divergence that was recognized by Rényi (1959) as characterizing the dimension of the set on which the limit measure was carried, thus suggesting the definition of the Rényi dimensions introduced around (13.6.1).

Intuitively, the infinite limits obtained from the discrete approximations can be regarded as stemming from the unreasonable requirement that observation of a real-valued random variable pins down its value precisely, that is, the observation specifies all the digits in its decimal representation, thus conveying infinite information.

One way of overcoming the apparent difficulties in linking discrete and continuous entropies is to consider each entropy relative to a reference measure on the relevant carrying space. Then (14.8.1a) is considered as an entropy relative to the discrete measure with unit masses at each integer, and the continuous version at (14.8.1b) is considered as an entropy relative to Lebesgue measure. This leads to the concept of the *relative* or *generalized entropy*, an approach that also permits the definition of the entropy of a distribution on a general probability space. Suppose that  $(\Omega, \mathcal{E}, \mu)$  is a measure space, and

$\mathcal{P} \ll \mu$  is a probability distribution on this space. Then the generalized entropy of  $\mathcal{P}$  with respect to the reference measure  $\mu$  is given by

$$H(\mathcal{P}; \mu) = - \int_{\Omega} \Lambda(\omega) \log \Lambda(\omega) \mu(d\omega) = - \int_{\Omega} \log \Lambda(\omega) \mathcal{P}(d\omega) = E_{\mathcal{P}}(-\log \Lambda), \quad (14.8.2)$$

where  $\Lambda(\omega) = (d\mathcal{P}/d\mu)$  is the Radon–Nikodym derivative of  $\mathcal{P}$  with respect to  $\mu$ . If  $\mathcal{P}$  is singular with respect to  $\mu$ , we set  $H(\mathcal{P}; \mu) = \infty$ .

If  $\mu = \mathcal{Q}$ , where  $\mathcal{Q}$  is again a probability measure, we can rewrite (14.8.2) in the form

$$-H(\mathcal{P}; \mathcal{Q}) = \int_{\Omega} \log \frac{d\mathcal{P}}{d\mathcal{Q}} \mathcal{P}(d\omega) \quad (14.8.3)$$

which, apart from the negative sign, identifies the generalized entropy with the expected value of the likelihood ratio  $(d\mathcal{P}/d\mathcal{Q})$  under the assumption that  $\mathcal{P}$  is the true distribution. This link to the likelihood ratio underlies the properties of the entropy scores introduced in Section 7.6. Indeed, the right-hand side of (14.8.3) is nothing other than the *Kullback–Leibler distance* between the two probability measures  $\mathcal{P}$  and  $\mathcal{Q}$ . Convexity of the function  $x \log x$ , when applied to the first form in (14.8.2), guarantees that this quantity is always nonnegative, and equals zero if and only if the two measures coincide. When both distributions are absolutely continuous with respect to a reference measure  $\mu$ , then in the terminology of Section 7.6, the right-hand side of (14.8.3) represents the (expected) *information gain*, that is, the expected value of the logarithm of the probability gain, resulting from scoring outcomes by  $-\log(d\mathcal{P}/d\mu)$  rather than  $-\log(d\mathcal{Q}/d\mu)$ , when the true distribution is really  $\mathcal{P}$ . In this formulation the reference measure  $\mu$  drops out of the comparison, or can be taken to be  $\mathcal{Q}$  itself as in (14.8.3).

Turning to point process entropies, we start from the entropy of a point process observed over a state space  $\mathcal{X}$  which we take to be a bounded region  $A \in \mathbb{R}^d$ . The distribution of the point process can then be regarded as a symmetric probability distribution on the countable union  $\mathcal{X}^{\cup}$ . Observation of a realization of the process conveys information of two kinds: the actual number of points observed, and the location of these points given their number. Assuming absolute continuity with respect to Lebesgue measure for the distribution of locations, and bearing in mind that the points are indistinguishable, we can write the probability density for the latter term in the form  $k! \pi_k^{\text{sym}}(x_1, \dots, x_k; A)$ , where  $\pi_k^{\text{sym}}$  denotes a symmetric probability density over  $A^{(k)}$ . This suggests defining the entropy of a realization  $\{x_1, \dots, x_N\}$  as

$$\begin{aligned} H &\equiv H(N; x_1, \dots, x_N) = H(N) + E[H(x_1, \dots, x_N \mid N)] \\ &= - \sum_{k=0}^{\infty} p_k \log p_k \\ &\quad - \sum_{k=1}^{\infty} p_k \int_{\mathcal{X}^{(k)}} \pi_k^{\text{sym}}(x_1, \dots, x_k) \log[k! \pi_k^{\text{sym}}(x_1, \dots, x_k)] dx_1 \dots dx_k. \end{aligned} \quad (14.8.4)$$

This follows the notation of Section 5.3, with the factor  $k!$  arising, just as in the discussion of likelihoods, from the fact that only unordered point sets can be distinguished, so that any given allocation of particles to points is repeated  $k!$  times. Rudemo (1964) and McFadden (1965b) introduced point process entropy effectively in this form. From an entropy viewpoint, the extra factor  $\log k!$  can also be regarded as the loss of information, for given  $k$  and a given set of locations, about which particle is located at which location. Under the assumption of indistinguishable points, there are  $k!$  permutations from which to choose, all of them equally likely, corresponding to a distribution with entropy  $\log k!$ . Notice that (14.8.4) can be written

$$\begin{aligned} H &= -E_{\mathcal{P}}(\log[p_N N! \pi_N^{\text{sym}}(x_1, \dots, x_N)]) \\ &= -E_{\mathcal{P}}(\log[j_N(x_1, \dots, x_N)]) = E_{\mathcal{P}}(-\log L), \end{aligned}$$

where  $L$  is the likelihood, identified as in Section 7.1 with the Janossy density.

As with the entropy of distributions with a continuous density considered earlier, the definition at (14.8.4) is scale-dependent: approximating each of the densities  $\pi_k$  by a discrete distribution on  $n$  points, results in a discrepancy which increases as  $(\sum p_k \log k) \log n = E(N) \log n$ . For this reason  $E(N)$  is sometimes regarded as the dimension of the distribution of a finite point process on  $\mathbb{R}$ .

The implicit reference measure here has  $d$ -dimensional Lebesgue measure on each constituent space  $\mathcal{X}^d \subseteq \mathbb{R}^d$  and unit mass at each nonnegative integer. The alternative is to proceed as in Section 14.4 and take the reference measure to be the probability distribution of some standard process, usually the Poisson process with unit rate, so that  $(d\mathcal{P}/d\mathcal{Q})$  reduces to the likelihood ratio relative to this standard process. When this is the Poisson process, and  $\mathcal{X}$  is a bounded set  $A \in \mathbb{R}^d$ , the net effect is merely to add an extra term  $\ell(A)$  to (14.8.4) (see Exercise 14.8.1).

The expected value of such a log likelihood ratio is just the *information gain* as introduced in Section 7.6, which therefore appears as the negative of the corresponding generalized entropy. It takes the same form whether the points are treated as distinguishable or indistinguishable, and is given by

$$G = \sum_{k=0}^{\infty} p_k \log \frac{p_k}{q_k} + \sum_{k=1}^{\infty} p_k \int_{A^{(k)}} \pi_k^{\text{sym}} \log \frac{\pi_k^{\text{sym}}}{q_k^{\text{sym}}} dx_1 \dots dx_k. \quad (14.8.5)$$

The two expressions (14.8.4) and (14.8.5) illustrate the difference in intent between an *absolute* and a *generalized* entropy, giving the entropy relative to a measure reflecting the structure of the space on which it is defined, or else the information gain, which compares the entropies of two probability measures defined within similar structural constraints.

For a simple or marked point process in time, entropy can be represented alternatively in terms of conditional intensities. Suppose that  $\mathcal{P}$  (which we use

as a shorthand for  $\mathcal{P}_T$ , the probability measure restricted to the events generated by the point process in  $[0, T]$  corresponds to a MPP on  $[0, T] \times \mathcal{K}$ , and that we represent the likelihoods in terms of conditional intensities  $\lambda^*(t, \kappa)$  relative to the internal history  $\mathcal{H}$  as in Definition 14.3.I(b). The generalized entropy and the information gain relative to an alternative probability measure  $\mathcal{P}^0$  can be found by taking expectations of the likelihood ratio as set out in (14.4.2). In particular, adopting the notation of (14.4.2), the expected information gain for MPPs over an interval  $(0, T)$  can be written in the form

$$G_T(\mathcal{P}; \mathcal{P}^0) = E_{\mathcal{P}}[d\mathcal{P}/d\mathcal{P}^0] \\ = E \left[ \int_{(0, T) \times \mathcal{K}} \log \mu(t, \kappa) N(dt \times d\kappa) - \int_0^T \int_{\mathcal{K}} [\mu(t, \kappa) - 1] \lambda^0(t, \kappa) dt \ell_{\mathcal{K}}(d\kappa) \right],$$

where  $\mu(\cdot)$  is the ratio of  $\mathcal{H}$ -conditional intensities  $\lambda^*(\cdot)$  and  $\lambda^0(\cdot)$  under  $\mathcal{P}$  and  $\mathcal{Q} \equiv \mathcal{P}^0$ , respectively. Taking predictable versions of the conditional intensity, the last expectation simplifies to give

$$G_T(\mathcal{P}; \mathcal{P}^0) = E \left[ \int_0^T \int_{\mathcal{K}} \lambda^*(t, \kappa) \log \mu(t, \kappa) dt \ell_{\mathcal{K}}(d\kappa) \right] - \int_0^T [m_g(t) - m_g^0(t)] dt, \quad (14.8.6)$$

where we have written  $m_g(\cdot)$  and  $m_g^0(\cdot)$  for the ground rates under  $\mathcal{P}$  and  $\mathcal{P}^0$  respectively. For simple point processes, a similar expression holds but without any marks  $\kappa$  or integral over  $\mathcal{K}$ .

To obtain the generalized entropy, it is necessary to identify the appropriate reference measure. When the point process is simple the reference measure  $\mathcal{Q}_T$  is the nonnormalized measure corresponding to  $e^T \text{Poi}(1, T)$ , where  $\text{Poi}(\lambda, T)$  is the probability measure of a Poisson process on  $(0, T)$  with constant rate  $\lambda$ . For an MPP the nonnormalized measure is a similar multiple of a compound Poisson process with unit rate and mark distribution  $\pi(d\kappa)$ . Adopting these conventions, we obtain the corresponding generalized entropy in the form

$$H_T = -E \left[ \int_0^T \int_{\mathcal{K}} \{ \lambda^*(t, \kappa) \log [\lambda^*(t, \kappa)] - [\lambda^*(t, \kappa) - 1] \} dt \pi(d\kappa) \right], \quad (14.8.7)$$

In general, the expressions (14.8.6–7) are not easy to evaluate explicitly, although they are usually straightforward to obtain from simulations. It does simplify, however, when the MPP is stationary. In this case we use the extension of the likelihood (14.4.2) to intrinsic conditional intensities (i.e., conditioned on some initial  $\sigma$ -algebra  $\mathcal{G}_0$ ) and we take  $\mathcal{G}_0 = \mathcal{H}_{-\infty}^0$ , so that for  $t > 0$ ,  $\lambda^{\mathcal{G}}(t)$  can be identified with the complete intensity  $\lambda^{\dagger}(t)$ . In this case, for all  $t$ , we have

$$E[\lambda^{\dagger}(t, \kappa)] = E[\lambda^{\dagger}(0, \kappa)] = m_g E[f^{\dagger}(\kappa \mid 0)],$$

where  $m_g = E[\lambda_g^{\dagger}(0)]$  is the overall mean rate, which we assume finite. Then, for example, the expected information gain (14.8.6) takes the form

$$G_T = T E \left[ \int_{\mathcal{K}} \lambda^{\dagger}(0, \kappa) \log \mu^{\dagger}(0, \kappa) \pi(d\kappa) \right] - [m_g - m_g^0] \equiv T G, \quad (14.8.8a)$$

where  $G$  is the *expected information gain per unit time* of<sup>2</sup> (7.6.14). It is often more usefully written as  $G$  being equal to

$$\mathbb{E}[\lambda_g^\dagger(0) \log \lambda_g^\dagger(0)] - [m_g - m_g^0] + \mathbb{E}\left[\log \lambda_g^\dagger(0) \int_{\mathcal{K}} f^\dagger(\kappa | 0) \log \frac{f^\dagger(\kappa | 0)}{\bar{f}(\kappa)} \pi(d\kappa)\right]. \quad (14.8.8b)$$

Then the first two terms represent the information gain due to the times of the points, and the third term the conditional information gain due to the marks of the points, given their occurrence times.

Similarly we can define an *entropy rate* by

$$\begin{aligned} H &\equiv \frac{H_T}{T} = -\mathbb{E}\left[\int_{\mathcal{K}} \lambda^\dagger(0, \kappa) \log \lambda^\dagger(0, \kappa) \pi(d\kappa)\right] + [m_g - 1] \\ &= -\mathbb{E}\left[\lambda_g^\dagger(0) \left(\log \lambda_g^\dagger(0) + \int_{\mathcal{K}} f^\dagger(\kappa | 0) \log f^\dagger(\kappa | 0) \pi(d\kappa)\right)\right] + [m_g - 1]. \end{aligned} \quad (14.8.9)$$

For reference, the last results are summarized in the proposition below.

**Proposition 14.8.I.** *Let  $N$  be a simple or marked stationary point process on  $\mathbb{R}$ , with complete intensity function (determined by the internal history and relative to a reference distribution  $\pi(\cdot)$  on the mark space)  $\lambda^\dagger(t, \kappa) = \lambda_g^\dagger(t) f^\dagger(\kappa | t)$ . Suppose that  $\mathbb{E}[\lambda_g^\dagger(0) \log \lambda_g^\dagger(0)] < \infty$ , so that  $N$  has finite ground rate  $m_g = \mathbb{E}[\lambda_g^\dagger(0)]$ . Then the expected information gain  $G$ , relative to a compound Poisson process with unit ground rate and strictly positive mark density  $\bar{f}(\kappa)$  is given by (14.8.8), and the entropy rate  $H$  by (14.8.9).*

**EXAMPLE 14.8(a).** *Simple, mixed, and compound Poisson processes.* The simplest example is a simple Poisson process of constant rate  $\mu$ . In this case  $G = \mu \log \mu - (\mu - 1)$ , representing the expected information gain per unit time over the unit rate Poisson process. Note the implicit ratio in the first term: each  $\mu$  is really the ratio  $\mu/1$  of rates between the true and reference processes. As a function of  $\mu$ ,  $G$  is 0 when  $\mu = 1$ , that is, when the true process coincides with the reference process, and is otherwise strictly positive. The mean entropy rate is obtained by changing signs and omitting the  $-1$  in the final bracket:  $\bar{H} = \mu - \mu \log \mu$ . Exercise 14.8.2 shows that amongst all point processes with given rate  $\lambda$ , the Poisson process has maximum entropy rate.

In the case of a mixed Poisson process, where the rate  $\lambda$  is a random variable with distribution function  $F(\cdot)$  say, the conditioning on  $\mathcal{G}_0$  must be taken into account, and the information gain becomes

$$G = \int_{\mathbb{R}_+} \lambda [\log \lambda - 1] dF(\lambda).$$

<sup>2</sup> Equation (7.6.14) in Volume I contains two errors: in the second term  $+m_g \mathbb{E}[\cdot \cdot \cdot]$  a factor  $f_{k|0}^\dagger$  is omitted and it is wrongly assumed that the process has unpredictable marks, so

that the term should instead read  $+\mathbb{E}\left[\sum_{k=1}^K \lambda_g^\dagger(0) f_{k|0}^\dagger \log \frac{f_{k|0}^\dagger}{f_k}\right]$ . In the ensuing display,  $G$  equals the right-hand side but the middle expression is also flawed and should be deleted.

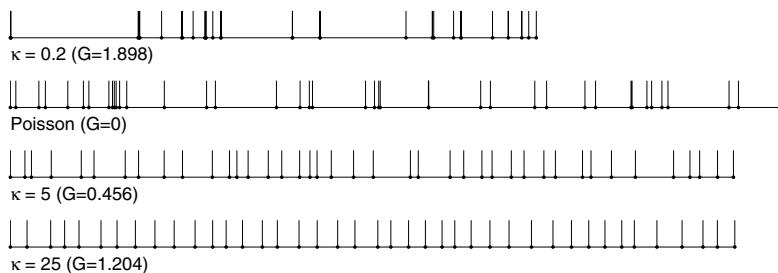


Figure 14.2

First 40 points of realizations of four different gamma-distributed renewal processes, same mean but shape parameters  $\kappa = 0.2, 1, 5, 25$ , and (expected) information gains  $G$  as shown.

Consider finally a compound Poisson process with ground rate  $\mu$  and mark distribution  $\{p_k: k \in \mathbb{Z}_+\}$ . The mark distribution is here discrete, and the reference distribution on the mark space can be any distribution  $\bar{f}_k$  on the non-negative integers for which all terms are strictly positive. Because the marks are chosen independently of the previous history of the process, (14.8.6) simplifies to

$$G = \mu \log \mu - (\mu - 1) + \mu \sum_{k=0}^{\infty} p_k \log(p_k / \bar{f}_k).$$

As does (14.8.6) more generally, this equation represents a decomposition of the expected information gain into two components: the gain due to modelling the counts, and the gain due to modelling the marks.  $\square$

EXAMPLE 14.8(b) *Renewal processes with gamma interevent distributions.* It is shown in Example 7.6(b) that the information gain for a renewal process with lifetime density  $f(\cdot)$ , relative to a Poisson process with the same mean rate  $m$ , takes the form

$$G = m \left[ 1 + \int_0^{\infty} f(y) \log \frac{f(y)}{m} dy \right]. \quad (14.8.10)$$

Figure 14.2, adapted from Daley and Vere-Jones (2004), illustrates the way  $G$  can vary with the model as the character of the interevent distribution is changed. Here the mean rate is fixed but the shape parameter is allowed to vary. Details of the computations are sketched in Exercise 14.8.3.  $\square$

We turn next to approximations of point process entropies by entropies of systems of discrete trials, with the aim of consolidating the ideas introduced in Section 7.6; we follow broadly the treatment in Daley and Vere-Jones (2004).

Suppose first that the process is unmarked, and that the observation interval  $(0, T]$  is partitioned into subintervals  $A_i = (u_{i-1}, u_i]$ ,  $i = 1, \dots, k_n$ , with  $u_0 = 0$ ,  $u_{k_n} = T$  for which forecasts are required. We observe either the values  $X_i = N(A_i)$  or the indicator variables  $Y_i = I_{N(A_i) > 0}$ , but not the

individual points. Our aim is to show how the information gain and associated quantities for the point process can be approximated by the analogous quantities for the  $X_i$  or the  $Y_i$ , both of which may be thought of as sequences of dependent trials.

In order to embed this situation into a limit process, we suppose that the given partition is a member of a *dissecting family* of partitions, say  $\mathcal{T}_n$ ,  $n = 1, 2, \dots$ , as defined in Section A1.6. The crucial partition, however, is not the partition of the state space but the partition this induces on the probability space  $\Omega$ . For example, a partition of  $(0, T]$  into  $r$  subintervals induces a partition of the probability space into  $2^r$  events, each corresponding to a particular sequence of values of the indicator variables  $Y_i$ . We denote by  $\mathcal{A}_n$  the algebra of events generated in this way by the partition  $\mathcal{T}_n$ .

To consider the effect of refining the partition, take a particular partition  $\mathcal{T}_{n_0}$ , say, and define for each set  $A \in \mathcal{T}_{n_0}$  the sequence of associated processes  $\eta^{(n)}(A)$ , where for  $n < n_0$ ,  $\eta^{(n)}(A) = 0$ , and for  $n \geq n_0$ ,

$$\eta^{(n)}(A) = \sum_{i: A_i^{(n)} \subseteq A} Y_i^{(n)} = \sum_{i: A_i^{(n)} \subseteq A} I_{N(A_i^{(n)}) > 0}. \quad (14.8.11)$$

For  $n > n_0$ ,  $\eta^{(n)}(A)$  counts the numbers of subintervals of  $A$  which belong to  $\mathcal{T}_n$  and contain a point of the process. It is clear that, for increasing  $n$ , the  $\eta^{(n)}$  are nondecreasing. Indeed, because the partitions form a dissecting system, and the point process is assumed to be simple, each point of the process will ultimately be the sole contributor to one of the nonzero terms  $\eta^{(n)}(A)$ , so that  $\eta^{(n)}(A) \uparrow N(A)$ . Thus, any event  $\{N(A) = k\}$  can be approximated by the corresponding events  $\{\eta^{(n)}(A) = k\}$ . More generally, any event defined by the simple point process in  $(0, T]$  can be approximated by events determined by the processes  $\eta^{(n)}(\cdot)$ , or equivalently  $\mathcal{H}_{(0, T]} = \bigvee_{n=1}^{\infty} \mathcal{A}_n$ .

A similar argument holds also for marked point processes. Here we consider a dissecting family  $\mathcal{T}_n$  of partitions of the product space  $(0, T] \times \mathcal{K}$ , each of which is of product form  $\mathcal{V}_n \times \mathcal{W}_n$ , so that each element of  $\mathcal{T}_n$  is a rectangle  $V \times W$ , where  $V \in \mathcal{V}_n$ ,  $W \in \mathcal{W}_n$  respectively. A family of processes  $\zeta^{(n)}(\cdot)$  can be defined much as in (14.8.11), but with the  $A_i^{(n)}$  in (14.8.11) interpreted as rectangles from  $(0, T] \times \mathcal{K}$ . In this case, if  $(0, T]$  is partitioned into  $r$  subintervals, and  $\mathcal{K}$  into  $s$  components, the  $\sigma$ -algebra  $\mathcal{A}_n$  will be generated by  $(s+1)^r$  distinct events, each corresponding to a sequence of length  $r$ , each term in which can be any one of the  $s$  marks, or a special mark  $\phi$  to allow for the possibility that no events occur in the subinterval in question.

Once again we find that, for any rectangle set  $A \times K$  from one of the partitions  $\mathcal{T}_n$ ,  $\zeta^{(n)}(A \times K) \uparrow N(A \times K)$ , and consequently  $\mathcal{H}_{(0, T]} = \bigvee_{n=1}^{\infty} \mathcal{A}_n$ . The following lemma summarizes the conclusions.

**Lemma 14.8.II.** *Let  $N$  be a marked point process with simple ground process and let  $\zeta^{(n)}$  be defined as in (14.8.11) with respect to a dissecting family of partitions of  $(0, T] \times \mathcal{K}$ . Then the processes  $\zeta^{(n)}(A \times K)$  generate the internal history  $\mathcal{H}_{(0, T]}$ .*

To proceed further, we need to clarify the role of the underlying  $\sigma$ -algebra in defining the generalized entropy. Here we follow the treatment of Csiszár (1969) and Fritz (1969). Suppose given a measure space  $(\Omega, \mathcal{E})$ , a  $\sigma$ -algebra  $\mathcal{A} \subseteq \mathcal{E}$ , and two measures on  $(\Omega, \mathcal{E})$ , a probability measure  $\mathcal{P}$ , and a reference measure (not necessarily a probability measure)  $\mathcal{Q}$ ; suppose also that  $\mathcal{P} \ll \mathcal{Q}$  on  $\mathcal{A}$ , with Radon–Nikodym derivative  $(d\mathcal{P}/d\mathcal{Q})_{\mathcal{A}}$ . As before, but explicitly recording the  $\sigma$ -algebra, we define the generalized entropy on  $(\Omega, \mathcal{A})$  as

$$\begin{aligned} H(\mathcal{P}, \mathcal{Q}; \mathcal{A}) &= - \int_{\Omega} \left( \frac{d\mathcal{P}}{d\mathcal{Q}} \right)_{\mathcal{A}} \log \left( \frac{d\mathcal{P}}{d\mathcal{Q}} \right)_{\mathcal{A}} \mathcal{Q}(d\omega) \\ &= - \int_{\Omega} \log \left( \frac{d\mathcal{P}}{d\mathcal{Q}} \right)_{\mathcal{A}} \mathcal{P}(d\omega) = \mathbb{E}_{\mathcal{P}} \left[ - \log \left( \frac{d\mathcal{P}}{d\mathcal{P}^0} \right)_{\mathcal{A}} \right]. \end{aligned} \quad (14.8.12)$$

When  $\mathcal{Q}(\cdot)$  is itself a probability measure, we prefer to change the sign and refer instead to the information gain

$$G(\mathcal{P}, \mathcal{P}^0; \mathcal{A}) = \int_{\Omega} \left( \frac{d\mathcal{P}}{d\mathcal{P}^0} \right)_{\mathcal{A}} \log \left( \frac{d\mathcal{P}}{d\mathcal{P}^0} \right)_{\mathcal{A}} \mathcal{P}^0(d\omega) = \int_{\Omega} \log \left( \frac{d\mathcal{P}}{d\mathcal{P}^0} \right)_{\mathcal{A}} \mathcal{P}(d\omega), \quad (14.8.13)$$

with  $G(\mathcal{P}, \mathcal{P}^0; \mathcal{A}) \geq 0$ , and it is set equal to  $+\infty$  if in fact  $\mathcal{P} \not\ll \mathcal{P}^0$ ; note that the integral can diverge even when the absolute continuity condition holds.

The following two properties of the generalized entropy, taken from Csiszár (1969) as in Fritz (1969) but restated here in terms of information gains, are of crucial importance (for a sketch of the proof and some related material, see Exercises 14.8.4–6).

**Lemma 14.8.III.**

- (a) *The information gain  $G(\mathcal{P}, \mathcal{P}^0; \mathcal{A})$  is nondecreasing under refinement of the  $\sigma$ -algebra  $\mathcal{A}$ .*
- (b) *Let  $\{\mathcal{A}_{\alpha}\}$  be a family of  $\sigma$ -algebras generating  $\mathcal{A}$ , and such that, to every  $\mathcal{A}_{\alpha_1}, \mathcal{A}_{\alpha_2}$ , there exists  $\mathcal{A}_{\alpha_3} \in \{\mathcal{A}_{\alpha}\}$  for which  $\mathcal{A}_{\alpha_1} \cup \mathcal{A}_{\alpha_2} \subseteq \mathcal{A}_{\alpha_3}$ . Then*

$$G(\mathcal{P}, \mathcal{P}^0; \mathcal{A}) = \sup_{\alpha} \{G(\mathcal{P}, \mathcal{P}^0; \mathcal{A}_{\alpha})\}.$$

Suppose, in particular, that  $\mathcal{A}^{(n)}$  is the  $\sigma$ -algebra derived from the  $\zeta^{(n)}(A)$  associated with a dissecting family of partitions as in Lemma 14.8.II. Then the condition of part (b) of Lemma 14.8.III holds as a consequence of the nested property of the partitions in a dissecting family, and because the family of  $\sigma$ -algebras  $\mathcal{A}_n$  is also monotonic increasing with limit  $\mathcal{H}_{(0,T]}$ , the two lemmas together imply the first part of the following proposition, which provides an extension and minor strengthening of Proposition 7.6.II.

**Proposition 14.8.IV.** *Consider an MPP  $N$  defined on  $(0, T] \times \mathcal{K}$ , with distribution  $\mathcal{P}$  on  $(\Omega, \mathcal{E})$ , and internal history  $\mathcal{H}$ , and let  $\mathcal{P}^0$  be an alternative probability measure on  $(\Omega, \mathcal{E})$ . Also let  $\mathcal{T}_n = \mathcal{V}_n \times \mathcal{W}_n$  be a dissecting family*



of partitions of  $(0, T] \times \mathcal{K}$ , and  $\mathcal{A}_n$  the  $\sigma$ -algebra of events induced by  $\mathcal{T}_n$  through the approximations  $\zeta^{(n)}$  of Lemma 14.8.II.

- (a) If  $G = G(\mathcal{P}_T, \mathcal{Q}_T; \mathcal{H})$  and  $G_n = G(\mathcal{P}_T, \mathcal{Q}_T; \mathcal{A}_n)$  denote the expected information gains associated with  $N$  and  $\zeta^{(n)}$ , respectively; then  $G_n \leq G$  and  $G_n \uparrow G$  as  $n \rightarrow \infty$ .
- (b) The same conclusions hold for the information gains relative to the intrinsic histories  $\mathcal{G} = \mathcal{H} \vee \mathcal{G}_0$  and  $\{\mathcal{G}_n = \mathcal{A}_n \vee \mathcal{G}_0\}$ .

PROOF. Part (a) is a direct consequence of Lemmas 14.8.II–III. Part (b) follows by first conditioning on  $\mathcal{G}_0$  and applying part (a), then taking expectations over  $\mathcal{G}_0$ .  $\square$

An important extension to these results is to situations where the point process evolves alongside another, stochastically related, process which can provide predictive information about the point process. It is in situations of this kind that more general conditional intensities  $\lambda^{\mathcal{F}}$  arise, supposing that the joint history is available for observation, but only the point process needs predicting. For such situations it is also possible to consider information gains of the type (14.8.6–10), even though they lose their strict interpretation as expected values of likelihood ratios. We attempt only an informal sketch of this development.

We consider just the case where a simple point process  $N(t)$  evolves alongside a second cumulative process  $W(t)$ , both processes being observed over  $(0, T]$ , or in terms of their increments over a family of partitions of  $(0, T)$ . Specifically, let  $Y_i^{(n)}, W_i^{(n)}$  denote the observed values of the increments of the processes  $N(t)$ ,  $W(t)$  over the set  $V_i^{(n)} \in \mathcal{V}_n$  defined as in the previous discussion, let  $p_i^{(n)}(Y \mid \mathcal{F}_{i-1}^{(n)})$  denote the conditional distribution of  $Y_i^{(n)}$  given observations  $(Y_j^{(n)}, W_j^{(n)})$  for  $0 \leq j \leq i-1$ , and  $p_i^{(n)}(Y \mid \mathcal{H}_{i-1}^{(n)})$  the distribution of  $Y_i^{(n)}$  given  $Y_j^{(n)}$  ( $0 \leq j \leq i-1$ ) only. This gives an information gain

$$\tilde{G}_n = E \left[ \sum_i p_i^{(n)}(Y_i^{(n)} \mid \mathcal{F}_{i-1}^{(n)}) \log \frac{p_i^{(n)}(Y_i^{(n)} \mid \mathcal{F}_{i-1}^{(n)})}{p_i^{(n)}(Y_i^{(n)} \mid \mathcal{H}_{i-1}^{(n)})} \right], \quad (14.8.14)$$

and the inequality  $\tilde{G}_n \geq 0$  holds essentially as a result of the well-known inequality for entropies,  $H(X \mid Y) \leq H(X)$ .

Lemma 14.8.III can now be applied to show that, in this situation also, refinement of the partitions can only increase the information gain. Introducing a further partition point  $\theta$  inside a subinterval  $(u_{i-1}, u_i]$  does not affect the information available to the two histories at the beginning of the first new subinterval,  $(u_{i-1}, \theta]$ , but in the second new subinterval,  $(\theta, u_i]$ , the full process obtains new information about both the point process and the explanatory variables, whereas the reference process obtains only the additional information about the point process. Hence the information gain can only increase with the introduction of the new partition.

If now we let  $n \rightarrow \infty$  in (14.8.14), and suppose that conditional intensities  $\lambda^{\mathcal{F}}(t, \kappa)$ ,  $\lambda^{\mathcal{H}}(t, \kappa)$  exist, we obtain the limit, which again is also an upper bound,

$$G(\mathcal{P}, \mathcal{P}^0; \mathcal{A}_{\mathcal{F}}) = \mathbb{E} \left[ \int_0^T \int_{\mathcal{K}} \log \left( \frac{\lambda^{\mathcal{F}}(t, \kappa)}{\lambda^{\mathcal{H}}(t, \kappa)} \right) \lambda^{\mathcal{F}}(t, \kappa) dt \ell_{\mathcal{K}}(d\kappa) \right],$$

the second term in (14.8.6) disappearing because the processes have the same expected rates. In particular, if the two processes are stationary, the information gain per unit time  $G$  equals

$$\mathbb{E} \left[ \lambda_g^{\mathcal{F}}(0) \log \frac{\lambda_g^{\mathcal{F}}(0)}{\lambda_g^{\mathcal{H}}(0)} + \log \lambda_g^{\mathcal{F}}(0) \int_{\mathcal{K}} f^{\mathcal{F}}(\kappa | 0) \log \frac{f^{\mathcal{F}}(\kappa | 0)}{f^{\mathcal{H}}(\kappa | 0)} \ell_{\mathcal{K}}(d\kappa) \right]. \quad (14.8.15)$$

Notice that these equations have the same form as if we had taken expectations of a likelihood ratio, but in fact no true likelihood ratio exists here because the  $\mathcal{F}$ -intensity by itself does not define a process but only the way the point process component is determined by the full process.

See Exercise 14.8.7 for a hidden Markov model example.

The last question we take up in this section is that of approximating the expected entropy rate from finite samples, as in MacMillan's theorem for a discrete ergodic source with finite alphabet [see, e.g., Billingsley (1965)]. Here, the basic statements assert the convergence of the log likelihoods  $T^{-1} \log L_T$ , either a.s. or in  $L_1$  norm. We consider simple point processes only, and follow Papangelou (1978) in deriving the  $L_1$  version of this result. Extensions to the MPP context are sketched in Exercise 14.8.11.

The main problem is that, although we have derived expressions for the entropy rate from a form of likelihood involving the complete intensity function, in reality the best that is likely to be available is the conditional intensity based on the internal history, starting from an empty history at time 0. Thus there are two main steps in the proof: first, establish the convergence of the pseudolikelihoods in which the complete intensity plays the role of the intrinsic intensity over a finite interval, and second, show that the difference between the true and pseudolikelihoods is asymptotically negligible. These two steps are set out in the next two lemmas.

**Lemma 14.8.V.** *Suppose  $\mathbb{E}[\lambda^\dagger(0) \log \lambda^\dagger(0)] < \infty$  for a simple stationary point process  $N$ . Then as  $T \rightarrow \infty$ ,*

$$\frac{1}{T} \int_0^T \log \lambda^\dagger(t) dN(t) \rightarrow \mathbb{E}[\lambda^\dagger(0) \log \lambda^\dagger(0) | \mathcal{I}] \quad (14.8.16)$$

*both a.s. and in  $L_1$  norm, where  $\mathcal{I}$  is the  $\sigma$ -algebra of invariant events.*

**PROOF.** Because  $N$  is stationary, the process  $\lambda^\dagger(t)$  is stationary. Also,  $\lambda^\dagger(t)$  is  $\mathcal{H}^\dagger$ -predictable so the set function

$$\xi(A) = \int_A \log \lambda^\dagger(u) N(du) \quad (\text{bounded } A \in \mathcal{B})$$

may be regarded as a stationary random signed measure with mean density  $m = E[\lambda^\dagger(0) \log \lambda^\dagger(0)]$ . The result (14.8.16) then follows from the a.s. and  $L_1$  ergodic results of Proposition 12.2.IV. [This involves noting that the two processes

$$\xi_+(A) = \int_A [\log \lambda^\dagger(u)]_+ N(du) \quad \text{and} \quad \xi_-(A) = \int_A [\log \lambda^\dagger(u)]_- N(du)$$

are both nonnegative measures to which the theorems apply directly, with the finiteness of  $E[\lambda^\dagger(0) \log \lambda^\dagger(0)]$  following from  $x \log x \geq -e^{-1}$  ( $x \geq 0$ ) and the finiteness assumption in the lemma.]  $\square$

In the next result the monotone family of  $\sigma$ -algebras  $\{\mathcal{H}_{(-T,0)}\}$  generated by  $\{N(t): -T < t < 0\}$ , which increase as  $T \rightarrow \infty$  to  $\mathcal{H}_{0-}$ , plays a key role. First, the nonnegativity of  $\lambda \equiv \lambda^\dagger(0)$  and assumed finiteness of  $E(\lambda \log \lambda)$  ensure that  $E(\lambda) < \infty$ . Next, the family  $\{\lambda_T: 0 < T < \infty\}$ , where

$$\lambda_T = E[\lambda^\dagger(0) \mid \mathcal{H}_{(-T,0)}], \quad (14.8.17)$$

constitutes a martingale, which by definition and the finiteness of  $E(\lambda)$  is uniformly integrable, so  $\lambda_T \rightarrow \lambda$  both a.s. and in  $L_1$  norm by the martingale convergence theorem. But the function  $x \log x$  is convex in  $x$ , so by Jensen's inequality,

$$\infty > E(\lambda \log \lambda) = E[E(\lambda \log \lambda \mid \mathcal{H}_{(-T,0)})] \geq E(\lambda_T \log \lambda_T),$$

and because  $x \log x \geq -e^{-1}$  (all  $x \geq 0$ ), it follows that  $E(\lambda_T \log \lambda_T)$  is well-defined and finite.

**Lemma 14.8.VI.** *Under the conditions of Lemma 14.8.V, with  $\lambda_T$  as at (14.8.17),*

$$E[\lambda^\dagger(0) \log \lambda^\dagger(0) - \lambda^\dagger(0) \log \lambda_T] \rightarrow 0 \quad (T \rightarrow \infty). \quad (14.8.18)$$

PROOF. Because  $\lambda_T \rightarrow \lambda$  in distribution and  $\infty > E(\lambda \log \lambda) \geq E(\lambda_T \log \lambda_T)$ , and also  $0 \leq \lambda \log(\lambda_T/\lambda) I\{\lambda_T > \lambda\} \leq \lambda(\lambda_T/\lambda - 1) = \lambda_T - \lambda$  for which  $\lambda_T \rightarrow \lambda$  in  $L_1$  norm, it is enough to show that  $E(\lambda_T \log \lambda_T) \rightarrow E(\lambda \log \lambda)$ . Let  $x \geq 1$  be a continuity point of the distribution of  $\lambda$ ; then for sufficiently large  $x$  we can certainly make

$$E(\lambda I_{\{\lambda > x\}} \log \lambda) < \epsilon$$

for arbitrary  $\epsilon > 0$ . Now

$$\begin{aligned} & E[\max(x \log x, \lambda_T \log \lambda_T)] \\ &= E[\max(x \log x, E(\lambda \mid \mathcal{H}_{(-\infty,0)}) \log[E(\lambda \mid \mathcal{H}_{(-\infty,0)})])] \\ &\leq E[\max(x \log x, \lambda \log \lambda)] \end{aligned}$$

by Jensen's inequality because  $\max(x \log x, y \log y)$  is convex in  $y > 0$  for  $x \geq 0$ . Because  $x$  is a continuity point for  $\lambda$  with  $x \geq 1$ ,

$$\begin{aligned} 0 &\leq E(\lambda_T I_{\{\lambda_T > x\}} \log \lambda_T) \\ &= E[\max(x \log x, \lambda_T \log \lambda_T)] - x \log x \Pr\{\lambda_T \leq x\} \\ &\leq E[\max(x \log x, \lambda \log \lambda)] - x \log x \Pr\{\lambda_T \leq x\} \\ &\rightarrow E[\max(x \log x, \lambda \log \lambda)] - x \log x \Pr\{\lambda \leq x\} \\ &= E(\lambda I_{\{\lambda > x\}} \log \lambda) < \epsilon, \end{aligned}$$

with the convergence holding uniformly for  $T$  sufficiently large.  $\square$

**Theorem 14.8.VII.** *Let the simple stationary point process  $N$  admit  $\mathcal{H}$ -predictable complete intensity  $\lambda^\dagger(t)$  and  $\mathcal{H}$ -predictable conditional intensity  $\lambda^*(t)$  on  $t \geq 0$  and be such that*

$$H \equiv -E(\lambda^\dagger(0)[\log \lambda^\dagger(0) - 1])$$

*is finite, so that  $m = E[\lambda^\dagger(0)]$  is finite also. Then as  $T \rightarrow \infty$ ,*

$$\frac{H_{(0,T]}}{T} \rightarrow H \quad \text{a.s.} \quad (14.8.19)$$

and

$$\frac{\log L_{(0,T)}}{T} \rightarrow E(Z \mid \mathcal{I}) \quad \text{in } L_1 \text{ norm,} \quad (14.8.20)$$

where  $Z = \lambda^\dagger(0)[\log \lambda^\dagger(0) - 1]$  and  $\mathcal{I}$  denotes the  $\sigma$ -algebra of invariant events for  $N$ .

PROOF. Convergence as in (14.8.19) follows from the definition of  $H_{(0,T]}$ ,  $H$ , and the conditions and results of the last two lemmas. To prove (14.8.20), consider the difference

$$E \left| \frac{\log L_{(0,T)}}{T} - E(Z \mid \mathcal{I}) \right|, \quad (14.8.21)$$

which by virtue of the triangle inequality is dominated by  $T_1 + T_2 + T_3 + T_4$ , where

$$\begin{aligned} T_1 &= \frac{1}{T} E \left| \int_0^T \log \lambda^*(t) dN(t) - \int_0^T \log \lambda^\dagger(t) dN(t) \right|, \\ T_2 &= \frac{1}{T} E \left| \int_0^T \lambda^*(t) dt - \int_0^T \lambda^\dagger(t) dt \right|, \\ T_3 &= E \left| \frac{1}{T} \int_0^T \log \lambda^\dagger(t) dN(t) - E(\lambda^\dagger(0) \log \lambda^\dagger(0) \mid \mathcal{I}) \right|, \\ T_4 &= E \left| \frac{1}{T} \int_0^T \lambda^\dagger(t) dt - E(\lambda^\dagger(0) \mid \mathcal{I}) \right|. \end{aligned}$$

Here,  $T_3 \rightarrow 0$  by Lemma 14.8.V, and applying the ergodic theorem to the stationary process  $\lambda^\dagger(t)$  implies that  $T_4 \rightarrow 0$ . By assumption  $\lambda^*(t)$  and  $\lambda^\dagger(t)$  are predictable on  $\mathbb{R}_+$  and  $\mathbb{R}$ , respectively, so both are  $\mathcal{H}$ -predictable on  $\mathbb{R}_+$ , and thus  $T_1$  is dominated by

$$\frac{1}{T} \mathbb{E} \int_0^T |\log \lambda^*(t) - \log \lambda^\dagger(t)| \lambda^\dagger(t) dt.$$

Recall from the projection Theorem 14.2.II that  $\lambda^*(t)$  can be replaced by a suitably chosen version of  $\mathbb{E}[\lambda^*(t) | \mathcal{H}_{t-}]$  without altering the value of the integrals in  $T_1$  and  $T_2$ . Using stationarity, replace  $(0, T)$  by  $(-T, 0)$ , which leads to

$$\begin{aligned} T'_1 &= \frac{1}{T} \mathbb{E} \left[ \int_{-T}^0 |\log \mathbb{E}(\lambda^\dagger(t) | \mathcal{H}_{(-T,t)}) - \log \lambda^\dagger(t)| \lambda^\dagger(t) dt \right], \\ T'_2 &= \frac{1}{T} \mathbb{E} \left[ \int_{-T}^0 |\mathbb{E}(\lambda^\dagger(t) | \mathcal{H}_{(-T,t)}) - \lambda^\dagger(t)| dt \right]. \end{aligned}$$

For each fixed  $t$ , the expectation of the first integrand  $\rightarrow 0$  as  $T \rightarrow \infty$  by Lemma 14.8.VI and stationarity, so the  $(C, 1)$  mean also converges to zero. In the proof of the same lemma, the expectation of the second integrand also converges to zero, so the  $(C, 1)$  mean does also.  $\square$

A different approach is to ask for entropy rates associated with the point process on  $\mathbb{R}$  in its dual form as a stationary process of intervals. Extensions of McMillan's theorem from its original context of a discrete time finite-alphabet source (i.e., from a finite state space stochastic process) to stationary sequences of random variables with arbitrary distributions were developed by Perez (1959) and can be applied directly to the process of intervals. In particular, Perez showed that if  $\{X_n: n = 0, \pm 1, \dots\}$  is a stationary sequence of r.v.s taking their values on the c.s.m.s.  $\mathcal{X}$ , if  $\Pi$  is a fixed totally finite or  $\sigma$ -finite measure on  $\mathcal{X}$ , and if the fidi distributions  $F_k$  of order  $k$  of the sequence are absolutely continuous with respect to the  $k$ -fold product measure  $\Pi^{(k)} = \Pi \times \dots \times \Pi$  on  $(\mathcal{X}^{(k)}, \mathcal{B}(\mathcal{X}^{(k)}))$ , then

$$-\frac{1}{n} \log \left( \frac{dF_k(X_1, \dots, X_n)}{d\Pi^{(k)}(X_1, \dots, X_n)} \right) \rightarrow \mathbb{E}(Z | \mathcal{I}),$$

where the invariant r.v.  $Z$  has expectation (finite or infinite)

$$\begin{aligned} \mathbb{E}(Z) &= \mathbb{E} \int_0^\infty \frac{dF(x | \mathcal{H}_{(-1)}^*)}{\Pi(dx)} \log \left( \frac{dF(x | \mathcal{H}_{(-1)}^*)}{\Pi(dx)} \right) \Pi(dx) \\ &= \mathbb{E} \int_0^\infty \log \left( \frac{dF(x | \mathcal{H}_{(-1)}^*)}{\Pi(dx)} \right) dF(x | \mathcal{H}_{(-1)}^*), \end{aligned}$$

where  $F(\cdot \mid \mathcal{H}_{(-1)}^*)$  is a regular version of the conditional distribution of  $X_0$  given the sequence of past values  $\{X_{-1}, X_{-2}, \dots\}$  which generate  $\mathcal{H}_{(-1)}^*$ . This result can be applied directly to our context if we take  $\Pi(dx) = \mu e^{-\mu x} dx$  ( $x \geq 0$ ), that is, the stationary interval distribution on  $\mathbb{R}_+$  associated with the Poisson process with constant mean rate  $\mu$ . This leads to the result that

$$H_I(\mathcal{P}; \mathcal{P}_\mu) = -E_0 \int_0^\infty f(x \mid \boldsymbol{\tau}) \log f(x \mid \boldsymbol{\tau}) dx + \log \mu - \frac{\mu}{m},$$

where  $f(\cdot \mid \boldsymbol{\tau})$  is the conditional intensity introduced in Corollary 14.3.VI,  $E_0$  is used to denote expectations over the vector of past intervals  $\boldsymbol{\tau}$ , and  $H_I$  denotes the ‘interval entropy rate.’ If, in particular, we take  $\mu = 1$  and use  $\mathcal{Q}$  to denote the measure corresponding to  $\Pi(dx) = e^{-x+1/m} dx$ , we have similarly that

$$H_I \equiv H_I(\mathcal{P}; \mathcal{Q}) = -E_0 \int_0^\infty f(x \mid \boldsymbol{\tau}) \log f(x \mid \boldsymbol{\tau}) dx. \quad (14.8.22)$$

This interval entropy rate is easily related to the entropy rate  $H$  by appealing to Corollary 14.3.VI. As in the proof of Proposition 14.3.V, for any function  $h(T, \boldsymbol{\tau})$  of the backward recurrence time  $T$  and the past sequence  $\boldsymbol{\tau}$ ,

$$\begin{aligned} E h(T, \boldsymbol{\tau}) &= m E_0 \int_0^{\tau_0} h(\tau_0 - x, \boldsymbol{\tau}) dx = m E_0 \int_0^\infty f(y \mid \boldsymbol{\tau}) dy \int_0^y h(x, \boldsymbol{\tau}) dx \\ &= m E_0 \int_0^\infty h(x, \boldsymbol{\tau}) [1 - F(x \mid \boldsymbol{\tau})] dx. \end{aligned}$$

Now by taking for  $h(T, \boldsymbol{\tau})$  the function

$$\frac{f(T \mid \boldsymbol{\tau})}{1 - F(T \mid \boldsymbol{\tau})} \log \frac{f(T \mid \boldsymbol{\tau})}{1 - F(T \mid \boldsymbol{\tau})},$$

it follows from Corollary 14.8.V that

$$\begin{aligned} E(\lambda^\dagger(0) \log \lambda^\dagger(0)) &= m E_0 \left[ \int_0^\infty f(x \mid \boldsymbol{\tau}) \log f(x \mid \boldsymbol{\tau}) dx - \int_0^\infty \log[1 - F(x \mid \boldsymbol{\tau})] f(x \mid \boldsymbol{\tau}) dx \right] \\ &= m E_0 \left[ \int_0^\infty f(x \mid \boldsymbol{\tau}) \log f(x \mid \boldsymbol{\tau}) dx + 1 \right], \end{aligned}$$

and hence

$$\begin{aligned} H &= -E(\lambda^*(0) \log \lambda^*(0) - \lambda^*(0)) \\ &= -m E_0 \int_0^\infty f(x \mid \boldsymbol{\tau}) \log f(x \mid \boldsymbol{\tau}) dx = m H_I. \end{aligned}$$

Thus,  $H(\mathcal{P}; \mathcal{P}_\mu) = m H_I(\mathcal{P}; \mathcal{P}_\mu)$ , which leads to the following statement.

**Proposition 14.8.VIII.** *For a simple stationary point process with mean rate  $m$ , the entropy rate per unit time equals  $m$  times the entropy rate per interval.*

## Exercises and Complements to Section 14.8

14.8.1 Show that, in comparison with (14.8.4), the relative entropy relative to the Poisson process contains the additional term  $\ell(A)$ .

[Hint: Substitute for the Poisson probabilities in the expression at (14.8.5) for  $E_{\mathcal{P}}(\log(d\mathcal{P}/d\mathcal{Q}))$ , so the term  $\log q_k^{\text{sym}}$  reduces to  $\log([\ell(A)]^n)$ .]

14.8.2 (a) Show that, subject to the conditions  $\sum_{k=0}^{\infty} p_k = 1$ ,  $\sum_{k=1}^{\infty} k p_k = \mu$  and  $p_k \geq 0$ , the Poisson distribution maximizes the sum  $-\sum_{k=0}^{\infty} p_k \log(k! p_k)$ .

(b) Deduce that for a regular point process on a bounded interval  $D \subset \mathbb{R}^d$  with  $E[N(D)] = \mu = \text{const.}$ , the point process entropy (14.8.4) is maximized when the process is Poisson with uniform mean rate over  $D$ .

[Hint: Start by writing (14.8.4) in the form

$$-H = \sum p_k \log(k! p_k) + \sum p_k \int_{D^{(k)}} \pi_k(\mathbf{y}) \log \pi_k(\mathbf{y}) d\mathbf{y}.$$

Now use (a) together with the fact that, conditional on  $k$ , the integral is maximized, subject to  $\int \pi_k(\mathbf{y}) d\mathbf{y} = 1$ , when  $\pi_k(\mathbf{y})$  reduces to a uniform distribution over  $D^{(k)}$ .]

14.8.3 *Entropy of renewal process with gamma lifetimes.* For the gamma density function  $f_{\kappa}(x; a) = e^{-ax} (ax)^{\kappa-1} a / \Gamma(\kappa)$ , with shape parameter  $\kappa$  and mean  $\kappa/a$ , so mean rate  $m = a/\kappa$ , the information gain  $G$  at (14.8.10) equals  $m(1 - \log m + \int_0^{\infty} f_{\kappa}(x; a) \log f_{\kappa}(x; a) dx)$ . Show that the integral equals  $-\kappa - \log \Gamma(\kappa) + \log a + (\kappa - 1)\psi(\kappa)$ , where  $\psi(z) = \Gamma'(z)/\Gamma(z)$  is the digamma function. Known expansions for  $\psi$  yield

$$G = m\left(\frac{1}{2} + \frac{1}{2} \log(\kappa/2\pi) + \frac{1}{3}\kappa^{-1} + \frac{1}{12}\kappa^{-2} + \frac{1}{90}\kappa^{-3} - \frac{1}{120}\kappa^{-4} + \dots\right).$$

[Hint: Daley and Vere-Jones (2004) give more detail.]

14.8.4 Let  $(\Omega, \mathcal{E}, \mu)$  be a measure space and  $\mathcal{P}$  a probability measure on  $(\Omega, \mathcal{E})$  with  $\mathcal{P} \ll \mu$ . Let  $\{\mathcal{A}_{\alpha}\}$  be a family of finite or countable subalgebras of  $\mathcal{E}$ . Define the generalized entropies  $H(\mathcal{P}; \mu)$ ,  $H(\mathcal{P}_{\alpha}; \mu_{\alpha})$  by

$$\begin{aligned} -H(\mathcal{P}; \mu) &= \int \frac{d\mathcal{P}}{d\mu} \log \frac{d\mathcal{P}}{d\mu} \mu(d\omega) = \int \log \frac{d\mathcal{P}}{d\mu} \mathcal{P}(d\omega), \\ -H(\mathcal{P}_{\alpha}; \mu_{\alpha}) &= \sum_j \mathcal{P}(U_{\alpha j}) \log \frac{\mathcal{P}(U_{\alpha j})}{\mu(U_{\alpha j})}, \end{aligned}$$

where  $\{U_{\alpha j}\}$  is an irreducible (and countable) partition generating  $\mathcal{A}_{\alpha}$ .

(i) If  $\mathcal{A}_{\alpha} \subseteq \mathcal{A}_{\beta}$  then  $H(\mathcal{P}_{\alpha}; \mu_{\alpha}) \geq H(\mathcal{P}_{\beta}; \mu_{\beta})$ .

(ii) If  $\{\mathcal{A}_{\alpha}\}$  generates  $\mathcal{E}$ , then

$$-H(\mathcal{P}; \mu) = \inf_{\alpha} \{-H(\mathcal{P}_{\alpha}; \mu_{\alpha})\}.$$

(iii) Let  $\mathcal{P}_T, \mathcal{Q}_T$  be defined as in Proposition 14.8.IV, and suppose that  $\{\mathcal{T}_{\alpha}\}$  is a family of finite or countable partitions of the interval  $(0, T]$ , so that  $\mathcal{T}_{\alpha} = \{A_{\alpha i}; i = 1, \dots, K_{\alpha}\}$  such that  $\{\mathcal{T}_{\alpha}\}$  generates  $\mathcal{B}((0, T])$ . Let  $\mathcal{A}_{\alpha}$  be the algebra of events generated by  $Z_{\alpha i} = N(A_{\alpha i})$ . Show that

$$-H(\mathcal{P}_T; \mathcal{Q}_T) = \inf_{\alpha} \{-H(\mathcal{P}_{\alpha T}; \mathcal{Q}_{\alpha T})\},$$

where

$$-H(\mathcal{P}_{\alpha T}; \mathcal{Q}_{\alpha T}) = \sum_{\mathbf{n}} p_{\alpha \mathbf{n}} \log \frac{p_{\alpha \mathbf{n}}}{q_{\alpha \mathbf{n}}},$$

$\mathbf{n} = (n_1, \dots, n_{K_\alpha})$ ,  $p_{\alpha \mathbf{n}} = \mathcal{P}\{Z_{\alpha i} = n_i, i = 1, \dots, K_\alpha\}$ , and  $q_{\alpha \mathbf{n}} = \prod_{i=1}^{K_\alpha} [\ell(A_{\alpha i})]^{n_i}$ . [This result provides a discrete approximation to generalized entropy. See Csiszar (1969) and Fritz (1973).]

14.8.5 (a) Show that  $H(\mathcal{P}_T; \mathcal{Q}_T)$  can also be characterized as

$$-H(\mathcal{P}_T; \mathcal{Q}_T) = \inf_{\alpha} \left\{ \sum p_{\alpha \mathbf{n}} \log p_{\alpha \mathbf{n}} - T \log \delta_{\alpha} \right\},$$

where  $\delta_{\alpha} = \max_i \mu(A_{\alpha i})$ .

(b) Show also that

$$-H(\mathcal{P}_T; \mathcal{Q}_T) = \inf_{\alpha} \left\{ \pi_{\alpha \mathbf{n}} \log \frac{\pi_{\alpha \mathbf{n}}}{q_{\alpha \mathbf{n}}} + \sum_{i=1}^{K_{\alpha}} \mathcal{P}\{Z_{\alpha i} \geq 1\} \log \mu(A_{\alpha i}) \right\},$$

where  $\pi_{\alpha \mathbf{n}} = \mathcal{P}\{Y_{\alpha i} = n_i, i = 1, \dots, K_{\alpha}\}$  for  $n_i = 0$  or  $1$  and  $Y_{\alpha i} = I_{\{N(A_{\alpha i}) > 0\}}$ . [Hint: Show first that if  $\mathcal{T}_{\alpha} \subseteq \mathcal{T}_{\beta}$  and each set  $A_{\alpha i}$  of  $\mathcal{T}_{\alpha}$  is a union of no more than  $r$  sets of  $\mathcal{T}_{\beta}$ , then

$$\sum p_{\alpha \mathbf{n}} \log p_{\alpha \mathbf{n}} \leq \sum p_{\beta \mathbf{n}} \log p_{\beta \mathbf{n}} \leq \sum p_{\alpha \mathbf{n}} \log p_{\alpha \mathbf{n}} + T \log r.$$

See Fritz (1973) for further details. This result shows that in some sense  $E(N(0, T])$  plays the role of a dimension for the process, as in Rényi's (1959) discussion of 'dimensional entropy.']

14.8.6 Generalize the results of Exercises 14.8.4–5 to the case of a general state space  $\mathcal{X}$  with Poisson distribution having nonatomic parameter measure  $\mu(\cdot)$ .

14.8.7 *Information gain in a Markov modulated point process (MMPP).*

- Let  $X(t)$  be a stationary continuous-time Markov process on finite state space  $\mathbb{X}$ , with stationary distribution  $\pi_i = \Pr\{X(t) = i\}$  (all  $t, i \in \mathbb{X}$ ), so the rate of jumps between states equals  $\sum_{i \in \mathbb{X}} \pi_i q_i = \lambda$  say, where the  $q_i$  are as in Section 10.3 below (10.3.2). Write down an expression for the information gain of the point process consisting of jumps of the Markov process relative to a Poisson process at rate  $\lambda$ .
- Suppose that, when  $X(t) = i$ , points occur in a Poisson process at rate  $\lambda_i$  [i.e., the point process is a Cox process directed by the states of the hidden Markov process  $X(\cdot)$ ]. Find the information gain of this point process compared with a Poisson process at rate  $\sum_{i \in \mathbb{X}} \pi_i \lambda_i$ .
- Compare the results of (a) and (b) and interpret.
- Consider extensions to (b) when it is a MPP that is directed by  $X(\cdot)$ , where both the mark distribution and the frequency of observed points may depend on  $X(\cdot)$ .

14.8.8 Calculate the entropy rate for a stationary renewal process whose lifetime p.d.f. has unit mean and

- is exponential on  $(0, \infty)$ ; or else, for some  $a$  in  $(0, 1)$ , is
  - uniform on  $(1 - a, 1 + a)$ , or
  - triangular on  $(1 - a, 1 + a)$ .
- (d) Interpret the limits from (b) and (c) when  $a \rightarrow 0$ .



- 14.8.9 The entropy rate of a stationary renewal process with absolutely continuous lifetime d.f. with density  $f(\cdot)$  equals  $\int_0^\infty f(x) \log f(x) dx$ , where  $m^{-1} = \int_0^\infty xf(x) dx$ . One technique used in approximating the behaviour of a stationary point process  $N$  on  $\mathbb{R}$  (e.g., for simulation purposes) is to replace  $N$  by a renewal process whose lifetime d.f. coincides with the stationary interval distribution. Use Proposition 14.8.VIII to show that the entropy rate of such an approximating stationary renewal process is larger than that of  $N$ .
- 14.8.10 Investigate extensions to MPPs of the arguments used to prove Lemmas 14.8.V–VI and Theorem 14.8.VII. Show also that the entropy rate for a stationary MPP can be given in a form analogous to that of Proposition 14.8.VIII by considering the bivariate sequence  $\{(\tau_i, \kappa_{i-1})\}$ , where  $\kappa_i$  is the mark associated with the point initiating the interval  $\tau_i$ , and applying the corresponding version of Perez's result noted before Proposition 14.8.VIII.

## CHAPTER 15

# Spatial Point Processes

15.1	Descriptive Aspects: Distance Properties	458
15.2	Directional Properties and Isotropy	466
15.3	Stationary Line Processes in the Plane	471
15.4	Space–Time Processes	485
15.5	The Papangelou Intensity and Finite Point Patterns	506
15.6	Modified Campbell Measures and Papangelou Kernels	518
15.7	The Papangelou Intensity Measure and Exvisibility	526

This last chapter provides an introduction to spatial point processes, meaning for the most part results for point processes in  $\mathbb{R}^2$  and  $\mathbb{R}^3$  where the order properties of the real line, which governed the development in the preceding chapter, are no longer available.

During the last few decades, the rapid growth of interest in image processing has brought about substantial treatments of spatial models, including both engineering and statistical aspects; see in particular Ripley (1981), Baddeley (1998), and van Lieshout (1995). At the same time, the collection of improved quality spatial data in ecology, geography, forestry, geophysics, and astronomy, has maintained a steady demand for spatial statistical models, for which Stoyan, Kendall and Mecke’s text SKM (1995) is an extensive general reference.

The material we present falls into two main components. In the first four sections we review mainly descriptive properties, distinguishing between distance and directional properties of spatial point patterns, starting from finite models, moving on to the moment properties of line processes, and then revisiting space–time models, where time reappears so that many of the modelling concepts in Chapter 14 are again available, but spatial patterns also play an important role.

The three final sections of the chapter provide an introduction to modelling centred around the concept of the Papangelou intensity; we provide some background and motivation from the statistical and physical settings, then attempt an introduction to the more mathematical theory.

This chapter also includes an introduction, mainly through the treatment of line processes in Section 15.3, to the rich and diverse territory of stochastic geometry, pioneered by Rollo Davidson, David Kendall, and others in the 1970s [see especially Kendall's (1974) introduction to Harding and Kendall (1974)]. For present purposes, we may take stochastic geometry to mean the study of families of geometric objects randomly located in one-, two- or three-dimensional Euclidean space, where to qualify as an 'object' all we demand is that the entity can be specified by a finite (or perhaps countably infinite) set of real parameters that describe aspects such as location, size, and shape.

To each such object there corresponds a point in a Euclidean parameter space of suitably high dimension, and random families of such objects can be defined as point processes on this parameter space as state space. Because a characteristic feature of geometric objects is their invariance under rigid motions such as translation, rotation, and reflection, a key question is the implication of such invariance properties on the first and second moment properties of the process.

The results for isotropic point processes considered in Section 15.2, and those for line processes in Section 15.3, both illustrate this general theme. The results are still applications of the factorization Lemma A2.7.II or equivalent disintegration results, but as the objects become more complex, so also do the disintegrations become more varied and more intricate.

The final three sections of the chapter introduce a rather different aspect of the theory of spatial point processes, where the underlying endeavour is to use the concepts of interior/exterior to provide some kind of weak counterpart to the ideas of past/future on the time axis. Central to this endeavour is the concept of the Papangelou intensity, which underlies recent developments in inference for spatial point processes, such as pseudolikelihood methods and point process residuals. For finite point processes, the properties of the Papangelou intensity can be developed in a relatively elementary manner from the theory of Janossy densities outlined in Chapter 5, and this is undertaken in Section 15.5.

The extension to general processes is altogether more demanding, requiring a combination of deep concepts from statistical mechanics and general point process theory. An introduction to this material is contained in Sections 15.5 and 15.6, and centred round the concept of exterior conditioning, meaning a conditioning of the point process on its behaviour outside a bounded set. In this sense the theory can be thought of as a kind of dual to the Palm theory, which is concerned with conditioning on the behaviour within a bounded set, as the dimensions of that set shrink to zero.

## 15.1. Descriptive Aspects: Distance Properties

Faced with a realization of a point process within a bounded region of  $\mathbb{R}^2$ , or a *spatial point pattern* as we generally describe it in this chapter, a statistician's first reaction is likely to be to seek some numerical characteristics with

which to describe its salient features. Spatial point patterns being, in general, objects of some complexity, a variety of different statistics has been developed for this purpose. In this section it is our aim to give a brief overview of some of these quantities, without getting too deeply involved with technical issues such as consistency, unbiasedness, or numerical stability. It is our concern rather to identify and place in context the model characteristics to which the statistics refer. More comprehensive introductions to spatial statistics, including point process models in particular, can be found in Ripley (1981), Diggle (1983, 2003), SKM (1987) and its second edition SKM (1995), Cressie (1991, 1993), van Lieshout (2000), Baddeley *et al.* (2005), and a broad collection of case studies in Baddeley *et al.* (2006).

Crudely speaking, the models that are available to describe point processes in  $\mathbb{R}^d$  for  $d = 2, 3, \dots$  are derived from Poisson processes, Gibbs processes, or (deterministic) lattice processes that may have undergone modification by translation, clustering, or inhibition. To the extent that features of these modelling mechanisms have been discussed earlier in the book we should have little more to say. Nevertheless it is worth noting briefly some properties that have been developed to describe how particular models and/or datasets may deviate from the simplest underlying structure meaning, most commonly, the ‘complete randomness’ as for a Poisson process.

Some of the earliest characteristics to be studied relate to nearest-neighbour distances, which have long been used to assist both in estimating areal densities and in classifying cluster properties. Indeed, they relate to some of the earliest applications of point process ideas in forestry, ecology, and elsewhere [see, e.g., Matérn (1960) and Warren (1962, 1971)]. The functions most frequently used relate to stationary point processes (Definition 12.1.II), often called homogeneous point processes in purely spatial contexts, in which case descriptions in terms of the Palm probability measure  $\mathcal{P}_0$  can also be used. Homogeneity in space is a major simplifying factor in conceptual models, but is a rare phenomenon in the real world, so that from a practical point of view, a rudimentary understanding of the behaviour of characteristics to be expected when the true model departs in different ways from homogeneity is also important.

The first such quantity we consider is a particular example of the avoidance function or avoidance probability [equations (2.3.1), (9.2.11) and Example 5.4(a)], namely,

$$1 - F(r) = \mathcal{P}\{N(S_r) = 0\},$$

where in this section we mostly write  $S_r = S_r(0)$  for the circle (or in  $\mathbb{R}^d$ , the sphere) of radius  $r$  with centre at the origin. When  $N$  is stationary, the function  $F(r)$  is also known as the *spherical contact distribution* [denoted  $H_s(r)$  in SKM (1995, pp. 72, 80)], or *empty space function*, because  $F(r) = \mathcal{P}\{N(S_r) > 0\}$  is the probability that a sphere of radius  $r$  makes contact with a point of  $N$ . It is also the distribution function of the distance from the origin to the point  $x^*(N)$  of (13.3.7).

The other function which plays a central role in this context that  $N$  is stationary is the *nearest-neighbour function* itself,

$$G(r) = \mathcal{P}_0\{N(S_r \setminus \{0\}) > 0\},$$

denoted  $D(r)$  in SKM (1995). It is the distribution function of the distance from an arbitrary point of the process, selected as origin, to the nearest other point of the process, or equivalently the Palm probability version of the spherical contact distribution. Its form for the Neyman–Scott process is given in Exercise 6.3.10 [see also Example 15.1(a) and Exercise 15.1.3].

The ratio of the two survivor functions,

$$J(r) = \begin{cases} \frac{1 - G(r)}{1 - F(r)} & \text{if } F(r) < 1, \\ 1 & \text{if } F(r) = 1, \end{cases} \quad (15.1.1)$$

is an indicator, relative to a Poisson process for which  $J(r) = 1$  (all  $r$ ), of clustering (when  $< 1$ ) or ‘regularity’ or inhibition (when  $> 1$ ) at varying distances  $r$ . The notation  $J(\cdot)$  follows van Lieshout and Baddeley (1996); Diggle (1983) uses  $q^*(\cdot)$  for the same function in the setting of a Poisson cluster process as in Example 15.1(a), but the concept dates at least to Warren (1971). Note too that for  $F(r) < 1$ , (15.1.1) has the alternative expression

$$J(r) = \frac{\mathcal{P}_0\{N(S_r \setminus \{0\}) = 0\}}{\mathcal{P}\{N(S_r) = 0\}}. \quad (15.1.1')$$

The similar ratios

$$\frac{\mathcal{P}_0\{N(S_r \setminus \{0\}) = k\}}{\mathcal{P}\{N(S_r) = k\}} \quad \text{and} \quad \frac{\mathcal{P}_0\{N(S_r \setminus \{0\}) \leq k\}}{\mathcal{P}\{N(S_r) \leq k\}} \quad (15.1.2)$$

are both identically 1 in  $r > 0$  for every  $k$  in any space  $\mathbb{R}^d$  when  $N$  is Poisson; deviations from 1 for different  $k$  may indicate more detail than  $J(\cdot)$  concerning clustering or inhibition.

More generally, as in SKM (1995, Section 4.1), take a convex compact set  $B \ni 0$  and define the *contact distribution function*  $H_B$  by

$$H_B(r) = \mathcal{P}\{N(rB) > 0\},$$

irrespective of  $\mathcal{P}$  being stationary or not. We consider these contact distributions only for the spherical case underlying  $J(r)$ . In assessing the behaviour of empirical estimates of the  $J$ -function, it is important to bear in mind the possibility of non-stationarity, where the value may depend on the spatial origin.

As a general rule, explicit expressions for the  $J$ -function are rather difficult to obtain. The first example below summarizes the more tractable results for Poisson cluster processes. Mase (1986, 1990) reviews difficulties in approximating Gibbs processes. Baddeley *et al.* (2005) and van Lieshout (2006a) examine a range of further results and examples.

EXAMPLE 15.1(a) *Two-dimensional stationary Poisson cluster process; centre-satellite process* [see Example 6.3(a) and Exercise 6.3.10]. Equation (6.3.14) of Proposition 6.3.III gives the empty space function for a general Poisson cluster process in a form which here reduces to

$$1 - F(r) = \exp \left( - \int_{\mathbb{R}^2} p_{S_r}(y) \mu_c \, dy \right), \quad (15.1.3)$$

where  $\mu_c$  is the intensity for the Poisson process of cluster centres (here assumed stationary) and  $p_{S_r}(y)$  is the probability that a cluster with centre at  $y$  has no members within  $S_r$ .

To find the  $G$ -function, suppose given a point of the process at the origin, and consider separately the distance to the nearest point from the same cluster, and to the nearest point from a different cluster. For any given cluster structure, there will be a well-defined distribution function tail,  $Q_{cl}(r)$  say, for the probability that within a distance  $r$  of some given point of a cluster there is no other point of the same cluster. The distance to the nearest point in a different cluster, however, has the same distribution  $F(r)$  as in (15.1.3). This implies [Diggle (1983, equation (4.6.5))] that

$$1 - G(r) = Q_{cl}(r) [1 - F(r)], \quad (15.1.4)$$

and hence that  $J(r) = Q_{cl}(r)$ . Thus, for a stationary Poisson cluster process,  $J(r)$  is equal to the probability that no two points from the same cluster lie within a distance  $r$  of each other, and therefore satisfies  $1 \geq J(r) \downarrow (0 \leq r \uparrow)$ .

Under more specific assumptions, the functions  $p_{S_r}(y)$  and  $Q_{cl}(r)$  can be evaluated explicitly. For example, for a Neyman–Scott process with isotropic normal distributions about the cluster centre, the function  $p_{S_r}(y)$  is evaluated in Exercise 6.3.10. If  $q_j$  denotes the probability that a cluster contains  $j$  points, and  $m_{cl} \equiv \sum_{j=1}^{\infty} j q_j < \infty$ , then  $Q_{cl}(\cdot)$  is evaluated over a cluster with  $j$  points with probability  $j q_j / m_{cl}$ ; see Exercise 15.1.3 and Warren (1971).

As a further instructive example, albeit special, consider a stationary Poisson cluster process whose clusters consist of exactly two points, one at the cluster centre and the other uniformly distributed on a circle of radius  $R$  around the cluster centre. Then it is impossible for two points from the same cluster to lie within a circle of radius less than  $\frac{1}{2}R$ , and certain that they will do so for some circle of any larger radius. Because the clusters consist of exactly two points, it follows immediately that  $Q_{cl}(r) = 0$  or 1 according as  $r < \text{or } \geq R$ . For circles of radius  $< \frac{1}{2}R$ , the process looks like a Poisson process. Some further details are given in Exercise 15.1.4.

This example can be adapted to furnish counterexamples in the case of anisotropy (cf. the next section).  $\square$

When the point process is stationary, we can find expressions for the  $F$ - and  $G$ -functions in terms of the local Janossy measures of Definition 5.4.IV. Thus, the empty space function  $F(x)$  is exactly the function  $J_0(\emptyset \mid S_x(0))$ ,

and this allows  $F(x)$  to be expressed in terms of the factorial moment densities (when these exist) as in (5.4.14):

$$\begin{aligned} F(x) = J_0(\emptyset \mid S_x(0)) &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \int_{S_x} \cdots \int_{S_x} m_{[k]}(y_1, \dots, y_k) dy_1 \dots dy_k \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} M_{[k]}[(S_x)^{(k)}]. \end{aligned} \quad (15.1.5)$$

A similar expansion holds for the nearest-neighbour function  $G(x)$  in terms of the empty space function and moment densities of the Palm distribution; note that such densities exclude the point at the origin. For a stationary process, this reduces to an expansion in terms of the reduced moment densities of the original process (cf. Proposition 13.2.VI). Thus for  $G(x)$ , (15.1.5) continues to hold with  $\check{M}_{[k]}[(S_x)^{(k)}] = m^{-1} \check{M}_{[k+1]}[(S_x)^{(k)}]$ , where  $m$  is the mean rate, in place of  $M_{[k]}[(S_x)^{(k)}]$ . A careful discussion of these and related expansions is given in van Lieshout (2006b), relating the moment measures for the Palm process to the Papangelou intensities through the Georgii–Nguyen–Zessin formula; see also the discussion of these topics in Section 15.5.

The  $F$ -,  $G$ -, and  $J$ -functions can be extended to MPPs provided due care is taken to specify the marks of the points appearing in the definitions. The situation here is analogous to that encountered in defining the Palm distributions  $\mathcal{P}_{(0,\kappa)}$  of an MPP, where it is necessary to specify the mark  $\kappa$  of the point at the origin. Thus, in specifying the empty space function for an MPP we need to distinguish between the empty space function for the ground process,  $F_g(x)$  say, which determines the distance from an arbitrary origin to any point of the process, regardless of its mark, and the more general family of functions  $F_B(x)$  determining the distance from such an origin to the first point with mark in the subset  $B \in \mathcal{B}_\mathcal{K}$ .

For nearest-neighbour distances there are in principle four different options to consider: the distance from a point of the process with arbitrary mark to the nearest point with arbitrary mark (giving the nearest-neighbour distribution function  $G_g(x)$  for the ground process); the distance from a point with arbitrary mark at the origin to the nearest neighbour with mark in a specified set  $B$  [giving the distribution  $G_{(g,B)}(x)$ , say]; the distance from a point at the origin with specified mark  $\kappa$  to the nearest point of the process regardless of its mark [giving  $G_{(\kappa,g)}(x)$  say]; and the distance from a point with mark  $\kappa$  at the origin to the nearest point with mark in the subset  $B \in \mathcal{B}_\mathcal{K}$  [giving  $G_{(\kappa,B)}(x)$ ]. The next example examines these options for the simplest case of an MPP with independent marks.

**EXAMPLE 15.1(b) Processes with independent marks** [see Definition 6.4.III]. Since the mark on the point at the origin is independent of the marks and locations of all further points, it has no effect on the nearest-neighbour distances. Thus, for independent marks we find

$$G_{(\kappa,g)}(x) = G_g(x); \quad J_{(\kappa,g)}(x) = J_g(x),$$

and similarly

$$G_{(\kappa, B)}(x) = G_{(g, B)}(x); \quad J_{(\kappa, B)} = J_{(g, B)}(x).$$

The nearest distance from the origin to a point within the mark set  $B$  here corresponds to the same distance for a point process obtained from the original by independent thinnings (Section 11.2) with thinning probability  $p = \int_B F(dx)$ . The effect is to multiply the  $k$ th factorial density in (15.1.5) by the factor  $p^k$ , so that we obtain

$$\begin{aligned} F_B(x) &= \sum_{k=0}^{\infty} \frac{(-p)^k}{k!} \int_{S_x} \cdots \int_{S_x} m_{[k]}^g(y_1, \dots, y_k) dy_1 \cdots dy_k \\ &= \sum_{k=0}^{\infty} \frac{(-p)^k}{k!} M_{[k]}^g[(S_x)^{(k)}], \end{aligned}$$

where the moment densities refer to those of the ground process.

A similar modification occurs for the corresponding  $G$ -function, irrespective of the mark at the origin.  $\square$

Practical estimation of the  $F$ - and  $G$ -functions raises the usual problems of allowing for edge effects and possible biases arising from nonhomogeneity. Ripley (1988) and Stoyan and Stoyan (1994) are among the several texts which examine such problems in depth. Here we mention only the edge correction for estimates of the nearest-neighbour distribution proposed in Hanisch (1984). This has the advantage of preserving the monotonicity of the estimate as a function of  $r$ . It replaces the naïve estimate

$$\hat{G}(r) = \frac{1}{N(W)} \sum_{k=1}^{N(W)} I_{N[(S_r(x_k))=0]}$$

with the form

$$\hat{G}_H(r) = \frac{\ell(W)}{N(W)} \sum_{k=1}^{N(W)} \frac{I_{N[(S_r(x_k)) \cap W = 0]}}{\ell(W^{-d(x_k, \partial W)})} \quad (15.1.6)$$

where  $d(x, \partial W)$  is the distance from the point  $x$  to the boundary of the observation region  $W$ , and  $A^{-\epsilon} = \{x \in A: \rho(x, A^c) > \epsilon\} = ((A^c)^\epsilon)^c$  denotes the  $\epsilon$ -interior of  $A$  [cf. the  $\epsilon$ -halo set at (A2.2.2);  $A^{-\epsilon}$  is defined for convex  $A$  below (12.2.12)]. The interpretation is that when a point  $x$  is too close to the boundary of  $W$  for the ball  $S_r(x)$  to be wholly contained in  $W$ , the count from  $S_r(x) \cap W$  is inflated by the weight factor  $\ell(W)/\ell(W^{-d(x_k, \partial W)})$ . Monotonicity and unbiasedness properties are outlined in Exercise 15.1.8.

The other quantities we briefly mention in this section are the distributions of point-to-point distances whose properties are summarized in the moment measures, especially the second-order or two-point moment measure. Again we assume stationarity, so that the quantities of principal importance are the



reduced moment measures of Section 12.6, or their equivalent representations as moment measures of one order lower for the Palm distributions, as in Proposition 13.2.VI. For spatial processes these reduced moment measures are functions of vector differences  $u = x_1 - x_2$ , and so can be represented in terms of polar or spherical polar coordinates. Thus, for example, if  $u = (r, \theta)$ ,  $\check{M}_2(u)$  can be factorized (disintegrated) into a marginal measure  $\check{K}_2(dr)$  and a family of conditional distributions  $\Gamma(d\theta | r)$  describing the distribution of the angle  $\theta$  for given  $r$ .

Decompositions of this kind are examined in more detail in the next section. Here we note that Ripley's  $K$ -function at (8.1.21), namely,

$$K(r) = (1/m^2)\check{K}_2(r) = (1/m^2)\check{M}_2(S_r \setminus \{0\}) = (1/m)\check{M}_1(S_r \setminus \{0\}), \quad (15.1.7)$$

is widely used alongside the  $F$ -,  $G$ -, and  $J$ -functions as a useful descriptive characteristic of spatial point patterns. It measures the rate of growth of the reduced second moment measure with distance  $r$  from the origin, and can be defined for both isotropic and anisotropic processes, although the former is often assumed. The behaviour of  $\check{K}_2(r)$  for the isotropic Neyman–Scott process referred to in Example 15.1(a) is described in Example 8.1(b); see also Exercise 15.1.2.

## Exercises and Complements to Section 15.1

15.1.1 The Slivnyak–Mecke Theorem 13.1.VII implies that for a Poisson process,  $J(r) = 1$  for all  $r$ . Investigate whether there are analogues of the constructions in Exercises 2.3.1 and 4.5.12 showing that there exist non-Poisson processes with this property.

15.1.2 Divide  $\mathbb{R}^2$  into unit squares. Independently to each square allocate  $N'$  points with the common distribution  $\Pr\{N' = j\} = \frac{1}{10}, \frac{8}{9}, \text{ and } \frac{1}{90}$  for  $j = 0, 1, 10$ , respectively, and for each square distribute its  $N'$  points uniformly over the square. Check that for any Borel set  $A$ , the first two moment measures  $M(A)$  and  $M_2(A)$  for the process are the same as for a Poisson process at unit rate, and hence conclude that Ripley's  $K$ -function is the same for both processes. [Hint: Baddeley and Silverman (1984) give this example with a plot of a realization, remarking that the plot is visually quite different from that of a Poisson process. They reference several other similar counterexamples.]

15.1.3 Consider a stationary Poisson cluster process in  $\mathbb{R}^2$  as in Example 15.1(a) when a typical cluster member is as in a Neyman–Scott process consisting of  $j$  points with probability  $q_j$  such that  $m_{cl} \equiv \sum_{j=0}^{\infty} j q_j < \infty$ , each point being i.i.d. about the cluster centre with a radially symmetric distribution for which  $\Pr\{\text{point lies within distance } r \text{ of centre}\} = \int_0^r u f(u) du$ . Show that

$$Q_{cl}(r) = \sum_{k=1}^{\infty} \frac{k q_k}{m_{cl}} \int_0^{\infty} [1 - P(r | y)]^{k-1} y f(y) dy,$$

where

$$P(r | y) = \begin{cases} \int_0^{r-y} f(x) dx + \int_{r-y}^{r+y} f(x) l_r(x, y) dx & (r > y), \\ \int_{y-r}^{y+r} f(x) l_r(x, y) dx & (r < y), \end{cases}$$

denotes the probability that a particular point of the cluster lies within a circle of radius  $r$  and centre at distance  $y$  from the cluster centre and  $l_r(x, y)$  is the length of the segment of a circle of radius  $x$  and centre 0 intersected by a circle of radius  $r$  and centre at distance  $y$  from 0.

For the spatial Neyman–Scott process of Exercise 6.3.10, conclude that the function  $Q_{\text{cl}}(\cdot)$  of (15.1.4) can be represented in the form

$$Q_{\text{cl}}(r) = \sum_{k=1}^{\infty} \frac{kq_k}{m_{\text{cl}}} \left[ \int_0^{\infty} e^{-x^2/2} x l_r(x, y) dx \right]^{k-1}.$$

- 15.1.4 Show that for the Neyman–Scott process of Example 6.3(a), the function  $r_2(x)$  in (8.1.19), a standardized conditional intensity for a point at  $x$  given a point at the origin, is given by

$$r_2(x) = 1 + \frac{m_{[2]}}{m} \int_{\mathcal{X}} f(x+u)f(u) du$$

when the distribution of cluster points from the cluster centre has density function  $f(\cdot)$  and  $m, m_{[2]}$  are the first two factorial moments of the cluster size distribution [cf. also (6.3.19)].

- 15.1.5 Consider the particular Poisson cluster process  $N$  with two-point clusters described at the end of Example 15.1(a), setting  $R = 1$ . Regard  $N = N_c + N_s$  as the superposition of two stationary dependent Poisson processes, each at rate  $\mu_c$ ,  $N_c$  consisting of the cluster centres and  $N_s$  of the points of the clusters at unit distance and random orientation relative to the centres.

- (a) Show that the empty space function  $F(r)$  is given by

$$\begin{aligned} 1 - F(r) &= \Pr\{N_c(S_r) = 0, N_s(S_r) = 0\} \\ &= \Pr\{N_c(S_r) = 0\} \Pr\{N_s(S_r) = 0 \mid N_c(S_r) = 0\}. \end{aligned}$$

The first term in the product equals  $e^{-\mu_c \pi r^2}$ . The other term is the same for  $r \leq \frac{1}{2}$  but for larger  $r$  we must evaluate  $\Pr\{\text{no cluster centre outside } S_r \text{ has a component point inside } S_r\}$ . Use standard Poisson process arguments to exclude points inside  $S_r$  from centres at distance  $y$  from the centre of  $S_r$ , with  $r < y < r + 1$ , to conclude that

$$1 - F(r) = \begin{cases} e^{-2\mu_c \pi r^2} & (0 < r \leq \frac{1}{2}), \\ e^{-\mu_c \pi r^2} \exp\left(-\mu_c \int_r^{r+1} \frac{2 \arccos \frac{1+y^2-r^2}{2y}}{2\pi} 2\pi y dy\right) & (\frac{1}{2} < r). \end{cases}$$

As a check, evaluate the latter case in the limit  $r \downarrow \frac{1}{2}$ .

- (b) Show that the first and second moment measures for this process are given by

$$\begin{aligned} \Pr\{N(dx_1) = 1\} &= 2\mu_c \ell(dx_1), \\ \Pr\{N(dx_1) = 1, N(dx_2) = 1\} &= \begin{cases} 4\mu_c^2 \ell(dx_1) \ell(dx_2) & \text{if } |x_1 - x_2| \neq 1, \\ 2\mu_c \ell(dx_1) I_{|x_1 - x_2| = 1} \frac{d\theta}{2\pi} & \text{otherwise,} \end{cases} \end{aligned}$$

and in the latter case,  $x_2$  is expressed in polar coordinates relative to  $x_1$ .

- (c) Evaluate the ratios at (15.1.2) for  $k = 1, 2, \dots$  and compare with  $J(r)$ .

[Remark: Since the clusters have two points, this is a Gauss–Poisson process.]

15.1.6 For a stationary point process in  $\mathbb{R}^d$ , express the function  $G(\cdot)$  in (15.1.1) as

$$G(r) = \lim_{\delta \downarrow 0} \frac{\bar{F}_\delta(r) - \bar{F}(r)}{F(\delta)},$$

where  $F(r) = \Pr\{N(S_r) > 0\}$ ,  $F_\delta(r) = \Pr\{0 < N(S_\delta) = N(S_r \setminus S_\delta)\}$ , and  $\bar{F}(r) = 1 - F(r)$  [cf. Chapter 3, Ambartzumian (1972), Paloheimo (1971)].

15.1.7 Consider Matérn's Model I in  $\mathbb{R}^2$  [see Example 8.1(c) and Exercise 8.1.8]. Show that for  $0 < r < R$ ,  $G(r) = 0$  and hence  $J(r) > 1$  for such  $r$ . [The same holds true for any hard-core model as described in Example 5.3(c).]

15.1.8 *Hanisch-type edge corrections* [Hanisch (1984)] are of the form

$$\hat{M}_0(A) = \sum_{x_k \in W} \frac{N[(x_k + A) \cup W]}{\ell(W - d(x_k, \partial W))},$$

where  $d(x, \partial W)$  is the distance from the point  $x$  to the boundary of the observation region  $W$ ,  $A$  is a test set, and the function being estimated is the first-order moment of the Palm distribution.

(a) Show that for the edge-corrected estimate (15.1.6) of the nearest-neighbour distribution, and assuming a simple, stationary, homogeneous process with finite intensity  $\lambda$  (= the mean rate for a simple process),

(i) for  $r_1 < r_2$ ,  $\hat{G}_H(r_1) \leq \hat{G}_H(r_2)$ ; and

(ii)  $E\left(\sum_{k=1}^{N(W)} \frac{I_{\{N[(S_r(x_k)) \cap W] = 0\}}}{\ell(W - d(x_k, \partial W))}\right) = \lambda G(r)$ .

[Hint: For (ii), write the left-hand side as the expected value of an integral against  $N$  and use the basic formula (13.3.2) for the Palm distribution.]

(b) Define analogous edge-corrected estimates of the marked versions  $G_g(r)$ ,  $G_{(g,B)}(r)$  defined above Example 15.1(b), and show that they have similar monotonicity and unbiasedness properties. [Hint: van Lieshout (2006b).]

(c) Investigate estimates of the same type for other functionals of the Palm process.

## 15.2. Directional Properties and Isotropy

Consider first a point process in  $\mathbb{R}^2$  whose probability structure is invariant under rotations about a fixed point in the plane. Such a process may represent, for example, the distribution of seedlings about a parent plant or of animals or other organisms about a nest or point of release (Byth, 1981). It is natural in such a case to take the fixed point as origin and to represent the points on the plane in terms of polar coordinates  $(r, \theta)$ , with  $0 \leq r < \infty$  and  $0 < \theta \leq 2\pi$ . By omitting the origin, which plays a special role, it can be represented as a product  $\mathbb{R}_0^+ \times \mathbb{S}$ , where  $\mathbb{R}_0^+ = (0, \infty)$  and  $\mathbb{S}$  denotes both the circle group and its representation as  $(0, 2\pi]$ .

Assuming that a.s. there are no points at the origin (and, we hasten to add, there is little difficulty in incorporating the contribution of an atom at

the origin if so desired), we have a process with the same kind of structure as a stationary MPP in time. The distance from the origin constitutes the mark and the angular distance  $\theta$  from a fixed reference axis corresponds to the time coordinate. The factorization Lemma A2.7.II applies in a similar fashion and leads to the following representation of the first and second moment measures, analogous to Proposition 8.3.II; the proof is left to Exercise 15.2.1.

**Proposition 15.2.I.** *Let  $N(\cdot)$  be a point process in the plane  $\mathbb{R}^2$ , invariant under rotations about the origin with  $N(\{0\}) = 0$  a.s., and having boundedly finite first and second moment measures. Then the first and second factorial moment measures have the respective factorizations*

$$M_1(dr \times d\theta) = K_1(dr) d\theta/2\pi,$$

$$M_{[2]}(dr_1 \times dr_2 \times d\theta_1 \times d\theta_2) = \check{M}_{[2]}(dr_1 \times dr_2 \times d\phi) d\theta_1/2\pi,$$

where  $\phi \equiv \theta_2 - \theta_1 \pmod{2\pi}$ ; these factorizations correspond to the integral relations, valid for bounded measurable  $h(\cdot)$  with bounded support, namely,

$$\begin{aligned} \mathbb{E} \int_{\mathbb{R}_+ \times \mathbb{S}} h(r, \theta) N(dr \times d\theta) &= \int_0^{2\pi} \frac{d\theta}{2\pi} \int_0^\infty h(r, \theta) K_1(dr), \\ \int_{(\mathbb{R}_+ \times \mathbb{S})^{(2)}} h(r_1, r_2, \theta_1, \theta_2) M_{[2]}(dr_1 \times dr_2 \times d\theta_1 \times d\theta_2) \\ &= \int_0^{2\pi} \frac{d\theta}{2\pi} \int_{\mathbb{R}_+^{(2)} \times \mathbb{S}} h(r_1, r_2, \theta, \theta + \phi) \check{M}_{[2]}(dr_1 \times dr_2 \times d\phi). \end{aligned}$$

Even without isotropy, for such a ‘centred’ process it is frequently convenient to use polar coordinates and hence to represent the first moment measure (assuming it is boundedly finite) in the form  $M_1(dr \times d\theta)$ . Writing

$$K_1(r) = \mathbb{E}[N(S_r)] = \int_0^r \int_0^{2\pi} M_1(ds \times d\theta)$$

for the expected number of points within a distance  $r$  of the origin, we can then define a *directional rose* as the Radon–Nikodym derivative

$$\Gamma(d\theta \mid r) = M_1(dr \times d\theta)/K_1(dr). \quad (15.2.1)$$

Observe that  $\Gamma$ , in contrast to  $K_1$ , is necessarily a probability distribution. In these terms, isotropy embodies two features: the directional rose is uniform over all angles (and equal to  $1/2\pi$ ), and independent of the radius  $r$ .

When densities exist, we may wish to express  $M_1(\cdot)$  in Cartesian coordinates rather than polar coordinates. The densities in these two representations are related by

$$m(x, y) = r^{-1} k_1(r) \gamma(\theta \mid r),$$

where  $r = \sqrt{x^2 + y^2}$ ,  $\theta = \arctan(y/x)$ ,  $k_1(r) = dK_1(r)/dr$ , and  $\gamma(\theta | r) = d\Gamma(\theta | r)/d\theta$ . In the isotropic case,  $\gamma(\theta | r) = 1/2\pi$ .

For the second-order analysis, we can introduce a factorial moment measure for the counting process on centred spheres  $N(S_r)$  by setting

$$K_{[2]}(dr_1 \times dr_2) = \int_{\mathbb{S}^{(2)}} M_{[2]}(dr_1 \times dr_2 \times d\theta_1 \times d\theta_2).$$

In the isotropic case we can then introduce a second-order directional rose  $\Gamma_2(d\phi | r_1, r_2)$  as the Radon–Nikodym derivative

$$\Gamma_2(d\phi | r_1, r_2) = \check{M}_{[2]}(dr_1 \times dr_2 \times d\phi) / K_{[2]}(dr_1 \times dr_2). \quad (15.2.2)$$

$K_{[2]}(dr_1 \times dr_2)$  represents the expected numbers of pairs of points located with one point at distance  $(r_1, r_1 + dr_1)$  from the origin and the other at distance  $(r_2, r_2 + dr_2)$ , and  $\Gamma_2(d\phi | r_1, r_2)$  gives the conditional probability distribution of the angular separation  $\phi$  between the points. The symmetry properties of the second-order moments again lead to

$$\Gamma_2(d\phi | r_1, r_2) = \Gamma_2(2\pi - d\phi | r_2, r_1),$$

but these quantities are in general different from  $\Gamma_2(d\phi | r_2, r_1)$ .

**EXAMPLE 15.2(a) Isotropic centred Poisson process.** Let  $N(\cdot)$  be a Poisson process in  $\mathbb{R}^2$  whose rate parameter has density  $\mu(x, y) = \mu(r)$  for  $r = \sqrt{x^2 + y^2}$ . Then

$$\begin{aligned} k_1(r) &= 2\pi r \mu(r), \\ k_{[2]}(r_1, r_2) &= 4\pi^2 r_1 r_2 \mu(r_1) \mu(r_2), \\ \gamma(\theta | r) &= 1/2\pi = \gamma_2(\phi | r_1, r_2). \end{aligned}$$

Isotropy here implies that the angular separation (at 0) between pairs of points is uniformly distributed. It is also easy to verify that the counting process on centred spheres is a Poisson process on  $\mathbb{R}_+$  with areal density  $2\pi r \mu(r)$ .  $\square$

**EXAMPLE 15.2(b) Isotropic centred Gauss–Poisson process.** Define a Gauss–Poisson process, of pairs of points in  $\mathbb{R}^2$ , by supposing that parent points are located around the origin  $O$  according to a Poisson process with density  $\mu(\cdot)$ , as in Example 15.2(a), and that with each such parent point there is associated an offspring or secondary point whose location relative to a parent point on the circle of radius  $r_1$  is governed by the probability distribution with density function  $f(r_2, \phi | r_1)$ , where  $r_2$  is the distance of the secondary point from  $O$  and  $\phi$  its angular separation (at  $O$ ) from the parent point; for isotropy, suppose that this angular separation is independent of the parent’s angular coordinate. The overall intensity  $k_1(r)$  at distance  $r$  from  $O$  is then the sum of two components, an intensity  $2\pi r \mu(r)$  of parent points and an intensity of secondary points obtained by averaging over all locations of parent points:

$$k_1(r) = 2\pi r \mu(r) + \int_0^\infty 2\pi s \mu(s) ds \int_0^{2\pi} f(r, \phi | s) d\phi.$$

Similarly, for the second-order radial moment measure density we find

$$k_{[2]}(r_1, r_2) = k_1(r_1)k_1(r_2) + \int_0^{2\pi} [2\pi r_1 \mu(r_1) f(r_2, \phi \mid r_1) + 2\pi r_2 \mu(r_2) f(r_1, 2\pi - \phi \mid r_2)] d\phi.$$

The first-order directional rose is of course uniform, but not so the second-order rose which in general depends on the form of the density function  $f(r_2, \phi \mid r_1)$ . If the factorization  $f(r_2, \phi \mid r_1) = f(r_2 \mid r_1)g(\phi)$  holds, then

$$k_{[2]}(r_1, r_2) = k_1(r_1)k_1(r_2) + 2\pi[r_1\mu(r_1)f(r_2 \mid r_1) + r_2\mu(r_2)f(r_1 \mid r_2)],$$

$$\gamma_2(\phi \mid r_1, r_2) = p(r_1, r_2)/2\pi + q(r_1, r_2)g(\phi),$$

where  $p(r_1, r_2) = k_1(r_1)k_2(r_2)/k_{[2]}(r_1, r_2)$  and  $q(r_1, r_2) = 1 - p(r_1, r_2)$ . Thus, the second-order directional rose is a mixture of two components: it reflects the relative proportions of pairs of points with radii  $r_1, r_2$  coming from independent point pairs and parent-offspring point pairs, respectively.

Note that if any given parent point at distance  $r_1$  from the origin has an offspring point with probability  $p(r_1)$ , then it is enough to change  $f(r_2, \phi \mid r_1)$  into a subprobability density function with

$$\int_{\mathbb{R}_+ \times \mathbb{S}} f(r_2, \phi \mid r_1) dr_2 d\phi = p(r_1). \quad \square$$

We turn now to consider those planar point processes that are both stationary and isotropic, so they are invariant under the group of rigid body motions in  $\mathbb{R}^2$ . By Proposition 15.2.I the first moment measure is a multiple of Lebesgue measure in  $\mathbb{R}^2$  by virtue of stationarity alone, so the effect of isotropy shows up first on the second moment measure, that is, on pairs of points. The only property of a pair of points that is invariant under rigid body motions is the distance between them, so a natural coordinate transformation of  $\mathbb{R}^2 \times \mathbb{R}^2$  to consider is of the form

$$(x_1, y_1, x_2, y_2) \mapsto (x_1, y_1, x_1 + r \cos \theta, y_1 + r \sin \theta),$$

where  $0 \leq r < \infty$  and  $0 < \theta \leq 2\pi$ , with  $r$  the distance between the two points and  $\theta$  the angle between the directed line joining them and a fixed reference axis. Assuming the point process to be simple and considering just the factorial and cumulant measures,  $\{r = 0\}$  has zero probability and so the second factorial cumulant measure can be represented as a measure on the space  $\mathbb{R}^2 \times \mathbb{R}_0^+ \times \mathbb{S}$ , corresponding to the coordinates  $x, y, r, \theta$  just introduced. Stationarity implies invariance with respect to shifts in the first two coordinates and so yields the usual representation in terms of a reduced factorial measure, which we write in the form  $\tilde{M}_{[2]}(dr \times d\theta)$ .

Without yet assuming isotropy, introduce a radial measure  $\check{K}_2(dr)$  and a second-order directional rose  $\Gamma_2(d\theta | r)$  via

$$\begin{aligned}\check{K}_2(dr) &= \int_0^{2\pi} \check{M}_{[2]}(dr \times d\theta), \\ \Gamma_2(d\theta | r) &= \check{M}_{[2]}(dr \times d\theta) / \check{K}_2(dr).\end{aligned}$$

The function

$$\check{K}_2(r) = \int_0^r \check{K}_2(ds)$$

can now be interpreted as the expected number of pairs of points separated by a distance  $r$  or less and for which the first point of the pair lies within a region of unit area. The directional rose  $\Gamma_2(d\theta | r)$  then represents the probability that, given the separation is  $r$ , the directed line joining the first point to the second makes an angle with the fixed reference axis falling in the interval  $(\theta, \theta + d\theta)$ . A more natural interpretation of  $\check{K}_2(\cdot)$  is in terms of the first moment measure of the Palm measure for the process: it equals the product of the mean density  $m$  and the expected number of points in a circle of radius  $r$  about the origin given a point at the origin, as noted around (15.1.7) and at (8.1.22) in the discussion of what is there called Ripley's  $K$ -function.

Consider now the implication of isotropy. A rotation through  $\alpha$  transforms the angle  $\theta$  to a new angle  $\theta'$  that depends in general on  $x, y$  as well as  $\alpha$ . Given any  $\theta$  and  $\theta'$ , we can find  $(x, y)$  and  $\alpha$  such that  $\theta$  is transformed into  $\theta'$ . Rotational invariance implies therefore that  $\Gamma_2(d\theta | r)$  must be invariant under arbitrary shifts  $\theta \mapsto \theta'$  and so reduces to the uniform distribution on  $\mathbb{S}$ . We summarize all this as follows.

**Proposition 15.2.II.** *Let  $\check{M}_{[2]}(\cdot)$  denote the reduced second factorial moment measure of a simple stationary point process in the plane. Then  $\check{M}_{[2]}(\cdot)$  can be expressed as*

$$\check{M}_{[2]}(dr \times d\theta) = \check{K}_2(dr) \Gamma_2(d\theta | r), \quad (15.2.3)$$

corresponding to the integral representation of the second factorial moment measure  $M_{[2]}(\cdot)$  (for bounded measurable  $h$  with bounded support)

$$\begin{aligned}& \int_{\mathbb{R}^2 \times \mathbb{R}^2} h(x_1, y_1, x_2, y_2) M_{[2]}(dx_1 \times dy_1 \times dx_2 \times dy_2) \\ &= \int_{\mathbb{R}^2} dx \, dy \int_{\mathbb{R}_+ \times \mathbb{S}} h(x, y, x + r \cos \theta, y + r \sin \theta) \check{K}_2(dr) \Gamma_2(d\theta | r),\end{aligned}$$

where  $\check{K}_2(\cdot)$  is a boundedly finite measure on  $\mathbb{R}_+$  and for each  $r > 0$ ,  $\Gamma_2(\cdot | r)$  is a probability measure on  $\mathbb{S}$ . If the process is also isotropic then

$$\Gamma_2(d\theta | r) = d\theta / 2\pi \quad (\text{all } r, \theta).$$

The symmetry properties of  $\check{M}_{[2]}(\cdot)$  here in the stationary case imply that  $\Gamma_2(d\theta | r) = \Gamma_2(\pi + d\theta | r)$ , which is the analogue for the representation being used here of the property noted for  $\Gamma_2(\cdot | r_1, r_2)$  below (15.2.2).

## Exercises and Complements to Section 15.2

- 15.2.1 Identify  $\mathbb{R}^2 \setminus \{0\}$  with  $\mathbb{R}_0^+ \times \mathbb{S}$  and consider mappings that lead to invariance of these component factors. Now apply the factorization Lemma A2.7.II and hence complete the proof of Proposition 15.2.I.  
 [Hint: In part (a) of the proposition identify  $\mathbb{R}^2 \setminus \{0\}$  with the product  $\mathbb{R}_0^+ \times \mathbb{S}$  and consider invariance under the actions of  $\mathbb{S}$ . For the second moment in part (b) use a diagonal factorization of the components of  $\mathbb{S} \times \mathbb{S}$ .]
- 15.2.2 Let  $N$  be a point process on state space the surface of a cone with semiangle  $\alpha$  (the extremes  $\alpha = 0$  and  $\frac{1}{2}\pi$  correspond to a cylinder and a disc, respectively).  
 (a) Use  $\mathbb{R}_0^+ \times \mathbb{S}$  to describe the points by the distance from the apex of the cone and the angle relative to a fixed plane through the axis of the cone subtended by a plane through the point and the axis (i.e., the ‘longitude’ of a point on the cone). Describe the first and second moment structure of a process invariant under rotations of the cone.  
 (b) For an alternative parameterization, cut the cone down a straight line from the apex and ‘unwrap’ it onto a plane, so that it fills the plane apart from a sector of angle  $2\pi - \theta$ , where  $\theta = 2\pi \sin \alpha$ . Rotations of the cone correspond to rotations in the plane modulo  $\theta$ , where the two edges of the missing sector are identified. Rephrase the results in (a) in terms of this parameterization.  
 [See Byth (1981) who uses the term  $\theta$ -stationary process.]
- 15.2.3 Exercise 8.1.7 gives some results for the isotropic case of a bivariate Neyman–Scott cluster process. Using the notation from there but now for the non-isotropic case, the directional rose has a density  $\gamma_2(\theta | r)$  proportional to

$$2\pi m_1 + \frac{m_{[2]} \exp[-r^2 g(\theta, \Sigma)/4(1 - \rho^2)]}{4\pi \sigma_1 \sigma_2 (1 - \rho^2)^{1/2}},$$

where

$$g(\theta, \Sigma) = \frac{\cos^2 \theta}{\sigma_1^2} - \frac{2\rho \cos \theta \sin \theta}{\sigma_1 \sigma_2} + \frac{\sin^2 \theta}{\sigma_2^2},$$

that is,  $\gamma_2(\theta | r) = p(r)/2\pi + q(r) \exp\{\cdot\cdot\}$ , where  $p(r)$ ,  $q(r)$  are nonnegative functions, involving a Bessel function arising from the normalizing condition  $\int_0^\infty \gamma_2(\theta | r) d\theta = 1$ , and  $p(r) \rightarrow 1$  as  $r \rightarrow \infty$ .

## 15.3. Stationary Line Processes in the Plane

Stationary line processes constitute the paradigm for many of the recent developments in stochastic geometry. In particular, Davidson’s conjecture that all stationary isotropic line processes are doubly stochastic and his imaginative early investigations of this question inspired important further studies by Krickeberg, Papangelou, Kallenberg, and others, and these in turn laid the foundation for recent work on the relations between conditional intensities and Gibbs potentials in the models of statistical physics. Here we give a brief introduction to the properties of line processes, based largely on the early



sections of Harding and Kendall (1974) and the work of Roger Miles. The same circle of ideas is introduced in Stoyan and Mecke (1983, Chapter 7), and more extensively in SKM (1995, Chapters 8–9).

It is convenient to characterize a directed line in the plane by its coordinates  $(p, \theta)$ , where  $\theta$  satisfying  $0 < \theta \leq 2\pi$  is the angle between the line and a fixed reference direction, and  $p$  is the signed perpendicular distance from the line to a fixed origin, being positive when, in looking in the direction of the line, the origin is to its left. We define a *random process of directed lines* to be a point process on the cylinder  $\mathbb{R} \times \mathbb{S}$ , each (random) point on the cylinder being identified with the (random) line in  $\mathbb{R}^2$  via its specification  $(p, \theta)$ . Thus, by the distribution of a stochastic line process, we mean a probability measure on the point process in  $\mathbb{R} \times \mathbb{S}$ . In this text we assume this point process (and hence the line process) to be simple.

For example, by a Poisson process of lines in the plane at rate  $\lambda$ , we mean a Poisson process on  $\mathbb{R} \times \mathbb{S}$  at rate  $\lambda/2\pi$ ,  $\{(p_i, \theta_i)\}$  say, representing directed lines whose directions are i.i.d. uniformly over  $(0, 2\pi]$  and whose signed perpendicular distances from a fixed origin form a Poisson process on  $\mathbb{R}$  at rate  $\lambda$  [see also Exercise 15.3.1(a)]. Note that, given a planar Poisson process at rate  $\lambda$  and locating through each point a line with direction uniformly distributed over  $(0, 2\pi)$ , we obtain infinitely many lines intersecting any unit interval with probability one (cf. Exercise 15.3.2).

A process of undirected lines can be treated as a point process on either  $\mathbb{R}_+ \times \mathbb{S}$  or  $\mathbb{R} \times (0, \pi]$ : the latter fits more easily into our discussion and follows, for example, Stoyan and Mecke (1983) [see also Exercise 15.3.1(b)].

Another representation of a line process is as a point process in  $\mathbb{R}^2$  in which the first coordinate  $x$  say, denotes the intercept by the line on a fixed reference line, and the second equals  $x \cot \theta$ , its intercept on a line orthogonal to the reference line, where  $\theta \in (0, \pi]$  is the angle that the line makes with the reference line [in terms of  $(p, \theta)$  these two intercepts are  $(p \sec \theta, p \operatorname{cosec} \theta)$ ]. Clearly, when  $p = 0$  the direction  $\theta$  is not determined by these two intercepts. However, either representation can be used to describe stationary processes (see below) for which the event  $p = 0$  has probability zero.

It will be evident that these two representations of a given line process lead to different distributions on different spaces.

As in Section 15.2 the principal questions we study relate to the effects of stationarity or isotropy on the moment structure of the process. By these conditions we mean of course invariance of the process of lines under translations and rotations in the plane, so our first task is to examine the effect of these motions on the coordinates  $(p, \theta)$  of a line. Rotation through an angle  $\alpha$  about the fixed origin corresponds to rotation of the cylinder through the same angle,  $(p, \theta) \mapsto (p, \theta + \alpha)$ . Translation of the plane a distance  $d$  in a direction making an angle  $\phi$  with the fixed reference axis induces the transformation

$$(p, \theta) \mapsto (p - d \sin(\theta - \phi), \theta) \quad (-\infty < d < \infty, 0 < \phi \leq \pi) \quad (15.3.1)$$

on the cylinder, corresponding to a shear whereby points on the cylinder are

displaced parallel to its axis through a distance varying (when  $d = 1$ ) from  $-1$  at  $\theta = \phi + \frac{1}{2}\pi$  through  $0$  at  $\theta - \phi = 0 \bmod \pi$  to  $+1$  at  $\theta = \phi - \frac{1}{2}\pi$  (here, addition of angles is taken modulo  $2\pi$ ).

We start by showing that any Borel measure on the cylinder that is invariant under the action of the shears (but not necessarily under rotations) has the product form  $\ell(dp) G(d\theta)$ , where  $\ell(\cdot)$  denotes Lebesgue measure on  $\mathbb{R}$ . This is the first occasion where we encounter an invariance result that cannot be handled via the factorization Lemma A2.7.II: this is so because shears do not generate translations of the cylinder along its axis. Nevertheless, the result we require is still a simple corollary of the more general theorems about the decomposition of invariant measures, which can be established via the general theory of the disintegration of measures, as for example in Krickeberg (1974b, Theorem 2). For the sake of completeness we sketch a simplified version of the theorem as it applies here.

**Lemma 15.3.I.** *Let  $M(dp \times d\theta)$  be a boundedly finite Borel measure on the cylinder  $\mathbb{R} \times \mathbb{S}$ , and let  $M(\cdot)$  be invariant with respect to the action of the shears at (15.3.1). With  $\ell(\cdot)$  denoting Lebesgue measure on  $\mathbb{R}$ , there then exists a totally finite Borel measure  $G$  on  $\mathbb{S}$  such that*

$$M(dp \times d\theta) = \ell(dp) G(d\theta). \quad (15.3.2)$$

PROOF. In outline, we find a factorization of  $M$  of the form  $K(dp | \theta) G_1(d\theta)$ , and then show that  $K(dp | \theta)$  factorizes as  $\lambda(\theta) \ell(dp)$ . To start with, there exists a function  $f(p)$ , as, for example,

$$f(p) = e^{-|n|} / \max(1, M_n) \quad (n < p \leq n+1, n = 0, \pm 1, \dots),$$

where  $M_n = \int_{(n, n+1] \times \mathbb{S}} M(dp \times d\theta)$ , such that  $f(p) > 0$  for all  $p$  and  $\int_{\mathbb{R} \times \mathbb{S}} f(p) M(dp \times d\theta) < \infty$ . Introduce the measure

$$G_1(d\theta) = \int_{p \in \mathbb{R}} f(p) M(dp \times d\theta) \quad (\theta \in \mathbb{S}).$$

For all Borel sets  $A$ ,  $M(A \times d\theta) \ll G_1(d\theta)$ , so by appealing to the usual arguments leading to the existence of regular conditional probabilities, we deduce the existence of a kernel  $K(dp | \theta)$  such that  $K(A | \theta)$  is a measurable function of  $\theta$  for each bounded Borel set  $A$ ,  $K(\cdot | \theta)$  is a Borel measure on  $\mathbb{R}$  for  $G_1$ -almost all  $\theta$ , and  $M(dp \times d\theta) = K(dp | \theta) G_1(d\theta)$ .

Invariance under the action of a given shear (15.3.1), with parameters  $(d, \phi)$  say, implies that for any bounded measurable  $h(\cdot)$  of bounded support,

$$\begin{aligned} \int_{\mathbb{R} \times \mathbb{S}} h(p, \theta) M(dp \times d\theta) &= \int_{\mathbb{R} \times \mathbb{S}} h(p, \theta) K(dp | \theta) G_1(d\theta) \\ &= \int_{\mathbb{R} \times \mathbb{S}} h(p + d \cos(\theta - \phi), \theta) K(dp | \theta) G_1(d\theta) \\ &= \int_{\mathbb{R} \times \mathbb{S}} h(p, \theta) K(dp - d \cos(\theta - \phi) | \theta) G_1(d\theta). \end{aligned}$$

Because this is true for all such  $h$ , the measure  $K(\cdot \mid \theta)$  must coincide with its shifted version  $K(\cdot - d \cos(\theta - \phi) \mid \theta)$  for  $G_1$ -almost all  $\theta$ . By choosing two appropriate values of  $d$  and  $\phi$ , we infer that for such  $\theta$ , the measure  $K(\cdot \mid \theta)$  is invariant under the action of two incommensurate shifts. This in turn implies that  $K(\cdot \mid \theta)$  is a multiple of Lebesgue measure,  $\lambda(\theta) \ell(\cdot)$  say (see Exercise 15.3.3 for details). Thus,  $K(A \mid \theta) = \ell(A) \lambda(\theta)$ , where the left-hand side is a measurable function of  $\theta$ , hence there is a measurable version  $\lambda^*(\theta)$  of  $\lambda(\theta)$  such that  $\lambda^*(\theta) = \lambda(\theta)$  for  $G_1$ -almost all  $\theta$ . Setting  $G(d\theta) = \lambda^*(\theta) G_1(d\theta)$  proves the lemma.  $\square$

**Corollary 15.3.II.** *Let a stationary line process have first moment measure  $M$  on  $\mathbb{R} \times \mathbb{S}$ . Then  $M$  factorizes in the form (15.3.2).*

Because the measure  $G$  is totally finite, it can be normalized to give a first-order directional rose  $\Pi(\cdot)$  on  $\mathbb{S}$

$$\Pi(d\theta) = \frac{G(d\theta)}{\int_{\mathbb{S}} G(d\theta)} = \frac{G(d\theta)}{G(\mathbb{S})}.$$

$\Pi(d\theta)$  may be interpreted as the probability that an arbitrary line has orientation  $\theta$ , and the total measure  $m \equiv G(\mathbb{S})$  has an interpretation as the *mean density of the line measure* induced by the process. To explain this idea, observe that for any line  $W$  and any closed bounded convex set  $A \subset \mathbb{R}^2$ , there exists a well-defined length  $\ell(W \cap A)$ . Given any configuration  $\{W_i\}$  of lines on the plane, we can introduce a corresponding line measure

$$Z(A) = \sum_i \ell(W_i \cap A).$$

This set function  $Z(\cdot)$  is clearly countably additive and extends to a measure on arbitrary Borel sets in the plane. Furthermore, if  $W_i$  has coordinates  $(p_i, \theta_i)$  in the cylinder  $\mathbb{R} \times \mathbb{S}$ , then the mapping  $(p_i, \theta_i) \mapsto \ell(W_i \cap A)$  is measurable, so that if the  $\{W_i\}$  constitute a realization of a stochastic line process, each  $Z(A)$  is a random variable. From Proposition 9.1.VIII it follows that  $Z(\cdot)$  is a random measure, which we call the *line measure* associated with the original line process.

**Proposition 15.3.III.** *Let  $Z$  be the line measure associated with a stationary line process  $W$  in  $\mathbb{R}^2$ . Then  $Z$  is a stationary random measure on  $\mathbb{R}^2$ , and if  $W$  has finite first moment measure  $M(\cdot)$ ,  $Z$  has mean density*

$$m = G(\mathbb{S}) = \int_{(0,1] \times \mathbb{S}} M(dp \times d\theta). \quad (15.3.3)$$

PROOF. Writing  $\ell_A(p, \theta) = \ell(W(p, \theta) \cap A)$  for a line with coordinates  $(p, \theta)$  and any bounded Borel set  $A \subset \mathbb{R}^2$ , we can express

$$Z(A) = \int_{\mathbb{R} \times \mathbb{S}} \ell_A(p, \theta) N(dp \times d\theta)$$

in terms of the point process  $N$  on the cylinder. Because  $\ell_A \geq 0$ , we have

$$E[Z(A)] = \int_{\mathbb{R} \times \mathbb{S}} \ell_A(p, \theta) M(dp \times d\theta) = \int_{\mathbb{S}} G(d\theta) \int_{\mathbb{R}} \ell_A(p, \theta) \ell(dp).$$

For fixed  $\theta$ , the integral over  $\mathbb{R}$  is simply the area of  $A$  evaluated as the integral of its cross-sections perpendicular to the direction  $\theta$ . Writing  $\ell_2$  for Lebesgue measure in  $\mathbb{R}^2$ , we have

$$E[Z(A)] = \int_{\mathbb{S}} G(d\theta) \ell_2(A) = m \ell_2(A),$$

establishing (15.3.3). Stationarity of  $Z(\cdot)$  follows from the stationarity of the line process defining  $Z(\cdot)$ .  $\square$

Given a configuration  $\{W_i\}$ , an alternative to the line measure  $Z(A)$  is the number of lines hitting  $A$ . This set function is subadditive but not in general additive over sets, and thus not a measure, although for each convex set  $A$  it defines a random variable whose distribution and moments can be investigated. If, however,  $A$  is itself a line, we obtain the point process on the line formed by its intersections with the lines  $W_i$ , which is a random measure.

**Proposition 15.3.IV.** *Let a stationary line process in  $\mathbb{R}^2$  be given with mean density  $m$  and directional rose  $\Pi(\cdot)$ .*

- (i) *Let  $V$  be a fixed line in  $\mathbb{R}^2$  with coordinates  $(p_V, \alpha)$ , and let  $N_V(\cdot)$  be the point process on  $V$  generated by its intersections with the line process. Then  $N_V$  is a stationary point process on  $V$  with mean density  $m_V$  given by*

$$m_V = m \int_{\mathbb{S}} |\cos(\theta - \alpha)| \Pi(d\theta). \quad (15.3.4)$$

*If the line process is isotropic,  $m_V$  is independent of  $V$  with*

$$m_V = 2m/\pi \quad (\text{all } V). \quad (15.3.4')$$

- (ii) *Let  $A$  be a closed bounded convex set in  $\mathbb{R}^2$ , and let  $Y(A)$  be the number of distinct lines of the line process intersecting  $A$ . If the line process is isotropic then*

$$E[Y(A)] = mL(A)/\pi, \quad (15.3.5)$$

*where  $L(A)$  is the length of the perimeter of  $A$ .*

PROOF. For any bounded measurable function  $h$  of bounded support in  $\mathbb{R}$ , we have

$$\int_{\mathbb{R}} h(x) N_V(dx) = \int_{\mathbb{R} \times \mathbb{S}} h(x(p, \theta)) N(dp \times d\theta),$$

where  $x = x(p, \theta)$  denotes the distance from a fixed origin on  $V$  to a point of intersection of a line with coordinates  $(p, \theta)$ , and  $N(\cdot)$  refers to the point

process on the cylinder representing the given line process. Because of stationarity, there is no loss of generality in taking the fixed origin on  $V$  as the origin for the cylindrical  $(p, \theta)$  coordinates. Then  $x(p, \theta) = p \sec(\theta - \alpha)$ , and on taking expectations, we obtain

$$\mathbb{E} \int_{\mathbb{R}} h(x) N_V(dx) = \int_{\mathbb{R} \times \mathbb{S}} h(p \sec(\theta - \alpha)) \ell(dp) m \Pi(d\theta). \quad (15.3.6)$$

Substituting  $u = p \sec(\theta - \alpha)$ ,  $\ell(dp) = |\cos(\theta - \alpha)| \ell(du)$ , from which (15.3.4) follows because  $N_V$ , being stationary, has  $\mathbb{E} N_V(dx) = m_V \ell(dx)$ .

In the isotropic case,  $\Pi(d\theta) = d\theta/2\pi$  and integration at (15.3.4) leads to  $2m/\pi$  as asserted.

To prove (ii), suppose first that  $A$  is a convex polygon. Apply the result of (i) to each side of the polygon in turn, so that adding over all sides shows that the expected number of intersections of lines from the line process with the perimeter of  $A$  equals  $2mL(A)/\pi$ . Convexity implies that each line intersecting the polygon does so exactly twice (except possibly for a set of lines of zero probability), so the factor 2 cancels and (15.3.5) is established for convex polygons. A limiting argument extends the result to any closed bounded convex set  $A$ .  $\square$

Propositions 15.3.III and 15.3.IV can be extended to processes of random hyperplanes and more generally random ‘flats’ in  $\mathbb{R}^d$ : see Exercises 15.3.4–5 for some preliminary results and the extensive series of papers by Miles (1969, 1971, 1974). Krickeberg (1974b) sets out a general form of the required theory of moment measures.

When further distributional aspects are specified, the results can be sharpened as in the basic example below. For extensions to higher dimensions see Exercise 15.3.6 and the cited papers by Miles.

EXAMPLE 15.3(a). We define a *stationary Poisson process of lines* in terms of the associated point process  $N(\cdot)$  on the cylinder  $\mathbb{R} \times \mathbb{S}$ . For the line process to be stationary in  $\mathbb{R}^2$ , the point process  $N(\cdot)$  must be invariant under shears of the cylinder, and its first moment measure must decompose as at (15.3.2). But the first moment measure of a Poisson process coincides with its parameter measure  $\mu(\cdot)$ , so

$$\mu(dp \times d\theta) = \mu \ell(dp) \Pi(d\theta) \quad (15.3.7)$$

must hold for some density  $\mu$  and probability distribution (here, the directional rose of the line process)  $\Pi(\cdot)$ . We can therefore write for the p.g.fl. of the point process  $N(\cdot)$ , with suitable functions  $h(\cdot)$ ,

$$\begin{aligned} G[h] &= \mathbb{E} \left[ \exp \left\{ \int_{\mathbb{R} \times \mathbb{S}} \log h(p, \theta) N(dp \times d\theta) \right\} \right] \\ &= \exp \left\{ \mu \int_{\mathbb{R}} \ell(dp) \int_{\mathbb{S}} (h(p, \theta) - 1) \Pi(d\theta) \right\}. \end{aligned}$$

Thus, for example, the second factorial moment measure is given by

$$\begin{aligned} M_{[2]}(dp \times dp' \times d\theta \times d\theta') &= \mu^2 \ell(dp) \ell(dp') \Pi(d\theta) \Pi(d\theta') \\ &= M(dp \times d\theta) M(dp' \times d\theta'), \end{aligned}$$

which is of course of the product form expected for a Poisson process.

The p.g.fl. for the point process  $N_V$  of intersections of the line process on a fixed line  $V$  follows from an extension of the reasoning leading to (15.3.6). In the notation used there,

$$\begin{aligned} G_{N_V}[h] &= \mathbb{E} \left[ \exp \left\{ \int_{\mathbb{R}} \log h(x) N_V(dx) \right\} \right] \\ &= \mathbb{E} \left[ \exp \left\{ \int_{\mathbb{R} \times \mathbb{S}} \log h(p \sec(\theta - \alpha)) N(dp \times d\theta) \right\} \right] \\ &= \exp \left\{ \mu \int_{\mathbb{R} \times \mathbb{S}} [h(p \sec(\theta - \alpha)) - 1] \ell(dp) \Pi(d\theta) \right\} \\ &= \exp \left\{ \mu \int_{\mathbb{S}} |\cos(\theta - \alpha)| \Pi(d\theta) \int_{\mathbb{R}} (h(u) - 1) du \right\}, \end{aligned}$$

which we recognize as the p.g.fl. of a stationary Poisson process on  $V$  with density  $m_V$  as at (15.3.4). In particular, for a stationary isotropic Poisson line process, the density  $m_V = 2\mu/\pi$  is independent of the orientation  $\alpha$  of  $V$ , and the number of lines crossing a closed convex set  $A$  has a Poisson distribution with parameter  $\mu L(A)/\pi$  with  $\mu$  equal to the mean line density.  $\square$

We now discuss second-order properties of line processes, confining our attention to stationary isotropic processes. Now it is clear that one invariant of a pair of lines, under both translations and rotations on the plane, is the angle  $\phi$  between them where  $0 < \phi \leq 2\pi$  (this allows for directed lines), so we take coordinates in the form

$$(p, \theta, p', \theta') \mapsto (p, \theta, p', \theta + \phi).$$

Invariance under rotations then implies that the second factorial moment measure  $M_{[2]}$  of the point process  $N$  representing the stationary isotropic line process in  $\mathbb{R}^2$  factorizes into a product

$$M_{[2]}(dp \times dp' \times d\theta \times d\theta') = \check{M}_{[2]}(dp \times dp' \times d\phi) d\theta/2\pi.$$

To handle the term  $\check{M}_{[2]}$  we proceed much as in Proposition 15.2.II to deduce that

$$\check{M}_{[2]}(dp \times dp' \times d\phi) = K_{[2]}(dp \times dp' \mid \phi) G_{[2]}(d\phi).$$

Invariance of  $N$  under shears implies that for almost all  $\phi$  and at least for (say) rational  $r$  and  $\psi = \theta - \alpha$ ,

$$K_{[2]}(dp \times dp' \mid \phi) = K_{[2]}((dp - r \cos \psi) \times (dp' - r \cos(\psi + \phi)) \mid \phi).$$

Provided  $\phi \neq 0$  or  $\pi$ , the equations  $r \cos \psi = u$  and  $r \cos(\psi + \phi) = v$  have unique solutions  $(r, \psi)$  for all real  $u$  and  $v$ , and therefore  $K_{[2]}(dp \times dp' \mid \phi)$  is invariant under at least a countable dense family of translations  $(p, p') \mapsto (p + u, p' + v)$ , including incommensurate pairs  $(u, v), (u', v')$ . For such values of  $\phi$  then,  $K_{[2]}(dp \times dp' \mid \phi)$  reduces to a multiple of Lebesgue measure  $\ell_2$  in  $\mathbb{R}^2$ ,  $\lambda(\phi) \ell_2(dp \times dp')$  say, where as earlier we can take  $\lambda(\phi)$  to be a measurable function of  $\phi$ .

The exceptional cases  $\phi = 0$  and  $\pi$  correspond to the occurrence of pairs of parallel and antiparallel lines, respectively. In both cases, the signed distance  $y$  between the lines is a further invariant of the motion. In the first case, invariance under translation implies that

$$K_{[2]}(dp \times dp' \mid 0) = K_{[2]}((dp + r \cos(\theta - \alpha)) \times (dp' + r \cos(\theta - \alpha)) \mid 0),$$

so that setting  $p' = p + y$ , the measure factorizes into the form

$$K_{[2]}(dp \times dp' \mid 0) = \tilde{K}_+(dy) \ell(dp)$$

for some boundedly finite measure  $\tilde{K}_+$  on  $\mathbb{R}$ . Similarly, for  $\phi = \pi$ , the measure  $K_{[2]}(dp \times dp' \mid \pi)$  is invariant under the transformations

$$(p, p') \mapsto (p + r \cos(\theta - \alpha), p' - r \cos(\theta - \alpha))$$

so that setting now  $p' = y - p$ , where  $y$  is the distance between parallel lines oriented in opposite senses,

$$K_{[2]}(dp \times dp' \mid \pi) = \tilde{K}_-(dy) \ell(dp)$$

for boundedly finite  $\tilde{K}_-$ . Finally, because  $M_{[2]}$  is symmetric under the transformation

$$(p, p', \theta, \theta') \mapsto (p', p, \theta', \theta),$$

all three of  $G_{[2]}$ ,  $\tilde{K}_+$  and  $\tilde{K}_-$  are symmetric under reflection in their respective origins. We have proved the following result.

**Proposition 15.3.V** [Davidson (1974a), Krickeberg (1974a, b)]. *Let  $M_{[2]}$  be the second factorial moment measure of a stationary isotropic line process in  $\mathbb{R}^2$ . Then  $M_{[2]}$  admits a representation in terms of the factors:*

- (i) *a totally finite symmetric measure  $G_{[2]}(d\phi)$  on  $(0, \pi) \cup (\pi, 2\pi)$  governing the intensity of pairs of lines intersecting at an angle  $\phi$ ;*
- (ii) *a boundedly finite symmetric measure  $\tilde{K}_+(dy)$  on  $\mathbb{R}$  governing the intensity of pairs of parallel lines distance  $y$  apart; and*
- (iii) *a similar measure  $\tilde{K}_-(dy)$  governing the intensity of pairs of antiparallel lines a distance  $y$  apart.*

The representation is realized by the integral relation, valid for bounded measurable nonnegative functions  $h$  of bounded support on  $\mathbb{R} \times \mathbb{R} \times \mathbb{S} \times \mathbb{S} = \mathbb{R}^2 \times \mathbb{S}^2$ ,

$$\begin{aligned} & \int_{\mathbb{R}^2 \times \mathbb{S}^2} h(p, p', \theta, \theta') M_{[2]}(dp \times dp' \times d\theta \times d\theta') \\ &= \int_{(0, \pi) \cup (\pi, 2\pi)} G_{[2]}(d\phi) \int_{\mathbb{R}^2 \times \mathbb{S}} h(p, p', \theta, \theta + \phi) dp dp' d\theta \\ & \quad + \int_{\mathbb{R}} \tilde{K}_+(dy) \int_{\mathbb{R} \times \mathbb{S}} h(p, p + y, \theta, \theta) dp d\theta \\ & \quad + \int_{\mathbb{R}} \tilde{K}_-(dy) \int_{\mathbb{R} \times \mathbb{S}} h(p, y - p, \theta, \theta + \pi) dp d\theta. \end{aligned} \quad (15.3.8)$$

A similar representation holds for the factorial covariance measure.

As Rollo Davidson (1974a) showed, many remarkable corollaries follow from the representation (15.3.8).

**Corollary 15.3.VI.** *With probability 1, a stationary isotropic line process in  $\mathbb{R}^2$  either has no pairs of parallel or antiparallel lines, or has infinitely many pairs of parallel lines, or has infinitely many pairs of antiparallel lines, according as  $\tilde{K}_+(\mathbb{R}) = \tilde{K}_-(\mathbb{R}) = 0$ , or  $\tilde{K}_+(\mathbb{R}) > 0$ , or  $\tilde{K}_-(\mathbb{R}) > 0$ , respectively.*

PROOF. Let  $A$  be a bounded Borel set in  $\mathbb{R}$  and  $V$  a fixed line in the plane. Then by the preceding discussion and (15.3.4') we can interpret  $2\tilde{K}_+(A)/\pi$  as the mean density of the stationary point process on  $V$  generated by its intersections with those lines of the process which have other lines of the process parallel to them and at separation  $y \in A$ . If  $\tilde{K}_+(\mathbb{R}) = 0$ , any such process of interaction has zero mean density and is therefore a.s. empty. Letting  $A \uparrow \mathbb{R}$ , we deduce that with probability 1 no line of the process has another line of the process parallel to it. Conversely, if  $\tilde{K}_+(\mathbb{R}) > 0$ , we can find a bounded Borel set  $A$  with  $\tilde{K}_+(A) > 0$  and a line  $V$  such that the process of associated points on  $V$  is stationary with positive mean density and therefore has an infinite number of points (see Proposition 12.1.VI).

The argument concerning antiparallel lines is similar.  $\square$

**Corollary 15.3.VII.**  *$M_{[2]}$  is invariant under reflections if and only if the process has a.s. no pairs of antiparallel lines, in which case it is also invariant under translations  $p \mapsto p + y$  of the cylinder parallel to its axis.*

PROOF. If  $\tilde{K}_-(\mathbb{R}) = 0$ , it follows from (15.3.8) that  $M_{[2]}$  is invariant under the transformation

$$(p, p', \theta, \theta') \mapsto (p, p', -\theta, -\theta')$$

on account of the symmetry properties of  $G_{[2]}$ . This mapping corresponds to reflection in the reference axis. Hence, by isotropy and stationarity, it



is invariant under any other reflection. A similar conclusion holds for the transformation

$$(p, p', \theta, \theta') \mapsto (p + y, p' + y, \theta, \theta').$$

Conversely, if  $\tilde{K}_-(\mathbb{R}) > 0$ , we can choose  $h(\cdot)$  in (15.3.8) so that a contradiction arises if we assume that  $M_{[2]}$  is invariant under either of these transformations.  $\square$

Provided that  $T^{-1}\tilde{K}_+((0, T])$  and  $T^{-1}\tilde{K}_-((0, T])$  both vanish in the limit as  $T \rightarrow \infty$ , we can show that the measure  $G_{[2]}$  of Proposition 15.3.V is positive definite: we proceed under the more restrictive assumption that  $\tilde{K}_+(\mathbb{R}) = \tilde{K}_-(\mathbb{R}) = 0$ , that is, that there are a.s. no parallel or antiparallel pairs of lines.

Let  $a(\theta)$  be a bounded measurable function on  $(0, 2\pi)$  and in (15.3.8) set

$$h(p, p', \theta, \theta') = I_{(0, T]}(p) I_{(0, T]}(p') a(\theta) a(\theta') = h(p, \theta) h(p', \theta').$$

In place of  $M_{[2]}$ , consider the ordinary second moment measure  $M_2$  for which  $M_2(A \times B) = M_{[2]}(A \times B) + M_1(A \cap B)$ . We have

$$\begin{aligned} 0 &\leq \mathbb{E} \left[ \left( \int_{\mathbb{R} \times \mathbb{S}} h(p, \theta) N(dp, d\theta) \right)^2 \right] \\ &= \int_{(\mathbb{R} \times \mathbb{S})^2} h(p, \theta) h(p', \theta') M_{[2]}(dp \times dp' \times d\theta \times d\theta') \\ &\quad + \int_{\mathbb{R} \times \mathbb{S}} h^2(p, \theta) M_1(dp \times d\theta) \\ &= T^2 \int_{\mathbb{S}} \int_{\mathbb{S}} a(\theta) a(\theta + \phi) G_{[2]}(d\phi) d\theta + \int_{\mathbb{S}} \frac{a^2(\theta) m T d\theta}{2\pi}. \end{aligned}$$

Division by  $T^2$  and rearrangement yield

$$\int_{\mathbb{S}} \int_{\mathbb{S}} a(\theta) a(\theta + \phi) G_{[2]}(d\phi) d\theta \geq - \int_{\mathbb{S}} \frac{a^2(\theta) m d\theta}{2\pi T} \rightarrow 0 \quad (T \rightarrow \infty);$$

that is, the measure  $G_{[2]}(\cdot)$  is positive definite (equivalently, in the terminology of Definition 8.6.I, it is a p.p.d. measure).

This result immediately suggests asking whether  $G_{[2]}(\cdot)$  can be interpreted as a covariance measure; accordingly, we look for some random measure  $Y$  on  $\mathbb{S}$  with which  $G_{[2]}(\cdot)$  may be associated. The appropriate candidate (as we now show) is the ergodic limit of the original point process on  $\mathbb{R} \times \mathbb{S}$  with respect to translations of the cylinder parallel to its axis, or, equivalently, the conditional expectation  $\mathbb{E}(N(\cdot) \mid \mathcal{I}_{\mathbb{S}})$  of the original process with respect to the invariant  $\sigma$ -algebra  $\mathcal{I}_{\mathbb{S}}$  generated by these translations. That such ergodic

limits exist follows directly from Theorem 12.2.IV, and we then have, for  $A, B \in \mathcal{B}(\mathbb{S})$ ,

$$\begin{aligned} E(Y(A)Y(B)) &= \lim_{T \rightarrow \infty} \frac{1}{T^2} \int_{(\mathbb{R} \times \mathbb{S})^2} M_{[2]}(dp \times dp' \times d\theta \times d\theta') \\ &\quad \times I_{(0,T]}(p)I_{(0,T]}(p')I_A(\theta)I_B(\theta') dp dp' d\theta d\theta' \\ &= \int_{\mathbb{S}^2} I_A(\theta)I_B(\theta + \phi) G_{[2]}(d\phi) d\theta. \end{aligned}$$

An even more surprising corollary is the following. Consider the Cox process  $N^*$  on the cylinder  $\mathbb{R} \times \mathbb{S}$  directed by the random measure  $\ell \times Y$ . It is readily checked that  $N^*$  is invariant under rotations and translations of the cylinder and that for  $A, B \in \mathcal{B}(\mathbb{S})$  and  $\mathbb{U}$  the unit interval,

$$\begin{aligned} E[N^*(\mathbb{U} \times A)] &= m\ell(A)/2\pi = EY(A), \\ M_{[2]}^*(\mathbb{U} \times \mathbb{U} \times A \times B) &= E(Y(A)Y(B)). \end{aligned}$$

Thus,  $N$  and  $N^*$  have the same first and second moment measure. We summarize this as follows.

**Proposition 15.3.VIII.** *If a stationary isotropic line process has boundedly finite second moment measure and has a.s. no parallel or antiparallel lines, then the reduced moment measure  $G_{[2]}(\cdot)$  of Proposition 15.3.V is positive definite and can be identified with the second moment measure of the random measure on  $\mathbb{S}$*

$$Y(A) = E(N(\mathbb{U} \times A) \mid \mathcal{I}) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{p \in (0,T]} N(dp \times A) \quad \text{a.s.,}$$

where  $\mathcal{I}$  is the invariant  $\sigma$ -algebra associated with shifts of the cylinder parallel to its axis. Furthermore,  $N$  has the same second moment measure as the Cox process  $N^*$  directed by  $\ell \times Y$ .

This proposition, coupled with his failure to find counterexamples, led Davidson to formulate his celebrated conjecture [‘the big problem’ of Davidson (1974b, p. 70)] that any stationary isotropic line process with a.s. no parallel or antiparallel lines and boundedly finite second moment measure must be a Cox process. Davidson showed that no counterexample could be constructed by taking a point process on a line and putting lines through its points [cf. Proposition 15.3.IV(i)], nor by taking a stationary point process in  $\mathbb{R}^2$  and putting lines through its points, nor seemingly ‘by tinkering with a Poisson line process.’ The structure of stationary isotropic line processes, as well as of the more general stationary hyperplane processes in spaces  $\mathbb{R}^{2d}$ , differs radically from those of stationary point processes in  $\mathbb{R}^1$ . That Davidson’s conjecture is false was shown by Kallenberg (1977b) in which the main idea is

the construction of a process from a lattice configuration in a parametrization with respect to a fixed line.

To describe this example, it is necessary to adopt the alternative representation of a line as a pair  $(x, y) \in \mathbb{R}^2$ . Here  $x$  is the  $x$ -coordinate of the intercept of the line with the  $x$  axis, and  $y = x \cot \theta$  as described earlier in this section. (Our attempt in the first edition to apply similar arguments to a lattice in the cylinder representation contained a fundamental flaw: see Exercise 15.3.10.)

**EXAMPLE 15.3(b) *Kallenberg's randomized lattice.*** We start from the line system specified in the alternative representation above by the square lattice of points in the plane with integer coordinates. We randomize the location of the lattice by translating it by a vector  $X$  uniformly distributed over the unit square, and rotating it by an angle  $\Phi$  uniformly distributed over the interval  $(0, 2\pi]$ . The resulting point process has unit mean density, and the average density of points in a realization ('sample density') is also a.s. unity. It is still of lattice type, and is invariant under arbitrary rigid motions of the plane.

The crucial requirement, however, is to produce from this randomized lattice a point process that is invariant under shears,  $\Sigma_\alpha: (x, y) \mapsto (x, x + \alpha y)$  say, for these correspond to translations in the space of lines. Because the process is already invariant under rotations, it is enough to consider just the shears  $\Sigma_\alpha$  parallel to the  $y$ -axis. To this end we consider a sequence of further randomizations: first select  $\alpha$  uniformly over the interval  $(-n, n)$ , then let  $n \rightarrow \infty$ .

This yields a sequence of point processes in the plane, which become more and more nearly invariant under shifts as  $n \rightarrow \infty$ . Moreover, each such process is still of lattice type (although no longer a square lattice), is invariant under translations and rotations, has mean density 1, and has mean sample density a.s. equal to 1. Boundedness of the mean densities implies that the sequence of point processes is tight in the topology of weak convergence (see Exercise 11.1.2). Thus, there exists at least one weakly convergent subsequence. However this argument does not preclude the possibility that the resulting limit measure might be null. To eliminate this last possibility, Kallenberg considers the corresponding Palm measures, and shows that these are tight, implying, because a Palm measure necessarily has a point at the origin, that the resultant limit is nonzero.

The resultant line process is not locally bounded, but the line process obtained from considering just the points in a vertical strip will be so and will still be invariant under vertical shears (corresponding to translations in the space of lines) and to vertical shifts (corresponding to rotations). The limit is not a Poisson process because it still has an a.s. lattice character, but from the construction it is invariant under both rotations and shears. Finally, its second factorial moment measure exists, and a further analysis shows that  $\tilde{K}_+(\mathbb{R}) = \tilde{K}_-(\mathbb{R}) = 0$ , and  $G_{[2]}(\cdot)$  is uniformly distributed on  $\mathbb{S}$ . This is the same second moment structure as for the simple Poisson process itself. Because it is not the Poisson process, it cannot be a Cox process either.

The corresponding line process is thus invariant under translations and rotations of the plane, has finite first and second moment measures, has a.s. no parallel or antiparallel lines, and is not a Cox process. It therefore refutes Davidson's conjecture.  $\square$

Details of the proof and further remarkable properties of the process in the plane constructed in this way are given in Kallenberg (1977b) (which includes a further comment by Kingman), and in Mecke's (1979) subsequent paper. In particular, Mecke gives both an explicit construction for the process and an algebraic characterization of it as the unique process having lattice character and invariant under all affine translations of the plane. The singular character is clearly evident from SKM (1995, Figure 8.3) illustrating the process.

### Exercises and Complements to Section 15.3

- 15.3.1 (a) The line process in  $\mathbb{R}^2$  represented as a Poisson process on  $\mathbb{R} \times \mathbb{S}$  is defined initially with respect to a specified origin and reference direction in  $\mathbb{R}^2$ . Show that the line process in fact is homogeneous and isotropic in  $\mathbb{R}^2$ . [Hint: Consider the transformation on  $(p, \theta)$  effected by moving the origin as underlying (15.3.1). The Poisson process on the cylinder is preserved under both this transformation and change of the reference direction.]
- (b) For each of the three point process representations below of a *process of undirected lines*, describe the point process that represents a line process that is (a) homogeneous; or (b) isotropic; or (c) both.
- Take the distance  $p > 0$  from the origin to the line as one parameter and the angle made by the line and a fixed reference axis as the other. The line process is represented as a point process on  $\mathbb{R}_+ \times \mathbb{S}$ .
  - As in (a) except that the distance is signed and the angle is restricted to the range  $(0, \pi]$ . Then the line process is a point process on  $\mathbb{R} \times (0, \pi]$ .
  - Describe a line by its intercepts on the  $x$  and  $y$  axes as parameters. Then a line process is represented as a point process in  $\mathbb{R}^2$ .
- 15.3.2 Suppose given a Poisson process in  $\mathbb{R}^2$  at unit rate; let  $W_\delta$  denote those of its points  $(x, y)$  lying in the wedge  $0 < y/x < \tan \delta$  with  $\delta < \frac{1}{2}\pi$ . Independently through each point construct a line with orientation  $\theta$  uniformly distributed on  $(0, 2\pi)$ . Let  $S_\epsilon$  denote the circle with centre at the origin and radius  $\epsilon > 0$ . Show that  $\Pr\{\text{no line through a point of } W_\delta \text{ intersects } S_\epsilon\} = 0$ . Conclude that with probability one, infinitely many lines intersect  $S_\epsilon$ .
- 15.3.3 Let  $\mu$  be a measure on  $\mathbb{R}$  invariant under shifts  $T_a$  and  $T_b$  for incommensurate  $a$  and  $b$ . Set  $F(x) = \mu((0, x])$  for  $x > 0$ ,  $= -\mu((x, 0])$  for  $x \leq 0$ , and let  $U$  be the set of points  $u \in \mathbb{R}$  such that  $\mu$  is invariant under shifts  $T_u$ . Show that  $u \in U$  implies  $-u \in U$ , and that for  $u, v \in U$ ,  $F(u+v) = F(u) + F(v)$ , so that  $U$  is an additive group and thus contains all points of the form  $ja + kb$  for positive or negative integers  $j, k$ . Deduce that  $F(x) = \alpha x$  for all  $x$  and some  $\alpha \geq 0$ . [This result, like Exercise 12.1.8, is a variant on the Hamel equation at (3.6.3).]
- 15.3.4 *Random hyperplanes.* A hyperplane is a  $(d-d')$ -dimensional linear subspace of  $\mathbb{R}^d$  shifted through some vector  $x \in \mathbb{R}^d$  for some positive integer  $d' < d$ .

- (a) Show that a directed hyperplane is uniquely specified by a pair  $(p, \theta)$ , where  $\theta$  lies on the  $d$ -dimensional unit ball  $S^d$ ,  $p \in \mathbb{R}$ , and the sense of the hyperplane (whether the normal to the origin is directed toward or away from the hyperplane) is determined by the sign of  $p$ .
- (b) A process of random  $d - d'$  hyperplanes can be represented as a point process in  $S^d \times \mathbb{R}$ . Rotation of the original plane corresponds to rotation by an element of  $S^d$ ; translation of the original plane corresponds to the transformation  $(p, \theta) \mapsto (p + \langle x, \theta \rangle, \theta)$ , where  $\langle x, \theta \rangle$  is the inner product in  $\mathbb{R}^d$ .
- (c) Such a process is homogeneous and isotropic if and only if the point process is invariant under both rotations and generalized shears as defined above, and its first moment measure, if it exists, is then a multiple of Lebesgue measure on  $S^d \times \mathbb{R}$ ,  $m\ell_d(\cdot)$  say.
- (d) Define a random measure  $\xi(A)$  for bounded Borel  $A$  in  $\mathbb{R}^d$  as the sum of the hypervolumes  $A \cap S_i$ , where the particular hyperplanes of the process are denoted by  $\{S_i\}$ . Then  $m\ell_d(\cdot)$  is the mean density of this random measure.
- (e) If  $L$  is an arbitrary fixed line in  $\mathbb{R}^d$ , the points of intersection of  $L$  with  $\{S_i\}$  form a stationary point process with mean density  $m\Gamma(\frac{1}{2}d)/2\pi^{1/2}$ .

[Hint. See references preceding Example 15.3(a).]

- 15.3.5 The special case of *random lines* in  $\mathbb{R}^3$  uses the representation of such lines as points in  $S^2 \times \mathbb{R}^2$ , where the component in  $S^2$  determines the direction of the line and the point in  $\mathbb{R}^2$  its point of intersection with the orthogonal plane passing through the origin. Find analogues to (d) and (e) of Exercise 15.3.4 for the ‘line density’ of the process (a random measure in  $\mathbb{R}^3$ ) and the point process generated by the points of intersection of the lines with an arbitrary plane in  $\mathbb{R}^3$ .
- 15.3.6 Extend the result of Example 15.3(a) to the context of Exercises 15.3.4(e) and 15.3.5 (i.e., show that if the original process is Poisson so are the induced processes on the line and plane, respectively).
- 15.3.7 Given a homogeneous isotropic Poisson directed line process in  $\mathbb{R}^2$ , form a ‘clustered’ line process with pairs of lines in each cluster in one of the following three ways.
- (a) *Railway line process* (i): To each line  $(p_i, \theta_i)$  of the process, add the line  $(p_i + d, \theta_i)$  for some fixed positive  $d$ . In the notation of Proposition 15.3.V,  $G_{[2]}$  and  $K_-$  are null, and  $K_+$  has an atom at  $d$ . The process is invariant under translation, rotation, and reflection.
  - (b) *Davidson’s railway line process* (ii): To each line  $(p_i, \theta_i)$  add the antiparallel line  $(-p_i - d, \pi + \theta_i)$ . Then  $G_{[2]}$  and  $K_+$  are null,  $K_-$  has an atom at  $d$ , and because of the built-in handedness, this process is not invariant under reflections of the plane.
  - (c) To each line  $(p_i, \theta_i)$  add the line  $(p_i, \theta_i + \alpha)$  for some fixed  $\alpha$  in  $0 < \alpha < \pi$ . The resulting line process is no longer translation invariant.

[Each process here is a possible analogue of the Poisson process of deterministic cluster pairs as in Bartlett (1963, p. 266) or Daley (1971, Example 5).]

- 15.3.8 (a) Show that two distinct lines  $(p, \theta)$  and  $(p', \theta')$  intersect in a point inside the unit circle if and only if  $|p| < 1$ ,  $|p'| < 1$  and

$$p^2 + p'^2 - 2pp' \cos(\theta - \theta') < \sin^2(\theta - \theta').$$

The expected number of line pairs intersecting within the circle is thus found by integrating the second factorial moment measure over the region defined by these inequalities.

- (b) More generally, the first moment measure  $M$  of the process of intersections is found from integrals of the form

$$\int_{\mathbb{R}^2} h(x, y) M(dx \times dy) = \int_{\mathbb{R}^2 \times \mathbb{S}^2} h(x(\mathbf{p}), y(\mathbf{p})) M_{[2]}(dp \times dp' \times d\theta \times d\theta'),$$

where  $h(\cdot)$  is a bounded measurable function of bounded support, and  $\mathbf{p}$  denotes the vector of coordinates  $p, p', \theta, \theta'$  and the two lines  $(p, \theta)$  and  $(p', \theta')$  intersect in the point  $(x(\mathbf{p}), y(\mathbf{p}))$ .

- (c) When the line process is homogeneous and isotropic,  $M_{[2]}$  reduces to the form described in Proposition 15.3.V. Assuming there are a.s. no parallel or antiparallel lines, use the representation of (b) to show that the intersection process is stationary and has mean density given by

$$4\pi \int_{(0, \pi)} \sin \phi G_{[2]}(d\phi).$$

- 15.3.9 Show that if a stationary isotropic line process has a.s. no parallel or antiparallel lines, then it cannot be a Poisson cluster process. [*Hint*: Consider the form of the second factorial moment measure; a Poisson cluster process with nontrivial cluster distribution cannot factorize in the form of Proposition 15.3.V.]

- 15.3.10 (a) Consider a line process represented by a lattice on the cylinder  $\mathbb{R} \times \mathbb{S}$ . Show that its properties are quite different from those of a line process represented by a lattice in  $\mathbb{R}^2$  using the alternative representation using the intercepts  $(x, y) = (p \sec \theta, p \operatorname{cosec} \theta)$ .
- (b) Let  $\{(p_i, \theta_i)\}$  denote a stationary line process in  $\mathbb{R}^2$ . Investigate whether the point process in  $\mathbb{R}$  with realizations  $\{p_i\}$  is stationary [i.e., invariant under all shears at (15.3.1)]. [The question is due to Dietrich Stoyan.]

## 15.4. Space–Time Processes

Space–time models combine elements from the evolutionary processes studied in Chapter 14 and from the descriptive properties of spatial patterns covered earlier in this chapter. Because the spatial location can always be considered as one component of a multi-dimensional mark, some aspects, such as the likelihood theory based on conditional intensity functions, are essentially special cases of the more general discussion of Chapter 14. For applications, on the other hand, the evolution of spatial features with time is often of special

interest. In this section we review some basic features of space–time point processes, trying to select those that most warrant more careful examination.

The earliest statistical models for space–time processes of which we are aware were prompted by an agricultural setting. As agricultural trials continued on experimental stations, fluctuations in soil fertility were studied, and it was observed that the spatial correlations decayed remarkably slowly. Pioneer studies by Whittle (1954, 1962), using diffusion methods, showed that such long-term correlations could be caused by a sequence of perturbations (applications of fertilizer or other treatments) followed by gradual diffusion. Whittle also observed that space–time models are likely to be more insightful, by penetrating farther into the physical processes generating a spatial point pattern, than a purely static model for that pattern.

Despite such considerations, studies of space–time models have lagged well behind those of simple temporal models, and even those of purely spatial models. No doubt the reasons have been largely practical, notably the difficulty of compiling good space–time datasets and the heavy computations needed to analyze them. Their importance, however, can only grow as time goes on and these difficulties are overcome.

Another point to bear in mind about space–time models is their diversity. The models on which we focus here—models for earthquakes form a paradigm example—are for events which can be regarded as points in both time and space dimensions. With earthquakes and forest fires, a point pattern can be obtained only by accumulating events over time (e.g., the fires which have occurred over the last year). Models for particles moving through space constitute a different class. Although they can be viewed as spatial point patterns evolving in time, in space–time they form families of trajectories rather than families of points. A similar situation arises for models for storm centres or rainfall cells within a storm; in the discussion in Wheater *et al.* (2000), these phenomena are treated as points that persist a while until they disappear.

A great deal of flexibility is added by moving from simple space–time point processes to space–time point processes with an associated mark. The spatial location itself may be viewed as a mark for a simple point process in time, thereby providing one route to likelihood analyses of space–time models. Further characteristics, such as magnitude, spatial extent, or even duration, can be added as additional marks. Deft use of this procedure, such as has been employed for some decades in applications to queueing systems and networks, can be very helpful in making complex models more tractable.

Finally it should be observed that behind many observed space–time point processes lie evolving but unobserved spatial fields: earthquakes may be regarded as a response to some evolving stress field, forest fires as a response to some underlying spatial field determining the ignition potential, and so on. Thus, the study of space–time point processes leads almost inevitably to the more general study of evolving spatial fields, although practical modelling in this direction is still limited and very subject-specific.

We turn to a more systematic study of space–time point processes and focus

on the most commonly occurring situation, namely, the process is stationary in time but not necessarily homogeneous<sup>1</sup> in space or in the mark distribution.

The two main aspects we discuss are the first- and second-order moment properties, thereby revisiting and elaborating the discussion in Sections 8.3 and 12.3, and the extension to space–time processes of the conditional intensity and likelihood arguments of Chapter 14. For simplicity of exposition we suppose throughout that space here refers to Euclidean space  $\mathbb{R}^2$ ; other options, such as point processes on the circle or the sphere, are indicated briefly in Exercises 15.4.1–2. For further examples, discussion and references, see Vere-Jones (2007).

We start with first moments.

**Lemma 15.4.I.** *If a stationary marked space–time point process has finite overall ground rate  $m_g$ , then there exist a distribution  $\Phi(dx)$  in space, and a family of conditional distributions  $\Psi(d\nu \mid x)$  for the residual mark  $\nu$ , such that the first moment measure can be decomposed as*

$$M(dt \times dx \times d\nu) = m_g dt \Phi(dx) \Psi(d\nu \mid x). \quad (15.4.1)$$

PROOF. We know from Proposition 8.3.II that, if an MPP is stationary and has finite ground intensity  $m_g$ , then its first moment measure can be written in the general form

$$M(dt \times d\kappa) = m_g dt \Pi(d\kappa).$$

The mark  $\kappa$  here has two components, the location  $x$  and a residual mark,  $\nu$  say, so the measure  $\Pi$  here is a bivariate distribution on the product of the location space and the space of the residual mark. The decomposition (15.4.1) is then just the standard disintegration of  $\Pi(\cdot)$  into the marginal distribution in space and a family of conditional distributions for the residual mark, given the spatial location.  $\square$

In this lemma, we allow the distribution of the residual mark  $\nu$  to depend on the spatial location  $x$ , but not (from stationarity) on the time  $t$ . The lemma implies in particular that a stationary space–time Poisson process must have an intensity measure  $\Lambda$  which can be disintegrated as in (15.4.1). Assuming densities exist, its intensity in (space–time, mark) space will be of the form

$$\lambda(t, x, \nu) = \lambda_g \phi(x) \psi(\nu \mid x),$$

where  $\lambda_g$  is the overall intensity (ground rate). [We follow convention by calling the process ‘space–time’, and write the components of a typical ‘space–time’ point  $(t, x)$  in reverse order.] The process can be otherwise interpreted as a space–time compound Poisson process with spatially varying space–time intensity  $\lambda(t, x) = \lambda_g \phi(x)$  and spatially dependent mark distribution with density  $\psi(\nu \mid x)$ .

---

<sup>1</sup> For a space–time process on, e.g.,  $\mathbb{R}_+ \times \mathbb{R}^2$  we adopt the convention of using *stationarity* to refer to invariance with respect to time-shifts, and *homogeneity* to refer to invariance with respect to shifts in space. Thus, a process that is invariant under shifts in both time and space is an homogeneous stationary space–time process.



EXAMPLE 15.4(a) *Models for persistent points; spatial M/G/ $\infty$  queue.* This example illustrates in simple form some of the issues which have been mentioned. Although in essence it is a model for particles that persist in time, it can be reduced to a marked space–time point process, in the narrow sense in which we have defined it, by treating the duration, as well as the location, as part of the mark. It can be regarded as a spatial version of an M/G/ $\infty$  queue, but many models for population and other processes have a similar general structure: particles which arrive at times  $t_i$  and locations  $x_i$ , persist for some time  $\tau_i$  and then die or otherwise disappear. The spatial birth-and-death process of Example 10.4(e) is a more complex example, where the duration  $\tau_i$  may depend on the evolving history of the whole set of particles. It in turn is a special case of the wider class of branching diffusion models, which has a considerable literature of its own (see Section 13.5).

One basic approach to the process is to consider it as an MPP in time, say  $N(dt \times d\kappa)$ , with time points  $t_i$  and marks  $\kappa_i = (x_i, \tau_i)$  embracing both the locations  $x_i$  and the durations  $\tau_i$ . Equally, it may be regarded as a space–time point process, with locations  $(t_i, x_i)$  say, and associated marks  $\tau_i$ . Once this underlying process has been specified, all other characteristics should be derivable from it.

Observations, however, may be restricted to snapshots of the time-varying spatial point pattern  $N_t(dx)$  representing the locations of the particles extant at time  $t$ . Here  $N_t$  may be regarded as a stochastic process taking values in  $\mathcal{X}^\cup$ ; the locations of the points at time  $t$  can also be represented as a vector  $\mathbf{x}_t$ , anticipating the notation to be used in Section 15.5. A slightly different approach is to fix sets  $A_i$  in  $\mathcal{X}$  and look at the joint evolution of the processes  $X_i(t) = N_t(A_i)$  as a multivariate time series. In any case, one initial question is to find a representation of the spatial processes  $N_t$  in terms of the underlying process  $N$ , and to examine how far the structure of  $N$  can be reconstructed from observations on the  $N_t$ .

The basic relation is of simple linear form:

$$N_t(A) = \int_{s=-\infty}^t \int_A \left[ \int_{\tau=t-s}^{\infty} N(ds \times dx \times d\tau) \right].$$

This representation is immediately useful in obtaining the first moment measure for  $N_t(\cdot)$ . Suppose that the MPP  $N$  is stationary in time, and, adopting notation similar to Lemma 15.4.I, that its first moment measure has a density which can be written in the form

$$m(t, x, \tau) = m_g \phi(x) \psi(\tau | x),$$

where  $\phi(x)$  is a time-invariant probability density over the spatial region of interest, and  $\psi(\tau | x)$  is the time-invariant probability density function for the life of a particle started at location  $x$ . Taking expectations in the expression for  $N_t$  we obtain  $E[N_t(dx)] = m_t(x) dx$ , where

$$m_t(x) = m_g \phi(x) \int_{-\infty}^t \int_{t-s}^{\infty} \psi(\tau | x) d\tau ds = m_g \phi(x) \int_0^{\infty} \tau \psi(\tau | x) d\tau$$

and  $\int_0^\infty \tau \psi(\tau | x) d\tau \equiv L(x)$  is the mean lifetime of a particle started at location  $x$ . A similar (albeit more involved) representation for the second moment measure of  $N_t$  is outlined in Exercise 15.4.2.

Only the mean of the lifetime distribution can be obtained from the above expression, even in the case that the spatial mean  $L(x)$  is independent of  $x$  and the ground rate  $m_g$  could be independently estimated. If the aim were to obtain further information about the lifetime distribution, it would be necessary to combine observations on  $N_t$  over a sequence of values of  $t$ .

Let us then consider, as a second step, the first moment structure for observations on two snapshots  $N_{t_1}$  and  $N_{t_2}$ . The combined observations can be treated as a single realization of a multivariate point process on the location space  $\mathcal{X}$ , with points of three different types: Type 1 is observed at  $t_1$  but not  $t_2$ , Type 2 at both times, and Type 3 at  $t_2$  but not  $t_1$ . (We assume, here and above, that the particles themselves cannot be distinguished by their ages: they are either present or not present.) Arguing much as above, and writing  $\Delta = t_2 - t_1$ , we obtain for the first moment measures of the three components

$$\begin{aligned} m_1(x) &= m_g \phi(x) \int_{-\infty}^{t_1} \int_{t_1-s}^{t_2-s} \psi(\tau | x) d\tau ds = m_g \phi(x) A_1(\Delta | x), \\ m_2(x) &= m_g \phi(x) \int_{-\infty}^{t_1} \int_{t_2-s}^{\infty} \psi(\tau | x) d\tau ds = m_g \phi(x) A_2(\Delta | x), \\ m_3(x) &= m_g \phi(x) \int_{t_1}^{t_2} \int_{t_2}^{\infty} \psi(\tau | x) d\tau ds = m_g \phi(x) A_3(\Delta | x), \end{aligned}$$

where

$$\begin{aligned} A_1(\Delta | x) &= A_3(\Delta | \sigma) = \int_0^\Delta \sigma \psi(\sigma | x) d\sigma + \Delta \int_\Delta^\infty \psi(\sigma | x) d\sigma, \\ A_2(\Delta | x) &= \int_\Delta^\infty \sigma \psi(\sigma | x) d\sigma - \Delta \int_\Delta^\infty \psi(\sigma | x) d\sigma. \end{aligned}$$

Similar decompositions can be obtained for the first moments of larger numbers of snapshots, and begin to piece together information about  $\psi$ .

Because of the simple linear relation between  $N_t$  and  $N$ , it is possible to extend the moment results into results for p.g.fl.s. For  $h \in \mathcal{V}$  we obtain

$$\begin{aligned} G_t[h] &= \mathbb{E} \left[ \prod_{t_i < t} \left( I(t - t_i < \tau_i) \cdot 1 + I(t - t_i \geq \tau_i) h(x_i) \right) \right] \\ &= \mathbb{E} \left[ \prod_{t_i < t} \mathbb{E} \left( 1 - I(t - t_i \geq \tau_i) [1 - h(x_i)] \mid (t_i, x_i) \right) \right], \end{aligned}$$

provided the  $\tau_i$ , which may depend on  $x_i$ , are independent of the past of the process up to  $t_i$ . We can then take expectations conditional on the  $(t_i, x_i)$  to obtain, using the Heaviside function  $H(u) = 0$  or  $1$  as  $u < 0$  or  $\geq 0$ ,

$$G_t[h] = \mathbb{E} \left[ \prod_i \left( 1 - H(t - t_i) [1 - h(x_i)] \Psi(t - t_i | x_i) \right) \right] = G^-[h_t],$$

where  $\Psi(u \mid x) = \int_0^u \psi(v \mid x) dv$ ,  $G^-$  is the p.g.fl. of the times and locations only, and  $h_t$  is the function

$$h_t(u, x) = 1 - H(t - u)[1 - h(x)]\Psi(t - u \mid x).$$

In view of stationarity, the last expression for  $G_t[h]$  is independent of  $t$ , because we can replace  $t - u$  by  $v$  without altering the value of the expectation.

In simple cases, the p.g.fl. can be evaluated explicitly. If the initiating points  $(t_i, x_i)$  form a constant rate Poisson process with intensity  $\lambda_g \phi(x)$ , then the previous expression can be evaluated as

$$\begin{aligned} G_t[h] &= \exp \left( -\lambda_g \int_{\mathcal{X}} [1 - h(x)] \phi(x) dx \int_0^\infty [1 - \Psi(v \mid x)] dv \right) \\ &= \exp \left( -\lambda_g \int_{\mathcal{X}} [1 - h(x)] L(x) \phi(x) dx \right). \end{aligned}$$

The interpretation here is that, independently of  $t$ , the process of extant particles is a Poisson process over  $\mathcal{X}$  with intensity  $\lambda(t, x) = \lambda_g L(x) \phi(x)$ .  $\square$

Even in the case of the first moment measure, finding nonparametric estimates for the measure is a problem which needs to be approached with some caution. In addition to the boundary problems which inevitably arise in spatial processes, a particular difficulty in the space-time context is the problem of distinguishing between transient features, such as random clusters, and long-term spatial inhomogeneities. Suppose that some form of kernel estimate is adopted, say

$$\hat{m}(t, y) = \int_{t-s \in A, y-x \in B} h_1(t-s) h_2(y-x) N(ds \times dx),$$

where the temporal and spatial sets  $A$  and  $B$  must be selected to reflect appropriate ‘bandwidths’ for the two smoothing kernels  $h_1$  and  $h_2$ . Unless the bandwidths are chosen with particular care (and possibly even then) an estimate of the above kind will either reflect the transient clusters (bandwidths too small) or smooth over true inhomogeneities (bandwidths too large). This in turn means that some knowledge of the cluster structure is required before the bandwidths are chosen. On the other hand, determining the cluster structure equally requires some knowledge of the spatial or temporal inhomogeneities.

This dilemma can be resolved only partially in practical situations, for example by making a preliminary estimate of the clustering and then using this to determine an initial choice of bandwidth. Vere-Jones (1992) and Musmeci and Vere-Jones (1987) describe two different *ad hoc* approaches to this problem, both requiring a preliminary estimate of the extent of clustering based on local variance/mean ratios. Moving to variable bandwidths generally leads to better visual representations of the data, but does not eliminate, indeed may only intensify, the above dilemma, and is in itself a nontrivial exercise.

Such difficulties raise the issue of whether, in fact, estimating the first moment measure in such complex situations is even a desirable goal. In many cases, the immediate concern is to find an informative visual display of the data, and many forms of kernel smoothing will achieve that. The question of which, if any, features are persistent spatial inhomogeneities may be better explored at a later stage, through the fitting of exploratory models. In this connection, Zhuang *et al.* (2002, 2004) describe a powerful technique, *stochastic declustering*, based on the assumption that the observed process can be approximated by a space–time ETAS model [Example 6.4(d)]. Suppose in the first instance that the process of initiating events is stationary in time but varying in space, and that the cluster parameters are constant in both time and space; then the fitted model can be used to estimate, for each observed event, the probability that that event is an initiating event (i.e., not the offspring of some earlier event). Smoothing these probabilities gives a first estimate of the first moment measure, not of the process as a whole, but of the process of initiating events. This estimate in turn can be used to give improved estimates of the cluster parameters, and the steps iterated until convergence is achieved. The resulting estimate for the first moment measure of the initiating events may be a more useful function than the first moment measure of the overall process. For example, it can play an important role as a diagnostic tool in identifying areas or time periods in which the process departs from its normal behaviour.

This technique can be extended to provide, for every event, the probability that the event is an offspring of any given preceding event; in turn, these probabilities can be used as the basis for simulating the detailed cluster structure of the data (*stochastic reconstruction*). In the earthquake context, these procedures lead to a technique for determining the clusters that is free of the subjective criteria used in other procedures for such declustering. On the other hand, the reconstruction is not unique, but depends on the particular simulation, and is based on the assumption that the underlying process has an ETAS structure. Nevertheless, it has already proven of value as an analytical and diagnostic tool.

We turn next to a consideration of second moment measures in the space–time context. These can become quite complicated even in the simple case that the process is stationary and has no additional marks. Several equivalent representations are possible, all variations on the basic form given in Proposition 8.3.II, namely,

$$M_2(dt_1 \times d\kappa_1 \times dt_2 \times d\kappa_2) = \check{M}_2(du \times d\kappa_1 \times d\kappa_2) dt_1, \quad (15.4.2)$$

where  $u = t_2 - t_1$  and  $\check{M}_2$  is ‘reduced’ with respect to the time variable only. Suppose for definiteness that space here is  $\mathbb{R}^2$ , so that each  $\kappa$  can be interpreted as a point  $x \in \mathbb{R}^2$ . Then this reduced measure can itself be looked at and standardized in several different ways, as indicated in the next proposition, which both specializes and refines the earlier discussion around Lemma 8.3.III.

**Proposition 15.4.II.** (a) *If a simple, stationary, space–time point process has boundedly finite ground process with finite second moment measure, then its reduced moment measure  $\check{M}_2$  as at (15.4.2) can be represented in either of the two forms*

$$\check{M}_2(du \times dx_1 \times dx_2) = \check{M}_2^g(du) \Pi_2(dx_1 \times dx_2 \mid u) \quad (15.4.3)$$

where  $\check{M}_2^g(\cdot)$  is the reduced second moment measure for the ground process, and  $\Pi_2(\cdot \times \cdot \mid u)$  is a bivariate probability distribution for the locations, given that the occurrence times are separated by an interval of length  $u$ ; or

$$\check{M}_2(du \times dx_1 \times dx_2) = \check{M}_2(du \mid x_1, x_2) \Pi(dx_1) \Pi(dx_2), \quad (15.4.4)$$

where  $\Pi(\cdot)$  is the stationary distribution in space, and  $\check{M}_2(du \mid x_1, x_2)$  is a reduced cross-moment measure for the occurrence of points at distinct locations  $x_1, x_2$ .

(b) *If the process is an homogeneous stationary space–time process with finite mean intensity  $m$  per space–time unit, then*

$$M_2(dt_1 \times dx_1 \times dt_2 \times dx_2) = m \mathring{M}_1(du \times dy) dt_1 dx_1 \quad (15.4.5)$$

where  $u = t_2 - t_1$ ,  $y = x_2 - x_1$  and  $\mathring{M}_1(\cdot)$  is the first-moment measure of the Palm distribution at (13.2.15).

PROOF. (15.4.3) is the form already treated in Lemma 8.3.III; it is a disintegration of  $\check{M}_2$  with respect to its marginal measure in  $u$  and the conditional distribution of the marks given  $u$ . Similarly (15.4.4) is a disintegration with respect to the product measure  $\Pi \times \Pi$  in  $(x_1, x_2)$ , justified by the absolute continuity of  $\check{M}_2(A \times dx_1 \times dx_2)$  with respect to  $\Pi(dx_1) \times \Pi(dx_2)$  for all bounded  $A$ . In both these representations, the assumption that the ground process exists implies that  $N(A \times \mathbb{R}^2) < \infty$  a.s. for bounded  $A$ , so that either the spatial coordinates are themselves restricted to a bounded set, or the occurrence rate drops off rapidly away from the spatial origin.

The final representation (15.4.5) is not constrained in this way, and is a consequence of the comments regarding moments of the Palm distribution summarized in Proposition 13.2.VI and in the discussion around (13.4.4). In general the local Palm distribution is conditioned by the value of the mark (here the location) at the point selected as origin, but when the process is homogeneous in space as well as time, the Palm distribution is independent of both the time and space coordinates of the point selected as origin, leading to the form (15.4.5) for its first moment measure.  $\square$

Before leaving (15.4.3–5) we note some further aspects of these representations. In (15.4.3),  $\Pi_2$  is not in general symmetric in  $(x_1, x_2)$  (the time sequence in which the two marks occur is important), nor do its marginals reduce in general to the stationary mark distribution. Rather, it provides a useful descriptive summary of how the spatial or mark distribution changes with the

time interval between the two points. For ergodic processes and large  $u$ , it should approximate to the product form  $\Pi(dx_1) \times \Pi(dx_2)$ ; for small  $u$ , and for cluster processes in particular, it may be highly concentrated about the diagonal  $x_1 = x_2$ . Interpreting the marks as spatial coordinates, Example 8.3(e) illustrates some types of behaviour which can occur in a space-time cluster process under rather general conditions. A further illustration, focussing on second-order properties, is in Example 15.4(c) below.

The decomposition (15.4.4) is the natural form to use in connection with spectral analysis, for its Fourier transform with respect to  $u$  gives a quantity, loosely defined in the ergodic case as

$$\gamma(\omega \mid x_1, x_2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega u} [\check{M}_2(du \mid x_1, x_2) - m_g^2 du],$$

which has the character of a cross-spectral density of the point processes associated with infinitesimal regions around the locations  $x_1$  and  $x_2$ . Indeed, if integrated over two disjoint spatial regions  $A, B$  it gives precisely such a cross-spectrum, namely that between occurrences in regions  $A$  and  $B$ :

$$\gamma_{A,B}(\omega) = \int_A \int_B \gamma(\omega \mid x_1, x_2) \Pi(dx_1) \Pi(dx_2). \quad (15.4.6)$$

Some examples and further discussion are given following Propositions 8.2.I and 8.2.III, the latter giving the spectral representation for an isotropic point process in  $\mathbb{R}^2$ .

The third decomposition (15.4.5) is the natural form to use when the stationary process is (spatially) homogeneous. It describes the expectation of finding a second point of the process after time lag  $u$  at distance  $y$  from a first point located at the space-time origin. If the process is also isotropic, then the first-order Palm measure  $\check{M}_1(du \times dy)$  depends only on the length  $\|y\|$ . This leads to the possibility of defining a family of Ripley  $K$ -functions,  $K(r \mid u)$  say, indexed by the lag  $u$  and defined by a decomposition of the type

$$\check{M}_1(du \times S_r) = K(r \mid u) du \quad (15.4.7)$$

( $S_r$  is the disc of radius  $r$  about the origin in  $\mathbb{R}^2$ ).

Such functions can be used to examine the change in the characteristics of the  $K$ -function with increasing time lags; for large lag  $u$  and ergodic processes, the initial choice of spatial origin is irrelevant and the form approximates that for the first moment, in this case the same as for an homogeneous Poisson process.

An alternative procedure for examining the same question is to consider the behaviour of the density  $\check{m}_1(u, y)$  as a function of the distance  $\|y\|$  for increasing values of  $u$ , that is, corresponding roughly to the behaviour of the second moment measure in annular regions about the spatial origin at increasing separations in time.

Without homogeneity, or in the presence of additional marks, there is no uniquely defined Palm distribution, but rather a family of Palm distributions indexed by the location, or by the value of the mark at the point selected as time origin. In such a situation, the most useful surrogate may be the average  $\bar{\mathcal{P}}_0(\cdot)$  of the Palm distributions relative to a point at the origin having a specified mark, as discussed around (13.4.2b).

**EXAMPLE 15.4(b) A random walk renewal process.** We consider a point process in time and space, with successive points represented as pairs  $(t_n, x_n)$ . It is assumed that the time points  $t_n$  form a renewal process (so that successive intervals  $t_{n+1} - t_n$  are i.i.d., with finite mean length  $\mu$  say), and the  $x_n$  (independently) form a random walk, so that the differences  $x_{n+1} - x_n$  also form an i.i.d. sequence, independent of the sequence of time intervals.

Although this process is not stationary (see Exercise 15.4.3), it is Palm stationary; indeed, its Palm structure, meaning the structure relative to a point pair  $(t_n, x_n)$  as origin, is very simple: it is just a random walk in the product space (time  $\times$  space), with the selected point as origin. From this observation, the first moment measure  $\dot{M}_1(du \times dy)$  of the Palm distribution which figures in (15.4.5) can be deduced, even though (15.4.5) itself is not meaningful. Let  $f(\cdot)$ ,  $g(\cdot)$  denote density functions for the time and space intervals respectively. As in the case of the simple renewal process studied in Chapter 4, the probability density of a second point having time and space coordinates  $(u, y)$ , given the occurrence of a point at the origin, can be written as a sum of convolution terms, so that for the density  $\dot{m}_1$  of  $\dot{M}_1$  we have

$$\dot{m}_1(u, y) = \sum_{n=1}^{\infty} f^{n*}(u) g^{n*}(y), \quad (15.4.8)$$

$f^{n*}(\cdot)$ ,  $g^{n*}(\cdot)$  denoting the  $n$ th convolution powers of  $f$  and  $g$ , respectively.

If  $h(u) = \sum_{n=1}^{\infty} f^{n*}(u)$  denotes the ordinary renewal density function in time, obtained by integrating out the spatial components, then we can write

$$\dot{m}_1(u, y) = h(u) \pi(y | u) = h(u) \sum_{n=1}^{\infty} \pi_n(u) g^{n*}(y),$$

where  $\pi(y | u)$  denotes the conditional probability density that, if a second point is observed at time  $u$ , then its location is at  $y$ , and the final sum exhibits this density as the sum of terms  $\pi_n(u) = f^{n*}(u)/h(u)$  for the probability that this second point corresponds to the  $n$ th term in the random walk. For example, if the renewal process in time reduces to a Poisson process of rate  $\lambda$ , the weights are simply the Poisson probabilities that precisely  $n - 1$  points have occurred in the time interval  $(0, t)$  (see Exercise 15.4.3).

Although this discussion is quite general, the asymptotic behaviour of the second moment measure depends crucially on the character of the underlying space. If this is  $\mathbb{R}^2$ , the random walk gradually diffuses away from the spatial

origin, and  $h(t, x) \rightarrow 0$  for all  $x$  as  $t \rightarrow \infty$ , even though  $h(t) \rightarrow 1/\mu$ . We have here an example for which the moment of the Palm distribution has a well-defined stationary form although no corresponding stationary point process exists.

If the space is bounded, in particular if  $\mathbb{R}^2$  is replaced by the circumference of the unit circle  $\mathbb{S}$  or the surface of the unit sphere, the behaviour is quite different, and a stationary version of the process does exist. In the case of  $\mathbb{S}$  for example, the convolution powers of a distribution  $g$  converge towards the uniform distribution on  $\mathbb{S}$ , and the function  $h(t, \theta)$  converges to the positive constant  $1/(2\pi\mu)$ , corresponding to the first moment density of the stationary version of the process. See also Exercise 15.4.4.  $\square$

EXAMPLE 15.4(c) *Second-order properties of space–time Poisson cluster processes.* As in the case of distance properties, the general results on Poisson cluster processes outlined in Proposition 6.3.III can be used as a starting point for examining the second-order properties of cluster processes in both spatial and space–time contexts. Let us assume that the process of cluster centres is a space–time Poisson process with spatially varying but time-stationary intensity

$$\Lambda(dt \times dx) = \lambda_c(x) dt dx,$$

and that, conditional on a cluster centre at time  $t$  and location  $x$ , the second factorial moment density for the cluster member process, in time and space, can be described by the function

$$\rho_{[2]}(s, y_1, y_2 \mid t, x) = \check{\rho}(s - t; y_1, y_2 \mid x).$$

These are the natural conditions for a process stationary in time but not necessarily homogeneous in space.

Then from the general results of Proposition 6.3.III, we obtain for the density of the function  $\check{M}_2(\cdot)$  of (15.4.2),

$$\check{m}_2(u; y_1, y_2) = \int_{\mathcal{X}} \rho_{[2]}(u; y_1, y_2 \mid x) \lambda_c(x) dx.$$

To obtain the decomposition at (15.4.3), note that, for  $u \neq 0$ , and assuming the integrals are finite,

$$\check{M}_2^g(du) = \check{m}_2^g(u) du = \int_{\mathcal{X} \times \mathcal{X}} \check{m}(u; y_1, y_2) dy_1 dy_2,$$

so that the bivariate kernel  $\Pi_2$  of (15.4.3) has a density  $\pi_2$  given by

$$\pi_2(y_1, y_2 \mid u) = \check{m}_2(u; y_1, y_2) / \check{m}_2^g(u) \quad (u \neq 0).$$

To obtain the further decomposition at (15.4.4), we need first to determine the overall ground rate  $m_g$  and the stationary distribution  $\Pi(dx)$  for the



spatial locations. Let  $m(y, u \mid x)$  denote the mean density of cluster elements at location  $y$  and after time lag  $u$ , from a cluster centre at location  $x$ . Then the mean cluster size for a cluster with centre  $x$  is given by

$$\mu(x) = \int_{\mathbb{R} \times \mathcal{X}} m(y, u \mid x) \, du \, dy, \quad \text{and} \quad m_g = \int_{\mathcal{X}} \lambda_c(x) \mu(x) \, dx.$$

For the density  $\pi$  of the stationary distribution  $\Pi$  of locations we have

$$\pi(y) = \frac{1}{m_g} \left[ \int_{\mathbb{R} \times \mathcal{X}} \lambda_c(x) m(y, u \mid x) \, du \, dx \right].$$

Then the kernel  $\check{M}_2(u \mid x_1, x_2)$  of (15.4.4) has density

$$\check{m}_2(u \mid x_1, x_2) = \check{m}_2(u, x_1, x_2) / [\pi(x_1)\pi(x_2)].$$

Whether these functions have convenient explicit forms depends on the particular assumptions for the cluster structure and the functions  $\rho_{[2]}(s, y_1, y_2 \mid t, x)$  and  $m(y, u \mid x)$  which then arise. A space-time Neyman-Scott example is outlined in Exercise 15.4.5, and a space-time analogue of the Bartlett-Lewis model, based on a finite version of Example 15.4(b), in Exercise 15.4.6.  $\square$

Nonparametric estimation of the second-order moment structure is again a difficult exercise, which we review only briefly. Suppose that the process can be assumed to be both stationary in time and homogeneous in space, so that an elementary estimate of the mean rate, such as  $N(W)/\ell(W)$ , where  $W$  is the space-time observation region, is available, and one can proceed to the estimation of second-order properties.

Under these assumptions, the most convenient description of the second order properties is through the first moment measure  $\hat{M}_1(du \times dy)$  of the Palm distribution, as in (15.4.5). The following is a general approach to its estimation. Let  $E$  be a small test set in space-time, for example,  $I_\delta(\tau) \times S_\epsilon(x)$ , where  $I_\delta(\tau)$  is the interval  $(\tau - \delta, \tau + \delta)$ , and  $S_\epsilon(x)$  the sphere with centre  $x$  and radius  $\epsilon$ . Estimate the Palm moment measure  $\hat{M}_1(E)$  by the average

$$\hat{\hat{M}}_1(E) = \frac{1}{N(W)} \sum_{i=1}^{N(W)} N(T_{z_i} E), \quad (15.4.9)$$

where  $T_z$  denotes a shift through  $z$  and the  $z_i = (t_i, x_i)$  denote the observed points. Dividing by the measure  $2\delta\pi\epsilon^2$  of the test set  $E$ , and varying the parameters  $\tau$  and  $x$  determining the location of  $E$  relative to the origin, one can build up a picture of the behaviour of the density of the Palm distribution over dimensions up to the order of some fraction (perhaps a quarter) of the size of the observation region.

In an early study of earthquake patterns in New Zealand, Chong (1983) used this kind of technique to examine how occurrence rates varied with time

and distance about a typical event characterized by its magnitude. By taking the test set  $E$  to be of annular form, a rough estimate was made of how the occurrence rate varied over various times and distance ranges about an event (earthquake) at the origin with magnitude in a specified small range. Instead of using annular sets, a disc could have been used, of radius  $r$  say, thus giving an indication of how the form of the  $K$ -function varied with the time separation from the origin, as in (15.4.8).

In addition to the usual problems of boundary effects and departures from spatial homogeneity, a particular difficulty in Chong's study is that the data are dominated by the clusters initiated by one or two large events, so that no very clear picture of the stationary behaviour can be obtained.

A superficially different (but ultimately equivalent) approach is to plot all pairs  $(t_i, x_i), (t_j, x_j)$  ( $i \neq j$ ) as points in the product space  $\mathcal{X} \times \mathcal{X}$  and estimate the first moment measure of this product counting measure, using a kernel estimate. A recent development of this type of procedure is given in Tanaka and Ogata (2005), who assume that the pairs can be treated as the realization of a nonhomogeneous Poisson process in the product space, and use the Palm intensity for an isotropic process as a surrogate for the Poisson intensity.

Estimation of the Bartlett spectrum for space–time processes assumes stationarity in time, and starts from the sample periodogram, which in turn can be obtained from finite Fourier transforms such as

$$J_i(\omega) = \int_0^T e^{-i\omega t} [N_{A_i}(dt) - m_{A_i} dt],$$

where  $m_{A_i}$  denotes the mean rate of occurrence (in time) of points in  $A_i$ . For each partition of the spatial region into sets  $A_i$  we can then form the matrix of periodogram estimates

$$I_{ij}(\omega) = \frac{1}{2\pi T} J_i(\omega) \bar{J}_j(\omega).$$

Each such quantity may be regarded as an estimate of the type of cross periodogram defined in (15.4.6). Further analysis can then proceed by smoothing the periodogram in the frequency domain, and by refining the partition and then smoothing the resulting spectral functions in the spatial domain.

In all cases, boundary effects cause major problems. As already mentioned, a variety of procedures for correcting the biases arising from edge effects have been developed for spatial point processes, and many of these can be adapted for use in the space–time context. Perhaps the most generally useful, even if somewhat inefficient in terms of data use, is to carry out the analysis within an extended space–time region, comprising the inner observation region and a buffer zone surrounding it in space and preceding it in time. Averaging is carried out over points  $(t_i, x_i)$  within the observation region only; use of the extended region ensures that full contributions can be obtained even from test sets centred on points near the boundary.

In a parametric model, the second-order moments will in general be functions of the parameters of interest, so that in principle estimates of the parameters can be obtained by fitting the theoretical forms for the moments to their nonparametric estimates. This corresponds to a form of moment estimation for the model parameters. The difficulty, however, is that little is known about the efficiency of such estimates, and indeed one would expect the efficiency to depend crucially on the sensitivity of the moment density to changes in the parameter values. Channelling the estimation through the spectrum, as suggested many years ago by Whittle (1951), has the advantage of reducing the estimate to linear combinations of approximately independent terms, and was adapted to a point process context by Ogata and Katsura (1991), but even here the variance properties are not easy to determine.

We turn finally to methods based on the conditional intensity. As soon as time appears as a governing variable, the possibility arises of representing the spatial coordinate as a mark, and hence appealing to the likelihood, simulation, and prediction methods described in Chapters 7 and 14.

In the space-time context, the conditional intensity  $\lambda^*(t, x)$  becomes a function of two variables with the property that for every spatial Borel set  $A$ , the quantity  $\lambda_A^*(t) = \int_A \lambda^*(t, x) dx$  is a conditional intensity for the point process of points with locations in  $A$ ,  $N_A(t) = N_A(0, t] = \int_{(0, t] \times A} N(du \times dx) = N((0, t] \times A)$ . A suitable reference model in this context is the bivariate Poisson process with constant overall rate  $\lambda$  and independently distributed spatial locations with strictly positive density  $f(x)$ . The log likelihood ratio then becomes

$$\log \frac{L_1}{L_0} = \sum_{i=1}^{N_g(T)} \log \frac{\lambda^*(t_i, x_i)}{\lambda f(x_i)} - \int_0^T \int_V [\lambda(t, x) - \lambda f(x)] dt dx, \quad (15.4.10)$$

where  $V$  is the spatial region under consideration. If additional marks are needed, then the conditional intensity becomes a function of three variables  $\lambda_{\mathcal{F}}^*(t, x, \kappa)$ , and a Poisson process with intensity  $\lambda_g f(x) g(\kappa | x)$  could be used for the reference process.

For computational purposes, the space-time likelihood is often written more conveniently in terms of the *ground intensity*  $\lambda_g^*(t)$  for the ground process and the conditional mark (spatial) distribution  $f^*(x | t)$ , so that  $\lambda^*(t, x) = \lambda_g^*(t) f^*(x | t)$ . The star here indicates that both quantities are conditional on the history  $\mathcal{F}_t$  up to time  $t$ . Provided  $f^*$  is normalized to a probability density for any given past history, we can write

$$\log \frac{L_1}{L_0} = \left[ \sum_{i=1}^{N(T)} \log \frac{\lambda_g^*(t_i)}{\lambda} - \int_0^T [\lambda_g(t) - \lambda] dt \right] + \sum_{i=1}^{N_g(T)} \log \frac{f^*(x_i | t_i)}{f(x_i)}.$$

In many models, the parameters appearing in the two terms have no common variables, in which case optimization can be carried out for the two terms separately.

As in the estimation of moment measures, parameter estimates obtained via likelihood maximization can be seriously biased by problems associated with both initial values and boundary effects. In the rare cases that the data permit, a buffer zone around the observation region in both time (preceding) and space (surrounding) may be introduced, and the likelihood ratio computed just for the data points lying within the observation region; the buffer zone here provides the information needed to compute the conditional intensity within the observation region. In principle it could be replaced by a form of stationary distribution for the boundary effects, analogous to the forward recurrence time distribution for a renewal process, but finding the analytical form for such boundary terms is a problem of similar difficulty to the Ising problem (see also Section 15.6).

Both Bayesian and non-Bayesian approaches to parameter estimation exist. Software routines for likelihood estimation, based on direct maximization of the likelihood or the posterior density, are incorporated in (among other packages) David Harte’s SSlib routines [Harte (2003), Brownrigg and Harte (2005)], where the procedures extend to point processes with i.i.d. marks.

Some of the most notable studies in the Bayesian direction have been made by Ogata and colleagues in Tokyo, following the general approach to nonstationary modelling for time series methods (ABIC) suggested by Akaike. Recent descriptions of their methods applied to non-stationary versions of the ETAS model are in Ogata *et al.* (2003), Ogata (2004), and Ogata and Zhuang (2006), and briefly described below.

EXAMPLE 15.4(d) *Space–time ETAS model* [see Example 6.4(d)]. After the Poisson model, this has become the best-known model in seismicity studies, where it is used as a first approximation to a wide range of catalogue data. In particular, local departures from the fit of an overall ETAS model have become an important diagnostic tool in identifying anomalous regions or time intervals.

The complete conditional intensity  $\lambda^\dagger$  for the spatial ETAS model has the three-dimensional form

$$\lambda^\dagger(t, x, M) = \beta e^{-\beta(M-M_0)} \left[ \mu f(x) + A \sum_{i:t_i < t} e^{\alpha(M_i-M_0)} g(t-t_i) h(x-x_i \mid M_i) \right], \quad (15.4.11)$$

where  $\mu$  is the arrival rate of ‘immigrants’,  $f(x)$  is the probability density for the location of a newly arrived immigrant,  $g$  and  $h$  are probability densities (in time and space, respectively) for the time and space coordinates of an ‘offspring’ event about its parent, the first exponential term  $\beta e^{-\beta(M-M_0)}$  describes the distribution of the ‘magnitude’ of a newly occurring event above a fixed threshold  $M_0$ ,  $A$  is a constant determining the criticality of the process, and the other exponential term  $e^{\alpha(M_i-M_0)}$  describes the factor by which the parent’s magnitude  $M_i$  inflates the expected number of offspring. In typical earthquake applications [e.g., Ogata (1998)],

$$g(t) = \frac{p-1}{c} \left(1 + \frac{t}{c}\right)^{-p} \quad (t > 0),$$

and

$$h(x \mid M) = \frac{q-1}{\pi D e^{\alpha M}} \left(1 + \frac{\|x\|^2}{D e^{\alpha M}}\right)^{-q}.$$

The condition for stability is  $\rho < 1$ , where

$$\rho = A\beta \int_{M_0}^{\infty} e^{\alpha(M-M_0)} e^{-\beta(M-M_0)} dM = \frac{A\beta}{\beta - \alpha}$$

can be interpreted as the expected number of direct offspring per ancestor, averaged over the ancestor's magnitude.

In a semiparametric version of this model, model parameters such as  $p, c, D, \alpha, q$  are fitted separately to distinct time intervals and geographic regions. Then the ABIC methods are used to select the parameters in the prior distributions for the model parameters. Here the prior distributions ('hyperparameters') control the smoothness of the fitted parameters (i.e., the rate of change of their values with changes in time or space), and the AIC procedure is used to select the hyperparameters which give an optimal degree of smoothing. Ogata's most recent studies use a Delaunay tessellation into cells of the space-time observation region, with each cell containing just one point of the observed process; three-dimensional splines are then used to effect the linking between cells. The general method follows from Tanemura *et al.* (1983); recent applications are in the three papers cited earlier.  $\square$

A rather different approach to space-time models is illustrated by the linked stress release model of Example 7.3(d). Here the observation region is subdivided into a small number of subregions, and the process is analyzed as a multivariate point process, the linkage between subregions being controlled in this example by a transfer to neighbouring regions of some proportion of the stress released by the occurrence of an event in any given region. Analogous models can be envisaged involving the spatial spread of infection for (spatial) epidemic models, or the extent of fire risk in forest fire models.

The next example illustrates a further type of model structure in which external variables are incorporated into the conditional intensity.

**EXAMPLE 15.4(e) *Mutually exciting model for electric signals data.*** The model used here is straightforward; it is a version of the mutually exciting *Lin-Lin* model described, for example, in Utsu and Ogata (1997). It points to some of the practical issues which may arise in attempting to develop explanatory models for spatially distributed data. The data consisted of a roughly 25-year list of times and locations of moderate size earthquakes,  $M \geq 4$ , occurring within a 200 km radius of Beijing, together with records of ultra-low frequency electric signals from five stations at essentially arbitrary locations  $z_r$  ( $r = 1, \dots, 5$ ) within the region. To simplify the analysis, both sets of

data were converted to daily  $\{0, 1\}$  values, and modelled as mutually exciting discrete-time point processes. The main question of interest is whether the electric signals show any predictive power.

A general, continuous, version of the model used to model the earthquakes can be represented through a conditional intensity

$$\lambda(t, x) = \mu(x) + \lambda_S(t, x) + \lambda_E(t, x), \quad (15.4.12)$$

where  $\mu(x)$  is an underlying spatial density of background events,  $\lambda_E(t, x)$  and  $\lambda_S(t, x)$  are the earthquake-clustering and signal-generated components of the intensity at time  $t$  and location  $x$ . It is assumed these can be expressed in the form

$$\lambda_S(t, x) = \int_{-\infty}^t h_S(t - u, x - w) N_S(du \times dw),$$

where  $N_S$  is the count process for the target signals, and the response function  $h_S$  can be represented parametrically as (for example) a sum of Laguerre polynomials; there is a similar representation for  $\lambda_E$  via  $h_E$  and  $N_E$ .

The first difficulty that arises is that although the earthquakes are well approximated by a space–time point process, the source of the signals is unknown, as indeed is the physical mechanism which causes them, if indeed they are associated with earthquakes. Thus the only available information about the signals is their occurrence or nonoccurrence at each of the recording stations. A visual inspection of the data suggests at most a loose association between the occurrence of a signal at one of the stations and the distance from that station to any temporally nearby earthquake. This suggests either ignoring the spatial dependence in the response term (making  $h_E$  a function of time only) or possibly introducing a weakly decaying spatial component with contributions from each station,  $h_E(t, x) = f(t) \sum_{r=1}^5 g(\|x - z_r\|)$ , thus assuming some cumulative effect if signals are observed simultaneously at several stations.

In the analysis described in Zhuang *et al.* (2005), locations were ignored in the initial model formulation, but this preliminary model was used as a diagnostic tool to examine the ability of the model to predict earthquake events in different parts of the study region, first using each station in turn as the source of the electric signals data, and comparing the log probability gains for different classes of events, and then using signals from different combinations of the stations. The overall analyses showed that, although earthquake clustering (described by the  $h_E$  term) made the largest contribution to the conditional intensity, the electric signals terms gave significant additional predictive power. The results reinforced the suggestion of a weak distance effect, but suggested that there were other effects, probably of greater importance, which masked any such dependence. On the other hand, if the roles of signals and earthquakes were reversed, the earthquakes showed no significant predictive power for the electric signals.  $\square$

In Section 7.5 we outlined one of the major advantages of the formulation in terms of conditional intensities, namely its role as the basis of simulation and prediction procedures, as well as in estimation and model selection. The simulation procedures apply as well to space–time models, regarded as MPPs, as to simple point processes in time (cf. Algorithm 7.5.V). Here we add a few further comments on the use of conditional intensity and simulation-based methods in model testing.

Leaving aside likelihood ratio tests and associated AIC procedures, which apply to point processes as to any other class of stochastic models, we turn to a class of diagnostic procedures which have been developed more specifically for point processes in recent years, first in the time domain, then more generally.

We outlined Ogata’s residual method, based on the time-change theorem, in Section 7.4. In principle it can be extended to multivariate and MPPs by the extended time-change results of Section 14.6. A more general family of diagnostic tests for space–time processes is based on checking for discrepancies in expressions of the form

$$R(h) \equiv \int_{\mathbb{R} \times \mathbb{R}^2} h(t, \kappa) [N(dt \times d\kappa) - \lambda^*(t, \kappa) dt d\kappa],$$

where  $h(t, \kappa)$  is a bounded, left-continuous, or more generally  $\mathcal{F}$ -predictable function which vanishes outside a bounded set. When  $\lambda^*$  is the true conditional intensity, the terms in square brackets are martingale increments, so any such integral has expected value 0 and from a quadratic variation argument [cf. Exercise 14.1.16 and Zhuang (2006)]

$$\text{var } R(h) = \int_{\mathbb{R} \times \mathbb{R}^2} [h(t, \kappa)]^2 \lambda^*(t, \kappa) dt d\kappa.$$

Consequently,  $R(h)$  can be made the basis for a rough test. Particular choices for  $h(\cdot)$  are typically in the form of a space–time window  $I_{(T_1, T_2] \times A}(t, \kappa)$  weighted to give raw residuals (no further weighting), or Pearson residuals ( $1/\sqrt{\lambda^*(t, \kappa)}$  weighting), or inverse- $\lambda$  residuals ( $1/\lambda^*(t, \kappa)$  weighting), and so on. Corresponding methods for spatial processes, using the Papangelou intensity in place of the conditional intensity, are described in the next section.

In the next section, we describe the corresponding residuals for spatial processes, which use the Papangelou intensity in place of the conditional intensity. In both contexts, the residual methods focus on local discrepancies between the data and the fitted model, such as might not show up in a global analysis or model selection procedure. Many further details are given for spatial processes in Baddeley *et al.* (2005); see also Schoenberg (2002) and Zhuang (2006), who extends the approach to second-order residuals that consider local departures from the expected behaviour of pairs of points.

One drawback to these procedures in their current form is that the data are most commonly compared to a conditional intensity based on the fitted

model, so that the procedures ignore the bias that comes from using the same data for both fitting and testing.

One final procedure we mention is to use the entropy score, or average log likelihood ratio, as a test statistic. For observations over a finite time window, the log likelihood ratio is computed from the data for the given model against a reference model such as a constant rate Poisson process. The same log likelihood ratio is then computed from data simulated over the same window from the given model. From the simulations, a histogram of values for the simulated log likelihood ratios can be obtained, and the observed value (from the data) located within this histogram. If the observed value lies in the extreme tails, the model can be rejected. It is usually preferable to average the likelihood ratio by dividing by the number of data points, or the space–time volume of the observation window, to give an entropy score per observation or per space–time unit, as discussed for example in Daley and Vere-Jones (2004) or Bebbington (2005). This gives a type of overall, portmanteau, test, in contrast to the residual methods which highlight localized areas of disagreement.

We conclude this section with a rather different space–time model, in which the role of time is simply to allow points to grow into sets.

**EXAMPLE 15.4(f) *Lilypond protocol models.*** A diverse range of models in stochastic geometry stem from a protocol initially studied in Häggström and Meester (1996) and Daley, Stoyan, and Stoyan (1999) as germ–grain models (cf. examples of particle systems in Section 6.4). Any particular model is in fact determined totally via a protocol acting on a point process  $N_g$  (in some c.s.m.s.  $\mathcal{X}$ ) so as to construct particles via a uniform growth mechanism that leads to a marked point process  $N = \{(x_i, \kappa_i)\}$  in which the marks  $\kappa_i$  are totally determined by the point set  $N_g$ . This structure contrasts starkly with cluster mechanisms in Section 6.2 and many other MPPs we have studied where the marks are often independent of the process  $N_g$ .

In the simple lilypond germ–grain model of the cited references,  $N_g$  is a stationary Poisson process in  $\mathbb{R}^d$ . At time zero, there start growing  $d$ -dimensional hyperspheres (‘grains’), one around each point as centre, at the same unit rate for each and every grain; any particular grain stops growing when it touches another grain which itself may have ceased growing earlier or else, because it touches the particular grain at the same instant, also then ceases growing. Ultimately, this leads to an MPP  $\{(x_i, \kappa_i)\}$  whose components are, respectively, the centres and radii of the grains. Because every grain touches at least one other grain, questions arise about how far this ‘touching’ mechanism extends, in addition to more obvious questions about the distribution of the size of a typical grain and when each realization of  $N_g$  is a boundedly finite set of points, the fraction of  $\mathbb{R}^d$  that is covered by some grain. The former question leads us to define ‘clusters’ of grains: each grain belongs to a unique cluster consisting of the union of itself and all grains that it touches or are touched by some member of the cluster. Häggström and Meester showed that in  $\mathbb{R}^d$ ,



no matter what finite  $d$ , every cluster is of finite extent; that is, there is no infinite cluster or, in their language, there is no percolation. Daley, Stoyan, and Stoyan report numerical studies of distributional properties of the radii  $\{\kappa_i\}$  for dimensions  $d = 1, 2, 3$ ; Daley, Mallows, and Shepp (2000) give algebraic formulae describing several properties when  $d = 1$  (cf. Exercise 15.4.5), as, for example, when  $N_g$  is a Poisson process on  $\mathbb{R}$  at unit rate, a generic grain has size  $V$  (i.e., length, so  $V_i = 2\kappa_i$  for all  $i$ ) with distribution

$$\Pr\{V > y\} = e^{-y} \exp(e^{-y} - 1). \quad (15.4.13)$$

Such germ–grain models can equally start from ground processes  $N_g$  other than Poisson, and may well have grains that are not necessarily spherical: they could, for example, be similarly oriented hypercubes [Daley (2004)]. Or,  $N_g$  could be some point set that is a minor perturbation of a lattice set, in which case the resulting grains may all be of about the same size. Or, again, for  $\mathcal{X} = \mathbb{R}^2$ , the grains could be randomly oriented finite line-segments  $L_i$  say that are grown centrally about  $x_i$  and cease growth when one of their growing tips contacts another line-segment. Simulation studies of these models with  $N_g$  a Poisson process indicate that the form of the distribution (15.4.13) holds approximately if  $V$  is interpreted as the  $d$ -dimensional volume of the grain or (in the case of line-segments) the area of the circle with the line-segment length as diameter. The fact that such properties persist, albeit approximately, across dimensions and geometrical shapes, points to some further manifestation of the strong influence of ‘pure randomness’ associated with Poisson processes.

Observe that because the grain-radii  $\{\kappa_i\}$  are determinate given the germs  $\{x_i\}$  of  $N_g$ , the only model-fitting that might be possible would concern the distribution of  $N_g$  on the basis of the distribution of these radii. When  $N_g$  is Poisson the only parameter to estimate is its intensity.  $\square$

## Exercises and Complements to Section 15.4

- 15.4.1 (a) In the M/G/ $\infty$  queueing model of Example 15.4(a), Poisson arrivals are served by an unlimited number of servers with common service time distributions. Recast this in the form of the example to obtain expressions for the mean number and p.g.fl. of the number of occupied servers. [Hint: Take  $\mathcal{X}$  to be  $\mathbb{Z}_+$  and identify the delay with the service time distribution.]
- (b) Similarly recast Example 10.4(e) in the form of Example 15.4(a) in the special case when the birth and death probabilities of an individual may depend on the location of the individual but not on the rest of the configuration.
- 15.4.2 (*Continuation*). Extend the arguments for the first moment measure in Example 15.4(a) to find an expression for the second factorial moment measure of the process  $N_t$  on the location space  $\mathcal{X}$  in terms of the second-order moment measures of the underlying process  $N$ . Consider in particular the situation where the locations and lifetimes are independent of each other and i.i.d. [Hint: In any case, the aim is to find an expression for the probability that at time  $t$ , two particles are extant in disjoint subsets  $dx_1, dx_2$  of  $\mathcal{X}$ . This requires both particles to have been born at times  $s_1, s_2$  before  $t$ .]

15.4.3 *Space–time renewal process: moment measures* [cf. Example 15.4(b)].

- (a) Use the discussion in Example 15.4(b) to show that although the process there has stationary ground process, it is not stationary when  $\mathcal{X} = \mathbb{R}^2$ . [Hint: Write the first moment density in the form  $m_g \pi(x | t)$  and show that  $\pi(x | t) \rightarrow 0$  as  $t \rightarrow \infty$ .]
- (b) Show that, by contrast, when  $\mathcal{X} = \mathbb{S}$  (circle) a stationary version of the process does exist. Find the stationary mark distribution, the forms of the quantities  $\check{M}_2^g(du)$  and  $\Pi_2(dx_1 \times dx_2 | u)$  of (15.4.3), and the quantity  $\check{M}_2(du | x_1, x_2)$  of (15.4.4). [Hint: Use the results of Exercise 12.1.4.]

15.4.4 *Space–time cluster models* [special cases of Example 15.4(b)].

- (a) *Neyman–Scott process* [continued from Example 6.3(a)]. Suppose the number of elements in the cluster has first and second factorial moments  $\mu_1$  and  $\mu_{[2]}$ , and that individuals are distributed relative to the cluster centre according to a distribution with density  $f(u, x)$  on  $\mathbb{R} \times \mathbb{R}^2$ , where  $u$  represents the time delay and  $x$  the displacement from the cluster centre. Check that the second cumulant measure for the resultant process still has density given by (6.3.19), namely,

$$\check{c}_{[2]}(u) = \mu_c \mu_{[2]} \int_{\mathbb{R}^d} \check{f}(y + u) \check{f}(y) dy,$$

with suitable interpretations of  $\check{f}$ ,  $y$ , and  $u$ . Find expressions for the mean spatial density and space–time covariance density, and compare with the forms given around Example 15.4.c. [Hint: Here the dependence on locations reduces to a function of the difference  $y_1 - y_2$  between the spatial locations. The process will not be homogeneous unless, in addition, the process of cluster centres is homogeneous.]

- (b) *Bartlett–Lewis process* [continued from Example 6.3(b)]. Suppose the cluster process has the form of a finite space–time random walk as in Example 15.4(b), where the component distributions are constant in time and space. Show that (6.3.23) continues to hold with an appropriate interpretation of the functions  $F$ . Use this to examine the second-order properties as in (a) above.
- (c) Investigate extensions of the previous two examples to situations where the component distributions are location-dependent.

15.4.5 *Lilypond systems*.

- (a) An algebraic definition of a lilypond system  $\{(x_i, \kappa_i): i = 1, 2, \dots\}$  for  $(x_i, \kappa_i) \in \mathbb{R}^d \times \mathbb{R}_+$  is that the pairs should satisfy, with probability one,
- (i)  $\{x_i\}$  is a boundedly finite point set;
  - (ii)  $|\kappa_i + \kappa_j| \leq d(x_i, x_j)$  (all  $i \neq j$ ); and
  - (iii) for every  $i$ , equality holds in (ii) for at least one and at most two  $j$ .
- (b) A lilypond system  $\{(x_i, \kappa_i)\}$  has a *descending chain* if there exists a sequence of germs  $\{x'_j\}$  such that  $|x'_j - x_{j+1}| < |x'_{j-1} - x'_j|$  for  $j = 2, 3, \dots$ . Show that a system that has no descending chain cannot percolate. [Hint: See Daley and Last (2005) and Heveling and Last (2006).]
- (c) In Example 15.4(e) with  $d = 1$  and  $N_g$  a Poisson process at unit rate, first consider  $N_g$  only on  $\mathbb{R}_+$  and let  $Q(t) = \Pr\{\text{origin is uncovered after } t\}$ .

time  $t$  of growth}. Show that  $1 - Q(t) = \int_0^t e^{-u} \cdot e^{-u} Q(u) du$ ; the corresponding differential equation has solution  $Q(t) = \exp(-\frac{1}{2}(1 - e^{-2t}))$ . By considering i.i.d. systems on  $\mathbb{R}_-$  and  $\mathbb{R}_+$ , deduce that for  $N_g$  on the whole of  $\mathbb{R}$ ,  $\Pr\{\text{no grain covers the origin}\} = [Q(\infty)]^2 = e^{-1}$ . The generic grain-length  $V$  at (15.4.13) has mean satisfying  $1 - E(V) = e^{-1}$ : interpret.

## 15.5. The Papangelou Intensity and Finite Point Patterns

As already mentioned more than once in this chapter, the salient difficulty in handling spatial processes is the lack of a time-like dimension through which anything resembling process dynamics can be described. Much effort has gone into circumventing the consequent modelling problems, and in this section we focus on what is perhaps the most important approach, based around the notion of the *Papangelou intensity*  $\rho(y | N)$ , introduced in Papangelou (1974b), following earlier work on lattice processes [see Besag (1974)].

The Papangelou intensity has statistical mechanical connections [see Example 15.5(a) and Section 15.6]. For our purposes here we start from Definition 10.4.I, which characterizes the appearance of  $\rho(\cdot | \cdot)$  in the analysis of finite point patterns; it is given there in terms of the Janossy densities of Section 5.3. For a pattern<sup>2</sup>  $\mathbf{x} = \{x_1, \dots, x_k\}$  containing exactly  $k$  points, we write

$$\rho(y | \mathbf{x}) = \frac{j_{k+1}(y, x_1, \dots, x_k)}{j_k(x_1, \dots, x_k)} = \frac{j_{k+1}(y, \mathbf{x})}{j_k(\mathbf{x})} = \frac{j(y, \mathbf{x})}{j(\mathbf{x})}, \quad (15.5.1a)$$

provided the test location  $y \notin \mathbf{x}$ . Roughly speaking,  $\rho(y | \mathbf{x})$  can be interpreted as the conditional intensity for finding a point of the realization at  $y$ , given the realization of the process throughout the remainder of the state space, which in this section we denote  $\mathcal{W}$  to emphasize that it is most commonly a bounded observation region ( $\mathcal{W}$  for Window) within two- or three-dimensional space. In any case,

$$\begin{aligned} & \Pr\{\text{realization contains } k+1 \text{ particles in } (y, y + dy), \\ & \qquad \qquad \qquad (x_i, x_i + dx_i), (i = 1, \dots, k)\} \\ &= (1 + o(1)) j_{k+1}(y, x_1, \dots, x_k) dy dx_1 \dots dx_k, \\ & \Pr\{\text{realization contains } k+1 \text{ particles in } (y, y + dy), (\mathbf{x}, \mathbf{x} + d\mathbf{x})\} \\ &= (1 + o(1)) j_{k+1}(y, \mathbf{x}) dy d\mathbf{x}, \end{aligned}$$

<sup>2</sup> We make much use of  $\mathbf{x}$  below, as earlier in Section 10.4, where elsewhere in the book we have used  $N$ , stressing the (integer-valued) random measure representation: the choice is a matter of convenience and emphasis towards readability. If needed, we write  $n(\mathbf{x})$  for  $\text{card}(\mathbf{x}) = \#(\mathbf{x})$  without comment. Thus,  $j_k(x_1, \dots, x_k) = j_{n(\mathbf{x})}(\mathbf{x})$  no matter what finite  $k$ , and write this last as  $j(\mathbf{x})$  where no ambiguity may arise, as in  $j(y \cup \mathbf{x}) = j_{n(y \cup \mathbf{x})}(y \cup \mathbf{x})$  and (15.5.1a). Also, in Section 10.4 we called  $\rho(\cdot | \cdot)$  the Papangelou conditional intensity; the briefer term is more convenient here.

with a similar interpretation for  $j_k$ , so in heuristic terms it is clear that (15.5.1a) represents the required conditional density.

In applications to inference, however, we are often concerned precisely with the value of  $\rho$  at observed points; in this case we consider ‘one fewer points’ and define the Papangelou intensity as the function, for  $x_k \in \mathbf{x} = \{x_1, \dots, x_k\}$ , by

$$\rho(x_k \mid \mathbf{x}) = \frac{j_k(x_1, \dots, x_k)}{j_{k-1}(x_1, \dots, x_{k-1})} = \frac{j(\mathbf{x})}{j(\mathbf{x} \setminus x_k)}. \quad (15.5.1b)$$

For an empty realization  $\mathbf{x} = \emptyset$ , we set  $\rho(x \mid \emptyset) = j_1(x)/j_0 = \rho(x \mid x)$ .

The explicit dependence of  $\rho$  on the number  $k$  of points in the realization, which is a feature of its definition via Janossy densities, can be subsumed in its general dependence on the sample, so that  $\rho$  can be regarded as a sample-based function, using either the notation  $\rho(y \mid \mathbf{x})$  as above, or  $\rho(y \mid N)$  when it is more convenient to consider the realization as a counting measure. Keeping to the former notation, and adopting a similar notation for the Janossy measures themselves, we can write the definition of both first- and higher-order Papangelou intensities in the single form [cf. (15.5.1a) and (15.5.1b)], as we use shortly and which perhaps best reveals its underlying character,

$$\rho(\mathbf{y} \mid \mathbf{x}) = \frac{j(\mathbf{x} \cup \mathbf{y})}{j(\mathbf{x} \setminus \mathbf{y})} = \rho(\mathbf{y} \setminus \mathbf{x} \mid \mathbf{x}) \rho(\mathbf{y} \cap \mathbf{x} \mid \mathbf{x}). \quad (15.5.2)$$

EXAMPLE 15.5(a) *Gibbs process with pairwise interactions* [cf. Examples 5.3(c) and 7.1(b)]. In Example 5.3(c) a Gibbs process on a bounded region  $\mathcal{W}$  is specified through Janossy densities of the form (but the present notation)

$$j_{n(\mathbf{x})}(\mathbf{x}) = C \exp \left( \sum_{i=1}^{n(\mathbf{x})} \psi_1(x_i) + \sum_{i=2}^{n(\mathbf{x})} \sum_{j=1}^{i-1} \psi_2(x_i, x_j) \right), \quad (15.5.3)$$

where  $\psi_1(\cdot)$  represents a potential energy due to an external force field and  $\psi_2(\cdot)$  represents an *interaction potential*, whereas the expression in the exponent is some measure of the energy of the system, under equilibrium conditions, when it happens that the system has  $k = N(\mathcal{W}) = n(\mathbf{x})$  particles and  $\mathbf{x}$  describes their locations. For this model we have

$$\begin{aligned} \rho(y \mid \mathbf{x}) &= \exp \left( \psi_1(y) + \sum_{i=1}^{N(\mathcal{W})} \psi_2(y, x_i) \right) & (y \notin \mathbf{x}), \\ \rho(x_j \mid \mathbf{x}) &= \exp \left( \psi_1(x_j) + \sum_{i \neq j} \psi_2(x_j, x_i) \right) & (x_j \in \mathbf{x}). \end{aligned}$$

Similarly, for the second-order Papangelou intensity we have for example for  $y_1, y_2 \notin \mathbf{x}$ ,

$$\rho(y_1, y_2 \mid \mathbf{x}) = \exp \left( \psi(y_1) + \psi(y_2) + \sum_{j=1}^2 \sum_{i=1}^{N(\mathcal{W})} \psi_2(y_j, x_i) + \psi_2(y_1, y_2) \right). \quad \square$$

It is as a sample function that the Papangelou intensity appears most naturally in statistical applications, particularly in questions relating to the likelihood. Given a realization  $\mathbf{x}$ , we know already that the likelihood is just the Janossy density  $j_{n(\mathbf{x})}(\mathbf{x})$ . Taking some arbitrary enumeration  $\{x'_1, \dots, x'_{n(\mathbf{x})}\}$  of the points in the sample, writing  $\mathbf{x}'_r$  for the set of the first  $r$  elements in this enumeration, and repeatedly using the ratio in (15.5.1b), we have

$$\log L = \sum_{i=1}^{n(\mathbf{x})} \log \rho(x'_i \mid \mathbf{x}'_{i-1}) + \log j_0, \quad (15.5.4)$$

This form is quite simple, but it has the disadvantage of treating the observed process as a finite process defined wholly within the observation region. In fact the observations are more commonly the restriction to the observation region of a point process defined within a much larger region. As already discussed around Definition 7.1.II, the Janossy densities that in principle should be used in such a situation are the local Janossy densities of Definition 5.4.IV. Their determination even in such a relatively simple model as the Gibbs process just considered is a formidable task. Even in treating a Gibbs process as a finite point process in its own right, a significant problem arises in the calculation of the last term in (15.5.4), which reduces to the logarithm of the partition function.

For these and other related reasons, parameter estimation via maximization of the exact likelihood of a spatial point process is commonly replaced by the maximization of the pseudolikelihood, introduced already in Example 7.1(b), and defined by analogy with the likelihood<sup>3</sup> for a spatial Poisson process by

$$\log L^\dagger(\mathbf{x}) = \sum_{x_i \in \mathbf{x}} \log \rho(x_i \mid \mathbf{x}) - \int_{\mathcal{W}} \rho(y \mid \mathbf{x}) dy. \quad (15.5.5)$$

Here the difficult term  $j_0$  disappears and the pseudolikelihood is relatively easily calculated for the Gibbs models above and some Markov spatial processes of Section 10.4. See Exercise 15.5.1.

For both time and space-time point processes, replacing the Poisson intensity by the conditional intensity still leads to an exact likelihood, but unfortunately the same is not true for spatial point processes when the Poisson intensity is replaced by the Papangelou intensity. Nevertheless many useful properties have been shown to hold for maximum pseudolikelihood estimates in the point process context, including consistency and asymptotic normality (Section 9.2 of Møller and Waagepetersen (2004) has a good survey). Moreover they are quite easily modified if data from a buffer region are available, much as discussed for estimation of the moment measures in Section 15.2. In such a case the Papangelou intensity is calculated for the realization in the

---

<sup>3</sup> Equation (15.5.5) is correct; the equation for  $L^\dagger(\mathbf{x})$  in Volume I (first impression) p. 217 is incomplete: it is missing a term  $\exp(-\int_{\mathcal{W}} \rho(y \mid \mathbf{x}) dy)$ .

extended region, and this sharpened form for the Papangelou intensity is then substituted into (15.5.5) which is evaluated over the observation region only. Similar modifications can be incorporated for Hanisch-type edge corrections.

One context where this difficulty is avoided is for point processes on a circle or sphere, for here the process can be at once both a.s. finite and stationary (invariant under rotations), and also there are no edges.

EXAMPLE 15.5(b) *Pairwise interaction process on a circle or sphere* [Billiot and Goulard (2001)]. The study quoted in the heading was prompted by a desire to describe and interpret the variation between the angles of the separate strands in the root system of a maize plant. For this purpose, repeated observations on the solid angles between strands were recorded for a sample of maize plants.

The data clearly show some tendency towards regular spacing, so as a basic model the authors suggest a pairwise interaction process as in Example 15.5(a), but assume in addition that the structure is invariant under rotations of the sphere (i.e., it is homogeneous). This implies that the first-order term must reduce to a constant, and that the pairwise interaction term must be a function only of the angle (length of great circle arc)  $\theta_{ij}$  between the pair of points  $x_i, x_j$  under consideration. Now any pair of points can be joined by two great circle arcs, one longer than the other; we suppose in the sequel that  $0 < \theta_{ij} \leq \pi$ , so that the smaller distance (angle) is always chosen. For definiteness we speak of points on a sphere, but it is clear (because no more than a pair of points needs to be considered in the first- and second-order potential functions) that a similar analysis could be carried out for points on a sphere in any finite-dimensional Euclidean space.

Under the additional conditions quoted, the Papangelou intensity of the Gibbs process of Example 15.5(a) takes the form

$$\rho(x \mid \mathbf{x}) = \exp \left( \phi_1 + \sum_{x_i \in \mathbf{x}; x_i \neq x} \phi_2(x - x_i) \right), \quad (15.5.6)$$

where  $x$  is a general point on the sphere,  $\phi_1$  is a constant, and  $\phi_2(\cdot)$  is a function of the angular separation  $\theta$ , and we use  $x - y$  to denote the angular separation of the points  $x, y$ . Billiot and Goulard suggest a semiparametric approach, expanding  $\phi_2$  in a finite cosine series

$$\phi_2(\theta) = \sum_{m=1}^M \alpha_m \cos m\theta,$$

with the first-order term  $\phi_1$ , and  $M$  and the coefficients  $\alpha_m$  to be determined.

A number of difficulties arise in this example. If only one realization is available, it is clearly necessary to suppose  $M \ll N(\mathbb{S})$ , the number of points in the realization. This suggests using an exact likelihood approach and using a criterion such as AIC to determine  $M$ . The main difficulty here is in the calculation of the normalization constant, although with  $M$  small enough this may still be feasible through numerical or Monte Carlo simulation. In the

situation considered by Billiot and Goulard, which is probably more typical of biological models, the repeat samples provide a greater abundance of data, and some version of the pseudolikelihood or Takacs–Fiksel methods (see the comments below and Exercise 15.5.2) looks more attractive. In fact, the authors allude to some difficulties with the usual pseudolikelihood estimates [see Billiot (1994)] and propose instead a variation based on a discrete approximation scheme derived from the pooled data from all sample members. We refer to their paper for computational and numerical details.  $\square$

Further properties of the Papangelou intensity for finite spatial point patterns can be deduced quite readily from the definition in terms of Janossy densities at (15.5.2) which, it will be noted, encompasses Papangelou intensities of higher order. We review these properties as a prelude to discussing the general case in Sections 15.6–7. As always, we suppose that the definition of Papangelou intensities is restricted to situations where the denominators in the ratios at (15.5.2) are positive. The properties of greatest interest, which indicate the main results to be expected in the general case, are set out in the Proposition below. We assume that the point process is regular (Definition 7.1.I) so that Janossy densities exist for all orders.

**Proposition 15.5.I.** *Let  $N$  be a regular, finite point process defined on an observation space  $\mathcal{W} \subset \mathbb{R}^2$ . Then the Papangelou intensities for  $N$  satisfy the following relationships.*

(i) Multiplicative relation: *For mutually disjoint  $\mathbf{x}, \mathbf{u}, \mathbf{v}$ ,*

$$\rho(\mathbf{u} \cup \mathbf{v} \mid \mathbf{x}) = \rho(\mathbf{u} \mid \mathbf{x} \cup \mathbf{v}) \rho(\mathbf{v} \mid \mathbf{x}), \quad (15.5.7a)$$

*and, more generally, with  $\rho(\mathbf{y} \mid \mathbf{x}) = j(\mathbf{x} \cup \mathbf{y})/j(\mathbf{x} \setminus \mathbf{y})$  as in (15.5.2),*

$$\begin{aligned} \rho(\mathbf{u} \cup \mathbf{v} \mid \mathbf{x}) &= \rho(\mathbf{u} \mid \mathbf{x} \cup \mathbf{v}) \rho(\mathbf{v} \setminus \mathbf{u} \mid \mathbf{x} \setminus \mathbf{u}) \\ &= \rho((\mathbf{u} \cup \mathbf{v}) \setminus \mathbf{x} \mid \mathbf{x}) \rho((\mathbf{u} \cup \mathbf{v}) \cap \mathbf{x} \mid \mathbf{x}). \end{aligned} \quad (15.5.7b)$$

(ii) Conditional probability interpretation: *For any bounded Borel set  $B \subset \mathcal{W}$ ,  $\mathbf{y} = \{y_1, \dots, y_{n(\mathbf{y})}\} \subset B$  and  $\mathbf{x} = \{x_1, \dots, x_{n(\mathbf{x})}\} \subset B^c$ , and with  $B^c N$  denoting  $N$  restricted to  $B^c$ ,  $\rho(\mathbf{y} \mid \mathbf{x}) d\mathbf{y}$  equals*

$$\frac{\Pr\{N(B) = n(\mathbf{y}) \text{ and } N(dy_j) = 1 \ (j = 1, \dots, n(\mathbf{y})) \mid B^c N = \sum_{i=1}^{n(\mathbf{x})} \delta_{x_i}\}}{\Pr\{N(B) = 0 \mid B^c N = \sum_{i=1}^{n(\mathbf{x})} \delta_{x_i}\}}. \quad (15.5.8)$$

(iii) Relation to Palm densities:

$$\rho(\mathbf{y} \mid \mathbf{x}) = q(\mathbf{x} \mid \mathbf{y})/q(\emptyset \mid \mathbf{y}), \quad (15.5.9)$$

*where  $q(\mathbf{x} \mid \mathbf{y})$  is the Janossy density on  $(\mathcal{W} \setminus \mathbf{y})^U$  of the higher-order Palm distribution  $\mathcal{P}_{\mathbf{y}}^{(l)}(\cdot)$  defined as in Exercise 13.1.11.*

(iv) Integral relations: for bounded Borel functions  $f(\mathbf{y}, \mathbf{x})$ ,

$$\int_{\mathcal{W}^\cup \times \mathcal{W}^\cup} f(\mathbf{y}, \mathbf{x}) \rho(\mathbf{y} \mid \mathbf{x}) j(\mathbf{x}) d\mathbf{x} d\mathbf{y} = \int_{\mathcal{W}^\cup \times \mathcal{W}^\cup} f(\mathbf{y}, \mathbf{x}) j(\mathbf{y} \cup \mathbf{x}) d\mathbf{x} d\mathbf{y}. \quad (15.5.10)$$

PROOF. The multiplicative relations at (15.5.7a) and (15.5.7b) are direct consequences of the definition at (15.5.2).

The probabilistic interpretation (15.5.8) becomes clear once it is recalled that the Janossy densities specify precisely the number of points in the realization. Thus, if the realization  $\mathbf{x}$  has  $k$  points and is therefore specified by  $j_k(\mathbf{x})$ , and if  $\mathbf{x} \subset B^c$ , there can be no other points in  $\mathcal{W}$  and hence none in  $B$ ; similarly, if the realization has  $j+k$  points in all and  $k$  of them are in  $B^c$ , then there must be exactly  $j$  in  $B$ .

To understand (15.5.9), consider first the Janossy density for the usual (first-order) Palm distribution for a finite point process, supposing that the realization consists of a point at the origin  $y$  and a vector  $\mathbf{x}$  of dimension  $n(\mathbf{x})$  of further points. In terms of the Janossy densities of the original process, we can write

$$q(\mathbf{x} \mid y) = j(\mathbf{x} \cup y) / m(y) \quad (y \notin \mathbf{x}),$$

where  $m(y)$  is the local mean density at  $y$  (i.e., the density of the first-order moment measure), which can be written here as

$$m(y) = \sum_{k=1}^{\infty} \frac{1}{k!} \int_{\mathcal{X}^{(k)}} j_{k+1}(x_1, \dots, x_k, y) \prod_{i=1}^k dx_i = \int_{\mathcal{W}^\cup} \frac{j(\mathbf{x} \cup y)}{n(\mathbf{x})!} d\mathbf{x}. \quad (15.5.11)$$

Because also  $q(\emptyset \mid y) = j_1(y) / m(y)$ , equation (15.5.9) in the case  $\ell = 1$  follows from

$$\rho(y \mid \mathbf{x}) = j(\mathbf{x} \cup y) / j(y) = q(\mathbf{x} \mid y) / q(\emptyset \mid y).$$

The higher-order expressions can be derived in a similar way. From this point of view, the Papangelou intensities are rescaled versions of the Janossy densities of the Palm distribution.

The final integral relation also has a portmanteau character, as it subsumes a series of relationships for fixed  $k$ , each of which follows directly from the ratio form of the definition in (15.5.2).  $\square$

By recasting the integrals in the last of these properties in terms of expectations, we recover the Georgii–Nguyen–Zessin formula [Georgii (1976), Nguyen and Zessin (1979b)] for finite point processes. It is a key point of the theory and lies behind many of the statistical applications of the Papangelou intensity.

**Proposition 15.5.II** (Georgii–Nguyen–Zessin Equation). *Let  $N$  be a finite point process on  $\mathcal{W}$  satisfying the conditions of Proposition 15.5.I,  $K$  a fixed positive integer, and  $h(\cdot, \cdot)$  a nonnegative, measurable, integrable function*



of  $(\mathbf{u}, \mathbf{v})$ , where  $\mathbf{u}$  has dimension  $K$ . For a given realization  $\mathbf{x}$  of  $N$ , with  $\text{card}(\mathbf{x}) \geq K$ , partition  $\mathbf{x}$  into components  $\mathbf{u}$  and  $\mathbf{v} = \mathbf{x} \setminus \mathbf{u}$  with  $\text{card}(\mathbf{u}) = K$ . Let  $N^{[K]}(\cdot)$  denote the modified form of the product counting measure omitting terms along the diagonals. Then, setting both expressions under the expectation signs equal to zero when the realization contains fewer than  $K$  points,

$$\mathbb{E} \left[ \int_{\mathcal{W}^{(K)}} h(\mathbf{u}, \mathbf{x} \setminus \mathbf{u}) N^{[K]}(d\mathbf{u}) \right] = \mathbb{E} \left[ \int_{\mathcal{W}^{(K)}} h(\mathbf{u}, \mathbf{x}) \rho(\mathbf{u} \mid \mathbf{x}) d\mathbf{u} \right]. \quad (15.5.12)$$

PROOF. We again start with the first-order result ( $K = 1$ ), assuming that  $N$  has a mean density  $m(x)$  as in (15.5.11) and Papangelou intensity  $\rho(x \mid \mathbf{y})$ . Then, with  $n(\mathbf{v}) = k - 1$ , the left-hand side of (15.5.12) can be written as

$$\mathbb{E} \left[ \sum_{u \in \mathbf{x}} h(u, \mathbf{x} \setminus u) \right] = \sum_{k=1}^{\infty} \frac{1}{k!} \int_{\mathcal{W} \times \mathcal{W}^{(k-1)}} k j_k(u \cup \mathbf{v}) h(u, \mathbf{v}) du d\mathbf{v}, \quad (15.5.13)$$

where the multiplier  $k$  arises because each element of  $\mathbf{x}$  can appear in turn as  $u$ , giving (from symmetry) equal contributions to the integral. For  $u \in \mathbf{v}$ ,  $j_k(u \cup \mathbf{v}) = \rho(u \mid \mathbf{v}) j_{k-1}(\mathbf{v})$ , so that writing  $\ell = k - 1$ , the expectation becomes

$$\sum_{\ell=0}^{\infty} \frac{1}{\ell!} \int_{\mathcal{W}} \int_{\mathcal{W}^{(\ell)}} \rho(u \mid \mathbf{v}) h(u, \mathbf{v}) j_{\ell}(\mathbf{v}) d\mathbf{v} du = \mathbb{E} \left[ \int_{\mathcal{W}} \rho(u \mid \mathbf{x}) h(u, \mathbf{x}) du \right].$$

In the higher-order forms, the combinatorial factor  $k^{[K]}$  replaces the term  $k$  but the argument is otherwise similar.  $\square$

The Georgii–Nguyen–Zessin formula was originally given for a finite Gibbs point process. The setting above, for a regular finite point process, is in fact no more general as Example 5.3.7 shows. The formula appears in the literature in a variety of guises, mostly associated with the fact that the left-hand side of (15.5.12) can be written in terms of the modified Campbell measure [Definition 13.1.I(b)]. Thus, for  $k = 1$ ,

$$\mathbb{E} \left[ \int_{\mathcal{W}} h(u, N \setminus u) N(du) \right] = \int_{\mathcal{W} \times \mathcal{N}_{\mathcal{W}}^{\#*}} h(u, N) C_{\mathcal{P}}^! (du \times dN)$$

so that in this case (15.5.12) becomes

$$\int_{\mathcal{W} \times \mathcal{N}_{\mathcal{W}}^{\#*}} h(u, N) C_{\mathcal{P}}^! (du \times dN) = \mathbb{E} \left[ \int_{\mathcal{W}} h(u, N) \rho(u \mid N) du \right], \quad (15.5.14)$$

immediately suggesting a definition of the Papangelou intensity as a Radon–Nikodym derivative, an approach we take up in Section 15.6.

Furthermore, if the first moment measure  $M$  exists, then the modified Campbell measure of Definition 13.1.I(b) can in turn be expressed in terms of modified local Palm distributions  $\mathcal{P}_u^!$ , namely,

$$\int_{\mathcal{W} \times \mathcal{N}_{\mathcal{W}}^{\#*}} h(u, N) C_{\mathcal{P}}^!(du \times dN) = \int_{\mathcal{W} \times \mathcal{N}_{\mathcal{W}}^{\#*}} h(u, N) \mathcal{P}_u^!(dN) M(du).$$

Finally, if the process is stationary—strictly speaking this is impossible in the context of this section unless the state space is the circle or other compact group, but is covered in the more general context of the next section—then  $M(du) = m du$  and the left-hand side simplifies further as in (13.2.5), yielding

$$\mathbb{E} \left[ \int_{\mathcal{W}} h(u, N) \rho(u | N) du \right] = m \int_{\mathcal{W} \times \mathcal{N}_{\mathcal{W}}^{\#*}} h(u, N) \mathcal{P}_0^!(dS_{-u}N) \ell(du).$$

If  $h(u, \mathbf{v})$  is a function of  $u$  only, another, suggestive, way of writing (15.5.12) in the case  $k = 1$  is as

$$\mathbb{E} \left[ \int_{\mathcal{W}} h(u) [N(du) - \rho(u | \mathbf{v}) du] \right] = 0, \quad (15.5.15)$$

so that in this sense the *residual process* [‘innovations process’ in Baddeley *et al.* (2005)],

$$\nu(B) = \int_B [N(du) - \rho(u | \mathbf{v}) du], \quad (15.5.16)$$

plays a role analogous to that of the martingale  $N - A$  in the temporal case. Note, however, that we cannot extend (15.5.15) to general functions  $h(u, \mathbf{v})$  with complete impunity, just because the form in which  $h$  enters the left- and right-hand sides of (15.5.12) is different. This constraint is analogous to the requirement that in the relation (14.2.3) defining the compensator as the dual predictable projection, we cannot replace the predictable process  $Y(\cdot)$  in that equation by a completely arbitrary process. Evidently, the necessary and sufficient condition to justify the replacement is that

$$\mathbb{E} \left[ \int_{\mathcal{W}} h(u, N) N(du) \right] = \mathbb{E} \left[ \int_{\mathcal{W}} h(u, N) \rho(u | N) du \right].$$

This idea is developed more fully in Section 15.7 where the concept of *exvisibility* is introduced as a spatial analogue of predictability. Meanwhile we use Proposition 15.5.II and (15.5.15) to establish the properties of some inference procedures for spatial point patterns in circumstances where  $h(u, \mathbf{v})$  satisfies the extended form of (15.5.15), namely,

$$\mathbb{E} \left[ \int_{\mathcal{W}} h(u, \mathbf{v}) [N(du) - \rho(u | \mathbf{v}) du] \right] = \mathbb{E} \left[ \int_{\mathcal{W}} h(u, \mathbf{v}) \nu(du) \right] = 0. \quad (15.5.15')$$

The class of such functions includes not only functions of  $u$  only, but also functions where the dependence on  $\mathbf{v}$  enters only through the conditional intensity  $\rho(u \mid \mathbf{v})$  (see Corollary 15.7.VI).

We mention first the *Takacs–Fiksel estimation procedure* [Takacs (1983), Fiksel (1988)], which may be regarded as a variation on the method of moments adapted to estimating the parameters in a spatial point process. It is based on equating the two sides of (15.5.12), or one of its equivalent forms such as (15.5.14), for selected functions  $h(u, \mathbf{v})$ , using the locations of the observed points to evaluate the left-hand side, and the parametric form for the Papangelou intensity, also dependent on the sample through the second element  $\mathbf{v}$  in  $h$ , to evaluate the right-hand side. Enough functions are selected to determine the parameters uniquely. The proper choice of estimating functions requires some care, and depends on the structure of the model. Edge effects again constitute a significant nuisance factor. The techniques have been applied especially for Gibbs and other Markov spatial point processes, because it is for just such processes that the Papangelou intensity provides a natural and accessible characterisation of the process. The case of pairwise interactions in Example 15.5(b) is typical. Diggle *et al.* (1994) has a useful review from an applied statistics perspective; Billiot (1997) establishes consistency and related asymptotic properties of the estimates, including also the case of repeated samples.

We turn next to diagnostic tests for spatial point pattern models. These have been developed recently [see, in particular, Baddeley *et al.* (2005) for an excellent review with examples and illustrations], and are based on the fact that the expectation of (15.5.16) is zero. Thus a significant deviation from zero of the sample equivalent of the integral in (15.5.15) or (15.5.15') can be used as an indication of a departure from the model under test.

Typically, the function  $h$  in such diagnostic tests has the form  $I_A(u)g(u, \mathbf{v})$  where  $A \in \mathcal{B}_W$  is a subset of the observation region, and  $g$  is interpreted as a weight function characterizing one of several possible types of residuals. Significance of the observed value of the resulting sample function can be determined approximately by comparing the sample value to its estimated standard deviation. The next lemma, where  $\nu$  is defined as in (15.5.16), is key both to such tests and to determining approximate confidence regions for the Takacs–Fiksel procedure.

**Lemma 15.5.III.** *Let  $h$  be as in Proposition 15.5.II and such that (15.5.15') holds. Then*

$$\begin{aligned} \text{var} \left( \int_W h(u, \mathbf{v}) \nu(du) \right) &= \mathbb{E} \left[ \int_W [h(u, \mathbf{v})]^2 \rho(u \mid \mathbf{v}) du \right] \\ &+ \mathbb{E} \left[ \int_{W \times W} h(u_1, \mathbf{v} \cup u_2) [h(u_2, \mathbf{v} \cup u_1) - 2h(u_2, \mathbf{v})] \rho(u_1, u_2 \mid \mathbf{v}) du_1 du_2 \right] \\ &+ \mathbb{E} \left[ \int_{W \times W} h(u_1, \mathbf{v}) h(u_2, \mathbf{v}) \rho(u_1 \mid \mathbf{v}) \rho(u_2 \mid \mathbf{v}) du_1 du_2 \right]. \end{aligned}$$

PROOF. Write  $\text{var} \left( \int_{\mathcal{W}} h(u, \mathbf{v}) \nu(du) \right) = E[X^2] + E[Y^2] - 2E[XY]$ , where  $X = \int_{\mathcal{W}} h(u, \mathbf{v}) N(du)$  and  $Y = \int_{\mathcal{W}} h(u, \mathbf{v}) \rho(u | \mathbf{v}) du$ . To evaluate  $E[X^2]$  we use the second-order version of Proposition 15.5.II, using as argument in the left-hand side of (15.5.12) the function of  $u_1, u_2$  and  $\mathbf{v} \setminus \{u_1, u_2\}$  equal to

$$h(u_1, \mathbf{v} \setminus u_1) h(u_2, \mathbf{v} \setminus u_2),$$

which gives

$$\begin{aligned} E \left[ \sum_{i \neq j} h(x_i, \mathbf{x} \setminus x_i) h(x_j, \mathbf{x} \setminus x_j) \right] \\ = E \left[ \int_{\mathcal{W} \times \mathcal{W}} h(u_1, \mathbf{v} \cup u_2) h(u_2, \mathbf{v} \cup u_1) \rho(u_1, u_2 | \mathbf{v}) du_1 du_2 \right]. \end{aligned}$$

The sum on the left-hand side omits the terms for which  $i = j$ ; these can be evaluated directly using (15.5.12) as

$$E \left( \sum_i [h(x_i, \mathbf{x} \setminus x_i)]^2 \right) = E \left[ \int_{\mathcal{W}} [h(u, \mathbf{v})]^2 \rho(u | \mathbf{v}) du \right].$$

Also

$$E[Y^2] = E \left[ \int_{\mathcal{W} \times \mathcal{W}} h(u_1, \mathbf{v}) h(u_2, \mathbf{v}) \rho(u_1 | \mathbf{v}) \rho(u_2 | \mathbf{v}) du_1 du_2 \right].$$

To evaluate the product term we write

$$\begin{aligned} E[XY] &= E \left[ \int_{\mathcal{W}} h(u_1, \mathbf{x} \setminus u_1) N(du_1) \int_{\mathcal{W}} h(u_2, \mathbf{x}) \rho(u_2 | \mathbf{x}) du_2 \right] \\ &= E \left[ \int_{\mathcal{W}} h(u_1, \mathbf{x}) \rho(u_1 | \mathbf{x}) du_1 \int_{\mathcal{W}} h(u_2, \mathbf{x} \cup u_1) \rho(u_2 | \mathbf{x} \cup u_1) du_2 \right] \end{aligned}$$

on appealing to both (15.5.15') and (15.5.12) in the last step. The double integral can be written as a product integral over  $\mathcal{W} \times \mathcal{W}$ , and the product of the terms  $\rho(\cdot)$  equals  $\rho(u_1, u_2 | \mathbf{v})$  by the multiplicative relation (15.5.7a). Collecting terms completes the proof.  $\square$

Residuals over a set  $A \subset \mathcal{W}$  correspond to setting  $h(u, \mathbf{v}) = I_A(u) g(u, \mathbf{v})$ , with  $g(u, \mathbf{v}) \equiv 1$  for *raw residuals*,  $g(u, \mathbf{v}) = 1/\rho(u | \mathbf{v})$  for *Stoyan–Grabarnik* or *inverse  $\lambda$  residuals*, and  $g(u, \mathbf{v}) = 1/\sqrt{\rho(u | \mathbf{v})}$  for *Pearson residuals*. When (15.5.15') holds, these residuals have zero means; their variances (cf. Lemma 15.5.III) are detailed in Example 15.5(c) and Exercise 15.5.5. When the model is parametrized by  $\theta$ , *pseudoscore residuals* are obtained by taking  $g_\theta(u, \mathbf{v}) = \partial \log \rho_\theta(u | \mathbf{v}) / \partial \theta$  (the name comes from the analogy with the score statistic in conventional statistical analysis). To see this, take the partial derivative in  $\theta$  of the log pseudolikelihood as in (15.5.5) of a realization of a point process with model parametrized by  $\theta$ ; straightforward algebra shows that this pseudoscore is expressible

$$\int_{\mathcal{W}} \frac{\partial}{\partial \theta} \log \rho(u | \mathbf{x}) \nu(du), \quad (15.5.17)$$

which is of the form of the integral in (15.5.15'). Accordingly, this quantity is called a pseudoscore residual; we denote it (with integral over  $A$ ) by  $R_\psi(A)$ .

EXAMPLE 15.5(c) *Stoyan–Grabarnik residuals* [Stoyan and Grabarnik (1991)]. We can write the residual  $R_{SG}(A)$  as

$$\sum_{x_i \in \mathbf{x} \cap A} \frac{1}{\rho(x_i \mid \mathbf{x} \setminus x_i)} - \int_A \frac{\rho(u, \mathbf{x})}{\rho(u, \mathbf{x})} du = \sum_{x_i \in \mathbf{x} \cap A} \frac{1}{\rho(x_i \mid \mathbf{x} \setminus x_i)} - \ell(A).$$

Using Lemma 15.5.III, its variance is given by

$$\begin{aligned} \text{var } R_{SG}(A) &= \text{var} \left( \sum_{x_i \in \mathbf{x} \cap A} \frac{1}{\rho(x_i \mid \mathbf{x} \setminus x_i)} \right) \\ &= \mathbb{E} \left[ \int_A \frac{1}{\rho(u \mid \mathbf{v})} du \right] + \mathbb{E} \left[ \int_{A \times A} \frac{\rho(u_1 \mid \mathbf{v}) \rho(u_2 \mid \mathbf{v}) - \rho(u_1, u_2 \mid \mathbf{v})}{\rho(u_1, u_2 \mid \mathbf{v})} du_1 du_2 \right]. \end{aligned}$$

In simple cases, we can evaluate this expression explicitly. For example, when  $N$  is Poisson,  $\rho(u \mid \mathbf{v}) = \lambda(u)$  independently of the rest of the realization  $\mathbf{v}$ , and because the second-order term  $\rho(u_1, u_2 \mid \mathbf{v}) = \lambda(u_1)\lambda(u_2)$ , the expression reduces to

$$\text{var } R_{SG}(A) = \int_A \frac{du}{\lambda(u)}.$$

For a Gibbs process with pairwise interactions as in Example 15.5(a),

$$\begin{aligned} \text{var } R_{SG}(A) &= \mathbb{E} \left[ \int_A \exp \left( -\psi_1(u) - \sum_{i=1}^{N(A)} \psi_2(u, x_i) \right) du \right] \\ &\quad - \int_{A \times A} [1 - e^{-\psi_2(u_1, u_2)}] du_1 du_2. \end{aligned}$$

It should be noted that in practice, the expressions used for the Papangelou intensities will be estimates only, whereas the above expressions assume the exact forms are known. This must result in some reduction of the variances; Baddeley *et al.* (2005) refer to the exact forms as *innovations* and reserve the term *residuals* for the estimated forms, giving several examples to illustrate the differences in the variances.  $\square$

## Exercises and Complements to Section 15.5

### 15.5.1 Papangelou intensities and pseudolikelihood ratios.

- Evaluate the Papangelou intensity (i) for a simple Poisson process over  $\mathcal{W} \subset \mathbb{R}^2$ ; and (ii) for a Poisson process with linear spatial drift. Write down the pseudolikelihood ratio for testing the one against the other, and show that it coincides with the ordinary likelihood ratio.
- Extend to the likelihood ratio for a pairwise interaction process, as in Example 15.5(a), including a drift, against the alternative of a Poisson process with drift.

15.5.2 *Second- and higher-order correlation functions; factorial cumulant densities.*

In the physics literature, ‘correlation’ has a wider meaning than that of a standardized centred bivariate product moment as in probability and statistics. For example, the two-point correlation function in Martínez and Saar (2002) is precisely the second-order product density  $m_{[2]}(\cdot, \cdot)$  of Section 5.4, and Buchler *et al.* (1998) speak of long-range correlations and the correlation length. Martínez and Saar discuss both a range of estimation procedures for this pair correlation function and higher-order correlation functions (these coincide with higher-order factorial product densities  $m_{[k]}(\cdot)$ ) that have been used in astrophysical studies.

15.5.3 *Covariance of residuals.* Let  $f$ ,  $g$ , and  $h$  be nonnegative measurable functions satisfying appropriate integrability conditions.

(a) Imitate the proof of Lemma 15.5.III to show that

$$\begin{aligned} \text{cov} \left( \sum_i h(x_i, \mathbf{x} \setminus x_i), \sum_j g(x_j, \mathbf{x} \setminus x_j) \right) \\ = \mathbb{E} \left[ \int_{\mathcal{W}} h(u, \mathbf{x}) g(u, \mathbf{x}) \rho(u \mid \mathbf{x}) du \right] \\ + \mathbb{E} \left[ \int_{\mathcal{W} \times \mathcal{W}} h(u, \mathbf{x} \cup v) g(v, \mathbf{x} \cup u) \rho(u, v \mid \mathbf{x}) du dv \right] \\ - \mathbb{E} \left[ \int_{\mathcal{W}} h(u, \mathbf{x}) \rho(u \mid \mathbf{x}) du \right] \mathbb{E} \left[ \int_{\mathcal{W}} g(v, \mathbf{y}) \rho(v \mid \mathbf{y}) dv \right]. \end{aligned}$$

Conclude that  $\text{var} \left( \int_{\mathcal{W}} h(u, \mathbf{x}) \nu(du) \right)$  equals

$$\begin{aligned} \text{var} \left( \int_{\mathcal{W}} h(u, \mathbf{x}) N(du) \right) + \left( \mathbb{E} \left[ \int_{\mathcal{W}} h(u, \mathbf{x}) \rho(u \mid \mathbf{x}) du \right] \right)^2 \\ + \mathbb{E} \left[ \int_{\mathcal{W} \times \mathcal{W}} h(u, \mathbf{x}) h(v, \mathbf{x}) \rho(u \mid \mathbf{x}) \rho(v \mid \mathbf{x}) du dv \right] \\ - 2 \mathbb{E} \left[ \int_{\mathcal{W} \times \mathcal{W}} h(u, \mathbf{x}) h(v, \mathbf{x} \cup u) \rho(u, v \mid \mathbf{x}) du dv \right]. \end{aligned}$$

(b) Use this expression to find the covariance of two integrals of the form  $\int_{\mathcal{W}} h(u, \mathbf{x}) \nu(du)$ ,  $\int_{\mathcal{W}} g(v, \mathbf{x}) \nu(dv)$ , and write down conditions for the two integrals to be uncorrelated.

(c) Show also that  $\text{var} \left( \sum_i h(x_i, \mathbf{x} \setminus x_i) - \int_{\mathcal{W}} f(u, \mathbf{x}) du \right)$  equals

$$\begin{aligned} \text{var} \left( \sum_i h(x_i, \mathbf{x} \setminus x_i) \right) + \int_{\mathcal{W} \times \mathcal{W}} \text{cov}(f(u, \mathbf{x}), f(v, \mathbf{x})) du dv \\ - 2 \int_{\mathcal{W}} \mathbb{E} \left[ h(u, \mathbf{x}) \rho(u \mid \mathbf{x}) \int_{\mathcal{W}} (f(v, \mathbf{x} \cup u) - \mathbb{E}[f(v, \mathbf{y})]) dv \right] du \end{aligned}$$

[Hint: Recall that  $\text{var} \left( \int_A X(u) du \right) = \int_{A \times A} \text{cov}(X(u), X(v)) du dv$ .]

15.5.4 *Takacs–Fiksel Method* [see also Stoyan and Stoyan (1994, pp. 330–331)].

(a) Write down the Takacs–Fiksel equations in the special case of a Gibbs process for which the potential function  $U(\cdot)$  [see Example 5.3(c)] is  
(i) limited to first- and second-order interaction terms; and then either  
(ii) spatially homogeneous, or (iii) isotropic.

- (b) Check the specific form of the estimating equations when the Gibbs process is of either hard-core or general Strauss form, and

$$h_k(x, \mathbf{v}) = N(S_{r_k}(x)) \exp[U(x, \mathbf{v})],$$

for an increasing sequence of radii  $r_k$ . [Hint: See Fiksel (1988).]

- (c) Use the results of Exercise 15.5.3(b) to find two uncorrelated test statistics to use in the Takacs–Fiksel method for estimating parameters in a hard-core model.

#### 15.5.5 Residual variances [Baddeley *et al.* (2007)].

- (a) Use Lemma 15.5.III and  $\nu(\cdot)$  as in Example 15.5(c) to verify the following expressions for the residual variances.

(i) *Raw residual*:

$$\begin{aligned} \text{var } \nu(A) &= \mathbb{E} \left[ \int_A \rho(u \mid N) \, du \right] \\ &\quad + \mathbb{E} \left[ \int_{A \times A} [\rho(u_1 \mid N) \rho(u_2 \mid N) - \rho(u_1, u_2 \mid N)] \, du_1 \, du_2 \right]. \end{aligned}$$

(ii) *Pearson residual*  $R_{\text{Psn}}(A) = \int_A [1/\sqrt{\rho(u \mid N)}] \nu(du)$ :

$$\begin{aligned} \text{var } R_{\text{Psn}}(A) &= \ell(A) \\ &\quad + \int_{A \times A} \mathbb{E} \left[ \frac{(\sqrt{\rho(u_1 \mid N) \rho(u_2 \mid N)} - \sqrt{\rho(u_1, u_2 \mid N)})^2}{\sqrt{\rho(u_1 \mid N) \rho(u_2 \mid N)}} \right] \, du_1 \, du_2. \end{aligned}$$

- (b) Show that for a Poisson process with intensity  $\lambda(x)$ ,

$$\text{var } \nu(A) = \int_A \lambda(u) \, du, \quad \text{and} \quad \text{var } R_{\text{Psn}}(A) = \ell(A),$$

and that when  $\lambda(u) = e^{\theta\eta(u)}$  the pseudoscore statistic  $R_\psi(A)$  at (15.5.17) has

$$\text{var } R_\psi(A) = \int_A \eta^2(u) e^{\theta\eta(u)} \, du.$$

Investigate the form of these three variances for a Gibbs process with pairwise interactions [cf. Examples 15.5(a), (c)].

## 15.6. Modified Campbell Measures and Papangelou Kernels

The material in these last two sections may be regarded as an introduction to the general theory to be found, for example, in MKM (1982, Chapter 9) and subsequent developments in Kallenberg (1983a, Chapters 12–14) (some misprints are corrected in the 1986 reprinting). Connections with statistical mechanics are explored in Nguyen and Zessin (1979b), Matthes, Warmuth,

and Mecke (1979), Glötzl (1980), Rauchenschwandtner (1980), and Georgii (1988), among others. Our treatment has been much influenced by Kallenberg's (1978) work, especially the informal account in Kallenberg (1984).

In the discussion of finite spatial point processes in Section 15.5, the results are essentially extensions and refinements of the basic theory outlined in Chapter 5. In particular this is true of processes defined via interaction potentials as in Example 15.5(a): in the finite case such a description is equivalent to a description in terms of Janossy densities [see Example 5.3(c) and Exercise 5.3.7]. The situation changes radically, however, if we try to specify the distribution of an infinite particle system in terms of interaction potentials. Certainly, equations such as (15.5.3) can no longer be used to describe the distribution of the point process, because the Janossy densities, which exist in a local sense, no longer exist in a global sense, and in any case sums such as those in the exponent of (15.5.3) in general diverge for  $k \rightarrow \infty$ .

The approach adopted by physicists in this situation has been to take a bounded subset,  $B$  say, of the state space, and to suppose that the particles in  $B$  are in equilibrium with both the 'external' interaction forces exerted on them by the particles outside  $B$  and the 'internal' interaction forces generated among themselves. Equations such as (15.5.3) can then be used to describe the conditions for local equilibrium, conditional on the configuration of particles outside  $B$ . Taking expectations over all possible exterior configurations leads to a family of balance equations that must be satisfied by the overall equilibrium distribution of the process, if in fact such a distribution exists (and this last proviso is an important qualifier).

For example, suppose that exactly one particle lies in  $B$  and that the process is to be specified through one- and two-point interaction potentials  $\psi_1$  and  $\psi_2$  as in Example 15.5(a). Then, conditional on the external configuration, which we denote  $B^cN$ , the local Janossy densities for the process on  $B$  must be of the form

$$j_1^B(y \mid B^cN) = C_B \exp \left( \psi_1(y) + \sum_{x_i \in \text{supp}(B^cN)} \psi_2(y, x_i) \right) \quad (y \in B), \quad (15.6.1)$$

where the normalization constant is now  $C_B$ , and  $\int_{B^c} \psi_2(y, x) dN(x)$  is required to converge for all  $N$  in the space  $\mathcal{N}^{\#*}(B^c)$  of boundedly finite simple point processes with support in  $B^c$ . To obtain a convenient form for the balance equations, multiply by a function  $f(y, \eta)$  mapping  $(y, \eta) \in B \times \mathcal{N}^{\#*}(B^c)$  into  $\mathbb{R}$  and take expectations over  $B^cN$ . This leads to

$$\begin{aligned} & \int_{\mathcal{N}^{\#*}(B^c)} \int_B f(y, \eta) j_1^B(y \mid \eta) \mathcal{P}_{B^c}(d\eta) \\ &= \int_{\mathcal{N}^{\#*}(\mathcal{X})} \int_B f(y, B^cN) N(dy) I_{\{N(B)=1\}}(N) \mathcal{P}(dN), \end{aligned} \quad (15.6.2)$$

where  $\mathcal{P}$  is the equilibrium distribution assumed to exist for the process as a whole, and  $\mathcal{P}_{B^c}$  is its projection onto  $\mathcal{N}^{\#*}(B^c)$ .



Given functions  $\psi_i(\cdot)$ , it is far from obvious whether there exists an equilibrium distribution  $\mathcal{P}$  satisfying (15.6.2) and related equations. Indeed, it is nothing other than a general version of the *Ising problem*, which in its original and special form referred to the existence of an equilibrium distribution for a process on a one-dimensional lattice specified by interactions of pairs of points involving only nearest neighbours. The general problem, even in the lattice case, is still unsolved, although many partial results are available [see, e.g., Preston (1976) for a general formulation of the Ising problem and a review of results known then]. In particular, it is known that even when an equilibrium distribution exists, it may not be unique (leading to the possibility of ‘phase transitions’), and that even when the interaction potentials are spatially stationary in character (i.e., they depend only on the relative positions of the particles), the resulting solutions may not be stationary (this is known as ‘symmetry breakdown’).

Within the class of Gibbs processes specified by pairwise interaction potentials [Example 15.5(a)], the existence and uniqueness of stationary versions of the process are known under rather strong constraints, in particular a sufficient condition is the existence of a finite interaction range  $R$  such that  $\psi_2(x_i, x_j) = 0$  for  $\|x_i - x_j\| > R$ . It is not our intention, however, to pursue the Ising problem as such, but rather to consider the question of how quantities analogous to the Papangelou intensities of Section 15.5 can be introduced and related to other characteristics of the point process.

One special feature of (15.6.2) is worth noting at this stage. For given location  $y$  and  $\{x_i: i = 1, 2, \dots\}$ , the function at (15.6.1) is dependent on the set  $B$  only through the normalization constant  $C_B$  and the requirements that  $y \in B$ ,  $x_i \in B^c$  ( $i = 1, 2, \dots$ ). We are therefore led to let  $B \downarrow \{y\}$  to recover, in the finite case, the functions  $\rho(y | \mathbf{x})$  that we could there interpret as the conditional intensity for the occurrence of a particle at  $y$  given the realization of the process throughout the remainder of the state space, that is, in  $\mathcal{X} \setminus \{y\}$ . From the interaction potential viewpoint, the term in the exponent at (15.6.2) can be related to the work required to introduce a new particle into the position  $y$  keeping the locations of the existing particles fixed. But although these ideas lead to a straightforward definition in the finite case, the situation is more complicated in the general case. It is necessary to distinguish three different random measures, each of which embodies some aspect of the conditional intensity  $\rho(\cdot | \cdot)$  of Section 15.5. These quantities are as follows.

- (i) The *Papangelou kernel*  $R(\cdot | \cdot)$  defined by integral relations extending (15.5.10).
- (ii) A random measure  $\pi(\cdot)$  describing, loosely speaking, the *atomic part* of these kernels.
- (iii) The *Papangelou intensity measures*  $\zeta(\cdot)$  as originally introduced by Papangelou (1974b) in terms of the limit

$$\zeta(B) = \lim_{n \rightarrow \infty} \sum_{i=1}^{k_n} \mathbb{E}(N(I_{ni}) | I_{ni}^c N), \quad (15.6.3)$$

where  $\mathcal{T} = \{\{I_{ni}: i = 1, \dots, k_n\}: n = 1, 2, \dots\}$  is a fixed dissecting system of partitions of  $B$  as in Definition A1.6.I.

In all approaches to this topic, Condition  $\Sigma$  below plays a fundamental role, and we work under it unless explicitly stated otherwise.

**Definition 15.6.I.** *The simple point process on the c.s.m.s.  $\mathcal{X}$  satisfies Condition  $\Sigma$  if for all bounded Borel sets  $B$ ,*

$$\mathcal{P}\{N(B) = 0 \mid B^c N\} > 0 \quad \text{a.s.} \quad (15.6.4)$$

This requirement generalizes the assumption of Section 15.5 that the Janossy densities be positive everywhere. Its essential role is to preclude situations where the behaviour inside  $B$  is deterministically controlled by the behaviour outside  $B$ , as can occur in (1°) and (2°) of Example 15.6(a).

EXAMPLE 15.6(a) *On Condition  $\Sigma$ .* (1°) Let  $N$  be a point process on  $\mathcal{X}$  for which  $\mathcal{P}\{N(\mathcal{X}) = r\} = 1$  for some fixed integer  $r$ . Then for any nonempty set  $A \in \mathcal{B}_{\mathcal{X}}$ ,  $\mathcal{P}\{N(A) = 0 \mid N(A^c) < r\} = 0$ , and thus Condition  $\Sigma$  is violated.

(2°) Let  $N_1$  be a point process with exactly one point uniformly distributed over the bounded state space  $\mathcal{X} \in \mathcal{B}(\mathbb{R}^d)$ , for example, a circle of unit area. Let  $N_2$  be a Poisson process at unit rate on  $\mathcal{X}$  with  $N_2$  independent of  $N_1$ . Then the point process  $N = N_1 + N_2$  violates Condition  $\Sigma$  because for any Borel set  $A \subseteq \mathcal{X}$  of positive Lebesgue measure,

$$\mathcal{P}\{N(A) = 0 \mid N(A^c) = 0\} = 0.$$

(3°) With  $N_1$  and  $N_2$  as in (2°), the process  $N$  equal to  $N_1$  with probability  $p$  and to  $N_2$  with probability  $q = 1 - p$ , with  $pq > 0$ , satisfies Condition  $\Sigma$  (details are left to the reader).  $\square$

In order to set down a general form of the integral equations (15.5.10) for the Papangelou kernel, denoted by  $R(A \mid N)$  for  $A \in \mathcal{B}_{\mathcal{X}}$  and  $N \in \mathcal{N}_{\mathcal{X}}^{\#*}$ , we start from (15.5.12) and (15.5.14), treating only the first-order case. As in (15.5.14), we can rewrite the left-hand side of this equation in terms of modified Campbell measure  $C_{\mathcal{P}}^!$  of Definition 13.1.I(b) to yield

$$\begin{aligned} \int_{\mathcal{X} \times \mathcal{N}_{\mathcal{X}}^{\#*}} h(u, N) C_{\mathcal{P}}^!(du \times dN) &\equiv \int_{\mathcal{X} \times \mathcal{N}_{\mathcal{X}}^{\#*}} h(u, N - \delta_u) N(du) \mathcal{P}(dN) \\ &= \int_{\mathcal{X} \times \mathcal{N}_{\mathcal{X}}^{\#*}} h(u, N) \rho(u \mid N) du \mathcal{P}(dN) \\ &= \int_{\mathcal{X} \times \mathcal{N}_{\mathcal{X}}^{\#*}} h(u, N) R(du \mid N) \mathcal{P}(dN). \end{aligned} \quad (15.6.5)$$

For the finite case considered in Proposition 15.5.III, this exhibits  $R(\cdot \mid N) du$  as being derived from a disintegration of the second component of  $C_{\mathcal{P}}^!$  with respect to  $\mathcal{P}$ , namely, for all bounded  $A \in \mathcal{B}_{\mathcal{X}}$  and  $U \in \mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#*})$ ,

$$\int_U R(A \mid N) \mathcal{P}(dN) = C_{\mathcal{P}}^!(A \times U), \quad (15.6.6)$$

and leads us to seek such a disintegration in general.

To justify such a disintegration, we use the absolute continuity condition that

$$C_{\mathcal{P}}^!(A \times \cdot) \ll \mathcal{P}(\cdot) \quad \text{for each fixed } A \in \mathcal{B}_{\mathcal{X}}. \quad (15.6.7)$$

The proof of this condition in Lemma 15.6.III below gives us an immediate illustration of the role that Condition  $\Sigma$  plays. The proof also makes use of the following results, in which  $\sigma\{N\}, \sigma\{B^c N\}$  denote the  $\sigma$ -algebras generated by random variables  $\{N(A), A \in \mathcal{B}_{\mathcal{X}}\}, \{N(A), A \in \mathcal{B}_{B^c}\}$ , and so on.

**Lemma 15.6.II.** (a) For any  $U \in \sigma\{N\} \subseteq \mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#*})$ , and any  $B \in \mathcal{B}_{\mathcal{X}}$ , there exists  $U^* \in \sigma\{B^c N\}$  such that

$$U \cap \{N(B) = 0\} = U^* \cap \{N(B) = 0\}. \quad (15.6.8)$$

(b) For any  $\sigma\{N\}$ -measurable function  $g(N)$ , and any  $B \in \mathcal{B}_{\mathcal{X}}$ , there exists a  $\sigma\{B^c N\}$ -measurable function  $g_0(N)$  such that

$$g(N) I_{\{N(B)=0\}}(N) = g_0(N) I_{\{N(B)=0\}}(N). \quad (15.6.9)$$

Furthermore, if  $E|g(N)| < \infty$ , then for any bounded  $\sigma\{B^c N\}$ -measurable  $Y$ ,

$$E(Y(N) g(N) I_{\{N(B)=0\}}(N)) = E(Y(N) g_0(N) \mathcal{P}\{N(B) = 0 \mid B^c N\}). \quad (15.6.10)$$

PROOF. Any  $U^* \in \sigma\{B^c N\}$  is generated by sets of the form  $\{N(A_i) = j_i: A_i \in \mathcal{B}(B^c), \text{ nonnegative integers } j_i\}$ , and for any  $U \in \sigma\{N\}$ ,  $U \cap \{N(B) = 0\}$  is generated by sets of the form  $\{N(B) = 0, N(A_i) = j_i: A_i \in \mathcal{B}(B^c), j_i = 0, 1, \dots\}$ . This proves (a).

Part (b) follows from (a) by a standard extension argument: start from indicator functions  $I_U(N)$  for  $U \in \sigma\{N\}$ , for which  $I_{U \cap \{N(B)=0\}} = I_{U^*} I_{\{N(B)=0\}}$  and  $I_{U^*}$  is  $\sigma\{B^c N\}$ -measurable.  $\square$

**Lemma 15.6.III.** Let  $N$  be a simple point process on the c.s.m.s.  $\mathcal{X}$  satisfying Condition  $\Sigma$ . Then for all bounded  $A \in \mathcal{B}_{\mathcal{X}}$ , the absolute continuity condition (15.6.7) holds.

PROOF. We have to show that for any  $U \in \mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#*})$  such that  $\mathcal{P}(U) = 0$ ,  $C_{\mathcal{P}}^!(A \times U) = 0$  (all bounded  $A \in \mathcal{B}_{\mathcal{X}}$ ).

Suppose first that  $U \subseteq \{N(A) = 0\}$ , so that  $U = U \cap \{N(A) = 0\}$ , and thus by Lemma 15.6.II(a),  $I_U(N) = I_{U^*}(N) I_{\{N(A)=0\}}(N)$  for some  $U^* \in \sigma\{A^c N\}$ . Noting that for  $y \in A$ ,  $N - \delta_y \in U^*$  if and only if  $N \in U^*$  (i.e., the behaviour of  $N$  inside  $A$  is irrelevant to whether  $N \in U^*$ ),

$$\begin{aligned} C_{\mathcal{P}}^*(A \times U) &= \int_{\mathcal{N}_{\mathcal{X}}^{\#*}} \int_A I_U(N - \delta_y) N(dy) \mathcal{P}(dN) \\ &= \int_{\mathcal{N}_{\mathcal{X}}^{\#*}} \int_A I_{\{N(A)=0\}}(N - \delta_y) N(dy) I_{U^*}(N) \mathcal{P}(dN) \\ &= \int_{\mathcal{N}_{\mathcal{X}}^{\#*}} I_{\{N(A)=1\}}(N) I_{U^*}(N) \mathcal{P}(dN) \leq \mathcal{P}(U^*). \end{aligned}$$

Equally, using (15.6.10),

$$0 = \mathcal{P}(U) = \mathbf{E}(I_{U^*}(N) \mathcal{P}\{N(A) = 0 \mid A^c N\}).$$

By Condition  $\Sigma$ , the coefficient of the bounded function  $I_{U^*}(N)$  is positive a.s., so we have a contradiction unless  $I_{U^*}(N) = 0$  a.s.; that is,  $\mathcal{P}(U^*) = 0$ , and hence  $C_{\mathcal{P}}^1(A \times U) = 0$  for such  $U$ .

Suppose next that  $\mathcal{P}(U) = 0$  for some  $U \subseteq \{N: N(A) \leq k\}$  for some given integer  $k \geq 1$ , and let  $\mathcal{T} = \{\{A_{ni}: i = 1, \dots, k_n\}: n = 1, 2, \dots\}$  be a dissecting family of partitions for  $A$ . Write also

$$U_{ni} = U \cap \{N(A_{ni}) = 0\} \quad \text{and} \quad U'_{ni} = U \setminus U_{ni}.$$

Then for  $n = 1, 2, \dots$  we have

$$\begin{aligned} C_{\mathcal{P}}^1(A \times U) &= \sum_{i=1}^{k_n} C_{\mathcal{P}}^1(A_{ni} \times U) = \sum_{i=1}^{k_n} [C_{\mathcal{P}}^1(A_{ni} \times U_{ni}) + C_{\mathcal{P}}^1(A_{ni} \times U'_{ni})] \\ &= \sum_{i=1}^{k_n} C_{\mathcal{P}}^1(A_{ni} \times U'_{ni}) \end{aligned}$$

because by the earlier argument,  $C_{\mathcal{P}}^1(A_{ni} \times U_{ni}) = 0$ . For the last sum write

$$\begin{aligned} \sum_{i=1}^{k_n} C_{\mathcal{P}}^1(A_{ni} \times U'_{ni}) &= \sum_{i=1}^{k_n} \int_{\mathcal{N}_{\mathcal{X}}^{\#}} \int_{A_{ni}} I_{U \setminus U_{ni}}(N - \delta_y) N(dy) \mathcal{P}(dN) \\ &\leq k \sum_{i=1}^{k_n} \mathcal{P}\{N(A_{ni}) \geq 2, N(A) \leq k\} \end{aligned}$$

(for this last step,  $N(A_{ni}) \leq N(A) \leq k$  for  $N \in U$ , and any  $y \in A_{ni}$  that is an atom of  $N$  can contribute to the integral only if also  $(N - \delta_y)(A_{ni}) \geq 1$ , and thus  $N(A_{ni}) \geq 2$  for such  $y$ ).

The assumption that  $N$  is simple implies that the last sum  $\rightarrow 0$  as  $n \rightarrow \infty$ , and hence that  $C_{\mathcal{P}}^1(A \times U) = 0$ .

To complete the proof, use monotone convergence to deduce that, whenever  $\mathcal{P}(U) = 0$ ,

$$C_{\mathcal{P}}^1(A \times U) = \lim_{k \rightarrow \infty} C_{\mathcal{P}}^1(A \times (U \cap \{N: N(A) \leq k\})) = 0. \quad \square$$

Standard arguments based on this result can now be used to establish the existence and uniqueness properties of the disintegration of the modified Campbell measure, leading to the following theorem whose proof is left to Exercise 15.6.1. Kallenberg (1983a, Chapter 13) gives a more extended treatment; part (iv) of the theorem follows Glötzl (1980).

**Theorem 15.6.IV.** Suppose given a simple point process  $N$  defined on the c.s.m.s.  $\mathcal{X}$ , with probability measure  $\mathcal{P}$ , and satisfying Condition  $\Sigma$ . Then there exists a unique kernel  $R(A | N)$  satisfying

- (i) for each bounded  $A \in \mathcal{B}_{\mathcal{X}}$ ,  $R(A | \cdot)$  is a Borel-measurable function on  $\mathcal{N}_{\mathcal{X}}^{\#*}$ ;
- (ii) for each  $N \in \mathcal{N}_{\mathcal{X}}^{\#*}$ ,  $R(\cdot | N)$  is a bounded finite Borel measure on  $\mathcal{B}_{\mathcal{X}}$ ; and
- (iii) for all nonnegative, measurable functions  $h(u, N): \mathcal{X} \times \mathcal{N}_{\mathcal{X}}^{\#*} \mapsto \mathbb{R}_+$ , vanishing for  $u$  outside a bounded Borel set of  $\mathcal{X}$ ,

$$\int_{\mathcal{X} \times \mathcal{N}_{\mathcal{X}}^{\#*}} h(u, N) C_{\mathcal{P}}^!(du \times dN) = \int_{\mathcal{X} \times \mathcal{N}_{\mathcal{X}}^{\#*}} h(u, N) R(du | N) \mathcal{P}(dN). \quad (15.6.11)$$

(iv) If also

$$C_{\mathcal{P}}^!(du \times dN) \ll \ell \times \mathcal{P} \quad \text{on} \quad \mathcal{B}_{\mathcal{X}} \times \mathcal{B}_{\mathcal{N}_{\mathcal{X}}^{\#*}}, \quad (15.6.12)$$

then there exists a  $\mathcal{B}_{\mathcal{X}} \times \mathcal{B}_{\mathcal{N}_{\mathcal{X}}^{\#*}}$ -measurable function  $\rho(u | N)$  such that (15.6.5) holds and for all  $A \in \mathcal{B}_{\mathcal{X}}$ ,

$$R(A | N) = \int_A \rho(x | N) \ell(dx) \quad (\mathcal{P}\text{-a.s. in } N). \quad (15.6.13)$$

**Definition 15.6.V.** When they exist, the kernel  $R(\cdot | \cdot)$  defined by (15.6.11) of Theorem 15.6.IV is the Papangelou kernel associated with the point process  $N$  of the Theorem, and the density  $\rho(x | N)$  of (15.6.13) is the Papangelou intensity of  $N$ .

Strictly speaking, the kernel  $R(\cdot | \cdot)$  is the first-order Papangelou kernel associated with  $N$ , because higher-order kernels can be defined via higher-order modified Campbell measures  $C_{\mathcal{P},k}^!$ . These we now introduce and discuss briefly, setting

$$\begin{aligned} C_{\mathcal{P},k}^!(A_1 \times \cdots \times A_k \times U) \\ = \int_{\mathcal{N}_{\mathcal{X}}^{\#}} \int_{A_1 \times \cdots \times A_k} I_U \left( N - \sum_{i=1}^k \delta_{y_i} \right) N^{[k]}(dy_1 \times \cdots \times dy_k) \mathcal{P}(dN) \end{aligned} \quad (15.6.14)$$

for bounded  $A_1, \dots, A_k \in \mathcal{B}_{\mathcal{X}}$  and  $U \in \mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#*})$ , where  $N^{[k]}(\cdot)$  denotes the  $k$ -fold factorial product measure defined by  $N(\cdot)$  as above Proposition 9.5.VI. Much as in Lemma 15.6.III, it can be shown that under Condition  $\Sigma$ ,  $C_{\mathcal{P},k}^!(A_1 \times \cdots \times A_k \times \cdot) \ll \mathcal{P}(\cdot)$ , and hence a  $k$ th-order Papangelou kernel  $R_k(A_1 \times \cdots \times A_k | N)$  is well-defined  $\mathcal{P}$ -a.s. for bounded  $A_1, \dots, A_k \in \mathcal{B}_{\mathcal{X}}$  by

$$\begin{aligned} \int_U R_k(A_1 \times \cdots \times A_k | N) \mathcal{P}(dN) \\ = C_{\mathcal{P},k}^!(A_1 \times \cdots \times A_k \times U) \quad (\text{all } U \in \mathcal{B}(\mathcal{N}_{\mathcal{X}}^{\#*})). \end{aligned} \quad (15.6.15)$$

Furthermore, regarding  $C_{\mathcal{P},k+\ell}^!$  as a measure on  $\mathcal{X}^{(\ell)} \times (\mathcal{X}^{(k)} \times \mathcal{N}_{\mathcal{X}}^{\#\#})$ , it can be shown that  $C_{\mathcal{P},k+\ell}^! \ll C_{\mathcal{P},k}^!$  with respect to subsets of  $\mathcal{X}^{(k)} \times \mathcal{N}_{\mathcal{X}}^{\#\#}$ . The corresponding kernel can be identified, up to the usual equivalence, with the kernel function

$$R_{\ell} \left( A_1 \times \cdots \times A_{\ell} \mid N + \sum_{i=1}^k \delta_{x_i} \right),$$

thus justifying the extension of the multiplicative relation (15.5.7) to the form (using the simplicity of  $N$  in an essential way)

$$\begin{aligned} R_{\ell} \left( dy_1 \times \cdots \times dy_{\ell} \mid N + \sum_{i=1}^k \delta_{x_i} \right) R_k(dx_1 \times \cdots \times dx_k \mid N) \\ = R_{k+\ell}(dy_1 \times \cdots \times dy_{\ell} \times dx_1 \times \cdots \times dx_k \mid N). \end{aligned} \quad (15.6.16)$$

Finally, defining  $R_0(\cdot \mid N) = 1$  for  $N(\mathcal{X}) = 0$  and zero otherwise, the Papangelou kernels of all orders can be combined into a portmanteau kernel on the space  $\mathcal{X}^{\cup*}$  when  $N(\mathcal{X}) < \infty$  via the equation

$$G(V \mid N) = \sum_{k=0}^{\infty} R_k(V \cap \mathcal{X}^{(k)} \mid N) / k! \quad (V \in \mathcal{B}(\mathcal{X}^{\cup*})). \quad (15.6.17)$$

Under Condition  $\Sigma$ , this *Gibbs kernel*  $G(\cdot \mid \cdot)$  is a density for the portmanteau Campbell measure defined much as in (15.6.14) but allowing  $k$  to vary, so the resultant set  $A$  is any Borel subset of  $\mathcal{X}^{\cup*}$ . Many properties of the Papangelou kernels can be assumed under a general treatment of the Gibbs kernel: see Kallenberg (1983a, Chapter 13) for details. Moreover, we can take the disintegrations the other way, assuming for any given  $k$  that the  $k$ th-order factorial moment measure exists, and disintegrating the  $k$ th-order Campbell measure relative to this moment measure in just the same way as for the ordinary Palm measures in Chapter 13. Then all these disintegrations can be combined to give a decomposition of the portmanteau Campbell measure into a family of Palm measures and an associated portmanteau factorial moment measure. Again we refer to Kallenberg (1983a, 1984) for further details.

We conclude this section with another property of the first-order Papangelou kernel; it is an extension of the conditional probability relation (15.5.8).

**Proposition 15.6.VI.** *Let  $N$  be a simple point process defined on the c.s. m.s.  $\mathcal{X}$  satisfying Condition  $\Sigma$ . Then for any bounded  $A, B \in \mathcal{B}_{\mathcal{X}}$  with  $A \subseteq B$ ,*

$$R(A \mid N) = \frac{\mathcal{P}\{N(A) = N(B) = 1 \mid B^c N\}}{\mathcal{P}\{N(B) = 0 \mid B^c N\}} \quad \text{a.s. on } \{N: N(B) = 0\}. \quad (15.6.18)$$

PROOF. In (15.6.5) substitute

$$h(u, N) = I_A(u) I_{\{N(B)=0\}}(N) I_U(N + \delta_u),$$

where  $U \in \sigma\{B^c N\}$ . When  $u \in A \subseteq B$ ,  $N + \delta_u \in U$  if and only if  $N \in U$ , so (15.6.5) yields

$$\begin{aligned} \int_U R(A | N) I_{\{N(B)=0\}}(N) \mathcal{P}(dN) &= \int_U N(A) I_{\{N(B)=1\}}(N) \mathcal{P}(dN) \\ &= \int_U E(N(A) I_{\{N(B)=1\}} | B^c N) \mathcal{P}(dN) \\ &= \int_U \mathcal{P}\{N(A) = N(B) = 1 | B^c N\} \mathcal{P}(dN). \end{aligned}$$

On the other hand, using Lemma 15.6.II(b), noting that  $U \in \sigma\{B^c N\}$ , we can write

$$E(R(A | N) I_{\{N(B)=0\}}(N)) = E(R_0(A | N) \mathcal{P}\{N(B) = 0 | B^c N\}),$$

where  $R_0(A | N)$  is  $\sigma\{B^c N\}$ -measurable. Because we have equality for all  $U$  on  $\{N(B) = 0\}$ ,

$$R_0(A | N) \mathcal{P}\{N(B) = 0 | B^c N\} = \mathcal{P}\{N(A) = N(B) = 1 | B^c N\}.$$

Moreover, by Condition  $\Sigma$ , the coefficient of  $R_0(A | N) > 0$  a.s., and on  $N(B) = 0$ ,  $R_0(A | N) = R(A | N)$ , so (15.6.18) follows.  $\square$

## Exercises and Complements to Section 15.6

15.6.1 Prove Theorem 15.6.IV. [Hint: Cf. the proof of Lemma 15.6.III.]

## 15.7. The Papangelou Intensity Measure and Exvisibility

In this final section we turn to an investigation of the two other measures,  $\pi$  and  $\zeta$ , referred to in the discussion around (15.6.3), and their relation to the first-order Papangelou kernel  $R(\cdot | \cdot)$ . Theorem 15.6.IV(iv) exhibits the Papangelou intensity as the Radon–Nikodym derivative w.r.t. Lebesgue measure of the kernel  $R(\cdot | \cdot)$ , under the condition that  $C_{\mathcal{P}}^1$  is absolutely continuous with respect to both  $\ell \times \mathcal{P}$  and  $\mathcal{P}$  (the latter property holds by definition). Atoms of  $\pi$  are precisely what hinders the existence of a Papangelou intensity in general. If the atomic component is absent, then the Papangelou intensity of the previous section coincides with the intensity defined via exvisibility, that is, as a density of  $\zeta$ .

Thus far, we have considered  $R(\cdot | N)$  as a kernel on the canonical space  $\mathcal{N}_{\mathcal{X}}^{\#*}$ . It is more convenient for this section to regard  $R(A | N)$  as defined on the space  $(\Omega, \mathcal{E}, \mathcal{P})$  itself. By an abuse of notation we sometimes write

$$R(A) = R(A, \omega) = R(A | N(\omega)), \quad (15.7.1)$$

and treat the quantity on the left-hand side as a *random measure*. That it is

a random measure follows from the measurability of  $R(A \mid N)$  as a function of  $N$ , which implies that  $R(A)$  is a random variable for each  $A \in \mathcal{B}_{\mathcal{X}}$  and hence that the requirements of Proposition 9.1.VIII are satisfied.

As we show shortly, this random measure has some properties in common with the compensator for temporal processes, but difficulties arise with the atoms in particular, and we start on a different tack, following broadly the approach adopted by Papangelou (1974b) and Kallenberg (1983a).

Suppose given a fixed dissecting system of partitions for  $\mathcal{X}$  [recall (15.6.3)]  $\mathcal{T} = \{\{I_{nj}: j = 1, \dots, k_n\}, n = 1, 2, \dots\}$ . We need the following lemma.

**Lemma 15.7.I.** *For bounded Borel sets  $A, B$  with  $A \subseteq B$ ,*

$$\mathcal{P}\{\cdot \mid A^c N\} = \frac{\mathcal{P}\{\cdot \cap \{N(B \setminus A) = 0\} \mid B^c N\}}{\mathcal{P}\{N(B \setminus A) = 0 \mid B^c N\}} \quad \text{a.s. on } \{N(B) = 0\}, \quad (15.7.2)$$

the denominator being a.s. positive on  $\{N(B) = 0\}$ .

PROOF.  $E[I_{\{N(B)=0\}}(N) \mathcal{P}\{N(B) = 0 \mid B^c N\}] = E[\mathcal{P}\{N(B) = 0 \mid B^c N\}]$  a.s. If  $N \in \{N(B) = 0\}$ , then either  $N$  is in a set of measure zero, or else  $\mathcal{P}\{N(B) = 0 \mid B^c N\} > 0$ , so that in either case the last probability is a.s. positive on  $\{N(B) = 0\}$ , and therefore the denominator in (15.7.2) is a.s. positive on  $\{N(B) = 0\}$ . The relation itself is just a version of  $P(U \mid V \cap W) = P(U \mid W)/P(V \mid W)$  for  $V \subseteq W$ , where because  $A \subseteq B$ , we can take  $P$  to be  $\mathcal{P}(\cdot \mid B^c N)$  and  $W$  to be  $\{N(B \setminus A) = 0\}$ .  $\square$

In what follows, we take  $A, B$  to be elements of  $\mathcal{T}$  and note that, because  $\mathcal{T}$  is countable, we can assume that (15.7.2) holds simultaneously a.s. for all such choices of  $A, B \in \mathcal{T}$ . First we examine any atomic component of  $R$ .

**Proposition 15.7.II.** *Let  $N$  be a simple point process on the c.s.m.s.  $\mathcal{X}$ . Let  $\mathcal{T}$  be a dissecting system of partitions of  $\mathcal{X}$ ,  $x$  a general point of  $\mathcal{X}$ , and  $\{I_n(x)\}$  a sequence of elements of  $\mathcal{T}$  with  $I_n(x) \downarrow \{x\}$  ( $n \rightarrow \infty$ ). Then the limit*

$$\pi\{x\} = \lim_{n \rightarrow \infty} \mathcal{P}\{N(I_n(x)) \geq 1 \mid I_n^c(x)N\} \quad (15.7.3)$$

exists a.s., is independent of  $\mathcal{T}$ , and can be identified a.s. on the set  $\{N(I_n(x) \setminus \{x\}) = 0\}$  with the ratio

$$\pi\{x\} = \frac{\mathcal{P}\{N\{x\} = N(I_n(x)) = 1 \mid I_n^c(x)N\}}{\mathcal{P}\{N(I_n(x) \setminus \{x\}) = 0 \mid I_n^c(x)N\}}, \quad (15.7.4)$$

the ratio being interpreted as 1 when both numerator and denominator vanish.

The equation

$$\pi(A) = \pi(A, N) \equiv \sum_{x_i \in \text{supp}(N) \cap A} \pi\{x_i\} = \sum \pi\{x_i\} \delta_{x_i}(A) \quad (15.7.5)$$

defines a random measure on  $\mathcal{X}$ .



When Condition  $\Sigma$  holds,  $\pi\{x\} < 1$  for all  $x$ , the atoms of  $\pi$  include the atoms of  $R$ , and

$$R\{x\} = \pi\{x\} / (1 - \pi\{x\}) \quad (15.7.6)$$

unless  $N\{x\} = 1$ , in which case  $R(\{x\}) = 0$ .

PROOF. In Lemma 15.7.I take  $A = I_n(x)$ ,  $B = I_m(x)$  with  $n > m$ . Omitting the dependence on  $x$  for notational brevity, (15.7.2) implies that a.s. on  $\{N(I_m) = 0\}$ ,

$$\begin{aligned} \mathcal{P}\{N(I_n) > 0 \mid I_n^c N\} &= 1 - \mathcal{P}\{N(I_n) = 0 \mid I_n^c N\} \\ &= 1 - \frac{\mathcal{P}\{N(I_m) = 0 \mid I_n^c N\}}{\mathcal{P}\{N(I_m \setminus I_n) = 0 \mid I_n^c N\}}. \end{aligned} \quad (15.7.7)$$

For increasing  $n$ , the numerator here remains fixed, and the denominator decreases monotonically to the limit  $\mathcal{P}\{N(I_m \setminus \{x\}) = 0 \mid I_m^c N\}$ . We deduce that the limit exists a.s. on the set  $\{N(I_m \setminus I_n) = 0\}$  for every  $n > m$ , hence a.s. on  $\{N(I_m \setminus \{x\}) = 0\}$ , and hence a.s. because  $N(I_m \setminus \{x\}) \rightarrow 0$  a.s. as  $m \rightarrow \infty$ . Also the ratio equals (15.7.4) except possibly for realizations where the denominator vanishes. If the latter holds, the ratio for finite  $n$  will tend to  $\infty$  unless in the limit the numerator also vanishes. In this case, we have  $\mathcal{P}\{N(I_n) = 0 \mid I_n^c N\} \rightarrow 0$ ; that is,  $\mathcal{P}\{N(I_n) \geq 1 \mid I_n^c N\} \rightarrow 1$ , implying that  $\pi\{x\} = 1$ , in accordance with the interpretation here that ' $0/0 = 1$ '.

Next, let  $\mathcal{T}_1$  and  $\mathcal{T}_2$  be dissecting systems with  $\mathcal{T}_1 \subseteq \mathcal{T}_2$ . Then the limits  $\pi_1\{x\}$ ,  $\pi_2\{x\}$  say exist for each system, and taking  $\{I_{n_j}(x)\} \subseteq \mathcal{T}_j$  with  $I_{n_j}(x) \downarrow \{x\}$  ( $n_j \rightarrow \infty$ ) for  $j = 1, 2$ , so  $\{I_{n_1}(x)\} \subseteq \mathcal{T}_1 \subseteq \mathcal{T}_2$ , it follows that  $\pi_1\{x\} = \pi_2\{x\}$  because we can always find  $I_{n'}(x) \downarrow \{x\}$  with successive terms taken alternately from  $\{I_{n_1}\}$  and  $\{I_{n_2}\}$ . In general, any two systems  $\mathcal{T}_1, \mathcal{T}_2$  generate by their intersection a third system  $\mathcal{T}_3$  with  $\mathcal{T}_3 \supseteq \mathcal{T}_1$  and  $\mathcal{T}_3 \supseteq \mathcal{T}_2$ , so  $\pi\{x\}$  is independent of  $\mathcal{T}$ .

For bounded  $A \in \mathcal{B}_{\mathcal{X}}$ ,  $N(A) < \infty$ , so the defining sum at (15.7.5) can be expressed as a finite sum of limits as at (15.7.3) over disjoint sets  $I_{n_j}$  for sufficiently large  $n$ , and thus it is a random variable. From Proposition 9.1.VIII, we deduce that  $\pi(A, N(\omega))$  is a random measure.

When Condition  $\Sigma$  holds, if  $\pi\{x\} = 1$  for some  $x$ , then for this  $x$  and all  $n$

$$\mathcal{P}\{N(I_n(x)) \geq 1 \mid I_n^c(x)N\} = 1 \quad \text{a.s.}$$

on account of the monotonicity of the denominator in (15.7.7). Consequently,  $\mathcal{P}\{N(I_n(x)) = 0 \mid I_n^c(x)N\} = 0$  a.s., contradicting Condition  $\Sigma$ . Thus,  $\pi\{x\} < 1$ .

To demonstrate (15.7.6), refer to (15.6.18). By putting  $A = I_m(x)$  and  $B = I_n(x)$ , deduce that on  $\{N(I_n) = 0\}$ ,

$$R(I_m(x) \mid N) = \frac{\mathcal{P}\{N(I_m) = 1 = N(I_n) \mid I_n^c N\}}{\mathcal{P}\{N(I_n) = 0 \mid I_n^c N\}}.$$

Here,  $R(\cdot | N)$  is a measure and  $I_m(x) \downarrow \{x\}$  as  $m \rightarrow \infty$ , so the left-hand side  $\rightarrow R(\{x\})$ . On the right-hand side,

$$\{N: N(I_m) = 1 = N(I_n)\} \downarrow \{N: N\{x\} = 1 = N(I_n)\},$$

so on  $\{N(I_n) = 0\}$ ,

$$R\{x\} \equiv R(\{x\}) = \frac{\mathcal{P}\{N\{x\} = 1 = N(I_n) \mid I_n^c N\}}{\mathcal{P}\{N(I_n) = 0 \mid I_n^c N\}}.$$

The numerator here coincides with that of (15.7.4), and the denominator equals

$$\mathcal{P}\{N(I_n \setminus \{x\}) = 0 \mid I_n^c N\} - \mathcal{P}\{N(I_n) = N\{x\} \geq 1 \mid I_n^c N\}.$$

Because  $N$  is simple, the last term equals  $\mathcal{P}\{N(I_n) = N\{x\} = 1 \mid I_n^c N\}$ . Then (15.7.6) follows from (15.7.4) whenever  $\{N(I_n) = 0\}$  for sufficiently large  $n$ . On the complementary event,  $x$  is an atom of  $N$  because

$$\bigcap_n \{N(I_n) \geq 1\} = \{N\{x\} \geq 1\}.$$

In this case, we choose some positive integer  $k$  and substitute

$$h(y, N) = I_A(y) I_{\{N(B) \leq k\}}(N) I_{\{N(A) \geq 1\}}(N)$$

in the relation (15.6.5) with  $B = I_n(x)$ ,  $A = I_m(x)$  with  $m \geq n$ . The left-hand side yields

$$E(R(I_m) I_{\{N(I_n) \leq k\}}(N) I_{\{N(I_m) \geq 1\}}(N)),$$

and the right-hand side is bounded above by  $(k+1)\mathcal{P}\{N(I_m) \geq 2, N(I_n) \leq k+1\}$ . Now repeat this with  $I_m$  replaced by a dissecting partition  $\{I_{nj}\}$  for  $I_m$ , and sum over  $j$ . Proceeding to the limit, the sum  $\rightarrow 0$  by simplicity, and the sum of the left-hand side converges to  $\sum R\{x\}N\{x\}$ , the sum being taken over all atoms lying in  $I_m$ . Consequently,  $R\{x\} = 0$  whenever  $N\{x\} > 0$ .  $\square$

The following simple examples may help illustrate the nature of atoms of  $R$  and  $N$ , and especially the role played by  $\pi(\cdot)$ .

EXAMPLE 15.7(a) *On Condition  $\Sigma$  [continued from Example 15.6(a)].* Let  $N$  be a point process on  $\mathcal{X}$  for which  $\mathcal{P}\{N(\mathcal{X}) = r\} = 1$  for some fixed integer  $r \geq 1$  as earlier in (1°). For  $x \in \text{supp}(N)$  and a given dissecting system  $\mathcal{T}$  let  $\{I_n\} \equiv \{I_n(x)\}$  be a sequence of elements of  $\mathcal{T}$  contracting to  $\{x\}$ . Then

$$\begin{aligned} \mathcal{P}\{N(I_n) \geq 1 \mid N(I_n^c) \leq r-1\} &= 1 = \lim_{n \rightarrow \infty} \mathcal{P}\{N(I_n) \geq 1 \mid N(I_n^c) \leq r-1\} \\ &= \pi\{x\}. \end{aligned}$$

The set  $\{N: N(I_n - \{x\}) = 0\} \cap \{N(I_n^c) \leq r-1\}$  consists precisely of those realizations  $N$  for which  $N(I_n^c) = r-1$ ,  $N(I_n) = N\{x\} = 1$ . If we assume that

the points of  $N$  are  $r$  points i.i.d. over  $\mathcal{X}$ , which is a bounded Borel subset of  $\mathbb{R}^d$ —for example, the interior of a circle or a sphere—then assuming the sets  $I_n$  have positive Lebesgue measure, we should have

$$\begin{aligned}\mathcal{P}\{N(I_n) = 1 \mid N(I_n^c) = r - 1\} &= 1 \\ &\neq \mathcal{P}\{N(I_n) = 1 = N\{x\} \mid N(I_n^c) = r - 1\} = 0 \\ &= \mathcal{P}\{N(I_n - \{x\}) = 0 \mid N(I_n^c) = r - 1\}.\end{aligned}$$

Thus, this example justifies the interpretation of  $\pi\{x\} = \pi(\{x\}, N)$  as unity when the expression at (15.7.4) equals ‘0/0’.  $\square$

**EXAMPLE 15.7(b)** *A particular Gauss–Poisson process.* Let there be given a Gauss–Poisson process in the plane, which, in its Poisson cluster process representation, consists of a cluster centre process that is Poisson at unit rate, and for which any point  $x$  in the centre process produces clusters of either zero or one additional points, with probability  $p$  for there being one point, which is then located at  $x + a$  for some fixed position  $a$  relative to the cluster centre. Consider such a process on a bounded subset  $\mathcal{X} \in \mathcal{B}(\mathbb{R}^2)$ . It can be checked that this process satisfies Condition  $\Sigma$ .

Suppose the state space  $\mathcal{X}$  contains all of  $x$ ,  $x - a$  and  $x + a$  for some  $x$ ; consider realizations  $N$  for which  $N\{x\} = 1$ , and let  $\{I_n\} \equiv \{I_n(x + a)\}$  be a sequence of sets of positive Lebesgue measure belonging to some dissecting system for  $\mathcal{X}$  and  $\downarrow \{x + a\}$ . On realizations  $N$  for which  $N\{x\} = 1 = N\{x - a\}$  and  $N(\mathcal{X} \setminus (\{x\} \cup \{x - a\} \cup I_n)) = 0$ ,  $\mathcal{P}\{N(I_n) \geq 1 \mid I_n^c N\} = \ell(I_n)(1 + o(1))$ , so the ratio at (15.7.4)  $\rightarrow 0$ . On realizations for which  $N\{x\} = 1 = N(I_n^c)$ ,

$$\mathcal{P}\{N(I_n) \geq 1 \mid I_n^c N\} = p + \ell(I_n)(1 + o(1))$$

and

$$\mathcal{P}\{N(I_n \setminus \{x + a\}) = 0 \mid I_n^c N\} = 1 - \ell(I_n)(1 + o(1)),$$

so the ratio at (15.7.4)  $\rightarrow p = \pi\{x\} \equiv \pi(\{x\}, N)$ .  $\square$

**EXAMPLE 15.7(c)** *Discrete Bernoulli process* [cf. Example 7.2(d)]. Consider a simple point process  $N$  on the finite set  $\mathcal{X} = \{x_1, \dots, x_n\}$  satisfying, independently for each point,  $N\{x_i\} = 0$  or 1 with probabilities  $q$  and  $p = 1 - q$ , respectively. The Janossy measures are purely atomic, with  $J_0 = q^n$  and

$$J_k(x_{r_1}, \dots, x_{r_k}) \equiv J_k(\{x_{r_1}\}, \dots, \{x_{r_k}\}) = p^k q^{n-k}$$

for any subset  $S_k$  of  $k$  distinct points in  $\mathcal{X}$ . Observe that

$$\sum_{\text{all } S_k} J_k(S_k) = \binom{n}{k} p^k q^{n-k}.$$

The integral equation defining  $R(\cdot \mid \cdot)$  reduces to a definition as a ratio of Janossy measures

$$R(y \mid x_1, \dots, x_k) = J_{k+1}(y, x_1, \dots, x_k) / J_k(x_1, \dots, x_k) = p/q.$$

Also, if  $B = \{y, x_1, \dots, x_k\}$ ,

$$\frac{\mathcal{P}\{N\{y\} = N(\{y, x_1, \dots, x_k\}) = 1 \mid B^c N\}}{\mathcal{P}\{N(\{y, x_1, \dots, x_k\}) = 0 \mid B^c N\}} = \frac{pq^k}{q^{k+1}} = \frac{p}{q},$$

consistent with (5.6.18). On the other hand,

$$\pi\{y\} = \frac{\mathcal{P}\{N\{y\} = N(\{y, x_1, \dots, x_k\}) = 1 \mid B^c N\}}{\mathcal{P}\{N(\{x_1, \dots, x_k\}) = 0 \mid B^c N\}} = \frac{pq^k}{q^k} = p,$$

and  $\pi\{y\}/(1 - \pi\{y\}) = p/(1 - p) = p/q$ .

Finally, if  $N\{y\} = 1$ ,  $\pi\{y\}$  is unchanged, but for  $R(\{y\} \mid \cdot)$  we should have a ratio of Janossy measures with the argument  $y$  repeated in the numerator, which is zero on account of the process being simple.  $\square$

The last of the three quantities mentioned around (15.6.3), the *Papangelou measure*  $\zeta(\cdot)$ , is arguably the one that has the most important applications. Papangelou (1974b) devised it primarily as a means of tackling certain problems in stochastic geometry quite distinct from the present context (see Kallenberg (1983b) and related discussion for an informal account and references). As we show shortly, its importance in statistical applications is that under weak conditions it has a density which can be identified with the Papangelou intensity.

We start by establishing its existence under Condition  $\Sigma$ , relating it to the random measures  $R(\cdot \mid \cdot)$  and  $\pi(\cdot)$  discussed already.

**Proposition 15.7.III.** *Let  $N, \mathcal{X}, \mathcal{T}, \pi$ , and  $R$  be as in Proposition 15.7.II, and suppose that Condition  $\Sigma$  holds. Then as  $n \rightarrow \infty$ , the limit*

$$\zeta(B) = \lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} \mathcal{P}\{N(B_{nj}) = 1 \mid B_{nj}^c N\}, \quad B_{nj} \equiv B \cap I_{nj}, \quad (15.7.8)$$

exists a.s. for all bounded  $B \in \mathcal{B}_{\mathcal{X}}$  and defines a random measure given a.s. by

$$\zeta(\cdot) = \pi(\cdot) + R_d(\cdot \mid \cdot), \quad (15.7.9)$$

where  $R_d(\cdot \mid \cdot)$  is the diffuse (i.e., nonatomic) component of the random measure  $R(\cdot \mid \cdot)$ .

PROOF. Without loss of generality assume that  $B \in \mathcal{T}$ ; then each  $B_{nj} = I_{nj} \in \mathcal{T}$  (although we continue to write  $B_{nj}$ ). Write

$$\begin{aligned} \sum_{j=1}^{k_n} \mathcal{P}\{N(B_{nj}) = 1 \mid B_{nj}^c N\} &= \sum_{j=1}^{k_n} [\mathcal{P}\{N(B_{nj}) = 1 \mid B_{nj}^c N\} I_{\{N(B_{nj})=0\}} \\ &\quad + \mathcal{P}\{N(B_{nj}) \geq 1 \mid B_{nj}^c N\} I_{\{N(B_{nj}) \geq 1\}} \\ &\quad - \mathcal{P}\{N(B_{nj}) \geq 2 \mid B_{nj}^c N\} I_{\{N(B_{nj}) \geq 1\}}] \\ &\equiv \Sigma_0(n) + \Sigma_1(n) - \Sigma_2(n), \quad \text{say.} \end{aligned}$$

For  $\Sigma_0(n)$ , observe from (15.6.18) that

$$\mathcal{P}\{N(B_{nj}) = 1 \mid B_{nj}^c N\} = R(B_{nj} \mid B_{nj}^c N) \mathcal{P}\{N(B_{nj}) = 0 \mid B_{nj}^c N\},$$

so we can rewrite  $\Sigma_0(n)$  in the form

$$\Sigma_0(n) = \int_B I_{B_n}(x) h_n(x) R(dx \mid B_{nj}^c N),$$

where, for fixed  $N$ ,  $B_n$  is the union of those  $B_{nj}$  where  $N(B_{nj}) = 0$ , and  $h_n(x) = \mathcal{P}\{N(B_{nj}) = 0 \mid B_{nj}^c N\}$  on  $B_{nj}$ .  $\{B_n\}$  is a monotonic increasing sequence of sets, with limit  $B \setminus \{\text{supp } N\}$ , and  $h_n(x) \uparrow 1 - \pi\{x\}$  a.s. by Proposition 15.7.II. By monotone convergence, therefore,

$$\Sigma_0(n) \rightarrow \int_{B \setminus \{\text{supp } N\}} (1 - \pi\{x\})(R_a(dx) + R_d(dx))$$

with  $R_a$  denoting the atomic component of  $R$ . By using (15.7.3),  $R_a \ll \pi_a$  a.s., so the first term here equals  $\pi(B \setminus \{\text{supp } N\})$ , and because  $\pi\{x\} = 0$   $R_d$ -a.s., and  $R\{x\} = 0$  for  $x \in \{\text{supp } N\}$ , the second term equals  $R_d(B)$ .

For  $\Sigma_1(n)$ , for given  $N$ , the sum reduces for  $n$  sufficiently large to a sum over sets  $B_{nj}$  containing exactly one of the atoms of  $N(\cdot)$  in  $B$ . By Proposition 15.7.II again, the limit as  $n \rightarrow \infty$  reduces to  $\pi(B \cap \{\text{supp } N\})$ . Thus,

$$\Sigma_0(n) + \Sigma_1(n) \rightarrow R_d(B) + \pi(B) = \zeta(B) \quad \text{a.s.},$$

and it remains to prove that  $\Sigma_2(n) \rightarrow 0$  a.s. Just as for  $\Sigma_1(n)$ , the sum reduces for  $n$  sufficiently large to a sum,  $\sum'$  say, over precisely  $N(B)$  sets  $B_{nj}$ , where  $N(B_{nj}) = 1$ ; that is,

$$\begin{aligned} \Sigma_2(n) &= \sum' \mathcal{P}\{N(B_{nj}) \geq 2 \mid B_{nj}^c N\} \\ &= \sum' \frac{\mathcal{P}\{N(B_{nj}) \geq 2, N(B \setminus B_{nj}) = 0 \mid B^c N\}}{\mathcal{P}\{N(B \setminus B_{nj}) = 0 \mid B^c N\}} \end{aligned}$$

on using Lemma 15.7.I. As  $n$  increases, each of the  $N(B)$  terms in the numerator  $\rightarrow 0$  because  $N$  is simple, and for the denominator, which is decreasing to  $\mathcal{P}\{N(B \setminus \{x\}) = 0 \mid B^c N\}$  for some  $x \in \{\text{supp } N\}$ , Condition  $\Sigma$  implies that

$$0 < \mathcal{P}\{N(B) = 0 \mid B^c N\} \leq \mathcal{P}\{N(B \setminus \{x\}) = 0 \mid B^c N\}. \quad \square$$

The Papangelou intensity measure is related to a first moment in much the same way as the first moment and intensity measures coincide (Propositions 9.3.IX–X).

**Corollary 15.7.IV.** *Suppose the first moment measure  $EN(\cdot)$  exists. Then for all bounded  $B \in \mathcal{B}_X$ ,  $\zeta(B)$  at (15.7.8) also has the representation*

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} E(N(B_{n_j}) \mid B_{n_j}^c N) = \zeta(B) \quad \text{a.s. and in } L_1 \text{ mean.} \quad (15.7.10)$$

PROOF. For the a.s. convergence, write

$$\begin{aligned} E(N(B_{n_j}) \mid B_{n_j}^c N) \\ = \mathcal{P}\{N(B_{n_j}) = 1 \mid B_{n_j}^c N\} + E(N(B_{n_j}) I_{\{N(B_{n_j}) \geq 2\}}(N) \mid B_{n_j}^c N), \end{aligned}$$

and use an extension of Lemma 15.7.II to write

$$\begin{aligned} \sum_{j=1}^{k_n} E(N(B_{n_j}) I_{\{N(B_{n_j}) \geq 2\}}(N) \mid B_{n_j}^c N) \\ = \sum_{j=1}^{k_n} \frac{E(N(B_{n_j}) I_{\{N(B_{n_j}) \geq 2\}} I_{\{N(B \setminus B_{n_j}) = 0\}} \mid B^c N)}{\mathcal{P}\{N(B \setminus B_{n_j}) = 0 \mid B^c N\}} \\ \leq \frac{E(\sum_{j=1}^{k_n} N(B_{n_j}) I_{\{N(B_{n_j}) \geq 2\}}(N) \mid B^c N)}{\mathcal{P}\{N(B) = 0 \mid B^c N\}}. \end{aligned}$$

This conditional expectation is bounded above by  $E(N(B) \mid B^c N) < \infty$  a.s., and  $N$  being simple implies that each term in the sum  $\rightarrow 0$  a.s. (cf. also Exercise 9.3.10), so by dominated convergence we have the required result.

To establish  $L_1$  convergence, observe in the proof of the proposition that  $\Sigma_0(n)$  increases monotonically to its limit, which has expectation bounded by  $E(N(B))$ , so its a.s. convergence implies its  $L_1$  convergence. Also,

$$\begin{aligned} \Sigma_1(n) - \Sigma_2(n) &= \sum_{j=1}^{k_n} \mathcal{P}\{N(B_{n_j}) = 1 \mid B_{n_j}^c N\} I_{\{N(B_{n_j}) \geq 1\}} \\ &\leq \sum_{j=1}^{k_n} I_{\{N(B_{n_j}) \geq 1\}} \leq N(B), \end{aligned}$$

so here the a.s. convergence of  $\Sigma_1(n) - \Sigma_2(n)$  implies its  $L_1$  convergence by the dominated convergence theorem. Finally,

$$\begin{aligned} E\left(\sum_{j=1}^{k_n} E(N(B_{n_j}) I_{\{N(B_{n_j}) \geq 2\}}(N) \mid B_{n_j}^c N)\right) &= E\left(\sum_{j=1}^{k_n} N(B_{n_j}) I_{\{N(B_{n_j}) \geq 2\}}\right) \\ &\rightarrow 0 \quad (n \rightarrow \infty) \end{aligned}$$

from simplicity and the assumption that  $E(N(B)) < \infty$ . □

So far the development in this chapter has been based mainly on disintegrations and limits, having much in common with the material of Chapter 13. It is possible to base the derivation of the Papangelou intensity measure on arguments much closer to those used in the discussion of the compensator and its density, the conditional intensity  $\lambda^*(\cdot, \omega)$ . With a state space  $\mathcal{X}$  more general than  $\mathbb{R}$  or  $\mathbb{R}_+$  as in Chapter 14, the concept of predictability used there is replaced by that of *exvisibility* due to Van der Hoeven (1982) [we follow the terminology of Kallenberg (1983a)].

Write  $\bar{\sigma}\{B^c N\}$  for the completion of the  $B$ -external  $\sigma$ -algebra  $\sigma\{B^c N\}$  with respect to the null sets of  $\sigma\{N\}$  of the process  $N$ . Then on the product space  $\mathcal{X} \times \Omega$  (or, more specifically,  $\mathcal{X} \times \mathcal{N}_{\mathcal{X}}^{\#}$  in the canonical set-up), define the *exvisible*  $\sigma$ -algebra  $\mathcal{Z}$  to be the  $\sigma$ -algebra generated by sets of the form  $B \times U$ , where  $B \in \mathcal{B}_{\mathcal{X}}$ ,  $U \in \bar{\sigma}\{B^c N\}$ . A stochastic process on  $(\Omega, \mathcal{E}, P)$  is then *exvisible* if it is measurable with respect to  $\mathcal{Z}$  on  $\mathcal{X} \times \Omega$ .

Given a random measure  $\xi$ , a ‘dual exvisible projection’ of  $\xi$  can be introduced as the unique random measure satisfying conditions (i)–(iii) of Proposition 15.7.V below. A direct proof of this assertion requires arguments from the general theory of processes analogous to those needed to give a direct proof of the properties of the compensator  $A(\cdot)$  in Section 14.1 [see Van der Hoeven (1982, 1983)]. However, when  $\xi$  is a point process satisfying the special conditions assumed in this chapter (namely, it is simple and satisfies Condition  $\Sigma$ ), it is not too difficult to see that the dual exvisible projection is nothing other than the Papangelou intensity measure  $\zeta$  itself: we conclude with a formal statement and proof of this result.

**Proposition 15.7.V.** *Under the conditions of Proposition 15.7.III, and assuming  $EN(\cdot)$  exists, the Papangelou intensity measure  $\zeta$  is the unique (up to equivalences) random measure satisfying the conditions*

- (i)  $\zeta$  is determined by the point process  $N$  (i.e.,  $\zeta(B)$  is  $\sigma\{N\}$ -measurable for every  $B \in \mathcal{B}_{\mathcal{X}}$ );
- (ii) the process  $Z(x) \equiv \zeta\{x\}$  is exvisible; and
- (iii) for every nonnegative exvisible process  $Y$  and bounded  $B \in \mathcal{B}_{\mathcal{X}}$ ,

$$E\left(\int_B Y(x) \zeta(dx)\right) = E\left(\int_B Y(x) N(dx)\right). \quad (15.7.11)$$

PROOF. The function  $\zeta$  as defined at (15.7.8) is the limit of a  $\sigma\{N\}$ -measurable r.v., and therefore (i) holds for  $\zeta$ .

To prove (ii), suppose  $x$  is an atom of  $\zeta$ . Then in the notation used in the proof of Proposition 15.7.III,

$$\zeta\{x\} = 1 - \pi\{x\} = \lim_{n \rightarrow \infty} (1 - h_n(x)) \quad \text{a.s.,}$$

where  $h_n(x) = \sum_{j=1}^{k_n} \mathcal{P}\{N(B_{nj}) = 0 \mid B_{nj}^c N\} I_{B_{nj}}(x)$  is clearly exvisible. The limit is thus a.s. equal to an exvisible process, and if the  $\sigma$ -fields are

complete we can allow modifications on sets of measure zero without upsetting measurability, so all versions are exvisible.

For (iii), take  $Y$  in (15.7.11) to have the special form

$$Y(x, \omega) = I_B(x) I_U(\zeta) \quad \text{for } U \in \sigma\{B^c N\}.$$

Then Corollary 15.7.IV implies that because  $U \in \sigma\{B_{nj}^c N\}$  for every  $B_{nj}$ ,

$$E(I_U N(B)) = E\left(I_U \sum_{j=1}^{k_n} E(N(B_{nj}) \mid B_{nj}^c N)\right) \rightarrow E(I_U \zeta(B)).$$

Because the left-hand side here is fixed, (15.7.11) follows for this particular function  $Y$ , and then for processes  $Y$  as described by standard extension arguments.

Thus,  $\zeta$  satisfies (i)–(iii): suppose  $\eta$  is some other random measure satisfying the conditions. Whenever  $U \in \sigma\{B_{nj}^c N\}$ , (15.7.11) implies that

$$\begin{aligned} E(I_U E[\eta(B_{nj}) \mid B_{nj}^c N]) &= E(I_U \eta(B_{nj})) = E(I_U N(B_{nj})) \\ &= E(I_U E(N(B_{nj}) \mid B_{nj}^c N)), \end{aligned}$$

from which it follows that

$$E(\eta(B_{nj}) \mid B_{nj}^c N) = E(N(B_{nj}) \mid B_{nj}^c N) \quad \text{a.s.},$$

and hence that

$$\lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} E(\eta(B_{nj}) \mid B_{nj}^c N) = \lim_{n \rightarrow \infty} \sum_{j=1}^{k_n} E(N(B_{nj}) \mid B_{nj}^c N) = \zeta(B) \quad \text{a.s.}$$

Each of these two sums may be further analysed by the same procedure as used in forming the sums  $\Sigma_0(n), \Sigma_1(n), \Sigma_2(n)$  in the proof of Proposition 15.7.III. In particular, using the  $\sigma\{B_{nj}^c N\}$ -measurability of  $\eta(\cdot)$  on  $\{N(B_{nj}) = 0\}$ ,

$$\begin{aligned} &\sum_{j=1}^{k_n} E(\eta(B_{nj}) I_{\{N(B_{nj})=0\}}(N) \mid B_{nj}^c N) I_{\{N(B_{nj})=0\}} \\ &= \sum_{j=1}^{k_n} \eta(B_{nj}) \mathcal{P}\{N(B_{nj}) = 0 \mid B_{nj}^c N\} I_{\{N(B_{nj})=0\}} \\ &\rightarrow \int_{B \setminus \{\text{supp } N\}} (1 - \pi\{x\}) \eta(dx) = \eta_d(B) + \sum (1 - \pi\{x_i\}) \eta\{x_i\}, \end{aligned}$$

where in the second step we have used the limit behaviour of the function  $h_n(x)$  as in the proof of the earlier result,  $\eta_d(\cdot)$  denotes the diffuse component of  $\eta$ , and summation is over the atoms in  $B$  of  $\eta(\cdot)$ . Thus, we have a.s.

$$\eta_d(B) + \sum (1 - \pi\{x_i\}) \eta\{x_i\} = \eta_d(B) + \sum (1 - \pi\{x_i\}) \zeta\{x_i\}. \quad (15.7.12)$$



Now it follows from conditions (ii) and (iii) that the atomic parts of  $\eta$  and  $\zeta$  must be equal, because if we let  $V = \{(x, \omega) \in \mathcal{X} \times \Omega: Z(x, \omega) > \zeta(\{x\}, \omega)\}$ , then  $V$  is an exvisible set and from (15.7.11),

$$\int_V (Z(x, \omega) - \zeta(\{x\}, \omega)) (\eta(dx, \omega) - \zeta(dx, \omega)) \mathcal{P}(d\omega) = 0.$$

Only atoms of  $\eta$  and  $\zeta$  contribute to this integral, and indeed only those for which  $\eta\{x_i\} - \zeta\{x_i\} > 0$ . This leads to a contradiction unless  $\eta\{x_i\} \leq \zeta\{x_i\}$  a.s., and by reversing the argument we deduce that  $\eta\{x_i\} = \zeta\{x_i\}$  a.s. for all atoms; that is,  $\eta$  and  $\zeta$  have the same atoms and the atoms are of the same size a.s. Then (15.7.12) implies that the diffuse components agree a.s.; that is,  $\eta$  and  $\zeta$  coincide except possibly on a set of measure zero.  $\square$

Now suppose that Condition  $\Sigma$  holds and that the point process  $N$  admits a Papangelou intensity  $\rho(x | N)$ . It means that the random measure  $R$  is a.s. absolutely continuous with  $\rho$  as its density. Because its atomic component  $\pi$  is null, it then follows from (15.7.9) that  $\zeta$  also has no atomic component, and that the diffuse components of  $\zeta$  and  $R$  coincide. Hence,  $\zeta$  also is a.s. absolutely continuous with density  $\rho$ . This gives the following corollary of Proposition 15.7.V, in which we repeat (15.7.11) for convenience.

**Corollary 15.7.VI.** *If the simple point process  $N$  admits a Papangelou intensity  $\rho$  and satisfies Condition  $\Sigma$ , then the random measures  $R$  and  $\zeta$  coincide, and both have density  $\rho$ . Moreover, in this case, for all nonnegative, exvisible,  $\mathcal{F}_N$ -measurable processes  $Y$*

$$\mathbb{E} \left( \int_B Y(x) N(dx) \right) = \mathbb{E} \left( \int_B Y(x) \zeta(dx) \right) = \mathbb{E} \left( \int_B Y(x) \rho(x | N) dx \right).$$

This equation may be compared with the extended form (15.5.15') of the Georgii–Nguyen–Zessin formula. There the role of the exvisible process  $Y$  is taken by the function  $h(x, \mathbf{v})$ , where  $\mathbf{v}$  denotes (points of) a realization  $N$ , so  $h$  must embody any condition of  $\mathcal{F}_N$ -measurability. If  $h$  does not depend on  $\mathbf{v}$ , or depends on it only through the function  $\rho(u | \mathbf{v})$ , which is itself exvisible, then (15.5.15') will hold. A further illustration is in Exercise 15.7.1.

## Exercises and Complements to Section 15.7

15.7.1 Show that if a process  $Y(x)$  is a.s. continuous and  $\mathcal{F}_N$ -measurable, and vanishes outside a bounded set, then it is exvisible.

# References with Index

[At the end of each reference entry is the page number or numbers where it is cited in this volume. A bibliography of about 600 references up to about 1970, although excluding much of the historical material of Chapter 1, is given in D.J. Daley and R.K. Milne (1972), *The theory of point processes: A bibliography*, *Int. Statist. Rev.* **41**, 183–201.]

- Aldous, D. and Thorisson, H. (1993). Shift coupling. *Stoch. Proc. Appl.* **44**, 1–14. [222]
- Ambartsumian, R.V. (1972). Palm distributions and superpositions of independent point processes in  $\mathbb{R}^n$ . In Lewis, P.A.W. (Ed.), *Stochastic Point Processes*, Wiley, New York, pp. 626–645. [466]
- Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York. [355]
- Ash, R.B. (1972). *Real Analysis and Probability*. Academic Press, New York. [Revised ed., with C. Doléans-Dade (2000), *Probability and Measure Theory*. Harcourt/Academic Press, San Diego.] [25]
- Asmussen, S. (1987). *Applied Probability and Queues*. John Wiley, Chichester. [355]
- (2003). = Asmussen (1987), 2nd ed. Springer-Verlag, New York. [96, 355]
- Asmussen, S., Nerman, O. and Olsson, M. (1996). Fitting phase-type distributions via the EM Algorithm, *Scand. J. Statist.* **23**, 419–441. [110]
- Asplund, E. and Bungart, L. (1966). *A First Course in Integration*. Holt, Rinehart and Winston, New York. [376]
- Baccelli, F. and Brémaud, P. (1994). *Elements of Queueing Theory: Palm-Martingale Calculus and Stochastic Recurrences*. Springer, Berlin. [222, 269, 328, 355]
- and —— (2003). = Baccelli and Brémaud (1994), 2nd ed. [177, 355]

- Baddeley, A.J. (1998). Spatial sampling and censoring. In Kendall, W., Lieshout, M.N.M. and Barndorff-Nielsen, O. (Eds.), *Stochastic Geometry: Likelihood and Computation*, Chapman and Hall, Boca Raton, FL, pp. 37–78. [457]
- , Gregori, P., Mateu, J., Stoica, R. and Stoyan, D. (Eds.) (2006). *Case Studies in Spatial Point Process Modeling* (Lecture Notes in Statistics **185**). Springer, New York. [459]
- and Møller, J. (1989). Nearest-neighbour Markov point processes and random sets. *Internat. Statist. Review* **57**, 89–121. [118, 128]
- , — and Pakes, A.G. (2007). Properties of residuals for spatial point processes. *Ann. Inst. Statist. Math.* **59** (to appear). [518]
- and Silverman, B.W. (1984). A cautionary example on the use of second-order methods for analyzing point patterns. *Biometrics* **40**, 1089–1094. [464]
- , Turner, R., Møller, J. and Hazelton, M. (2005). Residual analysis for spatial point processes (with Discussion). *J. Roy. Statist. Soc. Ser. B* **67**, 617–666. [419, 459, 502, 513–516]
- and van Lieshout, M.N.M. (1995). Area-interaction point processes. *Ann. Inst. Statist. Math.* **46**, 601–619. [124, 130]
- , — and Møller, J. (1996). Markov properties of cluster processes. *Adv. Appl. Probab.* **28**, 346–355. [129]
- Ball, F. and Milne, R.K. (2005). Simple derivations of properties of counting processes associated with Markov renewal processes. *J. Appl. Probab.* **42**, 1031–1043. [99, 116]
- Barbour, A.D. and Hall, P. (1984). On the rate of Poisson convergence. *Math. Proc. Cambridge Philos. Soc.* **95**, 473–480. [163]
- Barndorff-Nielsen, O. and Yeo, G.F. (1969). Negative binomial processes. *J. Appl. Probab.* **6**, 633–647. [10]
- Bartlett, M.S. (1955). *An Introduction to Stochastic Processes*. Cambridge University Press, Cambridge [2nd ed. 1966; 3rd ed. 1978]. [100]
- (1963). The spectral analysis of point processes (with Discussion). *J. Roy. Statist. Soc. Ser. B* **29**, 264–296. [484]
- Basawa, I.V. and Scott, D.J. (1983). *Asymptotic Inference for Non-ergodic Models* (Springer Lecture Notes Statist. **17**). Springer-Verlag, New York. [416]
- Baum, L.E. and Eagon, J.A. (1967). An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology. *Bull. Amer. Math. Soc.* **73**, 360–363. [104]
- and Petrie, T. (1966). Statistical inference for probabilistic functions of a Markov chain. *Ann. Math. Statist.* **37**, 1554–1563. [104]
- Bebbington, M.S. (2005). Information gains for stress release models. *Pure Appl. Geophys.* **162**, 2299–2319. DOI: 10.1007/s00024-005-2777-5, 2005. [503]
- and Borovkov, K. (2003). A stochastic two-node stress transfer model reproducing Omori’s law. *Pure Appl. Geophys.* **160**, 1429–1455. [229]
- Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall, New York. [250]
- Berbée, H.C.P. (1979). *Random Walks with Stationary Increments and Renewal Theory* (Mathematical Centre Tracts **112**). Mathematisch Centrum, Amsterdam. [229]

- Berman, M. (1978). Regenerative multivariate point processes. *Adv. Appl. Probab.* **10**, 411–430. [333]
- Bertoin, J. (1996). *Lévy Processes*. Cambridge University Press, Cambridge. [Paperback edition, 1998, reprinted 2002.] [12, 21, 82, 83]
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with Discussion). *J. Roy. Statist. Soc. Ser. B* **36**, 192–236. [118, 122, 506]
- Billingsley, P. (1965). *Ergodic Theory and Information*. Wiley, New York. [194, 449]
- (1968). *Convergence of Probability Measures*. Wiley, New York. [27, 214]
- Billiot, J.-M. (1994). Asymptotique de la méthode du maximum de pseudo-vraisemblance pour les processus ponctuels de Gibbs. (Thesis, Université de Montpellier II). [510]
- (1997). Asymptotic properties of Takacs–Fiksel estimation method for Gibbs point processes. *Statistics* **30**, 69–89. [514]
- and Goulard, M. (2001). An estimation method of the pair potential function for Gibbs point processes on spheres. *Scand. J. Statist.* **28**, 185–203. [509]
- Bingham, N.H., Goldie, C.M. and Teugels, J. (1987). *Regular Variation*. Cambridge University Press, Cambridge. Paperback edition (corrected printing), 1989. [160]
- Block, H.W., Borges, W.S. and Savits, T.H. (1985). Age-dependent minimal repair. *J. Appl. Probab.* **22**, 370–385. [235]
- 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. [375, 405]
- Bol’shakov, I.A. (1969). *Statistical Problems in Isolating a Stream of Signals from Noise* (in Russian). Sovyetskoye Radio, Moscow. [65]
- Brandt, A., Franken, P. and Lisek, B. (1990). *Stationary Stochastic Models*. Wiley, Chichester. [269]
- Breiman, L. (1963). The Poisson tendency in traffic distribution. *Ann. Math. Statist.* **34**, 111. [174]
- Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics*. Springer-Verlag, New York. [101, 375, 399–419]
- and Jacod, J. (1977). Processus ponctuels et martingales: résultats récents sur le modélisation et le filtrage. *Adv. Appl. Probab.* **9**, 362–416. [400]
- and Massoulié, L. (1994). Imbedded construction of stationary point processes and sequences with a random memory. *Queueing Systems* **17**, 213–234. [224, 427]
- and —— (1996). Stability of non-linear Hawkes processes. *Ann. Probab.* **24**, 1563–1588. [222, 224, 427–439]
- and —— (2001). Hawkes branching point processes without ancestors. *J. Appl. Probab.* **38**, 122–135. [141, 145]
- , —— and Ridolfi, A. (2005). Power spectra of random spike fields and related processes. *Adv. Appl. Probab.* **37**, 1116–1146. [376]
- , Nappo, G. and Torrisi, G.L. (2002). Rate of convergence to equilibrium of marked Hawkes processes. *J. Appl. Probab.* **39**, 123–136. [428]
- Brillinger, D.R. (1972). The spectral analysis of stationary interval functions. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1**, 403–431. [237]
- Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Adv. Appl. Probab.* **31**, 929–953. [83, 84]

- Brown, T.C. (1978). A martingale approach to the Poisson convergence of simple point processes. *Ann. Probab.* **6**, 615–628. [384, 419]
- (1982). Poisson approximations and exchangeable random variables. In Koch, G. and Spizzichino, F. (Eds.), *Exchangeability in Probability and Statistics*, North-Holland, Amsterdam, pp. 177–183. [384]
- (1983). Some Poisson approximations using compensators. *Ann. Probab.* **11**, 726–744. [384]
- and Nair, M.G. (1988). A simple proof of the multivariate random time change theorem for point processes. *J. Appl. Probab.* **25**, 210–214. [419, 425]
- Browne, S. and Sigman, K. (1992). Work-modulated queues with applications to storage processes. *J. Appl. Probab.* **29**, 699–712. [228, 235]
- Brownrigg, R. and Harte, D.S. (2005). Using R for statistical seismology. *R News* **5**(1), 31–35. URL: [cran.r-project.org/doc/Rnews/Rnews.2005-1.pdf](http://cran.r-project.org/doc/Rnews/Rnews.2005-1.pdf) [499]
- Buchler, J.R., Dufty, J.W. and Kandrup, H.E. (Eds) (1998). *Long-range Correlations in Astrophysical Systems* (Ann. New York Acad. Sci. **848**), New York. [517]
- Byth, K. (1981).  $\theta$ -stationary point processes and their second order analysis. *J. Appl. Probab.* **18**, 864–878. [466, 471]
- Campbell, N.R. (1909). The study of discontinuous phenomena. *Proc. Cambridge Philos. Soc.* **15**, 117–136. [66, 271]
- Chin, Y.C. and Baddeley, A.J. (1999). On connected component Markov point processes. *Adv. Appl. Probab.* **31**, 279–282. [129]
- and — (2000). Markov interacting component processes. *Adv. Appl. Probab.* **32**, 597–619. [129]
- Chong, F.S. (1983). Time-space-magnitude interdependence of upper crustal earthquakes in the main seismic region of New Zealand. *New Zealand J. Geol. Geophys.* **26**, 7–24. [496]
- Chung, K.L. (1967). *Markov Chains with Stationary Transition Probabilities*, 2nd ed. Springer-Verlag, New York. [115]
- (1974). *A Course in Probability Theory*, 2nd ed. Academic Press, New York. [48, 146, 209]
- Çınlar, E. (1975). *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs, NJ. [96, 115]
- Clifford, P. (1990). Markov random fields in statistics. In Grimmett, G.R. and Welsh, D.J.A. (Eds.), *Disorder in Physical Systems*, Clarendon Press, Oxford, pp. 19–32. [121, 122]
- Cramér, H. and Leadbetter, M.R. (1967). *Stationary and Related Stochastic Processes*. Wiley, New York. [10, 159]
- Cressie, N.A.C. (1991). *Statistics for Spatial Data*. Wiley, New York. [459]
- (1993). = Cressie (1991), revised ed. [459]
- Csiszár, I. (1969). On generalized entropy. *Stud. Sci. Math. Hungar.* **4**, 401–419. [447, 455]
- Cutler, C.D. (1991). Some results on the behaviour and estimation of the fractal dimensions of distributions on attractors. *J. Statist. Phys.* **62**, 651–708. [341, 351]
- Daley, D.J. (1971). Weakly stationary point processes and random measures. *J. Roy. Statist. Soc. Ser. B* **33**, 406–428. [205, 484]

- Daley, D.J. (1972). Asymptotic properties of stationary point processes with generalized clusters. *Z. Wahrs.* **21**, 65–76. [251]
- (1974). Various concepts of orderliness for point processes. In Harding and Kendall (1974), pp. 148–161. [47, 51]
- (1982a). 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**), pp. 313–320. [51]
- (1982b). Infinite intensity mixtures of point processes. *Math. Proc. Cambridge Philos. Soc.* **92**, 109–114. [51]
- (1987). The variation distance between Poisson distributions. (Unpublished). [Abstract in *Bull. Inst. Math. Statist.* **16**, 41.] [166, 175]
- (1999). The Hurst index of long-range dependent renewal processes. *Ann. Probab.* **27**, 2035–2041. [254]
- (2004). Further results for the lilypond model. In Baddeley, A., Gregori, P., Mateu, J., Stoica, R. and Stoyan, D. (Eds.), *Spatial Point Process Modelling and its Applications* (Treballs d'Informàtica i Tecnologia **20**), Universitat Jaume I, Castelló, pp. 55–65. [504]
- (2007). Long-range dependence in a Cox process directed by an alternating renewal process. (Submitted). [254]
- and Goldie, C.M. (2005). The moment index of minima (II). *Statist. Probab. Letters* **76**, 831–837. [254]
- and Last, G. (2005). Descending chains, the lilypond model, and mutual nearest-neighbour matching. *Adv. Appl. Probab.* **37**, 604–628. [506]
- , Mallows, C.L. and Shepp, L. (2000). A one-dimensional Poisson growth model with non-overlapping intervals. *Stoch. Proc. Appl.* **90**, 223–241. [504]
- , Rolski, T. and Vesilo, R. (2007). Long-range dependence in a Cox process driven by a Markov renewal process. *J. Appl. Math. Decision Sci.* (Special Issue *Statistics and Applied probability: A tribute to Jeffrey J. Hunter*) (to appear). [254]
- , Stoyan, D. and Stoyan, H. (1999). The volume fraction of a Poisson germ model with maximally non-overlapping spherical grains. *Adv. Appl. Probab.* **31**, 610–624. [503]
- and Vere-Jones, D. (1987). The extended probability generating functional, with application to mixing properties of cluster point processes. *Math. Nachr.* **131**, 311–319. [64, 213]
- and — (2004). Scoring probability forecasts for point processes: the entropy score and information gain. In Gani, J. and Seneta, E. (Eds.), *Stochastic Methods and their Applications* (Papers in honour of Chris Heyde; J. Appl. Probab. **41A**), pp. 297–312. [445, 454, 503]
- and Vesilo, R. (1997). Long-range dependence of point processes, with queueing examples. *Stoch. Proc. Appl.* **70**, 265–282. [249]
- and — (2000). Long-range dependence of inputs and outputs of classical queues. In McDonald, D.R. and Turner, S.R.E. (Eds.), *Analysis of Communication Networks: Call Centres, Traffic and Performance* (Fields Institute Communications **28**), pp. 179–186. [252]

- Darroch, J.N. and Morris, K.W. (1967). Some passage-time generating functions for discrete-time and continuous-time finite Markov chains. *J. Appl. Probab.* **4**, 496–507. [116]
- Davidson, R. (1974a). *Stochastic processes of flats and exchangeability* (Part II, Ph.D. thesis, Cambridge University, 1968). In Harding and Kendall (1974), pp. 13–45. [478, 479]
- (1974b). 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.] [481]
- (1974c). Exchangeable point-processes. In Harding and Kendall (1974), pp. 46–51. [73]
- Davis, M.H.A. (1976). The representation of martingales of jump processes. *SIAM J. Control* **14**, 623–638. [375]
- Davison, A.C. and Ramesh, N.I. (1993). A stochastic model for times of exposure to air pollution from a point source. In Barnett, V.D. and Turkman, K.F. (Eds.), *Statistics for the Environment*, Wiley, Chichester. [105]
- Debes, H., Kerstan, J., Liemant, A. and Matthes, K. (1971). Verallgemeinerungen eines Satzes von Dobrushin, III. *Math. Nachr.* **50**, 99–139. [175]
- Delasnerie, M. (1977). Flots mélangements et mesures de Palm. *Ann. Inst. Henri Poincaré, Sec. B* **8**, 357–369. [269, 330]
- Dellacherie, C. and Meyer, P.-A. (1978). *Probabilities and Potential*. Hermann, Paris, and North-Holland, Amsterdam. [356]
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with Discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 1–38. [104]
- Deng, L. and Mark, J.W. (1993). Parameter estimation for Markov-modulated Poisson processes via the E–M algorithm with time discretization. *Telecommun. Systems* **1**, 321–338. [104]
- Derman, C. (1955). Some contributions to the theory of denumerable Markov chains. *Trans. Amer. Math. Soc.* **79**, 541–555. [173, 175]
- Diggle, P.J. (1983). *Statistical Analysis of Spatial Point Patterns*. Academic Press, London. [459–461]
- (2003). = Diggle (1983), 2nd ed. Oxford University Press, Oxford. [459]
- , Fiksel, T., Grabarnik, P., Ogata, Y., Stoyan, D. and Tanemura, M. (1994). On parameter estimation for pairwise-interaction point processes. *Internat. Statist. Review* **62**, 99–117. [514]
- Dobrushin, R.L. (1956). On the Poisson law for distributions of particles in space (in Russian). *Ukr. Mat. Zh.* **8**, 127–134. [174]
- (1968). The description of a random field by means of conditional probabilities and conditions of its regularity. *Teor. Veroyat. Primenen.* **13**, 201–229. [Translation in *Theory Probab. Applic.* **13**, 197–224.] [118]
- Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2**, 183–201. [23]
- Doléans-Dade, C. (1970). Quelques applications de la formule de changement de variables pour les semimartingales. *Z. Wahrs.* **16**, 181–194. [419]
- Doob, J.L. (1948). Renewal theory from the point of view of the theory of probability. *Trans. Amer. Math. Soc.* **63**, 422–438. [268]

- Dubins, L.E. and Freedman, D.A. (1967). Random distribution functions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **2** (Pt. 1), 183–214. [50]
- Dudley, R.M. (1969). Random linear functionals. *Trans. Amer. Math. Soc.* **136**, 1–24. [53]
- Elliott, R.J. (1976). Stochastic integrals for martingales of a jump process with partially accessible jump times. *Z. Wahrs.* **36**, 213–226. [370]
- (1982). *Stochastic Calculus and its Applications*. Springer-Verlag, New York. [370]
- , Aggoun, L. and Moore, J.B. (1995). *Hidden Markov Models*. Springer, New York. [104]
- Falconer, K. (1990). *Fractal Geometry: Mathematical Foundations and Applications*. Wiley, Chichester. [341]
- Feller, W. (1966). *An Introduction to Probability Theory and its Applications*, Vol.2. Wiley, New York. [12, 25, 146, 259, 409]
- (1968). *An Introduction to Probability Theory and its Applications*, Vol.1, 3rd ed. Wiley, New York. [115]
- (1971). = Feller (1966), 2nd ed. [160]
- Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230. [12]
- Fiksel, T. (1988). Estimation of interaction potentials of Gibbsian point processes. *Statistics* **19**, 77–86. [514, 518]
- Fischer, D.R. and Meier-Hellstern, K.S. (1993). The Markov-modulated Poisson process (MMPP) cookbook. *Performance Eval.* **18**, 149–171. [105]
- Fleischmann, K. (1978). Mixing properties of cluster-invariant distributions. *Litovsk. Mat. Sb.* **18**(3), 191–199. [338]
- and Pohn, U. (1974). Ein Grenzwertsatz für subkritische Verzweigungsprozesse mit endlich vielen Typen von Teilchen. *Math. Nachr.* **64**, 357–362. [339]
- and — (1975). Subkritische räumlich homogene Verzweigungsprozesse. *Math. Nachr.* **70**, 231–250. [339]
- Franken, P. (1963). Approximation durch Poissonsche Prozesse. *Math. Nachr.* **26**, 101–114. [146]
- (1975). Einige Anwendungen der Theorie zufälliger Punktprozesse in der Bedienungstheorie, I. *Math. Nachr.* **70**, 303–319. [269]
- , König, D., Arndt, U. and Schmidt, V. (1981). *Queues and Point Processes*. Akademie-Verlag, Berlin. [222, 269, 300, 316, 328, 333]
- Fritz, J. (1969). Entropy of point processes. *Stud. Sci. Math. Hungar.* **4**, 389–399. [447]
- (1973). An approach to the entropy of point processes. *Period. Math. Hungar.* **3**, 73–83. [455]
- Galchuk, L.I. and Rosovskii, B.L. (1971). The ‘disorder’ problem for a Poisson process. *Teor. Veroyat. Primen.* **16**, 729–734. [Translation in *Theory Probab. Appl.* **16**, 712–717.] [101]
- Gel’fand, I.M. and Vilenkin, N.Ya. (1964). *Generalized Functions*, Vol.4. Academic Press, New York. [54]



- Georgii, H.-O. (1976). Canonical and grand canonical Gibbs states for continuum systems. *Commun. Math. Phys.* **48**, 31–51. [511]
- (1988). *Gibbs Measure and Phase Transitions*. (Studies in Mathematics **9**). de Gruyter, Berlin. [118, 519]
- Glötzl, E. (1980). Lokale Energien und Potentiale für Punktprozesse. *Math. Nachr.* **96**, 195–206. [519, 524]
- Grassberger, P. and Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D* **9**, 189–208. [346]
- Grigelionis, B. (1963). On the convergence of sums of random step processes to a Poisson process (in Russian). *Teor. Veroyat. Primen.* **8**, 189–194. [Translation in *Theory Probab. Appl.* **8**, 177–182.] [146]
- (1971). On the representation of integer-valued random measures by means of stochastic integrals with respect to Poisson measure. *Litovsk. Mat. Sb.* **11**, 93–108. [427, 428]
- Gross, D. and Miller, D.R. (1984). The randomization technique as a modeling tool and solution procedure for transient Markov processes. *Operat. Res.* **32**, 345–361. [117]
- Häggström, O. and Meester, R. (1996). Nearest neighbour and hard sphere models in continuum percolation. *Random Struct. Algor.* **9**, 295–315. [503]
- Halmos, P. (1950). *Measure Theory*. Van Nostrand, Princeton NJ. [20]
- Hammersley, J.M. and Clifford, P. (1971). Markov fields on finite fields and lattices (unpublished). See Besag (1974), Clifford (1990). [118]
- , Lewis, J.W.E. and Rowlinson, J.S. (1975). Relationships between the multinomial and Poisson models of stochastic processes, and between the canonical and grand canonical ensembles in statistical mechanics, with illustrations and Monte-Carlo methods for the penetrable sphere model of liquid–vapour equilibrium. *Sankhyā A* **37**, 457–491. [124]
- Hanisch, K.-H. (1984). Some remarks on estimators of the distribution function of nearest-neighbour distance in stationary spatial point-patterns. *Math. Operat. Statist., Ser. Statistics* **15**, 409–412. [463, 466]
- Harding, E.J. and Kendall, D.G. (Eds.) (1974). *Stochastic Geometry*. Wiley, Chichester. [458, 472]
- Harris, T.E. (1963). *The Theory of Branching Processes*. Springer-Verlag, Berlin. [1, 26, 334, 338]
- Harte, D.S. (1998). Dimension estimates of earthquake epicentres and hypocentres. *J. Non-linear Sci.* **8**, 581–618. [346]
- (2001). *Multifractals: Theory and Applications*. Chapman and Hall/CRC, Boca Raton, FL. [341–352]
- (2003). *Package “PtProcess”: Time Dependent Point Process Modelling*. R statistical program routines. Statistics Research Associates, Wellington. URL: [homepages.paradise.net.nz/david.harte/SSLib](http://homepages.paradise.net.nz/david.harte/SSLib) [499]
- Hawkes, J. and Oakes, D. (1974). A cluster representation of a self-exciting process. *J. Appl. Probab.* **11**, 493–503. [229, 232]
- Heveling, M. and Last, G. (2005). Characterization of Palm measures via bijective point shifts. *Ann. Probab.* **33**, 1698–1715. [269, 309–317]
- and — (2006). Existence, uniqueness, and algorithmic computation of general lilypond systems. *Random Struct. Algor.* **29**, 338–350. [506]

- Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **38** 1163–1174. [346]
- Ibragimov, I.A. and Linnik, Yu.V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Gröningen. [Translated from original in Russian (1965), Nauka, Moscow.] [167]
- Isham, V. (1981). An introduction to spatial point processes and Markov random fields. *Internat. Statist. Rev.* **49**, 21–43. [118]
- and Westcott, M. (1979). A self-correcting point process. *Stoch. Proc. Appl.* **8**, 335–347. [418]
- Jacobsen, M. (1982). *Statistical Analysis of Counting Processes* (Springer Lecture Notes Statist. **12**). Springer-Verlag, New York. [396]
- Jacod, J. (1975). Multivariate point processes: Predictable projections, Radon–Nikodym derivatives, representation of martingales. *Z. Wahrs.* **31**, 235–253. [399, 405]
- Jarupskin, B.D.S. (1984). *Maximum Likelihood and Related Estimation Methods in Point Processes*. Ph.D. thesis, University of California, Berkeley. [412, 416]
- Jelinski, Z., and Moranda, P. (1972). Software reliability research. In Freiberger, W. (Ed.), *Statistical Computer Performance Evaluation*, Academic Press, New York, pp. 465–484. [100]
- Jowett, J. and Vere-Jones, D. (1972). The prediction of stationary point processes. In Lewis, P.A.W. (Ed.), *Stochastic Point Processes*, Wiley, New York, pp. 405–435. [101, 410]
- Kagan, Y.Y. (1981a). Spatial distribution of earthquakes: The three-point correlation function. *Geophys. J. Roy. Astr. Soc.* **67**, 697–717. [249]
- (1981b). Spatial distribution of earthquakes: The four-point correlation function. *Geophys. J. Roy. Astr. Soc.* **67**, 719–733. [249]
- and Knopoff L., (1980). Spatial distribution of earthquakes: The two-point correlation function. *Geophys. J. Roy. Astr. Soc.* **62**, 303–320. [249]
- and Vere-Jones, D. (1995). Problems in the modelling and statistical analysis of earthquakes. In Heyde, C.C., Prohorov, Yu.V., Pyke, R. and Rachev, S.T. (Eds.), *Athens Conference on Applied Probability and Time Series, Volume 1: Applied Probability* (Springer Lecture Notes in Statistics **114**), Springer, New York, pp. 398–425. [249]
- Kallenberg, O. (1973). Characterization and convergence of random measures and point processes. *Z. Wahrs.* **27**, 9–21. [33]
- (1975). *Random Measures*. Akademie-Verlag, Berlin, and Academic Press, London. [3rd ed. 1983, reprinted with corrections as 4th ed. 1986]. [33–39, 95, 146, 153, 178, 272]
- (1977a). Stability of critical cluster fields. *Math. Nachr.* **77**, 7–43. [336, 337]
- (1977b). A counterexample to R. Davidson’s conjecture on line processes. *Math. Proc. Cambridge Philos. Soc.* **82**, 301–307. [481, 483]
- (1978). On conditional intensities of point processes. *Z. Wahrs.* **41**, 205–220. [519]
- (1983a). = Kallenberg (1975), 3rd ed. [55, 86, 178, 270, 272, 518–534]

- Kallenberg, O. (1983b). On random processes of flats with special emphasis on the invariance problem. *Bull. Int. Statist. Inst.* **50**(2), 854–862, and **50**(3), 383–392. [531]
- (1984). An informal guide to the theory of conditioning in point processes. *Int. Statist. Review* **52**, 151–164. [519, 525]
- (1986). = Kallenberg (1975), 4th ed. [518]
- Kaplan, E.L. (1955). Transformations of stationary random sequences. *Math. Scand.* **3**, 127–149. [268, 299]
- Karatzas, I., and Shreve, S.E. (1988). *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York. [370]
- Karbe, W. (1973). *Konstruktion einfacher zufälliger Punktfolgen*. Diplomarbeit, Friedrich-Schiller-Universität Jena, Sektion Mathematik. [33]
- Keiding, N. (1975). Maximum likelihood estimation in birth and death processes. *Ann. Statist.* **3**, 363–372. [416–418]
- Kelly, F.P. and Ripley, B.D. (1976). A note on Strauss's model for clustering. *Biometrika* **63**, 357–360. [123]
- Kendall, D.G. (1963). Extreme-point methods in stochastic analysis. *Z. Wahrs.* **1**, 295–300. [83, 86]
- (1974). Foundations of a theory of random sets. In Harding and Kendall (1974), pp. 322–376. [33, 458]
- Kerstan, J. (1964a). Verallgemeinerung eines Satzes von Prochorov und Le Cam. *Z. Wahrs.* **2**, 173–179. [163]
- (1964b). 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. [224, 427, 430, 439]
- and Debes, H. (1969). Zufällige Punktfolgen und Markoffsche Übergangsmatrizen ohne stationäre Verteilungsgesetze. *Wiss. Z. Friedrich-Schiller-Universität Jena* **18**, 349–359. [175]
- and Matthes, K. (1964). Stationäre zufällige Punktfolgen II. *Jber. Deutsch. Math.-Verein.* **66**, 106–118. [91]
- and — (1967). Ergodische unbegrenzt teilbare stationäre Punktfolgen. *Trans. Fourth Prague Conf. Inf. Theory Stat. Dec. Functions Random Proc.*, 399–415. [217, 218]
- , — and Mecke, J. (1974). *Unbegrenzt Teilbare Punktprozesse*. Akademie-Verlag, Berlin. [2nd, 3rd ed. = MKM (1978, 1982) in English, Russian.] [268]
- Khinchin, A.Ya. (1955). *Mathematical Methods in the Theory of Queueing* (in Russian). *Trudy Mat. Inst. Steklov* **49**. [Translated (1960) Griffin, London.] [47, 146, 268]
- Kijima, M. (1989). Some results for repairable systems. *J. Appl. Probab.* **26**, 89–102. [228, 235]
- Kingman, J.F.C. (1967). Completely random measures. *Pacific J. Math.* **21**, 59–78. [77–82]
- Klemm, A., Lindemann, C. and Lohmann, M. (2003). Modeling IP traffic using the batch Markovian arrival process. *Performance Eval.* **54**, 149–173. [110–113]
- König, D. and Matthes, K. (1963). Verallgemeinerung der Erlangschen Formeln, I. *Math. Nachr.* **26**, 45–56. [269]

- König, D., Matthes, K. and Nawrotzki, K. (1967). *Verallgemeinerung der Erlang-schen und Engset'schen Formeln* (Eine Methode in der Bedienungstheorie). Akademie-Verlag, Berlin. [269]
- Kraft, C.H. (1964). A class of distribution function processes which have derivatives. *J. Appl. Probab.* **1**, 385–388. [50]
- Krickeberg, K. (1965). *Probability Theory*. Addison-Wesley, Reading, MA. [378]
- (1974a). Invariance properties of the correlation measure of line-processes. In Harding and Kendall (1974), pp. 76–88. [478]
- (1974b). Moments of point processes. In Harding and Kendall (1974), pp. 89–113. [237, 473–478]
- Kulkarni, V.G. (1995). *Modeling and Analysis of Stochastic Systems*. Chapman & Hall, London. [96]
- Kummer, G. and Matthes, K. (1970). Verallgemeinerung eines Satzes von Sliwnjak, II. *Rev. Roumaine Math. Pures Appl.* **15**, 845–870. [268, 271]
- Kurtz, T.G. (1974). Point processes and completely monotone set functions. *Z. Wahrs.* **31**, 57–67. [33–36]
- Kutoyants, Y. (1979). Local asymptotic normality for Poisson type processes. *Izv. Akad. Nauk Arm. SSR Ser. Mat.* **14**(1), 3–20 and 72. [412]
- (1980). *Estimation of Parameters of Stochastic Processes* (in Russian). Armenian Academy of Science, Erevan. [412, 415]
- (1984a). Parameter estimation for processes of Poisson type. *Izv. Akad. Nauk Arm. SSR Ser. Mat.* **19**, 233–241 [415]
- (1984b). *Parameter Estimation for Stochastic Processes*. Heldermann, Berlin. [Translated by B.L.S. Prakasa Rao and revised from Kutoyants (1980).] [412, 415]
- Lanford, O.E. and Ruelle, D. (1969). Observables at infinity and states with short range correlations in statistical mechanics. *Comm. Math. Phys.* **13**, 194–215. [215]
- Last, G. (1996). Coupling with compensators. *Stoch. Proc. Appl.* **65**, 147–170. [429]
- (2004). Ergodicity properties of stress-release, repairable system and workload models. *Adv. Appl. Probab.* **36**, 471–498. [222, 228, 235, 429]
- and Brandt, A. (1995). *Marked Point Processes on the Real Line*. Springer, New York. [355, 368, 380, 382, 400–405]
- and Szekli, R. (1998). Asymptotic and monotonicity properties of some repairable systems. *Adv. Appl. Probab.* **30**, 1089–1110. [228, 235]
- Lawrance, A.J. (1971). Selective interaction of a Poisson and a renewal process: The dependency structure of the intervals between the responses. *J. Appl. Probab.* **8**, 170–173. [247]
- Le Cam, L. (1960). An approximation theorem for the Poisson binomial distribution. *Pacific J. Math.* **10**, 1181–1197. [165]
- Leadbetter, M.R. (1968). On three basic results in the theory of stationary point processes. *Proc. Amer. Math. Soc.* **19**, 115–117. [39]
- (1972). On basic results of point process theory. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **3**, 449–462. [39, 275, 283, 308]

- Leadbetter, M.R., Lindgren, G. and Rootzen, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York. [159]
- Lee, M.-L.T. and Whitmore, G. (1993). Stochastic processes directed by randomized time. *J. Appl. Probab.* **30**, 302–314. [83]
- Lee, P.M. (1964). A structure theorem for infinitely divisible point processes. Address to I.A.S.P.S., Berne (unpublished) [cf. Lee, P.M. (1964), The superposition of point processes (Abstract), *Ann. Math. Statist.* **35**, 1406–1407]. [91]
- (1967). Infinitely divisible stochastic processes. *Z. Wahrs.* **7**, 147–160. [91]
- Lewis, P.A.W. and Shedler, G.S. (1976). Simulation of nonhomogeneous Poisson processes with log linear rate function. *Biometrika* **63**, 501–506. [427]
- Liemant, A. (1969). Invariante zufällige Punktfolgen. *Wiss. Z. Friedrich-Schiller-Universität Jena* **18**, 361–372. [337]
- (1975). Verallgemeinerungen eines Satzes von Dobruschin, V. *Math. Nachr.* **70**, 387–390. [337]
- , Matthes, K. and Wakolbinger, A. (1988). *Equilibrium Distributions of Branching Processes*. Akademie-Verlag, Berlin. [334]
- Liggett, T.M. (1985). *Interacting Particle Systems*. Springer, New York. [431]
- Lindvall, T. (1988). Ergodicity and inequalities in a class of point processes. *Stoch. Proc. Appl.* **30**, 121–131. [222, 229, 439]
- (1992). *Lectures on the Coupling Method*. Wiley, New York. [132, 222, 229]
- Liptser, R.S. and Shiryaev, A.N. (1974). *Statistics of Random Processes* (in Russian). Nauka, Moscow. [Translation = — and — (1977, 1978).] [358]
- and — (1977). *Statistics of Random Processes, I: General Theory*. Springer-Verlag, New York. [358]
- and — (1978). *Statistics of Random Processes, II: Applications*. Springer-Verlag, New York. [358, 404, 405]
- and — (2000). = Liptser and Shiryaev (1977, 1978), retypeset ed. [358]
- Loève, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, New York. [146]
- MacDonald, I.L. and Zucchini, W. (1997). *A Hidden Markov and other Models for Discrete-Valued Time Series*. Chapman and Hall, London. [104, 116]
- Mandelbrot, B. (1982). *The Fractal Geometry of Nature*, 2nd ed. W.H. Freeman, San Francisco. [20, 256, 340]
- Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics*. Wiley, Chichester.
- = 2nd ed. of Mardia, K.V. (1972) *Statistics of Directional Data*. Academic Press, London. [189]
- Martínez, V.J. and Saar, E. (2002). *Statistics of the Galaxy Distribution*. Chapman & Hall/CRC, Boca Raton, FL. [21, 517]
- Maruyama, G. (1955). On the Poisson distribution derived from independent random walks. *Nat. Sci. Rep. Ochanomiza Univ.* **6**, 1–6. [174]
- Mase, S. (1986). On the possible form of size distributions for Gibbsian processes of mutuallu non-intersecting discs. *J. Appl. Probab.* **23**, 646–659. [461]
- (1990). Mean characteristics of Gibbsian point processes. *Ann. Inst. Statist. Math.* **42**, 203–220. [461]
- Massoulié, L. (1998). Stability results for a general class of interacting point process dynamics, and applications. *Stoch. Proc. Appl.* **75**, 1–30. [222, 427–439]

- Matérn, B. (1960). *Spatial Variation*. Meddelanded Stat. Skogsforsk. **49** (5), 1–144. [2nd ed. (1986). Lecture Notes in Statistics **36**, Springer-Verlag, New York.] [459]
- Matthes, K. (1972). Infinitely divisible point processes. In *Stochastic Point Processes* (P.A.W. Lewis, ed.), Wiley, New York, pp. 384–404. [66]
- , Kerstan J. and Mecke, J. (1974). See Kerstan, Matthes and Mecke (1974). [269]
- , — and —. (1978). *Infinitely Divisible Point Processes*. Wiley, Chichester. See MKM (1978). [33, 269]
- , — and —. (1982). *Bezgranichno Delimye Tochechnye Protsessy*. Nauka, Moscow. See MKM (1982). [269]
- , Warmuth, W. and Mecke, J. (1979). Bemerkungen zu einer Arbeit von Nguyen Xuan Xanh und Hans Zessin. *Math. Nachr.* **88**, 117–127. [519]
- McFadden, J.A. (1965a). The mixed Poisson process. *Sankhya A* **27**, 83–92. [73]
- (1965b). The entropy of a point process. *J. SIAM* **13**, 988–994. [442]
- McMillan, B. (1953). Absolutely monotone functions. *Ann. Math.* **60**, 467–501. [33]
- Mecke, J. (1967). Stationäre zufällige Masse auf lokal-kompakten abelschen Gruppe. *Z. Wahrs.* **9**, 36–58. [281–292]
- (1968). Eine charakteristische Eigenschaft der doppelt stochastischen Poisson-Prozesse. *Z. Wahrs.* **11**, 74–81. [166]
- (1975). Invarianzeigenschaften allgemeiner Palmscher Maße. *Math. Nachr.* **65**, 335–344. [269, 285–317]
- (1979). An explicit description of Kallenberg’s lattice-type point process. *Math. Nachr.* **89**, 185–195. [483]
- Mertens, J.-F. (1972). Théorie des processus stochastiques généraux: applications aux martingales. *Z. Wahrs.* **22**, 45–68. [399]
- Métivier, M. (1971). Sur la construction de mesures aléatoires presque sûrement absolument continues par rapport à une mesure donnée. *Z. Wahrs.* **20**, 332–344. [50]
- Meyer, P.A. (1971). Démonstration simplifiée d’un théorème Knight. In *Séminaire de Probabilités V* (Springer Lecture Notes in Mathematics **191**), Springer, New York, pp. 191–195. [419]
- Meyn, S.P. and Tweedie, R.L. (1993a). *Markov Chains and Stochastic Stability*. Springer-Verlag, London. [226, 228]
- and — (1993b). Stability of Markovian processes II: Continuous time processes and sampled chains. *Adv. Appl. Probab.* **25**, 487–517. [226–235]
- Mikosch, T. and Wang, Q. (1995). A Monte Carlo method for estimating the correlation exponent. *J. Statist. Physics*, **78**, 799–813. [346]
- Miles, R.E. (1969). Poisson flats in Euclidean spaces, Part I: A finite number of random uniform flats. *Adv. Appl. Probab.* **1**, 211–237. [476]
- (1971). Poisson flats in Euclidean spaces, Part II: Homogeneous Poisson flats and the complementary theorem. *Adv. Appl. Probab.* **3**, 1–43. [476]
- (1974). On the elimination of edge effects in planar sampling. In Harding and Kendall (1974), pp. 228–247. [476]
- Milne, R.K. (1971). *Stochastic analysis of multivariate point processes*. Ph.D. thesis, Australian National University. [89]

- MKM (1974). = Kerstan, Matthes and Mecke (1974). [269]
- MKM (1978). = Matthes, Kerstan and Mecke (1978). [33, 51, 66, 131, 146, 162, 174–178, 198, 213–222, 269, 298, 339]
- MKM (1982). = Matthes, Kerstan and Mecke (1982). [15, 215–221, 339, 519]
- Møller, J. and Waagepetersen, R.P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*, Chapman and Hall/CRC, Boca Raton, FL. [118–130, 508]
- Mönch, G. (1971). Verallgemeinerung eines Satzes von A. Rényi. *Stud. Sci. Math. Hungar.* **6**, 81–90. [33, 35]
- Moran, P.A.P. (1968). *An Introduction to Probability Theory*. Clarendon Press, Oxford. [20, 66]
- (1973). Necessary conditions for Markovian processes on a lattice. *J. Appl. Probab.* **10**, 605–612. [118]
- Moyal, J.E. (1962). The general theory of stochastic population processes. *Acta Math.* **108**, 1–31. [1, 5, 26]
- Musmeci and Vere-Jones (1987). A variable grid algorithm for smoothing of clustered data. *Biometrics* **42**, 483–494. [490]
- Nair, M.G. (1990). Random space change for multiparameter point processes. *Ann. Probab.* **18**, 1222–1231. [419]
- Neuts, M. (1978). Renewal processes of phase type. *Naval Research Logistics Quarterly* **25**, 445–454. [110, 111]
- (1979). A versatile Markovian point process. *J. Appl. Probab.* **16**, 764–779. [110, 111]
- (1989). *Structured Markov Chains of the M/G/1 Type and Their Applications*. Marcel Dekker, New York. [110]
- Neveu, J. (1968). Sur la structure des processus ponctuels stationnaires. *C. R. Acad. Sci. Paris Ser. A* **267**, 561–564. [269]
- (1976). Processus ponctuels. In *Springer Lecture Notes Math.* **598**, Springer, New York, pp. 249–445. [269]
- Nguyen, X.X. and Zessin, H. (1976). Punktprozesse mit Wechselwirkung. *Z. Wahrs.* **37**, 91–126. [15, 242]
- and — (1979a). Ergodic theorems for spatial point processes. *Z. Wahrs.* **48**, 133–158. [198]
- and — (1979b). Integral and differential characterizations of Gibbs processes. *Math. Nachr.* **88**, 105–115. [511, 518]
- Nieuwenhuis, G. (1994). Bridging the gap between a stationary point process and its Palm distribution. *Statist. Neerlandica* **48**, 37–62. [332]
- Ogata, Y. (1988). Statistical models for earthquake occurrence and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83**, 9–27. [419]
- (1998). Space–time point process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50**, 379–402. [499]
- (2004). Space-time model for regional seismicity and detection of crustal stress changes. *J. Geophys. Res.* **109**, B03308, doi:10.1029/2003JB002621 [501]
- and Katsura, K. (1991). Maximum likelihood estimates of the fractal dimension for random spatial patterns. *Biometrika* **78**, 463–474. [21, 253, 498]

- Ogata, Y., Katsura, K. and Tanemura, M. (2003). Modelling of heterogeneous space-time seismic activity and its residual analysis. *Appl. Stat.* **52**, 499–509. [499]
- and Zhuang, J. (2006). Space-time ETAS models and an improved extension. *Tectonophysics* **413**, 13–23. [499]
- Ososkov, G.A. (1956). A limit theorem for flows of similar events (in Russian). *Teor. Veroyat. Primen.* **1**, 274–282. [Translation in *Theory Probab. Appl.* **1**, 248–255.] [146]
- Palm, C. (1943). Intensitätsschwankungen im Fernspreverkehr. *Ericsson Techniks* **44**, 1–189. [146, 268]
- Paloheimo, J.E. (1971). On a theory of search. *Biometrika* **58**, 61–75. [466]
- Papangelou, F. (1970). The Ambrose–Kakutani theorem and the Poisson process. In *Springer Lecture Notes Math.* **160**, 234–240. [268, 269]
- (1972). Integrability of expected increments of point processes and a related random change of scale. *Trans. Amer. Math. Soc.* **165**, 483–506. [420]
- (1974a). On the Palm probabilities of processes of points and processes of lines. In Harding and Kendall (1974), pp. 114–147. [268]
- (1974b). The conditional intensity of general point processes and an application to line processes. *Z. Wahrs.* **28**, 207–226. [506, 520, 527, 531]
- (1978). On the entropy rate of stationary point processes and its discrete approximation. *Z. Wahrs.* **44**, 191–211. [449]
- Parthasarathy, K.R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York. [37, 55, 146–154]
- Penttinen, A.K. (1984). *Modelling Interaction in Spatial Point Patterns: Parameter Estimation by the Maximum Likelihood Method*. Jyväskylä Studies in Computer Science, Economics, and Statistics, **7**. University of Jyväskylä. [130]
- Perez, A. (1959). Information theory with an abstract alphabet (Generalized forms of McMillan's limit theorems for the case of discrete and continuous time). *Teor. Veroyat. Primen.* **4**, 105–109. [Translation in *Theory Probab. Appl.* **4**, 99–102.] [452]
- Preston, C.J. (1976). *Random Fields* (Springer Lecture Notes Math. **534**). Springer-Verlag, New York. [118, 520]
- (1977). Spatial birth-and-death processes. *Bull. Internat. Statist. Inst.* **46**, 371–391. [126]
- Qian, W. and Titterton, D.M. (1990). Parameter estimation for hidden Gibbs chains. *Statist. Probab. Letters* **10**, 49–58. [104]
- Ramaswami, V. (1980). The N/G/1 queue and its detailed analysis. *Adv. Appl. Probab.* **12**, 222–261. [110]
- Rauchenschwandtner, B. (1980). *Gibbsprozesse und Papangeloukerne* (Dissertations of the Johannes Kepler University of Linz, **17**). Verb. Wiss. Gesellsch. Österreichs, Wien. [519]
- Rebolledo, R. (1980). Central limit theorem for local martingales. *Z. Wahrs.* **51**, 269–286. [412]



- Rényi, A. (1956). A characterization of Poisson processes (in Hungarian; Russian and English summaries). *Magyar Tud. Akad. Mat. Kutato Int. Kozl.* **1**, 519–527. [Translation in (1976) *Selected Papers of Alfred Rényi*, Vol. 1 (P. Turan, ed.), pp. 622–628, Akadémiai Kiadó, Budapest.] [165]
- (1959). On the dimension and entropy of probability distributions. *Acta Math. Acad. Sci. Hungar.* **10**, 193–215. [340, 440, 455]
- (1967). Remarks on the Poisson process. *Stud. Sci. Math. Hungar.* **5**, 119–123. [33, 35]
- Resnick, S.I. (1986). Point processes, regular variation and weak convergence. *Adv. Appl. Probab.* **18**, 66–138. [161]
- (1987). *Extreme Values, Regular Variation, and Point Processes*, Springer-Verlag, New York. [161]
- Ripley, B.D. (1976). The second-order analysis of spatial point processes. *J. Appl. Probab.* **13**, 255–266. [191]
- (1977). Modelling spatial patterns (with Discussion). *J. Roy. Statist. Soc. Ser. B* **39**, 172–212. [126]
- (1981). *Spatial Statistics*. John Wiley & Sons, Chichester. [457, 459]
- (1988). *Statistical Inference for Spatial Processes*. Cambridge University Press, Cambridge. [463]
- and Kelly, F.P. (1977). Markov point processes. *J. London Math. Soc.* **15**, 188–192. [118, 122]
- Roberts, W.J.J., Ephraim, Y. and Dieguez, E. (2006). On Rydén’s EM algorithm for estimating MMPPs. *IEEE Signal Process. Lett.* **13**, 373–376. [109]
- Romanowska, M. (1978). Poisson approximation of some probability distributions. *Bull. Acad. Polon. Sci., Ser. Sci. Math. Astr. Phys.* **26**, 1023–1026. [163]
- Rudemo, M. (1964). Dimension and entropy for a class of stochastic processes. *Magyar Tud. Akad. Mat. Kutato Int. Kozl.* **9**, 73–87. [442]
- (1972). Doubly stochastic Poisson processes and process control. *Adv. Appl. Probab.* **4**, 318–338. [101, 410]
- (1973). Point processes generated by transitions of Markov chains. *Adv. Appl. Probab.* **5**, 262–286. [101, 109, 410]
- Ruelle, D. (1969). *Statistical Mechanics: Rigorous Results*. W.A. Benjamin, Reading MA. [118]
- Rydén, T. (1994). Parameter estimation for Markov-modulated Poisson processes. *Stoch. Models* **10**, 795–829. [105]
- (1996). An EM algorithm for estimation in Markov-modulated Poisson processes. *Comp. Statist. Data Anal.* **2**, 431–447. [105–111]
- Ryll-Nardzeweki, C. (1961). Remarks on processes of calls. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **2**, 455–465. [268, 299]
- Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman and Hall, New York. [83]
- Sampath, G. and Srinivasan, S.K. (1977). *Stochastic Models for Spike Trains of Single Neurons* (Springer Lecture Notes Biomath. **10**). Springer-Verlag, Berlin. [247]

- Samuels, S.M. (1965). On the number of successes in independent trials. *Ann. Math. Statist.* **36**, 1272–1278. [163]
- Schoenberg, F.P. (1999). Transforming spatial point processes into Poisson processes. *Stoch. Proc. Appl.* **81**, 155–164. [419]
- (2002). On rescaled Poisson processes and the Brownian bridge. *Ann. Inst. Statist. Math.* **54**, 445–457. [502]
- Serinko, R.J. (1994). A consistent approach to least squares estimation of correlation dimension in weak Bernoulli dynamical systems. *Ann. Appl. Probab.* **4**, 1234–1254. [351, 352]
- 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.] [254]
- Sigman, K. (1995). *Stationary Marked Point Processes*. Chapman and Hall, New York. [15, 222–231, 269, 322–333]
- Sinai, Ya.G. (Ed.) (2000). *Dynamical Systems, Ergodic Theory and Applications*, Second, Expanded and Revised Edition. Springer, Berlin. [196]
- Singpurwalla, N.D. and Wilson, S.P. (1999). *Statistical Methods in Software Engineering, Reliability and Risk*. Springer-Verlag, New York. [100]
- SKM (1987). = Stoyan *et al.* (1987). [269, 459]
- SKM (1995). = SKM (1987) 2nd ed. [244, 269, 457–460, 472, 483]
- Slivnyak, I.M. (1962). Some properties of stationary flows of homogeneous random events. *Teor. Veroyat. Primen.* **7**, 347–352. [Translation in *Theory Probab. Appl.* **7**, 336–341.] [268, 281, 299, 308]
- (1966). Stationary streams of homogeneous random events. *Vest. Harkov. Gos. Univ. Ser. Mech. Math.* **32**, 73–116. [268, 299, 308]
- Smith, W.L. (1955). Regenerative stochastic processes. *Proc. Roy. Soc. London Ser. A* **232**, 6–31. [333]
- Snyder, D.L. (1972). Filtering and detection for doubly stochastic Poisson processes. *IEEE Trans. Inf Theory* **IT-18**, 97–102. [101, 409]
- (1975). *Random Point Processes*. Wiley, New York. [409]
- and Miller, M.I. (2000). *Random Point Processes in Time and Space*. Springer-Verlag, New York. [= Snyder (1975), 2nd ed.] [400]
- Stam, A.J. (1967a). On shifting iterated convolutions, I. *Compositio Math.* **17**, 268–280. [174]
- (1967b). On shifting iterated convolutions, II. *Compositio Math.* **18**, 201–228. [174]
- Stone, C.J. (1968). On a theorem of Dobrushin. *Ann. Math. Statist.* **39**, 1391–1401. [169, 174]
- Stoyan, D. (2006). Personal communication, June 2006. [485]
- and Grabarnik, P. (1991). Second-order characteristics for stochastic structures connected with Gibbs point processes. *Math. Nachr.* **151**, 95–100. [516]
- , Kendall, W.S. and Mecke, J. (1987). *Stochastic Geometry and its Applications*. John Wiley & Sons, Chichester. [= SKM (1987); 2nd ed. = SKM (1995).] [269]
- , — and — (1995). = SKM (1995), = Stoyan, Kendall and Mecke (1987) 2nd ed. [269]
- and Mecke, J. (1983). *Stochastische Geometrie: Eine Einführung*. Akademie-Verlag, Berlin. [Precursor of SKM (1987, 1995).] [269, 472]

- Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields*. John Wiley & Sons, Chichester. [20, 463, 517]
- Straf, M. (1972). Weak convergence of stochastic processes with several parameters. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2**, 187–222. [145]
- Takacs, R. (1983). Estimator for the pair potential of a Gibbsian point process. *Johannes Kepler Univ. Linz, Inst. für Math. Inst. Ber.* 238. [514]
- Takens, F. (1985). On the numerical determination of the dimension of an attractor. In Braaksma, B.L.J., Broer, H.W. and Takens, F. (Eds.), *Dynamical Systems and Bifurcations: Proceedings of a Workshop held in Groningen, Netherlands, April 16–20, 1984*. Lecture Notes in Mathematics **1125**, Springer-Verlag, Berlin, pp. 99–106. [346]
- Tanaka, U. and Ogata, Y. (2005). Estimation of parameters for the Neyman–Scott spatial cluster model. Talk to 73rd annual meeting of the Japan Statistical Society. [497]
- Tanemura, M., Ogawa, T. and Ogata, Y. (1983). A new algorithm for three-dimensional Voronoi tessellation. *J. Comput. Phys.* **51**, 191–207. [500]
- Tempel'man, A.A. (1972). Ergodic theorems for general dynamical systems. *Trudy Moskov. Mat. Obsc.* **26**, 95–132. [Translation in *Trans. Moscow Math. Soc.* **26**, 94–132.] [196, 202]
- (1986). *Ergodic Theorems on Groups* (in Russian). Mosklas, Vilnius. [Translated and Revised (1992). *Ergodic Theorems for Group Actions*. Kluwer, Dordrecht.] [196]
- Thedéén, T. (1964). A note on the Poisson tendency in traffic distribution. *Ann. Math. Statist.* **35**, 1823–1824. [174]
- Thorisson, H. (1994). Shift-coupling in continuous time. *Probab. Theory Related Fields* **99**, 477–483. [222–231]
- (1995). On time and cycle stationarity *Stoch. Proc. Appl.* **55**, 183–209. [309, 332]
- (2000). *Coupling, Stationarity, and Regeneration*. Springer, New York. [132, 222–230, 309–316]
- Timár, A. (2004). Tree and grid factors for general point processes. *Electron. Commun. Probab.* **9**, 53–59. [269]
- Torrise, G.L. (2002). A class of interacting marked point processes: rate of convergence to equilibrium. *J. Appl. Probab.* **39**, 137–160. [428]
- Turner, T.R., Cameron, M.A. and Thomson, P.J. (1998). Hidden Markov chains in generalized linear models. *Canadian J. Statist.* **26**, 107–125. [104, 107]
- Utsu, T. and Ogata, Y. (1997). Statistical analysis of seismicity. In *Algorithms for Earthquake Statistics and Prediction*. IASPEI Software Library (Internat. Assoc. Seismology and Physics of the Earth's Interior) **6**, 13–94. [500]
- Van der Hoeven, P.C.T. (1982). Une projection de processus ponctuels. *Z. Wahrs.* **61**, 483–499. [534]
- (1983). *On Point Processes* (Mathematical Centre Tracts **167**). Mathematisch Centrum, Amsterdam. [534]
- van Lieshout, M.N.M. (1995). *Stochastic Geometry Models in Image Analysis and Spatial Statistics*. CWI Tract **108**, Amsterdam. [457]

- van Lieshout, M.N.M. (2000). *Markov Point Processes and their Applications*. Imperial College Press, London. [118–126, 459]
- (2006a). Markovianity in space and time. In D. Denteneer, F. den Hollander and E. Verbitsky (Eds.), *Dynamics & Stochastics: Festschrift in Honour of M.S. Keane* (Lecture Notes—Monograph Series **48**), Institute for Mathematical Statistics, Beachwood, pp. 154–168. [129, 460]
- (2006b). A  $J$ -function for marked point patterns. *Ann. Inst. Statist. Math.* **58**, 235–259. [462, 466]
- and Baddeley, A.J. (1996). A nonparametric measure of spatial interaction in point patterns. *Statist. Neerlandica* **50**, 344–361. [460]
- Vere-Jones, D. (1968). Some applications of probability generating functionals to the study of input/output streams. *J. Roy. Statist. Soc. Ser. B* **30**, 321–333. [174, 429, 431]
- (1971). The covariance measure of a weakly stationary random measure. *J. Roy. Statist. Soc. Ser. B* **33**, 426–428. [Appendix to Daley (1971).] [237]
- (1975). On updating algorithms and inference for stochastic point processes. In Gani, J. (Ed.), *Perspectives in Probability and Statistics*, Applied Probability Trust, Sheffield, and Academic Press, London, pp. 239–259. [101, 103]
- (1992). Statistical methods for the description and display of earthquake catalogs. In Walden, A.T. and Guttorp, P. (Eds.), *Statistics in the Environmental & Earth Sciences*, Edward Arnold, London, pp. 220–246. [490]
- (1999). On the fractal dimensions of point patterns. *Adv. Appl. Probab.* **31**, 643–663. [341–354]
- (2005). A class of self-similar random measure. *Adv. Appl. Probab.* **37**, 908–914. [263]
- (2007). Some models and procedures for space–time point processes. *Environ. Ecol. Statist.* **14** [Special issue on forest fires] (to appear). [487]
- and Ogata, Y. (1984). On the moments of a self-correcting process. *J. Appl. Probab.* **21**, 335–342. [418]
- and Schoenberg, F.R. (2004). Rescaling marked point processes. *Aust. N. Z. J. Statist.* **46**, 133–143. [423]
- Volkonski, V.A. (1960). An ergodic theorem on the distribution of fades. *Teor. Veroyat. Primen.* **5**, 357–360. [Translation in *Theory Probab. Appl.* **5**, 323–326.] [45]
- Warren, W.G. (1962). *Contributions to the study of spatial point processes*. Ph.D. thesis, University of North Carolina, Chapel Hill (Statist. Dept. Mimeo Series 337). [459]
- (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, 87–118. [459–461]
- Watanabe, S. (1964). On discontinuous additive functionals and Lévy measures of a Markov process. *Japanese J. Math.* **34**, 53–70. [365, 418]
- Wegmann, H. (1977). Characterization of Palm distributions and infinitely divisible random measures. *Z. Wahrs.* **39**, 257–262. [272, 282]
- Westcott, M. (1970). Identifiability in linear processes. *Z. Wahrs.* **16**, 39–46. [74]
- (1971). On existence and mixing results for cluster point processes. *J. Roy. Statist. Soc. Ser. B* **33**, 290–300. [213]

- Westcott, M. (1972). The probability generating functional. *J. Aust. Math. Soc.* **14**, 448–466. [59, 75, 210]
- (1976). A simple proof of a result on thinned point processes. *Ann. Probab.* **4**, 89–90. [156]
- Wheater, H.S., Isham, V.S., Cox, D.R., Chandler, R.E., Kakou, A., Oh, L., Onof, C. and Rodroguéz-Iturbe, I. (2000). Spatial-temporal rainfall fields: Modelling and statistical aspects. *Hydrol. Earth Syst. Sci.* **4**, 581–601. [486]
- Whittle, P. (1951). *Hypothesis Testing in Time Series Analysis*. Thesis, Uppsala University, Almqvist and Wicksell. [498]
- (1954). On stationary processes in the plane. *Biometrika* **41**, 434–449. [486]
- (1962). Topographic correlation, power-law covariance functions, and diffusion. *Biometrika* **49**, 305–314. [486]
- Widder, D.V. (1941). *The Laplace Transform*. Princeton University Press, Princeton, NJ. [205]
- Widom, B. and Rowlinson, J.S. (1970). A new model for the study of liquid–vapor phase transitions. *J. Chem. Phys.* **52**, 1670–1684. [124]
- Yaglom, A.Ya. (1961). Second-order homogeneous random fields. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **2**, 593–622. [54]
- Yashin, A. (1970). Filtering of jump processes. *Avtomat. i Telemekh.* **1970**(5), 52–58. [Translation in *Automation and Remote Control* **1970**, 725–730.] [101]
- Zähle, U. (1988). Self-similar random measures. I. Notion, carrying Hausdorff dimension, and hyperbolic distribution. *Probab. Theory Related Fields* **80**, 79–100. [256]
- (1990a). Self-similar random measures. II. A generalization to self-affine measures. *Math. Nachr.* **146**, 85–98. [256]
- (1990b). Self-similar random measures. IV. The recursive construction model of Falconer, Graf, and Mauldin and Williams. *Math. Nachr.* **149**, 285–302. [256]
- (1991). Self-similar random measures. III. Self-similar random processes. *Math. Nachr.* **151**, 121–148. [256]
- Zheng, X. (1991). Ergodic theorems for stress release processes. *Stoch. Proc. Appl.* **37**, 239–258. [228]
- Zhuang, J. (2006). Second-order residual analysis of spatiotemporal point processes and applications in model evaluation. *J. Roy. Statist. Soc. Ser. B* **68**, 635–653. [502]
- , Ogata, Y. and Vere-Jones, D. (2002). Stochastic declustering of space–time earthquake occurrences. *J. Amer. Statist. Assoc.* **97**, 369–380. [491]
- , — and — (2004). Analyzing earthquake clustering features by using stochastic declustering. *J. Geophys. Res.* **109**, B05301 doi:10.1029/2003JB002879. [491]
- , Vere-Jones, D., Guan, H., Ogata, Y. and Li, M. (2005). Preliminary analysis of observations on the ultra-low frequency electric field in the Beijing region. In Vere-Jones, D., Ben-Zion, Y. and Zúñiga, R. (Eds.), *Statistical Seismology (Pure Appl. Geophys. (PAGEOPH) 162* (6, 7)), pp. 1367–1396. [501]

# Subject Index

[Page references in slanted font, such as 382, are to Volume I,  
but these references are not intended to be comprehensive.]

- Adjacency relation**  $\sim$ , 120, 127
  - binary relations  $\preceq$ ,  $\succsim$ , 125
  - set-dependent adjacency ( $\tilde{\sim}$ ), 128
- Area-interaction point process**, 124
  - as equilibrium birth-and-death process, 130
- Asymptotic independence**
  - mixing, 206
  - weakly mixing, 206
  - $\psi$ -mixing, 206
  - ergodicity, 206
  - expressed via functionals, 210
- Asymptotic stationarity**, 223
  - $(C, 1)$ -asymptotic stationarity, 223
  - shift-coupling sufficient, 230
  - weak and strong coincide, 326
  - conditions for convergence of moment measures, 236
  - strong asymptotic stationarity, 223
- Atomic measure**, 382
  - counting measure correspondence, 4
  - moment measure characterization, 66
- Avoidance function**, 2, 33, 459
  - determines distribution of simple point process, 35
  - of Cox process, 38
  - use in limit theorems, 166
- Avoidance probability**, 459
- Bartlett spectrum**, 303
  - atom at origin, 205
- Batch Markovian arrival process** (BMAP), 110
  - convergence to equilibrium, 228
  - E–M algorithm for, 114
  - $Q$ -matrix structure, 110
  - representations, 111
  - stationary interval distribution and correlation properties, 118
- Bayesian-type formulae**
  - likelihood ratio, 406
- Bernoulli process**, discrete
  - Papangelou kernel properties, 530
- Bijective point map**, 309
- Binomial probability bounds**, 166
- Birth-and-death process**, 99
  - conditional mark distribution, 100
  - death process in reliability, 100
  - spatial, 126, 130
    - as space–time process, 488
- Blackwell renewal theorem**, 83
  - generalized, 331

- BMAP, 110, *see* Batch Markovian arrival process
- Bonferroni inequalities
  - p.g.fl. bounds, 71
- Borel measure, 384
- Boundedly finite Borel measure, 2
- Bounded measurable function space
  - $\text{BM}(\mathcal{X})$ , 52
- $B$ -selective point map, 311
- $(C, 1)$ -asymptotic stationarity, 326
  - see also* Asymptotic stationarity
- Campbell measure, 268
  - , basic properties
    - characterization, 272, 282
    - definition via extension, 270
    - factorization, 269
    - first moment measure relation, 273
    - for a.s. atomic random measure, 275
    - structure is singular, 275
    - characterization for point process, 275
  - for KLM measure  $\tilde{Q}$ , 272, 282
  - for random measures, 284
  - higher-order analogues, 272
  - 'modified' *v.* 'reduced', 270
  - Radon–Nikodym approach, 270
  - refinement of first moment, 270
  - terminology origin, 271
  - , invariant, 285, 293
    - factorization, 293
  - Palm measure characterization, 294
  - stationary random measure
    - invariance characterization, 285
  - , marked
    - from marked cumulative process, 379
    - of MPP, reduced, 331
    - product of predictable kernel and
      - Campbell measure for  $N_g$ , 380
    - semi-Markov process example, 381
- Campbell theorem, 271
  - original, 66
    - Campbell measure precursor, 66
  - refined, 288
- Cantor dust, 20
- Central limit properties
  - simple point process martingale, 412
    - with random scaling, 413
- Characteristic functional, 54
  - Taylor series moment expansion, 68
  - remainder terms, 71
- Characteristic functions
  - continuity condition, 63
  - convergence condition, 64
- Cliques, 121
  - properties, 121, 129
  - maximal, 121
- Cluster iterates, 334
  - Poisson case, 334
    - infinitely divisible, 334
    - method of backward trees, 336
  - stable/unstable dichotomy, 336
    - for critical cluster member process, 337
    - for Neyman–Scott process,
      - determined by random walk, 337
    - for stable cluster member processes,
      - weakly singular infinitely divisible limit, 338
- Cluster process, stationary
  - cluster components, construction, 193
  - from nonstationary components, 192
  - stochastic condition to be
    - well-defined, 192
- Compensator, 241, 358
  - absolute continuity from conditional d.f.s, 364
  - characterizations
    - Cox process, 419
    - Poisson process, 420
  - continuous
    - equals quadratic variation for simple point process, 370
    - variance of integrated martingale, 376
  - convergent sequence of
    - condition for Poisson/Cox limit, 384
    - dependent-thinning example, 387
  - dual predictable projection, 378
    - Campbell measure derivation, 377
  - integral of conditional intensity, 390
  - on whole of  $\mathbb{R}$ , 394
    - complete conditional intensity, 394
  - one-point process, 358
  - unbounded, 376
  - uniquely defined, determines process, 365
- Complete conditional intensity, 394
  - Hawkes process, 395
  - renewal process with density, 395
  - stationary process, history  $\mathcal{H}^\dagger$ , 396
    - hazard function representation, 397

- Complete separable metric space  
(c.s.m.s.), 124, 2
- Completely independent MPP
  - when ground process simple, 86
- Completely monotone set function, 35
  - rôle in determining point process, 36
- Completely random measure, 77
  - fidi distribution structure, 85
  - Laplace functional representation, 86
  - moments, 67
  - sample path structure, 79
    - Laplace transform, 82
  - stationary
    - Laplace–Stieltjes transform representation, 85
- Compound Poisson process
  - extended, 161
  - from random time change, 422
  - moment measures, 72
  - stationary, 180
    - MPP formulation, 181
- Concentration function of distribution, 167
  - tight convergence rate, 175
- Conditional intensity function 211, 231
  - existence of MPP, 429, 432
    - stationary solution, 434
  - weak and strong asymptotic stationarity, 434
  - law of evolution, 432
  - Lipschitz conditions, 430
- Conditional intensity
  - complete (for stationary process), 443, *see also* Complete conditional intensity
  - $\mathcal{F}$ -mark-predictable Radon–Nikodym derivative of marked Campbell measure, 391
  - integral as  $\mathcal{F}$ -predictable intensity of ground process, 391
  - $\mathcal{F}$ -predictable conditional mark distribution, 391
  - in information gain, 443
  - on coarser  $\mathcal{G}$  w.r.t.  $\mathcal{F}$ -conditional intensity, 393
  - for mixed Poisson process, 393
- Conditional intensity measure (for MPP), 399
  - product form for kernel, 399
- Condition  $\Sigma$ , 521
  - examples, 521, 529
- Configurations of points, ergodicity, 202
- Contact distribution function, 460
  - spherical, 459
- Continuity condition for fidi distributions, 28
- Controlled thinning of point process, 387
- Convergence
  - modes of, 131
  - of Campbell measures, 297
  - of fidi distributions, 132
  - of KLM measures, 147
  - of Laplace functionals
    - weak convergence conditions, 138
  - of moments, 141
    - of moment measures, 144
  - of p.g.f.s
    - weak convergence conditions, 138
  - of point processes
    - of probability distributions strong, 132
      - implies weak convergence, 132
    - weak, 132
      - does not imply strong, 134
- Convergence to equilibrium, 223
  - as limit distributions
    - interpretation
      - via inversion theorem, 325
      - via weak convergence, 326
    - Palm, from limit of  $n$ th point, 323
    - stationary, from  $t \rightarrow \infty$ , 323
  - conditional intensity conditions, 427
    - via Poisson embedding, 428
  - Foster–Lyapunov conditions, 228
  - in variation norm, 223
  - space  $\mathcal{Z}$  of initial conditions, 226
    - from Palm distribution, 323
  - see also* Asymptotic stationarity
- Convex averaging sequence, 196
- Copy or version of process
  - same fidi distributions, 11
- Counting measure on  $\mathbb{R}$ 
  - one-to-one correspondence with sequence of intervals, 24
- Coupling, 132, 229
  - conditions for, 231
  - coupling inequality, 133, 229
  - coupling time, 132
  - equivalences, 231
  - with stationary process



- implies strong asymptotic stationarity, 230
  - see also* Shift-coupling
- Covariance density, 69
- Covariance measure of random measure, 69
  - condition to be singular, 70
- Covariant mapping, 309
- Cox process
  - avoidance function, 38
  - class invariant under rarefaction, 166
  - contraction of thinned process, 157
  - convergence to
    - conditions on compensators, 384
    - via dependent-thinning, 387
  - directed by Markov chain, 101
  - likelihood, 102
  - E–M algorithm, 103
  - finite state space, 108
  - MPP extension, 117
  - directed by Markov diffusion, 409
  - directed by partially observed Markov process, 410
  - Neyman–Scott example, 410
  - directed by stationary  $G$ -process, 87
  - from scale-contraction of  $\mathbb{R}^d$ , 157
  - stationary, 181
    - iff directing measure stationary, 192
    - long-range dependent, 254
    - preserved under thinning and translation, 181
- C.s.m.s. (= Complete separable metric space), 124, 2
- Cumulant measures, 69
- Cumulative process, 356
  - adapted to right-continuous history  $\mathcal{F}$ 
    - has unique  $\mathcal{F}$ -compensator, 367
  - family for MPP, 356
  - weak convergence, 143
  - with density, 358
- Cycle stationarity
  - point-stationarity,  $\mathbb{R}^1$  analogue, 308
- Davidson’s ‘big problem’, 481
  - Kallenberg counterexample, 482
- Determining class of set functions, 372, 27
- Deterministic component
  - of random measure, 86
- Deterministic  $\mathcal{H}$ -intensity characterizes Poisson process, 420
- Deterministic lattice process in  $\mathbb{R}^d$ , 192
- Deterministic map of bounded set into itself, fractal dimension, 351
  - moment growth conditions, 352
  - consistent if mixing, 352
- Deterministic point process, 76, 137
  - in  $\mathbb{R}^d$ , 192
  - unit rate, 137
- Diagonal shifts and stationarity, 182
  - reduced measure, 183
- Dirac measure, 382, 3
- Directed lines, random process of, 472
  - see also* Line process
- Directional rose, 467
  - Ripley’s  $K$ -function, 467
  - process in  $\mathbb{R}^2$ , moment factorization, 467
- Dirichlet distribution, 24
- Dirichlet process, 11
  - random probability distribution, 11
  - moment measures, 74
- Discrete-time renewal process
  - $\mathcal{F}$ -compensator, 373
- Dissecting system, 382
  - use in entropy approximation, 446
  - monotone under refinement, 454
  - information gain in limit, 454
  - tiling, infinite analogue of, 311
  - tool to study sample paths, 39
- Dobrushin’s lemma, 45
- Doob–Meyer decomposition, 241, 430, 355
- Dust, 20
  - see also* Cantor, Lévy, 20
- Edge-effects
  - in nearest-neighbour estimation, 463
  - Hanisch-type correction, 466
- E–M algorithm, 101
  - numerical procedures, 103
  - uniformization algorithm, 109, 117
- Empty space ( $F$ )-function, 2, 459
  - estimation, edge-effects, 463
  - Hanisch-type corrections, 466
- Poisson cluster process, 461
  - via local Janossy measure, 461
  - see also* Avoidance function

- Entropy, 440
  - atomic distribution, 440
  - continuous distribution, 440
  - generalized or relative, 440
  - Poisson has maximum entropy, 454
- Entropy dimension, 341
- Entropy rate (process in  $\mathbb{R}^1$ ), 444
  - finite sample approx'n, 449
  - convergence conditions, 450
  - $L_1$  and strong convergence, 451
  - mixed and compound Poisson, 444
  - renewal process, 455
- Entropy rate for intervals, 452
  - relation to entropy rate, 453
- Enumeration, measurable, 14
- Ergodic theorem
  - averaging over group, 196
  - convex averaging sequence, 196
  - for MPP, 197
  - for random measure, 197
  - for weighted averages, 201
  - general, 199
  - higher order, 242
  - individual, 196
  - statistical versions of, 204
  - statistical, 197
- ETAS model, 203
  - nonlinear, 437
  - as space-time model, 499
- Event- (= interval-)stationarity, 327
- Extended Laplace functional, 57
  - sequence of, convergence, 58
- Extended MPP, 7
  - MPP counterexamples, 22
  - purely atomic random measure, 8
- Extended p.g.fl., 60
  - convergence, 64
- Exvisibility (spatial predictability), 513
  - of Papangelou intensity measure, 534
- Factorial moment measure, 133, 69**
  - advantages, 70
- Family of probability measures
  - uniformly tight, 136
- $\mathcal{F}$ -adapted process, 236
  - on  $\mathbb{Z}_+$ , 372
- $\mathcal{F}$ -predictable process, 425, 358
  - on  $\mathbb{Z}_+$ , 372
- $F$ -function, see Empty space function
- Fidi distributions, 2, 25
  - consistent family, 26
  - convergence of random measures, 135
  - equivalent to weak convergence, 135, 137
- Filtering problem, 400
- Finite-dimensional (fidi) distributions, 25, see Fidi distributions
- Finite point process
  - existence, 32
- Fixed atom of random measure, 39
  - at most a countable infinity, 39
  - component of random measure, 86
- Flow, 269
  - on probability space, 177, 179
  - stationary Cox process, 182
- Fractal dimension, Rényi, 340
  - controlled Palm growth, 354
  - deterministic map of set into itself, 351
  - from small-scale clustering, 349
  - $k$ th order,
    - controlled diagonal growth, 347
    - mean-square consistent estimate, 347
  - multinomial measures, 352
- Functionals, linear, 52
- Gamma random measure, 167**
  - stationary case, 162, 11, 30
- Gaussian measures on Hilbert space, 54
- Gauss–Poisson process, 185, 465
  - as limit of u.a.n. array, 154
  - isotropic centred, 468
  - KLM measure structure, 94
  - marked, 332
    - reduced moment densities, 332
  - Papangelou kernel aspects, 530
- Generalized Blackwell theorem, 331
- Generalized compound Poisson process, 61
- Generalized entropy
  - expected likelihood ratio, 441
  - Kullback–Leibler distance, 441
  - reference measure in, 441
- Generalized functions, random, 53
- Georgii–Nguyen–Zessin formula, 462
  - GNZ equation, 511
    - extended form, 513, 536

- Germ–grain models, 503
  - see also* Lilypond model,
  - Particle process, 205
- $G$ -function, 460
  - see also* Nearest-neighbour function
- Gibbs kernel, 525
  - portmanteau Campbell measure, 525
- Gibbs process with pairwise interactions, 507
  - interaction potential, 507
  - process on circle or sphere, 509
- $G$ -random measure, 83
  - directing measure of Cox process, 84
  - Lévy representation, 83
  - shot-noise, 84, 87
- Grassberger–Procaccia estimates, 355
- Ground measure, 3
- Ground process, 194, 7
  - condition to be well-defined, 22
- Haar measure**, 408
  - in transformation invariance, 409, 188
  - MPPs as canonical example, 190
- Hamel equation, 64, 192
  - application in  $\mathbb{R}^d$ , 192
  - variant of, 482
- Hammersley–Clifford theorem, 122
  - extended, 129
- Hawkes process, 183
  - complete conditional intensity, 395
  - convergence to equilibrium, 232, 236
  - strong asymptotic stationarity, 234
  - exponential decay, 99
  - moments, 145
  - nonlinear, 437
    - Lipschitz condition, 437
  - without ancestors, 141, 145
- Hazard function, conditional
  - rôle in likelihoods, 402
- Hereditary class of subsets, 122
- Hereditary function, 122, 128
- Hidden Markov model (HMM), 101, 400
  - Cox process, 103, 400
  - Neyman–Scott example, 410
  - Poisson observations for, 106
  - state estimation, 105
- Higher order ergodic theorem, 242
- Hill estimate, 346
- History, 357
  - minimal  $\mathcal{H}$ , 357
  - right-continuous, 357
  - $\mathcal{F}_{(+)}$ , 373
    - counterexample, 373
- HMM, *see* Hidden Markov model
- Homogeneous point process, 459
- Hougarde process, 83
- Hurst index, 250
  - cluster process, 251
  - Cox process, 252, 254
  - of superpositions, 252
  - stationary renewal process, 254
- Hyperplanes, directed, intersections of, 484
- IHF**, *see* Integrated hazard function
- Independence of random measures, 63
- Infinitely divisible distributions,
  - nonnegative, 82
- Infinitely divisible MPP, 94
  - cluster process representation, 94
  - ground process, 94
    - infinite divisibility, 94
  - KLM measure, 94
  - structure, 94
- Infinitely divisible point process, 87
  - a.s. finite characterization, 92
  - convergence, 147
    - equivalent KLM  $Q$ -weak convergence, 147
  - KLM measure, 89
  - regular/singular components, 92
  - representations
    - cluster process, 89
    - p.g.fl., 91
    - Poisson randomization, 89
  - transformation invariant iff KLM measure invariant, 221
  - , stationary
  - iff KLM measure stationary, 216
  - KLM measure
    - ergodicity, mixing conditions, 218
    - Palm factorization, 295
  - regular/singular dichotomy
    - if regular then ergodic, 217
    - regular iff Poisson cluster process representation, 216
    - singular if Poisson randomization representation, 216
  - weak/strong singular dichotomy, 220

- strongly singular iff Poisson randomization superposition exists, 221
  - weakly singular example, 185
- Infinitely divisible random measure, 87
  - fidi distributions infinitely divisible, 88
  - representation, 95
- Laplace functional representation, 93
- Infinite particle systems
  - external configuration, 519
- Information gain, expected, 276
  - generalized entropy, 441
  - in discrete approx'n to entropy, 447
  - limit of refinement, 454
  - mixed and compound Poisson, 444
- MPP
  - decomposition (points/marks), 444
  - renewal process, 445
  - rôle of conditional intensity, 443
  - simplifies when stationary, 443
  - two processes evolving simultaneously, 448
- Inhomogeneous Poisson process, as
  - Markov point process, 121
- Innovations ('residual') process, 513
- Integrated hazard function (IHF), 108, 359
  - for one-point process, 359
- Integration by parts for Lebesgue–Stieltjes integral, 107, 376
- Intensity measure of point process, 44
  - Korolyuk equation, 46
- Interacting point processes, 244
- Interaction function, 122
- Interaction potential
  - in Gibbs process, 127, 507
- Internal history  $\mathcal{H}$ , 234, 357
  - $\mathcal{H}$ -predictability of process, 374
- Intervals, process of, 302
  - definition of point process, 13
  - marked, stationary, 333
- Interval-stationarity, 268, 299
- Intrinsic history, 234, 357
- Invariant  $\sigma$ -field  $\mathcal{I}$ 
  - trivial
  - invariant functions constant, 204
- Inversion formulae, Palm measure
  - stationary point process, 291, 300
- Isotropy, 297, 466
  - moment measure factorizations, 467
  - rotationally invariant probability, 466
  - MPP analogue, 467
- Iterated convolutions
  - conditions for Poisson limit, 169
  - see also Cluster iterates, Random translations
- Janossy density, 125
  - in Papangelou intensity, 507
  - Markov point process density, 119
- $J$ -function, 460
  - clustering/regularity indicator, 460
  - Poisson cluster process, 461
  - Poisson process, 464
- Jordan–Hahn decomposition, 374, 38, 252
- Khinchin existence theorem, 45
- Khinchin orderliness, 47
- $K$ -function, 464
  - see Ripley's  $K$ -function
- KLM measure, 91
  - extended, convergence of, 147
  - Gauss–Poisson process, 94
  - infinitely divisible MPP, 94
  - Palm factorization for stationary
    - infinitely divisible point process, 295
    - regular, strongly or weakly singular, 295
  - $Q$ -weak convergence, 147
  - see also Infinitely divisible point process
- Kolmogorov consistency conditions, 27
- Kolmogorov existence theorem
  - for point process, 30
  - for random measure, 28
- Korolyuk's theorem, 45
  - generalized Korolyuk equation, 46
  - purely atomic random measure, 50
- Kullback–Leibler distance
  - expected information gain, 441
  - generalized entropy, 441
- Laplace functional, 57
  - characteristic functional analogue, 64
  - completely random measure, 86
  - convergence conditions, 64
  - expansion, first-order, 75
  - Taylor series, moments, 75
  - extended, 57

- Palm kernel characterization, 280
  - random measure characterization, 57
- Lebesgue–Stieltjes integration by parts, 107, 376
- Level-crossing, Poisson limit, 159
- Lévy dust, 20
- Lévy process
  - nonnegative, 82
  - subordinator, 20
- Lifted operator, 178
- Likelihood, MPP, point process
  - existence when  $\mathcal{H}$ -adapted, 401
  - from ground process intensity and conditional mark density, 403
  - reference probability measure, 401
- Likelihood ratio, point process
  - as Radon–Nikodym derivative, 411
  - from  $\mathcal{H}$ -conditional intensity, 402
  - $\mathcal{H}$ -local martingale, 404
  - in Bayesian-type formulae, 406
  - Poisson process with mixed atomic and continuous parameter, 411
- Lilypond protocol models, 503
  - algebraic specification, 505
  - germ–grain model, 503
    - absence of percolation, 504
    - volume fraction, 504
- Linear functionals, 52
- Linear process of random measure, 74
  - shot-noise example, 74
- Line process
  - clustered, 484
  - coordinate representations, 472
  - directed/undirected, 472, 483
  - in  $\mathbb{R}^3$ , 484
  - line measure, 474
  - Poisson, 472
    - stationary, 476
      - no parallel/antiparallel lines, 485
  - railway line process, 484
  - , stationary
    - directional rose, 474
    - isotropic, 474
      - second-order properties, 477
    - parallel/antiparallel lines, 478
    - reflection invariant, 478
  - mean density, 474
  - moment structure, 473
  - shear-invariant, 473, 482
    - moment factors, 474
- Line segment process
  - from lilypond model, 504
  - yielding random measure, 23
- Lipschitz conditions on conditional intensity functions, 430
- Local martingale, 358
  - integral w.r.t. martingale, 373
- Local Palm distribution  $\mathcal{P}_{(x,\kappa)}$ , 273
  - element of Palm kernel, 273
  - for modified Campbell measure, 298
  - for MPP, 317
  - for nonsimple point process, 282
  - for second-order Campbell measure, 298
  - for stationary process, 274
  - higher-order family, 281, 282, 284
  - random measure with density, 274
- Long-range dependence, 250
  - renewal process, 250, 254
  - cluster process, 251
  - covariance measure decomposition, 252
    - deterministic process, 253
  - ‘power-law decay’, 254
- MAP, *see* Markovian arrival process
- Mark distribution, stationary
  - in ergodic theorem for MPP, 197
- Marked Campbell measure, 317, 379
- Marked cumulative process, 378
  - marked Campbell measure, 379
- Marked point process (MPP), 194, 7
  - embedded regeneration points, 328
  - Palm moment inequalities, 333
  - ergodic theorem, 199
    - stationary mark distribution, 197
  - extended MPP, 7
  - individual ergodic theorem, 318
  - nonergodic case extension, 321
    - for marked random measure, 322
  - information gain, 443
  - points *v.* marks decomposition, 444
  - likelihood
    - existence when  $\mathcal{H}$ -adapted, 401
- Palm distributions for
  - mean Palm distribution  $\bar{\mathcal{P}}_0$ , 323
  - local Palm distributions  $\mathcal{P}_{(x,\kappa)}$ , 278, 317
    - for ground process, 279, 283
- representation on  $\mathcal{X} \times \mathcal{K}^\cup$ , simple, 23
  - with simple ground process, 22

- random rescaling to compound Poisson, 422
- rescaling counterexample, 425
- rescaling to two-dimensional Poisson process, 424
- , identified processes
  - completely random, 84
  - with simple ground process, 84
  - structure of, 85
- exponentially distributed intervals, 332
- marked Gauss–Poisson process, 332
  - reduced moment densities, 332
- on  $\mathbb{S}$ , stationary, 190
  - conditions for stationarity, 191
- , stationary, 197
- asymptotic stationarity, 223
  - $(C, 1)$ -asymptotic stationarity, 223
  - weak and strong coincide, 231
  - inversion theorem analogue, 327
- averaged (= mean) Palm distribution  $\bar{\mathcal{P}}_0$ , 319, 323
- complete conditional intensity, 399
- convergence to equilibrium, 323
- coupling, shift-coupling equivalences, 230
- ergodic,  $L_2$  convergence for, 204
- higher-order mark distributions, 321
- independent unpredictable marks,
  - ergodic limits, 319
- inversion theorem for Palm and stationary distributions, 324
- $k$ th order stationarity, 237
- mark-dependent Palm distributions, 317
- on  $\mathbb{R}^d \times \mathcal{K}$ 
  - a.s. zero–infinity dichotomy, 205
- $\mathcal{P}, \mathcal{P}_0$  invariant  $\sigma$ -algebras equivalent, 333
- reduced Campbell measure, 318, 331
- reduced moment measures, 238
  - higher-order ergodic theorem, 247
- stationary mark distribution, 197, 317
- Markov density function, 120
- Markov modulated Poisson process (MMPP), 101
  - convergence to equilibrium, 228
- Markov point process, 118
  - density function for, 119
  - Janossy densities, 119
  - Hammersley–Clifford theorem, 122
    - extended, 129
    - products of Markov density functions, 125
  - Papangelou conditional intensity, 120
  - simple finite point process with density, 120
- Markov renewal function, 115
  - mean Palm, subadditive, 332
- Markov renewal process (MRP), 96
  - as MPP, 98
  - point process properties, 98
  - ground process, renewal function, 332
  - factorial moment densities, 99
- Markov renewal function, 115
- Markov renewal operator, 98
- observed on subset, 115
- renewal function analogue, 99, 332
  - subadditivity, 332
- semi-Markov process equivalence, 97
  - see also Semi-Markov process
- Markovian arrival process (MAP), 110
  - $Q$ -matrix structure, 110
- Martingale, 427
  - integral w.r.t., local martingale property, 373
  - quadratic variation of, 368
    - increments as conditional variances, 369
  - see also Point process martingale
- Maximal clique, 121
- Mean Palm distribution  $\bar{\mathcal{P}}_0$  for MPP, 323
- Measurable enumeration, 14, 24
- Measure, decomposition of, 4
- Method of backward trees for cluster iterates, 336
- Method of reduced trees for cluster iterates, 336
- Metric transitivity
  - implies ergodicity, 204
  - of transformation, 194
  - ergodic theorem for, 201
- M/G/ $\infty$  queue, spatial
  - MPP in time, 488
  - space–time model, 488
- Minimal history  $\mathcal{H}$ , 357

- Mixed Poisson process
  - conditional intensities on  $\mathcal{G} \subset \mathcal{F}$ , 393
  - conditions to be ordinary, 51
  - contraction of thinned process, 156
  - defined by moments, condition, 73
  - fidi distributions, 62
  - limit of points moving with random velocities, 172
  - limit of random translations of non-ergodic process, 171
  - moment measures, 72
  - overall Palm probability in ergodic limit, 323
  - p.g.fl., 61
  - random translation invariant, 172
  - , stationary
    - mixing iff Poisson, 212
- Mixed random measure, moments, 67
- Mixing process, 206
  - $k$ th order mixing, 215
  - trivial tail  $\sigma$ -algebra, 215
- MMPP, *see* Markov modulated Poisson process
- Modified Campbell measure, 271
  - analysis of Papangelou kernel, 521
  - disintegration of, 524
  - in GNZ equation, 512
  - local Palm distributions for, 298
  - ‘reduced’ terminology, 270
- Moment measure, 65
  - Campbell theorem, 66
  - expectation measure, 65
  - higher-order, symmetric, 75
  - isotropy, factorization, 467
  - Ripley  $K$ -function and directional rose, 467
  - $k$ th order, 66
  - of diagonal of power set, 75
    - bound on multiplicity, 75
  - stationary
    - diagonal factorization, 237
    - reduced moment estimation, 244
- MPP, *see* Marked point process
- MRP, *see* Markov renewal process
- Multidimensional random measure, 7
- Multivariate characteristic function, 54
  - characteristic functional, 54
- Multivariate infinite divisibility, 94
- Multivariate Poisson process
  - independent components, as limit, 152
- Multivariate random measure, 7
- Mutual nearest-neighbour matching, 311
- Natural history, 357
- Nearest-neighbour distances, 459
  - stationary MPP, 462
  - independent marks, 462
- Nearest-neighbour ( $G$ -)function, 460
  - as Palm–Khinchin limit, 466
  - estimation, edge-effects, 463
  - Hanisch-type corrections, 466
- Matérn’s Model I, 466
  - stationary MPP, 462
  - independent marks, 462
  - via local Janossy measure, 461
- Nearest-neighbour matching, 311
- Negative binomial process, 10, 23, 73
  - defined via two mechanisms, 73
  - local properties, 73
- Neighbourhood relation, *see* Adjacency
- Neyman–Scott cluster process, 181
  - analogue on circle  $\mathbb{S}$ , 189
  - bivariate, nonisotropic, isotropic, 471
  - hidden structure, 410
  - in  $\mathbb{R}^2$ , 464
    - conditional intensity, 465
  - space–time, singular components, 349
    - different orders of growth, 350, 355
  - stationary
    - regular representation, 185
- Nonergodic process
  - Bartlett spectral mass at origin, 205
- Nonnegative increment process, 11
- Nonnegative infinitely divisible distributions, 82
- Nonnegative Lévy process, 82
- Nonnegative stable processes, 83
  - Laplace–Stieltjes transform, 85
- Nonstationary point process counter-examples
  - avoidance function stationary, 191
  - MMPP, 455
  - one-dimensional distributions
    - stationary, 191
- One-point process
  - convergence, 137
  - $\mathcal{H}$ -compensator, 358
    - absolutely continuous or continuous, 360
  - randomized hazard function, 362

- square integrable
  - quadratic variation, 371
- with prior  $\sigma$ -algebra  $\mathcal{F}_0$ 
  - compensator w.r.t. intrinsic history, 360
- , MPP
  - compensator, 389
  - for ground process, 389
- Orderly point process, 46
  - equivalent conditions, 51
  - Khinchin orderly, 47
  - $\mu$ -orderly, 46
  - terminology: orderly, ordinary, 51
- Ordinary point process, 46
  - simple, 51
- Ord's process, 130
- Palm, theory, 268**
  - queueing theory applications, 269
  - stochastic geometry applications, 269
- Palm distributions, measure, 273
  - for MPP, 277
  - from Campbell measure, 273
  - local, 273
  - Palm kernel, 273
- , moment measures, 292
  - as Radon–Nikodym derivatives, 293, 298
  - relation to stationary moments, 292
- Palm kernel, for Poisson process, 280
  - Laplace functional characterization, 280
  - see also Local Palm distribution
- Palm measure,  $\sigma$ -finite
  - Campbell measure characterization, 293
  - factorization characterization, 294
- , for MPP  $\mathcal{P}_{(x,\kappa)}$ , 317
  - mean ('average')  $\bar{\mathcal{P}}_0$ , 323
  - Sigman's alternative version, 332
- , stationary point process
  - inversion formulae, 291, 300
- Palm probabilities as rates, 290
- Palm–Khinchin equations
  - generalization, 302
  - point process in  $\mathbb{R}^d$ , 308
- Papangelou (conditional) intensity, 120, 506
  - conditional probability interpretation, 507
  - defined via Janossy densities, 507
  - density of Papangelou kernel, 524
  - density of Papangelou measure, 536
  - higher-order, 507
  - integral relations, 508
  - Markov density function rôle, 120
  - multiplicative relations, 510
    - conditional Papangelou kernels, 525
    - relation to Palm densities, 507
- Papangelou (intensity) measure
  - exvisibility property, 534
  - relation to first moment, 533
- Papangelou kernel, 520
  - atomic part, 520, 527
  - discrete Bernoulli example, 530
  - Gauss–Poisson example, 530
  - higher-order, 525
    - Gibbs (portmanteau) kernel, 525
  - Papangelou (intensity) measure, 520
  - Papangelou intensity as density, 524
  - viewed as random measure, 526
- Penetrable spheres model, 124
- P.g.fl., 59
  - see Probability generating functional
- PH-distributions, 111
  - representation  $(Q, \pi)$ , 111
  - Laplace transforms, 117
- Planar point process (= point process in  $\mathbb{R}^2$ )
  - isotropic, 466
    - Poisson, centred, 468
    - Gauss–Poisson, centred, 468
  - moments, polar coordinates and factorization, 467
  - stationary
    - isotropic, 469
      - counterexample, 205
    - moment measure factorization, 470
    - Ripley  $K$ -function and directional rose, 470
- Point map
  - bijective, 309
  - $B$ -selective, 311
  - covariant mapping, 309
  - for point-stationarity, 308
  - inverse, 309
  - properties, 317
- Point process (see also individual entries)
  - , basic properties
    - distribution, 26
      - determined by fidi distributions, 26



- ergodic limit a.s. constant
  - trivial invariant  $\sigma$ -algebra, 206, 208
  - metrically transitive, 206
- formal definition, 7
- integer-valued random measure, 7
- intensity measure, 44
- interval sequence definition, 13, 302
- Kolmogorov consistency conditions, 30
- nonnull, 188
- orderly, ordinary, 46
- random variable mapping characterization, 8
- simple, simplicity of, 47, 7, 43
- unaligned, 312
- , general properties
  - asymptotic independence
    - expressed via functionals, 210
    - mixing, 206
    - $\psi$ -mixing, 206
    - weakly mixing, 206
  - conditional intensity
    - local Poisson character, 370
  - entropy
    - discrete trial approximation, 445
    - dissecting partitions, 446
  - filtering, 400, 407
    - Cox process and HMM, 400
    - unit rate Poisson as reference, 407
    - mixed Poisson example, 407
    - updating formulae, 407
    - time series analogies, 401
  - likelihood
    - existence when  $\mathcal{H}$ -adapted, 401
  - likelihood ratio
    - from  $\mathcal{H}$ -conditional intensity, 402
    - $\mathcal{H}$ -local martingale, 404
  - martingale, for simple process, 247, 368, 412
    - $\mathcal{F}$ -compensator via IHFs of conditional d.f.s, 364
    - quadratic variation, 370
    - randomly scaled to normal limit, 415
    - (mixed) Poisson example, 416
  - , stationary, in  $\mathbb{R}$ , 178
    - complete conditional intensity, 396
    - MPP extension, 399
  - $k$ th order moments and  $k$ -point configurations, 244
  - $k$ th order stationarity, 237
    - reduced moment measures, 238
  - long-range dependence, 250
    - Hurst index, 250
  - Palm measure
    - a.s. zero-infinity dichotomy, 299
    - infinite intensity possible, 302
  - , stationary, in  $\mathbb{R}^d$ , 304
    - points enumerated by distance from origin, 305
      - a.s. no points equidistant from origin, 304
  - Palm measure bijective point map invariant, 312
  - , identified processes
    - controlled thinning, 387
    - in  $\mathbb{R}^2$ , see Planar point process
    - multivariate, 7
      - quadratic variation, 370
      - see also MPP
  - on circle  $\mathbb{S}$ 
    - stationary, 289
    - reduced moment measures, 298
  - on surface of cone, 471
    - two parameterizations, 471
  - on  $\mathbb{Z}_+$ , 372
    - $\mathcal{F}$ -adapted, -predictable, 372
    - $\mathcal{F}$ -compensator, 372
    - discrete-time renewal process, 373, 375
    - quadratic variation, 375
- Point-stationarity, 269, 299, 312
- Poisson approximants
  - inequality, 153
  - tool for convergence conditions, 154
- Poisson cluster process
  - convergence to equilibrium, 236
    - strong and geometric asymptotic stationarity, 236
  - empty space function, 461
  - stationary, regular representation, 183
    - nonuniqueness, 184
    - Khinchin measures, 184
    - Neyman-Scott process, 185
- Poisson distribution
  - maximum entropy, 454
- Poisson embedding
  - for general history, 438
  - history-dependent thinning, 426
- Poisson probabilities
  - bounds, 166

- Poisson process, convergence to
  - conditions on compensators, 384
  - dependent-thinning example, 387
- Poisson process
  - , basic properties
    - existence, 31
    - infinitely divisible, 91
    - isotropic centred, 468
    - moment measures, 72
    - ordinary iff simple, 51
    - p.g.fl., 60
    - stationary, 180
  - , characterization
    - class invariant under rarefaction, 165
    - deterministic  $\mathcal{H}$ -intensity, 420
    - inverse compensator scaling, 421
    - Palm kernel and Slivnyak–Mecke theorem, 281
    - Watanabe’s, 418
  - , general properties
    - configurations of points, 202
    - information gain, 444
    - maximum entropy rate, 454
    - Palm kernel, 280
    - Poisson property lost under random shift, 316
  - , limit properties
    - conditions on compensators, 384
    - dependent-thinning example, 387
    - contraction of thinned process, 155
    - multivariate, 151
    - of high level-crossing, 159
    - of dilated superposition, 150
    - extensions, 154
    - variation norm limit, 161
- Poisson randomization
  - representation for infinitely divisible process, 89
  - KLM measure totally finite, 93
- Power-law growth, 342
- Predictability, 355
- Probability generating functional
  - (p.g.fl.), 144, 59
  - characterization for point process, 60
  - convergence conditions, 64
  - counterexample, 64
  - expansions of
    - via factorial moments, 70
    - factorial cumulant measure, 71
    - moments, 70
  - extended p.g.fl., 60
    - condition for convergence, 64
    - infinitely divisible point process, 91
- Prohorov metric, 145
- $\psi$ -mixing stationary random measure
  - central limit property, 214
- Purely atomic measure, 3
  - see also Random measure, purely atomic
- Q**uadratic random measure, 9
  - integral as random measure, 42
  - moments, 67
- Quadratic variation process of martin-gale, 368
  - atomic and continuous components, 370, 375
  - equals compensator if continuous for simple point process, 370
  - for simple and multivariate point processes, 370
  - square integrable one-point process, 371
- $Q$ -weak convergence, 147
- R**ailway line process, 484
- Random distribution, 11, 23
  - absolutely continuous, 50
  - nonatomic condition, 75
- Random linear functionals
  - strict and broad sense, 53
- Random Markov shifts, 173
- Random measure, 6
  - , basic properties
    - characterization
      - characteristic functional, 56
    - distribution, 26
      - determined by fidi distributions, 26
      - Laplace functional, 57
      - random variables, 54
    - sample path components, 86
      - deterministic, 86
      - fixed atom, 39, 86
        - condition to be free of, 39
        - nonatomic a.s., condition, 41
      - set-indexed family of r.v.s, 17
    - tail  $\sigma$ -algebra, 208
  - , general properties
    - superposition, 63

- , identified
  - completely random,
    - see Completely random measure
  - integral of nonnegative process, 23
    - as random sum of discs or lines, 23
  - linear process of, 74
    - shot-noise example, 74
  - multidimensional, 7
  - multivariate, 7
  - nonnull, 188
  - on  $\mathbb{R}_+$ ,
    - history for  $\mathcal{F}$ -adapted process, 357
    - internal (natural, minimal), 357
  - with  $\chi^2$  density, 9, 42
- , purely atomic, 7
  - characterization
    - as point process, 275
    - via Campbell measure, 275
    - via construction, 49
  - extended MPP, 8
  - finitely many atoms, example, 277
    - Palm kernel, 277
  - Korolyuk equation for, 50
  - moment condition for, 66
  - on countable state space, 282
    - Campbell measure for, 282
- , stationary, 178
  - Bartlett spectrum atom at origin, 205
    - nonzero iff process nonergodic, 205
  - Campbell measure characterizations, 285
  - $k$ th order stationarity, 237
  - long-range dependence, 250, 254
    - Hurst index, 250
  - moment properties when absolutely continuous, 248
  - on  $\mathbb{R}^d$ , zero-infinity dichotomy, 187
  - Palm measure, distribution, 288
  - reduced moment measures, 238
- Random probability distribution,
  - see Random distribution
- Random process of directed lines, 472
  - see also Line process
- Random Schwartz distributions, 53
- Random time change (= Randomly rescaled point process)
  - Poisson, 421
  - proofs using p.g.fl., 419
  - proofs via interval distributions, 426
  - stationary compound Poisson process, 422
    - see also Poisson embedding
- Random thinning, see Thinning
- Random translations, 166
  - conditions for Poisson limit, 169
  - iterated convolution, u.a.n. property, 167
- Random variable mapping
  - characterization of point process or random measure, 8
- Random walk, point process boundedly finite, 24
- Random walk renewal process
  - as space-time process, 494
  - moment measures, 505
- Random walk on circle  $\mathbb{S}$ , 193
- Rarefaction, see Thinning
- Rayleigh-Lévy dust or flight, 21
- Reduced measure
  - by diagonal factorization, 183
  - Campbell measure for MPP, 331
  - moment measures
    - for stationary random measures, 238
    - properties, 239
    - scale factors in, 238
  - see also Modified Campbell measure
- Reference measure
  - in generalized entropy, 441
  - in likelihood, 401
- Refined Campbell theorem, 288
- Regeneration points in MPP, 328
- Regenerative measure
  - stationary, 186
- Regular conditional probabilities 380
  - on product space, 25
- Regular infinite divisibility, 92
  - a.s. finite, 92
  - KLM measure totally finite, 92
  - Poisson cluster process
    - representation, 93
- Relative entropy, 440
- Renewal process, 67
  - class invariant under rarefaction, 165
  - compensator and martingale for, 366
  - convergence to equilibrium, 227
  - entropy, gamma lifetimes, 454
  - not a Markov point process, 127
  - on circle  $\mathbb{S}$ , 193
  - phase type (PH-distributions), 111
  - Poisson limit under rarefaction, 165
  - recurrence relation for p.g.fl., 65

- stationary, 186
  - Palm–Khinchin equations, 302
  - reduced moment, cumulant measures, 247
  - with density
    - complete conditional intensity, 395
- Renewal theorem, 83
  - Blackwell, generalized, 331
  - for random walk on  $\mathbb{S}$ , 193
- Rényi dimensions, 341
  - correlation integrals, 341
  - quasi* factorial moment estimator, 343
  - consistent estimators, 343
  - discrete entropy approximation, 440
  - multifractal, 341
  - consistent estimator for, 343
  - unifractal, 341
- Rényi’s dimensional entropy, 455
- Rényi–Monch theorem, 35
  - Poisson limit from one-dimensional distributions, 162
  - use of variation norm, 162
- Repairable system model, 235
- Residual or ‘innovations’ process, 513
  - exvisibility, 513
- Right-continuity of histories, 357
  - $\mathcal{F}_{(+)}$ , 373
  - counterexample, 373
- Ripley’s  $K$ -function, 297, 464
  - radial component of moment factorization of process in  $\mathbb{R}^2$ , 467
- Scale-invariance, 255
- Schwartz distributions, random, 53
- Self-similarity, 83, 255
  - in point configurations, 249
  - of random measure, 83
  - stable random measures, 83
- Semi-Markov process, 96
  - compensator components, 381
  - conditional intensity and mark distribution, 381
  - Markov process simplification, 382
  - equivalent Markov renewal process, 97
  - likelihood, 116
  - point process properties, 98
  - see also* Markov renewal process
- Set function,
  - subadditive under refinement, 50
- Set-dependent adjacency ( $\tilde{\sim}$ ), 128
- Shift-coupling, 229
  - in limit theorems, 269
  - with stationary process
    - implies strong  $(C, 1)$ -asymptotic stationarity, 230
  - see also* Coupling
- Shift transformations, 177
  - shift operator, 178
- Shot-noise process, 163, 170
  - as linear process, 74
  - from  $G$ -random measure, 84, 87
- Signed random measure, 19
  - Wiener motion counterexample, 19
- Simple birth process
  - maximum likelihood estimator
  - asymptotically mixed normal, 416
- Simple counting measure, 3
- Simple point processes
  - distribution determined by avoidance function, 35
  - sample path property, 43
  - second moment measure sufficient condition, 66
  - sequence of, sufficient conditions for convergence, 140
- Singular infinite divisibility, 92
- Skorohod metric, 145
- Slivnyak–Mecke Poisson process
  - characterization theorem, 281
- Smoothing problem, 400
- Space of counting measures  $\mathcal{N}_{\mathcal{X}}^{\#}$ , 3
  - as c.s.m.s., 6
  - closed subspace of  $\mathcal{M}_{\mathcal{X}}^{\#}$ , 6
- Space of measures, as c.s.m.s., 6
- Space–time process
  - , general 485
  - estimation in, 490
  - evolving spatial field, 486
  - residual analysis diagnostics, 502
  - second moment estimation, 496
    - Bartlett spectrum, 497
    - boundary (edge) effects, 497
    - conditional intensity, 498
  - stochastic declustering, 491
  - stochastic reconstruction, 491
  - variety of processes, 486
  - with associated mark, 486

- , models 505
  - cluster, Bartlett–Lewis, 505
  - cluster, Neyman–Scott, 505
  - ETAS model, 499
  - M/G/ $\infty$  queue, 488
  - Poisson process, 348
  - spatial birth-and-death, 488
- , stationary
  - family of Palm distributions, 494
  - first moments, 487
  - Poisson, 487
    - Poisson cluster process, 495
    - terminology, 487
  - reduced second moments, 491
    - alternative representations, 492
    - Fourier transforms, 493
    - simplified when homogeneous, 492
  - stationary-time, homogeneous-space, 487
- Spatial birth-and-death process, 126
- Spatial point pattern, 458
  - models, 459
  - statistics, 459
    - diagnostic tests, 514
    - residual variances, 516, 518
    - reduced moment measures, 464
- Spherical contact distribution, 459
- Stable convergence, 419
  - identifying Poisson/Cox limit of convergent compensators, 384
  - dependent-thinning example, 387
  - point process martingale, 412
  - randomly scaled to normal limit, 415
- Stable random measures, 83
  - self-similarity, 83
- Stationarity, 178
  - on half-line
    - extension to whole line, 223, 235
    - preserved by random thinning, 181
    - preserved by random translation, 181
  - strict *v.* weak, 178
  - see also Asymptotic stationarity
- Stationary cluster process
  - asymptotic independence
    - determined by cluster centre process, 213
- Stationary cubic lattice process, 192
- Stationary gamma random measure, 162, 11, 30
- Stationary independent increment process, 81
- Stationary isotropic planar point process, 469
- Stationary MPP, 179
- Stationary point process
  - on  $\mathbb{R}^d$ , 178
    - a.s. zero–infinity dichotomy, 187
    - MPP extension, 205
    - functional equivalence, 180
  - on circle  $\mathbb{S}$ , 188
    - extension to MPP, 190
  - Palm measure inversion formulae, 291, 300
- Stationary random measure
  - ergodicity implies nontrivial mixture impossible, 217
  - on  $\mathbb{R}^d$ , 178
    - a.s. zero–infinity dichotomy, 187
    - functional equivalence, 180
- Stein–Chen methods, 163
- Stochastic continuity sets, 134
  - form an algebra, 135
- Stochastically continuous, 11
- Stochastic declustering, space–time model, 491
  - stochastic reconstruction, 491
- Stochastic integral w.r.t. Poisson process, 428
  - stochastic d.e. driven by Poisson process, 428
  - thinning construction, 427
- Stopping time sequence
  - properties of limits, 373
- Strauss process, 123
  - characterization, 123
  - extended, 130
- Stress-release model, 239
  - limit properties of estimators, 418
  - MPP variant, 235
- Subadditive set function, 43, 50
  - under refinement, 50
- Sums of independent random measures, 62
- Superposition of point processes, 146
  - convergence, 65
    - p.g.fl. condition for convergence, 154
- Superposition of random measures, 63
  - conditions for weak limit, 153
  - convergence, 65
- Sup metric on space of d.f.s, 145

- Support counting measure, 4
- Tail  $\sigma$ -algebra**, 208
  - tail event, 208
  - trivial, implies mixing, 209
- Takacs–Fiksel estimation procedure, 514
- Telegraph signal process, 101
- Thinning of point process, 155
  - condition for Poisson limit, 155
  - Cox process as limit, 157, 387
  - dependent-thinning, limit via convergent compensators, 387
  - random, nonstationary, with stationary output, 192
- Tiling of c.s.m.s., 15, 311
  - ‘infinite’ dissecting system, 311
- Triangular array, 146
  - conditions for Poisson limit, 150
  - independent array, 146
  - u.a.n. condition, 146
- U.a.n.**,
  - see Uniform asymptotic negligibility
- Unaligned point set, 312
- Uniform asymptotic negligibility (u.a.n.), 146
  - sufficient for infinite divisibility, 149
- Uniform integrability
  - rôle in convergence, 141
- Uniform random measure, 9
- Uniform tightness of probability measures, 136
- Updating formulae, likelihoods 407
  - estimation/detection separation, 407
  - simplified in Markovian case, 408
- Vacuity function**, 2
  - see also Avoidance function
- Variation norm, 144
  - Poisson limit property, 162
- Version or copy of process
  - same fidi distributions, 11
- Voronoi polygon, 305
  - about point at origin of  $N_0$ , 306
- Watanabe’s theorem** (Poisson characterization), 365
  - analogue of proof, 390
  - basic form, 420
  - extension to Cox process, 419
- Weak convergence, 390
  - totally finite measures, 2
  - weak-hash ( $w^\#$ ) convergence, 2
    - equivalence and non-equivalence, 3
    - example not weakly convergent, 22
  - , of random measures, 132
  - convergence of fidi distributions, 134
  - equivalent convergence modes, 135, 137
  - Laplace functional conditions, 138
  - p.g.fl. conditions, 138
- Weakly asymptotically uniformly distributed measures, 175
- Weighted averages, ergodic theorem, 201
- Wiener’s homogeneous chaos, 19
- Wold process, 92
  - complete conditional intensity, 399
  - convergence to equilibrium, 227
  - recurrence relation for p.g.fl., 65
  - stationary
    - Palm–Khinchin equations, 302
- Workload-dependent queueing process, 235
- Zero–infinity dichotomy**, a.s.
  - stationary random measure/point process on  $\mathbb{R}^d$ , 187
  - for MPP on  $\mathbb{R}^d \times \mathcal{K}$ , 205