

STOCHASTIC GEOMETRY FOR WIRELESS NETWORKS

MARTIN HAENGGI

CAMBRIDGE

CAMBRIDGE

more information www.cambridge.org/9781107014695

Stochastic Geometry for Wireless Networks

Covering point process theory, random geometric graphs, and coverage processes, this rigorous introduction to stochastic geometry will enable you to obtain powerful, general estimates and bounds of wireless network performance, and make good design choices for future wireless architectures and protocols that efficiently manage interference effects.

Practical engineering applications are integrated with mathematical theory, with an understanding of probability the only prerequisite. At the same time, stochastic geometry is connected to percolation theory and the theory of random geometric graphs, and is accompanied by a brief introduction to the R statistical computing language.

Combining theory and hands-on analytical techniques, this is a comprehensive guide to the spatial stochastic models essential for modeling and analysis of wireless network performance.

MARTIN HAENGGI is a Professor of Electrical Engineering, and a Concurrent Professor of Applied and Computational Mathematics and Statistics, at the University of Notre Dame, Indiana. He has conducted extensive research on the use of stochastic geometry for the analysis and design of modern wireless systems, publishing over 100 articles on the subject. He is the recipient of a US NSF CAREER award and the 2010 IEEE Communications Society Best Tutorial Paper Award, and is a Senior Member of the IEEE and ACM.

“Stochastic geometry is a rigorous mathematical basis for a number of applications. It has recently been applied to wireless networking concepts and design, and it is fair to say that it forms a valuable anchor of scientific support for the somewhat chaotic field of ad hoc networking. This monograph does a superior job in explaining the theory and demonstrating its use. It is the most complete, readable, and useful document to date that illuminates the intricate web of wireless networks and transforms it from a ‘dark art’ to a solid engineering discipline with a scientific foundation.”

Anthony Ephremides, University of Maryland

Stochastic Geometry for Wireless Networks

MARTIN HAENGGI

University of Notre Dame, Indiana



C A M B R I D G E U N I V E R S I T Y P R E S S
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org

Information on this title: www.cambridge.org/9781107014695

© Cambridge University Press 2013

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2013

Printed and bound in the United Kingdom by the MPG Books Group

A catalog record for this publication is available from the British Library

Library of Congress Cataloging in Publication data

Haenggi, Martin, author.

Stochastic geometry for wireless networks / Martin Haenggi, University of Notre Dame, Indiana.

pages cm

Includes bibliographical references and index.

ISBN 978-1-107-01469-5 (Hardback)

1. Wireless communication systems—Mathematics. 2. Stochastic models. I. Title.

TK5102.83.H33 2013

621.39'80151922—dc23

2012034887

ISBN 978-1-107-01469-5 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

For Roxana and my parents

Contents

Preface

Notation

Part I Point process theory

1 Introduction

- 1.1 What is stochastic geometry?
- 1.2 Point processes as spatial models for wireless networks
- 1.3 Asymptotic notation
- 1.4 Sets and measurability

Problems

2 Description of point processes

- 2.1 Description of one-dimensional point processes
- 2.2 Point process duality
- 2.3 Description of general point processes
- 2.4 Basic point processes
- 2.5 Distributional characterization
- 2.6 Properties of point processes
- 2.7 Point process transformations
- 2.8 Distances
- 2.9 Applications

Bibliographical notes

Problems

3 Point process models

- 3.1 Introduction
- 3.2 General finite point processes
- 3.3 Cox processes
- 3.4 Cluster processes
- 3.5 Hard-core processes

- 3.6 Gibbs processes
- 3.7 Shot-noise random fields
- Bibliographical notes
- Problems

4 Sums and products over point processes

- 4.1 Introduction
- 4.2 The mean of a sum
- 4.3 The probability generating functional
- 4.4 The Laplace functional
- 4.5 The moment-generating function of sums over Poisson processes
- 4.6 The probability generating and Laplace functionals for the Poisson point process
- 4.7 Summary of relationships
- 4.8 Functionals of other point processes
- Bibliographical notes
- Problems

5 Interference and outage in wireless networks

- 5.1 Interference characterization
- 5.2 Outage probability in Poisson networks
- 5.3 Spatial throughput in Poisson bipolar networks
- 5.4 Transmission capacity
- 5.5 Temporal correlation of the interference
- 5.6 Temporal correlation of outage probabilities
- Bibliographical notes
- Problems

6 Moment measures of point processes

- 6.1 Introduction
- 6.2 The first-order moment measure
- 6.3 Second moment measures
- 6.4 Second moment density
- 6.5 Second moments for stationary processes
- Bibliographical notes
- Problems

7 Marked point processes

- 7.1 Introduction and definition
- 7.2 Theory of marked point processes

7.3 Applications
Bibliographical notes
Problems

8 Conditioning and Palm theory

8.1 Introduction
8.2 The Palm distribution for stationary processes
8.3 The Palm distribution for general point processes
8.4 The reduced Palm distribution
8.5 Palm distribution for Poisson processes and Slivnyak's theorem
8.6 Second moments and Palm distributions for stationary processes
8.7 Palm distributions for Neyman–Scott cluster processes
8.8 Palm distribution for marked point processes
8.9 Applications
Bibliographical notes
Problems

Part II Percolation, connectivity, and coverage

9 Introduction

9.1 Motivation
9.2 What is percolation?

10 Bond and site percolation

10.1 Random trees and branching processes
10.2 Preliminaries for bond percolation on the lattice
10.3 General behavior of the percolation probability
10.4 Basic techniques
10.5 Critical threshold for bond percolation on the square lattice
10.6 Further results in bond percolation
10.7 Site percolation
Bibliographical notes
Problems

11 Random geometric graphs and continuum percolation

11.1 Introduction
11.2 Percolation on Gilbert's disk graph
11.3 Other percolation models
11.4 Applications
Bibliographical notes
Problems

12 Connectivity

- 12.1 Introduction
 - 12.2 Connectivity of the random lattice
 - 12.3 Connectivity of the disk graph
 - 12.4 Connectivity of basic random geometric graphs
 - 12.5 Other graphs
- Bibliographical notes
Problems

13 Coverage

- 13.1 Introduction
 - 13.2 Germ–grain and Boolean models
 - 13.3 Boolean model with fixed disks
 - 13.4 Applications
- Bibliographical notes
Problems

Appendix Introduction to R

References

Index

Preface

The performance of wireless systems depends strongly on the locations of the users or nodes. In modern networks, these locations are subject to considerable uncertainty and thus need to be modeled as a stochastic process of points in the two- or three-dimensional space.

The area of mathematics providing such models and methods to analyze their properties is *stochastic geometry*, in particular point process theory. Hence wireless network modeling and analysis is a very natural application of stochastic geometry, and, indeed, the last decade has witnessed a significant growth in this area. The goal of this book is to make the mathematical theory accessible to graduate students, researchers, and practitioners who are working in the field of wireless networks. This not only includes a coherent presentation of the theory as it applies to wireless networks, but also enables the reader to understand the related research articles and to define and solve new problems. The field is young enough to leave many opportunities for exciting and relevant new results. Indeed, not all the theoretical concepts covered in this book have found applications to wireless networks yet.

It is assumed that the reader has a solid background in basic probability and perhaps has had some exposure to point processes in one dimension, most likely in the form of traffic models for queueing theory.

While being rigorous, the book is not pedantic and does not dwell on measure-theoretic details. The interested reader can always study these intricacies from the mathematical literature; others may simply take measure theory for granted. For example, while the Radon–Nikodým theorem is mentioned on several occasions, it is not essential to follow the exposition. The many examples should illustrate the theoretical concepts and help develop a good intuition. To assist in this process, problems are included at the end of each chapter, some of which are based on the R statistical software for simulation and numerical studies. Also, many chapters have a dedicated applications section at the end, where the theory in that chapter is used to solve pertinent problems in wireless networking.

To the extent that it exists, standard terminology is used. Unfortunately, the notation is hardly consistent in the literature, which is a consequence of the fact that researchers from many different areas have made important contributions to the theory. Consequently, to help the reader understand books and articles with their

specific terminology and definitions, different names and conventions are mentioned as appropriate.

The first part of the book gives an introduction to stochastic geometry, in particular point process theory. In order to be consistent with the application, the focus is on simple point processes on the Euclidean space \mathbb{R}^d . Particular care is given to the functionals and their relationships, to higher-order statistics, and Palm distributions, covered in [Chapters 4, 6, and 8](#). These are the topics that are perhaps the most difficult to learn from the mathematical literature. [Chapter 5](#) is entirely devoted to important applications in wireless networks, in particular those modeled using Poisson point processes.

The second part of the book discusses percolation theory, connectivity, and coverage. It uses the material from [Part I](#) to model the locations of the points or nodes, but then focuses on how or whether they are connected or whether a certain region is covered by a set of nodes if each node can cover a certain small area around itself. Connectivity is closely tied to percolation, which is the question of the existence of giant components in a network. Many results for the continuous case (for which points sit in \mathbb{R}^d) are based on arguments from discrete percolation, i.e., percolation on trees or lattices, which is discussed in [Chapter 10](#). The next chapter introduces random geometric graphs and continuum percolation, while the last two chapters provide an introduction to connectivity and coverage problems.

The book is suitable both for self-learners and as a textbook for a graduate course, e.g., for a one-semester course on point process theory (Part I only), or for a course that covers both parts. In the latter case, some chapters in [Part I](#) will probably have to be discussed in less detail. For example, it is possible to leave out the sections on Gibbs processes, Janossy measures, and the Papangelou conditional intensity without losing the context. In a quarter-based system, each part could serve as a basis for a course.

Anticipating the use of the book as a textbook for a graduate course, I have refrained from using fonts that are not easily reproduced on a black- or white-board. As a consequence, x may denote a generic (deterministic) location in \mathbb{R}^d and also an element of a point process (and thus a random variable). It should always be clear what is meant from the context and the use of the term location on the one hand and point on the other.

I would like to thank my friends and collaborators Jeff Andrews, Radha Krishna Ganti, Nihar Jindal, Amites Sarkar, and Steven Weber. This book benefited greatly from our discussions. Also, I am grateful to Phil Meyler from Cambridge University Press for encouraging me to undertake this endeavor. Special thanks go to Jeff, Radha, and Amites for providing detailed comments on parts of a draft. Of course, I am fully responsible for all mistakes that may still be present.

M.H.

Notation

General notation

\emptyset	the empty set
\mathbb{N}	natural numbers $\{1, 2, \dots\}$
\mathbb{N}_0	$\mathbb{N} \cup \{0\}$
$[n]$	the set $\{1, 2, \dots, n\}$
\mathbb{R}^d	d -dimensional Euclidean space
\mathbb{R}^+	non-negative real numbers
\circ	concatenation of functions
\otimes	product of measures
o	origin of \mathbb{R}^d
$ \cdot \equiv v_d$	Lebesgue measure (of appropriate dimension)
$\ x\ $	Euclidean metric of $x \in \mathbb{R}^d$
$\lfloor x \rfloor$	largest integer smaller than or equal to $x \in \mathbb{R}$
$\#\{\cdot\}$	number of elements in set
Φ	point process on \mathbb{R}^d as random countable set or counting measure
ϕ	locally finite countable subset of \mathbb{R}^d or counting measure
\mathcal{N}	space of counting measures
N	point process as random counting measure
\mathfrak{N}	σ -algebra of counting measures
\mathbb{M}	mark space
$\hat{\Phi}$	marked point process; point process on $\mathbb{R}^d \times \mathbb{M}$
$\hat{\mathcal{N}}$	space of counting measures on $\mathbb{R}^d \times \mathbb{M}$
$\hat{\mathfrak{N}}$	σ -algebra of $\mathbb{R}^d \times \mathbb{M}$
\mathcal{B}^d	Borel σ -algebra in \mathbb{R}^d
\mathcal{B}	$= \mathcal{B}^1$
\mathbb{P}, P, P	probability measures
P_o	Palm distribution
$P_o^!$	reduced Palm distribution
Ψ	non-negative random measure (possibly a point process)
\mathcal{M}	space of non-negative random measures

\mathfrak{M}	σ -algebra of random measures
$1(\cdot)$	indicator function
$1_A(x)$	$\equiv 1(x \in A)$, indicator function of condition $x \in A$
$\delta(x)$	Dirac delta function
$\delta_x(A)$	$\equiv 1_A(x) \equiv \int_A \delta(y - x)dy$, Dirac measure at point x
$b(x, r)$	d -dimensional (closed) ball of radius r centered at x
c_d	$\triangleq b(o, 1) $, volume of d -dimensional unit ball
Λ	intensity measure (or first-order moment measure)
Λ	intensity measure of marked point process
$\lambda(x)$	intensity function
$\lambda(x, \Phi)$	Papangelou conditional intensity
\mathcal{V}	$[0, 1]$ -valued functions v with $1 - v$ of bounded support
\mathcal{U}	non-negative-valued functions of bounded support
$\mathcal{L}_X(s)$	Laplace transform $\mathbb{E}(e^{-sX})$ of random variable X
$\mu^{(2)}$	second-order moment measure
$\alpha^{(2)}$	second-order factorial moment measure
$\varrho^{(2)}$	second moment density (or second-order product density)
g	pair correlation function
$\ell, \bar{\ell}$	path loss function, radial path loss function
κ	reduced second moment measure
K	Ripley's K function
$S[f]$	sum of $f(x)$, where $x \in \Phi$
$G[v]$	probability generating functional for $v \in \mathcal{V}$
$L_\Psi[u]$	Laplace functional for $u \in \mathcal{U}$ of random measure Ψ
$V(B)$	vacancy indicator of $B \in \mathcal{B}^d$
$V(x)$	Voronoi cell of point x
$\sum_{x,y \in A}^\neq$	double sum $\sum_{x \in A} \sum_{y \in A; x \neq y}$
A_x	translation of set A by x : $A_x \triangleq \{y \in A: y + x\}$
$A \oplus B$	Minkowski addition $\{x \in A, y \in B: x + y\}$
$A \star B$	Cartesian product of sets including only distinct points
$\ x - A\ $	minimum distance $\min_{y \in A} \{\ x - y\ \}$
p_c	critical probability in percolation models
\triangleq	definition
$\stackrel{d}{=}$	equality in distribution

Abbreviations

PPP	Poisson point process
BPP	binomial point process
a.s.	almost surely (with probability 1)
a.a.s.	asymptotically almost surely
iid	independent and identically distributed
fidi	finite-dimensional
pdf	probability density function
cdf	cumulative distribution function
pgfl	probability generating functional
FKG (inequality)	inequality named after Fortuin, Kasteleyn, and Ginibre
BK (inequality)	inequality named after van den Berg and Kesten
LR (crossing)	left-right (crossing) in percolation models
TB (crossing)	top-bottom (crossing) in percolation models
RGG	random geometric graph
MAC	medium access control (for channel access)

Part I

Point process theory

1 Introduction

1.1 What is stochastic geometry?

Stochastic geometry (sometimes used synonymously with the older term geometric probability) deals with random spatial patterns. Random point patterns or *point processes* are the most basic and important such objects, hence *point process theory* is often considered to be the main sub-field of stochastic geometry.

Stochastic geometry provides answers to questions such as the following.

- How can one describe a (random) collection of points in one, two, or higher dimensions?
- How can one derive statistical properties of such a collection of points?
- How can one calculate statistical averages over all the possible realizations of such a random collection?
- How can one condition on having a point at a fixed location?
- Given an empirical set of points, which statistical model is likely to produce this point set?
- How can one describe more general random geometric objects such as a “random line” or a “random triangle”?

Throughout this book, we will use *point processes* to model the distributions of nodes (users, wireless terminals) in a wireless network where node locations are subject to uncertainty. In [Part II](#), we will also encounter random geometric graphs to address the connectivity of wireless networks and random regions in the context of coverage problems.

1.2 Point processes as spatial models for wireless networks

Loosely speaking, a point process is a random collection of points that reside in some space. In this book, we will focus on the one-, two-, and three-dimensional Euclidean spaces \mathbb{R} , \mathbb{R}^2 , and \mathbb{R}^3 , since, in our applications, the points represent the

locations of wireless nodes in the real world.

Point process models permit statements about entire classes of wireless networks, instead of just about one specific configuration of the network. In some cases, distributions over the point process can be calculated (for example for the interference), in others, *spatial averaging* is performed, which yields expected values of certain performance metrics, such as the likelihood of transmission success.

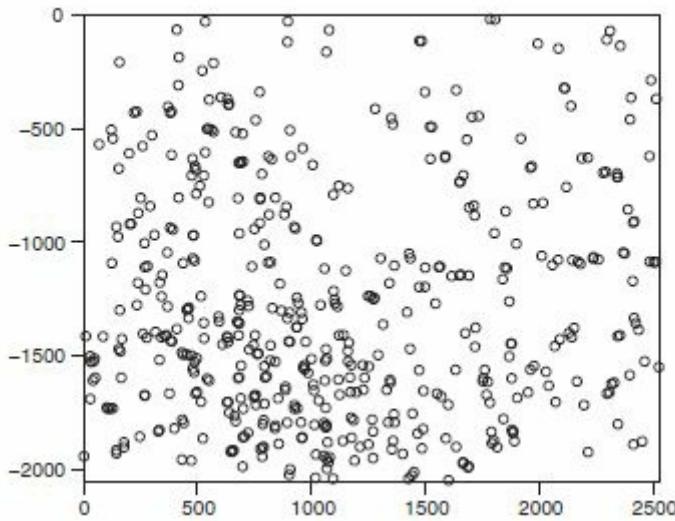


Figure 1.1 Locations of 493 cellular base stations in an area of about 5 km² in central London.

Figure 1.1 shows the locations of all cellular base stations in an area of 2.5 km × 2 km in a central part of London. It illustrates that even the base stations (let alone the mobile users) themselves do not form a nice deterministic pattern or lattice. A random approach to modeling these locations is thus sensible.

In some applications, nodes are naturally clustered (attracted to each other), in others, they are separated (repelled from each other), and there exist models for these cases. In many situations, such as when we model interference, the relevant point process is the one of the *transmitters* only, not the point process of all nodes. It is determined by the spatial configuration of nodes together with the channel access (MAC) scheme, which performs a *thinning* procedure on the set of all nodes to produce the set of transmitters.

Emerging classes of large wireless systems such as ad hoc and sensor networks and cellular networks with coverage extensions such as relays or micro-base stations have been the subject of intense investigation over the last decade. Classical methods of communication theory are insufficient to analyze these new types of networks for the following reasons. (i) The performance-limiting metric is the signal-to-interference-plus-noise ratio (SINR) rather than the signal-to-noise ratio (SNR). (ii) The interference depends on the path loss and fading

characteristics, which, in turn, are functions of the *network geometry*. (iii) The amount of uncertainty present in large wireless networks far exceeds the amount present in point-to-point systems: it is impossible for each node to know or predict the locations and channels of all but perhaps a few other nodes.

Two main tools have recently proved most helpful in circumventing the above difficulties: stochastic geometry and random geometric graphs. Stochastic geometry allows one to study the average behavior over many spatial realizations of a network whose nodes are placed according to some probability distribution. Random geometric graphs capture the distance dependence and randomness in the connectivity of the nodes. This book provides an introduction to these mathematical tools and discusses some important applications to problems in wireless networking.

1.3 Asymptotic notation

We will make use of the standard asymptotic notation, summarized in Box 1.1. While the notation $f(x) = O(g(x))$ is common, it is not strictly correct since it implies a symmetry that does not exist. For example, $O(x^7) = O(e^x)$ as $x \rightarrow \infty$, but this does not imply that $O(e^x) = O(x^7)$, since not every function bounded by the exponential is bounded by the monomial (but the converse is true). A set-theoretic notation is more rigorous, where $O(g(x))$ denotes the class of functions which remain bounded when they are divided by $g(x)$. The above statement then reads $O(x^7) \subset O(e^x)$, which is obviously not invertible.

Since “remains bounded” is weaker than “goes to zero” $o(\cdot)$ implies and is stronger than $O(\cdot)$, i.e., $f(x) = o(g(x)) \Rightarrow f(x) = O(g(x))$, or $o(g(x)) \subset O(g(x))$.

Box 1.1 Asymptotic notation

Let x tend to a . We write

$f(x) = O(g(x))$	if the ratio $f(x)/g(x)$ remains bounded
$f(x) = o(g(x))$	if the ratio $f(x)/g(x)$ goes to 0
$f(x) = \omega(g(x))$	if $g(x) = o(f(x))$
$f(x) = \Theta(g(x))$	if $f(x) = O(g(x))$ and $g(x) = O(f(x))$
$f(x) \sim g(x)$	if the ratio $f(x)/g(x)$ approaches 1.

More formally:

$$f(x) = O(g(x)) \text{ as } x \rightarrow a \Leftrightarrow \limsup_{x \rightarrow a} \left| \frac{f(x)}{g(x)} \right| < \infty$$

It is important to always indicate the limit point.

Example 1.1 Examples of asymptotic notation:

- $\ln x = O(x)$, $x^4 = O(e^x)$, $\sin x = O(1)$ as $x \rightarrow \infty$;
- $x^2 = O(x)$, $\sin x = O(x)$ as $x \rightarrow 0$;
- Taylor expansion (at $x = 0$): $e^x = 1 + x + O(x^2)$, $e^x - 1 = x + o(x)$, $e^x - 1 - x = o(x^{3/2})$, $e^x = 1 + x + \Theta(x^2)$.

1.4 Sets and measurability

1.4.1 Borel sets and σ -algebra

Here we give a brief overview of the set-theoretic concepts used in the book and define measurability. Starting with the half-open intervals of \mathbb{R} , given by

$$[u, v) = \{x \in \mathbb{R} : u \leq x < v\},$$

we obtain the Borel sets by taking the complement and countable unions and intersections of such intervals. The collection of all Borel sets forms the *Borel σ -algebra* or σ -field, denoted by \mathcal{B} . It contains all the singletons $\{x\}$ since

$$\{x\} = \bigcap_{n=1}^{\infty} \left[x, x + \frac{1}{n} \right)$$

and thus also all closed and open intervals since $(u, v) = [u, v) \setminus \{u\}$ and $[u, v] = [u, v) \cup \{v\}$. In fact, \mathcal{B} can also be defined starting with closed or open intervals instead of the half-open ones. The restriction to *countable* unions and intersections is important, otherwise *all* subsets of \mathbb{R} could be constructed simply by taking the union of the corresponding singletons.

The concept of the Borel σ -algebra is easily extended to the d -dimensional space \mathbb{R}^d , denoted by \mathcal{B}^d , by starting with (hyper)rectangles whose coordinate sets belong to \mathcal{B} . It further generalizes to all metric spaces (sets equipped with a distance or metric defined between their elements).

If a set consists of a countable number of singletons, it is a countable (finite or countably infinite) set. Conversely, all open sets are infinite sets, as are all half-open intervals of \mathbb{R} .

Letting $b(o, r) \triangleq \{x \in \mathbb{R}^d : \|x\| \leq r\}$ be the (closed) ball of radius r centered at the origin o , a set $A \subseteq \mathbb{R}^d$ is *bounded* if there is a ball $b(a, r)$ such that $A \subseteq b(a, r)$. It is *compact* if it is also closed. Since \mathcal{B}^d includes all closed sets it is a superset of

the compact sets. Most sets encountered in this book will be Borel or compact.

The Cartesian product of sets is denoted by \times .

1.4.2 Measurability

While we will not concern ourselves with the intricacies of measurability, we need to clarify what “measurable” means. Generally, an ordered pair (A, \mathcal{A}) forms a *measurable space* if \mathcal{A} is a σ -algebra of subsets of A . An important example is $(\mathbb{R}^d, \mathcal{B}^d)$ as defined before.

A function $f : \mathbb{R}^d \mapsto \mathbb{R}$ is \mathcal{B}^d -*measurable* if and only if the pre-image of $B \in \mathcal{B}$ is an element of \mathcal{B}^d , i.e.,

$$f^{-1}(B) \triangleq \{x \in \mathbb{R}^d : f(x) \in B\} \in \mathcal{B}^d.$$

More generally, if (N, \mathfrak{N}) is another measurable space, $f : N \mapsto \mathbb{R}$ is \mathfrak{N} -measurable if and only if the pre-image of $B \in \mathcal{B}$ is an element of \mathfrak{N} , i.e.,

$$f^{-1}(B) \triangleq \{\varphi \in \mathcal{N} : f(\varphi) \in B\} \in \mathfrak{N}.$$

This is the notion of measurability that we will employ in the context of point processes, where \mathcal{N} denotes the space of point patterns.

A *measure* ν is a function $\mathcal{A} \mapsto \mathbb{R} \cup \{\infty\}$, where \mathcal{A} is a σ -algebra, with the property of countable additivity, i.e.,

$$\nu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \nu(A_i)$$

for all pairwise disjoint $A_i \in \mathcal{A}$ and $n \in \mathbb{N} \cup \{\infty\}$ and satisfying $\nu(\emptyset) = 0$.

A fundamental measure is the *Lebesgue measure*, denoted by ν_d in the d -dimensional case or $|\cdot|$ if the number of dimensions is clear from the context. For intervals, it is defined as their length, i.e., $\nu_1((u, v]) = v - u$, and in higher dimensions, it is the area or (hyper)volume. For example, in two dimensions, the area of a disk of radius 2 is $|b(o, 2)| = 4\pi$. The Lebesgue measure is a *diffuse measure* since it gives zero mass to every singleton and thus any countable set. Other measures we will use frequently are probability measures, Dirac measures, and counting measures. The product of measures is denoted by \otimes .

Problems

1.1 Which of the following are true?

As $x \downarrow 0$:

- (a) $\cosh x = o(1)$
- (b) $\sin x = o(1)$
- (c) $x^{-1} = O(\log x)$.

As $x \rightarrow \infty$:

- (d) $O(2^x) = O(2^{ax})$
- (e) $e^{\Theta(1)} = \Theta(e^x)$
- (f) $O(\log x) = O(\log(x^a))$
- (g) $f_1(x) = O(g_1(x)), f_2(x) = O(g_2(x)) \Rightarrow f_1(x)f_2(x) = O(g_1(x)g_2(x)).$

1.2 Let X_1, X_2, \dots, X_k be independent and identically distributed (iid) random variables with cumulative distribution function (cdf) $F(x)$. What is the cdf of the minimum $\min_i\{X_i\}$?

1.3 Let X be a Poisson random variable with parameter Y , where Y is itself a Poisson random variable with parameter μ . Show that the generating function of $X + Y$ is

$$G_{X+Y}(x) \triangleq \mathbb{E}(x^{X+Y}) = \exp\{\mu(xe^{x-1} - 1)\}.$$

1.4 Let X_1, \dots, X_N be N iid random variables with cdf $F(x)$, where N is Poisson with mean μ . Calculate the cdf $G(x)$ of the *maximum* of the N random variables.

1.5 Let Z be the distance between two points picked independently (uniformly) at random in a disk of radius a . Show that $\mathbb{E}(Z^2) = a^2$.

1.6 Let a , b , and c be iid exponential. Show that the probability that the polynomial $ax^2 + bx + c$ has real roots is $1/3$.

1.7 A goose lays N eggs, where N is Poisson with mean λ . Each egg hatches with probability p , independently of the other eggs. Let K be the number of baby geese. Find $\mathbb{E}(K | N)$, $\mathbb{E}K$, and $\mathbb{E}(N | K)$.

2 Description of point processes

2.1 Description of one-dimensional point processes

There are several ways to describe a collection of points (x_1, x_2, \dots) in one dimension. Here are four of them, assuming the points lie on \mathbb{R}^+ .

1. Direct characterization of the points (or *arrival times* if the axis is a time axis) x_i .
2. Using the increasing step function $A(t) = \sum_{i=1}^{\infty} \mathbf{1}(x_i \in [0, t))$.
3. Using the *interarrival intervals* $S_i = x_{i+1} - x_i$, $i \in \mathbb{N}$. Here it is assumed that the points are ordered, i.e., $x_1 \leq x_2 \leq \dots$. In the case of *renewal processes*, the increments S_i are independent.
4. *Counting* the number of points falling in a set $B \subset \mathbb{R}$:

$$N(B) = \sum_{i=1}^{\infty} \mathbf{1}(x_i \in B).$$

Method 1 may be convenient if the total number of points n is finite and fixed and if the points are distributed independently and identically, i.e., if the random variables x_1, x_2, \dots, x_n are iid. Method 2 is a special case of method 4, with B restricted to an interval $[0, t)$. From $A(t)$, $N(B)$ can be calculated (in much the same way as the complete distribution of a random variable can be obtained from just the distribution function). Such cumulative functions are common in networking to count the number of packets that have arrived up to time t . Since method 3 relies on an ordering of the points, it is restricted to the one-dimensional case. Processes with independent interarrival intervals are the subject of study in the field of renewal theory. The most important case is where the S_i are exponentially distributed. This interarrival interval distribution results in a Poisson process.

An important dichotomy in point process theory is whether a point process is permitted to have multiple points at the same location. If not, it can be expressed as a *random set* $\{x_1, x_2, \dots\}$; if yes, the point process is not a random set, and a

random measure formalism as in method 4 is usually employed. We discuss these dual views of point processes in detail in the next section, before exploring point processes in general dimensions.

2.2 Point process duality

We first give a formal definition of a point process.

DEFINITION 2.1 (Point process) A point process is a countable random collection of points that reside in some measure space, usually the Euclidean space \mathbb{R}^d . The associated σ -algebra consists of the Borel sets \mathcal{B}^d , and the measure is the Lebesgue measure.

Such a point process can be described using two formalisms, a *random set formalism* (Box 2.1) and a *random measure formalism* (Box 2.2).

Box 2.1 Random set formalism

In this formalism, the point process is regarded as a *countable random set* $\Phi = \{x_1, x_2, \dots\} \subset \mathbb{R}^d$ consisting of random variables $x_i \in \mathbb{R}^d$ as its elements.

Box 2.2 Random measure formalism

In this formalism, a point process is characterized by counting the number of points falling in sets $B \subset \mathbb{R}^d$. The number of points in B is denoted by $N(B)$. Hence $N(B)$ is a random variable that assumes values from the non-negative integers \mathbb{N}_0 . N is called a (random) *counting measure*.

Similarly to the particle–wave duality in physics, we may speak of a duality in point process theory: On the one hand, point processes can be characterized as random sets, on the other hand, they can be regarded as random counting measures.

If Φ is given, N is obtained by

$$N(B) = \#\{\Phi \cap B\}.$$

Conversely, Φ is retrieved from N by

$$\Phi = \{x \in \mathbb{R}^d : N(\{x\}) = 1\}.$$

Hence there is a one-to-one mapping between Φ and N .

By considering the point process as a set $\Phi = \{x_1, x_2, \dots\}$, we have implicitly assumed that only one point can exist at a given location, since a set can only contain one instance of each element. Such a point process is called *simple*. The random measure formalism is more general; it permits the characterization of point processes that may have multiple co-located points.¹ It is used to formally define simple point processes.

DEFINITION 2.2 (Simple point process) A simple point process is a point process with counting measure

$$N(\{x\}) \in \{0, 1\} \text{ a.s. } \forall x \in \mathbb{R}^d.$$

So singletons have measure 0 or 1 almost surely. Sometimes a simple point process is also called an orderly point process.

If we use the random set formalism to describe a point process, we automatically restrict ourselves to simple point processes. Since Φ is a countable set, simple point processes belong to the more general class of *random closed sets*.

Since wireless nodes are usually not co-located, we focus mostly on simple point processes. There are, however, situations in which it makes sense to allow co-located points, for example when approximating two nearby transmitters, say in a node equipped with multiple antennas, as co-located, since the resulting non-simple point process may be more tractable than the original one.

For simple point processes, we can always adopt either the random measure or the random set description, whichever is more convenient. It also makes sense to assume that only a finite number of points can be placed in a finite area. Such point processes are called *locally finite*. Formally, a point process is locally finite if and only if

$$|B| < \infty \implies N(B) < \infty \text{ a.s.}$$

Such point processes are also called *boundedly finite*.

In the rest of the chapter, we will discuss how to probabilistically describe the random set Φ and the counting measure N and introduce some basic properties, models, and transformations.

2.3 Description of general point processes

Method 1 from Section 2.1 may be generalized to higher dimensions; it is usually

restricted to binomial point processes, which will be formally introduced in [Section 2.4.4](#). It is possible in principle to apply method 2 by replacing the interval $[0, t)$ by the ball $b(o, r)$. While the resulting function would certainly not contain the complete information about the point process, it may still reveal some useful properties, especially if the point process has some symmetry properties. We will discuss such a radially increasing function in [Section 6.5](#). More generally, (hyper)rectangles $[0, t_1] \times [0, t_2] \times \dots \times [0, t_d]$ could be used. As mentioned before, method 3 is not applicable to higher dimensions due to the lack of a natural ordering, but method 4 generalizes in a straightforward manner since B can be a subset of any space.

This indicates that the random measure formalism is the most natural and powerful one in higher dimensions.

2.3.1 Counting and intensity measures

Measurability

We cannot expect that $N(B)$ be defined for all subsets of \mathbb{R}^d . A sufficiently rich set of subsets is the Borel σ -algebra \mathcal{B}^d . So a natural condition for measurability of $N(B)$ is to require $B \in \mathcal{B}^d$.

Measurable decomposition

The counting measure can be decomposed into elementary measures as follows.

DEFINITION 2.3 (Measurable decomposition) The *measurable decomposition* of a simple point process Φ is the decomposition of the counting measure into Dirac measures:

$$N = \sum_{x \in \Phi} \delta_x. \tag{2.1}$$

$\delta_x(B) = 1_B(x)$ is the Dirac measure at x for $B \in \mathcal{B}^d$. It is 1 if $x \in B$ and 0 otherwise. The Dirac measure and thus N are *atomic measures*, i.e., they are concentrated on a countable collection of locations.² A counting measure for a simple point process is a special atomic measure that gives each location a mass of zero or one. Like the Lebesgue measure, it is a *Radon measure* since it is finite on finite sets. In fact, every translation-invariant Radon measure on $(\mathbb{R}^d, \mathcal{B}^d)$ is a constant multiple of the Lebesgue measure.

The measurable decomposition (2.1) provides another way to relate Φ and N . We may call N the counting measure *induced by the random set* Φ . To explicitly indicate this relationship, we may write N_Φ . It is, however, often more convenient to simply use Φ itself as the counting measure, i.e., to write $\Phi(B)$ for the number of

points falling in B . In this notation, the same symbol, Φ , is used as the counting measure *and* the random set (see [Box 2.3](#)). We will frequently make use of it.

Atoms

The set B in $N(B)$ or $\Phi(B)$ may be a singleton. For many types of point processes, the probability of having a point at a particular location is 0, but for some, this probability is positive for some locations. Such locations are called atoms.

DEFINITION 2.4 (Atom) An atom of a point process is a location $x_0 \in \mathbb{R}^d$ for which $\mathbb{P}(N(\{x_0\}) > 0) > 0$.

This definition shows that whether a location is an atom or not does not depend on the particular realization of a point process, only on its statistics, i.e., the set of atoms of a point process is a deterministic set.

Box 2.3 Duality of notation

The symbol Φ refers to the random set $\{x_1, x_2, \dots\}$ *and* to the associated counting measure N . The symbol N , on the other hand, always denotes a counting measure. A concrete realization of a point process is denoted by ϕ , which, by duality, can be viewed either as a set of points or as a counting measure. Hence we may write, for example, $\phi = \{x, y\}$ to indicate that there are two points at locations x and y in the realization, or we may write $\phi = \delta_x + \delta_y$ for the same set of points. The definition of $\phi(B)$ as the number of points in B is based on the latter interpretation as the counting measure.

A disadvantage of the measure formalism is that there is no compact standard and rigorous notation for the counting measure pertaining to $\Phi \setminus \{x\}$. Sometimes this measure is denoted as $N - \delta_x$, but strictly speaking this is valid only if x is a point of the process, i.e., if $N(\{x\}) > 0$. Alternatively, $N \setminus x$ or $N \setminus \{x\}$ may be used to indicate the counting measure of $\Phi \setminus \{x\}$.

Intensity measure

The deterministic counterpart of the random counting measure is the *intensity measure*, which is the expected number of points in a set B .

DEFINITION 2.5 (Intensity measure) The intensity measure Λ is defined as

$$\Lambda(B) \triangleq \mathbb{E}N(B), \quad \forall B \in \mathcal{B}^d.$$

At an atom x_0 , $\Lambda(\{x_0\}) > 0$, whereas for all other locations x , $\Lambda(\{x\}) = 0$. Other basic properties of point processes that are based on the counting or intensity measures are defined as follows.

DEFINITION 2.6 (Diffuse, atomic, and discrete point process) We call a point process *diffuse* if its intensity measure is a diffuse measure. If it has one or more atoms, it is called *atomic*. If there exists a *countable* set X of locations such that $N(\mathbb{R}^d \setminus X) = 0$ the process is called *discrete* or *purely atomic*.

Hence in a discrete process, the entire mass of the intensity measure is concentrated on atoms. Since any locally finite measure is uniquely decomposable into a purely atomic and a diffuse measure, any point process is the superposition (or union) of a diffuse and a discrete point process.

In one dimension, the intensity measure of a diffuse point process is continuous, whereas it is discontinuous if the point process has one atom or more. In higher dimensions, a diffuse process may still have a discontinuous intensity measure. This happens if mass is concentrated on a lower-dimensional subset. An example of such a point process will be given later in this chapter, in Example 2.10 and Fig. 2.7. Conversely, the intensity measure of a point process on \mathbb{R}^d is continuous if and only if $N(B) = 0$ for all $B \in \mathcal{B}^d$ with $|B| = 0$ (where $|\cdot|$ is the Lebesgue measure in d dimensions).

Example 2.1 defines a discrete point process that is not relevant for applications but will help illustrate some important theoretical concepts.

Example 2.1 Here we introduce the *die point process*, which models the locations of the pips (dots) of a die. Take the generic probability space $([0, 1], \mathcal{B}, |\cdot|)$ and denote the outcome by ω . Let X be the numerical random variable defined as $X(\omega) = 1 + \lfloor 6\omega \rfloor \in [6]$, where $\lfloor y \rfloor$ denotes the largest integer smaller than or equal to y and $[6] = \{1, 2, \dots, 6\}$. X may be interpreted as the outcome of the roll of a fair die. Define the two-dimensional *discrete* point process

$$N^\omega \triangleq \begin{cases} \delta_o & \text{if } X = 1 \\ \delta_{(-1,1)} + \delta_{(1,-1)} & \text{if } X = 2 \\ \delta_{(-1,1)} + \delta_o + \delta_{(1,-1)} & \text{if } X = 3 \\ \delta_{(-1,1)} + \delta_{(1,1)} + \delta_{(1,-1)} + \delta_{(-1,-1)} & \text{if } X = 4 \\ \delta_{(-1,1)} + \delta_{(1,1)} + \delta_{(1,-1)} + \delta_{(-1,-1)} + \delta_o & \text{if } X = 5 \\ \delta_{(-1,1)} + \delta_{(1,1)} + \delta_{(1,-1)} + \delta_{(-1,-1)} + \delta_{(-1,0)} + \delta_{(1,0)} & \text{if } X = 6. \end{cases}$$

This point process visualizes the face of the die when a certain number X is rolled.

The notation N^ω is used since $N(\omega)$ may lead to confusion with $N(B)$, the number of points in B . The corresponding random set representation of this point process is

$$\Phi = \begin{cases} \{o\} & \text{if } X = 1 \\ \{(-1, 1), (1, -1)\} & \text{if } X = 2 \\ \{(-1, 1), o, (1, -1)\} & \text{if } X = 3 \\ \{(-1, 1), (1, 1), (1, -1), (-1, -1)\} & \text{if } X = 4 \\ \{(-1, 1), (1, 1), (1, -1), (-1, -1), o\} & \text{if } X = 5 \\ \{(-1, 1), (1, 1), (1, -1), (-1, -1), (-1, 0), (1, 0)\} & \text{if } X = 6. \end{cases}$$

Hence $\Phi \subset A$ for $A \triangleq \{-1, 0, 1\}^2$. We have $\mathbb{E}N([-2, 2]^2) = 7/2$, $\mathbb{E}N([-1/2, 2]^2) = 7/6$, and $\mathbb{E}N(\{o\}) = 1/2$. Also, $\Phi(A) = X$.

2.3.2 Vacancies

A simpler characterization than the counting measure is provided by the *vacancy indicators* $V(B)$, which are defined as

$$V(B) \triangleq 1(N(B) = 0).$$

For simple point processes, the vacancies V fully describe the random set Φ , since Φ is the complement of the union of all vacant regions:

$$\Phi = \mathbb{R}^d \setminus \bigcup \{B \subset \mathbb{R}^d : V(B) = 1\}.$$

Counting measures are additive, while vacancies are multiplicative, in the following sense.

- If $N_1(B)$ and $N_2(B)$ are the numbers of nodes in B for two point processes Φ_1 and Φ_2 , then $N(B) = N_1(B) + N_2(B)$ is the number of nodes in B of the superimposed process.
- If $V_1(B)$ and $V_2(B)$ are the vacancy indicators, then $V(B) = V_1(B)V_2(B)$.

Also, generally, for a point process N with vacancies V ,

$$N(B_1) + N(B_2) = N(B_1 \cup B_2) + N(B_1 \cap B_2)$$

and

$$V(B_1)V(B_2) = V(B_1 \cup B_2) = \min\{V(B_1), V(B_2)\}.$$

In principle, a point process could be defined by a random function V that satisfies these properties almost surely.

2.4 Basic point processes

In this section, we introduce several basic yet important point process. We start with the ubiquitous Poisson point process.

2.4.1 One-dimensional Poisson processes

DEFINITION 2.7 (One-dimensional Poisson process) The one-dimensional Poisson point process (PPP) with *uniform intensity* λ is a point process in \mathbb{R} such that

- for every bounded interval $[a, b]$, $N([a, b])$ has a *Poisson* distribution with mean $\lambda(b - a)$:

$$\mathbb{P}(N([a, b]) = k) = e^{-\lambda(b-a)} \frac{(\lambda(b-a))^k}{k!};$$

- if $[a_1, b_1], [a_2, b_2], \dots, [a_m, b_m]$ are disjoint bounded intervals, then $N([a_1, b_1]), N([a_2, b_2]), \dots, N([a_m, b_m])$ are independent random variables.

This process is called *uniform* or *homogeneous*, since its intensity does not depend on the location. Instead of the half-open intervals, the closed or open intervals may also be used in the definition.

In fact, if the first property holds for all Borel sets, then the second property follows from it, as shown in Rényi (1967).

Since the real line has a natural ordering, the points are often numbered using not the natural numbers but the integers, such that

$$\dots, x_{-2} < x_{-1} < x_0 < x_1 < x_2 < \dots \text{ a.s.}$$

Other useful properties of the one-dimensional homogeneous PPP are the following.

1. The interarrival intervals S_i are iid exponential with parameter λ .
2. The generalized interarrival intervals $x_{k+i} - x_i$ for $k > 0$ are Erlang (gamma) distributed with parameters k and λ for all i .
3. The probability of two or more arrivals in a given interval is asymptotically of smaller order than the length of the interval:

$$\mathbb{P}(N([a, a+h)) \geq 2) = o(h), h \rightarrow 0.$$

4. With x_1 being the closest positive point to o , the joint pdf of the first n points is

$$f_{(x_1, \dots, x_n)}(t_1, \dots, t_n) = \lambda^n e^{-\lambda t_n} \mathbf{1}((t_1, \dots, t_n) \in \mathbb{R}^{+n}), \quad (2.2)$$

where

$$\mathbb{R}^{+n} = \{(t_1, \dots, t_n) \in \mathbb{R}^n : 0 < t_1 < t_2 < \dots < t_n\}$$

is the order cone (a hyperoctant or orthant) in n dimensions.

The most direct way to establish the last property is to use the fact that the interarrival intervals are exponential, i.e.,

$$f_{x_i - x_{i-1}}(t_i - t_{i-1}) = \lambda e^{-\lambda(t_i - t_{i-1})}, \quad i = 2, 3, \dots, n,$$

and independent, from which

$$\begin{aligned} f_{(x_1, \dots, x_n)}(t_1, \dots, t_n) &= f_{x_1, x_2 - x_1, \dots, x_n - x_{n-1}}(t_1, t_2 - t_1, \dots, t_n - t_{n-1}) \\ &= (\lambda e^{-\lambda t_1})(\lambda e^{-\lambda(t_2 - t_1)}) \dots (\lambda e^{-\lambda(t_n - t_{n-1})}) \end{aligned}$$

follows, which equals (2.2).

We may also define the *inhomogeneous* or *non-uniform* PPP, where the intensity is a function of the location, denoted as $\lambda(t) \geq 0$, in a similar way. For this process, the average number of points in any bounded interval $[a, b]$ is

$$\mathbb{E}N([a, b]) = \int_a^b \lambda(t) dt,$$

and the numbers of arrivals in disjoint intervals are (again) independent random variables.

2.4.2 Uniform Poisson processes in general dimensions

DEFINITION 2.8 (Homogeneous or uniform Poisson point process) The homogeneous or uniform PPP, with intensity λ , is a point process in \mathbb{R}^d such that

- for every compact set B , $N(B)$ has a Poisson distribution with mean $\lambda|B|$;
- if B_1, B_2, \dots, B_m are disjoint bounded sets, then $N(B_1), N(B_2), \dots, N(B_m)$ are independent random variables.

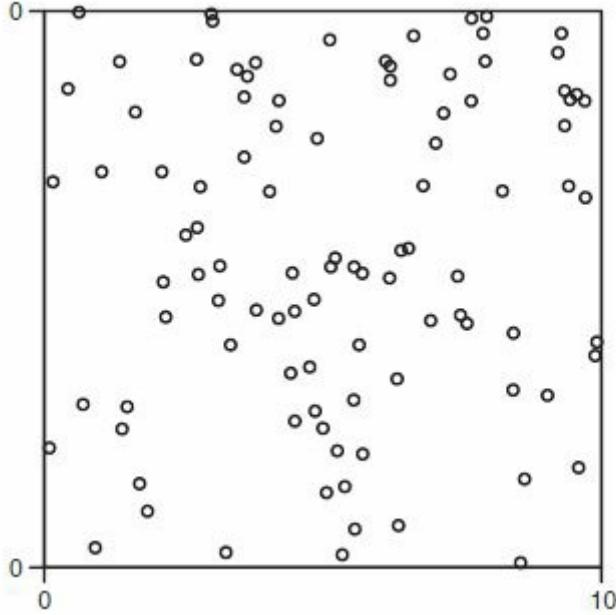


Figure 2.1 A (realization of a) uniform PPP of intensity $\lambda = 1$ on $[0, 10]^2$.

The intensity λ is the expected number of points of the process per unit area or volume.

A two-dimensional example of a uniform PPP is shown in Fig. 2.1. More precisely, what is displayed is a *realization* of the PPP, which is a deterministic set of points or a *point pattern*.

The following theorem provides an important connection between the Poisson process and the uniform distribution.

THEOREM 2.9 (Conditional property for Poisson point processes) *Consider a homogeneous PPP in \mathbb{R}^d with intensity $\lambda > 0$. Let $W \subset \mathbb{R}^d$ be any subset with $0 < |W| < \infty$. Given that $N(W) = n$, the conditional distribution of $N(B)$ for $B \subseteq W$ is binomial:*

$$\mathbb{P}(N(B) = k \mid N(W) = n) = \binom{n}{k} p^k (1-p)^{n-k},$$

where $p = |B|/|W|$.

Proof Let $0 \leq k \leq n$. We have

$$\begin{aligned}
\mathbb{P}(N(B) = k \mid N(W) = n) &= \frac{\mathbb{P}(N(B) = k, N(W) = n)}{\mathbb{P}(N(W) = n)} \\
&= \frac{\mathbb{P}(N(B) = k, N(W \setminus B) = n - k)}{\mathbb{P}(N(W) = n)} \\
&\stackrel{(a)}{=} \frac{\mathbb{P}(N(B) = k)\mathbb{P}(N(W \setminus B) = n - k)}{\mathbb{P}(N(W) = n)},
\end{aligned}$$

where (a) follows from the independence of the counts in disjoint regions. Now we can insert the probabilities from the Poisson distribution, which yields

$$\begin{aligned}
\mathbb{P}(N(B) = k \mid N(W) = n) &= \frac{e^{-\lambda|B|} \frac{(\lambda|B|)^k}{k!} e^{-\lambda|W \setminus B|} \frac{(\lambda|W \setminus B|)^{n-k}}{(n-k)!}}{e^{-\lambda|W|} \frac{(\lambda|W|)^n}{n!}} \\
&= \frac{n!}{k!(n-k)!} \left(\frac{|B|}{|W|} \right)^k \left(\frac{|W \setminus B|}{|W|} \right)^{n-k} \\
&= \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}$$

This theorem provides us with a method to simulate PPPs on a region A . First, draw a Poisson distributed number with mean $\lambda|A|$, call it n , and then place n points uniformly at random in A . If A is a rectangle, this is straightforward. Otherwise a rectangle that is larger than A needs to be chosen, and the points falling outside A are to be ignored. The only distinction between a binomial process and a PPP in W is that different realizations of the PPP consist of a different number of points.

2.4.3 General Poisson point processes

A general PPP on \mathbb{R}^d is defined as follows.

DEFINITION 2.10 (General Poisson point process) The PPP on \mathbb{R}^d with intensity measure Λ is a point process such that

- for every compact set $B \subset \mathbb{R}^d$, $N(B)$ has a Poisson distribution with mean $\Lambda(B)$. If Λ admits a density λ , we may write

$$\mathbb{P}(N(B) = k) = \exp \left(- \int_B \lambda(x) dx \right) \cdot \frac{\left(\int_B \lambda(x) dx \right)^k}{k!};$$

- if B_1, B_2, \dots, B_m are disjoint compact sets, then $N(B_1), N(B_2), \dots, N(B_m)$ are independent.

The uniform or homogeneous PPP is a special case where $\Lambda(B) = \lambda|B|$. An example of a non-uniform PPP on \mathbb{R}^2 with density $\lambda(x, y) = 2\exp(-y/6)(1 + \sin x)$ is shown in Fig. 2.2.

Even if the support³ of Λ is the entire space \mathbb{R}^d , the total number of points is still

finite a.s. if $\int_{\mathbb{R}^d} \Lambda(x)dx = \Lambda(\mathbb{R}^d)$ is finite. An important such process is the Poisson process with Gaussian intensity.

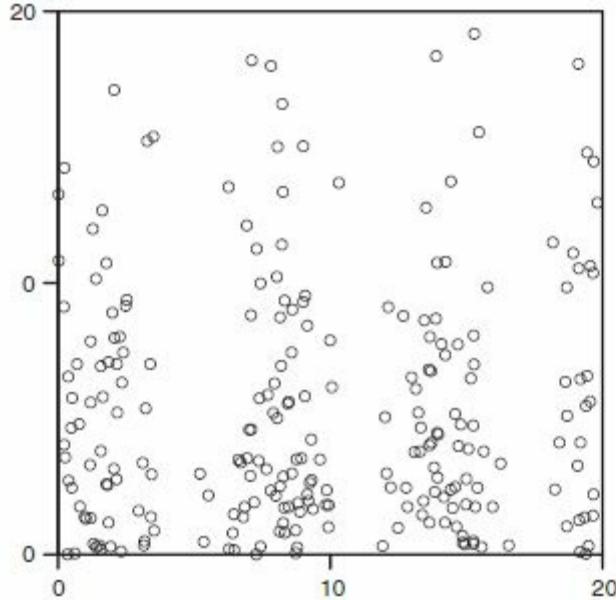


Figure 2.2 A non-uniform PPP of intensity $\lambda(x, y)=2\exp(-y/6)(1 + \sin x)$ on $[0, 20]^2$. It is apparent that the intensity of this PPP is periodic along the x -axis, while it is decreasing along the y -axis.

Example 2.2 The Gaussian Poisson process⁴ on \mathbb{R}^2 is a PPP with intensity function

$$\lambda(x) = \frac{n}{2\pi\sigma^2} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right),$$

where $n = \Lambda(\mathbb{R}^2)$ is the mean total number of points and σ^2 is the variance. When chocolate powder is added to a cappuccino with cup radius r , the chocolate particles may approximately form a truncated Gaussian Poisson process with $\lambda(x)=0$ for $\|x\| > r$, where $r \approx 6$ cm, as illustrated in Fig. 2.3.

Remark 2.1 The general PPP is not necessarily restricted to be diffuse; it may have atoms. If the intensity measure Λ has an atom of mass λ_0 at x_0 then $N(\{x_0\})$ is Poisson distributed with mean λ_0 . If it has at least one atom, then it cannot be a

simple point process, since there is a positive probability that two points (or more) are located at the position of the atom. Conversely, the PPP is simple if and only if it has no atoms or, equivalently, if the intensity measure has no discrete component.

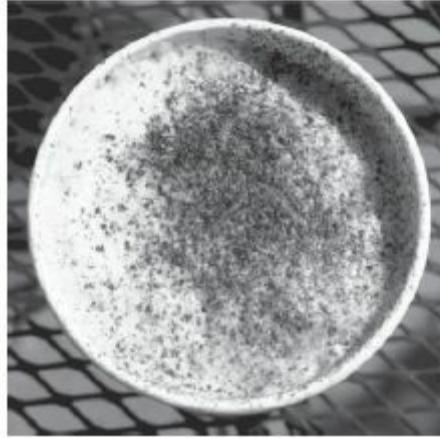


Figure 2.3 Cappuccino chocolate powder point process.

2.4.4 The binomial point process

The binomial point process (BPP) is a basic process whereby a fixed number of nodes n is identically and independently distributed on a compact set $W \subset \mathbb{R}^d$. In this case, the point process may be regarded as a random vector in W^n . The distribution may be uniform or general.

DEFINITION 2.11 (Uniform binomial point process) Let $\Phi = \{x_1, \dots, x_n\} \subset W$ be a point process with a fixed and finite number n of points on the compact set W . Φ is a uniform BPP if and only if Φ , viewed as a random vector, is uniformly distributed in W^n .

From the independence assumption it follows that for Borel subsets A_1, \dots, A_n of W ,

$$\begin{aligned}\mathbb{P}(x_1 \in A_1, \dots, x_n \in A_n) &= \mathbb{P}(x_1 \in A_1) \cdots \mathbb{P}(x_n \in A_n) \\ &= \frac{|A_1| \cdots |A_n|}{|W|^n}.\end{aligned}$$

The name of this process comes from the fact that the number of points in $A \subset W$ is binomially distributed with parameters $n = \Phi(W)$ and $p = |A| / |W|$. Hence the intensity is $\lambda(x) = n \mathbf{1}_W(x) / |W|$, and the intensity measure is

$$\Lambda(A) = n \frac{|A \cap W|}{|W|}, \quad \forall A \in \mathcal{B}^d.$$

Since the total number of points is fixed, the numbers of points in $\Phi(A)$ and $\Phi(W \setminus A)$ are *not independent*, in contrast to the PPP. Clearly, if $\Phi(A) = m$, we must have $\Phi(W \setminus A) = n - m$.

The uniform binomial point process cannot be visually distinguished from the uniform PPP with the same intensity. By Theorem 2.9, the PPP reduces to the binomial point process if its number of nodes is fixed.

In the general case, the spatial distribution over W is arbitrary, according to some pdf f .

DEFINITION 2.12 (General binomial point process) Let f be a probability density with support at most $W \subset \mathbb{R}^d$. A (general) BPP with n points on W is a set of iid random variables $\{x_1, \dots, x_n\}$, each with pdf f . The intensity measure is $\Lambda(A) = n \int_A f(x) dx$.

If we take the number of nodes n to be Poisson distributed with mean $\Lambda(W)$, we obtain a PPP. Conversely, the BPP can be viewed as a conditional PPP. For example, the Gaussian PPP conditioned on having n points is a binomial point process with pdf $f(x) = \exp(-\|x\|^2/(2\sigma^2))/(2\pi\sigma^2)$.

Box 2.4 Equivalence of binomial point process and conditional PPP

The binomial point process with n points and the conditional PPP with n points are equivalent.

2.4.5 The mixed Poisson point process

The mixed Poisson point process is a doubly stochastic point process in the sense that the intensity of the process is itself a random variable, say L . To obtain a realization, λ is first drawn as an instance of L , and then a PPP of intensity λ is generated. More formally, let L be a non-negative random variable defined on Ω . Given $L = \lambda$, Φ is a homogeneous PPP of intensity λ .

The statistics of this mixed PPP are obtained by first conditioning on L , using the Poisson statistics of this process, and then unconditioning on L . As an example, the intensity measure is

$$\mathbb{E}\Phi(B) = \mathbb{E}(\mathbb{E}\Phi(B) | L) = \mathbb{E}(L)|B|.$$

As we shall see, the mixed PPP is non-ergodic, which is to be expected, since a single realization does not reveal the statistics of L . In Section 3.3, we will see that

the mixed PPP is an instance of the larger class of Cox processes.

2.4.6 Randomly translated and perturbed lattices

A lattice is an atomic point process; it can be turned into a diffuse process by a random translation or a perturbation.

Random translation

To define randomly translated lattices, it is useful to first introduce the concept of Voronoi cells and Voronoi tessellations.

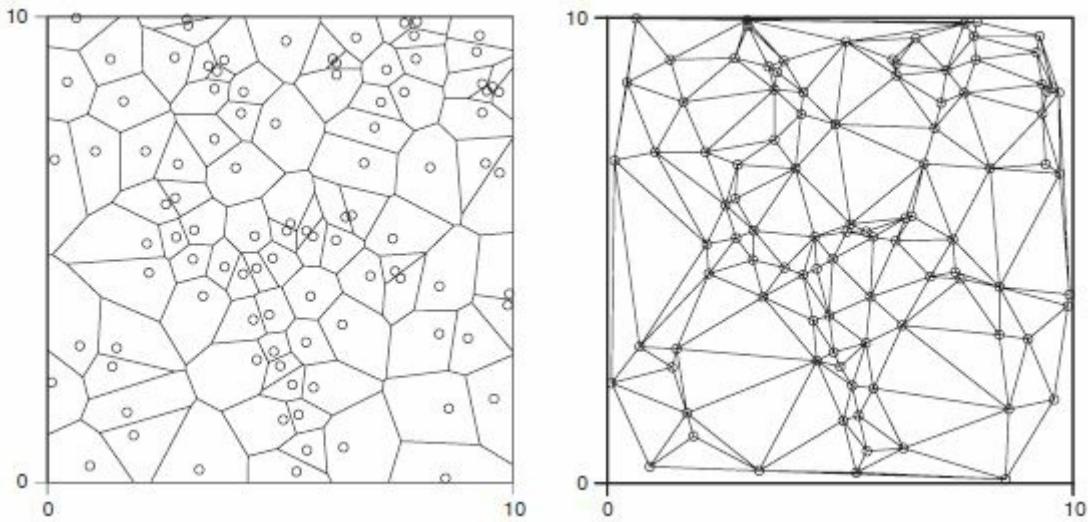


Figure 2.4 (Left) Voronoi tessellation of the PPP of intensity 1 from Fig. 2.1. The line segments represent those locations that are equidistant from two points of the PPP, and their intersections are the ones that are equidistant from three points of the PPP. (Right) Delaunay triangulation of the same point set. Here it is assumed that there are no points outside the $[0, 10]^2$ box. The union of the triangles is the convex hull of the point set.

DEFINITION 2.13 (Voronoi cell) The Voronoi cell $V(x)$ of a point x of a general point process $\Phi \subset \mathbb{R}^d$ consists of those locations of \mathbb{R}^d whose distance to x is not greater than their distance to any other point in Φ , i.e.,

$$V(x) \triangleq \{y \in \mathbb{R}^d : \|x - y\| \leq \|z - y\| \quad \forall z \in \Phi \setminus \{x\}\}.$$

This can be written more compactly as

$$V(x) = \{y \in \mathbb{R}^d : \|x - y\| \leq \|y - \Phi\|\}.$$

DEFINITION 2.14 (Voronoi tessellation) The Voronoi tessellation (or Voronoi diagram) is a decomposition of the space into the Voronoi cells of a general point

process.

The Voronoi cells do not partition the space, since the boundaries of each cell belong to the neighboring cell also.

The *Delaunay triangulation* is obtained by connecting by a straight line segment all pairs of points whose Voronoi cells share a common boundary. Figure 2.4 shows an example of a Voronoi tessellation and a Delaunay triangulation of a PPP. The Delaunay triangulation induces a natural, parameter-free, definition of the neighbors of a point x , namely those points connected to x by a line in the triangulation. Other definitions of neighborhood require a parameter such as a distance.

Thus equipped, we can return to random translations of lattices. General lattices are defined as

$$\mathbb{L} \triangleq \{u \in \mathbb{Z}^d : \mathbf{G}u\},$$

where $\mathbf{G} \in \mathbb{R}^{d \times d}$ is an arbitrary matrix with $\det \mathbf{G} \neq 0$, the so-called *generator matrix*. In this general case, the random translation vector is a random variable uniformly distributed over the Voronoi cell $V(o)$ of the point at the origin. Its volume $|V(o)|$ is $|\det \mathbf{G}|$. The randomly translated lattice is then defined as follows.

DEFINITION 2.15 (Randomly translated lattice) Let \mathbb{L} be a lattice and $V(o)$ the Voronoi cell of the origin. The randomly translated lattice is

$$\Phi \triangleq \{u \in \mathbb{Z}^d : \mathbf{G}u + X\},$$

where X is uniformly distributed over $V(o)$.

In the case of the square lattice with spacing η denoted by $\mathbb{L}_\eta^d \triangleq \{\eta \mathbb{Z}^d\}$, for which $\mathbf{G} = \eta \mathbf{I}$, where \mathbf{I} is the $d \times d$ identity matrix, X is uniformly distributed on $[0, \eta]^d$. The randomly translated lattice is $\Phi = \mathbb{L}_\eta^d + X$, and its density is $\lambda = \eta^{-d}$. The standard case is the randomly translated two-dimensional integer lattice, defined as $\Phi \triangleq \mathbb{Z}^2 + (U, V)$, where U and V are uniform random variables over the unit interval $[0, 1]$. In one dimension, using the standard probability space $([0, 1], \mathcal{B} \cap [0, 1], \nu_1)$, the randomly translated lattice may also be characterized by the counting measure

$$N = \sum_{n=-\infty}^{\infty} \delta_{\eta(n+\omega)}. \quad (2.3)$$

Perturbed lattice

Here, each point is perturbed by an independent random vector.

DEFINITION 2.16 (Perturbed lattice) Let (X_u) , $u \in \mathbb{Z}^d$, be a family of iid random variables with pdf f . The perturbed lattice is defined as

$$\Phi \triangleq \{u \in \mathbb{Z}^d : Gu + X_u\}.$$

A standard case is the Gaussian disturbance, where f is a Gaussian pdf.

2.4.7 The Bernoulli lattice process

Let $\tilde{\mathbb{L}}_\eta^d = \mathbb{L}_\eta^d + X$ be the randomly translated square lattice with spacing $\eta > 0$. A Bernoulli lattice process $\Phi_{\eta, p}$ is a point process in which each point of $\tilde{\mathbb{L}}_\eta^d$ is retained independently with probability p . Hence, given X , $\Phi_{\eta, p}(A)$ is binomially distributed with parameters $n = \#\{A \cap \tilde{\mathbb{L}}_\eta^d\}$ and p , and $\Phi_{\eta, p}(A_1)$ and $\Phi_{\eta, p}(A_2)$ are independent if A_1 and A_2 are disjoint. Averaging over X , we obtain

$$\mathbb{E}\Phi_{\eta, p}(A) = p \cdot \mathbb{E}(\#\{A \cap \tilde{\mathbb{L}}_\eta^d\}) = p\eta^{-d}|A|$$

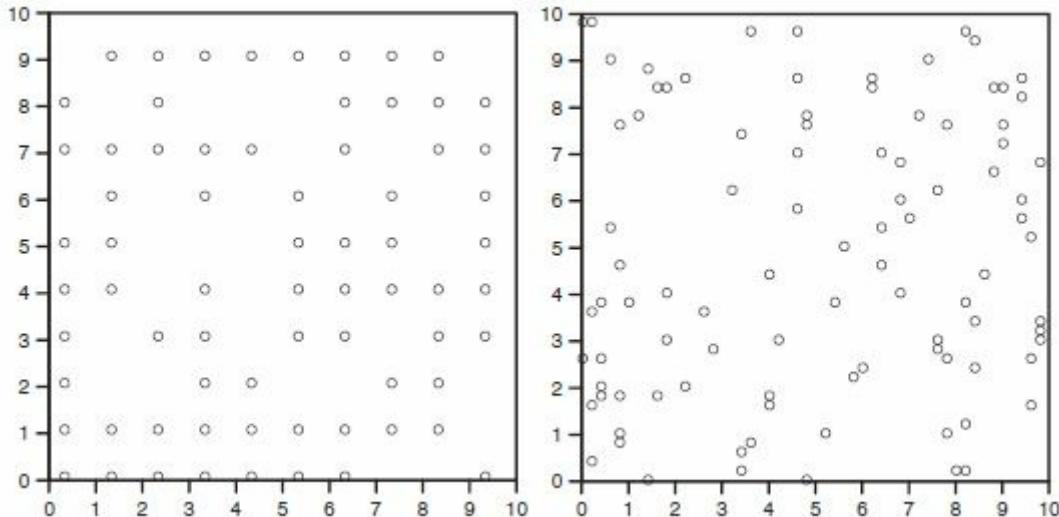


Figure 2.5 Bernoulli lattice processes on $[0, 10]^2$. (Left) $\eta = 1, p = 0.7$. (Right) $\eta = 0.2, p = 0.04$.

for the expected number of points in A . Two realizations of such processes with different parameters are shown in Fig. 2.5.

What is interesting about this process are its limiting properties as η and p go to zero in such a way that $p\eta^{-d} = \lambda$ is held constant. If A is open and bounded, the

distribution of $\Phi_{\eta, p}(A)$ tends to a Poisson distribution with mean $\lambda|A|$. Since the lattice gets dense, the points will be uniformly spread on A , and the process appears to approach a Poisson point process.

One way to define a sequence of denser and denser processes with constant intensity is as follows. Consider the sequences of lattices for a fixed X for which $\eta k = 2^{-k}$. Clearly $\tilde{\mathbb{L}}_{\eta_1}^d$ is a subset of $\tilde{\mathbb{L}}_{\eta_2}^d$, which is a subset of $\tilde{\mathbb{L}}_{\eta_3}^d$, and so on. Now let $p_k = 2^{-kd}\lambda$, so that we have $p_k \eta_k^{-d} = \lambda$ for all k . Intuitively this sequence of processes converges in distribution to a limiting process Φ , which is, in fact, a PPP of intensity λ .

In Fig. 2.5 (right), where $\eta = 0.2$ and $p = 0.04$, the process starts to look similar to a PPP of intensity 1. For $\eta = 0.1$ and $p = 0.01$, it would be visually indistinguishable from the PPP.

2.5 Distributional characterization

Like any random process, a point process can be described in statistical terms by defining the space of possible outcomes and then specifying the probabilities of different events.

2.5.1 The distribution of a point process

The space of realizations of a point process in \mathbb{R}^d , denoted by \mathcal{N} , is the set of all counting measures on \mathbb{R}^d . A counting measure is a non-negative, integer-valued measure that is finite on compact sets. Alternatively, when simple point processes are viewed as random sets, \mathcal{N} is the space of locally finite countable subsets of \mathbb{R}^d . We will usually denote its elements, the counting measures or countable sets on \mathbb{R}^d , also called point patterns, by φ .

DEFINITION 2.17 (Basic events) A basic event $E_{B, k}$ about the point process is the event that there are exactly k points in the region B , for compact $B \subset \mathbb{R}^d$ and $k \in \mathbb{N}_0$

$$E_{B, k} \triangleq \{N(B) = k\} = \{\varphi \in \mathcal{N} : \varphi(B) = k\}. \quad (2.4)$$

DEFINITION 2.18 (Canonical space) Let \mathcal{N} be the set of all counting measures on \mathbb{R}^d and \mathfrak{M} be the σ -field of subsets of \mathcal{N} generated by all events of the form $E_{B, k}$. The space \mathcal{N} equipped with its σ -field \mathfrak{M} is called the *canonical space* or *outcome space* for a point process in \mathbb{R}^d .

The σ -field \mathfrak{M} is actually the Borel σ -algebra on \mathcal{N} .⁵

DEFINITION 2.19 (Point process) A point process Φ is a measurable map $N : \Omega \rightarrow \mathcal{N}$ from a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ to the outcome space $(\mathcal{N}, \mathfrak{N})$.

Each elementary outcome $\omega \in \Omega$ determines an outcome $N^\omega \in \mathcal{N}$ for the entire point process.

DEFINITION 2.20 (Point process distribution) The *distribution* of N_Φ or Φ is the probability measure pertaining to the outcome measure space:

$$\mathbb{P}(E) \triangleq \mathbb{P} \circ N^{-1}(E) \equiv \mathbb{P}(N \in E) \quad \forall E \in \mathfrak{N}.$$

The symbol \circ stands for the concatenation of functions, e.g., $(f \circ g)(t) = f(g(t))$. Measurability requires that $N^{-1}(E) \in \mathcal{A}$.

The counting random variables are indexed by the Borel sets B . They form an \mathbb{N}_0 -valued family of random variables (or a random process) on \mathcal{B}^d ,

$$\mathcal{F}_N \triangleq \{B \in \mathcal{B}^d : N(B)\},$$

which is usually called a *random field*. If $B_1 \cap B_2 \neq \emptyset$, then clearly $N(B_1)$ and $N(B_2)$ cannot be independent. If B_1 and B_2 are disjoint, then $N(B_1)$ and $N(B_2)$ may be independent, as in the case of the PPP, but in many cases they are not. In fact, the random structure of the point process induces an intricate dependence of the counting random variables.

A general random variable (not necessarily numeric) is also referred to as a *random element*. So, a point process is a random element $(\Omega, \mathcal{A}, \mathbb{P}) \mapsto (\mathcal{N}, \mathfrak{N})$.

Remark 2.2 A possible source of confusion is the fact that in the case of numerical random variables, the Borel sets B are events. In the case of point processes, the Borel sets are indices of the counting random process, or arguments if the counting random variables are viewed as functions on Borel sets. Also, all $N(B)$ are themselves numerical (\mathbb{N}_0 -valued) random variables.

An event in the point process space is often referred to as a *property*. This is intuitive, since properties such as having no point in some area, or three points on a line, are indeed events.

2.5.2 The distribution of a discrete point process

Let $\varphi = \delta_x$ and consider the point process distribution

$$\mathbb{P}(E) = \begin{cases} 1 & \text{if } E = \{\varphi\} \\ 0 & \text{if } E \neq \{\varphi\}. \end{cases}$$

What kind of point process does this distribution characterize? The only counting measure that is possible as a realization is δ_x , corresponding to the (deterministic) point set $\{x\}$. We can write this degenerate distribution in a more compact form by noting that $\mathbb{P}(E) = \delta_\varphi(E)$, and thus $\mathbb{P} = \delta_{\delta_x}$. So whenever only a single realization δ is possible, $\mathbb{P} = \delta_\varphi$ in much the same way as, for a random variable, $\mathbb{P}(X \in B) = \delta_x(B)$ indicates that the only possible outcome is $X = x$.

If $\varphi = \delta_x + \delta_y$ still deterministic, we have $\mathbb{P} = \delta_{\delta_x + \delta_y}$. This process is the superposition of the two processes $\varphi_1 = \delta_x$ and $\varphi_2 = \delta_y$, and we write the distribution of $\Phi = \Phi_1 + \Phi_2$ as the *convolution*

$$\mathbb{P} = \mathbb{P}_1 * \mathbb{P}_2. \quad (2.5)$$

To see why this notation makes sense, compare it with the distribution of two deterministic (degenerate) numerical random variables: For $X_1 = 1$ and $X_2 = 2$, the distributions are $\mathbb{P}(X_1 \in B) = \delta_1(B)$ and $\mathbb{P}(X_2 \in B) = \delta_2(B)$ and the distribution of the superposition $X_1 + X_2$ is $\mathbb{P}(X_1 + X_2) = (\delta_1 * \delta_2)(B) = \delta_{1+2}(B)$.

The convolution expression (2.5) is used for general point processes (not just deterministic or atomic ones). If \mathbb{P}_1 is the distribution of an arbitrary point process Φ_1 and $\mathbb{P}_2 = \delta_{\delta_x}$ then $\mathbb{P} = \mathbb{P}_1 * \delta_{\delta_x}$ is the distribution of the point process $\Phi_1 \cup \{x\}$, or, as a random measure, $N = N_1 + \delta_x$. Adding δ_x to the random element N_1 corresponds to a convolution of the distribution by δ_{δ_x} .

If two realizations are possible, say φ_1 with probability 1/3 and φ_2 with probability 2/3, we have $\mathbb{P} = \frac{1}{3}\delta_{\varphi_1} + \frac{2}{3}\delta_{\varphi_2}$. The distribution may also be written in the form

$$\mathbb{P} = \frac{1}{3}\mathbf{1}(\varphi_1 \in \cdot) + \frac{2}{3}\mathbf{1}(\varphi_2 \in \cdot),$$

which is to be interpreted as

$$\mathbb{P}(E) = \frac{1}{3}\mathbf{1}(\varphi_1 \in E) + \frac{2}{3}\mathbf{1}(\varphi_2 \in E), \quad \forall E \in \mathfrak{N}.$$

Since discrete point processes have at most a countably infinite number of realizations, we can write down explicitly the distribution of all discrete point processes.

Example 2.3 Let Φ be the point process on \mathbb{R} where the two points -1 and 1 are

retained independently with probability 1/2. There are four possible outcomes, $\varphi_1 = 0$, $\varphi_2 = \delta_{-1}$, $\varphi_3 = \delta_1$, and $\varphi_4 = \delta_{-1} + \delta_1$, each one with probability 1/4. So

$$P = \frac{1}{4}(\delta_0 + \delta_{\delta_1} + \delta_{\delta_{-1}} + \delta_{\delta_1 + \delta_{-1}}).$$

The measure δ_0 refers to the empty point process (with no points) $\Phi = \emptyset$.

2.5.3 Comparison with numerical random variables

It is instructive to compare the definitions of point processes and standard (numerical) random variables.

Numerical random variables

Let X be a random variable on (Ω, \mathcal{A}, P) , i.e., an \mathcal{A} -measurable function $\Omega \mapsto \mathbb{R}$.

The *distribution* of X is the measure

$$P \triangleq P \circ X^{-1}$$

on $(\mathbb{R}, \mathcal{B})$, given by

$$P(B) = P \circ X^{-1}(B) = P(X \in B) \quad \forall B \in \mathcal{B},$$

where the set $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ is the *pre-image* of B . Elementary events in \mathcal{B} are the intervals $[a, b]$, $b > a$, and \mathcal{B} is their σ -algebra, and *measurability* is the requirement that $X^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$.

For a numerical random variable X , we almost always restrict ourselves to the *distribution function* (cdf) of X , where $B = (-\infty, x]$:

$$F(x) \triangleq P((-\infty, x]) = P(X \leq x).$$

The distribution function is right-continuous, always exists, and fully describes the distribution.

Table 2.1 shows a detailed comparison of numerical random variables with point processes. We write N^ω , since $N(\omega)$ may be confused with $N(B)$.

Most often, the underlying probability space can be taken to be the generic one on the unit interval: $(\Omega, \mathcal{A}, P) = ([0, 1], \mathcal{B} \cap [0, 1], \nu_1)$.

2.5.4 Integration over the space of simple sequences

As mentioned previously, a point process can be interpreted as a random measure or a random set or sequence. Let φ denote a locally finite countable subset of \mathbb{R}^d (or a point pattern) *and* the associated (deterministic) counting measure:

$$\varphi = \{x_1, x_2, \dots\}; \quad \varphi(B) = \sum_{x \in B} \mathbf{1}_B(x) \quad \forall B \in \mathcal{B}^d.$$

Table 2.1 Comparison of numerical random variables and point processes as random elements

	Numerical random variable	Point process
Probability space	$(\Omega, \mathcal{A}, \mathbb{P})$	$(\Omega, \mathcal{A}, \mathbb{P})$
Measurable space	$(\mathbb{R}, \mathcal{B})$	$(\mathcal{N}, \mathfrak{N})$
Random element	$X(\omega) \in \mathbb{R}$	$N^\omega \in \mathcal{N}$
Event	$B \in \mathcal{B}$	$E \in \mathfrak{N}$
Pre-image of event	$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$	$N^{-1}(E) = \{\omega \in \Omega : N^\omega \in E\}$
Measurability condition	$X^{-1}(B) \in \mathcal{A}$	$N^{-1}(E) \in \mathcal{A}$
Distribution	$P(B) = \mathbb{P} \circ X^{-1}(B) = \mathbb{P}(X \in B)$	$P(E) = \mathbb{P} \circ N^{-1}(E) = \mathbb{P}(N \in E)$
Measure space	$(\mathbb{R}, \mathcal{B}, P)$	$(\mathcal{N}, \mathfrak{N}, \mathbb{P})$
Normalization	$P(\mathbb{R}) = 1$	$\mathbb{P}(\mathcal{N}) = 1$
Differential element	dx	$d\phi$
Distribution function	$F(x) = P((-\infty, x])$	—
Expectation	$\mathbb{E}f(X) = \int_{\mathbb{R}} f(x)P(dx)$	$\mathbb{E}f(N) = \int_{\mathcal{N}} f(\phi)P(d\phi)$

If φ is viewed as a counting measure, then the outcome space \mathcal{N} is the set of all φ , and a point process Φ is a random choice of one of the φ in \mathcal{N} . Since the outcome space has probability 1, i.e., $\mathbb{P}(\mathcal{N}) = 1$, necessarily

$$\int_{\mathcal{N}} \mathbb{P}(d\varphi) = 1,$$

in the same way that for numerical random variables with distribution function (cdf) $F(x)$ and its derivative, the density function (pdf) $f(x)$, we have

$$\int_{\mathbb{R}} P(dx) = \int_{\mathbb{R}} dF(x) = \int_{\mathbb{R}} f(x)dx = 1.$$

To make the connection between the original probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and the outcome probability space $(\mathbb{R}, \mathcal{B}, P)$ explicit, we may write the expected value as

$$\mathbb{E}(X) = \int_{\Omega} X(\omega)\mathbb{P}(d\omega) = \int_{\mathbb{R}} xP(dx) = \int_{\mathbb{R}} x dF(x)$$

for numerical random variables. For point processes, the mapping is from $(\Omega, \mathcal{A}, \mathbb{P})$ to $(\mathcal{N}, \mathfrak{N}, \mathbb{P})$, thus the intensity measure can be expressed as

$$\Lambda(B) = \mathbb{E}N(B) = \mathbb{E}\left(\int_B N(dx)\right) = \int_{\Omega} N^{\omega}(B)\mathbb{P}(d\omega) = \int_{\mathcal{N}} \varphi(B)\mathbb{P}(d\varphi).$$

A general expectation over the point process is given by

$$\mathbb{E}f(\Phi) = \int_{\mathcal{N}} f(\varphi)\mathbb{P}(d\varphi)$$

for $f : \mathcal{N} \mapsto \mathbb{R}$ (when f is a function of the point process) or

$$\mathbb{E}f(\Phi(B)) = \int_{\mathcal{N}} f(\varphi(B))\mathbb{P}(d\varphi)$$

for $f : \mathbb{N}_0 \mapsto \mathbb{R}$ (when f is a function of the number of points in B). We will make extensive use of this notation later.

2.5.5 Finite-dimensional distributions and capacity functional

On specializing the point process distribution to events of the form

$$\bigcap_{i=1}^m \{\varphi \in \mathcal{N} : \varphi(B_i = k_i)\} = \{N(B_1) = k_1, \dots, N(B_m) = k_m\},$$

where m is finite, we arrive at the *finite-dimensional distribution*, often abbreviated to fidi distribution.

DEFINITION 2.21 (Finite-dimensional distribution) The finite-dimensional (fidi) distributions of a point process are the joint probability distributions of

$$(N(B_1), \dots, N(B_m))$$

for all finite integers $m > 0$ and all compact B_1, B_2, \dots, B_m .

Fidi distributions measure events that are intersections of a finite number of basic events as defined in Definition 2.17.

Example 2.4 The fidi distributions of the PPP for disjoint sets B_1, \dots, B_k are given by multiplication of the probabilities $\mathbb{P}(\Phi(B_i) = n_i)$:

$$\mathbb{P}(\Phi(B_1) = n_1, \dots, \Phi(B_k) = n_k)$$

$$= \frac{\lambda^{n_1 + \dots + n_k} |B_1|^{n_1} \dots |B_k|^{n_k}}{n_1! \dots n_k!} \exp\left(-\sum_{i=1}^k \lambda |B_i|\right).$$

This permits the calculation of the joint probabilities $\mathbb{P}(\Phi(B_1) = n_1, \dots, \Phi(B_k) = n_k)$ for generally (possibly overlapping) B_1, \dots, B_k (see Problem 2.3).

Example 2.5 For a uniform binomial point process on $W \subset \mathbb{R}^d$ with partition A_1, \dots, A_k and $n_1 + n_2 + \dots + n_k = n$, the fidi distributions are the multinomial distributions

$$\mathbb{P}(\Phi(A_1) = n_1, \dots, \Phi(A_k) = n_k) = \frac{n!}{n_1! \dots n_k!} \cdot \frac{|A_1|^{n_1} \dots |A_k|^{n_k}}{|W|^n}.$$

DEFINITION 2.22 (Capacity functional) The capacity functional T of a simple point process Φ is defined as

$$T(K) \triangleq \mathbb{P}(N(K) > 0), K \text{ compact.}$$

Hence the capacity functional is the complement of the void probability: $T(K) = 1 - \mathbb{P}(N(K) = 0)$. The importance of the fidi distributions and the capacity functional is illustrated by the following theorem.

THEOREM 2.23 (Fidi distributions and void probabilities)

- If the fidi distributions of two point processes are identical, then they have the same distribution.
- If the capacity functionals (or void probabilities) of two simple point processes are identical, then they have the same distribution.

We do not provide a proof of this theorem but give a proof sketch of Renyi's theorem, which focuses on PPPs.

THEOREM 2.24 (Rényi's theorem) Let $\Phi \in \mathcal{P}(W)$ be a point process and $\lambda: W \rightarrow \mathbb{R}^+$ be a function such that $\Lambda(B) = \int_B \lambda(x) dx < \infty$ for all bounded B . If

$$\mathbb{P}(\Phi(B) = 0) = \exp(-\Lambda(B)) \tag{2.6}$$

for any Borel B then Φ is Poisson with intensity function λ .

Proof (Sketch) The proof idea is the following. Consider boxes of the form

$$B_{\mathbf{k}}(n) = \prod_{i=1}^d (k_i 2^{-n}, (k_i + 1) 2^{-n}) \quad \text{for } n \geq 1 \text{ and } \mathbf{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d.$$

Let $C_k(n) = \mathbf{1}(\Phi(B_k(n)) > 0) = T(B_k(n))$. These indicators are independent due to (2.6). Letting

$$T_n(B) = \sum_{\mathbf{k} \in \mathbb{Z}^d} (\mathbf{1}(B_{\mathbf{k}}(n)) \subset B) \cdot C_{\mathbf{k}}(n),$$

we have $\Phi(B) = \lim_{n \rightarrow \infty} T_n(B)$. Since $T_n(B)$ is the sum of independent random variables, its probability generating function can be written explicitly, and monotone convergence implies that this function approaches that of $\Phi(B)$ as $n \rightarrow \infty$ which can be shown to be that of the Poisson distribution with mean $\Lambda(B)$.

The proof technique indicates that it is sufficient that (2.6) holds for finite unions of (hyper)rectangles.

A direct consequence of this theorem is the fact that the superposition $\Phi = \Phi_1 \cup \Phi_2$ of two independent PPPs with intensity measures Λ_1 and Λ_2 is a PPP with intensity measure $\Lambda_1 + \Lambda_2$. To show this, the simple calculation

$$\mathbb{P}(\Phi_1(B) = 0, \Phi_2(B) = 0) = e^{-\Lambda_1(B)} e^{-\Lambda_2(B)} = e^{-(\Lambda_1(B) + \Lambda_2(B))}$$

suffices.

The last result in this section establishes the connection between general PPPs and general binomial point processes.

THEOREM 2.25 (Conditional property for general PPPs) *Let $\Phi \subset \mathbb{R}^d$ be a general PPP with intensity function $\lambda(x)$. Take B such that $0 < \Lambda(B) < \infty$. Conditioned on $\Phi(B) = n$, the n points in B are independent and distributed with probability density*

$$\frac{\lambda(x)}{\Lambda(B)}.$$

Proof In the homogeneous case (constant λ), we have seen that, conditioned on $\Phi(B) = n$, the n points form a binomial point process. This is a straightforward generalization to the inhomogeneous case. The intensity $\Phi(x)$ needs to be normalized by $\Lambda(B)$ in order for it to be a proper probability density. \square

In general, the intensity (function) $\lambda(x)$ need not exist. If it does, we can retrieve it from the intensity measure by calculating the limit

$$\lambda(x) = \lim_{r \rightarrow 0} \frac{\mathbb{E}\Phi(b(x, r))}{|b(x, r)|}.$$

The denominator is the volume of the d -dimensional ball of radius r , which is $c_d r^d$, where

$$c_d = \frac{\pi^{d/2}}{\Gamma(d/2 + 1)} \quad (2.7)$$

is the volume of the unit ball.

2.6 Properties of point processes

In this section, we introduce and define several important properties of point processes. First, we introduce notation for a translated point process.

DEFINITION 2.26 (Translated point process and events) If $\Phi = \{x_1, x_2, \dots\}$ is a point process, then $\Phi_x \triangleq \{x_1 + x, x_2 + x, \dots\}$ is the point process translated by $x \in \mathbb{R}^d$. Similarly, for an event E , the translated event E_x is defined as

$$E \triangleq \{\varphi \in \mathcal{N}: \varphi_x \in E\} \quad E \in \mathfrak{N}.$$

So, for example, if E is the event that $\Phi \subset \mathbb{R}^2$ does not have any points in the ball $b(o, r)$, then $E_{(1, 0)}$ is the event that Φ does not have any points in the ball $b((1, 0), r)$. Analogously, we may denote the translated counting measure by N_x :

$$N_x = \mathbf{N}(\cdot - x).$$

This notation means that $N_x(B) = N(B - x)$ for all $B \in \mathcal{B}$, where $B - x = \{y \in B: y - x\}$.

DEFINITION 2.27 (Stationarity) A point process on \mathbb{R}^d is *stationary* if its distribution is translation-invariant, i.e., $P(E) \equiv P(E_x)$, for all $E \in \mathfrak{N}$ and $x \in \mathbb{R}^d$.

Equivalently, stationarity may be defined as the property that $\mathbb{P}(\Phi \in E) = \mathbb{P}(\Phi_x \in E)$, again for all E and x .

An atomic point process cannot be stationary. While stationarity implies that the intensity function is constant, the converse is not true: A constant intensity function does *not* imply stationarity.

DEFINITION 2.28 (Isotropy) A point process on \mathbb{R}^d is *isotropic*, if its distribution is rotationally invariant with respect to rotations about the origin o , i.e., $P(E) = P(rE)$, where r is an arbitrary rotation about o in \mathbb{R}^d , i.e.,

$$\mathbf{r}E = \{\varphi \in \mathcal{N} : \mathbf{r}^{-1}\varphi \in E\}, \quad E \in \mathfrak{N}.$$

DEFINITION 2.29 (Motion-invariance) A stationary and isotropic point process is called *motion-invariant*.

The class of motion-invariant point processes is important for their analytical tractability. Among the non-stationary point processes, the most tractable ones are certainly the general PPPs.

For stationary point processes, the intensity measure is translation-invariant:

$$\Lambda(B) = \mathbb{E}\Phi(B) \stackrel{\text{st}}{=} \mathbb{E}\Phi_x(B) = \mathbb{E}\Phi(B_{-x}) = \Lambda(B_{-x}), \quad \forall x \in \mathbb{R}^d.$$

The equality denoted by $\stackrel{\text{st}}{=}$ holds only for stationary point processes, the others hold in general. It follows that the intensity measure is then just a multiple of the Lebesgue measure, with λ being the proportionality constant:

$$\Lambda(B) = \lambda|B|.$$

Example 2.6 Examples of isotropic, non-stationary point processes:

- The inhomogeneous PPP $\Phi \subset \mathbb{R}^d$ with *radially symmetric* intensity function $\lambda(x) = f(\|x\|)$.
- The binomial point process of constant intensity defined on $b(o, r)$.

Example 2.7 An example of a stationary, non-isotropic point process: The randomly translated square lattice $\Phi = \mathbb{Z}^2 + (U, V)$, where U and V are uniform random variables over the unit interval $(0, 1)$. To demonstrate anisotropy, let B be a square of side length $11/10$. Then $\mathbb{P}(\Phi(B) = 0) = 0$, while for the square (with the same area) rotated by $\pi/4$, denoted by B_\diamond , we have $\mathbb{P}(\Phi(B_\diamond) = 0) > 0$, since the rotated square fits in the lattice without intersecting with any lattice point.

Another important property is ergodicity. If a point process is ergodic, it suffices to analyze one realization of the point process on an appropriately large window to obtain statistically meaningful results, since, loosely speaking, the spatial average corresponds to the ensemble average. Formally, ergodicity is defined as follows.

DEFINITION 2.30 (Ergodicity) Let $W_n \subset \mathbb{R}^d$ be a sequence of averaging windows with the following properties.

1. Each W_n is convex and compact.
2. $W_n \subset W_{n+1}$.

3. $\sup\{r \geq 0 : b(x, r) \subset W_n \text{ for some } x\} \rightarrow \infty$ as $n \rightarrow \infty$.

The point process is *ergodic* if

$$\frac{1}{|W_n|} \int_{W_n} P(E \cap Y_x) dx \rightarrow P(E)P(Y), \quad n \rightarrow \infty, \quad (2.8)$$

for all events $E, Y \in \mathfrak{N}$.

The third property ensures that W_n grows without bounds. The condition (2.8) of ergodicity is hard to verify. A stronger condition that is easier to work with is mixing.

DEFINITION 2.31 (Mixing) A stationary point process is *mixing* if

$$P(E \cap Y_x) \rightarrow P(E)P(Y) \text{ as } \|x\| \rightarrow \infty, \quad \forall E, Y \in \mathfrak{N}$$

If a point process is mixing, it is also ergodic. The PPP is mixing since the event E_x becomes independent of Y if the translation vector is long enough.

An equivalent condition to ergodicity is metrical transitivity.

DEFINITION 2.32 (Metrical transitivity) A stationary distribution P on $(\mathcal{N}, \mathfrak{N})$ is said to be *metrically transitive* if

$$P(E) \text{ is 0 or 1 for all invariant sets } E \in \mathfrak{N}$$

Here an *invariant set* E in \mathcal{N} is one for which

$$P(E \setminus E_x \cup E_x \setminus E) = 0 \quad \text{for all } x \in \mathbb{R}^d.$$

A point process is ergodic if and only if it is metrically transitive, which may be a more intuitive property; it means that events that make no explicit or implicit reference to the location of Φ are either true a.s. or false a.s. Examples of such events are

$$E_1 = \{\varphi \in \mathcal{N} : \text{there are three points in a ball of radius 1 somewhere in } \mathbb{R}^d\}$$

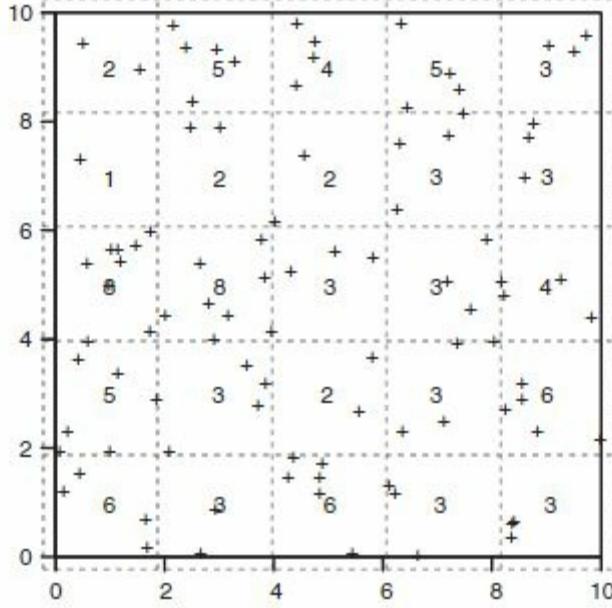


Figure 2.6 A uniform PPP of intensity $\lambda = 1$ on $[0, 1]^2$ and *quadrat counts* (number of points) in 25 squares of side length 2. Since the PPP is ergodic, the empirical distribution of the counts C_i is close to Poisson, with mean $100/25 = 4$ and variance 4. In this realization, there are 96 points, and the variance of the quadrat counts is 3.39.

and

$$E_2 = \{\varphi \in \mathcal{N} : \text{there are three points on a line somewhere in } \mathbb{R}^d\}.$$

For a PPP (with almost arbitrary intensity function), $P(E_1) = 1$ while $P(E_2) = 0$. The stationary PPP is ergodic whereas the mixed Poisson point process is not.

If Φ is an ergodic point process then

$$\lim_{W_n \uparrow \mathbb{R}^d} \frac{\Phi(W_n)}{|W_n|} = \lambda \quad \text{a.s.}$$

for the sequence of sets introduced in Definition 2.30. This is the law of large numbers applied to point processes. It can be proved using the Chebyshev inequality and the Borel–Cantelli lemma.

These properties have implications for simulations, since instead of generating many instances of a point process, it is equivalently possible to simulate a point process over a large enough area or volume and inspect different sub-regions. An example is shown in Fig. 2.6, where, for a PPP of intensity 1, the numbers of points falling into 2×2 sub-squares are indicated. Since these squares B_1, B_2, \dots are disjoint, the counts $\varphi(B_1), \varphi(B_2), \dots$ have the same statistics as the sequence $\varphi_i(B_1), \varphi_2(B_1), \dots$ for different realizations φ_i of the PPP. Their empirical distribution

tends to the Poisson distribution, with mean $100/25 = 4$. Such statistics are useful when an empirical data set is available, and the question is whether the data set is a realization of a PPP or not.

2.7 Point process transformations

2.7.1 Displacement

The first transformation we consider is the displacement of the points of the point process by a random translation vector,

$$\Phi' = \{x \in \Phi : x + V_x\}, \quad (2.9)$$

where the random variables V_x are independent but their distribution may depend on x . When applied to a PPP, the resulting process is again a PPP.

THEOREM 2.33 (Displacement theorem) *Let $\Phi \subset \mathbb{R}^d$ be a general PPP with intensity function $\lambda(x)$. If all points are independently displaced such that the distribution of the displaced location of a point at x has probability density $\rho(x, \cdot)$, the displaced points form a PPP Φ' with intensity function*

$$\lambda'(y) = \int_{\mathbb{R}^d} \lambda(x) \rho(x, y) dx. \quad (2.10)$$

In particular, if $\lambda(x) \equiv \lambda$ and $\rho(x, y)$ is a function only of $y - x$, then $\lambda'(y) \equiv \lambda$.

In this formulation, $\rho(x, \cdot)$ is the pdf of the new location $x + V_x$ in (2.9), also called the *displacement kernel*, and the second statement in the theorem asserts that if V_x does not depend on x and Φ is uniform with intensity λ , then Φ' is also uniform with the same intensity. If ρ is a function of $y - x$ only, (2.10) is a convolution integral. We defer the proof to [Section 7.2.3](#).

Example 2.8 Let Φ be a PPP on the plane with $\lambda(x) = n\mathbf{1}_{[0, 1]^2}(x)$. The displacement kernel $\rho(x, y) = \exp(-\|y\|/(2\sigma^2))/(2\pi\sigma^2)$ yields the Gaussian Poisson process from [Example 2.2](#). In this case, the new location of each point is independent of its original location.

Setting $\rho(x, y) = \delta(x - v)$ for a general PPP displaces all the points by the deterministic vector v . In this case, $\lambda'(y) = \lambda(y - v)$.

Lastly, on the plane again, $\rho(x, y) = \mathbf{1}_{b(x, r)}(y)/(\pi r^2) = \mathbf{1}_{b(o, r)}(y - x)/(\pi r^2)$

displaces each point by a random vector uniformly on $b(o, r)$.

The displacement theorem shows that the PPP is persistent, in the sense that independent displacements preserve the Poisson-ness, which makes it an attractive model when nodes in a wireless network move around independently.

2.7.2 Mapping

Point processes may be transformed by mapping each point of the process to another point, possibly in a space of different dimension. When applied to Poisson point processes, the resulting process is still Poisson in many cases.

THEOREM 2.34 (Mapping theorem) *Let Φ be a PPP on \mathbb{R}^d with intensity Λ and intensity function λ , and let $f: \mathbb{R}^d \mapsto \mathbb{R}^s$ be a measurable function with the property that $\Lambda(f^{-1}\{y\}) = 0$, $\forall y \in \mathbb{R}^s$, such that f does not shrink a (non-singleton) compact set to a singleton. Then*

$$\Phi' = f(\Phi) \triangleq \bigcup_{x \in \Phi} \{f(x)\}$$

is a PPP with intensity measure

$$\Lambda'(B') = \Lambda(f^{-1}(B')) = \int_{f^{-1}(B')} \lambda(x) dx \quad \text{for all compact } B' \subset \mathbb{R}^s.$$

Proof $\Phi'(B')$ is Poisson distributed with mean $\Lambda(f^{-1}(B'))$. If B'_1, \dots, B'_n are pairwise disjoint, so are the pre-images $f^{-1}(B'_1), \dots, f^{-1}(B'_n)$. Hence, the random variables

$$\Phi'(B'_1) = \Phi(f^{-1}(B'_1)), \quad \dots, \quad \Phi'(B'_n) = \Phi(f^{-1}(B'_n))$$

□

are independent.

The mapping function f does not need to be bijective. In some cases, Φ' is not locally finite, i.e., $\Lambda'(B')$ is infinite for some or all $B' \subset \mathbb{R}^s$ with $v_s(B') > 0$. For example, applying the mapping $f: \mathbb{R}^2 \mapsto \mathbb{R}$ with $f(x) = x \cdot (1, 0)$ (projection onto the first axis) to a uniform PPP, we have $\Lambda'(B') = \infty$ as soon as B' is non-empty and not atomic.

Specializing to linear mappings with $d = s$, we obtain the following corollary.

COROLLARY 2.35 (Linear mapping) *Let $\Phi \subset \mathbb{R}^d$ be a stationary PPP of intensity λ , and let $A: \mathbb{R}^d \mapsto \mathbb{R}^d$ be a non-singular linear mapping, given by a transformation matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$. Then $A(\Phi) = \{x \in \Phi: \mathbf{Ax}\}$ is a stationary PPP with intensity*

$$\lambda \det(\mathbf{A}^{-1}).$$

Example 2.9 If $\Phi = \{x_i\}$ is a homogeneous PPP of intensity λ on \mathbb{R}^2 , what is the intensity (function) $\lambda_1(x)$ of the one-dimensional PPP $\Phi_1 = \{\|x_i\|\}$? What is the intensity function $\lambda_2(x)$ of $\Phi_2 = \{\|x_i\|^2\}$?

Solution

In the first case, $f(x) = \|x\|$. Let $B' = [0, r]$. Then $f^{-1}(B') = b(o, r)$, and we have

$$\Lambda_1([0, r]) = \Lambda(b(o, r)) = \int_{b(o, r)} \lambda dx = \lambda \pi r^2,$$

from which the density follows as

$$\lambda_1(x) = \frac{d}{dx} \Lambda_1([0, x]) = 2\lambda \pi x, \quad x \geq 0.$$

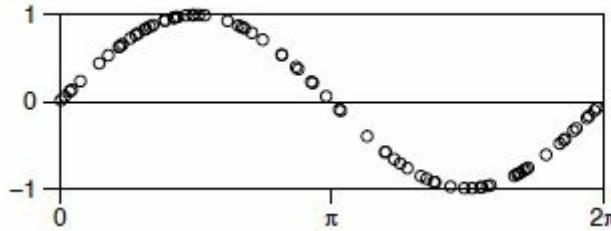


Figure 2.7 A uniform PPP $\{x_i\} \subset \mathbb{R}$ of intensity $\lambda = 2$ is transformed into the PPP $\{(x_i, \sin x_i)\} \subset \mathbb{R}^2$ with intensity measure concentrated on the one-dimensional curve $\{x \in \mathbb{R}: (x, \sin x)\}$.

Hence this is an inhomogeneous one-dimensional PPP with intensity function $\lambda_1(x) = 2\lambda \pi x \mathbf{1}(x \geq 0)$.

In the second case, $f(x) = \|x\|^2$. The pre-image of the interval $[0, r]$ is the ball $b(o, \sqrt{r})$, and we find $\Lambda_2([0, r]) = \lambda \pi r$ and $\lambda_2(x) = \lambda \pi \mathbf{1}(x \geq 0)$. Generally, for a uniform PPP on \mathbb{R}^d , $\{\|x_i\|^d\}$ is uniform on \mathbb{R}^+ with intensity $c_d \lambda$.

Example 2.10 Let $\Phi = \{x_i\} \subset \mathbb{R}$ be a uniform PPP and $\Phi' = \{(x_i, \sin x_i)\} \subset \mathbb{R}^2$. The intensity measure Λ' is concentrated on $\{x \in \mathbb{R}: (x, \sin x)\}$, i.e., the points of Φ' are located in a one-dimensional subset of \mathbb{R}^2 . An illustration is shown in Fig. 2.7, where Φ has intensity 2.

2.7.3 Thinning

Thinning is another basic and important transformation. In general, thinning is the process of removing certain points from a point process, usually according to a probabilistic rule. If the removal event is independent for all points, the thinning is called independent. If independent thinning is applied to a PPP, we have a particularly nice result.

THEOREM 2.36 (Independent thinning of a PPP) *Let $g: \mathbb{R}^d \rightarrow [0, 1]$ be a thinning function and apply it to a stationary PPP Φ by deleting each point x with probability $1 - g(x)$, independently of all other points. This thinning procedure generates an inhomogeneous PPP with intensity function $\lambda g(x)$.*

Proof Let Φ be the original process and $\tilde{\Phi}$ the process after thinning. Independence follows from the construction. The distribution of $\tilde{\Phi}(B)$ is given by

$$\mathbb{P}(\tilde{\Phi}(B) = k) = \sum_{i=k}^{\infty} \mathbb{P}(\Phi(B) = i) \mathbb{P}(\tilde{\Phi}(B) = k | \Phi(B) = i).$$

Given $\Phi(B) = i$, the i points of Φ in B are uniformly distributed. Thus the conditional probability of $\tilde{\Phi}(B) = 1$ given $\Phi(B) = 1$ is just

$$\mathbb{P}(\tilde{\Phi}(B) = 1 | \Phi(B) = 1) = |B|^{-1} \int_B g(x) dx,$$

and the complete conditional distribution is

$$\mathbb{P}(\tilde{\Phi}(B) = k | \Phi(B) = i) = \binom{i}{k} \left(|B|^{-1} \int_B g(x) dx \right)^k \left(1 - |B|^{-1} \int_B g(x) dx \right)^{i-k}.$$

Hence, with $G \triangleq \int_B g(x) dx$,

$$\begin{aligned} \mathbb{P}(\tilde{\Phi}(B) = k) &= \sum_{i=k}^{\infty} e^{-\lambda|B|} \frac{(\lambda|B|)^i}{i!} \binom{i}{k} \left(|B|^{-1} G \right)^k \left(1 - |B|^{-1} G \right)^{i-k} \\ &= e^{-\lambda|B|} \frac{(\lambda G)^k}{k!} \sum_{i=k}^{\infty} \frac{(\lambda|B|(1 - |B|^{-1} G))^{i-k}}{(i-k)!} \\ &= e^{-\lambda|B|} \frac{(\lambda G)^k}{k!} e^{\lambda|B|(1 - |B|^{-1} G)} \\ &= \frac{(\lambda G)^k}{k!} e^{-\lambda G}. \end{aligned}$$

In wireless networks, the point process of concurrent transmitters is a thinned version of the point process comprising all nodes. However, the thinning is independent only if ALOHA is used as the channel access (MAC) scheme. In fact, ALOHA may be defined as the MAC scheme that performs independent thinning of the nodes to obtain the set of transmitters. We will discuss point processes that stem from *dependent* thinning in the next chapter.

2.8 Distances

Nearest-neighbor and contact distances

We first define the nearest-neighbor operator NN_φ for a point set φ with at least two points as

$$\text{NN}_\varphi(x) \triangleq \arg \min_{y \in \varphi \setminus \{x\}} \|y - x\|, \quad x \in \varphi. \quad (2.11)$$

If the nearest neighbor is not unique, the operator picks one of the nearest neighbors uniformly at random. It may be applied to an arbitrary location $u \in \mathbb{R}^d$, in which case it is more appropriately called the nearest-point operator. Further, let $\|x - B\|$, $B \subset \mathbb{R}^d$, denote the minimum distance between x and B :

$$\|u - B\| \triangleq \min_{y \in B} \|u - y\|, \quad u \in \mathbb{R}^d, \quad B \subset \mathbb{R}^d.$$

DEFINITION 2.37 (Contact distance) The *contact distance* at location u of a point process Φ is $\|u - \Phi\|$.

DEFINITION 2.38 (Contact distribution function or empty space function) The contact distribution function or empty space function F^u is the cdf of $\|u - \Phi\|$:

$$F^u(r) \triangleq \mathbb{P}(\|u - \Phi\| \leq r) = \mathbb{P}(N(b(u, r)) > 0) = T(b(u, r)).$$

If Φ is stationary F^u does not depend on u and is just denoted by F . $F^u(r) = T(b(u, r))$ is the special case of the capacity function where balls are used for the argument of the capacity function.

Example 2.11 Let Φ be a homogeneous PPP of intensity λ on \mathbb{R}^d . $\mathbb{P}(\|u - \Phi\| \leq r) = \mathbb{P}(N(b(u, r)) > 0) = 1 - e^{-\lambda c_d r^d}$, where $c_d = |b(o, 1)|$ is the volume of the d -dimensional unit ball. So $V = c_d \|u - \Phi\|^d$ is the volume of the largest ball that fits in before hitting a point in Φ , and its distribution is $\mathbb{P}(V \leq v) = 1 - e^{-\lambda v}$.

DEFINITION 2.39 (Nearest-neighbor distance and distribution function) The *nearest-neighbor distance* is the distance from a point $x \in \Phi$ to its nearest neighbor. It is given by

$$\|x - \text{NN}_\Phi(x)\| = \|x - \Phi \setminus \{x\}\|, \quad x \in \Phi.$$

The corresponding distribution function $G^x(r) = \mathbb{P}(\|x - \text{NN}_\Phi(x)\| \leq r)$ is the *nearest-neighbor distance distribution function*.

Again if the process is stationary G^x does not depend on x and we write just G .

For the PPP, the contact distribution function and nearest-neighbor distance distribution function are identical, in the sense that

$$F^u(r) = G^u_{\Phi \cup \{u\}}(r),$$

where $G^u_{\Phi \cup \{u\}}$ is the nearest-neighbor distance distribution at u of the process $\Phi \cup \{u\}$. This follows from the independence property of the PPP: Conditioning on the PPP having a point at x does not affect the distribution of the other points, and the distribution of the two distance functions does not depend on x if the PPP is uniform. A rigorous discussion of conditioning on such events of probability 0 follows in [Chapter 8](#).

The J function

A useful measure of how close a process is to a PPP is the *J function*, which was introduced in van Lieshout & Baddeley (1996).

DEFINITION 2.40 (J function) For motion-invariant point processes, the *J function* is the ratio of the complementary nearest-neighbor distance and contact distributions:

$$J(r) \triangleq \frac{1 - G(r)}{1 - F(r)} \quad \text{for } r \geq 0.$$

For the uniform PPP, $J(r) \equiv 1$.

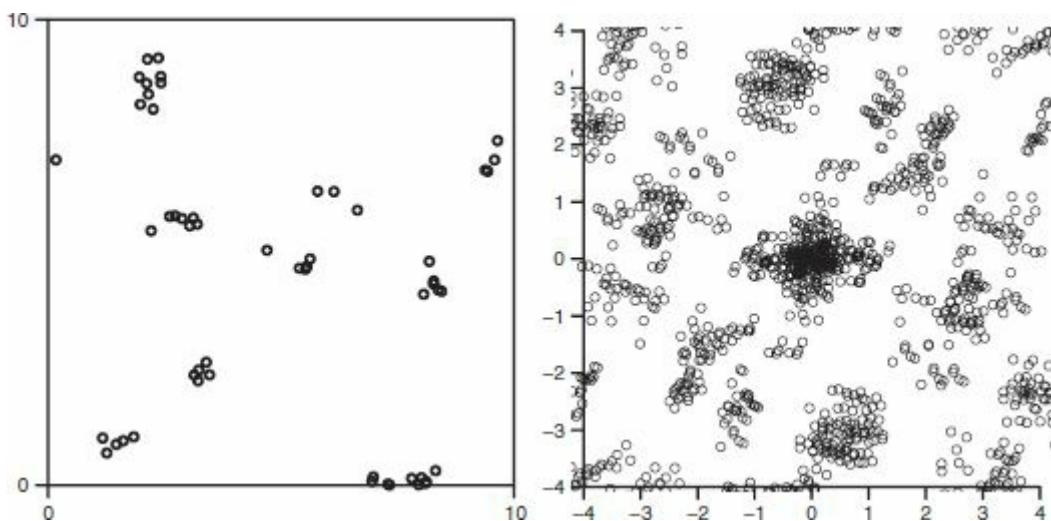


Figure 2.8 (Left) Clustered point process on $[0, 10]^2$. This realization has 55 points. (Right) ts Fry plot restricted to $[-4, 4]^2$. The entire Fry plot has $\binom{55}{2} = 1485$ points. The diameter of the high-density area near the origin is indicative of the mean diameter of a cluster.

Fry plots

The Fry plot or Patterson plot of a finite point process with n nodes is a plot of the

vectors $x_i - x_j$ for all distinct pairs of points. The Fry plot visualizes certain properties of the point process, such as whether it is clustered or more regular than a PPP. An important class of regular processes is constituted by hard-core processes, where a minimum distance between all pairs of points is imposed. They will be discussed in detail in [Section 3.5](#). To avoid boundary effects, the window through which the Fry points are observed is usually chosen smaller than the maximum difference between the points. For example, on $[0, 10]^2$, a window of $[-4, 4]^2$ may be chosen instead of the full $[-10, 10]^2$. Two examples are given in [Fig. 2.8](#) (for a clustered process) and [Fig. 2.9](#) (for a more regular process).

2.9 Applications

2.9.1 Distances in Poisson networks

For the capacity and performance analysis and comparison of protocols and algorithms for wireless networks, it is important that the distribution of the distances between the terminals be known. In the case where the nodes form a uniform PPP, the distances from a given location or a node to its n th nearest neighbor, $n \in \mathbb{N}$, can be characterized in a straightforward manner.

The main observation is that the n th nearest node is at a distance larger than r if there are at most $n - 1$ nodes in the ball of radius r around the node under consideration. Denoting the distribution of the distance to the n th nearest node by G_n and focusing on the origin, we have

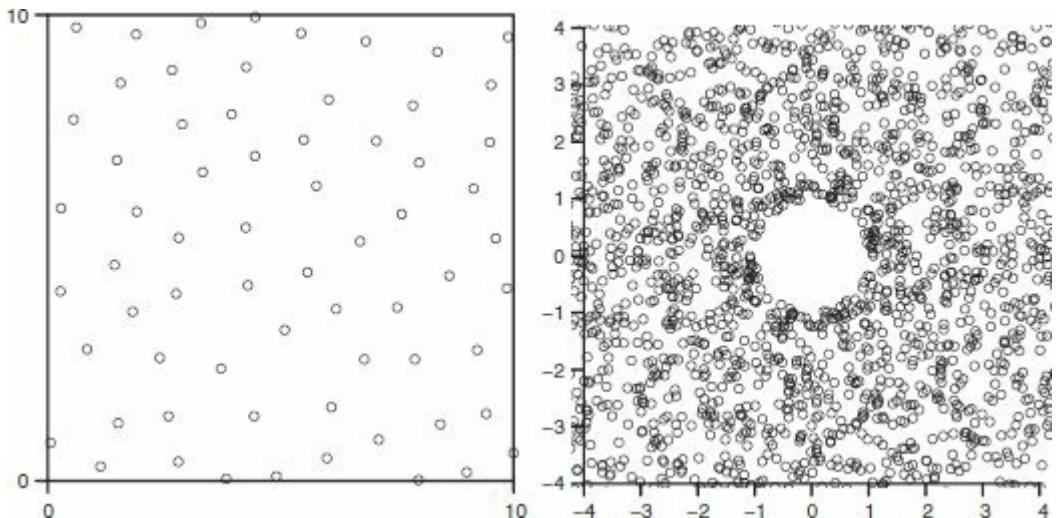


Figure 2.9 (Left) Hard-core process on $[0, 10]^2$ with minimum distance 1 between the points. (Right) Corresponding Fry plot on $[-4, 4]^2$. The hole in the center clearly indicates the hard-core distance 1.

$$\begin{aligned}
G_n(r) &= 1 - \mathbb{P}(\Phi(b(o, r)) < n) \\
&= 1 - \exp(-\lambda c_d r^d) \sum_{k=0}^{n-1} \frac{(\lambda c_d r^d)^k}{k!},
\end{aligned} \tag{2.12}$$

where c_d is given in (2.7). When taking the derivative, all terms in the sum but the one for $n - 1$ cancel out, so the pdf is the generalized gamma distribution

$$g_n(r) = \exp(-\lambda c_d r^d) \frac{d(\lambda c_d r^d)^n}{r \Gamma(n)}.$$

In two dimensions, it is

$$g_n(r) = \frac{2}{\Gamma(n)} (\lambda \pi)^n r^{2n-1} \exp(-\lambda \pi r^2).$$

2.9.2 Combining path loss and fading

Here we introduce a point process where the points represent the (power) path loss between the origin and randomly placed nodes in a wireless network. We model the path loss as a combination of a large-scale component, given by $\ell(r) = r^{-\alpha}$ for a path loss exponent α , and a fading component, which acts multiplicatively on the power.

Let the set $\{y_i\}$, $i \in \mathbb{N}$, consist of the points of a stationary Poisson point process in \mathbb{R}^2 of intensity 1 and define the path loss point process (before fading) as $\Phi = \{x_i\} \triangleq \{\|v_i\|^{\alpha}\}$. For the fading, let (h_i) be a family of iid random variables with distribution F .

Define $\Xi = \{\xi_i\} \triangleq \{x_i/h_i\}$ to be the path loss process with fading. This way, the power received at the origin if node y_i transmits at unit power is ξ_i^{-1} . Since Ξ is obtained by combining a deterministic mapping of a PPP followed by an independent displacement, both Φ and Ξ are PPPs, and our goal is to determine their intensities.

To obtain concrete results, we use the fairly general Nakagami- m (power) fading model whose distribution and density are

$$\begin{aligned}
F(x) &= 1 - \frac{\Gamma_{ic}(m, mx)}{\Gamma(m)}, \\
f(x) &= \frac{m^m x^{m-1} \exp(-mx)}{\Gamma(m)},
\end{aligned} \tag{2.13}$$

where

$$\Gamma_{ic}(n, a) \triangleq \int_a^\infty x^{n-1} e^{-x} dx$$

is the upper incomplete gamma function. This distribution is a single-parameter

version of the gamma distribution where both parameters are the same such that the mean is 1. For $m = 1$, it reduces to the exponential distribution, corresponding to Rayleigh fading.

First, we determine the intensity function of Φ using the mapping theorem. Proceeding as in Example 2.9, we find $\mathbb{E}\Phi([0, x)) = \pi x^\delta$ and thus

$$\lambda(x) = \pi \delta x^{\delta-1} \mathbf{1}(x \geq 0),$$

where $\delta = 2/\alpha$.

Next, we use the displacement theorem (Theorem 2.33) to determine the intensity function μ of Ξ . A point of Φ at x gets displaced to $\xi = x/h$, hence

$$\mathbb{P}(x/h \leq y) = 1 - F(x/h),$$

and the displacement kernel follows as

$$\rho(x, y) = \frac{d}{dy}(1 - F(x/y)) = \frac{x}{y^2} f(x/y).$$

Applying the theorem, we obtain for $y \geq 0$

$$\begin{aligned} \mu(y) &= \int_0^\infty \pi \delta x^{\delta-1} \frac{x}{y^2} f(x/y) dx \\ &= \pi \delta y^{\delta-1} \mathbb{E}(h^\delta) \\ &= \pi \delta y^{\delta-1} \frac{\Gamma(\delta + m)}{m^\delta \Gamma(m)}. \end{aligned}$$

So, remarkably, the intensities of Φ and Ξ are proportional with factor $\mathbb{E}(h^\delta)$. Whether the process with fading is denser depends on δ . Since in most cases $\alpha > 2$, we have $\delta < 1$, and $\mathbb{E}(h^\delta) < 1$ for all finite m . As $m \rightarrow \infty$, $F(x) \rightarrow 1(x \geq 1)$, i.e., there is no fading asymptotically, and $\mathbb{E}(h^\delta) \rightarrow 1$.

This result gives an indication how fading affects the connectivity of a network. If the node at the origin can successfully decode messages from neighbors with path loss (including fading) smaller than some threshold, it shows that fading generally decreases the number of such neighbors.

Bibliographical notes

The most up-to-date and authoritative reference on point process theory is the two-volume monograph by Daley & Vere-Jones (Daley & Vere-Jones 2003, 2008). Volume I describes basic point process models and discusses conditional intensities and second-order properties. It also has three appendices that review the mathematical background. Volume II sets out the basic theory of random measures and point processes in a unified setting and continues with the more theoretical

topics of the first volume: limit theorems, ergodic theory, Palm theory, and evolutionary behavior via martingales and conditional intensity. It also includes a detailed discussion of marked point processes and, in particular, of the structure of spatial point processes. Most relevant for the material covered in this chapter are [Chapters 1–4](#) in Volume I.

Another standard reference is Stoyan *et al.* (1995), which is less technical and includes point process theory as well as spatial statistics and a chapter on general random sets. Part of the material of this chapter is contained in [Chapters 1](#) and [2](#).

The book by Cressie (1993) focuses on spatial statistics but also provides a fairly detailed introduction to point process theory in Part III. A more recent book on spatial point process theory and statistics is Illian *et al.* (2008). Its applications stem from physics and ecology.

A concise discussion of Poisson point processes is provided in Kingman (1993). Basic properties are the topic of [Chapters 1](#) and [2](#). Chapter 1 of Baddeley *et al.* (2007) gives a condensed overview of much of the material covered in [Part I](#) of this book, while the other chapters discuss more specific topics in stochastic geometry.

In contrast, Mathai (1999) focuses on more general random objects such as lines and triangles.

The J function in Definition [2.40](#), which was introduced in van Lieshout & Baddeley (1996), was extended to inhomogeneous point processes in van Lieshout (2011). A natural question is whether $J(r) \equiv 1$ implies that the process is a PPP. This conjecture has been disproved in Bedford & van den Berg (1997), where a one-dimensional counter-example was carefully crafted.

The idea of combining large-scale path loss and fading in wireless networks into a new point process was proposed in Haenggi (2008a). Applications include connectivity and broadcast capacity.

Problems

2.1 For the uniform PPP on \mathbb{R} with intensity λ , derive the pdf of the interarrival interval containing the origin o . What do you observe?

2.2 Let X_m be a Poisson random variable with mean m . If M is a gamma distributed random variable, show that the distribution of X_M is negative binomial.

2.3 Let Φ be a uniform PPP of intensity λ on \mathbb{R} . Determine $\mathbb{P}(\Phi(B_1) = n_1, \Phi(B_2) = n_2)$ for $B_1 = [0, 2]$ and $B_2 = [1, 3]$.

2.4 Let (X_i) , $i \in \mathbb{N}$, be a family of independent exponential random variables with mean $1/\lambda$. Determine

$$\mathbb{P}(X_1 > t) + \sum_{k=1}^{\infty} \mathbb{P}(X_1 + \dots + X_k \leq s; X_1 + \dots + X_{k+1} > t)$$

and give an interpretation of the result as the void probability of a certain point process.

2.5 Can the outcome space \mathcal{N} be defined as

$$\mathcal{N} \triangleq \{\varphi \in \mathbb{R}^d : |\varphi| = 0\}?$$

2.6 Radial simulation. Take a one-dimensional PPP $\Psi = \{y_i\} \subset \mathbb{R}^+$ of intensity 1 on the positive real line. Assume the points are ordered, i.e., $y_1 < y_2 < \dots$. Given Ψ , we would like to produce a realization of a homogeneous PPP $\Phi = \{(r_i, \phi_i)\}$ of intensity λ on \mathbb{R}^2 , where (r_i, ϕ_i) is the polar representation of a point. The magnitude set $\{r_i\}$ should also be ordered. Find the function from y_i to r_i that maps the points in one dimension to the magnitudes in two dimensions with the desired intensity.

2.7 Generating function. Let X_1, X_2, \dots be a sequence of iid random variables $\Omega \rightarrow \mathbb{N}_0$ with generating function G_X . Let N be a random variable, independent of the X_i , also taking values in \mathbb{N}_0 . Show that the sum

$$S = X_1 + X_2 + \dots + X_N$$

has the generating function $G_S(s) = G_N(G_X(s))$ and derive $G_S(s)$ if N is Poisson with mean λ .

2.8 Find a representation analogous to the one in (2.3) for the counting measure of the two-dimensional randomly translated lattice.

2.9 Determine $\Lambda(\{x\})$ for all $x \in \mathbb{R}^d$ for the die point process defined in Example 2.1.

2.10 Write out explicitly the distribution of the die point process.

2.11 The point process in Example 2.3 is the superposition of two (even more) elementary point processes. Find them and verify that $P = P_1 * P_2$ using the standard rules for convolution.

2.12 Consider the Bernoulli die process, defined as follows: Start with the original die process, but then remove each point with probability q independently of the other points (if any). Such a point process could also be called a die process with erasures, with erasure probability q . Determine $\Lambda(\{o\})$ and $\Lambda(\mathbb{R}^2)$.

2.13 Let $s\Phi_{\text{die}}$ be the die process Φ_{die} scaled by s . Let

$$\Phi' = \bigcup \left\{ x \in \mathbb{Z}^2 : x + \frac{1}{3}\Phi_{\text{die}} \right\},$$

where $x + \frac{1}{3}\Phi_{\text{die}}$ is a translation of the (scaled) die process by x , and consider the stationary point process $\Phi \triangleq \Phi'|_Y$, where Y is uniform on $[0, 1]^2$. Find the intensity of Φ .

2.14 Let $\Phi \subset \mathbb{R}$ be a PPP with intensity function $\lambda(x) = \mathbf{1}(x \in [0, 1])$. What is the

intensity of the PPP that results from displacing each point independently by a random amount uniform on $[-1/2, 1/2]$?

2.15 Let $\Phi = \{x_1, \dots\} \subset \mathbb{R}$ be a PPP with intensity function $\lambda(x) = \mathbf{1}(0 \leq x < 2\pi)$. Let $\Phi' = \{\sin x_i\}$. Determine the intensity function λ' of Φ' .

2.16 Let Φ be a Gaussian Poisson process with $\sigma = 1$ and $n = 1$, and let B^- be the left half plane of \mathbb{R}^2 (all points with negative first coordinate). Determine $\Lambda(B^-)$ for Φ and for $\Phi_{(1, 0)}$. *Hint.* Relate the second part of the problem to a bit error probability problem in point-to-point communication.

2.17 Figure 2.2 shows a realization of a certain non-homogeneous PPP. Determine the intensity measure $\Lambda([0, 10]^2)$ for this process.

2.18 Is the randomly translated lattice (see Definition 2.15) ergodic?

2.19 Let Φ be a homogeneous PPP in \mathbb{R}^2 with intensity 1. Calculate the fraction of points for which $\text{NN}(\text{NN}(x)) = x$. These are the nodes belonging to a mutual-nearest-neighbor pair. How does the fraction depend on the intensity? For comparison, determine the fraction for a lattice.

2.20 Let Φ denote the randomly translated square integer lattice.

(a) Show that the intensity $\lambda \equiv 1$.

(b) Let B_\diamond be a square of area A with $1 < A < 2$, rotated by $\pi/4$ such that opposing corner points share one of their coordinates. Find $\mathbb{P}(\Phi(B_\diamond) = 0)$ and $\mathbb{P}(\Phi(B_\diamond) > 2)$ as a function of A .

2.21 We would like to produce realizations of an isotropic PPP $\Phi \subset \mathbb{R}^2$ with Gaussian intensity

$$\lambda(x) = \frac{M}{2\pi\sigma^2} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right), \quad x \in \mathbb{R}^2.$$

Let the simulation area be $W = [-5, 5]^2$.

- (a) Describe two ways of doing this, one by mapping and one by thinning. Implement them in R or MATLAB and include two plots in the homework for $\sigma^2 = 4$ and $M = 100$.
- (b) How many points do you find inside W on average, averaged over many runs of your program? How many do you expect to find analytically?
- (c) After having obtained the points, how could you verify that they do indeed stem from a PPP with the given intensity?

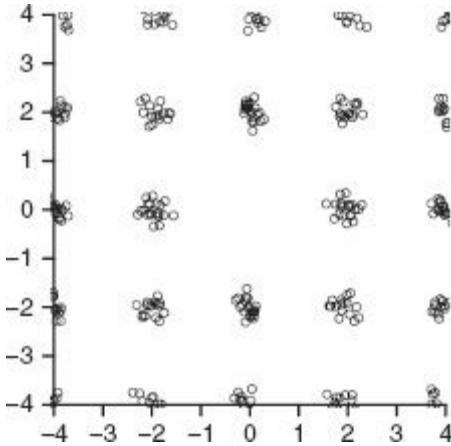


Figure 2.10 Fry plot for Problem 2.23.

2.22 Directional empty space function. Let $\Phi \subset \mathbb{R}^2$ be a PPP of intensity 1. Let $\{(r_i, \phi_i)\}$ be the polar representation of the point set. Let $n \in \mathbb{N}$ and define

$$R_k = \min \left\{ (r_i, \phi_i) \in \Phi, k \frac{2\pi}{n} < \phi_i \leq (k+1) \frac{2\pi}{n}; r_i \right\}, \quad k = 0, 1, \dots, n-1.$$

R_k is the distance of the nearest node to the origin in the k th sector of width $2\pi/n$.

We are interested in the cdf of the minimum of the R_k . Without calculation, what do you expect this cdf to be?

To verify find the distribution function of the minimum of the k random variables R_k using the result from Problem 1.2.

2.23 Describe the point process pertaining to the Fry plot in Fig. 2.10.

¹ In principle, point processes could be viewed also as *multisets*, which are sets where elements are allowed to appear more than once.

² We use the term *location* instead of point when talking about a general element in \mathbb{R}^d , to avoid confusion with the point of a point process.

³ The support of a function $f: \mathbb{R}^d \mapsto \mathbb{R}^+$ is defined as the set $\{x \in \mathbb{R}^d : f(x) > 0\}$.

⁴ Not to be confused with the Gauss–Poisson process, which will be introduced in Example 3.8.

⁵ The distance that turns \mathbb{N} into a (completely separable) metric space is based on the Prokhorov distance, which is a metric for weak convergence (Daley & Vere-Jones 2003, Appendix 2).

3 Point process models

3.1 Introduction

In this chapter, we introduce four additional point process models: cluster processes, hard-core processes, Cox processes, and Gibbs processes. Poisson point processes exhibit *complete spatial randomness* due to their independence property. Cox processes model less regular spatial distributions – they are *overdispersed* relative to PPPs, which means that the ratio of the variance of the number of nodes in a set to its mean is larger than 1. Gibbs processes, on the other hand, may be overdispersed or underdispersed.

Figure 3.1 illustrates the two directions along which a point process model may depart from the point of complete spatial randomness, the PPP. In terms of the J function introduced in Definition 2.40, regular processes have J values larger than 1, since the probability of having a nearby neighbor is small; conversely, clustered processes have J values smaller than 1.

3.2 General finite point processes

A general finite point process is a generalization of the binomial point process to a process where the total number of points is itself a random variable. Compared with a random vector, there are three differences: The randomness of the number of points and the facts that the point process is unordered and simple and thus is better represented as a set.

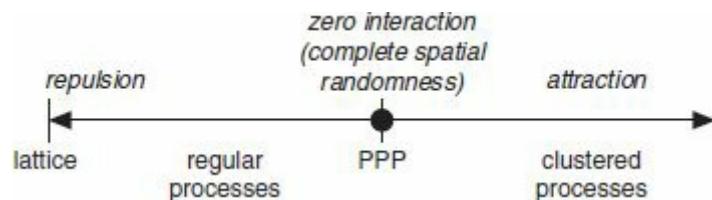


Figure 3.1 Point process taxonomy: The PPP is the point process with zero interaction between the points. On moving to the left on this axis, points start to repel each other, which leads to processes that are more regular. In the extreme case, a lattice is obtained. On

moving to the right, points attract each other, resulting in clustered processes.

DEFINITION 3.1 (General finite point process) A finite point process is defined using two distributions. First, a discrete distribution (p_k) , $k \in \mathbb{N}_0$, on the number of points n , with $\sum_{k \in \mathbb{N}_0} p_k = 1$. Second, for each $n \in \mathbb{N}$, a joint distribution $F_n(\cdot)$ given on the Borel sets of \mathbb{R}^{nd} ,

$$F_n(B_1 \times \cdots \times B_n) = \mathbb{P}(x_1 \in B_1, \dots, x_n \in B_n), \quad (3.1)$$

that determines the position of the points given that their total number is n .

This is a constructive definition in the sense that it provides a direct recipe for the simulation of the point process: First, draw the number n from (p_k) , then place the n points according to F_n . This is also the way PPPs are simulated.

Since the points are unordered, we stipulate that F_n should give equal weight to all $n!$ permutations of the coordinates, i.e., that it is *symmetric*. If the original distribution is not already symmetric, we can easily introduce a symmetric version by setting for any partition (B_1, \dots, B_n) of \mathbb{R}^d

$$F_n^{\text{sym}}(B_1 \times \cdots \times B_n) \triangleq \frac{1}{n!} \sum_{\text{perm}} F_n(B_{i_1} \times \cdots \times B_{i_n}),$$

where the sum is taken over all $n!$ permutations (i_1, \dots, i_n) of the integers $[n]$ and the normalizing factor $1/n!$ ensures that the resulting measure still has total mass unity, i.e., that $F_n^{\text{sym}}(\mathbb{R}^{nd}) = 1$ for all n .

The joint probability density pertaining to F_n is denoted by f_n .

Example 3.1 A simple example of a finite point process is the (uniform) binomial point process on $W \subset \mathbb{R}^d$ with n nodes, for fixed n . Hence $p_k = 1(k = n)$, and the points are distributed with probability density

$$f_n(x_1, \dots, x_n) = \left(\frac{1}{|W|} \right)^n, \quad (x_1, \dots, x_n) \in W^n.$$

This density is permutation-invariant.

Example 3.2 For the die process from Example 2.1, the situation is reversed. Here the number of points is random, while the distributions F_n , $n \in [6]$, are degenerate (deterministic). The first two are $F_1(B) = \delta_o(B)$ and $F_2(B_1 \times B_2) = \delta_{(-1,1)}(B_1) \cdot \delta_{(1,-1)}(B_2)$. These are not symmetric for $n > 1$. The symmetric version is $F_2^{\text{sym}}(B_1 \times B_2) = \frac{1}{2} (\delta_{(-1,1)}(B_1) \cdot \delta_{(1,-1)}(B_2) + \delta_{(1,-1)}(B_2) \cdot \delta_{(-1,1)}(B_1))$.

If the number of points of a BPP is not fixed, the process is called a *mixed binomial point process*, since now its intensity is a random variable – analogously to the mixed PPP. Hence a general finite point process is a mixed BPP if, given the number of points n , the joint distribution (3.1) factorizes to $\prod_{i=1}^n \mathbb{P}(x_i \in B_i)$. A finite PPP is a special case of a mixed BPP.

Related to the joint distribution F is the *Janossy measure*, defined as

$$J_n(B_1 \times \cdots \times B_n) \triangleq n! p_n F_n^{\text{sym}}(B_1 \times \cdots \times B_n), \quad n \in \mathbb{N}_0. \quad (3.2)$$

It follows that $J_0 = p_0$. The Janossy measure is not a probability measure but it sometimes offers advantages in combinatorial formulations. If it admits a density j_n with respect to the Lebesgue measure, the density is called the *Janossy density*. For $x_i \neq x_j$ for $i \neq j$,

$$j_n(x_1, \dots, x_n) dx_1 \dots dx_n$$

is the probability that there are exactly n points in the process, one in each of the n distinct infinitesimal regions dx_i located at x_i .

Example 3.3 The finite PPP with $\Lambda(\mathbb{R}^d) < \infty$ is defined by (p_k) being the Poisson distribution with mean $\Lambda(\mathbb{R}^d)$ and, given n ,

$$F_n(B_1 \times \cdots \times B_n) = \frac{1}{\Lambda(\mathbb{R}^d)^n} \prod_{k \in [n]} \Lambda(B_k),$$

which is already in symmetric form. If the distribution of each point is denoted by F ,

$$F_n(B_1 \times \cdots \times B_n) = \prod_{k \in [n]} F(B_k).$$

For the corresponding densities, given by $F(B) = \int_B f(x) dx$, we have

$$f_n(x_1, \dots, x_n) = f(x_1) \dots f(x_n),$$

which is already in symmetric form. The Janossy density is

$$j_n(x_1, \dots, x_n) = p_n n! f(x_1) \dots f(x_n).$$

This last expression holds more generally for mixed BPPs (finite point processes with iid located nodes).

Example 3.4 Consider the point process consisting of the first n points on the positive side of o of a uniform PPP on \mathbb{R} of intensity λ . The joint density is given in (2.2) as

$$f_n(t_1, \dots, t_n) = f_{(x_1, \dots, x_n)}(t_1, \dots, t_n) = \lambda^n e^{-\lambda t_n} \mathbf{1}((t_1, \dots, t_n) \in \mathbb{R}^{+n}).$$

These points are ordered. To obtain a symmetric pdf, the probability mass needs to be spread over all $n!$ hyperoctants. For $n = 2$,

$$f_n^{\text{sym}}(t_1, t_2) = \begin{cases} \frac{1}{2}\lambda^2 e^{-\lambda t_2} & \text{if } t_2 > t_1 \\ \frac{1}{2}\lambda^2 e^{-\lambda t_1} & \text{if } t_2 \leq t_1. \end{cases}$$

In general, denoting by $t_{(n)}$ the maximum of the n locations,

$$f_n^{\text{sym}}(t_1, \dots, t_n) = \frac{1}{n!} \lambda^n e^{-\lambda t_{(n)}}.$$

So the joint distribution depends only on the extreme value $t_{(n)}$. Since $p_n = 1$ (there are always n points), the Janossy density follows as

$$j_n(t_1, \dots, t_n) = \lambda^n e^{-\lambda t_{(n)}}.$$

If u is a bounded or non-negative functional on a general finite point process Φ , we can express the mean $\mathbb{E}(u(\Phi))$ as

$$\mathbb{E}(u(\Phi)) = \sum_{n=0}^{\infty} p_n \int_{\mathbb{R}^{nd}} u(\delta_{x_1} + \dots + \delta_{x_n}) F_n(dx_1 \times \dots \times dx_n)$$

or, using the Janossy measure (see (3.2)), as

$$\mathbb{E}(u(\Phi)) = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\mathbb{R}^{nd}} u(\delta_{x_1} + \dots + \delta_{x_n}) J_n(dx_1 \times \dots \times dx_n). \quad (3.3)$$

3.3 Cox processes

A generalization of the Poisson process is to allow for the intensity measure itself being random. The resulting process is then Poisson conditional on the intensity measure. It is called a *doubly stochastic Poisson process* or *Cox process*.

The intensity measure is a realization of a non-negative locally finite random measure, defined as follows.

DEFINITION 3.2 (Non-negative locally finite random measure) A non-negative locally finite measure is a measure $\psi: \mathcal{B}^d \mapsto \mathbb{R}^+$ such that

$$0 \leq \psi(B) < \infty \text{ for compact } B.$$

Let \mathcal{M} be the set of all such measures, and \mathfrak{m} be the smallest σ -algebra on \mathcal{M} that makes $\psi(B)$ measurable for all compact B .

A random non-negative locally finite measure Ψ is a random element that maps $(\Omega, \mathcal{A}, \mathbb{P})$ to the space $(\mathcal{M}, \mathfrak{m})$.

Remark 3.1 The set \mathcal{M} of non-negative locally finite measures is a superset of the set \mathcal{N} of locally finite counting measures. Counting measures are \mathbb{N}_0 -valued, while measures in \mathcal{M} are \mathbb{R}^+ -valued. So every point process Φ (or N) can also be interpreted as an element of \mathcal{M} – it is, in fact, a random (non-negative locally finite) measure. If non-simple point processes are included, then the mapping between point processes and \mathbb{N}_0 -valued locally finite measures is one-to-one.

Now we are in a position to define the Cox process.

DEFINITION 3.3 (Cox process) Let Q be a distribution on $(\mathcal{M}, \mathfrak{m})$, P_Λ the distribution of the Poisson point process with intensity measure Λ , and Ψ a random measure with distribution Q . Then the point process distribution

$$P(Y) = \int_{\mathcal{M}} P_\Lambda Q(d\Lambda), \quad \text{for } Y \in \mathcal{N}, \tag{3.4}$$

defines a *Cox process* Φ with driving measure Ψ . To ensure that Φ is simple, it is assumed that Q is concentrated on the subset of \mathcal{M} of diffuse measures, so that all realizations Λ are diffuse a.s.

Thus conditioned on Ψ , the Cox process is a PPP with intensity measure $\Lambda = \Psi$. From (3.4) follows that for disjoint B_1, \dots, B_n ,

$$\begin{aligned} P\left(\bigcap_{i=1}^n \{N(B_i) = k_i\}\right) &= \mathbb{E}\left(\prod_{i=1}^n \frac{\Psi(B_i)^{k_i}}{k_i!} \exp(-\Psi(B_i))\right) \\ &= \int_{\mathcal{M}} \prod_{i=1}^n \frac{\Lambda(B_i)^{k_i}}{k_i!} \exp(-\Lambda(B_i)) Q(d\Lambda). \end{aligned}$$

Using the continuous version of the law of total probability, this can also be written as

$$P\left(\bigcap_{i=1}^n \{N(B_i) = k_i\}\right) = \int_{\mathcal{M}} P\left(\bigcap_{i=1}^n \{N(B_i) = k_i\} \mid \Psi = \Lambda\right) Q(d\Lambda),$$

where

$$\mathbb{P} \left(\bigcap_{i=1}^n \{N(B_i) = k_i\} \mid \Psi = \Lambda \right) = \prod_{i=1}^n \frac{\Lambda(B_i)^{k_i}}{k_i!} \exp(-\Lambda(B_i)).$$

Specializing to the void probabilities, we obtain

$$\mathbb{P}(\Phi(B) = 0) = \mathbb{E}(\exp(-\Psi(B))) = \int_{\mathcal{M}} \exp(-\Lambda(B)) Q(d\Lambda).$$

Other statistical properties of the Cox process are obtained similarly by averaging over the distribution Q of the intensity measure. For all Cox processes, the number of points in bounded sets is *super-Poissonian*, which means that their variances always exceeds those for stationary PPPs of the same intensity¹. This is intuitive since the varying intensity increases the randomness.

Often, the construction of a Cox process is based on a random field $\zeta(x)$, $x \in \mathbb{R}^d$, called the *intensity field* of the Cox process. The random measure Ψ is then defined using the concatenation with the Lebesgue measure ν_d :

$$\Psi(B) \triangleq \zeta \circ \nu_d(B) = \int_B \zeta(x) dx \quad \text{for Borel } B.$$

This random measure is diffuse, since the Lebesgue measure is diffuse. The realizations of ζ are measurable and non-negative functions, such that $\Psi(B) < \infty$ a.s. for compact B . If $\zeta(x)$ is stationary, i.e., if

$$\zeta(x) \stackrel{d}{=} \zeta(x + y) \quad \text{for all } x, y \in \mathbb{R}^d$$

then $\Psi(B)$ is stationary, and $\mathbb{E}\Psi(B) = \Lambda(B) = |B|\mathbb{E}\zeta(o)$ if $\mathbb{E}\zeta(o) > 0$.

Example 3.5 The simplest example of a Cox process is the mixed Poisson point process introduced in [Section 2.4.5](#). Here the driving random measure Ψ is the random measure $\Psi = X\nu_d$, where the non-negative random variable X denotes the intensity. This is a typical example of a non-ergodic Cox process. The intensity is $\lambda = \mathbb{E}X$, and

$$\text{var } \Phi(B) = |B|\mathbb{E}X + |B|^2 \text{var } X$$

since

$$\begin{aligned} \text{var } \Phi(B) &= \mathbb{E}(\Phi(B)^2) - (\mathbb{E}\Phi(B))^2 \\ &\stackrel{(a)}{=} \int_{\mathbb{R}} (x|B| + (x|B|)^2) dF_X(x) - \left(\int_{\mathbb{R}} x|B| dF_X(x) \right)^2 \\ &= |B|\mathbb{E}X + |B|^2 \mathbb{E}(X^2) - (|B|\mathbb{E}X)^2, \end{aligned}$$

where (a) follows since $\Phi(B)$ is conditionally Poisson with mean $X|B|$. This confirms that this Cox process is super-Poissonian.

If X has a two-point distribution with $\mathbb{P}(X = \lambda_1) = 1-p$ and $\mathbb{P}(X = \lambda_2) = p$, $0 < p < 1$, then the intensity of Φ is

$$\lambda = (1-p)\lambda_1 + p\lambda_2,$$

and the nearest-neighbor distance distribution follows as

$$G(r) = \frac{1}{\lambda}(\lambda_1(1-p)(1 - e^{-\lambda_1 c_d r^d}) + \lambda_2 p(1 - e^{-\lambda_2 c_d r^d})).$$

Example 3.6 Random thinning. Take a basic PPP Φ_b with intensity measure Λ_b and intensity function $\lambda_b(x)$. A point from Φ_b is retained with probability $p(x)$, where $0 \leq p(x) \leq 1$ for $x \in \mathbb{R}^d$. If $p(x)$ is deterministic, the resulting process is simply a PPP with intensity function $\lambda_b(x)p(x)$. If $p(x)$ itself is random, drawn from a random field $T(x)$, a realization of φ of Φ is obtained by taking a realization φ_b of Φ_b and applying $p(x)$ -thinning to φ_b , where $p(x)$ is a realization of the random field. So given $T(x) = p(x)$ and given $\Phi_b = \varphi_b$, the probability that a point $x \in \Phi_b$ also belongs to Φ is $p(x)$. This process is a Cox process with driving random measure

$$\Psi(B) = \int_B T(x)\Lambda_b(dx).$$

Example 3.7 Poisson hole process. Let $\Phi_1, \Phi_2 \subset \mathbb{R}^2$ be independent uniform PPPs of intensities λ_1 and λ_2 , respectively. Further let

$$\Xi_r \triangleq \bigcup\{x \in \Phi_1 : b(x, r)\}$$

be the union of all disks of radius r centered at a point of Φ_1 . The random set Ξ_r is a *germ-grain model* (see [Section 13.2.1](#)). The Poisson hole process is

$$\Phi = \{x \in \Phi_2 : x \notin \Xi_r\} = \Phi_2 \setminus \Xi_r,$$

i.e., each point in Φ_1 carves out a hole of radius r from Φ_2 . The Poisson hole process is a Cox process with driving measure

$$\Psi = \lambda_2 v_d(\cdot \setminus \Xi_r).$$

This Cox process has applications in cognitive networks, where Φ_1 are the primary users and Φ_2 are the secondary users ([Lee & Haenggi 2012](#)). The process Φ can be interpreted as the secondary users that are allowed to transmit since they are at least at distance r from all primary users and therefore do not cause harmful interference. What is the intensity of Φ ? Since the probability of a point being

retained is the probability that it has no point of the PPP of intensity λ_1 within distance r , it is

$$\lambda = \lambda_2 \exp(-\lambda_1 \pi r^2).$$

Using the intensity field formulation, the Poisson hole process is obtained by setting $\zeta(x) = \lambda_2(1 - 1_{\Xi_r}(x))$. In three dimensions, the Poisson hole process is sometimes referred to as a “Swiss cheese model.” If instead one considers the complementary process $\Phi = \Phi_2 \cap \Xi_r$, which consists of only those points of Φ_2 that are within distance r of a point of Φ_1 , one may speak of an “inner-city model.”

3.4 Cluster processes

3.4.1 Definition

A general cluster process is generated by taking a parent point process and daughter point processes, one per parent, and translating the daughter processes to the position of their parent. The cluster process is then the union of all the daughter points. Denote the parent point process by $\Phi_p = \{x_1, x_2, \dots\}$, and let $n = \#\Phi_p \in \mathbb{N} \cup \{\infty\}$ be the number of parent points. Further, let (Φ_i) , $i \in \mathbb{N}$, be a family of finite points sets, the untranslated clusters or daughter processes. The cluster process is then the union of the translated clusters:

$$\Phi \triangleq \bigcup_{i \in [n]} \Phi_i + x_i.$$

In terms of random counting measures, letting (N_i) be the family of counting measures for the clusters, the cluster process is given by

$$N \triangleq \sum_{i \in [n]} N_i + x_i.$$

Alternatively, the counting measure may be expressed using the *cluster field* $N_c(\cdot | y)$ for $y \in \mathbb{R}^d$, which is sampled at the points of the parent process, i.e.,

$$N(A) = \int_{\mathbb{R}^d} N_c(A | y) \Phi_p(dy) = \sum_{y \in \Phi_p} N_c(A | y). \quad (3.5)$$

In this representation, $N_c(\cdot | y_i)$, which is indexed by the cluster centers or parent points, assumes the role of Φ_i , which is indexed by the cardinal i .

If the parent process is a lattice, the process is called a *lattice cluster process*. Analogously, if the parent process is a PPP, the resulting process is a *Poisson cluster process*.

3.4.2 Homogeneous independent cluster processes

If the parent process is stationary and the daughter processes are independent (among themselves and also from the parent process) finite point sets and have the same distribution, the procedure is called *homogeneous independent clustering*. In this case, only the statistics of one cluster need to be specified, which is usually done by referring to the *representative cluster*, denoted by N_0 or Φ_0 , with intensity measure Λ_0 . The intensity measure of the homogeneous independent cluster process follows as

$$\Lambda(B) = \int_{\mathbb{R}^d} \Lambda_0(B - x) \Lambda_p(dx). \quad (3.6)$$

If the integral is finite for all compact B , the cluster process is locally finite a.s. If the parent process is stationary with intensity λ_p , the cluster process is stationary also, with intensity

$$\lambda = \lambda_p \Lambda_0(\mathbb{R}^d).$$

This follows from $\lambda = \Lambda([0, 1]^d)$ and specializing (3.6) to $B = [0, 1]^d$.

Example 3.8 The planar Gauss–Poisson process was introduced in Newman (1970). It is a Poisson cluster process on \mathbb{R}^2 with homogeneous independent clustering. Each cluster has zero, one, or two points, with probabilities p_0 , p_1 , and $p_2 = 1 - p_0 - p_1$, respectively. In its standard form, the location of the points is given as follows. If a cluster consists of one point, that point is at the parent's location. If it has two points, the two points are separated by unit distance and have the parent's location as their midpoint (or, equivalently, as the location of one of the two points). The orientation of the axis of the two points is uniformly chosen, so that the resulting process is isotropic (motion-invariant, in fact). A realization is shown in Fig. 3.2. More generally, the distribution of the two points in the two-point clusters can be described by general density functions. In any case, the mean number of points per cluster is $\bar{c} = p_1 + 2p_2$.

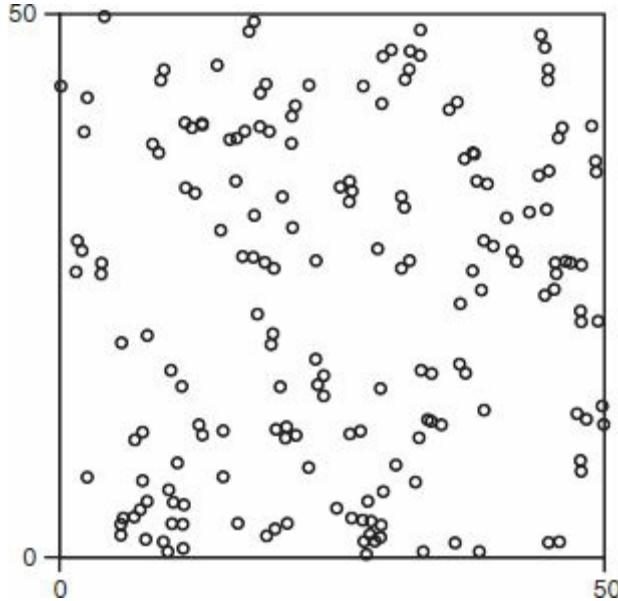


Figure 3.2 A standard Gauss–Poisson process with parent intensity $\lambda_p = 0.05$, $p_0 = 0$, $p_1 = 0.6$, and $p_2 = 0.4$ on $[0, 50]^2$. The expected number of points is $50^2 \lambda(p_1 + 2p_2) = 175$. This realization has 181 points.

3.4.3 Neyman–Scott processes

If the daughter points of a Poisson cluster process are iid (but random in number), the process is called a *Neyman–Scott process* (Neyman & Scott 1952).

DEFINITION 3.4 (Neyman–Scott cluster process) A Neyman–Scott (cluster) process is a cluster process where the parents form a uniform PPP of intensity λ_p and the daughter points in each cluster are random in number, independent of each other, and identically distributed.

Expressed differently, it is a cluster process where iid mixed BPPs are translated to the points of a uniform PPP. It is always stationary, and, if the distribution of the clusters is isotropic, it is motion-invariant. If \bar{c} is the mean number of daughter points per parent, i.e., $\bar{c} = \mathbb{E}(\#\Phi_0) = \Lambda_0(\mathbb{R}^d)$, its intensity is

$$\lambda = \bar{c}\lambda_p.$$

This holds also if Φ_p is not Poisson but a general stationary process.

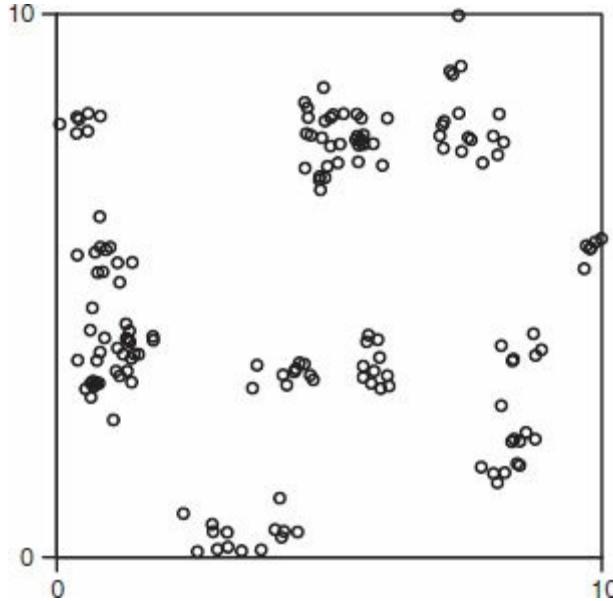


Figure 3.3 Thomas process with $\lambda_p = 1/4$, $\sigma = 0.3$, and $\bar{c} = 8$. The density is $\lambda = \lambda_p \bar{c} = 2$. This realization has 157 points.

If each cluster is itself a (finite) PPP, the resulting process is also called a *doubly Poisson cluster process* or *doubly Poisson point process*, where each cluster has intensity measure Λ_0 and mean number of points $\bar{c} = \Lambda_0(\mathbb{R}^d)$. Such processes are, in fact, Cox processes with driving random measure

$$\Psi(B) = \sum_{x \in \Phi_p} \Lambda_0(B - x),$$

where Φ_p is the parent PPP. Given that there are n points in a cluster, the points (before translation) are iid with pdf $\lambda_0(x) / \bar{c}$.

Two important concrete Neyman-Scott processes (in fact, doubly Poisson point processes) are the *Thomas cluster process* and the *Matérn cluster process*.

DEFINITION 3.5 (Thomas cluster process) The Thomas process is a doubly Poisson cluster process, where the intensity function of a cluster is given by

$$\lambda_0(x) = \frac{\bar{c}}{(2\pi)^{d/2}\sigma^d} \exp\left(-\frac{x^T x}{2\sigma^2}\right), \quad (3.7)$$

i.e., the daughter points are normally scattered with variance σ^2 around each parent point, and the mean number of daughter points is \bar{c} . The intensity of the process is $\lambda = \lambda_p \bar{c}$.

In two dimensions, each cluster is a Gaussian Poisson process as introduced in Example 2.2. An example of a Thomas process is shown in Fig. 3.3.

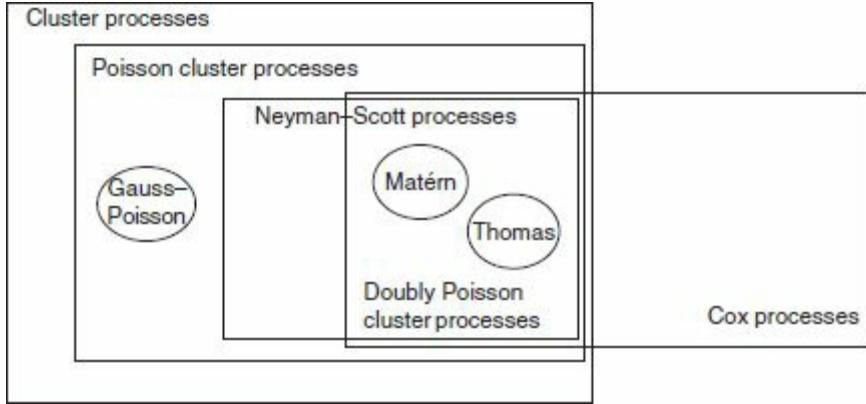


Figure 3.4 Relationship between different cluster processes. If a Neyman–Scott process is also a Cox process, then it is a doubly Poisson cluster process.

DEFINITION 3.6 (Matérn cluster process) The Matérn cluster process is a doubly Poisson cluster process, where

$$\lambda_0(x) = \frac{\bar{c}}{c_d r^d} \mathbf{1}_{b(o,r)}(x), \quad (3.8)$$

i.e., the daughter points are uniformly scattered on the ball of radius r centered at each parent point, and the mean number of daughter points is \bar{c} . The intensity of the process is $\lambda = \lambda_p \bar{c}$.

The driving random measure for the Matérn cluster process, viewed as a Cox process, is

$$\Psi(B) = \bar{c} \sum_{x \in \Phi_p} \nu_d(B \cap b(x, r)).$$

On the other hand, the Gauss–Poisson process from Example 3.8 is not a Cox process, since the number of points per cluster is not Poisson distributed. It is not even a Neyman–Scott process, since the clusters do not form mixed BPPs, i.e., a general finite point process as per Definition 3.1 is needed to describe a cluster. Figure 3.4 shows a Venn diagram that illustrates how the different classes of cluster processes are related.

3.5 Hard-core processes

Hard-core processes are point processes where points are forbidden to be closer than a certain minimum distance. In general, they are more regular (less clustered) than the PPP. One way to achieve such a minimum distance between points is to start with a point process that has no such restriction and then remove points that violate the condition.

Matérn hard-core processes

Two such models are the Matérn hard-core processes of type I and II, introduced in Matérn (1986).

DEFINITION 3.7 (Matern hard-core process of type I) Starting with a basic uniform PPP Φ_b with intensity λ_b , first flag for removal all points that have a neighbor within distance r . Then remove all flagged points.

DEFINITION 3.8 (Matérn hard-core process of type II) Starting with a basic uniform PPP Φ_b with intensity λ_b , add to each point x an independent random variable $m(x)$, called a mark, uniformly distributed on $[0, 1]$. Flag for removal all points that have a neighbor within distance r that has a smaller mark. Then remove all flagged points. Formally,

$$\Phi \triangleq \{x \in \Phi_b : m(x) < m(y) \text{ for all } y \in \Phi_b \cap b(x, r) \setminus \{x\}\}.$$

In both types of hard-core processes, all points are removed simultaneously, so even points that are thinned out can eliminate other points. A denser packing could be achieved for both types if points were removed one by one and the conditions re-checked after each removal. Such a process is sometimes referred to as a Matérn process of type III, but it is significantly less tractable than processes of the other two types.

The intensity of the type I process is $\lambda = \lambda_b \exp(-\lambda_b c_d r^d)$. Given r (and d), it is maximized at $\lambda_b = 1/(c_d r^d)$, so the highest density that can be achieved is $\lambda_{\max} = (e c_d r^d)^{-1}$.

To determine the intensity of the type II process, we first condition on a point having a given mark t . This point is retained with probability $\exp(-t \lambda_b c_d r^d)$, since $t \lambda_b$ is the density of points with marks smaller than t . Unconditioning on t , we obtain

$$\lambda = \lambda_b \int_0^1 \exp(-t \lambda_b c_d r^d) dt = \frac{1 - \exp(-\lambda_b c_d r^d)}{c_d r^d}.$$

In this case, the maximum density is achieved as $\lambda \rightarrow \infty$, yielding $\lambda_{\max} = (c_d r^d)^{-1}$, which is a factor of e higher than for the type I process. For $d = 2$ and $r = 1$, the maximum density is $1/\pi \approx 0.318$. In comparison, the densest deterministic arrangement, the triangular lattice, achieves a density of $2/\sqrt{3} \approx 1.15$. Examples of type I and type II processes are shown in Fig. 3.5.

Simple sequential inhibition processes

Another natural hard-core process model is the *simple sequential inhibition process*, often abbreviated to SSI process, where new points are generated sequentially uniformly in a domain $W \subset \mathbb{R}^d$. If a new point is within distance r of any existing point, it is rejected, and a new point is generated. The process terminates when no further points can be added. This process achieves a higher density than the Matérn hard-core processes, since a point is rejected only by (accepted) points of the process. On the plane, the process achieves slightly more than $1/3$ of the densest packing (triangular lattice). An example for $r = 1/3$ is shown in Fig. 3.6. Another example is Fig. 2.9, where $r = 1$.

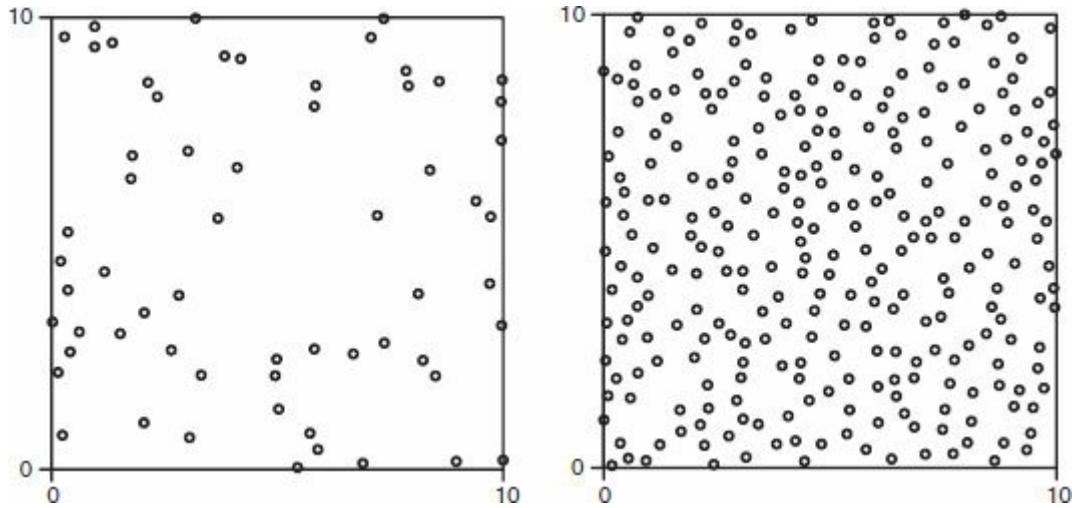


Figure 3.5 Matérn hard-core processes. (Left) Type I process with $\lambda_b = 8$ and $r = 1/3$. The density of this process is $\lambda = 8\exp(-8\pi/9) \approx 0.49$. (Right) Type II process with the same parameters. In this case the density is significantly higher: $\lambda = (1 - \exp(-8\pi/9))/(\pi/9) \approx 2.69$. This is close to the maximum $\lambda_{\max} = 9/\pi \approx 2.86$. The realization has 280 points.

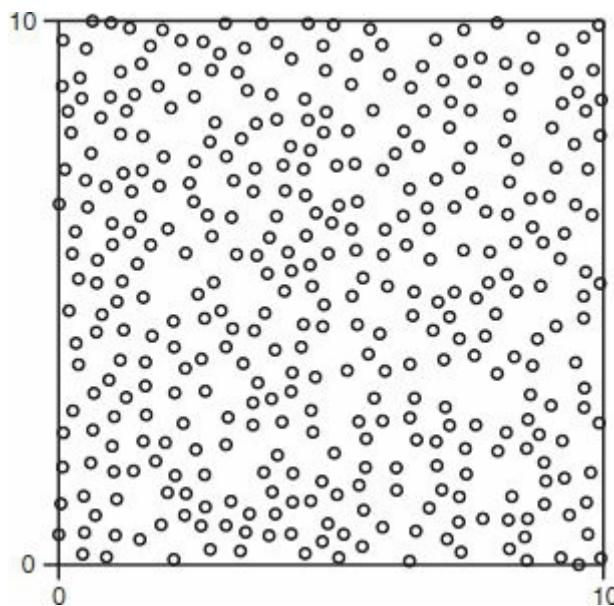


Figure 3.6 Simple sequential inhibition process with $r = 1/3$. This process achieves an even

higher density than the Matern hard-core process of type II. This realization has 99 more points on $[0,10]^2$ than the type II realization in Fig. 3.5 (right).

3.6 Gibbs processes

3.6.1 Introduction and definition

Gibbs (point) processes are related to Gibbs distributions in statistical physics. The main idea is to shape the distribution of a basic point process, usually a PPP, using a density on the space of counting measures \mathcal{N} , as illustrated by an introductory example.

Example 3.9 Let Ψ be a PPP with intensity measure Λ such that $\Lambda(\mathbb{R}^d) = 1$ and denote its distribution by Q . Define a new point process distribution $P_{[f]}$ by

$$P_{[f]}(Y) = \int_Y f_\lambda(\varphi) Q(d\varphi) \quad \text{for } Y \in \mathfrak{N}, \quad (3.9)$$

where $f_\lambda : \mathcal{N} \mapsto \mathbb{R}^+$ is given by

$$f_\lambda(\varphi) = \lambda^{\varphi(\mathbb{R}^d)} \exp(1 - \lambda).$$

The new distribution $P_{[f]}$ defines a PPP of intensity measure $\lambda\Lambda$. To see this, let P be the distribution of a PPP Φ of intensity measure $\lambda\Lambda$, and compare the measures that P and $P_{[f]}$ give to the event

$$Y_K = \{\varphi \in \mathcal{N} : \varphi(\mathbb{R}^d) = n, \varphi(K) = 0\}.$$

Since $P(Y_K) \equiv P_{[f]}(Y_K)$ for all $n \in \mathbb{N}_0$ and compact K , the void probabilities agree, so, by Theorem 2.23, the distributions are equal.

The function $f(\varphi)$ is a density with respect to the distribution of the Poisson process; it can be interpreted as the relative likelihood that Φ assumes the realization φ compared with Ψ assuming φ . It is therefore also called the *likelihood ratio* or *likelihood function*. Of course, it needs to be guaranteed that the resulting distribution is indeed the distribution of a point process, i.e., that $P(\mathcal{N}) = 1$. However, while f is a density, it is *not* a probability density on \mathcal{N} , so $\int_{\mathcal{N}} f(\varphi) d\varphi \neq 1$ in general.

The function f is called *hereditary* if $f(\varphi) > 0$ implies that $f(\varphi') > 0$ whenever φ'

$\subseteq \varphi$. So hereditary means that if a realization φ occurs with positive probability, so do all realizations with one or more points removed. Conversely, if $f(\varphi) = 0$, then all supersets of φ must also have zero probability.

DEFINITION 3.9 (Gibbs process) Let Q be the distribution of a PPP with intensity Λ and $f: \mathcal{N} \mapsto \mathbb{R}^+$ be a hereditary function such that

$$\int_{\mathcal{N}} f(\varphi) Q(d\varphi) = 1. \quad (3.10)$$

Then

$$P(Y) = \int_Y f(\varphi) Q(d\varphi), \quad \forall Y \in \mathfrak{N}, \quad (3.11)$$

is a probability measure on $(\mathcal{N}, \mathfrak{N})$ and thus the distribution of a point process. This point process is called a *Gibbs process*.

The expression (3.11) shows that P is absolutely continuous with respect to Q , so, formally f is the Radon-Nikodým derivative $f = dP/dQ$. So f is the density of P with respect to the probability measure of the PPP. This is a concept that we have previously tacitly used, with the Lebesgue measure on \mathbb{R}^d instead of the point process probability measure on \mathcal{N} : We say that a function g is the density of a measure G with respect to the Lebesgue measure if $G(B) = \int_B g(x) dx$, or, conversely g is the Radon-Nikodým derivative $g = dG/dv$.

In applications, frequently a measurable function f' is chosen such that $0 < F' < \infty$, where

$$F' = \int_{\mathcal{N}} f'(\varphi) Q(d\varphi).$$

f' is then normalized by F' to obtain a valid probability measure P by setting $f = f'/F'$. In many cases it is difficult to determine F' explicitly, so often the pseudo-likelihood function f' is considered instead of f .

The underlying PPP is usually referred to as the *reference Poisson process*, and we may write P_{ref} for its distribution instead of Q . We will for simplicity assume that the intensity measure of this reference PPP is given by $\Lambda(B) = |B \cap W|$ for a compact set W . This guarantees that the PPP has a finite number of points a.s., a property which is inherited by the Gibbs process. The distribution of such a PPP can be written explicitly as

$$P_{\text{ref}}(Y) = Q(Y) = e^{-|W|} \sum_{n=0}^{\infty} \frac{1}{n!} \int_{W^n} \mathbf{1}(\varphi \in Y) dx_1 \cdots dx_n, \quad (3.12)$$

where $\varphi = \{x_1, \dots, x_n\}$. To see this, partition \mathcal{N} into the events

$$Y_n = \{\varphi \in \mathcal{N} : \varphi(W) = n\}$$

that the point process has exactly n points. We know that for the PPP of unit intensity

$$\begin{aligned}\mathbb{P}_{\text{ref}}(Y_n) &= \int_{Y_n} \mathbb{P}_{\text{ref}}(\mathrm{d}\varphi) = e^{-|W|} \frac{|W|^n}{n!} \\ &= \frac{e^{-|W|}}{n!} \int_{W^n} \mathrm{d}x_1 \cdots \mathrm{d}x_n.\end{aligned}$$

Decomposing the general event into the events $Y \cap Y_n$ and applying the total probability law yields (3.12), since each term in the sum is the probability of an event $\mathbb{P}_{\text{ref}}(Y \cap Y_n)$.

Since $\mathbb{P}_{\text{ref}}(Y) = \mathbb{E}1(\Phi_{\text{ref}} \in Y)$, the distribution of the Gibbs process follows as

$$\begin{aligned}\mathbb{P}(Y) &= \mathbb{E}(1(\Phi_{\text{ref}} \in Y) f(\Phi_{\text{ref}})) \\ &= e^{-|W|} \sum_{n=0}^{\infty} \frac{1}{n!} \int_{W^n} 1(\varphi \in Y) f(\varphi) \mathrm{d}x_1 \cdots \mathrm{d}x_n,\end{aligned}\tag{3.13}$$

and the normalization condition (3.10) can be expressed as

$$\mathbb{P}(\mathcal{N}) = \mathbb{E}f(\Phi_{\text{ref}}) = 1.\tag{3.14}$$

The fact that f is the density of the Gibbs process with respect to the unit-intensity PPP may also be expressed as follows: For measurable $h: \mathcal{N} \mapsto \mathbb{R}$, we have

$$\mathbb{E}(h(\Phi)) = \mathbb{E}(h(\Phi_{\text{ref}}) f(\Phi_{\text{ref}})).$$

3.6.2 Strauss processes

An important class of Gibbs processes is constituted by the Strauss processes, defined as follows.

DEFINITION 3.10 (Strauss process) Let $a > 0$, $R > 0$, and $b \in \mathbb{R}^+ \cup \{\infty\}$. A Gibbs process is called a Strauss process on W if the density f has the form

$$f(\varphi) = c a^{\varphi(W)} \exp(-bt_R(\varphi)), \quad \forall \varphi \in \mathcal{N},$$

where

$$t_R(\varphi) = \frac{1}{2} \sum_{x,y \in \varphi}^{\neq} \mathbf{1}_{W \times W}(x, y) \mathbf{1}_{(0,R)}(\|x - y\|)$$

denotes the number of point pairs $\{x, y\}$ of φ whose distance is smaller than R . The notation \sum^{\neq} indicates a multi-sum over a set, where none of the elements of the set may be taken more than once. For example,

$$\sum_{m,n \in [5]} 1 = 25, \quad \text{whereas} \quad \sum_{m,n \in [5]}^{\neq} 1 = 20.$$

c is a normalizing constant so that (3.10) holds, i.e.,

$$c^{-1} = \int_{\mathcal{N}} a^{\varphi(W)} \exp(-bt_R(\varphi)) Q(d\varphi).$$

The parameters a , b , and R are called *activity*, *interaction parameter*, and *interaction radius*, respectively. The interaction parameter b determines the strength of repulsion between the points. If $b = \infty$, in particular,

$$f(\varphi) = \begin{cases} 0 & \text{if } t_R(\varphi) > 0 \\ ca^{\varphi(W)} & \text{if } t_R(\varphi) = 0. \end{cases} \quad (3.15)$$

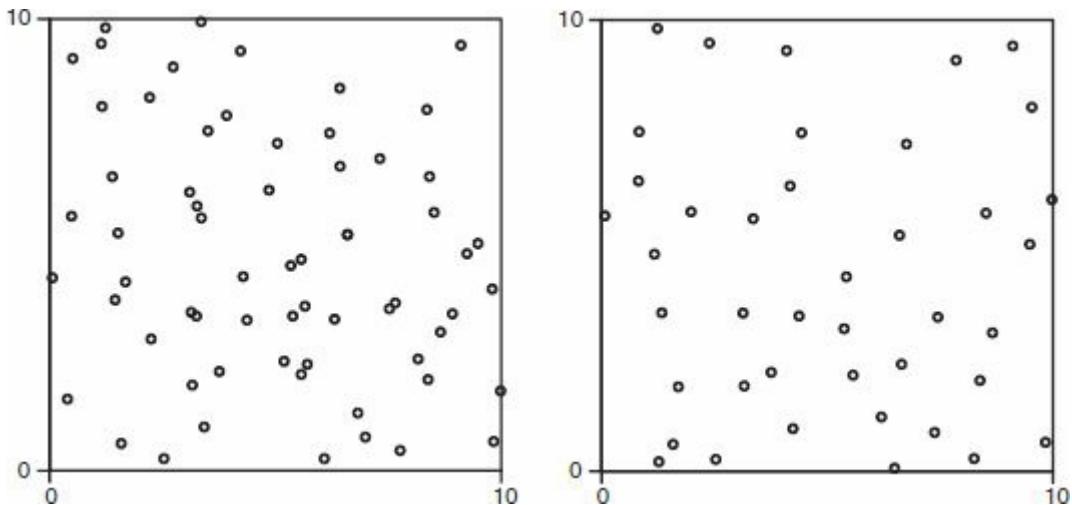


Figure 3.7 Illustration for Strauss processes with $a = 2$ and $R = 1$. (Left) $b = 1$. This realization has 65 points. (Right) $b = 5$. In this case, patterns with point pairs within distance 1 are penalized more severely; as a result, this realization has only 41 points.

This likelihood function describes a hard-core process with hard-core radius R , since f assigns probability 0 to all point sets that have two points closer than R . This process is often called a *Poisson hard-core process*. At the other extreme, if $b = 0$, then $f(\varphi) = ca^{\varphi(W)}$, and it is easily verified that the resulting Strauss process is a uniform (on W) PPP of intensity a . So in this case there is no repulsion at all between points. Figure 3.7 shows two Strauss processes with $a = 2$ for $b = 1$ and $b = 5$. For $b = 0$, the process would be a PPP of intensity 2 and contain 200 points on the $[0, 10]^2$ square on average.

A slight generalization is the *Strauss hard-core process*.

DEFINITION 3.11 (Strauss hard-core process) Let $a > 0$, $R' > R > 0$, and $b \in \mathbb{R} \cup \{\infty\}$. A Gibbs process is called a Strauss hard-core process on W if the density f has the form

$$f(\varphi) = \begin{cases} 0 & \text{if } t_R(\varphi) > 0 \\ ca^{\varphi(W)} \exp(-bt_{R'}(\varphi)) & \text{if } t_R(\varphi) = 0, \end{cases}$$

where t_R , is defined as in the standard Strauss process.

In this case, R is the hard-core radius, while R' is the interaction radius. The interaction parameter b here can assume arbitrary real values (including ∞) and describes the strength of the repulsion of point pairs whose distance lies in $[R, R']$. Interestingly, although the process is by construction hard-core, its appearance may be clustered: see Fig. 3.8. The parameter R imposes a strict minimum distance, which led to the name Strauss hard-core process.

Strauss processes are Gibbs processes with pairwise interaction only, i.e., there are no higher-order interactions, such as interactions between triplets of points.

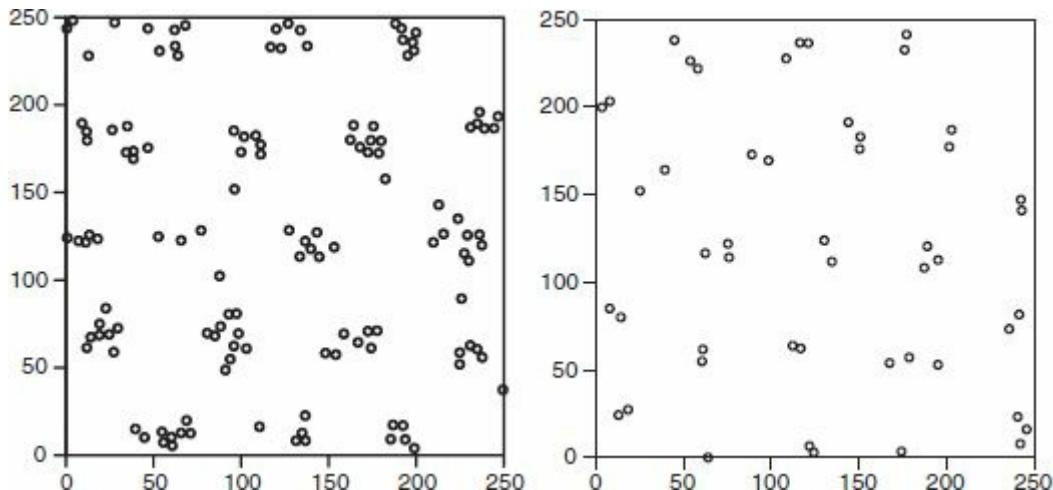


Figure 3.8 Realizations of Strauss hard-core processes with $a = 50$, $R = 4$, and $R' = 50$. (Left) $b = 1$. This realization has 130 points. (Right) $b = 4$. In this case, patterns with point pairs with distance in $[4, 50]$ are penalized more severely; as a result, this realization has only 49 points.

3.6.3 Area-interaction processes

Area-interaction processes, which were introduced in Baddeley & van Lieshout (1995), belong to the class of Gibbs processes with higher-order interactions. Their density f has the form

$$f(\varphi) = ca^{\varphi(W)} \gamma^{-|A_\varphi^{(r)}|},$$

where

$$A_\varphi^{(r)} = b(o, r) \oplus \varphi = \bigcup_{i=1}^{\varphi(W)} b(x_i, r)$$

is the region formed by putting disks of radius r at each point in φ ; \oplus denotes Minkowski addition: $A \oplus B = \{x \in A, y \in B: x + y\}$. The parameters of the area-interaction process are a , r , $\gamma > 0$, and c is the normalizing constant. If $\gamma < 1$, patterns with large areas $|A_\varphi^{(r)}|$ are preferred, which results in more regular processes. For $\gamma = 1$, there is no interaction, and the process is a PPP with intensity a . For $\gamma > 1$, the patterns are clustered, since they result in smaller $|A_\varphi^{(r)}|$. The normalization condition (3.14) yields

$$c^{-1} = \mathbb{E}(a^{\Phi_{\text{ref}}(W)} \gamma^{-|A_{\Phi_{\text{ref}}}^{(r)}|}), \quad (3.16)$$

which in general is difficult to compute.

The appeal of the area-interaction process lies in the fact that it can exhibit regular (repulsive) or clustered (attractive) behavior as a function of a single parameter, γ . Hence, in terms of the axis in Fig. 3.1, area-interaction processes span the entire range from very regular processes (small γ) to heavily clustered processes (large γ).

The effect of γ on the behavior of the process can also be observed from the J function (see Definition 2.40). It can be shown that $J(s)$ is given by

$$J(s) = \frac{1 - G(s)}{1 - F(s)} = \frac{a\gamma^{-\pi r^2}}{\lambda}, \quad s > 2r,$$

where λ is the intensity of the process. For $\gamma = 1$, we have $\lambda = a$, and thus $J(s) = 1$, as expected. $J(s) < 1$ for clustered patterns, which corresponds to $\gamma > 1$, and vice versa. This behavior is consistent with the observation made at the beginning of the chapter.

3.6.4 Conditional Gibbs processes

Definition and distribution

Here we focus on the conditional distribution of Gibbs processes given that they have a fixed number $n \geq 1$ of points. This situation is often referred to as the *canonical ensemble* in statistical physics.

DEFINITION 3.12 (Conditional Gibbs process) Let the intensity of the reference PPP be $\Lambda(B) = |B \cap W|$ for a compact set W , let $Y_n = \{\varphi \in \mathcal{N}: \varphi(W) = n\}$ be the class of point patterns with n points in W , and let P denote the distribution of the Gibbs process as per (3.11). Then

$$\mathbb{P}_n(Y) = \frac{\mathbb{P}(Y \cap Y_n)}{\mathbb{P}(Y_n)}, \quad \forall Y \in \mathfrak{N},$$

is the distribution of a conditional Gibbs process with n points on W for all n for which $\mathbb{P}(Y_n) > 0$.

From this definition follows an explicit expression for the distribution. We will henceforth denote a point set with n elements by φ_n .

THEOREM 3.13 (Distribution of conditional Gibbs process) *The distribution of the Gibbs process, conditioned on having n points on the domain $W \in \mathcal{B}^d$, is given by*

$$\begin{aligned} \mathbb{P}_n(Y) &= \int_{W^n} \mathbf{1}(\varphi_n \in Y) f_n(x_1, \dots, x_n) d(x_1, \dots, x_n) \\ &= \int_{[Y]} f_n(x_1, \dots, x_n) d(x_1, \dots, x_n) \quad \text{for } Y \in \mathfrak{N}. \end{aligned} \quad (3.17)$$

The second integral is over $[Y] = \{(x_1, \dots, x_n) \in W^n : \{x_1, \dots, x_n\} \in Y\}$, and $f_n : \mathbb{R}^{nd} \mapsto \mathbb{R}^+$ is a joint density function on W^n given by

$$f_n(\varphi_n) = \frac{f(\varphi_n)}{|W|^n \int_{\mathcal{N}} f(\varphi) Q_n(d\varphi)}, \quad (3.18)$$

where f is the (unconditioned density) of the Gibbs process as introduced in (3.11), and $Q_n(Y) = Q(Y \cap Y_n)/Q(Y_n)$ is the conditional distribution of the reference PPP given that it has n points in W .

Proof To see how (3.17) and (3.18) follow from Definition 3.12 we proceed as follows:

$$\begin{aligned} \mathbb{P}_n(Y) &= \frac{\mathbb{P}(Y \cap Y_n)}{\mathbb{P}(Y_n)} = \frac{\int_{Y \cap Y_n} f(\varphi) Q(d\varphi)}{\int_{Y_n} f(\varphi) Q(d\varphi)} \\ &= \frac{\int_{Y \cap Y_n} f(\varphi) Q(d\varphi) / Q(Y_n)}{\int_{Y_n} f(\varphi) Q(d\varphi) / Q(Y_n)} \\ &= \frac{\int_Y f(\varphi) Q_n(d\varphi)}{\int_{\mathcal{N}} f(\varphi) Q_n(d\varphi)} \\ &\stackrel{(a)}{=} \frac{\int_{[Y]} f(x_1, \dots, x_n) d(x_1, \dots, x_n)}{|W|^n \int_{\mathcal{N}} f(\varphi) Q_n(d\varphi)} \\ &= \int_{[Y]} f_n(x_1, \dots, x_n) d(x_1, \dots, x_n); \end{aligned}$$

(a) follows since the conditional distribution of a uniform PPP on W given that it has n points is uniform over W^n , hence the density is $|W|^{-n} \mathbf{1}((x_1, \dots, x_n) \in W^n)$.

A quick sanity check for $f(\varphi) \equiv 1$ shows that the resulting density f_n is indeed the uniform one over W^n , as expected. An alternative way of expressing f_n is by normalization. Since f_n is a pdf on W^n , given Y_n ,

$$f_n(x_1, \dots, x_n) = \frac{f(x_1, \dots, x_n)}{\int_{W^n} f(x_1, \dots, x_n) dx_1 \cdots dx_n}. \quad (3.19)$$

The function f_n does not depend on the ordering of the points in the sequence (x_1, \dots, x_n) , thus we may write $f_n(\varphi n)$ instead of $f_n(x_1, \dots, x_n)$.

We observe that in the conditional case the density f_n is a density with respect to the Lebesgue measure on \mathbb{R}^{nd} , since the conditional distribution of the unit-intensity PPP given that it has n nodes reduces to this Lebesgue measure. Indeed, the expression (3.17) reveals the connection to the general finite point processes, as introduced in Section 3.2. The difference is that in the Gibbsian framework the density f_n is usually expressed in the form of interaction potentials between points, as is common in statistical mechanics.

Standard description

It is one of the fundamental principles of statistical physics that in equilibrium the probability density of a configuration is inversely proportional to the exponential of the potential energy of the configuration. As a consequence, a standard form used to describe f_n is

$$f_n(\varphi_n) = f_n(x_1, \dots, x_n) = \frac{\exp(-U(x_1, \dots, x_n))}{Z},$$

where $U: \mathbb{R}^{nd} \mapsto \mathbb{R} \cup \{\infty\}$ is the *energy function* and Z is a normalizing constant called the *configurational partition function* that ensures that f_n is a valid pdf.

It is given by

$$Z = \int_{W^n} e^{-U(x_1, \dots, x_n)} dx_1 \cdots dx_n,$$

which is hard to compute usually. Frequently U is chosen to be of the form

$$U(x_1, \dots, x_n) = \sum_{i=1}^n \left(\varphi(x_i) + \sum_{k=i+1}^n \theta(x_i, x_k) \right), \quad \forall (x_1, \dots, x_n) \in W^n.$$

Here $\varphi: W \mapsto \mathbb{R} \cup \{\infty\}$ determines the intensity function of the Gibbs process, and $\theta: W^2 \mapsto \mathbb{R} \cup \{\infty\}$ is the *pair potential* with $\theta(x, y) = \theta(y, x)$ for all $x, y \in W$. As for Strauss processes, this choice of U is restricted to pairwise interactions. Indeed, we can define the Strauss process in terms of the functions φ and θ : For the Strauss

process (with n points), $\varphi(x) = -\log a$ for all $x \in W$, such that

$$f_n(x_1, \dots, x_n) = ca^n \exp\left(-\sum_{i=1}^n \sum_{k=i+1}^n \theta(x_i, x_k)\right),$$

and the pair potential is given by

$$\theta(x, y) = \begin{cases} b & \text{if } \|x - y\| \leq R \\ 0 & \text{if } \|x - y\| > R. \end{cases}$$

For $b = \infty$ and $a = 1$, we obtain the special case of the Poisson hard-core process with radius R :

$$f_n(x_1, \dots, x_n) = \begin{cases} 0 & \text{if } d_{\min} \leq R \\ c & \text{if } d_{\min} > R, \end{cases}$$

where $d_{\min} = \min_{i, k \in [n], i \neq k} \|x_i - x_k\|$ is the minimum distance between any pair of points. The number of points n cannot be chosen too large for a given W , since otherwise f_n will be zero for (almost) all (x_1, \dots, x_n) .

3.6.5 General finite Gibbs processes and Papangelou conditional intensity

Here, the number of points is variable also. In statistical physics, this is referred to as the *grand canonical ensemble*.

Direct description

We have defined Strauss and area-interaction processes, which are characterized directly for general numbers of points. As an example, we explore in detail how uniform PPPs of arbitrary intensity can be obtained.

Example 3.10 We re-consider Example 3.9 in a slightly modified form. We would like to verify that

$$f(\varphi) = \lambda^{\varphi(W)} e^{(1-\lambda)|W|} \tag{3.20}$$

is indeed the likelihood function yielding a PPP of intensity λ . We have from (3.13)

$$\begin{aligned} \mathbb{P}(\Phi(W) = n) &= \mathbb{P}(Y_n) = \frac{e^{-|W|}}{n!} \int_{W^d} \lambda^n e^{(1-\lambda)|W|} dx_1 \cdots dx_n \\ &= e^{-\lambda|W|} \frac{(\lambda|W|)^n}{n!}, \end{aligned}$$

which is the desired Poisson distribution with mean $\lambda|W|$. Also, conditioned on having n points, the normalization in (3.19) takes the form

$$\frac{f(\varphi)}{\int_{W^n} f(\varphi) dx_1 \cdots dx_n},$$

which yields the uniform distribution $|W|^{-n}$ on W . Hence (3.20) does indeed specify a PPP of intensity λ on $|W|$.

This calculation also provides the pre-constant c of the general (unconditioned) area-interaction process for $\gamma = 1$. In this case, the process is a PPP of intensity a , thus we have $c = e^{(1-a)|W|}$ from (3.20). From (3.16), we have the alternative expression $c^{-1} = \mathbb{E}a^{\Phi_{\text{ref}}(W)}$. Thus c^{-1} is the generating functional of the Poisson distribution with mean $|W|$, which is $e^{|W|(a-1)}$, so both ways of calculating c yield the same result, as required.

Conditional intensity

The likelihood function is related to another characteristic of Gibbs processes, the *Papangelou conditional intensity*, which is sometimes called just the conditional intensity.

DEFINITION 3.14 (Papangelou conditional intensity) Let $W \subset \mathbb{R}^d$ be compact, φ be a finite point pattern, and $f: \mathcal{N} \mapsto \mathbb{R}^+$ be hereditary. Then the *Papangelou conditional intensity* $\lambda(u, \varphi)$ is defined for $u \in W, \varphi \in \mathcal{N}$ as

$$\lambda(u, \varphi) \triangleq \frac{f(\varphi \cup \{u\})}{f(\varphi \setminus \{u\})} \quad \text{if } f(\varphi \setminus \{u\}) > 0. \quad (3.21)$$

If $f(\varphi \setminus \{u\}) = 0$, $\lambda(u, \varphi) = 0$ (in this case the numerator is 0 also since f is hereditary).

This definition does not depend on whether $u \in \varphi$ or not, since $\lambda(u, \varphi) = \lambda(u, \varphi \setminus \{u\})$. As a consequence, there is a one-to-one mapping from $\lambda(u, \varphi)$ to $f(\varphi)$. The conditional intensity may be interpreted as

$$\lambda(u, \varphi) du = \mathbb{P}(\Phi(du) = 1 | \Phi \cap (du)^c = \varphi \cap (du)^c),$$

which is the conditional probability that there is a point of the process Φ in the infinitesimal region of area du containing u given that $\Phi = \varphi$ outside of du .

For the PPP, the Papangelou conditional intensity reduces to the standard intensity due to its independence property, i.e., $\lambda(u, \varphi) = \lambda(u)$. Generally $\lambda(u) = \mathbb{E}\lambda(u, \Phi)$. These relationships will be explored in detail in Section 8.4.2.

A major advantage of the conditional intensity is that it does not depend on the

pre-constant of the likelihood f , which is often impossible to calculate.

The Papangelou conditional intensity can also be expressed using the Janossy densities. Assuming φ has n points,

$$\lambda(u, \varphi) = \frac{j_{n+1}(\varphi \cup \{u\})}{j_n(\varphi)}, \quad u \notin \varphi.$$

A comparison with (3.21) indicates that the Janossy density and the conditional intensity are proportional. Indeed, it can be shown that

$$\frac{j_n(x)}{f_n(x)} = e^{-|W|} \quad (3.22)$$

holds, where f_n is the density with respect to the distribution of the unit-rate PPP. A comparison of (3.3) and (3.13), which is of the form $\mathbb{E}u(\Phi)$, makes this relationship plausible.

Example 3.11 For the Poisson hard-core process with density given in (3.15), we have

$$\lambda(u, \varphi) = 1(t_R(\varphi \cup \{u\}) = 0),$$

where t_R is the number of point pairs closer than R , as defined in Definition 3.10.

Example 3.12 For the area-interaction process,

$$\lambda(u, \varphi) = a\gamma^{-|B(u, \varphi)|},$$

where $B(u, \varphi) = b(u, r) \setminus A_\varphi^{(r)}$ is the part of the disk of radius r centered at u that is not covered by any of the other disks $b(x_i, r)$. If $\gamma < 1$, this implies that a random point is less likely to occur when this uncovered area is small. Conversely, for $\gamma > 1$, the conditional density is largest if $|B(u, \varphi)| = 0$, i.e., if u is already covered by $A_\varphi^{(r)}$. This results in clustering. In Fig. 3.9, two conditional intensities for area-interaction processes with different values of γ are shown.

As mentioned previously, since it is often impossible to calculate the normalizing constant, the density f is often specified only as an unnormalized density $f' \propto f$, where f' is hereditary also. Usually f' is taken to be *locally stable*, in the sense that there is a constant C such that

$$f'(\varphi \cup \{u\}) \leq Cf'(\varphi) \text{ for all } u \in W, \varphi \subset W.$$

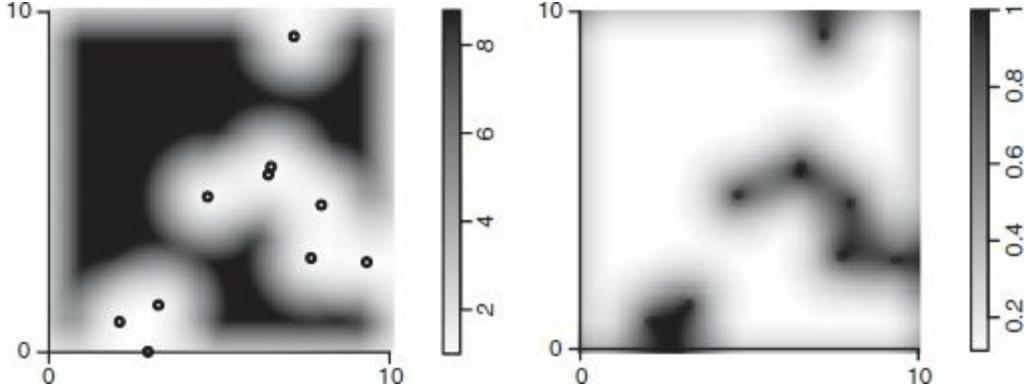


Figure 3.9 Papangelou conditional intensity for an area-interaction process with $a = 1$ and $r = 1$ on the $[0, 10]^2$ square. Ten uniformly distributed points are present. They are shown in the plot also. (Left) $\gamma = 1/2$. Here, the minimum conditional intensity is 1, and the maximum is $\gamma^{-\pi r^2} \approx 8.8$. The high-intensity (black) areas are the regions with no existing points, which indicates a regular process. (Right) $\gamma = 2$. In this case, the minimum intensity is $\gamma^{-\pi r^2} \approx 0.11$, and the maximum is 1. The high-intensity areas are the ones close to existing points, which indicates clustering.

Local stability implies that f' is hereditary and ensures integrability with respect to the unit-rate PPP.

The methods considered here for finite point processes cannot be used directly for infinite point processes. One reason is that a uniform PPP has a density with respect to another PPP only if their intensities are the same, as illustrated in the following example.

Example 3.13 This is a counter-example that shows that a uniform PPP $\Phi^{(1)}$ cannot be obtained from another uniform PPP $\Phi^{(2)}$ if the two have different intensities $\lambda_1 \neq \lambda_2$. Let

$$Y_\lambda = \{\varphi \in \mathcal{N}: \varphi(K_n)/|K_n| \rightarrow \lambda\}$$

for a sequence (K_n) of compact sets such that $K_n \uparrow \mathbb{R}^d$. Then $\mathbb{P}(\Phi^{(1)} \in Y_{\lambda_1}) = 1$ but $\mathbb{P}(\Phi^{(1)} \in Y_{\lambda_2}) = 0$ and vice versa, so the distribution of $\Phi^{(1)}$ cannot be obtained by shaping the distribution of $\Phi^{(2)}$ by a transformation of the form (3.9). An analogous situation can be observed in linear system theory. Consider a linear time-invariant system that shapes the frequency content of an input signal. If a certain frequency component is not present in the input signal, it cannot occur in the output signal. Similarly, for a certain event to have a nonzero probability in a Gibbs process, it must have a non-zero probability in the underlying process already. For two

(infinite) PPPs with different intensities, this is not the case.

On the other hand, for two PPPs on a bounded set W , the two probability measures, denoted as P_{λ_1} and P_{λ_2} , are mutually absolutely continuous with likelihood function

$$f(\varphi) = \frac{dP_{\lambda_1}}{dP_{\lambda_2}}(\varphi) = \left(\frac{\lambda_1}{\lambda_2}\right)^{\varphi(W)} e^{-(\lambda_2 - \lambda_1)|W|}. \quad (3.23)$$

A direct proof will be given in Example 4.3. This is a generalization of (3.9) in Example 3.9.

3.6.6 Markov point processes

Markov point processes are a more general class of Gibbs processes. A point process on $W \subset \mathbb{R}^d$ is called Markov if its conditional distribution on $B \subset W$ given its behavior in $W \setminus B$ depends only on its behavior on B^* , where

$$B^* = \{x \in W \setminus B : \exists y \in B \text{ with } \|x - y\| < r\} = (B \oplus b(o, r)) \setminus B$$

for some interaction radius r . For the Strauss process, the interaction radius is R , for the hard-core Strauss process, it is R' , and for the area-interaction process, it is $2r$. For an interaction radius R , the Papangelou conditional intensity satisfies

$$\lambda(u, \varphi) = \lambda(u, \emptyset) \text{ for all } \varphi \text{ with } \varphi \cap b(u, R) = \emptyset,$$

since points further away from u than R do not contribute to the conditional intensity at u .

3.6.7 Simulation of Gibbs processes

A straightforward method to simulate a Gibbs process with a given density f is based on rejection sampling.

1. Produce a realization φ of a uniform PPP on the domain W .
2. Produce a realization x of a uniformly distributed random variable on $(0, \sup f)$, where \sup is the supremum.
3. If $x < f(\varphi)$, then accept φ , otherwise go back to step 1.

This method is analogous to the method of producing realizations of numerical random variables with distribution f on $[0, w]$ by first producing a realization u of

a uniformly distributed random variable on $[0, w]$, then a realization x of a uniformly distributed random variable on $(0, \sup f)$, and accepting u if $f(u) > x$.

This method is inefficient if most realizations φ have $f(\varphi) = 0$. For example, the probability of producing an acceptable sample for the Poisson hard-core process is about $\exp(-\lambda^2 \pi r^2)$, since the number of pairs within distance r is about $\lambda^2 \pi r^2$, and they are approximately Poisson. So, for $\lambda = 2$ and $r = 2$, this probability is about 1.5×10^{-22} , which is of course unacceptable.

A more efficient simulation procedure is based on *spatial birth–death processes*, which are stochastic spatiotemporal processes that are designed such that they converge to the desired equilibrium distribution. One method to generate a realization with a fixed number of points n is as follows.

1. Produce a uniform binomial point process with n points on W .
2. Choose one of the points at random and delete it.
3. Denoting the current realization by φ , add a point randomly according to the Papangelou conditional intensity i.e., with density (proportional to) $f(\varphi \cup \{x\})/f(\varphi)$.
4. Go back to step 2, unless a termination criterion is met.

Step 3 is implemented by rejection sampling: A uniformly randomly chosen point on W is retained with probability $f(\varphi \cup \{x\})/(M f(\varphi))$; otherwise it is rejected, in which case a new point is drawn. M is chosen such that the probability is smaller than 1 for all $x \in W$.

The question here is how long it takes for the spatiotemporal process to converge to f . One way is to examine the statistical properties of the process and terminate when they reach values within some confidence interval of the expected ones. Convergence can be accelerated by producing initially a realization that is closer to the desired one than the uniform one. For example, if a hard-core process is to be produced, then a lattice may be a better starting point than a BPP.

3.7 Shot-noise random fields

A shot-noise random field is not a point process model, but it is an important random geometrical structure that is based on an underlying point process. It results from the superposition of random impulses or effects generated by the points of a point process Φ . These impulses are usually assumed to be homogeneous in the sense that they depend only on the difference $y - x$, where $x \in \Phi$ and y is the location where the random field is measured. Denoting by $g(y)$ the contribution of the point at the origin to the random field, $g(y - x)$ is the contribution of the point x . The function g can be viewed as an impulse response, i.e., the response of a linear

system to an impulse. Usually its value decreases over distance, so it is also called the *attenuation function*.

DEFINITION 3.15 (Shot noise) Associated with a point process $\Phi \subset \mathbb{R}^d$, a *shot-noise random field* or *shot-noise process* ζ is defined as

$$\zeta(y) \triangleq \sum_{x \in \Phi} g(y - x), \quad y \in \mathbb{R}^d, \quad (3.24)$$

where $g: \mathbb{R}^d \mapsto \mathbb{R}$. If Φ is a PPP, the random field is called *Poisson shot noise*.

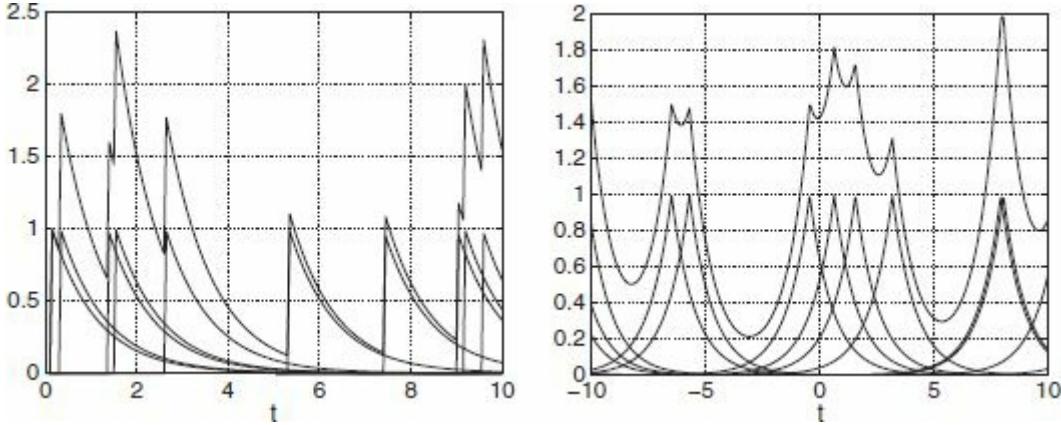


Figure 3.10 Exponential Poisson shot noise in one dimension. The underlying PPP has intensity 1. The individual impulses are shown, together with the shot noise (their sum). (Left) One-sided. Here $g(t) = e^{-t} \mathbf{1}(t \geq 0)$, and the PPP is restricted to \mathbb{R}^+ . (Right) Two-sided. Here $g(t) = e^{-|t|}$.

Example 3.14 Historically, shot noise was first observed in electric circuits, as spontaneous current fluctuations in electric conductors. Here the underlying point process was a one-dimensional uniform Poisson process, on the positive time axis \mathbb{R}^+ , and the impulse response was exponential, i.e., each spike would induce noise that decreases exponentially. In this case, $\Phi = \{t_i\} \subset \mathbb{R}^+$, and $\zeta(t) = \sum_{i=1}^{\infty} e^{t_i - t} \mathbf{1}(t \geq t_i)$.

If the effect were two-sided, the shot noise would be given by $\zeta(t) = \sum_{i=1}^{\infty} e^{-|t_i - t|}$. Illustrations for both cases are provided in Fig. 3.10.

In the context of wireless networks, the most important shot-noise field is the *interference field*, where Φ models the location of transmitters, and g (or ℓ) is the path loss law. Since the path loss is usually modeled using a power law of the form

$\ell(x) = \|x\|^{-\alpha}$, the resulting shot noise is called *power-law shot noise*.

In the next chapter, we will discuss how to find the statistical properties of a shot-noise random field, and in Chapter 5 we will apply these methods to characterize the interference field in a wireless system.

Bibliographical notes

In Daley & Vere-Jones (2003), finite point processes are covered in Chapter 5, Cox and cluster processes in Chapter 6, and conditional intensities in Chapter 7; certain aspects of spatial point processes are discussed in Chapter 15 of Daley & Vere-Jones (2008). The representation in (3.5) is the one used in Daley & Vere-Jones (2008).

Chapter 5 of Stoyan *et al.* (1995) is dedicated to the construction of point process models and Chapter 7 to random measures. Chapter 3 of Illian *et al.* (2008) focuses on finite processes, while Chapter 6 discusses stationary point process models and their simulation.

Cox processes are discussed in Chapter 6 of Kingman (1993) and in detail in the monograph by Grandell (1976), which is dedicated entirely to this topic.

The Neyman–Scott process was introduced by Neyman and Scott in two papers in 1952 and 1958 (Neyman & Scott 1952, 1958) as a model for the distribution of galaxies and has since become very popular for many applications.

Matérn hard-core processes are described in Matérn’s own book (Matérn 1986), and a detailed account of Markov point processes and their applications is provided in van Lieshout (2000).

The shot-noise process has been studied for more than a century already. In 1909, Campbell characterized the mean and variance of exponential Poisson shot noise (Campbell 1909), and in 1918 Schottky modeled spontaneous current fluctuations as shot noise (Schottky 1918). Later, Rice investigated the distribution of the shot noise (Rice 1944). Powerlaw shot noise was considered in Lowen & Teich (1990). They found that, in contrast to exponential shot noise, power-law shot noise does not converge to a Gaussian in the high-density limit (of the underlying point process). Such convergence issues were studied in detail in Heinrich & Schmidt (1985).

Problems

3.1 Determine the Janossy measure and Janossy density for the binomial point process in Example 3.1.

3.2 Consider the finite point process $\Phi \subset \mathbb{R}^2$ with $p_1 = 1/2$, $p_2 = 1/2$, $f_1(x) = 1_{W+}$

(x), and

$$f_2(x_1, x_2) = \left(\frac{1}{3} \mathbf{1}_{W^-}(x_1) + \frac{2}{3} \mathbf{1}_{W^+}(x_1) \right) \left(\frac{1}{3} \mathbf{1}_{W^-}(x_2) + \frac{2}{3} \mathbf{1}_{W^+}(x_2) \right)$$

for $W^+ = [0, 1]^2$ and $W^- = [-1, 0] \times [0, 1]$. Find the Janossy densities j_1 and j_2 and calculate $\mathbb{E}u(\Phi)$ for $u_1(\Phi) = \Phi(B)$ (for general $B \subset \mathbb{R}^2$) and $u_2(\Phi) = 2^{\Phi(\mathbb{R}^2)}$ with formula (3.3).

3.3 Find the intensity measure of a general finite point process as a function of (p_k) and F_n as defined in Definition 3.1.

3.4 Consider a one-dimensional Cox process obtained by random thinning in Example 3.6, where the basic PPP is uniform with intensity λ_b and the random field T is given by $T(x) = e^{-L|x|}$ with L a random variable.

- (a) Let $\mathbb{P}(L = 1) = 3/4$ and $\mathbb{P}(L = 2) = 1/4$. Determine $\Lambda([0, y))$ for all $y \in \mathbb{R}$, $\Lambda(\mathbb{R})$, and the value of the intensity function at 0, $\lambda(0)$.
- (b) Let the pdf of L be $f_L(x) = 2x\mathbf{1}(0 \leq x \leq 1)$. Again determine $\Lambda([0, y))$, $\Lambda(\mathbb{R})$, and $\lambda(0)$.

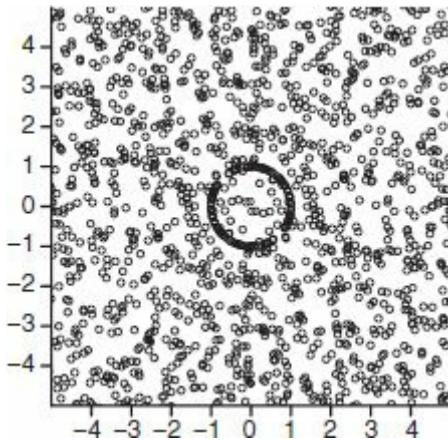


Figure 3.11 Fry plot for Problem 3.13.

3.5 For the germ–grain model in Example 3.7, consider the Cox process with intensity field

$$\zeta(x) = \lambda_a \mathbf{1}(x \in \Xi_r) + \lambda_b \mathbf{1}(x \notin \Xi_r).$$

Determine the intensity of the process and show that the Matérn cluster process can be modeled using this intensity field.

3.6 Consider the Gauss–Poisson process with $p_2 = 1$ (all clusters have exactly two points) with parent intensity λ . We would like to calculate the empty space function $F(r)$. Consider the cluster process as the superposition of the process of

cluster centers Φ_c and the process of the other points in each cluster Φ_1 , both uniform PPPs of intensity λ . They are dependent since each point in Φ_c has a point in Φ_1 at distance 1 in random orientation. Show that the empty space function can be written in the form

$$\begin{aligned} 1 - F(r) &= \mathbb{P}(\Phi_c(b(o, r)) = 0, \Phi_1(b(o, r)) = 0) \\ &= \mathbb{P}(\Phi_c(b(o, r)) = 0) \mathbb{P}(\Phi_1(b(o, r)) = 0 \mid \Phi_c(b(o, r)) = 0) \end{aligned}$$

and determine $F(r)$ for $r \leq 1/2$. For $r > 1/2$, show that

$$1 - F(r) = e^{-\lambda\pi r^2} \exp\left(-\lambda \int_r^{r+1} \frac{2 \arccos[(1+y^2-r^2)/(2y)]}{2\pi} 2\pi y \, dy\right).$$

3.7 Simulate the Poisson hole process in R and verify the intensity.

3.8 For the Gibbs processes, verify by explicit calculation of $P_\Phi(Y_K)$ and $P_\lambda(Y_K)$ (see (3.9)) that the void probabilities of the two processes agree.

3.9 Prove (3.22). *Hint.* See Example 3.3.

3.10 Find the pair potential θ corresponding to the Strauss hard-core process (with fixed number of points n) for given R and R' .

3.11 In Fig. 3.8 it appears that all “clusters” have two or three points. Explain this observation.

3.12 For the general finite Gibbs process, can the density $f(\varphi)$ (with respect to the PPP with intensity $\Lambda = |\cdot \cap W|$) be expressed as

$$f(\varphi_n) = P(Y_n) f_n(x_1, \dots, x_n), n \in \mathbb{N}_0,$$

where $Y_n = \{\varphi \in \mathcal{N}: \varphi(W) = n\}$ and f_n is the conditional density given that the process has n points?

3.13 Describe the point process pertaining to the Fry plot in Fig. 3.11.

¹ This statement is equivalent to saying that Cox processes are overdispersed, as mentioned at the beginning of the chapter.

4 Sums and products over point processes

4.1 Introduction

4.1.1 Motivation

Sums and products of functions evaluated at the points of a point process are important for many applications in wireless networks. First and foremost, the interference at a receiver is the sum of all the interfering signal powers emitted at the (random) locations of the undesired transmitters. Ideally we would like to have a complete distributional characterization of the interference. While this is possible only in very few cases, as will be discussed in [Chapter 5](#), we may be able to obtain transforms of random variables, such as the Laplace transform, which still uniquely define the distribution. Such transforms are related to *functionals* of point processes, which, in turn, have an interpretation as expected products over point processes.

4.1.2 Notation

Let $\Phi = \{x_i\} = \{x_1, x_2, \dots\} \subset \mathbb{R}^d$ be a point process and $f: \mathbb{R}^d \mapsto \mathbb{R}$ a measurable function. A sum of $f(x)$ over Φ can be alternatively written as

$$\sum_{x \in \Phi} f(x) = \int_{\mathbb{R}^d} f(x)\Phi(dx) = \int_{\mathbb{R}^d} f(x)p(x)dx, \quad (4.1)$$

where

$$p(x) = \sum_{y \in \Phi} \delta(x - y).$$

$\delta(\cdot)$ is the Dirac delta function. The mean value of the sum can be written as

$$\mathbb{E}\left(\sum_{x \in \Phi} f(x)\right) = \int_N \sum_{x \in \varphi} f(x)\mathbb{P}(d\varphi) = \int_N \int_{\mathbb{R}^d} f(x)\varphi(dx)\mathbb{P}(d\varphi).$$

The different ways of writing these expressions reflect the variety of approaches to the theory.

Example 4.1 The number of points in B is written as

$$\Phi(B) = \sum_{x \in \Phi} \mathbf{1}_B(x) = \int_B \Phi(dx).$$

Its mean value can be written in the following ways:

$$\begin{aligned}\mathbb{E}\Phi(B) &= \mathbb{E}\left(\sum_{x \in \Phi} \mathbf{1}_B(x)\right) \\ &= \int_N \varphi(B)P(d\varphi) \\ &= \int_N \sum_{x \in \varphi} \mathbf{1}_B(x)P(d\varphi) \\ &= \int_N \int_{\mathbb{R}^d} \mathbf{1}_B(x)\varphi(dx)P(d\varphi) \\ &= \int_N \int_B \varphi(dx)P(d\varphi).\end{aligned}$$

4.2 The mean of a sum

The most celebrated result about the expectation of a sum over a point process is the following.

THEOREM 4.1 (Campbell's theorem for sums) *Let Φ be a point process on \mathbb{R}^d and $f: \mathbb{R}^d \mapsto$ be a measurable function. Then the random sum*

$$S = \sum_{x \in \Phi} f(x)$$

is a random variable with mean

$$\mathbb{E}S = \int_{\mathbb{R}^d} f(x)\Lambda(dx),$$

provided that the right side is finite. If Φ has an intensity function λ ,

$$\mathbb{E}S = \int_{\mathbb{R}^d} f(x)\lambda(x)dx.$$

This formula also applies to point processes that are not simple.

Proof We prove that the theorem holds for simple f , i.e., if f is a step function

$$f = \sum_{i=1}^m c_i \mathbf{1}_{B_i}$$

for compact $B_i \subset \mathbb{R}^d$ and $c_i \in \mathbb{R}$. We have

$$S = \sum_{x \in \Phi} f(x) = \sum_{x \in \Phi} \sum_{i=1}^m c_i \mathbf{1}_{B_i}(x) = \sum_{i=1}^m c_i \Phi(B_i)$$

and thus

$$\mathbb{E}S = \mathbb{E} \left(\sum_{i=1}^n c_i \Phi(B_i) \right) = \sum_{i=1}^m c_i \mathbb{E} \Phi(B_i) = \sum_{i=1}^m c_i \Lambda(B_i) = \int_{\mathbb{R}^d} f(x) \Lambda(dx).$$

□

The result for general f follows by monotone approximation.

Written differently, Campbell's theorem says

$$\int_N \int_{\mathbb{R}^d} f(x) \varphi(dx) \mathbb{P}(d\varphi) = \int_{\mathbb{R}^d} f(x) \Lambda(dx).$$

For stationary point processes, it reduces to an integral of f over \mathbb{R}^d , scaled by the intensity.

COROLLARY 4.2 (Campbell's theorem for sums over stationary point processes) *If $\Phi \subset \mathbb{R}^d$ is stationary with intensity λ , the sum $S = \sum_{x \in \Phi} f(x)$ is a random variable with mean*

$$\mathbb{E}S = \lambda \int_{\mathbb{R}^d} f(x) dx.$$

If we need to emphasize that the function over which the sum is taken is f , we use $S[f]$.

4.3 The probability generating functional

4.3.1 Definition

The generating function G_X of a non-negative integer-valued random variable X is given by

$$G_X(t) \triangleq \mathbb{E}(t^X) = \sum_{n=0}^{\infty} t^n \mathbb{P}(X = n) \quad \text{for } t \in [0, 1].$$

The mean and variance of X can be derived from G_X :

$$\mathbb{E}(X) = G'_X(1), \quad \text{var}(X) = G''_X(1) + G'_X(1)(1 - G'_X(1)),$$

where

$$G'_X(1) = \frac{dG_X(t)}{dt} \Big|_{t=1} \quad \text{and} \quad G''_X(1) = \frac{d^2G_X(t)}{dt^2} \Big|_{t=1}.$$

The *generating functional* G of a point process $\Phi = \{x_i\}$ is defined by analogy

with G_X .

DEFINITION 4.3 (Probability generating functional of a point process) Let v be the family of all measurable functions $v: \mathbb{R}^d \mapsto [0,1]$ such that $1 - v$ has bounded support.¹ For $v \in v$, the *probability generating functional* (pgfl) of the point process Φ is defined as

$$G[v] \triangleq \mathbb{E} \left(\prod_{x \in \Phi} v(x) \right) = \int_{\mathcal{N}} \prod_{x \in \varphi} v(x) P(d\varphi).$$

Since

$$G[v] = \mathbb{E} \left[\exp \left(\sum_{x \in \Phi} \log v(x) \right) \right], \quad (4.2)$$

an equivalent definition, using (4.1), is

$$G[v] \triangleq \mathbb{E} \left[\exp \left(\int_{\mathbb{R}^d} \log v(x) \Phi(dx) \right) \right]. \quad (4.3)$$

Remark 4.1 The condition $v \in \mathcal{V}$ is sufficient, but not necessary. The pgfl exists for a larger class of functions. Two other sufficient conditions are

- (i) the point process has a finite number of points a.s. and
- (ii) $v: \mathbb{R}^d \mapsto [0,1]$, and the intensity measure satisfies

$$\int_{\mathbb{R}^d} |\log v(x)| \Lambda(dx) < \infty.$$

The second condition implies $\int_{\mathbb{R}^d} [1 - v(x)] \Lambda(dx) < \infty$. This is certainly satisfied if $v \in \mathcal{V}$, but not necessarily vice versa. So the condition is less restrictive.

To reveal the connection to the generating function, consider the joint generating function $G_{\Phi}(B_1, \dots, B_n)$ of the random vector $(\Phi(B_1), \dots, \Phi(B_n))$. For bounded Borel sets (B_1, \dots, B_n) , it is given by

$$G_{\Phi(B_1), \dots, \Phi(B_n)}(t_1, \dots, t_n) = \mathbb{E}(t_1^{\Phi(B_1)} \cdots t_n^{\Phi(B_n)}),$$

which is exactly $G[v]$ for $v(x) = \prod_{i \in [n]} v_i(x)$ and $v_i(x) = 1 + (t_i - 1)\mathbf{1}_{B_i}(x)$, where t_1, \dots, t_n all belong to $[0,1]$. Thus, in particular, for $n = 1$,

$$G_{\Phi(B)}(t) = G[v] \quad v(x) = 1 + (t - 1)\mathbf{1}_B(x).$$

Also, analogously to the generating function, the distribution P of Φ is determined uniquely by $G[v]$. Sometimes the pgfl is written as $G[v] = \mathbb{E}(v^{\Phi})$, a notation that emphasizes the connection to the generating function.

For finite point processes, the pgfl can be expressed using the Janossy measure as

$$G[v] = J_0 + \sum_{n \in \mathbb{N}} \frac{1}{n!} v(x_1) \cdots v(x_n) J_n(dx_1 \times \cdots \times dx_n).$$

Further specializing to an iid cluster of points (or mixed BPP), where each point has distribution F ,

$$G[v] = \mathbb{E} \left(\left(\int_{\mathbb{R}^d} v(x) F(dx) \right)^n \right).$$

Infinitely divisible point processes constitute an important class of point processes, in particular for the analysis of the limiting properties of superpositions of point processes. Their definition is related to the pgfl.

DEFINITION 4.4 (Infinitely divisible point process) A point process is *infinitely divisible* if, for all $k \in \mathbb{N}$, it can be represented as the superposition of k iid point processes.

Hence a point process Φ is infinitely divisible if for every k we can write

$$\Phi = \bigcup_{i \in [k]} \Phi_i^{(k)}$$

for iid $\Phi_i^{(k)}$. Denoting by $G_{1/k}$ the pgfl of each component point process, we may write the condition for infinite divisibility in terms of pgfls as

$$\forall k \in \mathbb{N}, \exists G_{1/k} \text{ s.t. } G[v] = (G_{1/k}[v])^k,$$

which indicates that, for any k , the (uniquely defined) non-negative k th root of its pgfl is again a pgfl. For the Poisson process, it is immediately seen that this is the case. A larger class of infinitely divisible point processes is constituted by the Poisson cluster processes.

4.3.2 Relationship to the moment-generating function and Laplace transform

With (4.2), we can relate the pgfl to the sum

$$S[\log v] = \sum_{x \in \Phi} \log v(x)$$

by

$$G[v] \equiv \mathbb{E}(e^{S[\log v]}),$$

which has the form of a moment-generating function. For a random variable X , the moment-generating function is defined as

$$M_x(t) \triangleq \mathbb{E}(e^t), \quad t \in \mathbb{R},$$

whenever this expectation exists. Replacing X by $S[\log v]$, we obtain $M_{S[\log v]}(t) \triangleq \mathbb{E}(e^{tS[\log v]})$, and it follows that $G[v] = M_{S[\log v]}(1)$.

Using the Laplace transform, defined for a random variable X as

$$\mathcal{L}_X(s) = \mathbb{E}(e^{-sX}), \quad s \in \mathbb{C}$$

we note that the pgfl is the Laplace transform of the random variable $S[\log v]$ for $s = -1$, i.e., $G[v] = \mathcal{L}_{S[\log v]}(-1)$.

Since $v \in [0,1]$, these quantities are all well defined.

Example 4.2 Consider the (very) simple point process on \mathbb{R} consisting of exactly one point x_1 , which is located at -1 with probability p and at 1 with probability $1-p$. Let $v(x) = 1 - 1(-2 < x < 0)$. We can evaluate $G[v]$ directly in this case. We have $G[v] = \mathbb{E}v(x_1) = 1 - p$. Using the approach via the moment-generating function of $S[\log v]$, we need the distribution of this sum, which is

$$\mathbb{P}(S[\log v] = -\infty) = p, \quad \mathbb{P}(S[\log v] = 0) = 1 - p.$$

So $\mathbb{E}S = -\infty$ if $p > 0$, and

$$\mathbb{E}(e^{s[\log v]}) = pe^{-\infty} + (1-p)e^0 = 1 - p = G[v].$$

4.4 The Laplace functional

As in the case of the generating function and the pgfl, there exists a functional that is related to the Laplace transform, the *Laplace functional*. In contrast to the pgfl, which is often restricted to random counting measures, it is defined for general non-negative random measures (see Definition 3.2).

DEFINITION 4.5 (Laplace functional) Let Ψ be a non-negative random measure on \mathbb{R}^d , and let \mathcal{U} be the set of all bounded non-negative measurable functions u of bounded support. The *Laplace functional* is defined as

$$L_\Psi[u] \triangleq \mathbb{E} \left[\exp \left(- \int_{\mathbb{R}^d} u(x) \Psi(dx) \right) \right], \quad \text{for } u \in \mathcal{U}. \quad (4.4)$$

Since u has bounded support, the integral is finite, so $L_\Psi[u] > 0$, and since u is non-negative, $L_\Psi[u] \leq 1$. The Laplace functional does not depend on the random measure Ψ itself, only on its distribution, denoted by Q . So we may write L_Q instead of L_Ψ .

The Laplace transform of the random vector $(\Psi(B_1), \dots, \Psi(B_n))$ is related to the Laplace functional by

$$\mathbb{E}(\exp(-s_1\Psi(B_1) - \dots - s_n\Psi(B_n))) = L_\Psi[s_1\mathbf{1}_{B_1} + \dots + s_n\mathbf{1}_{B_n}]$$

for $s_1, \dots, s_n \geq 0$.

If the Laplace functional is applied to a point process, i.e., a random counting measure, say Φ , we will drop the subscript if there is no danger of confusion.

Box 4.1 Probability generating and Laplace functionals

For a point process Φ , the probability generating functional and the Laplace functional are related by

$$L[u] \equiv G[e^{-u}] \quad \text{for } u \in \mathcal{U}$$

or

$$G[v] \equiv L[-\log v] \quad \text{for } v \in \mathcal{V}.$$

If a function u belongs to \mathcal{U} , then $e^{-u} \in \mathcal{V} \in \mathcal{V}$. Conversely for $v \in \mathcal{V}$, we have $-\log v \in \mathcal{U}$.

4.5 The moment-generating function of sums over Poisson processes

In the Poisson case, there is a simple formula for these functionals, given by another theorem by Campbell.

THEOREM 4.6 (Campbell's theorem for Poisson point processes) *Let Φ be a uniform PPP of intensity λ on \mathbb{R}^d and $f: \mathbb{R}^d \mapsto \mathbb{R}$ a measurable function. Then the sum*

$$S = \sum_{x \in \Phi} f(x)$$

is absolutely convergent a.s. if and only if

$$\int_{\mathbb{R}^d} \min(|f(x)|, 1) dx < \infty. \tag{4.5}$$

If it is, the moment-generating function is given by

$$\mathbb{E}(e^{tS}) = \exp\left(\lambda \int_{\mathbb{R}^d} (e^{tf(x)} - 1) dx\right) \tag{4.6}$$

for any complex t for which the integral converges, and, in particular, when t is purely imaginary.

Proof Consider (again) a simple function (step function) f that assumes finitely many non-zero values f_1, \dots, f_k and is zero outside some bounded region. Let

$$A_j = \{x: f(x) = f_j\}, \quad j = 1, \dots, k$$

be the level sets. Since the A_j are disjoint, the random variables $N_j = \Phi(A_j)$ are iid Poisson with mean $\lambda|A_j|$. We also have

$$S = \sum_{j=1}^k f_j N_j.$$

Since, for a Poisson random variable X with mean μ ,

$$\mathbb{E}(e^{tX}) = \exp(\mu(e^t - 1)),$$

the moment-generating function is

$$\begin{aligned} \mathbb{E}(e^{tS}) &= \prod_{j=1}^k \mathbb{E}(e^{tf_j N_j}) \\ &= \prod_{j=1}^k \exp(\lambda|A_j|(e^{tf_j} - 1)) \\ &= \exp\left(\sum_{j=1}^k \int_{A_j} \lambda(e^{tf(x)} - 1) dx\right) \\ &= \exp\left(\int_{\mathbb{R}^d} \lambda(e^{tf(x)} - 1) dx\right). \end{aligned} \quad \square$$

DEFINITION 4.7 (Characteristic functional) Expressing (4.6) as a function of f (and setting $t = 1$), we obtain the *characteristic functional* of the PPP

$$\mathbb{E}(e^{S[f]}) = \mathbb{E}\left(\exp\int_{\mathbb{R}^d} f(x)\Phi(dx)\right) = \exp\left(\lambda\int_{\mathbb{R}^d} (e^{f(x)} - 1) dx\right).$$

As the name suggests, the characteristic functional defines the PPP.

Example 4.3 As an application, we provide here a direct proof of the Radon-Nikodým derivative of two uniform PPPs Φ_1 and Φ_2 of intensities λ_1 and λ_2 on a bounded domain W given in (3.23). The claim is that the two probability measures P_1 and P_2 pertaining to the two processes are related by

$$P_1(Y) = \int_Y \left(\frac{\lambda_1}{\lambda_2}\right)^{\varphi(W)} e^{-(\lambda_2 - \lambda_1)|W|} P_2(d\varphi).$$

It is sufficient to show that

$$\begin{aligned}\mathbb{E}_{\Phi_1} \left[\exp \left(\int_W f(x) \Phi_1(dx) \right) \right] &= \\ \mathbb{E}_{\Phi_2} \left[\left(\frac{\lambda_1}{\lambda_2} \right)^{\Phi_2(W)} e^{-(\lambda_1 - \lambda_2)|W|} \exp \left(\int_W f(x) \Phi_2(dx) \right) \right]\end{aligned}$$

for measurable f . From the characteristic functional, we have on the left-hand side

$$\begin{aligned}\mathbb{E}_{\Phi_1} \left[\exp \left(\int_W f(x) \Phi_1(dx) \right) \right] &= \exp \left(\lambda_1 \int_{\mathbb{R}^d} (e^{f(x)} - 1) dx \right) \\ &= e^{-\lambda_1|W|} \exp \left(\lambda_1 \int_{\mathbb{R}^d} e^{f(x)} dx \right).\end{aligned}$$

For the right-hand side, we first move the term $(\lambda_1/\lambda_2)^{\Phi_2(W)}$ into the integral inside the exponential and then use the characteristic functional:

$$\begin{aligned}e^{-(\lambda_1 - \lambda_2)|W|} \mathbb{E}_{\Phi_2} \left[\exp \left(\int_W (f(x) + \log(\lambda_1/\lambda_2)) \Phi_2(dx) \right) \right] &= \\ e^{-(\lambda_1 - \lambda_2)|W|} \exp \left(\lambda_2 \int_W \left(\frac{\lambda_1}{\lambda_2} e^{f(x)} - 1 \right) dx \right) &= \\ e^{-(\lambda_1 - \lambda_2)|W|} \exp \left(\lambda_1 \int_W e^{f(x)} dx \right) e^{-\lambda_2|W|}.\end{aligned}$$

This proves the identity.

Campbell's theorem also holds for non-stationary PPPs. In this case, the convergence condition for the sum is

$$\int_{\mathbb{R}^d} \min(|f(x)|, 1) \Lambda(dx) < \infty,$$

and the result is

$$\mathbb{E}(e^{tS}) = \exp \left(\int_{\mathbb{R}^d} (e^{tf(x)} - 1) \Lambda(dx) \right).$$

The convergence condition is (4.5), since the integral

$$\int (1 - e^{-tf(x)}) \lambda(x) dx, \quad t > 0, \tag{4.7}$$

converges if and only if this condition holds. An illustrative example in one dimension is the case $f(x) = x^{-2} \mathbf{1}(x > 0)$, where the integral $\int f(x) dx$ diverges, but (4.5) holds, and the integral (4.7) evaluates to $\sqrt{t\pi}$.

Since the moment-generating function defines the moments, we have the following corollary.

COROLLARY 4.8 (Mean and variance for PPPs) *Let $S = \sum_{x \in \Phi} f(x)$ for a Poisson point process Φ on \mathbb{R}^d . Then*

$$\mathbb{E}S = \int_{\mathbb{R}^d} f(x)\Lambda(dx)$$

and

$$\text{var}(S) = \int_{\mathbb{R}^d} f^2(x)\Lambda(dx).$$

Proof Using the properties of the moment-generating function,

$$\mathbb{E}S = \frac{\partial}{\partial t}\mathbb{E}\exp(tS)\Big|_{t=0}$$

and

$$\text{var } S = \frac{\partial^2}{\partial t^2}\mathbb{E}\exp(tS)\Big|_{t=0} - (\mathbb{E}S)^2. \quad \square$$

The expression for the mean $\mathbb{E}S$ was obtained for general processes before, but the formula for the variance is new. It holds only for PPPs.

4.6 The probability generating and Laplace functionals for the Poisson point process

THEOREM 4.9 (Probability generating functional for the Poisson point process) *Let $v \in \nu$ be measurable, and let Φ be a Poisson process with intensity measure Λ . Then*

$$G[v] \triangleq \mathbb{E}\left(\prod_{x \in \Phi} v(x)\right) = \exp\left(-\int_{\mathbb{R}^d} [1 - v(x)]\Lambda(dx)\right). \quad (4.8)$$

Proof This follows from the characteristic functional by setting $v(x) = e^{f(x)}$, with the understanding that $f(x) \leq 0$. An alternative, direct proof is as follows. Consider

$$v(x) = 1 - \sum_{i=1}^n (1 - z_i)1_{C_i}(x)$$

for $z_i \in [0,1]$ and C_1, \dots, C_n pairwise disjoint and compact. Then

$$\begin{aligned} G[v] &= \mathbb{E}\left(\prod_{x \in \Phi} v(x)\right) \\ &= \int_N z_1^{\varphi(C_1)} \cdots z_n^{\varphi(C_n)} \mathbb{P}(d\varphi) \\ &= \mathbb{E}\left(z_1^{\Phi(C_1)} \cdots z_n^{\Phi(C_n)}\right) \\ &= \mathbb{E}\left(z_1^{\Phi(C_1)}\right) \cdots \mathbb{E}\left(z_n^{\Phi(C_n)}\right). \end{aligned}$$

With $\mathbb{E}(z_1^{\Phi(C_1)}) = \exp(-\Lambda(C_1)(1 - z_1))$ by the moment-generating (or probability generating) function for the Poisson distribution,

$$\begin{aligned}
G[v] &= \exp(-\Lambda(C_1)(1 - z_1)) \cdots \exp(-\Lambda(C_n)(1 - z_n)) \\
&= \exp\left(-\sum_{i=1}^n \int_{C_i} (1 - z_i) \Lambda(dx)\right) \\
&= \exp\left(-\int_{\mathbb{R}^d} (1 - v(x)) \Lambda(dx)\right).
\end{aligned}$$

This holds for piecewise constant v . The result for general v follows again by application of standard arguments from measure theory.

Example 4.4 Consider a uniform PPP of intensity λ . Let $v(x) = 1 - \mathbf{I}_B(x)$. Then the event $\Phi(B) = 0$ can be expressed as

$$\prod_{x \in \Phi} v(x) = 1.$$

It occurs with probability

$$\mathbb{P}(\Phi(B) = 0) = \mathbb{E}\left(\prod_{x \in \Phi} v(x)\right) = G[v] = \exp(-\lambda|B|),$$

as expected.

Since $L[u] \equiv G[e^{-u}]$, the Laplace functional follows directly from the pgfl:

$$L[u] = \exp\left(-\int_{\mathbb{R}^d} (1 - e^{-u(x)}) \Lambda(dx)\right).$$

This expression can also be obtained from Campbell's theorem, Theorem 4.6, specializing the expression for the moment-generating function to the case $f \geq 0$ and $t = -1$:

$$\mathbb{E}(e^{-S[f]}) = \exp\left(-\int_{\mathbb{R}^d} (1 - e^{-f(x)}) \Lambda(dx)\right). \quad (4.9)$$

But this is exactly the Laplace functional $L[f]$, since for point processes the integral $\int u(x)\Phi(dx)$ in (4.4) is just the sum $S[u]$.

4.7 Summary of relationships

Here we summarize the relationships between the different expectations and functionals introduced in this chapter. For notational simplicity when discussing PPPs, we focus on the homogeneous case.

1. Let

$$S[f] = \sum_{x \in \Phi} f(x) \quad \text{and} \quad P[v] = \prod_{x \in \Phi} v(x)$$

be the sum of $f(x)$ over a point process Φ and the product of $v(x)$ over Φ , respectively. Then $S[f]$ and $P[v]$ are related as by

$$P[v] = e^{S[\log v]} \quad \text{and} \quad S[f] = \log P[e^f].$$

In particular,

$$\mathbb{E}P[v] = \mathbb{E}(e^{S[\log v]}).$$

2. The expectation $\mathbb{E}(e^{S[f]})$ is the moment-generating function of S for $t = 1$ or the Laplace transform of S for $s = -1$:

$$\mathbb{E}(e^{S[f]}) = M_s(1) = \mathcal{L}_s(-1).$$

The moment-generating function for PPPs is given by Campbell's theorem, Theorem 4.6.

3. The expected product $\mathbb{E}P[v]$ is called the *probability generating functional* (pgfl) of the point process:

$$G[v] \equiv \mathbb{E}P[v].$$

4. Campbell's theorem says that, for PPPs,

$$\mathbb{E}(e^{S[f]}) = \exp \left(\lambda \int_{\mathbb{R}^d} (e^{f(x)} - 1) dx \right) \tag{4.10}$$

if the integral converges. This is the moment-generating function for $t = 1$.

5. If the integral diverges in (4.10) to $-\infty$, the sum $S = -\infty$ a.s., and the moment-generating function is 0. So the moment-generating function is well defined unless the integral diverges to $+\infty$. A sufficient condition to avoid divergence to ∞ is to focus on $f \leq 0$.
6. Equivalently, focus on $f \geq 0$ but consider the moment-generating function for $t = -1$, i.e., use

$$\mathbb{E}(e^{S[f]}) \equiv \mathbb{E}(e^{-s[-f]}).$$

In this way, we obtain the *Laplace functional* defined as $L[f] \triangleq \mathbb{E}(e^{-S[f]})$ for $f \geq 0$. For the PPP,

$$L[f] = \mathbb{E}(e^{-S[f]}) = \exp \left(-\lambda \int_{\mathbb{R}^d} (1 - e^{-f(x)}) dx \right), \quad f \geq 0.$$

7. The Laplace functional is related to the pgfl by setting $v(x) = e^{-f(x)}$, i.e.,

$$L[f] \equiv G[e^{-f}].$$

This yields

$$G[v] = \mathbb{E}P[v] = \exp\left(-\lambda \int_{\mathbb{R}^d} (1 - v(x)) dx\right),$$

where $v: \mathbb{R}^d \mapsto [0,1]$.

4.8 Functionals of other point processes

4.8.1 Cox processes

For Cox processes, there exists a simple relationship between the pgfl of the point process and the Laplace functional of the driving random measure.

THEOREM 4.10 (Pgfl of the Cox process) *The pgfl of the Cox process is*

$$G[v] = L_\Psi[1 - v], \quad (4.11)$$

where L_Ψ is the Laplace functional of the driving measure Ψ of Φ .

Proof Denoting the distribution of Ψ by Q , this follows from

$$\begin{aligned} G[v] &= \mathbb{E}G_\Psi[v] \\ &= \int_{\mathcal{M}} G_\Lambda[v] Q(d\Lambda) \\ &= \int_{\mathcal{M}} \exp\left(-\int_{\mathbb{R}^d} (1 - v(x)) \Lambda(dx)\right) Q(d\Lambda) \\ &= L_\Psi[1 - v], \end{aligned}$$

□

where G_Λ is the pgfl of the PPP with intensity measure Λ .

As a special case, consider the mixed Poisson process with intensity measure $X\Lambda$. Direct calculation yields

$$\begin{aligned} G[v] &= \mathbb{E}\left[\exp\left(-X \int_{\mathbb{R}^d} (1 - v(x)) \Lambda(dx)\right)\right] \\ &= \mathcal{L}_X\left(\int_{\mathbb{R}^d} (1 - v(x)) \Lambda(dx)\right). \end{aligned} \quad (4.12)$$

Since the mixed Poisson process is a Cox processes with driving measure $X\Lambda$, we can retrieve the pgfl for mixed Poisson point processes: The Laplace functional of $X\Lambda$ follows directly from (4.4), and it is easily verified that $L_{X\Lambda}[1 - v]$ corresponds to (4.12).

4.8.2 Cluster processes

We consider homogeneous independent cluster processes, as introduced in Section

[3.4.2](#), and denote the representative cluster by Φ_0 . Let $G_0^{[x]}$ denote the pgfl of $\Phi_0 + x$, i.e., the pgfl of the representative cluster translated by x :

$$G_0^{[x]}[v] \triangleq \mathbb{E} \left(\prod_{y \in \Phi_0 + x} v(y) \right) = \mathbb{E} \left(\prod_{y \in \Phi_0} v(y + x) \right).$$

THEOREM 4.11 (Pgfl of homogeneous independent cluster process) *The pgfl of the homogeneous independent cluster process with parent process Φ_p is*

$$G[v] = \mathbb{E} \left(\prod_{x \in \Phi_p} G_0^{[x]}[v] \right), \quad \text{for all } v \in \mathcal{V}. \quad (4.13)$$

Proof Denoting the distribution of the parent process by P_p and the cluster associated with parent point $x \in \Phi_p$ by $\Phi^{[x]}$, we have

$$\begin{aligned} G[v] &= \mathbb{E} \left(\prod_{x \in \Phi} v(x) \right) \\ &= \mathbb{E} \left(\prod_{x \in \Phi_p} \prod_{y \in \Phi^{[x]}} v(y) \right) \\ &= \int_{\mathcal{N}} \mathbb{E} \left(\prod_{x \in \varphi} \prod_{y \in \varphi^{[x]}} v(y) \right) P_p(d\varphi) \\ &\stackrel{(a)}{=} \int_{\mathcal{N}} \prod_{x \in \varphi} \mathbb{E} \left(\prod_{y \in \varphi^{[x]}} v(y) \right) P_p(d\varphi) \\ &= \int_{\mathcal{N}} \prod_{x \in \varphi} G_0^{[x]}[v] P_p(d\varphi) \\ &= \mathbb{E} \left(\prod_{x \in \Phi_p} G_0^{[x]}[v] \right), \end{aligned}$$

□

where (a) follows from the independence of the individual clusters.

COROLLARY 4.12 (Pgfl for Poisson cluster process) *If the parent process is a PPP with intensity measure Λ_p ,*

$$G[v] = \exp \left(\int_{\mathbb{R}^d} (G_0^{[x]}[v] - 1) \Lambda_p(dx) \right). \quad (4.14)$$

Proof Expression (4.13) can be viewed as

$$G[v] = G_p[G^{[\cdot]}_0[v]],$$

where G_p is the pgfl of the parent point process. Replacing v by $G_0^{[x]}[v]$ in (4.8) would yield the result. However, care must be taken since it is not guaranteed that $G_0^{[x]}[v] \in v$ if $v \in v$. To resolve this, let

$$\tilde{v}_r(x) = G^{[x]}_0[v] \mathbf{1}(x \in b(o, r)) + \mathbf{1}(x \notin b(o, r)) \in v$$

and note that $G[v] = \lim_{r \rightarrow \infty} G_p[\tilde{v}_r]$. The result then follows from monotone convergence. \square

Further simplification is possible for the Neyman–Scott process, where the daughter points of the representative cluster are distributed in an iid fashion with probability density f .

COROLLARY 4.13 (Pgfl for Neyman–Scott process) *For Neyman–Scott processes with uniform parent PPP of intensity λ_p ,*

$$G[v] = \exp \left(-\lambda_p \int_{\mathbb{R}^d} \left(1 - g_0 \left(\int_{\mathbb{R}^d} v(x+y) f(y) dy \right) \right) dx \right),$$

where g_0 is the generating function of the number of points in each cluster.

Proof Let the number of nodes in the representative cluster be $c = \#\Phi_0$. We have

$$\begin{aligned} G_0^{[x]}[v] &= \mathbb{E} \left(\prod_{y \in \Phi_0+x} v(y) \right) \\ &= \mathbb{E} \left[\left(\int_{\mathbb{R}^d} v(x+y) f(y) dy \right)^c \right]. \end{aligned} \quad \square$$

For the doubly Poisson cluster process, where the number of points per cluster is itself Poisson distributed with mean \bar{c} , $g_0(z) = \exp(-\bar{c}(1-z))$. In this case, the Neyman–Scott process is also a Cox process, and we can verify formula (4.11). We have

$$\begin{aligned} G[v] &= \mathbb{E} \left(\prod_{x \in \Phi_p} \prod_{y \in \Phi^{[x]}} v(y) \right) \\ &\stackrel{(a)}{=} \mathbb{E} \left[\prod_{x \in \Phi_p} \exp \left(-\bar{c} \int_{\mathbb{R}^d} (1 - v(x+y)) f(y) dy \right) \right] \\ &\stackrel{(b)}{=} \mathbb{E} \left[\exp \left(- \sum_{x \in \Phi_p} \bar{c} \int_{\mathbb{R}^d} (1 - v(y)) f(y-x) dy \right) \right] \\ &= \mathbb{E} \left[\exp \left(- \int_{\mathbb{R}^d} (1 - v(y)) \sum_{x \in \Phi_p} \bar{c} f(y-x) dy \right) \right], \end{aligned}$$

where (a) holds since each cluster is a PPP, and (b) follows by the substitution $y \leftarrow y + x$. This is the Laplace functional of $1 - v$ pertaining to the random measure Ψ whose density is given by

$$\zeta(y) = \sum_{x \in \Phi_p} \bar{c} f(y-x)$$

such that

$$\Psi(dy) = \sum_{x \in \Phi_p} \bar{c} f(y - x) dy.$$

This confirms the dual point of view of doubly Poisson (cluster) processes as Cox and as Neyman-Scott cluster processes.

Example 4.5 For the Thomas cluster process and the Matérn cluster process, the respective densities f are implied by the cluster intensity functions (3.7) and (3.8). The pgfl is given by

$$G[v] = \exp \left(-\lambda_p \int_{\mathbb{R}^d} \left[1 - \exp \left(\bar{c} \left[\int_{\mathbb{R}^d} v(x+y) f(y) dy - 1 \right] \right) \right] dx \right),$$

where $f(y) = \lambda_0(y)/\bar{c}$.

The next chapter is devoted to applications of these techniques to problems in wireless networking.

Bibliographical notes

The pgfl for finite point processes is introduced in Section 5.5 of Daley & Vere-Jones (2003) and extended to general processes and other functionals in Section 9.4 of Daley & Vere-Jones (2008). The paper by Westcott (1972) provides concise introduction to the pgfl of one-dimensional point processes and gives the conditions for its existence. It also states Campbell's theorem for the sum as an application of Fubini's theorem. In Chapter 4 of Stoyan *et al.* (1995), Campbell's theorem for sums is included; it is simply called the Campbell theorem, as is often the case. We add "for sums" to distinguish it from Campbell's theorem in Chapter 3 of Kingman (1993), which is our Campbell theorem for the PPP. This chapter in Kingman's book is concerned with sums over PPPs.

A detailed account of infinitely divisible point processes is given in the monograph by Matthes *et al.* (1978). They are also discussed in Section 10.2 of Daley & Vere-Jones (2008).

Problems

- 4.1 Complete the proof of Corollary 4.8 by explicitly calculating the derivatives of the moment-generating function.
- 4.2 Let $v(x) = \min\{1, \frac{1}{2}\|x\|^2\}$. Determine the pgfl $G[v]$ of the die process from Example 2.1.
- 4.3 Compare $\mathbb{E}e^{-s[f]}$ and $\exp(-\mathbb{E}S[f])$.
- 4.4 Show that the pgfl of the Poisson hole process, introduced in Example 3.7,

with parameters λ_1 , λ_2 , and r , is

$$G[v] = \mathbb{E} \left[\exp \left(-\lambda_2 \int_{\mathbb{R}^2 \setminus \Xi_r} (1 - v(x)) dx \right) \right].$$

Determine $G[v]$ for $v = 1 - \mathbf{1}_B(x)$, $\lambda_1 = 1$, and $r = 1$. This yields the void probability.

4.5 The *secrecy graph* is defined as follows. Let Φ be a PPP of intensity 1 and Ψ an independent PPP of intensity λ . There is a directed edge $x \rightarrow y$ for $x, y \in \Phi$ if y is closer to x than any node in Ψ , i.e., if $\|x - y\| < \|x - \Psi\|$. Compute the mean out-degree of $x \in \Phi$.

4.6 Let Φ be a Gauss-Poisson process where the representative cluster is $\Phi_0 = \{o\}$ with probability $1 - p$ and $\Phi_0 = \{o, x\}$ with probability p , where x is distributed with density f . It is assumed that the density is chosen such that $\mathbb{P}(x = o) = 0$. Show that the pgfl is

$$G[v] = \exp \left(\lambda_p \int_{\mathbb{R}^d} \left((1 - p)v(x) + p v(x) \int_{\mathbb{R}^d} v(x + y) f(y) dy - 1 \right) dx \right).$$

¹ So $v \in v$ means that $1 - v(x)$ vanishes outside a bounded set.

5 Interference and outage in wireless networks

In this chapter, we discuss several important applications of sums and functionals of point processes in wireless networks. One of the most important applications of point process theory is certainly the characterization of the interference. By interference, we always mean the interference *power*.

5.1 Interference characterization

5.1.1 Interference as shot noise

Let $\ell : \mathbb{R}^d \mapsto \mathbb{R}^+$ be a path loss function. If a point process Φ models the locations of transmitters, all transmitting at unit power, the sum power measured at location y is given by

$$I(y) = \sum_{x \in \Phi} \ell(y - x), \quad y \in \mathbb{R}^d, \tag{5.1}$$

provided that there is no fading in the channel. By comparison with (3.24), the interference $I(y)$ is a *shot-noise random field*. More precisely, since the path loss law is typically of the form $\ell(x) = \|x\|^{-\alpha}$ (or a bounded version thereof), it is a power-law shot-noise random field. An example is shown in Fig. 5.1, where the interference resulting from a PPP is plotted along a line on the plane.

5.1.2 Mean interference in stationary point processes

First, we derive the mean interference for a power-law path law $\ell(x) = \|x\|^{-\alpha}$ for a stationary transmitter point process $\Phi \subset \mathbb{R}^d$ of intensity λ . Since the expectation is taken over the point process, it follows from stationarity that $\mathbb{E}I(y) = \mathbb{E}I(o)$ (in fact, $I(y) \stackrel{d}{=} I(o)$) for all $y \in \mathbb{R}^d$, so we can focus on the origin, where

$$\mathbb{E}I = \mathbb{E}\left(\sum_{x \in \Phi} \|x\|^{-\alpha}\right).$$

By Campbell's formula for sums (Theorem 4.1),

$$\mathbb{E}I = \lambda \int_{\mathbb{R}^d} \|x\|^{-\alpha} dx.$$

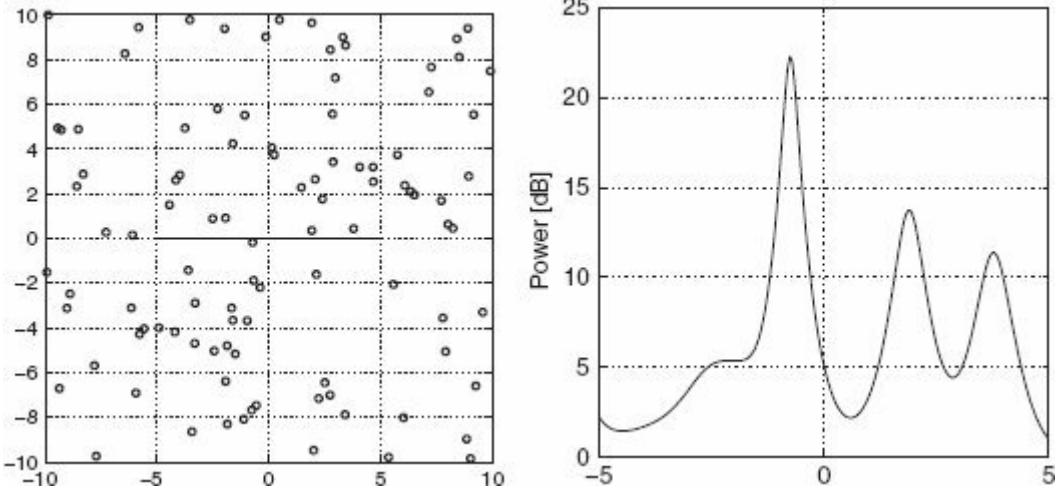


Figure 5.1 (Left) Realization of a uniform PPP of intensity 1/4. (Right) Interference field along the line segment $(-5, 0)$ to $(5, 0)$ in the realization on the left for a path loss law $\ell(x) = \|x\|^{-3}$. Since the interference varies by two orders of magnitude, it is shown in dB. The peak near 0 is due to the proximity of a node to the line segment near the origin.

In two dimensions,

$$\lambda \int_{\mathbb{R}^2} \|x\|^{-\alpha} dx = \lambda \int_0^{2\pi} \int_0^\infty r^{-\alpha} r dr d\varphi = \begin{cases} 2\pi \frac{\lambda}{2-\alpha} r^{2-\alpha} \Big|_0^\infty, & \alpha \neq 2, \\ 2\pi \lambda \log r \Big|_0^\infty, & \alpha = 2. \end{cases}$$

For the mean to be finite, we need $2 - \alpha < 0$ due to the upper integration bound and $2 - \alpha > 0$ due to the lower one; for $\alpha = 2$, it does not converge anyway. So *the mean interference for the power path loss law is infinite in two dimensions for all values of the path loss exponent*.

In three dimensions, we find

$$\mathbb{E}I = \lambda \int_{\mathbb{R}^3} \|x\|^{-\alpha} dx = 4\pi \frac{\lambda}{3-\alpha} r^{3-\alpha} \Big|_0^\infty, \quad \alpha \neq 3.$$

Again, there is no value of α for which the integral converges. We may conjecture that this holds in any number of dimensions. To investigate further, we focus on the uniform PPP and apply the mapping theorem (Theorem 2.34) to project it onto one

dimension.

1. Map to one dimension using the function $f(x) = \|x\|$. The new point process Φ' has intensity measure $\Lambda'([0, r)) = \lambda c_d r^d$ and intensity function $\lambda'(r) = \lambda c_d r^{d-1}$, where c_d is the volume of the d -dimensional unit ball.
2. Apply Campbell's theorem for the mean of a sum:

$$\begin{aligned}\mathbb{E}I &= \mathbb{E} \left(\sum_{r \in \Phi'} r^{-\alpha} \right) = \int_0^\infty r^{-\alpha} \lambda'(r) dr \\ &= \lambda c_d \frac{d}{d-\alpha} r^{d-\alpha} \Big|_0^\infty, \quad \alpha \neq d.\end{aligned}$$

So, for all d , there is no α for which the mean exists. If $\alpha \leq d$, the upper bound is the culprit, which means that all the interferers far away contribute most of the interference. In other words, the mean interference becomes infinite since the number of nodes is infinite. If $\alpha \geq d$, the problem is the lower integration bound, indicating that the nodes very close to the origin cause high interference due to the singularity of the path loss law.

While the mean interference diverges both for $\alpha \leq d$ and for $\alpha \geq d$, the two types of divergence are qualitatively very different. If $\alpha > d$, condition (4.5) is met, so the interference is finite a.s. On the other hand, if $\alpha \leq d$, (4.5) is violated, indicating that the interference is infinite a.s. So in the first case we can expect the interference to have a well-defined distribution (albeit with a heavy tail, so that its mean diverges), whereas in the second case its distribution does not exist, since we have not only $\mathbb{E}I = \infty$, but also $I = \infty$ a.s.

The divergence of the mean due to the nearby interferers clearly is a modeling artefact, since no receiver ever gets more power than was transmitted. If the path loss function is replaced by a more accurate one, for example

$$\ell(x) = \min\{1, \|x\|^{-\alpha}\} \quad \text{or} \quad \ell(x) = (1 + \|x\|)^{-\alpha},$$

the issue at $r \rightarrow 0$ is solved, condition (4.5) is always satisfied, and the mean interference is finite as soon as $\alpha > d$.

Example 5.1 Calculate the mean interference for the path loss law $\ell(x) = \min\{1, \|x\|^{-\alpha}\}$ for a uniform PPP of intensity λ and for a PPP of intensity $\lambda \mathbf{1}(\|x\| < R)$.

Solution

The mean interference in a PPP of intensity λ on the plane is

$$\mathbb{E}I = \lambda\pi + \frac{2\lambda\pi}{\alpha-2}, \quad \alpha > 2.$$

If the diameter of the network is bounded to $R > 1$,

$$\mathbb{E}I = \lambda\pi + \frac{\lambda\pi}{\alpha-2}(1 - R^{2-\alpha}), \quad \forall \alpha > 0.$$

Remark 5.1

- Why can we not use $\|x\|^{-\alpha}$ directly as the mapping function? The reason is that the resulting point process would no longer be locally finite: We would have $\Lambda'([0, 1]) = \Lambda([1, \infty]) = \infty$, which violates the finiteness of the mean measure.
- The mean is identical for all point processes with the same intensity. Also, due to stationarity, the interference is the same no matter where on \mathbb{R}^d we measure. We will discuss in [Chapter 8](#) how to determine the interference at a point or near a point of the process, rather than at an arbitrary location.

In the context of wireless networks, the uniform PPP is often termed a *Poisson network*.

5.1.3 Variance of the interference in Poisson networks

In the Poisson case, we can apply Corollary [4.8](#) to find the variance of the interference. Let $\ell(x) = \min\{r_0^{-\alpha}, \|x\|^{-\alpha}\}$ for $r_0 > 0$. For a homogeneous PPP on \mathbb{R}^d , we have

$$\begin{aligned} \text{var } I &= \lambda \int_{\mathbb{R}^d} \ell^2(x) dx = \lambda c_d r_0^{d-2\alpha} + \frac{\lambda c_d d}{d-2\alpha} r_0^{d-2\alpha} \Big|_{r_0}^{\infty} \\ &= \lambda c_d r_0^{d-2\alpha} \left(\frac{2\alpha}{2\alpha-d} \right), \quad \text{for } 2\alpha > d. \end{aligned}$$

The condition $2\alpha > d$ is usually satisfied. If the network is finite and we let $r_0 \rightarrow 0$, we would need $\alpha < d/2$ for finite variance, a range of α that is very unlikely.

5.1.4 Interference from a nearby transmitter in Poisson networks

In some cases we may be interested in the interference from just a single nearby transmitter. We focus first on the nearest transmitter to the origin and denote its interference by I_1 . In fact, this signal power may actually represent the desired signal, since it is coming from a nearby node; in this case, I_1 would be the desired

power.

Using the distance to the nearest point to the origin calculated in Example 2.11, we find

$$\mathbb{P}(I_1 \leq x) = \mathbb{P}(R^{-\alpha} \leq x) = \mathbb{P}(R > x^{-1/\alpha}) = \exp(-\lambda c_d x^{-\delta}),$$

where $\delta \triangleq d/\alpha$. The mean is

$$\mathbb{E}I_1 = c_d^{1/\delta} \Gamma(1 - 1/\delta).$$

If $\delta < 1$ then this does not exist. So, with the singular path loss law, the mean interference from only the nearest node is already infinite. This is due to the heavy tail of the distribution for small δ :

$$\mathbb{P}(I_1 > x) \sim \lambda c_d x^{-\delta}, x \rightarrow \infty.$$

If $\delta < 1$, $\int \mathbb{P}(I_1 > x) dx$ diverges and thus the mean does not exist. Generally, $\mathbb{E}(I_1^p)$ exists for $p < \delta$.

Analogously, we have for the pdf

$$f_{I_1}(x) \sim \lambda c_d \delta x^{-\delta-1}, x \rightarrow \infty.$$

Next we generalize the expression to the interference I_n from the n th nearest interferer. The complementary cdf of the distance to the n th nearest neighbor R_n is from (2.12)

$$\mathbb{P}(R_n > r) = \frac{\Gamma_{ic}(n, \lambda c_d r^d)}{\Gamma(n)}.$$

So, for $n = 2$,

$$\mathbb{P}(I_2 < x) = \exp(-\lambda c_d x^{-\delta})(1 + \lambda c_d x^{-\delta})$$

and

$$\mathbb{P}(I_2 > x) \sim \frac{1}{2}(\lambda c_d)^2 x^{-2\delta}.$$

So we need $2\delta > 1$ for $\mathbb{E}I_2$ to exist. For general n ,

$$\mathbb{P}(I_n < x) = \exp(-\lambda c_d x^{-\delta}) \sum_{i=0}^{n-1} \frac{(\lambda c_d x^{-\delta})^i}{i!}.$$

For the tail probability we need to sum from n to ∞ , so the dominant term will be the one for $i = n$ as $x \rightarrow \infty$. Therefore

$$\mathbb{P}(I_n > x) \sim \frac{1}{n!} (\lambda c_d)^n x^{-n\delta}.$$

This means that $\mathbb{E}(I_n^p)$ exists for $p < n\delta$. Hence if interference cancellation techniques are employed to achieve a finite second moment, $k > \alpha$ interferers need to be cancelled out in two-dimensional networks.

Although we can find the distribution of I_n for all n , it is difficult to obtain the distribution of the total interference this way, since the I_n are neither independent nor identically distributed. We proceed differently, treating the cases without and with fading separately.

5.1.5 Interference distribution in Poisson networks without fading

In this subsection we focus on the case of two-dimensional networks and assume there is no fading. Since $\ell(x)$ is assumed isotropic, we also use its one-dimensional version $\tilde{\ell} : \mathbb{R}^+ \mapsto \mathbb{R}^+$, so that $\tilde{\ell}(\|x\|) \equiv \ell(x)$. It is assumed that $\tilde{\ell}(x)$ is strictly monotonically decreasing (invertible), and that $\lim_{r \rightarrow \infty} \tilde{\ell}(r) = 0$.

Our goal is to find the characteristic function of the interference and from that, if possible, the distribution. We follow a basic yet powerful technique, used in Sousa & Silvester (1990). It consists of two steps.

1. Consider first a finite network, say on a disk of radius a centered at the origin, and condition on having a fixed number of nodes in this finite area. The nodes' locations are then iid.
2. Then de-condition on the (Poisson) number of nodes and let the disk radius go to infinity.

Step 1. Consider the interference from the nodes located within distance a of the origin:

$$I_a = \sum_{x \in \Phi \cap b(o, a)} \tilde{\ell}(\|x\|).$$

In the limit $a \rightarrow \infty$, $I_a \rightarrow I$. Let \mathcal{F}_{I_a} be the characteristic function (Fourier transform) of I_a , i.e.,

$$\mathcal{F}_{I_a}(\omega) \triangleq \mathbb{E}(e^{j\omega I_a}), j = \sqrt{-1}.$$

Conditioning on having k nodes in the disk of radius a ,

$$\mathcal{F}_{I_a}(\omega) = \mathbb{E}(\mathbb{E}(e^{j\omega I_a} | \Phi(b(o, a)) = k)).$$

Given that there are k points in $b(o, a)$, these points are independent and uniformly distributed on the disk with radial density

$$f_R(r) = \begin{cases} 2r/a^2 & \text{if } 0 \leq r \leq a \\ 0 & \text{otherwise,} \end{cases}$$

and the characteristic function is the product of the k individual characteristic functions:

$$\mathbb{E}(e^{j\omega I_a} | \Phi(b(o, a)) = k) = \left(\int_0^a \frac{2r}{a^2} \exp(j\omega \tilde{\ell}(r)) dr \right)^k. \quad (5.2)$$

Step 2. The probability of finding k nodes in $b(o, a)$ is given by the Poisson distribution, hence

$$\mathcal{F}_{I_a}(\omega) = \sum_{k=0}^{\infty} \frac{\exp(-\lambda\pi a^2)(\lambda\pi a^2)^k}{k!} \mathbb{E}(e^{j\omega I_a} | \Phi(b(o, a)) = k). \quad (5.3)$$

Inserting (5.2), summing over k , and interpreting the sum as the Taylor expansion of the exponential function, we obtain

$$\mathcal{F}_{I_a}(\omega) = \exp \left(\lambda\pi a^2 \left(-1 + \int_0^a \frac{2r}{a^2} \exp(j\omega \tilde{\ell}(r)) dr \right) \right).$$

Integration by parts, substituting $r \rightarrow \tilde{\ell}^{-1}(x)$, where $\tilde{\ell}^{-1}$ is the inverse of $\tilde{\ell}$, and letting $a \rightarrow \infty$ yields

$$\lim_{a \rightarrow \infty} a^2 \left(-1 + \int_0^a \frac{2r}{a^2} \exp(j\omega \tilde{\ell}(r)) dr \right) = \int_0^\infty (\tilde{\ell}^{-1}(x))^2 j\omega e^{j\omega x} dx$$

so that

$$\mathcal{F}_I(\omega) = \exp \left(j\lambda\pi\omega \int_0^\infty (\tilde{\ell}^{-1}(x))^2 e^{j\omega x} dx \right).$$

To get more concrete results, we need to specify the path loss law. For the standard power law $\tilde{\ell}(r) = r^{-\alpha}$, we obtain

$$\mathcal{F}_I(\omega) = \exp \left(j\lambda\pi\omega \int_0^\infty x^{-2/\alpha} e^{j\omega x} dx \right). \quad (5.4)$$

For $\alpha \leq 2$, the integral diverges, indicating that the interference is infinite almost

surely. For $\alpha > 2$,

$$\mathcal{F}_I(\omega) = \exp\left(-\lambda\pi\Gamma(1-2/\alpha)\omega^{2/\alpha}e^{-j\pi/\alpha}\right), \quad \omega \geq 0. \quad (5.5)$$

The values for negative ω are determined by the symmetry condition $\mathcal{F}_I^*(-\omega) = \mathcal{F}_I(\omega)$. For $\alpha = 4$,

$$\mathcal{F}_I(\omega) = \exp\left(-\lambda\pi^{3/2} \exp(-j\pi/4)\sqrt{\omega}\right).$$

This case is of particular interest, since it is the only one where a closed-form expression for the density exists:

$$f_I(x) = \frac{\pi\lambda}{2x^{3/2}} \exp\left(-\frac{\pi^3\lambda^2}{4x}\right). \quad (5.6)$$

This is the so-called Lévy distribution, which can also be viewed as an inverse gamma distribution, or as the inverse Gaussian distribution with infinite mean. For other values of α , the densities may be expressed in an infinite series (Sousa & Silvester 1990, Eqn. (22)).

5.1.6 Stable distributions

In order to interpret the characteristic function (5.5), we briefly introduce here the class of *stable distributions*.

DEFINITION 5.1 (Stable distribution) A random variable X is said to have a *stable* distribution if for all $a, b > 0$, there exists $c > 0, d$, such that

$$aX_1 + bX_2 \stackrel{d}{=} cX + d, \quad (5.7)$$

where X_1 and X_2 are iid with the same distribution as X . If (5.7) holds with $d = 0$, the distribution is *strictly stable*.

THEOREM 5.2 *For any stable random variable X , there is a parameter δ in the range $0 < \delta \leq 2$ such that the number c in the definition satisfies $c^\delta = a^\delta + b^\delta$.*

The parameter δ is the *characteristic exponent*, which is sometimes also referred to as the index of stability. For Gaussian random variables, $\delta = 2$ since $aX_1 + bX_2$ is Gaussian with mean $(a + b)\mu$ and variance $(a^2 + b^2)\sigma^2$, i.e., the definition holds with $c = \sqrt{a^2 + b^2}$ and $d = (a + b - c)\mu$. Generally, the characteristic function of a stable distribution is given by

$$\mathbb{E}(e^{itX}) = \begin{cases} \exp(jt\mu - \gamma|t|^\delta(1 - j\beta \operatorname{sgn}(t)\tan(\pi\delta/2))) & \delta \neq 1 \\ \exp(jt\mu - \gamma|t|(1 + j(2\beta/\pi)\operatorname{sgn}(t)\log(|t|))) & \delta = 1, \end{cases} \quad (5.8)$$

where $\operatorname{sgn}(t)$ is the sign of t , $\beta \in [-1, 1]$ is the skew parameter, μ is the drift, and γ is the dispersion parameter. On comparing (5.8) with (5.5), we observe that (5.5) is indeed the characteristic function of a stable random variable, with characteristic exponent $\delta = 2/\alpha$, skew 1, drift 0, and dispersion

$$\gamma = \frac{\lambda\pi\Gamma(1 - 2/\alpha)}{\cos(\pi/\alpha)} = \frac{\lambda\pi\Gamma(1 - \delta)}{\cos(\pi\delta/2)}.$$

For $\alpha = 4$, in particular, $\gamma = \lambda\pi^{3/2}/\sqrt{2}$. As mentioned previously, this case, where $\delta = 1/2$, is the only case for which a pdf of the interference power exists.¹

Skew 1 means that the support of the random variable is restricted to \mathbb{R}^+ , which of course must be the case since we are considering the interference power.

The Laplace transform of a stable random variable X with skew parameter 1 and drift 0 is of the compact form

$$\mathbb{E}(e^{-sX}) = e^{-\kappa s^\delta}. \quad (5.9)$$

The parameter κ is related to γ by $k = \gamma \cos(\pi/\alpha)$, and we will also refer to it as dispersion or the scale parameter. This does not violate common terminology, since, unfortunately, there exist many different definitions of dispersion and scale for stable distributions. In the form (5.9), the dispersion equals half the variance for Gaussian random variables. For other stable distributions ($\delta < 2$), the variance does not exist, but the dispersion plays a similar role, since it indicates how spread the distribution is.

The theory of stable distributions includes a result that relates PPPs and stable distributions.

THEOREM 5.3 (Series representation) *Let $\{\tau_i\} \subset \mathbb{R}$ be the arrival times of a PPP of intensity λ and let (h_i) be a family of iid random variables that are independent of τ_i . If the infinite sum*

$$\sum_{i=1}^{\infty} \tau_i^{-1/\delta} h_i \quad (5.10)$$

converges a.s., then it converges to a stable random variable.

The representation (5.10) is called the LePage series representation.

This is consistent with what we found, since, in our case, the distances raised to the d th power, $r^d_i = \|x_i\|^d$, constitute a homogeneous PPP of intensity λc_d . So our sum is

$$\sum_{i=1}^{\infty} (r_i^d)^{-\alpha/d},$$

and $\delta = d/\alpha$.

A consequence of Theorem 5.3 is the following.

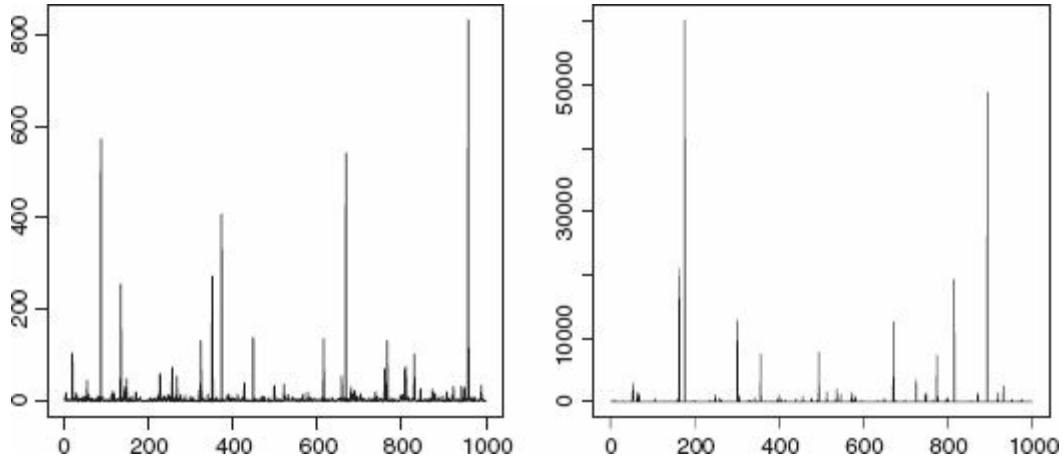


Figure 5.2 Sequence of 1000 realizations of stable random variables with dispersion 1. (Left) $\delta = 0.9$. (Right) $\delta = 0.5$. Typical for stable random variables with characteristic exponents less than 1 is that they assume huge values rarely. These peaks become more dominant as δ becomes small.

COROLLARY 5.4 (Interference scaling) *Let $I(\lambda)$ be the interference in a uniform PPP $\Phi \subset \mathbb{R}^d$ of intensity λ with path loss function $\ell(x) = \|x\|^{-\alpha}$. Then*

$$aI(\lambda) \stackrel{d}{=} I(a^\delta \lambda), \quad \forall a > 0, \text{ where } \delta = d/\alpha.$$

This may appear counter-intuitive, but it is consistent with (5.7). Since $I(\lambda_1 + \lambda_2) \stackrel{d}{=} I(\lambda_1) + I(\lambda_2)$ from the superposition property of PPPs, we have from the corollary

$$aI_1(\lambda) + bI_2(\lambda) \stackrel{d}{=} I_1(a^\delta \lambda) + I_2(b^\delta \lambda) \stackrel{d}{=} I((a^\delta + b^\delta)\lambda) \stackrel{d}{=} cI(\lambda),$$

where $c = (a^\delta + b^\delta)^{1/\delta}$ and I , I_1 , and I_2 stem from independent PPPs of the respective densities.

From the corollary we can obtain directly the Laplace transform of the interference if the transmit power is scaled by P . Since $I \propto P$, we have $PI(\lambda) = I(P^\delta \lambda)$, i.e., scaling the power by P has the same effect as scaling the density by P^δ . Conversely, if the density is scaled by a factor a , the interference is scaled by $a^{1/\delta}$. So, perhaps counter-intuitively, *the interference is not proportional to the density λ if it has a stable distribution*.

Stable random variables with $\delta < 1$ have the property that the arithmetic mean

$(X_1 + \dots + X_n)/n$ for iid X_k increases with n and is equal in distribution to $X_k n^{1/\delta-1}$. The sum and the maximum of n such stable random variables scale in the same way, i.e., the sum is dominated by the occasional large value.

Figure 5.2 shows realizations of stable random variables that model the interference in a PPP with power-law path loss. The corresponding (complementary) distribution functions are displayed in Fig. 5.3.

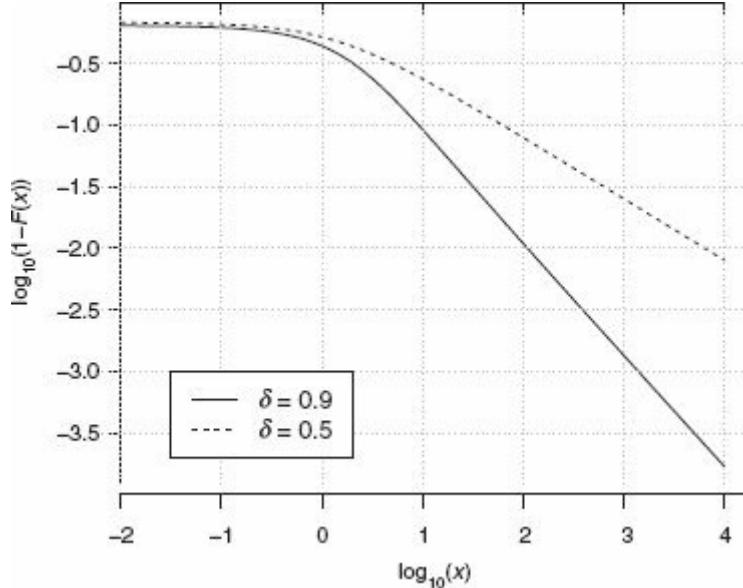


Figure 5.3 Complementary distribution functions of stable random variables with dispersion 1 on a log–log scale. This indicates, for example, that for $\delta = 1/2$, one out of every 100 realizations exceeds 10^4 (which is consistent with Fig. 5.2 (right)), while for $\delta = 9/10$, it is one out of 5000. On the other hand, for $\delta = 1/2$, 11 of 12 realizations are smaller than 100, while for $\delta = 9/10$, 99 out of 100 are below 100. The asymptotic slope is $-\delta$.

The corresponding Laplace transform is

$$\mathcal{L}_I(s) = \exp(-\lambda\pi\Gamma(1 - 2/\alpha)s^{2/\alpha}). \quad (5.11)$$

Stable distributions with characteristic exponents less than unity do not have any finite moments. In particular, the mean interference diverges, which is due to the singularity of the path loss law at the origin. This also follows immediately from the fact that $\mathbb{E}(I) = -(\text{d}/\text{d}s)\log(\mathcal{L}_I(s))|_{s=0} = \lim_{s \rightarrow 0} cs^{2/\alpha-1} = \infty$.

The method of conditioning on a fixed number of nodes, using the iid property of the node locations, and de-conditioning with respect to the Poisson distribution is applicable to many other problems.

5.1.7 Interference distribution with fading

With fading, the power from each transmitter x gets multiplied by a fading

coefficient h_x , assumed iid. Hence the interference is given by the sum

$$I = \sum_{x \in \Phi} h_x \ell(x),$$

and we aim at calculating its Laplace transform

$$\mathcal{L}(s) = \mathbb{E} e^{-sI} = \mathbb{E} \left(\prod_{x \in \Phi} e^{-sh\ell(x)} \right).$$

Since the fading is iid,

$$\mathcal{L}(s) = \mathbb{E}_\Phi \left(\prod_{x \in \Phi} \mathbb{E}_h (e^{-sh\ell(x)}) \right).$$

Mapping the PPP to one dimension, we know that $\lambda(r) = \lambda c_d dr^{d-1}$. Again let $\tilde{\ell}(\|x\|) \equiv \ell(x)$. Now we use the pgfl for $v(r) = \mathbb{E}_h (e^{-sh\ell(r)})$ to obtain

$$\mathcal{L}(s) = \exp \left\{ - \int_0^\infty \mathbb{E}_h [1 - e^{-sh\tilde{\ell}(r)}] \lambda(r) dr \right\}$$

for the one-dimensional PPP. Conditioned on h , we have

$$\begin{aligned} \int_0^\infty (1 - \exp(-sh\tilde{\ell}(r))) \lambda(r) dr &= \lambda c_d \int_0^\infty (1 - \exp(-sh r^{-\alpha})) dr^{d-1} dr \\ &\stackrel{(a)}{=} \lambda c_d \int_0^\infty (1 - \exp(-sh/y)) \delta y^{\delta-1} dy \\ &\stackrel{(b)}{=} \lambda c_d \int_0^\infty (1 - \exp(-shx)) \delta x^{-\delta-1} dx \\ &\stackrel{(c)}{=} \lambda c_d \int_0^\infty x^{-\delta} sh \exp(-shx) dx \\ &= \lambda c_d (hs)^\delta \Gamma(1-\delta), \quad 0 < \delta < 1, \end{aligned}$$

where (a) follows from the substitution $y \leftarrow r^{-1/\alpha}$, (b) from $x \leftarrow y^{-1}$, and (c) from integration by parts. With the expectation over h , we obtain

$$\mathcal{L}(s) = \exp(-\lambda c_d \mathbb{E}(h^\delta) \Gamma(1-\delta) s^\delta).$$

So with fading the interference also has a stable distribution with characteristic exponent δ . The dispersion κ changes by a factor $\mathbb{E}(h^\delta)$ to $\lambda c_d \mathbb{E}(h^\delta) \Gamma(1-\delta)$.

In the Rayleigh fading case, where h is exponential, $\mathbb{E}(h^\delta) = \Gamma(1+\delta)$, and

$$\begin{aligned}
\mathcal{L}(s) &= \exp(-\lambda c_d \Gamma(1+\delta) \Gamma(1-\delta) s^\delta) \\
&= \exp\left(-\lambda c_d \frac{\pi \delta}{\sin(\pi \delta)} s^\delta\right) \\
&= \exp\left(-\frac{\lambda c_d}{\text{sinc } \delta} s^\delta\right).
\end{aligned}$$

As $\delta \rightarrow 1$, $\sin(\pi\delta) \sim \pi(1-\delta)$, so in the limit we have

$$\mathcal{L}(s) \approx \exp\left(-\lambda c_d s^\delta \frac{\delta}{1-\delta}\right),$$

which indicates that the dispersion increases sharply as the path loss exponent approaches the number of network dimensions (from above).

5.2 Outage probability in Poisson networks

Definitions

DEFINITION 5.5 (Signal-to-interference-plus-noise ratio (SINR)) The SINR is

$$\text{SINR} \triangleq \frac{S}{W+I}, \quad (5.12)$$

where S is the desired signal power, W is the noise power, and I is the interference power. If noise is neglected ($W=0$), the SINR reduces to the signal-to-interference ratio SIR.

For a fixed modulation and coding scheme and with interference treated as noise, e.g., by using a simple linear receiver, a well-accepted model for packetized transmissions is that they succeed if the SINR exceeds a certain threshold θ . So we define the success probability as follows.

DEFINITION 5.6 (Transmission success probability)

$$p_s(\theta) \triangleq \mathbb{P}(\text{SINR} > \theta). \quad (5.13)$$

Its complement $1 - p_s$ is the *outage probability*.

The threshold θ is related to the (physical-layer) rate of transmission R . From Shannon's capacity formula, the two are related by $\theta = 2^R - 1$; due to practical limitations, however, the threshold has to be chosen slightly larger in order to accommodate rate R .

Outage in a Poisson network with Rayleigh fading

Here we assume that the receiver under consideration is attempting to decode a

message from a desired transmitter at distance r . With Rayleigh fading, the desired signal strength S at the receiver is exponential with mean $r^{-\alpha}$. Let $r = 1$ first. What is the success probability $\mathbb{P}(\text{SIR} > \theta)$ in a Poisson field of interferers of intensity λ that are subject to Rayleigh fading?

This is exactly the Laplace transform:

$$p_s \triangleq \mathbb{P}(\text{SIR} > \theta) = \mathbb{P}(S > I\theta) = \mathbb{E}_I(\exp(-\theta I)) = \exp(-c_d \lambda \theta^\delta \Gamma(1 + \delta) \Gamma(1 - \delta)).$$

If the desired link has distance r ,

$$p_s(r) = \mathbb{E}(\exp(-\theta r^\alpha I)) = \exp(-c_d \lambda \theta^\delta r^d \Gamma(1 + \delta) \Gamma(1 - \delta)). \quad (5.14)$$

So, in two dimensions, the success probability decays only with r^2 (inside the exponential), although the path loss is r^α .

The outage probability is the cdf of the SIR – so, quite remarkably, we have a closed-form expression for the SIR distribution in Poisson networks with Rayleigh fading! We state this result formally.

Box 5.1 Laplace transform of interference and success probability

If the desired signal power in a wireless network is exponentially distributed, such as in Rayleigh fading, the success probability of the transmission is given by the Laplace transform of the interference measured at the receiver. The Laplace transform of the interference is a probability generating functional of the point process of interferers.

THEOREM 5.7 (SIR distribution in Poisson networks with Rayleigh fading) *The pdf is*

$$f_{\text{SIR}}(x) = c \delta x^{\delta-1} e^{-cx^\delta},$$

where $c = c_d \lambda r^d \Gamma(1 + \delta) \Gamma(1 - \delta)$. This is a Weibull distribution. Its mean is

$$\mathbb{E}(\text{SIR}) = \frac{\Gamma(1/\delta)}{\delta c^{1/\delta}} = \frac{1}{r^\alpha} \frac{\Gamma(1 + 1/\delta)}{(\lambda C(\alpha))^{1/\delta}}, \quad C(\alpha) = c_d \Gamma(1 + \delta) \Gamma(1 - \delta).$$

While the success probability is a function of r^d , the mean SIR is proportional to $r^{-\alpha}$, as expected, since the received signal power S decays with $r^{-\alpha}$, while the interference does not depend on r .

The outage expression is valid also when the interferers are not subject to Rayleigh fading. As long as the desired signal is exponentially distributed, we have

$$p_s = \mathbb{E}(\exp(-\theta I)) = \exp(-c_d \lambda \theta^\delta r^d \mathbb{E}(h^\delta) \Gamma(1 - \delta)),$$

where the interferers' fading distribution can be arbitrary (as long as it has a finite δ th moment).

Including noise is straightforward. With Rayleigh fading, noise simply adds an exponential factor to the success probability. Denoting the success probability with noise and interference by p_s^{W+I} ,

$$p_s^{W+I} = \mathbb{P}(S > \theta(W + I)) = \mathbb{E}(\exp(-\theta r^\alpha (W + I))) = \exp(-\theta r^\alpha W) p_s^I.$$

With transmit power P , W is replaced by W/P . The interference-only success probability does not depend on the (common) transmit power.

5.3 Spatial throughput in Poisson bipolar networks

DEFINITION 5.8 (Poisson bipolar network) A Poisson bipolar network consists of a PPP of transmitters $\{x_i\} \subset \mathbb{R}^2$ and a set of receivers $\{y_i\}$ at fixed distance r and uniformly randomly chosen orientation from their transmitters. Hence $\|x_i - y_i\| = r$ for all i .

By the displacement theorem (Theorem 2.33), the point process of receivers $\{y_i\}$ is itself a PPP. So the Poisson bipolar network consists of two *dependent* PPPs, namely a transmitter PPP and a receiver PPP.

5.3.1 Spatial throughput

The success probability p_s can be viewed as a throughput metric for a particular link. To quantify the performance of an entire network, the throughput needs to be averaged over all links. Assume that each transmitter in a Poisson bipolar network decides to transmit with probability p and stay silent with probability $1 - p$. This is the so-called ALOHA channel access scheme. In a Poisson bipolar network of intensity λ , the spatial throughput may be defined as

$$T(p) \triangleq p \lambda p_s(p, r),$$

where $p_s(p, r)$ is the success probability of a transmission over distance r for transmit probability p . For a given r , this is maximized at

$$p_{\text{opt}} = \frac{1}{\lambda \theta^\delta \pi \Gamma(1 - \delta) \Gamma(1 + \delta) r^2}. \quad (5.15)$$

Of course, p_{opt} cannot exceed 1. If the expression on the right side yields a value larger than 1, this indicates that the network density λ could be increased to achieve a higher throughput. Similarly, the link distances r could be included in the metric, which would result in a transport capacity, and optimized.

5.3.2 Shannon throughput

The spatial throughput is an outage-based metric, since some transmissions will not be successful. In contrast, if the transmitter can react quickly to the SINR condition and adjust its rate of transmission, another quantity may be relevant, termed the *Shannon throughput* in Baccelli & Blaszczyszyn (2009). Considering $p_s(\theta)$ as a function of the threshold θ , the Shannon throughput (expressed in nats²) is

$$\begin{aligned} \mathbb{E} \log(1 + \text{SINR}) &= - \int_0^\infty \log(1 + \theta) dp_s(\theta) \\ &\stackrel{(a)}{=} \int_0^\infty p_s(e^x - 1) dx. \end{aligned}$$

Here (a) follows since the complementary cdf of the random variable $X = \log(1 + \text{SINR})$ is $\mathbb{P}(X > x) = \mathbb{P}(\text{SINR} > e^x - 1) = p_s(e^x - 1)$ and the expectation of a random variable with support \mathbb{R}^+ can be expressed as the integral over the complementary cdf. Multiplied by the density of concurrent transmitters, this yields another type of spatial throughput – the best one could hope for while treating interference as noise.

5.4 Transmission capacity

The transmission capacity, which was introduced in Weber *et al.* (2005), is a performance metric for wireless networks that measures the spatial intensity of successful transmissions, subject to a constraint on the permissible outage probability (where outage occurs when the SINR at a receiver is below a threshold). Bounding the outage probability has the advantage of enforcing a certain quality of the links, which is necessary in many applications. In contrast, maximizing the spatial throughput as in (5.15) may result in success probabilities as low as e^{-1} , which implies several retransmissions of each packet and, in turn, unacceptable delays. Also, in contrast to other metrics, the transmission capacity is relatively tractable; in certain cases it results in simple closed-form expressions, while in many others tight bounds can be found.

DEFINITION 5.9 (Transmission capacity) To a Poisson bipolar network of intensity λ , add a reference receiver at the origin and a desired transmitter at distance r . Let $p_s(\lambda)$ denote the success probability at this receiver in the presence of interference from all transmitters in the bipolar network. Given a target outage probability $\epsilon \in (0, 1)$, the *transmission capacity* λ_{tc} is defined as

$$\lambda_{tc}(\epsilon) \triangleq (1 - \epsilon) \max \{ \lambda : p_s(\lambda) \geq 1 - \epsilon \}$$

Since the success probability $p_s(\lambda)$ is strictly monotonically decreasing in λ , the transmission capacity may be expressed as

$$\lambda_{tc}(\epsilon) = (1 - \epsilon) p_s^{-1}(1 - \epsilon)$$

where $p_s^{-1}(1 - \epsilon)$ yields the intensity λ that results in a success probability $1 - \epsilon$.

Example 5.2 Find the transmission capacity of a two-dimensional network if all transmitters use unit power, all links are subject to Rayleigh fading, and noise is neglected.

Solution

From (5.14), we know that

$$p_s(\lambda) = \exp(-\pi\lambda\theta^\delta r^2\Gamma(1 + \delta)\Gamma(1 - \delta)),$$

where $\delta = 2/\alpha$. Defining $\gamma \triangleq \pi\theta^\delta r^2\Gamma(1 + \delta)\Gamma(1 - \delta)$ we obtain

$$\lambda_{tc}(\epsilon) = -\frac{1}{\gamma}(1 - \epsilon)\log(1 - \epsilon)$$

by inverting $p_s(\lambda)$. Since ϵ is usually small, we may use a Taylor expansion of the logarithm to obtain

$$\begin{aligned}\lambda_{tc}(\epsilon) &= \frac{\epsilon - \epsilon^2/2}{\gamma} + O(\epsilon^3) \\ &= \frac{\epsilon}{\gamma} + O(\epsilon^2),\end{aligned}$$

which gives good upper bounds as $\epsilon \rightarrow 0$.

The parameter γ introduced in the example is called the *spatial contention*

parameter in Haenggi (2009). It is in general defined as the (negative) slope of the success probability as a function of the transmitter intensity λ at $\lambda = 0$,

$$\gamma \triangleq -\frac{dp_s(\lambda)}{d\lambda} \Big|_{\lambda=0},$$

and thus captures how quickly the success probability decays as more links are spatially packed in the network. So, generally, the asymptotic behavior of the transmission capacity is always

$$\lambda_{tc}(\epsilon) \sim \frac{\epsilon}{\gamma}, \quad \epsilon \rightarrow 0.$$

The transmission capacity is not monotonically increasing with ϵ for larger ϵ , though, due to the factor $1 - \epsilon$; it reaches a maximum of $1/(e\gamma)$ at $\epsilon = 1 - 1/e$.

As a function of the link distance r , it is apparent that $\lambda_{tc} \propto r^{-2}$, which reflects the fact that a transmission over distance r requires an area proportional to r^2 if it is subject to an outage constraint.

5.4.1 Networks without fading

Without fading, there is one case for which the transmission capacity can be calculated in quasi-closed-form. This case is the one where the path loss exponent is $\alpha = 4$, for which we have found the pdf of the interference in (5.6). The corresponding cdf is

$$\mathbb{P}(I \leq x) = 1 - \text{erf}\left(\frac{\pi^{3/2}\lambda}{2\sqrt{x}}\right),$$

where erf denotes the error function

$$\text{erf } x \triangleq \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

The desired signal power is r^{-4} , thus

$$p_s(\lambda) = \mathbb{P}(r^{-4} > \theta I) = \mathbb{P}(I < 1/(\theta r^4)) = 1 - \text{erf}\left(\frac{\pi^{3/2}\lambda r^2 \sqrt{\theta}}{2}\right).$$

It follows that

$$\lambda_{tc}(\epsilon) = \frac{2}{\pi^{3/2} r^2 \sqrt{\theta}} \text{erf}^{-1}(\epsilon).$$

The spatial contention in this case is $\gamma = \pi r^{2\sqrt{\theta}}$ thus

$$\lambda_{tc}(\epsilon) \sim \frac{\epsilon}{\pi r^2 \sqrt{\theta}}, \quad \epsilon \rightarrow 0. \quad (5.16)$$

So the dependence on r and θ is exactly as in the case with Rayleigh fading, the only change in the asymptotic expression is in the pre-constant.

For other path loss exponents, bounds can be derived. An upper bound on the success probability and thus the transmission capacity is obtained by considering only the dominant interferers, which are those that by themselves can cause an outage. If Φ denotes the PPP of all interferers, the dominant ones are

$$\Phi_{\text{dom}} \triangleq \Phi \cap b(o, r\theta^{1/\alpha}),$$

since the interference from any of these nodes is at least $(r\theta^{1/\alpha})^{-\alpha} = r^{-\alpha}/\theta$, which results in an SIR of less than θ . A necessary condition for transmission success is that there are no dominant interferers, so

$$p_s(\lambda) < \mathbb{P}(\Phi_{\text{dom}} = \emptyset) = \exp(-\lambda\pi r^2\theta^\delta)$$

and

$$\begin{aligned} \lambda_{tc}(\epsilon) &< \frac{-(1-\epsilon)\log(1-\epsilon)}{\pi r^2 \theta^\delta} \\ &= \frac{\epsilon}{\pi r^2 \theta^\delta} + O(\epsilon^2), \quad \epsilon \rightarrow 0. \end{aligned}$$

A comparison with (5.16) reveals that this bound becomes asymptotically tight for $\alpha = 4$. It is unclear, however, whether this is the case also for smaller α , where the contribution of the non-dominant nodes becomes more significant.

To find a lower bound, we decompose the outage event $\{S < \theta I\}$ as

$$\begin{aligned} \{S < \theta I\} &= \left\{ I_1 > S/\theta \right\} \cup \left\{ I_2 > S/\theta \right\} \\ &\cup \left\{ I_1 \leq S/\theta, I_2 \leq S/\theta, I_1 + I_2 > S/\theta \right\}, \end{aligned}$$

where I_1 is the interference from the dominant nodes and $I_2 = I - I_1$ is the interference from the non-dominant nodes. $S = r^{-\alpha}$ is the desired signal power. Now $I_1 < S/\theta$ occurs only if there are no dominant nodes, so $\{I_1 < S/\theta\} = \{I_1 = 0\}$, thus the probability of the third event is 0, and it follows that

$$\begin{aligned} \mathbb{P}(S < \theta I) &= \mathbb{P}(I_1 > S/\theta) + \mathbb{P}(I_2 > S/\theta) - \mathbb{P}(I_1 > S/\theta)\mathbb{P}(I_2 > S/\theta) \\ &= (1 - \exp(-\lambda\pi r^2\theta^\delta)) + \exp(-\lambda\pi r^2\theta^\delta)\mathbb{P}(I_2 > S/\theta). \end{aligned}$$

If we find an upper bound for the last probability $\mathbb{P}(I_2 > S/\theta)$, we have an upper bound on the outage probability and thus a lower bound on the success probability. A simple bound is obtained from the Markov inequality, which states that $\mathbb{P}(I_2 > S/\theta) \leq \theta r^\alpha \mathbb{E}(I_2)$. From Campbell's theorem for sums, we have

$$\mathbb{E}I_2 = \lambda 2\pi \int_{r\theta^{1/\alpha}}^{\infty} t^{-\alpha} t dt = \frac{2\pi \lambda r^{2-\alpha} \theta^{\delta-1}}{\alpha - 2}.$$

Hence $\mathbb{P}(I_2 > S/\theta) \leq 2\pi r^2 \theta^\delta / (\alpha - 2)$ and

$$\mathbb{P}(S < \theta I) \leq 1 - \exp(-\lambda \pi r^2 \theta^\delta) + \exp(-\lambda \pi r^2 \theta^\delta) \frac{2\pi r^2 \theta^\delta}{\alpha - 2}.$$

Using $1 - e^{-x} \leq x$ and $e^{-x} < 1$ for $x > 0$ yields for the success probability

$$\begin{aligned} p_s(\lambda) &> 1 - \lambda \pi r^2 \theta^\delta - \frac{2\lambda \pi r^2 \theta^\delta}{\alpha - 2} \\ &= 1 - \frac{\lambda \pi r^2 \theta^\delta \alpha}{\alpha - 2}. \end{aligned}$$

A lower bound on the transmission capacity follows as

$$\begin{aligned} \lambda_{tc}(\epsilon) &> (1 - \delta) \frac{\epsilon(1 - \epsilon)}{\pi r^2 \theta^\delta} \\ &= (1 - \delta) \frac{\epsilon}{\pi r^2 \theta^\delta} + O(\epsilon^2), \quad \epsilon \rightarrow 0. \end{aligned}$$

So the lower and upper bounds differ by a factor of $1 - \delta$. Tighter lower bounds can be obtained using the Chebyshev inequality or the Chernoff bound.

5.4.2 Networks with fading

For the case of independent Rayleigh fading, we have found the transmission capacity in closed form in Example 5.2. For general fading, we obtain a bound by again concentrating only on dominant interferers. In this case,

$$\Phi_{\text{dom}} = \{x \in \Phi : hr^{-\alpha}/h_x \|x\|^{-\alpha} < \theta\},$$

where h is the fading coefficient of the desired link and h_x is that of interferer x . As derived in Section 2.9.2, the intensity function of the (one-dimensional) path loss process Ξ with fading is

$$\mu(x) = \lambda \pi \delta x^{\delta-1} \mathbb{E}(h^\delta).$$

Hence, given h , the probability of the absence of any dominant interferer is the probability that there is no point of Ξ within $r^\alpha \theta/h$ of the origin, i.e.,

$$\mathbb{P}(\Phi_{\text{dom}} = \emptyset \mid h) = \exp\left(-\int_0^{r^\alpha \theta/h} \mu(x)dx\right) = \exp(-\lambda \pi \mathbb{E}(h^\delta) r^2 \theta^\delta h^{-\delta})$$

and thus

$$p_s(\lambda) < \mathbb{E}_h(\exp(-\lambda \pi \mathbb{E}(h^\delta) r^2 \theta^\delta h^{-\delta})).$$

Applying Jensen's inequality in the form

$$\mathbb{E}_h(\exp(-\lambda \pi \mathbb{E}(h^\delta) r^2 \theta^\delta h^{-\delta})) > \exp(-\lambda \pi \mathbb{E}(h^\delta) r^2 \theta^\delta \mathbb{E}(h^{-\delta}))$$

does not yield the desired upper bound since the inequality goes in the wrong direction. It provides, however, the approximation

$$p_s(\lambda) \approx \exp(-\lambda \pi r^2 \theta^\delta \mathbb{E}(h^\delta) \mathbb{E}(h^{-\delta}))$$

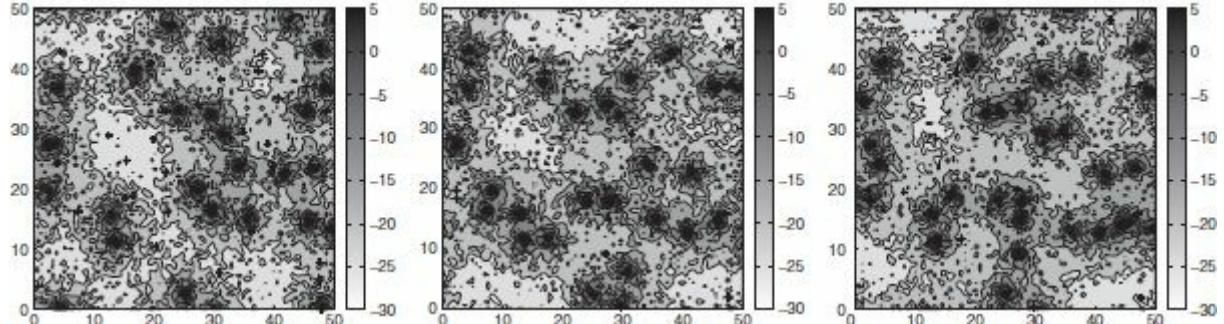


Figure 5.4 Contour plots of an interference random field in three different time slots. The interference power is measured in dB, as per the (gray-scale) color bar to the right of each plot. The network is formed by a Poisson point process of intensity $\lambda = 0.03$, indicated by the asterisks in all three plots. In each slot, a subset is chosen as transmitters with ALOHA probability $p = 1/3$. In the $[0, 50]^2$ square shown, there are $2500\lambda p = 25$ transmitters on average. The path loss law is r^{-3} , and fading is independent Rayleigh. Since the transmitters are chosen from a common set of nodes, the interference is spatially and temporally correlated despite the fading. For example, at the top left, the interference is always low since there are few nodes in that region.

and, in turn,

$$\lambda_{\text{tc}}(\epsilon) \approx \frac{-(1-\epsilon)\log(1-\epsilon)}{\pi r^2 \theta^\delta \mathbb{E}(h^\delta) \mathbb{E}(h^{-\delta})} \sim \frac{\epsilon}{\pi r^2 \theta^\delta \mathbb{E}(h^\delta) \mathbb{E}(h^{-\delta})}.$$

Interestingly, this is the *exact* expression for Rayleigh fading, i.e., the effects of upper bounding the success probability using dominant interferers and lower bounding it using Jensen's inequality precisely offset each other in this case.

Since $E(h^\delta)E(h^{-\delta}) > 1$ (except for the case of no fading), fading appears to have a negative effect on the transmission capacity.

5.5 Temporal correlation of the interference

Until now, we have characterized the interference only at a single location on the plane, and there was no temporal component to it. While the distribution of the interference does not depend on the location in a stationary point process, the interferences measured at two nearby locations are correlated, since they depend on the same point process and are thus subject to common randomness. So interference is *spatially correlated*. This is illustrated in Fig. 5.4, which shows three snapshots of an interference field in a Poisson network with ALOHA transmit probability $p = 1/3$.

Also, in a static network, a channel access or MAC scheme chooses a varying subset of nodes as transmitters. Say in time slot k , $\Phi_k \subset \Phi$ is the set of transmitters, and let I_k denote the interference at the origin measured at time k . Then I_k and I_j are correlated, again due to the common randomness Φ . Hence there is also *temporal correlation* in the interference field.

Both types of correlations, spatial and temporal, are worth exploring, since they affect the joint success probabilities of multi-hop communication and repetition-based transmission schemes, where a transmitter re-sends a message that was not received.

We focus on an interference field induced by a PPP of potential transmitters of intensity λ on \mathbb{R}^2 , which employ ALOHA as the MAC scheme. The path loss law $\ell(x)$ is assumed to have the following properties.

1. It depends only on $\|x\|$.
2. It monotonically decreases with $\|x\|$.
3. It is integrable, i.e.,

$$\int_{\mathbb{R}^2} \ell(x) dx < \infty. \quad (5.17)$$

For example, a valid path loss model is given by

$$\ell_\epsilon(x) = \frac{1}{\epsilon + \|x\|^\alpha}, \quad \epsilon \in (0, \infty), \quad \alpha > 2. \quad (5.18)$$

The standard singular path loss model $\ell(x) = \|x\|^{-\alpha}$ is obtained by taking the limit $\lim_{\epsilon \rightarrow 0} \ell_\epsilon(x)$. Since in ALOHA, each node makes an independent decision to transmit at every time instant, the set of transmitters at a time instant is a thinned version of the original PPP Φ , and the resulting transmitter process at time instant k , Φ_k , is again a PPP. This can be easily verified by considering the void probability. For any bounded set $B \subset \mathbb{R}^2$, we have

$$\begin{aligned}\mathbb{P}(\Phi_k(B) = 0) &= \mathbb{E} \prod_{x \in \Phi \cap B} \mathbf{1}(x \text{ not a transmitter at time } k) \\ &\stackrel{(a)}{=} \prod_{x \in \Phi \cap B} \mathbb{P}(x \text{ not a transmitter at time } k) \\ &= \mathbb{E} \left[(1 - p)^{\Phi(B)} \right] \\ &\stackrel{(b)}{=} \exp(-\lambda p |B|).\end{aligned}$$

Here (a) follows since nodes decide independently of each other whether or not to transmit. Since Φ is a PPP, $\Phi(B)$ is a Poisson random variable with mean $\lambda|B|$, and (b) follows from the moment-generating function of the Poisson random variable. From the above we see that the void probability of the transmitter process at time k is $\exp(-p\lambda|B|)$, which corresponds to that of a PPP. We also observe that Φ_k has density $p\lambda$, which is intuitive. The interference at time instant k and location z is given by

$$I_k(z) = \sum_{x \in \Phi} \mathbf{1}(x \in \Phi_k) h_{xz}(k) \ell(x - z), \quad (5.19)$$

where h_{xz} is the fading between locations x and z , which is assumed to be independent across time (block fading) and space.

Focusing on ALOHA with transmit probability p , our goal is to calculate the correlation coefficient of the random variables, $I_k(u)$ and $I_l(v)$, $k \neq l$, $u \neq v$.

Remark 5.2 With ALOHA, any finite collection or vector $\mathbf{v} = (I_{k_1}(z), I_{k_2}(z), \dots, I_{k_n}(z))$ with $n \in \mathbb{N}$, $k_1 \neq k_2 \neq \dots \neq k_n$, $z \in \mathbb{R}^2$, is *exchangeable* or *symmetrically dependent*, since the joint distribution of \mathbf{v} does not change under permutation of its components. Uses of exchangeability are surveyed in Kingman (1978).

To calculate the joint moments of the random variables, we use the joint Laplace transform of the random variables $I_k(u)$ and $I_l(v)$.

THEOREM 5.10 (Joint Laplace transform of interference) *The joint Laplace transform of $I_k(u)$ and $I_l(v)$, $k \neq l$, is*

$$\mathcal{L}(s_1, s_2) = \exp \left(-\lambda \int_{\mathbb{R}^2} [1 - \xi(s_1, u-x) \xi(s_2, v-x)] dx \right),$$

where

$$\xi(s, x) = 1 - p + p \mathcal{L}_h(s \ell(x)),$$

and \mathcal{L}_h denotes the Laplace transform of the fading distribution.

Proof The interference at time k and location $u \in \mathbb{R}^2$ is given by

$$I_k(u) = \sum_{x \in \Phi_k} h_{xu}(k) \ell(x - u),$$

and the interference at $v \in \mathbb{R}^2$ at time l is

$$I_l(v) = \sum_{y \in \Phi_l} h_{yv}(l) \ell(y - v).$$

So the joint Laplace transform is

$$\mathcal{L}(s_1, s_2) = \mathbb{E} \exp \left[-s_1 \sum_{x \in \Phi_k} h_{xu}(k) \ell(x - u) - s_2 \sum_{y \in \Phi_l} h_{yv}(l) \ell(y - v) \right].$$

Rewriting this as a product,

$$\begin{aligned} & \mathcal{L}(s_1, s_2) \\ &= \mathbb{E} \prod_{x \in \Phi} \exp(-s_1 \mathbf{1}(x \in \Phi_k) h_{xu}(k) \ell(x - u)) \exp(-s_2 \mathbf{1}(x \in \Phi_l) h_{xv}(l) \ell(x - v)). \end{aligned}$$

Since in ALOHA each node decides to transmit independently of the other nodes and across time, taking the expectation with respect to ALOHA,

$$\begin{aligned} & \mathcal{L}(s_1, s_2) \\ &= \mathbb{E} \prod_{x \in \Phi} [1 - p + p \exp(-s_1 h_{xu}(k) \ell(x - u))] [1 - p + p \exp(-s_2 h_{xv}(l) \ell(x - v))]. \end{aligned}$$

Since fading is assumed to be independent across time and space, on moving the expectation with respect to fading inside we get

$$\begin{aligned}
\mathcal{L}(s_1, s_2) &= \mathbb{E} \prod_{x \in \Phi} [1 - p + p\mathcal{L}_h(s_1\ell(x-u))] [1 - p + p\mathcal{L}_h(s_2\ell(x-v))] \\
&= \mathbb{E} \prod_{x \in \Phi} \xi(s_1, x-u) \xi(s_2, x-v) \\
&= \exp \left(-\lambda \int_{\mathbb{R}^2} [1 - \xi(s_1, x-u) \xi(s_2, x-v)] dx \right).
\end{aligned}$$

□

The last step follows from the pgfl of the PPP.

By a similar approach to that in the theorem above, the joint Laplace transform of the random variables $I_k(u)$ and $I_l(v)$, i.e., interference at the same time instant but at different locations can be shown to be

$$\mathcal{L}_{I(u)I(v)}(s_1, s_2) = \exp \left(-\lambda \int_{\mathbb{R}^2} [1 - \mathcal{L}_h(s_1\ell(x-u)) \mathcal{L}_h(s_2\ell(x-v))] dx \right). \quad (5.20)$$

The mean of the product of $I_k(u)$ and $I_l(v)$ at times k and l , $k \neq l$, which is required for the calculation of the correlation coefficient, is

$$\begin{aligned}
\mathbb{E}[I_k(u)I_l(v)] &= \frac{\partial^2}{\partial s_2 \partial s_1} \mathcal{L}(s_1, s_2) \Big|_{(s_1, s_2)=(0,0)} \\
&= p^2 \lambda \int_{\mathbb{R}^2} \ell(x-u) \ell(x-v) dx + \lambda^2 p^2 \left(\int_{\mathbb{R}^2} \ell(x) dx \right)^2.
\end{aligned}$$

LEMMA 5.11 *The spatiotemporal correlation coefficient of the interferences $I_k(u)$ and $I_l(v)$, $k \neq l$, for ALOHA and path loss functions $\ell(x)$ satisfying (5.17) is*

$$\zeta(u, v) = \frac{p \int_{\mathbb{R}^2} \ell(x) \ell(x - \|u - v\|) dx}{\mathbb{E}[h^2] \int_{\mathbb{R}^2} \ell^2(x) dx}. \quad (5.21)$$

Proof Since $I_k(u)$ and $I_l(v)$ are identically distributed, we have

$$\zeta(u, v) = \frac{\mathbb{E}[I_k(u)I_l(v)] - \mathbb{E}[I_k(u)]^2}{\mathbb{E}[I_k(u)^2] - \mathbb{E}[I_k(u)]^2}.$$

By substituting with the above quantities we have

$$\begin{aligned}
\zeta(u, v) &= \frac{p \int_{\mathbb{R}^2} \ell(x-u) \ell(x-v) dx}{\mathbb{E}[h^2] \int_{\mathbb{R}^2} \ell^2(x) dx} \\
&\stackrel{(a)}{=} \frac{p \int_{\mathbb{R}^2} \ell(x) \ell(x - \|u - v\|) dx}{\mathbb{E}[h^2] \int_{\mathbb{R}^2} \ell^2(x) dx},
\end{aligned} \quad (5.22)$$

where (a) follows by substituting $y = x - u$ and the fact that $\ell(x)$ depends only on $\|x\|$.

We observe that the correlation coefficient does not depend on k and l and is equal for any two time instants. From (5.20) we obtain the following lemma on the spatial correlation.

LEMMA 5.12 *The spatial correlation coefficient of the interference, i.e., the correlation coefficient of $I_k(u)$ and $I_k(v)$, $u \neq v$, is*

$$\zeta_s(u, v) = \frac{\int_{\mathbb{R}^2} \ell(x) \ell(x - \|u - v\|) dx}{\mathbb{E}[h^2] \int_{\mathbb{R}^2} \ell^2(x) dx}. \quad (5.23)$$

By setting $\|u - v\| = 0$ in Lemma 5.11, we obtain the temporal correlation coefficient.

LEMMA 5.13 *The temporal correlation coefficient for ALOHA is*

$$\zeta_t = \frac{p}{\mathbb{E}(h^2)}, \quad (5.24)$$

independently of $\ell(x)$. When the fading is Nakagami- m (see (2.13) for the distribution), the correlation coefficient is $\zeta_t = pm/(m + 1)$. In particular, for $m = 1$ (Rayleigh fading), the temporal correlation coefficient is $p/2$, and for $m \rightarrow \infty$ (no fading), it is p .

Hence the correlation increases with increasing m , i.e., fading decreases correlation, which is intuitive. Note that in the above derivation $\int_{\mathbb{R}^2} \ell^2(x) dx$ is not defined for $\ell(x) = \|x\|^{-\alpha}$, but we can use $\ell_\epsilon(x)$ and take $\epsilon \rightarrow 0$.

5.6 Temporal correlation of outage probabilities

We use the same setup as in the previous section and, in addition, assume that a transmitter at the origin has a destination located at $z \in \mathbb{R}^2$. Let A_k denote the event that the origin is able to connect to its destination z at time instant k , i.e.,

$$\text{SIR} = \frac{h_{oz}(k)l(z)}{I_k(z)} > \theta.$$

We assume all links are subject to Rayleigh fading. We would like to find the joint probability of success $\mathbb{P}(A_k, A_l)$, $k \neq l$, which gives an indication of the correlation

of the success probabilities in two different time slots. The derivation is based on a nice application of the pgfl.

Letting $\theta_z = \theta / \ell(z)$, we have

$$\begin{aligned}
\mathbb{P}(A_k, A_l) &= \mathbb{P}(h_{oz}(k) > \theta_z I_k(z), h_{oz}(l) > \theta_z I_l(z)) \\
&\stackrel{(a)}{=} \mathbb{E}[\exp(-\theta_z I_k(z)) \exp(-\theta_z I_l(z))] \\
&= \mathbb{E}\left[\exp\left(-\theta_z \sum_{x \in \Phi} l(x)[\mathbf{1}(x \in \Phi_k) h_{xz}(k) + \mathbf{1}(x \in \Phi_l) h_{xz}(l)]\right)\right] \\
&\stackrel{(b)}{=} \mathbb{E}\left[\prod_{x \in \Phi} \left(\frac{p}{1 + \theta_z \ell(x)} + 1 - p\right)^2\right] \\
&\stackrel{(c)}{=} \exp\left(-\lambda \int_{\mathbb{R}^2} 1 - \left(\frac{p}{1 + \theta_z \ell(x)} + 1 - p\right)^2 dx\right). \tag{5.25}
\end{aligned}$$

Here (a) follows from the independence of $h_{oz}(k)$ and $h_{oz}(l)$, $k \neq l$, (b) follows by taking the average with respect to $h_{xz}(k)$, $h_{xz}(l)$ and ALOHA, and (c) follows from the pgfl of the PPP.

From the joint probability, the conditional probabilities (given that a previous transmission was successful or failed) can be found.

Bibliographical notes

The use of stochastic geometry for wireless network analysis started with Musa & Wasylkiwskyj (1978) and Takagi & Kleinrock (1984) and continued with the interference characterization in PPPs in Sousa & Silvester (1990), Sousa (1990), and Ilow & Hatzinakos (1998). Lattices and PPPs are considered in Mathar & Mattfeldt (1995) and Haenggi (2009). The interference in binomial point processes is analyzed in Srinivasa & Haenggi (2010).

The connection between the Laplace transform of the interference and the success probability in Rayleigh fading was first noted in Linnartz (1992) and Zorzi & Pupolin (1995), and is explored in detail in Baccelli *et al.* (2006), which also provides results on the impact of interference and outage on routing.

The two volumes of Baccelli & Blaszczyszyn (2009) give a rigorous introduction to stochastic geometry (Volume I) and present detailed applications to MAC design and routing in ad hoc networks (Volume II).

Interference in wireless networks modeled as point processes and the resulting outage probabilities are discussed in detail in the monograph by Haenggi & Ganti (2008), which also contains extensions beyond the Poisson model.

The transmission capacity was introduced in Weber *et al.* (2005) and is the

subject of the monograph by Weber & Andrews (2012), and the interference correlation was analyzed in Ganti & Haenggi (2009b).

Details on stable distributions are available in Samorodnitsky & Taqqu (1994).

Problems

5.1 Let Φ be a uniform PPP, and let I_r denote the interference measured at the origin o stemming from interferers $\Phi \cap b(o, r)$. Show that if the path loss law is $\ell(r) = r^{-3}$,

$$\mathbb{E}(I_{2r} - I_r) = \mathbb{E}(I_\infty - I_{2r}), \quad \forall r > 0.$$

5.2 Let the path loss function be $\ell(x) = \min\{\|x\|^{-\alpha}, 1\}$. Let Φ be a stationary PPP on \mathbb{R}^2 of intensity λ . The interference is

$$I = \sum_{x \in \Phi} h_x \ell(x)$$

for iid h_x with distribution function F_h .

Determine $\text{var}(I)$ and $\mathbb{E}(I^2)$ for $F_h(x) = \mathbf{1}\{x \geq 1\}$ and for $F_h(x) = 1 - \exp(-x)$. What is the condition on α for the variance to be finite?

Using inequalities that involve the first and second moments, what can you say about the distribution of I ?

5.3 Consider four independent PPPs of intensity 1 on the plane. Take the path loss law to be $\ell(x) = \|x\|^{-4}$, and denote the interferences from the four PPPs at the origin o by I, I_1, I_2 , and I_3 . Find the constant c such that

$$I_1 + I_2 + I_3 \stackrel{d}{=} cI.$$

Generalize from the sum of the interference from three PPPs to n PPPs.

5.4 Set up a simulation to verify the mean values predicted by the theory in the previous problem for $\lambda = 1$ for the same path loss function but without fading. Plot the mean and the variance for $\alpha \in \{2.1, 2.2, \dots, 5.0\}$, and include the theoretical curve. What difficulty do you encounter?

5.5 Throughput using analysis. Take a PPP $\Phi \subset \mathbb{R}^2$ of intensity λ . Assume Rayleigh fading and a slotted ALOHA with probability p as the channel access protocol. Define the throughput to be

$$T = p(1-p)p_s(p),$$

where $p_s(p)$ is the success probability of a transmission over distance r when considering interference only:

$$p_s = \mathbb{P}(\text{SIR} > \theta).$$

Why is this a meaningful throughput metric?

Find the optimum transmit probability as a function of θ , α , and r , and the resulting p_s and T .

If the radios had *full-duplex* capabilities, how would the throughput definition change? Derive the corresponding optimum p and compare.

Plot the throughput for $r = 1$, $\lambda = 1$, $\theta = 1$, and $\alpha = 4$ as a function of p both for the half-duplex case and for the full-duplex case.

5.6 Throughput using simulation. Write a simulation for the same scenario, with $r = 1$, $\lambda = 1$, $\theta = 1$, and $\alpha = 4$. Plot the simulated throughput curves (as a function of p again) together with the analytical ones and compare.

5.7 Again for the same scenario, assume that the interferers do not fade (but the channel from the desired transmitter does). How do you expect the success probability to change? Verify using simulation.

5.8 For Poisson networks with Rayleigh fading and ALOHA, it is shown in Lemma 5.13 that the temporal interference correlation is proportional to the transmit probability p . Derive the same result using the method from Section 5.1.5: Consider a finite network first and condition on the number of nodes, then de-condition on the (Poisson) number of nodes and let the network area grow to infinity. Start with a finite ϵ for the path loss function ℓ_ϵ and let $\epsilon \rightarrow 0$ at the end.

5.9 For the temporal correlation of the success probabilities considered in Section 5.6, show that

$$\frac{\mathbb{P}(A_k | A_l)}{\mathbb{P}(A_k)} > 1$$

which indicates that success events are positively correlated.

¹ In general, pdfs also exist for $\delta = 1$ (the Cauchy distribution) and $\delta = 2$ (the Gaussian distribution), but, for $\delta \geq 1$, the interference in a PPP with the given path loss law is infinite a.s.

² 1 nat is the unit of information if the natural logarithm is used instead of the binary logarithm, which would result in bits. $1 \text{ nat} = 1/\log_2 \text{ bits}$.

6 Moment measures of point processes

6.1 Introduction

Point process theory provides useful analogues to the mean, variance, and higher-order statistics of numerical random variables. Since point processes are random elements in the space of locally finite counting measures, the moments are replaced by *moment measures*. In this chapter, we define such moment measures and calculate them for different types of point processes.

6.2 The first-order moment measure

The first moment of a point process is its intensity measure $\Lambda(B) = \mathbb{E}\Phi(B)$ that we are already familiar with. It corresponds to the mean of a numerical random variable.

Example 6.1 Let U_1 and U_2 be independent and uniformly distributed on $[0,s]$. For the randomly translated lattice

$$\Phi = \{m, n \in \mathbb{Z} : (U_1 + ms, U_2 + ns)\},$$

we find

$$\Lambda(B) = \mathbb{E}\Phi(B) = \frac{1}{s^2}|B|.$$

If Φ is stationary on \mathbb{R}^d , then $\Lambda(B) \equiv \Lambda(B + v)$ for all $v \in \mathbb{R}^d$, indicating that the intensity measure is invariant under translations. But only multiples of the Lebesgue measure have this property.

THEOREM 6.1 *If ν is a translation-invariant measure on \mathbb{R}^d , then $\nu(B) = c|B|$ for some $c > 0$*

COROLLARY 6.2 *If Φ is a stationary point process on \mathbb{R}^d , then its intensity measure Λ is a constant multiple of the Lebesgue measure. The constant is called the intensity of Φ .*

If the intensity measure Λ satisfies

$$\Lambda(B) = \int_B \lambda(x) dx$$

for some $\lambda(x)$, we call λ the *intensity function*. It has the following interpretation: For some small region $dx \subset \mathbb{R}^d$ at x ,

$$\mathbb{P}(\Phi(dx) > 0) \sim \mathbb{E}\Phi(dx) = \Lambda(dx) = \lambda(x)dx.$$

As pointed out in Section 2.6, $\mathbb{E}\Phi(B) = c|B|$ does not imply stationarity

Example 6.2 A Poisson cluster process with parent intensity $\lambda(r) = \min\{1, r^{-1}\}$ and clusters forming mixed BPPs with average number of nodes per cluster $\bar{c}(r) = \max\{1, r\}$ has a constant intensity but is not stationary.

Example 6.3 Similarly, a hybrid PPP/BPP (superposition of a PPP and BPP), with a BPP with one node on $[0, 1]^2$ and a PPP on $\mathbb{R}^2 \setminus [0, 1]^2$ of intensity 1 has unit intensity everywhere but different statistics inside the unit square and outside.

6.3 Second moment measures

Simple instances of second moments in point processes are the variance of the point counts $\Phi(B)$, given by

$$\text{var}\Phi(B) = \mathbb{E}(\Phi(B)) - (\mathbb{E}\Phi(B))^2,$$

and the covariance

$$\text{cov}(\Phi(A), \Phi(B)) = \mathbb{E}(\Phi(A)\Phi(B)) - \mathbb{E}\Phi(A)\mathbb{E}\Phi(B).$$

They can be obtained from the *second moment measure*, defined as follows.

DEFINITION 6.3 (Second moment measure) Let Φ be a point process on \mathbb{R}^d . Then

$\Phi^{(2)} \triangleq \Phi \times \Phi$ is a point process on $\mathbb{R}^d \times \mathbb{R}^d$ consisting of all ordered pairs (x, x') of points $x, x' \in \Phi$. The intensity measure $\mu^{(2)}$ of $\Phi \times \Phi$, called the *second moment measure*, is a measure on \mathbb{R}^{2d} defined as

$$\mu^{(2)}(A \times B) \triangleq \mathbb{E}(\Phi(A)\Phi(B)).$$

If $A = B$, then $\mu^{(2)}(A^2) = \mathbb{E}(\Phi(A)^2)$, so this measure does indeed provide the variance

$$\text{var}(\Phi(A)) = \mu^{(2)}(A^2) - (\Lambda(A))^2$$

and the covariance

$$\text{cov}(\Phi(A), \Phi(B)) = \mu^{(2)}(A \times B) - \Lambda(A)\Lambda(B).$$

Campbell's formula for sums (Theorem 4.1) also applies to $\Phi \times \Phi$, hence

$$\mathbb{E} \left(\sum_{x \in \Phi} \sum_{y \in \Phi} f(x, y) \right) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x, y) \mu^{(2)}(dx, dy).$$

Example 6.4 For the uniform PPP of intensity λ in \mathbb{R}^d , the second moment measure is

$$\mu^{(2)}(A \times B) = \lambda^2 |A||B| + \lambda|A \cap B|.$$

To see this, let us write A and B as

$$A = (A \cap B) \cup (A \setminus B); \quad B = (A \cap B) \cup (B \setminus A).$$

Then

$$\begin{aligned} \mu^{(2)}(A \times B) &= \mathbb{E}(\Phi(A)\Phi(B)) \\ &= \mathbb{E}[(\Phi(A \cap B) + \Phi(A \setminus B)) \cdot (\Phi(A \cap B) + \Phi(B \setminus A))] \\ &= \mathbb{E}(\Phi(A \setminus B))\mathbb{E}(\Phi(B \setminus A)) + \mathbb{E}(\Phi(A \cap B))\mathbb{E}(\Phi(A \setminus B)) \\ &\quad + \mathbb{E}(\Phi(A \cap B)\mathbb{E}(\Phi(B \setminus A))) + \mathbb{E}((\Phi(A \cap B))^2) \\ &= \mathbb{E}(\Phi(A))\mathbb{E}(\Phi(B)) + \underbrace{\mathbb{E}((\Phi(A \cap B))^2) - (\mathbb{E}(\Phi(A \cap B)))^2}_{\text{var } \Phi(A \cap B)} \\ &= \Lambda(A)\Lambda(B) + \Lambda(A \cap B) \\ &= \lambda^2 |A||B| + \lambda|A \cap B|. \end{aligned}$$

So there is a constant density λ^2 on all of $\mathbb{R}^d \times \mathbb{R}^d$, plus a positive mass on the

diagonal $\{(x, x) : x \in \mathbb{R}^d\}$.

The fact that points inside $A \cap B$ are of special importance can also be explained as follows. Consider what happens if a point is added. If the additional point falls in $A \setminus B$ or $B \setminus A$, the change is only linear in $|A|$ or $|B|$, respectively. If it falls into $A \cap B$, however, the change is quadratic.

Figure 6.1 shows a realization of the point process $\Phi^{(2)}$, where Φ is a PPP.

If $A = B$, we obtain $\mu^{(2)}(A^2) = \mathbb{E}(\Phi^2(A)) \neq \mathbb{E}^2(\Phi(A))$. The difference is the mass on the diagonal. But it is also the variance. So

$$\text{var}(\Phi(A)) = \lambda|A|,$$

as expected – since $\Phi(A)$ is Poisson distributed.

The second moment measure can be extended to higher-order moment measures in a straightforward manner. The n th-order moment measure is given by

$$\mu^{(n)}(B_1 \times \cdots \times B_n) = \mathbb{E}(\Phi(B_1) \times \cdots \times \Phi(B_n)),$$

and if $B_1 = \cdots = B_n = B$, then $\mu^n(B^n) = \mathbb{E}(\Phi(B)^n)$ yields the n th moment of the number of points in B .

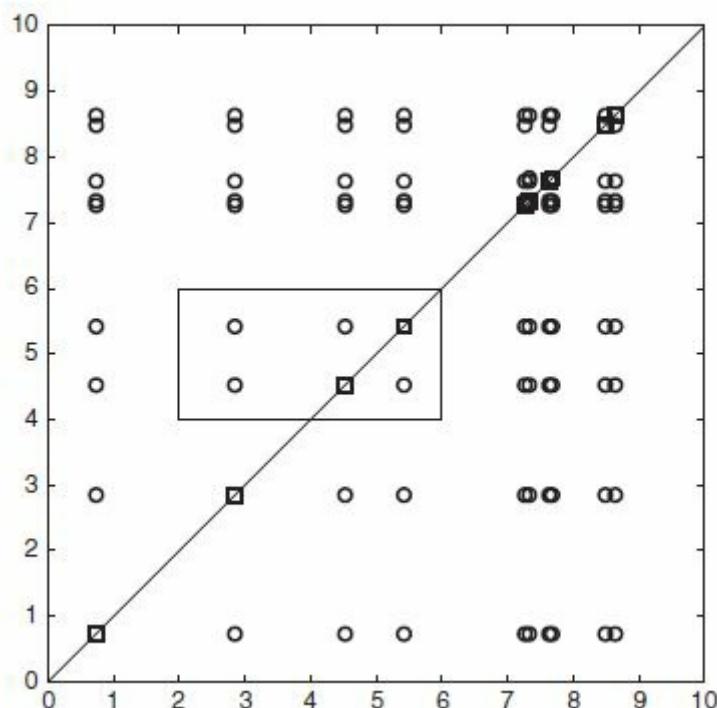


Figure 6.1 Illustration for second moment measure. Here $\Phi \subset \mathbb{R}^+$ is a PPP of intensity 1 on \mathbb{R}^+ . Shown is the product point process $\Phi^{(2)} = \Phi \times \Phi$ on $[0, 10]^2$. The points on the

diagonal, marked by \square , are the points (x, x) for $x \in \Phi$. The other points are marked by \circ . The box is the rectangle $C = A \times B$ with $A = [2, 6]$ and $B = [4, 6]$. For this realization, $\Phi^{(2)}(C) = 6$. If $A \cap B = \emptyset$, then no points on the diagonal lie in C . The expected number of points in C is $\mu^{(2)}(C) = 4 \cdot 2 + 2 = 10$.

In many instances, it is required or makes sense to remove the mass on the diagonal. This leads to the *second factorial moment measure*.

DEFINITION 6.4 (Second factorial moment measure) The second factorial moment measure is the intensity measure of the process $\Phi * \Phi$ of all ordered pairs of *distinct* points of Φ :

$$\alpha^{(2)}(A \times B) \triangleq \mathbb{E}(\Phi(A)\Phi(B)) - \mathbb{E}(\Phi(A \cap B)). \quad (6.1)$$

The set of points $\Phi * \Phi$ can be written as

$$\Phi * \Phi \triangleq \{(x, x') \in \Phi \times \Phi : x \neq x'\}.$$

Again from Campbell's formula, we know that the second factorial moment measure satisfies

$$\mathbb{E} \left(\sum_{x,y \in \Phi}^{\neq} f(x,y) \right) = \mathbb{E} \left(\sum_{(x,y) \in \Phi * \Phi} f(x,y) \right) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x,y) \alpha^{(2)}(dx, dy).$$

As before, the \neq symbol above the summation operator indicates that the sum is taken only over distinct point pairs.

If A and B are disjoint, then $\mu^{(2)}(A \times B) = \alpha^{(2)}(A \times B)$. Generally we have

$$\mu^{(2)}(A \times B) = \alpha^{(2)}(A \times B) + \Lambda(A \cap B).$$

The circles in Fig. 6.1 show the points of the product $\Phi * \Phi$.

We can expand (6.1) as follows:

$$\begin{aligned} \alpha^{(2)}(A \times B) &= \mathbb{E}(\#\{(x, y) : x \in \Phi \cap A, y \in \Phi \cap B, x \neq y\}) \\ &= \mathbb{E} \left(\sum_{x,y \in \Phi}^{\neq} \mathbf{1}_A(x)\mathbf{1}_B(y) \right). \end{aligned} \quad (6.2)$$

The difference between $\mu^{(2)}(A \times B)$ and $\alpha^{(2)}(A \times B)$ lies in the expectation of the sum

$$\sum_{x,y \in \Phi : x=y} \mathbf{1}_A(x)\mathbf{1}_B(y) = \sum_{x \in \Phi} \mathbf{1}_{A \cap B}(x).$$

The name “factorial” comes from the fact that

$$\begin{aligned}\alpha^{(2)}(A \times A) &= \mathbb{E}(\Phi(A)^2) - \mathbb{E}(\Phi(A)) \\ &= \mathbb{E}(\Phi(A)(\Phi(A) - 1)).\end{aligned}$$

Just like the second moment measure, the second factorial moment measure can be extended to higher-order measures. The n th-order factorial moment measure for $B_1 = \dots = B_n = B$ is given by

$$\alpha^{(n)}(B^n) = \mathbb{E}(\Phi(B)(\Phi(B) - 1) \cdot \dots \cdot (\Phi(B) - n + 1)).$$

Example 6.5 For a Poisson point process,

$$\alpha^{(2)}(A \times B) = \Lambda(A)\Lambda(B).$$

In the uniform case,

$$\alpha^{(2)}(A \times B) = \lambda^2 |A||B|.$$

Usually there exists a density (with respect to the Lebesgue measure) of the second factorial moment measure, the *second moment density*, which will be discussed in the next section. In contrast, the second moment measure $\mu^{(2)}$ does not have a density, since some of its mass is concentrated on a subset of smaller dimension. This is another principal motivation for using the factorial moment measure $\alpha^{(2)}$.

6.4 Second moment density

DEFINITION 6.5 (Second moment density (or second-order product density)) A point process Φ on \mathbb{R}^d is said to have *second moment density* $\varrho^{(2)}$ if

$$\alpha^{(2)}(C) = \alpha^{(2)}(A \times B) = \int_A \int_B \varrho^{(2)}(x, y) dy dx = \int_C \varrho^{(2)}(x, y) dx dy$$

for any compact $C = A \times B \subset \mathbb{R}^d \times \mathbb{R}^d$.

In differential form, $\alpha^{(2)}(dx, dy) = \varrho^{(2)}(x, y)dx dy$. Informally, $\varrho^{(2)}(x, y)$ is the joint probability that there are two points of Φ at two specified locations x and y in the

infinitesimal volumes dx and dy :

$$\mathbb{P}(\Phi(dx) > 0, \Phi(dy) > 0) \sim \mathbb{E}(\Phi(dx)\Phi(dy)) = \varrho^{(2)}(x,y)dx dy.$$

There exists also a first-order product density or first moment density, which is the density pertaining to the first moment measure or intensity measure. It is simply the intensity function $\lambda(x)$. The n th moment density or n th-order product density is the density pertaining to the n th-order factorial moment measure:

$$\alpha^{(n)}(B_1 \times \cdots \times B_n) = \int_{B_1} \int_{B_2} \cdots \int_{B_n} \varrho^{(n)}(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n$$

Example 6.6 It follows from the independence property of the PPP that the uniform PPP of intensity λ has $\varrho^{(2)}(x,y) = \lambda^2$. In the non-uniform case, $\varrho^{(2)}(x,y) = \lambda(x)\lambda(y)$. This generalizes to $\varrho^{(n)} = \lambda^n$, and it follows that for the uniform PPP

$$\alpha^{(n)}(B_1 \times \cdots \times B_n) = \lambda^n \prod_{k \in [n]} |B_k|.$$

If Φ is stationary then $\varrho^{(1)} = \lambda$ and $\varrho^{(2)}$ depends only on the difference of its arguments, i.e., there exists a $\varrho^{(2)}_{\text{st}} : \mathbb{R}^d \rightarrow \mathbb{R}^+$ such that

$$\varrho^{(2)}(x, y) \equiv \varrho^{(2)}_{\text{st}}(x - y) \quad \forall x, y \in \mathbb{R}^d$$

If Φ is motion-invariant then $\varrho^{(2)}_{\text{st}}$ depends only on the distance $r = \|x - y\|$, i.e., there exists a $\varrho^{(2)}_{\text{mi}} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ such that

$$\varrho^{(2)}(x, y) \equiv \varrho^{(2)}_{\text{st}}(x - y) \equiv \varrho^{(2)}_{\text{mi}}(\|x - y\|) = \varrho^{(2)}_{\text{mi}}(r), \quad \forall x, y \in \mathbb{R}^d.$$

In these cases, $\varrho^{(2)}_{\text{st}}$ and $\varrho^{(2)}_{\text{mi}}$ are also called second moment densities or second-order product densities. For example, in the uniform Poisson case, $\varrho^{(2)}_{\text{st}}(x) = \varrho^{(2)}_{\text{mi}}(\|x\|) = \lambda^2$

Example 6.7 Show that the second moment density for a (uniform) binomial point process with n points in W is

$$\varrho^{(2)}(x, y) = \frac{n(n-1)}{|W|^2} \mathbf{1}_W(x) \mathbf{1}_W(y).$$

Solution

For $A, B \subset W$, let $N_A = \Phi(A \setminus B)$, $N_B = \Phi(B \setminus A)$, and $N_C = \Phi(A \cap B)$; $p_A = |A \setminus B|/|W|$, $p_B = |B \setminus A|/|W|$, and $p_C = |A \cap B|/|W|$. The random variables N_A, N_B , and N_C follow a multinomial distribution with probabilities p_A, p_B , and p_C , respectively:

$$\begin{aligned}\alpha^{(2)}(A \times B) &= \mathbb{E}[(N_A + N_C)(N_B + N_C)] - \mathbb{E}N_C \\ &= \mathbb{E}(N_C^2) + \mathbb{E}(N_A N_C) + \mathbb{E}(N_B N_C) + \mathbb{E}(N_A N_B) - \mathbb{E}N_C \\ &\stackrel{(a)}{=} n(n-1) [p_C^2 + p_A p_C + p_B p_C + p_A p_B] \\ &= n(n-1) \frac{|A||B|}{|W|^2}.\end{aligned}$$

Here (a) follows because the mean and variance of N_C are np_C and $np_C(1-p_C)$, respectively, and $\text{cov}(N_A, N_B) = -np_A p_B$ so that $\mathbb{E}(N_C^2) - \mathbb{E}N_C = n(n-1)p_C^2$ and $\mathbb{E}(N_A N_B) = \text{cov}(N_A, N_B) + \mathbb{E}(N_A)\mathbb{E}(N_B) = n(n-1)p_A p_B$.

DEFINITION 6.6 (Pair correlation function) For a point process $\Phi \subset \mathbb{R}^d$ with intensity function $\lambda(x)$ and second moment density $\varrho^{(2)}(x, y)$, the pair correlation function is defined as

$$g(x, y) \triangleq \frac{\varrho^{(2)}(x, y)}{\lambda(x)\lambda(y)}.$$

For motion-invariant processes, it is

$$g_{\text{mi}}(r) \triangleq \frac{\varrho_{\text{mi}}^{(2)}(r)}{\lambda^2}.$$

The pair correlation function is identical to 1 in the uniform Poisson case, and it is smaller than 1 (for small r) if there is repulsion and larger than 1 if there is clustering.

For a BPP with n points on W , it is (from Example 6.7)

$$g(x, y) = \frac{n(n-1)}{|W|^2} \frac{|W|^2}{n^2} = 1 - \frac{1}{n}.$$

The pair correlation function measures the degree of correlation between the random variables $\Phi(A)$ and $\Phi(B)$ in a non-centered way. If $g(x, y) \equiv 1$, then $\text{cov}(\Phi(A), \Phi(B)) = 0$ for disjoint A and B .

Example 6.8 Let $\Phi \subset \mathbb{R}^d$ consist of n random points $\{x_1, \dots, x_n\}$, where x_i has marginal probability density $f_i(u)$, $u \in \mathbb{R}^d$. Then Φ has intensity

$$\lambda(x) = \sum_{i=1}^n f_i(x).$$

Let $f_{ij}(u, v)$ be the marginal *joint* density of (x_i, x_j) . Then Φ has second moment density

$$\varrho^{(2)}(x, y) = \sum_{i,j \in [n]}^{\neq} f_{ij}(x, y) \quad (6.3)$$

and pair correlation function

$$g(x, y) = \frac{\sum_{i,j \in [n]}^{\neq} f_{ij}(x, y)}{\left(\sum_{i \in [n]} f_i(x)\right) \left(\sum_{j \in [n]} f_j(y)\right)}.$$

For the simplest case $n = 2$, the joint probability of having one point in A and the other in B is

$$a^{(2)}(A \times B) = \int_A \int_B (f_{12}(x, y) + f_{21}(x, y)) dx dy,$$

which is consistent with (6.3).

The BPP is a special case with $f_i(x) = |W|^{-1}\mathbf{1}(x \in W)$ (if it is uniform) and $f_{ij}(x, y) = f_i(x)f_j(y) = |W|^{-2}\mathbf{1}(x \in W)$. Its intensity is $n/|W|$, and it follows from (6.3) that the second moment density is $n(n - 1)/|W|^2$, as expected.

Example 6.9 Consider a Poisson cluster process Φ , formed from a uniform parent PPP Φ_p with intensity λ_p by replacing each $x \in \Phi_p$ by a random finite cluster Φ_x . The clusters Φ_x for different $x \in \Phi_p$ are independent processes. Let Φ_x have intensity function $\lambda(u | x)$ and second moment density $h(u, v | x)$. Conditioned on Φ_p , the cluster process has intensity

$$\lambda(u | \Phi_p) = \sum_{x \in \Phi_p} \lambda(u | x).$$

The (unconditioned) intensity of the cluster process Φ is thus, by application of Campbell's formula,

$$\begin{aligned}
\lambda(u) &= \mathbb{E}(\lambda(u \mid \Phi_p)) \\
&= \mathbb{E} \sum_{x \in \Phi_p} \lambda(u \mid x) \\
&= \lambda_p \int_{\mathbb{R}^d} \lambda(u \mid x) dx.
\end{aligned}$$

If all clusters have the same intensity function $\lambda_0(x)$, i.e., $\lambda(u \mid x) \equiv \lambda_0(u - x)$, then $\lambda = \lambda_p \bar{c}$, where $\bar{c} = \Lambda_0(\mathbb{R}^d)$ is the mean number of points per cluster. The second moment density is

$$\varrho^{(2)}(u, v) = \lambda^2 + \lambda_p \int_{\mathbb{R}^d} h(u, v \mid x) dx.$$

There are two contributions to this second moment density, the one from pairs of points in different clusters and the one from pairs of points in the same cluster. The first contribution is λ^2 since clusters are independent. The second one follows from the second moment density of the clusters as

$$\mathbb{E} \left(\sum_{x \in \Phi_p} h(u, v \mid x) \right) = \lambda_p \int_{\mathbb{R}^d} h(u, v \mid x) dx.$$

As a quick sanity check, assume each parent had a daughter point at exactly the location of the parent. Then $g(u \mid x) = \delta_x(u)$ and $h(u, v \mid x) = \delta_x(u)\delta_x(v) = 0$ since $u \neq v$. So $\lambda(u) = \lambda_p \int \delta_x(u) dx = \lambda_p$, and we obtain the desired result.

In the *Matérn cluster process*, defined in Definition 3.6, the points in each cluster are distributed uniformly at random in a ball of radius R around their parent points, and each cluster contains a Poisson number of points with mean \bar{c} . Accordingly, the cluster intensity function is

$$\lambda_0(x) = \frac{\bar{c}}{c_d R^d} \mathbf{1}_{b(o, R)}(x).$$

In the two-dimensional case, the second moment density of the cluster Φ_x is $h(u, v \mid x) = \bar{c}^2 / (\pi^2 R^4)$ if $u, v \in b(x, R)$ and 0 otherwise. We have

$$\begin{aligned}
\int_{\mathbb{R}^2} \mathbf{1}(u, v \in b(x, R)) dx &= \int_{\mathbb{R}^2} \mathbf{1}_{b(x, R)}(u) \mathbf{1}_{b(x, R)}(v) dx \\
&= \int_{\mathbb{R}^2} \mathbf{1}_{b(u, R)}(x) \mathbf{1}_{b(v, R)}(x) dx \\
&= |b(u, R) \cap b(v, R)|.
\end{aligned}$$

The intersection area of two disks of radius R at distance r is

$$A_R(r) = 2R^2 \arccos \left(\frac{r}{2R} \right) - r \sqrt{R^2 - \frac{r^2}{4}}, \quad 0 \leq r \leq 2R, \quad (6.4)$$

and 0 for $r > 2R$. Hence the second moment density of the Matérn cluster process is

$$\varrho^{(2)}(u, v) = \lambda_p^2 \bar{c}^2 + \lambda_p \frac{\bar{c}^2}{\pi^2 R^4} A_R(\|u - v\|)$$

or, since this cluster process is motion-invariant, with $r = \|u - v\|$,

$$\varrho_{\text{mi}}^{(2)}(r) = \lambda_p^2 \bar{c}^2 + \lambda_p \frac{\bar{c}^2}{\pi^2 R^4} A_R(r).$$

The pair correlation function follows as

$$g_{\text{mi}}(r) = \frac{\varrho^{(2)}(r)}{\lambda_p^2 \bar{c}^2} = 1 + \frac{1}{\lambda_p} \frac{A_R(r)}{\pi^2 R^4}.$$

For $r \geq 2R$, this is just the pair correlation function of the PPP, since two points at distance larger than $2R$ cannot belong to the same cluster. For smaller r , $g(r) > 1$, as expected for a cluster process. Also, as $\lambda_p \rightarrow \infty$, this approaches the PPP's pair correlation function; in this case the (relatively few) point pairs belonging to the same cluster become irrelevant.

Generally, for Poisson cluster processes with a Poisson number of nodes per cluster and symmetric probability density function $f(x) = f(-x)$ for the points in each cluster,

$$\varrho_{\text{st}}^{(2)}(x) = \lambda^2 \left[1 + \frac{(f * f)(x)}{\lambda_p} \right], \quad (6.5)$$

where $\lambda = \lambda_p \bar{c}$ is the density of the cluster process. To show this, we focus on a single cluster and go back to the case of the finite point process. Given the number of nodes n , they are independently placed. Hence, given n , $\varrho^{(2)}(x, y | n) = n(n - 1)f(x)f(y)$ and

$$\varrho^{(2)}(x, y) = \bar{c}^2 f(x)f(y).$$

Still considering the case when x and y belong to the same cluster, but now averaging over the parent process Φ_p , yields

$$\begin{aligned}
\varrho_{\text{cl}}^{(2)}(u, v) &= \mathbb{E} \sum_{x \in \Phi_p} \bar{c}^2 f(u - x) f(v - x) \\
&= \lambda_p \bar{c}^2 \int_{\mathbb{R}^d} f(u - x) f(v - x) dx \\
&= \lambda_p \bar{c}^2 \int_{\mathbb{R}^d} f(z) f(v - u + z) dz \\
&= \lambda_p \bar{c}^2 (f * f)(u - v).
\end{aligned}$$

To this we need to add the λ^2 for the case when x and y belong to different clusters to obtain (6.5).

Example 6.10 In the (two-dimensional) Thomas cluster process defined in Definition 3.5, each cluster is a finite PPP with Gaussian intensity function

$$\lambda_0(x) = \frac{\bar{c}}{2\pi\sigma^2} \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right), \quad x \in \mathbb{R}^2.$$

From the convolution formula (6.5) it follows that

$$\varrho_{\text{mi}}^{(2)}(r) = \lambda_p^2 \bar{c}^2 + \lambda_p \frac{\bar{c}^2}{4\pi\sigma^2} \exp\left(-\frac{r^2}{4\sigma^2}\right), \quad (6.6)$$

and the pair correlation function is

$$g_{\text{mi}}(r) = 1 + \frac{1}{4\pi\lambda_p\sigma^2} \exp\left(-\frac{r^2}{4\sigma^2}\right).$$

Again we note that the pair correlation function approaches 1 from above as $\lambda_p \rightarrow \infty$.

6.5 Second moments for stationary processes

For a point process $\Phi \subset \mathbb{R}^d$, stationarity implies

$$\mathbb{E}[\Phi(A + v)\Phi(B + v)] = \mathbb{E}[\Phi(A)\Phi(B)]$$

for all $v \in \mathbb{R}^d$. Thus $\mu^{(2)}$ and $\alpha^{(2)}$ are invariant under *simultaneous* shifts.

Now apply the transform $T(x, y) = (x, y - x)$ from $\mathbb{R}^d \times \mathbb{R}^d$ onto itself. Under this transformation, the simultaneous shift becomes a shift in the first (d -dimensional) component only, i.e.,

$$(x, y) \rightarrow (x + v, y).$$

This process is illustrated in Fig. 6.2. So, for stationary processes, the image of $\alpha^{(2)}$ under the transformation T is a measure μ which is invariant under translations of the first coordinate. Since translation-invariant measures are multiples of the Lebesgue measure, it follows that the new measure μ can be written as

$$\mu = \lambda v_d \otimes \kappa$$

where λ is the intensity of the process, v_d the d -dimensional Lebesgue measure, and κ is a measure on \mathbb{R}^d called the *reduced second moment measure*. The symbol \otimes denotes the product of measures. Hence, under T , the second factorial moment measure *disintegrates* into the product of two parts, a Lebesgue component $\Lambda(dx)$ along the (original) diagonal $x = y$ and a reduced measure $\kappa(du)$ along $u = y - x$. Formally, starting with Campbell's formula for sums,

$$\begin{aligned}\mathbb{E} \left[\sum_{x,y \in \Phi}^{\neq} f(x,y) \right] &= \iint f(x,y) \alpha^{(2)}(dx,dy) \\ &\stackrel{(a)}{=} \iint f(x, x+u) \mu(dx, du) \\ &= \lambda \iint f(x, x+u) dx \kappa(du).\end{aligned}$$

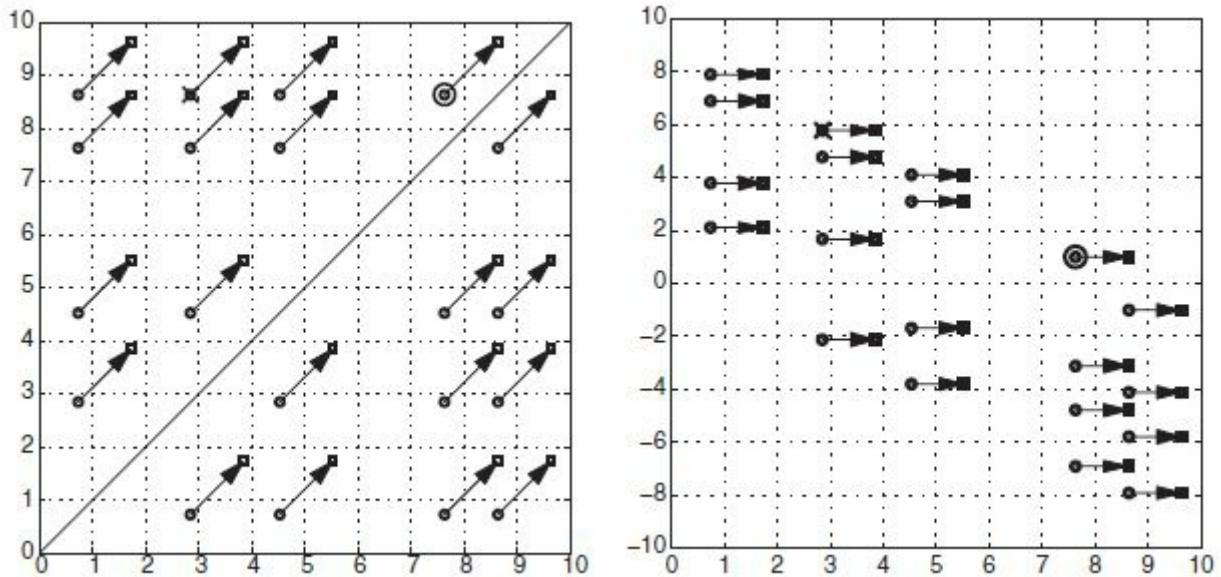


Figure 6.2 Illustration of coordinate transform $(x, y) \rightarrow (x, y - x)$. (Left) Original product point process with translation $v = 1$, which moves all points along the diagonal from \circ to \square . (Right) Transformed product point process with same translation. Here, the points are moved only horizontally. Two specific points are identified by a cross and a larger circle, respectively.

Here (a) follows from the application of the transform T . This leads us to the

definition of the reduced second moment measure.

DEFINITION 6.7 (Reduced second moment measure) Let Φ be a stationary point process on \mathbb{R}^d with intensity λ . Then there is a measure κ on \mathbb{R}^d such that for general measurable f ,

$$\mathbb{E} \left[\sum_{x,y \in \Phi}^{\neq} f(x,y) \right] = \lambda \iint f(x, x+u) dx \kappa(du).$$

κ is called the *reduced second moment measure* of Φ .

Example 6.11 For the uniform PPP, $\alpha^{(2)} = \lambda^2 v_d \otimes v_d$. The image μ of $\alpha^{(2)}$ under the transform T is also $\lambda^2 v_d \otimes v_d$, thus $\kappa = \lambda v_d$.

Choosing $f(x, y) = \mathbf{1}_A(x)\mathbf{1}_B(y-x)$, we have

$$\mathbb{E} \left[\sum_{x,y \in \Phi}^{\neq} \mathbf{1}_A(x)\mathbf{1}_B(y-x) \right] = \lambda \iint \mathbf{1}_A(x)\mathbf{1}_B(u) dx \kappa(du) = \lambda |A| \kappa(B). \quad (6.7)$$

Since $\lambda |A| = \mathbb{E}\Phi(A)$ and $\mathbf{1}_B(y-x) = \mathbf{1}_{B+x}(y)$, we obtain

$$\kappa(B) = \frac{\mathbb{E} \sum_{x \in \Phi \cap A} \Phi((B+x) \setminus \{x\})}{\mathbb{E}\Phi(A)}.$$

This expression can be interpreted as follows. For each point $x \in \Phi \cap A$, determine how many *other* points $y \in \Phi$ exist such that $y - x \in B$. Then take the average of that number over all points in A . Since Φ is stationary, this is equivalent to taking the average over all points of Φ .

If the second moment density $\varrho^{(2)}(x, y)$ exists, it depends only on the difference $x - y$, and we have the relationship

$$\kappa(B) = \frac{1}{\lambda} \int_B \varrho_{st}^{(2)}(u) du. \quad (6.8)$$

This can be seen by expanding the left side of (6.7) as follows:

$$\begin{aligned}
\mathbb{E} \left[\sum_{x,y \in \Phi}^{\neq} \mathbf{1}_A(x) \mathbf{1}_B(y-x) \right] &= \iint \mathbf{1}_A(x) \mathbf{1}_B(y-x) \varrho^{(2)}(x,y) dx dy \\
&= \int_A \int_{\mathbb{R}^d} \mathbf{1}_B(y-x) \varrho^{(2)}(x,y) dx dy \\
&= \int_A \int_{\mathbb{R}^d} \mathbf{1}_B(u) \varrho_{st}^{(2)}(u) du dx \\
&= \int_A \int_B \varrho_{st}^{(2)}(u) du dx \\
&= |A| \int_B \varrho_{st}^{(2)}(u) du,
\end{aligned}$$

which, by (6.7), is equal to $\lambda |A| \mathcal{K}(B)$. Hence (6.8) follows.

For motion-invariant processes, a simpler function is often useful, namely *Ripley's K function*, also called the *reduced second moment function* (Ripley 1976).

DEFINITION 6.8 (Ripley's *K* function) The *K* function is defined as

$$K(r) \triangleq \frac{1}{\lambda} \mathcal{K}(b(o,r)), \quad r \geq 0.$$

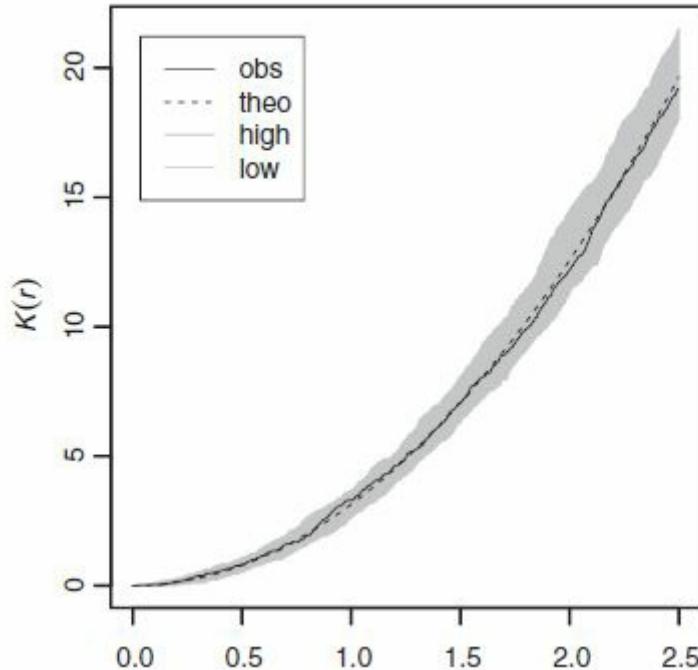


Figure 6.3 Empirical estimated *K* function of the uniform PPP (black solid line, called “obs” in the legend) and the theoretical $K(r) = \pi r^2$ (dashed, “theo”). The gray area shows the range between the pointwise maximum and minimum of 99 realizations.

So $\lambda K(r)$ is the mean number of points y of the process that satisfy $0 < \|y - x\| \leq r$ for a given point x of the process.

Example 6.12 For the uniform PPP on \mathbb{R}^d ,

$$K(r) = c_d r^d, \quad r \geq 0.$$

Figure 6.3 shows an empirical curve plus the theoretical one, together with the envelope of 99 realizations in two dimensions.

Example 6.13 For the Thomas process with parent PPP intensity λ_p ,

$$K(r) = \pi r^2 + \frac{1 - \exp(-r^2/(4\sigma^2))}{\lambda_p}.$$

As is typical for cluster processes, the K function consists of two parts, the first being the contribution from other clusters, and the second being that from the same cluster. If $F(r)$ denotes the cdf of the distance of two points in the same cluster,

$$K(r) = \pi r^2 + \frac{1}{\lambda} \bar{c} F(r).$$

For the Thomas process, $F(r) = 1 - \exp(-r^2/(4\sigma^2))$. For general Neyman–Scott processes, where the number of daughter points per parent M has an arbitrary distribution with mean $\bar{c} = \mathbb{E}M$,

$$\begin{aligned} K(r) &= \pi r^2 + \frac{1}{\lambda} \frac{\mathbb{E}(M(M-1))}{\bar{c}} F(r) \\ &= \pi r^2 + \frac{\mathbb{E}(M(M-1))F(r)}{\lambda_p \bar{c}^2}. \end{aligned}$$

The ratio $\mathbb{E}(M(M-1))/\mathbb{E}M = \text{var}M/\mathbb{E}M + \mathbb{E}M - 1$ is the mean number of other points in the cluster, averaged over all points in the cluster.

Example 6.14 Here we determine the K function of the Matérn hard-core process of type I, introduced in Definition 3.7, in two dimensions. Since the variable r is usually used in $K(r)$, we denote the hard-core distance by ρ .

We first calculate $\lambda K(r)$, the mean number of extra points within distance r of a given point x . A point of the underlying PPP of intensity λ_b at location y with distance $u = \|x - y\|$ is retained in the hard-core process if $u > \delta$ and $b(y, \rho) \setminus b(x, \rho)$ is free of other points of the basic process. (The disk $b(x, \rho)$ is already empty since there is a point at x that is retained.) The density of points in the PPP at distance u is $2\pi\lambda_b u$, so, if $r > \delta$,

$$\lambda K(r) = 2\pi\lambda_b \int_0^r \exp(-\lambda_b |\tilde{V}_\rho(u)|) u \, du,$$

where $\tilde{V}_\rho(u) = |\mathbf{b}((u, 0), \rho) \setminus b(o, \rho)|$. For $r \leq \delta$, $K(r) = 0$. Alternatively, this can be written as

$$\lambda K(r) = 2\pi\lambda_b \exp(\lambda_b\pi\rho^2) \int_0^r \exp(-\lambda_b|V_\rho(u)|) u du,$$

where $V_\rho(u) = |b((u, 0), \rho) \cup b(o, \rho)|$ is the *union* of the two disks. Dividing by the intensity $\lambda = \lambda_b \exp(-\lambda_b\pi\rho^2)$ yields the K function. It is usually written in the following form:

$$K(r) = 2\pi \exp(2\lambda_b\pi\rho^2) \int_0^r uk(u) du, \quad (6.9)$$

where

$$k(u) = \begin{cases} 0 & \text{if } u < \rho \\ \exp(-\lambda_b V_\rho(u)) & \text{if } u \geq \rho \end{cases} \quad (6.10)$$

is the probability that two points at distance u are both retained. It is easily verified that $K(r) \sim \pi r^2$ as $r \rightarrow \infty$, as is the case for all stationary point processes. $V_\rho(u)$ is given by

$$V_\rho(u) = 2\pi\rho^2 - 2\rho^2 \arccos\left(\frac{u}{2\rho}\right) + u\sqrt{\rho^2 - \frac{u^2}{4}}, \quad 0 \leq u \leq 2\rho. \quad (6.11)$$

For $u < 2\rho$, the union area is simply the area of the two disks, $2\pi\rho^2$.

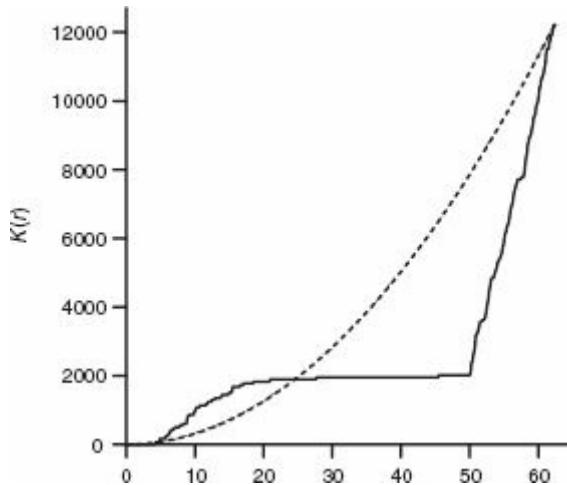


Figure 6.4 Empirical estimated K function of the Strauss hard-core process in Fig. 3.8 (right). The dashed line is the theoretical K function of the homogeneous PPP, i.e., $K(r) = \pi r^2$. It is observed that the process is locally clustered (up to about $r = 20$) and then exhibits hard-core behavior up to $r = 50$.

Example 6.15 A good approximation for the K function of the Strauss process (see Definition 3.10) with interaction radius R is

$$K(r) \approx \begin{cases} e^{-b}\pi r^2 & \text{if } r \leq R \\ \pi r^2 - (1 - e^{-b})\pi R^2 & \text{if } r \geq R. \end{cases}$$

This approximation is obtained by arguing that roughly a fraction $1 - e^{-b}$ of the points within distance R is deleted. These points are missing in the count $K(r)$ for $r > R$, resulting in the constant offset $(1 - e^{-b})\pi R^2$.

Example 6.16 The empirical K function of the Strauss hard-core process (see Definition 3.11) in Fig. 3.8 (right) is shown in Fig. 6.4. The K function is corrected for boundary effects using Ripley's isotropic correction, which is implemented in R's Kest command. It can be seen that the process exhibits clustered and hard-core properties, depending on the distance.

For a stationary point process on \mathbb{R}^d that has a second moment density,

$$K(r) = \frac{1}{\lambda^2} \int_{b(o,r)} \varrho_{st}^{(2)}(x) dx = \int_{b(o,r)} g_{st}(x) dx, \quad (6.12)$$

where g is the pair correlation function. If the process is on \mathbb{R}^2 and motion-invariant,

$$K(r) = 2\pi \int_0^r g(r') r' dr'.$$

We also have the following relationships.

LEMMA 6.9 *For a motion-invariant point process on \mathbb{R}^2 , the pair correlation function is given by*

$$g_{mi}(r) = \frac{1}{2\pi r} \frac{d}{dr} K(r).$$

LEMMA 6.10 (Invariance of K under thinning) *If Φ' is obtained by random (stationary) thinning from a stationary point process Φ , then the K functions of Φ and Φ' are identical.*

A close relative of the K function is the L function.

DEFINITION 6.11 (L function)

$$L(r) \triangleq \left(\frac{K(r)}{c_d} \right)^{1/d}.$$

The L function is sometimes preferred to the K function since it is simply $L(r) = r$ for the uniform PPP.

A convenient way to determine at which range a point process exhibits clustering or hard-core properties is to consider the function $L(r) - r = \sqrt{K(r)/\pi} - r$ (in two dimensions), which is above 0 for values of r at which the process is clustered and below 0 if it is hard-core.

An important question is whether second-order properties are sufficient to completely characterize a point pattern. The answer is no, as the following example illustrates.

Example 6.17 The Baddeley–Silverman cell process (Baddeley & Silverman 1984) is defined as follows. The plane \mathbb{R}^2 is divided into equal rectangular cells. In each cell, a random number M of points is placed, where M is distributed as

$$\mathbb{P}(M = 0) = 1/10; \quad \mathbb{P}(M = 1) = 8/9; \quad \mathbb{P}(M = 10) = 1/90.$$

The points are uniformly distributed in each cell, and the numbers of points in each cell are independent. All second-order properties of this process are identical to those of the uniform PPP (see Problem 6.7).

Bibliographical notes

The definition of the reduced second moment measure varies slightly throughout the literature. Here we adopted the definition in Baddeley *et al.* (2007). Some authors scale their $\kappa(B)$ by λ^{-1} (Stoyan *et al.* 1995), such that $K(r) = \kappa(b(o, r))$, or λ (Daley & Vere-Jones 2008).

Ripley's K function, however, is defined consistently. It was introduced in Ripley (1976). In fact, he denoted it by κ , but K is used more frequently. The approximation of the K function for Strauss processes was derived in Isham (1984).

The Baddeley–Silverman cell process was defined in Baddeley & Silverman (1984) and used to illustrate that second-order properties do not define a point process completely.

Problems

6.1 Show that

$$\text{var } \Phi(A) = \alpha^{(2)}(A \times A) + \Lambda(A) - (\Lambda(A))^2.$$

- 6.2** Use R to verify the accuracy of the approximation in Example 6.15.
- 6.3** Use the `x=clickppp(30)` command in R to enter 30 points manually. Try to produce a typical realization of a uniform PPP on $[0, 1]^2$. Use the Kest and envelope functions to verify whether the empirical K function matches the theoretical one.
- 6.4** Find the K function for Matérn cluster processes.
- 6.5** Show that for doubly Poisson cluster processes,
- $$K(r) = \pi r^2 + \frac{F(r)}{\lambda_p},$$
- where $F(r)$ is the distribution of the distance between two points of a cluster.
- 6.6** Consider the mixed Poisson process on \mathbb{R}^d with two intensities λ_1 and λ_2 . Letting the randomized intensity be L , assume $\mathbb{P}(L = \lambda_1) = p$ and $\mathbb{P}(L = \lambda_2) = 1 - p$. Determine $K(r)$.
- 6.7** Calculate the intensity function, the variance $\text{var } \Phi(B)$ for any B and the K function of the Baddeley–Silverman process in Example 6.17 for unit square cells.
- 6.8** Explain the relationship between the Fry plot introduced in Section 2.8 and the K function.
- 6.9** Interference. Repeat Problem 5.2 using your knowledge about second-order product densities. Write out the squared sum

$$I^2 = \left(\sum_{x \in \Phi} h_x \ell(x) \right)^2, \quad \ell(x) = \min\{\|x\|^{-\alpha}, 1\},$$

explicitly (as a “central” part with a sum over Φ and a “non-central” part with a double-sum over Φ).

- (a) Calculate its mean (i.e., the second moment of the interference) using the second-order product density of the uniform Poisson point process.
- (b) Find an example of a non-uniform PPP, where you can calculate the mean interference and the second moment.
- (c) What happens if the point process is not Poisson? Explain why Campbell’s formula for the variance in Corollary 4.8 does not hold for non-Poisson processes.

6.10 Second moment measures for one-dimensional cluster processes.

- (a) Let $\Phi \subset \mathbb{R}$ be a Matérn cluster process with parent intensity $\lambda_p = 1/3$, mean number of points per cluster $c = 6$, and radius $R = 1$. Create a plot of a realization of $\Phi \times \Phi$ on $[0, 10]^2$, analogously to Fig. 6.1. How many points are there in the $[0, 10]^2$ square of your realization? How many do

you expect? How many do you expect if you consider $\Phi \star \Phi$?

- (b) Derive the second moment density $\varrho^{(2)}(x)$ of the one-dimensional Matérn and Thomas cluster processes (as a function of the parent PPP intensity λ_P , the average number of points per cluster \bar{c} , and the radius R for Matérn or the variance σ^2 for Thomas).
 - (c) Use simulations to validate the second moment measure for the Matérn case for the same values as in (a), except for λ_P , which should vary from 0 to 10. Plot the simulated curve together with the analytical one obtained in (b).
-

7 Marked point processes

7.1 Introduction and definition

It is often useful not only to consider a single class of points, but also to differentiate between different types of points, or, more generally, to include additional information about the points in the model. This is the motivation behind marked point processes, where each point x_i is assigned a random variable, the *mark* m_i .

Example 7.1 Examples of marks.

- In a forest, m_i may denote the diameter of the tree at x_i .
- In a wireless network, m_i may denote the battery level, the transmission power, the number of antennas, the peak transmit power constraint, or the queue length of node i ; or, if the point process models the transmitters, the mark may indicate the location of the corresponding receiver or receivers.
- For a MAC scheme (such as ALOHA or CSMA), $m_i \in \{0,1\}$ could be the transmit or receive state of the node.
- In a fading environment, the sequence (m_{ij}^j) could indicate the channel state between nodes i and j . In this case, a symmetry condition $m_{ji}^j = m_{ij}^i$ may be enforced.
- In a cognitive network, the mark could indicate whether the node belongs to the primary user or the secondary users; more generally, it may denote the priority class of a node.
- In a cellular network, the marks can be used to distinguish different service providers. An example for central London is shown in Fig. 7.1.
- In a heterogeneous network, the mark may indicate the type of the node (say a base station, a femtocell, a mobile user, or a WiFi access point). An example for this type of marked process is presented in Fig. 7.2.

Thus a marked point process on \mathbb{R}^d is a random set $\hat{\Phi} = \{(x_i, m_i)\}$, where the points x_i constitute an (unmarked) point process, the so-called *ground process*, and the m_i are the marks corresponding to each x_i . The marks m_i may have a complicated structure. They belong to a given *space of marks* or *mark space* \mathbb{M} .

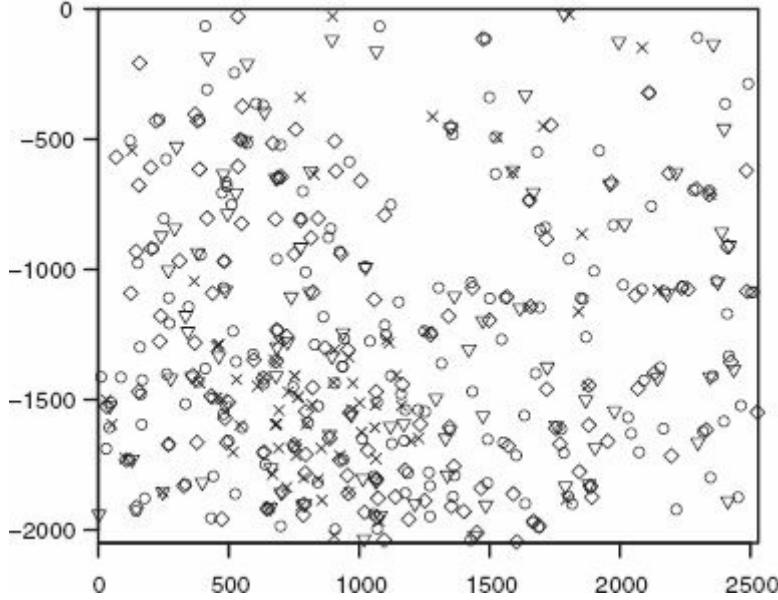


Figure 7.1 Marked point pattern of the locations of base stations in an area of about 5 km^2 in central London. The marks indicate the service provider; the symbols used are: \circ for O2 (179 points), \times for T-Mobile (77 points), ∇ for Orange (87 points), and \diamond for Vodafone (135 points). It can be seen that some base stations of different providers are co-located. As a result, the ground process is not a simple process in this example.

The Borel σ -algebra of \mathbb{M} is denoted by \mathcal{M} . In other words, in addition to the sequence (x_i) of points, we consider another sequence of random variables (m_i) , which are defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$, such that $m_i : \Omega \mapsto \mathbb{M}$.

The marks can be continuous variables, indicators, or represent random compact sets, as in the germ-grain model (see [Section 13.2.1](#)). The marked point process is said to be *independently marked* if the marks for different points are independent; they may depend on their location but not on the other points. So m_i may depend on x_i , but not on x_j or m_j for $i \neq j$ for independently marked processes. In this case, the *mark distribution* is denoted by M_x , which is a probability measure on $(\mathbb{M}, \mathcal{M})$.

If the marks are also identically distributed, we call the process *iid marked*. In this case, the mark distribution is independent of the location of the ground point and simply denoted by M .

It is always possible to interpret the marked point process as an ordinary point process on the product space $\mathbb{R}^d \times \mathbb{M}$. However, there is a particular feature of marked point processes which makes it worthwhile to consider them separately:

usually one defines Euclidean motions of marked point processes as transforms that move the points but leave the marks unchanged. So $\hat{\Phi}_x$, the translation of $\hat{\Phi}$ by x , is given by

$$\hat{\Phi}_x \triangleq \{(x_1 + x, m_1), (x_2 + x, m_2), \dots\}.$$

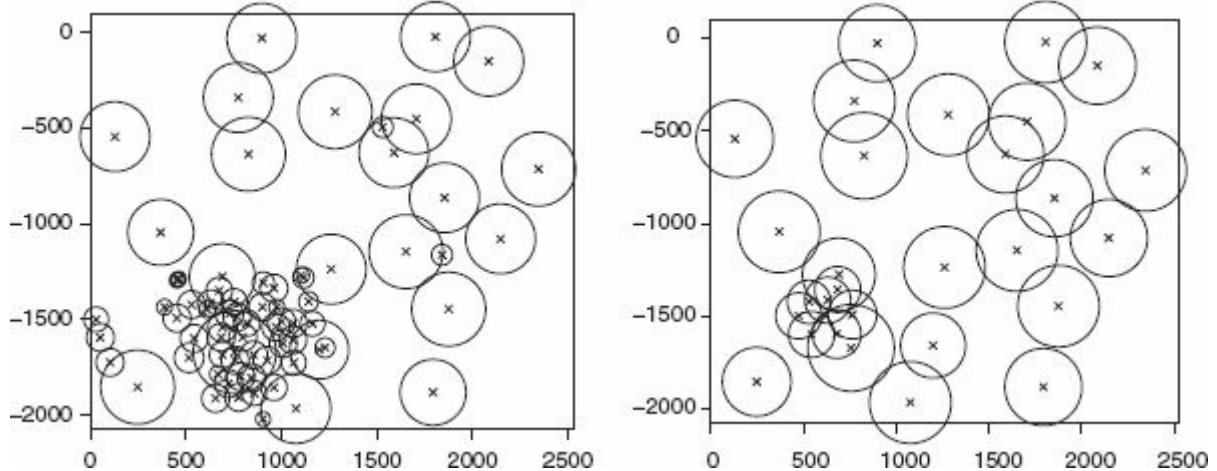


Figure 7.2 Marked point pattern of the locations of T-Mobile base stations in an area of about 5 km^2 in central London. Here the marks indicate the transmit power P ; the radius of the circle is proportional to $P^{1/4}$. (Left) GSM base stations. The power ranges from 0.8 W to 500 W. It is apparent that a small area is covered mostly by microcells. (Right) UMTS base stations. Here the minimum power is 6.3 W, and the maximum is 100 W.

Similarly, rotations act on marked point processes by rotating the points but *not* altering the marks.

The notion of a simple point process would change also if the marked process were considered as a regular one on the augmented space. By a simple marked point process, we mean that the points x_i are different (but different points may have the same mark). On the other hand, an ordinary point process on $\mathbb{R}^d \times \mathbb{M}$ would be simple even if $x_i = x_j$ for some i, j , as long as their marks m_i and m_j differ.

The formal definition is as follows.

DEFINITION 7.1 (Marked point process) A marked point process on \mathbb{R}^d with marks in \mathbb{M} is a point process $\hat{\Phi} = \{(x_i, m_i)\}$ on $\mathbb{R}^d \times \mathbb{M}$ such that $\hat{\Phi}(B \times \mathbb{M}) < \infty$ for all bounded Borel $B \subset \mathbb{R}^d$. The outcome space of $\hat{\Phi}$ is the space of all marked point sequences, denoted by \mathcal{N} , and \mathfrak{N} is the smallest σ -algebra on \mathcal{N} (the family of all marked point sequences) that makes every mapping

$$\hat{\varphi} \mapsto \hat{\varphi}(B \times L)$$

measurable for all Borel B and $L \in \mathcal{M}$. The distribution of a marked point process is denoted by $\hat{\Phi}: \mathfrak{N} \mapsto [0,1]$.

The marked point process resides in the space of counting measures on $\mathbb{R}^d \times \mathbb{M}$ or, equivalently, the space of sequences on $\mathbb{R}^d \times \mathbb{M}$ that are simple on \mathbb{R}^d . Expressed as a measure, we may write

$$\hat{N} = \hat{\Phi} = \sum_{i=1}^{\#\hat{\Phi}} \delta_{(x_i, m_i)},$$

where $\delta_{(x, m)}$ is the Dirac measure on $\mathbb{R}^d \times \mathbb{M}$.

It is sometimes more convenient to have the marks be indexed by their point x_i instead of the integer i . In this case, we write $\hat{\Phi} = \{(x_i, m_{x_i})\}$. Care must be taken if the point process undergoes Euclidean motions or other transformations. The marks will keep the original location of the point as their index: $\hat{\Phi}_x = \{(x_i + x, m_{x_i})\}$.

The notation for the random counting measure carries over from unmarked processes; for Borel B and $L \in \mathcal{M}$ the number of points in B with marks in L is denoted by $\hat{\Phi}(B \times L)$.

For each fixed $L \in \mathcal{M}$ the process

$$\Phi_{[L]} \triangleq \{x_i : (x_i, m_i) \in \hat{\Phi}, m_i \in L\} \quad (7.1)$$

is an unmarked process consisting of the points with marks from L , stripped of their marks. $\Phi_{[\mathbb{M}]}$ or $\Phi_{\mathbb{M}}$ is the complete unmarked process or ground process.

Remark 7.1 Expressions such as

$$\sum_{x_i \in \hat{\Phi}} x_i m_i$$

are notationally incorrect, since x_i is just the summation (dummy) variable, i.e., this is the same as $\sum_{x \in \Phi} xm_i$, so i is not defined. Instead, it is better to write

$$\sum_{i: x_i \in \hat{\Phi}} x_i m_i \quad \text{or} \quad \sum_{(x, m) \in \hat{\Phi}} xm.$$

Alternatively, if the ground point is used as the index of the mark,

$$\hat{\Phi} = \{(x_i, m_{x_i})\} = \{x \in \Phi : (x, m_x)\},$$

where $\Phi = \Phi_{\mathbb{M}}$ is the unmarked process. In this case, the sum is simply $\sum_{x \in \Phi} xm_x$.

Example 7.2 A uniform PPP on \mathbb{R}^3 cannot be viewed as a marked point process in

$\mathbb{R}^2 \times \mathbb{R}$ (with points in \mathbb{R}^2 and the third coordinate as the mark in \mathbb{R}), since $\hat{\Phi}(B \times \mathbb{R}) = \infty$ a.s. as soon as B is not empty.

Example 7.3 On the other hand, a PPP on $\mathbb{R}^2 \times [0, a]$ of intensity λ can be interpreted as a marked point process on \mathbb{R}^2 with marks from $\mathbb{M} = [0, a]$. The projected point process on \mathbb{R}^2 has intensity λa , since $\Lambda(B \times [0, a]) = \lambda a |B|$ for $B \subset \mathbb{R}^2$. The marks attached to each point are uniformly distributed on $[0, a]$.

Example 7.4 Let $\hat{\Phi} = \{(x_i, m_i)\}$ be an iid marked point process, where the ground process is a uniform PPP of intensity λ . The marks are drawn from a distribution function F_m on \mathbb{R}^+ and denote the random radius of a disk or ball centered at x_i . The set

$$\Xi = \bigcup_{i=1}^{\infty} b(x_i, m_i)$$

can then be viewed as a germ-grain model similar to the one in Example 3.7, but with random grain radii. Now let $\Phi' = \{x'_i\}$ be a uniform PPP of intensity λ' , and add binary marks $m'_i = 1_{\Xi}(x'_i)$ to form the marked process $\hat{\Phi}'$. Then the ground process $\Phi'_0 = \{x \in \mathbb{R}^d : (x, m) \in \hat{\Phi}', m = 0\}$ is an “inner city model” (see Example 3.7), whereas Φ'_1 is a Poisson hole process or “Swiss cheese model.” Both are Cox processes, with (random) intensity fields $\zeta_0(x) = \lambda'(1 - 1_{\Xi}(x))$ and $\zeta_1(x) = \lambda' 1_{\Xi}(x)$, respectively.

$\hat{\Phi}'$ is not independently marked, since the marks of nearby points are correlated.

Example 7.5 Can a cluster process be defined using an iid marked point process $\{(x_i, m_{x_i})\}$, with the parent points forming x_i and the marks being the daughter points?

Solution

Yes – if the cluster distribution does not depend on the location. In that case, $\mathbb{M} = \mathcal{N}$, and the marks denote the (untranslated) cluster points. If m_x is the cluster pertaining to the parent point $x \in \Phi_P$, then the cluster process is

$$\Phi = \bigcup_{x \in \Phi_P} m_x + x.$$

If the parents form a PPP, this cluster process is a Neyman–Scott process.

Example 7.6 Let $\Phi = \{x_i\}$ be a point process, and define the mark of point x as

$$m_x^{(\Phi)} = \{y \in \Phi : y - x\} = \Phi_{-x}. \quad (7.2)$$

This mark is called the *universal mark* of the point x , which is the point process seen from x . By construction, $o \in m_x^{(\Phi)}$ for all $x \in \Phi$ and all (non-empty) Φ . Also, if Φ is a hard-core process with hard-core radius r , for example, then $\#\{m_x^{(\Phi)} \cap b(o, r)\} = 1$ for all $x \in \Phi$.

The process $\hat{\Phi} = \{(x_i, m_{x_i})\}$ is not independently marked. Its significance will become apparent in the next chapter.

A marked process is said to be *stationary* if for all x the translated process Φ_x has the same distribution as Φ . It is *motion-invariant* if for all Euclidean motions (translations and rotations about the origin) \mathbf{m} , the process $\mathbf{m}\hat{\Phi}$ has the distribution of $\hat{\Phi}$. If the mark space is the finite set $[n]$ for some n , the marked point process is called a *multivariate* or *multitype* point process.

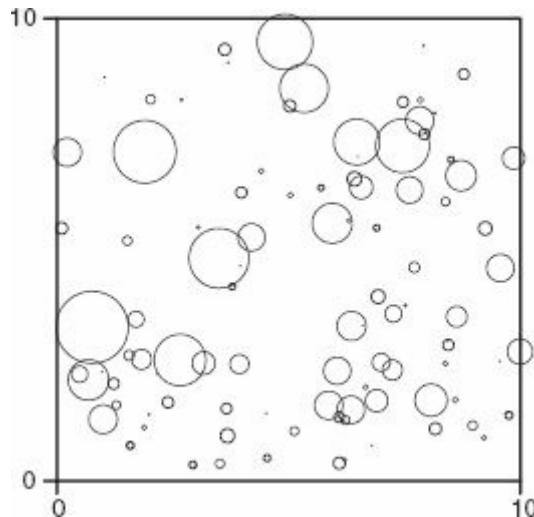


Figure 7.3 A marked uniform PPP with intensity 1 with independently exponentially distributed marks with mean $\bar{m} = 1$. The marks are indicated by the radii of the circles drawn around the points.

An example of an iid marked PPP with exponentially distributed marks is shown in Fig. 7.3.

7.2 Theory of marked point processes

7.2.1 Intensity measure and Campbell's theorem

The intensity measure $\hat{\Lambda}$ of a marked point process $\hat{\Phi}$ is defined as

$$\hat{\Lambda}(B \times L) \triangleq \mathbb{E}\hat{\Phi}(B \times L),$$

which is analogous to the (standard) intensity measure if $\hat{\Phi}$ was viewed as a (standard or non-marked) point process on $\mathbb{R}^d \times \mathbb{M}$. It can be expressed as

$$\hat{\Lambda}(B \times L) = \int_B M_x(L) \Lambda(dx), \quad B \subset \mathbb{R}^d, L \in \mathcal{M}.$$

This holds because

$$\begin{aligned} \hat{\Lambda}(B \times L) &\triangleq \mathbb{E} \left(\int_{\mathbb{R}^d} \int_{\mathbb{M}} \mathbf{1}(x \in B) \mathbf{1}(m \in L) \hat{\Phi}(dx, dm) \right) \\ &= \mathbb{E} \left(\int_{\mathbb{R}^d} \mathbf{1}(x \in B) M_x(L) \Phi(dx) \right) \\ &= \int_B M_x(L) \Lambda(dx). \end{aligned}$$

So $\hat{\Lambda}$ disintegrates as

$$\hat{\Lambda}(d(x, m)) = M_x(dm) \Lambda(dx), \quad (7.3)$$

which is consistent with the fact that the measure $\hat{\Lambda}(\cdot \times L)$ for fixed L in \mathbb{M} is absolutely continuous with respect to the intensity Λ of the ground process.

The marked counterpart of Campbell's theorem (for sums) follows from the definition of the intensity measure in a straightforward manner.

THEOREM 7.2 (Campbell's theorem for marked processes) *For any non-negative measurable function f on $\mathbb{R}^d \times \mathbb{M}$,*

$$\mathbb{E} \left(\sum_{(x, m) \in \hat{\Phi}} f(x, m) \right) = \int_{\mathbb{R}^d \times \mathbb{M}} f(x, m) \hat{\Lambda}(d(x, m)). \quad (7.4)$$

7.2.2 Stationary marked point processes

A marked point process $\hat{\Phi}$ is stationary if its ground process is stationary. If it is, then

$$\hat{\Lambda}(B \times L) = \mathbb{E}(\hat{\Phi}(B \times L)) = \mathbb{E}(\hat{\Phi}(B_x \times L)) = \hat{\Lambda}(B_x \times L) \quad \forall x \in \mathbb{R}^d.$$

If $L \in \mathcal{M}$ is fixed, $\hat{\Lambda}(\cdot \times L)$ is a translation-invariant measure. Since translation-invariant measures are proportional to the Lebesgue measure, we may write

$$\hat{\Lambda}(B \times L) = \lambda_{[L]} |B|,$$

where the quantity $\lambda_{[L]}$ is the intensity of $\Phi_{[L]}$ as defined in (7.1), i.e., the mean

number of points of $\hat{\Phi}$ per unit volume (of \mathbb{R}^d) with marks in L . Setting $L = \mathbb{M}$, we retrieve the intensity λ of the ground process:

$$\lambda_{\mathbb{M}} = \lambda.$$

As a function of L , $\lambda_{[L]}$ is a measure on $(\mathbb{M}, \mathcal{M})$; the quotients $\lambda_{[L]}/\lambda$ provide the mark distribution M of the stationary marked point process $\hat{\Phi}$:

$$M(L) = \frac{\lambda_{[L]}}{\lambda}. \quad (7.5)$$

Note that we have not assumed independent marking. We may still denote the conditional (marginal) distribution of mark m_x given x by M_x . Then the distribution M in (7.5) is the mark distribution of a “randomly chosen” point or *typical point*, in a sense that is made precise in the next chapter. The intensity measure satisfies

$$\hat{\Lambda} = \lambda \nu_d \otimes M. \quad (7.6)$$

If the ground process is ergodic, the mark distribution of the typical point can be expressed as

$$M(L) = \lim_{W_n \uparrow \mathbb{R}^d} \frac{\hat{\Phi}(W_n \times L)}{\hat{\Phi}(W_n \times \mathbb{M})} \quad (7.7)$$

for sequences (W_n) as in Definition 2.30.

In the stationary case, Campbell’s theorem for sums takes the following form.

THEOREM 7.3 (Campbell’s theorem for stationary marked processes) *For any measurable function $f: \mathbb{R}^d \times \mathbb{M} \mapsto \mathbb{R}^+$,*

$$\mathbb{E} \left(\sum_{(x,m) \in \hat{\Phi}} f(x, m) \right) = \lambda \int_{\mathbb{R}^d} \int_{\mathbb{M}} f(x, m) M(dm) dx.$$

In particular, for any Borel B with $|B| = 1$ and non-negative measurable functions h on \mathbb{M} ,

$$\mathbb{E} \left(\sum_{(x,m) \in \hat{\Phi}} 1_B(x) h(m) \right) = \lambda \int_{\mathbb{M}} h(m) M(dm).$$

If the marks are iid, Campbell’s theorem simplifies to the following corollary.

COROLLARY 7.4 (Campbell’s theorem for stationary iid marked processes) *For a stationary iid marked point process, (7.4) reduces to*

$$\mathbb{E} \left(\sum_{(x,m) \in \hat{\Phi}} f(x,m) \right) = \lambda \mathbb{E}_M \left(\int_{\mathbb{R}^d} f(x,m) dx \right), \quad (7.8)$$

where m is a random variable distributed with mark distribution M , and \mathbb{E}_M is the corresponding expectation.

In the case of real-valued marks the following characteristics are useful. The *mark distribution* F_M is given by

$$F_M(x) = \int_{\mathbb{R}} \mathbf{1}_{(-\infty, x]}(m) M(dm),$$

and the *mean mark* is

$$\bar{m} = \int_{\mathbb{R}} x dF_M(x).$$

If the marks are positive then the *mark sum measure* S_m can be considered:

$$S_m(B) = \sum_{(x,m) \in \hat{\Phi}} \mathbf{1}_B(x)m \quad \text{for Borel } B.$$

The mean of this random measure satisfies

$$\mathbb{E} S_m(B) = \lambda \bar{m} |B|.$$

7.2.3 The marking theorem

THEOREM 7.5 (Marking theorem for Poisson point processes) *Let $\hat{\Phi}$ be a marked point process on $\mathbb{R}^d \times \mathbb{M}$, and let Φ be the ground process. Then the following two statements are equivalent.*

1. Φ is a Poisson process on \mathbb{R}^d with intensity measure Λ , and, given Φ , the marks (m_x) are independent with distribution M_x on \mathbb{M} .
2. $\hat{\Phi}$ is a Poisson process on $\mathbb{R}^d \times \mathbb{M}$ with intensity measure $\hat{\Lambda} = \Lambda \otimes M_{(\cdot)}$, i.e., for $A \subset \mathbb{R}^d \times \mathbb{M}$,

$$\hat{\Lambda}(A) = \iint_{(x,m) \in A} \Lambda(dx) M_x(dm).$$

Proof We need to show that $\hat{\Phi}$ is a Poisson process on $\mathbb{R}^d \times \mathbb{M}$ with the given intensity measure. To this end, we determine the characteristic functional of the sum

$$S = \sum_{x \in \Phi} f(x, m_x)$$

for any measurable f . Given Φ , S is the sum of independent random variables $f(x, m_x)$, hence

$$\begin{aligned}\mathbb{E}(e^{-S} | \Phi) &= \prod_{x \in \Phi} \mathbb{E}(e^{-f(x, m_x)} | \Phi) \\ &= \prod_{x \in \Phi} \int_M e^{-f(x, m)} M_x(dm).\end{aligned}$$

Letting

$$\tilde{f}(x) = -\log \int_M e^{-f(x, m)} M_x(dm),$$

we have $\mathbb{E}(e^{-S} | \Phi) = \exp(-\sum_{x \in \Phi} \tilde{f}(x))$, and we can use the characteristic functional of the PPP (see Definition 4.7) to obtain

$$\begin{aligned}\mathbb{E}(e^{-S}) &= \exp\left(-\int_{\mathbb{R}^d} (1 - e^{-\tilde{f}(x)}) \Lambda(dx)\right) \\ &= \exp\left(-\int_{\mathbb{R}^d} \int_M (1 - e^{-f(x, m)}) \Lambda(dx) M_x(dm)\right) \\ &= \exp\left(-\int_{\mathbb{R}^d \times M} (1 - e^{-f(\hat{x})}) \hat{\Lambda}(d\hat{x})\right),\end{aligned}$$

where $d\hat{x}$ is the differential element of $\mathbb{R}^d \times M$. This shows that $\hat{\Phi}$ is a PPP with intensity measure $\hat{\Lambda}$. \square

The theorem also holds if the marks depend on the location of their point, as long as they remain independent. A consequence of the theorem is that the marks (m_x) themselves also form a PPP,¹ and that each sub-process $\Phi_{[L]}$, which includes the points x for which $m_x \in L$, is also a PPP for all $L \in \mathcal{M}$. Moreover, for disjoint L_i , the sub-processes $\Phi_{[L_i]}$ are independent.

As an application, we prove the displacement theorem for PPPs (Theorem 2.33). *Proof* (of Theorem 2.33) Let Φ be the original point process and denote by y_x the new position of the point originally at $x \in \Phi$. Recall that the displacement vectors are independent, and that the pdf of the new position x is given by $\rho(x, \cdot)$. From these assumptions, the marked process

$$\hat{\Phi} = \{(x, y_x) : x \in \Phi\}$$

satisfies the first statement in the marking theorem and thus also the second one, which means that it is a PPP on \mathbb{R}^{2d} . The mapping theorem implies that $\Phi' = \{y_x : x \in \Phi\}$ is a PPP on \mathbb{R}^d . The likelihood that a point of Φ' is located in the

infinitesimal volume dy at y if its original location was x is $\lambda(x)\rho(x, y)dy$, so the intensity measure of the displaced process is

$$\Lambda'(B) = \int_B \int_{\mathbb{R}^d} \lambda(x)\rho(x, y)dx dy,$$

and the intensity is given by (2.10). If $\lambda(x) = \lambda$ and $\rho(x, y) = h(x - y)$,

$$\begin{aligned}\lambda'(y) &= \lambda \int_{\mathbb{R}^d} h(y - x)dx = \lambda \int_{\mathbb{R}^d} h(x)dx \\ &= \lambda \int_{\mathbb{R}^d} \rho(0, x)dx = \lambda,\end{aligned}$$

□

since $\rho(x, \cdot)$ is a pdf. So in this case the intensity remains unchanged.

7.2.4 Moment measures

Moment measures for marked point processes can be defined in a fashion similar to the unmarked case. The second-order factorial moment measure is defined as

$$\hat{\alpha}^{(2)}(B_1 \times B_2 \times L_1 \times L_2) = \int_{\hat{\mathcal{N}}} \sum_{\substack{(x, m) \in \hat{\Phi} \\ (x', m') \in \hat{\Phi}}}^{\neq} \mathbf{1}_{B_1}(x)\mathbf{1}_{B_2}(x')\mathbf{1}_{L_1}(m)\mathbf{1}_{L_2}(m')\hat{P}(d\hat{\varphi}), \quad (7.9)$$

where the sum is taken over all pairs (x, m) and (x', m') in $\hat{\Phi}$ but excluding the cases where $x = x'$ since the summation is Σ^{\neq} . $\hat{\mathcal{N}}$ is the space of simple sequences (or counting measures) on $\mathbb{R}^d \times \mathbb{M}$.

A second-order statistic specific to marked processes is the *mark-correlation function* k_{mm} for a motion-invariant marked point process $\hat{\Phi} = \{(x_i, m_i)\}$ with marks from $\mathbb{M} \subseteq \mathbb{R}$. Heuristically, $k_{mm}(r)$ is a normalized mean of the product of marks at two positions separated by a distance r , under the condition that there are indeed points of $\hat{\Phi}$ in these two positions. Because of the motion-invariance it suffices to consider the positions o and x , with $r = \|x\|$. So

$$k_{mm}(r) \triangleq \frac{\mathbb{E}_{o,r}(m_o m_x)}{\bar{m}^2}, \quad (7.10)$$

where $\mathbb{E}_{o,r}$ is the expectation subject to the conditioning that $\hat{\Phi}$ has points at positions o and x with marks m_o and m_x , and \bar{m} is the mean mark. Note that k_{mm} differs from standard correlation coefficients in that it is always positive since the means are not subtracted in the numerator (it is not a covariance). So the value of k_{mm} if the marks at m_o and m_x are iid is 1 – in stark contrast to the correlation coefficient, where 1 indicates full correlation.

Let $\alpha^{(2)}$ be the second-order factorial moment measure of the unmarked point

process $\Phi = \{x_i\}$ and $\hat{\alpha}^{(2)}$ be the corresponding measure for the marked process, given by (7.9). Then, for fixed Borel subsets L_1 and L_2 of $[0, \infty)$, the measure $\hat{\alpha}^{(2)}(\cdot \times \cdot \times L_1 \times L_2)$ is absolutely continuous with respect to $\alpha^{(2)}$. If $\alpha^{(2)}$ is σ -finite, then according to the Radon-Nikodým theorem there is a function M_{x_1, x_2} such that

$$\hat{\alpha}^{(2)}(B_1 \times B_2 \times L_1 \times L_2) = \int_{B_1 \times B_2} M_{x_1, x_2}(L_1 \times L_2) \alpha^{(2)}(d(x_1, x_2)).$$

For fixed x_1 and x_2 , M_{x_1, x_2} can be interpreted as a measure on $\mathbb{M} \times \mathbb{M}$, the “two-point mark distribution,” and $M_{x_1, x_2}(d(m_1, m_2))$ is the joint pdf of the two marks pertaining to the points x_1 and x_2 .

Under the assumption of motion-invariance, M_{x_1, x_2} depends only on $r = \|x_1 - x_2\|$, and the simpler notation M_r is used. In this case,

$$k_{mm}(r) = \int_{\mathbb{M}^2} \frac{m_1 m_2 M_r(d(m_1, m_2))}{\bar{m}^2}.$$

Example 7.7 In a marked point process modeling a wireless network, where the ground points represent transmitter locations and the marks indicate the transmit power levels, chosen according to some power control algorithm. If the mark-correlation function $k_{mm}(r)$ is small at small distances r , that means that nearby transmitters choose small power levels (smaller than the average).

If the marks are binary and indicate whether a node is transmitting or not, $k_{mm}(r)$, again for small r , reveals whether two nearby nodes are more or less likely to transmit. As r increases, $k_{mm}(r)$ usually approaches 1, since marks tend to become less correlated over larger distances in many applications.

7.3 Applications

7.3.1 Modeling non-simple processes as simple marked ones

Instead of permitting multiple points at a single location in a point process, we can turn a non-simple process into a simple one by adding a mark from mark space $\mathbb{M} = \mathbb{N}_0$ to indicate the multiplicity of points at a location.

Example 7.8 Take a point process $\Phi \subset \mathbb{R}^2$ and denote by u_x and v_x the projection of $x \in \Phi$ onto the x - and y -axes, respectively, such that $x = (u_x, v_x)$. Now quantize

it to a point process N_q , expressed as a random counting measure, by setting

$$N_q = \sum_{x \in \Phi} \delta_{(\lfloor u_x \rfloor, \lfloor v_x \rfloor)},$$

where $\lfloor u \rfloor$ denotes the largest integer smaller than or equal to u . Since N_q is not a simple process, we cannot represent it as a random set. But we can transform it into a simple marked process with ground process $\Phi_q = \mathbb{Z}^2$ and marks

$$m_x = N_q(\{x\}), \quad \text{for each } x \in \Phi_q,$$

which denote the multiplicity of the point x in N_q .

7.3.2 Independent and dependent thinning

Let $\hat{\Phi}$ be a marked point process and Φ its ground process, with $\Phi_{[L]}$ as defined in (7.1).

In wireless networks, the set of transmitters in a given time slot is a subset of all nodes Φ . In point process terminology, the MAC scheme performs thinning on the point process of all nodes to obtain the point process of transmitters. If the points are equipped with binary marks denoting the transmit/receive state, the transmitter process is $\Phi_{[\{1\}]}$. Interference calculations can then be carried out using sums or products over $\Phi_{[\{1\}]}$.

Independent thinning

ALOHA performs independent thinning, so the marks (m_i) are independent Bernoulli random variables with mean p .

Dependent thinning

If the transmitters form a hard-core process with radius r , the transmit marks may be defined as

$$m_x = 1(\Phi \cap b(x, r) = \{x\}).$$

The corresponding MAC scheme is a CSMA-type scheme, which avoids having nearby nodes transmit concurrently.

Inspired by the Matérn hard-core process of type II, a process of higher density with the same hard-core radius can be defined using an additional mark, so that $\hat{\Phi} = \{(x_i, t_{x_i}, m_{x_i})\} \subset \mathbb{R}^2 \times [0,1] \times \{0,1\}$. Let (t_x) , $x \in \Phi$, be independent marks,

uniform on $[0,1]$, and set

$$m_x = \mathbf{1} \left(t_x = \min_{y \in \Phi \cap b(x,r)} \{t_y\} \right). \quad (7.11)$$

Then $\Phi_{[[0,1] \times \{1\}]}$ is the corresponding hard-core process of transmitters.

7.3.3 Bipolar network models

The Poisson bipolar network was introduced in Definition 5.8. Since each transmitter has its associated receiver, it can be conveniently described as a marked point process, where the transmitters form the ground process, and the receivers' locations are given by the marks. In principle, the process could be defined as $\hat{\Phi} = \{(x_i, m_{x_i})\}$ with the mark denoting the (absolute) location of the receiver. The disadvantage is that the receiver would not move with the transmitter if $\hat{\Phi}$ were translated. Hence it is better to use the mark to define the difference vector between receiver and transmitter, which results in an iid marked process. Letting (m_x) be iid with the pdf $f: \mathbb{R}^2 \rightarrow \mathbb{R}^+$, the receivers then form the point process $\{x_i + m_{x_i}\}$. Since f is general, this is a general bipolar model. To retrieve the original Poisson bipolar model, we choose the ground process $\Phi = \{x_i\}$ to be a PPP and f to be uniform on the circle of radius r .

Bibliographical notes

The monograph by Jacobsen (2006) is dedicated to marked point processes; it mainly focuses on the one-dimensional case. Other references include Sections 6.4 and 7.3 of Daley & Vere-Jones (2003) and Section 13.4 of Daley & Vere-Jones (2008). Chapter 5 of Kingman (1993) discusses marked PPPs and establishes connections to coloring.

In Chapter 2 of Baccelli & Blaszczyszyn (2009), there is an introduction to marked point processes that also includes a definition of shot noise and interference based on marked processes, and Chapter 10 in the same volume focuses on stationary marked point processes. Chapter 16 defines the Poisson bipolar model, and uses it to analyze the performance of regular and opportunistic ALOHA.

Problems

7.1 Define the K function (for stationary processes) using the universal mark $m_x^{(\Phi)}$.

7.2 Determine the mark distribution of the marked point process in Example 7.8 if Φ is a uniform PPP of intensity λ on the plane. Repeat for the case where Φ is a perturbed square lattice (see Definition 2.16) with iid perturbation vectors that are uniform on $b(o, 1/2)$.

7.3 Let Φ be a PPP of intensity λ and equip it with marks

$$m_x = \Phi(b(x, R)) - 1, \quad x \in \Phi$$

for $R > 0$ to obtain a marked point process. Show that the mark-correlation function is

$$k_{mm}(r) = 1 + \frac{|b(o, R) \cap b(r, R)|}{\pi r^2}.$$

7.4 For the marked process described by (7.11), describe the properties of $\Phi_{[[0,p] \times \{1\}]}$ if the ground process Φ is a PPP of intensity λ . What MAC scheme would result in this process?

7.5 Give a formal proof of Campbell's theorem for stationary marked point processes (7.8).

7.6 Describe the superposition of the transmitter and receiver points in the Poisson bipolar network model as a cluster process.

7.7 Consider a wireless network where the transmitters form a stationary point process of intensity λ . Each transmitter has a receiver at unit distance, and the channel to that receiver is subject to Rayleigh fading, which is independent for all transmitter–receiver pairs. The transmitter knows the channel state and chooses its transmit power such that the received power is 1 (channel inversion). The power level is attached to the transmitter point to form an iid marked process $\hat{\Phi} = \{(x_i, p_i)\}$, where p_i is the power at transmitter x_i . Determine the mean interference power at each receiver, assuming a path loss law $\ell(x) = \min\{1, \|x\|^{-4}\}$. Repeat for fractional channel inversion, where $p_i = 1/\sqrt{h_i}$ for iid exponential random variables (h_i) with mean 1.

¹ If the mark distribution is such that $m_x = m_y$ with non-zero probability for $x \neq y$, this PPP is not diffuse and therefore not simple, which is why we write (m_x) instead of $\{m_x\}$.

8 Conditioning and Palm theory

8.1 Introduction

8.1.1 Conditioning and the typical point

The *Palm probability* or *Palm measure* in point process theory is the probability of an event given that the point process contains a point at some location. It also formalizes the notion of the “typical point” of the process. Informally, the typical point results from a selection procedure in which every point has the same chance of being selected. This idea needs to be made mathematically precise, especially in infinite point processes. For example, a point chosen according to some sampling procedure, such as the one closest to the origin, is *not* typical, because it has been selected in a specific, deterministic manner. Intuitively, the Palm distribution is the conditional point process distribution given that a point (the typical point) exists at a specific location.

This type of conditioning is sometimes referred to as *interior conditioning*, since the conditioning is on $x \in \Phi$ and the question is how the point process looks outside of x . In contrast, the Papangelou conditional intensity is based on *exterior conditioning*, since the conditioning is on $\mathbb{R}^d \setminus \{x\}$, and the question is how likely it is to have a point at x . The two concepts are dual to each other.

If the point process is atomic, as the die process, conditioning on having a point at the location of one of the atoms causes no difficulty.

Example 8.1 Denote by P_o the distribution of the die process from Example 2.1 conditioned on a point at the origin $o \in \mathbb{R}^2$. Let E be the event that the total number of points $\Phi(\mathbb{R}^2)$ is odd. Then we have $P(E) = 1/2$ but $P_o(E) = 1$. Similarly, denoting the expectation with respect to P_o by \mathbb{E}_o , we have $\mathbb{E}\Phi(b((1, 1), \in)) = 1/2$ and $\mathbb{E}_o\Phi(b((1, 1), \in)) = 1/3$ for $0 < \in < 1$.

The difficulty arises when considering diffuse point processes (or when conditioning on a non-atomic location in an atomic point process that is a superposition of a diffuse and a purely atomic one), since, in this case, the event that we condition on has probability 0. Even then, such conditioning can be quite intuitive, as the following example illustrates.

Example 8.2 Consider the probability space $(\Omega, \mathcal{B}^2, \nu_2)$ with $\Omega = \{(\omega_1, \omega_2) \in [0, 1]^2\}$. The event $\omega_2 = 0$ has probability zero, but conditioning on it simply results in the probability mass being concentrated on the line segment $\hat{\Omega} = \{\omega_1 \in [0, 1] : (\omega_1, 0)\}$, which is akin to just the uniform distribution in the interval $[0, 1]$. So for instance we have $\mathbb{P}((\omega_1, \omega_2) \in [1/3, 2/3] \times 0) = 0$, while $\mathbb{P}((\omega_1, \omega_2) \in [1/3, 2/3] \times \{0\} \mid \omega_2 = 0) = 1/3$.

Applying the same reasoning to point processes, if we want to condition on a point at the origin, we can express the effect of conditioning by stating that the Palm measure P_x , which is obtained by conditioning on a point at x , is concentrated on the set

$$\hat{\Omega} = \{\omega \in \Omega : N^\omega(\{x\}) > 0\}.$$

Despite this concentration, there will in general still be uncountably many possible realizations of a diffuse point process – in the very same way as there is an uncountable set of outcomes in Example 8.2 after conditioning. An exception is the randomly translated lattice. If conditioned on a point at o , the location of the lattice is fixed, and only a single (deterministic) realization is possible.

Since an event $Y \in \mathfrak{N}$, such as having three points in the ball $b(o, 1)$ can be viewed as a *property* of the point process, the two terms event and property may be used interchangeably. In this chapter, we mainly use the term property.

Notation

We will use the symbol \parallel as a shortcut when conditioning on a point at a certain location:

$$\begin{aligned}\mathbb{P}(\Phi \text{ has property } Y \parallel x) &\triangleq \mathbb{P}(\Phi \text{ has property } Y \mid x \in \Phi) \\ &= \mathbb{P}(\Phi \in Y \mid x \in \Phi) \\ &= \mathbb{P}(Y \mid x \in \Phi) \\ &= \mathbb{P}(Y \parallel x) \\ &= \mathbb{P}_x(Y).\end{aligned}$$

For stationary processes,

$$\mathbb{P}(\Phi \in Y \| x) = \mathbb{P}(\Phi_x \in Y \| o),$$

where $\Phi_x \triangleq \{x_1 + x, x_2 + x, \dots\}$. Note that $\mathbb{P}(\Phi_x \in Y \| o)$ is to be understood as $\mathbb{P}(\Phi_x \in Y | o \in \Phi)$, not $\mathbb{P}(\Phi_x \in Y | o \in \Phi_x)$. So, for stationary point processes, conditioning may be restricted to $o \in \Phi$ without loss of generality, and the only Palm measure we need is P_o .

For $Y \in \mathcal{N}$, let

$$Y_x \triangleq \{\varphi \in \mathcal{N}: \varphi_{-x} \in Y\}$$

be the translated event or property. Then

$$\mathbb{P}(\Phi_x \in Y \| o) = \mathbb{P}(\Phi \in Y_{-x} \| o).$$

Example 8.3 In a stationary point process, the nearest-neighbor distance distribution may be expressed as

$$G(r) = \mathbb{P}(\Phi(b(o, r)) > 1 \| o) = 1 - \mathbb{P}(\Phi(b(o, r)) = 1 \| o),$$

and, with $Y_r = \{\varphi \in \mathcal{N}: \varphi(b(o, r)) = 1\}$,

$$G(r) = 1 - P_o(Y_r \| o) = 1 - P_o(Y_r).$$

By conditioning on the point at o , $P_o(\Phi(b(o, r)) = 0) = 0$ for all $r > 0$.

Expectations with respect to the Palm distribution are called Palm expectations and denoted by \mathbb{E}_x :

$$\mathbb{E}_x f(\Phi) \triangleq \int_{\mathcal{N}} f(\varphi) P_x(d\varphi).$$

8.1.2 Heuristic derivation

Let $h: \mathbb{R}^d \times \mathcal{N} \rightarrow \mathbb{R}^+$ be a measurable function. We wish to evaluate

$$\mathbb{E} \left(\sum_{x \in \Phi} h(x, \Phi) \right) = \int_{\mathcal{N}} \sum_{x \in \varphi} h(x, \varphi) \mathbb{P}(\mathrm{d}\varphi). \quad (8.1)$$

If \mathbb{R}^d is partitioned into domains D_1, D_2, \dots , each with non-zero volume, then we can write

$$\mathbb{E} \left(\sum_{x \in \Phi} h(x, \Phi) \right) = \sum_k \mathbb{E} \left(\sum_{x \in \Phi \cap D_k} h(x, \Phi) \mid \Phi(D_k) > 0 \right) \cdot \mathbb{P}(\Phi(D_k) > 0).$$

Now suppose $D_k \rightarrow dx$. Then $\mathbb{P}(\Phi(D_k) > 0) \rightarrow \Lambda(dx)$, and the conditional mean should converge to the mean $\mathbb{E}(h(x, \Phi) \parallel x)$ of $h(x, \varphi)$ with respect to $\mathbb{P}(\cdot \parallel x)$. Hence we obtain

$$\mathbb{E} \left(\sum_{x \in \Phi} h(x, \Phi) \right) = \int_{\mathbb{R}^d} \mathbb{E}(h(x, \Phi) \parallel x) \Lambda(dx).$$

For stationary Φ , $\Lambda = \lambda \nu_d$, so we have

$$\begin{aligned} \mathbb{E} \left(\sum_{x \in \Phi} h(x, \Phi) \right) &= \lambda \int_{\mathbb{R}^d} \mathbb{E}(h(x, \Phi_x) \parallel o) dx \\ &= \lambda \int_{\mathbb{R}^d} \int_{\mathcal{N}} h(x, \varphi_x) \mathbb{P}_o(\mathrm{d}\varphi) dx \end{aligned} \quad (8.2)$$

since Φ_x has a point at x if Φ has a point at o . This result is called the Campbell-Mecke theorem, which will be formally stated and proved in the next section.

Example 8.4 We would like to calculate the number of points in a set B whose neighbors are all at distance at least r using (8.2). To do that, we let

$$Y \triangleq \{\varphi \in \mathcal{N} : \varphi(b(o, r)) = 1\}$$

and for the function h in (8.1), we set

$$h(x, \varphi) \triangleq \mathbf{1}_B(x) \mathbf{1}_Y(\varphi_{-x}) = \mathbf{1}_B(x) \mathbf{1}_{Y_x}(\varphi)$$

for some bounded Borel set B . Since $\mathbf{1}_Y(\varphi_{-x}) = 1$ if the point x does not have any neighbor within distance r , we may write the mean number of points in B whose neighbors are all at distance at least r as

$$\mathbb{E} \left(\sum_{x \in \Phi} h(x, \Phi) \right) = \int_{\mathcal{N}} \sum_{x \in \varphi} h(x, \varphi) \mathbb{P}(\mathrm{d}\varphi),$$

and, using (8.2), we obtain

$$\int_{\mathcal{N}} \sum_{x \in \varphi \cap B} \mathbf{1}_Y(\varphi_{-x}) \mathbb{P}(d\varphi) = \lambda |B| \mathbb{P}_o(Y).$$

Next we need to discuss how to calculate the Palm measure \mathbb{P}_o , i.e., we need to construct a distribution on $(\mathcal{N}, \mathfrak{N})$ with the desired behavior of \mathbb{P}_o . There are two approaches to the Palm distribution, a local and a global one.

8.1.3 The local approach

As discussed previously, the conditioning event $o \in \Phi$ has probability zero in diffuse point processes. Nonetheless, if the limit

$$\lim_{\epsilon \downarrow 0} \frac{\mathbb{P}(Y \cap Y_\epsilon)}{\mathbb{P}(Y_\epsilon)}, \quad (8.3)$$

where $Y_\infty = \{\varphi \in \mathcal{N}: \varphi(b(o, \infty)) = 1\}$, exists, it corresponds to the Palm probability $\mathbb{P}_o(Y)$. Only in a few cases, however, can this limit be calculated explicitly.

Example 8.5 For the uniform PPP, the conditional probability

$$G_\infty(r) = 1 - \mathbb{P}(\Phi(b(o, r) \setminus b(o, \infty)) = 0 \mid \Phi(b(o, \infty)) = 1)$$

is well defined for $0 < \infty < r$, since

$$\mathbb{P}(\Phi(b(o, \infty)) = 1) = \lambda c_d \epsilon^d \exp(-\lambda c_d \epsilon^d) > 0.$$

We have

$$\begin{aligned} G_\infty(r) &= 1 - \frac{\mathbb{P}(\Phi(b(o, r) \setminus b(o, \infty)) = 0) \mathbb{P}(\Phi(b(o, \infty)) = 1)}{\mathbb{P}(\Phi(b(o, \infty)) = 1)} \\ &= 1 - \mathbb{P}(\Phi(b(o, r) \setminus b(o, \infty)) = 0) \\ &= 1 - \exp(-\lambda c_d(r^d - \infty^d)). \end{aligned}$$

By (8.3), $G(r) = \lim_{\infty \rightarrow 0} G_\infty(r)$, so

$$G(r) = 1 - \exp(-\lambda c_d r^d), r \geq 0.$$

This confirms that the spherical contact distribution function (or empty space function) and nearest-neighbor distance distribution for the stationary PPP are identical.

8.1.4 The global approach

If the point process is finite (and thus non-stationary), we may arrive at a definition of the Palm distribution by considering all points and averaging.

Example 8.6 We use the global approach to determine the nearest-neighbor distance distribution for the binomial point process. Let $\Phi = \{x_1, \dots, x_n\}$ be a uniform BPP on $W \subset \mathbb{R}^2$, and let $R_x, x \in \Phi$, be the distance to x 's nearest neighbor in $\Phi \setminus \{x\}$. What is $G(r) \triangleq \mathbb{P}(R_x \leq r \parallel x)$?

Solution

For each $i \in [n]$,

$$\begin{aligned}\mathbb{P}(R_x \leq r \mid x_i = x) &= 1 - \mathbb{P}(R_x > r \mid x_i = x) \\ &= 1 - \mathbb{P}(\Phi^{!x_i}(b(x, r)) = 0),\end{aligned}$$

where $\Phi^{!x_i} = \Phi \setminus \{x_i\}$ is a binomial point process with $n - 1$ points. Hence

$$\mathbb{P}(R_x \leq r \mid x_i = x) = 1 - \left(\frac{|W \setminus b(x, r)|}{|W|} \right)^{n-1}.$$

This is independent of the index i , as expected. It is plausible to interpret it as the conditional probability $\mathbb{P}(R_x \leq r \parallel x)$ and thus the nearest-neighbor distance distribution function $G(r)$.

For a stationary point process, we may average over the points falling into a test set B with $|B| > 0$. Then we explore whether the point process, seen from each point $x \in \Phi \cap B$, has property Y , i.e., we check whether $\Phi_{-x} \in Y$.

The Palm probability for the property Y is then defined as the fraction of points x expected to fall in B such that $\Phi_{-x} \in Y$. By stationarity this does not depend on B if $|B|$ is fixed. Analogously, using the idea of the universal mark (7.2), we may define the Palm distribution as the distribution of this mark. The resulting probabilities

quantify how likely it is that the point process viewed from the typical point has a given property.

For example, if under the Palm distribution the distance from o to the nearest point has an exponential distribution, this is equivalent to saying that the distance of the typical point to its nearest neighbor is exponentially distributed.

Now we are in a position to formally define the Palm distribution, first for stationary processes, then for the general case.

8.2 The Palm distribution for stationary processes

8.2.1 Definition

Let Φ be a stationary point process with finite intensity $\lambda < 0$.

DEFINITION 8.1 (Palm distribution for stationary point processes) The *Palm distribution* or *Palm measure* (at o) of a stationary point process Φ is a distribution defined on $(\mathcal{N}, \mathfrak{N})$ by

$$\begin{aligned} P_o(Y) &\triangleq \frac{1}{\lambda|B|} \mathbb{E} \left(\sum_{x \in \Phi} \mathbf{1}_Y(\Phi_{-x}) \mathbf{1}_B(x) \right) \quad \text{for } Y \in \mathfrak{N} \\ &= \frac{1}{\lambda|B|} \int_{\mathcal{N}} \sum_{x \in \varphi \cap B} \mathbf{1}_Y(\varphi_{-x}) P(d\varphi). \end{aligned} \tag{8.4}$$

This definition does not depend on B – it can be an arbitrary Borel set of finite positive volume. All points in B get translated to the origin of the coordinate system – this is expressed by Φ_{-x} – and it is checked whether the translated realization belongs to Y . If it is, it is counted in the sum. The expected sum is normalized by the mean number of points in B .

Another interpretation for this definition can be given using marked point processes. To each point $x \in \Phi$ give a mark 1 or 0 depending on whether the shifted process Φ_{-x} belongs to Y or not. For example, consider (again) $Y = \{\varphi \in \mathcal{N} : \varphi(b(o, r)) = 1\}$. Then x has mark 1 precisely when its nearest neighbor is further than r away. The result is a stationary marked point process with mark distribution M on $\{0, 1\}$, and we can write

$$P_o(Y) = M(\{1\}) = \frac{\lambda_{[1]}}{\lambda},$$

where $\lambda_{[1]}$ is the intensity of the points whose mark is 1. The mean number of points of Φ with mark 1 in B is $\lambda_{[1]}|B|$, so the above definition makes $P_o(Y)$ independent of the test set B .

Alternatively, the notation $\lambda(Y)$ may be used to denote the density of points x for which $\Phi_{-x} \in Y$. So we may write

$$\mathbb{P}_o(Y) = \frac{\lambda(Y)}{\lambda}.$$

8.2.2 The Campbell–Mecke theorem

The derivation leading to (8.2) suggests that the following result, the so-called *Campbell–Mecke theorem* (or *refined Campbell theorem*), holds for stationary point processes.

THEOREM 8.2 (Campbell–Mecke theorem) *For any non-negative measurable function h on $\mathbb{R}^d \times \mathcal{N}$,*

$$\mathbb{E} \left(\sum_{x \in \Phi} h(x, \Phi) \right) = \lambda \int_{\mathbb{R}^d} \int_{\mathcal{N}} h(x, \varphi_x) \mathbb{P}_o(d\varphi) dx.$$

We defer the proof to the next section. The reason why the Palm distribution becomes necessary in the integral expression becomes apparent when writing the theorem in the form

$$\int_{\mathcal{N}} \sum_{x \in \varphi} h(x, \varphi) \mathbb{P}(d\varphi) = \lambda \int_{\mathbb{R}^d} \int_{\mathcal{N}} h(x, \varphi_x) \mathbb{P}_o(d\varphi) dx.$$

The sum on the left-hand side is over $h(x, \varphi)$, where $x \in \varphi$. Without the Palm distribution, this connection between x and φ would be lost on the right-hand side, since x would be just a generic location on \mathbb{R}^d , while φ would be a generic counting measure.

Taking $h(x, \varphi) = f(x)$ reproduces the standard Campbell theorem (for sums). The theorem can also be written in the form

$$\mathbb{E} \left(\sum_{x \in \Phi} h(x, \Phi) \right) = \lambda \mathbb{E}_o \left(\int_{\mathbb{R}^d} h(x, \Phi_x) dx \right),$$

where \mathbb{E}_o is the Palm expectation.

8.2.3 Inversion formula

In the stationary case, the point process distribution can be retrieved from the Palm distribution.

THEOREM 8.3 (Palm inversion formula) *Let Φ be a stationary point process of intensity λ , and let $V(o)$ be the Voronoi cell (see Definition 2.13) of the point at o (given that there is a point at o). Then*

$$\mathbb{E}f(\Phi) = \lambda \mathbb{E}_o \left(\int_{\mathbb{R}^d} f(\Phi_{-x}) \mathbf{1}(x \in V(o)) dx \right).$$

Proof Let

$$h(x, \varphi) = \mathbf{1}(\varphi(\tilde{b}(o, \|x\|)) = 0),$$

where $\tilde{b}(o, r)$ is the *open* ball of radius r with center o . We have

$$\int_{\mathbb{R}^d} h(x, \Phi) \Phi(dx) = \sum_{x \in \Phi} h(x, \Phi) = 1 \quad \text{a.s.},$$

since a.s. there is a unique nearest point to the origin in a simple stationary process. It follows that

$$\begin{aligned} \mathbb{E}f(\Phi) &= \mathbb{E} \left(\int_{\mathbb{R}^d} f(\Phi_x - x) h(x, \Phi_x - x) \Phi(dx) \right) \\ &\stackrel{(a)}{=} \lambda \int_{\mathbb{R}^d} \mathbb{E}_o(f(\Phi_x) h(x, \Phi_x)) dx \\ &= \lambda \int_{\mathbb{R}^d} \mathbb{E}_o(f(\Phi_{-x}) h(-x, \Phi_{-x})) dx \\ &\stackrel{(b)}{=} \lambda \mathbb{E}_o \left(\int_{\mathbb{R}^d} f(\Phi_{-x}) \mathbf{1}(x \in V(o)) dx \right), \end{aligned}$$

where (a) follows from the Campbell–Mecke theorem and (b) since $h(-x, \Phi_{-x}) = 1$ for those locations $x \in \mathbb{R}^d$ that are closer to the origin than any point in Φ_{-x} . But with a point added at o , these locations correspond to the Voronoi cell of this point.

Example 8.7 Setting $f(\varphi) \equiv 1$, we obtain $1 = \lambda \mathbb{E}_o |V(o)|$, so the mean area or volume of the *typical* Voronoi cell is $1/\lambda$.

Example 8.8 Let V_φ^o be the Voronoi cell of φ that contains the origin and set $f(\varphi) = |V_\varphi^o|^{-1}$. Using the theorem,

$$\begin{aligned} \mathbb{E}(|V_\Phi^o|^{-1}) &= \lambda \mathbb{E}_o(|V(o)|^{-1} \cdot |V(o)|) \\ &= \lambda \\ &= (\mathbb{E}|V(o)|)^{-1}, \end{aligned}$$

which, with Jensen's inequality, shows that $\mathbb{E}|V_\Phi^o| \geq \mathbb{E}|V(o)|$, which is a spatial

incarnation of the *waiting time paradox*: The origin is more likely to fall into a large Voronoi cell than into a small one. The waiting time paradox is often stated in conjunction with Poisson models, but our derivation shows that it is a general phenomenon.

8.3 The Palm distribution for general point processes

For general point processes, we need a different approach to define the Palm distribution. It is based on the Campbell measure \mathcal{C} and the Radon–Nikodým theorem.

8.3.1 The Campbell measure

DEFINITION 8.4 (Campbell measure) The *Campbell measure* \mathcal{C} is defined as a measure on $(\mathbb{R}^d \times \mathcal{N}, \mathcal{B}^d \times \mathfrak{N})$ by

$$\mathcal{C}(B \times Y) \triangleq \mathbb{E}(\Phi(B)\mathbf{1}_Y(\Phi)),$$

for bounded Borel sets B and $Y \in \mathfrak{N}$.

With the Campbell measure, the Campbell–Mecke theorem may be expressed as

$$\int_{\mathcal{N}} \sum_{x \in \varphi} f(x, \varphi) \mathbb{P}(\mathrm{d}\varphi) = \int_{\mathbb{R}^d \times \mathcal{N}} f(x, \varphi) \mathcal{C}(\mathrm{d}(x, \varphi)).$$

Proof (of the Campbell–Mecke theorem) First let $f: \mathbb{R}^d \times \mathcal{N} \mapsto \mathbb{R}^+$ be the indicator function $f(x, \varphi) = 1((x, \varphi) \in B \times Y)$ for some bounded Borel set B and some $Y \in \mathfrak{N}$. Then

$$\begin{aligned} \mathbb{E} \left(\sum_{x \in \Phi} f(x, \Phi) \right) &= \mathbb{E}(\Phi(B)\mathbf{1}(\Phi \in Y)) \\ &= \mathcal{C}(B \times Y) \\ &= \int_{\mathbb{R}^d} \int_{\mathcal{N}} \mathbf{1}((x, \varphi) \in B \times Y) \mathcal{C}(\mathrm{d}(x, \varphi)) \\ &= \int_{\mathbb{R}^d} \int_{\mathcal{N}} f(x, \varphi) \mathcal{C}(\mathrm{d}(x, \varphi)). \end{aligned}$$

Using the linearity of expectation and integration, the theorem holds for step functions. By the monotone convergence theorem, it also holds for functions of the form $\sum_{i=1}^n \beta_i \mathbf{1}((x, \varphi) \in A_i)$, where $A_i \in \mathcal{B}^d \times \mathfrak{N}$. By taking limits, this extends to \square

arbitrary non-negative f .

The Campbell measure is a generalization of the intensity measure Λ since $\mathcal{C}(B \times \mathcal{N}) = \Lambda(B)$. As a consequence, $\mathcal{C}(B \times Y) \leq \Lambda(B)$. For fixed Y , let $\mu_Y(B) \triangleq \mathcal{C}(B \times Y)$. Since μ_Y is a measure and $\mu_Y \leq \Lambda$, μ_Y is absolutely continuous with respect to Λ , i.e., $\mu_Y \ll \Lambda$, and by the Radon–Nikodým theorem, there exists a density f_Y such that

$$\mu_Y(B) = \int_B f_Y(x) \Lambda(dx)$$

for Borel B where $f_Y : \mathbb{R}^d \rightarrow \mathbb{R}^+$ is measurable (and unique up to equality Λ -a.s.). This density f_Y is the Radon–Nikodým derivative $d\mu_Y/d\Lambda$. It is used for the general definition of the Palm distribution.

8.3.2 General definition of the Palm distribution

DEFINITION 8.5 (Palm distribution for general point processes) The Palm distribution $P_x(Y)$ is defined as

$$P_x(Y) \triangleq f_Y(x)$$

where $f_Y(x)$ is the density pertaining to $\mathcal{C}(\cdot \times Y)$, such that

$$\mathcal{C}(B \times Y) = \int_B f_Y(x) \Lambda(dx) = \int_B P_x(Y) \Lambda(dx). \quad (8.5)$$

More compactly, the Palm distribution is the Radon–Nikodým derivative

$$P_x(Y) \triangleq \frac{d\mathcal{C}(\cdot \times Y)}{d\Lambda}.$$

For each x , $P_x(\cdot)$ is a distribution (probability measure) on $(\mathcal{N}, \mathfrak{N})$. The definition is consistent with the intuition we initially developed when interpreting the Palm measure as the limit (8.3). Indeed, as $\epsilon \rightarrow 0$, we have

$$\begin{aligned} \mathbb{P}(\Phi \in Y \mid \Phi(b(x, \epsilon)) > 0) &= \frac{\mathbb{P}(\Phi \in Y, \Phi(b(x, \epsilon)) > 0)}{\mathbb{P}(\Phi(b(x, \epsilon)) > 0)} \\ &\sim \frac{\mathbb{E}(\Phi(b(x, \epsilon)) \mathbf{1}_Y(\Phi))}{\mathbb{E}(\Phi(b(x, \epsilon)))} \\ &= \frac{\mathcal{C}(b(x, \epsilon) \times Y)}{\Lambda(b(x, \epsilon))} \\ &\sim P_x(Y). \end{aligned}$$

Application to stationary point processes

The general definition of course includes the stationary case, where we have

$$P_o(Y) = P_x(Y_x) \quad \text{for } Y \in \mathfrak{N},$$

which follows from

$$\begin{aligned} \lambda \int_B P_z(Y) dz &= \mathcal{C}(B \times Y) = \mathcal{C}(B_x \times Y_x) \\ &= \lambda \int_{B_x} P_y(Y_x) dy \\ &= \lambda \int_B P_{x+z}(Y_x) dz \end{aligned}$$

for all B , x , and Y . This implies $P_z(Y) = P_{x+z}(Y_x)$. Furthermore, we can retrieve the Campbell–Mecke theorem:

$$\begin{aligned} \mathbb{E} \left(\sum_{x \in \Phi} h(x, \Phi) \right) &= \int_{\mathbb{R}^d \times \mathcal{N}} h(x, \varphi) \mathcal{C}(dx, \varphi) \\ &= \lambda \int_{\mathbb{R}^d} \int_{\mathcal{N}} h(x, \varphi) P_x(d\varphi) dx \\ &= \lambda \int_{\mathbb{R}^d} \int_{\mathcal{N}} h(x, \varphi_x) P_o(d\varphi) dx. \end{aligned}$$

Example 8.9 Let $Y = \{\varphi \in \mathcal{N}: \varphi(\{o\}) > 0\}$ be the set of simple sequences that have a point at o . What is $P_o(Y)$?

Solution

Denoting by $\Phi_Y(B)$ the number of points $x \in \Phi \cap B$ for which $\Phi_{-x} \in Y$, we have

$$\begin{aligned} P_o(Y) &= \frac{\lambda(Y)}{\lambda} = \frac{\mathbb{E}\Phi_Y([0, 1]^d)}{\lambda} \\ &= \frac{\mathbb{E}\#\{x \in \Phi \cap [0, 1]^d: \mathbf{1}_Y(\Phi_{-x})\}}{\lambda} \\ &= \frac{\mathbb{E}\Phi([0, 1]^d)}{\lambda} \\ &= 1. \end{aligned}$$

This is of course to be expected, since we are conditioning on $o \in \Phi$.

For stationary point processes of intensity λ , λP_o is the density (with respect to the Lebesgue measure) pertaining to the Campbell measure.

If Φ is ergodic, we can obtain the Palm distribution from the limit

$$\mathbb{P}_o(Y) = \lim_{a \rightarrow \infty} \frac{\Phi_Y([-a, a]^d)}{\Phi([-a, a]^d)}.$$

$\Phi_y(B)$ is defined in Example 8.9.

Alternative definition of the Campbell measure

The Campbell measure can also be expressed in terms of the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ as a measure on $(\mathbb{R}^d \times \Omega, \mathcal{B}^d \times \mathcal{A})$. In this case,

$$C(B \times A) = \int_B \mathbb{P}_x(A) \Lambda(dx), \quad B \in \mathcal{B}^d, A \in \mathcal{A}, \quad (8.6)$$

where $\mathbb{P}_x(A)$ is the Palm measure on (Ω, \mathcal{A}) . This is more general than $\mathbb{P}_x(Y)$ since it is defined for all events $A \in \mathcal{A}$, not just for the ones in \mathcal{N} . This formulation is useful if the statistics of Φ depend on a random variable that is defined on the original probability space, as illustrated in the following example.

Example 8.10 Consider the mixed Poisson point process defined in Section 2.4.5. Let L be a non-negative random variable defined on Ω . Given $L = \lambda$, Φ is a homogeneous PPP of intensity λ . Let $A = \{L \leq v\}$ for some $v \geq 0$. For $B \subset \mathbb{R}^d$, we have

$$\begin{aligned} C(B \times A) &= \mathbb{E}(\Phi(B)\mathbf{1}(L \leq v)) \\ &= \mathbb{E}(\mathbb{E}(\Phi(B)\mathbf{1}(L \leq v) \mid L)) \\ &= \mathbb{E}(L|B|\mathbf{1}(L \leq v)) \\ &= |B|\mathbb{E}(L\mathbf{1}(L \leq v)). \end{aligned}$$

From (8.6),

$$\begin{aligned} C(B \times A) &= \int_B \mathbb{P}_x(A) \Lambda(dx) \\ &= \mathbb{P}_x(A)|B|\mathbb{E}L, \end{aligned}$$

and thus

$$\mathbb{P}_x(L \leq v) = \frac{\mathbb{E}(L\mathbf{1}(L \leq v))}{\mathbb{E}L}.$$

Hence the distribution of L under \mathbb{P}_x is skewed compared with its original distribution. *The event $L \leq v$ is less likely to have occurred if Φ has a point at x .*

As an extreme case, consider the case where L assumes only two values, 0 with

probability $1 - p$ and 1 with probability p . So $\mathbb{P}(L \leq v) = (1 - p)u(v) + pu(v - 1)$, where $u(x) = 1(x \geq 0)$ is the unit step function. In contrast, the Palm distribution is $\mathbb{P}_x(L \leq v) = u(v - 1)$. This is intuitive since if the PPP has a point at x , it cannot have intensity 0.

As another example, consider $\mathbb{P}(L \leq v) = 1 - \exp(-v)$. The Palm distribution of L is given by

$$\mathbb{P}_x(L \leq v) = 1 - \exp(-v) - v \exp(-v).$$

These are examples where the event A in the Campbell measure $\mathcal{C}(\cdot \times A)$ is an event in the original probability space.

Intensity conditioned on event

We have interpreted $P_x(Y)$ as the limit of the probability $P(Y | \Phi(b(x, \in)) > 0)$ as $\in \downarrow 0$. Using Bayes' theorem,

$$\mathbb{P}(\Phi(b(x, \epsilon)) > 0 | \Phi \in Y) = \frac{\mathbb{P}(\Phi(b(x, \epsilon)) > 0)}{\mathbb{P}(\Phi \in Y)} \mathbb{P}(\Phi \in Y | \Phi(b(x, \epsilon)) > 0)$$

and, as $\in \downarrow 0$,

$$\frac{\mathbb{P}(\Phi(b(x, \epsilon)) > 0 | \Phi \in Y)}{\mathbb{P}(\Phi(b(x, \epsilon)) > 0)} \rightarrow \frac{P_x(Y)}{P(Y)}.$$

If Φ has an intensity function $\lambda(x)$, $\mathbb{P}(\Phi(b(x, \in)) > 0) \sim \lambda(x)|b(x, \in)|$, and

$$\frac{\mathbb{P}(\Phi(b(x, \epsilon)) > 0 | \Phi \in Y)}{|b(x, \epsilon)|} \rightarrow \lambda(x) \frac{P_x(Y)}{P(Y)}.$$

So the right side can be interpreted as the conditional intensity of the point process given Y . This conditional intensity, however, must not be confused with the Papangelou conditional intensity, where the conditioning is on the point process; the two concepts are rather different, as shown in the next section.

Isotropy of Palm distributions

Clearly the Palm distribution P_o is never a stationary distribution since under P_o a point process must always contain o , i.e., it becomes atomic. However, if the point process Φ is motion-invariant then its Palm distribution is isotropic, i.e.,

$$P_o(Y) = P_o(rY) \text{ for } Y \in \mathfrak{N}$$

for every rotation r about the origin.

Example 8.11 Consider a randomly translated and rotated lattice Φ . Conditioned on $o \in \Phi$, the lattice is no longer stationary, but it is still isotropic. One of the four nearest neighbors of the origin is bound to lie at an angle between 0 and $\pi/2$, with uniform distribution. If a second point is added in the conditioning, say $(0,1) \in \Phi$ (if the underlying lattice is \mathbb{Z}^2), then the Palm distribution equals the (degenerate) distribution of \mathbb{Z}^2 . In general, such two-fold Palm distributions can be derived from higher-order Campbell measures.

8.4 The reduced Palm distribution

When calculating Palm probabilities, it is often more natural not to consider the point that the point process is conditioned on. In the context of wireless networks, for example, assuming all transmitting nodes form a point process, one may want to identify one of them as the desired transmitter, while the others are interferers. When conditioning on the location of the desired transmitter, its power should not be included in the interference.

8.4.1 Definition

In the reduced Palm distribution, denoted by $P_x^!$, the point at x on which we condition is not included in the distribution:

$$P_x^!(Y) \triangleq \mathbb{P}(\Phi \setminus \{x\} \in Y \mid x) \quad \text{for } Y \in \mathfrak{N}$$

In particular,

$$P_o^!(Y) = \mathbb{P}(\Phi \setminus \{o\} \in Y \mid o) \quad \text{for } Y \in \mathfrak{N}$$

The corresponding expectation is the expectation with respect to the reduced Palm distribution, denoted as $E_o^!$, i.e.,

$$\mathbb{E}_o^! f(\Phi) \triangleq \int_{\mathcal{N}} f(\varphi) P_o^!(d\varphi).$$

In the stationary case, the definition has a form analogous to the (non-reduced) Palm distribution, given in Definition 8.1.

DEFINITION 8.6 (Reduced Palm distribution for stationary processes)

$$\mathbb{P}_o^!(Y) = \frac{1}{\lambda|B|} \int_{\mathcal{N}} \sum_{x \in \varphi \cap B} \mathbf{1}_Y(\varphi_{-x} \setminus \{o\}) \mathbb{P}(d\varphi).$$

In the general case, we define first the so-called *reduced* version of the Campbell measure.

DEFINITION 8.7 (Reduced Campbell measure) The reduced Campbell measure $\mathcal{C}^!$ is defined as

$$\mathcal{C}^!(B \times Y) \triangleq \mathbb{E} \left(\sum_{x \in \Phi \cap B} \mathbf{1}(\Phi \setminus \{x\} \in Y) \right).$$

On replacing \mathcal{C} by $\mathcal{C}^!$ in (8.5), we arrive at the definition of the *reduced Palm distribution* $\mathbb{P}_x^!$.

DEFINITION 8.8 (Reduced Palm distribution) The reduced Palm distribution $\mathbb{P}_x^!$ is defined by the relationship

$$\mathcal{C}^!(B \times Y) = \int_B \mathbb{P}_x^!(Y) \Lambda(dx).$$

or, equivalently, as the Radon–Nikodým derivative

$$\mathbb{P}_x^!(Y) \triangleq \frac{d\mathcal{C}^!(\cdot \times Y)}{d\Lambda}$$

of the reduced Campbell measure with respect to the intensity measure.

Example 8.12 The nearest-neighbor distribution function can be expressed via \mathbb{P}_o or $\mathbb{P}_o^!$ by

$$\begin{aligned} G(r) &= 1 - \mathbb{P}_o(\{\varphi \in \mathcal{N}: \varphi(b(o, r)) = 1\}) \\ &= 1 - \mathbb{P}_o^!(\{\varphi \in \mathcal{N}: \varphi(b(o, r)) = 0\}). \end{aligned}$$

Example 8.13 Let Φ_n be a BPP with n nodes. Then the reduced Palm distribution of Φ_n is the distribution of Φ_{n-1} .

Box 8.1 Definition of Palm distribution and reduced Palm distribution

The Palm distribution is the Radon–Nikodým derivative of the Campbell measure, while the reduced Palm distribution is the Radon–Nikodým derivative of the reduced Campbell measure, each with respect to the intensity measure.

The reduced Campbell measure is sometimes also referred to as the modified Campbell measure. The corresponding reduced version of the (stationary) Campbell–Mecke theorem reads

$$\begin{aligned}\int_{\mathcal{N}} \sum_{x \in \varphi} f(x, \varphi \setminus \{x\}) P(d\varphi) &= \int_{\mathcal{N}} \sum_{x \in \varphi} f(x, \varphi - \delta_x) P(d\varphi) \\ &= \int_{\mathbb{R}^d \times \mathcal{N}} f(x, \varphi) \mathcal{C}^!(d(x, \varphi))\end{aligned}$$

for measurable $f: \mathbb{R}^d \times \mathcal{N} \mapsto \mathbb{R}^+$. Here, the point x is explicitly removed from φ before evaluating f . The notation $\varphi - \delta_x$ refers to the counting measure of the point pattern φ minus the point at x (see Box 2.3).

The Campbell–Mecke theorem (Theorem 8.2) is not restricted to stationary point processes. Here we give the general version without proof.

THEOREM 8.9 (Campbell–Mecke theorem for general point processes) *For general point processes Φ with mean measure Λ and measurable $h: \mathbb{R}^d \times \mathcal{N} \mapsto \mathbb{R}^+$,*

$$\mathbb{E} \left(\sum_{x \in \Phi} h(x, \Phi) \right) = \int_{\mathbb{R}^d} \mathbb{E}_x(h(x, \Phi)) \Lambda(dx).$$

The reduced version is

$$\mathbb{E} \left(\sum_{x \in \Phi} h(x, \Phi \setminus \{x\}) \right) = \int_{\mathbb{R}^d} \mathbb{E}_x^!(h(x, \Phi)) \Lambda(dx). \quad (8.7)$$

8.4.2 Connection to Papangelou conditional intensity

It can be shown that, under some mild conditions,¹ the reduced Campbell measure is absolutely continuous with respect to the point process distribution, i.e.,

$$\mathcal{C}^!(B \times \cdot) \ll P(\cdot) \quad \text{for each fixed } B.$$

This implies the existence of a measurable function $R(B, \cdot)$, the Radon–Nikodým derivative $d\mathcal{C}^!(B \times \cdot)/dP$, such that

$$\mathcal{C}^!(B \times Y) = \int_Y R(B, \varphi) \mathbb{P}(\mathrm{d}\varphi).$$

The function R is called the *Papangelou kernel*, whose density (with respect to the Lebesgue measure) is the Papangelou conditional intensity, i.e.,

$$R(B, \varphi) = \int_B \lambda(x, \varphi) \mathrm{d}x$$

Hence

$$\begin{aligned} \mathcal{C}^!(B \times Y) &= \int_B \int_Y \lambda(x, \varphi) \mathbb{P}(\mathrm{d}\varphi) \mathrm{d}x \\ &= \int_B \mathbb{E}(\lambda(x, \Phi) \mathbf{1}(\Phi \in Y)) \mathrm{d}x, \end{aligned}$$

which shows the connection between the reduced Campbell measure and the Papangelou conditional intensity.

As a consequence, for general non-negative measurable $h: \mathbb{R}^d \times \mathcal{N} \mapsto \mathbb{R}^+$,

$$\begin{aligned} \mathbb{E} \left(\sum_{x \in \Phi} h(x, \Phi \setminus \{x\}) \right) &= \int_{\mathbb{R}^d \times \mathcal{N}} h(x, \varphi) \mathcal{C}^!(\mathrm{d}(x, \varphi)) \\ &= \int_{\mathbb{R}^d \times \mathcal{N}} h(x, \varphi) R(\mathrm{d}x, \varphi) \mathbb{P}(\mathrm{d}\varphi) \\ &= \int_{\mathbb{R}^d} \mathbb{E}(h(x, \Phi) \lambda(x, \Phi)) \mathrm{d}x \\ &\stackrel{(a)}{=} \int_{\mathbb{R}^d} \mathbb{E}_x^!(h(x, \Phi)) \lambda(x) \mathrm{d}x, \end{aligned} \tag{8.8}$$

where (a) follows from the reduced version of the general Campbell–Mecke theorem given in (8.7). On comparing the last two lines, it follows that, for almost all x ,

$$\mathbb{E}_x^! h(x, \Phi) = \mathbb{E} \left(\frac{\lambda(x, \Phi)}{\lambda(x)} h(x, \Phi) \right), \tag{8.9}$$

or, writing out the expectations,

$$\int_{\mathcal{N}} h(x, \varphi) \mathbb{P}_x^!(\mathrm{d}\varphi) = \int_{\mathcal{N}} \frac{\lambda(x, \varphi)}{\lambda(x)} h(x, \varphi) \mathbb{P}(\mathrm{d}\varphi),$$

which shows that $\lambda(x, \Phi)/\lambda(x)$ is the Radon–Nikodým density of $\mathbb{P}_x^!$ with respect to \mathbb{P} . Indeed, setting $h(x, \varphi) = \mathbf{1}(\varphi \in Y)$ for an arbitrary $Y \in \mathfrak{N}$,

$$\int_Y \mathbb{P}_x^!(\mathrm{d}\varphi) = \mathbb{P}_x^!(Y) = \int_Y \frac{\lambda(x, \varphi)}{\lambda(x)} \mathbb{P}(\mathrm{d}\varphi).$$

Setting $h \equiv 1$ yields $\lambda(x) = \mathbb{E}\lambda(x, \Phi)$, which is intuitively plausible.

Equation (8.8) is usually referred to as the *Georgii–Nguyen–Zessin formula*. For stationary processes and a function h that depends only on the point process, it reduces to

$$\lambda \mathbb{E}_o^! h(\Phi) = \mathbb{E}(h(\Phi)\lambda(o, \Phi)). \quad (8.10)$$

These formulas, including (8.9), relate interior and exterior conditioning. On the left side, the mean value of $\lambda \mathbb{E}h(x, \Phi)$ is modified by the presence of the point at o (which is not considered), on the right, it is modified by the probability of having a point at o given Φ . The intuition is the following: If having a point at o increases the mean, then the mean is increased if it is more likely to have a point at o .

Example 8.14 For the BPP with n points, the Papangelou conditional intensity does not exist, since $P_x^!$ and P are not absolutely continuous but mutually singular. For example, the event of having exactly n points has probability 1 under P but probability 0 under $P_x^!$, and vice versa for the event of having $n - 1$ points.

Example 8.15 Letting $Y_r = \{\varphi \in \mathcal{N}: \varphi(b(o, r)) = 0\}$, the J function, introduced in Definition 2.40, can be expressed as

$$\begin{aligned} J(r) &= \frac{P_o^!(Y_r)}{P(Y_r)} \\ &\stackrel{(a)}{=} \frac{\mathbb{E}((\lambda(o, \Phi)/\lambda)1(\Phi \in Y_r))}{P(Y_r)} \\ &= \mathbb{E}\left(\frac{\lambda(o, \Phi)}{\lambda} \mid Y_r\right). \end{aligned}$$

Here (a) follows from the Georgii–Nguyen–Zessin formula. The last expression is often easier to evaluate than the ratio of the (complementary) nearest-neighbor distance and empty space functions, $(1 - G(r))/(1 - F(r))$.

8.5 Palm distribution for Poisson processes and Slivnyak's theorem

The independence property of the PPP states that the number of points in $b(x, \epsilon)$ is independent of the number of points in any region outside this ball, for arbitrarily small ϵ . This suggests that conditioning on a point at x does not change the distribution of the rest of the process. Slivnyak's theorem formalizes this statement.

THEOREM 8.10 (Slivnyak's theorem) *For the Poisson point process with intensity*

function λ , if $\lambda(o) > 0$,

$$\mathbb{P}(\Phi \in Y \mid o) = \mathbb{P}(\Phi \cup \{o\} \in Y).$$

More generally,

$$P_x = P * \delta_{\delta_x}, \quad \forall x \in \{y \in \mathbb{R}^d : \lambda(y) > 0\}.$$

More compactly,

$$P_x^! \equiv P.$$

Proof Let $\lambda(x) > 0$. Since P_x and $P * \delta_{\delta_x}$ are the distributions of simple processes, their equality is established if the corresponding void probabilities are equal, i.e., if

$$P * \delta_{\delta_x}(V_K) = P_x(V_K)$$

for all compact $K \subset \mathbb{R}^d$ where $V_K = \{\varphi \in \mathcal{N} : \varphi(K) = 0\}$.

Let B be any bounded Borel set. Then

$$\begin{aligned} \int_B P * \delta_{\delta_x}(V_K) \Lambda(dx) &= \int_{B \setminus K} P(V_K) \Lambda(dx) \\ &= P(V_K) \Lambda(B \setminus K) \\ &= \mathbb{E}(\mathbf{1}(\Phi(K) = 0)) \cdot \mathbb{E}(\Phi(B \setminus K)) \\ &= \mathbb{E}(\Phi(B \setminus K) \mathbf{1}(\Phi(K) = 0)) \\ &= \mathcal{C}((B \setminus K) \times V_K). \end{aligned}$$

Clearly $\mathcal{C}((B \cap K) \times V_K) = \mathbb{E}(\Phi(B \cap K) \mathbf{1}(\Phi(K) = 0)) = 0$. (If there was a point in the intersection the void indicator would be zero.) Hence

$$\int_B P * \delta_{\delta_x}(V_K) \Lambda(dx) = \mathcal{C}(B \times V_K) = \int_B P_x(V_K) \Lambda(dx)$$

using $\mathcal{C}(B \times Y) = \int_B P_x(Y) \Lambda(dx)$. □

The condition that the location x of the conditioning must be chosen from the support of λ is frequently written as “for Λ -almost all $x \in \mathbb{R}^d$.”

Slivnyak’s theorem is reversible, i.e., if $P_x^! \equiv P$ holds for a point process, it must be a PPP. This result is known as Mecke’s theorem. This important property is emphasized in Box 8.2.

Box 8.2 Conditioning property of the PPP

Conditioning on $x \in \Phi$ in a PPP is the same as adding a point at x . This can be extended to multiple points.

Slivnyak's theorem can be interpreted as

$$P_x(Y) = \mathbb{P}(\Phi \in Y \| x) = \mathbb{P}(\Phi \cup \{x\} \in Y) \quad \text{for } Y \in \mathfrak{N}$$

or

$$\int_{\mathcal{N}} f(\varphi) P_x(d\varphi) = \int_{\mathcal{N}} f(\varphi \cup \{x\}) \mathbb{P}(d\varphi)$$

for all measurable non-negative functions f .

8.6 Second moments and Palm distributions for stationary processes

For stationary processes Φ , the reduced Campbell measure can be expressed as

$$\mathcal{C}^!(B \times Y) = \mathbb{E}_o^!(\Phi(B) 1_Y(\Phi)),$$

and the distribution of Φ given that $x \in \Phi$ is the same as the distribution of Φ_x given that $o \in \Phi$, i.e.,

$$\Phi \| x \stackrel{d}{=} (\Phi \| o) + x = \Phi x \| o.$$

Using the Campbell–Mecke theorem,

$$\begin{aligned} \alpha^{(2)}(A \times B) &= \mathbb{E} \left(\sum_{x_1, x_2 \in \Phi}^{\neq} 1_A(x_1) 1_B(x_2) \right) \\ &= \int_{\mathcal{N}} \sum_{x \in \varphi} 1_A(x) \varphi(B \setminus \{x\}) \mathbb{P}(d\varphi) \\ &= \lambda \int_{\mathbb{R}^d} \int_{\mathcal{N}} 1_A(x) \varphi((B - x) \setminus \{o\}) \mathbb{P}_o(d\varphi) dx. \end{aligned}$$

With the reduced second moment measure, we can express this as

$$\alpha^{(2)}(A \times B) = \lambda \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} 1_A(x) 1_B(x + u) \mathcal{K}(du) dx \tag{8.11}$$

$$= \lambda \int_A K(B - x) dx. \quad (8.12)$$

So we may define the reduced second moment measure κ as follows.

DEFINITION 8.11 (Reduced second moment measure based on Palm distribution)

$$\kappa(B) \triangleq \int_N \varphi(B \setminus \{o\}) P_o(d\varphi) = \int_N \varphi(B) P_o^!(d\varphi) = E_o^! \Phi(B).$$

Hence, $\kappa(B)$ is the mean number of points in $\Phi \cap B \setminus \{o\}$ under the condition that $o \in \Phi$, i.e., it is the intensity measure of the reduced Palm distribution.

Box 8.3 Intensity measure of the reduced Palm distribution

The reduced second moment measure κ is the intensity measure pertaining to the reduced Palm distribution.

Accordingly, the K function can be defined compactly as

$$K(r) \triangleq (1/\lambda) E((\Phi b(o, r)) - 1 \mid o) = (1/\lambda) E_o^! \Phi(b(o, r)).$$

8.7 Palm distributions for Neyman–Scott cluster processes

Let Φ be a Neyman–Scott cluster process with distribution P and C_o the Palm distribution of the representative cluster Φ_0 . Then

$$P_o = P * C_o. \quad (8.13)$$

The conditioning here is on a point of the cluster process Φ at the origin, not a parent point (parent points are not included in the process). Equation (8.13) can be interpreted as an extension of Slivnyak's theorem. If the points of a cluster were located deterministically at $\{o, y_2, \dots, y_c\}$, we would have

$$P_o = P * \left(\delta_{\delta_o} + \sum_{i=2}^c \delta_{\delta_{y_i}} \right).$$

With the points of the representative cluster Φ_0 at $\{x_i\}$, $i \in [c]$, now random, the distribution of a cluster is

$$C = \sum_{i=1}^c \delta_{\delta_{x_i}},$$

and conditioning on $o \in \Phi_0$ requires the use of the corresponding Palm measure in (8.13).

If the distribution of the points in the representative cluster is atomic, the Palm distribution C_o is obtained by shifting each atom to the origin and taking the weighted average of the distributions of these shifted clusters. For example, consider the one-dimensional point process where the representative cluster has exactly two points at -1 and 1 . This case actually corresponds to a one-dimensional Gauss–Poisson process with $p_2 = 1$ (see Example 3.8). The Palm distribution C_o of such a cluster is the distribution of the point process with two outcomes, either $\{-2, 0\}$ or $\{0, 2\}$, each with probability $1/2$. So although the Gauss–Poisson process is not a Neyman–Scott process, (8.13) still applies.

Example 8.16 We calculate the nearest-neighbor distance distribution for the standard two-dimensional Gauss–Poisson process with parent intensity λ_p , where two-point clusters occur with probability p_2 and have the two points at distance 1 . Letting

$$Y = \{\phi \in \mathcal{N} : \phi(b(o, r)) = 1\}.$$

we have

$$C_o(Y) = \frac{p_1 + 2p_2 \mathbf{1}(r < 1)}{p_1 + 2p_2},$$

since, if $r < 1$, for sure there is only one point in the ball (given that $o \in \Phi_0$). If $r \geq 1$, it is $p_1/(p_1 + 2p_2) = p_1/\bar{c}$, where $\bar{c} = p_1 + 2p_2$ is the mean number of nodes per cluster. From (8.13), the nearest-neighbor distance distribution can in general be expressed as

$$1 - G(r) = P_o(Y) = \mathbb{P}(\Phi(b(o, r)) = 0) \cdot C_o(Y). \quad (8.14)$$

For the first term $\mathbb{P}(\Phi(b(o, r)) = 0)$ noting that

$$\mathbb{E}|\Phi_0 \oplus b(o, r)| = p_1 \pi r^2 + p_2 V_r(1),$$

where $V_r(1)$ is the area of the union of two disks of radius r at distance 1 , see (6.11), we have

$$\mathbb{P}(\Phi(b(o, r)) = 0) = \exp(-\lambda_p(p_1 \pi r^2 + p_2 V_r(1))).$$

8.8 Palm distribution for marked point processes

The definitions of these distributions follow those of unmarked point processes, using the Campbell measure and formula (8.5). For a marked point process $\hat{\Phi}$ with distribution \hat{P} , the Campbell measure is defined by

$$\hat{C}(B \times L \times Y) \triangleq \mathbb{E}(\hat{\Phi}(B \times L) \mathbf{1}_Y(\hat{\Phi})) = \int_{\hat{\mathcal{N}}} \hat{\varphi}(B \times L) \mathbf{1}_Y(\hat{\varphi}) \hat{P}(\mathrm{d}\hat{\varphi})$$

for Borel $B \subset \mathbb{R}^d$, $L \in \mathcal{M}$, and $Y \in \hat{\mathfrak{N}}$.

The measure $\hat{C}(\cdot \times \cdot \times Y)$ is for all Y absolutely continuous with respect to $\hat{\Lambda}$, hence there exists a measure $\hat{P}_{(x,m)}$ such that

$$\hat{C}(B \times L \times Y) = \int_{B \times L} \hat{P}_{(x,m)}(Y) \Lambda(\mathrm{d}(x, m)). \quad (8.15)$$

For fixed x and m , $\hat{P}_{(x,m)}$ is a distribution on $\hat{\mathfrak{N}}$. It can be interpreted as the probability of Y given that $\hat{\Phi}$ has a point at x of mark m .

In the stationary case, using (7.6), (8.15) takes the form

$$\hat{C}(B \times L \times Y) = \lambda \int_B \int_L \hat{P}_{(o,m)}(Y_{-x}) M(\mathrm{d}m) \mathrm{d}x, \quad (8.16)$$

where M is the mark distribution of $\hat{\Phi}$.

The Campbell–Mecke or refined Campbell theorem for a stationary marked point process is

$$\int_{\hat{\mathcal{N}}} \sum_{(x,m) \in \hat{\varphi}} h(x, m, \hat{\varphi}) \hat{P}(\mathrm{d}\hat{\varphi}) = \lambda \int_{\mathbb{R}^d} \int_M \int_{\hat{\mathcal{N}}} h(x, m, \hat{\varphi}_x) \hat{P}_{(o,m)}(\mathrm{d}\hat{\varphi}) M(\mathrm{d}m) \mathrm{d}x \quad (8.17)$$

for any non-negative measurable function h on $\mathbb{R}^d \times \mathbb{M} \times \hat{\mathcal{N}}$.

Again in the stationary case, suppose $L \in \mathcal{M}$. Then we can define the *Palm distribution of $\hat{\Phi}$ with respect to the mark set L* as \hat{P}_L , where

$$\hat{P}_L(Y) \triangleq \frac{1}{M(L)} \int_L \hat{P}_{(o,m)}(Y) M(\mathrm{d}m) \quad \text{for } Y \in \hat{\mathfrak{N}}.$$

This Palm distribution can be interpreted as the conditional distribution of $\hat{\Phi}$ given that $\hat{\Phi}$ has a point at o with mark in L .

The mark distribution M itself can be thought of as a Palm distribution. Set $Y_L = \{\hat{\varphi} \in \hat{\mathcal{N}} : (o, m) \in \hat{\varphi}, m \in L\}$. Then $\hat{P}(o, m | Y_L) = 1(m \in L)$ and thus

$$\hat{P}_M(Y_L) = \int_M \hat{P}_{(o,m)}(Y_L) M(dm) = M(L).$$

8.9 Applications

8.9.1 Mean interference in cluster processes

We would like to determine the mean interference at a receiver that is located in the vicinity of its *desired* transmitter, with all other transmitters contributing to the interference. This is a very natural application for the reduced Palm measure. Letting Φ be a stationary point process representing the locations of all transmitters, we condition on a transmitter at o and declare this transmitter to be the desired one, so its power is not included in the interference. Hence the mean interference is expressed as the expectation with respect to the reduced Palm distribution

$$\mathbb{E}_o^! I(z) = \mathbb{E}_o^! \sum_{x \in \Phi} \ell(x - z), \quad (8.18)$$

where z is the location of the receiver and the path loss function $\ell(x)$ is bounded and integrable, i.e., $\int_{\mathbb{R}^2} \ell(x) dx < \infty$. All transmit powers are set to unity. Fading could be included, but it does not change the mean interference as long as its mean is 1.

To get concrete results, we let Φ be a Thomas cluster process on the plane, and we compare the resulting interference with the interference in a uniform PPP of the same intensity.

By Campbell's theorem, we can express the sum (8.18) as an integral of the intensity measure. The intensity measure pertaining to the reduced Palm expectation is the reduced second moment measure, hence

$$\mathbb{E}_o^! \sum_{x \in \Phi} \ell(x - z) = \int_{\mathbb{R}^2} \ell(x - z) \mathcal{K}(dx) = \frac{1}{\lambda} \int_{\mathbb{R}^2} \ell(x - z) \varrho_{mi}^{(2)}(x) dx.$$

For the Thomas cluster process with \bar{c} points per cluster on average (see (6.6)),

$$\varrho_{mi}^{(2)}(x) = \lambda^2 + \frac{\lambda_p \bar{c}^2}{4\pi\sigma^2} \exp\left(-\frac{\|x\|^2}{4\sigma^2}\right),$$

so

$$\mathbb{E}_o^!(I(z)) = \lambda \int_{\mathbb{R}^2} \ell(x) dx + \frac{\bar{c}}{4\pi\sigma^2} \int_{\mathbb{R}^2} \ell(x - z) \exp\left(-\frac{\|x\|^2}{4\sigma^2}\right) dx.$$

The first term is the interference in the Poisson case. The second term is the contribution due to clustering, which is decreasing as $\|z\|$ grows, as expected. As

we get further away from the origin, we may or may not be inside a cluster, so we find the interference to be the same as under the regular (non-Palm) expectation, which is identical to the interference in the Poisson case:

$$\lim_{\|z\| \rightarrow \infty} \mathbb{E}_o^! I(z) = \mathbb{E} I(z) = \mathbb{E} I_{\text{PPP}}(z).$$

8.9.2 Mean interference in hard-core processes

Next we calculate the mean interference in hard-core processes. Here we focus on measuring the interference at the point of a transmitter, i.e., we do not have the displacement z we used in the clustered case. Again using the reduced second moment measure \mathcal{K} of the point process, we have

$$\mathbb{E}_o^!(I) = \int_{\mathbb{R}^2} \ell(x) \mathcal{K}(dx). \quad (8.19)$$

Since we are considering hard-core processes, the path loss function does not need to be bounded (at the origin), it merely needs to be integrable. For a radially symmetric path loss function $\tilde{\ell}(r)$ and isotropic point processes, a polar representation is more convenient:

$$\mathbb{E}_o^!(I) = 2\pi \int_0^\infty \tilde{\ell}(r) \mathcal{K}(r dr) = \lambda \int_0^\infty \tilde{\ell}(r) K'(r) dr. \quad (8.20)$$

Using Ripley's K function, we have $K'(r)dr = (2\pi/\lambda)\mathcal{K}(r dr)$. Next we define the *excess interference ratio* (EIR).

DEFINITION 8.12 The excess interference ratio (EIR) is the mean interference measured at the typical point of a stationary hard-core point process of intensity λ with minimum distance ρ relative to the mean interference in a Poisson process of intensity $\lambda(r) = \lambda 1_{[\rho, \infty)}(r)$:

$$\text{EIR} \triangleq \mathbb{E}_o^!(I)/\mathbb{E}_o^!(I_{\text{PPP}}). \quad (8.21)$$

We focus on the Matérn hard-core process of type I, where points from a stationary PPP of intensity λ_b are retained only if they are at distance at least ρ from all other points. Its K function was calculated in Example 6.14.

By inserting the derivative of (6.9), $K'(r)$, into (8.20), we obtain the mean interference

$$\mathbb{E}_o^!(I) = 2\pi\lambda_b \exp(\pi\lambda_b\rho^2) \int_\rho^\infty \tilde{\ell}(r) r \exp(-\lambda_b V_\rho(r)) dr.$$

We split the interference into two terms, comprising the interference from the nodes

closer than 2ρ and further than 2ρ , respectively: $I = I_{<2\rho} + I_{>2\rho}$. We focus first on $I_{<2\rho}$, i.e., the range $\rho \leq r \leq 2\rho$. In this range, $V_\rho(r)$ is increasing and concave, so it can be upper bounded by a first-order Taylor expansion at $r = 3\rho/2$ and lower bounded by connecting the points $V_\rho(\rho) = \rho^2(4\pi/3 + \sqrt{3}/2)$ and $V_\rho(2\rho) = 2\pi\rho^2$ by a straight line. This yields tight lower and upper bounds on $\mathbb{E}_o^!(I_{<2\rho})$

The interference from nodes outside $r > 2\rho$ is the same as in the (equi-dense) PPP:

$$\mathbb{E}_o^!(I_{>2\rho}) = 2\pi\lambda_b \frac{\exp(-\lambda_b\pi\rho^2)}{(2\rho)^{\alpha-2}(\alpha-2)}.$$

The total interference in the PPP is obtained by replacing the 2ρ in the denominator by ρ , $\mathbb{E}_o^!(I_{\text{PPP}}) = 2^{\alpha-2}\mathbb{E}_o^!(I_{>2\rho})$. For the excess interference ratio, we find

$$\text{EIR} = \frac{1}{2^{\alpha-2}} \left(\frac{\mathbb{E}_o^!(I_{<2\rho})}{\mathbb{E}_o^!(I_{>2\rho})} + 1 \right).$$

Bibliographical notes

[Chapter 13](#) of Daley & Vere-Jones (2008) gives a rigorous account of Palm theory, and brief introductions are given in [Section 4.4](#) of Stoyan *et al.* (1995) and [Section 9.2](#) of Baccelli & Blaszczyszyn (2009).

The Georgii-Nguyen-Zessin formula was introduced in Georgii (1976) and Nguyen & Zessin (1979).

The mean interference in Matérn hard-core processes of types I and II is calculated in Haenggi (2011). Palm theory is used extensively in Ganti & Haenggi (2009a) for the characterization of the interference and outage in clustered networks, in Giacomelli *et al.* (2011) for the outage analysis in general motion-invariant networks in the high-SIR regime, and in Ganti *et al.* (2011) for an extension of the transmission capacity framework to general point processes, also for high SIR. In Chapter 18 of Baccelli & Blaszczyszyn (2009), it is employed to analyze the performance of CSMA applied to PPPs.

Problems

8.1 Let Φ be the randomly rotated die process from Example 2.1, and let $B = [-1, 1]^2$. Determine $\mathbb{E}\Phi(B)$, $\mathbb{E}_o\Phi(B)$, and $\mathbb{E}_{(-1,1)}\Phi(B)$.

8.2 The die cluster process. Consider a Poisson cluster process with parent intensity λ_p , where each cluster is an independent and randomly rotated die process. Again, for $B = [-1, 1]^2$, find $\mathbb{E}\Phi(B)$, $\mathbb{E}_o\Phi(B)$, and $\mathbb{E}_{(-1,1)}\Phi(B)$ and sketch the K function

$K(r)$.

8.3 Consider the mixed Poisson point process on \mathbb{R}^2 , where λ is a Bernoulli random variable with mean p . Determine $\Lambda([0,1]^2) = \mathbb{E}\Phi([0, 1]^2)$ and $\mathbb{E}_o\Phi([0, 1]^2)$.

8.4 Is $\kappa(B) = \kappa(-B)$? Give evidence for or against.

8.5 Prove that for the (general) PPP, $\lambda(x, \Phi) \equiv \lambda(x)$.

8.6 Illustrate the stationary form of the Georgii-Nguyen-Zessin formula (8.10) using a hard-core process. Choose a function h for which it is apparent that the interior conditioning on the left side has the same effect as the scaling of the intensity.

8.7 Find the J function for Neyman-Scott cluster processes, expressed using the cluster Palm measure \mathcal{C}_o .

¹ Essentially, cases where the behavior of the point process inside B is deterministically controlled by the behavior outside B are excluded.

Part II

Percolation, connectivity, and coverage

9 Introduction

9.1 Motivation

In this part, we are concerned with certain global properties of a network or graph, such as the existence of a giant connected component. The main new mathematical tool that we discuss is percolation theory. Percolation theory started some 50 years ago as a mathematical framework to study the behavior of porous media. It has been used to address questions such as the following.

If a stone gets wet, does the water penetrate the stone?

If a material consists of two components, one of which is a perfect insulator, what is the probability that the resistance is finite?

If we drill for oil, what is the probability that a large number of oil chambers is connected to the one we drilled into?

What is the probability that a forest fire spreads across an entire forest?

What is the probability that a virus spreads globally?

What is the probability that most of a network is connected?

It turns out that certain key events, such as the existence of a giant connected component in a network, emerge rather suddenly as a network parameter is changed. Such phenomena are called *phase transitions*.

In the context of graphs or networks, percolation is related to connectivity and coverage. These are the other two topics in this part.

9.2 What is percolation?

The basic model for percolation is the two-dimensional lattice, i.e., the graph $\mathbb{L}^2 \triangleq G(V, E)$ with $V = \mathbb{Z}^2$ and $E = \{(x, y) \in V^2 : \|x - y\| = 1\}$. So $\deg v = 4 \quad \forall v \in V$. Now, we turn the lattice into a *random graph* by removing each edge with probability $1 - p$, independently of all other edges. Equivalently, we retain each edge with probability p . This is a so-called *bond percolation model*. If instead of

edges vertices are removed with probability $1 - p$ (together with their associated edges), the model is called a *site percolation model*.

The other ingredients we need are the (*connected*) *components*. For general graphs, they are defined as follows.

DEFINITION 9.1 (Connected component) A *connected component* (or often just *component*) is the maximal set of vertices and edges such that, for any two vertices x and y in the set, there exists an alternating sequence of distinct vertices and edges that starts with x and ends with y , and each vertex belongs to the previous and next edges in the sequence.

So x and y are in the same component if we can walk from one to the other using edges present in the graph.

Let $C(x)$ be the set of vertices in the component that contains vertex x and denote the number of elements by $|C(x)|$. Then the *percolation probability* is the probability that the cluster containing vertex x is infinitely large.

DEFINITION 9.2 (Percolation probability) Let $C \triangleq C(o)$. The *percolation probability* $\theta(p)$ at the origin is

$$\theta(p) \triangleq \mathbb{P}(|C| = \infty) = 1 - \sum_{k=1}^{\infty} \mathbb{P}(|C| = k).$$

It turns out that in many cases, there is a *critical probability* p_c such that for $p < p_c$, $\theta(p) = 0$, while for $p > p_c$, $\theta(p) > 0$. In the latter case, there exists an infinite cluster somewhere in the network with probability 1. This follows from Kolmogorov's zero–one law.

Similarly, there exists a critical probability p'_c for the mean cluster size

$$\chi_x(p) = \mathbb{E}|C(x)|$$

of the cluster or component containing x . Focusing on the origin, for $p < p'_c$, $\mathbb{E}|C| < \infty$, whereas for $p > p'_c$, $\mathbb{E}|C| = \infty$.

This discussion will be made precise in the next chapter.

Remark 9.1 In percolation terminology, frequently the model is defined not using the presence or absence of edges or vertices, but using the notion of *open* and *closed* bonds or sites. An open bond or site corresponds to an edge or vertex that is present in the graph, whereas a closed bond or site indicates absence. For someone used to thinking in terms of electric circuits, this terminology is counterintuitive. If one thinks of water pipes, for example, it is easy to remember what open and closed stand for.

10 Bond and site percolation

10.1 Random trees and branching processes

In problems involving graphs and their properties, it is often helpful to first consider graphs without cycles, i.e., trees.

10.1.1 Percolation on regular branching trees

DEFINITION 10.1 (k -regular trees and branching trees) A k -regular tree (or Bethe lattice) is a tree where all vertices have degree $k > 0$. A $(k - 1)$ -branching tree is obtained from a k -regular tree by declaring one of the vertices the root vertex and removing one of its edges.

Necessarily, k -regular trees contain an infinite number of vertices (and thus no leaf nodes). Two-regular trees, for example, are isomorphic to the graph with vertex set $V = \mathbb{Z}$ with edges between vertices i, j if $|i - j| = 1$. The root vertex in a $(k - 1)$ -branching tree has degree $k - 1$. Also, if we delete an arbitrary edge in a k -regular tree, it splits into two $(k - 1)$ -branching trees. For the purposes of branching and percolation, branching trees are the more natural objects to study than regular trees. We will denote the k -branching tree by \mathbb{T}_k . The 3-regular tree and 2-branching tree are shown in Fig. 10.1.

Let us define a percolation model on a k -branching tree $\mathbb{T}_k = (V, E)$ by keeping each edge of the tree with probability $p > 0$ (we delete it otherwise), independently of all other edges. What is left is a random, possibly infinite, subgraph of \mathbb{T}_k . Natural questions include the following. How large is the component at the origin? What is the critical probability for the existence of an infinite component anywhere in the tree? It turns out that in the case of trees, the critical probability is easy to calculate.

Before getting back to this model, we consider the more general model of branching processes.

10.1.2 Branching processes

In the bond percolation model on the tree \mathbb{T}_k , the distribution of the number of children of each vertex, the *offspring distribution*, is binomial with parameters p and k .

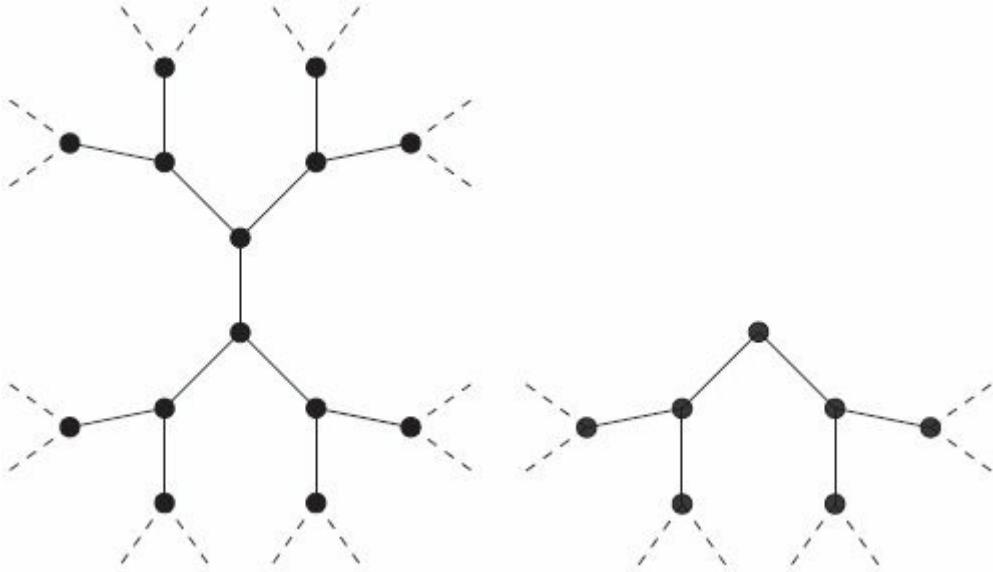


Figure 10.1 (Left) 3-regular tree. (Right) 2-branching tree. If any of the edges in the k -regular tree is removed, the remaining graph consists of two $(k - 1)$ -branching trees.

The model can be generalized to random branching trees with arbitrary offspring distributions, so-called *branching processes*. Let Z_n be the number of members in the n th generation. Each member x_i of the n th generation gives birth to a random independent number of children, $X_{n,i}$, which are the members of the $(n + 1)$ th generation. The (common) distribution of $X_{n,i}$ is called the *offspring distribution*. Assuming $Z_0 = 1$, the evolution of Z_i can be represented by a random tree rooted at Z_0 for which

$$Z_{n+1} = X_{n,1} + X_{n,2} + \dots + X_{n,Z_n} \quad (10.1)$$

holds. $(X_{n,i})$, $n \in \mathbb{N}_0$, $i \in \mathbb{N}$, forms a doubly infinite array of iid random variables. Such a branching process is called a *Galton–Watson branching process*. We will denote the mean number of children by μ , i.e., we let $\mu \triangleq \mathbb{E}(X_{n,i}) = \mathbb{E}(Z_1)$. If we take the offspring distribution to be binomial with parameters k and p , we obtain the original model of the random k -branching tree.

Example 10.1 Let the offspring distribution be $\mathbb{P}(X = 0) = 1/8$, $\mathbb{P}(X = 1) = 1/2$, \mathbb{P}

$(X = 2) = 3/8$. Hence $\mathbb{E}(X) = \mu = 5/4$. A possible realization of the first four generations of this tree may look as shown in Fig. 10.2.

What is the probability that a general random branching tree dies out? The answer is given by the following theorem.

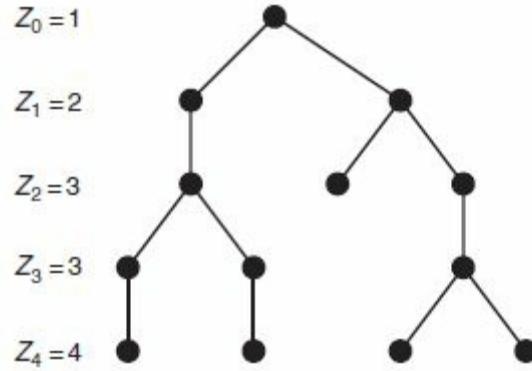


Figure 10.2 Simple branching process example. Z_k indicates the number of vertices at depth k .

THEOREM 10.2 (Growth of branching processes) *If $\mu \leq 1$, the branching process does not grow forever with probability 1, except when $\mathbb{P}(X = 1) = 1$. If $\mu > 1$, the branching process grows forever with positive probability.*

Before proving this theorem, we first show that the expected number of members in a branching process grows or decays exponentially fast, depending on the expected number of children per node. We will need generating functions for the proof. Recall that

$$G_X(s) \triangleq \mathbb{E}(s^X) = \sum_{n \in \mathbb{N}_0} \mathbb{P}(X = n) s^n$$

of a random variable $X \in \mathbb{N}_0$. We have $\mathbb{E}(X) = G'_X(1) = (\text{d}/\text{d}s)G(s)|_{s=1}$ and, for X and Y being mutually independent,

$$G_{X+Y}(s) = G_X(s)G_Y(s).$$

Further, all derivatives are non-negative on \mathbb{R}^+ , and the sum $S = X_1 + \dots + X_N$, for iid X_i and a random variable $N \in \mathbb{N}$ independent of the X_i , has the generating function

$$G_S(s) = G_N(G_X(s)).$$

Now let G_n be the generating function of Z_n and $G(s) = G_{X_i}(s)$ the generating function¹ of the offspring distribution. From (10.1) we have

$$G_{n+1}(s) = G_n(G(s))$$

and

$$G_n(s) = \underbrace{G(G(G(\dots(G(s))\dots)))}_{n \text{ times}}.$$

The first lemma we need is about the mean number of members in each generation.

LEMMA 10.3 (Mean growth) *The expected number of members evolves as follows:*

$$\mathbb{E}X = \mu \Rightarrow \mathbb{E}Z_n = \mu^n.$$

Proof We have $G_n(s) = G(G_{n-1}(s))$ and

$$\begin{aligned} \mathbb{E}Z_n &= G'_n(1) = G'(G_{n-1}(1))G'_{n-1}(1) \\ &= G'(1)G'_{n-1}(1) \\ &= \mu \mathbb{E}Z_{n-1}. \end{aligned}$$
□

Next we prove a lemma about the probability of extinction.

LEMMA 10.4 (Extinction probability) *The probability η that $Z_n = 0$ for some n is equal to the smallest non-negative root of the equation $G(s) = s$.*

Proof Approximating the probability of ultimate extinction by $\eta_n \triangleq \mathbb{P}(Z_n = 0)$, we have

$$\eta_n \rightarrow \eta \text{ as } n \rightarrow \infty.$$

Equivalently, we may write $\eta = \mathbb{P}(Z_n = 0 \text{ for some } n)$. η_n can be expressed as

$$\eta_n = \mathbb{P}(Z_n = 0) = G_n(0) = G(G_{n-1}(0)) = G(\eta_{n-1}).$$

Now let $n \rightarrow \infty$ and use the fact that G is continuous to obtain

$$\eta = G(\eta).$$

This tells us that η is indeed a root of $G(s) = s$, but the claim is that it is the *smallest non-negative* root. To verify this, suppose that r is any non-negative root

of $G(s) = s$. Since G is non-decreasing, we have

$$\eta_1 = G(0) \leq G(r) = r,$$

and

$$\eta_2 = G(\eta_1) \leq G(r) = r,$$

and so on, giving that $\eta_n \leq r$ for all n and hence $\eta \leq r$. See Fig. 10.3 for an illustration of this recursive procedure.

Alternatively, we can argue as follows. $G(s)$ is non-negative and convex increasing on $[0, 1]$, with $G(1) = 1$. If $G(0) = 0$, the situation is trivial, since each node will have at least one child for sure, which means the probability of extinction is 0. In this case, 0 is the smallest non-negative root. If $G(0) > 0$ and $\mu \leq 1$, there cannot be a root of $G(s)$ within $[0, 1)$, since $G'(1) = \mu \leq 1$. On the other hand, if $G(0) > 0$ and $\mu > 1$, there must be a root within $[0, 1)$, since, with $G(s) \geq 0$, $G(1) = 1$, and $G'(1) > 1$, it must hold that $G(1 - \epsilon) < 1 - \epsilon$ for some small ϵ .

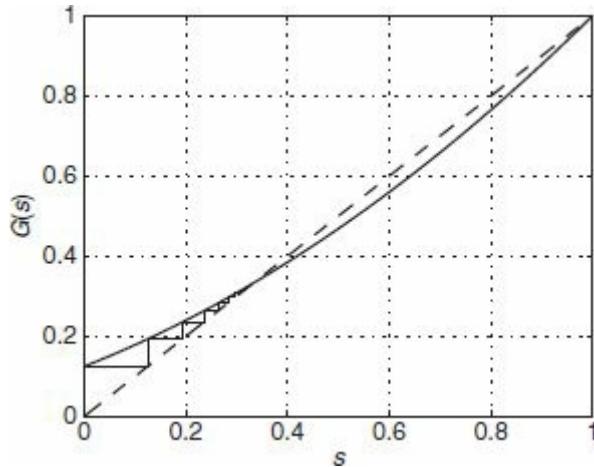


Figure 10.3 $G(s)$ as a function of s for the parameters used in Examples 10.1 and 10.2. In this case, $\mu = 5/4$, so there must be a root of $G(s) - s$ for $s \in (0, 1)$. It is at $s = 1/3$, so the extinction probability is $1/3$, which is the value to which the recursive procedure described in the proof of Lemma 10.2 converges.

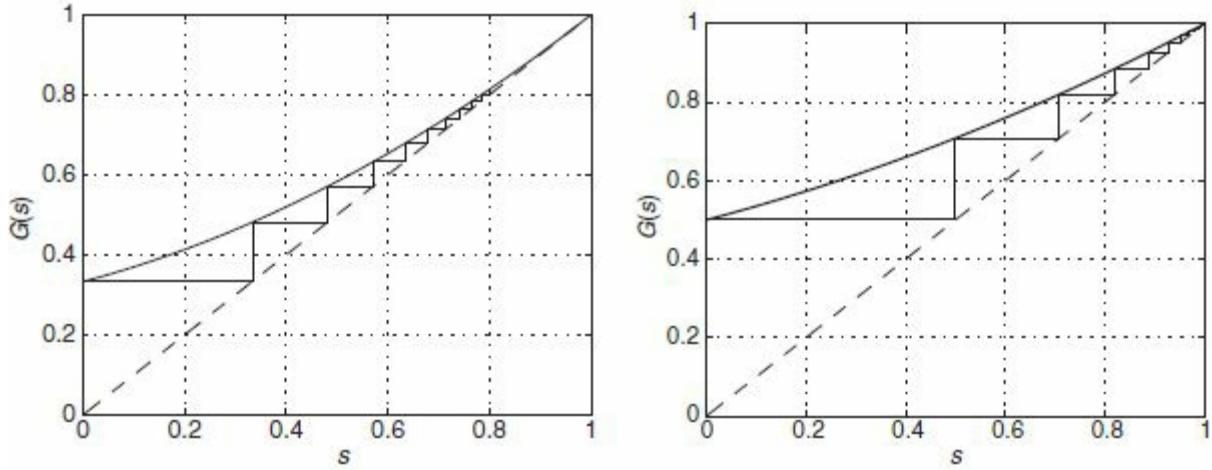


Figure 10.4 Scenarios in which extinction occurs with probability 1, i.e., the smallest (positive) solution to $G(s) = s$ is $s = 1$. (Left) $G(s) = s^2/3 + s/3 + 1/3$, so $\mu = G'(1) = 1$ and $\mathbb{P}(X=0) > 0$. (Right) $G(s) = s^2/6 + s/3 + 1/2$, so $\mu = G'(1) = 2/3$.

It follows from Descartes' sign rule that $G(s) - s$ can have at most two positive roots, i.e., only one besides $s = 1$. The rule says to count the number c of sign changes in the polynomial coefficients when they are ordered by increasing (or decreasing) exponent. Then the number of positive roots of the polynomial is at most c , or less by a multiple of 2. In our case, unless $\mathbb{P}(X=1) = 1$, $c = 2$. Figure 10.4 shows two cases in which extinction occurs almost surely.

Example 10.2 Consider the same random branching tree as in Example 10.1, where each node has zero, one, or two children with probabilities $1/8$, $1/2$, and $3/8$, respectively. The generating function is

$$G(s) = \frac{3}{8}s^2 + \frac{1}{2}s + \frac{1}{8}.$$

Solving $G(s) = s$ yields $s = 1/3$ and (of course) $s = 1$. The smallest nonnegative solution is $s = 1/3$, so the random tree is infinite with probability $2/3$. The expected number of children is $\mathbb{E}X = 1/2 + 2 \cdot 3/8 = 5/4$.

We are now equipped for the proof of Theorem 10.2.

Proof (of Theorem 10.2) Consider $\mu > 1$ first. Since $G'(1) = \mu$, we have $G'(1) > 1$. Since $G(1) = 1$, this means that there is some $s' < 1$ for which $G(s') < s'$. Since $G(0) \geq 0$ and G is continuous, there must be some points $0 \geq s'' \geq s'$ with $G(s'') = s''$, which implies that the smallest non-negative solution of $G(s) = s$ is strictly smaller than 1. Hence $1 - \eta > 0$, i.e., the process survives forever with positive probability.

Next consider $\mu \leq 1$ but $\mathbb{P}(X = 1) < 1$ (this implies that $G(0) > 0$). First note that $G'(s) > 0$ and $G''(s) \geq 0$, so $G(s)$ is strictly increasing and convex on \mathbb{R}^+ . Together with $G'(1) \leq 1$, it follows that $G'(s) < 1$ for $s \in [0, 1)$ and thus $G(s) > s$ for $s < 1$. Therefore the smallest non-negative solution of $G(s) = s$ is $s = 1$, which implies that $\eta = 1$, i.e., the process dies out for sure.

If $\mathbb{P}(X = 0) > 0$, the branching process can always stop even at its first iteration, so the probability of eventual extinction is positive.

If we specialize the offspring distribution to be binomial with parameters k and p , we get one step closer to the percolation model on the branching tree. Specializing Lemma 10.4 to binomial offspring, we have the following corollary.

COROLLARY 10.5 (Extinction probability for binomial offspring) *For general k , the generating function of the binomial offspring distribution is*

$$G^{(k)}(s) = (1 - p + ps)^k.$$

Let η be the smallest non-negative root of $G^{(k)}(s) - s$. We have $\eta < 1$ as soon as $p > 1/k$, so there is a positive chance of the tree growing to infinite size if $p > 1/k$.

Proof The expression for $G^{(k)}(s)$ is obtained by writing $\sum_{n=0}^k \mathbb{P}(X=n)s^n$ in binomial form. Since $\mu = \mathbb{E}X = pk$, $p > 1/k$ is sufficient for the probability of non-extinction to be positive. \square

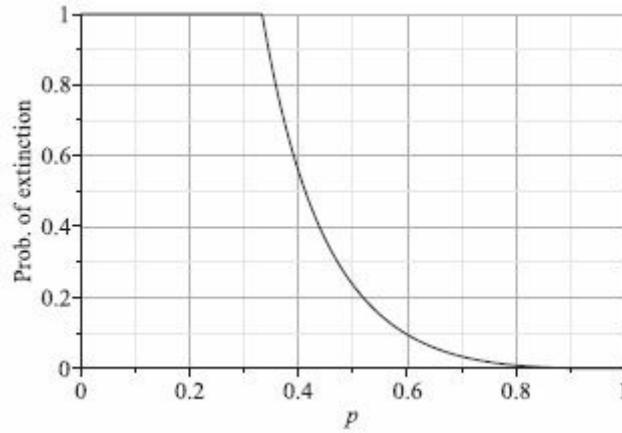


Figure 10.5 Extinction probability for Example 10.3.

As a consequence of the corollary, the percolation probability at the root node of the random branching tree with binomial offspring distribution with parameters k and p is

$$\theta(p) = 1 - \eta,$$

and η is the extinction probability.

Example 10.3 For $k = 3$, $G^{(3)}(s) = s$ yields the extinction probability

$$\eta = \min \left\{ 1, \frac{2p^2 - 3p + \sqrt{4p - 3p^2}}{2p^2} \right\}.$$

As can be seen in Fig. 10.5, $\eta < 1$ as soon as $p > 1/3$.

Now we return to the original percolation model on the k -branching tree, where each vertex has k children but each edge is deleted with probability $1 - p$. This model starts $k - m$ new processes every time a node generates $m < k$ children. If the average number of children $pk > 1$, then, each time we start a new process, there is a positive probability that an infinite tree is generated. It follows that an infinite tree is generated with probability 1 after a finite number of trials. *So a positive probability of non-extinction in the branching model implies the existence of an infinite (sub)tree rooted at some vertex and vice versa.* Conversely, if the extinction probability is 1, then no infinite (sub)tree can exist. Hence we have the following theorem.

THEOREM 10.6 (Percolation threshold for k -branching tree) *The critical probability for percolation on the k -branching tree is $p_c = 1/k$. For $p > p_c$, $\theta(p) = 1 - \eta$, where η is the smallest root of $(1 - p + ps)^k - s$. For $p \leq p_c$, $\theta(p) = 0$.*

10.1.3 Site percolation on the branching tree

In *site percolation*, the *vertices* are open or closed, instead of the edges. If a vertex is closed, all associated edges are closed. On k -regular branching trees, the critical probability of site percolation is equal to the critical probability for bond percolation, since closing an edge has the same effect as closing the vertex “below” that edge (if the root vertex is on top).

10.1.4 Mean cluster size

The expected cluster size of the component of vertex x in a percolation model with probability p is

$$\chi_x(p) \triangleq \mathbb{E}|C(x)|,$$

where $|C(x)| = \#C(x)$ is the number of elements in $C(x)$. The associated critical probability is

$$p'_c = \sup\{p: \chi_x(p) < \infty\} = \inf\{p: \chi_x(p) = \infty\}.$$

Since

$$\chi_x(p) = \infty \cdot \mathbb{P}(|C(x)| = \infty) + \sum_{n=1}^{\infty} n \mathbb{P}(|C(x)| = n),$$

$\chi_x(p) = \infty$ as soon as $p > p_c$. The converse, that the mean component size is finite if $p < p_c$, also holds for most graphs, but is harder to prove in general. In the case of the k -branching tree, $p_c = p'_c$ is easy to show. We have, for $C = C(o)$,

$$\chi(p) = \sum_{x \in \mathbb{T}_k} \mathbb{P}(x \in C) = \sum_{i=0}^{\infty} k^i p^i,$$

since a vertex at depth m of the tree is connected to the root with probability p^m and there are k^m vertices at that depth.

Hence the mean cluster size is finite and equal to $1/(1 - kp)$ if $p < 1/k$ and infinite otherwise, and the critical probability p'_c for infinite cluster size on the tree is also equal to $1/k$.

For the binary tree ($k = 2$), the complete cluster size distribution can be calculated. If $|C| = n$, there must be $n - 1$ open edges, all between vertices connected to o , and $n + 1$ closed edges, all emanating from vertices that are connected to o . So

$$\mathbb{P}(|C| = n) = c_n p^{n-1} (1-p)^{n+1}$$

for some coefficient c_n that counts the number of ways in which the $n - 1$ open and $n + 1$ closed edges can be arranged such that $|C| = n$.

To find c_n , let us encode the random tree using a sequence of symbols from the set $\{S, R, L\}$. The reason for the choice of the symbols S , R , and L is that the construction of the tree can be interpreted as a random walk on \mathbb{N}_0 whose position indicates the number of active nodes. The walk starts at 1 since initially only the root node is active. If an S is drawn, the walk stays at 1, so S indicates “stay.” If an R is drawn, the walk moves to the right, so R indicates “right.” Lastly, if an L is drawn, the walk moves to the left (L indicates “left”). The size of the tree corresponds to the first time the walk reaches 0. So c_n can be found from the theory of random walks. Here we use a direct method.

Start with the root and declare it active. Choose a symbol S , R , or L with the

following probabilities: $S = \{0, 1\}$ with probability $2p(1 - p)$, $R = \{1, 1\}$ with probability p^2 , and $L = \{0, 0\}$ with probability $(1 - p)^2$. If the outcome is L , terminate (both outgoing edges are closed). If the outcome is S , declare one of the edges open and the other closed, and let the vertex at the other end of the open edge be active. If the outcome is R , declare both edges open, and let both vertices be active. Finally, declare the root node inactive. Proceed with all active nodes as with the root node. The tree obtained once all vertices have been rendered inactive is a realization of the random branching tree that is fully described by the sequence of symbols from the alphabet S, R, L , and the length of that sequence equals the cluster size $|C|$. Replacing each symbol by its twodigit binary sequence, the realization of the tree is also encoded by a binary sequence of length $2|C|$ that contains exactly $|C| - 1$ ones and $|C| + 1$ zeros. Letting $|C| = n$, there are $\binom{2n}{n-1}$ ways of picking $n - 1$ ones from a sequence of length $2n$. But not all of them actually encode a tree of size n . In fact, the sequences consisting of S, R , and L cannot be shifted cyclically, since any cyclic shift of a valid sequence would lead to an invalid one.

For example, the sequence $SRLL$ encodes a tree of size 4, where the root node has degree 1 and the node at depth 1 has degree 2. A cyclic shift that produces a sequence starting with L would mean that the root node has degree 0, immediately terminating the tree. Also, the L at the end is necessary, since otherwise the tree would not be terminated (there would be more active nodes). As another example, the tree shown in Fig. 10.2, if interpreted as a binary tree of size 13, is encoded, for instance, by $RSRSSLRLSRL$. Other sequences are obtained if active nodes are explored in a different order.

So, since among the n cyclic shifts of each sequence, only one is valid, we have $c_n = (1/n) \binom{2n}{n-1}$ and thus

$$\mathbb{P}(|C| = n) = \frac{1}{n} \binom{2n}{n-1} p^{n-1} (1-p)^{n+1}. \quad (10.2)$$

Another option to derive the distribution is to start with the recursion

$$\begin{aligned} \mathbb{P}_p(|C| = 1) &= (1-p)^2 \\ \mathbb{P}_p(|C| = 2) &= 2p(1-p)^3 \\ &\vdots \\ \mathbb{P}_p(|C| = n) &= 2p(1-p)\mathbb{P}_p(|C| = n-1) + p^2 \sum_{k=1}^{n-2} \mathbb{P}_p(|C| = k)\mathbb{P}_p(|C| = n-k-1), \end{aligned} \quad (10.3)$$

where the last line is valid for $n \geq 3$. From these formulas, one can find the generating function $G_{|C|}(s)$, whose “inversion” yields (10.2).

For $p = 1/2$, it follows from Stirling’s formula that

$$\mathbb{P}(|C| = n) \sim \frac{1}{\sqrt{\pi n^3}}, \quad n \rightarrow \infty.$$

10.2 Preliminaries for bond percolation on the lattice

10.2.1 Probability space

We have defined $\mathbb{L}^2 = (\mathbb{Z}^2, E)$ where edges exist between all vertex pairs with distance 1. More generally, in d dimensions, $\mathbb{L}^d = (\mathbb{Z}^d, E)$. Let $0 \leq p \leq 1$ and $q = 1 - p$. Denote each edge in E to be *open* with probability p and *closed* with probability q . We consider the following probability space. As sample space we take

$$\Omega = \prod_{e \in E} \{0, 1\},$$

whose elements are represented as $\omega = (\omega(e): e \in E)$ and called *configurations*; the value $\omega(e) = 0$ corresponds to e being closed, and $\omega(e) = 1$ corresponds to e being open. We take \mathcal{F} to be the σ -field generated by the basic events

$$C_e(a) = \{\omega \in \{0, 1\}^E : \omega(e) = a\}, \quad a \in \{0, 1\}.$$

In technical terms, this σ -field consists of the subsets of Ω generated by the finite-dimensional cylinders, since Ω is a discrete product topology.

As probability measure, we take the product measure with density p on (Ω, \mathcal{F}) ; this is the measure

$$\mathbb{P}_p = \prod_{e \in E} \mu_e,$$

where μ_e is the Bernoulli measure on $\{0, 1\}$ given by

$$\mu_e(\omega(e) = 0) = q, \quad \mu_e(\omega(e) = 1) = p.$$

We drop the subscript p from \mathbb{P}_p whenever there is no ambiguity. Similarly, \mathbb{E} is the shortcut for the corresponding expectation operator \mathbb{E}_p .

There is a natural partial order on the set Ω of configurations, which is given by $\omega_1 \leq \omega_2$ if and only if $\omega_1(e) \leq \omega_2(e)$ for all $e \in E$. Further, there is a one-to-one correspondence between Ω and the set of subsets of E . For $\omega \in \Omega$, we define

$$K(\omega) \triangleq \{e \in E: \omega(e) = 1\};$$

thus $K(\omega)$ is the set of open edges of the lattice when the configuration is ω . Clearly, $\omega_1 \leq \omega_2$ if and only if $K(\omega_1) \subseteq K(\omega_2)$.

10.2.2 Coupling of bond percolation processes

Suppose that $(X(e): e \in E)$ is a family of independent random variables indexed by the edge set E , where each $X(e)$ is uniform on $[0, 1]$. We may couple all bond percolation processes on \mathbb{L}^d as p ranges over the interval $[0, 1]$ as follows. Define $\eta_p \in \Omega$ by

$$\eta_p(e) \triangleq \begin{cases} 1 & \text{if } X(e) < p \\ 0 & \text{if } X(e) \geq p \end{cases} \quad e \in E.$$

We say that the edge e is p -open if $\eta_p(e) = 1$. The random vector η_p has independent components and marginal distributions given by

$$\mathbb{P}(\eta_p(e) = 0) = 1 - p, \quad \mathbb{P}(\eta_p(e) = 1) = p.$$

So we have $\eta_{p_1} \leq \eta_{p_2}$ whenever $p_1 \leq p_2$. Generally as p increases, the configuration η_p runs through typical configurations of percolation processes with all edge probabilities. Coupling has the advantage that $\eta_{p_1} \leq \eta_{p_2}$ if $p_1 \leq p_2$. Without coupling, no such statement could be made.

Considering the random subgraph of \mathbb{L}^d containing all vertices but only the open edges, we denote the connected components as *open clusters*.

10.3 General behavior of the percolation probability

10.3.1 The d -dimensional case

We have defined the *percolation probability* $\theta(p)$ in Definition 9.2. Clearly we have $\theta(0) = 0$ and $\theta(1) = 1$. It is fundamental to percolation theory that there exists a critical value p_c of p such that

$$\theta(p) \begin{cases} = 0 & \text{if } p < p_c \\ > 0 & \text{if } p > p_c, \end{cases}$$

where p_c is called the *critical probability* defined by

$$p_c \triangleq \sup \{p: \theta(p) = 0\}.$$

We use $p_c(d)$ to denote the critical probability in d -dimensional lattices. Since in one dimension, an infinite component cannot occur as soon as $p < 1$, we have $p_c(1)$

= 1.

Generally it is apparent that $\theta_d(p)$ is non-decreasing in d since if the origin belongs to an infinite open cluster in \mathbb{L}^d it also belongs to an infinite open cluster in the “augmented” lattice \mathbb{L}^{d+1} . So

$$p_c(d+1) \leq p_c(d), \quad d \geq 1.$$

It is known that $\theta_d(p)$ is continuous except possibly at $p = p_c$. For $3 \leq d \leq 19$, the possibility of a discontinuity at $p_c(d)$ has not been ruled out.

THEOREM 10.7 *The probability $\psi(p)$ that there exists an infinite open cluster in the graph is*

$$\psi(p) = \begin{cases} 0 & \text{if } \theta(p) = 0 \\ 1 & \text{if } \theta(p) > 0. \end{cases}$$

Proof This follows from Kolmogorov’s zero–one law. The event of the existence of an infinite open cluster does not depend on the state of any finite collection of edges. By the zero–one law, $\psi \in \{0, 1\}$. If $\theta(p) = 0$, then by the union bound

$$\psi(p) \leq \sum_{x \in \mathbb{Z}^2} \mathbb{P}(|C(x)| = \infty) = 0.$$

On the other hand, if $\theta(p) > 0$, then

$$\psi(p) \geq \mathbb{P}(|C| = \infty) > 0,$$

so that $\psi(p) = 1$ by the zero–one law.

This can also be shown by noting that the existence of an infinite cluster is invariant under any translation of coordinates on the plane, and, since the model is ergodic, the event can only have probability 0 or 1.

Note that the theorem does not tell us how many infinite open clusters there are. We will address this question later.

The next theorem states that θ is a non-decreasing function of p .

THEOREM 10.8 (Non-decreasing percolation probability) *When $0 < p_1 < p_2 < 1$, we have that $\theta(p_1) \leq \theta(p_2)$.*

Proof Let

$$p_1 = p_2 \frac{p_1}{p_2},$$

where $p_1/p_2 < 1$. Take a realization of G_{P_2} and delete each edge independently with probability $1 - p_1/p_2$. The resulting graph is a realization of G_{P_1} that contains fewer edges than G_{P_2} . So, if there is an infinite cluster in G_{P_1} , then there must also be one in G_{P_2} . \square

The result also follows from the coupling introduced in [Section 10.2.2](#).

The qualitative behavior of the percolation probability is shown in Fig. 10.6.

10.3.2 Scaling theory

Scaling theory studies the behavior of critical quantities in percolation theory near the critical probability p_c . It assumes that

$$\theta(p) \xrightarrow{\log} (p - p_c)^\beta \quad \text{as } p \downarrow p_c, \quad (10.4)$$

$$\chi(p) \xrightarrow{\log} (p - p'_c)^{-\gamma} \quad \text{as } p \uparrow p'_c, \quad (10.5)$$

$$\mathbb{P}_{p_c}(n \leq |C| < \infty) \xrightarrow{\log} n^{-1/\delta}, \quad n \rightarrow \infty, \quad (10.6)$$

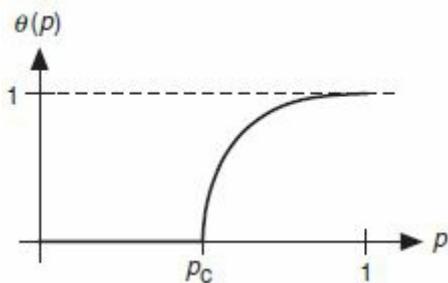


Figure 10.6 Qualitative behavior of the percolation probability $\theta(p)$.

where $g(p) \xrightarrow{\log} h(p)$ means $\log g(p) \sim \log h(p)$ for the given asymptotics and $|C(o)|$ is the size of the component containing vertex o . The *critical exponents* $\beta > 0$, $\gamma > 0$, and $\delta > 0$ depend on the dimension d . It can be shown that the critical exponents are related. In two dimensions, for example,

$$\gamma + 2\beta = \beta(\delta + 1). \quad (10.7)$$

Example 10.4 For the 3-branching tree with binomial offspring (see [Example 10.3](#)), we have from the Taylor expansion of $\theta(p) = 1 - \eta$

$$\theta(p) = 9 \left(p - \frac{1}{3} \right) - 45 \left(p - \frac{1}{3} \right)^2 + O \left(p - \frac{1}{3} \right)^3 \quad \text{as } p \downarrow 1/3.$$

Hence the critical exponent $\beta = 1$ in (10.4) for the 3-branching tree.

10.3.3 Simple bounds for the percolation probability on the square lattice

THEOREM 10.9 *There exists a $1/3 \leq p_c \leq 2/3$ such that $\theta(p) = 0$ for $p < p_c$ and $\theta(p) > 0$ for $p > p_c$.*

Proof First, we show that for $p < 1/3$, $\theta(p) = 0$. Let $\sigma(n)$ denote the number of self-avoiding (consisting of distinct vertices and edges) paths of length n in \mathbb{L}^2 starting at the origin. $\sigma(n)$ is unknown but a simple bound is

$$\sigma(n) \leq 4 \cdot 3^{n-1}$$

since at each step there are three choices of directions, except for the first step, where there are four.

Now let the random variable $N(n)$ be the number of open paths of length n in the random grid. Any such path is open with probability p^n , so

$$\mathbb{E}N(n) = p^n \sigma(n).$$

Now, if the origin belongs to an infinite open cluster, then for each n there must exist at least one open path of length n starting at o , so that

$$\theta(p) \leq P(N(n) \geq 1) \leq \mathbb{E}N(n) = p^n \sigma(n) \quad \forall n.$$

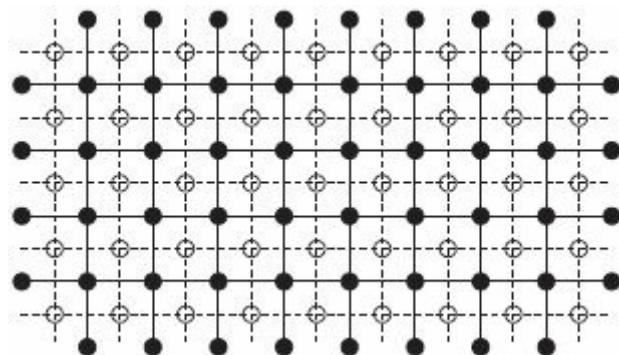


Figure 10.7 Square lattice (solid points and edges) and its dual (open points and dashed edges). The square lattice is *self-dual*, i.e., the dual to its dual is the original lattice.

Hence

$$\theta(p) \leq p^n 4 \cdot 3^{n-1} = 4p(3p)^{n-1} \quad \forall n.$$

So by choosing $p < 1/3$ and letting $n \rightarrow \infty$ we have proven the first part.

The second part is to show that for $p > 2/3$, $\theta(p) > 0$. This part is based on a counting argument known as *Peierls' argument*. We need the concept of the *dual lattice* or, more generally, the *dual graph*. For a planar graph G , drawn such that edges intersect at vertices only, replace each face by a point (including infinite faces, if they exist), and add edges to these points whenever the corresponding faces of G share a boundary edge in G . The dual lattice to \mathbb{L}^2 is the lattice where a vertex is put in the center of each cell in the original lattice, i.e., the translated lattice by $(1/2, 1/2)$ (see Fig. 10.7).

We can also construct a dual to the random lattice by declaring edges to be closed if they cross a closed edge of the original lattice, and open if they cross an open edge of the original lattice. An example is shown in Fig. 10.8.

Box 10.1 Open and closed paths in the original and dual percolation processes

By convention, the definition of *open* and *closed* bonds or edges is reversed in the dual lattice. In the dual lattice, the edges or paths considered are the closed ones. For example, in Fig. 10.8, the bonds drawn in the original lattice are the open ones, while the ones drawn in the dual lattice are closed.

A *circuit* is a closed² path, i.e., a path where the start and end vertices are the same and all vertices have degree 2. Any finite connected component in the random grid is surrounded by a closed³ circuit in the dual random grid. So the event $|C| < \infty$ is the equivalent to the event that o lies inside a closed circuit in the dual random grid.

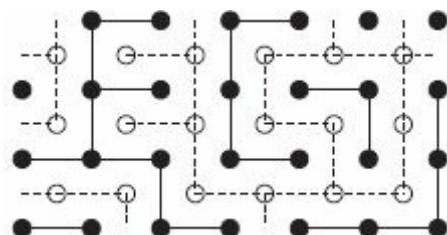


Figure 10.8 Dual lattice to the random lattice, where an edge is open whenever the

corresponding edge in the original lattice is open and closed whenever the edge in the original lattice is closed. Any finite component in the original lattice is surrounded by a closed circuit in the dual lattice.

Now consider first some deterministic quantities. Let $B(k) = [-k, k]^2 \cap \mathbb{Z}^2$, let \mathcal{C} denote the set of all circuits in the dual lattice that enclose the origin o , and let $\mathcal{C}_k \subset \mathcal{C}$ be the subset of circuits that surround $B(k)$. Finally, let $\rho(n)$ be the number of circuits of length n of the dual lattice that surround the origin. This deterministic quantity satisfies

$$\rho(n) \leq n\sigma(n-1),$$

since any circuit of length n surrounding the origin contains a path of length $n-1$ starting at some point $x = (k + 1/2, 1/2)$ for some $0 \leq k < n$.

Next we introduce some random quantities. We call the random grid G and its dual G' . So G includes all open edges in the original lattice, whereas G' includes all closed edges in the dual lattice. Let $\partial B(k)$ be the collection of vertices on the boundary of the box $B(k)$. Now observe that there is at least one vertex $x \in \partial B(k)$ with $|C(x)| = \infty$ if and only if there is no circuit in \mathcal{C}_k in the dual lattice G' :

$$\begin{aligned} \mathbb{P}\left(\bigcup_{x \in \partial B(k)} \{|C(x)| = \infty\}\right) &= \mathbb{P}\left(\bigcap_{\gamma \in \mathcal{C}_k} \gamma \not\subset G'\right) \\ &= 1 - \mathbb{P}\left(\bigcup_{\gamma \in \mathcal{C}_k} \gamma \subset G'\right) \\ &\geq 1 - \sum_{\gamma \in \mathcal{C}_k} \mathbb{P}(\gamma \subset G') \end{aligned} \tag{10.8}$$

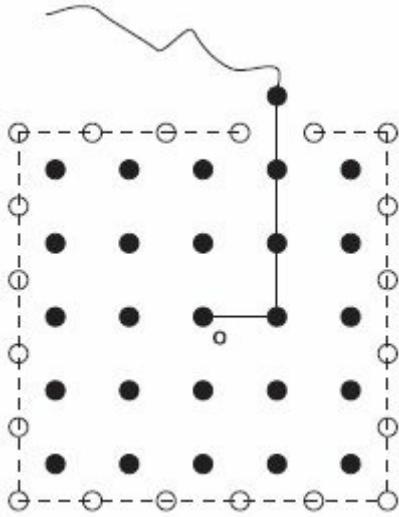


Figure 10.9 Box $B(2)$ and an open path (consisting of closed edges) in the dual lattice, such that there is a positive probability that o belongs to an infinite component. Conversely, if the origin's component is finite, there must be a closed circuit around it.

by the union bound. This holds for all $k > 0$. A dual circuit of length n in G' has probability of occurrence q^n , thus

$$\begin{aligned}
 \sum_{\gamma \in \mathcal{C}_k} \mathbb{P}(\gamma \subset G') &\leq \sum_{n=8k}^{\infty} n\sigma(n-1)q^n \\
 &\leq 4 \cdot 3^{-2} \sum_{n=8k}^{\infty} (3q)^n n \\
 &= (8k - 24kq + 3q) \frac{(3q)^{8k}}{(3q-1)^2}, \quad \text{for } q < 1/3 \\
 &< 1, \quad \text{for } k \text{ large enough.}
 \end{aligned} \tag{10.9}$$

Plugging (10.9) into (10.8), we conclude that for k large enough,

$$\mathbb{P}\left(\bigcup_{x \in \partial B(k)} \{|C(x)| = \infty\}\right) > 0.$$

If $|C(x)| = \infty$ for some $x \in B(k)$, then $|C| = |C(o)| = \infty$ with probability greater than p^{2k} . It follows that, for $q < 1/3$, we have $\mathbb{P}(|C| = \infty) > 0$, and the proof is complete.

So the question of whether the origin belongs to an infinite component with positive probability is equivalent to the question of whether there exists a k for which there is a positive probability that the box $B(k)$ is not surrounded by a circuit in the dual lattice. The theorem states that, for $q < 1/3$, this is indeed the case for large k .

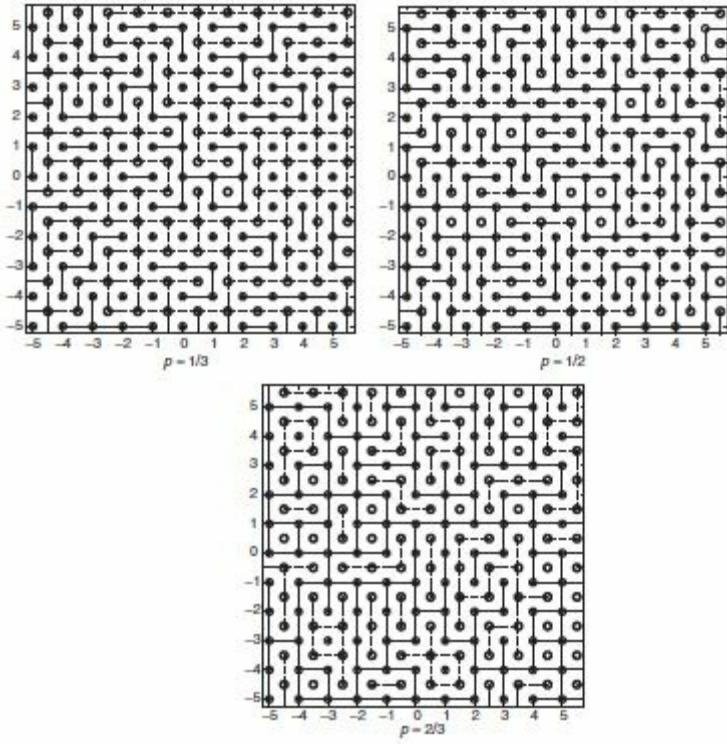


Figure 10.10 Examples for components in the original random lattice and the dual for $p = 1/3, 1/2$, and $2/3$.

The bound $\sigma(n) \leq 4 \cdot 3^{n-1}$ is relatively crude. Assuming we know what $\sigma(n)$ is, we define

$$\lambda(2) \triangleq \lim_{n \rightarrow \infty} \sigma(n)^{1/n}.$$

By the following logarithmic version of Fekete's lemma, this limit exists. It is called the *connective constant*.

LEMMA 10.10 (Fekete 1923) *Let (a_n) be a submultiplicative sequence of real numbers, i.e., $a_{n+m} \leq a_n a_m$ for $n, m \geq 1$. Then $\lim_{n \rightarrow \infty} a_n^{1/n}$ exists.*

Then, using the same technique as in the proof of Theorem 10.9, we obtain

$$\frac{1}{\lambda(2)} \leq p_c(2) \leq 1 - \frac{1}{\lambda(2)}.$$

Our previous bound was $\lambda \leq 3$. Noticing that among the 27 possibilities of taking three steps, 2 are prohibited, namely turning left three times in a row and turning right three times in a row, we can improve the bound to $\lambda \leq 25^{1/3} \approx 2.92$. The exact value of $\lambda(2)$ is unknown, but counting the number of paths of length 25 yields $\lambda \leq 2.736$, and the best known bounds are $2.6256 \leq \lambda(2) \leq 2.6792$.

10.3.4 Generalization of the lower bound

Using the same technique as in the two-dimensional case, we obtain for general d

$$\frac{1}{\lambda(d)} \leq p_c(d).$$

There is, however, no corresponding upper bound.

We have $\sigma(n) \leq 2d(2d-1)^{n-1}$. So we get the lower bound

$$p_c(d) \geq \frac{1}{2d-1}.$$

It is known that $p_c(d) \sim (2d)^{-1}$ as $d \rightarrow \infty$. So, for large d , bond percolation on \mathbb{L}^d behaves similarly to bond percolation on a $2d$ -branching tree. This is intuitive since, as the number of dimensions gets large, the growth process on \mathbb{Z}^d starting at o becomes less and less likely to hit an already visited site again, i.e., most random walks are self-avoiding.

In order to get sharper results on the percolation probability, we need to extend our toolbox by three important techniques, namely the *FKG inequality*, the *BK inequality*, and *Russo's formula*.

10.4 Basic techniques

10.4.1 Definitions

DEFINITION 10.11 (Increasing event) A random variable X is increasing on (Ω, F) if $X(\omega) \leq X(\omega')$ whenever $\omega \leq \omega'$. It is decreasing if $-X$ is increasing. An event $A \in F$ is increasing whenever its indicator function is an increasing variable, i.e., if $1_A(\omega) \leq 1_A(\omega')$ whenever $\omega \leq \omega'$.

If A is an event on the probability space for bond percolation, then, if A is an increasing event, $\mathbb{P}_p(A) \leq \mathbb{P}_{p'}(A)$ whenever $p \leq p'$. Alternatively, increasing events are called up-sets and defined as follows.

DEFINITION 10.12 (Up-set) $U \subset \Omega$ is an *up-set* if

$$a, b \in \Omega, a \in U, \text{ and } a \leq b \Rightarrow b \in U.$$

This is identical to the first definition since the first definition says $1_A(a) \leq 1_A(b)$ if $a \leq b$, and the condition in the second definition is $1_A(a) = 1$. Hence $1_A(b) = 1$

must hold also.

10.4.2 The FKG inequality

This inequality is named after Fortuin, Kasteleyn, and Ginibre. It was first proved by Harris in 1960 and is therefore also called Harris' lemma. It expresses the fact that increasing events can only be positively correlated. For our purposes, a special case is sufficient.

THEOREM 10.13 (FKG inequality) *If A and B are two increasing events or two decreasing events, then*

$$\mathbb{P}(A \cap B) \geq \mathbb{P}(A)\mathbb{P}(B).$$

proof We prove the theorem in the case where A and B depend on finitely many edges. Let $X = 1_A$ and $Y = 1_B$ be the indicators of the increasing events A and B , which are increasing random variables. We can then reformulate the FKG inequality as

$$\mathbb{E}(XY) \geq \mathbb{E}(X)\mathbb{E}(Y).$$

This holds for general increasing random variables X and Y with finite second moment.

Suppose that X and Y depend only on the state of edges e_1, e_2, \dots, e_n for some integer n . We proceed by induction.

Suppose first that $n = 1$, so that X and Y are functions only of the state $\omega(e_1)$. Pick any two states $\omega_1, \omega_2 \in \{0, 1\}$. Since both X and Y are increasing,

$$(X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2)) \geq 0$$

with equality if $\omega_1 = \omega_2$. If $\omega_1 \neq \omega_2$, then either both differences are negative or both are positive. Therefore,

$$\begin{aligned} 0 &\leq \sum_{\omega_1=0}^1 \sum_{\omega_2=0}^1 (X(\omega_1) - X(\omega_2))(Y(\omega_1) - Y(\omega_2))\mathbb{P}(\omega(e_1) = \omega_1)\mathbb{P}(\omega(e_1) = \omega_2) \\ &= \sum_{\omega_1=0}^1 X(\omega_1)Y(\omega_1)\mathbb{P}(\omega(e_1) = \omega_1) + \sum_{\omega_2=0}^1 X(\omega_2)Y(\omega_2)\mathbb{P}(\omega(e_1) = \omega_2) \\ &\quad - \sum_{\omega_1=0}^1 \sum_{\omega_2=0}^1 (X(\omega_1)Y(\omega_2) + X(\omega_2)Y(\omega_1))\mathbb{P}(\omega(e_1) = \omega_1)\mathbb{P}(\omega(e_1) = \omega_2) \\ &= 2(\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)). \end{aligned}$$

Note that $\mathbb{P}(\omega(e_1) = 0) + \mathbb{P}(\omega(e_1) = 1) = 1$.

Suppose now that the result is valid for values of n satisfying $n < k$, and that X and Y are increasing functions of the states $\omega(e_1), \omega(e_2), \dots, \omega(e_k)$ of the edges e_1, \dots, e_k . Then

$$\begin{aligned}\mathbb{E}(XY) &= \mathbb{E}\left(\mathbb{E}(XY \mid \omega(e_1), \dots, \omega(e_{k-1}))\right) \\ &\geq \mathbb{E}\left(\mathbb{E}(X \mid \omega(e_1), \dots, \omega(e_{k-1}))\mathbb{E}(Y \mid \omega(e_1), \dots, \omega(e_{k-1}))\right),\end{aligned}$$

since, for given $\omega(e_2), \dots, \omega(e_k)$, it is the case that X and Y are increasing in the single variable $\omega(e_k)$. Now both $\mathbb{E}(X \mid \omega(e_1), \dots, \omega(e_{k-1}))$ and $\mathbb{E}(Y \mid \omega(e_1), \dots, \omega(e_{k-1}))$ are increasing functions of the states of $k-1$ edges. Thus

$$\begin{aligned}\mathbb{E}(XY) &\geq \mathbb{E}\left(\mathbb{E}(X \mid \omega(e_1), \dots, \omega(e_{k-1}))\right)\mathbb{E}\left(\mathbb{E}(Y \mid \omega(e_1), \dots, \omega(e_{k-1}))\right) \\ &= \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}\quad \square$$

The proof for X and Y depending on infinitely many edges uses the martingale convergence theorem and the Cauchy–Schwartz inequality.

Example 10.5 Let $\Pi_1, \Pi_2, \dots, \Pi_k$ be families of paths in \mathbb{L}^2 , and let A_i be the event that there exists some path in Π_i that is open. The A_i are increasing events, and therefore

$$\mathbb{P}\left(\bigcap_{i=1}^k A_i\right) \geq \mathbb{P}(A_1)\mathbb{P}\left(\bigcap_{i=2}^k\right),$$

since the intersection of increasing events is increasing. We iterate this to obtain

$$\mathbb{P}\left(\bigcap_{i=1}^k A_i\right) \geq \prod_{i=1}^k \mathbb{P}(A_i).$$

The FKG inequality can be used to prove the following important theorem.

THEOREM 10.14 *Let $G = (V, E)$ be an infinite connected graph with countably many edges and consider bond percolation with edges being open with probability p . The value of the critical probability $p_c(x)$ is independent of the choice of $x \in V$.*

Proof For any vertex $x \in V$, we write $\theta_x(p)$ for the probability that x lies in an

infinite open cluster and

$$p_c(x) = \sup \{p : \theta_x(p) = 0\}$$

for the associated critical probability. Let $\{x \leftrightarrow y\}$ and $\{x \leftrightarrow \infty\}$ denote the events that x is connected to y and that x is connected to a node that is infinitely far away (and thus belongs to the infinite component), respectively. We have by the FKG inequality that, for all $y \in V$,

$$\theta_x(p) = \mathbb{P}(\{x \leftrightarrow \infty\}) \geq \mathbb{P}(\{x \leftrightarrow y\} \cap \{y \leftrightarrow \infty\}) \geq \mathbb{P}(x \leftrightarrow y)\theta_y(p),$$

so that $p_c(x) \geq p_c(y)$. This inequality also holds with x and y interchanged, so the two probabilities must be the same. \square

Example 10.6 Take a box $B(k) = [-k, k]^2 \cap \mathbb{Z}^2$. Let A be the event that there is an open path connecting a vertex of the top face to a vertex of the bottom face. This is a *top–bottom (TB) crossing*. Let B the event that there is an open path from a vertex of the left face to a vertex of the right face. This is a *left–right (LR) crossing*. The probability that there are both top–bottom and left–right crossings is at least the product of the probabilities that there is a TB crossing and an LR crossing.

10.4.3 The BK inequality

This inequality is named after van den Berg and Kesten, and was proved in 1985. It can be viewed as the reverse of the FKG inequality, but it applies to the event $A \circ B$ that two increasing events occur on *disjoint* sets of edges, rather than the larger event $A \cap B$ that A and B occur on any set of edges. $A \circ B$ is the set of configurations $\omega \in \Omega$ for which there are *disjoint* sets of open edges such that the first set guarantees the occurrence of A while the second set guarantees the occurrence of B .

The canonical setting is the following. Let G be a finite subgraph of \mathbb{L}^d , and let $A_G(x, y)$ be the event that there exists an open path $x \leftrightarrow y$ in G . Then $A_G(x, y) \circ A_G(u, v)$ is the event that there exist two edge-disjoint open paths in G , one joining x and y , the other joining u and v . Suppose now that we are given that $A_G(u, v)$ occurs, and we ask for the (conditional) probability of $A_G(u, v) \circ A_G(x, y)$. The conditioning on $A_G(u, v)$ amounts to knowing some information about the occurrence of open edges, but we are not allowed to use all such open edges in finding an open path from x to y disjoint from one of the open paths from u to v . Thus it is plausible that

$$\mathbb{P}\left(A_G(u, v) \circ A_G(x, y) \mid A_G(u, v)\right) \leq \mathbb{P}(A_G(x, y)).$$

This is essentially the assertion of the BK inequality.

DEFINITION 10.15 (Disjoint occurrence) Let A and B be two increasing events which depend on the state vector $\omega = (\omega(e_1), \dots, \omega(e_n))$ of n distinct edges e_1, \dots, e_n of \mathbb{L}^d . Each such ω is specified uniquely by the subset $K(\omega) = \{e_i : i \in [n], \omega(e_i) = 1\}$ of open edges among these n edges. Then $A \circ B$ is the set of ω for which there exists a subset $H \subset K(\omega)$ such that any ω' determined by $K(\omega') = H$ is in A and any ω'' determined by $K(\omega'') = K(\omega) \setminus H$ is in B .

In short, A and B occur disjointly if they occur on disjoint edge sets of \mathbb{L}^d .

THEOREM 10.16 (BK inequality) *If A and B are two increasing events or two decreasing events,*

$$\mathbb{P}(A \circ B) \leq \mathbb{P}(A)\mathbb{P}(B).$$

proof (Sketch) Consider the case where A and B denote the existence of two open paths between different pairs of vertices. Let $G = (V, E)$ be a finite subgraph of \mathbb{L}^d . Let

$$A = \{u \leftrightarrow v\}; B = \{x \leftrightarrow y\}, \quad u, v, x, y \in V.$$

Then $A \circ B$ is the event that there exist two edge-disjoint open paths from u to v and x to y . Let e be an edge in E . Replace e by two parallel edges e' and e'' with the same end vertices, each of which is open with the same probability p independently of each other and all other edges. The splitting of e can only make the search for two disjoint paths easier. With splitting, the probability of finding two disjoint open paths can therefore only increase or remain equal. We continue this splitting process, replacing every edge $f \in E$ by two parallel edges f' and f'' . At each stage we look for two open paths, the first one avoiding all edges marked " and the second avoiding all edges marked '. At each stage the probability of finding two disjoint paths can only increase or remain equal. When all edges of G have been split in two, we end up with two independent copies of G . In the first we look for an open path $u \leftrightarrow v$, and in the second we look for $x \leftrightarrow y$. Since such paths occur independently in each copy of G , the probability that they both occur is $\mathbb{P}(A)\mathbb{P}(B)$.

There exists a version of the BK inequality, *Reimer's inequality*, which holds for general events that are not necessarily increasing or decreasing. In fact, Reimer's inequality includes both the FKG and the BK inequalities. However, to date, there are no known applications of Reimer's inequality that could not be achieved

without it.

Example 10.7 Continuing Example 10.6, what is the probability that there is a TB open path and a disjoint LR open path? The probability of this event is smaller than the product of the probabilities that there is a TB open path and that there is an LR open path. So these events are negatively correlated.

10.4.4 Russo's formula

The third relation estimates the rate of change of the probability of occurrence of an increasing event A as p increases. We first need to introduce the notion of a *pivotal edge*. If A is increasing, an edge e is pivotal if and only if A occurs when e is open and does not occur when e is closed. A pivotal edge is thus a critical edge for A .

DEFINITION 10.17 (Pivotal edge) Let $A \in F$ and $\omega \in \Omega$. The edge e is *pivotal* for the pair (A, ω) if the occurrence of A critically depends on e , i.e., if $1_A(\omega) \neq 1_A(\omega')$ where ω' is the configuration such that $\omega'(e) = 1 - \omega(e)$ and $\omega'(f) = \omega(f)$ for all $f \in E \setminus \{e\}$.

The event “ e is pivotal for A ” is the set of all configurations ω for which e is pivotal for (A, ω) . Observe that this event is independent of the state of e itself but only depends on the state of the other edges.

For example, let A be the event that there is an LR open crossing of the box $B(n) \subset \mathbb{L}^2$. Any edge e of $B(n)$ is pivotal for A if, when e is open, there is an LR open crossing and when it is closed, there is no LR open crossing of $B(n)$. (This implies that one end vertex of e is joined to the left side of $B(n)$ while the other end vertex is joined to the right side of $B(n)$ by open paths.)

THEOREM 10.18 (Russo's formula) *Let A be an increasing event which depends on the state of finitely many edges of \mathbb{L}^d , and let $N(A)$ denote the number of edges that are pivotal for A . Then*

$$\frac{d}{dp}\mathbb{P}(A) = \mathbb{E}(N(A)).$$

proof (Sketch) Let $(X(e))$, $e \in E^d$, be a collection of iid random variables indexed by the edge set E^d of \mathbb{L}^d , uniformly distributed on $[0, 1]$. Consider the coupled percolation process introduced in Section 10.2.2. η_p is a random configuration, where edges are open with probability p , i.e., $\eta_p(e) = 1(X(e) < p)$, for all $e \in E^d$.

Hence $\mathbb{P}_p(A) = \mathbb{P}(\eta_p \in A)$. As A is an increasing event, we have for $\delta > 0$

$$\begin{aligned}\mathbb{P}_{p+\delta}(A) &= \mathbb{P}(\eta_{p+\delta} \in A) \\ &= \mathbb{P}\left((\{\eta_{p+\delta} \in A\} \cap \{\eta_p \notin A\}) \cup \{\eta_p \in A\}\right) \\ &= \mathbb{P}(\{\eta_{p+\delta} \in A\} \cap \{\eta_p \notin A\}) + \mathbb{P}(\eta_p \in A) \\ &= \mathbb{P}(\{\eta_{p+\delta} \in A\} \cap \{\eta_p \notin A\}) + \mathbb{P}_p(A).\end{aligned}\tag{10.10}$$

Now, if $\eta_p \notin A$ while $\eta_{p+\delta} \in A$, this means that there are some edges e on which A depends for which $\eta_p(e) = 0$ but $\eta_{p+\delta}(e) = 1$, or, equivalently, $p \leq X(e) < p + \delta$. Since A depends only on the state of finitely many edges, the probability that there is more than one edge e with $p \leq X(e) < p + \delta$ is negligible (of order $o(\delta)$) compared with the probability that there is one such edge, as $\delta \rightarrow 0$. If e is the only edge for which $p \leq X(e) < p + \delta$, then e must be pivotal for A , in the sense that $\eta_p \notin A$ but $\eta_{p'} \in A$, where $\eta_{p'}(e) = 1 = 1 - \eta_p(e)$ and $\eta_{p'}(e') = \eta_p(e')$ for all other edges $e' \neq e$. Therefore, as $\delta \rightarrow 0$,

$$\begin{aligned}\mathbb{P}(\{\eta_{p+\delta} \in A\} \cap \{\eta_p \notin A\}) &= \sum_{e \in E} \mathbb{P}(\{e \text{ pivotal for } A\} \cap \{p \leq X(e) < p + \delta\}) + o(\delta) \\ &\stackrel{(a)}{=} \sum_{e \in E} \mathbb{P}(\{e \text{ pivotal for } A\}) \mathbb{P}(p \leq X(e) < p + \delta) + o(\delta) \\ &= \delta \sum_{e \in E} \mathbb{P}(\{e \text{ pivotal for } A\}) + o(\delta),\end{aligned}$$

where (a) follows from the independence of the state of an edge with the state of being pivotal or not. Inserting this relation into (10.10) yields

$$\mathbb{P}_{p+\delta}(A) = \mathbb{P}_p(A) + \delta \sum_{e \in E} \mathbb{P}(\{e \text{ is pivotal for } A\}) + o(\delta).$$

Dividing by δ and letting $\delta \rightarrow 0$, we obtain

$$\frac{d}{dp} \mathbb{P}(A) = \sum_{e \in E} \mathbb{P}(e \text{ is pivotal for } A) = \mathbb{E}(N(A)).$$

Russo's formula is not always valid for events that depend on more than finitely many edges: Indeed $\mathbb{P}(A)$ might not be differentiable for all p . For example, for percolation in two dimensions, $(\theta(p) - \theta(p_c))/(p - p_c)$ equals 0 if $p < p_c$ but tends to ∞ as $p \downarrow p_c$.

We can also recast Russo's formula in an integral form.

COROLLARY 10.19 (Integral form of Russo's formula) *Let A be an increasing event,*

which depends on the state of finitely many edges of \mathbb{L}^d , and let $N(A)$ denote the number of edges that are pivotal for A . Then, for any $0 \leq p_1 < p_2 \leq 1$,

$$\mathbb{P}_{p_2}(A) = \mathbb{P}_{p_1}(A) \exp \left(\int_{p_1}^{p_2} \frac{1}{p} \mathbb{E}_p(N(A) | A) dp \right).$$

Proof

$$\begin{aligned} \frac{d}{dp} \mathbb{P}_p(A) &= \sum_{e \in E} \mathbb{P}(\{e \text{ is pivotal for } A\}) \\ &\stackrel{(a)}{=} \frac{1}{p} \sum_{e \in E} \mathbb{P}(\{e \text{ is pivotal for } A\} \cap \{e \text{ is open}\}) \\ &= \frac{1}{p} \sum_{e \in E} \mathbb{P}(\{e \text{ is pivotal for } A\} \cap A) \\ &= \frac{1}{p} \sum_{e \in E} \mathbb{P}(e \text{ is pivotal for } A | A) \mathbb{P}_p(A) \\ &= \frac{1}{p} \mathbb{E}(N(A) | A) \mathbb{P}_p(A). \end{aligned}$$

Here (a) follows (again) from the independence of the state of an edge and the state of being pivotal. Dividing by $\mathbb{P}_p(A)$ on both sides and integrating over $[p_1, p_2]$ yields the result.

Good estimates of $\mathbb{E}(N(A) | A)$ are often hard to obtain. The trivial bound

$$\mathbb{P}_p(e \text{ is pivotal for } A | A) \leq 1$$

leads to

$$\mathbb{P}_{p_2} \leq (p_2/p_1)^m \mathbb{P}_{p_1}(A) \quad \text{if } 0 \leq p_1 < p_2 \leq 1, \quad (10.11)$$

where m is the (finite) number of edges in terms of which A is defined. This equality is useful to show that $\mathbb{P}_p(A)$ cannot increase very fast – unless m is large.

10.4.5 The square-root trick

The so-called *square-root trick* is a consequence of the FKG inequality.

If A_1, \dots, A_t are up-sets whose union A has a very high probability then one of the A_i must have high probability. Indeed, the complements A_i^c are *down-sets* so that $\mathbb{P}(A_1^c \cap A_2^c) \geq \mathbb{P}(A_1^c) \mathbb{P}(A_2^c)$ and, in general,

$$\prod_{i=1}^t \mathbb{P}(A_i^c) \leq \mathbb{P} \left(\bigcap_{i=1}^t A_i^c \right) = \mathbb{P}(A^c)$$

by De Morgan's law. It follows that for some i we have

$$\mathbb{P}(A_i^c) \leq (\mathbb{P}(A^c))^t,$$

i.e.,

$$\mathbb{P}(A_i) \geq 1 - (1 - \mathbb{P}(A_1 \cup \dots \cup A_t))^{1/t}.$$

In the case where each A_i has the same probability, this inequality holds for every i . For $t = 2$, this is known as the “square-root trick”; for $t = n$, this is the “ n th-root trick.”

Now we are adequately prepared for the derivation of the critical probability for bond percolation on the square lattice.

10.5 Critical threshold for bond percolation on the square lattice

10.5.1 Subcritical phase

Here we show that, if $p < p_c$, the distribution of the radius of the cluster size decreases exponentially. Let $S(n)$ be the diamond of radius n (i.e., the ball of radius n with the Manhattan or L_1 distance):

$$S(n) \triangleq \{x \in \mathbb{Z}^2 : \|x\|_1 \leq n\}.$$

An example for $n = 3$ is shown in Fig. 10.11. Let $A_n = \{0 \leftrightarrow \partial S(n)\}$ be the event that there exists an open path connecting the origin to any vertex lying on the surface of $S(n)$. Defining the radius of the origin's cluster C by $\text{rad } C = \max_{x \in C} |x|_1$, we see that $A_n = \{\text{rad } C \geq n\}$.

THEOREM 10.20 (Upper bound on the cluster radius) *If $p < p_c$, there exists $\psi(p) > 0$ such that*

$$\mathbb{P}(\text{rad } C \geq n) = \mathbb{P}(A_n) < \exp(-n\psi(p)).$$

Proof (Sketch) We start with Russo's formula in integral form, which states that for any $0 \leq p_1 < p_2 \leq 1$ we have

$$\mathbb{P}_{p_2}(A_n) = \mathbb{P}_{p_1}(A_n) \exp \left(\int_{p_1}^{p_2} \frac{1}{p} \mathbb{E}_p(N(A_n) \mid A_n) dp \right),$$

and thus

$$\begin{aligned}\mathbb{P}_{p_1}(A_n) &= \mathbb{P}_{p_2}(A_n) \exp\left(-\int_{p_1}^{p_2} \frac{1}{p} \mathbb{E}_p(N(A_n) | A_n) dp\right) \\ &\leq \exp\left(-\int_{p_1}^{p_2} \mathbb{E}_p(N(A_n) | A_n) dp\right).\end{aligned}$$

The next step is to choose $p_1 < p_c$ and show that the mean number of pivotal edges, given that A_n occurs, grows roughly linearly with n when $p < p_c$. The idea is that, since $p < p_c$, $\mathbb{P}(A_n) \rightarrow 0$ as $n \rightarrow \infty$, so that if A_n occurs it must depend critically on more and more edges since there can only be very few open paths connecting 0 to $\partial S(n)$. It is plausible that the number of pivotal edges in paths from the origin to $\partial S(2n)$ is approximately twice the number of such edges in paths to $\partial S(n)$ since these sparse paths have to traverse twice the distance. So it seems plausible that $\mathbb{E}(N(A_n) | A) = \Theta(n)$. The complete proof is rather lengthy.

THEOREM 10.21 (Cluster size distribution) *If $0 < p < p_c$, there exists $\lambda(p) > 0$ such that*

$$\mathbb{P}(|C| \geq n) \leq \exp(-n\lambda(p)) \quad (10.12)$$

and there exists $0 < \zeta(p) < \infty$ such that

$$\mathbb{P}(|C| = n) \leq \frac{(1-p)^2}{p} n \exp(-n\zeta(p)), \quad n \in \mathbb{N}.$$

Proof We shall give a sketch of the proof of the first part. The first step is to show that

$$\mathbb{E}(|C| e^{t|C|}) \leq \chi(p) (1 - 2t\chi(p)^2)^{-1/2}, \quad 0 \leq t < \frac{1}{2\chi(p)^2},$$

where $\chi(p) = \mathbb{E}(|C|)$. This derivation uses the BK inequality and some graphtheoretic facts. Next, by the Markov inequality,

$$\mathbb{P}(|C| \geq n) = \mathbb{P}(|C| e^{t|C|} \geq n e^{tn}) \leq \frac{1}{n e^{tn}} \mathbb{E}(|C| e^{t|C|}), \quad \text{for } t \geq 0.$$

Combining the two results yields

$$\mathbb{P}(|C| \geq n) \leq \frac{\chi(p)}{n e^{tn} (1 - 2t\chi(p)^2)^{1/2}}, \quad 0 \leq t < \frac{1}{2\chi(p)^2}.$$

Now set

$$t = \frac{1}{2\chi(p)^2} - \frac{1}{2n}, \quad n > \chi(p)^2,$$

to obtain

$$\mathbb{P}(|C| \geq n) \leq (e/n)^{1/2} \exp(-\frac{1}{2}n/\chi(p)^2),$$

□

which shows (10.12).

10.5.2 Supercritical phase

Here we establish the uniqueness of the infinite open cluster for $p > p_c$.

THEOREM 10.22 (Uniqueness of the infinite open cluster) *If $p > p_c$, then*

$$\mathbb{P}(\{\text{there exists exactly one infinite open cluster}\}) = 1.$$

Proof (Sketch) The statement is trivial if $p = 1$, so we assume $0 \leq p_c < p < 1$. Let N be the number of infinite open clusters. As before, let $S(n)$ be the diamond of Manhattan radius n . Write $N(0)$ for the number of infinite open clusters when all edges in $S(n)$ are closed, and $N(1)$ for when they are open. Finally, $M(n)$ denotes the number of infinite open clusters which intersect $S(n)$.

Since the sample space $\Omega = \{0, 1\}^E$ is a product space with a natural family of translations, \mathbb{P} is a product measure on Ω , and N is translation-invariant on Ω , it is a.s. constant, i.e.,

$$\exists k \in \mathbb{N} \cup \{0, \infty\} \text{ s.t. } \mathbb{P}(N=k) = 1.$$

Naturally the value of k may depend on p .

We show first that $k \in \{0, 1, \infty\}$. Suppose to the contrary that $2 \leq k < \infty$. Since every configuration on $S(n)$ has a strictly positive probability, it follows by the almost sure constantness of N that

$$\mathbb{P}(N(0) = k) = \frac{\mathbb{P}(\{N = k\} \cap \{\text{all edges of } S(n) \text{ are closed}\})}{\mathbb{P}(\{\text{all edges of } S(n) \text{ are closed}\})} = 1,$$

since $\mathbb{P}(\{N = k\}) = 1$. The same holds for the case when all edges in $S(n)$ are open, so

$$\mathbb{P}(N(0) = N(1) = k) = 1.$$

Now $N(0) = N(1)$ if and only if $S(n)$ intersects at most one infinite open cluster, since, if it intersected two or more, then it would matter whether $S(n)$'s edges are open or not; if they were open the clusters would be joined. (This is where we use $k < \infty$. If $k = \infty$, the number of infinite components would not change if two of them were merged.) Therefore

$$\mathbb{P}(M(n) \geq 2) = 0.$$

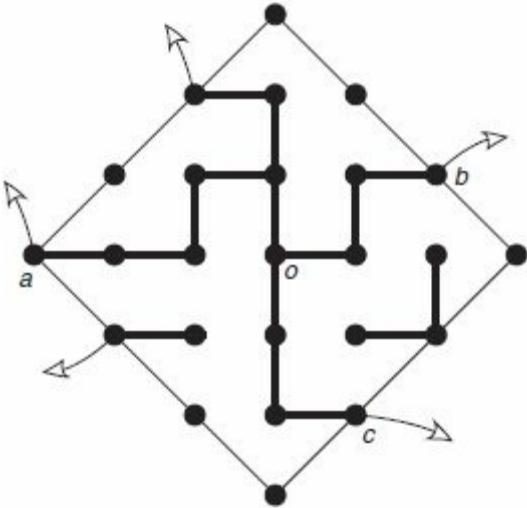


Figure 10.11 Illustration for Theorem 10.22. The diamond $S(n)$ is shown for $n = 3$. The arrows indicate connections to infinite open clusters that are distinct if edges in $S(n)$ do not connect them. Here, the vertex at the origin is a trifurcation if removing its three edges results in the partitioning of the infinite component it was connected to through the three boundary vertices a , b , and c , into three infinite open clusters. There is no other trifurcation in this configuration. A sufficient condition for o to be a trifurcation is if the three paths to a , b , and c are open and all other edges in $S(n)$ are closed. There can be no other trifurcation inside $S(n)$ that uses the same three boundary nodes as its gateways to the infinite cluster.

Since $S(n) \rightarrow \mathbb{Z}^2$ as $n \rightarrow \infty$, we have $M(n) \rightarrow N$ as $n \rightarrow \infty$, and

$$0 = \lim_{n \rightarrow \infty} \mathbb{P}(M(n) \geq 2) = \mathbb{P}(N \geq 2),$$

in contradiction with $\mathbb{P}(N = k) = 1$ for some $k \geq 2$.

Suppose next that $k = \infty$. This assumption also leads to a contradiction if the We have thus reached a contradiction, which means that mean number of *trifurcations* in a diamond $S(n)$ is calculated. A node x is called a trifurcation if it lies in an infinite open cluster, has (open) degree 3, and the deletion of x and its three open edges splits the infinite cluster into exactly three disjoint infinite clusters and no finite clusters. See Fig. 10.11 for an illustration. Letting T_x be the event that x is a trifurcation, it can be shown (under the assumption that $k = \infty$) that, for each $x \in \mathbb{Z}^2$, $\mathbb{P}(T_x) > 0$. So the mean number of trifurcations in $S(n)$ grows like $|S(n)| \sim n^2$. On the other hand, by construction, each trifurcation in $S(n)$ is connected to a unique vertex on $\partial S(n)$ which grows only linearly in n . This is the contradiction.

10.5.3 Critical threshold

The first result provides a lower bound on the critical probability.

THEOREM 10.23 (Harris' theorem)

$$\theta(1/2) = 0.$$

Proof We proceed (again) by contradiction. Suppose that $\theta(1/2) > 0$. Consider the box $B(n)$, and let $A^i, i \in \{t, b, l, r\}$ be the event that some vertex on the top, bottom, left, right side of $B(n)$ belongs to an infinite open path of \mathbb{L}_d^2 that uses no other vertex of $B(n)$. These are four increasing events of equal probability whose union is the event that some vertex of $\partial B(n)$ belongs to the infinite cluster. Since we assume $\theta(1/2) > 0$, there is a.s. an infinite cluster, and hence, as $n \rightarrow \infty$,

$$\mathbb{P}_{1/2} \left(\underbrace{\bigcup_{i \in \{t, b, l, r\}} A^i(n)}_{A(n)} \right) \rightarrow 1.$$

Now we are using the fourth-root trick (and the fact that the four probabilities are all equal) to obtain

$$\mathbb{P}_{1/2}(A^i(n)) \geq 1 - (1 - \mathbb{P}_{1/2}(A(n)))^{1/4}$$

so that

$$\mathbb{P}(A^i(n)) \rightarrow 1$$

as $n \rightarrow \infty$. Therefore there exists an n_0 large enough such that

$$\mathbb{P}(A^i(n_0)) > 7/8.$$

Now consider the dual box $B_d(n)$ defined as

$$B_d(n) = \{(i + 1/2, j + 1/2) : (i, j) \in B(n)\},$$

and let $A_d^i(n)$ be the event that some vertex on the respective side of $B_d(n)$ belongs to an infinite open path of \mathbb{L}_d^2 that uses no other vertex of $B_d(n)$. Each edge of \mathbb{L}_d^2 is open with probability $1/2$, so $\mathbb{P}_{1/2}(A_d^i(n)) = \mathbb{P}_{1/2}(A^i(n))$ for all n , and, in particular,

$$\mathbb{P}_{1/2}(A_d^i(n_0)) > 7/8.$$

Consider the event

$$Z = A^l(n_0) \cap A^r(n_0) \cap A_d^t(n_0) \cap A_d^b(n_0)$$

that there exist infinite open paths of \mathbb{L}^2 connecting to the left and right sides of $B(n_0)$ without using any other vertex of $B(n_0)$ and that there exist infinite closed paths connecting to some vertex on the top and bottom sides of $B_d(n_0)$ without using any other vertex of $B_d(n_0)$. Using the union bound,

$$\begin{aligned}\mathbb{P}_{1/2}(Z) &= 1 - \mathbb{P}(\bar{A}^l(n_0) \cup \bar{A}^r(n_0) \cup \bar{A}_d^t(n_0) \cup \bar{A}_d^b(n_0)) \\ &\geq 1 - (\mathbb{P}_{1/2}(\bar{A}^l(n_0)) + \mathbb{P}_{1/2}(\bar{A}^r(n_0)) + \mathbb{P}_{1/2}(\bar{A}_d^t(n_0)) + \mathbb{P}_{1/2}(\bar{A}_d^b(n_0))) \\ &> 1/2\end{aligned}$$

since $\mathbb{P}_{1/2}(\bar{A}^i(n_0)) < 1/8$. If Z occurs, there must be two infinite open clusters in $\mathbb{L}^2 \setminus B(n_0)$, one containing the infinite open path connected to the left side of $B(n)$ and the other containing the infinite open path connected to the right side of $B(n)$. These two infinite open clusters must be disjoint because they are separated by two infinite closed paths in $\mathbb{L}^2_d \setminus B_d(n_0)$. See Fig. 10.12 for an example.

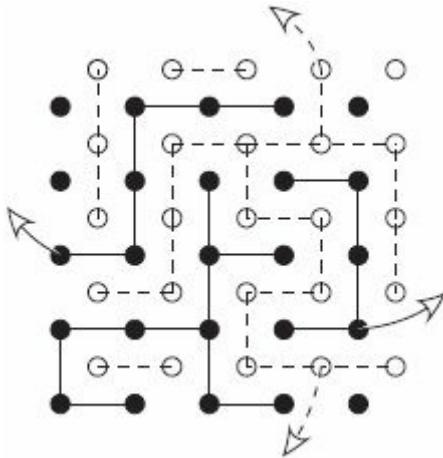


Figure 10.12 Illustration for Harris' theorem 10.23. The figure shows infinite open paths of $\mathbb{L}^2 \setminus B(n_0)$ for $n_0 = 2$ connecting some vertex on the left and right sides of $B(n_0)$ and infinite closed paths in $\mathbb{L}^2_d \setminus B_d(n_0)$ connecting some vertex on the top and bottom sides of $B_d(n_0)$. If there were two infinite closed paths in the dual lattice, as in this figure, this would guarantee that the left side and right side of $B(n_0)$ are connected to two separate infinite open clusters.

The same reasoning implies that there must be two disjoint infinite closed clusters in $\mathbb{L}^2_d \setminus B_d(n_0)$, one containing an infinite path starting from the top and the other from the bottom, and separated by the two infinite open paths of $\mathbb{L}^2 \setminus B(n_0)$. Now since $\theta(1/2) > 0$, the uniqueness theorem says that there is exactly one infinite open cluster. Therefore, there must be a left-right crossing within $B(n)$ which forms a barrier to any top-bottom closed crossing of $B_d(n)$. As a consequence there must

be at least two disjoint infinite closed clusters in \mathbb{L}_d^2 , but the probability that there are two infinite open clusters in \mathbb{L}_d^2 is the same as the probability that there are two infinite open clusters in \mathbb{L}^2 , which is zero.

We have thus reached a contradiction, which means that $\mathbb{P}_{1/2}(Z)$ has to be zero, and the assumption $\theta(1/2) > 0$ cannot hold. \square

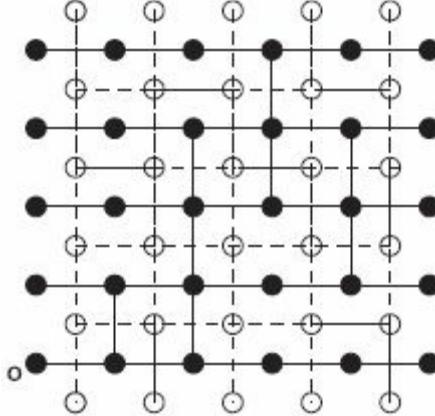


Figure 10.13 Illustration for Lemma 10.24. Shown are the rectangle $R(n)$ and its dual $R_d(n)$ for $n = 4$. The thin lines indicate all the edges inherited from \mathbb{L}^2 (solid) and \mathbb{L}_d^2 (dashed), respectively. The thick lines indicate open edges, solid for $R(n)$ and dashed for $R_d(n)$. There is always either a left-right crossing in $R(n)$ or a top-bottom crossing in $R_d(n)$, and the two events are mutually exclusive.

Harris' theorem implies that $p_c \geq 1/2$. The next lemma is the main step in showing the converse $p_c \leq 1/2$.

LEMMA 10.24 (Crossing the square for $p = 1/2$) *Let $LR(n)$ be the event that there is a left-right crossing of the rectangle $R(n) = [0, n + 1] \times [0, n] \cap \mathbb{Z}^2$. Then $\mathbb{P}_{1/2}(LR(n)) = 1/2$ for all n .*

Proof The rectangle $R(n)$ is the subgraph of \mathbb{L}^2 having vertex set $[0, n + 1] \times [0, n] \cap \mathbb{Z}^2$ and edge set comprising all edges of \mathbb{L}^2 joining pairs of vertices in $S(n)$ except those joining pairs $(i, j), (k, l)$ with either $i = k = 0$ or $i = k = n + 1$, i.e., edges in the left or right sides. Let $R_d(n)$ be the subgraph of \mathbb{L}_d^2 having vertex set $\{(i + 1/2, j + 1/2) : 0 \leq i \leq n, -1 \leq j \leq n\}$ and all edges inherited from \mathbb{L}_d^2 except for the one along the top and bottom sides. The two subgraphs are obtained by a 90 degree rotation. See Fig. 10.13 for an illustration.

Let us consider the following two events: $LR(n)$ is the event that there exists an open path of $R(n)$ joining a vertex on the left side of $R(n)$ to a vertex on the right side, and $TB_d(n)$ is the event that there exists an open path of $R_d(n)$ joining a vertex on the top to a vertex at the bottom.

If $\text{LR}(n) \cap \text{TB}_d(n) \neq \emptyset$, there is a left–right open path in $R(n)$ crossing a top–bottom closed path in $R_d(n)$. But at the crossing of these two paths an open edge of \mathbb{L}^2 would cross an open edge of \mathbb{L}^2_d , which is impossible. Hence $\text{LR}(n) \cap \text{TB}_d(n) = \emptyset$. On the other hand, either $\text{LR}(n)$ or $\text{TB}_d(n)$ must occur. Let D be the set of vertices that are reachable from the left side of $R(n)$ by an open path. Suppose that $\text{LR}(n)$ does not occur. Then there exists a top–bottom open path of \mathbb{L}^2_d crossing only edges of $R(n)$ contained in the edge boundary of D , and so $\text{TB}_d(n)$ occurs. Consequently, $\text{LR}(n)$ and $\text{TB}_d(n)$ form a partition of the sample space Ω and

$$\mathbb{P}(\text{LR}(n)) + \mathbb{P}(\text{TB}_d(n)) = 1.$$

Since $R(n)$ and $R_d(n)$ are isomorphic, flipping each edge of \mathbb{L}^2_d yields that $\mathbb{P}_p(\text{TB}_d(n)) = \mathbb{P}_{1-p}(\text{LR}(n))$, thus

$$\mathbb{P}_p(\text{LR}(n)) + \mathbb{P}_{1-p}(\text{LR}(n)) = 1.$$

□

Inserting $p = 1/2$ proves the lemma.

Now we are ready to prove one of the most celebrated results in percolation theory.

THEOREM 10.25 (Critical threshold for bond percolation on the square lattice) *The critical probability for bond percolation on the square lattice is*

$$p_c = 1/2.$$

Proof We know that $p_c \geq 1/2$. Suppose that $p_c > 1/2$. Then $p = 1/2$ belongs to the subcritical phase, and we know from Theorem 10.20 that there exists $\psi(1/2) > 0$ such that for all n

$$\mathbb{P}_{1/2}(0 \leftrightarrow \partial^r R(n)) \leq \mathbb{P}_{1/2}(0 \leftrightarrow \partial S(n)) < \exp(-n\psi(1/2)),$$

where $\{0 \leftrightarrow \partial^r R(n)\}$ is the event that the origin is connected by an open path to a vertex lying on the right side of $R(n)$. Consequently, since $\text{LR}(n)$ is the event that there exists an open path of $R(n)$ joining a vertex on the left side of $R(n)$ to a vertex on the right side,

$$\begin{aligned}
\mathbb{P}_{1/2}(\text{LR}(n)) &\leq \sum_{k=0}^n \mathbb{P}_{1/2}((0, k) \leftrightarrow \partial^r R(n)) \\
&\leq (n+1)\mathbb{P}_{1/2}(0 \leftrightarrow \partial S(n)) \\
&< (n+1)\exp(-n\psi(1/2)),
\end{aligned}$$

which yields that $\mathbb{P}_{1/2}(\text{LR}(n)) \rightarrow 0$ as $n \rightarrow \infty$ and therefore contradicts the crossing-the-square lemma. Consequently, $p_c \leq 1/2$, which completes the proof. \square

10.5.4 Steps of the proof

Here is a list of the steps we took to establish the result on the bond percolation in \mathbb{L}^2 .

- We have shown the uniqueness of the infinite open cluster in the supercritical phase (a.s. constant number of infinite open clusters; existence of trifurcations if that number were infinite).
- We have lower bounded the critical threshold: $p_c \geq 1/2$ since $\theta(1/2) = 0$. If $\theta(1/2) > 0$ we would have at least two infinite open clusters. Here we used the fourth-root trick, which follows from the FKG inequality.
- Next we proved the crossing-the-square lemma for $p = 1/2$: The probability of an LR crossing is $1/2$. Here we used the dual lattice and the fact that there is either an LR or a TB_d crossing.
- Finally we upper bounded the critical threshold by showing that $p_c \leq 1/2$. To do that, we assumed that $p = 1/2$ belongs to the subcritical phase. But then the exponentially decreasing tail of the radius of the mean cluster size in the subcritical phase, which we obtained by an application of Russo's formula in integral form, would imply that there cannot be an LR crossing in the rectangle $R(n)$, which we have shown to occur with probability $1/2$, independently of n . This contradiction proves that $p = 1/2$ is not subcritical.
- The lower and upper bounds together establish that $p_c = 1/2$.

Actually, in the square lattice, there is no need for the full trifurcation argument to establish the uniqueness of the infinite component. Instead, we may argue as follows. Consider a large box B centered at o . Since $p = 1/2$ is subcritical, a.s. there is no infinite open path intersecting B , so B is surrounded by a closed circuit in the dual lattice. Since \mathbb{L} and \mathbb{L}_d are dual, B is also surrounded by open circuits in the original lattice. By the coupling argument, this must also hold for $p > 1/2$. Since B can be arbitrarily large, such open circuits must intersect two hypothetical infinite components, which would join them to one. So there can be only one infinite component.

10.6 Further results in bond percolation

10.6.1 At the critical point

Does there exist an infinite cluster when $p = p_c$? The answer is negative when $d = 2$ (as we showed) or $d \geq 19$. In the two-dimensional case, we can bound the distribution of the component size for $p = 1/2$.

THEOREM 10.26 (Power law inequalities at the critical threshold)

$$\mathbb{P}_{1/2}(0 \leftrightarrow \partial B(n)) \geq \frac{1}{2\sqrt{n}} \quad \forall n \in \mathbb{N}$$

and

$$\mathbb{P}_{1/2}(|C| \geq n) \geq \frac{1}{2\sqrt{n}} \quad \forall n \in \mathbb{N}.$$

Proof Since any open path connecting the origin to $\partial B(n)$ contains at least n vertices,

$$\mathbb{P}_{1/2}(|C| \geq n) \geq \mathbb{P}_{1/2}(0 \leftrightarrow \partial B(n)),$$

so we need only prove the first statement.

As before, let $LR(2n - 1)$ be the event that there is an open path in the rectangle $R(2n - 1) = [0, 2n] \times [0, 2n - 1] \cap \mathbb{Z}^2$ connecting some vertex on its left side to some vertex on its right side. This path must cross the center line $\{(n, k) \in \mathbb{Z}^2 : 0 \leq k \leq 2n - 1\}$ in at least one vertex. That vertex thus is connected by two disjoint paths to the left and right sides of $R(2n - 1)$.

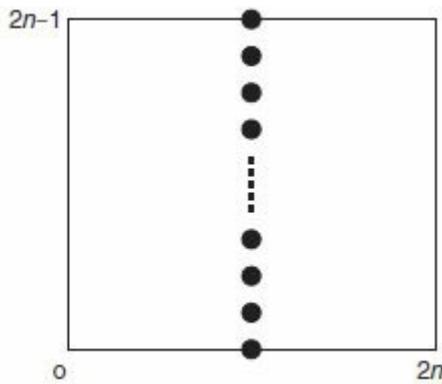


Figure 10.14 Illustration for Theorem 10.26. Any point on the middle line of the rectangle $R(2n - 1)$ must be connected by disjoint paths by the left and the right side of the rectangle by the crossing-the-square lemma (Lemma 10.24), which occurs with probability 1/2. This probability can be compared with the probability of one of these middle nodes being connected to the boundary of the $2n \times 2n$ box centered at these middle nodes.

Denoting by $A_n(k)$ the event that the vertex (n, k) is joined by an open path to the

surface $\partial B(n, (n, k))$ of the box $B(n, (n, k))$ centered at (n, k) of side length $2n$, we have by the union bound

$$\mathbb{P}_{1/2}(\text{LR}(2n-1)) \leq \sum_{k=0}^{2n-1} \mathbb{P}_{1/2}(A_n(k) \circ A_n(k)).$$

Applying the BK inequality, we obtain

$$\begin{aligned} \mathbb{P}_{1/2}(\text{LR}(2n-1)) &\leq \sum_{k=0}^{2n-1} \mathbb{P}_{1/2}^2(A_n(k)) \\ &= 2n \mathbb{P}_{1/2}^2(A_n(0)) \\ &= 2n \mathbb{P}_{1/2}^2(0 \leftrightarrow \partial B(n)). \end{aligned}$$

We know that $\mathbb{P}_{1/2}(\text{LR}(2n-1)) = 1/2$ for all $n \in \mathbb{N}$, so $1/2 \leq 2n \mathbb{P}_{1/2}^2(0 \leftrightarrow \partial B(n))$, and the result follows. The situation is depicted in Fig. 10.14.

Also we obtain directly the tail of the distribution of the radius $\text{rad } C$: From

$$\mathbb{P}_p(0 \leftrightarrow \partial B(n/2)) \leq \mathbb{P}_p(0 \leftrightarrow \partial S(n)) = \mathbb{P}_p(\text{rad } C \geq n)$$

it follows that

$$\mathbb{P}_{1/2}(\text{rad } C \geq n) \geq \frac{1}{\sqrt{2n}}.$$

So, at the critical threshold, the distribution of the size of the cluster at the origin is governed no longer by an exponential law, but rather by a power law.

10.6.2 Generalization to d dimensions

As discussed in Section 10.3.2, assuming that $\theta(p_c) = 0$, it is believed that

$$\mathbb{P}_{pc}(|C| \geq n) \sim n^{-1/\delta}, \quad n \rightarrow \infty,$$

i.e., the tail decays according to a power law with critical exponent $\delta > 0$.

For the critical probability itself, the following series expansion has been established:

$$p_c(d) = \frac{1}{2d} + \frac{1}{4d^2} + \frac{7}{16d^3} + O(d^{-4}).$$

10.7 Site percolation

In a site percolation model, vertices are open with probability p and closed with probability $1-p$. The definitions of bond percolation carry over. Site percolation is *more general than bond percolation* in the sense that every bond model can be recast as a site model, but not the reverse. To recast a bond model as a site model, we make use of the notion of a *covering graph* G_c of a graph G , which is obtained as follows. Place a vertex of G_c in the middle of each edge of G . Two vertices of G_c are declared adjacent if and only if the two corresponding edges of G share a common end vertex of G . Define now a bond percolation process on G , and declare a vertex of G_c to be open (closed) if the corresponding edge of G is open (closed). This results in a site percolation process on G_c . Any path of open edges of G corresponds to a path of open vertices of G_c , and vice versa. As a result,

$$p_c^b(G) = p_c^s(G_c).$$

For example, if $G = \mathbb{L}^2$, then $G_c = \mathbb{L}_c^2$ is the lattice shown in Fig. 10.15. Hence the site percolation threshold for this covering graph is $1/2$.

10.7.1 Site percolation on the triangular lattice

A triangular lattice \mathbb{T} (in two dimensions) is a regular lattice where each site (vertex) has degree 6. It can be obtained by adding one diagonal bond in each elementary square of a square lattice \mathbb{L}^2 , say the bottom-left to top-right diagonal. This shows that $p_c^b(\mathbb{T}) \leq p_c^b(\mathbb{L}^2)$. One can show that the equality is strict: $p_c(\mathbb{T}) < p_c(\mathbb{L}^2) = 1/2$; and, with more effort, that the critical probability for bond percolation on the triangular lattice is

$$p_c^b(\mathbb{T}) = 2\sin(\pi/18) \approx 0.347.$$

Turning to site percolation, the critical probability for site percolation on the triangular lattice is

$$p_c^s(\mathbb{T}) = 1/2.$$

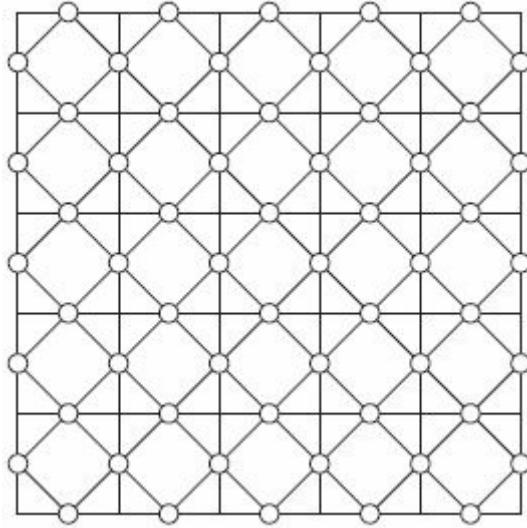


Figure 10.15 Covering graph of the lattice \mathbb{L}^2 .

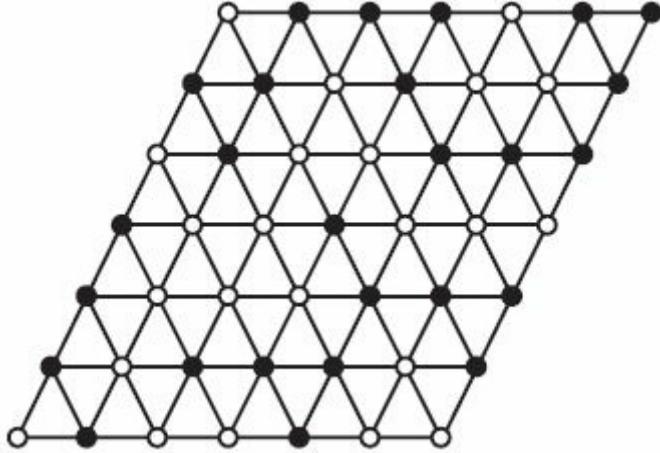


Figure 10.16 Site percolation on the triangular lattice \mathbb{T} . Each rhombus has a TB path or an LR path, and the two events are mutually exclusive.

The proof of this fact is based on a similar argument to that which we used for bond percolation on the square lattice. One important observation is that, considering a rhombus of a certain side length n , the probability of an open LR path is $1/2$ for all $n \in \mathbb{N}$ for $p = 1/2$. The probability of a TB path is also $1/2$, and the two events partition the sample space, see Fig. 10.16. So site percolation on the triangular lattice exhibits a self-duality property.

In the covering graph in Fig. 10.15, each site also has degree 6, and we know that the site percolation threshold is also $1/2$ – both these lattices with fixed degree 6 have the same site percolation threshold.

10.7.2 General results and bounds

THEOREM 10.27 *For any infinite connected graph G , we have $p_c^s(G) \geq p_c^b(G)$.*

Proof We have

$$\mathbb{P}_p^s(|C_x| \geq n) \leq p \mathbb{P}_p^b(|C_x| \geq n) \quad \forall n \in \mathbb{N}, \forall p \in (0, 1), \forall x \in G$$

since x is open with probability p and a closed site implies at least one closed edge. Letting $n \rightarrow \infty$, we obtain

$$\theta_x^s(G) \leq p \theta_x^b(G),$$

so if $\theta_x^b = 0$ then $\theta_x^s = 0$, hence $p_c^s(G) \geq p_c^b(G)$. The graph can also be a multigraph but it has to be locally finite (finite node degrees).

THEOREM 10.28 *For an infinite connected graph G with maximum degree $1 < \Delta < \infty$,*

$$\frac{1}{\Delta - 1} \leq p_c^b(G) \leq p_c^s(G) \leq 1 - (1 - p_c^b(G))^{\Delta - 1}.$$

The first inequality is obtained by comparison with a branching process. The second one is the previous theorem. The right side is the probability that a vertex of maximum degree has degree at least 2. A particular vertex does not help create an infinite component if it has degree only 1. It follows for site percolation on \mathbb{L}^2 that $p_c^s \leq 7/8$.

Next we state without proof two fairly general results that connect critical probabilities of certain related lattices.

THEOREM 10.29 *For any planar lattice (subject to some symmetry constraints) \mathbb{L}^2 and its dual \mathbb{L}_d^2 ,*

$$p_c^b(\mathbb{L}^2) + p_c^b(\mathbb{L}_d^2) = 1.$$

This result implies Theorem 10.25. The triangular lattice has $p_c^b = 2 \sin(\pi/18)$, so the hexagonal lattice has $p_c^b = 1 - 2\sin(\pi/18)$. Not surprisingly, the threshold for the triangular lattice is smaller since each vertex has degree 6. $2 \sin(\pi/18)$ is approximately $\pi/9 \approx 0.35$. It follows from Theorem 10.28 that the critical probability site percolation on the triangular lattice ($\Delta = 6$) is bounded by

$$p_c^s(\mathbb{T}) \leq 1 - (1 - 2 \sin(\pi/18))^5 \approx 0.88,$$

which is not particularly tight given that the exact value is $1/2$.

THEOREM 10.30 *Let \mathbb{L} be a symmetric planar lattice, and let \mathbb{L}^\times be the graph obtained from \mathbb{L} by adding all diagonals to all faces of \mathbb{L} . Then*

$$p_c^s(\mathbb{L}) + p_c^s(\mathbb{L}^\times) = 1.$$

Again this is derived from the fact that for a suitably chosen “rectangle,” there is either an LR open crossing in \mathbb{L} or a TB crossing in \mathbb{L}^\times . This theorem implies that $p_c^s(\mathbb{T}) = 1/2$, since all faces are triangles.

10.7.3 Numerical bounds

Consider site percolation on \mathbb{L}^2 . Take a square of side length n and consider the event that there is an LR crossing through the square. From the theory we know that $\lim_{n \rightarrow \infty} \mathbb{P}_p(\text{LR}(n)) = 0$ if $p < p_c$ while $\lim_{n \rightarrow \infty} \mathbb{P}_p(\text{LR}(n)) = 1$ for $p > p_c$. With increasing n the transition occurs more and more sharply. In particular the value $p(n)$ at which $\mathbb{P}_{p(n)}(\text{LR}(n)) = 1/2$ tends to p_c .

The problem is that the rate of convergence to p_c is unknown.

It has been established in Newman & Ziff (2000) using simulations that

$$p_c^s(\mathbb{L}^2) = 0.59274621 \pm 0.00000013.$$

Bibliographical notes

The central result in percolation theory that $0 < p_c^b(\mathbb{L}) < 1$ for bond percolation on \mathbb{L}^2 was proved in Broadbent & Hammersley (1957). This and subsequent publications by Hammersley initiated the study of percolation. The main result derived in this chapter, namely that $p_c^b(\mathbb{L}) = 1/2$, was first proved in Kesten (1980). It is often referred to as the Harris–Kesten result. Peierls’ argument was introduced in Peierls (1936), and the FKG inequality was proved in Fortuin *et al.* (1971). The uniqueness argument presented at the end of [Section 10.5.4](#) is due to Harris (1960).

The books by Grimmett (1999) and by Bollobás & Riordan (2006) provide an in-depth treatment of bond and site percolation, including complete proofs of the theorems not proved here. The two books disagree on whether edges in the dual graph should be called closed or open if they cross an open edge of the original graph – see [Box 10.1](#). Here we follow Grimmett’s convention.

In the book by Bollobás and Riordan, [Chapter 3](#) presents an alternative proof of Theorem 10.21 using dependent percolation. The results in [Section 10.7](#) can be found in [Chapter 5](#), which also includes a survey of the different methods which can

be used to prove the Harris–Kesten result.

A general account of critical exponents in percolation models is provided in Durrett (1985).

Athreya & Ney (2004) provided a unified treatment of branching processes, including multitype branching.

The method of coupling bond percolation processes described in Section 10.2.2 is an instance of the much more general *coupling method*, which is frequently used to establish inequalities or estimates of distances of probability distributions. The book by Lindvall (2002) gives an introduction and a broad set of examples on coupling.

Problems

10.1 Consider the 2-branching tree (also called binary tree) with binomial offspring distribution. Determine the percolation probability and plot it, and find the critical exponents β , γ , and δ in (10.4), (10.5), and (10.6).

10.2 Show that the sequence of the numbers of members (Z_n) in each generation of the binary tree forms a Markov chain and find the transition probabilities.

10.3 Find the probability of extinction η for the branching process whose offspring distribution is Poisson with mean λ .

10.4 Let T be the total progeny of the branching process, i.e.,

$$T \triangleq \sum_{n=0}^{\infty} Z_n,$$

and let its generating function be $G_T(s) = \mathbb{E}(s^T)$. Show that

$$G_T(s) = sG(G_T(s)),$$

where $G_{(s)} = \mathbb{E}(s^X)$ is the generating function of the offspring distribution.

Hint. Establish first that, conditioned on $Z_1 = i$,

$$T \stackrel{d}{=} 1 + \sum_{k=1}^i T_k,$$

where the T_k are iid random variables with the same distribution as T .

10.5 For a branching process with mean offspring $\mu < 1$, show that the expected total progeny is

$$\mathbb{E}T = \frac{1}{1-\mu}.$$

- 10.6** Explain the recursion formula (10.3) using the random walk argument.
- 10.7** Determine (analytically) the percolation threshold of the lozenge lattice in Fig. 10.17.
- 10.8** Pivotal edges. Consider the event $A = \{o \leftrightarrow \partial B(1)\}$ on a bond percolation model on \mathbb{L}^2 . Let the four edges of o be e_1, \dots, e_4 . Give an explicit expression for the event W_i that edge e_i , $i \in [4]$, is pivotal, and determine $\mathbb{P}_p(W_i)$. Then find $\mathbb{E}_p(N(A))$ and $\mathbb{E}_p(N(A) | A)$. How do the two expectations behave as $p \rightarrow 0$? Is this behavior the one that would be expected? Explain.
- 10.9** Russo's formula. First, verify your result from Problem 10.8 for $\mathbb{E}_p(N(A))$ using the original (differential) form of Russo's formula. Next, study the proof of Corollary 10.19 (the integral form of Russo's formula) and complete it by integrating the last equation in the proof. Next, using again the concrete event from Problem 10.8 and the fact that you know $\mathbb{P}_p(A)$, "reverse engineer" $\mathbb{E}_p(N(A) | A)$ from Russo's integral formula – and compare this result with your expression from Problem 10.8.
- 10.10** Prove that $p_c = p'_c$ for bond percolation on \mathbb{L}^2 , where

$$p'_c = \sup \{p : \chi(p) < \infty\}$$

and $\chi(p)$ is the expected size of the component at the origin (as defined in Section 10.1.4).

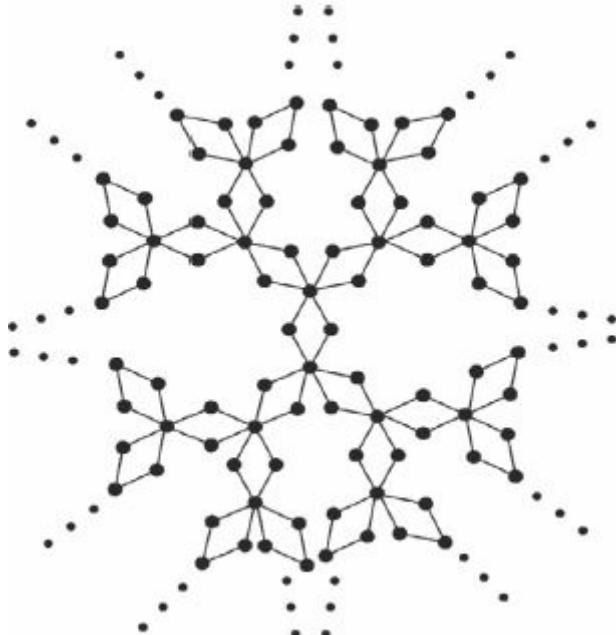


Figure 10.17 Lozenge lattice for Problem 10.7.

- 10.11** When proving that the number of infinite components cannot be infinite, we used the concept of trifurcations. Draw a configuration inside the diamond $S(n)$ for some n that has three trifurcations, and indicate the boundary vertices that are

assumed to be connected to an infinite open cluster. Explain why bifurcations would not be sufficient to prove that there cannot be an infinite number of infinite open clusters.

¹ Not to be confused with the prior use of G as the pgfl of a point process.

² Here, “closed” refers to the fact that the first and last vertices are identical, not to open or closed edges in the probabilistic bond percolation model.

³ Here, “closed” refers to the edges in the circuit being closed in the percolation model.

11.1 Introduction

In this chapter, we introduce *random geometric graphs* (RGGs) as generalizations of the lattice models considered in the previous chapter. These are graphs whose vertices are embedded in \mathbb{R}^d and whose edges depend on the spatial location of the vertices. The locations are usually modeled using a point process, and the edges may be formed deterministically or probabilistically given the locations. We will use the terms vertex and point interchangeably. For a specific class of graphs, we will bound the percolation probability. Since the vertices can assume arbitrary locations in some space, usually the Euclidean space \mathbb{R}^d , one speaks of *continuum percolation*.

DEFINITION 11.1 (Basic random geometric graph) Let $\Phi = \{x_i\} \subset \mathbb{R}^d$ be a (simple) point process. For each (unordered) pair of points $\{x, y\} \subset \Phi$, add an edge with probability $\xi(x - y)$, independently for each pair, where $\xi : \mathbb{R}^d \mapsto [0, 1]$ is the (translation-invariant) *connectivity function*. The *effective area* (or volume) of ξ is

$$\mathcal{A}_\xi \triangleq \int_{\mathbb{R}^d} \xi(y) dy,$$

and the resulting graph $G = (\Phi, E)$ is a random geometric graph.

Hence there are two sources of randomness in a random geometric graph, the location of the points (or vertices) and the existence of edges. The point process may be finite or infinite. If it is finite, a frequently used model is the (general) binomial point process, usually on the unit cube $[0, 1]^d$. If it is infinite, the typical model is the Poisson point process (PPP).

Finite models suffer from boundary effects, since vertices near the boundary typically have smaller degrees than do vertices near the center. A remedy to this problem is provided by graphs defined on a torus. In most cases, the unit torus $[0,$

$1]^d$ is used, with the distance metric denoted as $\|\cdot\|_t$ and defined by

$$\|x - y\|_t \triangleq \min\{\|x + z - y\| : z \in \mathbb{Z}^d\}.$$

We call the models defined with a connectivity function basic, since the existence of edges is independent for each vertex pair (given the distance). There exist other models, in which either the presence of an edge does not depend solely on the distance, or there is dependence between the existence of edges. Another class of non-basic RGGs consists of directed (basic) random geometric graphs, where directed edges $\vec{x}y$ are formed with probability $\xi(x - y)$, independently for each *ordered* pair (x, y) of points.

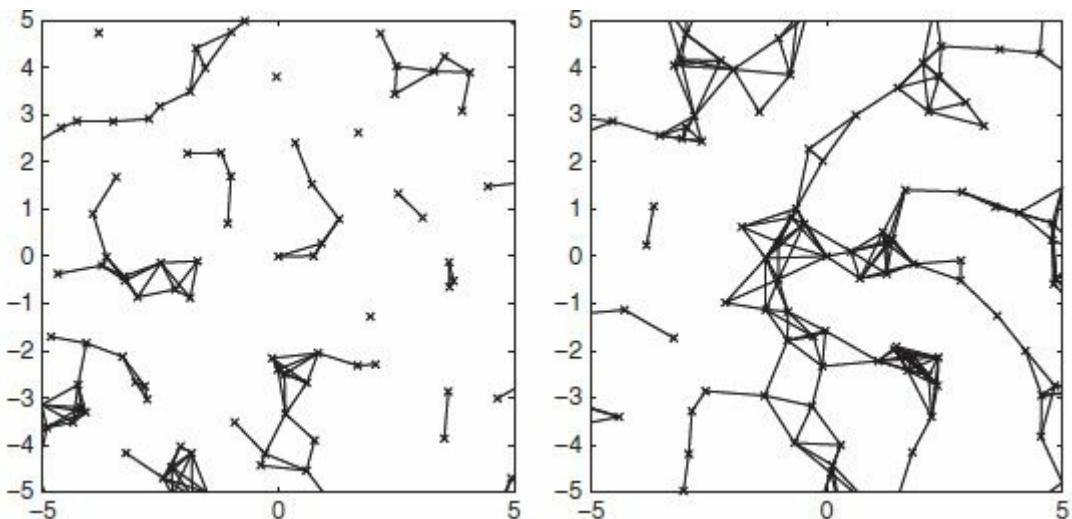


Figure 11.1 Gilbert disk graphs $G_{r,1}$ on the window $[-5, 5]^2$ with radii $r = 1$ (left) and $r = 1.3$ (right). For $r = 1$, there are several small components, whereas for $r = 1.3$, a larger component emerges.

Example 11.1 In the previous chapter, we have encountered several examples of basic RGGs already: The bond percolation model on the square lattice is an RGG with $\Phi = \mathbb{L}^2$ and $\xi(v) = p\mathbf{1}(\|v\| = 1)$; and the site percolation model on the square lattice is an RGG where Φ is a Bernoulli lattice process (see Section 2.4.7) and $\xi(v) = \mathbf{1}(\|v\| = 1)$.

The most basic and well-studied model is *Gilbert's disk graph*, which has a deterministic connectivity function.

DEFINITION 11.2 (Gilbert's disk graph) Gilbert's disk graph $G_{r,\lambda} = (\Phi, E)$ is a basic random geometric graph, where Φ is a uniform PPP of intensity λ on the plane and

the connectivity function is $\zeta(v) = \mathbf{1}(\|v\| \leq r)$, i.e., there is an edge between $x, y \in \Phi$ if $\|x - y\| \leq r$.

In the disk graph, the only source of randomness is the location of the vertices. Its effective area is πr^2 . Two realizations of disk graphs with $\lambda = 1$ are shown in Fig. 11.1. Owing to its importance, the next section is devoted to this model.

11.2 Percolation on Gilbert's disk graph

For the disk graph $G_{r,\lambda}$, by Definition 11.2, the node degree is Poisson with mean $\lambda\pi r^2$. The natural question is that of how to find the critical mean node degree $a \triangleq \lambda\pi r^2$ such that there is an infinite component in the graph. Assume that o is part of the PPP – which we can achieve by conditioning on $o \in \Phi$ or, equivalently, adding o to Φ . Then a basic problem in *continuum percolation* problem is that of how to find

$$a_c = \sup \{a : \theta(a) = 0\},$$

where $\theta(a)$ is the probability that the component of $G_{r,\lambda}$ containing o is infinite. Next we discuss several techniques to bound the critical mean degree.

11.2.1 Bounding using branching processes

The idea behind this technique is to start with the vertex o of Φ , grow the cluster containing o in “generations,” and compare the growing cluster to a branching process.

Assume that the origin o is a point of Φ . It forms generation 0. For the first generation, pick the points of Φ within distance r of o . The second generation consists of the points of Φ which are each within distance r of a first-generation point, but are not in the first generation themselves (i.e., they are not within distance r of o). The third generation consists of the points of Φ not belonging to the first two generations, but which are each within distance r of some second-generation point, and so on. The associated branching process is obtained by setting each offspring size distribution to be Poisson with mean $\lambda\pi r^2$, so that we are essentially growing the same cluster containing o , but ignoring the fact that the various disks we have scanned for points actually overlap. So the progeny (total number of nodes in the branching process, summed over generations 0 through n) stochastically dominate the corresponding number in the disk graph. If $\lambda\pi r^2 \leq 1$, the branching process dies out with probability 1, so the process on the disk graph dies out, too,

which means that the critical degree for percolation is at least 1. Hence

$$a_c = (\lambda \pi r^2)_c > 1.$$

When $\lambda \pi r^2 > 1$, the probability that the branching process dies out can be calculated (see Problem 10.3), which provides an upper bound on the percolation probability θ at o , i.e., the probability that o belongs to an infinite component.

11.2.2 Bounding using bond percolation on the square lattice

In this technique, we use what we know from bond percolation on the integer (square) lattice to find an upper bound. We divide the plane into squares of size $c = r/\sqrt{8}$. Each square corresponds to an edge of the lattice which is open if the square contains at least one point of the PPP. Thus, each edge is open with probability $p = 1 - \exp(-\lambda c^2)$ independently of all other edges, see Fig. 11.2.

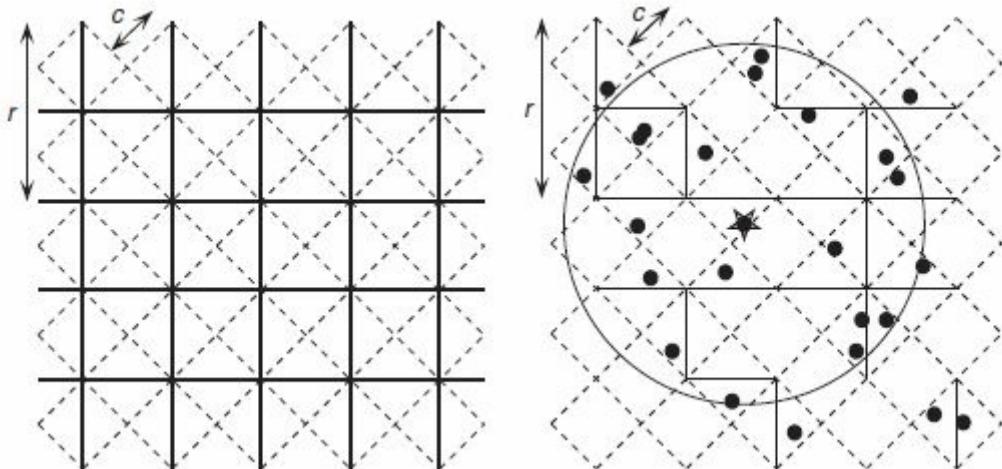


Figure 11.2 (Left) Partitioning of the plane into squares of side length $c = r/\sqrt{8}$. The bold edges are the edges of the lattice $(r/2)\mathbb{L}^2$. Each square corresponds to a bond. (Right) The bond is declared open if its square contains a point of the PPP, so $p = 1 - \exp(-\lambda c^2)$ for the bond percolation model. To illustrate the radius r of the disk graph, the circle of radius r around the point indicated by the star near the center is also drawn. If two points fall within squares that touch at least at one corner, they are within distance r . Thus the disk graph $G_{r,\lambda}$ percolates if the bond model does.

Since $p_c = 1/2$ for bond percolation, if

$$\lambda \geq \frac{\log 2}{c^2}$$

the bond percolation model contains an infinite cluster a.s. Now, if two edges are

adjacent and open, then by construction two points of the PPP are located in two squares that touch at a corner at least. Two such points cannot be further from each other than r , and they are connected in $G_{r,\lambda}$. Accordingly, an infinite collection of connected open edges corresponds to an infinite set of connected points in $G_{\lambda,r}$. So there exists an infinite open cluster a.s. if

$$\lambda \geq \frac{\log 2}{c^2} \implies a_c = (\lambda \pi r^2)_c \leq 8\pi \log 2 \approx 17.4.$$

If the density is fixed to $\lambda = 1$, we have $r_c < \sqrt{8 \log 2} \approx 2.35$.

So we have a lower and an upper bound on the critical mean degree. Next we try to improve the upper bound.

11.2.3 Bounding using face percolation on the hexagonal lattice

Here we exploit a known result for face percolation on the hexagonal lattice. This percolation model is equivalent to site percolation on the triangular lattice, as illustrated in Fig. 11.3, so the critical threshold for face percolation on the hexagonal lattice is also 1/2. To connect the model to the disk graph, we declare a face to be open if it contains one or more points of the PPP, as shown in Fig. 11.4.

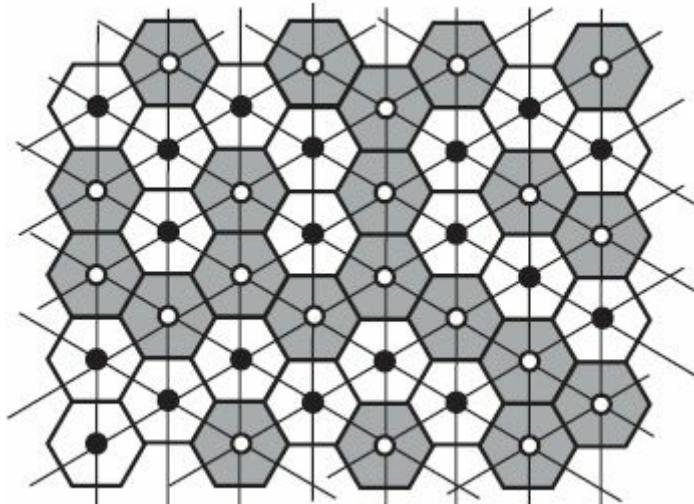


Figure 11.3 Illustration of face percolation in the hexagonal lattice. Two open faces (or Voronoi cells) are connected if they touch on one side, and the model is percolating if there is an infinite connected component, as usual. Also shown are the centers of each face, which form a triangular lattice. If an open face corresponds to an open site in the triangular lattice, there is a one-to-one mapping between the two percolation models.

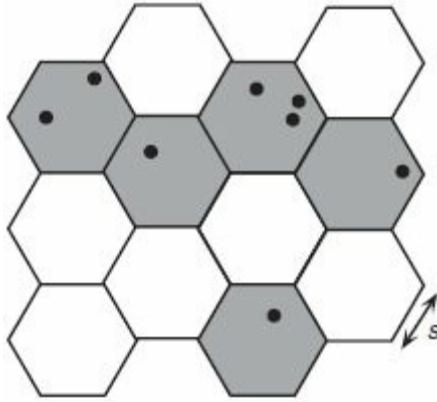


Figure 11.4 Faces containing one or more points are assumed open. The side length of each face is s .

Letting s denote the side length of the hexagonal cells, the probability of a face or cell being open is $1 - \exp(-\lambda A)$, where $A = \frac{3\sqrt{3}s^2}{2}$ is the cell area. So if $1 - \exp(-\lambda A) < 1/2$ we do not have face percolation, and if $\exp(-\lambda A) < 1/2$ we do. Now we need to connect this model to the disk graph. Any two points in neighboring faces are at distances of at most

$$s\sqrt{(2\sqrt{3})^2 + 1^2} = s\sqrt{13},$$

and any two points in non-neighboring faces are at distances of at least s . Consequently, if $1 - \exp(-\lambda A) < 1/2$ and $r < s$, we do not have percolation in $G_{r,\lambda}$, and if $1 - \exp(-\lambda A) > 1/2$ and $r > s\sqrt{13}$, then we do. Taking $\lambda = 1$, we find that if $a = \pi r^2 > 13\pi s^2$ and

$$A = \frac{3\sqrt{3}}{2}s^2 > \log 2,$$

the graph percolates. So percolation occurs if

$$a > 13 \frac{2\pi \log 2}{3\sqrt{3}} \approx 10.9.$$

On the other hand, if $a = \pi r^2 < \pi s^2$ and

$$A = \frac{3\sqrt{3}}{2}s^2 < \log 2$$

there is no percolation. So percolation does not occur if

$$a < \frac{2\pi \log 2}{3\sqrt{3}} \approx 0.84.$$

Hence we have established the bounds

$$\frac{2\pi \log 2}{3\sqrt{3}} \leq a_c \leq \frac{26\pi \log 2}{3\sqrt{3}} \approx 10.9.$$

The upper bound is quite close to the best one known, but the lower bound of 0.84 is worse than the bound 1 obtained from the branching argument.

11.2.4 A refined branching technique

We set out to improve the lower bound using a more refined branching technique than the simple one used in [Section 11.2.1](#). The idea is to try to take into account that two points that are connected are likely to have common neighbors. This is not considered in the simple branching argument used previously and leads to an improved bound.

THEOREM 11.3 (Lower bound for percolation on the disk graph)

$$a_c \geq \frac{6\pi}{2\pi + 3\sqrt{3}} \approx 1.642.$$

Proof Consider the graph $G_r = G_{r,1}$. Let C_o be the vertex set of the component of the origin. We shall construct a sequence of pairs of disjoint finite sets of points of the plane, $(D_0, L_0), (D_1, L_1), \dots$. The points in D_t are the points of C_o that are called *dead* at time t : They belong to C_o and so do all their neighbors. The points in L_t are *live* at time t : They belong to C_o but we have made no attempt yet to find their neighbors. To start the sequence, we set $D_0 \triangleq \emptyset$ and $L_0 \triangleq \{X_0\} \triangleq \{o\}$. Next, let N_0 be the set of neighbors of X_0 and we set $D_1 = \{X_0\}$ and $L_1 = N_0$.

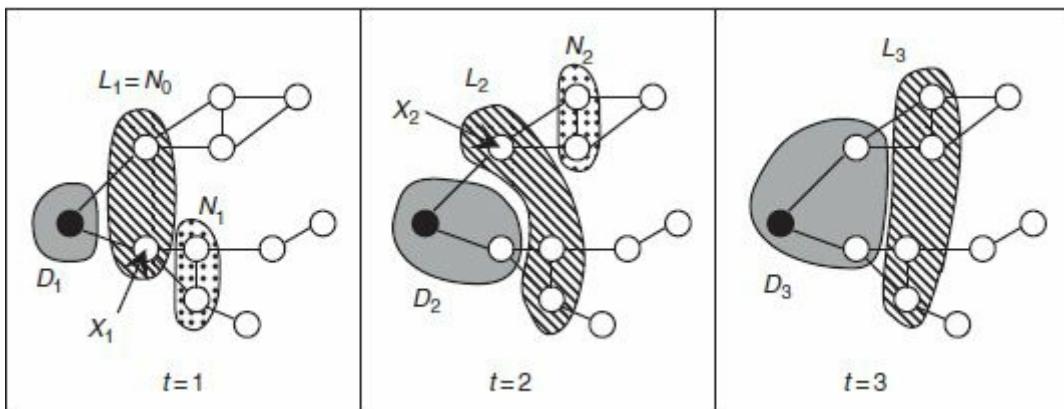


Figure 11.5 Illustration for the proof of Theorem 11.3.

Having found (D_t, L_t) , if $L_t = \emptyset$, we terminate the sequence; otherwise we pick a point X_t from L_t and define

$$D_{t+1} \triangleq D_t \cup \{X_t\} = \{X_0, X_1, \dots, X_t\} \text{ and } L_{t+1} \triangleq N_t \cup L_t \setminus \{X_t\},$$

where N_t is the set of neighbors of X_t that are not neighbors of any of the points in D_t . The process is illustrated in Fig. 11.5. The sets D_t and L_t are disjoint finite sets. By construction, $D_t \cup L_t \subset C_o$, and, if $L_k = \emptyset$, then $D_k = C_o$. Since

$$D_t \setminus \{X_0\} = \{X_1, X_2, \dots, X_{t-1}\} \subset \bigcup_{i=0}^{t-2} N_i,$$

we have

$$|D_t| - 1 = t - 1 \leq \sum_{i=0}^{t-2} |N_i|. \quad (11.1)$$

Let V_t be the disk of radius r with center X_t , and set $U_t = \bigcup_{s=0}^t V_s$. Conditioning on the points X_0, \dots, X_t , we find that $|N_t|$ is Poisson with mean $|V_t \setminus U_{t-1}|$. Since the center X_t of V_t is in a disk V_s , $s \leq t-1$, Fig. 11.6 tells us that

$$|V_t \setminus U_{t-1}| \leq \left(\frac{\pi}{3} + \frac{\sqrt{3}}{2} \right) r^2.$$

We denote the right side by b for short. To bound $|C_o|$, let Z_0, Z_1, \dots be independent Poisson random variables with $\mathbb{E}Z_0 = \pi r^2$ and $\mathbb{E}Z_i = b$ for $i \geq 1$. Then inequality (11.1) implies that

$$\mathbb{P}(|C_o| \geq k) \leq \mathbb{P} \left(\sum_{i=0}^{k-2} Z_i \geq k-1 \right).$$

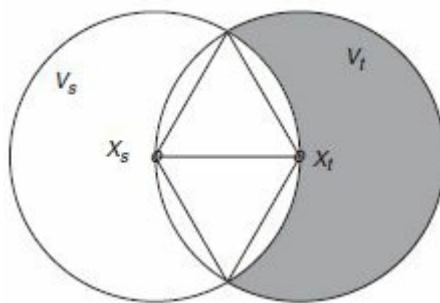


Figure 11.6 A moon-shaped region whose area $|V_t \setminus V_s|$ upper bounds the area in which can lie those neighbors of a vertex X_t that are not also neighbors of X_s . X_s and X_t are assumed connected.

If $b < 1$ it follows that $\mathbb{P}(|C_o| \geq k) \rightarrow 0$ as $k \rightarrow \infty$, so there is no percolation in this case. The lower bound follows since

$$b = \left(\frac{\pi}{3} + \frac{\sqrt{3}}{2} \right) r^2 < 1 \iff (2\pi + 3\sqrt{3})a < 6\pi.$$

So $a_c > 1.642$. □

The refined technique is based on the same approach as the simple one in [Section 11.2.1](#). The difference is merely that the mean number of children in the branching process is bounded more tightly, by removing common neighbors of a parent and child node.

A more elaborate multitype branching argument yields $a_c > 2.19$. The true value is $4.508 < a_c < 4.515$ with confidence 99.99%. For this value, about 1% of the nodes are isolated.

11.2.5 Uniqueness of the infinite component

As for the lattice model, the giant component is unique if it exists.

THEOREM 11.4 (Uniqueness of the infinite component in the disk graph) *In the disk model, there is at most one unbounded component a.s.*

The proof is quite involved. It is essentially a modification of the trifurcation argument used in the proof of [Theorem 10.22](#). It starts by noting that the disk graph is ergodic and that the number of unbounded components in an ergodic model is constant a.s. Then, as in bond percolation on the lattice, it can be shown that this constant can only be 0, 1, or ∞ . The last step is to rule out the possibility of an infinite number of infinite components.

11.3 Other percolation models

11.3.1 General connectivity functions

Here we consider RGGs with vertices forming a uniform PPP of intensity λ with a vertex added at o and a general connectivity function ξ . The effective area $A\xi = \int_{\mathbb{R}^d} \xi(y) dy$, and the mean degree is $\alpha = \lambda A\xi$. It is assumed that ξ is chosen such that $0 < A\xi < \infty$, otherwise the model is trivial.

As in the case of a fixed disk radius, we expect the existence of a critical density

$$\lambda_c = \inf\{\lambda: \theta(o) > 0\}.$$

If $\lambda > \lambda_c$, the model *percolates*, i.e., an infinite component exists a.s. As before, we may also speak of a critical mean degree $a_c = \lambda_c A_\xi$.

An interesting question is whether the critical intensity depends only on a or also on the shape of ξ . For example, if

$$\xi(y) = \frac{1}{2} \exp\left(-\frac{\|y\|}{r}\right),$$

then $A_\xi = \pi r^2$ as for the standard disk graph. But the variance of the node degree is now $\pi r^2/8$, as per Campbell's formula for the variance in a PPP (see Corollary 4.8).

To study the effect of a change in ξ while keeping A_ξ constant, consider the family of connection functions, derived from a parent function $\xi = \xi_1$, given by

$$\xi_p(y) = p\xi(\sqrt{p}y), \quad 0 < p \leq 1.$$

They all have the same A_ξ , but the critical intensity may decrease with decreasing p .

THEOREM 11.5 *For $0 < p < 1$,*

$$\lambda_c(\xi) \geq \lambda_c(\xi_p).$$

Proof Consider a third connection function that dominates ξ_p in the sense that its mean degree a is larger: $\hat{\xi}_p(y) = \xi(\sqrt{p}y)$. We can obtain a realization of an RGG G_{λ, ξ_p} by starting with a realization of $G_{\lambda, \hat{\xi}_p}$ and removing each edge with probability $1 - p$, irrespective of its length. This is a bond percolation model. We can also obtain a realization of the graph $G_{p\lambda, \hat{\xi}_p}$ by starting with $G_{\lambda, \hat{\xi}_p}$ and eliminating each vertex independently with probability p . This way, we obtain a *site percolation model*, which is, after re-scaling all distances with \sqrt{p} , a realization of G_{λ, ξ_p} .

From Theorem 10.27 we know that $p_c^s \geq p_c^b$, so bond percolation occurs if site percolation occurs. Applied to the above graphs, this means that if $G_{\lambda, \xi}$ percolates (this is the site percolation model), then G_{λ, ξ_p} percolates also. \square

This implies that unreliable connections are at least as valuable as reliable ones if the loss in reliability is compensated for by increased edge length (to keep a constant).

In view of the comparison with branching processes, there is another intuitive

explanation for this behavior. What matters is the overlap between a point's neighbors and its neighbors' neighbors. In the disk model, this overlap is significant, as a brief calculation shows. Take the intersection of two disks of the same radius r , if their distance is h , and h is distributed as $f_h(v) = 2/v^2 \mathbf{1}(v < r)$. If instead the “spread out” connectivity function $\xi_p(y) = p\mathbf{1}(\sqrt{p}\|y\| < r)$ is used, the overlap (as a fraction of the total area of the disk) is smaller by a factor of p . From this point of view, we would expect the critical mean degree to decrease to 1 as $p \rightarrow 0$. This is indeed the case, as shown in Franceschetti *et al.* (2005).

Another way to transform the connectivity function is to shift (and compress) it. Assuming that it is isotropic, i.e., $\xi(y) \equiv \xi(y')$ if $\|y\| = \|y'\|$, let $\xi_v(y) = \xi(b(\|y\| - v))\mathbf{1}(\|y\| \geq v)$, where b is chosen such that a stays constant. In this way, the hard-disk connectivity function $\xi(y) = \mathbf{1}(\|y\| \leq r)$ gets transformed into the annulus $\xi_v(y) = \mathbf{1}(v \leq \|y\| \leq r/b + v)$ with inner radius v and outer radius $r/b + v = \sqrt{v^2 + r^2}$. As v increases the annulus gets thinner, and the overlap of a parent's annulus and a child node's annulus becomes smaller and smaller. It is plausible that, as $v \rightarrow \infty$, the process of growth from the origin approaches that of an independent branching process.

11.3.2 Abstract random graphs

Here we put the results for RGGs in relationship to the results for abstract random graphs, where vertices are not embedded in a metric space. The most prominent examples are the *Gilbert model* and the *Erdős-Rényi* model.

DEFINITION 11.6 (Gilbert random graph) Given $n \in \mathbb{N}$ and $0 \leq p \leq 1$, a Gilbert random graph $G(n, p)$ is a graph with vertex set $\{1, \dots, n\}$ where each pair of vertices has an edge independently with probability p . The elements of the associated probability space are the $2^{\binom{n}{2}}$ graphs of n vertices, and the probability of a graph with m edges occurring is

$$\mathbb{P}(\{G(n, p) \text{ has } m \text{ edges}\}) = p^m(1-p)^{\binom{n}{2}-m}.$$

Owing to the binomial distribution, this model is also referred to as a *binomial random graph*.

DEFINTION 11.7 (Erdős-Rényi random graph) Given $n \in \mathbb{N}$ and $0 \leq M \leq \binom{n}{2}$, the Erdős-Rényi graph $G(n, M)$ is chosen uniformly at random from all graphs with n vertices and M edges. The probability of each such graph is thus

$$\mathbb{P}(G) = \left(\frac{\binom{n}{2}}{M}\right)^{-1}.$$

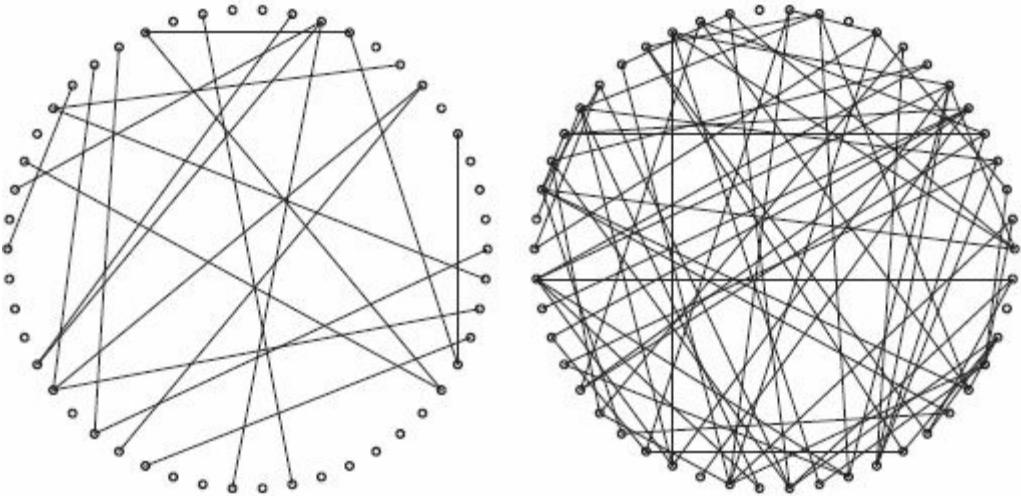


Figure 11.7 Gilbert random graphs with $n = 50$ vertices. (Left) $p = 1/50$. (Right) $p = 3/50$. This graph is almost fully connected.

Since this is a uniform distribution, this model is also referred to as a *uniform random graph*.

The two models are closely related in the sense that if $M \approx p(\frac{n}{2})$, they behave the same as $n \rightarrow \infty$. We focus on the Gilbert model. Realizations of $G(50,1/50)$ and $G(50,3/50)$ are shown in Fig. 11.7.

The node degree is binomial with parameters $n - 1$ and p . For $p = 1/2$, all possible graphs with n vertices have the same probability. For constant $p > 0$, $G(n,p)$ is connected asymptotically almost surely (a.a.s.), i.e.,

$$\lim_{n \rightarrow \infty} \mathbb{P} (\{G(n,p) \text{ connected}\}) = 1,$$

since the mean node degree $p(n - 1)$ increases linearly with n . The regime where the mean node degree stays almost constant is the most interesting. Let us consider, for example, the occurrence of triangles. The expected number of triangles is $p^3 \binom{n}{3}$. If $p = n^{-1.01}$, the expected number of triangles goes to zero. If $p = n^{-0.99}$, the expected number of triangles goes to infinity. So $p = 1/n$ is a *threshold function* for the appearance of triangles. The same argument can be repeated for cycles of any fixed size k : Their expected number is $p^k \binom{n}{k}$. Since $\binom{n}{k} = \Theta(n^k)$, $p = n^{-1}$ is the threshold for cycles of any size.

Let $\bar{m} = p(\frac{n}{2})$ denote the mean number of edges. Other threshold functions are as follows. For $p = o(n^{-2})$, $\bar{m} \rightarrow 0$, so there will be no edges asymptotically. At $p = n^{-2}$, edges start to appear. At $p = n^{-3/2}$, edges with a common vertex (trees with three vertices) appear. At $p = n^{-1-1/k}$, trees with $k + 1$ vertices appear (see Problem 11.2).

At $p = n^{-1}$, not only cycles of any size appear, the graph also forms a component

larger than $\Theta(\log n)$. To get some insight why this happens, we explore the size of the component containing vertex 1, denoted by $|C(1)|$. We grow the set $\{1\}$ by adding all neighbors of vertex 1, then adding the neighbors' neighbors etc., as before in the case of the disk graph. In the first step, the distribution of neighbors is binomial with parameters $n - 1$ and p . In the next step, fewer nodes are available as neighbors. We aim at bounding this growth process by comparing it with a branching process. To this end, let L_i denote the number of active or live vertices at time i . These are the vertices that have not been explored yet in the process. We have the following recursion:

$$\begin{aligned} L_0 &= 1, \\ L_i &= L_{i-1} + X_i - 1, \quad i \in \mathbb{N}, \end{aligned}$$

since, conditioned on L_{i-1} , X_i is binomially distributed with parameters $n - (i - 1) - L_{i-1}$ and p . Here $n - (i - 1) - L_{i-1}$ is the number of potential neighbors that are left at stage i . In each step, we pick a vertex from the live ones. The process terminates once $L_i = 0$, i.e.,

$$|C(1)| = \min\{i \in \mathbb{N} : L_i = 0\}.$$

Hence we have the following upper bound.

LEMMA 11.8 (Upper bound on component size) *Let T be the total number of progeny in a Galton–Watson branching process with a binomial offspring distribution with parameters n and p , and let $|C(1)|$ be the size of the component at vertex 1 in a random graph $G_{n,p}$. Then*

$$\mathbb{P}(|C(1)| \geq k) \leq \mathbb{P}_{n,p}(T \geq k).$$

In other words, $|C(1)|$ is stochastically dominated by T .

Proof The number of neighbors of each live vertex in the graph is stochastically dominated by the number of children in the binomial branching process.

Conversely, we have the following lower bound.

LEMMA 11.9 (Lower bound on component size) *Let T be the total number of progeny in a Galton–Watson branching process with a binomial offspring distribution with parameters $n - k$ and p . For any $k \in \mathbb{N}$,*

$$\mathbb{P}(|C(1)| \geq k) \geq \mathbb{P}_{n-k,p}(T \geq k).$$

Proof At any stage, there are at least $n - k$ potential neighbors for a live node, until the component has grown to size k . But then we already know that $|C(1)| \geq k$ occurred, so it is irrelevant that there are fewer vertices left afterwards. \square

Since the distribution on the right side depends on k , this does *not* imply stochastic dominance of T by $|C(1)|$.

So the distribution of the component behaves similarly to the progeny in the binomial branching process. In the limit as $n \rightarrow \infty$, the difference between the binomial distributions with parameters n and $n - k$ (and p) vanishes, so the upper and lower bounds become tight. Since the binomial branching process changes its behavior critically depending on the mean number of offspring np , we can expect the same in the random graph.

This is indeed the case, and the following rather sharp results can be obtained for the regime $p = \Theta(n^{-1})$.

THEOREM 11.10 (Small components for $pn < 1$) *If $c = pn < 1$, then the largest component of $G(n, p)$ has at most*

$$\frac{3}{(1-c)^2} \log n$$

vertices a.a.s.

The proof uses a similar procedure to what we used to establish the improved lower bound for the critical mean node degree for percolation on the disk graph.

THEOREM 11.11 (Unique giant component for $pn > 1$) *If $c = pn > 1$, then the largest component of $G(n, p)$ has $\Theta(n)$ vertices, and the second-largest component has at most*

$$\frac{16c}{(c-1)^2} \log n$$

vertices a.a.s.

Uniqueness is shown by establishing that the probability that any two vertices in giant components lie in different giant components goes to zero.

THEOREM 11.12 (Largest component at $pn = 1$) *If $c = pn = 1$, then the largest components of $G(n, p)$ have $\Theta(n^{2/3})$ vertices.*

For connectivity, $p(n)$ needs to decay a bit more slowly than with n^{-1} . We will explore this in the next chapter.

11.4 Applications

11.4.1 Spatiotemporal graph models

Gilbert's disk graph is a rather crude model of a wireless network; it models only a static power constraint and completely ignores fading, interference, half-duplex constraints, and changing sets of concurrent transmitters. As a result, the node isolation and connectivity defined on a static graph does not accurately reflect the connectivity properties of an actual wireless network. For example, due to half-duplex constraints, the network is, in each time slot, split into many small directed networks with star topology, with a transmitter in the center.¹ To get an accurate and meaningful account on the connectivity properties, a graph model needs to capture the directivity of the information flow as well as the time evolution.

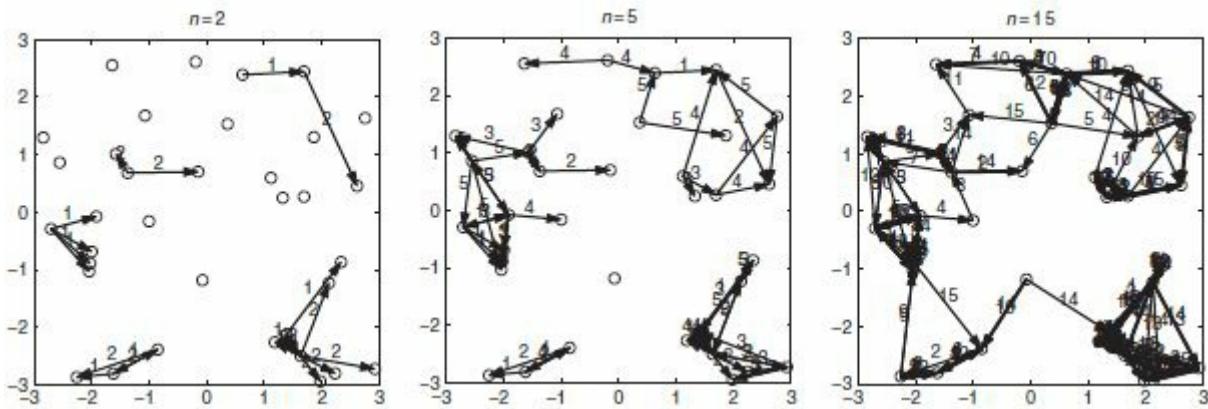


Figure 11.8 Examples of space–time SINR graphs \mathcal{G}^n_1 for $n = 2, 5$, and 15 on the same realization of Φ . The graph \mathcal{G}^{15}_1 is fairly well connected, although not all pairs of nodes are connected by a causal path. The cluster in the bottom right, for example, cannot reach a node outside the cluster. On the other hand, the cluster can be reached from the rest of the network, via the edge with weight 14.

The most complete random geometric graph for wireless networks is the *space–time SINR graph*. It is a weighted and directed multigraph defined as follows.

DEFINITION 11.13 (Space–time SINR multigraph) Let Φ be a point process that is partitioned into a transmitter process $\Phi_t(k)$ and a receiver process $\Phi_r(k)$ in each time slot $k \in \mathbb{Z}$. For $y \in \mathbb{R}^2$, let

$$I_y^{ls}(k) = \sum_{x \in \Phi_t(k) \setminus \{z\}} h_{xz}(k) \ell(x - z)$$

be the interference at location y excluding transmitter z and

$$\mathbf{1}_k(z \rightarrow y) = \mathbf{1} \left(\frac{h_{zy}(k)\ell(z-y)}{W + I_y^{!z}(k)} \geq \theta \right) \mathbf{1}(z \in \Phi_t(k)) \mathbf{1}(y \in \Phi_r(k)). \quad (11.2)$$

W is the thermal noise power, the fading process (h_{xy}) is iid across space and time, and the path loss function is integrable over $\mathbb{R}^2 \setminus b(o, \infty)$. So $\mathbf{1}_k(z \rightarrow y) = 1$ if and only if a transmission from $z \rightarrow y$ is successful in time slot k under the SINR model. Define a *snapshot graph* $G_k(\Phi, \vec{E}_k)$ by the directed edge set

$$\vec{E} = \{(z, y) \in \Phi^2 : \mathbf{1}_k(z \rightarrow y)\}$$

and add the weight k to each edge. Each snapshot graph is a weighted bipartite graph with directed edges from Φ_t to Φ_r .

The *space-time SINR graph* \mathcal{G}_m^n is the edge-union of all snapshot graphs from time m to n :

$$\mathcal{G}_m^n \left(\Phi, \bigcup_{k=m}^n \vec{E}_k \right).$$

Three space-time SINR multigraphs, \mathcal{G}_1^2 , \mathcal{G}_1^5 , and \mathcal{G}_1^{15} are shown in Fig. 11.8.

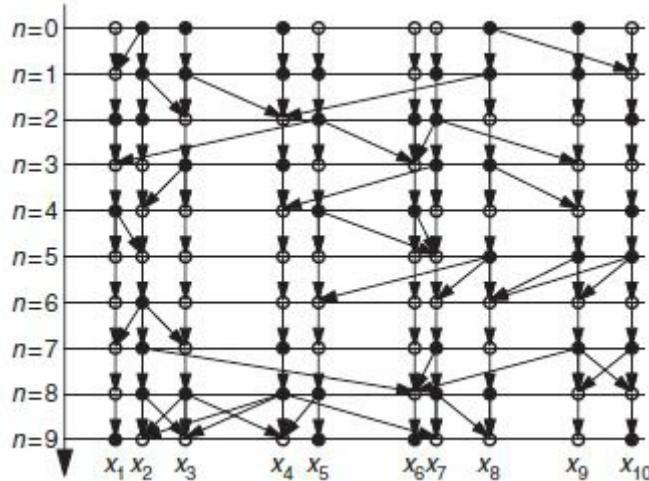


Figure 11.9 A realization of a space-time SINR graph $\mathbb{G}(\Phi \times \mathbb{Z}, \vec{E})$, where $\Phi \subset \mathbb{R}$. Transmitters are indicated by a \bullet , whereas \circ are receivers. The time axis is vertical. The vertical arrows model the node's memories, and the diagonal ones represent the successful transmissions.

The space-time SINR multigraph captures all the information on how messages may spread through the network; they can spread only along *causal paths* in the graph.

DEFINITION 11.14 (Causal path) A causal path is a directed path in \mathcal{G}_m^n with

strictly increasing edge weights.

Hence causality in this context means that, for a message to propagate over a number of hops, the first link needs to exist before the second, before the third, and so on.

An equivalent graph may be defined without multiple edges and weights if the vertex set is augmented by a time axis \mathbb{Z} .

DEFINITION 11.15 (Space–time SINR graph) The *Space–time SINR graph*, denoted as $\mathbb{G}(\Phi \times \mathbb{Z}, \bar{\mathbb{E}})$, has a directed edge from vertex (x, n) to vertex $(y, n+1)$ if $\mathbf{1}_n(x \rightarrow y) = 1$. It also has directed edges from (x, n) to $(x, n+1)$ for all $(x, n) \in \Phi \times \mathbb{Z}$, to account for the memory at node x .

In this model, messages propagate along directed edges. By definition, there are no loops since time always runs forward, and each path is causal. This graph is best visualized for a one-dimensional point process. An example is shown in Fig. 11.9.

Both types of graph capture the complete information about all transmissions and their successes in the network. Routing algorithms build paths that are subsets of the causal paths in the SINR graphs. So these graphs permit the formulation and (at least partial) solution of fundamental problems, irrespective of what routing algorithm may be used, including finding answers to the following questions.

- Can a node communicate with a neighbor in a finite amount of time on average?
- Denoting by $\Delta_{x,y}$ the time it takes for the first (causal) path to emerge from x to y , is the *asymptotic propagation speed*

$$v \triangleq \lim_{\|x-y\| \rightarrow \infty} \frac{\|x-y\|}{\mathbb{E}_{x,y} \Delta_{x,y}}$$

zero or non-zero? Here $\mathbb{E}_{x,y}$ is the expectation with respect to the two-fold Palm distribution of the point process, conditioned on having points at x and y .

The second problem is a *first-passage percolation* problem, where, in general, the question is that of how long it takes for a percolation process to reach a certain site. The typical model is to assign random passage times to the edges in a bond percolation model. In our case, the role of the passage times are assumed by the weights in \mathbb{G} and the time component in \mathbb{G} , respectively.

To address these questions, we introduce the notion of the *local delay*.

DEFINITION 11.16 (Local delay) For $x, y \in \Phi$ and $n \in \mathbb{Z}$, the *local delay* is defined as

$$D_{x,y}(n) \triangleq \min\{k > 0 : \mathbf{1}_{n+k}(x \rightarrow y)\}.$$

$D_{xy}(n)$ is at least 1, since it takes at least one time slot to communicate. The delay is measured from the end of the n th time slot until the end of the time slot during which the successful transmission took place.

LEMMA 11.17 (Finite conditional local delay) *Let Φ be a general point process, and let the fading distribution have infinite support. If the channel access scheme is ALOHA with probability $0 < p < 1$, then, given Φ , all local delays are a.s. finite geometric random variables.*

Proof Conditioned on Φ , all randomness is independent from one time slot to the next. So the variables $\mathbf{1}_k(x \rightarrow y)$ are iid Bernoulli and thus $D_{x,y}(n)$ is geometric. It remains to show that $\mathbb{E}\mathbf{1}_k(x \rightarrow y | \Phi) > 0$ for almost all Φ . We have

$$\mathbb{E}\mathbf{1}_k(x \rightarrow y | \Phi) = p(1-p)\mathbb{P}(h > \ell(\|x - y\|)(W + I_y^{\text{lx}})).$$

Since $\mathbb{P}(h > u) > 0$ for all $u > 0$ (due to the infinite support) and the interference I_y^{lx} is finite a.s., this probability is strictly positive.

It follows, under the assumption in the lemma, that there exist paths from x to y for all $x, y \in \Phi$. So the Space–time SINR graphs are connected in this sense, via causal paths from x to y in \mathcal{G}_1^∞ , or via paths from (x, n) to the set $\{m \in \mathbb{N} : (y, n+m)\}$ in \mathbb{G} .

To bound the delay of communication between two arbitrary nodes, let

$$D_x(n) \triangleq \min\{y \in \Phi \setminus \{x\} : D_{x,y}(n)\}$$

be the *exit delay* of node x at time n . This is the time it takes for node x to talk to *any* other node. Formally defining the *path formation time* $\Delta_{x,y}(n)$ as

$$\Delta_{x,y}(n) = \min\{k > 0 : \mathcal{G}_n^{n+k} \text{ has a causal path from } x \text{ to } y\},$$

we have

$$D_x(n) \leq \Delta_{x,y}(n) \leq D_{x,y}(n) \tag{11.3}$$

since reaching a fixed node y takes at least as long as reaching any node, and reaching a fixed node in a direct transmission takes at least as long as reaching it via an arbitrary path. It follows that all three random variables are finite a.s.

The next step towards answering our questions involves averaging over the point

process. We focus on the typical point in a PPP.

THEOREM 11.18 (Mean exit delay) *Consider a Poisson network with exponential fading, ALOHA, and power-law path loss. If $W > 0$, the mean exit delay $\bar{D} \triangleq \mathbb{E}_o D_o(0)$ is infinite irrespective of the transmit probability p . If $W = 0$, there exists a critical transmit probability below which the mean exit delay is finite.*

Proof (Sketch) $W > 0$: Here it is enough to consider noise only. Let $T = T_o(0)$ be the number of transmission attempts without interference. Clearly $T_o(0) \leq D_o(0)$. We have

$$\begin{aligned}\mathbb{P}_o(T > k \mid \Phi) &= \mathbb{P}_o\{\forall 0 \leq m \leq k, x \in \Phi_r(m) : h_{ox}(m) < \|x\|^\alpha \theta W\} \\ &= \prod_{x \in \Phi \setminus \{o\}} (p + (1-p)\mathbb{P}(h < \|x\|^\alpha \theta W))^k \\ &= \exp\left(k \sum_{x \in \Phi \setminus \{o\}} \log(p + (1-p)(1 - e^{-\|x\|^\alpha \theta W}))\right),\end{aligned}$$

where h denotes a generic (exponentially distributed) fading random variable. Next follows the averaging with respect to Φ , which is an application of the pgfl. The resulting expression can be bounded by $1/k$, thus $\mathbb{P}^o(T > k) \geq 1/k$, which proves the result.

For $W = 0$, we show that nearest-neighbor communication is possible in finite mean time if p is small enough. For a transmission over distance r in an interference field induced by the transmitters in Φ , the probability of success is (see (5.14)) $p_s = \mathbb{E}(e^{-\theta r^\alpha I})$, which is the Laplace transform of the interference at $s = \theta r^\alpha$. If the interference were independent from time slot to time slot, the local delay for this transmission would simply be geometric with mean p_s^{-1} . But the interference is only *conditionally independent* given Φ (cf. Lemma 11.17), so we need to calculate

$$\mathbb{E}_o(p_s^{-1}) = \mathbb{E}_o\left(\frac{1}{\mathbb{E}(e^{-\theta r^\alpha I} \mid \Phi)}\right),$$

where the expectation \mathbb{E}_o is over Φ . Letting

$$I \triangleq \sum_{x \in \Phi_t} h_{xo} \|x\|^{-\alpha}$$

and denoting by $\mathcal{L}_I(s \mid \Phi) = \mathbb{E}(e^{-sI} \mid \Phi)$ its conditional Laplace transform given Φ , we have

$$\begin{aligned}\mathbb{E}_o \left(\frac{1}{\mathcal{L}_I(s \mid \Phi)} \right) &= \exp \left(\lambda \int_{\mathbb{R}^2} \frac{ps}{s(1-p) + \|x\|^\alpha} dx \right) \\ &= \exp \left(\frac{p\lambda\pi^2\delta s^\delta}{(1-p)^{1-\delta} \sin(\pi\delta)} \right).\end{aligned}\tag{11.4}$$

So the mean local delay for a transmission over distance r is

$$\bar{D}(r) = \frac{1}{p(1-p)} \exp \left(\frac{p\lambda\pi^2\delta\theta^\delta r^2}{(1-p)^{1-\delta} \sin(\pi\delta)} \right) = \frac{1}{p(1-p)} \exp \left(\frac{p\lambda\gamma r^2}{(1-p)^{1-\delta}} \right)$$

with $\gamma = \pi\theta^\delta\pi\delta/\sin(\pi\delta)$. The factor $(p(1-p))^{-1}$ accounts for the fact that the source node under consideration needs to transmit, while the destination node needs to listen. Next we de-condition on the distance between the two nodes, assuming they are nearest neighbors (see Example 2.11). This yields

$$\begin{aligned}\bar{D}_{NN} &= \frac{1}{p(1-p)} 2\pi\lambda \int_0^\infty \exp \left(\frac{\lambda p\gamma r^2}{(1-p)^{1-\delta}} \right) r \exp(-\pi\lambda r^2) dr \\ &= \frac{1}{p(1-p)} \frac{\pi(1-p)^{1-\delta}}{\pi(1-p)^{1-\delta} - p\gamma}, \quad 0 < p < \frac{\pi(1-p)^{1-\delta}}{\gamma}.\end{aligned}$$

So the typical node can indeed transmit to its nearest neighbor in finite time on average, provided that p is not too large. Strictly speaking, the condition $p < (\pi(1-p)^{1-\delta})/\gamma$ is only sufficient, not necessary, since we have not considered that, if the node at x is o 's nearest neighbor, then necessarily there is no node in $b(o, \|x\|)$, which reduced the interference. So the actual range of p for finite \bar{D}_{NN} is slightly \square larger.

Lemma 11.17, together with (11.3), shows that, even with $W > 0$, given $x, y \in \Phi$, the mean local delay from x to y is finite, i.e.,

$$\mathbb{E}_{x, y} \Delta_{x, y}(n) < \infty \quad \forall n \in \mathbb{Z}.$$

The expectation with respect to the two-fold Palm distribution is easy to evaluate in the case of the PPP, since, by Slivnyak's theorem, for any property $Y \in \mathcal{N}$,

$$\mathbb{P}_{x, y}(\Phi \in Y) = \mathbb{P}(\Phi \cup \{x, y\} \in Y).$$

So the infinite mean local delay is due to the fact that a node is surrounded by an arbitrarily large void with positive probability in the PPP. With two nodes at fixed locations, this cannot happen, i.e., under the two-fold Palm probability $\mathbb{P}_{x, y}$ the nearest-neighbor distance of x is bounded by $\|x - y\|$.

However, even with noise only, using a bounding technique similar to the one in the proof for the noisy case, it can be shown that the mean local delay grows superlinearly in the distance $\|x - y\|$, so we have

$$v \triangleq \lim_{\|x-y\| \rightarrow \infty} \frac{\|x-y\|}{\mathbb{E}_{x,y} \Delta_{x,y}} = 0$$

in the noisy case.

Space-time disk-interference graph

In the disk-interference graph, the interference condition (11.2) is replaced by the condition that there is no interferer inside a disk of radius $\beta\|z-y\|$ around the receiver at y and the distance $\|z-y\| < \eta$ (noise condition), i.e.,

$$\begin{aligned} 1_k(z \rightarrow y) &= 1(b(y, \beta\|z-y\|) \cap (\Phi_t(k) \setminus \{z\}) = \emptyset) \cdot 1(\|z-y\| < \eta) \\ &= 1(\Phi_t(k) \cap b(y, \beta\|z-y\|) \subseteq \{z\}) \cdot 1(\|z-y\| < \eta). \end{aligned}$$

Using this interference condition, snapshot graphs G_k and Space-time disk-interference multigraphs \mathcal{G}_m^n can be defined as before. These graphs are easier to work with since the interference condition is localized.

Focusing again on the PPP, we can calculate the mean out-degree of the snapshot graph in closed form.

LEMMA 11.19 (Average out-degree of transmitter in disk-interference snapshot graph) *The average out-degree of the typical transmitting node is*

$$\mathbb{E}N_t = \frac{1-p}{p\beta^2} (1 - \exp(-\lambda p \pi \beta^2 \eta^2)).$$

Proof Consider the typical transmitter at the origin o . Dropping the time index since the mean degree does not depend on k , we have

$$\begin{aligned} \mathbb{E}N_t &= \mathbb{E} \left(\sum_{x \in \Phi_r} \mathbf{1}(o \rightarrow x) \right) \\ &\stackrel{(a)}{=} \lambda(1-p) \int_{\mathbb{R}^2} \mathbb{E}_{\Phi_t} \mathbf{1}(o \rightarrow x) dx \\ &\stackrel{(b)}{=} \lambda(1-p) \int_{b(o,\eta)} \exp(-\lambda p \pi \beta^2 \|x\|^2) dx \\ &= \frac{1-p}{p\beta^2} (1 - \exp(-\lambda p \pi \beta^2 \eta^2)), \end{aligned}$$

where (a) follows from Campbell's theorem for sums and the independence of Φ_t and Φ_r , and (b) follows since $\mathbf{1}(o \rightarrow x) = 1$ if and only if the ball $b(o, \beta\|x\|)$ is free of interferers and x is inside the ball $b(o, \eta)$. \square

Since the process of successful receivers is not an independent thinning of Φ_r , the out-degree is *not* Poisson. In the interference-limited (or noise-free) case ($\eta \rightarrow \infty$), $\mathbb{E}N_t = (1-p)/(p\beta^2)$, and in the noise-limited case ($\beta \rightarrow 0$), $\mathbb{E}N_t = \lambda(1-p)\pi\eta^2$.

For the exit delay, it is easy to see that $\mathbb{E}_o(D_o(0)) = \infty$, since the probability that there is no node in $b(o, \eta)$ is $\exp(-\lambda\pi\eta) > 0$. In the noise-free case, nearest-neighbor communication is possible if p is small enough, as in the SINR case.

In the general case, we can use subadditivity of the path formation random variables to establish that the asymptotic propagation speed exists. First we augment the PPP Φ by $\{o, x, y, y + x\}$ for arbitrary $x, y \in \mathbb{R}^2$ to avoid cumbersome notation of the Palm distributions. For all such x, y , we have the triangle-type inequality

$$\Delta_{o, y}(n) \leq \Delta_{o, x}(n) + \Delta_{x, y}(\Delta_{o, x}(n))$$

and thus

$$\Delta_{o, x+y}(n) \leq \Delta_{o, y}(n) + \Delta_{y, y+x}(\Delta_{o, y}(n)).$$

Since $\Delta_{x,y}(n) \stackrel{\text{d}}{=} \Delta_{x, y}(0)$ and, by stationarity $\Delta_{y, y+x}(n) \stackrel{\text{d}}{=} \Delta_{o, x}$, we obtain

$$\mathbb{E}\Delta_{o, x+y} \leq \mathbb{E}\Delta_{o, x} + \mathbb{E}\Delta_{o, y}.$$

It then follows from the theory of subadditive sequences that the limit $\|x\|/\mathbb{E}\Delta_{o, x}$ exists.

With noise (finite η), $v = 0$ as stated before. In the noise-free case, however, it can be shown that $v > 0$, so the path formation time scales linearly with the distance. Moreover, with noise present but η large enough that the disk graph on Φ with radius η percolates, two nodes in the giant component can communicate with positive message propagation speed. In the terminology of first-passage percolation, v^{-1} is called the time constant of the percolation model.

It is instructive to study the properties of the fastest paths between two nodes x and y in simulated realizations of the Space–time SINR (or disk-interference) multigraphs. For example, the average hop length of this fastest path as a function of $\|x - y\|$ provides insight into how routing schemes should choose the next-hop neighbors for efficient communication.

11.4.2 Percolation on the static SINR graph

The static SINR graph is a simpler model, where half-duplex constraints and fading are not considered, but interference is. It is assumed that all nodes always transmit, and that pairs of nodes $\{x, y\}$ are connected by an undirected link if the SINR condition is met at both nodes. The interference can be suppressed by a factor γ . Hence the condition for the existence of an edge (11.2) is replaced by the condition

$$\mathbf{1}(z \leftrightarrow y) = \mathbf{1} \left(\frac{\ell(z-y)}{W + \gamma I_y^{!z}} \geq \theta \right) \mathbf{1} \left(\frac{\ell(z-y)}{W + \gamma I_z^{!y}} \geq \theta \right) \quad (11.5)$$

for an undirected edge. For the case where the nodes form a PPP of intensity λ , percolation occurs provided that γ is small enough.

THEOREM 11.20 (Percolation on the static SINR graph) *Let $\ell(x)$ be a path loss function that is isotropic, strictly decreasing, integrable on $\mathbb{R}^2 \setminus b(o, \infty)$, with $\ell(0) > \theta W$. Let λ_c be the critical density when $\gamma = 0$. Then, for any density $\lambda > \lambda_c$, there exists a $\gamma^*(\lambda) > 0$ such that for $\gamma \leq \gamma^*(\lambda)$, the static SINR graph percolates.*

Proof (Sketch) The proof consists of the following steps.

1. For $\gamma = 0$ (no interference), the model is just a disk graph with radius $r_0 = \tilde{\ell}^{-1}(\theta W)$, where $\tilde{\ell}(u) = \ell(\|u\|)$. λ_c is the corresponding critical intensity. For $\lambda > \lambda_c$, let

$$r^*(\lambda) = \sqrt{\frac{\lambda_c}{\lambda}} r_0 < r_0.$$

The disk graph $G_{r, \lambda}$ with $r > r^*(\lambda)$ is supercritical.

2. In the square lattice $(d\mathbb{Z})^2$, consider the event A that there is a horizontal and two vertical crossings in the rectangle of size $3d/2 \times d/2$ surrounding an edge in the lattice. Invoking a result in Meester & Roy (1996), by choosing d large, the probability of A can be made large.
3. Consider the event B if the interference shot-noise process does not exceed some value M at any point in the rectangle. This is shown using a modified path loss function.
4. Denote as open an edge for which both A and B hold: There are crossings and the interference is bounded. We need to show that, for appropriate choices of M and d , there is a positive probability of an infinite component. To do this, we need an exponential bound on having n closed edges.
5. Event A is easily bounded exponentially. For B , a careful application of Campbell's theorem for the PPP is required, including an upper bound on the interference in a lattice process. Finally, the intersection AB is exponentially bounded by an application of Schwartz's inequality.
6. Using Peierls' argument, by showing that there is no closed path around the origin with positive probability, it is established that percolation occurs for certain values of M and d .
7. If two nodes are in the same rectangle and at distance less than r , with $r > r^*$ (see step 1) there is a value of γ such that they are connected in the SINR model. So all connections that exist in the Boolean disk model also exist in the SINR model.

8. Crossings of adjacent edges overlap, so they all belong to the same component. \square
 These crossings also exist in the SINR model, thus there is an infinite component.

This is another instance where a result in the continuum world is obtained using a result on a lattice. Uniqueness of the infinite component has not been established for this model.

11.4.3 Secrecy graphs

There has been growing interest in information-theoretic secrecy in wireless networks. The *secrecy graph* is a graph model of a network that includes only edges along which secure communication is possible. Since the so-called secrecy capacity is positive if the SNR at the intended receiver is higher than that at an eavesdropper, this graph includes only (directed) edges from a transmitter that are shorter than the distance to the nearest eavesdropper.

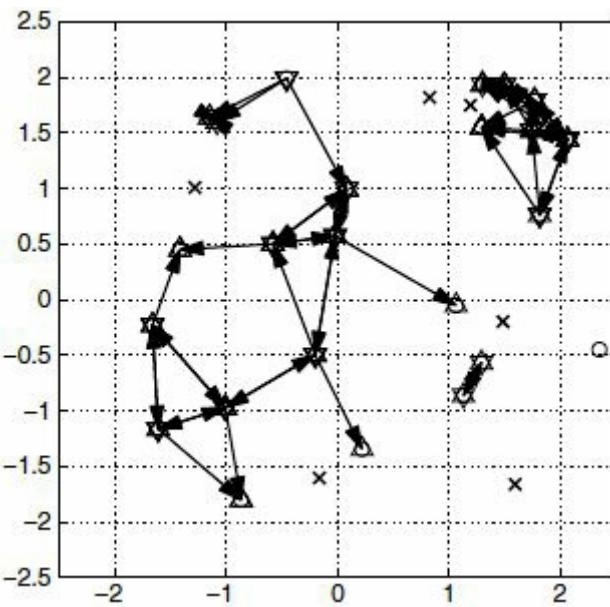


Figure 11.10 Example of a Poisson secrecy graph on a 5×5 square $\mathbb{G}_{1/4}$ (with eavesdropper density $\lambda = 0.25$). Eavesdroppers are shown as \times . Legitimate nodes are shown as \circ ; if they can receive from someone, a triangle Δ is added, and if they can talk to someone, a ∇ is added to the \circ .

Both the legitimate users (the “good guys”) and the eavesdroppers (“bad guys”) are modeled as point processes. The simplest model is the one in which both are PPPs.

DEFINITION 11.21 (Poisson secrecy graph) Let $\Phi \subset \mathbb{R}^d$ be a PPP of intensity 1 and $\Psi \subset \mathbb{R}^d$ an independent PPP of intensity λ . The Poisson secrecy graph is the directed graph $\mathbb{G}_\lambda(\Phi, \mathbb{E})$, where the edge set is given by

$$\vec{E} = \{(x, y) \in \Phi^2 : \|x - y\| < \|x - \Psi\|\}.$$

An example Poisson secrecy graph is shown in Fig. 11.10.

Since this is a directed graph, there are different ways of defining infinite components. If the typical node can reach an infinite number of nodes with positive probability, the graph *out-percolates*, i.e., there exists an infinite out-component. Conversely, if it can receive from an infinite number of nodes with positive probability, we have in-percolation and there is an infinite in-component. Further, we may consider two related undirected graphs:

- G_λ is defined as the undirected graph with an edge $\{x, y\}$ if $(x, y) \in \vec{E}$ or $(y, x) \in \vec{E}$; and
- G'_λ is the undirected graph with an edge $\{x, y\}$ if $(x, y) \in \vec{E}$ and $(y, x) \in \vec{E}$.

The two undirected graphs pertaining to the graph in Fig. 11.10 are shown in Fig. 11.11.

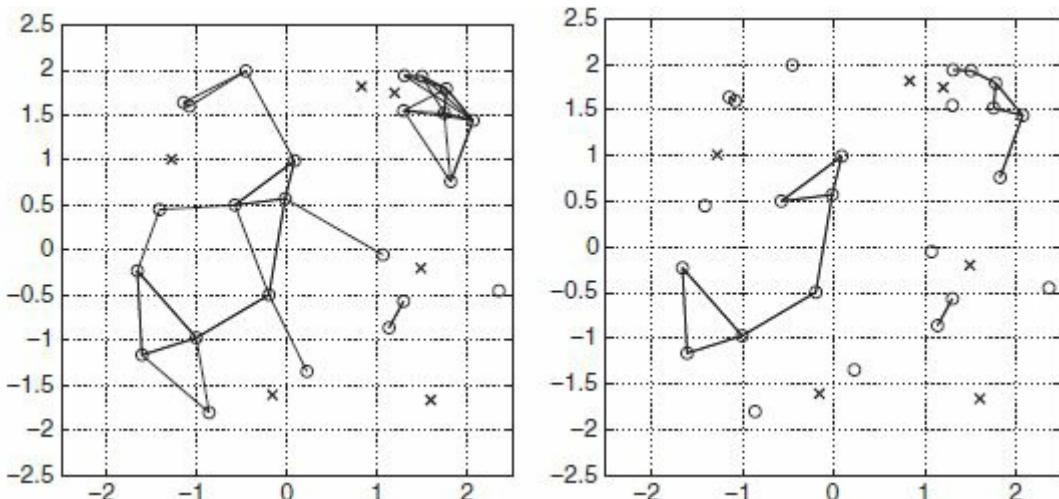


Figure 11.11 Undirected secrecy graphs pertaining to the (directed) secrecy graph in Fig. 11.10. (Left) The graph G_λ , where an edge $\{x, y\}$ is present if either (x, y) or (y, x) exists in the directed graph. (Right) The graph G'_λ , where an edge $\{x, y\}$ is present if both (x, y) and (y, x) exist in the directed graph.

For both these graphs, it is of interest to study the conditions for the emergence of a giant component. We expect in all cases the existence of a critical intensity λ_c , below which the corresponding infinite component exists a.s., and above which it does not a.s.

Here we show that out-percolation does *not* occur if $\lambda \geq 1$. Let $\theta_{\text{out}}(\lambda)$ be the probability that o belongs to an infinite out-component of \vec{G}_λ .

THEOREM 11.22 (Out-percolation on the Poisson secrecy graph) *If $\lambda \geq 1$, $\theta_{\text{out}}(\lambda) = 0$ for all dimensions d .*

Proof We compare the growth of the out-component originating at the typical node to the growth of a branching process. The out-degree in \mathcal{G}_λ is geometric with mean $1/\lambda$. This is easily seen if we consider the sequence of nearest neighbors of o in the combined process $\Phi \cup \Psi$. $N_{\text{out}} = n$ if the closest n points are in Φ and the $(n+1)$ st is in Ψ . Since these are independent events,

$$\mathbb{P}(N_{\text{out}} = n) = \frac{\lambda}{1+\lambda} \left(\frac{1}{1+\lambda} \right)^n. \quad (11.6)$$

The growth process of the out-component is stochastically dominated by the (independent) branching process with geometric distribution with mean $1/\lambda$. To see this, let $r(x) = \|x - \Psi\|$, $x \in \Phi$, and denote by C the set of nodes already explored and by

$$A = \bigcup_{x \in C} b(x, r(x))$$

the associated region. Now take one of the (out-)neighbors of a node in C , say y . Then there are two possibilities: $b(y, r(y)) \subset A$, in which case no new nodes are added to C ; and $b(y, r(y)) \setminus A \neq \emptyset$, in which case y 's out-degree is geometric with mean $1/\lambda$. Since the branching process dies out a.s. if $\lambda \geq 1$, so does the growth process. \square

Since the secrecy graph is a *dependent model*, where the existence of edges pertaining to nearby vertices are correlated events, it is difficult to get good bounds on the critical intensity. Simulations show that for out-percolation on \mathcal{G} , $\lambda_c \approx 0.15$, so, if the ratio of bad guys to good guys is higher than 0.15, it is impossible for a node to talk securely to an infinite number of other nodes.

Bibliographical notes

The most complete book on random geometric graphs is that by Penrose (2003). In particular, [Chapters 9 and 10](#) focus on percolation. There is a monograph dedicated to continuum percolation (Meester & Roy 1996); it includes in [Section 3.6](#) a proof of the uniqueness of the infinite component on the disk graph. The main objects of study are models based on PPPs. Franceschetti & Meester (2007) gives an introduction to different types of random geometric graphs and their percolation and connectivity properties with applications to communication networks. Random connectivity functions are also studied in Franceschetti *et al.* (2005), Balister *et al.*

(2004), and Balister *et al.* (2005).

Abstract random graphs are discussed rigorously and in detail in the books by Janson *et al.* (2000) and Bollobás (2001). A less technical account of random graphs is given by Spencer (2001).

The Space–time SINR multigraph was introduced first in Ganti & Haenggi (2007), and is analyzed in more detail in Ganti & Haenggi (2011). The Space–time SINR graph and its properties are studied in Baccelli *et al.* (2011) and in Chapter 22 of Baccelli & Blaszczyszyn (2009). The static SINR graph was introduced and studied in Dousse *et al.* (2006).

Secrecy graphs were introduced in Haenggi (2008b), and rigorous and high-confidence bounds on the different critical intensities were derived in Sarkar & Haenggi (2011).

Problems

11.1 Using a comparison with a Galton–Watson branching process, find a (numerical) bound on the percolation probability $\theta(r, \lambda)$ of Gilbert’s disk graph.

11.2 For the Gilbert random graph, show that $p = n^{-1-1/k}$ is the threshold function for the appearance of trees with $k + 1$ vertices.

11.3 In the Gilbert random graph, show that, if $np < 1$, the expected size of the component containing vertex 1 is bounded by

$$\mathbb{E}|C(1)| \leq \frac{1}{1 - np}.$$

11.4 For the Space–time SINR graph, find an upper bound on the in-degree of a vertex that depends only on the threshold θ .

11.5 Determine the mean in-degree of a receiver in the Space–time disk-interference snapshot graph. The out-degree of a transmitter was determined in Lemma 11.19. Use a similar calculation to that for the out-degree (Lemma 11.19). Then try to find a “shortcut,” where little to no calculation is needed.

11.6 For the Poisson secrecy graph, show that the in-degree N_{in} has the moment-generating function (mgf)

$$\mathbb{E}e^{tN_{\text{in}}} = \mathbb{E}(\exp(V(e^t - 1)/\lambda)),$$

where V is the volume of the Voronoi cell associated with the typical point in a PPP of intensity 1. For the one-dimensional case, give a concrete expression for the mgf, and determine $\mathbb{P}(N_{\text{in}} = k)$.

11.7 For out-percolation on the Poisson secrecy graph, show that

$$\theta(\lambda) \leq \max\{0, 1 - \lambda\},$$

where $\theta(\lambda)$ is the probability that the origin belongs to an infinite out-component in $\tilde{\mathcal{G}}_\lambda$.

11.8 For the undirected Poisson secrecy graph G_λ , find an upper bound on the critical intensity using a face percolation argument. The component at the origin is finite if it is surrounded by a circuit of closed faces. Define a face to be closed if it contains eavesdroppers in a configuration that makes it impossible for an edge to form from the inside of the circuit to the outside. If that configuration occurs with probability at least $1/2$, the component at o will be finite a.s. Note that the resulting bound will be relatively weak.

¹ If nodes have multi-packet reception capability, this topology can be more complicated.

12 Connectivity

12.1 Introduction

In some cases, percolation is not sufficient, since it merely guarantees that an infinite number of nodes is connected somewhere in a graph or network. In fact the *fraction* of connected nodes may be quite small; it could easily be that most nodes are disconnected from the infinite component. Connectivity is a much more stringent condition; it requires that all nodes are connected a.s.

Since the connectivity parameters, such as the radius of the disk graph, often need to be infinite for infinite graphs, we focus on finite graphs first, say on a square of area n , and then study the connectivity behavior of the graph as $n \rightarrow \infty$.

We start our discussion with the connectivity of the random lattice.

12.2 Connectivity of the random lattice

We consider a box $B(n)$ of size $n \times n$ vertices of the square lattice \mathbb{L}^2 . As in the bond percolation model, each edge is open with probability p . We would like to find the condition on p under which all vertices in $B(n)$ are connected, asymptotically as $n \rightarrow \infty$. We established that the critical probability for bond percolation on \mathbb{L}^2 is $p_c = 1/2$. Recall that $\theta(p)$ denotes the probability that the origin o belongs to the infinite component.

If $p > 1/2$, there exists a unique infinite component on the lattice, and each vertex in $B(n)$ is connected to it with probability $\theta(p)$. Hence there is a component in $B(n)$ of mean size $n^2\theta(p)$. Indeed, it can be shown that, in the limit $n \rightarrow \infty$, $\theta(p)$ is exactly the fraction of vertices in $B(n)$ that are connected.

Hence, for full connectivity, necessarily $p \rightarrow 1$. The question is how fast this needs to happen as n grows. Letting p be a function of n , the mean number of isolated nodes is asymptotically

$$\mathbb{E}N_{\text{isol}}(n) \sim n^2(1 - p(n))^4, \quad n \rightarrow \infty$$

since each (interior) node is isolated with probability $(1-p(n))^4$. Boundary nodes are asymptotically negligible, since their number is only $\Theta(n)$. Choosing

$$p(n) = 1 - \frac{c(n)}{\sqrt{n}}$$

results in $\mathbb{E}N_{\text{isol}}(n) \sim c(n)^4$. So $c(n) \rightarrow 0$ is necessary for asymptotic connectivity. The following theorem gives a sharper result.

THEOREM 12.1 (Isolated nodes in the random lattice) *Let $\mathbb{E}N_{\text{isol}}(n)$ be the number of isolated vertices in $B(n)$. Then $N_{\text{isol}}(n)$ converges to a Poisson random variable with mean $M > 0$ if and only if*

$$n^2(1-p(n))^4 \rightarrow M, \quad n \rightarrow \infty$$

Proof We use the so-called *Chen–Stein* method to show that the number of isolated vertices converges to a Poisson distribution under the stated condition. The particular result that we want to apply is the following.

LEMMA 12.2 (Chen–Stein bound on the total variation distance) *Let (A_i) , $i \in [n]$, be a collection of indicator random variables. Let $S = \sum_{i=1}^n A_i$, where $\mathbb{E}A_i = p_i$, and*

$$M \triangleq \sum_{i=1}^n p_i$$

with the assumption that $M < \infty$. Hence $\mathbb{E}S = M$. Then, if the A_i are increasing functions of independent random variables (X_k) ,

$$\|S - \text{Poi}(M)\|_{\text{TV}} \leq \frac{1 - e^{-M}}{M} \left(\text{var } S - M + 2 \sum_{i=1}^n p_i^2 \right), \quad (12.1)$$

where $\text{Poi}(M)$ is a Poisson random variable with mean M and $\|\cdot\|_{\text{TV}}$ is the total variation distance defined as

$$\|P - Q\|_{\text{TV}} \triangleq \sup \{Y \in \mathcal{F}: |\mathbb{P}(Q \in Y) - \mathbb{P}(P \in Y)|\}$$

for two random variables P and Q defined on the same probability space with σ -algebra \mathcal{F} .

In our situation, the underlying independent random variables (X_k) are the indicators that the edges in the dual lattice are closed, and the random variables A_i , or A_x , $x \in B(n)$, are the indicators that vertex x is isolated, which clearly are

increasing functions of the (X_k) . Further, S is the number of isolated nodes in $B(n)$, given as

$$N_{\text{isol}}(n) = \sum_{x \in B(n)} \mathbf{1}(x \text{ isolated}).$$

The four corner vertices are isolated with probability $(1 - p(n))^2$, the other boundary vertices with probability $(1 - p(n))^3$, and the interior nodes with probability $(1 - p(n))^4$. Adding up these probabilities and using $1 - p(n) = \Theta(1/\sqrt{n})$),

$$\begin{aligned} \mathbb{E}N_{\text{isol}}(n) &= \sum_{x \in B(n)} \mathbb{P}(x \text{ isolated}) \\ &= (n-2)^2(1-p(n))^4 + (4n-8)(1-p(n))^3 + 4(1-p(n))^2 \\ &= n^2(1-p(n))^4 + \Theta(1/\sqrt{n}) \\ &\rightarrow M, \quad n \rightarrow \infty. \end{aligned}$$

Let $p_x = \mathbb{P}(x \text{ isolated})$. Next, to apply (12.1), we need the sum $\sum_{x \in B(n)} P_x^2$ which it is easy to show goes to zero as $n \rightarrow \infty$. The last ingredient is the second moment $\mathbb{E}(N_{\text{isol}}(n)^2)$, which can be expressed as

$$\mathbb{E}(N_{\text{isol}}(n)^2) = \mathbb{E} \left(\sum_{x \in B(n)} \mathbf{1}(x \text{ isolated}) \right) + \mathbb{E} \left(\sum_{x,y \in B(n)} \mathbf{1}(x \text{ and } y \text{ isolated}) \right).$$

The first term is M again, and the second one is dominated by the (almost) n^4 vertices which do not share a common bond. So it is asymptotically $n^4(1-p(n))^8 \sim M^2$. It follows that

$$\lim_{n \rightarrow \infty} \text{var } N_{\text{isol}} = \lim_{n \rightarrow \infty} (\mathbb{E}(N_{\text{isol}}(n)^2) - \mathbb{E}(N_{\text{isol}}(n))^2) = M.$$

So the right side of (12.1) approaches 0 asymptotically, which shows that the total variation distance goes to 0, and thus M converges in distribution to the Poisson distribution. \square

We know that the Poisson distribution emerges naturally as the limiting distribution of the sum of n independent indicator random variables with low probability. The Chen–Stein bound extends this result to the situation of dependent random variables, where the dependence needs to be asymptotically negligible. Lemma 12.2 gives a precise sufficient condition for this.

From the theorem follows the next corollary.

COROLLARY 12.3 (Absence of isolated nodes in the random lattice) *If $c(n) \rightarrow c$ as n*

$\rightarrow \infty$, the number of isolated nodes is Poisson distributed with mean c^4 , and thus the probability of having no isolated nodes is $\exp(-c^4)$.

Next we need to show that the absence of isolated nodes implies that the lattice is connected.

THEOREM 12.4 (Connectivity of the random lattice) *Let $p(n) = 1 - c(n)/\sqrt{n}$. Then the vertices in $B(n)$ are connected if and only if $c(n) = o(1)$, i.e., if $c(n) \rightarrow 0$ as $n \rightarrow \infty$.*

Proof We show that if $c(n) \rightarrow c$, where $c \in (0, \infty)$, the vertices in $B(n)$ are either isolated or belong to a single component that contains all non-isolated nodes a.a.s. Again we need only focus on the interior nodes. For two (adjacent) nodes to be isolated, a self-avoiding path of length at least 6 needs to exist in the dual graph. The probability of this event is

$$P_6 \leq \sum_{k=6}^{\infty} 4 \cdot 3^{k-1} (1 - p(n))^k = \Theta(n^{-3}).$$

The order of such pairs of adjacent nodes is $\Theta(n^2)$, so asymptotically there will be no isolated components of size 2. Components of higher order are even less likely to occur, since they need longer self-avoiding paths in the dual graph. So as soon as the isolated nodes vanish, the graph is fully connected. \square

12.3 Connectivity of the disk graph

12.3.1 Simple connectivity

Again we first focus on isolated nodes. So the first question is as follows: How large does the mean degree a have to be in order that a.a.s. there are no isolated nodes in the disk graph? The answer is given in the following theorem.

THEOREM 12.5 (Isolated nodes in a disk graph) *Let $G^{(n)}_{r(n),1}$ be the subgraph containing the vertex set $\Phi \cap [0, \sqrt{n}]^2$ inherited from $G_{r(n),1} = (\Phi, E)$ and all the edges between the vertices present. If*

$$a(n) \triangleq \pi r^2(n) = \log n + c(n), \quad n \rightarrow \infty$$

where $c(n) = \omega(1)$, then a.a.s. there are no isolated nodes in $G^{(n)}_{r(n),1}$.

$c(n) = \omega(1)$ is any function diverging to infinity as $n \rightarrow \infty$.

Proof The expected number of isolated nodes on the square $[0, \sqrt{n}]^2$ is

$$\mathbb{E}N_{\text{isol}}(n, r) = n \exp(-\pi r^2).$$

If $a(n) = \pi r(n)^2 = \log n$, then $\mathbb{E}N_{\text{isol}} = 1$. If $a(n) = \log n + \omega(1)$, $\mathbb{E}N_{\text{isol}}(n, r) \rightarrow 0$. \square

As in the random lattice, if $c(n) \rightarrow c$, the number of isolated vertices is asymptotically Poisson with mean e^{-c} , and

$$\lim_{n \rightarrow \infty} \mathbb{P}(N_{\text{isol}}(n) = 0) = \exp(-e^{-c}). \quad (12.2)$$

This can be shown using another Chen–Stein-type bound.

Again the main obstacles to connectivity are isolated nodes, so we have the following result on the connectivity.

THEOREM 12.6 (Connectivity of a disk graph) *The disk graph $G^{(n)}_{r(n,1)}$ is connected a.a.s. if and only if $c(n) \rightarrow c = \infty$.*

The proof is rather lengthy and hence has been omitted.

On the other hand, if $c = -\infty$ (arbitrarily slowly) then the graph is disconnected a.a.s. This follows from (12.2), which shows that, in this case, the probability of having no isolated node goes to 0.

If $r^2(n)/\log n \rightarrow \infty$, the graph is said to be in the *superconnectivity regime*, whereas if $r^2(n)/\log n \rightarrow 0$, the graph is in the *subconnectivity regime*.

Applied to wireless networks, this result says that the radius of transmission (at a given rate) needs to increase with $\sqrt{\log n}$. In terms of transmit power, if the path loss exponent is at least 2, it means that the power needs to scale at least with $\log n$ in order to keep a network of size n connected.

12.3.2 Multiple connectivity

A more general question is the following: When does the disk graph become k -connected? k -connectivity means that the graph cannot be disconnected by the removal of $k - 1$ or fewer vertices. Equivalently, by Menger’s theorem, a graph is k -connected if there are at least k independent paths between all pairs of distinct vertices.

Again we start by calculating the number of vertices that are k -connected.

THEOREM 12.7 (Nodes with degree less than k) *Let $G^{(n)}_{r(n),1}$ be the subgraph containing the vertex set $\Phi \cap [0, \sqrt{n}]^2$ inherited from $G_{r(n),1} = (\Phi, E)$ and all the*

edges between the vertices present. If

$$\pi r^2 = \log n + (k-1)\log \log n - \log \Gamma(k) + c(n),$$

where $c(n) = \omega(1)$, as $n \rightarrow \infty$ the graph $G^{(n)}_{r(n),1}$ does not contain any node with degree less than with k a.a.s.

Proof Let N_{k-1} be the number of nodes with degree $k-1$. For $\pi r^2 = \log n + (k-1)\log \log n - \log(\Gamma(k))$,

$$\begin{aligned} \mathbb{E}(N_{k-1}) &= n \exp(-\pi r^2) \frac{(\pi r^2)^{k-1}}{\Gamma(k)} \\ &= n \cdot \frac{(\log n)^{1-k} \Gamma(k)}{n} \cdot \frac{[\log n + (k-1)\log \log n - \log \Gamma(k)]^{k-1}}{\Gamma(k)} \\ &= 1 + \Theta\left(\frac{\log \log n}{\log n}\right) = 1 + o(1). \end{aligned}$$

If we make πr^2 a little larger, by $c(n) = \omega(1)$, the number of nodes with degree $k-1$ (or smaller¹) goes to zero. \square

It turns out that, as in the case of single connectivity, the disk graph becomes k -connected as soon as there are no vertices of degree $k-1$ or lower. Hence the condition for k -connectivity follows from this result.

Remarkably, k -connectivity is achieved at small additional cost compared with simple connectivity. For $\pi r^2 = (1 + \epsilon) \log n$, k -connectivity is achieved asymptotically for any k , for all $\epsilon > 0$.

12.4 Connectivity of basic random geometric graphs

Here we briefly discuss the random geometric graph with connectivity function ξ , as defined in Definition 11.1. The vertices form a PPP Φ of intensity λ . We add the origin o to the point process. The mean node degree is

$$a = \lambda \int_{\mathbb{R}^d} \xi(x) dx.$$

We denote by $\theta(\lambda) = \mathbb{P}(|C(o)| = \infty)$ the probability that the component of the origin is infinite. It is plausible that

$$\lim_{\lambda \rightarrow \infty} \theta(\lambda) = 1.$$

As for the disk graph, it can be shown that there exists a critical density λ_c below

which $\theta(\lambda) = 0$ and above which $\theta(\lambda) > 0$. The fact that there is no percolation for small enough λ can (again) be proven by a comparison with a branching process: If $a \leq 1$, then there is no percolation. The fact that there is percolation for a large enough λ can be shown using a comparison with site percolation on the square lattice.

More strongly, it is also reasonable to guess that, given that a point's component is finite, the probability of that point being isolated tends to 1 as $\lambda \rightarrow \infty$, as in the disk graph. Indeed, we have

$$\lim_{\lambda \rightarrow \infty} \frac{-\log(1 - \theta(\lambda))}{a} = 1.$$

Written differently, this says that $1 - \theta(\lambda) \sim e^{-a}$, $\lambda \rightarrow \infty$, so the probability of being in a finite component equals the probability of being isolated in the limit, as in the disk graph. The intuition behind this result is the following. As λ grows, a finite component with k points occurs with higher probability if the k points lie close together, since then a smaller area around the component needs to be empty. Having k points inside a small ball is most likely if $k = 1$.

12.5 Other graphs

12.5.1 Nearest-neighbor graphs

Let Φ be a PPP of intensity 1 on $[0, \sqrt{n}]$ so that $\mathbb{E}|\Phi| = n$. Assume an undirected edge exists from each node to its k nearest neighbors and denote the resulting graph by G_k . Each node has degree at least k , since it “sends out” k edges. Most nodes will have a larger degree due to the edges they get since they are among the k nearest neighbors of another node. So $\mathbb{E}(\deg G_k) > k$. The natural question is the following: What is the minimum k that guarantees connectivity a.a.s.? The best known bounds are given in the following theorem.

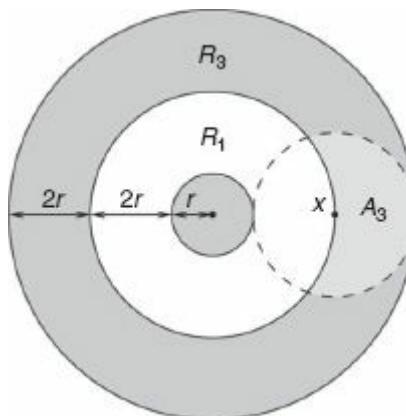


Figure 12.1 Illustration for Theorem 12.8. A likely scenario for the k -nearest-neighbor graph to be disconnected is when $k + 1$ vertices are located in the inner ball of radius r , the annulus R_1 is empty, and each vertex in the outer annulus R_3 has k neighbors in R_3 .

THEOREM 12.8 *The k -nearest-neighbor graph is connected a.a.s. if and only if*

$$0.4125 \log n < k(n) < 0.5139 \log n.$$

We do not give a proof here, since the constants are difficult to establish. It is not difficult, though, to show that $k(n) = \Theta(\log n)$ is the right scaling.

As usual in these situations, finding the most likely scenario for the graph to be disconnected gives the right scaling. In this case, the scenario is the following (see Fig. 12.1).

- The smallest disk contains a component of size at least $k + 1$.
- The annulus R_1 (of inner radius r and outer radius $3r$) is empty.
- None of the nearest k neighbors of a node in the annulus R_3 (of inner radius $3r$ and outer radius $5r$) is in the smallest disk.
- A node at x must have k neighbors inside A_3 (otherwise it may send an edge to a node in the smallest disk).

These events together ensure that the component inside the smallest disk is isolated. If $k = \log n / 8$ and $\pi r^2 = k + 1$, they are bound to occur as n grows. This establishes a lower bound of $\log n / 8 < k$.

12.5.2 Abstract random graphs

Here we explore the connectivity of the Gilbert random graph. From the prior results, we may guess that a mean degree of $\log n$ is necessary. Indeed, it turns out that, at $pn = \log n$, the graph becomes connected.

THEOREM 12.9 (Critical probability for connectivity) *If*

$$p = \frac{\log n}{n} + \frac{t}{n}, \quad t \in \mathbb{R},$$

then

$$\lim_{n \rightarrow \infty} \mathbb{P}(G(n, p) \text{ connected}) = \exp(-e^{-t}).$$

To make this plausible, set $p(n) = c/n$. Then asymptotically the node degree is Poisson with mean c , a vertex is isolated with probability e^{-c} , and the mean number

of isolated vertices is ne^{-c} . A necessary condition for connectivity is thus that $ne^{-c} \rightarrow 0$, which requires $c > \log n$, or $p > \log n/n$. This simple calculation gives the right order for $p(n)$ required for connectivity. For $p = \log n/n + t/n$, the mean number of isolated nodes is e^{-t} . It can be shown that the actual number is (again) Poisson distributed, so the probability of having no isolated node is $\exp(-e^{-t})$, as given in the theorem. Of course, we have not shown that the absence of isolated nodes is a *sufficient* condition for connectivity, but it turns out that it is.

Bibliographical notes

Most of the books listed in the previous chapter contain significant parts on connectivity, including Penrose (2003) and Franceschetti & Meester (2007) for the geometric setting, and Janson *et al.* (2000) and Bollobás (2001) in the abstract setting. In particular, the proof of Theorem 12.6 can be found in Penrose's book.

Recent results on the k -nearest-neighbor graphs can be found in Balister *et al.* (2005); the best currently known lower bound for connectivity (see Theorem 12.8) is proven in Walters (2012).

The Chen–Stein method is discussed in detail in Barbour & Chen (2005).

Problems

12.1 Without using any of the results in this chapter, show the following: For the disk graph,

$$\pi r^2(n) = 5\pi \log n$$

is a sufficient condition for a.a.s. connectivity.

Hint. Partition the square of area n into small squares of area $\sqrt{\log n}$. Define an event on each square and show that the probability of all events occurring tends to 1 as $n \rightarrow \infty$. Then prove that this joint event implies connectivity under the given condition for $r(n)$.

12.2 Estimate the probabilities of the events in the “likely scenario for disconnectedness” in the discussion of the lower bound in Theorem 12.8 and show that $k(n) > \log n/8$ is necessary.

12.3 Consider the 1-nearest-neighbor graph with vertex set Φ , where Φ is a PPP of intensity 1 on \mathbb{R}^d . Describe the properties of its components for general d . Then focus on $d = 1$ and make statements about the distribution of the size of the typical component.

12.4 Take a Gaussian PPP with $\sigma^2 = 1/2$ (see Example 2.2), so that the intensity is

$\lambda(r) = (n/\pi)e^{-r^2}$. Add a point at the origin, and define a disk graph by connecting two points within distance s of each other. Let ρ be the largest radius u such that all points inside $b(o, u)$ are connected to o . Show that, assuming $Rs \ll 1$, the number of isolated nodes in the ball $b(o, R)$ can be approximated as

$$\mathbb{E}N_{\text{isol}}(R) \approx \frac{\exp(-ns^2 e^{-R^2}) - \exp(-ns^2)}{s^2}.$$

Then, assuming $n \rightarrow \infty$ and $n^{-1/2} \ll s \ll (\log n)^{-1/2}$, show that

$$\rho \approx \log(ns^2/\log(1/s^2)).$$

¹ Asymptotically, $\mathbb{E}(N_{k-j}) \ll \mathbb{E}(N_{k-1})$, for $j \in \{2, 3, \dots, k\}$.

13 Coverage

13.1 Introduction

In a geometric setting, a natural counterpart to connectivity is coverage. Here it is assumed that a random set is attached to each point of a point process, and the main question is whether the union of all these random sets covers a given target area or volume, possibly the entire \mathbb{R}^d .

In the context of wireless networks, such questions naturally arise in cellular systems, where the goal is to provide a large region or country with cellular service, and in sensor networks, where the random sets are the regions covered by each individual sensor and the goal is to monitor a large area of interest.

13.2 Germ–grain and Boolean models

13.2.1 Germ–grain models

The germ–grain model is defined as the union of independent identically distributed compact random sets (grains) shifted by points (germs) of a point process.

DEFINITION 13.1 (Germ–grain model) Let $\Phi = \{x_i\}$ be a point process on \mathbb{R}^d , the *germs*, and (S_1, S_2, \dots) a collection of random non-empty sets, the *grains*. Then the union

$$\Xi = \bigcup_{i \in \mathbb{N}} x_i + S_i$$

is a *germ–grain model*. $\Xi \subset \mathbb{R}^d$ is the *occupied region* and $\mathbb{R}^d \setminus \Xi$ is the *unoccupied* or *vacant region*.

The collection of random sets $(x_i + S_i)$ is referred to as a *coverage process*. Alternatively, the random set S may be indexed by the grains, as in marked point

processes, i.e.,

$$\Xi = \bigcup_{x \in \Phi} x + S_x \subset \mathbb{R}^d.$$

A location $y \in \mathbb{R}^d$ is said to be *covered* if $y \in \Xi$.

If the sets S_x or S_i are countable, Ξ is a cluster (point) process, as described in Section 3.4, and $|S| = v_d(S) = 0$. To have any hope of covering an uncountable subset of \mathbb{R}^d , we need $|S| > 0$ with positive probability.

DEFINITION 13.2 (Vacancy) The *vacancy* $V(\mathcal{R})$ is the Lebesgue measure of the part of a region of interest $\mathcal{R} \subseteq \mathbb{R}^d$ that is not covered:

$$V(\mathcal{R}) \triangleq |\mathcal{R} \setminus \Xi| = \int_{\mathcal{R}} \chi(y) dy,$$

where

$$\chi(y) = \mathbf{1}(y \notin \Xi) = \prod_i \mathbf{1}(y \notin S_i + x_i).$$

Each connected subset of Ξ forms a *component*; in the context of coverage processes, components are also referred to as *clumps*.

DEFINITION 13.3 (Coverage) A set $\mathcal{R} \subseteq \mathbb{R}^d$ is said to be covered by the occupied region Ξ if $\mathbb{P}(V=0) = 1$.

Remark 13.1 The condition $\mathbb{P}(V=0) = 1$ does not imply that $\mathcal{R} \subseteq \Xi$, since a subset of \mathcal{R} of lower dimension (and thus Lebesgue measure 0 in d dimensions) might not be covered. So, strictly speaking, $\mathbb{P}(V=0) = 1$ is only a *necessary condition* for $\mathcal{R} \subseteq \Xi$. However, if the grains S are closed sets and the region \mathcal{R} to be covered is a bounded open set, the two are equivalent, i.e., $\mathbb{P}(V=0) = 1 \Leftrightarrow \mathcal{R} \setminus \Xi = \emptyset$.

Apart from the question of coverage, an important question is whether the covered region contains a clump of infinite size. If it does, the coverage process is said to percolate. The relationships between connectivity, coverage, and the respective percolation events are summarized in Box 13.1.

Box 13.1 Percolation, connectivity, and coverage

Let $C(o)$ be the vertices in the component containing the origin in a random geometric graph $G(\Phi, E)$ and $\Xi(o)$ the clump of a coverage process containing o .

The parallels between connectivity and coverage are related as follows.

<i>Random object</i>	<i>event or property</i>
Random geometric graph connectivity	$\equiv C(o) \cap \Phi = \Phi$
Coverage process	percolation $\equiv P(C(o) = \infty) > 0$ coverage $\equiv \Xi(o) \cap \mathcal{R} = \mathcal{R}$ percolation $\equiv P(\Xi(o) = \infty) > 0$

13.2.2 Boolean models

The most tractable germ-grain model is the *Boolean model*, defined as follows.

DEFINITION 13.4 (Boolean model) A Boolean model is a germ–grain model where the germ point process is a uniform PPP and the grains S_i are iid.

Sometimes, germ–grain models whose germs form a lattice are called lattice Boolean models. A typical case of a Boolean model is where the grains S_i are balls of random radii. The first important result states that each location is covered a Poisson number of times.

THEOREM 13.5 (Vacancy and coverage in a Boolean model) *In the Boolean model where the PPP has intensity λ , the probability that a location is not covered is $\exp(-\lambda \mathbb{E}|S|)$, which is also the fraction of \mathbb{R}^d that is not occupied. It follows that $\Xi \neq \mathbb{R}^d$ if $\mathbb{E}|S| < \infty$. Moreover, the number of times each location is covered is Poisson distributed with mean $\lambda \mathbb{E}|S|$.*

Proof Let M be the number of times the origin is covered:

$$M = \sum_{x \in \Phi} \mathbf{1}(o \in x + S_x). \quad (13.1)$$

Let $F(y) = \mathbb{P}(y \in S)$, such that $\mathbb{E}|S| = \int_{\mathbb{R}^d} F(y) dy$. Taking expectations,

$$\begin{aligned} EM &= \mathbb{E} \left[\sum_{x \in \Phi} \mathbb{E}(\mathbf{1}(o \in x + S_x) \mid \Phi) \right] \\ &= \mathbb{E} \left[\sum_{x \in \Phi} F(-x) \right] \\ &\stackrel{(a)}{=} \lambda \int_{\mathbb{R}^d} F(-y) dy \\ &= \lambda \mathbb{E}|S|, \end{aligned}$$

where (a) follows from Campbell's theorem for sums (Theorem 4.1). To show that M is Poisson distributed, we calculate the moment-generating function using Campbell's theorem for the PPP (Theorem 4.6), applied to the function $f(x) = \mathbb{1}(o \in x + S_x)$ as used in (13.1). We have

$$\begin{aligned}\mathbb{E}e^{tM} &= \mathbb{E} \exp \left(t \sum_{x \in \Phi} f(x) \right) \\ &\stackrel{(a)}{=} \exp \left(\lambda \int_{\mathbb{R}^d} (\mathbb{E}e^{tf(x)} - 1) dx \right) \\ &\stackrel{(b)}{=} \exp \left(\lambda \int_{\mathbb{R}^d} (e^t - 1) F(-x) dx \right) \\ &= \exp(\lambda \mathbb{E}(|S|)(e^t - 1)),\end{aligned}$$

where (a) follows from Campbell's theorem, Theorem 4.6, and (b) from the fact that

$$\mathbb{E}e^{t\mathbb{1}(o \in x + S_x)} = 1 - F(-x) + e^t F(-x).$$

So M has the moment-generating function of a Poisson random variable and thus is \square Poisson distributed.

Application of this theorem yields the first and second moments of the vacancy.

COROLLARY 13.6 (First and second moments of the vacancy) *For the Boolean model, the first two moments of the vacancy are given by*

$$\mathbb{E}(V(\mathcal{R})) = |\mathcal{R}| \exp(-\lambda \mathbb{E}|S|)$$

and

$$\mathbb{E}(V(\mathcal{R})^2) = \iint_{\mathcal{R}^2} \exp \left(-2\lambda \mathbb{E}|S| + \lambda \mathbb{E}(|(y_1 - y_2 + S) \cap S|) \right) dy_1 dy_2.$$

Proof The mean is

$$\mathbb{E}V = \int_{\mathcal{R}} \mathbb{P}(x \notin \Xi) dx = \int_{\mathcal{R}} \mathbb{E}\chi(x) dx = |\mathcal{R}| \exp(-\lambda \mathbb{E}|S|).$$

For the second moment, we have

$$\begin{aligned}\mathbb{E}(V^2) &= \iint_{\mathcal{R}^2} \mathbb{P}(y_1 \notin \Xi, y_2 \notin \Xi) dy_1 dy_2 \\ &= \iint_{\mathcal{R}^2} \mathbb{E}(\chi(y_1)\chi(y_2)) dy_1 dy_2,\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}(\chi(y_1)\chi(y_2)) &= \mathbb{P}(\forall i: y_1 \notin x_i + S_i, y_2 \notin x_i + S_i) \\
&\stackrel{(a)}{=} \mathbb{P}(\forall i: x_i \notin y_1 - y_2 + S_i, x_i \notin S_i) \\
&= \mathbb{P}(\forall i: x_i \notin (y_1 - y_2 + S_i) \cup S_i) \\
&\stackrel{(b)}{=} \exp(-\lambda \mathbb{E}(|(y_1 - y_2 + S) \cup S|)) \\
&= \exp(-2\lambda \mathbb{E}|S| + \lambda \mathbb{E}(|(y_1 - y_2 + S) \cap S|)). \tag{13.2}
\end{aligned}$$

□

Here (a) follows from stationarity, and (b) follows from Theorem 13.5.

If $\mathbb{E}(|S|) = \infty$, then the number of germs in each component is infinite a.s. if $\lambda > 0$, i.e., there is no critical density. Otherwise, the main questions are whether the component in Ξ containing the origin has a positive probability of being infinite (assuming the origin is part of the PPP) and whether \mathbb{R}^d is covered. As for the disk graph (see Theorem 11.4), it can be shown that the infinite component is unique a.s.

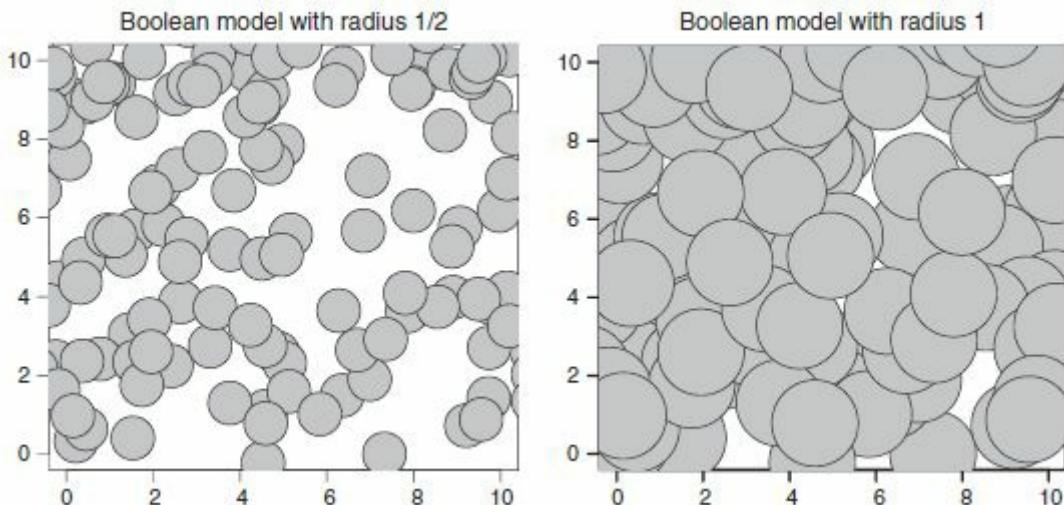


Figure 13.1 Boolean models on a PPP of intensity 1 with disks of radius r as grains. (Left) $r = 1/2$. The uncovered fraction is $\exp(-\pi/4) \approx 45\%$. (Right) $r = 1$. The uncovered fraction is $\exp(-\pi) \approx 4\%$.

13.3 Boolean model with fixed disks

13.3.1 Single coverage

The most studied model is the Boolean model with disks (or balls) of fixed radius r as grains. In this case we may write $\Xi = \Phi \oplus S$. This coverage process is then equivalent to Gilbert's disk graph. In fact, from the disk graph, we immediately have a result on the coverage percolation problem: If $r > r_c/2$, where r_c is the critical radius for percolation on the disk graph, there is an infinite component in

the Boolean model, since then two disks of the coverage process overlap if their vertices are connected in the disk graph.

A basic application of this model is in sensor networks. If the sensors are capable of detecting events within radius r , Ξ is the region covered by the sensor network. [Figure 13.1](#) shows two examples of these Boolean models.

We are interested in finding the condition on the disk radius r , so that a square of area n is covered a.a.s. as $n \rightarrow \infty$. We attack this problem in three steps. First we find a necessary condition that is based only on the mean vacancy. Second, we employ the variance of the vacancy to improve upon the necessary condition. Third, we show that this second necessary condition is also sufficient.

A first necessary condition for coverage of the plane

Focusing on the two-dimensional case, the most basic question is how large does r need to be in order that the entire plane is covered? Since we may always scale λ and r while keeping λr^2 constant, we fix $\lambda = 1$ without loss of generality.

The probability that the origin is not covered is the probability that there is no point of the PPP within distance r , i.e.,

$$\mathbb{E}_\chi(o) = \exp(-\pi r^2).$$

This holds for any point in \mathbb{R}^2 , so the expected vacancy of a square of area n is

$$\mathbb{E}V([0, \sqrt{n}]^2) = \mathbb{E} |[0, \sqrt{n}]^2 \setminus \Xi| = n \exp(-\pi r^2).$$

This also follows from [Corollary 13.6](#).

A necessary condition for coverage is $\mathbb{E}V([0, \sqrt{n}]^2) \rightarrow 0$ as $n \rightarrow \infty$. This is guaranteed by

$$\pi r^2 = \log n + \omega(1), \quad n \rightarrow \infty, \tag{13.3}$$

which is the same condition as for connectivity. However, $\mathbb{E}V \rightarrow 0$ is not *sufficient* for coverage, since it does not guarantee that $\mathbb{P}(V = 0) \rightarrow 1$. The non-implication $\mathbb{E}V(n) \rightarrow 0 \Rightarrow \mathbb{P}(V = 0) \rightarrow 1$ exemplifies that convergence in mean does not imply a.s. convergence.

A second necessary condition for coverage of the plane

A bound on $\mathbb{P}(V = 0)$ can be obtained if the second moment (or variance) of the vacancy is also known. Writing $V = V1(V > 0)$ and applying the Cauchy–Schwartz inequality, we have

$$\mathbb{E}V = \mathbb{E}(V1(V > 0)) \leq (\mathbb{E}(V^2)\mathbb{P}(V > 0))^{1/2},$$

and it follows that

$$\mathbb{P}(V > 0) \geq \frac{(\mathbb{E}V)^2}{\mathbb{E}(V^2)} = 1 - \frac{\text{var } V}{\mathbb{E}(V^2)} \implies \mathbb{P}(V = 0) \leq \frac{\text{var } V}{\mathbb{E}(V^2)}. \quad (13.4)$$

Hence we need to calculate the variance, for which we can use Corollary 13.6. The covariance of the vacancy indicators follows from (13.2) as

$$\text{cov}(\chi(y_1)\chi(y_2)) = \exp(-2\lambda\mathbb{E}|S|) (\exp(\lambda\mathbb{E}(|(y_1 - y_2 + S) \cap S|)) - 1),$$

and thus

$$\begin{aligned} \text{var } V(\mathcal{R}) &= \iint_{\mathcal{R}^2} \text{cov}(\chi(x_1), \chi(x_2)) dx_1 dx_2 \\ &= \exp(-2\pi r^2) \iint_{\mathcal{R}^2} (\exp(|b(o, r) \cap b(y_1 - y_2, r)|) - 1) dy_1 dy_2 \\ &\sim |\mathcal{R}| e^{-2\pi r^2} \int_{\mathcal{R}^2} (\exp(|b(o, r) \cap b(y, r)|) - 1) dy \quad \text{as } |\mathcal{R}| \rightarrow \infty \end{aligned}$$

since asymptotically the boundary effects vanish. The area of the intersection of two disks of radius r at distance h is (cf. (6.4))

$$A(h) = 2r^2 \arccos\left(\frac{h}{2r}\right) - \frac{h}{2} \sqrt{4r^2 - h^2}, \quad h < 2r.$$

It follows that

$$\text{var } V(\mathcal{R}) \sim |\mathcal{R}| e^{-2\pi r^2} 2\pi \int_0^{2r} h(e^{A(h)} - 1) dh.$$

Since $A(h) \geq \pi r^2 - 2rh$, the integral is tightly lower bounded as

$$\int_0^{2r} h(e^{A(h)} - 1) dh > \int_0^{\pi r^2/2} h(e^{\pi h^2 - 2hr} - 1) dh,$$

and hence

$$\begin{aligned} \text{var } V(\mathcal{R}) &\gtrsim |\mathcal{R}| e^{-2\pi r^2} 2\pi \int_0^{2r} h(e^{\pi h^2 - 2hr} - 1) dh \\ &= |\mathcal{R}| \pi e^{-\pi r^2} \frac{2 - e^{-\pi r^2} (\pi^2 r^4 + 2\pi r^2 + 2)}{4r^2}, \end{aligned}$$

where \gtrsim denotes asymptotic inequality. To obtain an upper bound, we note that $A(h) \leq \pi r^2 - \pi rh/2$ (for $h \leq 2r$), which yields

$$\text{var } V(\mathcal{R}) \lesssim |\mathcal{R}| 4e^{-\pi r^2} \frac{2 - e^{-\pi r^2} (\pi^2 r^4 + 2\pi r^2 + 2)}{\pi r^2}.$$

So there is a constant factor of $16/\pi^2$ between the two asymptotic bounds, and we have

$$\text{var } V([0, \sqrt{n}]^2) = \Theta\left(n \frac{e^{-\pi r^2}}{r^2}\right), \quad \text{as } n \rightarrow \infty.$$

As $n \rightarrow \infty$, necessarily $r \rightarrow \infty$ to have any hope for coverage (cf. (13.3)), thus we have dropped all the smaller terms in the bounds on the variance.

It follows that the lower bound on $\mathbb{P}(V > 0)$ in (13.4) needs to be zero for coverage, i.e., $(\mathbb{E}V)^2/\mathbb{E}(V^2) \rightarrow 0$. Since $(\mathbb{E}V)^2 = n^2 e^{-2\pi r^2}$, the condition is

$$\frac{(\mathbb{E}V)^2}{\text{var } V + (\mathbb{E}V)^2} \rightarrow 0 \iff nr^2 e^{-\pi r^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Owing to the factor r^2 , this is a more restrictive necessary condition than (13.3). Taking $\pi r^2 = \log n + \log \log n + c$, we have $nr^2 e^{-\pi r^2} \rightarrow e^{-c}/\pi$, so we need to replace the constant c by a function that is $\omega(1)$, i.e., our new necessary condition for coverage of the plane is

$$\pi r^2 = \log n + \log \log n + \omega(1). \quad (13.5)$$

The necessary and sufficient condition for coverage of the plane

To show that this tighter necessary condition (13.5) is also sufficient, we use an observation made by Gilbert in the case of open disks: *The square of area n is covered if all intersections of disk boundaries are covered, as are all intersections between disk boundaries and the border of the square of area n .*

If the disks are closed, the result is the same, but the formulation of Gilbert's condition is less elegant, since all intersections of disk boundaries now have to be covered three times (they are already covered by the two disks forming the intersection but need to be covered by another disk), while the intersections between disk boundaries and the border need to be covered twice (they are already covered once by the disk giving rise to the intersection and need to be covered by another disk). Figure 13.2 shows two realizations of PPPs on the unit square and the circles of radius r with their intersections.

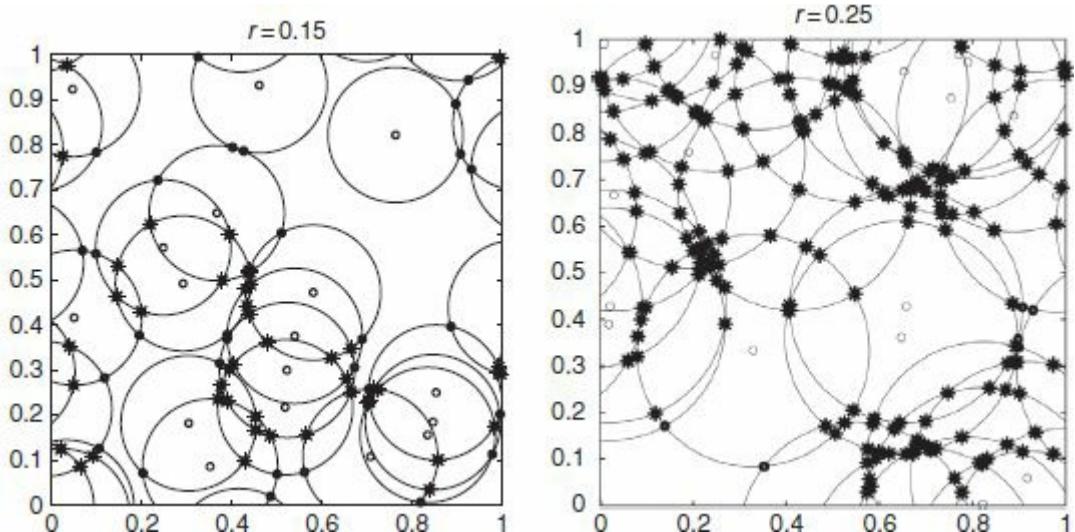


Figure 13.2 Realization of PPP of intensity 20 and circles of radius r around each point. Intersections are marked by \bullet , and covered intersections are marked by $*$. (Left) $r = 0.15$. 61% of the intersections are covered, while the covered fraction is $1 - \exp(-20\pi \cdot 0.15^2) \approx 76\%$. (Right) $r = 0.25$. 97% of the intersections are covered, while the covered fraction is $1 - \exp(-20\pi \cdot 0.25^2) \approx 98\%$.

Equipped with Gilbert's observation, we argue as follows.

1. For a disk radius r , the density of intersections is $4\pi r^2$, since a disk boundary intersects each disk boundary within distance $2r$.
2. The expected number of intersections in the square of area n is $4\pi r^2 n$.
3. To be covered, each intersection needs to have a point of the PPP at distance less than r .
4. The mean number of intersections that are not covered is $4\pi r^2 n e^{-\pi r^2}$, which goes to zero if (13.5) holds. So this condition is also sufficient.

Next we consider the case of multiple coverage.

13.3.2 Multiple coverage

Here we require each location of the plane to be covered k times at least. Gilbert's condition generalizes in a straightforward way to k -coverage: Each intersection of disk boundaries must be covered k times.

Arguing as in the case of single coverage, the necessary and sufficient condition is that the mean number of intersections that are covered $k-1$ (or fewer¹) times, denoted by I_{k-1} , goes to zero:

$$\mathbb{E}(I_{k-1}) \sim 4\pi r^2 n \exp(-\pi r^2) \frac{(\pi r^2)^{k-1}}{\Gamma(k)} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

THEOREM 13.7 (Condition for k -coverage with fixed disks) *The necessary and sufficient condition for $\mathbb{P}(V([0, \sqrt{n}]^2) = 0) \rightarrow 1$ as $n \rightarrow \infty$, i.e., k -coverage, is*

$$\pi r^2 = \log n + k \log \log n + \omega(1).$$

Proof For $\pi r^2 = \log n + k \log \log n + c$, $\mathbb{E}(I_{k-1}) \rightarrow e^{-c} \pi^{-k}$. □

As in the connectivity problem, k -coverage requires only a slightly larger radius than single coverage.

With more effort, it can be shown that, for $\pi r^2 = \log n + k \log \log n + c$,

$$\mathbb{P}(V([0, \sqrt{n}]^2) = 0) \rightarrow \exp\left(-\frac{e^{-c}}{\Gamma(k)}\right), \quad n \rightarrow \infty.$$

The idea behind the proof of this result is the observation that the uncovered regions essentially form their own Poisson process with mean $e^{-c}/\Gamma(k)$. To make these heuristics rigorous, the Chen-Stein method may be used.

13.4 Applications

The classical application of coverage processes in wireless networks is certainly in sensor networks. For this problem, the results established in the previous section are directly applicable. Here we discuss three applications that go beyond Boolean models.

13.4.1 SINR coverage in wireless networks

Let $\Phi = \{x_i\} \subset \mathbb{R}^2$ be a point process, and denote by $I^{!x}(y)$ the interference at location y if $x \in \Phi$ is the desired transmitter, i.e.,

$$I^{!x}(y) = \sum_{z \in \Phi \setminus \{x\}} h_{zy} \ell(\|z - y\|),$$

where h_{zy} is the fading between z and y . We define the SINR at y as

$$\text{SINR}(x, y) = \frac{h_{xy} \ell(\|x - y\|)}{W + I^{!x}(y)}$$

for a thermal noise power W . The coverage cell of transmitter x consists of all locations at which the SINR with x being the desired transmitter is equal to or larger than a threshold θ :

$$S_x = \{y \in \mathbb{R}^2 : \text{SINR}(x, y) \geq \theta\}.$$

Owing to the fading, the cell is very likely to be disconnected. The SINR coverage process is then

$$\Xi_{\text{SINR}} = \bigcup_{x \in \Phi} S_x.$$

If $\theta \geq 1$, each location cannot be covered by more than one transmitter, which does not have to be the closest one.

Noise-free coverage

If Φ is a uniform PPP of intensity λ and the fading is iid Rayleigh, the path loss exponent is α , and $W = 0$, then the coverage probability $p_x(y)$ that location y is covered by transmitter x follows from the success probability in [Section 5.2](#),

$$p_x(y) = \exp(-\lambda\pi\theta^\delta\Gamma(1+\delta)\Gamma(1-\delta)\|x-y\|^2),$$

with $\delta = 2/\alpha$. If $\theta \geq 1$, $p_x(y)$ is the covered area fraction, and we obtain for the mean coverage cell size

$$\mathbb{E}|S| = \int_{\mathbb{R}^2} p_o(y) dy = \frac{1}{\lambda\theta^\delta\Gamma(1+\delta)\Gamma(1-\delta)}.$$

Without noise and fading, the coverage cells converge to the Voronoi cells as the path loss exponent α increases.

Interference-free coverage

If just noise is present, the coverage process, now called an SNR coverage process, is a Boolean model (if Φ is a PPP). Without fading, the SNR cell is simply a disk of radius $(\theta W)^{-1/\alpha}$ since

$$\frac{\|x-y\|^{-\alpha}}{W} \geq \theta \iff \|x-y\| < \frac{1}{(\theta W)^{1/\alpha}}.$$

It follows that $|S| = \pi(\theta W)^{-\delta}$. The noise power and SNR threshold may be replaced by a single parameter. Using a transmit power P is equivalent to changing the noise power to $W' = W/P$. In this case, we can use the results from [Section 13.3](#) to find the conditions on P to keep a large area covered. With Rayleigh fading, we have

$$\mathbb{P}(\text{SNR} \geq \theta) = \mathbb{P}\left(\frac{h\|x-y\|^{-\alpha}}{W} \geq \theta\right) = \exp(-\theta W\|x-y\|^\alpha)$$

and

$$\mathbb{E}|S| = \frac{\pi\Gamma(1+\delta)}{(\theta W)^\delta}.$$

Normalization by the cell size in the non-fading case yields

$$\frac{\mathbb{E}|S|}{\pi(\theta W)^{-\delta}} = \Gamma(1+\delta),$$

which shows that Rayleigh fading *decreases* the mean cell size if $\alpha > 2$ since $\Gamma(1+\delta) < 1$ if $\delta < 1$.

Noise and interference

With Rayleigh fading, the coverage probability in the presence of both noise and interference is

$$p_o(y) = \exp(-\theta W\|y\|^\alpha - \lambda\pi\theta^\delta\Gamma(1+\delta)\Gamma(1-\delta)\|y\|^2)).$$

This is integrable for $\delta = 1/2$, but the resulting expression is unwieldy.

13.4.2 SIR coverage by nearest transmitter

Here we consider a slight modification of the previous model, where the desired transmitter is not given deterministically, but chosen to be the nearest transmitter, which is assumed to be at distance R . This changes the interference significantly, since now the disk $b(o, R)$ is guaranteed to be free of interferers, where the typical receiver is located at o . So the interference is given by

$$I = \sum_{x \in \Phi \setminus b(o, R)} h_x \ell(x)$$

for iid fading random variables h_x . Again we focus on the case where the transmitters form a PPP of intensity λ on the plane and fading is Rayleigh both for the interference and for the desired signal. Noise is ignored.

To find the Laplace transform of the interference given R , we can follow the steps in [Section 5.1.7](#), with the difference that $b(o, R)$ is empty.

$$\begin{aligned}
\mathcal{L}_I(s \mid R) &= \mathbb{E} \prod_{x \in \Phi \setminus b(o, R)} \mathbb{E}_h(e^{-sh\ell(x)}) \\
&= \mathbb{E} \prod_{x \in \Phi \setminus b(o, R)} \frac{s\ell(x)}{1 + s\ell(x)} \\
&\stackrel{(a)}{=} \exp \left(-\lambda \int_{\mathbb{R}^2 \setminus b(o, R)} \frac{1}{1 + s\ell(x)} dx \right) \\
&= \exp \left(-2\pi\lambda \int_R^\infty \frac{rs}{s + r^\alpha} dr \right) \\
&= \exp (\pi s^\delta \Gamma(1 + \delta) \Gamma(1 - \delta) - \pi R^2 H_\delta(R^\alpha/s)),
\end{aligned}$$

where (a) follows from the pgfl, $\delta = 2/\alpha$, and H_δ is the Gauss hypergeometric function

$$H_\delta(x) \triangleq {}_2F_1(1, \delta; 1+\delta; -x).$$

Using H_δ , the “complementary integral” from 0 to R can be expressed as

$$\int_0^R \frac{rs}{s + r^\alpha} dr = \frac{R^2}{2} H_\delta(R^\alpha/s).$$

The probability that o is covered by the nearest transmitter at distance R follows as

$$\begin{aligned}
p_o(R) &= \mathbb{P}(hR^{-\alpha} > \theta I) \\
&= \mathcal{L}_I(\theta R^\alpha) \\
&= \exp \left(-\lambda \pi R^2 (\theta^\delta \Gamma(1 + \delta) \Gamma(1 - \delta) - H_\delta(1/\theta)) \right).
\end{aligned}$$

Since R is the nearest-neighbor distance in the PPP, it is distributed as $f_R(r) = 2\lambda\pi r e^{-\lambda\pi r^2}$, and the coverage probability is

$$\begin{aligned}
p_o &= \mathbb{E}(p_o(R)) \\
&= \frac{1}{1 + \theta^\delta \Gamma(1 + \delta) \Gamma(1 - \delta) - H_\delta(1/\theta)}.
\end{aligned}$$

For $\alpha = 4$, since $H_{1/2}(1/\theta) = \sqrt{\theta} \arctan(1/\sqrt{\theta})$, we obtain the simple closed-form result

$$p_o = \frac{1}{1 + \sqrt{\theta}(\pi/2 - \arctan(1/\sqrt{\theta}))}. \quad (13.6)$$

13.4.3 Secrecy coverage

The secrecy graph, which includes only edges along which secure communication is possible, was introduced in Section 11.4.3. Here we study the corresponding coverage problem.

Base stations and eavesdroppers are distributed randomly on the plane, and the base stations can cover circular areas with radii determined by the distance to the nearest eavesdroppers. While these assumptions result in an analytically tractable model, they are quite realistic. In fact, cellular networks are now undergoing a major transition from carefully planned base station deployments to an irregular heterogeneous infrastructure that includes micro-base stations and femtocells. Even without such small base stations, it can be argued that the results obtained from a random model are as accurate as, or better than, those obtained from a lattice-based model for the base stations, see Figs. 7.1 and 7.2.

The reason why a base station can cover the circular area up to the nearest eavesdropper is that, given the distance to the furthest legitimate node within the disk, it can always choose a positive rate of transmission such that the secrecy capacity is positive. The locations of the eavesdroppers remain uncovered, i.e., the disks are open disks. The main question is what density of eavesdroppers can be accommodated while still guaranteeing that the entire area or volume of interest is covered securely? This would ensure that mobile stations could roam around everywhere and be reached securely by a base station. Hence the downlink is intrinsically secure, while the uplink (from the mobile to the base station) may have to be secured by transmission of a one-time pad via the downlink.

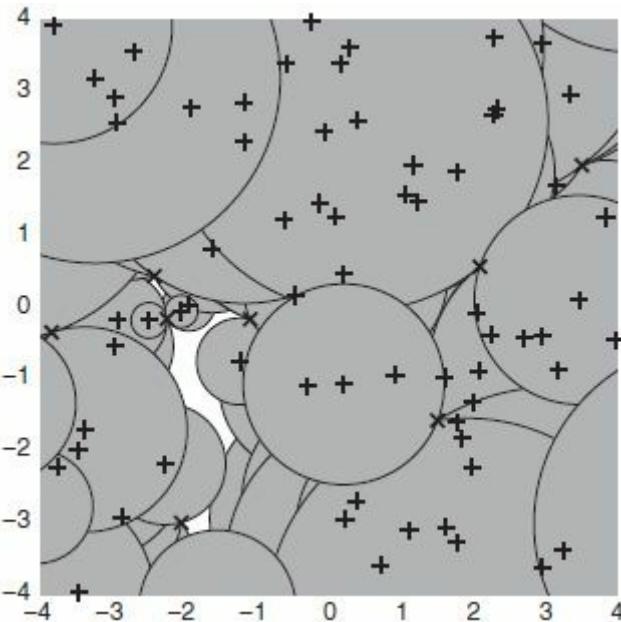


Figure 13.3 Example for secrecy coverage of an 8×8 square for eavesdropper density $\lambda = 0.1$. The base stations are marked by +, and the eavesdroppers by \times . The covered area is shaded gray.

We assume that the base stations and eavesdroppers form independent Poisson point processes of intensities 1 and λ , respectively, in \mathbb{R}^d , denoted by Φ and Ψ . An example for $\lambda = 0.1$ is shown in Fig. 13.3.

This coverage model behaves quite differently from the corresponding model

with *independent* disk radii with the same distribution. If the radii were independent, we know from (13.5) that

$$\mathbb{E}(\pi R^2) = (1 + \epsilon) \log n$$

would be sufficient for asymptotic coverage of a square of area n for any $\epsilon > 0$. In the secrecy coverage case, $\mathbb{E}(\pi R^2) = 1/\lambda$, so, if the radii were independent,

$$\lambda = [(1 + \epsilon) \log n]^{-1}$$

would be sufficient and λ would have to decrease only logarithmically in n for coverage. Owing to the dependence (one eavesdropper may determine the radius of several nearby coverage disks), λ has to decrease significantly faster, at a rate of about $n^{-1/d}$.

Secrecy coverage in one dimension

We determine the covered fraction of the line $\mathbb{P}(o \in \Xi) = 1 - \mathbb{E}\chi(o)$.

THEOREM 13.8 (Secrecy coverage in one dimension) *The covered line fraction is*

$$\mathbb{P}(o \in \Xi) = \frac{1 + 4\lambda}{(1 + 2\lambda)^2}.$$

Proof Let L be the event that o is covered by a point lying to the left of o , and let R be the event that o is covered by a point lying to the right of o . Then L and R are independent, and

$$\mathbb{P}(o \in \Xi) = 1 - (1 - \mathbb{P}(L))(1 - \mathbb{P}(R)) = 1 - (1 - \mathbb{P}(R))^2 = 2\mathbb{P}(R) - \mathbb{P}(R)^2$$

by symmetry. The event R occurs if, for some $t > 0$, the closest point of Φ (a legitimate node) is at distance t and there are no points of Ψ (eavesdroppers) in $[0, 2t]$. Hence

$$\mathbb{P}(R) = \int_0^\infty e^{-t} e^{-2\lambda t} dt = \frac{1}{2\lambda + 1}.$$

□

So the expected vacancy of an interval of length n is

$$\mathbb{E}V([0, n]) = n\mathbb{E}\chi(o) = n(1 - \mathbb{P}(o \in \Xi)) = \frac{4\lambda^2 n}{(1 + 2\lambda)^2}.$$

To achieve coverage of the real line ($n \rightarrow \infty$), we thus need $\lambda^2 n \rightarrow 0$. A more detailed analysis reveals that $\mathbb{P}(V=0) \sim e^{-4n\lambda^2}$ if $\lambda n \rightarrow \infty$, so the scaling condition

for coverage is indeed $\lambda = o(n^{-1/2})$. If $n\lambda^2 \rightarrow \infty$, $\mathbb{P}(V=0) \rightarrow 0$; and if $n\lambda^2 \rightarrow 0$, $\mathbb{P}(V=0) \rightarrow 1$.

Secrecy coverage in two dimensions

The two-dimensional case is more involved. The natural conjecture is that the critical scaling is $\lambda = o(n^{-1/3})$. Indeed, it can be shown that, if $n\lambda^3 \rightarrow \infty$, $\mathbb{P}(V=0) \rightarrow 0$. For the converse, however, the best known result is

$$n(\log n)^3 \lambda^3 \rightarrow \quad \Rightarrow \quad \mathbb{P}(V=0) \rightarrow 1,$$

corresponding to $\lambda = o(n^{-1/3}(\log n)^{-1})$, so there is a logarithmic gap.

13.4.4 Sentry selection

The problem of *sentry selection* is inspired by a practical issue in energy-efficient sensor networking. If a region of interest \mathcal{R} is known to be k -covered (each $y \in \mathcal{R}$ is covered by at least k sensors), can the k sensors be colored with k colors in such a way that \mathcal{R} is covered by the subset of sensors with a given color? If this is the case, then time can be divided into frames, such that in frame i only the nodes of color $1 + \text{mod}(i, k)$ are active, while the others can be put to sleep to conserve energy, while still guaranteeing coverage all the time.

Since k -coverage is achievable at little additional cost (small increase in sensing radius r – see Theorem 13.7), significant energy gains can be expected – provided that the sensors providing the k -cover can be partitioned into k single covers.

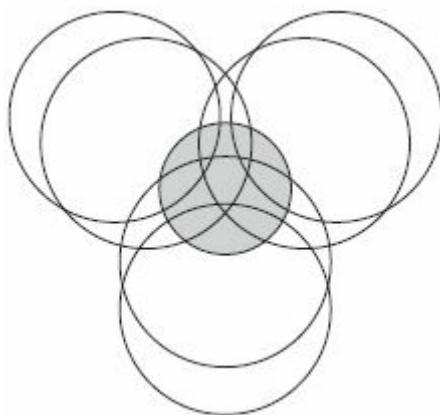


Figure 13.4 Example for an “obstacle” configuration for sentry selection. The gray region is 2-covered, but the six disks providing the two-coverage cannot be colored with two colors so that the disks of each color both provide single coverage.

Clearly, this is not possible in general. It is easy to construct a counter-example of a k -cover that is not partitionable.

Example 13.1 Let $n = 4$, $k = 2$. The set of subsets of $\{1, 2, 3, 4\}$ of size $k = 2$ has $\binom{4}{2} = 6$ elements:

$$S = \left\{ \{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\} \right\}.$$

Then let

$$\begin{aligned} S_1 &= \left\{ \{1, 2\}, \{1, 3\}, \{1, 4\} \right\}, \\ S_2 &= \left\{ \{1, 2\}, \{2, 3\}, \{2, 4\} \right\}, \\ S_3 &= \left\{ \{1, 3\}, \{2, 3\}, \{3, 4\} \right\}, \\ S_4 &= \left\{ \{1, 4\}, \{2, 4\}, \{3, 4\} \right\}. \end{aligned}$$

Each element of S is an element in exactly $k = 2$ sets S_i , but there is no way to partition S_i into two single covers.

The concrete setup is that of a PPP of intensity 1 of sensors and an area of interest $\mathcal{R} = [0, \sqrt{n}]^2$. Assume each sensor can cover a disk of radius r , and let Ξ be the resulting Boolean model. Let C_k be the event that \mathcal{R} is k -covered. We know that k -coverage is achieved asymptotically if $\pi r^2 = \log n + k \log \log n + \omega(1)$. Let P_k be the event that Ξ is k -partitionable. Then

$$\mathbb{P}(C_k \setminus P_k) \leq \frac{c_k}{\log n}$$

such that, if $\mathbb{P}(C_k) = \Theta(1)$, the conditional probability $\mathbb{P}(P_k | C_k) = 1 - o(1)$ as $n \rightarrow \infty$. The proof is rather lengthy. It has to be shown that “obstacles” to k -partitionability, such as the one shown in Fig. 13.4, occur with probability going to zero as $n \rightarrow \infty$. So asymptotically the k -cover can be partitioned a.s., and the proposed sentry selection scheme works.

A simpler way to prove the existence of a coloring is to show that random coloring achieves the partitioning with positive probability. This is an application of the *probabilistic method*. This method, together with the Lovász local lemma, yields the desired result if the level of coverage is $3k \log \log n$ everywhere.

This proof technique is non-constructive, i.e., it leaves open the (interesting) question of how to actually find the coloring. In practice, sensor nodes should choose a color by using solely local knowledge.

Bibliographical notes

A detailed introduction to coverage processes is given in the monograph by Hall (1985). The Boolean model is analyzed in Chapter 4 of Stoyan *et al.* (1995) and in Chapters 3–5 of Meester & Roy (1996). The probabilistic method is the subject of Alon & Spencer (2008).

SINR coverage is discussed in Part II of Baccelli & Blaszczyszyn (2009). The case of nearest-transmitter coverage is analyzed in detail in Andrews *et al.* (2011) and extended to multi-tier heterogeneous networks in Dhillon *et al.* (2012).

The problem of secrecy coverage is introduced and studied in Sarkar & Haenggi (2012), and the main result on sentry selection is presented in Balister *et al.* (2010).

Problems

13.1 Let Φ_1, Φ_2, \dots be independent uniform PPPs of intensity λ on \mathbb{R} , and let $1 \geq t_1 \geq t_2, \dots$ be a sequence of positive numbers converging to 0. For

$$\Xi = \bigcup_{i \in \mathbb{N}} \bigcup_{x \in \Phi_i} (x, x + t_i),$$

where $(x, x + t_i) \subset \mathbb{R}$ is the open interval from x to $x + t_i$, let $V = V(\mathbb{R}) = |\mathbb{R} \setminus \Xi|$ be the vacancy of this model. The uncovered set $\mathbb{R} \setminus \Xi$ is called the *random Cantor set*. Find the probability that a location $y \in \mathbb{R}$ is not covered by Ξ , and give the (necessary and sufficient) condition on the sequence (t_i) for $\mathbb{P}(V = 0)$.

13.2 For single coverage of the Boolean model with fixed disks of radius r , find the condition on $r(n)$ such that, for the vacancy $V = V([0, \sqrt{n}]^2)$, the mean goes to 0 but the probability of $V = 0$ is not 1, i.e.,

$$\mathbb{E}(V) \rightarrow 0 \quad \text{but} \quad \mathbb{P}(V = 0) \neq 1.$$

Is there an $r(n)$ such that $\mathbb{E}(V) \rightarrow 0$ and $\mathbb{P}(V = 0) = 0$?

13.3 The *Stienen model* is a two-dimensional coverage process where the germs form a uniform PPP and the grains are disks with radius equal to half the nearest-neighbor distance. Produce a realization of the Stienen model on a square $[0, 5]^2$ with the PPP having intensity 1. Determine the covered area fraction of the model.

13.4 In the standard Boolean model (with the germs forming a PPP), the giant component is unique if it exists. Show by counter-example that this is not necessarily true in models where the vertices form a stationary lattice.

13.5 For the nearest-transmitter coverage problem discussed in Section 13.4.2, find the coverage probability for $\alpha = 4$, as in (13.6), but with noise.

13.6 Secrecy coverage in one dimension. Consider a typical interarrival interval

between two eavesdroppers. Find the probability that this interval is covered as a function of the intensity of the eavesdropper PPP.

Hint. Show first that the coverage of the interval depends only on the location of two points near the midpoint of the interval.

Appendix Introduction to R

A.1 Overview

R is a software package for statistical computation and graphics. It is free and open-source and very flexible. It implements many modern statistical methods with good graphical support. The core of R is an interpreted computer language which allows branching and looping as well as modular programming using functions. It can be downloaded for different operating systems from the main website <http://www.r-project.org/>, which also contains links to manuals and documentation. For example, a relatively concise introduction to R is available at <http://cran.r-project.org/doc/manuals/R-intro.pdf>.

The R language is enhanced by numerous libraries. For the purposes of spatial statistics, the spatstat library has been developed. It supports the following:

Creation, manipulation, and plotting of point patterns

Exploratory data analysis, including intensity maps, empty space functions,

nearest-neighbor distance functions, and K , L , and J functions

Model fitting

All figures with point patterns in this book were created using the powerful commands of the spatstat library. To use it, it needs to be loaded after starting R by typing `library(spatstat)`. A demo can be run by `demo(spatstat)`, while typing `help(spatstat)` provides a help page. An introduction to spatstat is given in Baddeley & Turner (2005), and further documentation is available at www.spatstat.org.

R is equipped with all standard data types, such as scalars, vectors, arrays, and lists. Additional spatstat data types are point patterns and windows. A point pattern records the locations of the points observed in a certain region, and a window is a region in a two-dimensional space representing the “study area.” The point pattern is an object of the class `ppp`, which stands for planar point pattern, not Poisson point process. Each point pattern must have a window associated with it that describes the spatial region in which the points are observed. This window is of

type `owin`, which stands for observation window. For example,

```
x=runif(20,max=10)
y=runif(20,max=3)
u=ppp(x,y,c(0,10),c(0,3))
```

creates a `ppp` object `u` with a pattern of 20 random points uniform on the window $[0, 10] \times [0, 3]$. The command `c` is used to create vectors, e.g., `c(2:4,6,1)` denotes the vector $(2, 3, 4, 6, 1)$. Typing just `u` reveals basic information about the object, while `summary(u)` also gives the intensity. The coordinates of the points can be retrieved by `coords(u)`, or by `u$x` and `u$y`, respectively. To change the coordinates, the `coords` command should be used, rather than assigning values to `u$x` or `u$y` directly.

Marks can be added to and removed from point patterns using the functions `marks` and `unmark` or the operator `%mark%`. For example,

```
X=rpoispp(100)
marks(X)=rexp(X$n)
```

adds independent exponentially distributed marks to a PPP of intensity 100 on the unit square. Alternatively, `X=X %mark% M` can be used, where `M` is a vector of marks. A histogram of the marks is obtained by `hist(X$marks,20)`.

`spatstat` provides a number of point patterns that can be used for statistical analysis. For example, the `longleaf` marked point pattern is loaded by `data(longleaf)`. It includes 584 locations of longleaf pine trees on a 200×200 window, with their diameter as marks.

A.2 List of commands

Table A.1 lists some of the most common `spatstat` commands. In some commands, if no window is explicitly specified, the default window $[0, 1]^2$ is used. Note that R is case-sensitive.

Some commands accept functions as arguments. An important case is the `rpoispp` command, which produces Poisson point patterns. For example, Fig. 2.2 was created by

```
win=square(20)
X=rpoispp(function(x,y) {2*exp(-y/6)*(1+sin(x))},3,win)
```

and a Neyman–Scott process may be produced by

```

cfun=function(x,y,radius,n) {runifdisc(n, radius, centre=c(x,y))}
X=rNeymanScott(2,0.2,cfun,radius=0.2,n=5,square(4))

```

Here the parent PPP has intensity 2 on $[0, 4]^2$, the radius of a cluster is 0.2, and there are five points per cluster.

Table A.1 List of some useful commands to produce and analyze point processes.
Here X refers to a point process object (of type ppp)

summary(X)	summary information about X
X\$n or npoints(X)	number of points in X
X>window	window pertaining to X
X[c(1,10)]	point pattern containing the first 10 points of X
coords(X)	extracts coordinates of points
marks(X)=M	assigns marks M to X
is.marked(X)	is TRUE if X is a marked point pattern
unmark(X)	ground process of X
unique(X)	remove duplicate points from X
win=owin(c(-10,10),c(-10,10))	defines window $[-10, 10]^2$
win=square(10)	defines window $[0, 10]^2$
X=clickppp(10)	lets user enter 10 points manually
superimpose(X,Y)	union of point patterns X and Y
shift(X,c(-2,0))	shift X (and its window) by vector $(-2, 0)$
rsyst(win,10,10)	randomly translated lattice with 10×10 points
runifpoint(20)	20 uniformly distributed points on $[0, 1]^2$
rpoispp(100)	PPP on $[0, 1]^2$ with intensity 100 on default window $[0, 1]^2$
rMatClust(0.1,1,8,win)	Matérn cluster process on win with cluster density 0.1, cluster radius 1, and eight nodes per cluster on average
rNeymanScott(0.1,1,rfun,win)	Neyman–Scott cluster process with cluster density 0.1, maximum cluster radius 1, and random cluster function rfun
rGaussPoisson(0.1,0.2,0.4,win)	Gauss–Poisson process with cluster density 0.1, cluster diameter 0.2, and probability 0.4 that a cluster has two points
rMaternI(5,1/3,win)	Matérn hard-core process of type I with $\lambda_b = 5$, $r = 1/3$
rMaternII(5,1/3,win)	Matérn hard-core process of type II with $\lambda_b = 5$, $r = 1/3$
rcell(nx=15)	Baddeley–Silverman cell process with 15^2 cells
rjitter(X,1/3)	randomly perturb X uniformly on disk of radius $1/3$
nndist(X)	nearest-neighbor distances

<code>pairdist(X)</code>	distances between all pairs of points
<code>quadratcount(X,nx=4,ny=5)</code>	quadrat count for 4×5 rectangles
<code>plot(X,main="title",xlab="x")</code>	plots points of X with title “title”
<code>plot(Fest(X))</code>	plots estimate of the empty space function of X
<code>plot(Gest(X))</code>	plots estimate of the nearest-neighbor distance of X
<code>plot(Kest(X))</code>	plots estimate of the K function of X
<code>plot(Lest(X))</code>	plots estimate of the L function of X
<code>plot(Jest(X))</code>	plots estimate of the J function of X
<code>plot(allstats(X))</code>	plots the Fest, Gest, Kest, and Jest functions
<code>plot(pcf(X))</code>	plots estimate of pair correlation function of X
<code>plot(delaunay(X))</code>	plots Delaunay triangulation
<code>plot(dirichlet(X))</code>	plots Voronoi(–Dirichlet) tessellation
<code>plot(density(X,0.2))</code>	plots density estimate for X with resolution 0.2
<code>fryplot(X)</code>	plots Fry plot
<code>frypoints(X)</code>	extracts point pattern pertaining to Fry plot
<code>plot(envelope(X,Kest,nsim=99))</code>	plots estimated K function of X together with the envelope of 99 K functions of PPPs

References

- Alon, N. & Spencer, J. (2008), *The Probabilistic Method*, 3rd edn., Wiley.
- Andrews, J. G., Baccelli, F. & Ganti, R. K. (2011), A tractable approach to coverage and rate in cellular networks, *IEEE Transactions on Communications* **59**(11), 3122–3134.
- Athreya, K. B. & Ney, P. E. (2004), *Branching Processes*, Dover Publications.
- Baccelli, F. & Blaszczyszyn, B. (2009), *Stochastic Geometry and Wireless Networks*, NOW.
- Baccelli, F., Blaszczyszyn, B. & Mirsadeghi, M. O. H. (2011), Optimal paths on the space–time SINR random graph, *Advances in Applied Probability* **43**(1), 131–150.
- Baccelli, F., Blaszczyszyn, B. & Mühlethaler, P. (2006), An ALOHA protocol for multihop mobile wireless networks, *IEEE Transactions on Information Theory* **52**(2), 421–436.
- Baddeley, A. J. & Silverman, B. W. (1984), A cautionary example on the use of secondorder methods for analyzing point patterns, *Biometrics* **40**(4), 1089–1093.
- Baddeley, A. & Turner, R. (2005), Spatstat: an R package for analyzing spatial point patterns, *Journal of Statistical Software* **12**(6), 1–42.
- Baddeley, A. J. & van Lieshout, M. N. M. (1995), Area-interaction point processes, *Annals of the Institute of Statistical Mathematics* **47**(4), 601–619.
- Baddeley, A., Bárány, I., Schneider, R. & Weil, W. (2007), *Stochastic Geometry*, Springer.
- Balister, P., Bollobás, B., Sarkar, A. & Walters, M. (2005), Connectivity of random k -nearest-neighbor graphs, *Advances in Applied Probability* **37**, 1–24.
- Balister, P., Bollobás, B., Sarkar, A. & Walters, M. (2010), Sentry selection in wireless networks, *Advances in Applied Probability* **42**(1), 1–25.
- Balister, P., Bollobás, B. & Walters, M. (2004), Continuum percolation with steps in an annulus, *Annals of Applied Probability* **14**, 1869–1879.
- Balister, P., Bollobás, B. & Walters, M. (2005), Continuum percolation with steps

- in the square or the disc, *Random Structures and Algorithms* **26**(4), 392–403.
- Barbour, A. D. & Chen, L. H. Y. (2005), *Stein's Method and Applications*, Singapore University Press.
- Bedford, T. & van den Berg, J. (1997), A remark on the Van Lieshout and Baddeley J function for point processes, *Advances in Applied Probability* **29**(1), 19–25.
- Bollobás, B. (2001), *Random Graphs*, 2nd edn., Cambridge University Press.
- Bollobás, B. & Riordan, O. (2006), *Percolation*, Cambridge University Press.
- Broadbent, S. & Hammersley, J. (1957), Percolation processes I. Crystals and mazes, *Proceedings of the Cambridge Philosophical Society* **53**, 629–641.
- Campbell, N. (1909), The study of discontinuous phenomena, *Mathematical Proceedings of the Cambridge Philosophical Society* **15**, 117–136.
- Cressie, N. (1993), *Statistics for Spatial Data*, revised edn., Wiley.
- Daley, D. J. & Vere-Jones, D. (2003), *An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods*, 2nd edn., Springer.
- Daley, D. J. & Vere-Jones, D. (2008), *An Introduction to the Theory of Point Processes, Volume II: General Theory and Structure*, 2nd edn., Springer.
- Dhillon, H. S., Ganti, R. K., Baccelli, F. & Andrews, J. G. (2012), Modeling and analysis of K -tier downlink heterogeneous cellular networks, *IEEE Journal on Selected Areas in Communications* **30**(3), 550–560.
- Dousse, O., Franceschetti, M., Macris, N., Meester, R. & Thiran, P. (2006), Percolation in the signal-to-interference ratio graph, *Journal of Applied Probability* **43**(2), 552–562.
- Durrett, R. (1985), Some general results concerning the critical exponents of percolation, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **69**, 421–437. “
- Fekete, M. (1923), Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten, *Mathematische Zeitschrift* **17**(1), 228–249.
- Fortuin, C. M., Kasteleyn, P. W. & Ginibre, J. (1971), Correlation inequalities on some partially ordered sets, *Communications in Mathematical Physics* **22**(2), 89–103.
- Franceschetti, M. & Meester, R. (2007), *Random Networks for Communication: From Statistical Physics to Information Systems*, Cambridge University Press.
- Franceschetti, M., Booth, L., Cook, M., Bruck, J. & Meester, R. (2005), Continuum percolation with unreliable and spread out connections, *Journal of Statistical Physics* **118**(3/4), 721–734.
- Ganti, R. K. & Haenggi, M. (2007), Dynamic connectivity and packet propagation delay in ALOHA wireless networks, in *41st Asilomar Conference on Signals, Systems, and Computers (Asilomar '07)*, Pacific Grove, CA.
- Ganti, R. K. & Haenggi, M. (2009a), Interference and outage in clustered wireless ad hoc networks, *IEEE Transactions on Information Theory* **55**(9), 4067–4086.

- Ganti, R. K. & Haenggi, M. (2009b), Spatial and temporal correlation of the interference in ALOHA ad hoc networks, *IEEE Communications Letters* **13**(9), 631–633.
- Ganti, R. K. & Haenggi, M. (2012), Dynamic connectivity and path formation time in Poisson networks, *The Journal of Mobile Communication, Computation and Information*, <http://www.nd.edu/~mhaenggi/pubs/winet12.pdf>.
- Ganti, R. K., Andrews, J. G. & Haenggi, M. (2011), High-SIR transmission capacity of wireless networks with general fading and node distribution, *IEEE Transactions on Information Theory* **57**(5), 3100–3116.
- Georgii, H.-O. (1976), Canonical and grand canonical Gibbs states for continuum systems, *Communications in Mathematical Physics* **48**(1), 31–51.
- Giacomelli, R., Ganti, R. K. & Haenggi, M. (2011), Outage probability of general ad hoc networks in the high-reliability regime, *IEEE/ACM Transactions on Networking* **19**(4), 1151–1163.
- Grandell, J. (1976), *Doubly Stochastic Poisson Processes*, Springer.
- Grimmett, G. (1999), *Percolation*, 2nd edn., Springer.
- Haenggi, M. (2008a), A geometric interpretation of fading in wireless networks: theory and applications, *IEEE Transactions on Information Theory* **54**(12), 5500–5510.
- Haenggi, M. (2008b), The secrecy graph and some of its properties, in *2008 IEEE International Symposium on Information Theory* (ISIT '08), Toronto.
- Haenggi, M. (2009), Outage, local throughput, and capacity of random wireless networks, *IEEE Transactions on Wireless Communications* **8**(8), 4350–4359.
- Haenggi, M. (2011), Mean interference in hard-core wireless networks, *IEEE Communications Letters* **15**(8), 792–794.
- Haenggi, M. & Ganti, R. K. (2008), *Interference in Large Wireless Networks*, NOW.
- Hall, P. (1985), On the coverage of k -dimensional space by k -dimensional spheres, *The Annals of Probability* **13**(3), 991–1002.
- Harris, T. E. (1960), A lower bound on the critical probability in a certain percolation process, *Mathematical Proceedings of the Cambridge Philosophical Society* **56**, 13–20.
- Heinrich, L. & Schmidt, V. (1985), Normal convergence of multidimensional shot noise and rates of this convergence, *Advances in Applied Probability* **17**(4), 709–730.
- Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. (2008), *Statistical Analysis and Modeling of Spatial Point Patterns*, Wiley.
- Ilow, J. & Hatzinakos, D. (1998), Analytical alpha-stable noise modeling in a Poisson field of interferers or scatterers, *IEEE Transactions on Signal Processing* **46**(6), 1601–1611.
- Isham, V. (1984), Multitype Markov point processes: some approximations,

- Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **391**, 39–53.
- Jacobsen, M. (2006), *Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes*, Birkhäuser.
- Janson, S., Luczak, T. & Rucinski, A. (2000), *Random Graphs*, Wiley-Interscience.
- Kesten, H. (1980), The critical probability of bond percolation on the square lattice equals 1/2, *Communications in Mathematical Physics* **74**(1), 41–59.
- Kingman, J. F. C. (1978), Uses of exchangeability, *The Annals of Probability* **6**(2), 183–197.
- Kingman, J. F. C. (1993), *Poisson Processes*, Oxford Science Publications.
- Lee, C.-H. & Haenggi, M. (2012), Interference and outage in Poisson cognitive networks, *IEEE Transactions on Wireless Communications* **11**(4), 1392–1401.
- Lindvall, T. (2002), *Lectures on the Coupling Method*, Dover Publications.
- Linnartz, J.-P. M. G. (1992), Exact analysis of the outage probability in multiple-user radio, *IEEE Transactions on Communications* **40**(1), 20–23.
- Lowen, S. B. & Teich, M. C. (1990), Power-law shot noise, *IEEE Transactions on Information Theory* **36**(6), 1302–1318.
- Matérn, B. (1986), *Spatial Variation*, 2nd edn., Springer.
- Mathai, A. M. (1999), *An Introduction to Geometrical Probability*, Gordon and Breach Science Publishers.
- Mathar, R. & Mattfeldt, J. (1995), On the distribution of cumulated interference power in Rayleigh fading channels, *Wireless Networks* **1**(1), 31–36.
- Matthes, K., Kerstan, J. & Mecke, J. (1978), *Infinitely Divisible Point Processes*, John Wiley & Sons.
- Meester, R. & Roy, R. (1996), *Continuum Percolation*, Cambridge University Press.
- Musa, S. & Wasylkiwskyj, W. (1978), Co-channel interference of spread spectrum systems in a multiple user environment, *IEEE Transactions on Communications* **26**(10), 1405–1413.
- Newman, D. S. (1970), A new family of point processes which are characterized by their second moment properties, *Journal of Applied Probability* **7**(2), 338–358.
- Newman, M. E. J. & Ziff, R. M. (2000), Fast Monte Carlo algorithm for site or bond percolation, *Physical Review Letters* **85**(19), 4104–4107.
- Neyman, J. & Scott, E. L. (1952), A theory of the spatial distribution of galaxies, *Astrophysical Journal* **116**, 144–163.
- Neyman, J. & Scott, E. L. (1958), Statistical approach to problems of cosmology, *Journal of the Royal Statistical Society, Series B* **20**(1), 1–43.
- Nguyen, X. X. & Zessin, H. (1979), Integral and differential characterization of the Gibbs process, *Mathematische Nachrichten* **88**, 105–115.
- Peierls, R. (1936), On Ising's model of ferromagnetism, *Mathematical*

- Proceedings of the Cambridge Philosophical Society* **32**, 477–481.
- Penrose, M. (2003), *Random Geometric Graphs*, Oxford University Press.
- Rényi, A. (1967), Remarks on the Poisson process, *Studia Scientifica Mathematica Hungarica* **2**, 119–123.
- Rice, S. O. (1944), Mathematical analysis of random noise, *Bell System Technical Journal* **23**, 282–332.
- Ripley, B. D. (1976), The second-order analysis of stationary point processes, *Journal of Applied Probability* **13**, 255–266.
- Samorodnitsky, G. & Taqqu, M. S. (1994), *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman & Hall.
- Sarkar, A. & Haenggi, M. (2011), Percolation in the secrecy graph. ArXiv, <http://arxiv.org/abs/1107.4613>.
- Sarkar, A. & Haenggi, M. (2012), Secrecy coverage, *Internet Mathematics*. Available at <http://www.nd.edu/~mhaenggi/pubs/im12.pdf>.
- Schottky, W. (1918), Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern, *Annalen der Physik* **57**, 541–567.
- Sousa, E. S. (1990), Interference modeling in a direct-sequence spread-spectrum packet radio network, *IEEE Transactions on Communications* **38**(9), 1475–1482.
- Sousa, E. S. & Silvester, J. A. (1990), Optimum transmission ranges in a direct-sequence spread-spectrum multihop packet radio network, *IEEE Journal on Selected Areas in Communications* **8**(5), 762–771.
- Spencer, J. (2001), *The Strange Logic of Random Graphs*, Springer.
- Srinivasa, S. & Haenggi, M. (2010), Distance distributions in finite uniformly random networks: theory and applications, *IEEE Transactions on Vehicular Technology* **59**(2), 940–949.
- Stoyan, D., Kendall, W. S. & Mecke, J. (1995), *Stochastic Geometry and its Applications*, 2nd edn., John Wiley & Sons.
- Takagi, H. & Kleinrock, L. (1984), Optimal transmission ranges for randomly distributed packet radio terminals, *IEEE Transactions on Communications* **32**(3), 246–257.
- van Lieshout, M. N. M. (2000), *Markov Point Processes and their Applications*, Imperial College Press.
- van Lieshout, M. N. M. (2011), A J -function for inhomogeneous point processes, *Statistica Neerlandica* **65**(2), 183–201.
- van Lieshout, M. N. M. & Baddeley, A. J. (1996), A nonparametric measure of spatial interaction in point patterns, *Statistica Neerlandica* **50**(3), 344–361.
- Walters, M. (2012), Small components in k -nearest neighbour graphs, *Discrete Applied Mathematics*.
- Weber, S. & Andrews, J. G. (2012), *Transmission Capacity of Wireless Networks*, NOW.

- Weber, S., Yang, X., Andrews, J. G. & de Veciana, G. (2005), Transmission capacity of wireless ad hoc networks with outage constraints, *IEEE Transactions on Information Theory* **51**(12), 4091–4102.
- Westcott, M. (1972), The probability generating functional, *Journal of the Australian Mathematical Society* **14**, 448–466.
- Zorzi, M. & Pupolin, S. (1995), Optimum transmission ranges in multihop packet radio networks in the presence of fading, *IEEE Transactions on Communications* **43**(7), 2201–2205.

Index

- abstract random graph, 230
- ALOHA, 38, 106
- area-interaction process, 64, 69
- asymptotic notation, 5
- atom, 12
- Baddeley–Silverman process, 135
- basic event, 25
- Bethe lattice, *see* regular tree
- binomial point process, 20, 48, 126, 168
 - mixed, 48
 - uniform, 20
- binomial random graph, *see* Gilbert random graph
- birth–death process, 71
- BK inequality, 202
- bond percolation
 - coupling, 191
 - critical probability, 211
 - model, 179
 - probability space, 190
 - simple bounds, 193
- Boolean model, 257
- Borel σ -algebra, 6
- BPP, *see* binomial point process
- branching process, 181, 182, 243
 - offspring distribution, 182
 - total progeny, 219
- branching tree, 181
- Campbell measure, 160, 162
 - for marked point processes, 172

reduced, 165, 166, 170
Campbell's theorem for Poisson point processes, 83, 87, 257
Campbell's theorem for sums, 78, 93, 95, 109, 121, 127, 130, 239, 257
 for marked point processes, 144
 for stationary iid marked processes, 145
 for stationary marked processes, 145
 over stationary point processes, 79
Campbell–Mecke theorem, 158, 170
 for general point processes, 166
 for stationary marked point processes, 172
 for stationary point processes, 158
 reduced version, 166
canonical ensemble, 65
canonical space, 25
capacity functional, 29
causal path, 235
characteristic functional, 84
Chen–Stein method, 247, 263
clump, 256
cluster process, 53
 interference, 173
complete spatial randomness, 47
component, 180
 of coverage process, 256
 of random graph, 180
conditional property
 of Poisson point processes, 31
connected component, 180
connective constant, 197
connectivity
 disk graph, 249
 of Gilbert random graph, 252
 random lattice, 246
connectivity function, 221
contact distance, 38
contact distribution function, 39
continuum percolation, 221, 223
counting measure, 10, 12
 for randomly translated lattice, 23
 measurable decomposition, 12
coverage, 255

definition, 256
secrecy, 266
SINR, 263
covering graph, 215
Cox process, 50
intensity field, 51
critical probability, 180
of mean cluster size, 188
tree, 187

Delaunay triangulation, 22
density
with respect to Lebesgue measure, 61
with respect to Poisson point process distribution, 61
Descartes' sign rule, 185
die point process, 14, 152
die process, 92
Dirac measure, 12, 141
disjoint occurrence, 201
disk graph, 222
connectivity, 249
multiple connectivity, 250
percolation, 223
displacement kernel, 35
displacement theorem, 35
dual graph, 194
dual lattice, 194

effective area, 221
energy function, 67
Erdős–Rényi random graph, 230
ergodic point process, 33
excess interference ratio, 174
exchangeable random variables, 113
exit delay, 237
expectation
with respect to Palm measure, 154
with respect to reduced Palm measure, 165
exterior conditioning, 152
extinction probability, 184

face percolation, 224
factorial moment measure, 122
finite-dimensional distribution, 29
first-passage percolation, 236
FKG inequality, 199
Fry plot, 40

Galton–Watson branching process, 182
Gauss–Poisson process, 54, 92, 171
generating function, 79
Georgii–Nguyen–Zessin formula, 167
germ–grain model, 53, 142, 255
Gibbs process, 60
 conditional, 65
 distribution, 65
 definition, 60
 energy function, 67
 general finite, 67
 likelihood function, 60
 pair potential, 67
 partition function, 67
 rejection sampling, 71
 simulation, 71
Gilbert random graph, 230
 connectivity, 252
Gilbert’s disk graph, 222
grand canonical ensemble, 67
ground process, 138, 141

hard-core point process, 57, 174
Harris’ lemma, *see* FKG inequality
Harris–Kesten theorem, 212
hereditary function, 60
higher-order moment measure, 122
homogeneous independent clustering, 54

increasing event, 198
infinite divisibility, 81
intensity measure, 13
interference, 93
 as shot noise, 93

correlation, 111
distribution with fading, 102
distribution without fading, 97
in cluster processes, 173
in hard-core processes, 174
in type I Matérn hard-core processes, 174
joint Laplace transform, 113
Laplace transform, 102, 103
mean, 93
interference field, 73
interior conditioning, 152
isotropy, 32

J function, 39, 168

Janossy density, 49

Janossy measure, 49

K function, 131, 170

of Matérn hard-core process, 133

of Neyman–Scott process, 133

of uniform Poisson point process, 132

Kolmogorov's zero–one law, 180, 192

L function, 135

Laplace functional, 82

Laplace transform, 81

of interference, 102, 103

connection with outage, 105

of stable random variable, 100

lattice cluster process, 54

Lebesgue measure, 7, 12

LePage series, 100

local delay, 236

mapping theorem, 36

mark distribution, 139, 145

for stationary point processes, 144

mark-correlation function, 147

marked point process, 138, 140

iid marked, 139

independently marked, 139

stationary, 144
marking theorem, 145
Markov point process, 71
Matérn cluster process, 57, 127
Matérn hard-core process, 58
 K function, 133
 mean interference, 174
 type I, 58, 174
 type II, 58
mean cluster size, 188
mean interference, 93
 in cluster processes, 173
 in hard-core processes, 174
measurability, 6
measurable decomposition, 12
measure, 7
 atomic, 12
 diffuse, 7
 Dirac, 12
 intensity, 13
 Janossy, 49
 Lebesgue, 7, 119
 Radon, 12
Mecke's theorem, 169
metrical transitivity, 33
Minkowski addition, 64
mixed binomial point process, 48
mixing point process, 33
modified Campbell measure, *see* reduced Campbell measure
moment measure
 n th order, 122
 first, 119
 second, 120
moment-generating function, 81, 257
motion-invariance, 32
multiset, 10
multivariate point process, 143

Nakagami fading, 42
nearest-neighbor distance, 39, 154, 155, 165
 binomial point process, 156

Gauss–Poisson process, 171
nearest-neighbor graphs, 251
Neyman–Scott cluster process, 55
Neyman–Scott process, 142
 K function, 132

offspring distribution, 181
outage probability, 104
 for Rayleigh fading, 104
 temporal correlation, 115

pair correlation function, 125
 binomial point process, 125
 Poisson process, 125

pair potential, 67

Palm distribution, 161
 for marked point processes, 172
 for stationary processes, 157
 inversion formula, 158
 Poisson point process, 168
 reduced, 165

Palm expectation, 154

Palm measure, *see* Palm distribution

Papangelou conditional intensity, 68, 167

Papangelou kernel, 167

partition function, 67

path formation time, 237

Peierls’ argument, 194

percolation
 of coverage process, 256
 of static SINR graph, 240

percolation probability
 definition, 180

perturbed lattice, 23

pgfl, *see* probability generating functional

pivotal edge, 203

point pattern, 17, 25

point process
 atomic, 13
 Cox, 50
 definition, 25

diffuse, 13
discrete, 13
distribution, 25
ergodic, 144
ergodicity, 33
finite, 47
hard-core, 57
infinitely divisible, 81
locally finite, 11
marked, 138
metrical transitivity, 33
mixing, 33
multivariate, 143
orderly, 11
property, 26, 153
simple, 11
taxonomy, 47

Poisson bipolar network, 105, 150
Poisson cluster process, 54, 126
Poisson hard-core process, 63, 67, 69
Poisson hole process, 53, 92, 142
Poisson network, 96
Poisson point process, 15, 18
 K function, 132
 doubly stochastic, 50
 finite, 49
 nearest-neighbor distance, 155
 one-dimensional, 15
 probability generating functional, 86
 thinning, 37
 uniform, 16

Poisson shot noise, 72
power-law shot noise, 73
PPP, *see* Poisson point process
probability generating functional, 80
 for mixed binomial point process, 81
 mixed Poisson point process, 89
 of Cox process, 89
 of homogeneous independent cluster process, 89
 of Neyman–Scott process, 90
 of Poisson cluster process, 90

R statistical software, 272
list of commands, 274

Rényi's theorem, 30

Radon–Nikodým density, 167

Radon–Nikodým derivative, 61, 161, 165, 166

Radon–Nikodým theorem, 148, 160

random closed set, 11

random field, 25

random geometric graph, 221
basic, 221
directed, 222
spatiotemporal, 233

randomly translated lattice, 23

reduced Campbell measure, 165

reduced Palm distribution, 165
for stationary processes, 165

reduced second moment function, 131

reduced second moment measure, 130, 170

refined Campbell theorem, *see* Campbell–Mecke theorem

regular tree, 181

Reimer's inequality, 202

RGG, *see* random geometric graph

Ripley's K function, *see* K function

Russo's formula, 203
integral form, 204

second factorial moment measure, 122

second moment density, 124
Matérn cluster process, 128
Thomas cluster process, 129

second moment measure, 120
for stationary processes, 129
reduced, 170

secrecy coverage, 266

secrecy graph, 241
out-percolation, 243

sentry selection, 268

set, 6
bounded, 6
compact, 6

Shannon throughput, 106

shot noise, 72, 93
signal-to-interference-plus-noise ratio, 104
simple sequential inhibition process, 58
singleton, 6
SINR, *see* signal-to-interference-plus-noise ratio
SINR coverage, 263
site percolation, 188, 215
 in triangular lattice, 215
 model, 179
Slivnyak’s theorem, 168, 238
space–time SINR graph, 235
space–time SINR multigraph, 234
spatial birth–death process, 71
spatial contention, 108
spatial throughput, 106
spatiotemporal graph, 233
square-root trick, 205
stable distribution, 99
 characteristic exponent, 99
 characteristic function, 99
 dispersion, 100
 LePage series, 100
static SINR graph, 240
stationarity, 32
Stienen model, 270
Strauss process, 62, 67
 K function, 134
 hard-core, 63
subadditivity, 239
success probability, 104
Swiss cheese model, 53, 142

thinning, 37
Thomas cluster process, 56, 129, 173
transmission capacity, 107
 with fading, 110
 without fading, 108
trifurcation, 208
typical point, 144, 152, 157

uniform random graph, *see* Erdős–RényiC random graph

universal mark, 142, 157

up-set, 198

vacancy

in coverage problems, 256

moments, 258

in point processes, 14

of Boolean model, 258

vacancy indicator, 14

void probability, 30

Voronoi cell, 22, 159

Voronoi diagram, 22

Voronoi tessellation, 22

waiting time paradox, 159