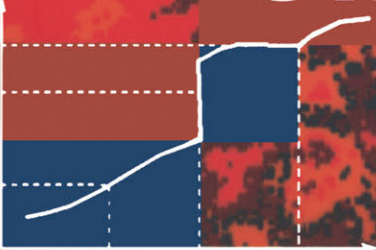


Christian Lantuéjoul

# Geostatistical Simulation



Models  
and  
Algorithms



Springer-Verlag Berlin Heidelberg GmbH

Christian Lantuéjoul  
Geostatistical Simulation

---

**Springer-Verlag Berlin Heidelberg GmbH**

Christian Lantuéjoul

# Geostatistical Simulation

Models and Algorithms

With 185 Figures and 5 Tables



Springer



Dr. Christian Lantuéjoul  
Centre de Géostatistique  
École des Mines  
35, rue Saint-Honoré  
77305 Fontainebleau  
France

ISBN 978-3-642-07582-7      ISBN 978-3-662-04808-5 (eBook)

DOI 10.1007/978-3-662-04808-5

CIP data applied for

Die Deutsche Bibliothek – CIP – Einheitsaufnahme

Lantuéjoul, Christian: Geostatistical simulation: models and algorithms; with 5 tables /  
Christian Lantuéjoul. – Berlin ; Heidelberg ; New York ; Barcelona ; Hongkong ; London ;  
Mailand ; Paris ; Singapur ; Tokio : Springer, 2002

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

<http://www.springer.de>

© Springer-Verlag Berlin Heidelberg 2002

Originally published by Springer-Verlag Berlin Heidelberg New York in 2002.

Softcover reprint of the hardcover 1st edition 2002

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Product liability: The publishers cannot guarantee the accuracy of any information about the application of operative techniques and medications contained in this book. In every individual case the user must check such information by consulting the relevant literature.

Camera ready by author

Cover design: E. Kirchner, Heidelberg

Printed on acid-free paper SPIN 106776976/3130/as 5 4 3 2 1 0

# Acknowledgements

This book proposes a unified presentation of the algorithms for conditionally simulating the spatial models currently used in geostatistics.

The first draft was prepared for postgraduate and undergraduate students at the Ecole des Mines. The course material was then expanded for several lecture courses given in Latin America and improved as a result of discussions with Gustavo Gedler from Intevep (Venezuela), Armando Remacre from Campinas University (Brazil), Marco Alfaro and Pedro Carrasco from Codelco (Chile). The next set of improvements came when I spent the 1995 summer term at the Department of Statistics of the University of Washington in Seattle with Paul Sampson and Adrian Raftery. There I had the privilege of attending a seminar by Julian Besag on exact simulations.

The book really started to take form back at the Centre de Géostatistique with continuing support and encouragement from the director, Michel Schmitt. I also would like to thank all my colleagues, especially Alain Galli, Christian Lajaunie, Didier Renard, Jacques Rivoirard and Hans Wackernagel, as well as Francis Maisonneuve from the Ecole des Mines and Ute Müller from Edith Cowan University (Australia) for their careful reading of the different chapters of this book. Their helpful criticisms have been greatly beneficial. Sincere thanks are also due to Margaret Armstrong and Ute Müller for their patience in correcting my English. Lastly I have really appreciated working with Janet Sterritt-Brunner, Susanna Pohl and Wolfgang Engel from Springer Verlag.

One last word. Many ideas presented in the book come more or less directly from Georges Matheron who was the founder of the Centre de Géostatistique, and who left us with a large inheritance of inestimable value.

# Table of Contents

<b>1. Introduction</b> .....	1
1.1 Simulation versus estimation .....	1
1.2 Models, simulations and algorithms .....	2
1.3 Outline of this book .....	4
1.3.1 The tools .....	4
1.3.2 The algorithms .....	4
1.3.3 The models .....	5

---

## Part I. The tools

---

<b>2. Investigating stochastic models</b> .....	9
2.1 Definition of a probability space .....	9
2.2 Random functions .....	10
2.3 Random sets .....	11
2.4 Random point processes .....	16
2.5 Random populations of objects .....	17
<b>3. Variographic tools</b> .....	21
3.1 The covariograms .....	21
3.1.1 The transitive covariogram .....	21
3.1.2 The geometric covariogram .....	22
3.2 The covariance function and the variogram .....	24
3.2.1 Definitions and general properties .....	24
3.2.2 The indicator variogram .....	26
<b>4. The integral range</b> .....	29
4.1 An estimation problem .....	29
4.2 The integral range .....	30
4.3 Practical assessment of the integral range .....	32
<b>5. Basic morphological concepts</b> .....	39
5.1 Dilation and erosion .....	39
5.2 Opening and Closing .....	41
5.3 Hausdorff distance .....	42

5.4	Object detection	43
<b>6.</b>	<b>Stereology: some basic notions</b>	<b>47</b>
6.1	The convex model	47
6.1.1	The Minkowski functionals	47
6.1.2	The formulae	49
6.2	The polyconvex model	51
6.2.1	Euler-Poincaré Characteristic	52
6.2.2	Extension of the Minkowski functionals	52
<hr/>		
<b>Part II. The algorithms</b>		
<hr/>		
<b>7.</b>	<b>Basics about simulations</b>	<b>57</b>
7.1	Uniform distributions	57
7.1.1	Uniform distribution in $]0, 1[$	57
7.1.2	Uniform distribution on a bounded domain of $\mathbb{R}^d$	60
7.2	Non-uniform distribution	61
7.2.1	Inversion method	61
7.2.2	Acceptance-Rejection method	62
7.2.3	Ad hoc methods	64
<b>8.</b>	<b>Iterative algorithms for simulation</b>	<b>67</b>
8.1	Some reminders on Markov chains	67
8.1.1	Markov chain	67
8.1.2	Convergence	68
8.1.3	Some features of a transition kernel	69
8.2	Metropolis algorithm	71
8.3	A Hit-and-Run algorithm	74
8.3.1	A stereological algorithm	74
8.3.2	A uniform Gibbs sampler	76
8.3.3	A more general hit-and-run algorithm	78
8.4	Application to conditional simulations	80
8.4.1	Restricting the transition kernel	80
8.4.2	Simulated annealing	82
<b>9.</b>	<b>Rate of convergence of iterative algorithms</b>	<b>87</b>
9.1	Rates of convergence	87
9.2	Minorization	88
9.3	Isofactorial representation	90
9.4	Empirical determination of the rate of convergence	95

<b>10. Exact simulations</b> .....	101
10.1 Propp and Wilson's algorithm .....	101
10.1.1 Principles .....	101
10.1.2 Condition for a singleton .....	104
10.1.3 Monotonic Monte Carlo .....	107
10.2 Fill's algorithm .....	111
<hr/>	
<b>Part III. The models</b>	
<hr/>	
<b>11. Point processes</b> .....	119
11.1 The Poisson point process .....	120
11.1.1 The homogeneous Poisson point process .....	120
11.1.2 The non-homogeneous Poisson point process .....	122
11.2 Conditional simulation of a Poisson point process .....	124
11.3 The Cox process .....	129
<b>12. Tessellations</b> .....	133
12.1 Statistical characterization of a tessellation .....	133
12.1.1 Partition point of view .....	133
12.1.2 Random function point of view .....	134
12.1.3 Cell population point of view .....	135
12.2 Voronoi tessellation .....	136
12.2.1 Definition and basic properties .....	136
12.2.2 Simulation .....	137
12.3 Poisson tessellation .....	144
12.3.1 Definition and basic properties .....	144
12.3.2 Simulation .....	146
<b>13. Boolean model</b> .....	153
13.1 Definition and basic properties .....	153
13.1.1 Definition .....	153
13.1.2 Avoiding functional .....	155
13.1.3 Stability properties .....	156
13.2 Simulation .....	157
13.2.1 Non conditional simulation .....	158
13.2.2 Conditional simulation .....	159
<b>14. Object based models</b> .....	167
14.1 Random token model .....	167
14.1.1 Definition and basic properties .....	167
14.1.2 Simulation .....	169
14.2 Boolean random function .....	171
14.2.1 Definition and basic properties .....	171
14.2.2 Simulation .....	173

Table of Contents

14.3	The dead leaves model	175
14.3.1	Definition and basic properties	175
14.3.2	Simulation	178
<b>15.</b>	<b>Gaussian random function</b>	<b>183</b>
15.1	Definition and basic properties	183
15.1.1	Results from probability	183
15.1.2	Definition of a gaussian random function	185
15.1.3	Textures of the realizations	186
15.2	Non conditional simulation	189
15.2.1	The dilution method	190
15.2.2	The tessellation method	191
15.2.3	The spectral method	191
15.2.4	The turning bands method	192
15.2.5	How many random functions?	197
15.3	Conditional Simulation	199
<b>16.</b>	<b>Gaussian variations</b>	<b>205</b>
16.1	Excursion set of a gaussian random function	205
16.1.1	Definition and basic properties	205
16.1.2	Conditional simulation	208
16.2	Plurigaussian random function	211
16.2.1	Definition and examples	211
16.2.2	Spatial distribution	214
16.2.3	Conditional simulation	216
<b>17.</b>	<b>Substitution random functions</b>	<b>221</b>
17.1	Definition and main properties	221
17.2	Some examples	224
17.2.1	A basic example	224
17.2.2	Geometry	227
17.2.3	Topology	230
17.3	Conditional simulation	232

# Notation

## Set notation

$x, y, z, \dots$	points of $\mathbb{R}^d$
$X, Y, Z, \dots$	subsets of $\mathbb{R}^d$
$\mathcal{X}, \mathcal{Y}, \dots$	classes of subsets of $\mathbb{R}^d$
$\mathfrak{X}, \mathfrak{Y}, \dots$	family of classes of subsets of $\mathbb{R}^d$
$X^c$	complement of $X$ in $\mathbb{R}^d$
$X \cup Y$	union of $X$ and $Y$
$X \cap Y$	intersection of $X$ and $Y$
$X \setminus Y = X \cap Y^c$	set difference of $X$ and $Y$
$\check{X} = \{-x : x \in X\}$	reflection of $X$
$X_x$	$X$ shifted by $\vec{ox}$ ( $o = \text{origin of } \mathbb{R}^d$ )
$X \oplus Y = \cup_{y \in Y} X_y$	Minkowski sum of $X$ and $Y$
$X \ominus Y = \cap_{y \in Y} X_y$	Minkowski difference of $X$ and $Y$
$X^Y = (X \ominus \check{Y}) \oplus Y$	morphological opening of $X$ w.r.t. $Y$
$X^{\check{Y}} = (X \oplus \check{Y}) \ominus Y$	morphological closure of $X$ w.r.t. $Y$
$X \downarrow S$	projection of $X$ onto subspace $S$
$\overset{\circ}{X}$	topological interior of $X$
$\overline{X}$	topological closure of $X$
$\partial X$	boundary of $X$
$ X $	volume of $X$ (case $X$ infinite)
$\#X$	cardinality of $X$ (case $X$ finite)
$B(x, r)$	ball with center $x$ and radius $r$
$B_d = B(o, 1)$	standard unit ball in $\mathbb{R}^d$
$\omega_d =  B_d $	volume of the unit ball in $\mathbb{R}^d$
$S_d = \partial B_d$	standard unit sphere in $\mathbb{R}^d$
$H(\alpha, p)$	hyperplane with direction $\alpha$ and location $p$
$Y^X$	set of all mappings from $X$ to $Y$

## Classes of subsets

$\mathcal{F}_d$	closed subsets of $\mathbb{R}^d$
$\mathcal{G}_d$	open subsets of $\mathbb{R}^d$
$\mathcal{K}_d$	compact subsets of $\mathbb{R}^d$
$\mathcal{K}_d^*$	nonempty compact subsets of $\mathbb{R}^d$
$\mathcal{C}_d$	convex subsets of $\mathbb{R}^d$
$\mathcal{R}_d$	polyconvex subsets of $\mathbb{R}^d$

## Analysis

$1_x$	indicator function of set $X$
$d(X, Y)$	Hausdorff distance between $X, Y \in \mathcal{K}_d^*$
$W_i(X)$	$i^{\text{th}}$ Minkowski functional of $X \in \mathcal{R}_d$
$\chi(X) = W_d(X)$	Euler-Poincaré characteristic of $X \in \mathcal{R}_d$
$L^2(X, p)$	set of scalar functions defined on $X$ and square integrable for $p$
$\langle f, g \rangle$	scalar product between $f, g \in L^2(X, p)$
$\ f\ _2$	norm of $f \in L^2(X, p)$
$\ \mu\ $	total variation of measure $\mu$

## Probability and statistics

$A \vee B$	union of the events $A$ and $B$
$A \wedge B$	intersection of the events $A$ and $B$
$\mathcal{D}(X)$	distribution of $X$
$X \sim p$	the distribution of $X$ is $p$
$X \stackrel{df}{\sim} F$	the distribution function of $X$ is $F$
$X \stackrel{\mathcal{D}}{\equiv} Y$	$X$ and $Y$ have the same distribution
$T_x$	hitting functional of $X \in \mathcal{F}_d$
$P(x, A)$	transition kernel
$P^{(n)}(x, A)$	$n^{\text{th}}$ iterate of the transition kernel $P$
$s^2(X Y)$	dispersion variance of $X$ in $Y$



**Distributions**

<i>Unif</i>	Uniform distribution on $]0, 1[$
<i>Unif(D)</i>	Uniform distribution on $D$
<i>Gauss</i>	Standard Gaussian distribution
<i>Gauss(m, <math>\sigma^2</math>)</i>	Gaussian distribution with mean value $m$ and variance $\sigma^2$
<i>Exp</i>	Exponential distribution with scale factor 1
<i>Exp(b)</i>	Exponential distribution with scale factor $b$
<i>Gamma(<math>\alpha, b</math>)</i>	Gamma distribution with parameter $\alpha$ and scale factor $b$
<i>Beta(<math>\alpha, \beta</math>)</i>	Beta distribution with parameters $\alpha$ and $\beta$
<i>Ber(p)</i>	Bernoulli distribution with parameter $p$
<i>Bin(n, p)</i>	Binomial distribution with index $n$ and parameter $p$
<i>Poisson(<math>\theta</math>)</i>	Poisson distribution with mean $\theta$
<i>Geom(p)</i>	Geometric distribution with parameter $p$
<i>Nbd(<math>\nu, p</math>)</i>	Negative binomial distribution with index $\nu$ and parameter $p$
<i>Sichel(<math>\theta, \alpha</math>)</i>	Sichel distribution with parameters $\alpha$ and $\theta$
<i>InvG(<math>\alpha, \beta</math>)</i>	Inverse Gaussian distribution with parameters $\alpha$ and $\beta$
<i>Bigauss(<math>\rho</math>)</i>	standard bigaussian distribution with correlation $\rho$
<i>Multb(n; <math>p_1, \dots, p_n</math>)</i>	Multinomial distribution with index $n$ and parameters $p_1, \dots, p_n$

**Miscellaneous**

$a \equiv b \pmod{m}$	$a$ and $b$ are congruent modulo $m$
$\left(\frac{a}{b}\right)$	Legendre coefficient of $a$ and $b$
$(X, \preceq)$	poset
$M_x = \{y \in X : x \preceq y\}$	upper bounds of $x$ in $(X, \preceq)$
$M^x = \{y \in X : y \preceq x\}$	lower bounds of $x$ in $(X, \preceq)$

# 1. Introduction

## 1.1 Simulation versus estimation

The following problem was raised by Alfaro (1979). A submarine cable has to be laid across the straits of Gibraltar. How can its length be predicted if the depth of the sea floor has been measured sparsely along its trajectory?



Fig. 1.1. Part of the actual trajectory and sample data points

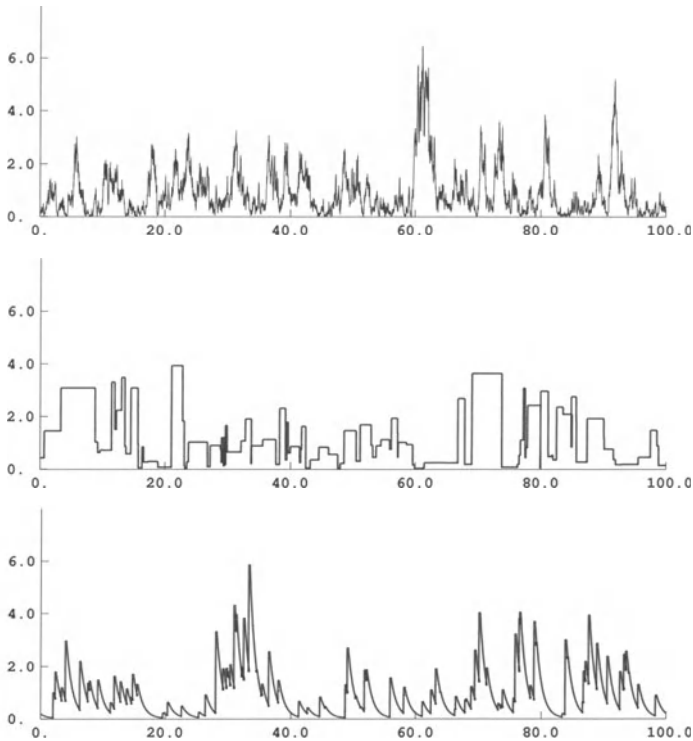
An exact determination of the length requires knowledge of the depth at each point of the trajectory. But these are mostly unknown. In a geostatistical setting, they are considered as random and can be estimated by linear regression starting from the available data points. This suggests estimating the actual length as the length of the estimated trajectory.

The results turn out to be disappointing. The length of the trajectory is seriously underestimated (see Figure 1.2). Clearly, the estimated trajectory is much smoother than the actual one.



Fig. 1.2. Part of the actual trajectory and its estimate from linear regression. In this particular example, the estimated trajectory is piecewise linear because the linear regression has been carried out using an exponential covariance function

What is really questionable in this procedure is not the construction of an estimator for the length starting from the depth estimator, but the depth estimator itself. Linear regression estimation requires only the mean and the covariance function. But the covariance function does not tell us much about the length of the trajectories. Figure 1.3 shows realizations of three different stochastic processes with the same exponential point distribution and the same exponential covariance function. The distributions of their trajectory lengths are totally different.



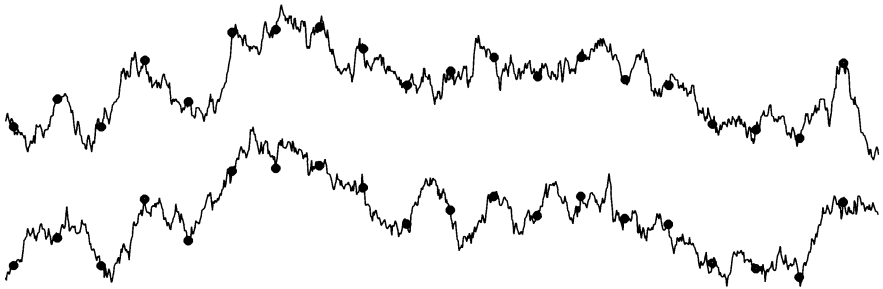
**Fig. 1.3.** Realizations of 3 different stochastic processes (from top to bottom: diffusion, tessellation, shot-noise) with the same exponential point distribution and the same exponential covariance function

## 1.2 Models, simulations and algorithms

Linear regression was chosen as an estimator because its implementation parameters (mean and covariance function) can be reasonably inferred from the available sample data. However, the formal expression of the length of a

trajectory<sup>1</sup> shows that its distribution, and even its mean value, depend on much more than a mean value or a covariance function. Finally, the quality of the estimator is poor because it is based on available but not relevant information. This is a rather common situation: the relevant information is not always available. In order to improve the estimation of the cable length, we have to make assumptions about the relevant information. These assumptions are summarized in a stochastic *model*. Of course, the more appropriate the model to the physical reality under study, the more accurate the resulting estimation. In some cases, the choice of the model is self-evident. For instance, the signal emitted by a micro-probe is blurred by a Poisson noise whose mean is proportional to the signal intensity (Daly, 1993). In other cases, there is a lot of freedom in its choice.

This book does not deal with the choice and the validation of a stochastic model (a book by Matheron (1989) is devoted to these issues). Here we are concerned with the way in which to use a model once it has been specified.



**Fig. 1.4.** Parts of the actual (top) and of a conditionally simulated (bottom) trajectories

In some cases, it is possible to devise an estimator which suits the problem addressed and the model chosen. An alternative and more general approach is to generate a realization of the stochastic model. Information is then extracted from it to produce an estimate of the required quantity. Of course, the

<sup>1</sup> Let  $f$  be a numerical function defined on  $[0, 1]$ . For every finite subdivision  $\sigma = (x_0 = 0 < x_1 < \dots < x_n = 1)$  of  $[0, 1]$  we put

$$\ell(\sigma) = \sum_{i=0}^{n-1} \sqrt{[f(x_{i+1}) - f(x_i)]^2 + (x_{i+1} - x_i)^2}$$

The length of  $f$  is then defined as the least upper bound of the  $\ell(\sigma)$

$$\ell = \sup_{\sigma} \ell(\sigma)$$

estimate will be more accurate if it is derived from a realization that honors the experimental data. Any realization of the model is called a *simulation*. A simulation that fits the available data is said to be *conditional*. By replicating the (conditional) simulations, many estimates can be obtained. This yields a distribution function with a mean, a variance and confidence limits.

A computer program is usually required to produce simulations from a given model. This is the practical implementation of an *algorithm*. So a simulation algorithm is designed starting from a model. Very often, a given model can be simulated via different algorithms. But they are not all equally efficient. The implementation of certain algorithms may lead to inefficient programs because computers have their own limitations, especially in regard to numerical precision, speed and memory.

### 1.3 Outline of this book

This book tries as much as possible to be self-contained. It is divided into three parts: the tools, the algorithms and the models.

#### 1.3.1 The tools

This part gives the statistical, geostatistical, morphological and stereological tools necessary for investigating, defining and describing a stochastic model.

It is well known from probability theory how to characterize the distribution of a random variable. But how can we characterize the distribution of more complicated stochastic models, such as random functions, random sets, random point processes, random populations of objects? Chapter 2 examines this question. It is followed by a reminder of the main variographic tools used in geostatistics (covariance function, variogram and covariograms). Chapter 4 introduces the integral range to predict how the realizations of a random function fluctuate statistically within any bounded domain. This knowledge is necessary to check the correctness of a simulation program. The next chapter presents two basic concepts in mathematical morphology (dilation and erosion) that can be used to investigate random sets and provide simple descriptions of many stochastic models. People working in the earth sciences are often faced with the problem of deducing 3-dimensional information starting from unidimensional (drill or bore holes) or bidimensional (outcrop) data. Chapter 6 provides some simple, yet not complete answers to that stereological problem.

#### 1.3.2 The algorithms

Some readers may be surprised that the algorithms are treated before the models. In fact, the algorithms presented in this part are either basic or generic.

Chapter 7 gives the main algorithms for simulating standard univariate distributions (uniform, gaussian, Poisson...). In the next chapter, algorithms based on markovian iterations are introduced. Their flexibility makes them particularly suitable for conditional simulations. In addition to a few basic algorithms (Metropolis, restriction of a Markov chain, simulated annealing), particular attention is focused on iterative algorithms for simulating uniformly distributed points in high-dimensional domains, as this is a basic step for many conditional simulation algorithms. In principle, infinitely many iterations must be run in order to simulate a model correctly. In practice, the algorithm has to be stopped after a finite number of iterations, which raises the difficult problem of determining its rate of convergence. This is addressed in chapter 9. In fact, this problem can be bypassed in some cases by considering "exact" algorithms that appeared recently in the literature after the pioneering work by Propp and Wilson (1995). Chapter 10 contains a short review of the topic.

### 1.3.3 The models

This book does not claim to be comprehensive. Except in chapters 16 and 17, the models considered here are prototypes. They have been chosen for their simplicity, their tractability and their range of applicability. All of them are stationary<sup>2</sup>, univariate<sup>3</sup>, and defined on continuous spaces<sup>4 5</sup>.

The models considered in this book can be organized according to a certain hierarchy. Point processes are at the bottom of this hierarchy. Two of these (*Poisson point process* and *Cox process*) are studied in chapter 11. The next chapter deals with the *Poisson flats* as well as the *tessellations* (Poisson and Voronoi) that can be built starting from a Poisson point process. Chapter 13 is concerned with the *boolean model*. It consists of taking the union of independent random compact sets implanted at the points of a Poisson process. More generally, instead of random compact sets, we can consider random

<sup>2</sup> Regarding the gaussian random function with stationary increments of order  $k$ , we refer to Matheron (1973), Chilès (1977, 1995) as well as Chilès and Delfiner (1999). The spatial deformation of a gaussian random function has been investigated by Sampson and Guttorp (1992), Monestiez, Meiring, Sampson and Guttorp (1997). See also Vedel-Jensen and Stougaard-Nielsen (1998), and Senoussi, Chadoeuf and Allard (2000) for the spatial deformation of a stationary point process.

<sup>3</sup> Good references for multivariate models and methods are Wackernagel (1995), as well as Chilès and Delfiner (1999).

<sup>4</sup> As mentioned by Hu, Blanc, and Touati (1998), models defined on continuous spaces are desirable when their realizations are used as input argument in finite element programs. The discretization required depends on the values of the realizations and cannot be given beforehand.

<sup>5</sup> Because we consider only models defined on continuous spaces, important models like the Markov random fields (Besag, 1974; Guyon, 1995) are not treated (even if continuous versions of this model are now available (Arak and Surgailis (1989)).

functions with compact support or converging to 0 rapidly at infinity. Replacing the union by the addition, the supremum or the superposition, we obtain three different object-based models respectively called the *dilution random function*, the *boolean random function*, or the *dead-leaves model*. These three models are presented in chapter 14. *Gaussian random functions* are then introduced via the central limit theorem. Two of their variations are presented in chapter 16. The first one is a random set (*excursion set of a gaussian random function*), the second one is a step function (*plurigaussian random function*). Finally, chapter 17 proposes the composition of a random function with stationary increments and a stationary stochastic process. The *substitution random function* thus obtained shows up a large flexibility regarding the geometry and the topology of its realizations.

Each of these models is defined. Its main properties are reviewed, and algorithms to simulate it conditionally or not are given and demonstrated.

A final comment: an abundant geostatistical literature is devoted to simulations. But the word simulation does not necessarily have the same meaning as in the present book. Sometimes the simulations are produced by an algorithm based on an incompletely specified model, or even no model at all. The outcomes produced by these algorithms seem to have been obtained from computer aided design with a stochastic flavour. Individual simulations are very nice, but there is very little variability between them. This tends to give users the mistaken impression that they have tighter control over the reality than they really do, and can lead to some appallingly wrong decisions. This book only presents algorithms that are based on sound stochastic models.

Part I

**The tools**



## 2. Investigating stochastic models

In this chapter we review the statistical description of several classes of stochastic models (random functions, random sets, random point processes and random populations of objects) that will be encountered throughout this book. This review is preceded by a brief reminder about probability calculus.

### 2.1 Definition of a probability space

A stochastic model is completely specified by a *probability space*, that is a triplet  $(\Omega, \mathcal{A}, P)$ , where

- $\Omega$  is the set of all possible *realizations* (or *states*) of the model.
- $\mathcal{A}$  is the set of *events*. An event is a family of realizations.
- $P$  is a *probability measure*. It measures the frequency of occurrence of each event.

For technical reasons that are not developed here, a family of realizations is not necessarily an event. The set  $\mathcal{A}$  must constitute a  $\sigma$ -*algebra*, which means that it includes  $\Omega$  and  $\emptyset$  as events, it is stable under complementation (if  $A$  is an event, then  $\Omega \setminus A$  is also an event), and it is stable under countable union and intersection (if  $(A_n)$  is a sequence of events, then  $\cup_n A_n$  and  $\cap_n A_n$  are events).

The probability measure  $P$  is defined on  $\mathcal{A}$ . It satisfies  $P(\emptyset) = 0$  and  $P(\Omega) = 1$ . Moreover,  $P(\Omega \setminus A) = 1 - P(A)$  for any event  $A$ . Finally,  $P$  is  $\sigma$ -additive, i.e.  $P(\cup_n A_n) = \sum_n P(A_n)$  for any sequence  $(A_n)$  of pairwise disjoint events. An immediate and useful consequence of the  $\sigma$ -additivity of  $P$  is that  $P(\cup_n A_n) = \lim_n P(A_n)$  for any increasing sequence  $(A_n)$  of events. By complementation, we also have  $P(\cap_n A_n) = \lim_n P(A_n)$  for any decreasing sequence  $(A_n)$  of events.

In practical applications,  $\mathcal{A}$  is often defined as the  $\sigma$ -algebra spanned by a family  $\mathcal{A}_0$  of elementary events, that is the smallest  $\sigma$ -algebra that contains  $\mathcal{A}_0$ . Similarly, the probability measure  $P$  can be completely specified by its

values on  $\mathcal{A}_0$ , provided that  $\mathcal{A}_0$  satisfies some stability properties<sup>1</sup>. Explicit examples will be given in the next sections.

## 2.2 Random functions

A *random function*  $Z$  is a family of random variables  $\{Z(x)\}$  where  $x$  belongs to  $\mathbb{R}^d$  or some subset of it. In the case  $d = 1$ , we prefer to speak of a *stochastic process*. Figure 2.1 shows a realization of a random function.

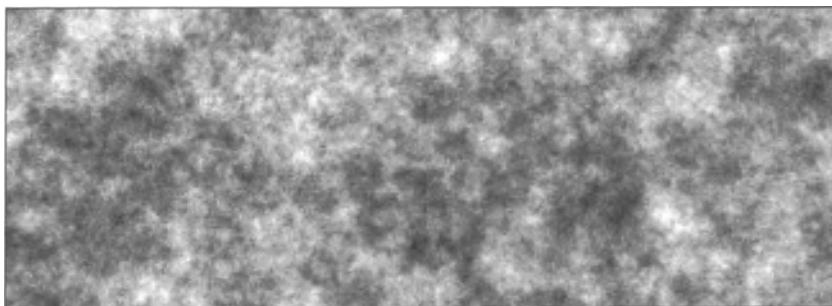


Fig. 2.1. Realization of a random function

Clearly, the statistical properties of a random function include those of the random vectors that constitute it. To each random vector  $(Z(x_1), \dots, Z(x_n))$ , a cumulative multivariate distribution function

$$F_{x_1, \dots, x_n}(z_1, \dots, z_n) = P\{Z(x_1) < z_1, \dots, Z(x_n) < z_n\}$$

can be associated. Letting  $n$  run through  $\mathbb{N}$  and  $(x_1, \dots, x_n)$  through  $\mathbb{R}^d \times \dots \times \mathbb{R}^d$ , we obtain a family of cumulative multivariate distribution functions. This family defines the *spatial distribution* of  $Z$ .

**Remark 2.2.1.** In the case when  $Z$  takes discrete values, it is often more convenient to define its spatial distribution starting from the multivariate distribution of its random vectors

$$F_{x_1, \dots, x_n}(z_1, \dots, z_n) = P\{Z(x_1) = z_1, \dots, Z(x_n) = z_n\}$$

---

<sup>1</sup> For instance,  $\mathcal{A}_0$  is called a semi-algebra if it contains  $\emptyset$  and  $\Omega$ , if it is stable under finite intersection, and if the complement of any element of  $\mathcal{A}_0$  can be written as a finite union of pairwise disjoint elements of  $\mathcal{A}_0$ . Any  $\sigma$ -additive mapping  $P$  from the semi-algebra  $\mathcal{A}_0$  to  $[0, 1]$  such that  $P(\Omega) = 1$  can be extended to a probability measure on the  $\sigma$ -algebra spanned by  $\mathcal{A}_0$ . Moreover, this measure is unique (Neveu, 1965).

By definition, the spatial distribution of  $Z$  measures the frequency of occurrence of events based on finite sets of points. Let  $\mathcal{A}_0$  denote this class of events. A classical theorem due to Kolmogorov (1933) states that the spatial distribution can be extended to a probability measure on the  $\sigma$ -algebra  $\mathcal{A}$  spanned by  $\mathcal{A}_0$ . Moreover, this measure is unique.

**Remark 2.2.2.** As the events of  $\mathcal{A}$  are based on only finite or countable sets of points, assertions like "the realizations are differentiable at  $x$ " or "the maximum of the random function over a given domain  $D$  exceeds the value  $\lambda$ " which require continuous sets of points are not events of  $\mathcal{A}$  and therefore cannot be measured.

### 2.3 Random sets

A *random set* is a stochastic model whose realizations are subsets of  $\mathbb{R}^d$ . Typical examples of random sets include those obtained by thresholding a random function  $Z$  at a given level  $\lambda$  (cf. Figure 2.2)

$$X = \{x \in \mathbb{R}^d \mid Z(x) \geq \lambda\}$$



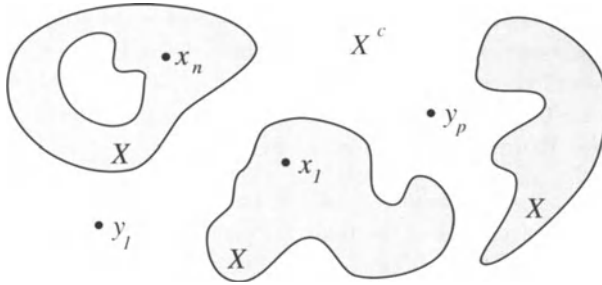
Fig. 2.2. Realization of a random set

Any random set  $X$  can be characterized by its *indicator function*, defined as

$$1_X(x) = \begin{cases} 1 & \text{if } x \in X \\ 0 & \text{otherwise} \end{cases}$$

This suggests specifying the statistical properties of the random set by using the spatial distribution of its indicator function. Using remark 2.2.1., it can be written

$$F_{x_1, \dots, x_n; y_1, \dots, y_p}(1, \dots, 1; 0, \dots, 0) = P\{x_1 \in X, \dots, x_n \in X; y_1 \notin X, \dots, y_p \notin X\}$$



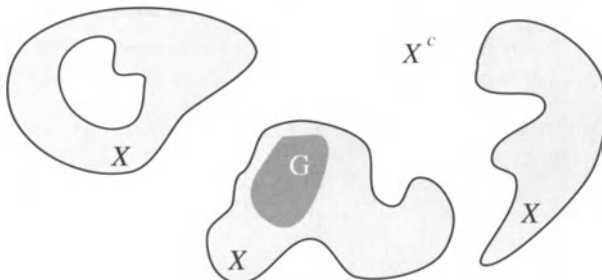
**Fig. 2.3.** Specifying the statistical properties of a random set using the spatial distribution of its indicator function amounts to probing the random set with finite sets of points

Unfortunately this promising idea does not always work. Consider for instance the random set consisting of a single point located in  $\mathbb{R}^d$  according to a gaussian distribution. Its spatial distribution equals zero irrespectively of the parameters of the gaussian distribution. In such a case, a probe made up of a finite set of points is too thin to determine the location of the point. Bigger probes are necessary. Matheron (1975) proposed to resort to open subsets.

Given an open set  $G$ , several events can be considered for describing the relative location of  $G$  w.r.t.  $X$ :

- (1)  $G$  is contained in  $X$  ( $G \subset X$ ).
- (2)  $G$  hits  $X$  ( $G \cap X \neq \emptyset$ ).
- (3)  $G$  hits  $X^c$  ( $G \cap X^c \neq \emptyset$ ).
- (4)  $G$  avoids  $X$  ( $G \cap X = \emptyset$  or equivalently  $G \subset X^c$ ).

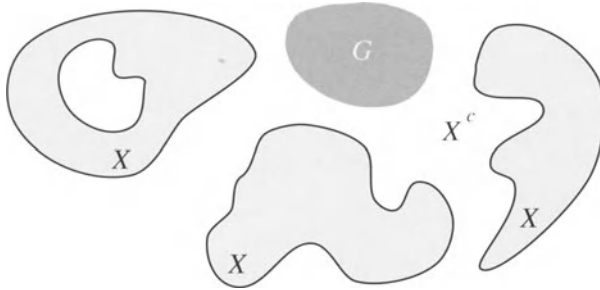
Since (1) and (3), as well as (2) and (4) are complementary, only two different types of events have to be considered, for instance (1) and (4). Let us examine them in turn.



**Fig. 2.4.** Probing the random set using the inclusion logic leads to the random open set theory

$G$  is contained in  $X$  iff it is contained in the interior  $\overset{\circ}{X}$  of  $X$ . Accordingly, probing a random set using the inclusion logic does not allow us to distinguish between a set and its interior. This leads to the random open set theory (cf. Figure 2.4).

Similarly,  $G$  avoids  $X$  iff it avoids the closure  $\overline{X}$  of  $X$ . Accordingly, probing a random set by avoidance does not allow us to distinguish between a set and its closure. This leads to the random closed set theory (cf. Figure 2.5).



**Fig. 2.5.** Probing the random set using the avoidance logic leads to the random closed set theory

The complement of an open set is a closed set, so both theories are dual w.r.t. complementation. However, we usually prefer using the random closed set theory rather than the open one, because random closed sets include several classes of models that are very useful for practical applications, such as point processes and line networks.

From now on, we will be mainly concerned with random closed sets (in brief RACS).

At this point, some notation must be introduced. Following Matheron (1975), the family of the closed subsets of  $\mathbb{R}^d$  is denoted by  $\mathcal{F}$ . If  $A \subset \mathbb{R}^d$ , we write  $\mathcal{F}_A$  for the family of the closed subsets that hit  $A$ , and similarly  $\mathcal{F}^A$  for those that avoid  $A$ .

$$\mathcal{F}_A = \{F \in \mathcal{F} \mid F \cap A \neq \emptyset\} \quad \mathcal{F}^A = \{F \in \mathcal{F} \mid F \cap A = \emptyset\}$$

$\mathcal{F}$  can be equipped with the  $\sigma$ -algebra  $\sigma_{\mathcal{F}}$  spanned by the events  $\mathcal{F}^G$  with  $G$  open. Matheron (1975) showed that probability measures do exist on  $(\mathcal{F}, \sigma_{\mathcal{F}})$ . The distribution of the RACS  $X$  is completely specified by such a probability measure.

Let  $K$  be a compact (i.e. closed and bounded) subset of  $\mathbb{R}^d$ . Note that  $\mathcal{F}^K = \cup_n \mathcal{F}^{G_n}$ , where  $G_n$  is the open set consisting of the points with distance less

than  $\frac{1}{n}$  from  $K$ . Consequently  $\mathcal{F}^K \in \sigma_{\mathcal{F}}$ . This makes it possible to introduce the *avoiding functional* of the RACS  $X$ . It is defined on the family  $\mathcal{K}$  of compact subsets of  $\mathbb{R}^d$  by the formula

$$Q(K) = P\{\mathcal{F}^K\} = P\{X \cap K = \emptyset\}$$

Rather than the avoiding functional, we sometimes prefer to consider the *hitting functional* which is its complement

$$T(K) = P\{\mathcal{F}_K\} = P\{X \cap K \neq \emptyset\}$$

The hitting functional satisfies the following properties:

i)  $0 \leq T \leq 1$  with  $T(\emptyset) = 0$ .

ii) clearly  $T(K) \leq T(K \cup K')$ . More generally, let  $(K_i, i \in I)$  be a finite sequence in  $\mathcal{K}$ . We have

$$P\left\{\mathcal{F}^K \wedge \left(\bigwedge_{i \in I} \mathcal{F}_{K_i}\right)\right\} \geq 0$$

where  $\bigwedge$  denotes the intersection of events. Using the inclusion-exclusion formula (Comtet, 1970), this inequality becomes

$$\sum_{J \subset I} (-1)^{\#J+1} T(K \cup K_J) \geq 0$$

where  $K_J = \bigcup_{j \in J} K_j$  and  $\#J$  denotes the cardinality of  $J$ .

iii) Let  $(K_n)$  be a monotonic decreasing sequence in  $\mathcal{K}$  with limit  $K$ . Then

$$\lim_{n \rightarrow +\infty} T(K_n) = T(K)$$

Our reason for studying the hitting functional of a RACS comes from the following fundamental result due to Choquet (1954), Kendall (1974) and Matheron (1975):

**Theorem 2.3.1.** *Let  $T$  be a functional defined on  $\mathcal{K}$ , that satisfies properties i), ii) and iii) given above. Then there exists a unique probability measure  $P$  on  $(\mathcal{F}, \sigma_{\mathcal{F}})$  such that*

$$P\{\mathcal{F}_K\} = T(K) \quad K \in \mathcal{K}$$

In other words, the functional  $T$  is the RACS equivalent of the distribution function of a random variable.

**Example 2.3.1.** Let  $X$  be the RACS made up of a single point uniformly located within a compact subset  $K_0$ . The hitting functional of  $X$  is

$$T(K) = P\{X \cap K \neq \emptyset\} = \frac{|K_0 \cap K|}{|K_0|} \quad K \in \mathcal{K}$$

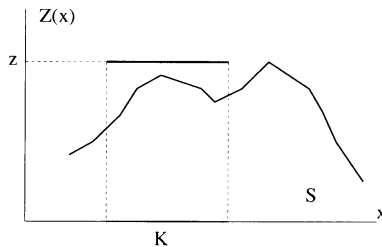
Conversely, theorem 2.3.1. implies that a RACS with the above hitting functional consists of a single point uniformly located within  $K_0$ .

**Remark 2.3.1.** Matheron (1975) has thoroughly investigated the connections between spatial distributions and hitting functionals. His results can be stated as follows. Let  $X$  be a RACS with hitting functional  $T$  and spatial distribution  $F$ . There exist infinitely many RACS's with spatial distribution  $F$ . Among those, only one, say  $X'$ , is *separable*, i.e. such that  $\overline{X' \cap D} = X'$  a.s. for some countable set  $D$  that is dense in  $\mathbb{R}^d$ . The hitting functional  $T'$  of  $X'$  satisfies  $T'(K) \leq T(K)$  for any  $K \in \mathcal{K}$ , with equality if  $K$  is a finite subset.

**Remark 2.3.2.** Let  $Z$  be a random function defined on  $\mathbb{R}^d$ , and let  $S$  denotes its subgraph

$$S = \{(x, z) \in \mathbb{R}^d \times \mathbb{R} \mid Z(x) \geq z\}$$

Suppose that  $Z$  is upper semicontinuous. Then  $S$  is a RACS of  $\mathbb{R}^d \times \mathbb{R}$ . Accordingly, if  $K \in \mathcal{K}$  and  $z \in \mathbb{R}$ , then  $S \in \mathcal{F}^{(K,z)}$  is an event which takes place if and only if  $\max_{x \in K} Z(x) < z$  (see Figure 2.6). Therefore the maximum of  $Z$  over  $K$  is measurable (Matheron, 1968). Similarly, the minimum of a lower semicontinuous function can be made measurable by considering its supergraph as a RACS.



**Fig. 2.6.** If  $Z$  is upper semicontinuous, its maximum over a compact subset is measurable

**Remark 2.3.3.** Instead of considering random open or closed sets, it is also possible to consider random *regular* sets. These sets satisfy

$$\overline{\overset{\circ}{X}} = \overline{X} \quad \overset{\circ}{\overline{X}} = \overset{\circ}{X}$$

i.e.  $X$  and its complement  $X^c$  do not present any infinitely thin components or any isolated points, which make them quite suitable for modeling physical realities such as porous media. This particular class of random set which has already been considered by Matheron (1975) is currently under investigation by Schmitt (2000).

**Remark 2.3.4.** Theorem 2.3.1. can be refined when further information is available about the RACS  $X$ . In the typical case when it is compact and convex, Vitale (1983) and Molchanov (1984) have shown that its distribution is uniquely determined by the values of  $P\{X \subset K\}$  for any convex compact subset  $K$  of  $\mathbb{R}^d$ .

## 2.4 Random point processes

A random point process in  $\mathbb{R}^d$  is a random set of  $\mathbb{R}^d$ , each realization of which is made up of a finite or countable number of points (cf. Figure 2.7).

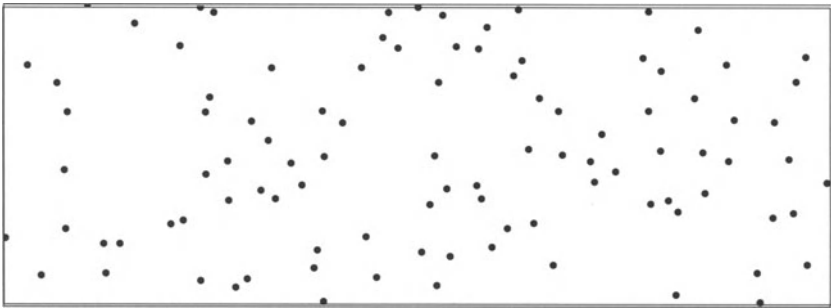


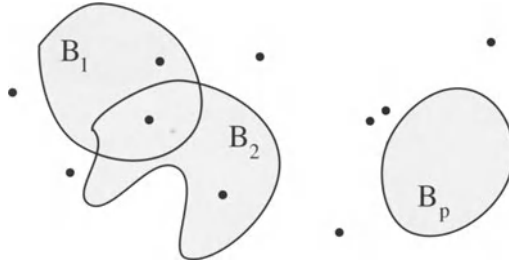
Fig. 2.7. Realization of a random point process.

In the case when each realization is closed (i.e. the limit of any sequence of points of the realization also belongs to the realization), the point process is a RACS. Accordingly, its statistical properties are given by its hitting (or avoiding) functional.

A *locally finite* point process is a typical example of a closed one. This is a point process for which the number of points falling within any compact subset is almost surely finite. In this case, it is also convenient to describe the statistical properties of the random point process by counting the number of points in certain subsets of space. Clearly, the family of probing sets cannot be arbitrary. The one usually considered is the family  $\mathcal{B}$  of Borel sets or **borelians**, that is the smallest  $\sigma$ -algebra spanned by the open subsets of  $\mathbb{R}^d$ . This family includes closed sets (by complementation) and compact subsets, just to mention a few. The bounded borelians are of special interest because the number of points they contain is almost surely finite. More



specifically, let  $N(B)$  be the random number of points of the process within the bounded borelian  $B$  (in short  $B \in \mathcal{B}_0$ ). Starting from  $N$ , it is possible to define the spatial distribution of the random point process as the family of the multivariate distributions of  $(N(B_1), \dots, N(B_p))$  for any  $p \in \mathbb{N}$  and for any  $B_1, \dots, B_p \in \mathcal{B}_0$  (cf. Figure 2.8).



**Fig. 2.8.** Spatial distribution of a locally finite point process

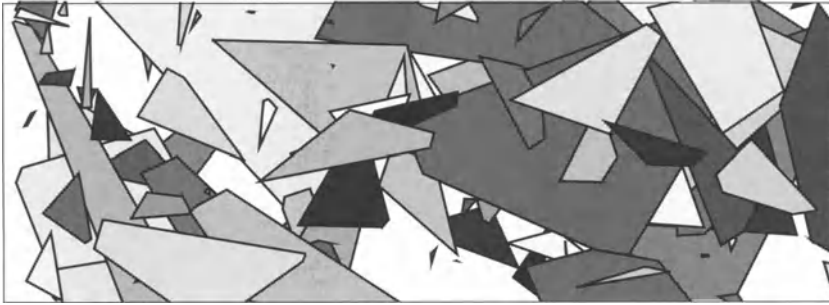
If  $K \in \mathcal{K}$ , the formula  $Q(K) = P\{N(K) = 0\}$  shows that the avoiding functional of a locally finite point process can easily be derived from its spatial distribution. This means that the spatial distribution provides at least as much information about the point process as the avoiding functional. In fact, they both provide exactly the same information. It should also be mentioned that it is easier to describe the properties of the point process using the spatial distribution, but it is easier to establish its mathematical properties using the avoiding functional.

**Remark 2.4.1.** The avoiding functional has a more general range of applicability because it can be used for closed point processes and not only for locally finite ones. Kendall (1974) gives the zeroes of a Brownian motion as an example of a closed point process which is not locally finite.

**Remark 2.4.2.** It is also possible to consider *weighted point processes* by assigning a value or weight to each individual point of the process. Formally, a weighted point process can be regarded as an ordinary point process defined on  $\mathbb{R}^d \times \mathbb{R}$ .

## 2.5 Random populations of objects

In this section, we denote by  $\mathcal{K}'$  the family of nonempty compact subsets of  $\mathbb{R}^d$ . The elements of  $\mathcal{K}'$  are called *objects*. A random population of objects is a point process on  $\mathcal{K}'$ . Each realization is made up of a finite or countable number of objects (cf. Figure 2.9).



**Fig. 2.9.** Realization of a random population of objects displayed by superposition. Arbitrary gray-tone values have been assigned to the objects in order to distinguish between them

Exactly as for point processes in  $\mathbb{R}^d$ , if the random population of objects is closed (and in particular locally finite), then its statistical properties can be given by the equivalent of a hitting functional. Of course, this assumes that  $\mathcal{K}'$  has been equipped with a topology. A natural topology for  $\mathcal{K}'$  is the one induced by the *Hausdorff distance*<sup>2</sup>. Several characterizations of compact subsets of  $\mathcal{K}'$  for the Hausdorff topology exist (Matheron, 1975; Schmitt, 1997), but they are difficult to handle.

Let  $K$  be an arbitrary compact subset of  $\mathbb{R}^d$ . An important property of a locally finite population is that the number of objects totally *contained* in  $K$  is finite. A stronger property can be considered, namely that the number of objects *hitting*  $K$  is finite. A random population of objects satisfying this property is said to be of *finite order*. In this case, it is possible to define the spatial distribution of the random population as the family of the multivariate distributions of  $(N(K_1), \dots, N(K_p))$ , where  $N(K)$  denotes the (random) number of objects hitting  $K$ .

**Remark 2.5.1.** In this approach, the objects in a population are indistinguishable. Indeed, the number of objects is finite or countable, so it is possible to order them. But any order relation would be arbitrary, and would not affect any statistics defined on the population. It can also be noticed that the description of a population as a point process in  $\mathcal{K}'$  does not allow two objects to be identical. If necessary, a weight, such as a multiplicity order, can be assigned to each object. An alternative approach is the concept of exponential spaces developed by Carter and Prenter (1972) in the case of finite populations.

---

<sup>2</sup> The Hausdorff distance between the two nonempty compact subsets of  $\mathbb{R}^d$  is the maximum euclidean distance separating their points. See chapter 5 for a precise definition.

## Exercises

- 1 Let  $A$  and  $B$  two arbitrary subsets of  $\mathbb{R}^d$ . Show that  $\mathcal{F}^{A \cup B} = \mathcal{F}^A \cap \mathcal{F}^B$ , but only  $\mathcal{F}_{A \cap B} \subset \mathcal{F}_A \cap \mathcal{F}_B \subset \mathcal{F}_{A \cup B}$ .
- 2 What is the hitting functional of a point randomly located in  $\mathbb{R}^d$  with p.d.f.  $f$ ?
- 3 A random set  $X$  is said to be *autodual* if  $X$  and its complement  $X^c$  have the same spatial distribution.
  - 1) Show that  $P\{x \in X\} = \frac{1}{2}$ .
  - 2) Show that the trivariate distributions of  $X$  are completely specified by its bivariate distributions.
- 4 Let  $\mathcal{X}$  and  $\mathcal{Y}$  be two non random populations of objects of finite order, and let  $n_x$  and  $n_y$  their hitting counting functional defined on  $\mathcal{K}$  as

$$n_x(K) = \sum_{X \in \mathcal{X}} 1_{X \cap K \neq \emptyset} \qquad n_y(K) = \sum_{Y \in \mathcal{Y}} 1_{Y \cap K \neq \emptyset}$$

Show that  $n_x = n_y$  implies  $\mathcal{X} = \mathcal{Y}$ . Does this implication remain true if the equality  $n_x = n_y$  holds only on the singletons of  $\mathcal{K}$ ?

### 3. Variographic tools

The investigation of a stochastic model using pairs of points yields significant geometric information. This chapter introduces the basic tools of transitive and intrinsic geostatistics. These are the covariograms, the covariance function and the variogram. Particular attention is paid to indicator variograms where specific properties and limitations are encountered.

#### 3.1 The covariograms

The cases of a function and a set will be considered in turn.

##### 3.1.1 The transitive covariogram

Let  $f$  be an integrable function defined on  $\mathbb{R}^d$ .

**Definition 3.1.1.** *The transitive covariogram of  $f$  is the mapping*

$$K(h) = \int_{\mathbb{R}^d} f(x) f(x+h) dx \quad h \in \mathbb{R}^d$$

Another possible way to define  $K$  is to express it as the convolution product  $K(h) = f \star \tilde{f}(h)$  where  $\tilde{f}(x) = f(-x)$ . The following proposition summarizes the main properties of the transitive covariogram.

**Proposition 3.1.1.** *The transitive covariogram of  $f$  satisfies*

i)  $K(0) = \int_{\mathbb{R}^d} f^2(x) dx.$

ii)  $K(-h) = K(h).$

iii)  $\int_{\mathbb{R}^d} K(h) dh = \left[ \int_{\mathbb{R}^d} f(x) dx \right]^2.$

iv) *for any finite sequence of points  $(x_\alpha)_{\alpha=1,n}$  and for any finite sequence of real numbers  $(\lambda_\alpha)_{\alpha=1,n}$ , we have*

$$\sum_{\alpha,\beta=1}^n \lambda_\alpha \lambda_\beta K(x_\alpha - x_\beta) \geq 0.$$

The proofs of i), ii) and iii) are straightforward. Regarding iv)

$$\begin{aligned} \sum_{\alpha, \beta=1}^n \lambda_\alpha \lambda_\beta K(x_\alpha - x_\beta) &= \sum_{\alpha, \beta=1}^n \lambda_\alpha \lambda_\beta \int_{\mathbb{R}^d} f(x) f(x + x_\alpha - x_\beta) dx \\ &= \sum_{\alpha, \beta=1}^n \lambda_\alpha \lambda_\beta \int_{\mathbb{R}^d} f(x + x_\alpha) f(x + x_\beta) dx \\ &= \int_{\mathbb{R}^d} \left[ \sum_{\alpha=1}^n \lambda_\alpha f(x + x_\alpha) \right]^2 dx \geq 0. \end{aligned}$$

### 3.1.2 The geometric covariogram

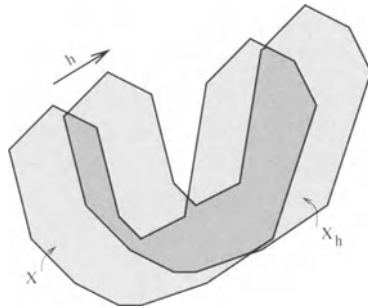
Suppose now that  $f$  is the indicator function of some subset  $X$  in  $\mathbb{R}^d$ . The volume  $|X|$  of  $X$  is finite because  $f$  is integrable. Let  $X_h$  denote the set  $X$  translated by the vector  $h$  (see Figure 3.1). Then we have

$$K(h) = \int_{\mathbb{R}^d} 1_X(x) 1_X(x+h) dx = \int_{\mathbb{R}^d} 1_X(x) 1_{X-h}(x) dx = |X \cap X_{-h}|$$

Noting that  $K(h) = K(-h)$ , we arrive at the following definition:

**Definition 3.1.2.** *The geometric covariogram of  $X$  is the mapping*

$$K(h) = |X \cap X_h| \quad h \in \mathbb{R}^d$$



**Fig. 3.1.** The area of the dark domain is the geometric covariogram of  $X$  at  $h$ .

**Example 3.1.1.** The geometric covariogram of a 3-dimensional ball of diameter  $a$  is

$$K(h) = \pi \frac{a^3}{6} \left( 1 - \frac{3}{2} \frac{|h|}{a} + \frac{1}{2} \frac{|h|^3}{a^3} \right) 1_{|h| \leq a}$$

The properties of the geometric covariogram of  $X$  include those of a transitive covariogram (Matheron, 1967; Serra, 1982):

**Proposition 3.1.2.** *The geometric covariogram of  $X$  satisfies*

i)  $K(0) = |X|.$

ii)  $K(-h) = K(h).$

iii)  $\int_{\mathbb{R}^d} K(h) dh = |X|^2.$

iv) *for any finite sequence of points  $(x_\alpha)_{\alpha=1,n}$  and for any finite sequence of real numbers  $(\lambda_\alpha)_{\alpha=1,n}$ , we have*

$$\sum_{\alpha,\beta=1}^n \lambda_\alpha \lambda_\beta K(x_\alpha - x_\beta) \geq 0.$$

v) *the  $(d-1)$ -volume of the boundary  $\partial X$  of  $X$  can be obtained by integrating the derivative  $K'_\alpha(0)$  of  $K$  in the direction  $\alpha$  at the origin over all possible directions*

$$- \int_{S_d} K'_\alpha(0) d\alpha = \omega_{d-1} |\partial X|$$

*In this formula,  $\omega_{d-1}$  stands for the  $(d-1)$ -volume of the unit ball of  $\mathbb{R}^{d-1}$  and  $S_d$  is the unit sphere in  $\mathbb{R}^d$ .*

For instance, using formula v), the perimeter  $p(X)$  of  $X$  if  $X \subset \mathbb{R}^2$ , or the surface area  $a(X)$  if  $X \subset \mathbb{R}^3$  are respectively given by

$$p(X) = -\frac{1}{2} \int_0^{2\pi} K'_\alpha(0) d\alpha \quad a(X) = -\frac{1}{\pi} \int_{S_3} K'_\alpha(0) d\alpha$$

**Remark 3.1.1.** It is also possible to define a probabilistic version of the geometric covariogram by putting

$$K(h) = E\{|X \cap X_h|\} \quad h \in \mathbb{R}^d$$

for a random set  $X$  with a.s. finite volume. Its properties are similar. Note in particular

$$- \int_{S_d} K'_\alpha(0) d\alpha = \omega_{d-1} E\{|\partial X|\}$$

## 3.2 The covariance function and the variogram

### 3.2.1 Definitions and general properties

**Definition 3.2.1.** *The random function  $Z$  is said to be second-order stationary if each random variable  $Z(x)$  has a finite mean  $m$  that is independent of  $x$ , and if the covariance between each pair of variables  $Z(x)$  and  $Z(y)$  is finite and depends only on  $x - y$*

$$\text{Cov}\{Z(x), Z(y)\} = C(x - y) \quad x, y \in \mathbb{R}^d$$

The constant  $m$  and the function  $C$  are called respectively the mean value and the covariance function of  $Z$ .

**Remark 3.2.1.** Rather than using the covariance function, people sometimes prefer resorting to a normalized version of it, called the *correlation function*:

$$\rho(h) = \frac{C(h)}{C(0)} \quad h \in \mathbb{R}^d$$

Some elementary properties of a covariance function are given below. A comprehensive review is given in Chilès and Delfiner (1999).

**Proposition 3.2.1.** *The covariance function  $C$  satisfies*

*i)  $C(0) \geq 0$ .*

*ii)  $C(-h) = C(h)$ .*

*iii) The integral of  $C$  over  $\mathbb{R}^d$  is non negative.*

*iv)  $C$  is positive definite. This means that for any finite sequence of points  $(x_\alpha)_{\alpha=1,n}$  and for any finite sequence of real numbers  $(\lambda_\alpha)_{\alpha=1,n}$ , we have*

$$\sum_{\alpha, \beta=1}^n \lambda_\alpha \lambda_\beta C(x_\alpha - x_\beta) \geq 0.$$

The properties of a covariance function are quite similar to those of a transitive covariogram. Indeed, it will be established in chapter 14 that any transitive covariogram is the covariance function of a second-order stationary random function. Statement *i*) holds since  $C(0) = \text{Var}\{Z(x)\}$ . Statement *iv*) is also straightforward:

$$\sum_{\alpha, \beta=1}^n \lambda_\alpha \lambda_\beta C(x_\alpha - x_\beta) = \text{Var} \left\{ \sum_{\alpha=1}^n Z(x_\alpha) \right\} \geq 0$$

The proof of *iii*) which requires spectral arguments (Yaglom, 1986) is omitted. See also the section of next chapter devoted to the integral range.

In general, it is difficult to establish directly that a function is positive definite. Fortunately a classical result by Bochner makes things easier at least for continuous functions.

**Theorem 3.2.1.** *A continuous function  $C$  defined on  $\mathbb{R}^d$  is positive definite if and only if it is the Fourier transform of a positive measure  $\mathcal{X}$  with finite mass*

$$C(h) = \int_{\mathbb{R}^d} e^{i\langle h, u \rangle} d\mathcal{X}(u) \quad h \in \mathbb{R}^d$$

The positive measure  $\mathcal{X}$  is called the *spectral measure* of  $C$ .

For estimation purposes, instead of working in terms of covariance function, we often prefer resorting to a more flexible tool called the variogram (Matheron, 1971; Chauvet, 1994).

**Definition 3.2.2.** *The variogram of  $Z$  is the mapping*

$$\gamma(x - y) = \frac{1}{2} \text{Var}\{Z(x) - Z(y)\} \quad x, y \in \mathbb{R}^d$$

This definition makes sense not only if  $Z$  is second-order stationary, but also if it has *stationary increments* (e.g. Brownian motion). In the stationary case considered here, the variogram and the covariance function of  $Z$  are related by the formula

$$\gamma(h) = C(0) - C(h)$$

from which we derive the main properties of the variogram.

**Proposition 3.2.2.** *The variogram  $\gamma$  satisfies*

i)  $\gamma(0) = 0$ .

ii)  $\gamma(-h) = \gamma(h)$ .

iii)  $\gamma$  is conditionally negative definite. In other words, for any finite sequence of points  $(x_\alpha)_{\alpha=1, n}$  and for any finite sequence of real numbers  $(\lambda_\alpha)_{\alpha=1, n}$  such that  $\sum_{\alpha=1, n} \lambda_\alpha = 0$ , we have

$$\sum_{\alpha, \beta=1}^n \lambda_\alpha \lambda_\beta \gamma(x_\alpha - x_\beta) \leq 0.$$

The proof of iii) relies on the positive definiteness of  $C$ .

$$\begin{aligned} \sum_{\alpha, \beta=1}^n \lambda_\alpha \lambda_\beta \gamma(x_\alpha - x_\beta) &= \sum_{\alpha, \beta=1}^n \lambda_\alpha \lambda_\beta [C(0) - C(x_\alpha - x_\beta)] \\ &= \left[ \sum_{\alpha=1}^n \lambda_\alpha \right]^2 C(0) - \sum_{\alpha, \beta=1}^n \lambda_\alpha \lambda_\beta C(x_\alpha - x_\beta) \\ &= - \sum_{\alpha, \beta=1}^n \lambda_\alpha \lambda_\beta C(x_\alpha - x_\beta) \leq 0 \end{aligned}$$

The following proposition gives a complete characterization of variograms (Schoenberg, 1938; Matheron, 1972; Dellacherie and Meyer, 1983).



**Proposition 3.2.3.**  $\gamma$  is a variogram if and only if for any  $t > 0$  the function  $\exp\{-t\gamma\}$  is a covariance function.

Let us also mention an equivalent of Bochner's theorem for variograms:

**Theorem 3.2.2.** A continuous function defined on  $\mathbb{R}^d$  and vanishing at the origin is conditionally negative definite if and only if it can be written

$$-\gamma(h) = Q(h) + \int_{\mathbb{R}^d} \frac{1 - \cos \langle u, h \rangle}{|u|^2} d\mathcal{X}(u)$$

where  $Q$  is a positive quadratic form and  $\mathcal{X}$  is a positive symmetric measure with no atom at the origin and satisfying

$$\int_{\mathbb{R}^d} \frac{d\mathcal{X}(u)}{1 + |u|^2} < \infty$$

### 3.2.2 The indicator variogram

In this section, the second-order stationary random function  $Z$  is the indicator function of a random set  $X$ . Its variogram can be written as

$$\gamma(x - y) = \frac{1}{2} E[1_X(x) - 1_X(y)]^2$$

or equivalently

$$\gamma(x - y) = \frac{1}{2} [P\{x \in X, y \notin X\} + P\{x \notin X, y \in X\}]$$

Let  $p$  be the mean value of  $Z$ , that is the proportion of points in  $X$ . Since

$$P\{x \in X, y \notin X\} = p - P\{x \in X, y \in X\} = P\{x \notin X, y \in X\}$$

we have

$$P\{x \in X, y \notin X\} = \gamma(x - y) = P\{x \notin X, y \in X\}$$

Suppose now  $x$  and  $y$  close to each other. Then the previous formula shows that both points lie necessarily near the boundary of the random set. This suggests that the variogram of  $X$  conveys information about its boundary. More precisely, let us define the *specific*  $(d-1)$ -volume  $\sigma^{(d)}$  of  $X$  (i.e. specific perimeter in  $\mathbb{R}^2$ , specific surface area in  $\mathbb{R}^3$ ) as the mean  $(d-1)$ -volume of the boundary of  $X$  per unit  $d$ -volume. Starting from formula  $v)$  of proposition 3.1.2 and using ergodicity arguments, Matheron (1967) showed that it can be written

$$\sigma^{(d)} = \frac{1}{\omega_{d-1}} \int_{S_d} \gamma'_\alpha(0) d\alpha,$$

where  $\gamma'_\alpha$  denotes the first derivative of the variogram at the origin in the direction  $\alpha$  and  $S_d$  stands for the unit sphere in  $\mathbb{R}^d$ . If the variogram has parabolic behaviour in the vicinity of the origin, then  $\gamma'_\alpha(0) = 0$  for any  $\alpha$ , and  $\sigma^{(d)} = 0$ . If we also assume  $X$  to be topologically regular, then this implies that the realizations of  $X$  are equal either to  $\mathbb{R}^d$  or to  $\emptyset$ . On the other hand, if  $\gamma'_\alpha(0) = +\infty$  for any  $\alpha$ , then  $\sigma^{(d)} = +\infty$ , which means that the random set is fractal.

The isotropic case when  $\gamma'_\alpha(0)$  does not depend on  $\alpha$  is also of special interest. Let  $\gamma'(0)$  be their common value. Since the  $(d - 1)$ -volume of  $S_d$  is  $d\omega_d$ , we obtain

$$\sigma^{(d)} = \frac{d\omega_d}{\omega_{d-1}}\gamma'(0)$$

This gives

$$\sigma^{(2)} = \pi\gamma'(0) \qquad \sigma^{(3)} = 4\gamma'(0)$$

in two and three dimensions.

Now we show that only certain variograms of random functions can be variograms of random sets. Starting from

$$[1_X(x) - 1_X(y)]^2 = |1_X(x) - 1_X(y)|$$

and using

$$|1_X(x) - 1_X(z)| \leq |1_X(x) - 1_X(y)| + |1_X(y) - 1_X(z)|$$

we have shown that the variogram of a random set must satisfy the triangle inequality

$$\gamma(x - z) \leq \gamma(x - y) + \gamma(y - z)$$

Consequently the gaussian variogram  $\gamma(h) = 1 - \exp\{-|h|^2\}$  cannot be the variogram of a random set. Taking  $x - y = y - z = h$  with  $|h|$  small, this would imply  $4|h|^2 \leq 2|h|^2$  as a first approximation, which is a contradiction. In fact, the triangle inequality is necessary but not sufficient for a conditionally negative definite function to be an indicator variogram. A conjecture has been proposed by Matheron (1993) to characterize the family of indicator variograms.

**Conjecture 3.2.1.** *Among the family of variograms, the indicator variograms fulfill the supplementary inequality:*

$$\sum_{\alpha, \beta=1}^n \epsilon_\alpha \epsilon_\beta \gamma(x_\alpha - x_\beta) \leq 0$$

for any finite sequence of points  $(x_\alpha)_{\alpha=1, n}$  and for any finite sequence of values  $(\epsilon_\alpha)_{\alpha=1, n}$  in  $\{-1, 0, 1\}$  such that  $\sum_{\alpha=1, n} \epsilon_\alpha = 1$ .

At the present time, it is difficult to know what variograms are indicator variograms. We have just seen that the Gaussian variogram cannot be an indicator variogram. In contrast to this, there exist random sets with the exponential variogram  $\gamma(h) = \exp\{-|h|\}$  (e.g. a Poisson tessellation, see chapter 12). Whether the spherical variogram is an indicator variogram is still unknown.

## Exercises

**3.1** What is the transitive covariogram of the function

$$f(x) = \exp\left\{-\frac{|x|^2}{a^2}\right\} \quad x \in \mathbb{R}^d$$

**3.2** Let  $X$  be a polygon, and let  $K$  be its geometric covariogram. Prove directly that

$$-\int_0^{2\pi} K'_\alpha(0) d\alpha = 2|\partial X|$$

**3.3** What is the geometric covariogram of a line segment whose length is exponentially distributed?

**3.4** Show that the random set  $X$  and its complement have the same indicator variogram.

**3.5** Let  $X$  and  $Y$  be two independent random sets with proportions  $p_X$ ,  $p_Y$  and variograms  $\gamma_X$ ,  $\gamma_Y$ .

- 1) What are the proportions of  $X \cap Y$ ,  $X \cup Y$  and  $X \setminus Y$ ?
- 2) Show that the variogram of  $X \cap Y$  is

$$\gamma_{X \cap Y} = p_X \gamma_Y + p_Y \gamma_X - \gamma_X \gamma_Y$$

- 3) Using the result of exercise **3.4**, deduce from 2) that the variogram of  $X \cup Y$  is

$$\gamma_{X \cup Y} = q_X \gamma_Y + q_Y \gamma_X - \gamma_X \gamma_Y$$

where  $q_X = 1 - p_X$  and  $q_Y = 1 - p_Y$ .

- 4) Writing  $X \setminus Y = X \cap Y^c$ , show that

$$\gamma_{X \setminus Y} = p_X \gamma_Y + q_Y \gamma_X - \gamma_X \gamma_Y$$

## 4. The integral range

The integral range is a simple and powerful tool that helps to quantify the statistical fluctuations of a stochastic model. Their knowledge is required to ensure the correctness of any computer program supposed to produce simulations of the model.

We start with the presentation of a very simple estimation problem in order to illustrate the problems encountered. The integral range is then defined starting from the variographic tools of chapter 3, and its main statistical properties are reviewed. Finally, a procedure is proposed for its practical assessment.

### 4.1 An estimation problem

Consider a second-order stationary random function  $Z = (Z(x), x \in \mathbb{R}^d)$  with mean  $m$ , variance  $\sigma^2$  and correlation function  $\rho$ . All these quantities are supposed to be unknown. All we have is one realization  $(z(x), x \in V)$  of  $Z$  in the bounded domain  $V \subset \mathbb{R}^d$ . We want to estimate  $m$ . To do this, we can consider the average of  $z$  over  $V$

$$z(V) = \frac{1}{|V|} \int_V z(x) dx,$$

which is a realization of the estimator

$$Z(V) = \frac{1}{|V|} \int_V Z(x) dx$$

Clearly, this estimator satisfies  $EZ(V) = m$ . In other words, it is unbiased. To calculate its variance we require the correlation between all pairs of points of  $V$ , namely

$$\text{Var } Z(V) = \frac{\sigma^2}{|V|^2} \int_V \int_V \rho(x - y) dx dy$$

We might expect the variance of  $Z(V)$  to vanish when the domain  $V$  becomes infinite. Unfortunately, this is not always the case as the following example shows. Define  $Z(x) = Y$  where  $Y$  is a non-constant random variable. Then

$Z(V) = Y$ , so  $\text{Var } Z(V) = \text{Var } Y > 0$  which does not tend to 0 when  $V$  becomes very large. This is the reason why the concept of ergodicity is introduced.

**Definition 4.1.1.** *The second-order stationary random function  $Z$  is said to be ergodic if*

$$\lim_{V \rightarrow \infty} \text{Var } Z(V) = 0$$

This definition<sup>1</sup> is not the only possibility (for other definitions, see Yaglom (1987), Matheron (1989), Cressie (1991), Chiles and Delfiner (1999)). The one considered here is very weak because it just says that  $Z(V)$  is an asymptotically good estimator of  $m$  in the quadratic mean sense. In particular, it does not indicate how large the domain  $V$  must be for the variance of  $Z(V)$  to become negligible. The following example shows that different, and sometimes unexpected, behaviours can be observed.

**Example 4.1.1.** The function

$$C_\mu(h) = \frac{1}{1 + |h|^\mu} \quad h \in \mathbb{R}$$

is a covariance function in  $\mathbb{R}^d$  when  $0 < \mu \leq 2$  (cf. chapter 15). Suppose that  $Z = (Z(x), x \in \mathbb{R})$  has the covariance function  $C_\mu$ . We are interested in the variance of  $Z([0, l])$  for large values of  $l$ . We can expect this variance to decrease faster and faster as  $\mu$  increases. Indeed this is effectively what happens for  $\mu < 1$  since  $\text{Var } Z([0, l])$  behaves as  $l^{-\mu}$ . But for  $\mu > 1$ ,  $\text{Var } Z([0, l])$  behaves as  $l^{-1}$ . In the intermediate case  $\mu = 1$ , a behaviour as  $\ln l/l$  takes place.

## 4.2 The integral range

**Definition 4.2.1.** *The quantity*

$$A = \lim_{V \rightarrow \infty} |V| \frac{\text{Var } Z(V)}{\sigma^2}$$

*is called the integral range of  $Z$ .*

This limit exists for all usual covariance functions. It is non negative but may be zero or infinite. Because of the term  $\sigma^2$  that appears in the denominator as a normalization factor, the integral range has the dimension of a  $d$ -volume (hence the terminology of correlation length in  $\mathbb{R}$ , or correlation area in  $\mathbb{R}^2$

<sup>1</sup> Strictly speaking, it is necessary to specify what the limit in  $V$  means. We say that  $\lim_{V \rightarrow \infty} f(V) = f_0$  if for any increasing sequence  $(V_n, n \in \mathbb{N})$  of compact and convex subsets such that  $\cup_n V_n = \mathbb{R}^d$ , we have  $\lim_{n \rightarrow +\infty} f(V_n) = f_0$ .

sometimes found in literature; see Yaglom, 1987). It has a simple physical interpretation (Matheron, 1989), at least in the case when it is nonzero and finite. Suppose  $|V|$  very large w.r.t.  $A$ . Then we can write

$$\text{Var } Z(V) \approx \frac{\sigma^2 A}{|V|}$$

Moreover, we can find an integer  $n$  such that

$$\frac{|V|}{A} \approx n$$

Combining these two approximations, we finally have

$$\text{Var } Z(V) \approx \frac{\sigma^2}{n}$$

In this formula, the right-hand side can be interpreted as the variance of an average of  $n$  independent observations with the same variance  $\sigma^2$ . This suggests that the integral range may be considered as a reference  $d$ -volume for predicting whether  $V$  is big enough for the estimation of  $m$  to be sufficiently precise.

The integral range has a simple expression in the case when the correlation function vanishes rapidly.

**Proposition 4.2.1.** *If the correlation function  $\rho$  is integrable, then the integral range can be written as*

$$A = \int_{\mathbb{R}^d} \rho(h) dh$$

**Proof:** Let  $(V_n, n \in \mathbb{N})$  be an increasing sequence of convex compact domains converging to  $\mathbb{R}^d$ . Note that

$$|V_n| \frac{\text{Var } Z(V_n)}{\sigma^2} = \frac{1}{|V_n|} \int_{V_n} \int_{V_n} \rho(x-y) dx dy = \int_{\mathbb{R}^d} \rho(h) \frac{K_n(h)}{K_n(0)} dh$$

where  $K_n$  denotes the geometric covariogram of  $V_n$  (cf. definition 3.1.2). The integrands are bounded by the integrable function  $|\rho|$ . Moreover, they converge to  $\rho(h)$  since the ratios  $K_n(h)/K_n(0)$  converge to 1 (by the convexity of the  $V_n$ 's). Therefore we can apply Lebesgue's dominated convergence theorem to obtain:

$$\lim_{n \rightarrow \infty} |V_n| \frac{\text{Var } Z(V_n)}{\sigma^2} = \int_{\mathbb{R}^d} \lim_{n \rightarrow \infty} \rho(h) \frac{K_n(h)}{K_n(0)} dh = \int_{\mathbb{R}^d} \rho(h) dh$$

A table of the integral ranges for the most common covariance functions is given in Appendix 2 at the end of this book.

### 4.3 Practical assessment of the integral range

In this section, we propose an algorithm to estimate the integral range of a second-order stationary, ergodic random function starting from a realization  $(z(x), x \in V)$ .

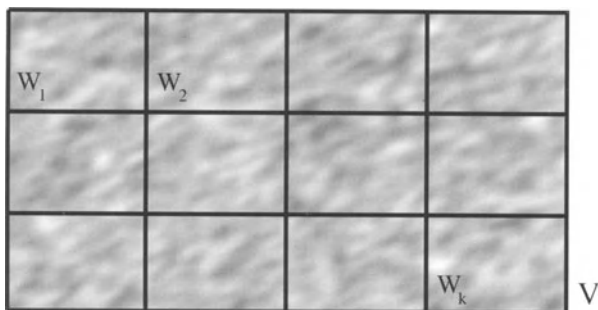
Suppose  $0 < A < \infty$ , and let  $W$  be an arbitrary subdomain of  $V$ . In the case when  $|W| \gg A$ , the variance of  $Z$  over  $W$  can be approximated as

$$\text{Var } Z(W) \approx \frac{\sigma^2 A}{|W|}$$

so that

$$\log \text{Var } Z(W) \approx \log(\sigma^2 A) - \log |W|$$

by taking logarithms. Consequently, the plot of  $\log \text{Var } Z(W)$  versus  $\log |W|$  should be linear with slope  $-1$  for large values of  $\log |W|$ . This is what is tested in practice.



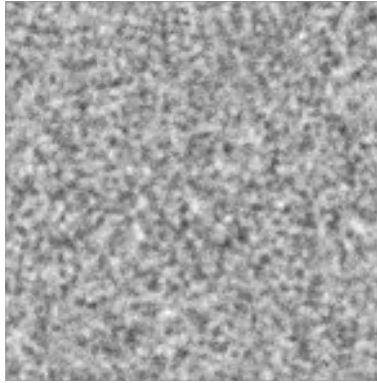
**Fig. 4.1.** The domain  $V$  has been divided into  $k$  subdomains  $W_1, \dots, W_k$ , all congruent to  $W$ . The experimental dispersion variance of the  $z(W_i)$ 's is an estimate of  $\text{Var } Z(W)$

The first question is how to evaluate  $\text{Var } Z(W)$  for each subdomain  $W$ . Suppose that the domain  $V$  can be decomposed into a union of disjoint subdomains  $W_1, \dots, W_k$ , all of the same size, shape and orientation as the domain  $W$  (in other words  $W$  "divides"  $V$ . See Figure 4.1). Then  $\text{Var } Z(W)$  can be estimated by the experimental *dispersion variance*<sup>2</sup> of the  $z(W_i)$ 's, defined as

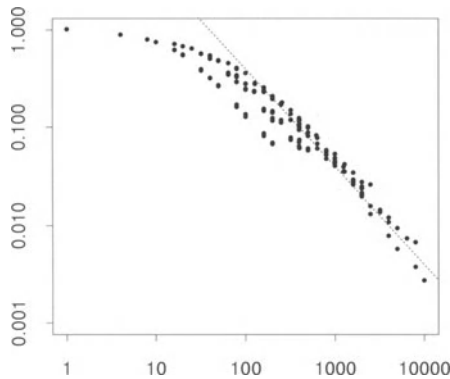
$$s^2(W|V) = \frac{1}{k} \sum_{i=1}^k [z(W_i) - z(V)]^2$$

<sup>2</sup> The corresponding estimator is  $S^2(W|V) = \frac{1}{k} \sum_{i=1}^k [Z(W_i) - Z(V)]^2$ . Its mean value  $ES^2(W|V) = \text{Var } Z(W) - \text{Var } Z(V)$  is precisely what Matheron (1970, 1978) calls a dispersion variance. In the case when  $V$  is very large, the term  $\text{Var } Z(V)$  can be neglected and we have  $ES^2(W|V) \approx \text{Var } Z(W)$ .

Consequently, for each division of  $V$  into subdomains, we begin by determining the average of  $z$  in each subdomain, then calculate their experimental dispersion variance and finally plot its logarithm against the logarithm of the size of the subdomains.



**Fig. 4.2.** Realization of a random function with a finite, nonzero, integral range (spherical covariance with a scale factor of 10). The simulation field is  $400 \times 400$



**Fig. 4.3.** Plot of the experimental dispersion variance versus the subdomain area in log-log coordinates. The experimental points tend to be aligned along a straight line with slope  $-1$ . In this case, the mean of the random function can be estimated with good precision

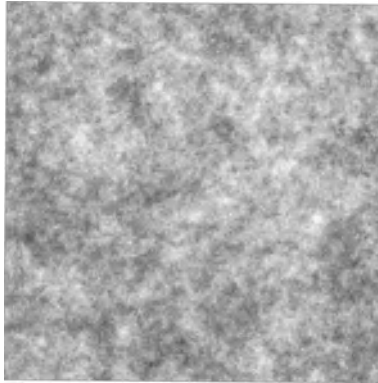
This algorithm has been applied to a realization of a two-dimensional random function with finite and nonzero integral range (see Figure 4.2). The experimental test is shown on Figure 4.3. Several points lie on the same verticals. They correspond to rectangular subdomains with the same area but with different side lengths. Note also that there are several sequences of points that



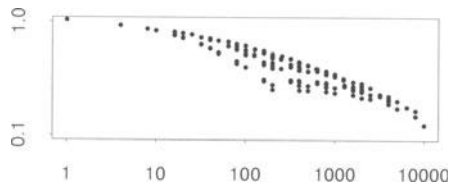
appear to lie on monotonically decreasing curves. This is the consequence of the additivity formula  $s^2(X|V) = s^2(X|W) + s^2(W|V)$  that holds if  $X$  divides  $W$  and  $W$  divides  $V$ . For  $|W|$  large enough, the points tend to lie on a straight line with slope  $-1$ . This plot does not contradict a conclusion that the simulation is part of a realization of a model with finite, nonzero integral range. From the equation  $\log \text{Var}\{Z(W)\} \approx 1.7 - \log |W|$  thus obtained for large values of  $\log |W|$ , we may deduce  $\sigma^2 A \approx 10^{1.7} = 50.12$  by putting  $|W| = 1$ . We then divide by the experimental dispersion variance of a point in  $V$  (0.98) to get 51.14 as the estimate of  $A$ . In the case considered here, the actual value of the integral range is 62.83.

Starting from the estimate of  $\sigma^2 A$ , it is not difficult to choose  $V$  so that  $\text{Var} Z(V)$  is smaller than a prespecified value.

It remains to show that the algorithm does not lead to misleading estimates when the condition  $0 < A < \infty$  is not satisfied.



**Fig. 4.4.** Realization of a random function with an infinite integral range (hyperbolic covariance with a scale factor of 10). The simulation field is  $400 \times 400$



**Fig. 4.5.** Plot of the experimental dispersion variance versus the subdomain area in log-log coordinates. The experimental points do not lie along a straight line with slope  $-1$ . The estimation variance of the mean decreases slowly as a function of the subdomain area. Starting from the simulation of Figure 4.4, the mean can be estimated with low precision

Figure 4.4 shows a realization of a random function with hyperbolic covariance function

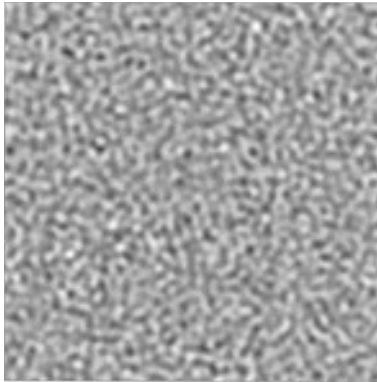
$$C(h) = \left(1 + \frac{|h|}{a}\right)^{-1} \quad h \in \mathbb{R}^2$$

This covariance has been chosen because it has an infinite integral range. The application of the algorithm to that simulation gives the experimental plot on Figure 4.5. The points obtained no longer lie on a straight line with slope  $-1$ . In this case two interpretations are possible. Firstly, the integral range is finite and nonzero, but the subdomains are not sufficiently large to estimate it properly. Secondly, the integral range is infinite (which is actually the case for this simulation). In either case, only a low-precision estimate of the mean is possible.

We have already mentioned the existence of random functions with a zero integral range. The realization on Figure 4.6 has been built using the covariance function

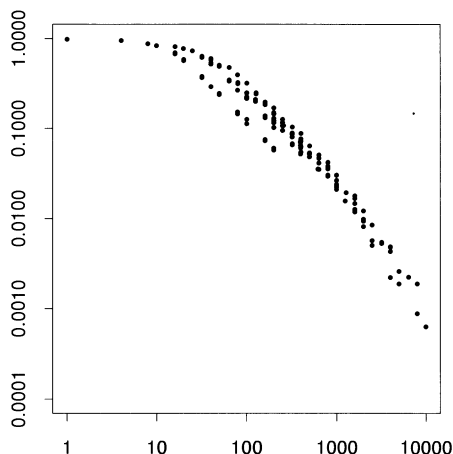
$$C(h) = \exp\left\{-\frac{|h|^2}{a^2}\right\} \left(1 - \frac{|h|^2}{a^2}\right) \quad h \in \mathbb{R}^2$$

with  $a = 10$ .



**Fig. 4.6.** realization of a random function with an zero integral range. The simulation field is  $400 \times 400$

Figure 4.7 shows the plot obtained by applying the algorithm. For large subdomains, the points lie on a line with a slope steeper than  $-1$ , which is very favourable for the estimation of the mean. Note however that if the subdomain area had been limited to 500, then a line with slope  $-1$  would have been observed, which would not have invalidated the existence of a finite, nonzero integral range. It would have resulted in a rather conservative estimation of the mean.



**Fig. 4.7.** Plot of the experimental dispersion variance versus the subdomain area in log-log coordinates. The experimental points lie along a straight line with a slope steeper than  $-1$ . The estimation variance of the mean decreases rapidly as a function of the subdomain area. From the simulation of Figure 4.6, the mean can be estimated with very good precision

We conclude this section with an example of a random function for which the second moment is easier to estimate than the mean.

**Example 4.3.1.** Let  $Z$  be a unidimensional standardized gaussian random function (see chapter 15) with the hyperbolic covariance function

$$C(h) = \frac{1}{1 + |h|} \quad h \in \mathbb{R}$$

Ten independent simulations have been carried out in the segment  $[0, l]$  with  $l = 1000$ . For each simulation, estimates of the first two moments

$$z([0, l]) = \frac{1}{l} \int_0^l z(x) dx \quad z_2([0, l]) = \frac{1}{l} \int_0^l z^2(x) dx$$

have been calculated. Here are the results

$z([0, l])$	$z_2([0, l])$
-0.088	0.916
-0.239	1.087
0.119	1.046
-0.032	1.023
0.119	1.071
-0.027	1.036
0.026	0.972
-0.108	1.037
0.044	0.979
0.220	1.118

It turns out that the mean estimates are rather erratic, whereas those of the second moment are not so dispersed. Why does this happen? The second moment of  $Z$  can be seen as the mean of  $Z^2$ . In the gaussian case,  $Z^2$  is a second-order stationary random function with mean 1 and covariance function

$$C_2(h) = \frac{2}{(1 + |h|)^2} \quad h \in \mathbb{R}$$

Now, it should be noted that the integral range of  $Z$  is infinite, whereas that of  $Z^2$  is finite ( $A_2 = 2$ ). This is what makes the difference.

## Exercises

**4.1** Let  $Z$  be a second-order stationary random function in  $\mathbb{R}^d$ , and let  $V$  be a domain in  $\mathbb{R}^d$  that is partitioned into  $n$  congruent subdomains  $(W_1, \dots, W_n)$ .  
1) Suppose  $W$  is a random subdomain uniformly selected among the  $W_i$ 's. Show that

$$E\{Z(W) | Z(V)\} = Z(V)$$

2) Deduce that the distribution  $F_W$  of  $Z(W)$  is less *dispersed* than that  $F_V$  of  $Z(V)$ , in the sense where

$$\int \varphi(z) dF_V(z) \leq \int \varphi(z) dF_W(z)$$

for any convex function  $\varphi$ . (Hint: use Jensen's inequality).

3) Derive some consequences of this inequality, in particular regarding the means and the variances of  $F_V$  and  $F_W$ .

**4.2** Suppose that the correlation function  $\rho$  is integrable.

- 1) Show that  $\rho$  admits a spectral density, say  $f$ .
- 2) Show that the integral range of  $\rho$  is equal to  $(2\pi)^d f(0)$ .

## 5. Basic morphological concepts

The purpose of this chapter is to introduce the two basic concepts of mathematical morphology, namely *dilations* and *erosions*. Three possible uses are given. Firstly, dilation and erosion can be combined to produce two other morphological concepts (*openings* and *closings*) which have rich structural content. Secondly, the Hausdorff distance between objects has a simple morphological interpretation. Finally, the chance of detecting an object by regular sampling can be simply expressed in terms of dilations.

For further reading on mathematical morphology, the reader can consult Matheron (1967, 1975) and Serra (1982).

Here is a brief summary of the notation that will be used throughout this chapter. The workspace is the  $d$ -dimensional euclidean space, and  $o$  stands for its origin. Let  $A \subset \mathbb{R}^d$ . Its complement in  $\mathbb{R}^d$  and its symmetrical set w.r.t.  $o$  are denoted by  $A^c$  and  $\check{A}$  respectively. Given  $x \in \mathbb{R}^d$ ,  $A_x$  is the set obtained from  $A$  by translation through vector  $x$ . If  $B$  is subset of  $\mathbb{R}^d$ , then

$$A \oplus B = \bigcup_{y \in B} A_y \qquad A \ominus B = \bigcap_{y \in B} A_y$$

are respectively the sum and the difference of  $A$  by  $B$  in the sense of Minkowski.

### 5.1 Dilation and erosion

**Definition 5.1.1.** *The point  $x$  is said to belong to the set  $A$  dilated by  $B$  if  $B_x$  hits  $A$ .*

The relation  $B_x \cap A \neq \emptyset$  holds if and only if there exists  $y \in B$  such that  $x + y \in A$ , or equivalently such that  $x \in A_{-y}$ . Consequently, a dilation can be expressed as a Minkowski sum

$$\{x \mid B_x \cap A \neq \emptyset\} = A \oplus \check{B}$$

Note that dilation is distributive w.r.t. set union

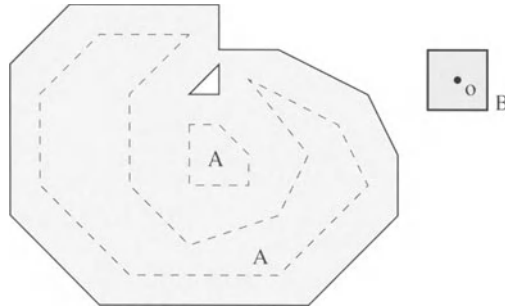


Fig. 5.1. Example of a set dilated by a square.

$$(A \cup A') \oplus \check{B} = (A \oplus \check{B}) \cup (A' \oplus \check{B})$$

but not w.r.t. intersection (two dilated sets can intersect even if the corresponding original sets are disjoint). Nonetheless the following inclusion holds

$$(A \cap A') \oplus \check{B} \subset (A \oplus \check{B}) \cap (A' \oplus \check{B})$$

**Definition 5.1.2.** *The point  $x$  is said to belong to the set  $A$  eroded by  $B$  if  $B_x$  is totally contained within  $A$ .*

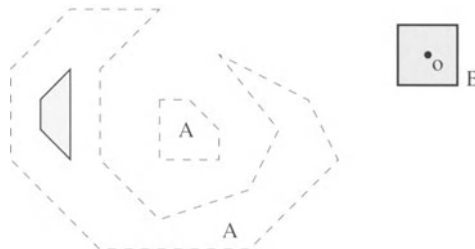


Fig. 5.2. Example of a set eroded by a square

Erosion can be written as a Minkowski difference

$$\{x \mid B_x \subset A\} = A \ominus \check{B}$$

Dilation and erosion are dual operation w.r.t. complementation:

$$(A \ominus \check{B})^c = A^c \oplus \check{B} \quad (A \oplus \check{B})^c = A^c \ominus \check{B}$$

Indeed, this amounts to writing  $B_x \subset A \iff B_x \cap A^c = \emptyset$ .

By duality, the dilation properties become

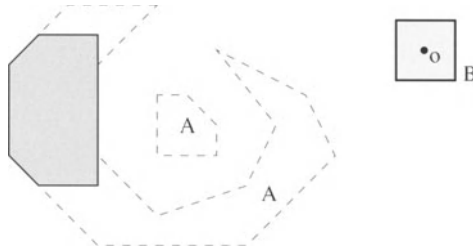
$$(A \cap A') \ominus \check{B} = (A \ominus \check{B}) \cap (A' \ominus \check{B})$$

$$(A \cup A') \ominus \check{B} \supset (A \ominus \check{B}) \cup (A' \ominus \check{B})$$

## 5.2 Opening and Closing

**Definition 5.2.1.** *The point  $x$  is said to belong to the set  $A$  opened by  $B$  if it belongs to a translate of  $B$  that is totally contained within  $A$ .*

In other words, the set  $A$  opened by  $B$  is the set swept out by all the translates of  $B$  included in  $A$ . It is denoted by  $A_B$ .



**Fig. 5.3.** Example of a set opened by a square

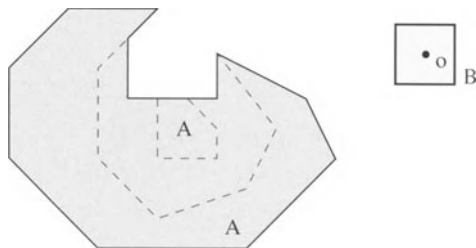
The definition of opening shows that this operation is

- *anti-extensive*  $A_B \subset A$
- *increasing*  $A \subset A' \implies A_B \subset A'_B$
- *idempotent*  $(A_B)_B = A_B$

Suppose  $x \in A_B$ . Then there exists  $y \in \mathbb{R}^d$  such that  $x \in B_y$  and  $B_y \subset A$ . This exactly means that  $y \in (\check{B})_x$  and  $y \in A \ominus \check{B}$ . Equivalently,  $(\check{B})_x \cap (A \ominus \check{B})$  is nonempty. Therefore an opening can be expressed as a combination of an erosion and a dilation:

$$A_B = (A \ominus \check{B}) \oplus B$$

**Remark 5.2.1.** Note in this formula the erosion by  $B$  and the dilation by  $\check{B}$ . This is necessary for recentering reasons: if  $B = \{h\}$ , then  $A \ominus \check{B} = A_{-h}$  and  $A_B = A_{-h+h} = A_o = A$ .



**Fig. 5.4.** Example of a set closed by a square

**Definition 5.2.2.** *The point  $x$  is said to belong to the set  $A$  closed by  $B$  if every translate of  $B$  that contains  $x$  hits  $A$ .*

The set  $A$  closed by  $B$  is denoted by  $A^B$ . Note that  $x \notin A^B$  if and only if every translate of  $B$  that contains  $x$  is totally contained within the complement  $X^c$  of  $X$ , that is  $x \in A_{B^c}$ . This shows that opening and closing are dual operations. Opening a set is equivalent to closing its complement:

$$(A_B)^c = (A^c)^B \qquad (A^B)^c = (A^c)_B$$

By duality, it is immediately clear that closing is

- *extensive*  $A \subset A^B$
- *increasing*  $A \subset A' \implies A^B \subset A'^B$
- *idempotent*  $(A^B)^B = A^B$

Moreover, a closing is a combination of a dilation and an erosion

$$A^B = (A \oplus \check{B}) \ominus B$$

### 5.3 Hausdorff distance

The aim of this section is to define a distance between objects (nonempty compact subsets) of  $\mathbb{R}^d$ . Let us start with the euclidean distance  $d(x, y)$  between any two points  $x$  and  $y$  of  $\mathbb{R}^d$ . The next step is to consider the distance from  $x$  from an object  $Y$

$$d(x, Y) = \min_{y \in Y} d(x, y)$$

Now, let  $X$  be another object. The maximum distance from a point of  $X$  to  $Y$  is

$$\max_{x \in X} d(x, Y) = \max_{x \in X} \min_{y \in Y} d(x, y)$$



**Definition 5.3.1.** *The Hausdorff distance between both objects  $X$  and  $Y$  is*

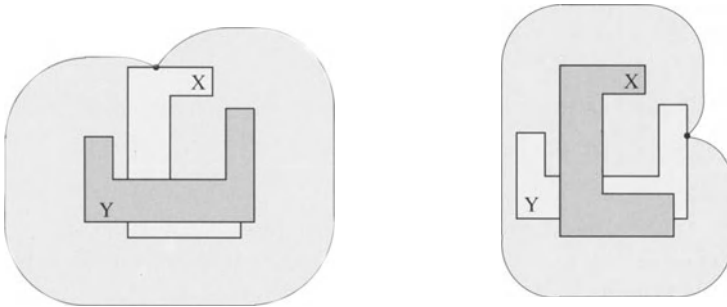
$$d(X, Y) = \max \left( \max_{x \in X} d(x, Y), \max_{y \in Y} d(y, X) \right)$$

It can be readily shown that  $d$  satisfies the properties of a distance function (positivity, symmetry, and triangle inequality). Moreover, the Hausdorff distance generalizes the euclidean distance, in the sense that  $d(\{x\}, \{y\}) = d(x, y)$  for any pair of points  $x$  and  $y$ . This is the reason why the same notation  $d$  has been used for the euclidean and for the Hausdorff distances.

Let  $rB$  be the ball of radius  $r$  centered at the origin. Observe that  $d(x, Y) = \min\{r \geq 0 \mid x \in Y \oplus rB\}$ . Accordingly  $\max_{x \in X} d(x, Y) = \min\{r \geq 0 \mid X \subset Y \oplus rB\}$ , and finally

$$d(X, Y) = \min\{r \geq 0 \mid X \subset Y \oplus rB, Y \subset X \oplus rB\}$$

which gives a geometric interpretation of the Hausdorff distance in terms of dilations by balls (see Figure 5.5).



**Fig. 5.5.** The diagram on the left shows the object  $Y$  dilated by a ball until it just covers the object  $X$ ; similarly on the right, the object  $X$  is dilated until it just covers  $Y$ . The Hausdorff distance between  $X$  and  $Y$  is the minimum of the radii of these two balls.

## 5.4 Object detection

The problem addressed in this section is the detection of an object  $A \subset \mathbb{R}^d$  from experimental information at the nodes of a regular grid. Assuming that its mesh sizes  $a = (a_1, \dots, a_d)$  have been fixed, then the nodes of the grid can be written

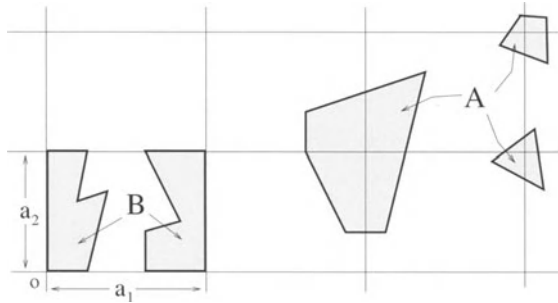
$$x + ka = (x_1 + k_1 a_1, \dots, x_d + k_d a_d)$$

with  $k = (k_1, \dots, k_d) \in \mathbb{Z}^d$ . The point  $x = (x_1, \dots, x_d)$  is the unique grid node contained within the set

$$D = \{(x_1, \dots, x_d) \mid 0 \leq x_1 < a_1, \dots, 0 \leq x_d < a_d\}$$

It is taken as the reference node of the grid. The grid itself is denoted by  $G_x$ . In particular  $G = G_o$  is the grid referenced at the origin.

The object  $A$  is said to be detected by the sampling grid  $G_x$  if  $G_x \cap A \neq \emptyset$ , or equivalently if  $x \in A \oplus \check{G}$ . Since  $x \in D$  and  $\check{G} = G$ , it follows that  $A$  is detected if and only if  $x \in B = (A \oplus G) \cap D$  (cf. Figure 5.6).



**Fig. 5.6.** The object  $A$  is detected if and only if the reference node of the grid is located within  $B$ .

Of course, the sampling grid has an arbitrary position over the object  $A$ . This can be expressed by assuming that the reference node  $x$  is uniformly located within  $D$ . Then the probability for  $A$  to be detected is

$$p(A) = \frac{|B|}{|D|}$$

It should be pointed out that this probability does not involve the  $d$ -volume of  $A$  but only that of  $B$ . This is particularly clear in the case of Figure 5.6 where the two small components of  $A$  do not improve its chance of detection. Indeed a point  $y$  belongs to  $B$  if there exists  $k \in \mathbb{Z}^d$  such that  $y + ka \in A$ . From this point of view,  $B$  can be interpreted as the set  $A$  modulo  $a$ .

It is also possible to estimate the volume of an object by counting the number of points of the sampling grid that fall within it. Transitive geostatistics has been developed for this purpose. For more information, the reader should consult Matheron (1965) or Ki u (1997).

## Exercises

**5.1** Suppose that the part of a set  $A$  located outside a domain  $D$  is unknown. How does this affect the result of the dilation of  $A$  w.r.t  $B$ ? More precisely, establish the following formula

$$(X \oplus \check{B}) \cap (D \ominus \check{B}) = ((X \cap D) \oplus \check{B}) \cap (D \ominus \check{B})$$

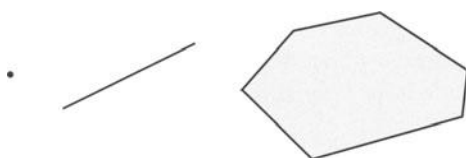
**5.2** What is the Hausdorff distance between two balls?

## 6. Stereology: some basic notions

It frequently happens that real-world objects can only be investigated through planar or even linear cross-sections (thin slides, polished sections, outcrops, drill-holes,...). Stereology develops formulae to retrieve 3D information from the available 1D or 2D data. A stereological formula is usually based on a model which makes assumptions about the geometric properties of the objects under study. Of course, the more specific the model, the more information can be retrieved. Two different models are presented here. The first one deals with convex objects, the second one with polyconvex objects. Both models are treated in Hadwiger (1957). Other models can be found in Santaló (1977) and Serra (1982).

### 6.1 The convex model

**Definition 6.1.1.** *A subset of  $\mathbb{R}^d$  is said to be convex if it contains the segment joining any pair of its points.*



**Fig. 6.1.** Some examples of convex sets

#### 6.1.1 The Minkowski functionals

Let  $\mathcal{C}_d$  be the family of the compact convex subsets of  $\mathbb{R}^d$ ,  $d = 1, 2, 3, \dots$ . This family is stable under intersection (but not under union), Minkowski sum and difference, projection onto an  $i$ -flat (affine subspace of dimension  $i$ ), and intersection with an  $i$ -flat. We are interested in scalar functions defined on  $\mathcal{C}_d$  that are

- increasing:  $K \subset K'$  implies  $\phi(K) \leq \phi(K')$ .
- additive:  $\phi(K \cup K') + \phi(K \cap K') = \phi(K) + \phi(K')$  if  $K \cup K'$  is convex.
- motion invariant.

The  $d$ -volume of a compact convex set or the  $(d - 1)$ -volume of its boundary are two simple examples. This class of functions has the structure of a convex cone. The following theorem due to Minkowski gives a generating family:

**Theorem 6.1.1.** *Any function  $\phi$  defined on  $\mathcal{C}_d$  which is increasing, additive and motion invariant, is necessarily a positive linear combination of  $d + 1$  functions denoted by  $W_0, W_1, \dots, W_d$*

$$\phi(K) = \sum_{i=0}^d \lambda_i W_i(K)$$

Moreover the functional  $W_i$  is homogeneous with degree  $d - i$ . If  $\lambda > 0$ , then  $W_i(\lambda K) = \lambda^{d-i} W_i(K)$ .

These scalar functions are usually called the *Minkowski functionals*. They are defined up to a multiplicative constant. Conventionally, normalization factors are chosen in such a way that all functionals assign the same value to the unit ball of  $\mathbb{R}^d$ , namely its volume

$$\omega_d = \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2} + 1)}$$

Here  $\Gamma$  is the classic gamma function

$$\Gamma(\alpha) = \int_0^{+\infty} e^{-t} t^{\alpha-1} dt \quad (\alpha > 0)$$

It satisfies  $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$ , with  $\Gamma(1) = 1$  and  $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ . Several Minkowski functionals have very simple physical interpretations:

- $W_0(K) = |K|$  is the  $d$ -volume of  $K$ .
- $W_1(K) = |\partial K|/d$ , where  $|\partial K|$  denotes the  $(d - 1)$ -volume of the boundary of  $K$  (perimeter in 2D, surface area in 3D).
- $W_{d-1}(K) = \omega_d b(K)/2$ , where  $b(K)$  stands for the *mean breadth* of  $K$ . If  $b(K, u)$  denotes the length of the projection of  $K$  onto a line with direction  $u$  (see Figure 6.2), we have

$$b(K) = \frac{1}{d\omega_d} \int_{S_d} b(K, u) du$$

- $W_d(K) = \omega_d$  if  $K \neq \emptyset$  and 0 otherwise.

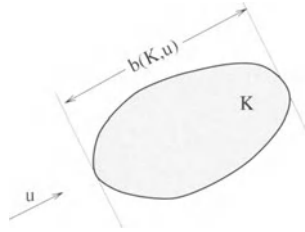


Fig. 6.2. Breadth of  $K$  in direction  $u$

**Example 6.1.1.** In the case  $d = 2$ , let us write  $W_1(K) = W_{2-1}(K)$ . This gives at once

$$b(K) = \frac{1}{\pi} |\partial K|,$$

which is a particular case of proposition 3.1.2.,  $v$ ). For a triangle with sides  $a, b$  and  $c$ , we obtain

$$b(K) = \frac{a + b + c}{\pi}$$

**6.1.2 The formulae**

**The projection formula.** Let  $\dot{S}_i$  be an  $i$ -subspace with a uniform orientation: all of the  $i$ -subspaces have the same chance to be selected. The projection  $K \downarrow \dot{S}_i$  of  $K \in \mathcal{C}_d$  onto  $\dot{S}_i$  is a random compact convex set. We denote by  $W_j^{(i)}(K \downarrow \dot{S}_i)$  its Minkowski functionals in  $\dot{S}_i$ . Their mean values are given by *Cauchy's formula*:

$$E \left\{ W_j^{(i)}(K \downarrow \dot{S}_i) \right\} = \frac{\omega_i}{\omega_d} W_{d+j-i}(K) \quad 0 \leq j \leq i < d$$

Here is a sketch of the proof. We start by proving that the mean value being sought is increasing, additive and motion invariant on  $\mathcal{C}_d$ . According to Minkowski's theorem, this function is a positive linear combination of the Minkowski functionals. The positive coefficients assigned to the functionals are identified by considering the special case when  $K$  is a ball.

**Remark 6.1.1.** The case  $j = 0$  gives the mean projected  $i$ -volume of  $K$

$$E \left\{ |K \downarrow \dot{S}_i| \right\} = \frac{\omega_i}{\omega_d} W_{d-i}(K) \quad 0 \leq i < d$$

Some authors (Miles, 1972) use this formula to define the Minkowski functionals.

**Remark 6.1.2.** The usual form of Cauchy's formula is

$$\frac{1}{C_{id}} \int W_j^{(i)}(K \downarrow S_i) dS_i = \frac{\omega_i}{\omega_{d-i}} W_{d+j-i}(K) \quad 0 \leq j \leq i < d$$

with

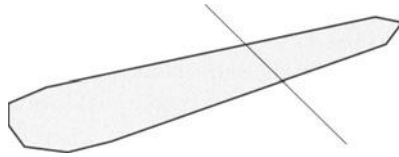
$$C_{id} = \frac{\omega_0}{\omega_d} \binom{d}{i} \frac{\prod_{j=0}^i \omega_{d-j}}{\prod_{j=0}^i \omega_j}$$

The integral is taken over all possible  $i$ -subspaces of  $\mathbb{R}^d$ . The measure  $dS_i$  is not described here (cf. Hadwiger, 1957; Santaló, 1976). Suppose  $K \neq \emptyset$ . In the case  $j = i$ ,  $W_j^{(i)}(K \downarrow S_i) = \omega_i$ ,  $W_{d+j-i}(K) = \omega_d$ , and this integral becomes

$$\frac{1}{C_{id}} \int dS_i = \frac{\omega_d}{\omega_{d-i}} \quad 0 \leq i < d$$

The expected value given above is just the ratio between the two integrals.

**The section formula.** All of the  $i$ -flats (affine subspaces of dimension  $i$ ) hitting  $K$  are now considered. Each of them can be given the same chance to be selected. We thus define a *uniform  $i$ -flat hitting  $K$* , denoted by  $\check{F}_i$ . Of course, the orientation of a uniform  $i$ -flat is not uniform. For instance, the elongated convex set of Figure 6.3 has more chance to be hit by a line orthogonal to its main axis than by a line parallel to it.



**Fig. 6.3.** In general, a uniform line hitting a set does not have a uniform orientation

Let  $\check{F}_i$  be a uniform  $i$ -flat hitting  $K$ . Crofton's formula (Hadwiger, 1957) gives the mean value of the Minkowski functionals of the sections of  $K$ :

$$E \left\{ W_j^{(i)}(K \cap \check{F}_i) \right\} = \frac{\omega_i \omega_{d-j}}{\omega_{d-i} \omega_{i-j}} \frac{W_j(K)}{W_i(K)} \quad 0 \leq j \leq i < d$$

**Example 6.1.2.** What is the mean linear section of  $K$ ? Let us apply Crofton's formula with  $i = 1$  and  $j = 0$ . This gives

$$E \left\{ W_0^{(1)}(K \cap \check{F}_1) \right\} = \frac{\omega_d}{\omega_{d-1}} \frac{W_0(K)}{W_1(K)}$$

The Minkowski functionals are then replaced by their physical interpretation, and we get

$$E \left\{ W_0^{(1)}(K \cap \check{F}_1) \right\} = \frac{d \omega_d}{\omega_{d-1}} \frac{|K|}{|\partial K|}$$

**Remark 6.1.3.** The standard form for Crofton’s formula is

$$\frac{1}{C_{id}} \int W_j^{(i)}(K \cap F_i) dF_i = \frac{\omega_i}{\omega_{d-i}} \frac{\omega_{d-j}}{\omega_{i-j}} W_j(K) \quad 0 \leq j \leq i < d,$$

It does not involve a mean value, but integration over all possible  $i$ -flats of  $\mathbb{R}^d$ . For more information, see Hadwiger (1957) or Santaló (1976).

**The dilation formula.** Let  $B \in \mathcal{C}_d$ . Define  $\dot{B}$  as the random set obtained from  $B$  by a uniform rotation. If  $K \in \mathcal{C}_d$ , then  $K \oplus \dot{B}$  is a random compact convex set. Its mean Minkowski functionals are given by Steiner’s formula

$$E\{W_i(K \oplus \dot{B})\} = \frac{1}{\omega_d} \sum_{j=0}^{d-i} \binom{d-i}{j} W_{i+j}(K) W_{d-j}(B)$$

In the particular case when  $B$  is the ball of radius  $\lambda$  centered at the origin (denoted by  $\lambda B$ ), the previous formula simplifies to:

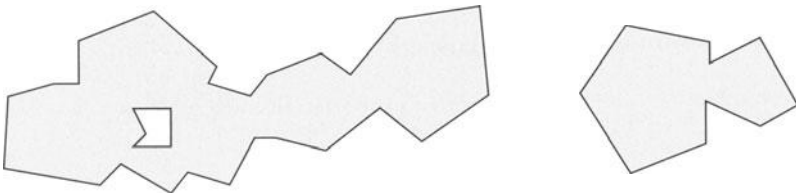
$$W_i(K \oplus \lambda B) = \sum_{j=0}^{d-i} \binom{d-i}{j} W_{i+j}(K) \lambda^j$$

**Remark 6.1.4.** As an application of Steiner’s formula, let  $K \in \mathcal{C}_{d-1}$ . We also have  $K \in \mathcal{C}_d$ , so its Minkowski functionals can be defined either in  $\mathbb{R}^d$  or in  $\mathbb{R}^{d-1}$ . Both sets of functionals are related by the formula

$$W_j^{(d)}(K) = \frac{j}{d} \frac{\omega_j}{\omega_{j-1}} W_{j-1}^{(d-1)}(K) \quad 1 \leq j \leq d$$

## 6.2 The polyconvex model

**Definition 6.2.1.** A subset of  $\mathbb{R}^d$  is said to be polyconvex if it is a finite union of convex sets.



**Fig. 6.4.** Example of a polyconvex set.



Let  $\mathcal{R}_d$  be the family of all compact, polyconvex subsets of  $\mathbb{R}^d$ .  $\mathcal{R}_d$  is stable under union and intersection, projection onto  $i$ -flats and dilation (but not erosion!).  $\mathcal{R}_d$  is sometimes called the *convex ring*.

Can the Minkowski functionals be extended to  $\mathcal{R}_d$ ? If so, are the previous formulae still valid? The answer to these questions is the object of the next sections.

### 6.2.1 Euler-Poincaré Characteristic

Let  $\mathcal{X}$  denote the function defined on  $\mathcal{C}_d$  by  $\mathcal{X}(K) = 1$  if  $K$  is not empty and 0 if not. This function can be extended to the convex ring in the following way. Let  $K \in \mathcal{R}_d$ . By the very definition of the convex ring, there exists a finite sequence  $(K_i, i \in I)$  in  $\mathcal{C}_d$  such that  $K = \bigcup_{i \in I} K_i$ . We put

$$\mathcal{X}(K) = \sum_{J \subset I} (-1)^{\#J-1} \mathcal{X}(K_J)$$

where  $K_J = \bigcap_{j \in J} K_j \in \mathcal{C}_d$ . One can verify that the value  $\mathcal{X}(K)$  thus obtained does not depend on the decomposition of  $K$  into convex sets.  $\mathcal{X}$  is called the *Euler-Poincaré characteristic* of  $K$ .

The function  $\mathcal{X}$  has been built to be additive

$$\mathcal{X}(K \cup K') + \mathcal{X}(K \cap K') = \mathcal{X}(K) + \mathcal{X}(K') \quad K, K' \in \mathcal{R}_d$$

Moreover, it has an interesting topological interpretation:

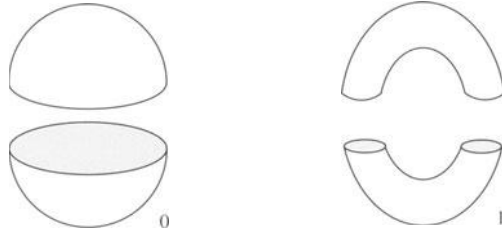
- In  $\mathbb{R}$ ,  $\mathcal{X}(K)$  is the number of connected components of  $K$ .
- In  $\mathbb{R}^2$ ,  $\mathcal{X}(K)$  is equal to the number of connected components of  $K$  minus the number of its holes.
- In  $\mathbb{R}^3$ ,  $\mathcal{X}(K)$  is equal to the number of connected components of  $K$  minus its genus, that is the maximum number of closed curves that can be drawn on its surface without changing its connectivity (see Figure 6.5).

### 6.2.2 Extension of the Minkowski functionals

The extension rests on the additivity property. Because of it, we should have

$$W_i(K) = \sum_{J \subset I} (-1)^{\#J-1} W_i(K_J)$$

for any  $K \in \mathcal{R}_d$ . It remains to show that the righthand side of the equation does not depend on the decomposition of  $K$  into convex sets.



**Fig. 6.5.** Let  $K$  be a three-dimensional compact subset with a smooth boundary. The genus of  $K$  is defined as the maximum number of closed curves that can be drawn on its surface without changing its connectivity. For instance, the genus of a ball is equal to 0. That of a torus is 1

In the case  $i = d$ , notice that  $W_d(K) = \omega_d \mathcal{X}(K)$  if  $K \in \mathcal{C}_d$ . As  $\mathcal{X}$  is additive, this formula is also valid in  $\mathcal{R}_d$ .

Suppose now  $i < d$ . Since  $K_J \in \mathcal{C}_d$ , Crofton's formula gives

$$W_i(K_J) = \frac{1}{\omega_i C_{id}} \int W_i^{(i)}(K_J \cap F_i) dF_i$$

by integrating over all  $i$ -flats of  $\mathbb{R}^d$ . But  $W_i^{(i)}(K_J \cap F_i) = \omega_i \mathcal{X}(K_J \cap F_i)$ .

Therefore

$$W_i(K_J) = \frac{1}{C_{id}} \int \mathcal{X}(K_J \cap F_i) dF_i$$

so that  $W_i(K)$  can be rewritten

$$\begin{aligned} W_i(K) &= \sum_{J \subset I} (-1)^{\#J-1} \frac{1}{C_{id}} \int \mathcal{X}(K_J \cap F_i) dF_i \\ &= \frac{1}{C_{id}} \int \sum_{J \subset I} (-1)^{\#J-1} \mathcal{X}(K_J \cap F_i) dF_i \\ &= \frac{1}{C_{id}} \int \mathcal{X}(K \cap F_i) dF_i, \end{aligned}$$

which does not depend on the decomposition of  $K$ . Therefore the following definition makes sense.

**Definition 6.2.2.** *The Minkowski functionals of  $K \in \mathcal{R}_d$  are defined by*

$$\begin{aligned} W_i(K) &= \frac{1}{C_{id}} \int \mathcal{X}(K \cap F_i) dF_i \quad 0 \leq i \leq d-1 \\ W_d(K) &= \omega_d \mathcal{X}(K) \end{aligned}$$

**Remark 6.2.1.** The extension of the Minkowski functionals to  $\mathcal{R}_d$  does not change their physical interpretation. For instance  $W_0(K)$  is the  $d$ -volume of  $K$  and  $dW_1(K)$  the  $(d-1)$ -volume of its boundary.

The extension of Crofton's formula to  $\mathcal{R}_d$  is now straightforward by additivity. If  $K \in \mathcal{R}_d$  and  $0 \leq j \leq i < d$ , then

$$\begin{aligned} \frac{1}{C_{id}} \int W_j^{(i)}(K \cap F_i) dF_i &= \sum_{J \subset I} (-1)^{\#J-1} \frac{1}{C_{id}} \int W_j^{(i)}(K_J \cap F_i) dF_i \\ &= \sum_{J \subset I} (-1)^{\#J-1} \frac{\omega_i}{\omega_{d-i}} \frac{\omega_{d-j}}{\omega_{i-j}} W_j(K_J) \\ &= \frac{\omega_i}{\omega_{d-i}} \frac{\omega_{d-j}}{\omega_{i-j}} W_j(K) \end{aligned}$$

The same additivity argument shows that Cauchy's formula is also valid in  $\mathcal{R}_d$ . For  $0 \leq j \leq i < d$ , we have

$$\frac{1}{C_{id}} \int W_j^{(i)}(K \downarrow S_i) dS_i = \frac{\omega_i}{\omega_{d-i}} W_{d+j-i}(K)$$

In contrast to this, Steiner's formula cannot be extended to  $\mathcal{R}_d$ . Dilation does not commute with intersection.

## Exercises

**6.1** What is the mean linear section of a triangle?

**6.2** Prove directly Steiner's formula for the case of a triangle.

**6.3** Let  $K \in \mathcal{C}_d$ . Suppose that  $K$  is open w.r.t. a ball of radius  $\lambda$ . Prove Steiner's formula for erosions (Matheron, 1978)

$$W_i(K \ominus \lambda B) = \sum_{j=0}^{d-i} \binom{d-i}{j} W_{i+j}(K) (-\lambda)^j$$

Part II

**The algorithms**

## 7. Basics about simulations

This chapter deals with the basic simulation algorithms that are indispensable for designing more elaborate ones. The case of the uniform distribution in  $]0, 1[$  is considered first. Then we turn to the simulation of standard univariate distributions (exponential, gaussian, poisson etc ...). Finally, we mention algorithms for several multivariate distributions (random point on a bounded domain, random direction, gaussian vector).

### 7.1 Uniform distributions

#### 7.1.1 Uniform distribution in $]0, 1[$

The probability density function (p.d.f.) of the uniform distribution on  $]0, 1[$  is defined as

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Despite its simple form, it is extremely difficult to simulate it properly. More or less satisfactory algorithms do exist nowadays. The objective of this section is not to give a definitive solution to this ongoing problem, but simply to explain some of the difficulties that it raises.

Let us start with a preliminary remark. It seems paradoxical to produce random values using a computer that runs deterministic operations according to pre-established rules. Things must be viewed statistically. A sequence of values is considered as simulated from  $f$  if it shares the same statistical properties as a sequence of realizations of mutually independent variables of distribution  $f$ . Independence and goodness of fit can be checked using standard statistical tests such as the chi-square test or the Kolmogorov-Smirnov test etc... (Knuth, 1969; Rohatgi, 1976; Rubinstein, 1981).

This section is devoted to the description of the multiplicative generator which rests on a congruential procedure inspired by the game of roulette (Lehmer, 1951). Given two integers  $a$  and  $m$ , one defines the inductive sequence by

$$u_{n+1} \equiv au_n \pmod{m}$$

starting from an initial integer  $u_0$  (the *seed*). Then the normation to  $]0, 1[$  is ensured by putting

$$x_n = \frac{u_n}{m}$$

Clearly, the sequence  $x_n$  takes at most  $m-1$  values. Matheron (1980) showed that the sequence is periodic with period  $m-1$  whatever the initial value  $u_0 \in \{1, \dots, m-1\}$  provided that the following three conditions are fulfilled:

- $m$  is a prime number.
- $p = \frac{m-1}{2}$  is also a prime number.
- $a^p \equiv -1 \pmod{m}$ .

For instance,  $a = 1000$  and  $m = 2001179$  satisfy these three conditions. But how can this be checked?

To verify that  $m$  is prime, it is sufficient to check that no prime number less than  $\sqrt{m}$  divides it (Eratosthenes' sieve). This method is rather fast provided that  $m$  is not too large (an efficiency indication is obtained by extrapolating an asymptotic result of number theory (Hardy and Wright, 1979): the number of prime numbers less than  $\sqrt{m}$  is around  $2\sqrt{m}/\ln m$ ). The third condition is more complicated. If  $a \not\equiv 0 \pmod{m}$ , Fermat's theorem says that  $a^{m-1} \equiv 1 \pmod{m}$ . This implies

$$a^{\frac{m-1}{2}} \equiv \pm 1 \pmod{m}$$

At this point, let us introduce the *Legendre coefficient* defined as

$$\left(\frac{a}{m}\right) = a^{\frac{m-1}{2}} \pmod{m}$$

Legendre coefficients satisfy the following properties (Hardy and Wright, 1979):

- They are invariant modulo  $(m)$

$$\left(\frac{a+km}{m}\right) = \left(\frac{a}{m}\right) \quad k \in \mathbb{N}$$

- They are multiplicative

$$\left(\frac{ab}{m}\right) = \left(\frac{a}{m}\right) \left(\frac{b}{m}\right)$$

- if  $a$  and  $m$  are odd prime numbers, the well known *law of quadratic reciprocity* by Gauss holds

$$\left(\frac{a}{m}\right)\left(\frac{m}{a}\right) = (-1)^{\frac{a-1}{2}\frac{m-1}{2}}$$

- In the case  $a = 2$ , we have

$$\left(\frac{2}{m}\right) = \begin{cases} +1 & \text{if } m \equiv \pm 1 \pmod{8} \\ -1 & \text{if } m \equiv \pm 3 \pmod{8} \end{cases}$$

Using these 4 formulae, let us carry out the calculations for  $a = 1000$  and  $m = 2001179$ .

$$\left(\frac{1000}{m}\right) = \left(\frac{2^3 5^3}{m}\right) = \left(\frac{2}{m}\right)^3 \left(\frac{5}{m}\right)^3 = \left(\frac{2}{m}\right)\left(\frac{5}{m}\right)$$

But

$$\left(\frac{2}{m}\right) = -1$$

since  $2001179 \equiv 3 \pmod{8}$ . On the other hand, the law of quadratic reciprocity gives

$$\left(\frac{5}{m}\right) = \left(\frac{m}{5}\right) (-1)^{\frac{5-1}{2}\frac{m-1}{2}} = \left(\frac{m}{5}\right) = \left(\frac{4}{5}\right) = \left(\frac{2}{5}\right)^2 = 1$$

Finally

$$\left(\frac{1000}{2001179}\right) = -1$$

as announced.

In fact, the multiplicative generator does not produce a sequence of numbers that are perfectly compatible with the model of independent variables uniformly distributed in  $]0, 1[$ . Starting from  $U_0$  uniform in  $\{1, \dots, m-1\}$ , we define the sequence  $U_{n+1} \equiv aU_n \pmod{m}$ , then we put  $X_n = U_n/m$ . The  $X_n$ 's are identically distributed with a mean equal to  $1/2$ . Their common variance  $(m-2)/12m$  is asymptotically equal to the model variance  $1/12$  as the modulus  $m$  becomes very large. Moreover, the correlation coefficient between  $X_n$  and  $X_{n+1}$  is approximately equal to  $1/a$  instead of 0.

From now onwards, we assume that we can generate independent uniform variables in  $]0, 1[$ . If  $U$  stands for one such variable, we write  $U \sim Unif$ .

### 7.1.2 Uniform distribution on a bounded domain of $\mathbb{R}^d$

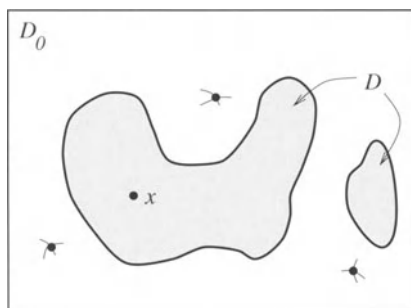
Let  $D$  be a bounded domain in  $\mathbb{R}^d$  with nonzero  $d$ -volume. Our objective is to generate a point  $x$  uniformly distributed in  $D$  (in brief  $x \sim Unif(D)$ ).

Suppose at first that  $D$  is the right parallelotope  $\prod_{i=1}^d ]a_i, a_i + \ell_i[$  with origin  $(a_1, \dots, a_d)$  and sides  $\ell_1, \dots, \ell_d$ . If  $U_1, \dots, U_d$  are independent and uniform in  $]0, 1[$ , then  $(a_1 + \ell_1 U_1, \dots, a_d + \ell_d U_d)$  is uniform in  $D$ .

Suppose now that  $D$  has a more complicated shape. As  $D$  is bounded, it can be enclosed within a right parallelotope  $D_0$ . Let  $x$  be a uniform point in  $D_0$ . If  $x \in D$ , then  $x$  is uniform in  $D$ . This simple remark provides the following *acceptance-rejection* algorithm:

**Algorithm 7.1.1.** (*Uniform point in a bounded domain*)

1. Generate  $x \sim Unif(D_0)$ .
  - 2.1. If  $x \in D$ , deliver  $x$ .
  - 2.2. Otherwise, go to 1.



**Fig. 7.1.** Simulation of a uniform point in a bounded domain

Let  $N$  be the number of attempts necessary to get a uniform point in  $D$ . Clearly, we have

$$P\{N = n\} = \frac{|D|}{|D_0|} \left(1 - \frac{|D|}{|D_0|}\right)^{n-1} \quad n > 0$$

so that

$$E\{N\} = \frac{|D_0|}{|D|}$$

This acceptance method is effective provided that  $D_0$  encloses  $D$  tightly.



## 7.2 Non-uniform distribution

### 7.2.1 Inversion method

Suppose that we wish to simulate a random variable  $X$  with distribution function  $F$  (in short  $X \stackrel{df}{\sim} F$ ). Remember that  $F$  is defined as

$$F(x) = P\{X < x\} \quad x \in \mathbb{R}$$

$F$  is non-decreasing and left continuous, but not necessarily bijective. It is nonetheless possible to assign it a pseudo-inverse  $F^{-1}$  by putting

$$F^{-1}(u) = \sup\{x \in \mathbb{R} \mid F(x) \leq u\} \quad u \in ]0, 1[$$

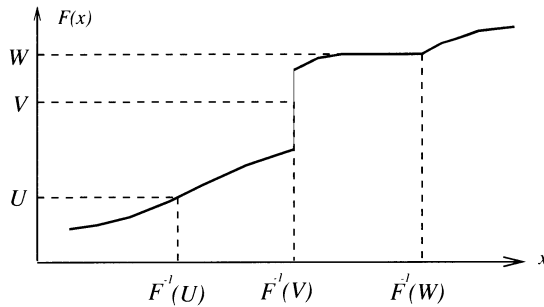
**Algorithm 7.2.1.** (*Inversion method*)

1. Generate  $U \sim Unif$ .
2. Deliver  $F^{-1}(U)$ .

**Proof:** Note first that  $F^{-1}(u) < x$  if and only if  $u < F(x)$ . Then replace  $u$  by  $U \sim Unif$ . In probabilistic terms, this gives

$$P\{F^{-1}(U) < x\} = P\{U < F(x)\} = F(x)$$

which proves that  $F^{-1}(U) \stackrel{df}{\sim} F$ .



**Fig. 7.2.** Simulation of a distribution using the inversion method

**Example 7.2.1.** We say that  $X$  follows an exponential distribution with parameter  $a > 0$  (in brief  $X \sim Exp(a)$ ) if its distribution function is  $F(x) = 1 - \exp(-ax)$ . Since  $F(x) = u$  if and only if  $x = -\ln(1 - u)/a$ , the inversion method amounts to simulating  $U \sim Unif$  and returning  $-\ln(1 - U)/a$ . Actually, returning  $-\ln U/a$  is sufficient since  $1 - U \sim Unif$ .

**Example 7.2.2.** A Bernoulli variable with mean  $p$  is a random variable that takes only the two values 1 and 0 with probability  $p$  and  $q = 1 - p$  respectively. The inversion method shows that if  $U \sim Unif$ , then  $1_{U \geq q} \sim Ber(p)$ .

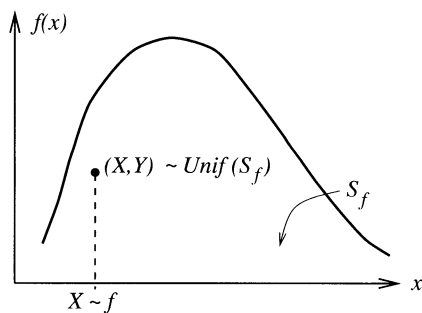
### 7.2.2 Acceptance-Rejection method

Let  $f$  be a non negative function defined in  $\mathbb{R}^d$  with finite integral. Then  $f$  can be regarded as p.d.f. up to a normalization factor. Our objective is to generate a random vector  $X$  having a p.d.f. proportional to  $f$  (in brief  $X \sim f$ ).

The simulation algorithm is based on the observation that simulating a vector with density  $f$  amounts simulating a uniform point in the subgraph  $S_f = \{(x, y) \mid 0 \leq y \leq f(x)\}$  of  $F$ . More precisely

**Proposition 7.2.1.** *Suppose that  $f$  is non-negative and has a finite integral.*

- 1) *If  $(X, Y) \sim Unif(S_f)$ , then  $X \sim f$ .*
- 2) *If  $X \sim f$  and  $U \sim Unif$ , then  $(X, Uf(X)) \sim Unif(S_f)$ .*



**Fig. 7.3.** If  $(X, Y)$  is uniform in the subgraph of  $f$ , then  $X$  follows a p.d.f. proportional to  $f$

**Proof:** Part 1 is straightforward: if  $B$  is a Borel subset of  $\mathbb{R}^d$ , we have

$$P\{X \in B\} = \frac{\int_B f(x) dx}{\int_{\mathbb{R}^d} f(x) dx}$$

which shows that  $X \sim f$ . Regarding 2, let  $B$  be a Borel subset of  $\mathbb{R}^d \times \mathbb{R}$  that is totally contained within  $S_f$ . If  $X = x$ , then  $(X, Uf(X)) \in B$  with probability  $|B \cap L_x|/f(x)$ , where  $L_x = \{(x, y), y \in \mathbb{R}\}$ . Thus we can write

$$P\{(X, Uf(X)) \in B\} = \int_{\mathbb{R}^d} f(x) \frac{|B \cap L_x|}{f(x)} dx = \int_{\mathbb{R}^d} |B \cap L_x| dx = |B|$$

and therefore  $(X, Uf(X)) \sim Unif(S_f)$ .  $\square$

Now let  $g$  be another non-negative function defined in  $\mathbb{R}^d$  and with finite integral. Suppose that  $f \leq g$  and that we know how to simulate  $g$  (by inverting its associated distribution function for instance). Let  $X \sim g$  and  $U \sim Unif$ . Then  $(X, Ug(X)) \sim Unif(S_g)$  (part 2 of Proposition 7.2.2). If  $(X, Ug(X)) \in S_f$ , then  $(X, Ug(X)) \sim Unif(S_f)$ , and consequently  $X \sim f$  (part 1 of Proposition 7.2.2). This leads to the following algorithm:

**Algorithm 7.2.2.** (*Acceptance-rejection method*)

1. Generate  $X \sim g$  and  $U \sim Unif$ .
  - 2.1. If  $Ug(X) \leq f(X)$ , deliver  $X$ .
  - 2.2. Otherwise, goto 1.

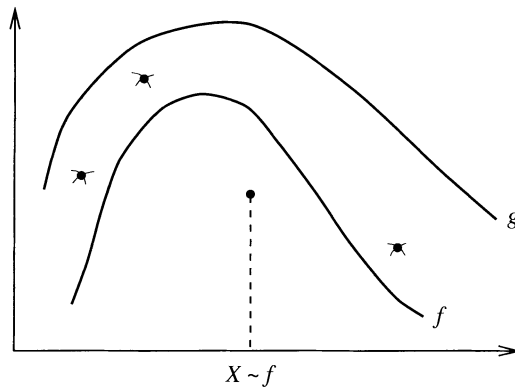


Fig. 7.4. Simulation of  $f$  using the acceptance-rejection method

Finding a function  $g$  that is a suitable upper bound is an important issue. Once  $g$  has been fixed, the mean number of attempts to get a simulated value is equal to

$$E\{N\} = \frac{|S_g|}{|S_f|} = \frac{\int_{\mathbb{R}^d} g(x) dx}{\int_{\mathbb{R}^d} f(x) dx}$$

**Example 7.2.3.** The p.d.f. of a standard gaussian distribution is

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \quad x \in \mathbb{R}$$

Note that

$$f(x) \leq g(x) = \sqrt{\frac{e}{2\pi}} \exp\{-|x|\}$$

and that if  $X \sim g$ , then  $|X| \sim Exp(1)$ . Writing  $|X| = -\ln V$  with  $V \sim Unif$  (see Example 7.2.1), the acceptance criterion  $Ug(X) \leq f(X)$  can be rewritten  $\ln U \leq -(\ln V + 1)^2$ . Hence the algorithm:

**Algorithm 7.2.3.** (*Standard gaussian distribution*)

1. Generate  $U, V \sim Unif$  independent.
  - 2.1. If  $\ln U \leq -(\ln V + 1)^2$ , then generate  $W \sim Ber(0.5)$  and deliver  $(2W - 1) \ln V$ .
  - 2.2. Otherwise, goto 1.

This method is rather fast: the mean number of attempts is

$$E\{N\} = \sqrt{\frac{2e}{\pi}} \approx 1.315$$

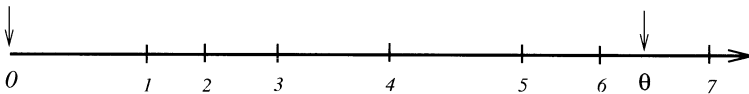
**Remark 7.2.1.** The acceptance-rejection method can be extended to discrete distributions. For instance, the probability system  $(p_n, n \in \mathbb{N})$  induces the p.d.f.  $f(x) = p_{\lfloor x \rfloor}$  for any  $x \in [0, +\infty[$ . If  $X \sim f$ , then  $\lfloor X \rfloor \sim p$ .

### 7.2.3 Ad hoc methods

**Poisson distribution.** The discrete variable  $N$  follows a Poisson distribution with mean  $\theta$ , in brief  $N \sim Poisson(\theta)$ , if

$$P\{N = n\} = \exp(-\theta) \frac{\theta^n}{n!} \quad n \in \mathbb{N}$$

A well known result in probability says that if independent exponentially distributed intervals with mean 1 are concatenated starting from the origin, then the number of intervals totally contained within  $[0, \theta]$  has a Poisson distribution with mean  $\theta$ .



**Fig. 7.5.** Simulation of a Poisson variable by concatenating intervals with independent exponential lengths. In this example the value generated is 6.

In this approach, the length of the  $i^{\text{th}}$  interval can be written as  $-\ln U_i$  with  $U_i \sim Unif$ . Accordingly, the value generated is the first index  $n$  such that  $-\sum_{i=1}^{n+1} \ln U_i \geq \theta$ , or equivalently  $\prod_{i=1}^{n+1} U_i \leq \exp(-\theta)$ . This leads to the following algorithm:

**Algorithm 7.2.4.** (*Poisson distribution*)

1. Set  $N = 0$  and  $T = 1$ .
2. Generate  $U \sim Unif$  and put  $T = UT$ .
  - 3.1. If  $T > \exp(-\theta)$ , set  $N = N + 1$  and go to 2.
  - 3.2. If not, deliver  $N$ .

Using this algorithm, the mean number of uniform variables to be simulated is equal to  $\theta + 1$ . Consequently, this algorithm is efficient provided that  $\theta$  is relatively small. For large  $\theta$  (typically  $\theta \geq 100$ ), the distribution of  $[\theta + 0.5 + \sqrt{\theta}Y]$  with  $Y$  standard gaussian is a good approximation of a Poisson distribution. There also exists a acceptance-rejection algorithm that is efficient whatever the  $\theta$  value (Ahrens and Dieter, 1982). The upper bound is rather complicated and is not presented here.

**Gaussian distribution.** Let  $X$  and  $Y$  be two independent standard gaussian variables. This gives a random point in  $\mathbb{R}^2$  which can be written as  $(R, \Theta)$  in polar coordinates. It can be shown that  $R$  and  $\Theta$  are independent.  $R^2$  follows an exponential distribution with mean 2, and  $\Theta$  is uniform in  $[0, 2\pi[$ . This leads to the algorithm proposed by Box and Muller (1958):

**Algorithm 7.2.5.** (*Standard gaussian distribution*)

1. Generate  $U, V \sim \text{Unif independent}$ .
2. Deliver  $\sqrt{-2 \ln U} \cos(2\pi V)$  or  $\sqrt{-2 \ln U} \sin(2\pi V)$ .

**Symmetric stable distributions.** The most convenient way to characterize a symmetric stable distribution is via its characteristic function

$$E\{e^{iuX}\} = e^{-|u|^\alpha}$$

This distribution depends on a parameter  $\alpha \in ]0, 2]$  which characterizes the heaviness of its tail. The case  $\alpha = 2$  corresponds to a gaussian distribution which has moments of any order. If  $\alpha \in ]1, 2[$ , the mean is equal to 0, but the variance is infinite. In the case  $\alpha \leq 1$ , even the mean does not exist. Despite the rather complicated expression of its p.d.f. (not reproduced here), the symmetric stable distribution can be simulated very simply by using the following algorithm due to Chambers et al. (1976)

**Algorithm 7.2.6.** (*Symmetric stable distribution*)

1. Generate  $U \sim \text{Unif}(-\frac{\pi}{2}, \frac{\pi}{2}[)$  and  $V \sim \text{Exp}(1)$  independent.
2. Deliver  $\frac{\sin(\alpha U)}{\cos^{\frac{1}{\alpha}} U} \left( \frac{\cos[(1-\alpha)U]}{V} \right)^{\frac{1-\alpha}{\alpha}}$ .

In the case  $\alpha = 2$ , this is just the Box-Muller algorithm for simulating a gaussian distribution.

**Uniform direction in  $\mathbb{R}^d$ .** Let  $(X_1, \dots, X_d)$  be  $d$  independent standard gaussian variables. Their p.d.f. is

$$f(x_1, \dots, x_d) = \frac{1}{(2\pi)^{\frac{d}{2}}} \exp \left\{ -\frac{x_1^2 + \dots + x_d^2}{2} \right\}$$

Since  $f$  depends only on the modulus of  $x = (x_1, \dots, x_d)$ , it is isotropic. Hence the algorithm:

**Algorithm 7.2.7.** (*Uniform direction in  $\mathbb{R}^d$* )

1. Generate a vector  $X = (X_1, \dots, X_d)$  whose components are independent standard gaussian variables.
2. Calculate  $|X| = \sqrt{X_1^2 + \dots + X_d^2}$ .
3. Deliver the vector  $X/|X|$ .

## Exercises

**7.1** Apply the inversion method to simulate the Cauchy distribution

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2} \quad x \in \mathbb{R}$$

**7.2** Let  $0 < p < 1$  and  $U \sim Unif$ . Show that the discrete random variable  $N = \lfloor \log_p(U) \rfloor$  follows the geometric distribution

$$P\{N = n\} = (1-p)p^n \quad n \in \mathbb{N}$$

**7.3** Let  $a \geq 1$ . Use the inequality  $\exp(-x^2/2) \leq x \exp(-x^2/2)$  which is valid for  $x \geq 1$  to derive a simulation algorithm for the truncated gaussian distribution having the p.d.f.

$$f_a(x) \propto \exp \left\{ -\frac{x^2}{2} \right\} 1_{x \geq a}$$

**7.4** Let  $f_\alpha$  be the p.d.f. of a gamma distribution of parameter  $\alpha > 0$

$$f_\alpha(x) = \frac{1}{\Gamma(\alpha)} e^{-x} x^{\alpha-1} \quad x > 0$$

- 1) How can we simulate  $f_\alpha$  in the case  $\alpha = 1$ ?
- 2) If  $\alpha < 1$ , give the simulation algorithm by acceptance-rejection method using

$$g(x) = x^{\alpha-1} 1_{x \leq 1} + e^{-x} 1_{x > 1}$$

How fast is the algorithm?

- 3) Same question as 2) in the case  $\alpha > 1$  with

$$g(x) = \frac{x^{\lambda-1}}{(x^\lambda + \alpha^\lambda)^2}$$

where  $\lambda = \sqrt{2\alpha - 1}$  (Cheng, 1977).

## 8. Iterative algorithms for simulation

The problem addressed in this chapter is the simulation of a distribution  $p$  on  $(\Omega, \mathcal{A})$ . In the case when the state-space  $\Omega$  has an abstract setting or a complicated shape, the algorithms presented in the previous chapter may not be directly applicable. More general algorithms are then required.

Since their introduction by Metropolis et al. in 1953, iterative algorithms have proved to be quite effective. They consist of generating a sequence of random states  $(X_n, n \in \mathbb{N})$  in such a way that the distribution of  $X_n$  tends to  $p$  as  $n$  becomes very large. For practical reasons, the sequence is often a Markov chain, so we give some reminders in the first part of this chapter. Then two prototypes of iterative algorithms are reviewed. The first one is the famous Metropolis algorithm. It is illustrated by the simulation of a Poisson distribution. The second one is a Hit-and-Run algorithm devised to simulate a uniform point in a domain in a high dimensional space. Variations of this algorithm are closely related to the famous Gibbs sampler (Geman and Geman, 1984). A general procedure is also proposed to turn a non conditional iterative algorithm into a conditional one. The chapter concludes with a short presentation of simulated annealing as a conditional simulation algorithm.

### 8.1 Some reminders on Markov chains

The objective of this section is to give the basic elements required for the Markov simulation. For more comprehensive reference on Markov chains, the reader can consult Meyn and Tweedie (1993), and also Tierney (1994).

#### 8.1.1 Markov chain

Let us denote by  $X_n$  the random state at the  $n^{\text{th}}$  iteration of the algorithm. In the general case,  $X_n$  depends on all the previous states  $X_0, \dots, X_{n-1}$ . But this is not practical from a computational standpoint, as it is not possible to keep a large number of states in memory. Therefore some restrictions have therefore to be considered. One possibility is to construct  $X_n$  as a function of  $X_{n-1}$  uniquely. In this case,  $X_n$  no longer depends on  $X_0, X_1, \dots, X_{n-2}$  once  $X_{n-1}$  is explicitly known. The sequence  $(X_n, n \in \mathbb{N})$  is then called a *Markov chain*.

A particularly useful feature of a Markov chain is that its simulation is straightforward. We start by simulating  $X_0$ , then  $X_1$  given  $X_0 = x_0$ , then  $X_2$  given  $X_1 = x_1$  etc ... This requires an initial distribution  $p_0$  for  $X_0$  and the conditional distributions to jump from  $X_0$  to  $X_1$ , from  $X_1$  to  $X_2$  ... Most of the time, these conditional distributions need not be dependent of the iteration order. All of them are then identical and can be written

$$P(x, A) = P\{X_{k+1} \in A \mid X_k = x\}$$

In this particular case, the Markov chain is said to be *homogeneous*, and the common conditional distribution is called a *transition kernel*. More precisely

**Definition 8.1.1.** *A function  $P$  from  $\Omega \times \mathcal{A}$  to  $[0, 1]$  such that*  
*i) for each  $A \in \mathcal{A}$ ,  $P(\cdot, A)$  is measurable*  
*ii) for each  $x \in \Omega$ ,  $P(x, \cdot)$  is a probability measure on  $(\Omega, \mathcal{A})$*   
*is called a transition kernel on  $(\Omega, \mathcal{A})$ .*

One consequence of homogeneity is that the conditional distribution from  $X_k$  to  $X_{k+n}$  does not depend on  $k$ . It is also a transition kernel on  $(\Omega, \mathcal{A})$  and can be denoted by

$$P^{(n)}(x, A) = P\{X_{k+n} \in A \mid X_k = x\}$$

Because of the Markov property, the  $P^{(n)}$  are related by

$$P^{(n+1)}(x, A) = \int_{\Omega} P^{(n)}(x, dy) P(y, A) = \int_{\Omega} P(x, dy) P^{(n)}(y, A)$$

with the convention  $P^{(0)}(x, dy) = \delta_x(dy)$  (Dirac measure at  $x$ ).

Finally, the distribution of a homogeneous Markov chain can be characterized by an initial distribution  $p_0$  and a transition kernel  $P$ . Since mainly homogeneous Markov chains are considered in this book, the word homogeneous will often be omitted.

### 8.1.2 Convergence

How can we express the idea that  $P^{(n)}(x, A)$  is close to  $p(A)$ ? Meyn and Tweedie (1993) consider the *total variation norm* of the signed measure  $P^{(n)}(x, \cdot) - p$ . This is defined as

$$\| P^{(n)}(x, \cdot) - p \| = 2 \sup_{A \in \mathcal{A}} \left| P^{(n)}(x, A) - p(A) \right|$$

Then one possible way to express the convergence of the distribution of  $X_n$  to  $p$  is to require that

$$\lim_{n \rightarrow \infty} \| P^{(n)}(x, \cdot) - p \| = 0$$



for all  $x \in \Omega$ . Halmos (1969) has shown that the total variation norm of  $P^{(n)}(x, \cdot) - p$  can also be written as

$$\| P^{(n)}(x, \cdot) - p \| = \sup_{|f| \leq 1} \left| \int_{\Omega} [P^{(n)}(x, dy) - p(dy)] f(y) \right|$$

so that the convergence implies

$$\lim_{n \rightarrow \infty} \int_{\Omega} P^{(n)}(x, dy) f(y) = \int_{\Omega} p(dy) f(y)$$

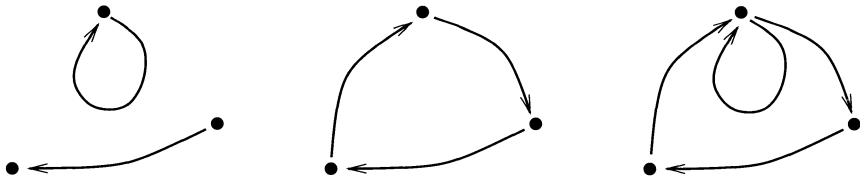
for any bounded measurable function  $f$  and for any  $x \in \Omega$ . The effects caused by the initial state  $x$  are no longer perceptible after a sufficiently large number of iterations.

### 8.1.3 Some features of a transition kernel

The convergence of  $P^{(n)}$  to  $p$  implies a number of consequences.

Let  $A \in \mathcal{A}$  be such that  $p(A) > 0$ . Because of the convergence, there exists  $n$  such that  $P^{(n)}(x, A) > 0$ . In other words, any state can communicate with any event of positive occurrence probability. The transition kernel is then said to be *p-irreducible*.

This result can be refined. If  $p(A) > 0$ , then the convergence says that  $P^{(n)}(x, A) > 0$  for  $n$  large enough. In particular, the greatest common divisor of the  $n$ 's such that  $P^{(n)}(x, A) > 0$  is 1. This is usually expressed by saying that  $P$  is *aperiodic*.



**Fig. 8.1.** This figure shows 3 Markov chains defined on the same state-space but with different transition kernels. Arrows indicate positive direct transitions. The left chain is reducible. The middle one is irreducible but periodic with period 3. It suffices to add a loop at one state to make it aperiodic (cf. right chain)

Let us now start from

$$P^{(n+1)}(x, A) = \int_{\Omega} P^{(n)}(x, dy) P(y, A)$$

and let  $n$  tend to infinity. By definition of the convergence, the left-hand side converges towards  $p(A)$ . Since  $P(\cdot, A)$  is measurable and bounded by 1, the

right-hand side converges to the integral of  $P(\cdot, A)$  for  $p$  (Halmos' property), and we finally end up with

$$p(A) = \int_{\Omega} p(dy)P(y, A)$$

The distribution  $p$  is then said to be *invariant* under the transition kernel  $P$ .

If  $p$  is invariant under  $P$ , then  $\|P^{(n)}(x, \cdot) - p\|$  is a monotonically decreasing function of  $n$ . Indeed, by applying the Markov property and the definition of invariance, we have

$$\begin{aligned} \|P^{(n+1)}(x, \cdot) - p\| &= \sup_{|f| \leq 1} \left| \int_{\Omega} [P^{(n+1)}(x, dy) - p(dy)] f(y) \right| \\ &= \sup_{|f| \leq 1} \left| \int_{\Omega} [P^{(n)}(x, dz) - p(dz)] \int_{\Omega} P(z, dy) f(y) \right| \\ &\leq \sup_{|g| \leq 1} \left| \int_{\Omega} [P^{(n)}(x, dz) - p(dz)] g(y) \right| \\ &= \|P^{(n)}(x, \cdot) - p\| \end{aligned}$$

But this does not necessarily imply the convergence of  $P^{(n)}(x, \cdot)$  to  $p$ . Other assumptions are required, such as the ones given by Tierney (1994).

**Theorem 8.1.1.** *Let  $P$  be a transition kernel and  $p$  a distribution. If  $P$  is  $p$ -irreducible, and aperiodic, and if  $p$  is invariant under  $P$ , then*

$$\lim_{n \rightarrow \infty} \|P^{(n)}(x, \cdot) - p\| = 0$$

for  $p$ -almost any  $x \in \Omega$ .

**Remark 8.1.1.** Under the further assumption that any event of positive occurrence is visited infinitely often by the Markov chain, i.e.

$$P \left\{ \sum_{n=1}^{\infty} 1_{X_n \in A} = \infty \mid X_0 = x \right\} = 1$$

for any state  $x$  and any event  $A$  such that  $p(A) > 0$ , Tierney (1994) has established that the convergence in theorem 8.1.1. takes place not only  $p$ -almost surely, but for every  $x \in \Omega$ . Suppose that  $X_0$  follows the distribution  $p_0$ . If we denote by  $P^{(n)}(p_0, \cdot)$  the distribution of  $X_n$ , then we can write

$$\lim_{n \rightarrow \infty} \|P^{(n)}(p_0, \cdot) - p\| = 0$$

Although it is not directly related to convergence, the concept of reversibility is emphasized because of its importance in the design of simulation algorithms. Suppose that  $p$  is invariant under  $P$ , and let  $(X, Y) \sim p(dx)P(x, dy)$ . Then  $Y \sim p$ , so that  $X$  and  $Y$  have the same distribution. Reversibility is stronger. It says that both vectors  $(X, Y)$  and  $(Y, X)$  have the same bivariate distribution.

**Definition 8.1.2.** *The transition kernel  $P$  is said to be reversible w.r.t. the distribution  $p$  if*

$$\int_{\Omega} \int_{\Omega} p(dx)P(x, dy)f(x, y) = \int_{\Omega} \int_{\Omega} p(dx)P(x, dy)f(y, x)$$

for any numerical function that is defined on  $\Omega \times \Omega$ , and is measurable and bounded.

In the case when  $f$  is a product of indicator functions

$$f(x, y) = 1_{x \in A} 1_{y \in B} \quad A, B \in \mathcal{A}$$

we obtain the following exchange formula

$$\int_A p(dx)P(x, B) = \int_B p(dx)P(x, A)$$

The invariance of  $p$  under  $P$  can be obtained from this by taking  $B = \Omega$ .

## 8.2 Metropolis algorithm

Suppose that we want to simulate the distribution  $p$  on  $(\Omega, \mathcal{A})$  as the limit distribution of a Markov chain. The first problem is to construct a transition kernel for  $p$ . Metropolis et al. (1953) proposed an elegant solution to this problem in the case of a discrete distribution<sup>1</sup>. In this case a transition kernel is completely characterized by the transition probabilities  $P(x, y)$  for all pairs of states  $x$  and  $y$ .

The Metropolis algorithm requires a candidate transition kernel  $Q$  that is symmetric (i.e.  $Q(x, y) = Q(y, x)$  for all  $x, y \in \Omega$ ). Starting from state  $x$ , the idea is to generate a candidate state  $y$  from  $Q(x, \cdot)$  and to accept it with a probability  $\alpha(x, y)$  which depends on both states  $x$  and  $y$ . Of course,  $y$  will be more often accepted if  $p(x)$  is small or  $p(y)$  large. Metropolis et al. propose the acceptance criterion

$$\alpha(x, y) = \min\left(\frac{p(y)}{p(x)}, 1\right)$$

---

<sup>1</sup> Continuous versions of Metropolis' solution can be found in the literature (Tierney, 1994; Gaetan and Guyon, 1997).

which amounts to considering the transition kernel

$$P(x, y) = Q(x, y) \alpha(x, y) = Q(x, y) \min\left(\frac{p(y)}{p(x)}, 1\right)$$

for  $x \neq y$  as well as

$$P(x, x) = 1 - \sum_{y \neq x} P(x, y)$$

It remains to show that a Markov chain with transition kernel  $P$  has the limit distribution  $p$ . According to theorem 8.1.1, it suffices to check that  $P$  is  $p$ -irreducible, aperiodic and admits  $p$  as an invariant distribution. Assuming  $p > 0$  on  $\Omega$ , then  $P$  is  $p$ -irreducible and aperiodic if this is the case for  $Q$ . Regarding the third property, it can be seen that the symmetry of  $Q$  implies the following *reversibility* formula

$$p(x)P(x, y) = Q(x, y) \min(p(y), p(x)) = Q(y, x) \min(p(x), p(y)) = p(y)P(y, x)$$

for  $x \neq y$ , which in turn implies the invariance

$$\sum_x p(x)P(x, y) = \sum_x p(y)P(y, x) = p(y) \sum_x P(y, x) = p(y)$$

In summary, here is the Metropolis algorithm to simulate a discrete, positive distribution  $p$  starting from a candidate transition kernel  $Q$  that is  $p$ -irreducible, aperiodic and symmetric. In this algorithm, the test  $p(y) < Up(x)$  is a mere reformulation of the acceptance criterion  $U < \alpha(x, y)$ . It is of course superfluous in the case  $p(y) \geq p(x)$ .

**Algorithm 8.2.1.** (*Metropolis algorithm*)

1. Let  $x \in \Omega$ .
2. Generate  $y \sim Q(x, \cdot)$  and  $U \sim Unif$ .
3. If  $p(x)U < p(y)$ , then put  $x = y$ .
4. Goto 2.

**Remark 8.2.1.** There are many versions of the Metropolis algorithm. For instance, another acceptance criterion has been proposed by Barker (1965)

$$\alpha(x, y) = \frac{p(y)}{p(x) + p(y)} \quad x \neq y$$

Hastings (1970) has also extended the Metropolis algorithm by considering a non-symmetric candidate transition kernel.

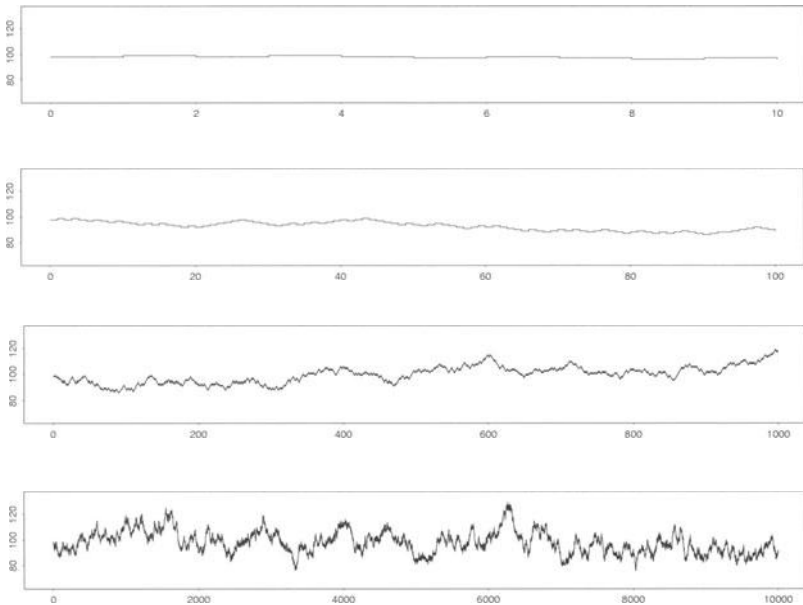
**Example 8.2.1.** Let us come back to the simulation of a Poisson distribution with parameter  $\theta$

$$p(x) = \exp(-\theta) \frac{\theta^x}{x!} \quad x \in \mathbb{N}$$

For the candidate transition kernel, we can take  $Q(x, x+1) = Q(x, x-1) = \frac{1}{2}$  if  $x \geq 1$  and  $Q(0, 1) = Q(0, 0) = \frac{1}{2}$ . It is immediately clear that  $Q$  is  $p$ -irreducible, aperiodic and symmetric. Choosing Barker's variation leads to the transition kernel

$$P(x, x+1) = \frac{\theta}{2(\theta + x + 1)} \quad P(x, x-1) = \frac{x}{2(\theta + x)}$$

and  $P(x, x) = 1 - P(x, x+1) - P(x, x-1)$ . This transition kernel has been used to simulate a Poisson distribution with mean  $\theta = 100$ . Figure 8.2 shows the evolution of the iterative simulation at various scales. The rate of convergence of this algorithm will be studied in the next chapter.



**Fig. 8.2.** Display at various scales (10, 100, 1000 and 10000) of some iterations during a Poisson simulation

### 8.3 A Hit-and-Run algorithm

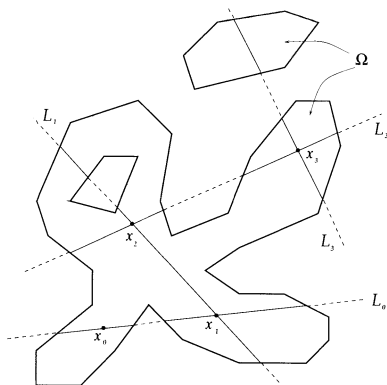
The problem considered here is the simulation of the uniform distribution in an open subset  $\Omega$  of  $\mathbb{R}^d$  with finite volume. The importance of this problem has been stressed in the last chapter (cf. Proposition 7.2.1). In the present case,  $\Omega$  has an arbitrary shape. It is not necessarily bounded, or convex, or even connected. To make things even more complicated, the dimension  $d$  is supposed to be large, typically several hundreds<sup>2</sup>. In cases like this, the standard acceptance-rejection method fails because the failure rate is too high. As an example, suppose that  $\Omega$  is a unit ball in  $\mathbb{R}^d$ . If it is enclosed within a hypercube of side 2, then the mean number of draws per success is  $2^d/\omega_d$ . For  $d = 100$ , this is equal to  $10^{70}$ .

#### 8.3.1 A stereological algorithm

The iterative algorithm given below has been considered by many authors (Turchin, 1971; Smith, 1983; Bélisle et al., 1993; Lantuéjoul, 1996):

**Algorithm 8.3.1.** (*Uniform point in  $\Omega$ . Stereological approach*)

1. Let  $x \in \Omega$ .
2. Generate a uniformly oriented line  $L$  passing through  $x$  (Alg. 7.2.7).
3. Generate a uniform point  $y \sim \text{Unif}(L \cap \Omega)$ .
4. Put  $x = y$  and goto 2.



**Fig. 8.3.** Principle of the stereological algorithm. Suppose that the current point is  $x_n$ . A uniformly oriented line  $L_n$  is generated through  $x_n$ . The next point  $x_{n+1}$  is then uniformly selected in  $L_n \cap \Omega$

<sup>2</sup> When carrying out conditional simulations, the dimension of the space is often equal to the number of conditioning data points.

Why does this algorithm work? Let  $X_n$  be the random point produced at the  $n^{\text{th}}$  iteration. The sequence  $(X_n, n \in \mathbb{N})$  is a homogeneous Markov chain. A direct calculation shows that its transition kernel has the probability density function

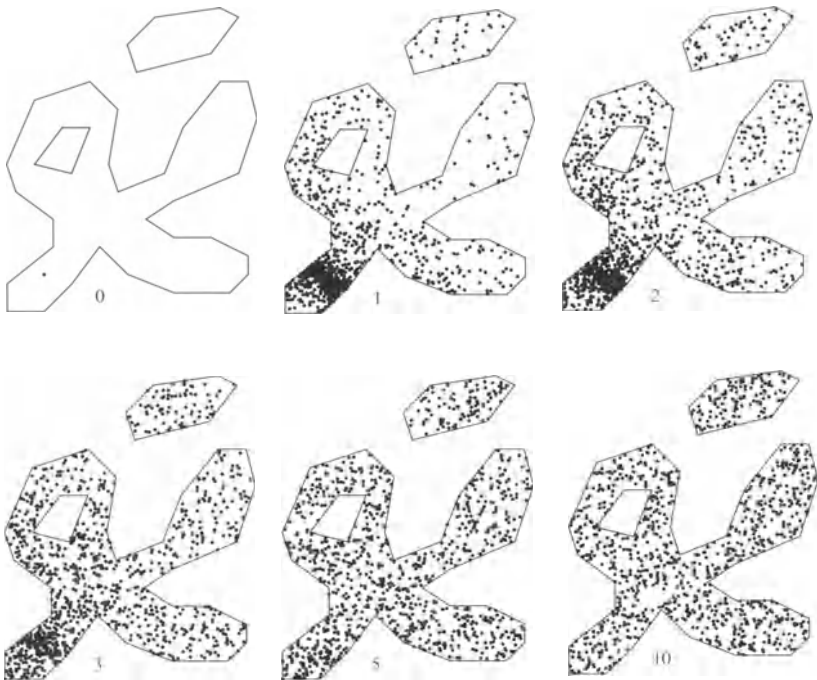
$$P(x, A) = \int_A f(x, y) dy \quad x \in \Omega, A \in \mathcal{A}$$

defined by

$$f(x, y) = \frac{2}{d\omega_d} \frac{1}{\ell(x, y, \Omega) |x - y|^{d-1}} \quad y \neq x,$$

where  $\ell(x, y, \Omega)$  stands for the length of the intersection between  $\Omega$  and the line passing through  $x$  and  $y$ . Note in particular that  $f$  is symmetric in  $x$  and  $y$ . This implies that the uniform distribution in  $\Omega$  (noted here  $u$ ) is invariant under  $P$ .

$$\int_{\Omega} u(x) P(x, A) dx = \int_{\Omega} \int_A f(x, y) dy \frac{dx}{|\Omega|} = \int_A \int_{\Omega} f(y, x) \frac{dx}{|\Omega|} dy = \frac{|A|}{|\Omega|} = u(A)$$



**Fig. 8.4.** 1000 independent simulations (algorithm 8.3.1) have been carried out from the same initial point (top left). The populations of points generated at iterations 1, 2, 3, 5 and 10 are displayed

Moreover,  $P$  is  $u$ -irreducible and aperiodic since any point is accessible from any other one at each iteration. Then, according to theorem 8.1.1,  $u$  is the limit distribution of the Markov chain.

Figure 8.4 illustrates how this algorithm works. We have taken  $\Omega$  to be a two-dimensional domain made up of two connected components. The initial point has been chosen in the "cloverleaf" component (top left). One thousand uniformly oriented lines have then been generated through this point, resulting in 1000 points at the first iteration (top middle). The distribution of these points is far from uniform. In particular there is a high density of points along the lines passing through the initial point and having a short intersection with  $\Omega$ . The 1000 points are used as input for a second iteration. Figure 8.4 shows the evolution of the simulation at iterations 2, 3, 5 and 10. A gradual convergence toward uniformity is observed. The rate of convergence seems to be quite fast since uniformity has practically been reached by the 5<sup>th</sup> iteration. This will be explicitly studied next chapter.

### 8.3.2 A uniform Gibbs sampler

Algorithm 8.3.1 has many variations. Since the intersection of  $\Omega$  by a line in an arbitrary direction may be difficult to determine (in practice, a uniform point in  $L \cap \Omega$  is usually simulated using the acceptance-rejection method), it may sometimes be preferable to generate lines in a finite set of directions only (cf. Figure 8.5). In the case when the directions considered are those of the coordinate axes, the successive points produced by the Markov chain vary in only one coordinate. This is just a Gibbs sampler (Geman and Geman, 1984).

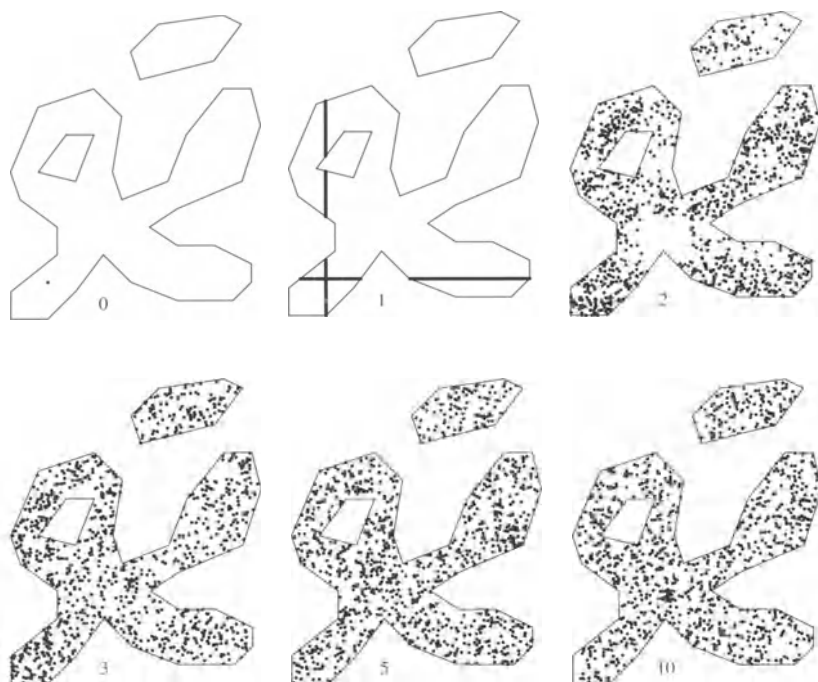
To simplify the presentation, let us assume that the  $d$  directions of the coordinate axes are independently and uniformly selected at each iteration. Then the transition kernel can be written

$$P(x, A) = \frac{1}{d} \sum_{i=1}^d P_i(x, A)$$

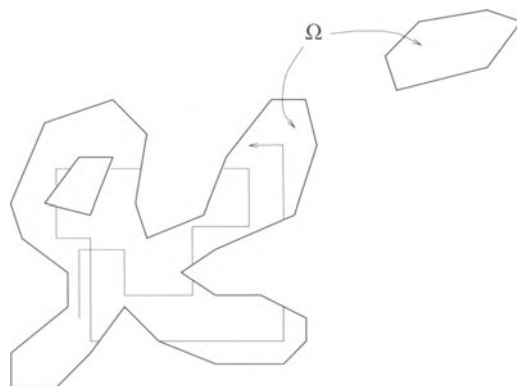
where  $P_i(x, A)$  is the probability that a change in the  $i^{\text{th}}$  coordinate  $x_i$  of  $x$  transforms it into a point of  $A$ . It can be shown easily that  $u$  is invariant under each  $P_i$ . Indeed, denoting by  $x^i$  the remaining coordinates of  $x$ , and by  $L_i(x^i)$  the line parallel to the  $i^{\text{th}}$  coordinate axis passing through  $x$ , we have

$$\begin{aligned} \int_{\Omega} u(x) P_i(x, A) dx &= \int_{\Omega} \frac{|L_i(x^i) \cap A|}{|L_i(x^i) \cap \Omega|} \frac{dx^i dx_i}{|\Omega|} \\ &= \int \frac{|L_i(x^i) \cap A|}{|L_i(x^i) \cap \Omega|} \frac{dx^i}{|\Omega|} |L_i(x^i) \cap \Omega| \\ &= \int |L_i(x^i) \cap A| \frac{dx^i}{|\Omega|} = \frac{|A|}{|\Omega|} = u(A) \end{aligned}$$





**Fig. 8.5.** 1000 independent simulations carried out with lines in two directions only. They are displayed at iterations 1, 2, 3, 5 and 10. The initial point is the same for all simulations



**Fig. 8.6.** In the case when the transitions take place only along the horizontal or the vertical direction, the process cannot jump from one component of  $\Omega$  to another. The Markov chain is not irreducible

Accordingly,  $u$  is invariant under  $P$ . Note also that  $P$  is aperiodic if it is  $u$ -irreducible (the same random direction can be selected twice in a row). But  $P$  may not be  $u$ -irreducible. As a counter-example, consider a domain  $\Omega$  partitioned into two subsets such that their points differ in two coordinates at least (cf. Figure 8.6). The possible reducibility of the Markov chain is the only theoretical limitation in using the Gibbs sampler for uniform distributions.

### 8.3.3 A more general hit-and-run algorithm

The hit-and-run algorithm presented above can be seen as a particular case of a more general algorithm. Suppose that each  $x \in \Omega$  is assigned a domain  $A(x) \subset \mathbb{R}^d$  with finite volume.  $A(x)$  needs not being totally included within  $\Omega$ . Consider the following algorithm:

**Algorithm 8.3.2.** (*Uniform point in  $\Omega$ . Local approach*)

1. Let  $x \in \Omega$ .
2. Generate  $y \sim \text{Unif}(A(x))$ .
3. If  $y \notin \Omega$ , goto 2.
4. Generate  $U \sim \text{Unif}$ . If  $U < \frac{|A(x)|}{|A(x)| + |A(y)|}$ , then put  $x = y$ .
5. Goto 2.

The transition kernel of this algorithm is

$$P(x, A) = \int_{A(x) \cap A} \frac{dy}{|A(x)|} \frac{|A(x)|}{|A(x)| + |A(y)|} + 1_{x \in A} \left( \frac{|A(x) \setminus \Omega|}{|A(x)|} + \int_{A(x) \cap \Omega} \frac{dy}{|A(x)|} \frac{|A(y)|}{|A(x)| + |A(y)|} \right)$$

or equivalently

$$P(x, A) = \int_{A(x) \cap A} \frac{dy}{|A(x)| + |A(y)|} + 1_{x \in A} \left( 1 - \int_{A(x) \cap \Omega} \frac{dy}{|A(x)| + |A(y)|} \right)$$

for any  $x \in \Omega$  and any  $A \in \mathcal{A}$ . It must be pointed out that in general  $P$  is not  $u$ -reversible. However, reversibility is ensured as soon as

$$x \in A(y) \iff y \in A(x)$$

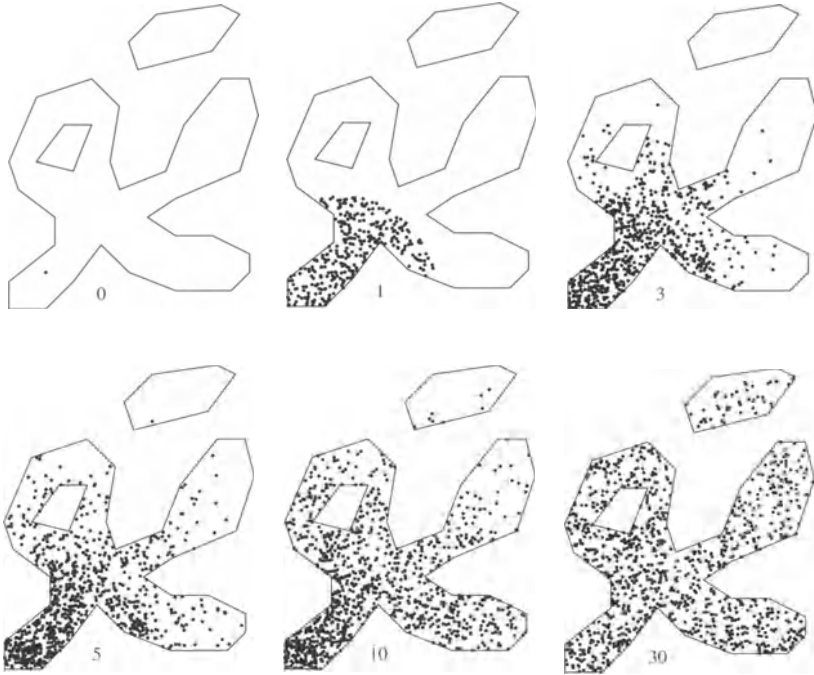
holds for any  $x, y \in \Omega$ . This is typically the case when the  $A(x)$  are balls of same radius associated with a given metric<sup>3</sup> (euclidean or geodesic for

<sup>3</sup> In the case of the euclidean metric, all the  $A(x)$  have the same volume. In this case, there is no risk in accepting any candidate point in  $\Omega$ . Steps 4 and 5 of algorithm 8.3.2 can be reduced to "Put  $x = y$  and goto 2".

instance). Another possibility is to take for each  $A(x)$  the subset of  $\Omega$  directly seen from point  $x$  (i.e. the *star* of  $x$  in  $\Omega$ ).

In the example of Figure 8.7, the  $A(x)$  are Euclidean disks. Their common radius has been fixed to a fifth of the width of the domain  $\Omega$ . It has been chosen small enough so that an important part of each  $A(x)$  is contained in  $\Omega$ , and large enough to allow transitions between both connected components of  $\Omega$ .

Because not all pairs of points communicate directly, the rate of convergence of Algorithm 8.3.2 is expected to be slower than that of 8.3.1. This is confirmed in Figure 8.7 where quasi-uniformity has not been completely reached after 30 iterations. As a matter of fact, Algorithm 8.3.2 can be improved in a number of ways. For instance, the  $A(x)$  can be modified at each iteration. This makes it possible to adapt their shape and size to the current location.



**Fig. 8.7.** 1000 independent simulations carried out using a local approach (cf. Algorithm 8.3.2). They are displayed at iterations 1, 3, 5, 10 and 30. The initial point is the same for all simulations

## 8.4 Application to conditional simulations

Let  $(\Omega, \mathcal{A}, p)$  be a probability space. Suppose that  $p$  can be simulated as the limit of a stationary Markov chain with known transition kernel  $P$ . Because of a set of conditions or constraints, not all states of  $\Omega$  are allowed. Only those located within a measurable subset  $\Omega_c$  of  $\Omega$  are permitted. Let  $\mathcal{A}_c$  be the  $\sigma$ -algebra induced by  $\mathcal{A}$  on  $\Omega_c$  ( $A \in \mathcal{A}_c$  if there exists  $B \in \mathcal{A}$  such that  $A = B \cap \Omega_c$ ). Assume that  $0 < p(\Omega_c) < 1$ . The problem is to simulate the distribution  $p_c$  on  $(\Omega_c, \mathcal{A}_c)$  defined by

$$P_c(A) = \frac{p(A)}{p(\Omega_c)} \quad A \in \mathcal{A}_c$$

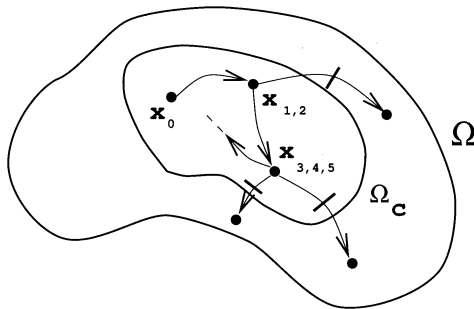
In what follows, two different algorithms will be considered. The first one consists of *restricting the transition kernel*  $P$  to  $\Omega_c \times \mathcal{A}_c$  (Lantuéjoul, 1996). The second one is the version of *simulated annealing* proposed by Hegstad et al. (1995).

### 8.4.1 Restricting the transition kernel

The principle is quite simple. Excursions outside  $\Omega_c$  are forbidden. Here is the algorithm:

**Algorithm 8.4.1.** (*Restriction of a Markov chain*)

1. Let  $x \in \Omega_c$ .
2. Generate  $y \sim P(x, \cdot)$ .
3. If  $y \in \Omega_c$ , put  $x = y$ .
4. Goto 2.



**Fig. 8.8.** Simulation by restricting a transition kernel. Starting from the current state  $x_n$ , a transition kernel on  $\Omega$  is run to generate a candidate state  $y$ . The new generated state is then  $x_{n+1} = y$  if  $y \in \Omega_c$  and  $x_n$  if not

The states generated constitute a realization of a Markov chain on  $(\Omega_c, \mathcal{A}_c)$  with the transition kernel

$$P_c(x, A) = P(x, A) + 1_{x \in A} P(x, \Omega \setminus \Omega_c) \quad x \in \Omega_c, A \in \mathcal{A}_c$$

The invariance of  $p$  under  $P$  by no means implies that of  $p_c$  under  $P_c$  (see exercise 8.3). However, if  $P$  satisfies the further assumption of *reversibility*, namely

$$\int_B p(dx) P(x, A) = \int_A p(dx) P(x, B) \quad A, B \in \mathcal{A}$$

then we have for any  $A, B \in \mathcal{A}_c$

$$\begin{aligned} \int_B p(dx) P_c(x, A) &= \int_B p(dx) P(x, A) + \int_{B \cap A} p(dx) P(x, \Omega \setminus \Omega_c) \\ &= \int_A p(dx) P(x, B) + \int_{A \cap B} p(dx) P(x, \Omega \setminus \Omega_c) \\ &= \int_A p(dx) P_c(x, B) \end{aligned}$$

Now the division by  $p(\Omega_c)$  shows that  $P_c$  is also reversible

$$\int_B p_c(dx) P_c(x, A) = \int_A p_c(dx) P_c(x, B),$$

which is sufficient to ensure that  $p_c$  is invariant under  $P_c$

The  $p_c$ -irreducibility of  $P_c$  cannot be guaranteed by algorithm 8.4.1 because it depends on the shape of  $\Omega_c$ . This has to be studied on a case by case basis.

Regarding the aperiodicity of  $P_c$ , we have the following result.

**Proposition 8.4.1.** *If  $P$  is reversible and  $P_c$   $p_c$ -irreducible, then  $P_c$  is aperiodic.*

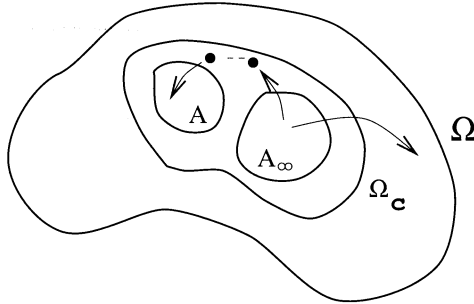
**Proof:** Let  $x \in \Omega_c$  and  $A \in \mathcal{A}_c$  such  $p_c(A) > 0$ . We have to show that  $P_c^{(n)}(x, A) > 0$  as soon as  $n$  is large enough. Consider the sequence  $(A_n, n \in \mathbb{N})$  defined as follows

$$A_n = \{y \in \Omega_c \mid P(y, \Omega \setminus \Omega_c) > 0, P_c^{(n)}(y, A) > 0\}$$

The  $A_n$ 's are increasing events of  $\mathcal{A}_c$ . Let  $A_\infty = \cup_n A_n$ . It is intuitively clear that  $p_c(A_\infty) > 0$  because  $\Omega_c$  and  $\Omega \setminus \Omega_c$  must communicate. A more formal proof is given in the appendix of this chapter.

As an immediate consequence, we have  $p_c(A_n) > 0$  for some index  $n$ . Now, as  $P_c$  is  $p_c$ -irreducible, there exists  $k$  such that  $P_c^{(k)}(x, A_n) > 0$ . Moreover, it stems from the definition of  $A_n$  that  $P^{(l)}(y, A) > 0$  for any  $y \in A_n$  and for any index  $l \geq n$ . Accordingly

$$P_c^{(k+l)}(x, A) \geq \int_{A_n} P_c^{(k)}(x, dy) P_c^{(l)}(y, A) > 0$$



**Fig. 8.9.**  $A_\infty$  is the set of states of  $\Omega_c$  that communicate with  $A$  in a number finite of iterations and from which a direct transition to  $\Omega \setminus \Omega_c$  takes place with positive probability

holds for any  $l \geq n$ , which establishes the proposition.  $\square$

Because of theorem 8.1.1, the conditions of proposition 8.4.1 are sufficient to ensure the correctness of the simulation algorithm obtained by restricting a transition kernel. It should be pointed out that the assumption  $P(\Omega_c) > 0$  is imperative. The algorithm cannot work if  $\Omega_c$  almost never occurs.

#### 8.4.2 Simulated annealing

Since its introduction by Kirkpatrick (1983) as a way for improving combinatorial optimization procedures, simulated annealing has proved to be a powerful algorithm for geostatistical simulations, especially when they are subject to complicated conditions or constraints (Hegstad et al., 1994). In this section we keep the same assumptions and notation as in the previous section. In particular the allowed set of states  $\Omega_c$  has a positive frequency of occurrence.

A convenient way to account for the conditions or the constraints is to resort to an *objective function*. Let us denote it by  $C$ .  $C$  is a measurable and positive function defined on  $\Omega$ . It characterizes  $\Omega_c$  in the sense that the value  $C(x)$  assigned to the state  $x$  is equal to 0 if and only if  $x$  is allowed. In general  $C$  is designed so that the smaller  $C(x)$ , the closer  $x$  to  $\Omega_c$ .

The next step is to incorporate the objective function into distributions on  $(\Omega, \mathcal{A})$ . Let us consider the following distributions

$$p_t(A) \propto \int_A p(dx) \exp \left\{ -\frac{C(x)}{t} \right\} \quad A \in \mathcal{A}$$

where  $t$  is a strictly positive parameter usually called temperature in reference to the metallurgical cooling process which inspired the algorithm. The reason for introducing this family of distributions is that  $p_t$  tends to  $p_c$  when the temperature becomes very low. More precisely:

**Proposition 8.4.2.**  $\lim_{t \rightarrow 0} \|p_t - p_c\| = 0$

**Proof:** For any bounded, measurable function  $f$ , we have

$$\int_{\Omega} p_t(dx) f(x) dx = \int_{\Omega} p(dx) \frac{\exp\left\{-\frac{C(x)}{t}\right\}}{\int_{\Omega} p(dy) \exp\left\{-\frac{C(y)}{t}\right\}} f(x)$$

and since the integral in the denominator is bounded from below by  $p(\Omega_c)$ , the dominated convergence theorem can be applied, and we have

$$\lim_{t \rightarrow 0} \int_{\Omega} p_t(dx) f(x) = \int_{\Omega} p(dx) \frac{1_{x \in \Omega_c}}{p(\Omega_c)} f(x) = \int_{\Omega_c} p_c(dx) f(x)$$

□

Proposition 8.4.2 suggests a simulation algorithm for  $p_c$  that consists of simulating the distribution  $p_t$  and letting the temperature  $t$  decrease simultaneously.

Using a Metropolis type algorithm, it is relatively easy to simulate  $p_t$ . Starting from  $x$ , a candidate state  $y$  is generated from  $P(x, \cdot)$  (the reversible transition kernel used to simulate  $p$ ) and accepted with probability

$$\alpha(x, y) = \min\left(\frac{h(y)}{h(x)}, 1\right)$$

where we have put  $h(x) = \exp\{-C(x)/t\}$  for short. This gives the transition kernel

$$P_t(x, A) = \int_A P(x, dy) \alpha(x, y) + 1_{x \in A} \int_{\Omega} P(x, dy) [1 - \alpha(x, y)]$$

Checking that  $p_t$  is invariant under  $P_t$  is straightforward. Replacing  $P_t$  by its definition, we have

$$\begin{aligned} \int_{\Omega} p_t(dx) P_t(x, A) &\propto \int_A p(dx) \int_{\Omega} P(x, dy) \min(h(x), h(y)) + \int_A p(dx) h(x) \\ &\quad - \int_{\Omega} p(dx) \int_A P(x, dy) \min(h(x), h(y)) \end{aligned}$$

and both double integrals cancel each other because  $P$  is reversible. Indeed, reversibility plays exactly the same role as symmetry in the standard Metropolis algorithm. A similar calculation shows that  $P_t$  is also reversible.

The irreducibility and the aperiodicity of  $P_t$  stem directly from those of  $P$ .

Regarding the cooling schedule, things are more complicated. Several authors (Geman and Geman 1984; Hajek, 1988; Winkler, 1995) have shown that decreasing the temperature too quickly may result in being trapped in a local minimum of the objective function. A typical recommended choice is a temperature proportional to  $1/\ln(n+1)$  at the  $n^{\text{th}}$  iteration.

**Remark 8.4.1.** In contrast to the restriction algorithm, the conditions are not satisfied at each iteration. Consequently, if the simulated annealing algorithm is stopped after  $n$  iterations, then the conditions may not be perfectly honored. In order to cope with this problem, one could consider resuming the iterations until an allowed state is reached. Unfortunately this does not work because even after a large number of iterations the first allowed state encountered may be biased.

**Remark 8.4.2.** Simulated annealing can also be applied in the case when  $p(\Omega_c) = 0$ , but care must be taken to avoid biases. Hwang (1980) and Lajaudine (1998) gives explicit and convincing examples.

## Appendix

This appendix completes the proof of proposition 8.4.1 by showing that  $p(A_\infty) > 0$ .

Suppose  $p(A_\infty) = 0$  and consider the sequence

$$I_n = \int_{\Omega_c} p(dy) P^{(n)}(y, \Omega \setminus \Omega_c) \quad n \geq 1$$

Clearly we have

$$I_1 = \int_{\Omega_c} p(dy) P(y, \Omega \setminus \Omega_c) = \int_{A_\infty} p(dy) P(y, \Omega \setminus \Omega_c) \leq p(A_\infty) = 0$$

More generally, since  $p(A_\infty) = 0$  and  $P(y, \Omega \setminus \Omega_c) = 0$  if  $y \in \Omega_c \setminus A_\infty$ , we can write

$$I_{n+1} = \int_{\Omega_c \setminus A_\infty} p(dy) P^{(n+1)}(y, \Omega \setminus \Omega_c) = \int_{\Omega_c \setminus A_\infty} p(dy) \int_{\Omega_c} P(y, dz) P^{(n)}(z, \Omega \setminus \Omega_c)$$

and the reversibility of  $P$  implies

$$I_{n+1} = \int_{\Omega_c} p(dy) P^{(n)}(y, \Omega \setminus \Omega_c) \int_{\Omega_c \setminus A_\infty} P(y, dz) \leq I_n$$

so that we finally have  $I_n = 0$  for any  $n \geq 1$ . When  $n$  becomes very large, the total variation convergence of  $P^{(n)}(y, \cdot)$  to  $p$  implies

$$0 = \lim_{n \rightarrow \infty} \int_{\Omega_c} p(dy) P^{(n)}(y, \Omega \setminus \Omega_c) = \int_{\Omega_c} p(dy) p(\Omega \setminus \Omega_c) = p(\Omega_c) p(\Omega \setminus \Omega_c)$$

which is a contradiction since  $0 < p(\Omega_c) < 1$ . □



## Exercises

**8.1** Let  $0 < \epsilon < 1$ . Show that if  $P$  is  $p$ -irreducible but not necessarily aperiodic, then

$$P_\epsilon(x, A) = (1 - \epsilon) \sum_{n=1}^{\infty} \epsilon^{n-1} P^{(n)}(x, A) \quad x \in \Omega, A \in \mathcal{A}$$

is  $p$ -irreducible and aperiodic.

**8.2** Show that the candidate transition  $Q$  of Metropolis algorithm needs not be symmetric if the acceptance criterion is replaced by

$$\alpha(x, y) = \min\left(\frac{p(y)Q(y, x)}{p(x)Q(x, y)}, 1\right)$$

This algorithm (and some other variations) is due to Hastings (1970).

**8.3** Let  $\Omega = \mathbb{N}$ , and let  $P$  the transition kernel defined by  $P(x, y) = \frac{1}{2}$  if  $y = x + 1$  or  $y = 0$ .

1) Show that the distribution  $p(x) = 2^{-(x+1)}$  is invariant for  $P$ , and that  $P$  is  $p$ -irreducible as well as aperiodic.

2) Apply the restriction simulation algorithm in the 3 following cases

- $\Omega_c = \mathbb{N} \setminus \{0\}$
- $\Omega_c = \{2x, x \in \mathbb{N}\}$
- $\Omega_c = \{2x + 1, x \in \mathbb{N}\}$

What conclusions can be drawn?

## 9. Rate of convergence of iterative algorithms

In the previous chapter, we showed how to simulate a distribution as the limit of an iterative algorithm. In practice, these algorithms cannot be run eternally; they have to be stopped after a finite number of iterations. Of course there is no reason for the distribution thus simulated to be identical to the limit distribution, and the number of iterations to carry out must be chosen so that their difference lies below a prescribed level of acceptability. This requires the study of the *rate of convergence* of the iterative algorithm.

A considerable amount of literature has been devoted to the determination of the rate of convergence of algorithms based on Markovian iterations (in particular Nummelin, 1984; Meyn and Tweedie, 1993; Tierney, 1994; Dufflo, 1996). This chapter only deals with two cases corresponding to two different assumptions on the transition kernel of the algorithm:

- If the transition kernel is *minorized* by a positive measure, then the rate of convergence is uniform (section 9.2).
- If the transition kernel admits an *isofactorial representation*, then a geometric rate of convergence is expected in many cases (section 9.3). The integral range introduced in chapter 4 can be used to determine it empirically (section 9.4).

### 9.1 Rates of convergence

Let  $p$  be a distribution on  $(\Omega, \mathcal{A})$  that can be simulated as the limit of a  $p$ -irreducible and aperiodic Markov chain with transition kernel  $P$ . As was seen in the last chapter, the distance between the  $n^{\text{th}}$  iterate  $P^{(n)}(x, \cdot)$  and  $p$  can be specified by the total variation of their difference

$$\| P^{(n)}(x, \cdot) - p \| = 2 \sup_{A \in \mathcal{A}} \left| P^{(n)}(x, A) - p(A) \right| \quad x \in \Omega$$

This distance depends on the initial state  $x$  and the iteration step  $n$ .

**Definition 9.1.1.** *The transition kernel  $P$  is said to be geometrically ergodic if there exist a positive measurable function  $\alpha$  on  $\Omega$  and a constant  $\epsilon$  with  $0 < \epsilon < 1$  such that*

$$\| P^{(n)}(x, \cdot) - p \| \leq \alpha(x) \epsilon^n$$

for any  $x \in \Omega$  and any  $n \in \mathbb{N}^1$ .

A simple observation shows that this rate of convergence is not very strong. Suppose that the Markov chain starts from an initial distribution  $p_0$  which is not concentrated on a unique point. Then after  $n$  iterations we have

$$\| P^{(n)}(p_0, \cdot) - p \| = \int_{\Omega} \| P^{(n)}(x, \cdot) - p \| p_0(dx) \leq \epsilon^n \int_{\Omega} \alpha(x) p_0(dx)$$

from which no conclusion can be drawn unless  $\alpha$  is  $p_0$ -integrable. Things become more interesting in the case when  $\alpha$  is bounded because it is automatically  $p_0$ -integrable. This leads to the following definition:

**Definition 9.1.2.** *The transition kernel  $P$  is said to be uniformly ergodic if there exist two constants  $\alpha > 0$  and  $0 < \epsilon < 1$  such that*

$$\| P^{(n)}(x, \cdot) - p \| \leq \alpha \epsilon^n$$

for any  $x \in \Omega$ .

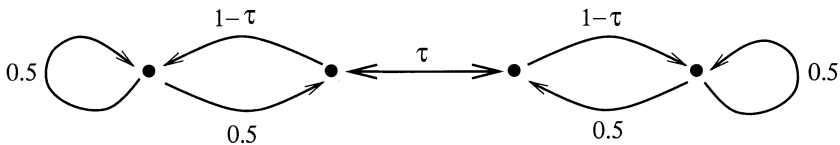
In this case, we also have

$$\| P^{(n)}(p_0, \cdot) - p \| \leq \alpha \epsilon^n$$

for any initial distribution  $p_0$ . Thus the rate of convergence is mainly governed by  $\epsilon$ . It is fast provided that this number is not too close to 1.

## 9.2 Minorization

What slows down the convergence of a Markov chain? Mainly, the presence of "bottlenecks" that prevent classes of states from communicating easily. Figure 9.1 shows an example of a bottleneck.



**Fig. 9.1.** A small value for  $\tau$  creates a bottleneck in the communication between two classes of states. Although not impossible, the transition from one class to the other is difficult, which makes the convergence rather slow

The following assumption guarantees that the bottlenecks, if any, have only a moderate influence:

<sup>1</sup> This definition is not standard. Some authors allow  $\alpha$  to take infinite values at some states but require it to be  $p$ -integrable (e.g. Tierney, 1994).

**Definition 9.2.1.** *The transition kernel  $P$  is minorized if there exist a probability measure  $q$  on  $(\Omega, \mathcal{A})$  and a constant  $\epsilon > 0$  such that*

$$P(x, A) \geq \epsilon q(A) \quad x \in \Omega, A \in \mathcal{A}$$

The minorization of  $P$  gives a minimum chance for a direct transition from any state  $x$  to any event  $A$ . Indeed, it also gives a maximum chance to that transition because

$$P(x, A) = 1 - P(x, A^c) \leq 1 - \epsilon q(A^c) = 1 - \epsilon + \epsilon q(A)$$

At this point, it can be noted that  $\epsilon \leq 1$  (use the inequality of definition 9.2.1 with  $A = \Omega$ ). The case  $\epsilon = 1$  necessarily leads to  $P(x, \cdot) = q$ . More importantly, the bounds of both inequalities do not depend on the initial state  $x$ , which suggests that the communication between the states tends to be uniformly favoured by minorization. Not surprisingly, a uniform rate of convergence is obtained.

**Theorem 9.2.1.** *If  $P$  is minorized by  $\epsilon q$  with  $\epsilon < 1$ , then*

$$\| P^{(n)}(x, \cdot) - p \| \leq 2(1 - \epsilon)^n$$

for any  $x \in \Omega$ .

A proof by induction of this theorem is given in the appendix of this chapter.

**Example 9.2.1.** Let us come back to the hit-and-run algorithm for simulating a uniform point within the open subset  $\Omega \subset \mathbb{R}^d$  (see section 8.3). The transition kernel admits a density function given by

$$f(x, y) = \frac{2}{d\omega_d} \frac{1}{\ell(x, y, \Omega) |x - y|^{d-1}} \quad y \neq x,$$

where  $\ell(x, y, \Omega)$  denotes the length of the intersection between  $\Omega$  and the line passing through  $x$  and  $y$ . Suppose that  $\Omega$  is *bounded* with diameter  $\delta = \sup_{x, y \in \Omega} |x - y|$ . Then  $\ell(x, y, \Omega) |x - y|^{d-1} \leq \delta^d$ , which gives the minorization property

$$P(x, A) = \int_A f(x, y) dy \geq \epsilon p(A) \quad \text{with } \epsilon = \frac{2}{d\omega_d} \frac{|\Omega|}{\delta^d} \quad \text{and } p(A) = \int_A \frac{dy}{|\Omega|}$$

It remains to show that the constant  $\epsilon$  is strictly less than 1. But a classical isoperimetric inequality states that the volume of  $\Omega$  is at most equal to  $\omega_d(\delta/2)^d$  (volume of a ball with diameter  $\delta$ ). Accordingly

$$\epsilon = \frac{2}{d\omega_d} \frac{|\Omega|}{\delta^d} \leq \frac{2}{d\omega_d} \frac{\omega_d}{\delta^d} \left(\frac{\delta}{2}\right)^d = \frac{1}{d2^{d-1}} < 1$$

**Remark 9.2.1.** In the case when  $\Omega$  is *unbounded*, the previous reasoning can no longer be applied. It is even doubtful that there exists  $\epsilon > 0$  such that the minorization condition  $P(x, A) \geq \epsilon p(A)$  still holds. The convergence of  $P^{(n)}(x, \cdot)$  to  $p$  is probably not uniform and not even geometric. A weaker mode of convergence, such as that proposed by Baxter and Rosenthal (1995), seems to be more appropriate.

**Remark 9.2.2.** Regarding the Gibbs sampler for uniform distributions there is no general result, even in the case when  $\Omega$  is bounded. Roberts and Rosenthal (1998) show that a uniform rate of convergence is expected when the boundary of  $\Omega$  is twice differentiable. In the same time, Bésislé (1997) exhibits examples where the convergence can be arbitrarily slow if the boundary is sufficiently irregular.

### 9.3 Isofactorial representation

To simplify the presentation of this section, we assume that  $\Omega$  is discrete and that  $p(x) > 0$  for any  $x \in \Omega$ . Let  $L^2(\Omega, p)$  be the vector space of the numerical functions that are defined on  $\Omega$  and are square integrable for  $p$

$$f \in L^2(\Omega, p) \iff \sum_{x \in \Omega} f^2(x)p(x) < \infty$$

A classical result is that  $L^2(\Omega, p)$  is a Hilbert space when endowed with the scalar product  $\langle f, g \rangle = \sum_{x \in \Omega} f(x)g(x)p(x)$ . The norm attached to the element  $f \in L^2(\Omega, p)$  is denoted by  $\|f\|_2 = \sqrt{\langle f, f \rangle}$ .

Let  $f \in L^2(\Omega, p)$ . It is not totally obvious that the quantity  $\sum_{y \in \Omega} P(x, y)f(y)$  is finite for any  $x \in \Omega$ . Indeed using Jensen's inequality<sup>2</sup> and the inequality  $p(x)P(x, y) \leq p(y)$  stemming from the invariance of  $p$  under  $P$ , we have

$$\left( \sum_{y \in \Omega} P(x, y)f(y) \right)^2 \leq \sum_{y \in \Omega} P(x, y)f^2(y) \leq \sum_{y \in \Omega} \frac{p(y)}{p(x)} f^2(y) = \frac{1}{p(x)} \|f\|_2^2 < \infty$$

So it is possible to introduce the function  $Pf$  defined on  $\Omega$  as follows

$$Pf(x) = \sum_{y \in \Omega} P(x, y) f(y) \quad x \in \Omega$$

Moreover, we have  $Pf \in L^2(\Omega, p)$ :

---

<sup>2</sup>  $\varphi\left(\sum_{x \in \Omega} xp(x)\right) \leq \sum_{x \in \Omega} \varphi(x)p(x)$  for any convex function  $\varphi$ .

$$\begin{aligned} \sum_{x \in \Omega} [Pf(x)]^2 p(x) &= \sum_{x \in \Omega} \left( \sum_{y \in \Omega} P(x, y) f(y) \right)^2 p(x) \\ &\leq \sum_{x \in \Omega} \sum_{y \in \Omega} P(x, y) f^2(y) p(x) = \sum_{y \in \Omega} f^2(y) \sum_{x \in \Omega} P(x, y) p(x) \\ &= \sum_{y \in \Omega} f^2(y) p(y) = \|f\|_2^2 < \infty \end{aligned}$$

Consequently, the transition kernel  $P$  induces an operator in  $L^2(\Omega, p)$  by  $f \rightarrow Pf$ . This operator is also denoted by  $P$ . It is clearly linear, and the inequality  $\|Pf\|_2 \leq \|f\|_2$  just established says that it is a contraction.

**Assumption 9.3.1.** *The operator  $P$  admits the isofactorial representation*

$$Pf(x) = \sum_{i=0}^{\infty} \lambda_i u_i(x) \langle f, u_i \rangle$$

where  $(\lambda_i, i \in \mathbb{N})$  is a sequence of real numbers and the  $u_i$ 's (the factors) constitute an orthonormal Hilbert basis of  $L^2(\Omega, p)$ . By convention  $u_0$  is the vector whose components are all equal to 1.

Isofactorial representations are associated with Hilbert-Schmidt operators<sup>3</sup>, and more generally with self-adjoint<sup>4</sup> and compact<sup>5</sup> operators (Dunford and Schwartz, 1967; Brézis, 1983). Isofactorial representations are not necessarily unique. For instance, the identity operator  $I_d$  can be expressed in an isofactorial form starting from any orthonormal basis of  $L^2(\Omega, p)$  (associated with  $\lambda_i = 1$  for each  $i \in \mathbb{N}$ ). As an immediate consequence, any linear combination of the identity and a self-adjoint, compact operator admits an isofactorial expansion even if it is not compact. Moreover it has the same factors as those of the compact operator from which it has been derived.

Explicit examples of isofactorial expansions will be given a little later on. For the time being, let us examine the implications of this definition by considering particular choices of  $f$ .

– If  $f = 1_{\cdot=y}$  (indicator function of state  $y$ ), we obtain

$$P(x, y) = Pf(x) = \sum_{i=0}^{\infty} \lambda_i u_i(x) u_i(y) p(y)$$

<sup>3</sup>  $P$  is a Hilbert-Schmidt operator if  $P$  admits an isofactorial expansion with  $\sum_i \lambda_i^2 < \infty$ .

<sup>4</sup>  $P$  is self-adjoint if  $\langle Pf, g \rangle = \langle Pg, f \rangle$  for any  $f, g \in L^2(\Omega, p)$ .

<sup>5</sup>  $P$  is compact if any bounded sequence  $(f_i, i \in \mathbb{N})$  of  $L^2(\Omega, p)$  admits a strongly convergent subsequence  $(Pf_{i_k}, k \in \mathbb{N})$ . If  $\Omega$  is countable, then the identity operator  $I_d$  is not compact. For instance, the orthonormal basis  $(u_i, i \in \mathbb{N})$  is a bounded sequence, but  $\|I_d u_i - I_d u_j\|_2 = \|u_i - u_j\|_2 = \sqrt{2}$ , so that  $(I_d u_i, i \in \mathbb{N})$  has no strongly convergent subsequence.

From this we derive  $p(x)P(x, y) = p(y)P(y, x)$ , which means that  $P$  is reversible. This in turn implies that  $P$  is self-adjoint.

- If  $f = u_i$ , then the orthonormality property gives at once

$$Pu_j(x) = \sum_{i=0}^{\infty} \lambda_i u_i(x) \langle u_j, u_i \rangle = \sum_{i=0}^{\infty} \lambda_i u_i(x) 1_{i=j} = \lambda_j u_j(x)$$

which shows that the factor  $u_j$  is an eigenvector of  $P$  for the eigenvalue  $\lambda_j$ .

- If  $f = P(\cdot, y)$ , then a direct calculation gives  $Pf(x) = P^{(2)}(x, y)$ . On the other hand

$$Pf(x) = \sum_{i=0}^{\infty} \lambda_i u_i(x) \langle P(\cdot, y), u_i \rangle$$

But  $\langle P(\cdot, y), u_i \rangle = \lambda_i u_i(y) p(y)$  because  $P$  is reversible and  $u_i$  is an eigenvector of  $P$ . We thus obtain

$$P^{(2)}(x, y) = \sum_{i=0}^{\infty} \lambda_i^2 u_i(x) u_i(y) p(y)$$

More generally, a simple reasoning by induction on the iteration order gives

$$P^{(n)}(x, y) = \sum_{i=0}^{\infty} \lambda_i^n u_i(x) u_i(y) p(y)$$

Here is a consequence regarding the eigenvalues of  $P$ .

**Proposition 9.3.1.** *All the eigenvalues of  $P$  lie between  $-1$  and  $+1$ . Moreover,  $\lambda_0 = 1$  is the only eigenvalue of modulus 1.*

**Proof:** Since  $P$  is a contraction, we can write

$$1 = \|u_i\|_2^2 \geq \|Pu_i\|_2^2 = \|\lambda_i u_i\|_2^2 = \lambda_i^2$$

It follows that  $|\lambda_i| \leq 1$ . Note also that  $\lambda_0 = 1$  because of the convention  $u_0 = 1$ . Suppose now that  $|\lambda_i| = 1$  for some index  $i > 0$ . Let us choose a state  $x$  such that  $u_i(x) \neq 0$ . Then

$$p(x) = \lim_{n \rightarrow \infty} P^{(2n)}(x, x) \geq \lim_{n \rightarrow \infty} (\lambda_0^{2n} u_0^2(x) + \lambda_i^{2n} u_i^2(x)) p(x) > p(x)$$

which is a contradiction. □

We now turn to the evaluation of the rate of convergence of the Markov chain with transition kernel  $P$ . Since  $\Omega$  is discrete, the total variation between  $P^{(n)}(x, \cdot)$  and  $p$  can be written as

$$\| P^{(n)}(x, \cdot) - p \| = \sum_{y \in \Omega} \left| \frac{P^{(n)}(x, y)}{p(y)} - 1 \right| p(y)$$

and because of Jensen's inequality

$$\| P^{(n)}(x, \cdot) - p \|^2 \leq \sum_{y \in \Omega} \left( \frac{P^{(n)}(x, y)}{p(y)} - 1 \right)^2 p(y) = \sum_{y \in \Omega} \frac{[P^{(n)}(x, y)]^2}{p(y)} - 1$$

Now, the reversibility formula  $p(x)P^{(n)}(x, y) = p(y)P^{(n)}(y, x)$  can be applied, which leads to

$$\| P^{(n)}(x, \cdot) - p \|^2 \leq \sum_{y \in \Omega} \frac{P^{(n)}(x, y) p(y) P^{(n)}(y, x)}{p(x)p(y)} - 1 = \frac{P^{(2n)}(x, x)}{p(x)} - 1$$

Expanding  $P^{(2n)}(x, x)$  and using the fact that  $\lambda_0 = 1$  and  $u_0 = 1$ , we obtain

$$\| P^{(n)}(x, \cdot) - p \|^2 \leq \sum_{i \neq 0} \lambda_i^{2n} u_i^2(x)$$

At this point, nothing further can be said without additional assumptions.

**Assumption 9.3.2.** *The set of eigenvalues has no accumulation points at 1 and  $-1$ .*

This assumption is not very strong. It is satisfied if  $P$  is self-adjoint and compact because its eigenvalues have at most one accumulation point at 0 (Dunford and Schwartz, 1967; Brézis, 1983) and more generally if  $P$  can be written as  $(1 - \tau)I_d + \tau Q$  where  $Q$  is a self-adjoint, compact transition kernel and  $0 < \tau \leq 1$ .

One way to express assumption 9.3.2 mathematically is to say that there exists a constant  $\epsilon$  with  $0 < \epsilon < 1$  such that  $|\lambda_i| \leq \epsilon$  for any  $i \neq 0$ . Indeed, we can take  $\epsilon = \max_{i \neq 0} |\lambda_i|$ . Then we have

$$\| P^{(n)}(x, \cdot) - p \|^2 \leq \epsilon^{2n} \sum_{i \neq 0} u_i^2(x) = \epsilon^{2n} \frac{1 - p(x)}{p(x)}$$

on account of the formula  $\sum_i u_i^2(x)p(x) = 1^6$ . Finally

**Theorem 9.3.1.** *If  $P$  admits an isofactorial expansion and if its eigenvalues have no accumulation points at  $\pm 1$ , then  $P$  is geometrically ergodic with*

$$\epsilon = \max_{i \neq 0} |\lambda_i| \quad \alpha(x) = \sqrt{\frac{1 - p(x)}{p(x)}}$$

---

<sup>6</sup> Expand  $1_{\cdot=x}$  in the basis and set  $\cdot = x$ .



**Example 9.3.1.** Here we take  $\Omega = \mathbb{N}$ .  $P$  is called a *Jacobi transition kernel* if only direct transitions between neighbouring states can take place. Then  $P$  is an infinite tridiagonal matrix satisfying  $P(x, y) = 0$  if  $|x - y| > 1$ . A slight extension of an argument by Akhiezer and Glazman (1966) shows that  $P$  is a compact operator on  $L^2(\mathbb{N}, p)$  if  $\lim_{x \rightarrow \infty} P(x, x - 1) = 1$ . It follows from theorem 9.3.1 that the rate of convergence of  $P^{(n)}(x, \cdot)$  to  $p$  is at least geometric. Actually, there is little chance for it to be uniform precisely because of the shape of the Jacobi transition kernel which does not facilitate the mixing between non neighbouring states. Generally, simulation algorithms based on such transition kernels are easy to design and to implement but a price has to be paid for that in terms of the rate of convergence.

As an illustration, let us consider the simulation of the Poisson distribution

$$p(x) = \exp(-\theta) \frac{\theta^x}{x!} \quad x \in \mathbb{N}$$

which is invariant under the Jacobi transition kernel defined as follows

$$P(x, x-1) = \frac{x}{\theta + x}, \quad P(x, x) = \frac{\theta}{(\theta + x)(\theta + x + 1)}, \quad P(x, x+1) = \frac{\theta}{\theta + x + 1}$$

Clearly,  $\lim P(x, x - 1) = 1$ , so that  $P$  is a compact operator. Therefore a geometric rate of convergence is expected. It remains to evaluate the constant  $\epsilon$  as well as the function  $\alpha$ .

In the present case, an explicit, although tedious determination of all eigenvalues of  $P$  is possible, so that an optimal value for  $\epsilon$  can be given

$$\epsilon = \max_{i \neq 0} |\lambda_i| = \sqrt{\frac{\theta}{\theta + 1}}$$

The smaller  $\theta$ , the faster the convergence.

In contrast to this, the evaluation of  $\alpha$  is straightforward. We find explicitly

$$\alpha(x) = \sqrt{\frac{1 - p(x)}{p(x)}} \leq \sqrt{\exp(\theta) \frac{x!}{\theta^x}}$$

For large  $x$ ,  $\alpha(x)$  can be very large. Accordingly, it is recommended to start running the simulation algorithm with a small initial  $x$  value.

**Remark 9.3.1.** The Jacobi transition kernel  $P$  is very similar to the transition kernel  $Q$  produced by the Metropolis-Barker procedure for simulating a Poisson distribution (see example 8.2.1). Indeed  $P$  and  $Q$  are related by

$$Q = \frac{1}{2}(I_d + P)$$

In other words, at each iteration a transition of  $Q$  has the same chance of being a transition of  $P$  or no change at all. This suggests that  $Q$  is half as fast as  $P$ . Indeed, a direct calculation gives

$$\| Q^{(n)}(x, \cdot) - p \| \leq \left( \frac{1 + \epsilon}{2} \right)^n \alpha(x)$$

The fact that  $Q$  is geometrically ergodic is not surprising since it admits an isofactorial expansion. Nonetheless  $Q$  is not compact; its eigenvalues have an accumulation point at 0.5.

## 9.4 Empirical determination of the rate of convergence

Suppose also here that the transition kernel  $P$  has the following isofactorial expression

$$P(x, y) = \sum_{i=0}^{\infty} \lambda_i u_i(x) u_i(y) p(y),$$

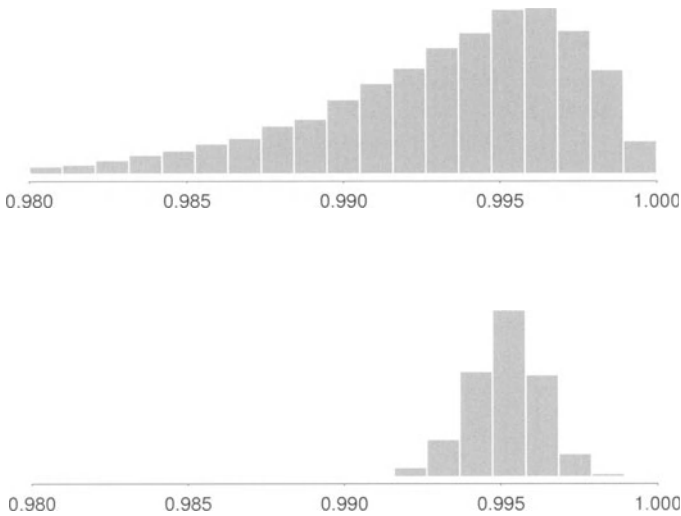
(assumption 9.3.1) with no accumulation of eigenvalues at  $\pm 1$  (assumption 9.3.2). The problem addressed here is to evaluate the quantity  $\epsilon = \sup_{i \neq 0} |\lambda_i|$  empirically.

A simple procedure is to carry out a simulation for a certain number of iterations, say  $N$ , then derive an experimental transition matrix and calculate its eigenvalues (and eigenvectors). In order to reduce the influence of the initial state, the states generated during the first iterations (the "burn-in" period) have to be discarded (Geyer, 1992). The reversibility of the Markov chain can also be exploited to improve the estimation of the transition kernel by recording not only the transitions between iterations  $n$  and  $n + 1$ , but also those between  $n$  and  $n - 1$ .

**Example 9.4.1.** This procedure has been tested by simulating the Poisson distribution with mean  $\theta = 100$  using the transition kernel of example 9.3.1. In order to get an idea of the statistical fluctuations on the estimates of  $\epsilon$ , two sets of 10000 simulations have been carried out. The size of the simulations varies from 1000 (first set) to 10000 (second set) iterations. Figure 9.2 presents the histograms obtained. The one based on large size simulations appears to be much less dispersed around the actual value of  $\epsilon$  (0.9950).

At this point, one realizes that  $\epsilon$  must be estimated with great precision. A change in  $\epsilon$  from 0.99 to 0.999 implies a rate of convergence 10 times slower. From this point of view, the estimation procedure cannot be considered as satisfactory. Indeed it is flawed in that it cannot say what should be the size of the simulation so as to produce a reliable estimate for  $\epsilon$ . In what follows, we propose a criterion using the integral range introduced in chapter 4 to answer that question.

Let us denote by  $X$  a stationary Markov chain with transition kernel  $P$ . Since  $P$  is reversible, there is no inconvenience in assuming  $X$  indexed by the



**Fig. 9.2.** Histogram of the estimates of  $\epsilon$  obtained from 10000 simulations of size 1000 (top) and 10000 (bottom)

relative integers. Let  $f \in L^2(\Omega, p)$ . Then  $f(X)$  is a second-order stationary stochastic process (but not necessarily markovian) with mean  $m_f = \langle f, u_0 \rangle$ , variance  $\sigma_f^2 = \sum_{i \neq 0} \langle f, u_i \rangle^2$ , and covariance function

$$C_f(n) = \sum_{i \neq 0} \lambda_i^{|n|} \langle f, u_i \rangle^2$$

Because the eigenvalues of  $P$  do not accumulate at  $\pm 1$  (assumption 9.3.2) the integral range of  $C_f$  is finite and equal to

$$A_f = \frac{1}{\sigma_f^2} \sum_{n=-\infty}^{+\infty} C_f(n) = \frac{1}{\sigma_f^2} \sum_{i \neq 0} \frac{1 + \lambda_i}{1 - \lambda_i} \langle f, u_i \rangle^2$$

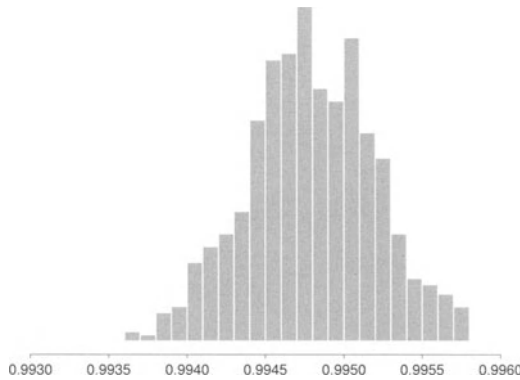
(cf. proposition 4.2.1). We note that  $A_f$  is maximum if  $f$  is proportional to the eigenvector  $u_1$  associated to the greatest eigenvalue  $\lambda_1$  different from 1.

$$A_f \leq \frac{1}{\sigma_f^2} \frac{1 + \lambda_1}{1 - \lambda_1} \sum_{i \neq 0} \langle f, u_i \rangle^2 = \frac{1 + \lambda_1}{1 - \lambda_1} = A_{u_1}$$

In what follows, we propose to estimate  $\lambda$  by using the same procedure as before. The only supplementary requirement is that the simulation size must be large enough to allow the estimation of  $A_{u_1}$ .

**Algorithm 9.4.1.** (*Estimation of  $\lambda$* )

1. Set  $N$  to some prespecified value and simulate the Markov chain  $X$  in  $V = \{0, \dots, N\}$ .
2. Estimate the experimental transition kernel  $P^*$  from the simulation.
3. Estimate the eigenvalues and the eigenvectors of  $P^*$ . Let  $u_1^*$  be the eigenvector associated to the greatest eigenvalue less than 1.
4. Consider the realisation of  $u_1^*(X)$  in  $V$ , and plot  $\log s^2(W|V)$  versus  $\log |W|$  for several subdomains  $W$  that divide  $V$ .
  - 5.1. If a straight line with slope  $-1$  results, then return the eigenvalue that has the greatest modulus less than 1.
  - 5.2. Otherwise, increase  $N$ , extend the simulation of  $X$  and goto 2.



**Fig. 9.3.** Histogram of the  $\epsilon$  values built from 1000 simulations

**Example 9.4.2.** This algorithm has been applied 1000 times on the Markov chain of example 9.4.1. The size of the simulations range from 300000 to 900000 iterations. Each simulation produces an estimate of  $\epsilon$ , and their histogram is displayed in Figure 9.3. It is interesting to compare this histogram with that at the bottom of Figure 9.2. The statistical fluctuations on the integral range of  $u_1(X)$  lead to variations on  $\epsilon$  between 0.9936 and 0.9958, with an average of 0.9948 and a standard deviation of 0.0007. As seen above, the actual value of  $\epsilon$  is 0.9950.

**Appendix:** Proof of theorem 9.2.1.

Let us put  $h^{(n)}(x, A) = P^{(n)}(x, A) - p(A)$ . We have to show that  $|h^{(n)}(x, A)| \leq (1 - \epsilon)^n$ . This will be done by induction.

Consider first the case  $n = 1$ . Since  $p$  is invariant under  $P$ , we have

$$p(A) = \int_{\Omega} p(dx)P(x, A) \geq \epsilon \int_{\Omega} p(dx)q(A) = \epsilon q(A)$$

Consequently, we can write  $p(A) - \epsilon q(A) = (1 - \epsilon)r(A)$  where  $r$  is another probability measure on  $(\Omega, \mathcal{A})$ . In terms of  $r$ , the minorization inequality may be written

$$h(x, A) \geq \epsilon q(A) - p(A) = -(1 - \epsilon)r(A)$$

By complementation, this inequality becomes

$$h(x, A) = -h(x, A^c) \leq (1 - \epsilon)r(A^c)$$

so that we obtain

$$-(1 - \epsilon)r(A) \leq h(x, A) \leq (1 - \epsilon)r(A^c).$$

In particular

$$|h(x, A)| \leq 1 - \epsilon$$

Suppose now that  $|h^{(n)}(x, A)| \leq (1 - \epsilon)^n$  for some  $n \geq 1$ . Since  $p$  is invariant under  $P^{(n)}$ , we have

$$h^{(n+1)}(x, A) = \int_{\Omega} h^{(n)}(x, dy) h(y, A)$$

Now, according to the Hahn decomposition (Halmos, 1969), the signed measure  $h^{(n)}(x, \cdot)$  is the difference of two positive measures with disjoint supports. More precisely, there exists  $\Omega_n \in \mathcal{A}$  such that

$$A \subset \Omega_n \implies h^{(n)}(x, A) \geq 0 \qquad A \subset \Omega_n^c \implies h^{(n)}(x, A) \leq 0$$

Let us separate the positive and the negative part of  $h^{(n)}(x, \cdot)$

$$h^{(n+1)}(x, A) = \int_{\Omega_n} h^{(n)}(x, dy) h(y, A) + \int_{\Omega_n^c} h^{(n)}(x, dy) h(y, A)$$

to get an upper bound for  $h^{(n+1)}(x, A)$

$$\begin{aligned} h^{(n+1)}(x, A) &\leq (1 - \epsilon)r(A^c)h^{(n)}(x, \Omega_n) - (1 - \epsilon)r(A)h^{(n)}(x, \Omega_n^c) \\ &= (1 - \epsilon)r(A^c)h^{(n)}(x, \Omega_n) + (1 - \epsilon)r(A)h^{(n)}(x, \Omega_n) \\ &= (1 - \epsilon)h^{(n)}(x, \Omega_n) \leq (1 - \epsilon)^{n+1} \end{aligned}$$

which involves the inequalities obtained for  $h(y, A)$  (first line), the fact that  $h^{(n)}(x, \cdot)$  has a zero integral (second line), and the induction hypothesis as well as the positivity of  $h^{(n)}(x, \Omega_n)$  (third line). Similarly,  $h^{(n+1)}(x, A)$  can be shown to be bounded from below

$$\begin{aligned}
h^{(n+1)}(x, A) &\geq -(1 - \epsilon)r(A)h^{(n)}(x, \Omega_n) + (1 - \epsilon)r(A^c)h^{(n)}(x, \Omega_n^c) \\
&= -(1 - \epsilon)r(A)h^{(n)}(x, \Omega_n) - (1 - \epsilon)r(A^c)h^{(n)}(x, \Omega_n) \\
&= -(1 - \epsilon)h^{(n)}(x, \Omega_n) \geq -(1 - \epsilon)^{n+1}
\end{aligned}$$

and finally

$$|h^{(n+1)}(x, A)| \leq (1 - \epsilon)^{n+1}$$

which is the desired result.  $\square$

## Exercises

**9.1** Let  $P$  be a irreducible, aperiodic and  $p$ -invariant transition kernel on  $(\Omega, \mathcal{A})$ . Suppose that a iterate  $P^{(n_0)}$  of  $P$  satisfies the minorization condition of definition 9.2.1. Show that  $P$  is uniformly ergodic.

**9.2** Let  $p$  be a distribution on  $\mathbb{N}$  that is the limit distribution of a Markov chain with Jacobi transition kernel  $P$ . For each  $n > 0$ , we put  $\mathbb{N}_n = \{n, n + 1, \dots\}$ . Let  $P_n$  be the transition kernel on  $\mathbb{N}_n \times \mathbb{N}_n$  defined by  $P_n(x, y) = P(x, y)$  for all  $x, y \in \mathbb{N}_n$  except  $P_n(n, n) = P(n, n) + P(n, n - 1)$ . Show that  $P_n$  is irreducible and aperiodic. What is its invariant distribution?

## 10. Exact simulations

The previous chapter highlighted the difficulties in determining the rate of convergence of an iterative Markov algorithm quantitatively. This is very frustrating as the only problem is to generate one state from the target distribution: then it suffices to run the transition kernel from that state to produce as many as desired.

A decisive breakthrough was made by Propp and Wilson in 1996. They demonstrated that knowing only the transition kernel it is possible to simulate the target distribution on a finite state-space. To achieve this goal, they use a technique called *coupling from the past*. This chapter describes it in terms of *backwards iterations* as in Diaconis and Freedman (1999).

Since the publication of Propp and Wilson's seminal paper, considerable literature has been devoted to exact simulations. New algorithms have come out as in Fill (1998). Exact simulations are now possible in infinite state-spaces (Murdoch and Green, 1998; Diaconis and Freedman, 1999; Mira et al., 1999; Fill et al., 2000). Stochastic geometry constitutes one of the most promising fields of applications where original algorithms are being set up for simulating point processes (Häggström et al., 1999; Kendall and Møller, 1999) or object based models (Kendall and Thönnies, 1999). A comprehensive review of exact simulation in stochastic geometry is Møller (2000).

### 10.1 Propp and Wilson's algorithm

#### 10.1.1 Principles

Let  $\Omega$  be a state-space, supposed here to be finite (unless stated explicitly). Let  $p$  be a distribution on  $\Omega$  such that  $p(x) > 0$  for any  $x \in \Omega$ . Suppose that  $p$  is the limit of an irreducible and aperiodic Markov chain with transition kernel  $P$

$$\lim_{n \rightarrow \infty} P^{(n)}(x, y) = p(y) \quad x, y \in \Omega$$

Let  $\Omega^\Omega$  be the set of all mappings from  $\Omega$  to  $\Omega$ . A random mapping  $F$  is a random element of  $\Omega^\Omega$ . Its distribution is given by a probability measure  $\mu$  on  $\Omega^\Omega$

$$\mu(f) = P\{F = f\} \quad f \in \Omega^\Omega$$

**Definition 10.1.1.** *The random mapping  $F$  generates the transition kernel  $P$  if*

$$P\{F(x) = y\} = P(x, y) \quad x, y \in \Omega$$

**Remark 10.1.1.** Given a transition kernel  $P$ , there always exists a random mapping that generates  $P$ . Take for instance

$$\mu(f) = \prod_{x \in \Omega} P(x, f(x)) \quad f \in \Omega^\Omega$$

The proof is the topic of exercise 10.1.

Let  $(F_n, n \geq 1)$  be a sequence of i.i.d. random mappings that generate  $P$ . Instead of composing the mappings in direct order  $(F_n \circ \dots \circ F_1)$ , we compose them in reverse order  $(F_1 \circ \dots \circ F_n)$ . Because the random mappings are independent, we have

$$P\{F_1 \circ \dots \circ F_n(x) = y\} = P^{(n)}(x, y) \quad x, y \in \Omega$$

Let us also put

$$\Omega_n = F_1 \circ \dots \circ F_n(\Omega) \quad n \geq 1$$

The  $\Omega_n$ 's constitute a decreasing sequence of nonempty random sets:

$$\Omega_{n+1} = F_1 \circ \dots \circ F_{n+1}(\Omega) = F_1 \circ \dots \circ F_n [F_{n+1}(\Omega)] \subset F_1 \circ \dots \circ F_n(\Omega) = \Omega_n$$

Because  $\Omega$  is finite, their intersection  $\Omega_\infty = \bigcap_{n \geq 1} \Omega_n$  cannot be empty. Moreover we have

**Theorem 10.1.1.** *If  $\#\Omega_\infty = 1$  a.s., then  $\Omega_\infty \sim p$ .*

**Proof:** Let  $y \in \Omega$ . We always have

$$P\{y \in \Omega_\infty\} = P\left\{\bigwedge_{n \geq 1} y \in \Omega_n\right\} = \lim_{n \rightarrow \infty} P\{y \in \Omega_n\},$$

and also

$$P\{y \in \Omega_n\} = P\{y \in F_1 \circ \dots \circ F_n(\Omega)\} \geq P\{y = F_1 \circ \dots \circ F_n(y)\} = P^{(n)}(y, y)$$

Therefore

$$P\{y \in \Omega_\infty\} \geq \lim_{n \rightarrow \infty} P^{(n)}(y, y) = p(y)$$

Now suppose that  $\#\Omega_\infty = 1$  holds almost surely. Then it follows  $P\{\Omega_\infty = \{y\}\} \geq p(y)$  for any  $y$ , as well as  $\sum_{y \in \Omega} P\{\Omega_\infty = \{y\}\} = 1$ . This is possible only if  $P\{\Omega_\infty = \{y\}\} = p(y)$  for any  $y \in \Omega$ .  $\square$

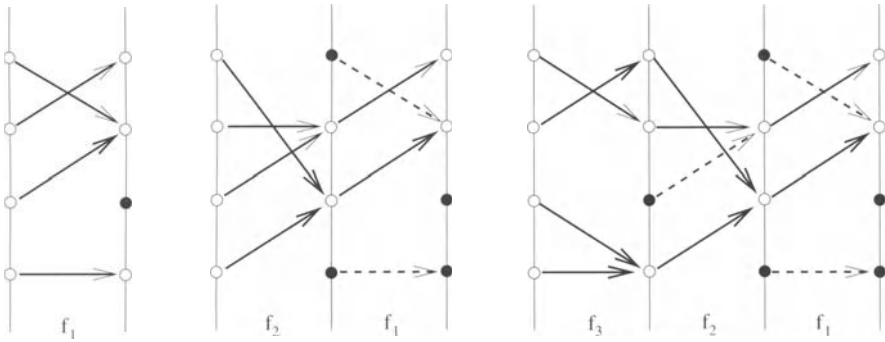
Because of theorem 10.1.1 the following algorithm works provided that  $\Omega_\infty$  is almost surely a singleton. In the initialization step  $Id$  stands for the identity mapping on  $\Omega$ .



**Algorithm 10.1.1.** (*Propp and Wilson*)

1. Set  $g = Id$ .
2. Generate  $f \sim \mu$  and put  $g = g \circ f$ .
3. If  $\#\mathcal{g}(\Omega) > 1$ , then goto 2.
4. Deliver  $g(\Omega)$ .

In order to see how this algorithm works in practice, we consider a state-space  $\Omega$  with 4 states. Figure 10.1 shows realizations of the first 3 random mappings and their composition.



**Fig. 10.1.** Running Propp and Wilson's algorithm for 3 iterations. Convergence has not occurred yet since  $\#\Omega_3 = 2$  (the two states of  $\Omega_3$  are depicted as white disks)

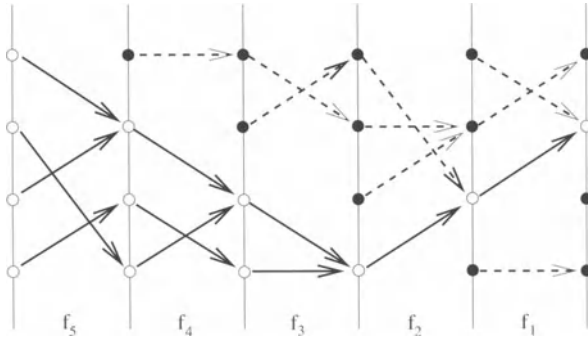
Assuming that a single state has been obtained after 5 iterations, Figure 10.2 shows a complete run of Propp and Wilson's algorithm. A closer look at the figure shows that all trajectories have already "coalesced" after running  $f_3$ . One might think that the state thus obtained can be regarded as a realization of the target distribution. However this is generally not the case because the distribution of the first coalescence state may be biased. A simple counterexample is proposed in exercise 10.2.

Still in the case when  $\Omega_\infty$  is almost surely a singleton, Propp and Wilson introduce the (random) number of iterations for coalescence

$$N = \min\{n \geq 1 : \#\Omega_n = 1\}$$

**Proposition 10.1.1.** *The complementary distribution of  $N$  is submultiplicative.*

$$P\{N > m + n\} \leq P\{N > m\} P\{N > n\}$$



**Fig. 10.2.** All trajectories have coalesced after 5 iterations. The common state (represented as a white disk) can be regarded as a realization of the target distribution

**Proof:** For each  $m, n \geq 1$  consider  $\Omega_{m,m+n} = F_{m+1} \circ \dots \circ F_{m+n}(\Omega)$ . Note that  $N \leq m+n$  is ensured if  $\Omega_m$  or  $\Omega_{m,m+n}$  is a singleton. Accordingly,  $N > m+n$  implies  $\#\Omega_m > 1$  and  $\#\Omega_{m,m+n} > 1$ . The result follows because  $\#\Omega_m > 1$  and  $\#\Omega_{m,m+n} > 1$  are independent events, and because  $P\{\#\Omega_{m,m+n} > 1\} = P\{\#\Omega_n > 1\} = P\{N > n\}$ .  $\square$

Other results related to the mean coalescence time can be found in Propp and Wilson (1996).

### 10.1.2 Condition for a singleton

The following example shows that  $\Omega_\infty$  is not always a singleton.

**Example 10.1.1.** Take  $\Omega = \{x, y\}$  and consider the random mapping that swaps  $x$  and  $y$  or keeps them unchanged with the same probability 0.5. The associated transition kernel is

$$P = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

$P$  is clearly irreducible, aperiodic and its invariant distribution is uniform. However  $\Omega_n$  cannot converge to a singleton because both states  $x$  and  $y$  are in the range of each realization of the random mapping.

Consequently, the fact that the random mapping generates  $P$  is not sufficient for  $\Omega_\infty$  to be a singleton. Other assumptions are required. This problem is now addressed.

The first step is to extend  $P$  to a transition kernel on the family  $\mathcal{P}^*(\Omega)$  of non-empty subsets of  $\Omega$  using the formula

$$P_F(A, B) = P\{F(A) = B\} \quad A, B \in \mathcal{P}^*(\Omega)$$

This new transition kernel is written with a subscript as it depends on the random mapping  $F$ . Of course  $P_F(\{x\}, \{y\}) = P(x, y)$ . Observe also that  $P(A, B)$  can be positive only if  $\#A \geq \#B$  because  $\#f(A) \leq \#A$  for each  $f \in \Omega^\Omega$  and each  $A \in \Omega$ . The iterated transition kernel  $P_F^{(n)}$  satisfies

$$P_F^{(n)}(A, B) = P\{F_1 \circ \dots \circ F_n(A) = B\}$$

Recall that a state  $A \in \mathcal{P}^*(\Omega)$  is said to be *recurrent* if a Markov chain with initial state  $A$  returns to  $A$  almost surely. Otherwise,  $A$  is said to be *transient*. The following characterization of transient or recurrent states is standard (e.g. Feller, 1968):

**Theorem 10.1.2.**  *$A$  is recurrent or transient depending on whether the series with general term  $P_F^{(n)}(A, A)$  diverges or converges. In the transient case, the series with general term  $P_F^{(n)}(B, A)$  is convergent whatever the state  $B$ .*

**Lemma 10.1.1.** *A recurrent state communicates only with states with the same cardinality.*

**Proof:** Suppose that  $A$  is recurrent. If  $B$  is accessible from  $A$ , then  $A$  is also accessible from  $B$  (if not,  $A$  would be transient). The existence of a path from  $A$  to  $B$  implies  $\#A \geq \#B$ . Similarly, the existence of a path from  $B$  to  $A$  implies  $\#B \geq \#A$ . Therefore  $\#A = \#B$ .  $\square$

The following lemma shows how the recurrent and transient states of  $\mathcal{P}^*(\Omega)$  are arranged.

**Lemma 10.1.2.** *If  $A$  is recurrent, any  $B \subset A$  is recurrent. If  $A$  is transient, any  $B \supset A$  is transient.*

**Proof:** Because both parts of the lemma are equivalent, only the first part needs to be established. Let  $A$  and  $B$  two non-empty subsets of  $\Omega$  with  $B \subset A$ . Denote by  $\mathcal{C}_B$  the family of all subsets of  $A$  with the same cardinality as  $B$ . If  $F_1 \circ \dots \circ F_n(A) = A$ , then there exists  $C \in \mathcal{C}_B$  such that  $F_1 \circ \dots \circ F_n(C) = B$ . Accordingly

$$P_F^{(n)}(A, A) \leq \sum_{C \in \mathcal{C}_B} P_F^{(n)}(C, B)$$

Suppose now that  $A$  is recurrent. Then we have by theorem 10.1.2

$$\infty = \sum_{n=0}^{\infty} P_F^{(n)}(A, A) \leq \sum_{n=0}^{\infty} \sum_{C \in \mathcal{C}_B} P_F^{(n)}(C, B).$$

This implies the existence of  $C \in \mathcal{C}_B$  such that

$$\sum_{n=0}^{\infty} P_F^{(n)}(C, B) = \infty,$$

and this formula is sufficient to ensure that  $B$  is recurrent (theorem 10.1.2).  
 $\square$

**Theorem 10.1.3.**  $\# \Omega_{\infty} = 1$  a.s. if and only if the only recurrent states of  $\mathcal{P}^*(\Omega)$  are the singletons.

**Proof:** Let  $A$  be a non-empty subset of  $\Omega$ . For each  $n \geq 1$  we have  $F_1 \circ \dots \circ F_n(A) \subset \Omega_n$ . If  $A$  is recurrent, then lemma 10.1.1 implies  $\#F_1 \circ \dots \circ F_n(A) = \#A$  a.s., so  $\#A \leq \# \Omega_n$  a.s. Consequently  $\#A \leq \min_{n \geq 1} \# \Omega_n = \# \Omega_{\infty}$ . Now  $\# \Omega_{\infty} = 1$  a.s. gives  $\#A = 1$  at once.

Conversely, suppose that the only recurrent states are the singletons. We are going to prove that  $P\{\# \Omega_{\infty} = 1\} = 1$ , or equivalently, that  $P\{\Omega_{\infty} = A\} = 0$  for any  $A \subset \Omega$  such that  $\#A > 1$ . Given the definition of  $\Omega_{\infty}$  and using Fatou's lemma<sup>1</sup>, we can write

$$P\{\Omega_{\infty} = A\} = P\{\liminf(\Omega_n = A)\} \leq \liminf P\{\Omega_n = A\}$$

But  $\liminf P\{\Omega_n = A\} = 0$  because  $A$  is transient: by application of theorem 10.1.2, the series of general term  $P_F^{(n)}(\Omega, A) = P\{\Omega_n = A\}$  is convergent.  
 $\square$

Let  $\mathcal{M}$  be the set of all non-negligible realizations of  $F$  ( $f \in \mathcal{M}$  if and only if  $\mu(f) > 0$ ), and let  $\mathcal{M}_c$  be the set of all finite compositions of elements of  $\mathcal{M}$ :  $f \in \mathcal{M}_c$  if and only there exist  $n \geq 1$  and  $f_1, \dots, f_n \in \mathcal{M}$  such that  $f = f_n \circ \dots \circ f_1$ . In particular  $\mathcal{M} \subset \mathcal{M}_c$ .

**Theorem 10.1.4.** *The following four assertions are equivalent:*

- (1)  $\# \Omega_{\infty} = 1$  a.s.
- (2) For all  $x, y \in \Omega$  there exists  $f \in \mathcal{M}_c$  such that  $f(x) = f(y)$ .
- (3) There exists a constant mapping  $f \in \mathcal{M}_c$ .
- (4) For each  $z \in \Omega$ , there exists  $f \in \mathcal{M}_c$  such that  $f(\Omega) = \{z\}$ .

**Proof:** The implications (4)  $\implies$  (3)  $\implies$  (2) are trivial. The proof is completed by showing (2)  $\implies$  (1) and (1)  $\implies$  (4).

Suppose that assertion (2) holds. Let  $x, y \in \Omega$  with  $x \neq y$ . There exists  $f \in \mathcal{M}_c$  such that  $f(x) = f(y)$ . We put  $z = f(x) = f(y)$ . The mapping  $f$  is a path from  $\{x, y\}$  to  $z$ , but there exists no return path from  $z$  to  $\{x, y\}$ . So the state  $\{x, y\}$  is transient. This shows that each state of  $\mathcal{P}^*(\Omega)$  with

---

<sup>1</sup> The inequality  $P\{\liminf A_n\} \leq \liminf P\{A_n\}$  holds for any sequence of events  $(A_n, n \in \mathbb{N})$ .

cardinality 2 is transient, from which we derive by lemma 10.1.2 that each state with cardinality  $> 1$  is transient. Therefore the only recurrent states are singletons, which is equivalent to  $\#\Omega_\infty = 1$  a.s. by theorem 10.1.3. This proves (2)  $\implies$  (1).

It remains to prove (1)  $\implies$  (4). Suppose that  $P\{\#\Omega_\infty = 1\} = 1$ . According to theorem 10.1.1 we know that  $P\{\Omega_\infty = \{z\}\} = p(z) > 0$  for all  $z \in \Omega$ . By Fatou's lemma, we have  $P\{\Omega_\infty = \{z\}\} \leq \liminf P\{\Omega_n = \{z\}\}$ , from which it follows that  $P\{\Omega_n = \{z\}\} > 0$  provided that  $n$  is large enough. But  $\Omega_n = F_1 \circ \dots \circ F_n(\Omega)$ , so there exist realizations  $f_1, \dots, f_n$  of  $F_1, \dots, F_n$  such that  $f_1 \circ \dots \circ f_n(\Omega) = \{z\}$ . Assertion (4) is thus established with  $f = f_1 \circ \dots \circ f_n$ .  $\square$

### 10.1.3 Monotonic Monte Carlo

We now come to a practical point. The original version of Propp and Wilson's algorithm becomes unpracticable when the number of states is very large<sup>2</sup>. Fortunately, algorithmic simplifications may occur in the case when the random mapping is increasing w.r.t. a partial order on the state-space. The resulting algorithm proves to be sufficiently efficient to make the simulation of Ising's model possible on a  $4200 \times 4200$  grid (the number of states of this model is  $2^{4200 \times 4200}$ ).

Suppose that  $\Omega$  is endowed with a *partial order* denoted by  $\preceq$ . (The pair  $(\Omega, \preceq)$  is often called a *poset*). Recall that a partial order relation is reflexive ( $x \preceq x$ ), antisymmetric (if  $x \preceq y$  and  $y \preceq x$ , then  $x = y$ ) and transitive (if  $x \preceq y$  and  $y \preceq z$ , then  $x \preceq z$ ). The relation  $x \preceq y$  can also be written  $y \succeq x$ , which amounts to saying that  $x$  is a predecessor of  $y$  or that  $y$  is a successor of  $x$ . If  $x$  is neither a predecessor nor a successor of  $y$ , then  $x$  and  $y$  are said to be incomparable.

**Definition 10.1.2.** *The mapping  $f$  from  $\Omega$  into  $\Omega$  is said to be increasing (in short  $f \nearrow$ ) if  $f(x) \preceq f(y)$  whenever  $x \preceq y$ .*

Increasing mappings can be characterized simply using the up-sets and the down-sets of  $\Omega$ :

**Definition 10.1.3.** *A subset  $X$  of  $\Omega$  is an up-set if it contains all successors of its elements ( $x \in X$  and  $y \succeq x$  implies  $y \in X$ ). Similarly  $X$  is a down-set if it contains all predecessors of its elements ( $x \in X$  and  $y \preceq x$  implies  $y \in X$ ).*

---

<sup>2</sup> Another related question is to know whether Propp and Wilson's algorithm can work when the state-space is infinite. Diaconis and Freedman (1999) exhibit one circumstance where this is possible: the space-state is equipped with a metric that makes the realizations of the random mapping lipschitzian.

Let  $\mathcal{U}$  and  $\mathcal{D}$  be the families of all up-sets and all down-sets of  $\Omega$ . Clearly  $\mathcal{U}$  and  $\mathcal{D}$  are stable under union and intersection. They contain the empty set  $\emptyset$  and  $\Omega$ . Moreover up-sets and down-sets are complementary<sup>3</sup>. The following proposition shows that up-sets and down-sets play the same role for increasing mappings as open sets and closed sets for continuous mappings.

**Proposition 10.1.2.** *The mapping  $f$  is increasing if and only if the inverse image of each up-set is an up-set, or if and only if the inverse image of each down-set is a down-set*

$$\forall U \in \mathcal{U} \quad f^{-1}(U) \in \mathcal{U} \iff f \nearrow \iff \forall D \in \mathcal{D} \quad f^{-1}(D) \in \mathcal{D}$$

**Proof:** Because up-sets and down-sets are complementary, it suffices to establish the proposition for up-sets only.

Suppose that  $f$  is increasing. We want to show that  $f^{-1}(U) \in \mathcal{U}$  for each  $U \in \mathcal{U}$ . Let  $x, y \in \Omega$  with  $x \in f^{-1}(U)$  and  $y \succeq x$ . Then we have  $f(x) \in U$  and  $f(y) \succeq f(x)$  ( $f \nearrow$ ). Since  $U$  is an up-set, it follows  $f(y) \in U$  or equivalently  $y \in f^{-1}(U)$ . Therefore  $f^{-1}(U)$  is an up-set.

Conversely, suppose that the inverse image by  $f$  of each up-set is an up-set. Let  $x, y \in \Omega$  such that  $x \preceq y$ , and consider the subset  $M_{f(x)} = \{z \in \Omega : z \succeq f(x)\}$ . Clearly  $M_{f(x)} \in \mathcal{U}$ , so  $f^{-1}(M_{f(x)}) \in \mathcal{U}$ . But  $x \in f^{-1}(M_{f(x)})$  and  $y \succeq x$ , so  $y \in f^{-1}(M_{f(x)})$ . Therefore  $f(y) \in M_{f(x)}$ , that is  $f(y) \succeq f(x)$ .  $\square$

Assume that the transition kernel  $P$  can be generated by a random mapping  $F$  that has all its realizations increasing. Assume further that  $\Omega$  admits a minimal state  $\hat{0}$  and a maximal state  $\hat{1}$  such that  $\hat{0} \preceq x \preceq \hat{1}$  for any  $x \in \Omega$ . Let  $(F_n, n \geq 1)$  be a sequence of independent copies of  $F$ . Because the  $F_n$ 's are increasing, we can write

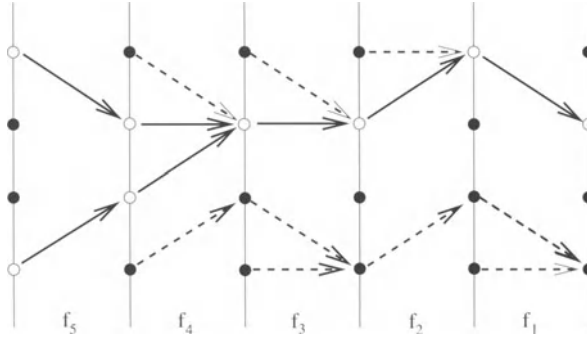
$$F_1 \circ \cdots \circ F_n(\hat{0}) \preceq F_1 \circ \cdots \circ F_n(x) \preceq F_1 \circ \cdots \circ F_n(\hat{1})$$

so that  $F_1 \circ \cdots \circ F_n(\hat{0}) = F_1 \circ \cdots \circ F_n(\hat{1}) = z$  implies  $F_1 \circ \cdots \circ F_n(x) = z$  for each  $x \in \Omega$ . We end up with the following algorithm where only the trajectories from  $\hat{0}$  and  $\hat{1}$  are considered:

**Algorithm 10.1.2.** (*Propp and Wilson for increasing mappings*)

1. Set  $g = \text{Id}$ .
2. Generate  $f \sim \mu$  and put  $g = g \circ f$ .
3. If  $g(\hat{0}) \neq g(\hat{1})$ , then goto 2.
4. Deliver  $g(\Omega)$ .

<sup>3</sup> Let  $U \in \mathcal{U}$ . Suppose that  $x \in \Omega \setminus U$  and let  $y \preceq x$ . If  $y \in U$ , this would imply  $x \in U$ , which is a contradiction. Therefore  $y \in \Omega \setminus U$  and  $\Omega \setminus U$  is a down-set.



**Fig. 10.3.** Exact simulation using Propp and Wilson's algorithm in the case of a poset with extremal elements. If the transition kernel is generated by an increasing random mapping, then the trajectories cannot intersect. The coalescence of the trajectories from the extremal elements guarantees that of all trajectories

Since  $P\{F_1 \circ \dots \circ F_n(\hat{0}) = \hat{1}\} = P^{(n)}(\hat{0}, \hat{1}) \xrightarrow{n \rightarrow \infty} p(\hat{1}) > 0$ , there exists  $f = f_1 \circ \dots \circ f_n \in \mathcal{M}_c$  such that  $f(\hat{0}) = \hat{1}$ . And because  $f$  is increasing, we have  $f(x) = \hat{1}$  for all  $x \in \Omega$ . It follows that  $f$  is a constant mapping, which is sufficient to ensure that  $\Omega_\infty$  is almost surely a singleton (theorem 10.1.4). The random number of iterations  $N$  required for coalescence is bounded above by  $\min(N_{0,1}, N_{1,0})$ , where  $N_{0,1}$  (resp.  $N_{1,0}$ ) is the number of iterations to jump from  $\hat{0}$  to  $\hat{1}$  (resp. from  $\hat{1}$  to  $\hat{0}$ ).

What are the properties of a transition kernel generated by an increasing random mapping?

**Proposition 10.1.3.** *If  $P$  is generated by an increasing random mapping, then  $P$  is stochastically monotonic in the sense that  $P(x, U) \leq P(y, U)$  for any up-set  $U$  and any  $x \preceq y$ . Equivalently  $P(x, D) \geq P(y, D)$  for any down-set  $D$  and any  $x \preceq y$ .*

**Proof:** Suppose that the increasing random mapping  $F$  generates  $P$ . Let  $U \in \mathcal{U}$  and  $x \preceq y$ . Then  $F(x) \in U$  implies  $F(y) \in U$ , so

$$P(x, U) = P\{F(x) \in U\} \leq P\{F(y) \in U\} = P(y, U),$$

which establishes the stochastic monotonicity for up-sets. Now let  $D \in \mathcal{D}$ . Then  $\Omega \setminus D$  is an up-set, so

$$P(x, D) = 1 - P(x, \Omega \setminus D) \geq 1 - P(y, \Omega \setminus D) = P(y, D)$$

which completes the proof.  $\square$

**Remark 10.1.2.** It should be pointed out that a random mapping that generates a stochastically monotonic transition kernel is not necessarily increasing. More surprisingly, there exist monotonic transition kernels that are

generated by none increasing random mapping Machida (1999). See exercise 10.5.

**Example 10.1.2.** We are concerned with the exact simulation of the binomial distribution with index  $k = 10$  and parameter  $r = 0.7$

$$p(x) = \binom{k}{x} r^x (1-r)^{k-x} \quad 0 \leq x \leq k$$

A transition kernel for  $p$  can be obtained using Metropolis algorithm (Barker variation). Explicitly, we find

$$\begin{aligned} P(x, x-1) &= \frac{(1-r)x}{r(k-x+1) + (1-r)x} \\ P(x, x+1) &= \frac{r(k-x)}{r(k-x) + (1-r)(x+1)} \end{aligned}$$

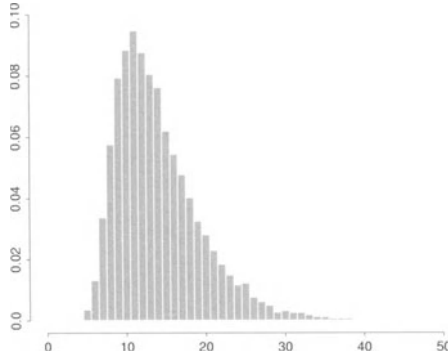
and  $P(x, x) = 1 - P(x, x-1) - P(x, x+1)$ . Because  $P$  is a tridiagonal matrix, it is not difficult to devise a random mapping that generates  $P$ . Take for instance

$$F(x) = \begin{cases} x-1 & \text{if } U < P(x, x-1) \\ x & \text{if } P(x, x-1) \leq U < 1 - P(x, x+1) \\ x+1 & \text{if } U \geq 1 - P(x, x+1) \end{cases}$$

where  $U$  is a uniform variable on  $]0, 1[$ . Note that  $F(x) = x+1$  and  $F(x+1) = x$  cannot take place for a given  $U$ . In such a case, we would have  $1 - P(x, x+1) \leq U < P(x+1, x)$ , which is not possible because the entries of  $P$  satisfy  $P(x, x+1) + P(x+1, x) = 1$ . Consequently,  $F$  is increasing and  $P$  is stochastically monotonic.

The random mapping  $F$  has been used in algorithm 10.1.2 to produce a set of 10000 exact simulations. A first observation is that the statistical properties of the binomial distribution have been properly reproduced. Special attention has also been paid to the number of iterations for coalescence. The values on the histogram on Figure 10.4 range from 5 to 53. Their mean is equal to 14.123 iterations, which is not much larger than the integral range (8.689) associated with the second greatest eigenvalue of the transition kernel (see section 9.4). Their standard deviation is 5.472 iterations, which gives a coefficient of variation of 0.387. It should be also mentioned that the generated value is usually not independent of the number of iterations to produce it. For instance, there is only one state (5) that can be generated in 5 iterations exactly. Note however that the correlation coefficient between the value generated and the number of iterations is quite small (0.07).





**Fig. 10.4.** Histogram of the number of iterations to simulate exactly a binomial distribution with parameter 0.7 and index 10 using Propp and Wilson's algorithm for monotonic Markov chains

### 10.2 Fill's algorithm

In the case where the state-space  $\Omega$  is partially ordered with extremal states, Fill's algorithm can be regarded as an alternative to Propp and Wilson's algorithm.

Let  $\tilde{P}$  be the matrix defined by

$$p(x) \tilde{P}(x, y) = p(y) P(y, x) \quad x, y \in \Omega$$

$\tilde{P}$  are clearly non-negative entries. Moreover, all the entries of each row add up to 1 as a consequence of the invariance of  $p$  under  $P$ :

$$\sum_{y \in \Omega} \tilde{P}(x, y) = \frac{1}{p(x)} \sum_{y \in \Omega} p(y) P(y, x) = \frac{1}{p(x)} p(x) = 1$$

Accordingly  $\tilde{P}$  is a transition kernel. It is called the *time-reversal* transition kernel associated with  $P$ . In the case when  $P$  is reversible, then  $\tilde{P} = P$ . A simple calculation shows that the time-reversal transition kernel associated with  $P^{(n)}$  is  $\tilde{P}^{(n)}$  for each  $n > 0$ .

From now on, we assume that  $\tilde{P}$  can be generated by an increasing mapping  $\tilde{F}$ . In that case,  $\tilde{P}^{(n)}$  is stochastically monotonic (see proposition 10.1.3), and we have  $\tilde{P}^{(n)}(\hat{1}, \hat{0}) \leq \tilde{P}^{(n)}(x, \hat{0})$  for each  $x \in \Omega$  because  $\{\hat{0}\}$  is a down-set. Accordingly

$$p(x) \tilde{P}^{(n)}(\hat{1}, \hat{0}) \leq p(x) \tilde{P}^{(n)}(x, \hat{0}) = p(\hat{0}) P^{(n)}(\hat{0}, x)$$

Therefore we can write (provided that  $n$  is large enough)

$$p(x) \leq P^{(n)}(\hat{0}, x) \frac{p(\hat{0})}{\tilde{P}^{(n)}(\hat{1}, \hat{0})} \quad x \in \Omega$$

This gives a way to simulate  $p$  using the acceptance-rejection method. At first a candidate state  $x$  is generated w.r.t. the distribution  $P^{(n)}(\hat{0}, \cdot)$ . Then it is accepted with probability

$$\alpha_n(x) = \frac{p(x)}{P^{(n)}(\hat{0}, x)} \frac{\tilde{P}^{(n)}(\hat{1}, \hat{0})}{p(\hat{0})} = \frac{\tilde{P}^{(n)}(\hat{1}, \hat{0})}{\tilde{P}^{(n)}(x, \hat{0})}$$

A candidate state is generated very simply by running the transition kernel  $P$   $n$  times. The following lemma is useful to simulate an event with probability  $\alpha_n(x)$ :

**Lemma 10.2.1.**

$$\alpha_n(x) = P\{\tilde{F}_1 \circ \dots \circ \tilde{F}_n(\hat{1}) = \hat{0} \mid \tilde{F}_1 \circ \dots \circ \tilde{F}_n(x) = \hat{0}\}$$

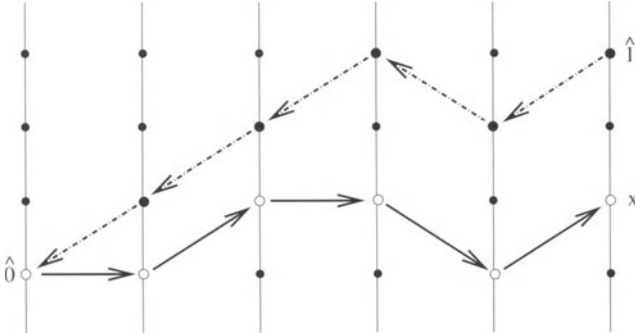
**Proof:** Because  $\tilde{F}_1 \circ \dots \circ \tilde{F}_n$  is increasing, the event  $\tilde{F}_1 \circ \dots \circ \tilde{F}_n(\hat{1}) = \hat{0}$  implies  $\tilde{F}_1 \circ \dots \circ \tilde{F}_n(x) = \hat{0}$ . Accordingly

$$\begin{aligned} \alpha_n(x) &= \frac{P\{\tilde{F}_1 \circ \dots \circ \tilde{F}_n(\hat{1}) = \hat{0} \wedge \tilde{F}_1 \circ \dots \circ \tilde{F}_n(x) = \hat{0}\}}{P\{\tilde{F}_1 \circ \dots \circ \tilde{F}_n(x) = \hat{0}\}} \\ &= \frac{P\{\tilde{F}_1 \circ \dots \circ \tilde{F}_n(\hat{1}) = \hat{0}\}}{P\{\tilde{F}_1 \circ \dots \circ \tilde{F}_n(x) = \hat{0}\}} \\ &= \frac{\tilde{P}^{(n)}(\hat{1}, \hat{0})}{\tilde{P}^{(n)}(x, \hat{0})} \quad \square \end{aligned}$$

Using lemma 10.2.1, it becomes easy to simulate an event of probability  $\alpha_n(z)$  because a trajectory from  $x$  to  $\hat{0}$  is already available. This is the one that has been produced from  $\hat{0}$  to  $x$  during the generation of the candidate state. We finally end up with the following algorithm<sup>4</sup>:

**Algorithm 10.2.1.** (Fill)

1. Set  $n = 1$ .
2. Put  $n = n + 1$ .
3. Generate  $x \sim P^{(n)}(\hat{0}, \cdot)$ . Save the states  $x_0 = \hat{0}, x_1, \dots, x_n = x$  generated between  $0$  and  $n$ .
4. Put  $y_n = \hat{1}$ . For  $i = n, n - 1, \dots, 1$  generate  $\tilde{f}_i \sim \mu \mid \tilde{f}_i(x_i) = x_{i-1}$  and put  $y_{i-1} = \tilde{f}_i(y_i)$ .
5. If  $y_0 \neq \hat{0}$ , then goto 2.
6. Deliver  $x$ .



**Fig. 10.5.** Exact simulation using Fill's algorithm. A candidate state  $x$  is generated by running the transition kernel  $P$ . Given the trajectory produced, a new trajectory is generated from  $\hat{1}$  using the time-reversal transition kernel  $\tilde{P}$ . The state  $x$  is a simulation from  $p$  if the new trajectory terminates at  $\hat{0}$

Figure 10.5 illustrates how this algorithm works.

This algorithm stops at the  $n^{th}$  iteration with probability

$$\sum_{x \in \Omega} P^{(n)}(\hat{0}, x) \alpha_n(x) = \frac{\tilde{P}^{(n)}(\hat{1}, \hat{0})}{p(\hat{0})}$$

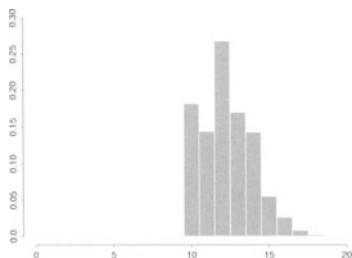
Let  $(X_i, i \geq 0)$  be the Markov chain with transition kernel  $P$ . A slightly more elaborate calculation (Møller and Schladitz, 1998) shows that

$$P\{X_n = x \wedge \tilde{F}_1 \circ \dots \circ \tilde{F}_n(\hat{1}) = \hat{0} \mid \wedge_{i=1}^n \tilde{F}_i(X_i) = X_{i-1}\} = p(x) \frac{\tilde{P}^{(n)}(\hat{1}, \hat{0})}{p(\hat{0})}$$

This formula is interesting because the right-hand side member is the product of the distribution of  $x$  and that of  $n$ . As a consequence, the state generated is independent of the number of iterations to get it. For that reason Fill's algorithm is said to be *interruptible* (Fill, 1998; Thönnies, 1998).

**Example 10.2.1.** Algorithm 10.2.1 has been used to carry out 10000 simulations of the binomial distribution of example 10.1.2. The histogram of the number of iterations obtained is reproduced on Figure 10.6. This simulation exercise shows that Fill's algorithm performs better than Propp and Wilson's algorithm in the case of the distribution considered. The mean number of iterations is slightly smaller (12.281 instead of 14.123). A bigger difference is observed in the standard deviations (1.667 instead of 5.472). The histogram produced by Fill's algorithm is much tighter. Its values range between 10 and

<sup>4</sup> Step 2 of Fill's algorithm is usually written as "Put  $n = 2n$ ". Here the step "Put  $n = n + 1$ " has been considered in order to facilitate the comparison of the performances between both algorithms.



**Fig. 10.6.** Histogram of the number of iterations to simulate exactly a binomial distribution with parameter 0.7 and index 10 using Fill's algorithm

19. The independence between the value generated and the number of iterations has been confirmed experimentally. Their correlation coefficient was found equal to  $-0.012$ .

### Exercises

**10.1** The aim of this exercise is to construct a random mapping that generates the transition kernel  $P$ .

1) Let  $\Xi$  be a non-empty subset of  $\Omega$ . Prove the "multinomial identity"

$$\sum_{f \in \Omega^\Xi} \prod_{x \in \Omega} P(x, f(x)) = \prod_{x \in \Xi} \sum_{y \in \Omega} P(x, y) = 1$$

2) Apply this identity to show that the random mapping  $F$  with distribution

$$\mu(f) = \prod_{x \in \Omega} P(x, f(x)) \quad f \in \Omega^\Omega$$

generates  $P$ .

3) Show that the coordinates of  $F$  are independent.

4) Show that  $\#\Omega_\infty = 1$  almost surely.

**10.2** When running the Propp and Wilson's algorithm, the first coalescence state may be privileged. This exercise illustrates this problem

Suppose  $\Omega = \{x, y\}$ , and consider the random mapping that either interchanges  $x$  and  $y$  with probability  $p$  or sends all states to  $x$  with the complementary probability  $q = 1 - p$  ( $p, q > 0$ ).

1) Determine the transition kernel and the limit distribution.

2) Show that  $x$  is always the first coalescence state.

**10.3** Let  $P$  be a transition kernel on the finite state-space  $\Omega$  with target distribution  $p$ . Assume that  $P$  satisfies the minorization condition

$$P(x, y) \geq \epsilon q(y) \quad x, y \in \Omega$$

for some probability measure  $q$  on  $\Omega$  and some scalar  $0 < \epsilon < 1$ .

1) Show that  $P$  can be written as a mixture of two transition kernels

$$P(x, y) = \epsilon Q(x, y) + (1 - \epsilon)R(x, y) \quad x, y \in \Omega$$

with  $Q(x, y)$  independent of  $x$ .

2) Using the fact that  $Q$  can be generated by a random constant mapping, show that the trajectories always coalesce when simulating exactly the target distribution  $P$  using Propp and Wilson's algorithm. (Observation by Murdoch and Green, 1998).

3) Give an upper bound for the complementary distribution function of number of iterations for coalescence.

**10.4** Show that in the case where the finite state-space  $\Omega$  is equipped with a metric  $d$ , then theorem 10.1.4 can be replaced by the more flexible statement:  $\#\Omega_\infty = 1$  a.s. if and only if for each pair of distinct points  $x$  and  $y$  in  $\Omega$  there exists  $f \in \mathcal{M}_c$  such that  $d[f(x), f(y)] < d(x, y)$ .

**10.5** Let  $\Omega = \{t, x, y, z\}$  with  $t \preceq x, y \preceq z, x$  and  $y$  being incomparable.

1) What are the up-sets of  $\Omega$ ?

2) Let  $P$  be the transition kernel defined as follows

$$P = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

Show that  $P$  is stochastically monotonic

3) Suppose that  $P$  is generated by an increasing random mapping  $F$ . Show that

$$P\{F(t) = x\} = P\{F(t) = x, F(x) = z, F(y) = x, F(z) = z\} = \frac{1}{2}$$

$$P\{F(z) = x\} = P\{F(t) = t, F(x) = t, F(y) = x, F(z) = x\} = \frac{1}{2}$$

and conclude to a contradiction (this counter-example was discovered by Machida, 1999).

**10.6** Let  $(\Omega, \preceq)$  be a poset. A mapping  $f$  from  $\Omega$  to  $\Omega$  is said to be decreasing if  $f(x) \succeq f(y)$  whenever  $x \preceq y$ .

1) Show that  $f$  is decreasing if and only if the inverse image by  $f$  of an up-set (resp. of a down-set) is a down-set (resp. an up-set).

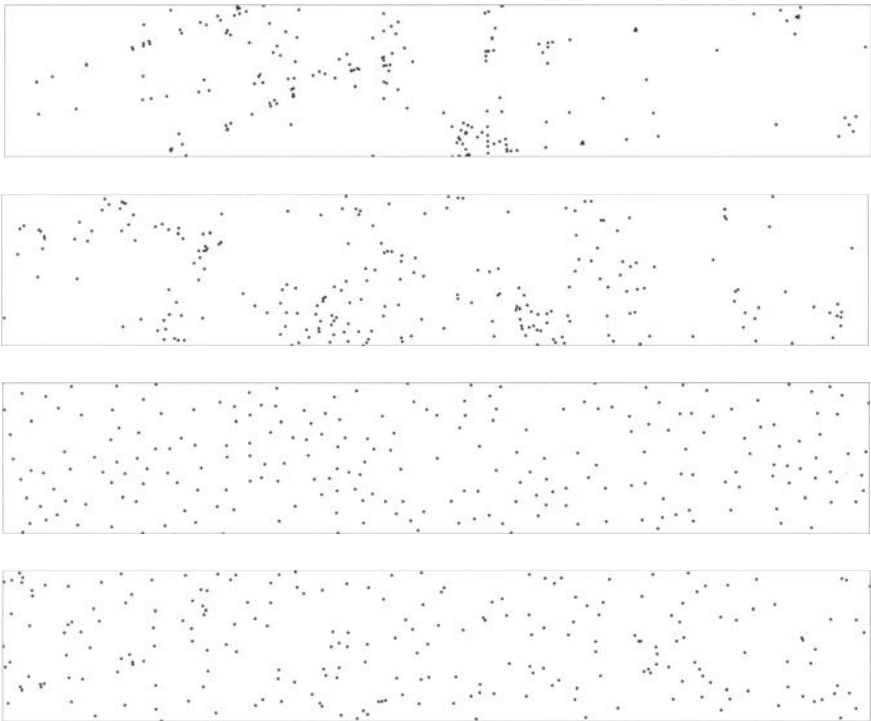
2) Let  $P$  be a transition kernel on  $\Omega$  with limit distribution  $p$ . Suppose that  $P$  is generated by a decreasing random mapping. In that case, show that algorithm 10.1.2 is still applicable to simulate the target distribution  $p$  exactly (Kendall 1996,1997).

## Part III

### **The models**

# 11. Point processes

There are at least three reasons for focussing attention on the particular class of random sets composed of the *point processes*. Firstly, despite the rudimentary character of their constitutive elements, point processes can exhibit rather complicated structures (cf. Figure 11.1). Secondly, they can be used for modelling a wide range of natural phenomena (trees in a forest, precious stone deposits, earthquake epicentres, stars in galaxies, populations in settlements...). Thirdly, point processes can act as basic ingredients for designing more complicated models.



**Fig. 11.1.** A few examples of point processes. From top to bottom, intersections of random lines, a cluster process, a repulsion process and purely random points.

One-dimensional point processes have been extensively studied and are the object of a vast amount of literature. The seminal work by Matérn (1960) still constitutes an excellent statistical reference, even if more up to date surveys are now available (Stoyan et al., 1987; Daley and Vere-Jones, 1988).

This chapter deals with two locally finite point processes, namely the Poisson point process and the Cox process. In addition to reviewing their main properties, the problem of their conditional simulation is addressed.

## 11.1 The Poisson point process

Because of its practical importance, the stationary case is considered first.

### 11.1.1 The homogeneous Poisson point process

As was seen in section 2.4, the statistical properties of this locally finite point process can be specified by its spatial distribution. In this chapter,  $N$  denotes the counting functional of the process. It is defined on the family  $\mathcal{B}_0$  of the bounded borelians of  $\mathbb{R}^d$ .

**Definition 11.1.1.** *The spatial distribution of a homogeneous Poisson point process with intensity  $\theta$  is characterized by the following two properties:*

*i) if  $A \in \mathcal{B}_0$ , then the number  $N(A)$  of points in  $A$  is a Poisson random variable with parameter  $\theta|A|$  where  $|A|$  denotes the  $d$ -volume of  $A$*

$$P\{N(A) = n\} = \exp\{-\theta|A|\} \frac{[\theta|A|]^n}{n!}$$

*ii) if  $(A_i, i \in I)$  is a finite family of pairwise disjoint elements of  $\mathcal{B}_0$ , then the random variables  $(N(A_i), i \in I)$  are mutually independent.*

Since the mean value of a Poisson distribution coincides with its parameter, the mean number of points within  $A$  is equal to  $\theta|A|$ . In particular,  $\theta$  is the mean number of points per unit  $d$ -volume.

**Remark 11.1.1.** It is not obvious that this definition completely characterizes a spatial distribution. Indeed, let  $(A_i, i \in I)$  be an arbitrary finite family of (not necessarily disjoint) elements of  $\mathcal{B}_0$ , and let  $(n_i, i \in I)$  be a finite family of integers. For each subset  $J$  of  $I$ , we denote by  $A_J$  the set of points that belong to all the  $A_j$ 's indexed by  $j \in J$  and only to them<sup>1</sup>. The  $A_J$ 's are pairwise disjoint. Accordingly, property ii) of the definition implies

<sup>1</sup> Formally,  $A_J = \bigcap_{j \in J} A_j \setminus \bigcup_{k \notin J} A_k$ . For instance,  $A_{\{i\}}$  is the set of all points contained only in  $A_i$ .



$$\begin{aligned}
 P\left\{\bigwedge_{i \in I} N(A_i) = n_i\right\} &= \sum P\left\{\bigwedge_{\emptyset \neq J \subset I} N(A_J) = n_J\right\} \\
 &= \sum \prod_{\emptyset \neq J \subset I} P\{N(A_J) = n_J\},
 \end{aligned}$$

in which the sums are over all sequences of integers  $(n_J, J \subset I)$  such that  $\sum_{J \ni i} n_J = n_i$  for any index  $i \in I$ . It follows that

$$P\left\{\bigwedge_{i \in I} N(A_i) = n_i\right\} = \sum \prod_{\emptyset \neq J \subset I} \exp\{-\theta|A_J|\} \frac{[\theta|A_J|]^{n_J}}{n_J!}$$

by property i).

The following theorem shows that the points of a homogeneous Poisson point process are independently and purely randomly located.

**Theorem 11.1.1.** *Let  $A \in \mathcal{B}_0$ . If  $N(A) = n$ , the  $n$  points are independent and uniform in  $A$ .*

**Proof:** Suppose  $N(A) = n$ , and consider the random closed set made up of these  $n$  points. Its avoiding functional (cf. section 2.3) can be written as

$$\begin{aligned}
 Q_n(K) &= P\{N(A \cap K) = 0 \mid N(A) = n\} \\
 &= \frac{P\{N(K \cap A) = 0 \wedge N(A \setminus K) = n\}}{P\{N(A) = n\}} \\
 &= \frac{P\{N(K \cap A) = 0\} P\{N(A \setminus K) = n\}}{P\{N(A) = n\}} \\
 &= \frac{\exp\{-\theta|K \cap A|\} \exp\{-\theta|A \setminus K|\} \frac{(\theta|A \setminus K|)^n}{n!}}{\exp\{-\theta|A|\} \frac{[\theta|A|]^n}{n!}} = \left[ \frac{|A \setminus K|}{|A|} \right]^n
 \end{aligned}$$

for each compact set  $K \in \mathcal{K}$ . This completes the proof because  $|A \setminus K|/|A|$  is the avoiding functional at  $K$  of a point uniformly located in  $A$  (cf. example 2.3.1).  $\square$

This theorem leads to a very simple algorithm for simulating a Poisson point process within  $A$ :

**Algorithm 11.1.1.** *(Simulation of a homogeneous Poisson point process)*

1. Set  $X = \emptyset$ .
2. Generate  $n \sim \text{Poisson}(\theta|A|)$ .
3. If  $n > 0$ , then for  $i = 1, \dots, n$  generate  $x_i \sim \text{Unif}(A)$  independently.
4. Deliver  $X = \{x_1, \dots, x_n\}$ .

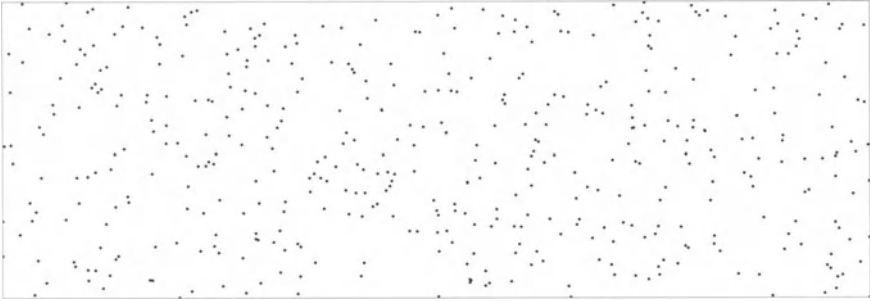


Fig. 11.2. Realization of a homogeneous Poisson point process.

### 11.1.2 The non-homogeneous Poisson point process

In this section, the Poisson intensity  $\theta$  is no longer constant. It varies through space. In other words,  $\theta = (\theta(x), x \in \mathbb{R}^d)$  is now a function that is supposed to be positive and locally integrable. For each  $A \in \mathcal{B}_0$ , we denote by  $\theta(A)$  the integral of  $\theta$  over  $A$

$$\theta(A) = \int_A \theta(x) dx$$

**Definition 11.1.2.** *The spatial distribution of a Poisson point process with intensity function  $\theta$  is characterized by the following two properties:*

i) if  $A \in \mathcal{B}_0$ , then the number  $N(A)$  of points in  $A$  is a Poisson random variable with parameter  $\theta(A)$ .

$$P\{N(A) = n\} = \exp\{-\theta(A)\} \frac{[\theta(A)]^n}{n!}$$

ii) if  $(A_i, i \in I)$  is a finite family of pairwise disjoint elements of  $\mathcal{B}_0$ , then the random variables  $(N(A_i), i \in I)$  are mutually independent.

Of course, this definition is merely an extension of 11.1.1. The same remark concerning its apparent lack of completeness can be made.

Suppose that  $A$  contains  $n$  points of the process. Where are they located? The answer is given by the following theorem:

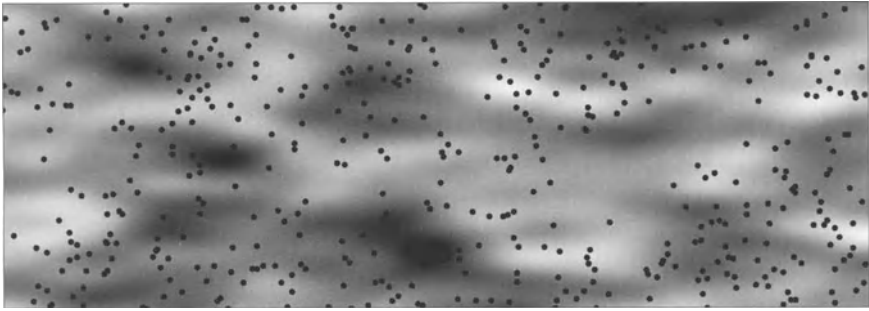
**Theorem 11.1.2.** *Let  $A \in \mathcal{B}_0$ . If  $N(A) = n$ , the  $n$  points are independent and identically distributed within  $A$  w.r.t. the density function  $\theta 1_{\in A}$ .*

From this theorem, we can immediately derive an algorithm to simulate the Poisson point process in  $A$ . This is basically the same algorithm as in the homogeneous case. The only difference is that the intensity function  $\theta$  is also required to simulate the location of the points within  $A$ , which is usually done by an acceptance-rejection procedure (cf. algorithm 7.2.2).

**Algorithm 11.1.2.** (*Simulation of a non-homogeneous Poisson point process*)

1. Set  $X = \emptyset$ .
2. Generate  $n \sim \text{Poisson}(\theta(A))$ .
3. If  $n > 0$ , then for  $i = 1, \dots, n$  generate  $x_i \sim \theta 1_{\in A}$  independently.
4. Deliver  $X = \{x_1, \dots, x_n\}$ .

Figure 11.3 shows a realization of a Poisson point process in superposition with its intensity function.



**Fig. 11.3.** Realization of a non-homogeneous Poisson point process. Its intensity function is depicted by the greytone background.

To conclude this section, let us calculate the covariance between the numbers of points in two elements  $A$  and  $B$  of  $\mathcal{B}_0$ . Writing

$$N(A) = N(A \cap B) + N(A \setminus B)$$

$$N(B) = N(A \cap B) + N(B \setminus A)$$

and using the independence property in disjoint domains, it follows that

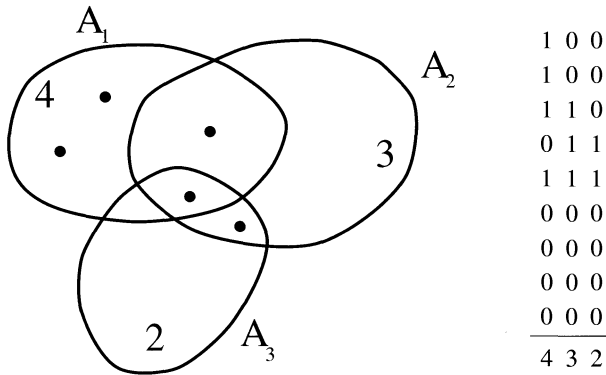
$$\text{Cov}\{N(A), N(B)\} = \text{Var}\{N(A \cap B)\}$$

Now  $N(A \cap B)$  is Poisson distributed with parameter  $\theta(A \cap B)$ , and its variance is  $\theta(A \cap B)$ . We have finally shown

**Proposition 11.1.1.** *If  $A, B \in \mathcal{B}_0$ , then  $\text{Cov}\{N(A), N(B)\} = \theta(A \cap B)$ .*

### 11.2 Conditional simulation of a Poisson point process

The objective of this section is to simulate a Poisson point process (with known intensity function  $\theta$ ) subject to the condition that  $N(A_i) = n_i$  for some finite family  $(A_i, i \in I)$  of elements of  $\mathcal{B}_0$  (see left of Figure 11.4). In the case when the cardinality  $\#I$  of  $I$  is small, it is possible to carry out nonconditional simulations and to keep only those that honour the conditions. Things are different when  $\#I$  is large and when the  $A_i$ 's are overlapping: the rejection rate becomes too high.



**Fig. 11.4.** How to carry out a conditional simulation of a Poisson point process in  $A_1 \cup A_2 \cup A_3$  given that  $N(A_1) = 4$ ,  $N(A_2) = 3$  and  $N(A_3) = 2$ ? The solution proposed requires a coding of the simulations that is represented on the right of the figure

Because of the independence property ii) of definition 11.1.2, the part of the simulations outside the union of the  $A_i$ 's is not affected by the conditions. As a consequence, the conditional simulation problem really arises only in  $A = \bigcup_{i \in I} A_i$ . Here  $A$  is chosen as the simulation field.

Let us start with two preliminary remarks. Firstly, the number of points of a simulation is bounded above by  $n = \sum_{i \in I} n_i$ . Secondly, the conditional simulation can be carried out in two steps<sup>2</sup>

<sup>2</sup> The formula

$$\begin{aligned}
 P\left\{N(K) = 0 \mid \bigwedge_{i \in I} N(A_i) = n_i\right\} &= \sum_{(n_J)} P\left\{\bigwedge_{\emptyset \neq J \subset I} N(A_J) = n_J \mid \bigwedge_{i \in I} N(A_i) = n_i\right\} \\
 &\quad \times P\left\{N(K) = 0 \mid \bigwedge_{\emptyset \neq J \subset I} N(A_J) = n_J\right\} \quad K \in \mathcal{K}
 \end{aligned}$$

shows that the simulation can be carried out in two steps. Regarding the second step, the fact that the points of the  $A_J$ 's have independent locations directly stems from the formula

- i) the joint simulation of the  $N(A_J)$ 's, given  $N(A_i) = n_i$  for each  $i \in I$ .
- ii) the independent simulation of the points in each  $A_J$  given  $N(A_J) = n_J$ .

The second step does not present any difficulties because of theorem 11.1.2. Only the first one needs special treatment.

Consider a realization of the Poisson point process such as the one in Figure 11.4. It can be regarded as a sequence of points  $(x_1, \dots, x_p)$  with  $p \leq n$ . A coding consisting of  $\#I$  bits can be associated with each point  $x_j$  by setting the bit of order  $i$  to 1 if  $x_j \in A_i$  and to 0 if not. This gives an array of  $p$  rows and  $\#I$  columns that is completed by  $n - p$  rows of 0, as shown on the right of Figure 11.4.

Let  $\Omega$  be the set of all arrays of size  $n \times \#I$  that can be obtained in this way. Each element  $\omega \in \Omega$  is assigned the probability

$$p(\omega) \propto n_\emptyset(\omega)! \prod_{\emptyset \neq J \subset I} [\theta(A_J)]^{n_J(\omega)}$$

where  $n_J(\omega)$  denotes the number of rows of  $\omega$  corresponding to points of  $A_J$  (in particular,  $n_\emptyset(\omega)$  is the number of rows of zeroes in  $\omega$ ). Using this probability measure, it is not difficult to show that the multivariate distribution of the numbers  $(N_J, \emptyset \neq J \subset I)$  of rows associated with the coding of points of  $A_J$  coincides with the conditional multivariate distribution of the  $N(A_J)$ 's<sup>3</sup>

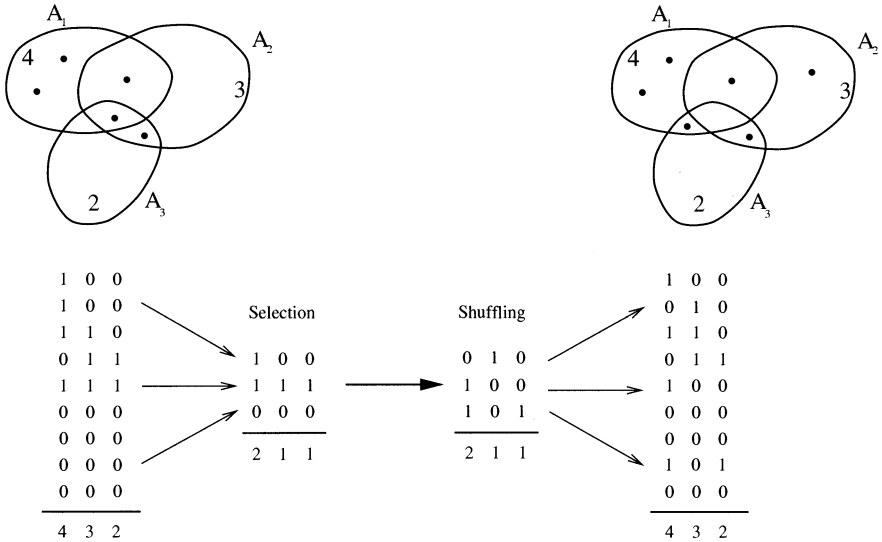
$$\begin{aligned} P\left\{ \bigwedge_{\emptyset \neq J \subset I} N_J = n_J \right\} &\propto \sum \binom{n}{\dots n_J \dots} n_\emptyset! \prod_{\emptyset \neq J \subset I} \theta(A_J)^{n_J} \\ &\propto \sum \prod_{\emptyset \neq J \subset I} \frac{\theta(A_J)^{n_J}}{n_J!} \\ &\propto P\left\{ \bigwedge_{\emptyset \neq J \subset I} N(A_J) = n_J \mid \bigwedge_{i \in I} N(A_i) = n_i \right\} \end{aligned}$$

Therefore the first step of the conditional simulation amounts to the simulation of distribution  $p$ . This can be done using the Metropolis algorithm (cf. section 8.2). Let  $\omega$  be the current array. For the next iteration, a candidate array is built by selecting uniformly and independently a certain number of rows of  $\omega$ , by permuting uniformly their bits indepently along each column and by replacing the selected rows by the new ones (see Figure 11.5). The candidate array  $\omega'$  thus obtained is then accepted with probability

$$\begin{aligned} P\left\{ N(K) = 0 \mid \bigwedge_{\emptyset \neq J \subset I} N(A_J) = n_J \right\} &= P\{N(K) = 0\} \times \\ &\prod_{\emptyset \neq J \subset I} P\{N(A_J \setminus K) = n_J \mid N(A_J) = n_J\} \end{aligned}$$

<sup>3</sup> In this formula, the summation is over all finite sequences  $(n_J, J \subset I)$  such that  $\sum_{J \ni i} n_J = n_i$  for each index  $i \in I$ . A multinomial coefficient is introduced to account for the fact that the rows of the  $\omega$ 's are ordered.

$$\alpha(\omega, \omega') = \frac{p(\omega')}{p(\omega) + p(\omega')}$$



**Fig. 11.5.** Generation of a candidate array from the current one. Here three rows are randomly selected. Their bits are independently permuted along each column. This produces three new lines that replace the selected ones.

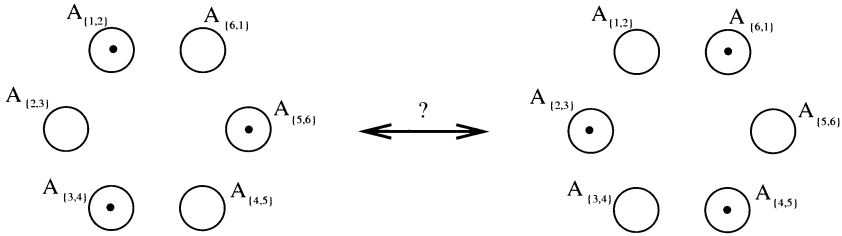
This algorithm depends on one parameter, namely the number of rows  $k$  to be selected. This number must be chosen carefully. If  $k$  is too large (for instance  $k = n$ ), then the acceptance rate is very small. Things improve by taking  $k$  smaller, but then some irreducibility problems may be encountered (see Figure 11.6). As the minimum value for  $k$  that ensures irreducibility is usually unknown, a reasonable solution is to take  $k$  random and to simulate it at each iteration. For instance, the rows can be independently selected with the same probability  $\nu$ . Then  $k$  has a binomial distribution with parameter  $\nu$  and index  $n$ .

Finally, here is the algorithm to conditionally simulate the numbers of points in the  $A_J$ 's.

**Algorithm 11.2.1.** (*Conditional simulation of a Poisson point process. First step*)

1. Let  $\omega \in \Omega$ .
2. Select independently each row of  $\omega$  with probability  $\nu$ .
3. Permute the bits of the selected rows uniformly along each column independently.

4. Replace the selected rows by the newly generated ones. Let  $\omega'$  be the candidate array thus obtained.
5. Let  $U \sim u$ . If  $(p(\omega) + p(\omega'))U < p(\omega)$ , then put  $\omega = \omega'$ .
6. Go to 2.



**Fig. 11.6.** Each of the 6 borelians  $A_1, \dots, A_6$  has 2 connected components, and each component is common to 2 of them. If exactly one point is allowed in each borelian, then only 2 configurations of points are possible, and at least 3 rows must be selected to ensure the transition between both configurations.

At this point, nothing has been said about how to start running the algorithm. Things are quite easy in the case when each  $A_{\{i\}}$  has a positive intensity because it suffices to set  $n_{\{i\}} = n_i$  for each  $i \in I$ .

**Remark 11.2.1.** Even in the case when  $\theta(A_{\{i\}}) = 0$  for some index  $i \in I$ , it is still possible to initialize the simulation by setting  $n_{\{i\}} = n_i$  for each  $i \in I$ . Because of Metropolis' acceptance criterion, the initial points are rapidly relocated, provided that the conditions are mutually compatible, that is  $\Omega \neq \emptyset$ . The non-emptiness of  $\Omega$  is related to the existence of solutions to the linear system of equations

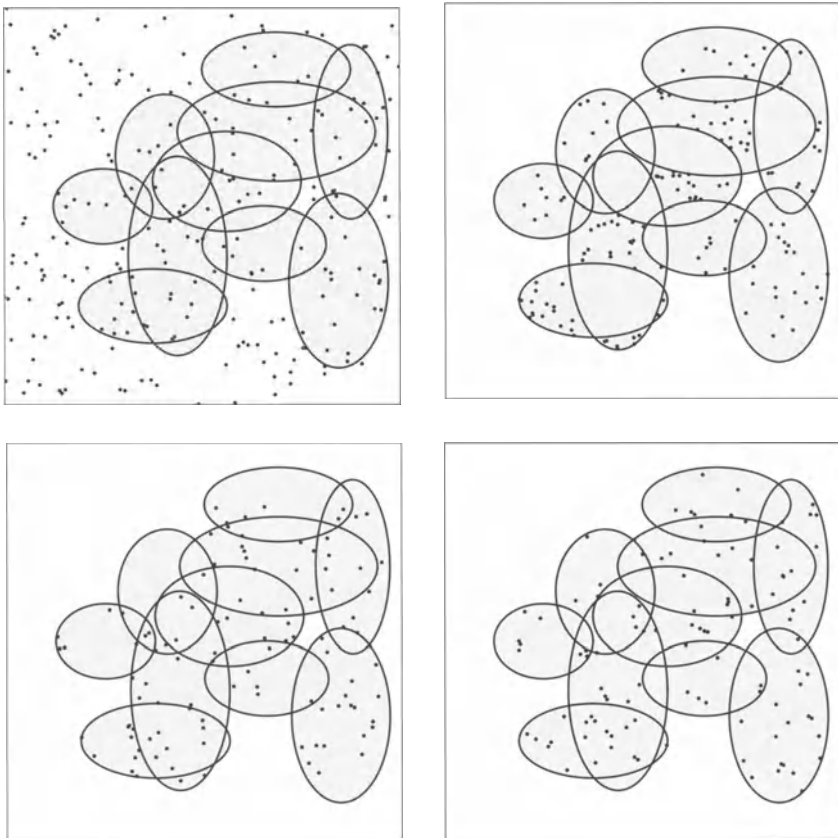
$$\sum_{J \ni i} n_J 1_{A_J \neq \emptyset} = n_i \quad i \in I$$

A clearly necessary condition is that the inequality  $\sum_J n_j \leq \sum_K n_k$  holds for any disjoint subfamilies  $J$  and  $K$  of  $I$  such that  $\bigcup_J A_j \subset \bigcup_K A_k$ . But this condition is not sufficient. As a counterexample, consider the three sets  $A_1 = \{y, z\}$ ,  $A_2 = \{x, z\}$  and  $A_3 = \{x, y\}$  where  $x, y$  and  $z$  are three distinct points of  $\mathbb{R}^d$ . The necessary condition is clearly satisfied whatever the values assigned to  $n_1, n_2$  and  $n_3$ , but there is no solution to the system in the case when  $n_1 = n_2 = n_3 = 1$ . More generally, let  $S(n_i, i \in I)$  be the number of solutions of the linear system of equations. Its generating function is given by

$$\sum_{(n_i, i \in I)} S(n_i, i \in I) \prod_{i \in I} s_i^{n_i} = \prod_{A_J \neq \emptyset} \frac{1}{1 - \prod_{j \in J} s_j} \quad (0 \leq s_i < 1, i \in I)$$

as can be easily shown by expanding the right hand side of this formula.

In order to illustrate how this algorithm works in practice, we have considered a non-conditional simulation of a homogeneous Poisson point process with intensity  $\theta = 0.0015$  in a field  $400 \times 400$  (see top left of Figure 11.7). Ten ellipses have been superimposed on the simulation to act as conditioning sets. They enclose 149 points and the content of individual ellipses varies from 7 to 26. Starting from the initial population depicted in the top right of Figure 11.7, the algorithm has been applied with a selection rate of 0.03, which means that on average 4.5 rows are selected at each iteration. After running the algorithm twice for 10000 iterations, the two bottom populations were obtained. In both cases, the number of candidate transitions accepted is about 3500. With a higher selection rate of 0.05, this number would have dropped to 1500.



**Fig. 11.7.** Conditional simulation of a Poisson point process. Top left, a non conditional simulation with the conditioning sets. Top right, the initial population. Bottom, two conditional simulations obtained after 10000 iterations.



**Remark 11.2.2.** The algorithm proposed requires the knowledge of the intensities  $(\theta(A_J), \emptyset \neq J \subset I)$ . This does not mean that these intensities must be permanently kept in memory, but simply that they must be calculated each time they are required. We have designed other algorithms that do not rely on the actual value of the  $\theta(A_J)$ 's, but all of them have turned out to be extremely inefficient.

### 11.3 The Cox process

Suppose now that the intensity function  $\theta$  is randomized. This gives the Cox process (Cox, 1955).

**Definition 11.3.1.** A Cox process is a Poisson point process whose intensity is a positive, a.s. locally integrable random function.

Let  $\Theta$  denote the random intensity function. By definition, the number  $N(A)$  of points in  $A \in \mathcal{B}_0$  has a Poisson distribution with (random) mean

$$\Theta(A) = \int_A \Theta(x) dx$$

The system of probabilities is therefore equal to

$$P\{N(A) = n\} = E \left\{ e^{-\Theta(A)} \frac{(\Theta(A))^n}{n!} \right\}$$

Unless  $\Theta(A)$  is a.s. constant,  $N(A)$  is not Poisson distributed.

**Example 11.3.1.** If  $\Theta(A)$  follows a gamma distribution

$$f(t) = \frac{b^\alpha}{\Gamma(\alpha)} e^{-bt} t^{\alpha-1} \quad \alpha, b > 0; \quad t > 0$$

then  $N(A)$  has a negative binomial distribution

$$P\{N(A) = n\} = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha) n!} \left( \frac{b}{b+1} \right)^\alpha \left( \frac{1}{b+1} \right)^n$$

**Example 11.3.2.** If  $\Theta(A)$  has a generalized inverse gaussian distribution (Jørgensen, 1982)

$$f(t) = \frac{\left(\frac{a}{b}\right)^{\frac{\alpha}{2}}}{2K_\alpha(2\sqrt{ab})} t^{\alpha-1} \exp \left\{ -at - \frac{b}{t} \right\} \quad \alpha \in \mathbb{R}, \quad a, b > 0; \quad t > 0$$

where  $K_\alpha$  stands for the modified Bessel function of the third kind of order  $\alpha$  (Abramovitz and stegun, 1970), then  $N(A)$  follows a generalized Sichel distribution

$$P\{N(A) = n\} = \frac{1}{n!} \frac{\left(\frac{a}{b}\right)^{\frac{\alpha}{2}}}{\left(\frac{a+1}{b}\right)^{\frac{\alpha+n}{2}}} \frac{K_{\alpha+n}\left(2\sqrt{(a+1)b}\right)}{K_{\alpha}\left(2\sqrt{ab}\right)}$$

The case  $\alpha = -\frac{1}{2}$  corresponds to the standard Sichel distribution. This distribution has been introduced for representing long-tailed frequency data in various fields such as linguistics (sentence-length in written prose) and the mining industry (stone density in alluvial diamond deposits) (Sichel, 1973; 1974).

Clearly, the mean number of points in  $A$  is  $E\{N(A)\} = E\{\Theta(A)\}$ . The variance has a more complicated form

$$\text{Var}\{N(A)\} = E\{\Theta(A)\} + \text{Var}\{\Theta(A)\}$$

Rather than proving this directly, let us get it as a by-product of the following proposition.

**Proposition 11.3.1.** *If  $A, B \in \mathcal{B}_0$ , then*

$$\text{Cov}\{N(A), N(B)\} = E\{\Theta(A \cap B)\} + \text{Cov}\{\Theta(A), \Theta(B)\}$$

**Proof:** Let us express the covariance as the difference of two terms

$$\text{Cov}\{N(A), N(B)\} = E\{N(A)N(B)\} - E\{N(A)\}E\{N(B)\}$$

To calculate the first term, we randomize over all possible realizations of  $\Theta$ . For each realization, the point process under study is a Poisson point process. Then proposition 11.1.1 can be applied to give

$$E\{N(A)N(B)\} = E\{E\{N(A)N(B) | \Theta\}\} = E\{\Theta(A \cap B) + \Theta(A)\Theta(B)\}$$

It remains to subtract  $E\{N(A)\}E\{N(B)\} = E\{\Theta(A)\}E\{\Theta(B)\}$  to obtain the desired result.  $\square$

Thus, the covariance between  $N(A)$  and  $N(B)$  is the sum of two terms. The first one results from the Poisson implantation of the points, once that the intensity has been fixed. The second one takes the spatial statistical fluctuations of the intensity into account. This covariance formula clearly brings out the *hierarchical* construction of the Cox process. One very interesting feature of this point process is that, in contrast to the Poisson point process, the numbers of points within disjoint domains are not necessarily independent.

Suppose now that  $n$  points of the Cox process are contained in  $A$ . How are they distributed spatially? The answer to that question is not as simple as the one in the case of a Poisson point process (cf. theorems 11.1.1 and 11.1.2). See for instance exercise 11.4. Note however that this does not have any impact

on the non conditional simulation of a Cox process because it can be achieved by simulating first the intensity function, then the Poisson point process with the intensity function generated. The problem of the conditional simulation is much tougher.

Here are the main lines of an algorithm to simulate the Cox process in  $B \in \mathcal{B}_0$  subject to the conditions  $(N(A_i) = n_i, i \in I)$ . Details of proof are omitted. In this algorithm, the union of the  $A_i$ 's is again denoted by  $A$ . There is no reason to take  $B = A$  because the range of the conditions can go far beyond  $A$ .

**Algorithm 11.3.1.** (*Conditional simulation of a Cox process*)

1. Simulate  $\Theta$  in  $A$  given  $(N(A_i) = n_i, i \in I)$ . Let  $(\theta(x), x \in A)$  be the resulting family of intensities.
2. Simulate  $\Theta$  in  $B$  given  $(\Theta(x) = \theta(x), x \in A)$ . Let  $(\theta(x), x \in B)$  be the resulting family of intensities.
3. Simulate a Poisson point process with intensity function  $(\theta(x), x \in B)$  in  $B$  given  $(N(A_i) = n_i, i \in I)$ .

The way of carrying out steps 1 and 2 depends largely on the model considered for  $\Theta$ .

In many applications, people are not directly interested in the exact location of all the points of the process. They simply want to know how many points can be found in some parts of the simulation field. In this case, it is sufficient to simulate the number of points in a family of non-overlapping and congruent subsets  $(B_j, j \in J)$  that cover  $B$ . Things are facilitated when each  $A_i$  coincides with one  $B_j$ . Then the algorithm takes the form

**Algorithm 11.3.2.** (*Conditional simulation of a Cox process. Simplified version*)

1. Simulate  $(\Theta(A_i), i \in I)$  given  $(N(A_i) = n_i, i \in I)$ . Let  $(\theta_i, i \in I)$  be the generated values.
2. Simulate  $(\Theta(B_j), j \in J \setminus I)$  given  $(\Theta(A_i) = \theta_i, i \in I)$ . Let  $(\theta_j, j \in J \setminus I)$  be the generated values.
3. For any  $j \in J \setminus I$ , generate  $N(B_j) \sim \text{Poisson}(\theta_j)$ .

The first two steps of this algorithm were used by Freulon (1992, 1994) to restore electron microprobe images of chemical concentration of steel samples blurred by Poisson noise. The Gibbs sampler (cf. section 8.3.2) was found quite appropriate for carrying out the first step. In the mining industry, the Cox process can also be used to model deposits where the mineralization takes place as discrete particles, such as gold grains, heavy minerals or precious stones (Kleingeld, 1987). The random intensity  $\Theta(A)$  associated with a region  $A$  represents all topographic, geologic and geographic factors that affect the

particle concentration and indicates the propensity for the region to be rich. In this context, conditional simulations and implementation details can be found in (Kleingeld et al., 1997).

## Exercises

**11.1** Consider a homogeneous Poisson point process with intensity  $\theta$  in  $\mathbb{R}$ . Show that the points of the process split  $\mathbb{R}$  into independent intervals of exponential length with mean value  $\theta^{-1}$ .

**11.2** Let  $x$  be a point of a Poisson point process with intensity function  $\theta$  in  $\mathbb{R}$ . What is the length distribution of the interval delimited by the process on the right of  $x$ ?

**11.3** Let  $X$  be a planar Poisson point process with constant intensity  $\theta$ . Let  $Y$  be the set of the distances between the points of  $X$  and the origin:

$$Y = \{d(x, o), x \in X\}$$

Show that  $Y$  is a Poisson point process on  $[0, \infty[$  with intensity

$$\theta_Y(r) = 2\pi\theta r \quad r \geq 0$$

**11.4** Let  $N$  be the counting functional of a Cox process with random intensity function  $\Theta$ . Given  $A, B \in \mathcal{B}_0$  with  $B \subset A$ , show that

$$P\{N(B) = 1 \mid N(A) = 1\} = \frac{E\{e^{-\Theta(A)}\Theta(B)\}}{E\{e^{-\Theta(A)}\Theta(A)\}}$$

**11.5** Suppose  $\Theta \sim \text{Gamma}(\alpha, b)$  and let  $N \sim \text{Poisson}(\Theta)$ . Show that

$$\Theta \mid N = n \sim \text{Gamma}(\alpha + n, b + 1)$$

## 12. Tessellations

A *tessellation* is a division of space into small units or *cells*. The cells are usually polytopes (polygons in  $\mathbb{R}^2$ , polyhedra in  $\mathbb{R}^3$ ), but this is not strictly necessary. Depending on the application considered, a tessellation can be regarded either as a partition of space or as a random function (by assigning each cell a value), or even as a population of cells. Of course, different interpretations lead to different statistical characterizations. A brief description of the possible interpretations is given in the first section of this chapter. Then we turn to the presentation and the (conditional) simulation of two well known tessellation models, namely the Voronoi and the Poisson tessellations.

### 12.1 Statistical characterization of a tessellation

#### 12.1.1 Partition point of view

As a first approximation, the cells of a tessellation can be modelled as a family of bounded subsets partitioning  $\mathbb{R}^d$ . However, this modeling is not satisfactory because it lacks precision. Whereas there is no difficulty to ascertain that a point belongs to the interior of a cell, attributing a point at the boundary of two cells to either cell is arbitrary. One way to overcome this difficulty is to make a clear distinction between the cells and their boundaries. In what follows, the cells are modelled by pairwise disjoint bounded open sets such that the union of their closures fills the whole space. Accordingly,  $\mathbb{R}^d$  is partitioned into a family of cells  $(X_i, i \in I)$  and the union of their boundaries  $\cup_{i \in I} \partial X_i$ . In the case when each cell is topologically connected, the tessellation is completely characterized by the cell boundaries. Since the cells are open, the union of all cell boundaries is closed and can be statistically described by its hitting functional. This approach was successfully applied by Mecke (1983) and Stoyan et al. (1985) to describe tessellations of convex polyhedra.

Unfortunately, this approach can not be applied if the cell are not connected (for instance, the dead-leaves model, see section 14.3). In that case, Mathéron's approach (1969) can be considered. It consists of partitioning  $\mathbb{R}^d$  into the cells  $(X_i, i \in I)$  and the *points* of the boundaries  $(\{x\}, x \in \cup_{i \in I} \partial X_i)$ . In

contrast with the previous approach, the boundaries of the cells are not seen as a whole, but as a family of individual points, hence the expression "atomizing the boundaries" used by Matheron. Let  $\mathfrak{P}$  be the set of partitions that can be defined in this way. If  $\mathcal{X}$  is such a partition, then an element of  $\mathcal{X}$  is either a cell or a point of the boundary of a cell.  $\mathfrak{P}$  can be equipped with a  $\sigma$ -algebra  $\sigma(\mathfrak{P})$  spanned by the events

$$\mathfrak{A}_G = \{\mathcal{X} \in \mathfrak{P} \mid G \subset X \text{ for some } X \in \mathcal{X}\}$$

for all open subsets  $G$  of  $\mathbb{R}^d$ . Matheron (1969) has shown that probabilities do exist on  $(\mathfrak{P}, \sigma(\mathfrak{P}))$ .

**Remark 12.1.1.** The assertion "the point  $x$  belongs to a cell" is an event. To see that, let us denote by  $G_n$  the open ball with centre  $x$  and radius  $\frac{1}{n}$ . Because the cells are open,  $x$  belongs to a cell if and only if there exists a neighbourhood of  $x$  contained in a cell. Accordingly

$$\mathfrak{A}_{\{x\}} = \bigcup_{n \geq 1} \mathfrak{A}_{G_n} \in \sigma(\mathfrak{P})$$

In what follows, we will assume  $P(\mathfrak{A}_{\{x\}}) = 1$  for any  $x \in \mathbb{R}^d$ .

**Remark 12.1.2.** Similarly, the assertion "both points  $x$  and  $y$  belong to the same cell" is also an event. The same reasoning holds if we take  $G_n$  equal to  $\{x, y\}$  dilated by the open ball centered at the origin and with radius  $\frac{1}{n}$ . Because  $\sigma(\mathfrak{P})$  is stable under intersection and set difference,  $(\mathfrak{A}_{\{x\}} \cap \mathfrak{A}_{\{y\}}) \setminus \mathfrak{A}_{\{x, y\}} \in \sigma(\mathfrak{P})$ , which means that "  $x$  and  $y$  belong to different cells" is an event. More generally, assertions like "the  $k$  points  $x_1, \dots, x_k$  belong to  $p$  different cells" are also events.

From the partition point of view, a random tessellation is characterized by a probability measure  $P$  on  $(\mathfrak{P}, \sigma(\mathfrak{P}))$ . It is said to be *stationary* if  $P$  is invariant under translation

$$P(\mathfrak{A}_{G_h}) = P(\mathfrak{A}_G) \quad G \in \mathcal{G}, h \in \mathbb{R}^d$$

### 12.1.2 Random function point of view

Here, each cell of the tessellation is assigned a value drawn from a distribution  $F$  with mean  $m$  and variance  $\sigma^2$ . The values do not depend on the cells and different cells have independent values. Because all points are assumed to belong to a cell almost surely (cf. remark 12.1.1), each point  $x$  can be assigned a random value  $Z(x)$ . This defines a random function  $Z$  whose statistical properties can be described by its spatial distribution

$$F_{x_1, \dots, x_k}(z_1, \dots, z_k) = P\{Z(x_1) < z_1, \dots, Z(x_k) < z_k\}$$

for all  $k$ , all set of points  $(x_1, \dots, x_k)$  and all set of values  $(z_1, \dots, z_k)$  (cf. remark 12.1.2). Suppose that the tessellation is stationary. If  $\rho(x - y)$  denotes the probability that both points  $x$  and  $y$  belong to the same cell, then we have

$$E\{Z(x)Z(y)\} = (\sigma^2 + m^2)\rho(x - y) + m^2[1 - \rho(x - y)]$$

so that

$$Cov\{Z(x), Z(y)\} = \sigma^2 \rho(x - y)$$

### 12.1.3 Cell population point of view

For a stationary tessellation, the mean  $d$ -volume of the cells, the mean  $(d - 1)$ -volume of their boundaries etc... are of interest. In order to calculate these mean values, we must at first be more specific about their definition.

One possible definition consists of assigning the same weight to each cell. This is intuitively simple, but technically difficult since a tessellation may have an infinite number of cells. Consider first a realisation of the random tessellation within a bounded domain. If the tessellation is *locally finite* the domain contains a finite number of cells. One can therefore assign the same weight to each cell within the domain and build the empirical histogram of any characteristic of the cells. Assume that this empirical histogram converges to some distribution as the size of the domain becomes very large. Assume moreover that the distribution thus obtained does not depend on the realisation considered. The limit distribution thus obtained is said to be *weighted in number*.

Another possible definition deals with the characteristics of a single cell only, for instance the one that contains the origin (this is possible because the origin belongs almost surely to a cell, see remark 12.1.1). Considering all possible realisations of the random tessellation, we get a distribution. This distribution does not coincide with the one weighted in number because the origin has more chance to belong to a large cell than to a small one. The distribution thus obtained is called *weighted in volume*.

The relationship between both types of distributions has been established by Miles (1961):

**Theorem 12.1.1.** *Let  $\psi$  be some characteristic of a cell with  $d$ -volume  $v$ . The distribution weighted in number  $F(d\psi, dv)$  and the distribution weighted in volume  $F_0(d\psi, dv)$  are linked by the formulae*

$$F_0(d\psi, dv) = \frac{v}{E\{V\}} F(d\psi, dv) \quad F(d\psi, dv) = \frac{v^{-1}}{E_0\{V^{-1}\}} F_0(d\psi, dv)$$

where the symbols  $E$  and  $E_0$  stand for the mean values in number and in volume.

The distribution weighted in volume is used when the characteristics of the cell containing the origin are of interest, while the distribution weighted in number is used for the statistical characterization of the total population of cells. The cell containing the origin is called the *fundamental cell* and is denoted by  $C_f$ . Although the distribution in number does not refer to a particular cell, it is nonetheless possible to assign its characteristics to a virtual cell, called the *typical cell* and denoted by  $C_t$ .

Ergodic arguments can be used to show that the probability  $\rho(h)$  that two points separated by vector  $h$  belong to the same cell, is related to the geometric covariogram  $K$  of the typical cell:

$$\rho(h) = \frac{K(h)}{K(0)} \quad h \in \mathbb{R}^d$$

## 12.2 Voronoi tessellation

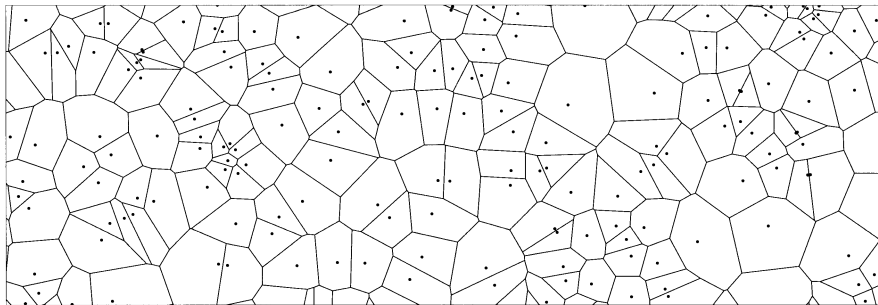
### 12.2.1 Definition and basic properties

Let  $X$  be a Poisson point process with intensity function  $\theta$  in  $\mathbb{R}^d$ . In order to simplify the description, the points of  $X$  are usually called germs. In a Voronoi tessellation, the cells are the zones of influence of the germs of  $X$ .

**Definition 12.2.1.** *The cell  $C(x, X)$  associated with the germ  $g$  of  $X$  is*

$$C(g, X) = \{x \in \mathbb{R}^d \mid |x - g| < \min_{h \in X} |x - h|\}$$

$C(g, X)$  is a convex polytope limited by a set of hyperplanes that are bisectors of pairs of Poisson germs. Figure 12.1 shows an example of a planar Voronoi tessellation.



**Fig. 12.1.** Example of a planar Voronoi tessellation

Now, let  $x$  be an arbitrary point of  $\mathbb{R}^d$ . If  $X \neq \emptyset$ , then  $x$  has almost surely a closest germ. This germ is denoted by  $g(x, X)$ . Therefore we have



$$|x - g(x, X)| = \min_{h \in X} |x - h|$$

Of course  $g(x, X) = x$  if  $x \in X$ . The cell that contains the point  $x \in \mathbb{R}^d$  is  $C[g(x, X), X]$ . It is convenient to denote it by  $C(x, X)$ .

The calculations involving Voronoi tessellations are rather difficult. The first and already impressive results obtained by Meijering (1953) in  $\mathbb{R}^3$  were extended by Miles (1974) and Møller (1989). Using different approaches, both obtain the expected value of the total  $s$ -volume  $F_s$  of the  $s$ -facets of the typical polytope

$$E\{F_s\} = \frac{2^{d-s+1} \pi^{\frac{d-s}{2}} \Gamma\left(\frac{d^2 - ds + s + 1}{2}\right) \Gamma\left(\frac{d}{2} + 1\right)^{d-s+\frac{s}{d}} \Gamma\left(d - s + \frac{s}{d}\right)}{d(d-s)! \theta^{\frac{s}{d}} \Gamma\left(\frac{d^2 - ds + s}{2}\right) \Gamma\left(\frac{d+1}{2}\right)^{d-s} \Gamma\left(\frac{s+1}{2}\right)}$$

In particular

$$E\{|C_t|\} = \frac{1}{\theta} \quad E\{|\partial C_t|\} = \frac{2\sqrt{\pi}(d-1)! \Gamma\left(\frac{d}{2} + 1\right)^{\frac{d-1}{d}} \Gamma\left(1 + \frac{d-1}{d}\right)}{\theta^{\frac{d-1}{d}} \Gamma\left(\frac{d+1}{2}\right) \Gamma\left(d - \frac{1}{2}\right)}$$

To the best of the author’s knowledge, little is known about the characteristics of the fundamental cell. Even its  $d$ -volume has no explicit formula. By conditioning w.r.t. the germ  $g$  nearest to the origin, we readily get

$$E\{|C_f|\} = \theta \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{-\theta[B(o, |g|) \cup B(x, |x - g|)]} dx dg$$

where  $o$  denotes the origin and  $B(u, r)$  stands for the ball with centre  $u$  and radius  $r$ .

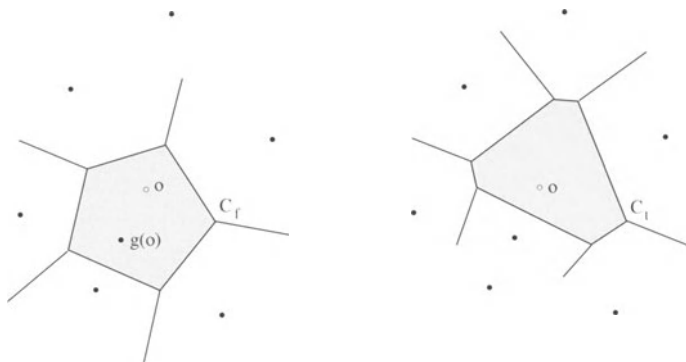
### 12.2.2 Simulation

**Fundamental and typical cells.** Here the tessellation is supposed to be stationary. As was said earlier, the fundamental cell is the cell that contains the origin. Miles and Maillardet (1982) observed that if a germ is added to the Poisson point process at the origin, then the cell of germ  $o$  thus obtained has exactly the same statistical properties as the typical cell. To summarize

$$C_f \stackrel{\mathcal{D}}{=} C(o, X) \quad C_t \stackrel{\mathcal{D}}{=} C(o, X \cup o)$$

Both cells can be explicitly constructed by generating the Poisson germs one after the other further and further from the origin. Beyond a certain

distance, both cells can no longer be modified by the generation of further germs. Figure 12.2 shows a simulation result. As can be seen, the explicit construction of the fundamental and the typical cells may not require the same germs of  $X$ .



**Fig. 12.2.** Simulation of a fundamental and a typical cell associated with a stationary planar Voronoi tessellation

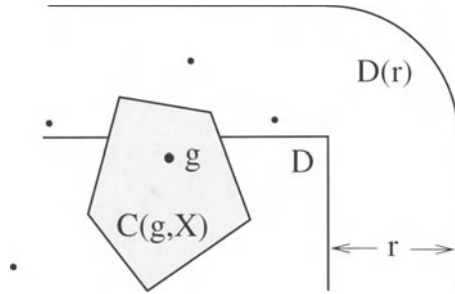
**Non conditional simulation.** The problem addressed here is the non conditional simulation of a tessellation with intensity function  $\theta$  within the convex field  $D \subset \mathbb{R}^d$ . The main difficulty is to handle the edge-effects. One possible way is to generate the Poisson germs one after the other further and further from  $D$  till all the cells hitting  $D$  can no longer be modified by the generation of further germs. In the following algorithm,  $D(r)$  denotes the set of points at a distance less than  $r$  from  $D$ . We have also assumed that we know how to construct the Voronoi diagram of a finite population of germs. Various procedures are available. Boissonnat and Yvinec (1998) give an iterative one, while Preparata and Shamos (1985) propose a recursive one.

**Algorithm 12.2.1.** (*Voronoi tessellation*)

1. Set  $r = 0$ . Simulate a Poisson point process  $X$  with intensity function  $\theta 1_{\in D}$  (cf. algorithm 11.1.2). If  $X = \emptyset$ , then goto 3.
2. If  $|y - g| < \inf_{z \in \partial D(r)} |y - z|$  for any  $g \in X$  such that  $C(g, X) \cap D \neq \emptyset$  and for any  $y \in C(g, X)$ , then deliver  $X$ .  $\square$
3. Generate  $U \sim \text{Unif}$ , and determine  $s > r$  such that  $\theta[D(s)] - \theta[D(r)] = -\ln U$ . Generate  $x \sim \theta 1_{\in \partial D(s)}$ .
4. Set  $r = s$  and  $X = X \cup x$ . Goto 2.

How can we run step 2? Let  $C(g, X)$  be a cell with germ  $g$  hitting  $D$ . A classical result states that  $C(g, X)$  is bounded if and only if  $g$  is contained in the interior of the convex hull of  $X \setminus g$ . In the case when it is bounded,  $C(g, X)$

is totally contained in the convex hull of its vertices. Then because  $D(r)$  is convex, the condition of step 2 is satisfied as soon as  $|v-g| < \inf_{z \in \partial D(r)} |v-z|$  for any vertex  $v$  of  $C(g, X)$  (see Figure 12.3).



**Fig. 12.3.** If the cell has its vertices closer to its germ than to the boundary of the dilated simulation domain  $D(r)$ , then its shape is not affected by the addition of germs outside  $D(r)$

**Remark 12.2.1.** Suppose that  $\theta$  is constant. If  $D = \{o\}$ , then the cell  $C(o, X)$  delivered by algorithm 12.2.1 is precisely the fundamental cell  $C_f$ . To get a typical cell,  $C(o, X)$  can also be considered, but step 1 of the algorithm must be replaced by "Set  $r = 0$  and  $X = \{o\}$ ".

**Conditional simulation.** We now turn to the problem of simulating the random function  $Z$  built by assigning independent values to the cells of a Voronoi tessellation (see section 12.1.2). Let  $\theta$  be the Voronoi intensity function, and let  $F$  be the valuation distribution. Some conditional data are available, namely  $(Z(c) = z(c), c \in C)$ .

In order to avoid the complications resulting from the introduction of a bounded simulation field, we assume here that the Poisson point process has an integrable intensity function

$$\int_{\mathbb{R}^d} \theta(x) dx = \vartheta < \infty$$

Firstly, it should be noted that the way to account for the conditioning data points depends on the type of valuation distribution considered:

- If the distribution is discrete, then different cells may have the same value, so two conditioning data points with the same value may belong to two different cells. In that case, the valuation distribution does affect the geometry of the cells. For instance, two points with the same value have all the more chance to belong to the same cell when their common value has a low frequency of occurrence.

– If the distribution is continuous, then different cells almost surely take different values, so that two conditioning data points with the same value necessarily belong to the same cell. Consequently, the valuation distribution does not affect the geometry of the cells.

Let us start by giving an iterative algorithm for simulating a Poisson population of germs with independent values. It is based on the Barker variation of the Metropolis algorithm (see algorithm 8.2.1 and example 8.2.1).

**Algorithm 12.2.2.** (*Voronoi tessellation. Iterative algorithm*)

1. Set  $X = \emptyset$ .
2. Generate a random variable  $U$  that takes the values  $+1, -1$  and  $0$  with probabilities

$$p_{+1} = \frac{\vartheta}{\vartheta + \#X + 1} \quad p_{-1} = \frac{\#X}{\vartheta + \#X} \quad p_0 = 1 - p_{+1} - p_{-1}$$

- 3.1. If  $U = +1$ , then generate  $x \sim \theta$  and  $z(x) \sim F$ . Set  $X = X \cup \{x\}$ .
- 3.2. If  $U = -1$ , then generate  $x \sim Unif(X)$ . Set  $X = X \setminus \{x\}$ .
4. Goto 2.

This non conditional algorithm describes the evolution of a Markov chain that is reversible (this property is inherited from the Metropolis algorithm). Moreover, let  $\Omega_c$  be the set of all allowed populations of valued germs. In the discrete case, it can be observed that  $\Omega_c$  is stable under union (because two conditioning data points with the same valuation do not have to belong to the same cell). This is sufficient to ensure that irreducibility is preserved if algorithm 12.2.2 is restricted to  $\Omega_c$ . Accordingly, the restriction technique of section 8.4.1 is applicable, and the algorithm 12.2.2 can be made conditional by requiring the conditions to be satisfied at each iteration. Here is the algorithm in which the notations  $X_x = X \cup \{x\}$  and  $X^x = X \setminus \{x\}$  are used for short:

**Algorithm 12.2.3.** (*Conditional Voronoi tessellation. Discrete case*)

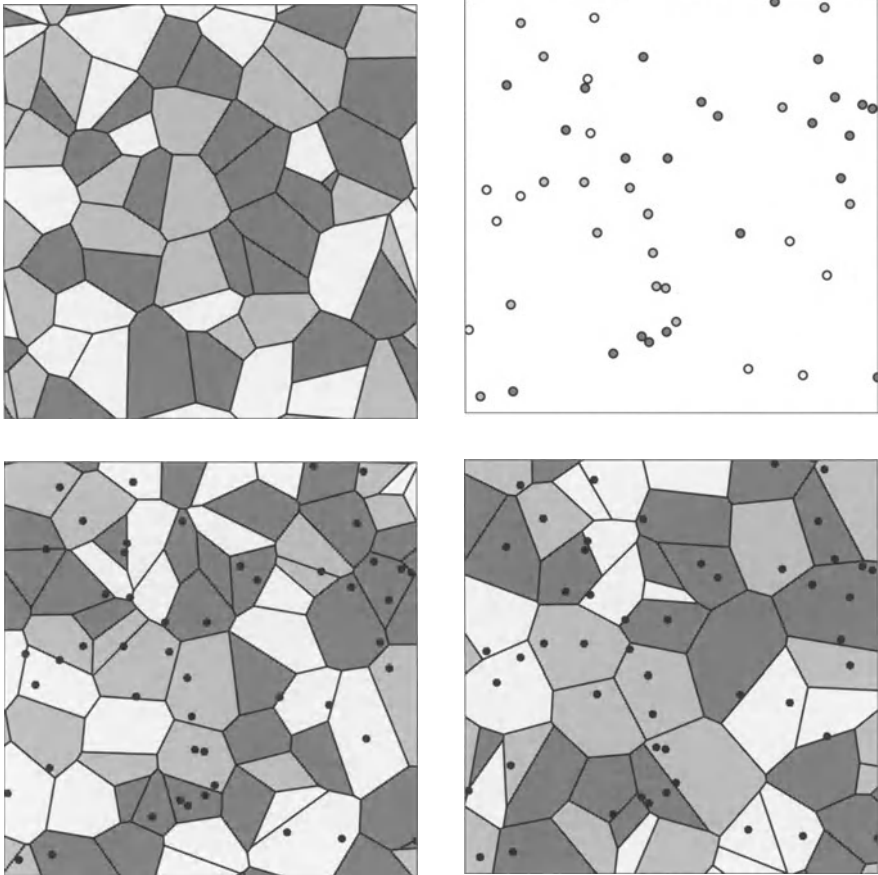
1. Set  $X = C$ .
2. Generate a random variable  $U$  that takes the values  $+1, -1$  and  $0$  with probabilities

$$p_{+1} = \frac{\vartheta}{\vartheta + \#X + 1} \quad p_{-1} = \frac{\#X}{\vartheta + \#X} \quad p_0 = 1 - p_{+1} - p_{-1}$$

- 3.1. If  $U = +1$ , then generate  $x \sim \theta$  and  $z(x) \sim F$ . If  $z(c) = z(x)$  for every  $c \in C \cap C(x, X_x)$ , then set  $X = X_x$ .
- 3.2. If  $U = -1$ , then generate  $x \sim Unif(X)$ . If  $z(c) = z(g(c, X^x))$  for every  $c \in C \cap C(x, X)$ , then set  $X = X^x$ .

*4. Goto 2.*

This algorithm has been tested on a tessellation with intensity function  $\theta = 12.5$  on the square  $[-2, 2] \times [-2, 2]$  and with a discrete uniform valuation distribution (3 possible valuations). Figure 12.4 illustrates the results obtained in the simulation field  $[-1, 1] \times [-1, 1]$ . From a non conditional simulation (top left), 50 data points have been uniformly and independently selected (top right). They have been used as conditioning data to produce two conditional simulations that are depicted in the second row. Both simulations have been stopped after 8000 iterations.



**Fig. 12.4.** Conditional simulation of a Voronoi tessellation with discrete values. Top left, a non-conditional simulation. Top right, the conditioning data set. Second row, two conditional simulations

The choice of a halting criterion must be based on the rate of convergence of the algorithm. One possible way to determine it is to compare the spatial distribution  $F_n$  of the tessellation produced at the  $n^{\text{th}}$  iteration and the conditional spatial distribution  $F_\infty$  that we want to simulate. Without entering into details, it can be said that this rate of convergence is dictated by the evolution of the number of germs of the tessellation. This number evolves according to a Markov chain with compact Jacobi transition kernel (see example 9.3.1). Let  $\lambda$  be its greatest eigenvalue with modulus strictly less than 1, it can be established that there exists a positive constant  $\alpha$  such that

$$|F_n(\mathcal{X}) - F_\infty(\mathcal{X})| \leq \alpha \lambda^n$$

for any event  $\mathcal{X}$  affecting the spatial distribution. As was seen in section 9.4, the value of  $\lambda$  can be obtained by the integral range technique. Explicitly, we find  $\lambda = 0.998$ , which corresponds to an integral range of about 1100 iterations.

The conditional simulation algorithm for the continuous case differs from the discrete one in three respects:

– Firstly, the initialization. Let us write  $C_i = z^{-1}(z_i)$  for each value  $z_i$  taken by  $z$  on  $C$ . The family  $(C_i, i \in I)$  constitutes a partition of  $C$ . The initialization problem amounts to assigning a point  $x_i \in \mathbb{R}^d$  to each component  $C_i$  in such a way that

$$\max_{c \in C_i} |x_i - c| < \min_{c \in C \setminus C_i} |x_i - c|$$

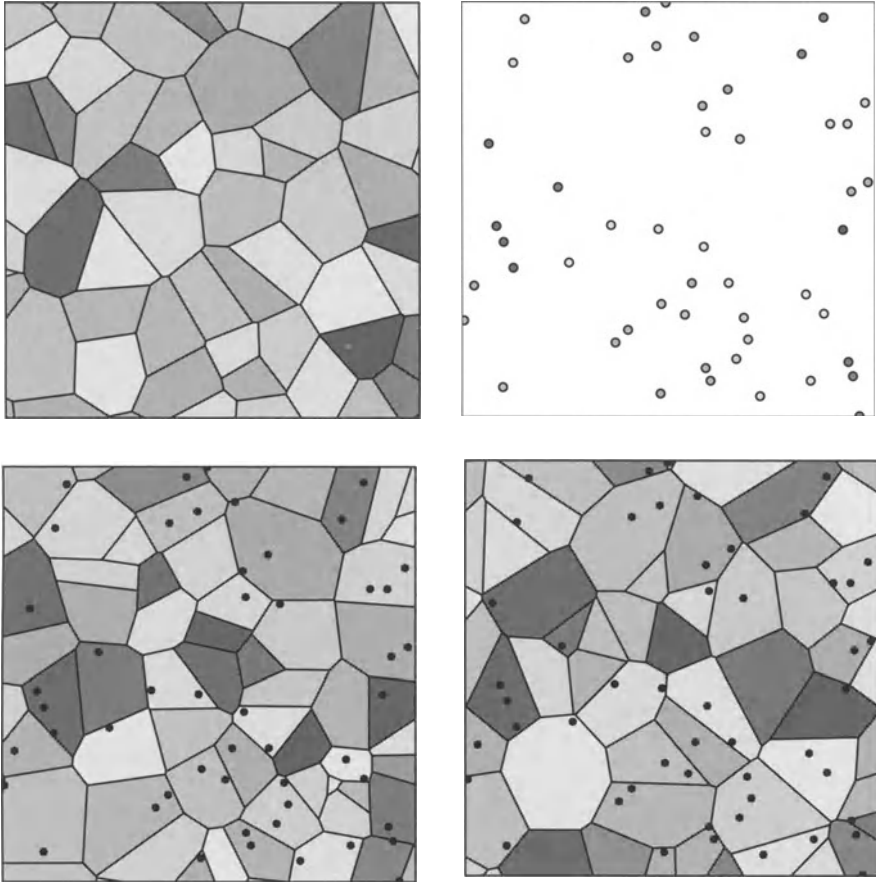
This can be achieved by picking  $x_i$  in the domain

$$X_i = \bigcap_{c \in C_i} C(c, C \setminus C_i)$$

Of course, if one of the  $X_i$ 's is empty, then no such point can be selected, which means that the conditioning data are not compatible with the tessellation model. This is exactly what happens if the convex hulls of several  $C_i$ 's overlap. As a result of the initialization, we put  $X_0 = \{x_i, i \in I\}$ .

– Because the cell geometry is independent of the valuation distribution, a valuation does not have to be generated each time a new germ is inserted. It suffices to simulate the germ values once the simulation of the cell geometry has been completed.

– Let  $\Omega_c$  be the set of all allowed populations of germs (in contrast to the discrete case, the germs are not assigned a value). Note that  $\Omega_c$  is *not* stable under union (two different cells cannot contain equivalued germs), which makes the irreducibility of the restricted Markov chain questionable. In order to cope with that difficulty, it is convenient to introduce a third operation



**Fig. 12.5.** Conditional simulation of a Voronoi tessellation with continuous values. Top left, a non-conditional simulation. Top right, the conditioning data set. Second row, two conditional simulations

beside the *insertion* of a new germ and the *removal* of an old one. This operation consists of replacing an old germ by a new one. Indeed, the *replacement* is just a combination of an insertion and a removal but both are carried out in one step.

These three differences lead to the following algorithm. It is presented using the short notations  $X_x = X \cup \{x\}$ ,  $X^x = X \setminus \{x\}$  and  $X_x^y = (X \cup \{x\}) \setminus \{y\}$ .

**Algorithm 12.2.4.** (*Conditional Voronoi tessellation. Continuous case*)

1. Set  $X = C$ .
2. Generate a random variable  $U$  that takes the values  $+1$ ,  $-1$  and  $0$  with probabilities

$$p_{+1} = \frac{1}{2} \frac{\vartheta}{\vartheta + \#X + 1} \quad p_{-1} = \frac{1}{2} \frac{\#X}{\vartheta + \#X} \quad p_0 = 1 - p_{+1} - p_{-1}$$

- 3.1. If  $U = +1$ , then generate  $x \sim \theta$ . If  $C \cap C(x, X_x) = \emptyset$ , or if there exists  $g \in X$  such that  $C \cap C(x, X_x) = C \cap C(g, X)$ , then set  $X = X_x$ .
  - 3.2. If  $U = 0$ , then generate  $x \sim \theta$  and  $y \sim \text{Unif}(X)$ . If  $C \cap C(x, X_x^y) = \emptyset$ , or if there exists  $g \in X$  such that  $C \cap C(x, X_x^y) = C \cap C(g, X)$ , and if  $C \cap C(y, X) = \emptyset$ , or if there exists  $g \in X_x^y$  such that  $C \cap C(g, X_x^y) = C \cap C(y, X)$ , then set  $X = X_x^y$ .
  - 3.3. If  $U = -1$ , then generate  $y \sim \text{Unif}(X)$ . If  $C \cap C(y, X) = \emptyset$ , or if there exists  $g \in X$  such that  $C \cap C(g, X^y) = C \cap C(x, X)$ , then set  $X = X^y$ .
4. Goto 2.

In order to show how this algorithm works, we have considered the same tessellation model as in the discrete case, except that the valuation distribution is now uniform over  $]0, 1[$ . The two conditional simulations at the bottom of Figure 12.5 were stopped after 10000 iterations. The rate of convergence of this algorithm has been determined using the integral range technique. It turns out to be slightly slower than in the discrete case. The critical eigen value has been found equal to 0.9984, corresponding to an integral range of 1200 iterations.

## 12.3 Poisson tessellation

### 12.3.1 Definition and basic properties

A hyperplane in  $\mathbb{R}^d$  is specified by two parameters, namely a *direction* parameter  $\alpha$  and a *location* parameter  $p$ :

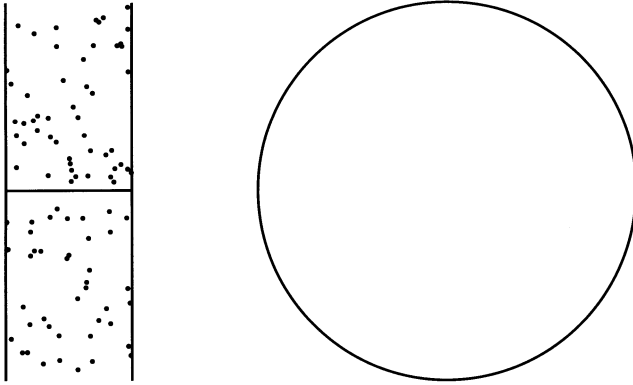
$$H(\alpha, p) = \{x \in \mathbb{R}^d \mid \langle x, \alpha \rangle = p\}$$

Notice that  $H(\alpha, p) = H(-\alpha, -p)$ . In order to avoid a double parametrization of the hyperplanes, it is common to take  $p \in \mathbb{R}$  and  $\alpha$  in a hemisphere, for instance in  $S_d^+$ , consisting of the points in  $S_d$  with positive  $d$ -coordinate. Using this parametrization, a hyperplane is just a point in  $S_d^+ \times \mathbb{R}$ .

**Definition 12.3.1.** *A network of Poisson hyperplanes is parametrized by a Poisson point process in  $S_d^+ \times \mathbb{R}$ .*

A Poisson hyperplane network is completely determined by one parameter, namely the intensity of the Poisson point process in the parameter space. It is called the *hyperplane intensity* and is denoted by  $\theta$ . In what follows,  $\theta$  is supposed to be a positive and locally integrable function. The case when  $\theta$  is constant is also worthwhile considering because it leads to interesting simplifications.





**Fig. 12.6.** Realization of a Poisson line network represented in the parameter space  $[0, \pi[ \times \mathbb{R}$  (left) and in Euclidean space (right). The polygons that the lines delineate are the Poisson polygons

**Definition 12.3.2.** *The polytopes delimited by a network of Poisson hyperplanes are called Poisson cells.*

The Poisson cells are convex as intersection of half-spaces. Their statistical properties depend only on a single parameter, the hyperplane intensity  $\theta$ .

Many results about the moments of the typical polytopes are available. Miles (1961, 1974) obtained the expected value of the total content of its  $s$ -facets (number of vertices and total edge length in  $\mathbb{R}^2$ , number of vertices, total edge length and total boundary area in  $\mathbb{R}^3$ )

$$E\{F_s\} = \binom{d}{s} \frac{2^d}{\omega_s} \frac{1}{(\omega_{d-1}\theta)^s} \quad 0 \leq s \leq d$$

while Matheron (1975) gave the mean value of its Minkowski functionals

$$E\{W_i(C_t)\} = \frac{\omega_i}{\omega_{d-i}} \left( \frac{2}{\omega_{d-1}\theta} \right)^{d-i} \quad 0 \leq i \leq d$$

Of course, both sets of formulae provide the same value for the mean  $d$ -volume of the typical cell and for the mean  $d - 1$ -volume of its boundary:

$$E\{|C_t|\} = \frac{1}{\omega_d} \left( \frac{2}{\omega_{d-1}\theta} \right)^d \quad E\{|\partial C_t|\} = \frac{d 2^d}{\omega_{d-1}} \frac{1}{(\omega_{d-1}\theta)^{d-1}}$$

Miles (1974) obtained a formula for the mean number of  $s$ -facets (number of vertices and edges in  $\mathbb{R}^2$ , number of vertices, edges and faces in  $\mathbb{R}^3$ ) of a typical Poisson cell

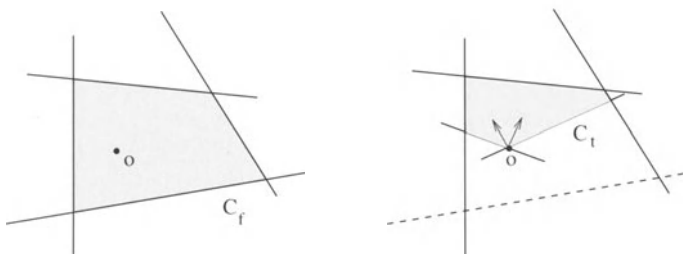
$$E\{N_s\} = 2^{d-s} \binom{d}{s} \quad 0 \leq s \leq d$$

In contrast to this, little is known about the moments of the fundamental Poisson cell. The first three moments of its volume were obtained by Mathéron (1975). In particular

$$E\{|C_f|\} = \frac{d! \omega_d}{(\omega_{d-1} \theta)^d}$$

### 12.3.2 Simulation

**Fundamental and typical cell.** In order to simulate the fundamental cell of a stationary Poisson tessellation, it suffices to generate the Poisson hyperplanes one after the other in decreasing order of proximity to the origin. The procedure is continued until the generation of additional hyperplanes no longer affects the cell containing the origin.



**Fig. 12.7.** Simulation of a fundamental and a typical cell associated with a stationary planar Poisson tessellation

The simulation of a typical cell is more tricky (see Figure 12.7). One possible way is to take the intersection between the fundamental cell and the domain limited by  $d$  random hyperplanes passing through the origin and with direction joint density function

$$f(\alpha_1, \dots, \alpha_d) \propto |\alpha_1, \dots, \alpha_d| \quad \alpha_1, \dots, \alpha_d \in S_d^+$$

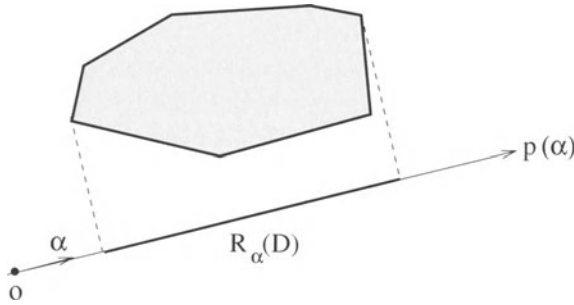
where  $|\alpha_1, \dots, \alpha_d|$  denotes the  $d$ -volume of the parallelotope with edges  $\alpha_1, \dots, \alpha_d$  (Miles (1974)). In 2D, this p.d.f. becomes

$$f(\alpha_1, \alpha_2) = \frac{1}{2\pi} \sin |\alpha_1 - \alpha_2| \quad \alpha_1, \alpha_2 \in [0, \pi[$$

If  $d$  is not too large, then  $f$  can be simulated using the acceptance-rejection method (see algorithm 7.2.2).

**Non conditional simulation.** The simulation of a Poisson network of hyperplanes crossing the convex field  $D$  amounts to the simulation of a Poisson point process in the domain  $R(D) \in S_d^+ \times \mathbb{R}$  associated with  $D$  in the parameter space. Algorithm 11.1.2 can be applied provided that we know the mean number of points in  $R(D)$ , that is the mean number of hyperplanes crossing  $D$ .

Consider a hyperplane with direction  $\alpha$ . It crosses  $D$  if and only if its location parameter belongs to a certain subset  $R_\alpha(D)$  of  $\mathbb{R}$  (cf. Figure 12.8), so that the domain  $R(D)$  can be written  $R(D) = \cup_{\alpha \in S_d^+} R_\alpha(D)$ .



**Fig. 12.8.** Specification of the location parameter of the hyperplanes of direction  $\alpha$  crossing  $D$

If  $\theta$  is constant, then the mean number of points of  $R(D)$  is

$$\theta[R(D)] = \theta \int_{S_d^+} |R_\alpha(D)| d\alpha$$

Because  $D$  is convex, the length of  $R_\alpha(D)$  is exactly the breadth of  $D$  in direction  $\alpha$  and therefore

$$\theta[R(D)] = \theta \frac{d\omega_d}{2} b(D)$$

where  $b(D)$  stands for the mean breadth of  $D$ .

In order to assign independent values to the cells, they must be identified individually starting from the hyperplane network. For this, we refer to Boissonat and Yvinec (1998) and references therein.

**Conditional simulation.** The problem is now to simulate a Poisson tessellation with hyperplane intensity  $\theta$  and valuation distribution  $F$  in the convex field  $D$ , given  $Z(c) = z(c)$  for all  $c \in C$ .

Because of the parametrization of the hyperplanes considered, this problem amounts to a conditional simulation of a Poisson point process with intensity

function  $\theta$  in  $R(D)$ . The algorithm developed in section 11.2 is not applicable because the conditions imposed are different.

Here too, the conditional simulation algorithm depends on whether the valuation distribution is discrete or continuous. Both cases will be considered in turn.

Let us start with the discrete case. The cells take their values in the finite or countable set  $\{z_j, j \in J\}$ . For each  $j \in J$ ,  $p_j$  is the frequency of occurrence of  $z_j$ . Once again, the idea is to design a conditional simulation algorithm by modifying a non conditional iterative algorithm. But things are more complicated for the Poisson tessellation because the insertion or the removal of a hyperplane may lead to the creation or the disappearance of more than one cell.

Let  $X = \{x_1, \dots, x_{n(X)}\}$  be a population of points that is a realization of the Poisson point process in  $R(D)$ . Its probability density function is

$$f(X) \propto \prod_{i=1}^{n(X)} \theta(x_i) 1_{x_i \in R(D)}$$

Let  $\Omega_c$  be the set of all allowed populations, and let  $n_j(X)$  be the number of cells of  $X$  containing conditioning points of  $C_j = \{c \in C | z(c) = z_j\}$ . Then the conditional distribution of the population  $X$  can be written as

$$f_c(X) \propto \prod_{i=1}^{n(X)} \theta(x_i) 1_{x_i \in R(D)} \prod_{j \in J} p_j^{n_j(X)} 1_{X \in \Omega_c}$$

This expression leads to the following algorithm that is a Metropolis algorithm with a twofold acceptance criterion. The first one specifies the number of points of the population. The second one validates the insertion of a new point or the removal of an old one in accordance with the shape of the valuation distribution. In this algorithm  $\vartheta$  is used as an abbreviation for the mean number of Poisson points of  $R(D)$ . This algorithm also assumes the existence of an initial population  $X_0$ , that can be any population of points associated with hyperplanes that separate the conditioning data points with different values (for instance, the bisectors separating the nearest points of all disjoint components  $C_j$  and  $C_k$  of  $C$ ). This initial population  $X_0$  does exist with positive probability provided that the convex hulls of the  $C_j$ 's are pairwise disjoint.

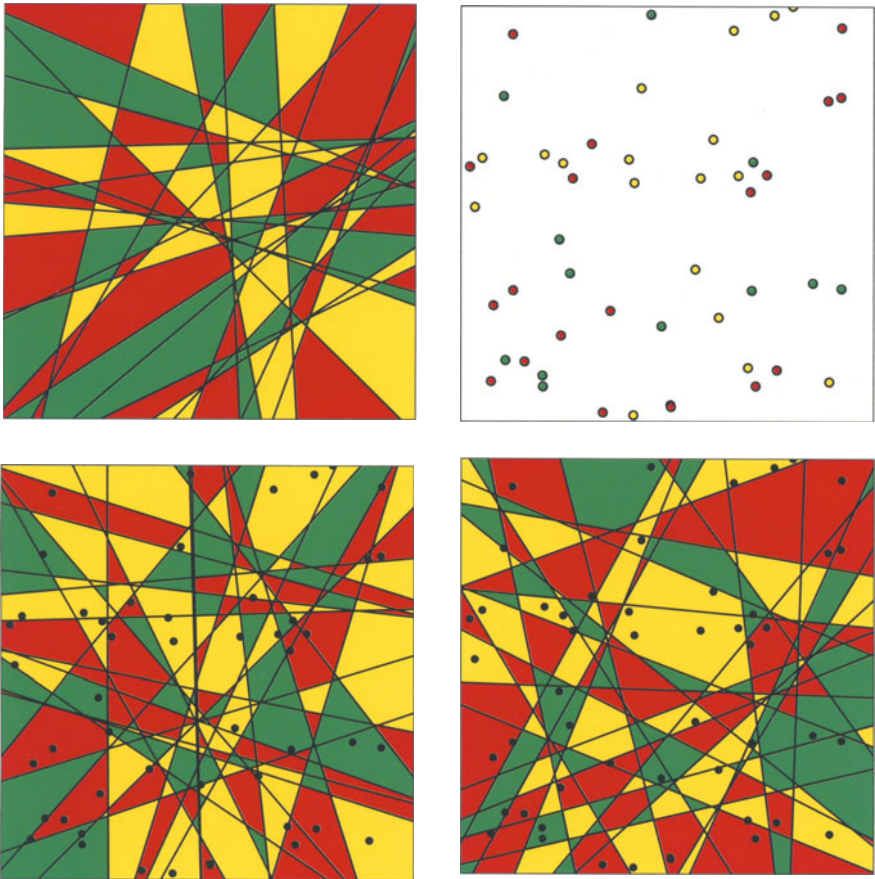
**Algorithm 12.3.1.** (*Conditional Poisson tessellation. Discrete case*)

1. Set  $X = X_0$ .
2. Generate a random variable  $U$  that takes the values  $+1, -1$  and  $0$  with probabilities

$$p_{+1} = \frac{\vartheta}{\vartheta + \#X + 1} \quad p_{-1} = \frac{\#X}{\vartheta + \#X} \quad p_0 = 1 - p_{+1} - p_{-1}$$

- 3.1. If  $U = +1$ , then generate  $x \sim \theta$  in  $R(D)$ . Set  $X = X \cup x$  with probability  $\min(\prod_{j \in J} p_j^{n_j(X \cup x) - n_j(X)}, 1)$ .
- 3.2. If  $U = -1$ , then generate  $x \sim \text{Unif}(X)$ . If the hyperplanes of  $X \setminus x$  separate all the components  $C_j$  of  $C$ , then set  $X = X \setminus x$  with probability  $\min(\prod_{j \in J} p_j^{n_j(X \setminus x) - n_j(X)}, 1)$ .
4. Goto 2.

**Remark 12.3.1.** This algorithm can also be applied to carry out a conditional simulation of a Voronoi tessellation with a discrete valuation distribution. Indeed, the acceptance criterion is the same as in algorithm 12.2.3. The only difference is that no value is assigned to the cells with no conditioning data points.



**Fig. 12.9.** Conditional simulation of a Poisson tessellation with discrete values. Top left, a non-conditional simulation. Top right, the conditioning data set. Second row, two conditional simulations

In order to illustrate algorithm 12.3.1, we have considered a stationary Poisson line network with 20 lines crossing the simulation field on average. The valuation distribution consists of three equiprobable values. The two conditional simulations at the bottom of Figure 12.9 have been obtained after running 2000 iterations. In this example, the rate of convergence is fast because the different components of  $C$  can be separated easily by random lines. Using algorithm 9.4.1, the integral range obtained is about 90 iterations, which corresponds to a critical eigenvalue of 0.98.

Let us now turn to the case of a Poisson tessellation with continuous valuation distribution. This case is much simpler because the values and the geometries of the cells are independent (as for a Voronoi tessellation with continuous values). It suffices to generate the points of  $R(D)$  iteratively and accept the insertion of a new point or the removal of an old one provided that all conditioning points with the same value must lie in the same cell, and that all pairs of conditioning points with different values are separated by a hyperplane. The proposed algorithm is basically the same as algorithm 12.2.4, except for two differences:

- The initialization procedure. In order to get an allowed population  $X_0$ , one possible way is to generate a sequence of independent points in  $R(D)$  (for instance with density  $\theta$ , but this is not strictly necessary). Let  $x$  be such a point, and let  $H(x)$  be its corresponding hyperplane. The point  $x$  is discarded if  $H(x)$  separates conditioning points with the same value, and kept if not. The procedure is continued till all pairs of conditioning points with different values are separated by hyperplanes. In order to avoid having too many points in the initial step, only those whose hyperplane is the first one to separate conditioning points with different values can be kept.
- Algorithm 12.2.4 works with a replacement step. In the case of the Poisson tessellation, this step is possible but not required, because the set of allowed populations of points of  $R(D)$  is stable under union, which makes the procedure automatically irreducible.

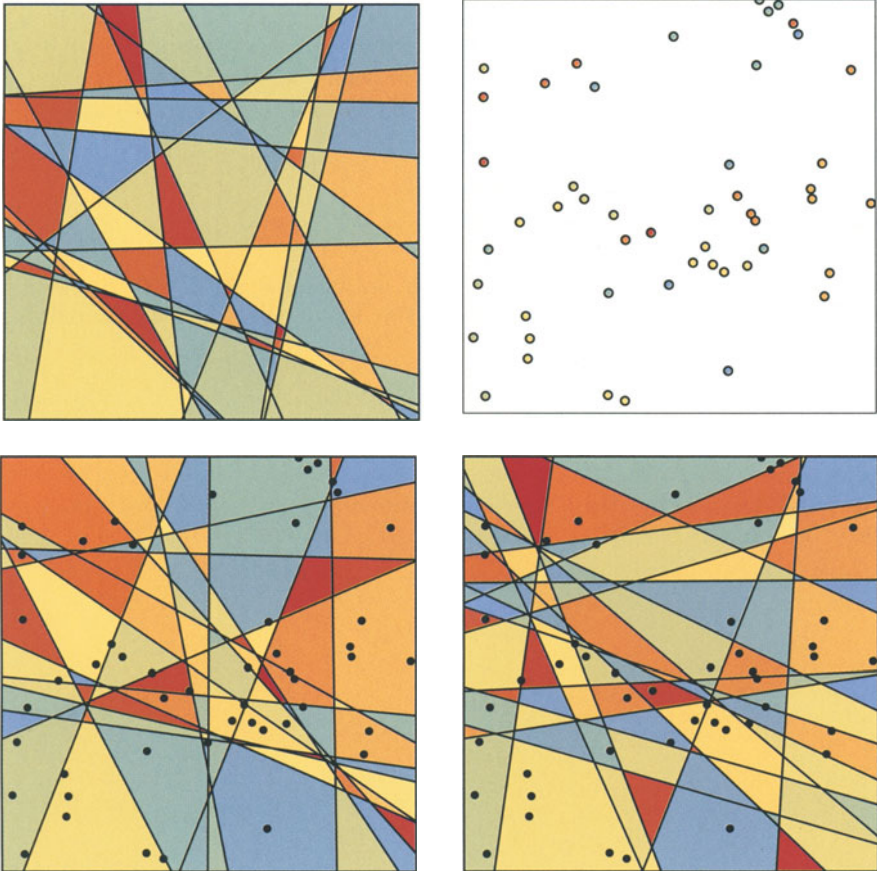
So here is the algorithm, in which the phrase "component of  $C$ " refers to any subset consisting of all points of  $C$  with the same value.

**Algorithm 12.3.2.** (*Conditional Poisson tessellation. Continuous case*)

1. Set  $X = X_0$ .
2. Generate a random variable  $U$  that takes the values  $+1, -1$  and  $0$  with probabilities

$$p_{+1} = \frac{\vartheta}{\vartheta + \#X + 1} \quad p_{-1} = \frac{\#X}{\vartheta + \#X} \quad p_0 = 1 - p_{+1} - p_{-1}$$

- 3.1. If  $U = +1$ , then generate  $x \sim \theta$  in  $R(D)$ . If  $H(x)$  does not split any component of  $C$ , then set  $X = X \cup x$ .
- 3.2. If  $U = -1$ , then generate  $x \sim \text{Unif}(X)$ . If the hyperplanes of  $X \setminus x$  separate all components of  $C$ , then set  $X = X \setminus x$ .
4. Goto 2.



**Fig. 12.10.** Conditional simulation of a Poisson tessellation with continuous values. Top left, a non-conditional simulation. Top right, the conditioning data set. Second row, two conditional simulations

Some results provided by this algorithm are depicted in Figure 12.10. The model considered is stationary with 20 lines crossing the simulation field on average, and the valuation distribution is uniform over  $]0, 1[$ . Fifty points have been uniformly and independently selected from a non conditional simulation to serve as conditioning data points. Despite the fact that several

distant points have the same value (see the 5 points in the yellow cell at the bottom), the algorithm converges reasonably fast. The critical eigenvalue is approximately 0.99, which corresponds to an integral range of about 200 iterations. The algorithm has been stopped after 2000 iterations.

## Exercises

**12.1** Consider a stationary tessellation with discrete valuation distribution. Given that the tessellation takes the same value at points  $x$  and  $y$ , what is the probability that both points belong to the same cell?

**13.2** Let  $X$  be a stationary planar Poisson tessellation with line intensity  $\theta$ , and let  $D$  be a convex compact subset.

1) Show that the number of lines hitting  $D$  is Poisson distributed with mean  $\theta |\partial D|$ .

2) Let  $D'$  be another convex compact subset included within  $D$ . Show that given one line of the process crosses  $D$ , this line crosses  $D'$  with probability  $|\partial D'| / |\partial D|$ .

**12.3** What is the geometric covariogram of the typical cell of a stationary Poisson tessellation in  $\mathbb{R}^d$  with hyperplane intensity  $\theta$ ?

**12.4** Show that the intersection of a stationary planar line process with intensity  $\theta$  and an arbitrary line is a stationary point process with intensity  $2\theta$ .



## 13. Boolean model

Random sets can be constructed by placing independent random objects at Poisson points and by taking their union. This simple idea underpins the definition of the boolean model. Despite its simplicity, it is methodologically and practically important, and has been the object of a considerable work (Matheron, 1967, 1975; Serra, 1982; Stoyan et al., 1987; Hall, 1988; Molchanov, 1997).

This chapter considers only boolean models that are random closed sets. First, the avoiding functional of a boolean model, stationary or not, is calculated (the simplifications that arise in the stationary case are reported in the appendix). From this, the algebraic and stereological stability properties of the model are easily established. Finally, an iterative algorithm for the conditional simulation of the boolean model is proposed.

### 13.1 Definition and basic properties

#### 13.1.1 Definition

Two basic and independent ingredients are required for the construction of a boolean model in  $\mathbb{R}^d$

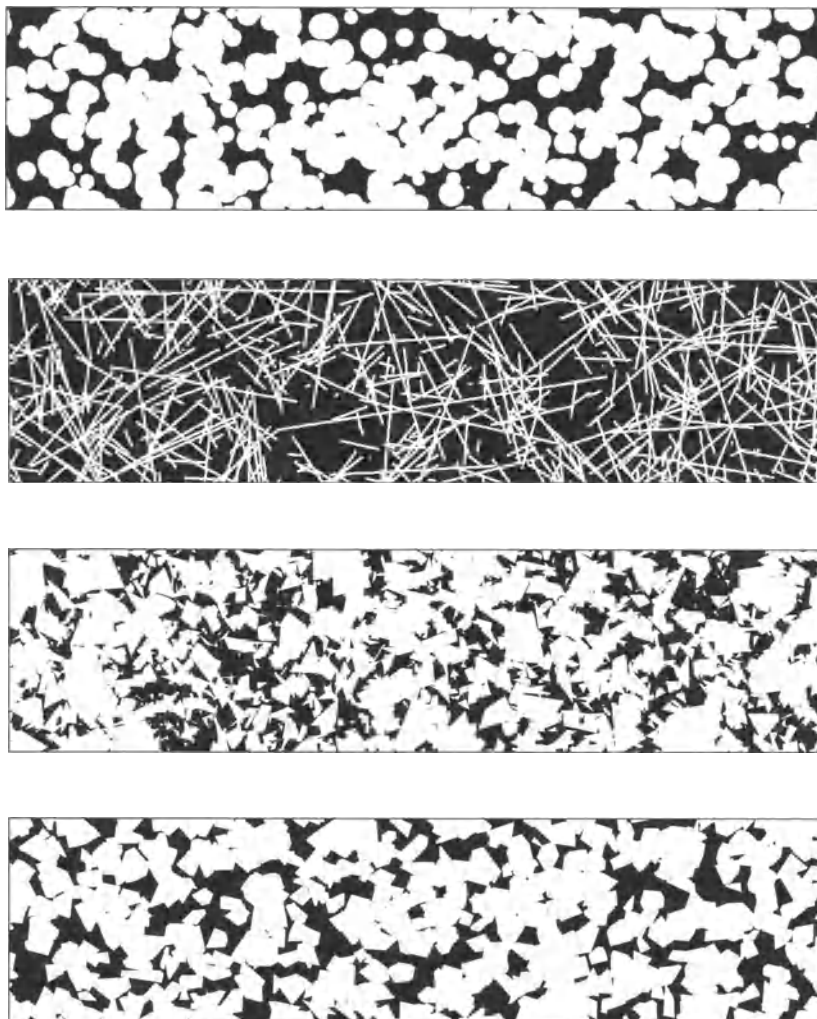
- i) a set of *germs*, that is, a Poisson point process  $\mathcal{P}$  with intensity function  $\theta = (\theta(x), x \in \mathbb{R}^d)$ .
- ii) a family  $(A(x), x \in \mathbb{R}^d)$  of independent random non-empty compact subsets of  $\mathbb{R}^d$ . The subset  $A(x)$  is called the *object implanted at  $x$*  and its hitting functional is denoted by  $T_x$ .

**Definition 13.1.1.** *A boolean model  $X$  is the union of all the objects implanted at the Poisson germs*

$$X = \bigcup_{x \in \mathcal{P}} A(x)$$

Four realizations of boolean models are displayed in Figure 13.1. Their appearance depends considerably on the shape of the objects. From top to bottom, these are uniform planar cross-sections of balls with constant radius,

segments of lines, typical Poisson polygons (see section 12.3.2), and typical Voronoi polygons (see section 12.2.2).



**Fig. 13.1.** Realizations of four boolean models. From top to bottom, the objects are random disks, segments of lines, Poisson polygons and Voronoi polygons

Because a boolean model is the union of possibly infinitely many objects, it is not guaranteed to be closed. This will certainly be the case if each point  $x \in \mathbb{R}^d$  has a neighbourhood hit only by finitely many objects a.s., and in particular if the population of objects  $(A(x), x \in \mathcal{P})$  is of finite order (cf.

section 2.5). More precisely, let  $N(K)$  be the number of objects hitting the compact subset  $K$

$$N(K) = \sum_{x \in \mathcal{P}} 1_{K \cap A(x) \neq \emptyset}$$

A simple calculation shows that the mean of  $N(K)$  is

$$\vartheta(K) = \int_{\mathbb{R}^d} \theta(x) T_x(K) dx \leq \infty$$

**Definition 13.1.2.** *The boolean model is said to be of finite order if*

$$\vartheta(K) < \infty \quad \forall K \in \mathcal{K}$$

For the rest of this chapter the boolean models under consideration are assumed to be of finite order.

### 13.1.2 Avoiding functional

Let us start by determining the distribution of the number of objects hitting the compact subset  $K$ .

**Proposition 13.1.1.**  *$N(K)$  is Poisson distributed with mean  $\vartheta(K)$ .*

**Proof:** For each  $D \in \mathcal{K}$ , let us consider the number  $N_D(K)$  of objects implanted in  $D$  and hitting  $K$

$$N_D(K) = \sum_{x \in \mathcal{P} \cap D} 1_{K \cap A(x) \neq \emptyset}$$

By definition 11.1.2, the number of points of  $\mathcal{P} \cap D$  follows a Poisson distribution with mean

$$\theta(D) = \int_D \theta(x) dx$$

Suppose that this number is equal to  $n$ . According to theorem 11.1.2, the  $n$  points are independently located within  $D$  with the same p.d.f.  $\theta(\cdot)/\theta(D)$ . Moreover, an object implanted at  $x$  hits  $K$  with the probability  $T_x(K)$  and avoids it with the complementary probability  $1 - T_x(K)$ . Consequently the generating function of  $N_D(K)$  is

$$E \left\{ s^{N_D(K)} \right\} = \sum_{n=0}^{+\infty} e^{-\theta(D)} \frac{(\theta(D))^n}{n!} \left( \int_D \frac{\theta(x)}{\theta(D)} [1 - T_x(K) + sT_x(K)] dx \right)^n$$

where  $0 \leq s \leq 1$ , and by summation

$$E \left\{ s^{N_D(K)} \right\} = \exp \left\{ (s - 1) \int_D \theta(x) T_x(K) dx \right\}$$

In order to extend this result to the whole space, let  $(D_n, n \in \mathbb{N})$  be an increasing sequence of compact sets covering  $\mathbb{R}^d$ . Then  $(N_{D_n}(K), n \in \mathbb{N})$  is also increasing and converges a.s. to  $N(K)$ . Therefore

$$E \left\{ s^{N(K)} \right\} = \lim_{n \rightarrow \infty} E \left\{ s^{N_{D_n}(K)} \right\} = \exp \left\{ (s-1) \int_{\mathbb{R}^d} \theta(x) T_x(K) dx \right\}$$

and the right hand-side can be recognized as the generating function of a Poisson distribution with mean  $\vartheta(K)$ .  $\square$

Because  $P\{X \cap K = \emptyset\} = P\{N(K) = 0\}$ , we deduce

**Corollary 13.1.1.** *The avoiding functional of the boolean model is*

$$P\{X \cap K = \emptyset\} = e^{-\vartheta(K)} \quad K \in \mathcal{K}$$

### 13.1.3 Stability properties

**Proposition 13.1.2.** *The following stability properties are satisfied*

- i) *the union of two independent boolean models is a boolean model.*
- ii) *a boolean model dilated by a non-empty compact subset of  $\mathbb{R}^d$  is a boolean model.*
- iii) *the intersection between a boolean model and a compact subset of  $\mathbb{R}^d$  is a boolean model.*
- iv) *the cross-section of a boolean model by an  $i$ -flat is a boolean model.*

**Proof:** Let  $X'$  and  $X''$  be two independent boolean models with intensity functions  $\theta'$  and  $\theta''$ , and object hitting functionals  $T'$  and  $T''$ . Because of the independence and corollary 13.1.1

$$\begin{aligned} P\{(X' \cup X'') \cap K = \emptyset\} &= P\{X' \cap K = \emptyset\} P\{X'' \cap K = \emptyset\} \\ &= \exp \left\{ - \int_{\mathbb{R}^d} [\theta'(x)T'_x(K) + \theta''(x)T''_x(K)] dx \right\} \\ &= \exp \left\{ - \int_{\mathbb{R}^d} \theta(x) T_x(K) dx \right\} \end{aligned}$$

with  $\theta = \theta' + \theta''$  and  $T_x$  defined as

$$T_x(K) = \frac{\theta'(x)}{\theta'(x) + \theta''(x)} T'_x(K) + \frac{\theta''(x)}{\theta'(x) + \theta''(x)} T''_x(K) \quad K \in \mathcal{K}$$

This is the hitting functional of a random object equal to  $A'(x)$  with probability  $\theta'(x)/(\theta'(x) + \theta''(x))$  and to  $A''(x)$  with the complementary probability  $\theta''(x)/(\theta'(x) + \theta''(x))$ . This proves i).

Let  $D$  be a nonempty compact subset of  $\mathbb{R}^d$ . ii) directly stems from

$$X \oplus \check{D} = \bigcup_{x \in \mathcal{P}} A(x) \oplus \check{D}$$

Regarding iii), one can also write  $X \cap D = \bigcup_{x \in \mathcal{P}} A(x) \cap D$ , but this formula is not easy to interpret because  $A(x) \cap D$  may be almost surely empty. Instead, let us write<sup>1</sup>

$$\begin{aligned} P\{(X \cap D) \cap K = \emptyset\} &= P\{X \cap (D \cap K) = \emptyset\} \\ &= \exp \left\{ - \int_{\mathbb{R}^d} \theta(x) T_x(D \cap K) dx \right\} \\ &= \exp \left\{ - \int_{\mathbb{R}^d} \theta(x) T_x(D) \frac{T_x(D \cap K)}{T_x(D)} dx \right\} \end{aligned}$$

But  $T_x(D \cap K)/T_x(D)$  is the hitting functional of  $A(x) \cap D$  given that this object is nonempty. Consequently,  $X \cap D$  is a boolean model with intensity function  $\theta T_x(D)$  and object hitting functional  $T_x(D \cap K)/T_x(D)$ .

Let  $F$  be an  $i$ -flat of  $\mathbb{R}^d$ . iv) can be deduced from iii) by considering an increasing sequence  $(D_n, n \in \mathbb{N})$  with limit equal to  $F$ .  $\square$

## 13.2 Simulation

The problem is to simulate the boolean model in  $D$ , subject to the conditions that two finite subsets  $C_0$  and  $C_1$  must be contained in  $X^C$  and in  $X$  respectively. This problem is important in the oil industry. Petroleum engineers require the reservoir geometry as input to run fluid flow simulation programs. An algorithm developed by Haldorssen (1983) has been notably improved by Chessa (1995). It consists of independently simulating the sandbodies (i.e. the objects) that intersect the wells and those that do not. Such a dichotomic approach is made possible because of the independence properties of the Poisson point process. The difficulty with this approach is that the distribution of an object intersecting wells depends not only on its implantation, but also on the number and the location of the wells that it intersects. Usually, things become intractable as soon as this number is greater than 1.

The iterative algorithm described below was designed after a private communication with Matheron in 1990. A preliminary work was then carried out by Gedler (1991) under the supervision of the author. Since this conditional simulation algorithm is a modified version of a non conditional simulation algorithm (by restricting a transition kernel, see section 8.4.1), the non conditional simulation algorithm is presented first.

<sup>1</sup> In the last line, the convention  $0/0 = 0$  is required.

### 13.2.1 Non conditional simulation

In the proof of proposition 13.1.2, we showed that  $X \cap D$  is a boolean model with intensity function  $\theta(\cdot)T_x(D)$ . Accordingly, the number of objects of  $X \cap D$  is Poisson distributed with mean

$$\vartheta(D) = \int_{\mathbb{R}^d} \theta(x) T_x(D) dx$$

The hitting functional of an object of  $X \cap D$  implanted at  $x$  is  $T_x(D \cap \cdot)/T_x(D)$ .

**Definition 13.2.1.** *A typical object of  $X \cap D$  is an object uniformly selected among the objects of  $X \cap D$ .*

A typical object of  $X \cap D$  is implanted according to the density function  $\theta(\cdot)T_x(D)$ . Its hitting functional is

$$T(K) = \int_{\mathbb{R}^d} \frac{\theta(x) T_x(D)}{\vartheta(D)} \frac{T_x(D \cap K)}{T_x(D)} dx = \frac{1}{\vartheta(D)} \int_{\mathbb{R}^d} \theta(x) T_x(D \cap K) dx$$

Note in particular that  $T(D) = 1$  and  $T(K) = 0$  if  $K$  is disjoint from  $D$ .

**Proposition 13.2.1.**  *$X \cap D$  has the same distribution as an union of  $N$  independent typical objects where  $N$  is Poisson distributed (mean  $\vartheta(D)$ ).*

**Proof:** Let  $Y$  be such a union. Its avoiding functional is

$$\begin{aligned} P\{Y \cap K = \emptyset\} &= \sum_{n=0}^{\infty} e^{-\vartheta(D)} \frac{\vartheta(D)^n}{n!} [1 - T(K)]^n \\ &= \exp\{-\vartheta(D) T(K)\} = \exp\left\{-\int_{\mathbb{R}^d} \theta(x) T_x(D \cap K) dx\right\} \end{aligned}$$

which is exactly the avoiding functional of  $X \cap D$ .  $\square$

As a consequence, a non conditional simulation of the boolean model can be obtained using the following algorithm:

**Algorithm 13.2.1.** *(boolean model)*

1. Set  $X = \emptyset$ .
2. Generate  $N \sim \text{Poisson}(\vartheta(D))$ .
3. If  $N = 0$ , then deliver  $X$ .  $\square$
4. Generate a typical object  $A \sim T$ .
5. Set  $X = X \cup A$ ,  $N=N-1$  and goto 3.

This requires the simulation of a typical object. The standard algorithm is

**Algorithm 13.2.2.** *(Typical object)*

1. Generate  $x \sim \theta(\cdot)T_x(D)$ .

2. Generate  $A \sim T_x(D \cap \cdot)/T_x(D)$ .
3. Deliver  $A$ .

If it is difficult to simulate an object implanted at  $x$  and hitting  $D$  directly, a rejection algorithm can be used as an alternative.

**Algorithm 13.2.3.** (*Typical object (rejection technique)*)

1. Generate  $x \sim \theta(\cdot)T_x(D)$ .
2. Generate  $A \sim T_x$ .
3. If  $A \cap D = \emptyset$ , then goto 2.
4. Deliver  $A \cap D$ .

Finally algorithm 13.2.1 is modified slightly to make it iterative. This does not present any difficulty because the Metropolis algorithm can be used to simulate a Poisson distribution (cf. example 8.2.1). The following algorithm simulates a boolean model iteratively. More precisely, it simulates the population of objects that constitutes the boolean model. In this algorithm, the number of objects of population  $\Phi$  is denoted by  $\#\Phi$ .

**Algorithm 13.2.4.** (*Population of objects of a boolean model*)

1. Set  $\Phi = \emptyset$ .
2. Generate a random variable  $U$  that takes the values  $+1, -1$  and  $0$  with probabilities

$$p_{+1} = \frac{\vartheta(D)}{\vartheta(D) + \#\Phi + 1} \quad p_{-1} = \frac{\#\Phi}{\vartheta(D) + \#\Phi} \quad p_0 = 1 - p_{+1} - p_{-1}$$

- 3.1. If  $U = +1$ , then generate  $A \sim T$  and set  $\Phi = \Phi \cup A$ .
- 3.2. If  $U = -1$ , then generate  $A \sim \text{Unif}(\Phi)$  and set  $\Phi = \Phi \setminus A$ .
4. Goto 2.

### 13.2.2 Conditional simulation

Let us start with two remarks. Firstly, the Markov chain in algorithm 13.2.4 is reversible (this property is inherited from Metropolis algorithm). Secondly, the restriction of this Markov chain to the set  $\Omega_c$  of all the populations of objects that honour the conditions on  $C_0$  and  $C_1$  is irreducible. This stems from the fact that  $\Omega_c$  is stable under concatenation<sup>2</sup>. As a consequence, the restriction technique of section 8.4.1 is applicable, and the algorithm 13.2.4 can be made conditional by requiring the conditions to be satisfied at each iteration. Here is the algorithm in question:

**Algorithm 13.2.5.** (*Conditional boolean model*)

<sup>2</sup> If  $\Phi, \Psi \in \Omega_c$ , then the population  $\Phi + \Psi$  made of all objects of  $\Phi$  and all objects of  $\Psi$  is also an element of  $\Omega_c$ .

1. Generate  $\Phi \in \Omega_c$ .
2. Generate a random variable  $U$  that takes the values  $+1, -1$  and  $0$  with probabilities

$$p_{+1} = \frac{\vartheta(D)}{\vartheta(D) + \#\Phi + 1} \quad p_{-1} = \frac{\#\Phi}{\vartheta(D) + \#\Phi} \quad p_0 = 1 - p_{+1} - p_{-1}$$

- 3.1. If  $U = +1$ , then generate  $A \sim T$ . If  $A \cap C_0 = \emptyset$ , then set  $\Phi = \Phi \cup A$ .
- 3.2. If  $U = -1$ , then generate  $A \sim \text{Unif}(\Phi)$ . If  $C_1 \subset \Phi \setminus A$ , then set  $\Phi = \Phi \setminus A$ .
4. Goto 2.

Of course, this algorithm must start with an allowed population  $\Phi \in \Omega_c$ . One possible way to get one is to simulate a sequence of independent typical objects. Each time an object hits  $C_0$ , it is automatically discarded. The procedure is continued until the remaining objects completely cover  $C_1$ . In order to avoid starting with too many objects, it is recommended to keep only the objects that are the first ones to cover points of  $C_1$ . We therefore have the following algorithm:

**Algorithm 13.2.6.** (*Initialization of the conditional boolean model*)

1. Set  $\Phi = \emptyset$  and  $C = C_1$ .
2. Generate  $A \sim T$ .
3. If  $A \cap C_0 \neq \emptyset$  or if  $A \cap C = \emptyset$ , then goto 2.
4. Set  $\Phi = \Phi \cup A$  and  $C = C \setminus A$ .
5. If  $C \neq \emptyset$ , then goto 2.
6. Deliver  $\Phi$ .

Let  $N_g$  be the number of objects generated to complete this procedure. In the case when  $C_1 \neq \emptyset$ , a simple calculation shows that the mean value of  $N_g$  is

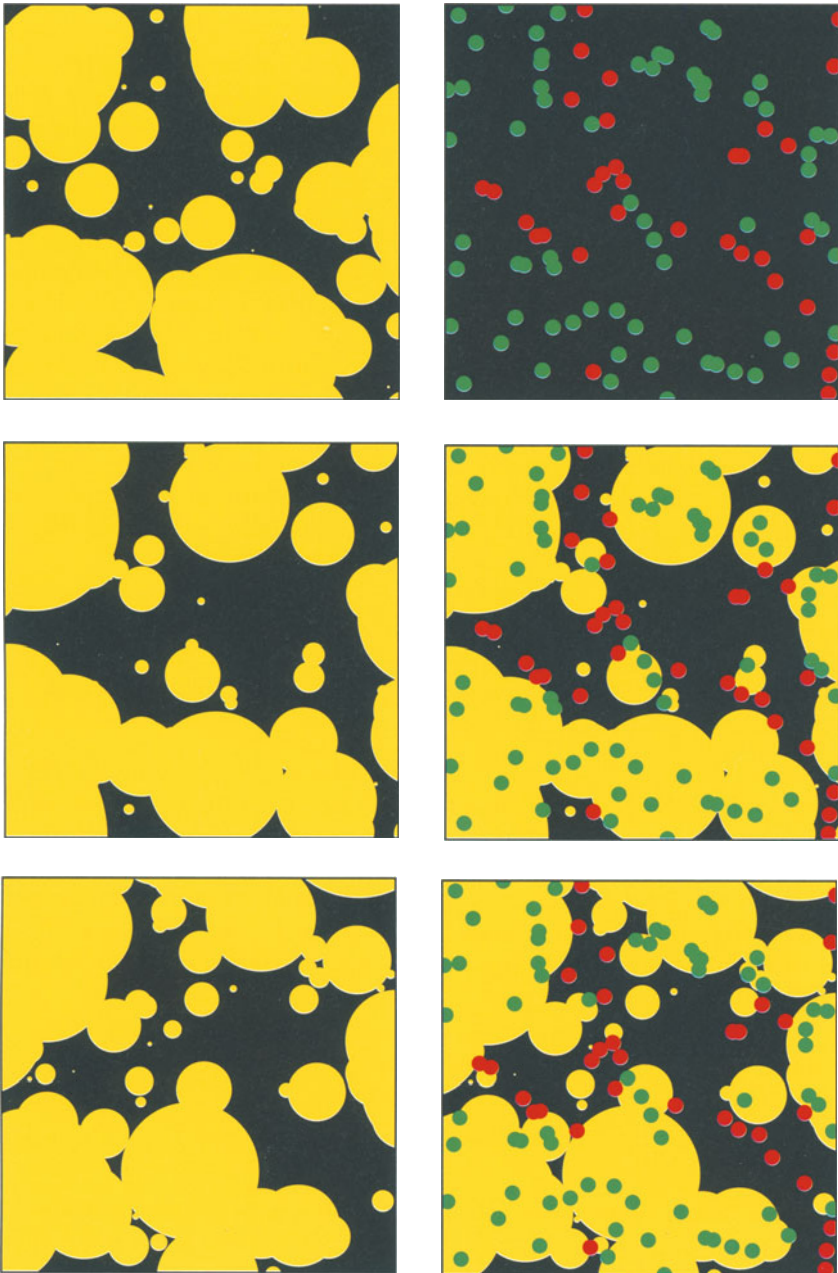
$$E\{N_g\} = \sum_{\substack{C \subset C_1 \\ C \neq \emptyset}} (-1)^{|C|+1} \frac{1}{T(C_0 \cup C) - T(C_0)}$$

This mean value is finite if and only if  $P\{C_0 \subset (X \cap D)^C, C_1 \subset X \cap D\} > 0^3$ .

<sup>3</sup> Clearly,  $E\{N_g\} < \infty$  if and only if  $T(C_0 \cup \{c\}) - T(C_0) > 0$  for every  $c \in C_1$ . In order to exploit these inequalities, let us write  $X \cap D$  as the union of  $\#\mathcal{C}_1$  i.i.d. boolean models  $(X(c), c \in C_1)$ . Their common intensity function is  $\theta(\cdot)T(D)/\#\mathcal{C}_1$ . Note that  $X \cap D$  contains  $C_1$  as soon as each  $X(c)$  contains  $c$ . Accordingly

$$\begin{aligned} P\{C_0 \subset (X \cap D)^C, C_1 \subset X \cap D\} &\geq \prod_{c \in C_1} P\{C_0 \cap X(c) = \emptyset, C_1 \subset X(c)\} \\ &= \exp\left\{-\frac{\vartheta(D)}{\#\mathcal{C}_1}T(C_0)\right\} - \exp\left\{-\frac{\vartheta(D)}{\#\mathcal{C}_1}T(C_0 \cup \{c\})\right\} > 0 \end{aligned}$$





**Fig. 13.2.** Conditional simulation of a boolean model. Top left, a non-conditional simulation. Top right, the conditioning data set. Second and third row, two conditional simulations, depicted without (left) and with (right) the conditioning data points

Figure 13.2 displays an example of conditional simulation. The boolean model considered here is stationary (intensity  $\theta = 0.00766$ ) and the objects are disks with exponential radius (mean value 5). The simulation field is  $100 \times 100$ . One hundred data points (top right) have been uniformly selected from a non conditional simulation (top left). They are used as conditioning data to produce two conditional simulations (cf. second and third rows) that are depicted without (left) and with (cf. right) the conditioning data points. Both simulations have been stopped after 2000 iterations.

The choice of a stopping criterion depends on the rate of convergence of the algorithm. One possible way to determine it is to compare the avoiding functional  $Q_n$  of the random set  $X_n$  produced at the  $n^{\text{th}}$  iteration and the conditional avoiding functional  $Q_\infty$  that we want to simulate. Without going into details, an argument similar to that of Lantuéjoul (1997) can be used. It exploits the fact that the number of objects of  $X_n$  evolves according to a Markov chain with a compact Jacobi transition kernel (see example 9.3.1). Denoting by  $\chi$  the normed eigenvector associated with the eigenvalue  $\lambda$  that has the greatest modulus strictly less than 1, it can be established that

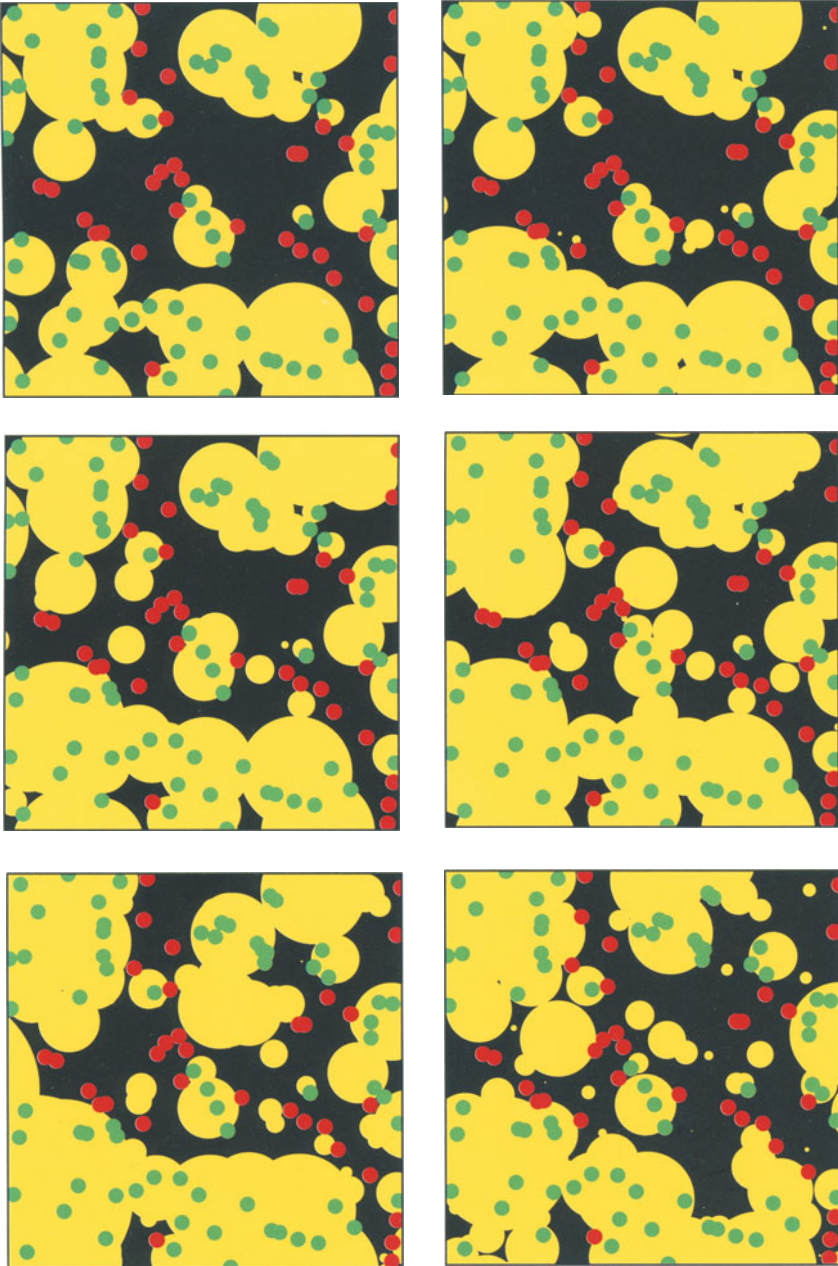
$$|Q_n(K) - Q_\infty(K)| \leq 2\lambda^n E\{\chi(\#\Phi_0)\}$$

where  $\Phi_0$  stands for the initial population of objects. As was seen in section 9.4, the value of  $\lambda$  can be obtained by the integral range technique. Explicitly, we find  $\lambda = 0.993$ , which corresponds to an integral range of about 320 iterations.

Figure 13.3 displays the output of the algorithm after 0, 100, 200, 400, 800 and 1600 iterations. Even if it honours all the conditions, the initial output (top left) is not a conditional simulation of the boolean model because the Poisson intensity has not been accounted for. In particular the number of objects is far too small (43 instead of 93 on the average). This number increases gradually and seems to stabilize after about 400 iterations (63, 76, 102, 101, 91). The output obtained at that step (right of second row) looks already very similar to a conditional simulation.

**Remark 13.2.1.** Kendall and Thönnnes (1998) succeeded in designing an exact conditional simulation algorithm in the case when the objects are bounded.

**Remark 13.2.2.** Algorithm 13.2.5 can be applied when the constraints are more general than  $C_0$  and  $C_1$ . As was seen earlier (see section 8.4.1), this algorithm can work provided that the set of allowed states  $\Omega_c$  is such that i)  $P\{\Omega_c\} > 0$  and ii) the non conditional Markov chain restricted to  $\Omega_c$  is irreducible. This remark is useful for reservoir engineers who want to incorporate more information than just well data in their simulations (dynamic data, assumptions from seismic data or even geological interpretations).



**Fig. 13.3.** Running the conditional simulation of a boolean model. From top to bottom and left to right, the simulations at iterations 0, 100, 200, 400, 800 and 1600

**Remark 13.2.3.** The boolean model is a special case of object-based models. A first extension consists of replacing the Poisson point process that specifies the location of the objects by a spatial birth-and-death process (Preston, 1977; Stoyan et al., 1987). It is also possible to lift the independence assumption between the objects. In that case, an object is inserted or removed as a function of the objects present (and not only their numbers). This leads to the concept of spatial birth-and-death process of objects. A typical example is the pairwise interaction model considered by Sylverseen and Omre (1994). The first version of this model included a fixed number of objects, which allowed its simulation using a Gibbs sampler. This restriction was removed in a subsequent paper (1997). Many other extensions are possible, such as the Markov jump process of objects.

## Appendix: The stationary case

In this appendix, we consider the case when

- i) the intensity function  $\theta$  is constant.
- ii) all objects are identically distributed up to a translation. To put things differently, let  $A$  denote the random object implanted at the origin. Then  $A(x)$  has the same hitting functional as  $A$  shifted to  $x$ :

$$P\{A(x) \cap K \neq \emptyset\} = P\{A_x \cap K \neq \emptyset\} = P\{A \cap K_{-x} \neq \emptyset\} = T(K_{-x})$$

Using these new assumptions, the avoiding functional of  $X$  becomes

$$P\{X \cap K = \emptyset\} = \exp \left\{ -\theta \int_{\mathbb{R}^d} T(K_{-x}) dx \right\}$$

But

$$\begin{aligned} \int_{\mathbb{R}^d} T(K_{-x}) dx &= E \left\{ \int_{\mathbb{R}^d} 1_{A \cap K_{-x} \neq \emptyset} dx \right\} \\ &= E \left\{ \int_{\mathbb{R}^d} 1_{-x \in A \oplus \tilde{K}} dx \right\} = E \{|A \oplus \tilde{K}|\} \end{aligned}$$

and finally

**Proposition 13.2.2.** *The avoiding functional of a stationary boolean model with intensity  $\theta$  and object  $A$  is*

$$P\{X \cap K = \emptyset\} = e^{-\theta E\{|A \oplus \tilde{K}|\}} \quad K \in \mathcal{K}$$

Let us consider some particular cases for  $K$ :

- If  $K = \{x\}$  is reduced to a single point, then we obtain the probability that  $x$  does not belong to any object, that is the background proportion

$$q = e^{-\theta E\{|A|\}}$$

- If  $K = \{x, x+h\}$  is a pair of points, then we obtain the bivariate distribution of the complement  $X^C$  of  $X$ . Since  $|A \oplus \check{K}| = |A \cup A_h| = 2|A| - |A \cap A_h|$ , it is convenient to introduce geometric covariogram  $K_A$  of  $A$  (cf. section 3.1), and we get

$$P\{x \in X^C, x+h \in X^C\} = q^2 e^{\theta K_A(h)}$$

From this, we derive

$$\begin{aligned} P\{x \in X, x+h \in X^C\} &= q - q^2 e^{\theta K_A(h)} \\ P\{x \in X, x+h \in X\} &= 1 - 2q + q^2 e^{\theta K_A(h)} \end{aligned}$$

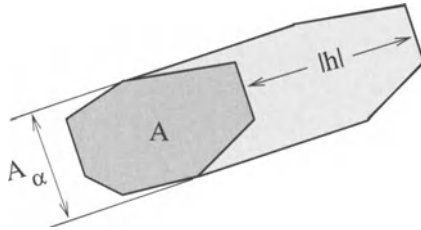


Fig. 13.4. Dilation of a convex set by a segment of line

- If  $K = [x, x+h]$  is a line segment, then  $|A \oplus \check{K}|$  is tractable only if  $A$  is almost surely convex. Let  $|h|$  and  $\alpha$  be the modulus and the direction of  $K$ . We have  $|A \oplus \check{K}| = |A| + |h| |A_\alpha|$ , where  $A_\alpha$  stands for the projected  $(d-1)$ -volume of  $A$  onto a hyperplane orthogonal to  $\alpha$  (see Figure 13.4). Consequently

$$P\{[x, x+h] \subset X^C\} = q e^{-\theta |h| E\{|A_\alpha|\}}$$

Assuming moreover that  $A$  is isotropic (that is its hitting functional is rotation invariant), then Cauchy's formula (cf. section 6.1.2) can be applied and the probability simplifies to

$$P\{[x, x+h] \subset X^C\} = q e^{-\theta |h| \frac{\omega_{d-1}}{d \omega_d} E\{|\partial A|\}}$$

This probability varies as an exponential function of the modulus of  $h$ .

– If  $K = B(x, r)$  is a ball with radius  $r$ , explicit calculations are also possible in the case when the objects are almost surely convex. In this case, Steiner’s formula (cf. section 6.1.2) can be applied

$$E\{|A \oplus \check{K}|\} = \sum_{i=0}^d \binom{d}{i} E\{W_i(A)\} r^i$$

and we get

$$P\{B(x, r) \subset X^C\} = \exp \left\{ -\theta \sum_{i=0}^d \binom{d}{i} E\{W_i(A)\} r^i \right\}$$

These formulae are very useful to check the compatibility of a set of experimental data with a boolean model, as well as to statistically infer its parameters.

### Exercises

**13.1** Let  $X$  be a boolean model in  $\mathbb{R}^d$  with constant intensity  $\theta$  and objects  $A(x) = [x, 2x]$ . Let  $X'$  be another boolean model with intensity  $\theta' = \theta 2^{-d}$  and objects  $A'(x) = [x/2, x]$ . Calculate the avoiding functionals of  $X$  and  $X'$ . What can be concluded about the parametrization of a non-stationary Boolean model?

**13.2** Let  $X$  be a stationary boolean model in  $\mathbb{R}$  with intensity  $\theta$ . The objects of  $X$  can be written  $A(x) = [x, x + L(x)]$  where the  $L(x)$  are independent and exponentially distributed with mean  $a$ .

- 1) What is the probability that both points 0 and 1 belong to  $X$ ?
- 2) What is the probability that an object covers both 0 and 1?
- 3) Suppose that  $A(x)$  covers 0 and 1. What is the distribution of  $L(x)$ ?

## 14. Object based models

There are various ways to construct a random function starting from a population of random objects. Everything depends on the way in which the different objects are combined. In this chapter, three combination rules are successively considered, namely the addition, the maximum and the superposition, which give rise respectively to the *random token model*, the *boolean random function* and the *dead leaves model*.

### 14.1 Random token model

#### 14.1.1 Definition and basic properties

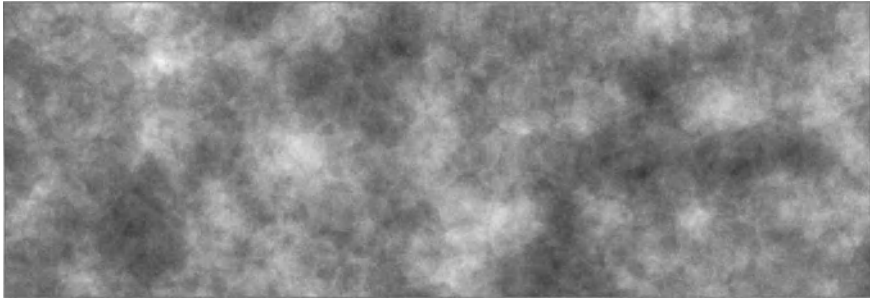
The random token model is built from three independent basic ingredients:

- i) a set of *germs* that consist of points from a stationary Poisson point process  $X$  with intensity  $\theta$  in  $\mathbb{R}^d$ .
- ii) a population of *objects*. This is a family  $(A(x), x \in \mathbb{R}^d)$  of random compact subsets of  $\mathbb{R}^d$ . The  $A(x)$  are independent, and each  $A(x)$  has the same distribution as a reference random compact subset  $A_0$  up to a translation by the vector  $\vec{ox}$ . Moreover  $A_0$  is assumed to have a finite mean  $d$ -volume.
- iii) a family of *weights*  $(\varepsilon(x), x \in \mathbb{R}^d)$  assigned to the objects. These weights are mutually independent. They attain only two values  $\pm 1$  with the same probability  $\frac{1}{2}$ .

**Definition 14.1.1.** *A random token model is a weighted sum of the indicator functions of the objects implanted at the Poisson germs*

$$Z(y) = \sum_{x \in X} \varepsilon(x) 1_{y \in A(x)} \quad y \in \mathbb{R}^d$$

**Remark 14.1.1.** The random token model is a slight extension of a model popularized by Alfaro (1978, 1980) as the *random coin model*, in which the objects are random balls. It is in turn a special case of a more general model introduced by Matérn (1960, 1986) and Serra (1968) and called *dilution random function*. In this model, the indicator functions of the objects are replaced by random functions with integrable transitive covariogram.



**Fig. 14.1.** Realization of a random token model. The objects are random disks (uniform planar cross-sections of balls with fixed diameter, equal to a quarter of the field height). They are barely visible because of their large density. Each point of the simulation is covered by 25 disks on average

The following proposition shows that the spatial distribution of a random token model can be expressed in terms of its Fourier transform:

**Theorem 14.1.1.** *Let  $J$  be a finite family of indices. The Fourier transform of  $(Z(y_j), j \in J)$  satisfies*

$$E \exp \left\{ i \sum_{j \in J} u_j Z(y_j) \right\} = \exp \left\{ \theta \sum_{K \subset J} [\cos u_K - 1] p_K \right\}$$

where

$$u_K = \sum_{k \in K} u_k \quad p_K = \sum_{K \subset L \subset J} (-1)^{\#L \setminus K} |A_0 \ominus \check{Y}_L|$$

and  $Y_L = \{y_l, l \in L\}$ .

Because of its length, the proof is given in the appendix. We now give some particular cases of this formula, depending on the choice of the support set  $Y = (y_j, j \in J)$ .

- If  $Y = \{y\}$ , we obtain the Fourier transform of  $Z(y)$

$$E \left\{ e^{i u Z(y)} \right\} = e^{\theta [\cos u - 1]} E|A_0|$$

Equivalently

$$P\{Z(y) = n\} = e^{-2\theta E|A_0|} I_n(2\theta E|A_0|) \quad n \in \mathbb{Z}$$

where  $I_n$  denotes the modified Bessel function of order  $n$  (Abramowitz and Stegun, 1970).

- If  $Y = \{y, z\}$ , the Fourier transform of  $(Z(y), Z(z))$  can be written using the geometric covariogram  $K$  of  $A_0$



$$E\left\{e^{i[uZ(y) + vZ(z)]}\right\} = e^\theta [\cos u + \cos v - 2][K(0) - K(h)] \\ \times e^\theta [\cos(u + v) - 1]K(h)$$

From this we easily derive that the covariance between  $Z(y)$  and  $Z(z)$  is proportional to the geometric covariogram  $K$  of  $A_0$  at  $y - z$ :

$$\text{Cov}\{Z(y), Z(z)\} = -\frac{\partial^2 \ln \Phi}{\partial u \partial v}(0, 0) = \theta K(y - z)$$

**Remark 14.1.2.** In the case of the dilution random function mentioned in remark 14.1.1, the covariance is proportional to the transitive covariogram of the random functions used for the objects.

### 14.1.2 Simulation

Because the nonconditional simulation of a random token model is similar to that of a boolean model, we directly proceed to the simulation of  $Z$  in a field  $D$  given the conditional data  $Z(c) = z(c), c \in C$ ?

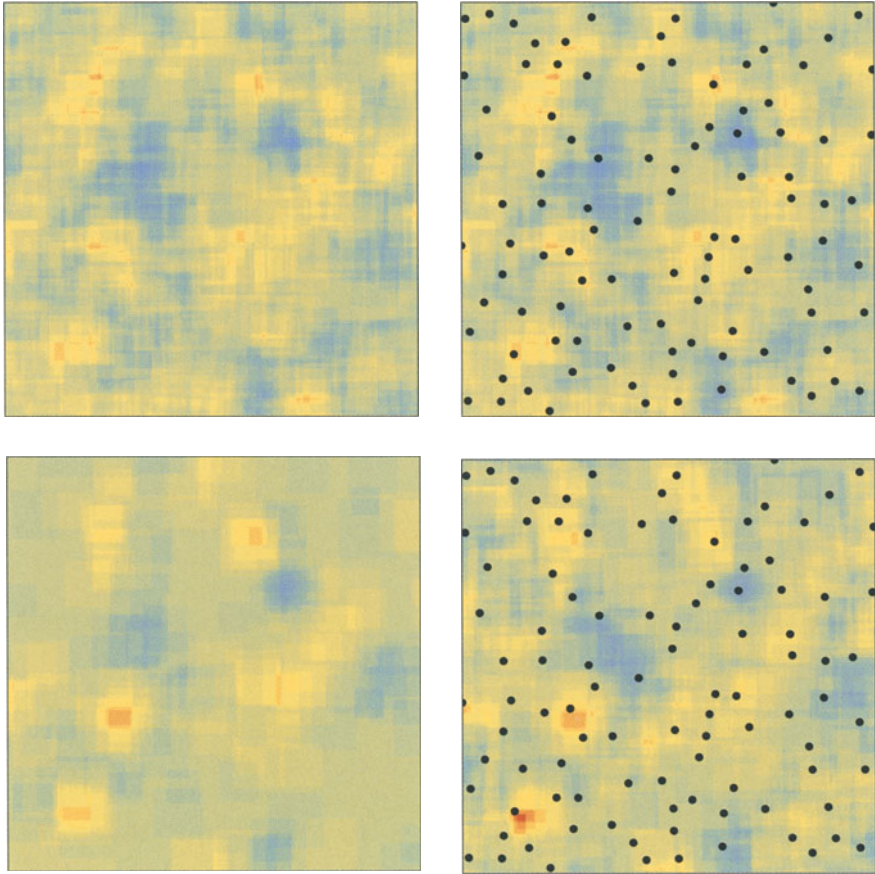
Once again, we are going to propose a procedure that preserves the conditions at each iteration (see section 8.4.1). The first step is to simulate an initial population of objects that is compatible with the data. In the following algorithm, objects are generated one after another in the vicinity of the conditioning data points. Only those are kept that reduce the differences between the values of the random function at the conditioning data points and the conditioning values. In other words, a newly generated object is kept if  $|Z(c) - z(c)|$  does not increase for any  $c \in C$ , and even decreases for some  $c_0 \in C$ . It is discarded otherwise. The procedure stops when all conditions are satisfied.

**Algorithm 14.1.1.** (*Random token model. Initialization of the conditional procedure*)

1. Set  $\mathcal{X}_0 = \emptyset$ .
2. If  $\sum_{c \in C} |z(c)| = 0$ , then deliver  $\mathcal{X}_0$ .  $\square$
3. Select  $c_0 \sim \text{Unif}\{c \in C \mid z(c) \neq 0\}$ , generate an object  $A(x)$  containing  $c_0$ , and put  $\varepsilon(x) = -\text{sign}(z(c_0))$ .
4. If  $|z(c) + \varepsilon(x)| \leq |z(c)|$  for all  $c \in C \cap A(x)$ , then insert  $(A(x), \varepsilon(x))$  to  $\mathcal{X}_0$  and put  $z(c) = z(c) + \varepsilon(x)$  for all  $c \in C \cap A(x)$ .
5. Goto 2.

**Remark 14.1.3.** Instead of step 1, we could have considered starting the procedure with a non conditional simulation. The other steps are unchanged.

This initialization procedure has been tested on a random token model with Poisson intensity  $\theta = 5$  and with square objects whose side length are uniform on  $]0, 1.5[$ . The simulation field is a square  $10 \times 10$  (see top left of Figure 14.2). Fifty conditioning data points have been uniformly selected for this exercise (see top right of Figure 14.2). The simulation of an object containing a given point  $c \in C$  is achieved using the acceptance-rejection method. At each attempt, an object is uniformly implanted within the square with centre  $c$  and side 1.5. In this exercise, 330 iterations were necessary to construct the initial population (see the bottom left of Figure 14.2).



**Fig. 14.2.** Conditional simulation of a random token model. Top left, a non conditional simulation. Top right, the conditioning data points superimposed on the non conditional simulation. Bottom left, the conditional simulation at the initial stage. Bottom right, a conditional simulation

The next step is to design the iterative procedure itself. Note at first that the insertion or the deletion of an object is not possible because they would disrupt the conditioning. The replacement of an object by another one is admissible but not fully satisfactory because a procedure based solely on such replacements would keep the number of objects in the population constant. This is the reason why we resort to a more general form of replacement. It consists of selecting a domain hitting the simulation field, and replacing all the objects implanted in that domain by new ones generated according to the distribution of the non conditional model. The new population of objects thus obtained is accepted provided that all conditions are satisfied.

There is a lot of flexibility with regard to the selection of the domains. The simplest one is to introduce a compact subset  $B$  of  $\mathbb{R}^d$ , and to consider all the translates of  $B$  that hit the simulation field. This leads to the following conditional simulation algorithm:

**Algorithm 14.1.2.** (*Random token model. Conditional simulation*)

1. Set  $\mathcal{X} = \mathcal{X}_0$  (cf. algorithm 14.1.1).
2. Generate  $x \sim \text{Unif}(D \oplus \check{B})$ . Set  $\mathcal{Y} = \{(A(y), \varepsilon(y)) \in \mathcal{X} \mid y \in B_x\}$ .
3. Generate  $N \sim \text{Poisson}(\theta|B|)$ . Generate a population  $\mathcal{Z}$  of  $N$  independent valued objects implanted in  $B_x$ .
4. If  $\sum_{(x \setminus \mathcal{Y}) \cup \mathcal{Z}} \varepsilon(x) 1_{c \in A(x)} = z(c)$  for all  $c \in C$ , then put  $\mathcal{X} = (\mathcal{X} \setminus \mathcal{Y}) \cup \mathcal{Z}$ .
5. Goto 2.

This algorithm has been used to get the conditional simulation displayed at the bottom right of Figure 14.2. In this example  $B$  is a square with side equal to 1.5. Combined with a Poisson intensity  $\theta = 5$ , this gives an average replacement of 11.25 objects. This is a high rate: only 1767 candidate transitions were accepted after running the simulation for 50000 iterations.

**Remark 14.1.4.** In the case when the objects are bounded, Algorithm 14.1.2 can be made faster by considering only the domains located in the vicinity of the conditioning data points. The rest of the simulation is not affected by the conditioning and can be simulated once and for all.

## 14.2 Boolean random function

### 14.2.1 Definition and basic properties

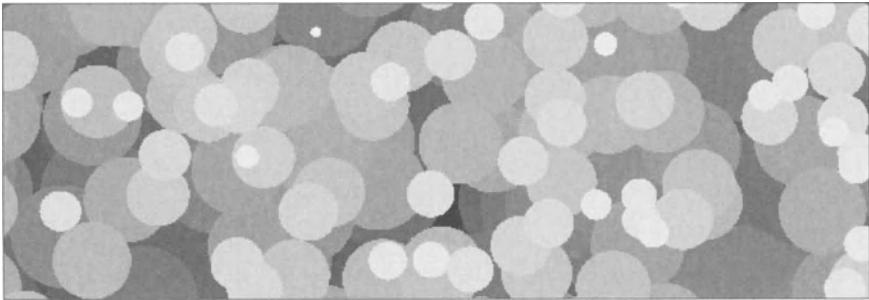
Introduced by Jeulin and Jeulin (1980) for representing rough surfaces, this model has been extensively studied and generalized by several authors (Serra, 1988; Préteux and Schmitt, 1988). The version presented here is a simple variation of the boolean islands (Serra, 1988). The ingredients of this model are

- i) a Poisson point process  $X$  with intensity  $\theta$  in  $\mathbb{R}^d$ .
- ii) a population of objects  $(A(x), x \in \mathbb{R}^d)$ . The objects are independent and have (up to a translation) the same distribution as a reference random compact set  $A_0$  with finite mean  $d$ -volume.
- iii) a family of weights  $(\varepsilon(x), x \in \mathbb{R}^d)$  assigned to the objects. These weights are identically distributed and take positive values  $z_1 < \dots < z_n$  with probability  $p_1, \dots, p_n$ . They are mutually independent, but may depend on the objects.

**Definition 14.2.1.** A boolean random function assigns to each point the maximum weight of the Poisson population of objects that contain it

$$Z(y) = \max_{\substack{x \in X \\ y \in A(x)}} \varepsilon(x) \quad y \in \mathbb{R}^d$$

In this definition we have written maximum and not supremum. Indeed, since the objects have finite mean volume, each point of  $\mathbb{R}^d$  is almost surely covered by a finite number of objects. Figure 14.3 shows a realization of a boolean random function.



**Fig. 14.3.** Realization of a boolean random function of disks

The union of the objects with value  $z_i$  is a stationary boolean model, say  $X_i$ , with intensity  $\theta_i = \theta p_i$ . The distribution of its reference object  $A_i$  coincides with that of  $A_0$  only if the objects and their values are independent. This suggests an alternative definition of a boolean random function:

$$Z(y) = \max_{\{i \mid y \in X_i\}} z_i \quad y \in \mathbb{R}^d$$

From this remark, we find that the maximum of  $Z$  over the compact subset  $K$  of  $\mathbb{R}^d$  is less than  $z$  if  $K$  belongs to the background of all the boolean models  $X_i$  associated with values  $z_i \geq z$ .

**Proposition 14.2.1.** *The distribution of the maximum of the boolean random function over a compact subset  $K$  of  $\mathbb{R}^d$  is given by*

$$P\left\{\max_{y \in K} Z(y) < z\right\} = \prod_{\{i \mid z_i \geq z\}} e^{-\theta_i} E|A_i \oplus \check{K}|$$

A slightly more general argument gives

**Theorem 14.2.1.** *The spatial distribution of the boolean random function is given by*

$$P\left\{\bigwedge_{i=1}^n Z(Y_i) = z_i\right\} = \prod_{i=1}^n P\{Y_i \subset X_i\} \prod_{i=2}^n P\{X_i \cap Y_{1,i-1} = \emptyset\}$$

for any family  $(Y_1, \dots, Y_n)$  of finite and pairwise disjoint subsets of  $\mathbb{R}^d$ , and where  $Y_{1,i-1}$  is a short notation for  $\cup_{j=1}^{i-1} Y_j$ . Explicitly

$$P\{Y_i \subset X_i\} = \sum_{Y \subset Y_i} (-1)^{\#Y} e^{-\theta_i} E|A_i \oplus \check{Y}|$$

$$P\{X_i \cap (\cup_{j=1}^{i-1} Y_j) = \emptyset\} = e^{-\theta_i} E|A_i \oplus \check{Y}_{1,i-1}|$$

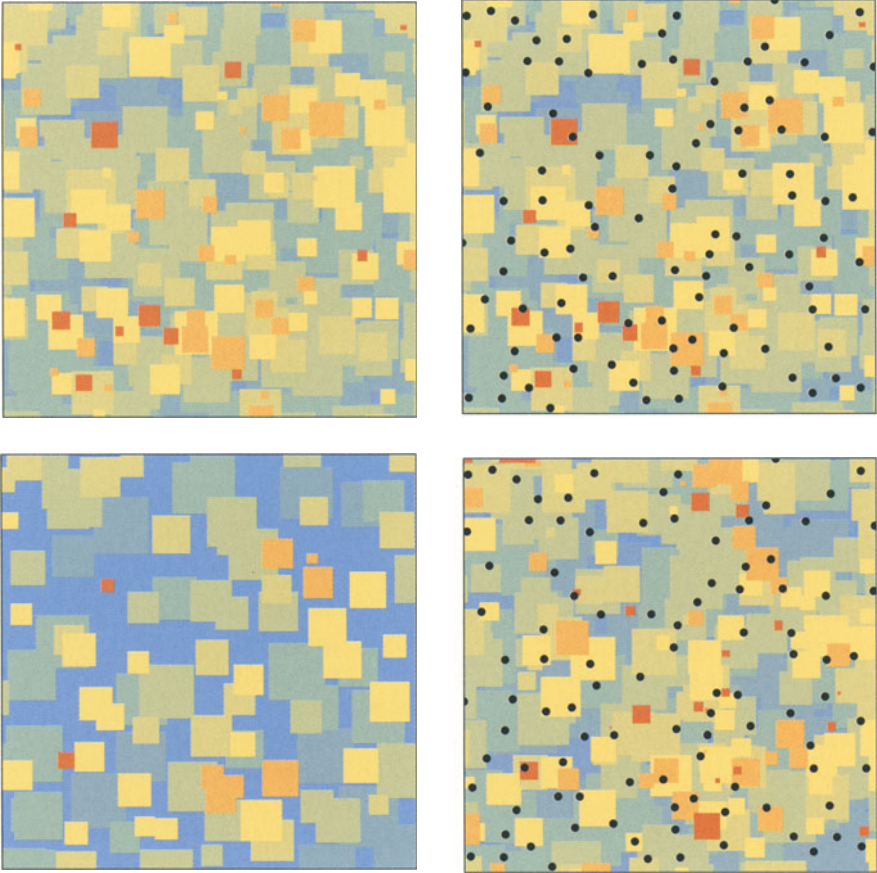
### 14.2.2 Simulation

The problem addressed here is the simulation of a boolean random function  $Z$  in the field  $D$ , given that  $Z(c) = z(c)$  for  $c$  contained in a finite subset  $C$  of  $\mathbb{R}^d$ . In order to avoid any edge effects, it is advantageous to consider  $Z$  in  $D$  as the maximum of  $N(D)$  typical objects, where

- $N(D)$  follows a Poisson distribution with mean  $\vartheta(D) = \theta E|D \oplus \check{A}_0|$
- a typical object of  $D$  can be written as the intersection between  $D$  and an object  $A(x)$  implanted w.r.t. density function  $x \rightarrow P\{(A_0)_x \cap D \neq \emptyset\}$  (cf. definition 13.2.1). The value assigned to the typical object is that of  $A(x)$ .

The conditional simulation algorithm that we propose consists of simulating the population  $\mathcal{X}$  of the objects that constitute the boolean random function. It is a variation of a non conditional iterative algorithm obtained by restricting the population to the set  $\Omega_c$  of all allowed populations (cf. section 8.4.1). At each step, the only changes permitted are either the insertion of a new object or the removal of an existing one. A new object can be inserted provided that its value does not exceed those of the conditioning data points that it contains. An existing object can be removed provided that the conditioning data points that it contains are honoured by the other objects of the population. An easy way to get an allowed population to start the algorithm is as follows. Typical objects are generated one after the other. We keep only those which are compatible with the conditioning data set and which are

the first ones to honour a data point. The initial procedure is completed as soon as all conditioning data points have been honoured. It is not detailed in Algorithm 14.2.1.



**Fig. 14.4.** Conditional simulation of a Boolean random function. Top left, a non conditional simulation. Top right, the same with the conditioning data set. Bottom left, the initial population used to start the conditional simulation algorithm. Bottom right, a conditional simulation obtained after 5000 iterations

**Algorithm 14.2.1.** (*Boolean random function. Conditional simulation*)

1. Generate  $\mathcal{X} \in \Omega_c$ .
2. Generate a random variable  $U$  that takes the values  $+1, -1$  and  $0$  with probability

$$p_{+1} = \frac{\vartheta(D)}{\vartheta(D) + \#\mathcal{X} + 1} \quad p_{-1} = \frac{\#\mathcal{X}}{\vartheta(D) + \#\mathcal{X}} \quad p_0 = 1 - p_{+1} - p_{-1}$$

- 3.1. If  $U = +1$ , then generate a typical valued object  $(A, \varepsilon)$ . If  $z(c) \geq \varepsilon$  for all  $c \in C \cap A$ , then set  $\mathcal{X} = \mathcal{X} \cup (A, \varepsilon)$ .
- 3.2. If  $U = -1$ , then generate  $(A, \varepsilon) \sim Unif(\mathcal{X})$ . If for every  $c \in A \cap C$  there exists  $(A', \varepsilon') \in \mathcal{X} \setminus (A, \varepsilon)$  such that  $z(c) = \varepsilon'$ , then set  $\mathcal{X} = \mathcal{X} \setminus (A, \varepsilon)$ .
4. Goto 2.

An example of this algorithm is given on Figure 14.4. Each object is a random square. The size of its side is  $1.5\sqrt{U}$  with  $U \sim Unif([0, 1])$ , and its value follows a conditional geometric distribution with parameter  $\sqrt{U}$  on  $\{1, \dots, 8\}$ . So the bigger the objects, the larger their values on average. The simulation field is  $10 \times 10$ . The top left of Figure 14.4 shows a non conditional simulation. On the top right, a set of 50 conditioning data points is superimposed on the non conditional simulation. Starting from this data set, the initialization procedure has been applied to produce the population depicted in the bottom left. The conditional simulation at the bottom right of Figure 14.4 has been obtained after running Algorithm 14.2.1 for 5000 iterations.

### 14.3 The dead leaves model

As its name indicates, this model has been devised to imitate falling leaves in autumn. The dead leaves overlap and gradually tessellate the ground.

#### 14.3.1 Definition and basic properties

This model has many variations. The one considered here is an extension by Jeulin (1979) of a model devised by Matheron (1968). It is based on the following considerations:

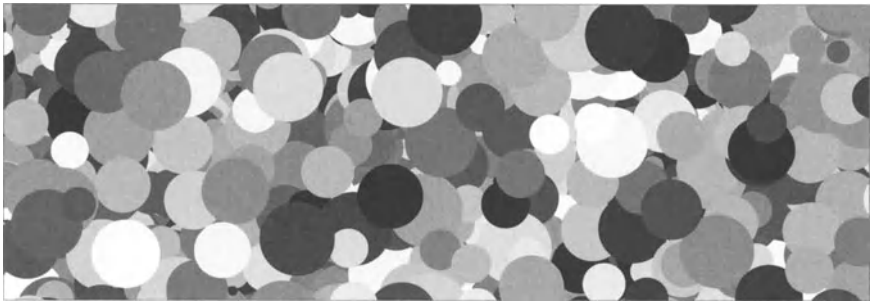
- i) the leaves are independent and nonempty random compact subset of  $\mathbb{R}^d$ .
- ii) the leaves fall according to a stationary Poisson point process with intensity  $\theta$  in  $\mathbb{R}^d \times ]-\infty, 0[$ . The leaf fallen on the point  $x$  at time  $t$  is denoted by  $A(x, t)$ .
- iii) a categorical value or *colour* is associated with each leaf. The set of all possible colours is finite and denoted by  $I$ . All leaves have the same chance  $p_i$  to be assigned the colour  $i \in I$ . Different leaves have independent colors.
- iv) all leaves with colour  $i$  have the same distribution (up to a translation) as a reference random compact subset  $A_i$  which has nonzero and finite mean  $d$ -volume.

As an immediate consequence, the leaves with colour  $i$  fall according to a Poisson point process with intensity  $\theta_i = \theta p_i$  in  $\mathbb{R}^d \times ]-\infty, 0[$ . Moreover,

each point  $x \in \mathbb{R}^d$  is almost surely covered by a leaf at time 0 because the leaves have a nonzero mean content. This makes the following definition possible:

**Definition 14.3.1.** *The dead leaves model associates to each point  $x \in \mathbb{R}^d$  the colour  $Z(x)$  of the most recently fallen leaf that covers  $x$ .*

**Remark 14.3.1.** It should be noted that the leaves fall according to a time reversible process. Therefore the direction of time can be inverted. This amounts to considering the leaves falling according to a Poisson point process in  $\mathbb{R}^d \times ]0, \infty[$ , and defining  $Z(x)$  as the colour of the first leaf that covers  $x$ . This point of view was first mentioned by Jeulin (1979) and further developed by Kendall and Thönnies (1999).



**Fig. 14.5.** Realization of a dead leaves model

Figure 14.5 depicts an example of a dead leaves model. The leaves are disks and their diameters are uniformly distributed between 0 and a maximal value (a quarter of the height of the image). On this realization, the uniform size of the diameters is not visible for two reasons. Firstly, what can be perceived are the areas of the disks and not their diameters. Secondly, the smaller a leaf, the more chance it has to be covered by a larger one.

Let  $Y$  be a finite subset of  $\mathbb{R}^d$ . What is the probability that  $Z$  takes some specific colour  $z(y)$  at each point  $y$  of  $Y$ ? Using remark 14.3.1, this probability can be calculated by randomizing on the first leaf  $A(x, t)$  that covers points of  $Y$ . Let  $L$  be the set of the points covered and let  $Y_1 = Y \setminus L$  its complement in  $Y$ . If the leaf has colour  $i$ , then  $L \subset A(x, t)$  and  $Y_1 \cap A(x, t) = \emptyset$  if and only if  $-x \in A_i \ominus \check{L}$  and  $-x \notin A_i \oplus \check{Y}$ . Accordingly we can write

$$P\left\{\bigwedge_{y \in Y} Z(y) = z(y)\right\} = \int_0^\infty e^{-\theta E|A \oplus \check{Y}|t} \alpha(t) dt$$

where  $\alpha(t)$  is defined as a sum over all indices  $i$  of  $I$  and to all nonempty subsets  $L$  of  $z^{-1}(i)$



$$\alpha(t) = \sum \theta_i E|(A_i \ominus \check{L}) \setminus (A_i \oplus \check{Y}_1)| P\left\{ \bigwedge_{y \in Y_1} Z(y) = z(y) \right\}$$

After integration, we obtain

$$P\left\{ \bigwedge_{y \in Y} Z(y) = z(y) \right\} = \frac{\sum \theta_i E|(A_i \ominus \check{L}) \setminus (A_i \oplus \check{Y}_1)| P\left\{ \bigwedge_{y \in Y_1} Z(y) = z(y) \right\}}{\theta E|A \oplus \check{Y}|}$$

Iterating this formula gives us

**Theorem 14.3.1.** *The spatial distribution of the dead leaves model is*

$$P\left\{ \bigwedge_{y \in Y} Z(y) = z(y) \right\} = \sum_{\mathcal{Y}} \prod_{k=0}^{ord(\mathcal{Y})} \frac{\theta_{i_k} E|(A_{i_k} \ominus \check{L}_k) \setminus (A_{i_k} \oplus \check{Y}_{k+1})|}{\theta E|A \oplus \check{Y}_k|}$$

in which the sum is extended to all finite sequences  $\mathcal{Y} = (Y_0, Y_1, \dots, Y_{n+1})$  satisfying

i)  $Y_0 = Y \supseteq \dots \supseteq Y_{n+1} = \emptyset$ . The integer  $n$  is called the order of  $\mathcal{Y}$  and denoted by  $ord(\mathcal{Y})$ .

ii) for each  $0 \leq k \leq ord(\mathcal{Y})$ , there is an index  $i_k \in I$  such that  $L_k = Y_k \setminus Y_{k+1} \in z^{-1}(i_k)$ .

Let us now see how this formula can be applied. To simplify the results, we will write  $a_i = \theta_i E|A_i|$  and  $a = \sum a_i$ .

- If  $Y = \{y\}$ , then  $\mathcal{Y} = \{\{y\}, \emptyset\}$  is the only sequence allowed. Of course,  $ord(\mathcal{Y}) = 0$ , and

$$P\{Z(y) = i\} = \frac{a_i}{a} \quad i \in I$$

- If  $Y = \{x, y\}$  with  $z(x) = i$  and  $z(y) = j$ , then two cases must be considered depending on whether  $i$  and  $j$  are equal or not. If  $i \neq j$ , there are only two possible sequences for  $\mathcal{Y}$ , namely  $\{\{x, y\}, \{x\}, \emptyset\}$  and  $\{\{x, y\}, \{y\}, \emptyset\}$ . The introduction of the geometric covariogram  $K_i$  of  $A_i$  leads to an interesting simplification

$$E|(A_i \ominus \{x, y\}) \setminus (A_i \oplus \{x\})| = K_i(0) - K_i(x - y)$$

from which we derive

$$P\{Z(x) = i, Z(y) = j\} = \frac{a_i(a_j - b_j) + a_j(a_i - b_i)}{a(2a - b)} \quad i, j \in I \quad i \neq j$$

after putting  $b_i = \theta_i K_i(x - y)$  and  $b = \sum b_i$ . In the case  $i = j$ , the calculations are similar, but a third sequence  $\{\{x, y\}, \emptyset\}$  has to be considered. This finally gives

$$P\{Z(x) = i, Z(y) = i\} = \frac{2a_i(a_i - b_i) + ab_i}{a(2a - b)} \quad i \in I$$

### 14.3.2 Simulation

Let us start with the non conditional simulation of the dead leaves model in the field  $D$ . In the same manner as in sections 13.2 and 14.2, typical leaves are introduced to cope with edge-effects. The leaves of colour  $i$  hitting the simulation field  $D$  are located according to a Poisson point process with intensity function  $\vartheta_i(x) = \theta_i P\{x \in D \oplus \check{A}_i\}$  in  $\mathbb{R}^d \times ]0, \infty[$ . Its integral over  $\mathbb{R}^d$  is clearly finite

$$\int_{\mathbb{R}^d} \vartheta_i(x) dx = \theta_i \int_{\mathbb{R}^d} P\{x \in D \oplus \check{A}_i\} dx = \theta_i E|D \oplus \check{A}_i| < \infty$$

Let  $\dot{x}$  be a random point in  $\mathbb{R}^d$  with density function  $\vartheta_i$ . The intersection of  $A(\dot{x}, t)$  and  $D$  is called a *typical  $i$ -leaf*. The hitting functional of a typical  $i$ -leaf is

$$T_i(K) = \int_{\mathbb{R}^d} \frac{\vartheta_i(x)}{\theta_i E|D \oplus \check{A}_i|} P\{x \in (D \cap K) \oplus \check{A}_i\} dx \quad K \in \mathcal{K}$$

Finally, the probability that a typical leaf is an  $i$ -leaf is

$$q_i = \frac{\theta_i E|D \oplus \check{A}_i|}{\sum_{i \in I} \theta_i E|D \oplus \check{A}_i|} \quad i \in I$$

Using this notation, the simulation algorithm that consists of assigning to each point of  $D$  the colour of the first leaf that covers it can be written as follows:

**Algorithm 14.3.1.** (*Dead leaves model*)

1. Set  $D_0 = D$  and  $Z = i_0 \notin I$  on  $D_0$ .
2. If  $D_0 = \emptyset$ , then deliver  $Z$ .  $\square$
3. Generate first  $i \sim q$ , and then a typical  $i$ -leaf  $A \sim T_i$ . Put  $Z(x) = i$  for all  $x \in A \cap Z^{-1}(i_0)$  and  $D_0 = D_0 \setminus A$ .
4. Goto 2.

**Remark 14.3.2.** The big advantage of this algorithm which was proposed by Kendall and Thönnies (1999) is to provide an exact simulation. An iterative algorithm directly based on definition 14.3.1 could be also considered. This would require the study of its rate of convergence as well as determining an appropriate halting criterion. Note in particular that the criterion that consists of stopping the algorithm as soon as the simulation field has been completely covered by leaves is biased. Indeed, the "ultimate leaf" that covers the last holes has more chance than normal to be large. To determine the order of magnitude of this bias, we carried out 10000 iterative simulations of a dead leaves model on a circle with perimeter 10. The leaves are arcs whose

length is exponentially distributed with mean value 1. Figure 14.6 shows the Q-Q plot of the length of a typical leaf versus the ultimate. The mean and variance of the ultimate leaf are respectively 2.46 and 2.54 (instead of 1 and 1 for the typical leaf).

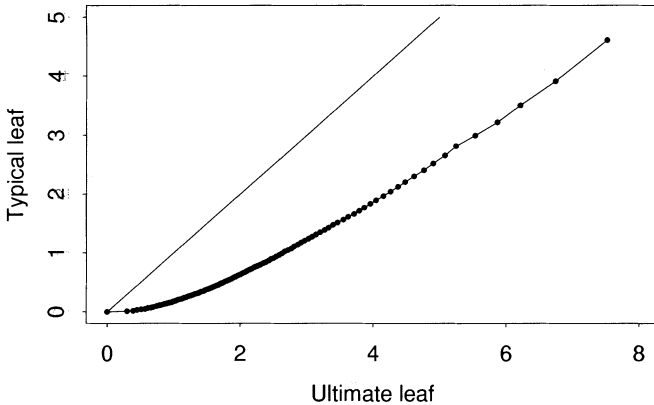


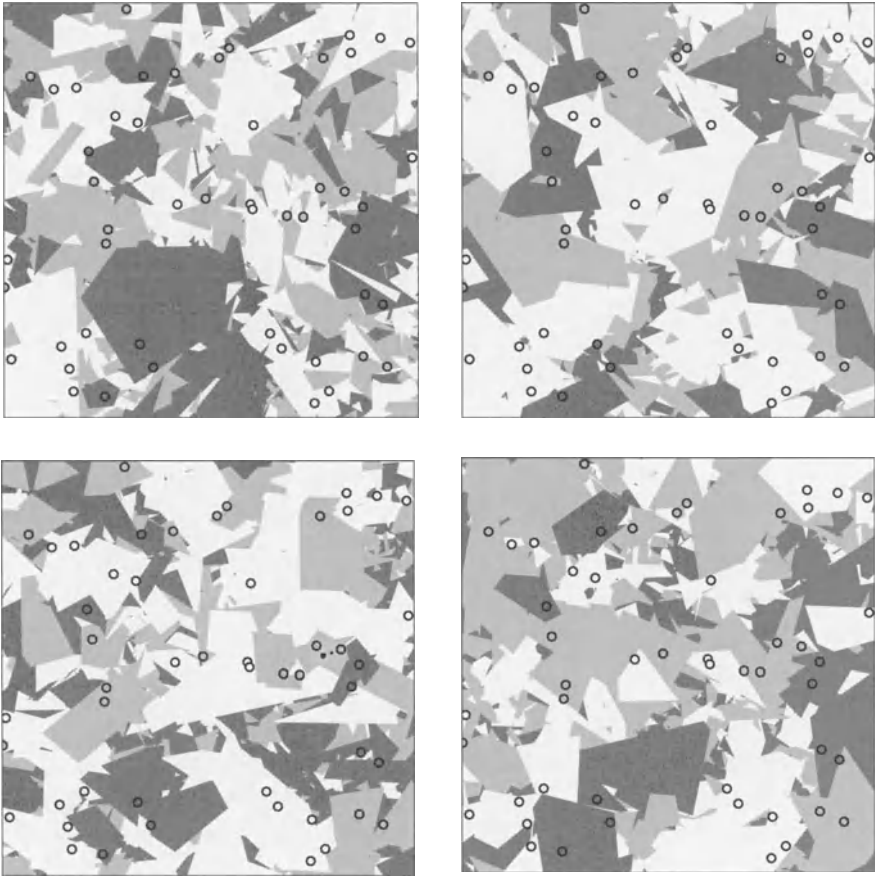
Fig. 14.6. Q-Q plot of the length of a typical leaf versus the ultimate

The problem of the conditional simulation of the dead leaves model is now considered. Let  $z(c), c \in C$  be the colours that the simulation must honour at a finite data subset  $C$  of  $D$ . Indeed, the exact algorithm proposed by Kendall and Thönnies remains basically valid. The only difference is that a typical leaf that falls on an uncovered conditioning data point of a different colour is systematically discarded. Here is the algorithm:

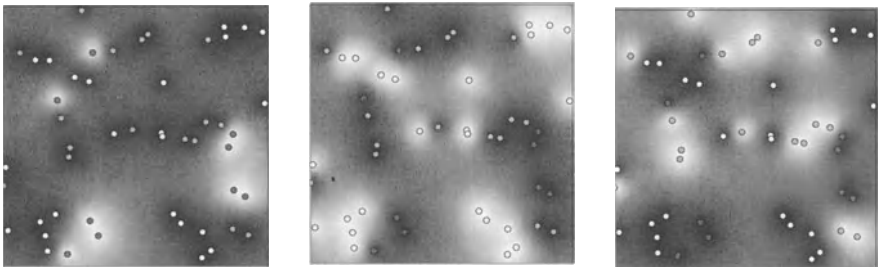
**Algorithm 14.3.2.** (*Dead leaves model. Conditional simulation*)

1. Set  $(D_0 := D)$  and  $Z = i_0 \notin I$  on  $D_0$ .
2. If  $(D_0 = \emptyset)$  then deliver  $Z$ .  $\square$
3. Generate first  $i \sim q$ , and then a typical  $i$ -leaf  $A \sim T_i$ .
4. If for any  $c \in A$ , either  $Z(c) \neq i_0$ , or  $Z(c) = i_0$  as well as  $z(c) = i$ , then put  $Z(x) = i$  for all  $x \in A \cap Z^{-1}(i_0)$  and  $D_0 = D_0 \setminus A$ .
5. Goto 2.

In order to illustrate this algorithm, we simulate a dead leaves model consisting of Poisson polygons. The simulation field is  $10 \times 10$  and the mean area of a leaf is 0.5. The top left of Figure 14.7 depicts a non conditional simulation, from which 50 independent and uniform points are selected to be used as a conditioning data set. Images of three conditional simulations are also shown. To generate each of them between 5000 and 10000 leaves were required. Of course many fewer leaves are visible. Their number ranges between 250 and 450 with a mean value of about 350.



**Fig. 14.7.** Conditional simulation of a dead leaves model. The leaves are typical Poisson polygons (same distribution and same proportion for the three colours). Top left, a non conditional simulation. Top right and bottom, three conditional simulations



**Fig. 14.8.** Conditional distribution of the colour of the points of the simulation field. From left to right, this gives the probability that a point is red, yellow or green

Replicating the simulations gives us an idea about the impact of the data set on the spatial distribution of the conditional model. Starting from a set of 1000 independent simulations, we have computed the frequency of occurrence of each point of the simulation field to belong to a red, a yellow or a green leaf (cf. Figure 14.8). From this we can derive an unbiased estimate of any conditional distribution based on points of  $D$ .

### Appendix: spatial distribution of a random token model

This appendix gives the proof of theorem 14.1.1. Let  $D$  be a compact subset of  $\mathbb{R}^d$ . We first consider the contribution of the Poisson points in  $D$

$$Z_D(y) = \sum_{x \in X \cap D} \varepsilon(x) 1_{y \in A(x)} \quad y \in \mathbb{R}^d$$

The Fourier transform of  $(Z_D(y_j), j \in J)$ , defined as

$$\Phi_D(u) = E \left\{ \exp \left\{ i \sum_{j \in J} u_j Z(y_j) \right\} \right\}$$

can be calculated by conditioning on the number and the location of the Poisson points in  $D$ . Explicitly, we get

$$\Phi_D(u) = \sum_{n=0}^{\infty} e^{-\theta|D|} \frac{[\theta|D|]^n}{n!} \left[ \frac{1}{|D|} \int_D E \left\{ \exp \left( i \varepsilon(x) \sum_{j \in J} u_j 1_{y_j \in A(x)} \right) \right\} dx \right]^n$$

which leads to

$$\ln \Phi_D(u) = \theta \int_D E \left\{ \exp \left( i \varepsilon(x) \sum_{j \in J} u_j 1_{y_j \in A(x)} \right) - 1 \right\} dx$$

Now, because of the distribution of  $\varepsilon(x)$ , this formula simplifies to

$$\ln \Phi_D(u) = \theta \int_D E \left\{ \cos \left( \sum_{j \in J} u_j 1_{y_j \in A(x)} \right) - 1 \right\} dx$$

Then we see that  $\sum_{j \in J} u_j 1_{y_j \in A(x)} = u_K$  for some  $K \subset J$  if and only if  $Y_K \subset A(x)$  and  $Y_{J \setminus K} \cap A(x) = \emptyset$ . Accordingly, if we introduce the notation

$$p_K(x) = P\{Y_K \subset A(x), Y_{J \setminus K} \cap A(x) = \emptyset\}$$

the previous formula becomes

$$\ln \Phi_D(u) = \theta \sum_{K \subset J} [\cos u_K - 1] \int_D p_K(x) dx$$

Expanding the domain  $D$  to  $\mathbb{R}^d$  gives

$$\ln \Phi(u) = \theta \sum_{K \subset J} [\cos u_K - 1] \int_{\mathbb{R}^d} p_K(x) dx$$

Finally, the integral can be explicitly calculated by starting from the definition of  $p_K$  and using the inclusion-exclusion formula

$$\begin{aligned} \int_{\mathbb{R}^d} p_K(x) dx &= \int_{\mathbb{R}^d} E \left\{ \prod_{k \in K} 1_{y_k \in A(x)} \prod_{l \in J \setminus K} [1 - 1_{y_l \in A(x)}] \right\} dx \\ &= \int_{\mathbb{R}^d} E \left\{ \sum_{L \supset K} (-1)^{\#L \setminus K} 1_{Y_L \subset A(x)} \right\} dx \\ &= \sum_{L \supset K} (-1)^{\#L \setminus K} E |A_0 \ominus \check{Y}_L| \end{aligned}$$

This completes the proof.  $\square$

## Exercises

**14.1** Let  $Z$  be a dilution random function in  $\mathbb{R}^d$  defined as

$$Z(y) = \sum_{x \in X} \varepsilon(x) f(y - x) \quad y \in \mathbb{R}^d$$

where

- $X$  is a Poisson point process with intensity  $\theta$
- $f$  is a deterministic function with finite integral
- $\varepsilon(x), x \in \mathbb{R}^d$  is a family of independent weights taking the values  $\pm 1$  with the same probability  $\frac{1}{2}$

What is the covariance function of  $Z$ ? What happens in the case when the integral of  $f$  is zero?

**14.2** The notation is the same as in the previous exercise, except that now  $X$  consists of the nodes of a cubic grid with mesh size  $a$  and uniform origin. The function  $f$  is also assumed to have its support contained within a ball of diameter  $a$ . What is the covariance function of  $Z$ ?

**14.3** Using the notation of section 14.3, what is the probability that the compact subset  $B$  is totally contained within a leaf of type  $i$ ?

# 15. Gaussian random function

An extremely useful consequence of the central limit theorem is the existence of a class of random functions whose spatial distribution depends only on their first two moments. These are *Gaussian random functions*. Their main statistical properties are reviewed, the texture of their realizations is examined, and algorithms are proposed to simulate them, conditionally or not.

## 15.1 Definition and basic properties

### 15.1.1 Results from probability

Let  $Y_1, \dots, Y_n, \dots$  be a sequence of independent and identically distributed random variables. If their mean  $m$  is finite, then the *strong law of large numbers* says that the average  $(Y_1 + \dots + Y_n)/n$  converges almost surely towards  $m$

$$\lim_{n \rightarrow +\infty} \frac{Y_1 + \dots + Y_n}{n} = m \quad a.s.$$

In the case when their variance  $\sigma^2$  is also finite, then the *central limit theorem* states that the distribution of  $(Y_1 + \dots + Y_n)/n$  tends to be gaussian around its mean value

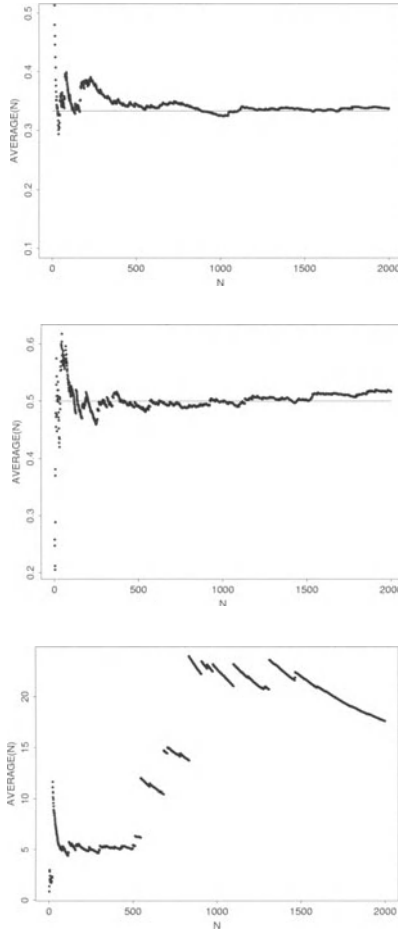
$$\lim_{n \rightarrow +\infty} P \left\{ \frac{Y_1 + \dots + Y_n - m}{\frac{n \sigma}{\sqrt{n}}} < y \right\} = G(y)$$

where  $G$  denotes the standard gaussian distribution function.

Going further, we can wonder from what value of  $n$  onwards the distribution of  $(Y_1 + \dots + Y_n)/n$  can be considered as gaussian for practical purposes. Provided that the variables have a finite third order absolute moment  $m_3 = E\{|Y_i - m|^3\}$ , an answer is given by the *Berry-Esséen theorem* (Feller, 1971; Rao, 1984). This theorem gives an upper bound for the difference between the standardized distribution of the average and the standard gaussian distribution:

$$\sup_{y \in \mathbb{R}} \left| P \left\{ \frac{Y_1 + \dots + Y_n - m}{\frac{\sigma}{\sqrt{n}}} < y \right\} - G(y) \right| < \frac{\alpha}{\sqrt{n}} \frac{m_3}{\sigma^3}$$

where  $\alpha$  is a number less than 1.32132.



**Fig. 15.1.** The consequences of the law of large numbers and the central limit theorem can be illustrated by plotting the average of  $n$  independent values as a function of  $n$ . The top distribution has a finite mean and a finite variance, the middle one has a finite mean but an infinite variance and the bottom one has an infinite mean as well as an infinite variance

These basic results are now illustrated by considering the family of p.d.f.'s  $(f_k, k \geq 1)$  defined as



$$f_k(x) = \frac{k}{(1+x)^{k+1}} \quad x \geq 0$$

Consider the case  $k = 4$ . The distribution  $f_4$  has a finite mean and a finite variance, so the law of large numbers and the central limit theorem are applicable. In Figure 15.1 (top), the average of  $n$  independently generated values has been plotted as a function of  $n$ . The curve obtained tends to the mean value of  $1/3$  and stays there.

For  $k = 3$ , the distribution  $f_3$  has a finite mean but an infinite variance. Consequently, the law of large numbers holds, but the central limit theorem does not. In that case, the curve (Figure 15.1, middle) tends to the mean value of  $1/2$  with occasional "bumps".

Finally for  $k = 2$ , the distribution  $f_2$  has an infinite mean. So even the law of large numbers is not applicable. As can be seen from Figure 15.1 (bottom), there are intervals on which the curve is decreasing, but from time to time a very large value is generated resulting in a sharp jump in the function values. In the long run, the curve tends to infinity (i.e. its mean).

### 15.1.2 Definition of a gaussian random function

Now let  $(Y_n, n \in \mathbb{N})$  be a sequence of independent and identically distributed (i.i.d.) second-order stationary random functions. To simplify the notation, the  $Y_n$ 's have been standardized: their mean is equal to 0 and their variance to 1. Consider the random function  $Y^{(n)}$  defined as

$$Y^{(n)} = \frac{Y_1 + \dots + Y_n}{\sqrt{n}}$$

Because of the central limit theorem, any linear combination of its variables

$$\sum_{j=1}^p \lambda_j Y^{(n)}(x_j) = \frac{\sum_{j=1}^p \lambda_j Y_1(x_j) + \dots + \sum_{j=1}^p \lambda_j Y_n(x_j)}{\sqrt{n}},$$

tends to follow a gaussian distribution as  $n$  becomes very large. This leads to the following definition:

**Definition 15.1.1.** *A random function is said to be gaussian if any linear combination of its variables follows a gaussian distribution.*

Let  $Y$  be a gaussian random function with mean  $m$  and covariance function  $C$ . The multivariate distribution of  $(Y(x_1), \dots, Y(x_p))$  is characterized by its Fourier transform

$$\Phi(x_1, \dots, x_p; u_1, \dots, u_p) = E \left\{ \exp \left( i \sum_{j=1}^p u_j Y(x_j) \right) \right\}$$

Since  $Y$  is gaussian,  $\sum_{j=1}^p u_j Y(x_j)$  follows a gaussian distribution with mean  $\sum_{j=1}^p u_j m(x_j)$  and variance  $\sum_{j,k=1}^p u_j u_k C(x_j, x_k)$ . Therefore

$$\Phi(x_1, \dots, x_p; u_1, \dots, u_p) = E \left\{ \exp \left( i \sum_{j=1}^p u_j m(x_j) - \frac{1}{2} \sum_{j,k=1}^p u_j u_k C(x_j, x_k) \right) \right\}$$

In particular,  $\Phi$  depends only on  $m$  and  $C$ . Consequently

**Theorem 15.1.1.** *The spatial distribution of a gaussian random function is totally characterized by its mean value and its covariance function.*

Now let  $(Y(y_1), \dots, Y(y_q))$  be another vector. Suppose that both vectors are uncorrelated, that is  $C(x_j, y_k) = 0$  for any  $j = 1, \dots, p$  and any  $k = 1, \dots, q$ . Then the Fourier transform of  $(Y(x_1), \dots, Y(x_p), Y(y_1), \dots, Y(y_q))$  can be written as the product of the Fourier transforms of  $(Y(x_1), \dots, Y(x_p))$  and  $(Y(y_1), \dots, Y(y_q))$ . It follows that

**Theorem 15.1.2.** *Two gaussian vectors are independent if and only if they are uncorrelated.*

### 15.1.3 Textures of the realizations

From now onwards,  $Y$  denotes a second-order stationary gaussian random function with mean 0, variance 1 and covariance function  $C$ . As was said in chapter 3,  $C$  cannot be arbitrary, since the variance of any linear combination of the variables of  $Y$  must be non negative. This was expressed in chapter 3 by saying that a covariance function is *positive definite*. Table 15.1 gives some examples of covariance functions. These have been chosen because they have totally different shapes. In particular their behaviours differ at the origin, near the range or at large scale. The consequences of these differences are apparent in the realizations shown on Figure 15.2.

How can we compare realizations derived from different covariances? The scale factor  $a$  of the spherical, exponential, stable and Gaussian covariances has been chosen so that the realizations behave similarly at large scale. This can be done by giving these four covariances the same integral range in two dimensions. On Figure 15.2, the simulation domain is 160 times larger than the integral range. This is not possible for the hyperbolic covariance which has an infinite integral range. As it can be observed, the realisation shows no sign of spatial homogeneity. In the hyperbolic case, the scale factor has been assigned a value that gives the covariance the same behaviour at the origin as the exponential covariance.

Spherical	$C(h) = \left(1 - \frac{3 h }{2a} + \frac{1 h ^3}{2a^3}\right) 1_{ h  \leq a}$
Exponential	$C(h) = \exp\left\{-\frac{ h }{a}\right\}$
Stable	$C(h) = \exp\left\{-\sqrt{\frac{ h }{a}}\right\}$
Hyperbolic	$C(h) = \frac{1}{1 + \frac{ h }{a}}$
Gaussian	$C(h) = \exp\left\{-\frac{ h ^2}{a^2}\right\}$
Cardinal sine	$C(h) = \frac{\sin\left\{\frac{ h }{a}\right\}}{\frac{ h }{a}}$

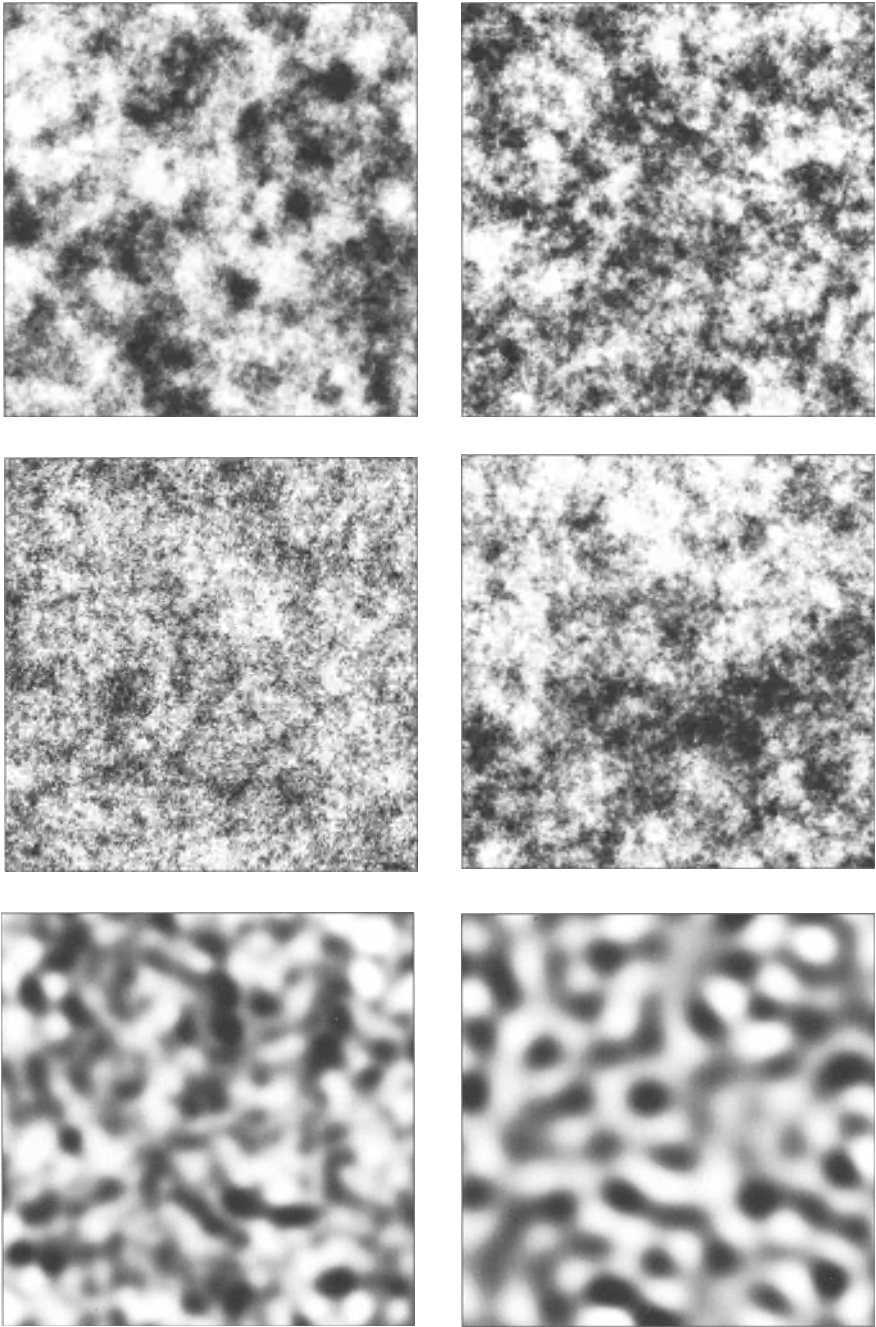
**Tab. 15.1.** Formula for some models of covariance function

The stable covariance has an infinite slope at the origin, which explains the grainy texture of its realization on Figure 15.2. The cardinal sine covariance, like the hyperbolic covariance, has an infinite integral range in two dimensions. Its scale factor has been chosen to give it the same behaviour at the origin as the gaussian covariance. The cardinal sine covariance is pseudoperiodic (but not strictly periodic: a covariance function cannot be isotropic and periodic in a workspace of dimension greater than one), which leads to realisations with a dendritic pattern (see the bottom right of Figure 15.2). The width of the dendritic limbs is closely related to the scale factor.

In all cases, the realizations look continuous. Indeed, Adler (1981) showed that the condition

$$1 - C(h) \leq \frac{C_0}{|\ln |h||^{1+\varepsilon}} \quad h \in \mathbb{R}^d$$

for some  $0 < C_0 < \infty$  and some  $\varepsilon > 0$  is sufficient for all realizations of  $Y$  to be continuous almost surely.

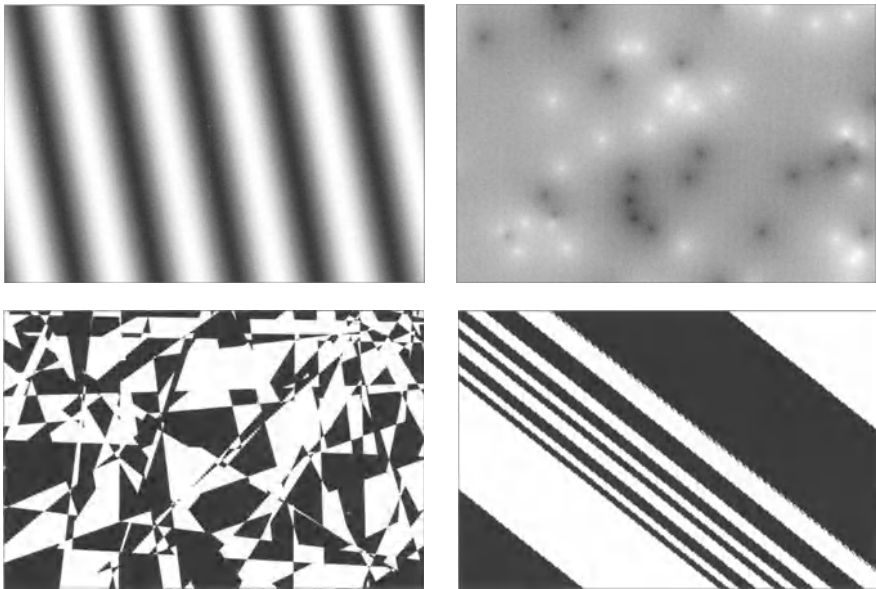


**Fig. 15.2.** Realizations of gaussian random functions with different covariance functions. From top to bottom and left to right, spherical, exponential, stable, hyperbolic, gaussian and cardinal sine covariances

## 15.2 Non conditional simulation

In this section, the problem of simulating a stationary gaussian random function with mean 0, variance 1 and covariance function  $C$  in a continuous domain  $D$  of  $\mathbb{R}^d$  is considered. Based on the central limit theorem, one possible procedure consists of simulating a large number of independent standardized, stationary random functions (not necessarily gaussian) with covariance  $C$ . This procedure raises two questions, firstly, how to simulate a random function with covariance  $C$  in general, and secondly, how many of these random functions have to be simulated so that the central limit theorem is applicable.

There are many different methods for simulating a random function with specified covariance. They include the dilution, tessellation, spectral and turning bands methods (see Figure 15.3). We will present them briefly and compare them from various standpoints such as their degree of generality, their range of validity and their practicality. Unfortunately, the second question cannot be answered by giving some no magic number. However a number of criteria are available to decide how many random functions to simulate. This will be discussed at the conclusion of this chapter.



**Fig. 15.3.** Examples of methods that can be used to simulate a gaussian random function with an exponential covariance. From top to bottom and left to right, the spectral method, the dilution method, the tessellation method and the turning bands method

### 15.2.1 The dilution method

This method can be applied in the case when the covariance is proportional to the geometric covariogram of a random object  $A$

$$C(h) = \theta E|A \cap A_h|$$

In that case, it suffices to consider

$$Y(y) = \sum_{x \in X} \varepsilon(x) 1_{y \in A(x)} \quad y \in D$$

where  $X$  is a stationary Poisson point process in  $\mathbb{R}^d$  with intensity  $\theta$ ,  $(A(x), x \in \mathbb{R}^d)$  is a family of independent random objects (each  $A(x)$  has the same distribution as  $A$  shifted by vector  $\vec{ox}$ ), and  $\varepsilon$  is a family of independent random weights taking only the two values  $\pm 1$  with probability  $\frac{1}{2}$  (see section 14.1). For instance, an exponential covariance can be generated using typical Poisson polytopes (see section 12.3.2). It is also possible to take random balls, but then the value assigned to a ball depends on its radius (Hammersley and Nelder (1955)).

From a practical point of view, the dilution method works well when the random object  $A$  is almost surely bounded. For instance, it is ideal for simulating a spherical covariance function because it is the geometric covariogram of a ball.

Another common case is to have the covariance proportional to the transitive covariogram of a function  $f$

$$C(h) = \theta \int_{\mathbb{R}^d} f(x) f(x+h) dx$$

In that case, we can take

$$Y(y) = \sum_{x \in X} \varepsilon(x) f(y-x) \quad y \in D$$

For instance, the exponential covariance with scale factor  $a$  is proportional to the transitive covariogram of the function

$$f(x) = \left(\frac{|x|}{a}\right)^\nu K_\nu\left(\frac{|x|}{a}\right)$$

where  $\nu = \frac{d-1}{4}$  and  $K_\nu$  is the modified Bessel function of order  $\nu$ . This approach is possible (see top right of Figure 15.3) but is not easy to implement because  $f$  is not bounded and does not have bounded support.

### 15.2.2 The tessellation method

In this method,  $\mathbb{R}^d$  is partitioned into a stationary population of random cells which are assigned independent random values. The covariance of the random function thus obtained is equal to the geometric covariogram of the cells times the variance of the random values (see section 12.1.2). For instance, the exponential covariance can be associated with a population of Poisson polytopes (see bottom left of Figure 15.3).

Provided that we can simulate the tessellation, this method is rather efficient. But its degree of generality is limited because many geometric covariograms cannot be associated with the typical cell of a stationary tessellation (e.g. the spherical covariogram). This raises the question of how to characterize the covariance function of a stationary tessellation. So far, the answer is unknown.

### 15.2.3 The spectral method

We know that the covariance  $C$  is a positive definite function. If it is also continuous, then Bochner's theorem states that it is the Fourier transform of a positive measure (the spectral measure), say  $\mathcal{X}$

$$C(h) = \int_{\mathbb{R}^d} e^{i \langle h, u \rangle} d\mathcal{X}(u)$$

Moreover  $C(0) = 1$ , so  $\mathcal{X}$  is a probability measure. A direct calculation shows that if  $V$  is a random vector with distribution  $\mathcal{X}$ , and if  $U$  is a variable with a uniform distribution over  $]0, 1[$  independent of  $V$ , then the random function defined as

$$Y(x) = \sqrt{2} \cos(\langle V, x \rangle + 2\pi U)$$

is standardized and has  $C$  as its covariance function. The spectral method, developed by Shinozuka and Jan (1972) among others, makes use of this result and is quite general. Algorithm 15.2.1 gives a way of implementing it.

**Algorithm 15.2.1.** (*Spectral method*)

1. Generate independently  $V_1, \dots, V_n \sim \mathcal{X}$  and  $U_1, \dots, U_n \sim Unif$ .
2. Compute  $\frac{\sqrt{2}}{\sqrt{n}} \sum_{k=1}^n \cos(\langle V_k, x \rangle + 2\pi U_k)$  for each  $x \in D$ .

The practical set-up of this method requires knowledge of the spectral measure  $\mathcal{X}$  and a criterion for deciding the number  $n$  of basis functions. The second issue is deferred to the next section. Regarding the first, things are simple if  $C$  is integrable. In this case,  $\mathcal{X}$  has a density<sup>1</sup>, say  $f$ , and the formula linking  $f$  and  $C$  can be inverted to give

<sup>1</sup>  $C^2 \leq |C|$  because  $-1 \leq C \leq 1$ . So  $C$  is square integrable as soon as it is integrable. A classical result (e.g Rudin (1966)) states that the Fourier transform of a square integrable functions is also square integrable. Moreover the correspondence is bijective provided that all functions almost everywhere equal are identified.

$$f(u) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i \langle u, h \rangle} C(h) dh$$

A list of covariance functions and their spectral distributions is given in Appendix of this book.

**Example 15.2.1.** The spectral distribution of an exponential covariance with scale factor  $a$  has the density

$$f(u) = \frac{\Gamma(\mu)}{\pi^\mu} \frac{a^d}{(1 + a^2|u|^2)^\mu}$$

with  $\mu = \frac{d+1}{2}$ . It can be simulated either by using the acceptance-rejection method, or by interpreting it as a gamma ( $\frac{1}{2}$ ) mixture of gaussian distributions

$$f(u) = \int_0^\infty \left(\frac{ta^2}{\pi}\right)^{\frac{d}{2}} e^{-ta^2|u|^2} \frac{1}{\sqrt{\pi}} e^{-t} t^{\frac{1}{2}-1} dt$$

The spectral method is general but not always efficient. Indeed, the spectral measure is easier to sample when it vanishes rapidly at infinity. Because the Fourier transform exchanges the properties at 0 and at infinity, the spectral method is especially recommended when the covariance function is smooth at the origin.

### 15.2.4 The turning bands method

Designed by Matheron (1972a, 1973), the turning bands method is a stereological device for reducing a multidimensional simulation into unidimensional ones. Let us start again with the spectral representation of  $C$ . By introducing the system of polar coordinates  $u = (\theta, \rho)$  where  $\theta$  is a direction parameter ( $\theta$  spans half a sphere, say  $S_d^+$ ), and  $\rho$  is a location parameter ( $-\infty < \rho < +\infty$ ), the spectral measure  $d\mathcal{X}(u)$  can be written as the product of the distribution  $d\varpi(\theta)$  of  $\theta$  and the conditional distribution  $d\mathcal{X}_\theta(\rho)$  of  $\rho$  given  $\theta$

$$d\mathcal{X}(u) = d\varpi(\theta) d\mathcal{X}_\theta(\rho)$$

Using this decomposition, the spectral expansion of the covariance  $C$  becomes

$$C(h) = \int_{S_d^+} \int_{-\infty}^{+\infty} \exp\{i\rho \langle h, \theta \rangle\} d\mathcal{X}_\theta(\rho) d\varpi(\theta)$$

This leads us to consider the unidimensional function

$$C_\theta(r) = \int_{-\infty}^{+\infty} \exp\{ir\rho\} d\mathcal{X}_\theta(\rho)$$

By Bochner's theorem,  $C_\theta$  is a covariance function, and we have



$$C(h) = \int_{S_d^+} C_\theta(\langle h, \theta \rangle) d\varpi(\theta)$$

The idea behind the turning bands method is to reduce the simulation of a gaussian random function with covariance  $C$  to the simulations of independent stochastic processes with covariances  $C_\theta$ . Let  $(\theta_n, n \in \mathbb{N})$  be a sequence of directions of  $S_d^+$ , and let  $(X_n, n \in \mathbb{N})$  be a sequence of independent stochastic processes with covariances  $C_{\theta_n}$ . The random function

$$Y^{(n)}(x) = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k(\langle x, \theta_k \rangle) \quad x \in \mathbb{R}^d$$

admits the covariance

$$C^{(n)}(h) = \frac{1}{n} \sum_{k=1}^n C_{\theta_k}(\langle h, \theta_k \rangle)$$

As  $n$  becomes very large, the central limit theorem implies that the spatial distribution of  $Y^{(n)}$  tends to become gaussian with covariance  $\lim_{n \rightarrow +\infty} C^{(n)}$ . This limit is exactly  $C$  in the case when  $\frac{1}{n} \sum_{k=1}^n \delta_{\theta_k}$  converges weakly to  $\varpi$ . The turning bands algorithm is the following one:

**Algorithm 15.2.2.** (*Turning bands method*)

1. Choose a set of directions  $\theta_1, \dots, \theta_n$  such that  $\frac{1}{n} \sum_{k=1}^n \delta_{\theta_k} \approx \varpi$ .
2. Generate independent standard stochastic processes  $X_1, \dots, X_n$  with covariance functions  $C_{\theta_1}, \dots, C_{\theta_n}$ .
3. Compute  $\frac{1}{\sqrt{n}} \sum_{k=1}^n X_k(\langle x, \theta_k \rangle)$  for any  $x \in D$ .

**Remark 15.2.1.** Both step 3 of algorithm 15.2.2 and step 2 of algorithm 15.2.1 stem from the application of the central limit theorem. Accordingly, they are similar. The only difference is that the spectral method uses stochastic processes exclusively built starting from cosine functions. In that sense, the turning band method can be seen as a generalization of the spectral method.

Algorithm 15.2.2 does not say how to generate the directions, how to determine the covariance  $C_\theta$ , and how to build a stochastic process with covariance  $C_\theta$ . These questions are now considered in turn.

**Generating the directions.** Suppose here that  $\varpi$  has a continuous density  $p$  on  $S_d^+$ . If  $\frac{1}{n} \sum_{k=1}^n \delta_{\theta_k}$  converges weakly to the uniform distribution on  $S_d^+$ , then  $\frac{|S_d^+|}{n} \sum_{k=1}^n \delta_{\theta_k} p(\theta_k)$  converges weakly to  $\varpi$ . Consequently, only the isotropic case needs being considered.

The simplest method is to arrange the  $\theta_k$ 's regularly on  $S_d^+$ . This approach works well if  $d = 2$ , but not otherwise. For instance, if  $d = 3$ , the maximum number of regular directions is equal to 15 (Guibal, 1972; Journel and Huijbregts, 1978).

Another method is to generate the  $\theta_k$ 's independently and uniformly in  $S_d^+$ . However this is not fully satisfactory because the convergence of  $\frac{1}{n} \sum_{k=1}^n \delta_{\theta_k}$  to the uniform distribution of  $S_d^+$  is rather slow.

A third approach is to resort to sequences with weak discrepancy (Bouleau, 1986). It consists of producing  $\theta_n$  as far as possible from  $\theta_1, \dots, \theta_{n-1}$  while filling in  $S_d^+$  as fast as possible. In three dimensions, Freulon (1992) started with the binary and the ternary expansions of each integer  $n = 1, 2, \dots$

$$n = a_0 + 2a_1 + \dots + 2^p a_p = b_0 + 3b_1 + \dots + 3^q b_q$$

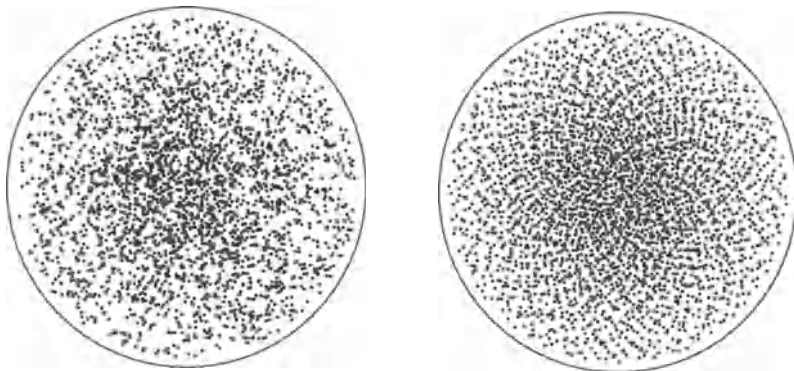
( $a_i = 0, 1$  and  $b_j = 0, 1, 2$ ) to get 2 numbers  $u_n$  and  $v_n$  between 0 and 1

$$u_n = \frac{a_0}{2} + \frac{a_1}{4} + \dots + \frac{a_p}{2^{p+1}} \qquad v_n = \frac{b_0}{3} + \frac{b_1}{9} + \dots + \frac{b_q}{3^{q+1}}$$

Then he put

$$\theta_n = \left( \cos(2\pi u_n) \sqrt{1 - v_n^2}, \sin(2\pi u_n) \sqrt{1 - v_n^2}, v_n \right)$$

Figure 15.4 compares 3000 independent and uniform points with 3000 points generated according to Freulon's algorithm.



**Fig. 15.4.** Stereographical projection of 3000 points generated uniformly (left) or according to a sequence with a weak discrepancy (right)

**Determining the unidimensional covariances.** The standard approach is to calculate the spectral measure  $\mathcal{X}$  of  $C$ , then derive the spectral measure  $\mathcal{X}_\theta$  of  $C_\theta$  and take its Fourier transform.

Considerable simplifications occur when  $C$  is isotropic, i.e. it can be written as  $C(h) = C_d(|h|)$  for some scalar function  $C_d$  defined on  $\mathbb{R}^+$ . In this case, all of the  $C_\theta$  are equal to a single covariance function, say  $C_1$ . Matheron (1972a) gives the relationship between  $C_1$  and  $C_d$

$$C_d(r) = 2 \frac{(d-1)\omega_{d-1}}{d\omega_d} \int_0^1 (1-t^2)^{\frac{d-3}{2}} C_1(tr) dt$$

where  $\omega_d$  stands, as usual, for the  $d$ -volume of the unit ball in  $\mathbb{R}^d$ . If  $d = 3$ , this formula reduces to

$$C_3(r) = \int_0^1 C_1(tr) dt$$

or equivalently

$$C_1(r) = \frac{d}{dr} (r C_3(r))$$

For  $d = 2$ , the relationship between  $C_1$  and  $C_2$  is more complicated

$$C_2(r) = \frac{1}{\pi} \int_0^\pi C_1(r \sin \theta) d\theta$$

Its general solution is

$$C_1(r) = 1 + r \int_0^{\pi/2} \frac{dC_2}{dr}(r \sin \theta) d\theta$$

(Brooker, 1985; Dietrich, 1995). Because this integral form is not easy to handle, Gneiting (1995, 1996) derives  $C_1$  explicitly for the most commonly used covariances  $C_2$ .

**Simulating the stochastic processes.** There is usually a wide choice of methods for simulating a stochastic process with a given covariance function  $C_1$ , which makes the turning bands method very flexible. Of course, the spectral method can be applied, but this just amounts to applying it directly in  $\mathbb{R}^d$ . In what follows, two examples encountered in three-dimensional isotropic simulations are discussed.

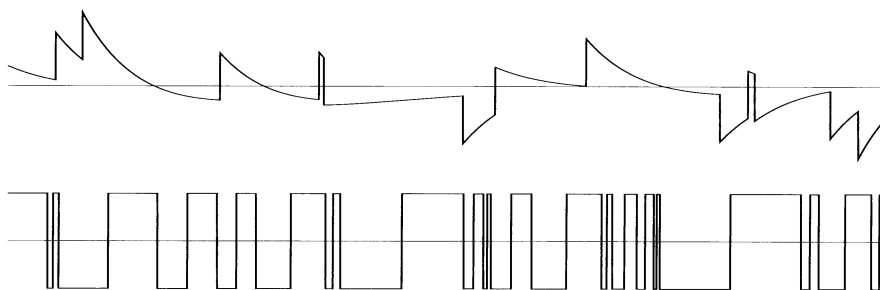
**Example 15.2.2.** Suppose that  $C_3$  is an exponential covariance with scale factor  $a$ . The associated unidimensional covariance is

$$C_1(r) = \left(1 - \frac{r}{a}\right) \exp\left\{-\frac{r}{a}\right\} \quad r \leq 0$$

To simulate  $C_1$ , we can consider the dilution stochastic process using the function

$$f(t) = \left(1 - \frac{t}{a}\right) \exp\left\{-\frac{t}{a}\right\} 1_{t \geq 0}$$

(Matheron, 1972; Journel and Huijbregts, 1978). The realization shown on top of Figure 15.5 is only approximate because  $f$  does not have bounded support. This is the reason why the following alternative process is preferred for simulating  $C_1$ . It is a special case of a migration process developed by Haas et al. (1967). It consists of generating a Poisson point process that partitions  $\mathbb{R}$  into independent exponential intervals of mean length  $2a$ . Each interval is split into two halves. The first half is set to  $-1$ , the second half to  $+1$  (see bottom of Figure 15.5).



**Fig. 15.5.** Simulation of a dilution process (top) and a migration process (bottom). Both processes can be used to simulate a 3D gaussian random function with an exponential covariance using the turning bands method

**Example 15.2.3.** Suppose that  $C_3$  is a spherical covariance with scale factor  $a$ . Then

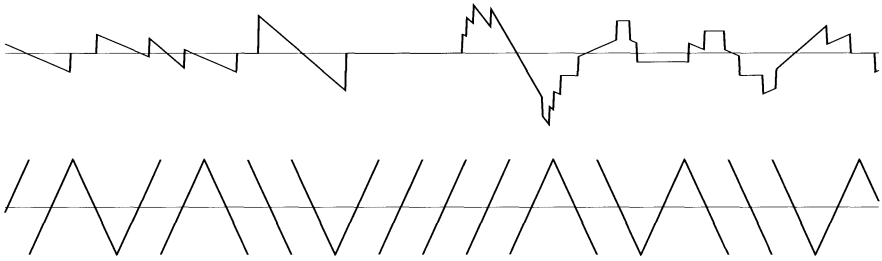
$$C_1(r) = \left(1 - 3\frac{r}{a} + 2\frac{r^3}{a^3}\right) 1_{0 \leq r \leq a}$$

$C_1$  can be simulated by dilution using the function

$$f(t) = t 1_{|t| \leq \frac{a}{2}}$$

(Matheron, 1972, Journel and Huijbregts, 1978). The main difference to the previous example is that  $f$  has bounded support. This makes the dilution stochastic process numerically suitable (see top of Figure 15.6).

Instead of considering a Poisson point process, it is also possible to implant  $f$  at the nodes of a regular grid with mesh size  $a$  and origin uniform in  $]0, a[$ . In that case  $f$  must be multiplied by  $\sqrt{3}$  to ensure a variance of 1, and this gives the simulation displayed at the bottom of Figure 15.6.



**Fig. 15.6.** Simulation of a standard dilution process (top) and a regular dilution process (bottom). Both processes can be used to simulate a 3D gaussian random function with a spherical covariance using the turning bands method

**15.2.5 How many random functions?**

Suppose that  $n$  is fixed. We want to know how close the spatial distribution of

$$Y^{(n)} = \frac{Y_1 + \dots + Y_n}{\sqrt{n}}$$

is to a standard gaussian spatial distribution? Equivalently, if  $\{x_1, \dots, x_p\}$  denotes an arbitrary set of points, and if  $\{\lambda_1, \dots, \lambda_p\}$  denotes an arbitrary set of real numbers, how far the distribution of  $\sum_{j=1}^p \lambda_j Y^{(n)}(x_j)$  is from a gaussian distribution with mean 0 and variance  $\sigma^2 = \sum_{j,k=1}^p \lambda_j \lambda_k C(x_j - x_k)$ ? There are many ways of addressing this problem. To avoid a lengthy description, only the distance approach and the moment approach are considered here.

**The distance approach.** This consists of assessing the distance between  $F^{(n)}$ , the distribution of  $\sum_{j=1}^p \lambda_j Y^{(n)}(x_j)$  and  $F$ , that of a gaussian variable with mean 0 and variance  $\sigma^2$ . Several distance functions can be considered, including

– the *Kolmogorov distance*

$$d_K(F^{(n)}, F) = \sup_{y \in \mathbb{R}} |F^{(n)}(y) - F(y)|$$

An upper bound of the Kolmogorov distance is provided by the Berry-Esséen theorem.

– the *Lévy distance*

$$d_L(F^{(n)}, F) = \inf \{ \epsilon > 0 : \forall y \in \mathbb{R} F^{(n)}(y - \epsilon) - \epsilon \leq F(y) \leq F^{(n)}(y + \epsilon) + \epsilon \}$$

Note that  $\sqrt{2}d_L(F^{(n)}, F)$  is the maximum distance between the graphs of  $F^{(n)}$  and  $F$ , measured along lines at 135°.

– the *total variation distance*

$$d_{TV}(F^{(n)}, F) = 2 \sup_{B \in \mathcal{B}} |F^{(n)}(B) - F(B)|$$

It can be established that  $d_L \leq d_\kappa \leq \frac{1}{2}d_{TV}$  (e.g. Huber, 1981).  
 – the *Matheron distance* (1984)

$$d_M(F^{(n)}, F) = \sqrt{-\int_{\mathbb{R}^2} [dF^{(n)}(x) - dF(x)] |x - y| [dF^{(n)}(y) - dF(y)]}$$

Note in particular the minus sign that stems from the fact that the mapping  $h \rightarrow -|h|$  is conditionally positive definite. The Matheron distance can also be written as

$$d_M(F^{(n)}, F) = \sqrt{\frac{1}{\pi} \int_0^\infty \left[ \frac{\hat{F}^{(n)}(u) - \hat{F}(u)}{u} \right]^2 du}$$

where  $\hat{F}^{(n)}$  and  $\hat{F}$  are the Fourier transforms of the distributions  $F^{(n)}$  and  $F$ .

Unfortunately, all these distances are difficult to handle analytically. But they can either be computed or estimated by simulation.

**The moment approach.** This consists of comparing the moments of the random variable  $\sum_{j=1}^p \lambda_j Y^{(n)}(x_j)$  with those of a gaussian variable with mean 0 and variance  $\sigma^2$ . As all the odd moments are equal to 0, and the variances are automatically equal, the simplest non trivial moment to consider is the  $4^{th}$  moment.

Because the  $4^{th}$  order moment of a centered gaussian variable is equal to 3 times the square of its variance, the difference between the  $4^{th}$  order moment  $\mu_4^{(n)}$  of  $\sum_{j=1,p} \lambda_j Y^{(n)}(x_j)$  and that of a centered gaussian variable with variance  $\sigma^2$  can be written  $\mu_4^{(n)} - 3\sigma^4$ . Tedious calculation shows that

$$\mu_4^{(n)} - 3\sigma^4 = \frac{1}{n} (\mu_4 - 3\sigma^4)$$

where  $\mu_4$  is the  $4^{th}$  moment common to each variable  $\sum_{j=1,p} \lambda_j Y_i(x_j)$ . It then suffices to choose  $n$  so that the difference between the fourth moments is less than a critical value.

**Example 15.2.4.** Suppose we want to simulate a standard gaussian random function  $Y$  with exponential covariance (scale factor  $a$ ). We test the convergence of the  $4^{th}$  moment of  $Y^{(n)}(x) - Y^{(n)}(x+h)$  to that of  $Y(x) - Y(x+h)$ . The latter is equal to  $3\sigma^4 = 12(1-\rho)^2$  with  $\rho = \exp\{-|h|/a\}$ . In the case of the tessellation method or the turning bands method (migration technique), we obtain

$$\mu_4^{(n)} - 3\sigma^4 = \frac{4}{n}(1-\rho)(3\rho-1)$$

The absolute value of this difference is less than  $\epsilon = 0.01$  for all  $0 < \rho \leq 1$  as soon as  $n$  is greater than 135. In the case of the spectral method, we obtain instead

$$\mu_4^{(n)} - 3\sigma^4 = \frac{3}{n}(1 - \rho)(3\rho - 1),$$

Consequently, only 100 functions only are needed to reach the desired precision.

The number of basic functions that has to be simulated is heavily dependent not only the simulation method, but also on the linear combination of interest.

**Example 15.2.5.** Consider the same model as in example 15.2.4, but the linear combination is now  $Y(x) - Y(x+h) + Y(x+2h)$ . In this case, we obtain

$$\mu_4^{(n)} - 3\sigma^4 = \frac{4}{n}(1 - \rho)(-9 + 31\rho - 21\rho^2 + 3\rho^3)$$

for the tessellation method or for the turning bands method, and

$$\mu_4^{(n)} - 3\sigma^4 = \frac{3}{n}(1 - \rho)(35 - 21\rho + 7\rho^2 - \rho^3)$$

for the spectral method. Based on the precision  $\epsilon = 0.01$  and provided that  $\rho \geq 0.3$ , this leads to  $n \approx 400$  and  $n \approx 2000$  respectively. It should also be noted that only 300 basic functions are required for using the spectral method when  $\rho$  is close to 0.

This demonstrates that there is no hope of finding a golden number for the minimal number of basic random functions under all circumstances. The material given in this section can only partially cover the topic and should be seen as a set of suggestions or clues for further work.

## 15.3 Conditional Simulation

Let  $Y$  be a stationary gaussian random function  $\in \mathbb{R}^d$  with mean 0, variance 1 and covariance function  $C$ . We want to carry out a simulation of  $Y$  that honours the conditions  $Y(c) = y(c)$  for any  $c$  within some finite subset  $C$  of  $\mathbb{R}^d$ .

**Definition 15.3.1.** *The simple kriging estimator (or linear regression) of  $Y(x)$  on the  $Y(c)$ 's is the linear combination*

$$Y^*(x) = \sum_{c \in C} \lambda_c Y(c)$$

*that minimizes the mean quadratic error  $E(Y^*(x) - Y(x))^2$ .*

It can be shown that the coefficients  $\lambda_c$ 's are the solutions of the system of linear equations

$$\sum_{c' \in C} \lambda_{c'} C(c, c') = C(c, x) \quad \forall c \in C$$

The corresponding mean quadratic error (the kriging variance) is equal to

$$E (Y^*(x) - Y(x))^2 = 1 - \sum_{c \in C} \lambda_c C(c, x)$$

The simulation algorithm is based on the following theorem.

**Theorem 15.3.1.**  *$Y^*$  and  $Y - Y^*$  are independent gaussian random functions.*

**Proof:** Any linear combination of variables of  $Y^*$  and of  $Y - Y^*$  is a linear combination of variables of  $Y$ , and hence follows a gaussian distribution. This establishes that the random functions  $Y^*$  and  $Y - Y^*$  are jointly gaussian. In order to prove the independence of their spatial distribution, by theorem 15.1.2 it is sufficient to show that any finite linear combination of variables of  $Y^*$  is uncorrelated with any finite linear combination of variables of  $Y - Y^*$ . This is so because  $Y^*(x)$  and  $Y(y) - Y^*(y)$  are uncorrelated whatever the points  $x$  and  $y$  of  $\mathbb{R}^d$ :

$$\begin{aligned} Cov\{Y^*(x), Y(y) - Y^*(y)\} &= Cov\left\{\sum_{c \in C} \lambda_c(x) Y(c), Y(y) - \sum_{d \in C} \lambda_d(y) Y(d)\right\} \\ &= \sum_c \lambda_c(x) \left(C(c, y) - \sum_d \lambda_d(y) C(c, d)\right) = 0 \end{aligned}$$

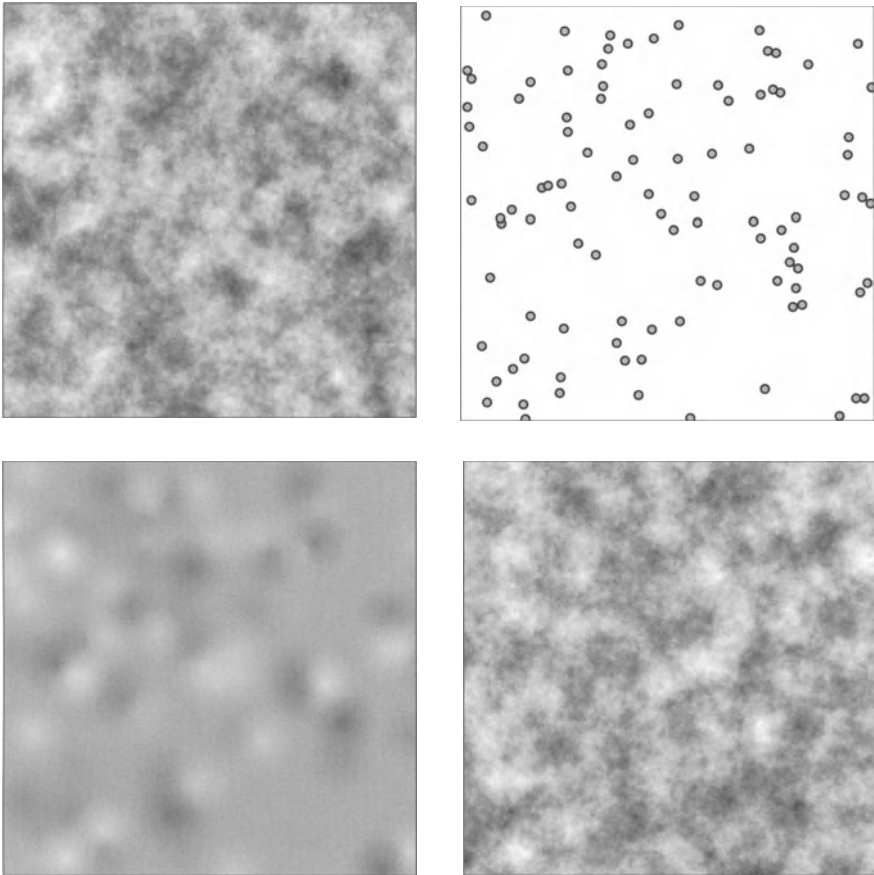
Therefore  $Y$  can be written as a sum of the two independent gaussian random functions  $Y^*$  and  $Y - Y^*$ . This leads to the following algorithm to simulate  $Y$  in the domain  $D$ :

**Algorithm 15.3.1.** *(Conditional simulation of a gaussian random function)*

1. Calculate the kriged estimates  $y^*(x) = \sum_c \lambda_c(x) y(c)$  for each  $x \in D$ .
2. Simulate a gaussian random function with mean 0 and covariance  $C$  in the domain  $D$  and at the conditioning points. Let  $(z(x), x \in D)$  and  $(z(c), c \in C)$  be the generated values.
3. Calculate the kriged estimates  $z^*(x) = \sum_c \lambda_c(x) z(c)$  for each  $x \in D$ .
4. Return  $(y^*(x) + z(x) - z^*(x), x \in D)$ .

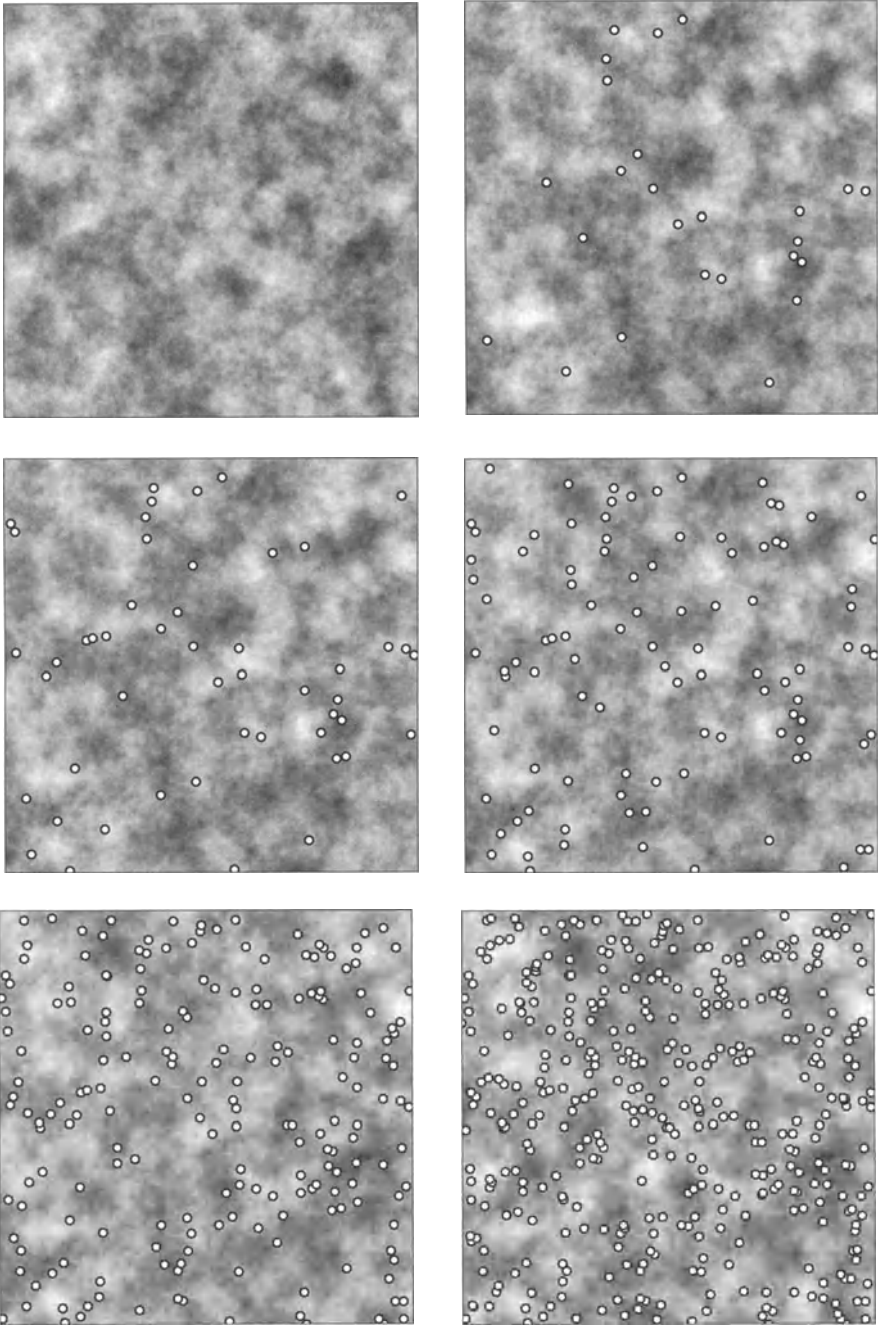


**Remark 15.3.1.** One can verify directly that the conditioning data are honoured. If  $c \in C$ , then  $y^*(c) = y(c)$  and  $z^*(x) = z(c)$ , so that  $y^*(c) + z(c) - z^*(c) = y(c) + z(c) - z(c) = y(c)$ . It can be also expected that if  $x \in D$  is beyond the range of the conditioning data points, then the conditional simulation at  $x$  should not be affected by the data, and therefore should be reduced to a non-conditional simulation. This is effectively what happens:  $y^*(x) \approx 0$  and  $z^*(x) \approx 0$ , so  $y^*(x) + z(x) - z^*(x) \approx 0 + z(x) - 0 \approx z(x)$ .



**Fig. 15.7.** Conditional simulation of a gaussian random function. Top left, a non conditional simulation. Top right, the conditioning data points. Bottom left, the kriged estimates. Bottom right, a conditional simulation

To illustrate this algorithm, we have considered a gaussian random function with a spherical covariance function (scale factor 40). The simulation domain is a square of side length 400. A non-conditional simulation is displayed as a



**Fig. 15.8.** Influence of the data density on the conditional simulation. Top left, a non conditional simulation. Top right, second and third rows, 5 simulations conditioned by 25, 50, 100, 200 and 400 points

reference image in the top left of Figure 15.7. This simulation has been sampled at random to produce a set of 100 conditioning data points (top right). These points have been used as input data for a kriging estimation (bottom left) and for a conditional simulation (bottom right). The difference between both results is striking. Whereas the kriging result is smooth, the conditional simulation has the same degree of spatial variability as the reference image. The difference is not as marked when the covariance function is smooth and has a long range.

As the number of conditioning data points increases, the conditional simulation looks more and more similar to the reference image. This can be observed in Figure 15.8 where the reference image (top left) is shown together with 5 conditional simulations obtained using 25, 50, 100, 200 and 400 data points. In order to facilitate the comparison between the conditional simulations, all of them have been constructed starting from the same non conditional simulation (step 2 of algorithm 15.3.1). Moreover the conditioning data sets are nested. It does not require many conditioning data points to get a conditional simulation close to the reference image. This example suggests that conditional simulation may be an interesting device for image compression.

## Exercises

### 15.1 Starting from

$$\frac{1}{1 + |h|} = \int_0^\infty e^{-t|h|} e^{-t} dt$$

derive an algorithm for simulating a 3D gaussian random function with hyperbolic covariance function.

**15.2** Let  $C$  be an isotropic covariance in 3D with a finite integral range. Show that the associated unidimensional covariance  $C_1$  has a zero integral range.

**15.3** Let  $Y$  be a standard, stationary gaussian random function with covariance function  $C$ . Let  $(x_\alpha)_{\alpha=1,n}$  be a finite set of points of  $\mathbb{R}^d$ . To simplify the notation, we put  $C_{xy} = Cov\{Y(x), Y(y)\}$ ,  $C_{x\alpha} = Cov\{Y(x), Y(x_\alpha)\}$  and  $C_{\alpha\beta} = Cov\{Y(x_\alpha), Y(x_\beta)\}$ .

1) Let  $Y^*(x) = \sum_{\alpha=1}^n \lambda_\alpha(x) Y(x_\alpha)$  and  $Y^*(y) = \sum_{\alpha=1}^n \lambda_\alpha(y) Y(x_\alpha)$  be the kriging estimators of  $Y(x)$  and  $Y(y)$  based on the  $Y(x_\alpha)$ 's. Prove that

$$Cov\{Y^*(x), Y^*(y)\} = \sum_{\alpha, \beta=1}^n C_{x\alpha} B_{\alpha\beta} C_{\beta y}$$

where  $B$  is the inverse of the  $n \times n$  covariance matrix of the  $Y(x_\alpha)$ 's.

2) Now let  $Y^S(x) = y^*(x) + Y(x) - Y^*(x)$  be the probabilistic version of a simulation of the gaussian random function given  $Y^S(x_\alpha) = y_\alpha$  for  $\alpha = 1, n$ . Show that

$$Cov\{Y^S(x), Y^S(y)\} = C_{xy} - \sum_{\alpha, \beta=1}^n C_{x\alpha} B_{\alpha\beta} C_{\beta y}$$

3) Verify that  $Cov\{Y^S(x), Y^S(y)\} = 0$  when  $x$  is a conditioning data point, say  $x_\alpha$  for instance.

**15.4** Prove that the dilution stochastic process with intensity  $\theta$  and function

$$f(t) = t \mathbf{1}_{0 \leq t \leq a}$$

has a spherical covariance function with scale factor  $a$  and variance  $\theta a^3/3$ .

**15.5** Let  $Y$  be a stationary random function with variogram  $\gamma$ . By definition, the (probabilistic version of the) regional variogram of  $Y$  in the domain  $D$  is

$$\Gamma_D(h) = \frac{1}{2 K(h)} \int_{D \cap D_h} [Y(x+h) - Y(x)]^2 dx$$

where  $K$  is the geometric covariogram of  $D$  (cf. section 3.1).

- 1) Show that  $\Gamma_D(h)$  is an unbiased estimator of  $\gamma(h)$ , i.e.  $E\{\Gamma_D(h)\} = \gamma(h)$ .
- 2) Show that in the gaussian case the variance of  $\Gamma_D(h)$  is

$$Var\{\Gamma_D(h)\} = \frac{1}{2 K^2(h)} \int_{\mathbb{R}^d} [\gamma(u+h) + \gamma(u-h) - 2\gamma(u)]^2 K(u) du$$

3) Let  $\Gamma_D^{(n)}$  be the regional variogram of  $Y^{(n)}$  in  $D$ . Show that its variance is a weighted average of the variance of  $\Gamma_D$  and of the variance of the (probabilistic version of the) regional variogram  $\Gamma_{D,i}$  of the  $Y_i$ 's, namely

$$Var\{\Gamma_D^{(n)}(h)\} = \left(1 - \frac{1}{n}\right) Var\{\Gamma_D(h)\} + \frac{1}{n} Var\{\Gamma_{D,i}(h)\}$$

## 16. Gaussian variations

This chapter is devoted to two models constructed from gaussian random functions. The first one is an excursion set obtained by thresholding a gaussian random function. The second one is a step function that results from the joint thresholding of two or more gaussian random functions. It is called a plurigaussian random function.

### 16.1 Excursion set of a gaussian random function

#### 16.1.1 Definition and basic properties

Let  $Y$  be a standard, stationary gaussian random function with covariance  $C$  and variogram  $\gamma = 1 - C$ . Let  $\lambda$  be a real number.

**Definition 16.1.1.** *The  $\lambda$ -level excursion set of  $Y$  is the random set<sup>1</sup>*

$$X_\lambda = \{x \in \mathbb{R}^d : Y(x) \geq \lambda\}$$

By construction the excursion sets form a monotonically decreasing family of random sets

$$\lambda \leq \mu \quad \implies \quad X_\lambda \supset X_\mu$$

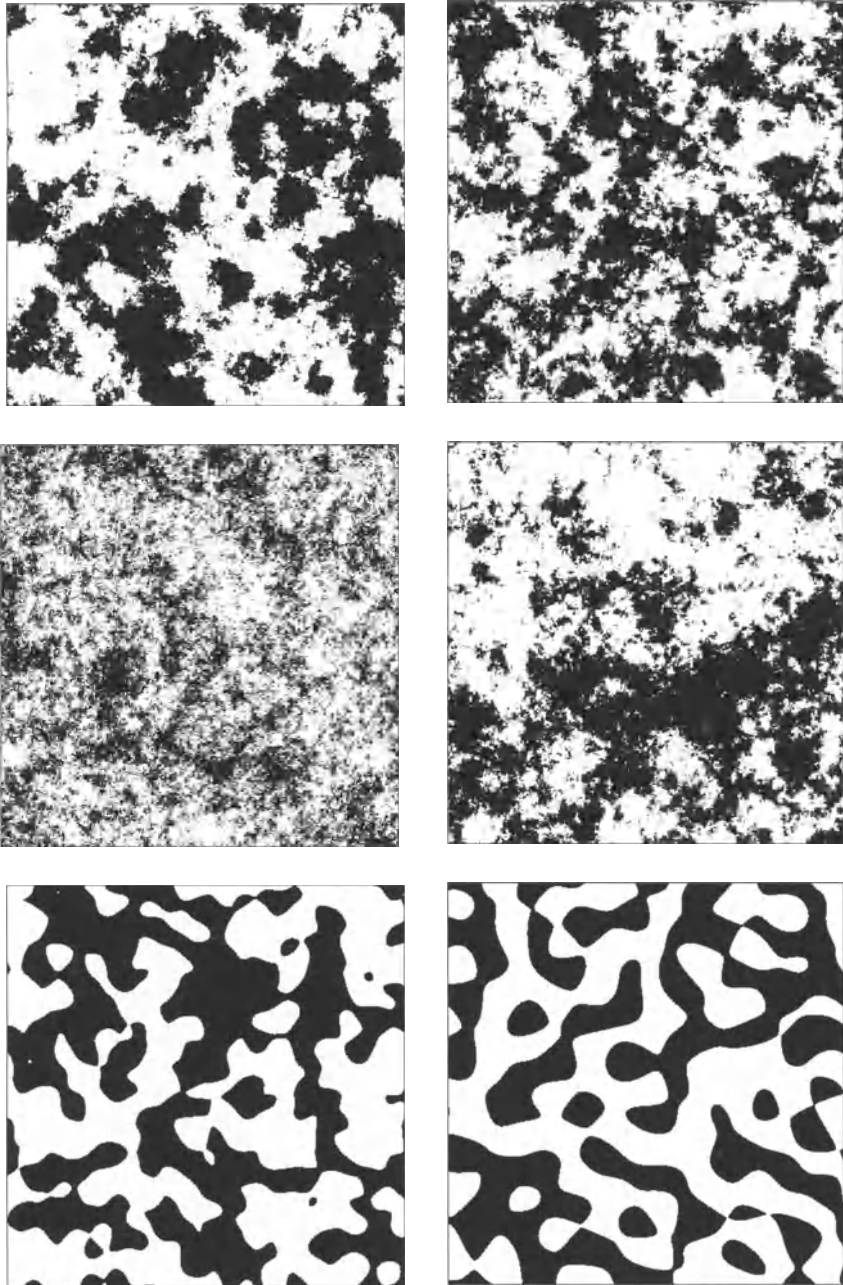
In the case when almost all realizations of  $Y$  are continuous (cf. Adler's sufficient condition in section 15.1.3, the  $X_\lambda$  are random closed sets. Moreover, a theorem by Ylvisaker (1965) (also mentioned in Wschebor (1985)) states that  $Y$  almost surely has no local extremum at level  $\lambda$  in any closed ball of  $\mathbb{R}^d$ . It follows that we almost surely have

$$\bigcap_{\mu < \lambda} X_\mu = X_\lambda \qquad \overline{\bigcup_{\mu > \lambda} X_\mu} = X_\lambda$$

Figure 16.1 depicts the excursion sets at level 0 for the six realizations shown in Figure 15.2. Totally different patterns are observed and their boundaries present a wide range of roughness. The following proposition can help in understanding the variety of the geometries obtained.

---

<sup>1</sup> The usual geostatistical terminology for this model is "truncated gaussian random function" (Matheron et al. (1987); de Fouquet et al. (1989)).



**Fig. 16.1.** Thresholding the six gaussian realizations shown in Figure 15.2 at level 0. These realizations were obtained starting from different covariance functions (from top to bottom and left to right, spherical, exponential, stable, hyperbolic, gaussian and cardinal sine

**Proposition 16.1.1.** *The variogram  $\gamma_\lambda$  of  $X_\lambda$  is equal to*

$$\gamma_\lambda(h) = \frac{1}{\pi} \int_0^{\arcsin \sqrt{\gamma(h)/2}} \exp\left(-\frac{\lambda^2}{2}(1 + \tan^2 t)\right) dt$$

**Proof:** let  $x, y \in \mathbb{R}^d$  with  $x - y = h$ . According to section 3.2.2, the indicator variogram of  $X_\lambda$  is

$$\gamma_\lambda(h) = P\{x \in X_\lambda, y \notin X_\lambda\} = \int_{-\infty}^\lambda \int_\lambda^{+\infty} g_\rho(u, v) du dv$$

where  $g_\rho$  is the probability density function of a bivariate gaussian distribution with correlation coefficient  $\rho = C(h) = 1 - \gamma(h)$

$$g_\rho(u, v) = \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left(-\frac{1}{2} \frac{u^2 + v^2 - 2\rho uv}{1 - \rho^2}\right)$$

By direct calculation, or by using Fourier arguments (Matheron, 1975a), it can be shown that  $g_\rho$  satisfies

$$\frac{\partial g_\rho}{\partial \rho} = \frac{\partial^2 g_\rho}{\partial u \partial v}$$

From this, it follows

$$\frac{\partial \gamma_\lambda}{\partial \rho}(h) = \int_{-\infty}^\lambda \int_\lambda^{+\infty} \frac{\partial^2 g_\rho}{\partial u \partial v}(u, v) du dv = -g_\rho(\lambda, \lambda)$$

By integration, this gives

$$\gamma_\lambda(h) = \int_\rho^1 g_r(\lambda, \lambda) dr = \frac{1}{2\pi} \int_\rho^1 \frac{1}{\sqrt{1 - r^2}} \exp\left(-\frac{\lambda^2}{1 + r}\right) dr$$

and the proof is completed by making the change of variable  $r = \cos(2t)$  with  $0 \leq t \leq \frac{\pi}{2}$ .  $\square$

In particular, this gives for  $\gamma(h) \approx 0$

$$\gamma_\lambda(h) \approx \frac{1}{\pi\sqrt{2}} \sqrt{\gamma(h)} \exp\left\{-\frac{\lambda^2}{2}\right\}$$

If  $\gamma$  has linear behavior at the origin (e.g. exponential or hyperbolic variograms), then  $\gamma_\lambda(h) \propto |h|^{0.5}$ , so that  $X_\lambda$  has an infinite specific perimeter (cf. section 3.2.2, which explains the noisiness of its realizations. In contrast to this, if  $\gamma$  has parabolic behavior at the origin (e.g. gaussian and cardinal sine variograms), then  $\gamma_\lambda(h) \propto |h|$ , and the specific perimeter of  $X_\lambda$  is finite. Its realizations have smooth boundaries.

**Remark 16.1.1.** In the case  $\lambda = 0$ , the formula of proposition 16.1.1 simplifies to

$$\gamma_0(h) = \frac{1}{\pi} \arcsin \sqrt{\frac{\gamma(h)}{2}}$$

**Remark 16.1.2.** It should also be pointed out that the formula of proposition 16.1.1 induces a bijective correspondence between  $\gamma_\lambda(h)$  and  $\gamma(h)$ . This implies that the mapping  $\gamma \rightarrow \gamma_\lambda$  is one-to-one. Notice that it cannot be onto because, as already shown in section 3.2.2, there exist variograms that are not indicator variograms. This raises the question of which variograms are associated with the excursion sets of gaussian random functions. Proposition 16.1.1 provides an answer to that question, but this answer is purely formal. More specifically, let  $\gamma_0$  be an indicator variogram. According to remark 16.1.1,  $\gamma_0$  is the variogram of a 0-level excursion set of a gaussian random function if the function  $\gamma$  defined as

$$\gamma(h) = 2 \sin^2(\pi\gamma_0(h))$$

is a variogram in  $\mathbb{R}^d$ . In general this is difficult to prove. As an example, the case when  $\gamma_0$  is an exponential variogram is considered in Appendix 1.

### 16.1.2 Conditional simulation

In this section, the parameters of the model ( $\gamma = 1 - C, \lambda$ ) are assumed to be known. Let  $A_0$  and  $A_1$  be disjoint finite subsets of  $\mathbb{R}^d$ . Our objective is to carry out simulations of  $X_\lambda$  in a field  $D$ , subject to the conditions  $a \notin X_\lambda$  for any  $a \in A_0$  and  $a \in X_\lambda$  for each  $a \in A_1$ . Such simulations exist since  $P\{A_0 \cap X_\lambda = \emptyset, A_1 \subset X_\lambda\} \neq 0$ .

In order to simplify the presentation of the algorithm developed by Matheron et al. (1987), concise notation is introduced. Let  $A = A_0 \cup A_1$ . For each  $a \in A$ , we put  $I_a = ]-\infty, \lambda[$  if  $a \in A_0$  and  $[\lambda, +\infty[$  if  $a \in A_1$ . Let us assume that  $A$  is totally ordered. For each mapping  $y$  on  $A$  we put by  $y_a = y(a)$  and define  $y_A$  to be the vector whose  $a^{th}$  component is  $y_a$ . We similarly define the random vector  $Y_A$  and the cartesian product  $I_A$ . In particular, we take  $y_A \in I_A$  to mean that  $y_a \in I_a$  for each  $a \in A$ .

The conditional simulation of  $X_\lambda$  corresponds to the simulation of  $Y$  subject to the constraint  $Y_A \in I_A$ . The statistical properties of  $Y$  are given by all of the multivariate distributions of the vectors  $Y_X$  that can be built from any finite and totally ordered subset  $X$  of  $\mathbb{R}^d$  (see section 2.2). The p.d.f. of  $Y_X$  given  $Y_A \in I_A$  can be written

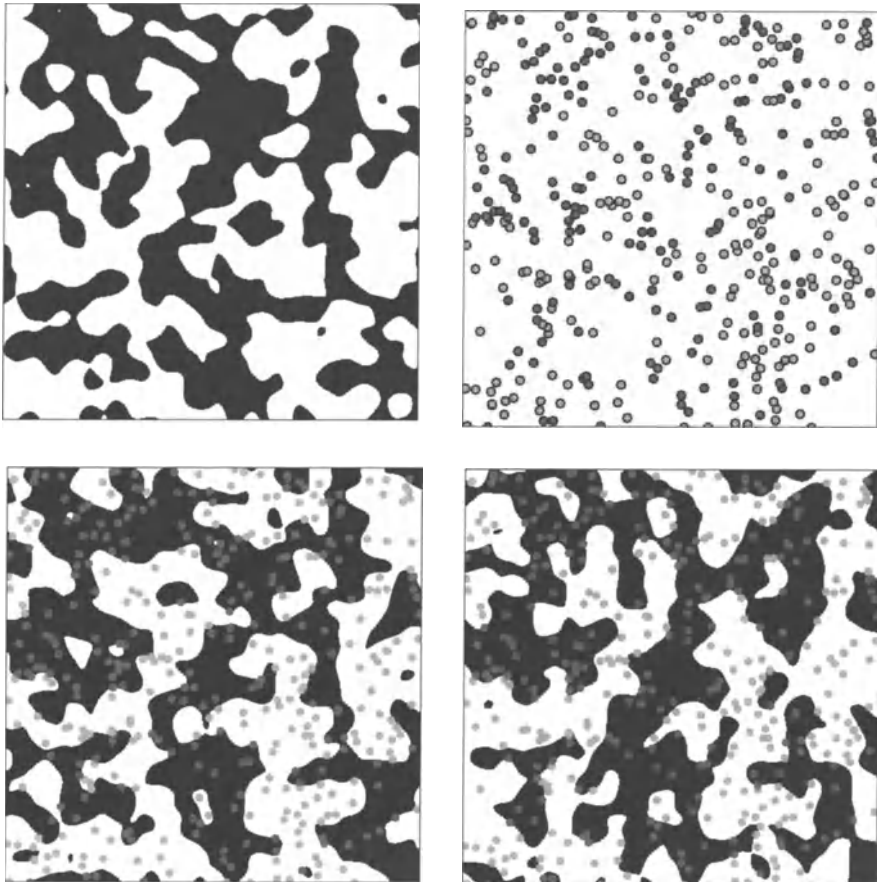
$$g(y_X \mid Y_A \in I_A) = \int g(y_X \mid Y_A = y_A) g(y_A \mid Y_A \in I_A) dy_A$$



where  $g(y_A \mid Y_A \in I_A)$  is the p.d.f. of  $Y_A$  given the constraints  $Y_A \in I_A$  and  $g(y_X \mid Y_A = y_A)$  that of  $Y_X$  given the conditions  $Y_A = y_A$ . This equation leads to the following algorithm for the simulation of  $X_\lambda$ :

**Algorithm 16.1.1.** (*Conditional simulation of an excursion set of a gaussian random function*)

1. Generate  $y_A \sim g(\cdot \mid Y_A \in I_A)$ .
2. For each  $x \in D$ , generate  $y(x) \sim g(\cdot \mid Y_A = y_A)$ .
3. Return  $\{x \in D : y(x) \geq \lambda\}$ .



**Fig. 16.2.** Conditional simulation of an excursion set of a gaussian random function. Top left, a non conditional simulation. Top right, the conditioning data points (the red points belong to the black phase, the green ones to the white) . Bottom, two conditional simulations

Since step 2 was studied in the previous chapter and step 3 is straightforward, only step 1 needs to be detailed. Its object is to produce realizations of  $Y_A$  given  $Y_A \in I_A$ . To achieve this, Freulon (1992, 1994) adopted an iterative procedure based on the Gibbs sampler (Geman and Geman, 1984). It consists of building a stationary discrete Markov chain  $(Y_A(n), n \geq 0)$  whose state space is  $\mathbb{R}^A$ , and whose equilibrium distribution is the one to be simulated, namely  $g(\cdot \mid Y_A \in I_A)$ . In what follows we denote by  $y_A^a$  the vector  $y_A$  without its  $a^{\text{th}}$  component.

A simple way to initialize the Markov chain is to take  $Y_A(0) = y_A \in I_A$ . An alternative is to simulate  $Y_a(0) \sim g(\cdot \mid I_a)$  (conditional Gaussian p.d.f.) for each  $a \in A$ .

Let  $(a_n, n \in \mathbb{N})$  be a sequence of elements of  $A$  such that each element is encountered infinitely many times (for instance, a periodic sequence or the realization of a positive recurrent Markov chain). At the  $n^{\text{th}}$  iteration, only the component of order  $a_n$  is modified. Suppose  $Y_A(n) = y_A$ . Then we put  $Y_A^a(n+1) = Y_A^a(n)$  and simulate  $Y_a(n+1) \sim g(\cdot \mid y_A^a, I_a)$ .

It can be shown that the p.d.f. of  $Y_A(n)$  tends to  $g(\cdot \mid I_A)$  as  $n$  becomes very large. However the rate of convergence toward equilibrium is not well known and is the object of ongoing research (Chan, 1993; Rosenthal, 1993; see also a general discussion by Galli and Gao on recent convergence results).

The p.d.f.  $g(\cdot \mid y_A^a, I_a)$  is the restriction to  $I_a$  of a gaussian p.d.f. with mean  $y_a^* = \sum_{b \neq a} \lambda_b^a y_b$  (kriged estimate of  $Y_a$  on the  $Y_b$ 's,  $b \neq a$ ) and variance  $\sigma_a^2$  (kriging variance). It can be simulated using the acceptance-rejection method.

The kriging weights and the kriging variances can be obtained simultaneously from the inverse  $M^{-1} = [M_{ab}^{-1}]_{a,b \in A}$  of the covariance matrix  $M = [C(a - b)]_{a,b \in A}$ . It can be shown that

$$\lambda_b^a = -\frac{M_{ab}^{-1}}{M_{aa}^{-1}} \quad \sigma_a^2 = \frac{1}{M_{aa}^{-1}} \quad (a, b \in A)$$

**Remark 16.1.3.** It is also possible to carry out step 1 by resorting to a hit-and-run algorithm (see section 8.3). This amounts to simulating a uniform point in the subgraph of the density function

$$h(y_A) = \exp\left\{-\frac{1}{2} y_A' M^{-1} y_A\right\} 1_{y_A \in I_A} \quad y_A \in \mathbb{R}^A$$

If the number of conditioning data points is very large, the matrix  $M$  cannot be inverted numerically and  $h$  cannot be evaluated. In fact, the calculation of  $M^{-1}$  can be avoided altogether. Instead of simulating  $Y_A$  directly, Galli and Gao simulate first  $z_A = M^{-1} Y_A$  and then derive  $Y_A = M z_A$ . This is possible because  $z_A$  has a p.d.f. proportional to

$$k(z_A) = \exp\left\{-\frac{1}{2} z_A' M z_A\right\} 1_{M z_A \in I_A} \quad z_A \in \mathbb{R}^A$$

which depends on  $M$  and not on its inverse.

An illustration of algorithm 16.1.1 is presented in Figure 16.2. The starting point is the non conditional simulation shown in the bottom left of Figure 16.1. From this realization, 400 uniform points have been selected independently to act as conditioning data (top right). The bottom of Figure 16.2 shows two conditional simulations. Because of the large number of conditioning data points, the variability of the conditional simulations is limited. Note however that their connectivity properties are not identical. Pairs of conditioning data points may be connected in one simulation and not in another one. Top to bottom percolation takes place in the black phase of the left hand-side simulation and in the white phase of the right hand-side simulation. Allard (1993, 1994) devised a conditional simulation algorithm which accepts the connectivity relationships between points as conditioning data.

## 16.2 Plurigaussian random function

The excursion sets of (gaussian) random functions are models designed for representing physical phenomena consisting of two distinct phases only. One way of extending these models to accommodate phenomena with more than two phases, say 3 phases, is to threshold the random function  $Y$  at two different levels  $\lambda < \mu$  and to define the 3 phases as  $\{Y < \lambda\}$ ,  $\{\lambda \leq Y < \mu\}$  and  $\{Y \geq \mu\}$ . This is not always satisfactory in practice because if  $Y$  has continuous realizations then both phases  $\{Y < \lambda\}$  and  $\{Y \geq \mu\}$  cannot be in contact. This observation prompted Galli et al. (1994) to design a more general model based on thresholding several gaussian random functions. This section gives a succinct description of the model and of its implications. For a more general presentation, the reader can consult the book by Armstrong et al. (2001). This book contains numerous examples from the geosciences.

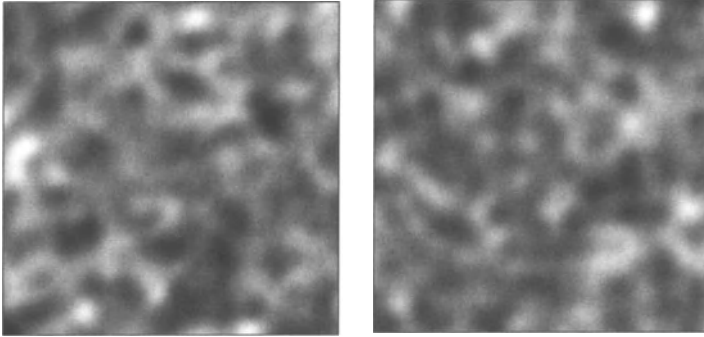
### 16.2.1 Definition and examples

The basic ingredients required for the construction of a plurigaussian random function are

- two standardized gaussian random functions  $Y$  and  $Z$  defined on  $\mathbb{R}^d$ , with respective covariance functions  $C_Y$  and  $C_Z$ . Both random functions are assumed to be independent, although dependency relationships between them can be considered in certain cases (Armstrong et al., (2001)).
- A family  $(D_i, i \in I)$  of subsets of  $\mathbb{R}^2$  constituting a partition (that is  $\cup_{i \in I} D_i = \mathbb{R}^2$  and  $D_i \cap D_j = \emptyset$  for  $i \neq j$ ). The indices are numerical or categorical. The set of indices  $I$  is usually finite, but may be countable.

**Definition 16.2.1.** *A plurigaussian random function with underlying random functions  $Y$  and  $Z$ , and partition  $(D_i, i \in I)$  is the random function  $X$  from  $\mathbb{R}^d$  to  $I$  that satisfies*

$$X(x) = i \quad \iff \quad (Y(x), Z(x)) \in D_i$$



**Fig. 16.3.** The realizations of both underlying gaussian random functions

Figure 16.3 shows realizations of two independent gaussian random functions with the same gaussian variogram. Its scale factor is a tenth of the width of the simulation field. Both gaussian realizations have been used to produce the realizations of the six different plurigaussian random functions depicted in Figure 16.4. Each plurigaussian random function has its own partition indicated as a flag below each realization.

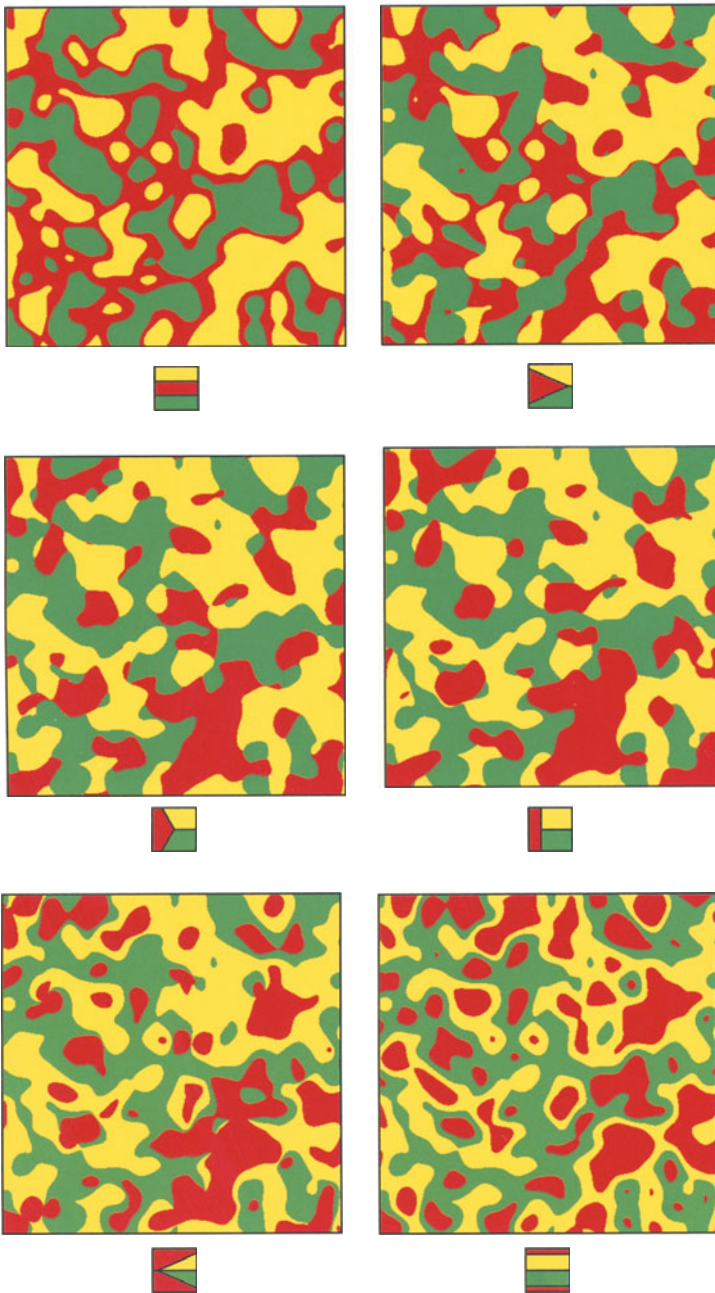
Each partition has three cells. Their size and shape have been designed so that all phase occur with the same frequency. Whereas the green and the yellow cells are symmetric about the  $x$ -axis, the red one is obtained by moving the triple point common to the three cells along the  $x$ -axis from  $+\infty$  to  $-\infty$ .

In the top left partition, the triple point is at  $+\infty$ . In this case, the red cell is a horizontal stripe between the yellow and the green one, thus preventing the yellow and the green phases to be in contact. Because the shape of the cells,  $(Y(x), Z(x))$  belongs to  $D_i$  irrespectively of  $Y(x)$ . The realization is equivalent to a realization obtained by thresholding  $Z$  at two different levels.

As the triple point moves backward, direct contacts between green and yellow cells become possible. Accordingly, the red phase gradually loses its separating power.

In the middle left partition, the cells are globally invariant under a  $120^\circ$  rotation around the triple point. The three phases turn out to have exactly the same statistical properties. Their joint distribution is invariant under permutation. There is no preferential contact between any two of them.

The red cell of the middle right partition is a vertical stripe. The associated phase is obtained by thresholding  $Y$  below a certain level. In the residual space, the green and the yellow phases are delimited by a threshold on  $Z$ . This hierarchichal construction is perceptible on the realization. The yellow and the green phases seem to be partly hidden by the red one.



**Fig. 16.4.** Realizations of six plurigaussian random functions. All of them have been built starting from the same realizations of the underlying gaussian random functions (see Figure 16.3). Each plurigaussian random function has its own partition represented by a flag below each realization

In the bottom left partition, the green and the yellow cells are contained in a small angle that the red cell starts to envelop. Little by little, the components of the red phase tend to split into blobs.

In the last example, the triple point is rejected to  $-\infty$ . This gives a partition with four horizontal stripes. The red cell is divided into two components located above and below the green and the yellow cells. The realization obtained does not depend on  $Y$ . It is obtained by thresholding  $Z$  at three different levels. The red phase is the union of two excursion sets of  $Z$ , the first one above the largest threshold and the second one below the smallest threshold. Because both thresholds have a large amplitude, these excursion sets appear as little blobs. Depending on the sign of the threshold, the blobs are either surrounded by the green phase or by the yellow one.

### 16.2.2 Spatial distribution

Let  $x \in \mathbb{R}^d$ . The distribution of  $X(x)$  is given by

$$P\{X(x) = i\} = P\{(Y(x), Z(x)) \in D_i\} = \int_{D_i} g(t)g(u) dt du$$

where  $g$  is the standard gaussian p.d.f.

Let  $x, y \in \mathbb{R}^d$ . In order to get the bivariate distribution of  $(X(x), X(y))$ , we can write

$$\begin{aligned} P\{X(x) = i, X(y) = j\} &= P\{(Y(x), Z(x)) \in D_i, (Y(y), Z(y)) \in D_j\} \\ &= \int_{(t,u) \in D_i} \int_{(v,w) \in D_j} g_v(t, v) g_\zeta(u, w) dt du dv dw \end{aligned}$$

where  $g_v$  and  $g_\zeta$  stand for the standardized bivariate gaussian p.d.f.'s with respective correlation coefficients  $v = C_Y(x - y)$  and  $\zeta = C_Z(x - y)$ . The next step is to expand  $g_v$  and  $g_\zeta$  in terms of *Hermite polynomials*<sup>2</sup>

$$\begin{aligned} g_v(t, v) &= \sum_{n=0}^{\infty} \frac{v^n}{n!} H_n(t) H_n(v) g(t) g(v) \\ g_\zeta(u, w) &= \sum_{p=0}^{\infty} \frac{\zeta^p}{p!} H_p(u) H_p(w) g(u) g(w) \end{aligned}$$

Let us also put

<sup>2</sup> The Hermite polynomial of degree  $n$  is defined here as

$$H_n(x) = \frac{1}{g(x)} \frac{d^n g}{dx^n}(x) \quad n \geq 0$$

They constitute an orthogonal basis of  $L^2(\mathbb{R}, g)$ .

$$\begin{aligned} \theta_{n,p}(i) &= \int_{D_i} H_n(t) H_p(u) g(t) g(u) dt du \\ \theta_{n,p}(j) &= \int_{D_j} H_n(v) H_p(w) g(v) g(w) dv dw \end{aligned}$$

This gives

$$P\{X(x) = i, X(y) = j\} = \sum_{n,p=0}^{\infty} \frac{v^n \zeta^p}{n! p!} \theta_{n,p}(i) \theta_{n,p}(j)$$

More generally, we have

**Theorem 16.2.1.** *The spatial distribution of a plurigaussian random function is given by*

$$P\left\{\bigwedge_{j \in J} X(x_j) = i_j\right\} = \sum_{m,n \in \mathbb{N}^{\mathcal{P}_2(J)}} \prod_{K \in \mathcal{P}_2(J)} \frac{\rho_K^{m_K} \zeta_K^{n_K}}{m_K! n_K!} \prod_{j \in J} \theta_{m_j, n_j}(i_j)$$

for every finite family  $(x_j, j \in J)$  of points of  $\mathbb{R}^d$  and for every finite family  $(i_j, j \in J)$  of elements of  $I$ . Here  $\mathcal{P}_2(J)$  denotes the family of two-element subsets of  $J$ , and  $\mathbb{N}^{\mathcal{P}_2(J)}$  is the set of mappings from  $\mathcal{P}_2(J)$  to  $\mathbb{N}$ .  $m_K$  and  $n_K$  stand for the integer values taken by  $m, n \in \mathbb{N}^{\mathcal{P}_2(J)}$  at  $K \in \mathcal{P}_2(J)$ . Moreover for each  $j \in J$ ,  $m_j$  and  $n_j$  are the sums of the  $m_K$  and the  $n_K$  extended over all  $K$  that contain  $j$ . Finally, if  $K = \{k, l\}$ , then  $\rho_K = C_Y(x_k - x_l)$  and  $\zeta_K = C_Z(x_k - x_l)$ .

**Proof:** Let us start with

$$P\left\{\bigwedge_{j \in J} X(x_j) = i_j\right\} = P\left\{\bigwedge_{j \in J} (Y(x_j), Z(x_j)) \in D_{i_j}\right\}$$

Let  $g_Y$  and  $g_Z$  be the multivariate gaussian p.d.f.'s of  $(Y(x_j), j \in J)$  and  $(Z(x_j), j \in J)$ . We have

$$P\left\{\bigwedge_{j \in J} X(x_j) = i_j\right\} = \int \dots \int g_Y(u) g_Z(v) \prod_{j \in J} 1_{(u_j, v_j) \in D_{i_j}} du dv$$

where  $u = (u_j, j \in J)$  and  $v = (v_j, j \in J)$ . It can be shown that  $g_Y$  and  $g_Z$  admit the hermitian expansions (see Appendix 2)

$$\begin{aligned} g_Y(u) &= \sum_{m \in \mathbb{N}^{\mathcal{P}_2(J)}} \prod_{K \in \mathcal{P}_2(J)} \frac{\rho_K^{m_K}}{m_K!} \prod_{j \in J} H_{m_j}(u_j) g(u_j) \\ g_Z(v) &= \sum_{n \in \mathbb{N}^{\mathcal{P}_2(J)}} \prod_{K \in \mathcal{P}_2(J)} \frac{\zeta_K^{n_K}}{n_K!} \prod_{j \in J} H_{n_j}(v_j) g(v_j) \end{aligned}$$

It then suffices to replace each multivariate gaussian p.d.f. by its hermitian expansion and to integrate w.r.t. each  $D_{i_j}$  to obtain the desired result.  $\square$

Of course, further simplifications occur in the case when the  $D_j$ 's are rectangles or stripes with their boundaries parallel to the coordinate axes.

### 16.2.3 Conditional simulation

The problem addressed now is the simulation of the plurigaussian random function  $X$  in the field  $D$  given the values  $(i_j, j \in J)$  taken at the data points  $(c_j, j \in J)$ .

The design of the conditional simulation algorithm is similar to that of an excursion set of a gaussian random function. As it is just a multivariate extension of algorithm 16.1.1, we give it without further comment.

**Algorithm 16.2.1.** (*Conditional simulation of a plurigaussian random function*)

1. Generate jointly  $(y(c_j), j \in J)$  and  $(z(c_j), j \in J)$  given  $(X(c_j) = i_j, j \in J)$ .
2. For each  $x \in D$ , generate independently  $y(x) \sim g(\cdot \mid Y(c_j) = y(c_j), j \in J)$  and  $z(x) \sim g(\cdot \mid Z(c_j) = z(c_j), j \in J)$ .
3. For each  $x \in D$  return  $\sum_{i \in I} i 1_{(y(x), z(x)) \in D_i}$ .

Step 1 can be performed either by running a Gibbs sampler or a hit-and-run algorithm. Step 2 consists of the conditional simulation of two independent gaussian random functions. To illustrate this algorithm, we have considered the plurigaussian random function, one realization of which is depicted at the middle right of Figure 16.4 (see the top left of Figure 16.5). From this non conditional simulation 400 points have been uniformly selected to act as conditioning data. The bottom of Figure 16.5) shows two conditional simulations.

## Appendix 1:

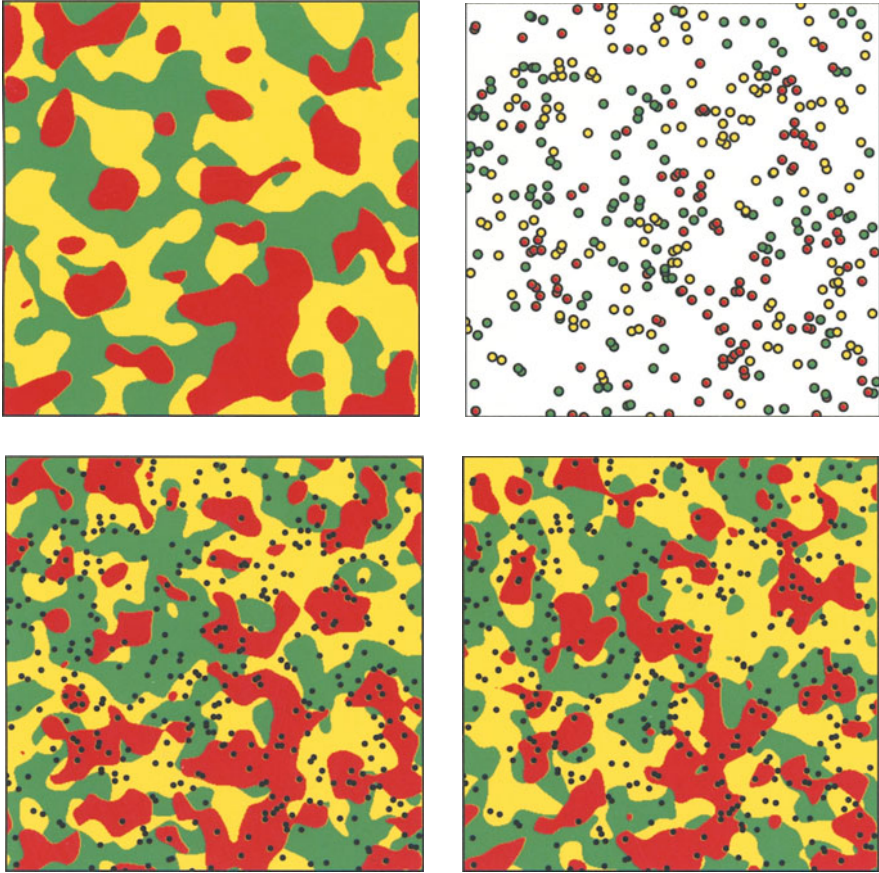
In this appendix we are going to show that the exponential variogram

$$\gamma_0(h) = \frac{1}{4} \left( 1 - e^{-|h|} \right)$$

is the variogram of an excursion set of a gaussian random function at level 0. According to remark 16.1.1, this amounts to showing that

$$\gamma(h) = 2 \sin^2 \left( \frac{\pi}{4} \left( 1 - e^{-|h|} \right) \right) = 1 - \cos \left( \frac{\pi}{2} \left( 1 - e^{-|h|} \right) \right) = 1 - \sin \left( \frac{\pi}{2} e^{-|h|} \right)$$





**Fig. 16.5.** Conditional simulation of an plurigaussian random function. The flag is that of median right of Figure 16.4. Top left, a non conditional simulation. Top right, the conditioning data points. Bottom, two conditional simulations

is a variogram, or equivalently that

$$C(h) = \sin\left(\frac{\pi}{2}e^{-|h|}\right)$$

is a covariance function. Because of proposition 3.2.3, it suffices to prove that  $-\ln C$  is a variogram. A starting point proposed by Matheron is the product expansion of the sine function

$$\sin x = x \prod_{k=1}^{\infty} \left(1 - \frac{x^2}{\pi^2 k^2}\right)$$

This gives

$$C(h) = \frac{\pi}{2} e^{-|h|} \prod_{k=1}^{\infty} \left(1 - \frac{e^{-2|h|}}{4k^2}\right)$$

and taking the logarithm

$$\ln C(h) = \ln \frac{\pi}{2} - |h| + \sum_{k=1}^{\infty} \ln \left( 1 - \frac{e^{-2|h|}}{4k^2} \right)$$

Letting  $h = 0$  gives

$$0 = \ln \frac{\pi}{2} + \sum_{k=1}^{\infty} \ln \left( 1 - \frac{1}{4k^2} \right)$$

Subtracting these two expressions, we obtain

$$\ln C(h) = -|h| + \sum_{k=1}^{\infty} \left[ \ln \left( 1 - \frac{e^{-2|h|}}{4k^2} \right) - \ln \left( 1 - \frac{1}{4k^2} \right) \right]$$

The logarithms are now expanded into series

$$\ln C(h) = -|h| + \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} \frac{1 - e^{-2n|h|}}{n4^n k^{2n}}$$

This expression can be simplified by using the Riemann zeta function  $\zeta(s) = \sum_{k=1}^{\infty} k^{-s}$

$$\ln C(h) = -|h| + \sum_{n=1}^{\infty} \frac{\zeta(2n)}{n4^n} \left( 1 - e^{-2n|h|} \right)$$

Differentiating w.r.t.  $|h|$  and taking  $h = 0$  we obtain

$$0 = -1 + 2 \sum_{n=1}^{\infty} \frac{\zeta(2n)}{n4^n}$$

The coefficient  $-1$  of  $|h|$  is replaced by this expression and we finally get

$$\ln C(h) = \sum_{n=1}^{\infty} \frac{\zeta(2n)}{n4^n} \left( -2n|h| + 1 - e^{-2n|h|} \right)$$

Therefore, to conclude the proof it remains to be shown that the function  $h \rightarrow |h| - (1 - e^{-|h|})$  is a variogram in  $\mathbb{R}^d$ . We do so by proving that  $\gamma$  admits a spectral representation with a positive density. Using the notation  $\nu = (d + 1)/2$ , the spectral representations of the linear and the exponential variograms can be written as

$$\begin{aligned} |h| &= \int_{\mathbb{R}^d} \frac{1 - \cos \langle u, h \rangle}{|u|^2} \frac{\Gamma(\nu)}{\pi^\nu} \frac{1}{|u|^{2\nu-2}} du \\ 1 - e^{-|h|} &= \int_{\mathbb{R}^d} \frac{1 - \cos \langle u, h \rangle}{|u|^2} \frac{\Gamma(\nu)}{\pi^\nu} \frac{|u|^2}{(1 + |u|^2)^\nu} du \end{aligned}$$

Hence

$$|h| - (1 - e^{-|h|}) = \frac{\Gamma(\nu)}{\pi^\nu} \int_{\mathbb{R}^d} \frac{1 - \cos \langle u, h \rangle}{|u|^{2\nu}} f(u) du$$

with

$$f(u) = \frac{1}{|u|^{2\nu-2}} - \frac{|u|^2}{(1 + |u|^2)^\nu} = \frac{1}{|u|^{2\nu-2}} \left[ 1 - \left( \frac{|u|^2}{1 + |u|^2} \right)^\nu \right] > 0$$

### Appendix 2:

In this appendix, we calculate the hermitian expansion of the multivariate p.d.f. of the gaussian vector  $(Y(x_j), j \in J)$ .

Let us start with its Fourier transform

$$\widehat{g}_Y(u) = \exp \left( -\frac{1}{2} \sum_{j,k \in J} u_j u_k \rho_{jk} \right)$$

where  $\rho_{jk}$  stands for the correlation between  $Y(x_j)$  and  $Y(x_k)$ . In the double sum, we distinguish between the terms associated with identical indices and the other ones. After a minor change, the Fourier transform can be rewritten

$$\widehat{g}_Y(u) = \exp \left( -\frac{1}{2} \sum_{j \in J} u_j^2 \right) \exp \left( \sum_{j \neq k} (-iu_j) (-iu_k) \rho_{jk} \right)$$

Now the right hand side exponential term is expanded

$$\widehat{g}_Y(u) = \exp \left( -\frac{1}{2} \sum_j u_j^2 \right) \sum_{n=0}^\infty \frac{1}{n!} \left( \sum_{j \neq k} (-iu_j) (-iu_k) \rho_{jk} \right)^n$$

and so is the power term

$$\widehat{g}_Y(u) = \exp \left( -\frac{1}{2} \sum_j u_j^2 \right) \sum_{n=0}^\infty \sum_{j \neq k} \prod \frac{\rho_{jk}^{n_{jk}}}{n_{jk}!} (-iu_j)^{n_{jk}} (-iu_k)^{n_{jk}}$$

In this formula, the inner sum is extended over all  $n_{jk} \geq 0$  such that  $\sum_{j \neq k} n_{jk} = n$ . The sum over  $n$  is now eliminated. To do this, note that for each  $j \in J$  the term  $(-iu_j)$  appears raised to a power that is the sum of the  $n_{kl}$  with either  $k = j$  or  $l = j$ . Let us denote it by  $n_j$ .

$$\widehat{g}_Y(u) = \exp \left( -\frac{1}{2} \sum_j u_j^2 \right) \sum_{j \neq k} \sum_{n_{jk}=0}^\infty \prod \frac{\rho_{jk}^{n_{jk}}}{n_{jk}!} \prod_j (-iu_j)^{n_j}$$

But

$$\begin{aligned}
(-iu_j)^{n_j} e^{-\frac{u_j^2}{2}} &= (-iu_j)^{n_j} \hat{g}(u_j) \\
&= \int_{-\infty}^{+\infty} e^{iu_j x_j} \frac{d^{n_j} g}{dx^{n_j}}(x_j) dx_j \\
&= \int_{-\infty}^{+\infty} e^{iu_j x_j} H_{n_j}(x_j) g(x_j) dx_j
\end{aligned}$$

It follows

$$\widehat{g_Y}(u) = \sum_{j \neq k} \sum_{n_{jk}=0}^{\infty} \prod_{j \neq k} \frac{\rho_{jk}^{n_{jk}}}{n_{jk}!} \prod_{j \in J} \int_{-\infty}^{+\infty} e^{iu_j x_j} H_{n_j}(x_j) g(x_j) dx_j$$

The Fourier transform is inverted and we finally obtain

$$g_Y(y) = \sum_{j \neq k} \sum_{n_{jk}=0}^{\infty} \prod_{j \neq k} \frac{\rho_{jk}^{n_{jk}}}{n_{jk}!} \prod_{j \in J} H_{n_j}(y_j) g(y_j)$$

For instance, if  $J$  has three elements denoted by 1, 2 and 3, then

$$g_Y(y) = \sum_{i,j,k=0}^{\infty} \frac{\rho_{12}^k}{k!} \frac{\rho_{13}^j}{j!} \frac{\rho_{23}^i}{i!} H_{j+k}(y_1) H_{i+k}(y_2) H_{i+j}(y_3) g(y_1) g(y_2) g(y_3)$$

## 17. Substitution random functions

A convenient way to introduce *substitution random functions* is by analogy with mapping techniques. Roughly speaking, a map displays spatial information (e.g. relief) coded in a conventional way (e.g. green for plains, blue for seas etc...). More formally, it can be expressed as  $Y \circ T$  where  $T$  denotes a spatial information function and  $Y$  stands for a coding process.

Similarly a substitution random function is defined as  $Y \circ T$  where  $T$  is a random function (usually called the *directing function*), and  $Y$  is a stochastic process (the *coding process*). There are some limitations on both these ingredients to ensure that a substitution random function is second-order stationary (see section 17.1). However these limitations are mild, and the directing function and the coding process can often be chosen to produce realizations with a prespecified geometry or topology (see section 17.2). The last section deals with the conditional simulation of substitution random functions.

### 17.1 Definition and main properties

**Definition 17.1.1.** *A random function  $Z$  defined on  $\mathbb{R}^d$  is called substitution random function with directing function  $T$  and coding process  $Y$  if it can be written as*

$$Z(x) = Y[T(x)] \quad x \in \mathbb{R}^d$$

*The directing random function  $T$  is defined on  $\mathbb{R}^d$  and has stationary increments. The coding process  $Y$  is a stationary stochastic process.  $T$  and  $Y$  are independent<sup>1</sup>.*

Let  $x, y \in \mathbb{R}^d$ . If  $T$  has stationary increments, then the distribution of the difference  $T(x) - T(y)$  depends only on the vector  $x - y$ . More precisely

$$T(x) - T(y) \stackrel{\mathcal{D}}{=} T(x - y) - T(o)$$

---

<sup>1</sup> This definition may remind readers of the subordination technique used to transform a Markov chain into another one (Feller, 1971). In that case, the directing function  $T$  is unidimensional and has positive increments.

A random function with stationary increments is often defined up to a random variable. So there is no loss of generality in assuming  $T(o) = 0$ , and we can write

$$T(x) - T(y) \stackrel{\mathcal{D}}{=} T(x - y)$$

For  $x = o$ , this gives

$$-T(y) \stackrel{\mathcal{D}}{=} T(-y)$$

From this it follows  $T(y) \stackrel{\mathcal{D}}{=} T(-y)$  provided that the distribution of  $T(y)$  is symmetric around the origin.

More generally, to say that  $T$  has stationary increments means that for any finite sequence  $(x_i, i \in I \cup \{0\})$  of points of  $\mathbb{R}^d$  the two random vectors  $(T(x_i) - T(x_0), i \in I)$  and  $(T(x_i - x_0), i \in I)$  have the same multivariate distribution.

To simplify the presentation, the following properties of the substitution random functions will be established only in the case when the directing function and the coding process are discrete. The generalization to the continuous case is immediate.

**Proposition 17.1.1.** *Z is stationary.*

**Proof:** We have to show that

$$P \left\{ \bigwedge_{i \in I} Y[T(x_i + h)] = z_i \right\} = P \left\{ \bigwedge_{i \in I} Y[T(x_i)] = z_i \right\}$$

for any finite sequence  $(x_i, i \in I)$  of points of  $\mathbb{R}^d$ , any finite sequence  $(z_i, i \in I)$  of values and for any vector  $h \in \mathbb{R}^d$ . Putting  $y_i = x_i + h$  for any  $i \in I$ , we first apply the independence of  $T$  and  $Y$  to obtain

$$\begin{aligned} P \left\{ \bigwedge_{i \in I} Y[T(y_i)] = z_i \right\} &= \sum_{(t_i)} P \left\{ \bigwedge_{i \in I} T(y_i) = t_i, Y(t_i) = z_i \right\} \\ &= \sum_{(t_i)} P \left\{ \bigwedge_{i \in I} T(y_i) = t_i \right\} P \left\{ \bigwedge_{i \in I} Y(t_i) = z_i \right\} \end{aligned}$$

Because  $T$  has stationary increments we can also write

$$\begin{aligned} P \left\{ \bigwedge_{i \in I} T(y_i) = t_i \right\} &= P \left\{ \bigwedge_{i \in I} T(x_i) - T(-h) = t_i \right\} \\ &= \sum_t P \left\{ \bigwedge_{i \in I} T(x_i) = t_i + t, T(-h) = t \right\} \end{aligned}$$

On the other hand, the stationarity of  $Y$  implies

$$P \left\{ \bigwedge_{i \in I} Y(t_i) = z_i \right\} = P \left\{ \bigwedge_{i \in I} Y(t_i + t) = z_i \right\}$$

for any  $t \in \mathbb{R}$ . Consequently, putting  $u_i = t_i + t$  for any  $i \in I$ , we can write

$$\begin{aligned} P \left\{ \bigwedge_{i \in I} Y[T(y_i)] = z_i \right\} &= \sum_{(u_i), t} P \left\{ \bigwedge_{i \in I} T(x_i) = u_i, T(-h) = t \right\} \\ &\quad \times P \left\{ \bigwedge_{i \in I} Y(u_i) = z_i \right\} \end{aligned}$$

It remains to apply once again the independence of  $T$  and  $Y$  to obtain

$$\begin{aligned} P \left\{ \bigwedge_{i \in I} Y[T(y_i)] = z_i \right\} &= \sum_{(u_i), t} P \left\{ \bigwedge_{i \in I} T(x_i) = u_i, Y(u_i) = z_i, T(-h) = t \right\} \\ &= P \left\{ \bigwedge_{i \in I} Y[T(x_i)] = z_i \right\} \end{aligned}$$

which completes the proof.

**Proposition 17.1.2.**  *$Z$  has the same point distribution as  $Y$ .*

**Proof:** It directly stems from the independence of  $T$  and  $Y$  and the stationarity of  $Y$ :

$$\begin{aligned} P\{Y[T(x)] = z\} &= \sum_t P\{T(x) = t, Y(t) = z\} \\ &= \sum_t P\{T(x) = t\} P\{Y(t) = z\} \\ &= P\{Y(\cdot) = z\} \sum_t P\{T(x) = t\} \\ &= P\{Y(\cdot) = z\} \end{aligned}$$

Suppose now that  $Y$  admits a finite second-order moment. Proposition 17.1.2 shows that the second-order moment of  $Z$  is also finite. In that case, both  $Y$  and  $Z$  have covariance functions, denoted respectively by  $C_Y$  and  $C_Z$ . The relationship between them is specified by the following proposition, the proof of which is left to the reader.

**Proposition 17.1.3.** *The covariance functions of  $Y$  and  $Z$  (if they exist) are related by the formula*

$$C_Z(h) = E\{C_Y(T(h))\} \quad h \in \mathbb{R}^d$$

**Example 17.1.1.** If the coding process is defined as  $Y(t) = \sqrt{2} \cos(\omega t + \Phi)$  where  $\Phi$  is uniform on  $[0, 2\pi[$ , then  $C_Y(t) = \cos(\omega t)$ , so that

$$C_z(h) = E\{\cos(\omega T(h))\}$$

Moreover, if the distribution of  $T(h)$  is symmetric around the origin, then  $E\{\sin(\omega T(h))\} = 0$  and we find that the covariance function of  $Z(x)$  and  $Z(x + h)$  is the value of the Fourier transform of  $T(h)$  at the point  $\omega$

$$C_z(h) = E\left\{e^{i\omega T(h)}\right\}$$

Another interesting feature is isofactorial permanence.

**Proposition 17.1.4.** *If the coding process is a discrete Markov chain with self-adjoint, compact transition kernel, then the bivariate distributions of the substitution random function are isofactorial. Moreover, the factors are the same as those of the transition kernel of the Markov chain.*

**Proof:** Suppose that  $Y$  has its states in  $\mathbb{N}$ . According to section 9.3, its bivariate distributions can be written as

$$P\{Y(s) = i, Y(t) = j\} = \sum_{n=0}^{\infty} \lambda_n^{|s-t|} u_n(i) u_n(j) p(i) p(j)$$

for any  $s, t \in \mathbb{R}$  and any  $i, j \in \mathbb{N}$ . In this formula,  $p$  is the point distribution of  $Y$ . The  $\lambda_n$ 's are the eigenvalues of the transition kernel and the family  $(u_n, n \in \mathbb{N})$  of their normed eigenvectors constitute an orthonormal basis of  $L^2(\mathbb{N}, p)$ . It follows

$$P\{Z(x) = i, Z(y) = j\} = \sum_{n=0}^{\infty} E\{\lambda_n^{|T(x-y)|}\} u_n(i) u_n(j) p(i) p(j)$$

for any  $x, y \in \mathbb{R}^d$  and any  $i, j \in \mathbb{N}$ .  $\square$

## 17.2 Some examples

### 17.2.1 A basic example

In this example, we use the construction by Chentsov (1957) for the directing function. Let  $\mathcal{H}$  be a network of Poisson hyperplanes in  $\mathbb{R}^d$  (see section 12.3). Consider one hyperplane  $H$  in the network. All the points on one side of  $H$  are assigned the same value  $+1$ , and all the points on the other side the opposite value  $-1$ . This defines a random function on  $\mathbb{R}^d$ , denoted by  $V_H$ , which attains only the two values  $\pm 1$  (the points of  $H$  need not be valued because  $H$  has measure 0 in  $\mathbb{R}^d$ ). We proceed in the same way independently for all the hyperplanes of the network. Then we put



$$T(x) = \sum_{\substack{H \in \mathcal{H} \\ H \cap [o, x] \neq \emptyset}} V_H(x)$$

for all  $x \in \mathbb{R}^d$ . Note that  $V_H(x) = \mp V_H(o)$  depending on whether  $H$  splits  $]o, x[$  or not. Accordingly

$$T(x) = \frac{1}{2} \sum_{H \in \mathcal{H}} [V_H(x) - V_H(o)]$$

This implies

$$T(x) - T(y) = \frac{1}{2} \sum_{H \in \mathcal{H}} [V_H(x) - V_H(y)]$$

for any  $x, y \in \mathbb{R}^d$ . In this formula, the summation needs to be extended only to the hyperplanes separating  $x$  and  $y$ . Therefore

$$T(x) - T(y) = \sum_{\substack{H \in \mathcal{H} \\ H \cap [x, y] \neq \emptyset}} V_H(x)$$

Now, the number of hyperplanes hitting  $[x, y]$  is Poisson distributed with mean  $\theta \omega_{d-1} |x - y|$  where  $\theta$  is the Poisson intensity of  $\mathcal{H}$  and  $\omega_{d-1}$  is as usual the  $(d - 1)$ -volume of the unit ball in  $\mathbb{R}^{d-1}$  (see section 12.3.2 and remark 6.1.4). It follows that the variogram of  $T$  is

$$\frac{1}{2} \text{Var}\{T(x) - T(y)\} = \frac{1}{2} \theta \omega_{d-1} |x - y|$$

Figure 17.1 shows a two-dimensional realization of Chentsov’s model with line intensity  $\theta = 10$ . This realization is displayed in 6 nested squares with sides of length 12.5, 25, 50, 100, 200 and 400, all centered on the origin. The Poisson lines are barely visible because of the large number of lines used. About 16000 lines have been used to produce the bottom right image).

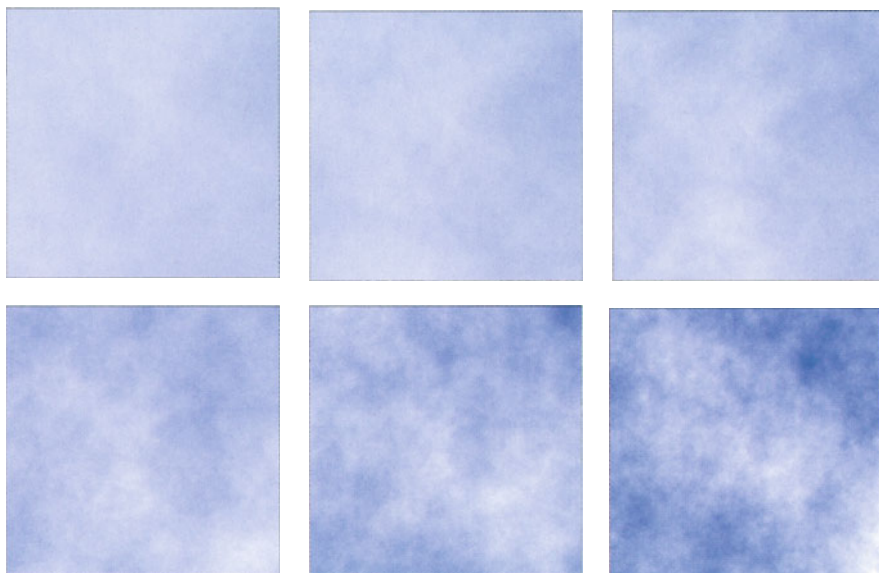
It should also be pointed out that the Poisson intensity acts as a scale factor. In the variogram formula it is numerically equivalent to move  $y$  away from  $x$ , or to increase the Poisson intensity  $\theta$ . Indeed, Chentsov’s construction is one prototype for *self-similar models*.

As for the coding process  $Y$ , we consider a reversible Markov chain on  $(0, 1)$  with transition kernel

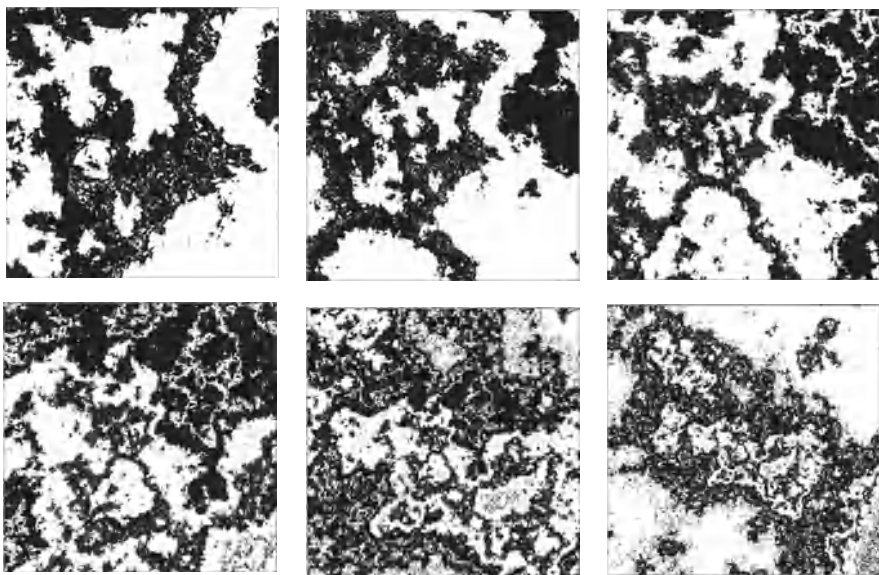
$$P = \begin{pmatrix} p & q \\ q & p \end{pmatrix}$$

Its stationary distribution is uniform.

The coding process has been simulated with  $p = 0.85$ , and the realization produced has been composed with that of Figure 17.1. A realization of the substitution random function is thus obtained (see Figure 17.2).



**Fig. 17.1.** Realization of a directing random function using Chentsov's construction. It is displayed at six different scales



**Fig. 17.2.** Realization of a substitution random function displayed at six different scales

This figure calls for several comments. Firstly, the realization does not present any sign of spatial homogeneity. Indeed, the directing function may assign close values to distant points of  $\mathbb{R}^d$ , so their neighbourhoods have a non-negligible chance of being coded similarly. In other words, the covariance function of the substitution random function has a slow rate of decay. Analytically speaking, this covariance function is equal to

$$Cov\{Z(x), Z(y)\} = \frac{1}{4} e^{-\lambda} \sum_{n \in \mathbb{Z}} I_n(\lambda) (p - q)^{|n|}$$

where  $\lambda = \theta \omega_{d-1} |x - y|$  is the mean number of hyperplanes separating  $x$  and  $y$ , and where the  $I_n$ 's stand for the modified Bessel functions (see the Appendix 1 of this chapter). Since  $I_n(z) \approx e^z / \sqrt{2\pi z}$  when  $z \rightarrow +\infty$  (Abramovitz and Stegun, 1965), we find that

$$Cov\{Z(x), Z(y)\} \approx \frac{1}{4} \frac{p}{q} \frac{1}{\sqrt{2\pi\theta \omega_{d-1} |x - y|}}$$

It follows that the substitution random function has an infinite integral range.

If the intensity  $\theta$  is large, then two neighbouring points of  $\mathbb{R}^d$  are separated by many Poisson hyperplanes. Accordingly, the directing function may assign them very different values and there is a good chance that the coded values are independent. This explains why the realization of Figure 17.2 has a fragmented appearance. However this random set model is not fractal because its specific  $(d - 1)$ -volume is finite (explicitly,  $d\omega_d\theta q/2$ ).

### 17.2.2 Geometry

Is it possible to compel the realizations of a substitution random function to respect some prespecified geometry? This question is important in the earth sciences where structural information is often available and has to be taken into account in simulations. Of course it is not possible to provide a general answer to that question because the word geometry encompasses many features including size, shape and structure. Two examples are now given.

**Stratification.** Anisotropies can be generated by considering a variation of Chentsov's construction. Each Poisson hyperplane of the network splits  $\mathbb{R}^d$  into two half-spaces. The upper half-space (that contains the points with greatest  $d$ -coordinates) is assigned the value  $+1$  and the lower half-space the value  $-1$ . Keeping the same coding process, realizations like the one in Figure 17.3 are obtained.

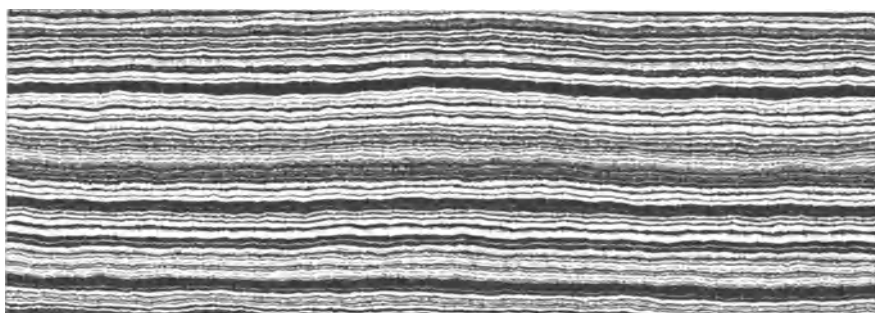
The covariance function of this substitution random function takes the form

$$Cov\{Z(x), Z(y)\} = \frac{1}{4} e^{-\lambda} \sum_{n \in \mathbb{Z}} \left(\frac{r}{1-r}\right)^{n/2} I_n\left(2\lambda\sqrt{r(1-r)}\right) (p - q)^{|n|}$$

In this formula, the vector  $x - y$  appears in two places, firstly via the mean number  $\lambda$  of hyperplanes separating  $x$  and  $y$ , and secondly via the probability  $r$  that  $x$  belongs to the upper half-space limited by a uniform hyperplane hitting  $[x, y]$  (see Appendix 1). This coefficient  $r$  is responsible for the anisotropy of the covariance. In two dimensions, we have

$$r = \frac{1}{2} \left( 1 + \langle (0, 1), \frac{x - y}{|x - y|} \rangle \right)$$

If  $x$  and  $y$  lie in the same horizontal hyperplane (i.e.  $\langle x - y, (0, \dots, 0, 1) \rangle = 0$ ), then  $r = \frac{1}{2}$  and the value of  $Cov\{Z(x), Z(y)\}$  coincides with the one given in the previous section where the half-spaces are isotropically valued. In contrast to this, if  $x$  and  $y$  belong to the same vertical line (i.e.  $\langle x - y, (0, \dots, 0, 1) \rangle = \pm|x - y|$ ), then either  $r = 0$  or  $r = 1$ . A limited expansion of the Bessel functions in the vicinity of 0 or 1 shows that  $Cov\{Z(x), Z(y)\}$  is an exponential function of  $|x - y|$ . This is not surprising because the directing function  $T$  is monotonic along the vertical, which makes the substitution random function markovian along it.



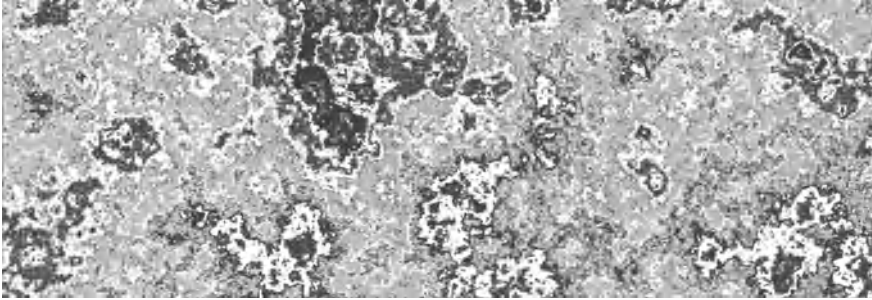
**Fig. 17.3.** Realization of an anisotropic substitution random function

**Isopartition.** In the previous two examples, the substitution random function took only two values 0 and 1. In order to extend this set of values, it suffices to extend the range of the coding process. To illustrate this, let us keep the same directing function as in 17.2.1, and let us choose for the coding process a Markov chain on  $(0, 1, 2)$  with transition kernel

$$P = \begin{pmatrix} p & q & q \\ q & p & q \\ q & q & p \end{pmatrix}$$

The kernel  $P$  has been designed in such a way that it assigns to all states the same mean sojourn time. Moreover, when leaving a state, every other state has the same chance of being accessed. This results in a substitution model

where the three facies  $Z^{-1}(0)$ ,  $Z^{-1}(1)$  and  $Z^{-1}(2)$  have exactly the same statistical properties, and where the interactions between different facies do not depend on the facies considered. Such a model is called an *isopartition*.



**Fig. 17.4.** Realization of an isopartition obtained by substitution. It may suggest a polished section of metamorphic rock

The statistical symmetry of this model is not perceptible in the realization shown in Figure 17.4. A close look at the model parameters can help clarify the situation. The directing function has been built using the Poisson line intensity  $\theta = 5$ , so the mean number of lines hitting the simulation field  $1150 \times 400$  is 15500. The values of the directing function generated range between  $-82$  and  $87$ , which seems quite small in comparison with the number of Poisson lines involved. However, assuming the simulation field centered at the origin (which is the case here), then the directing function has maximum variability at the vertices of the rectangular field. If  $v$  denotes one such vertex, then the distribution of  $T(v)$  is centered and approximatively normal with variance  $\theta\omega_{2-1}|v| \approx 6087$ . This yields a standard deviation of  $78$ , that is compatible with the values generated. Regarding the transition kernel, the sojourn rate  $p$  has been set to  $0.8$ , so on average, the coding process remains at the same state for  $5$  steps. A comparison between the total variability of the directing function ( $87 - (-82) = 169$ ) and the mean sojourn time ( $5$ ) suggests that not many transitions take place in the simulation field. This is confirmed if we consider the probability  $F(\ell)$  for a segment of length  $\ell$  to be totally contained within a single phase. This probability can be assessed numerically starting from its formal expression

$$F(\ell) = \sum_{n=0}^{\infty} \frac{2p^n(1-p)^2}{n+2} \sum_{k=1}^{n+1} \sin^2 \frac{k\pi}{2} \cot^2 \frac{k\pi}{2(n+2)} \exp\left(-4\theta\ell \sin^2 \frac{k\pi}{2(n+2)}\right)$$

(cf. Appendix 2). This gives for instance  $F(10) = 0.054$ . For  $\ell = 40$ , this probability becomes quite small ( $0.004$ ) but remains far from negligible.

### 17.2.3 Topology

Suppose, as before, that the coding process has a finite set of states, say  $(0, 1, \dots, n-1)$ . Then the substitution random function can be characterized by the  $n$  random sets  $Z^{-1}(0), Z^{-1}(1), \dots, Z^{-1}(n-1)$ . These random sets cannot be independent because they partition  $\mathbb{R}^d$ . The dependence relationships between them are particularly apparent if they are topological. For instance, two random sets may not be in contact, or one random set must be surrounded by another one. Is it possible to design models that fulfill these types of topological relationships?

To answer that question, a simple idea is to choose a coding process where direct transitions between state  $i$  and state  $j$  can take place if and only if both random sets  $Z^{-1}(i)$  and  $Z^{-1}(j)$  are allowed to be in contact. This idea can be easily implemented in the case when the coding process is markovian but requires the directing function to have continuous realizations. This implies in turn that continuous time coding processes must be considered instead of discrete ones.

Let  $T_\theta$  be a directing function based on Chentsov's model with Poisson intensity  $\theta$ . It is not difficult to establish that the spatial distribution of  $T_\theta/\sqrt{\theta}$  tends to be a centered gaussian with the variogram  $\gamma(h) = \frac{1}{2}\omega_{d-1}|h|$  as  $\theta$  tends to the infinity<sup>2</sup>. According to Adler's criterion (1981), a gaussian random function with a linear variogram has continuous realizations.

Figure 17.5 shows simulations of four different substitution random functions. All of them have been built using the same realization of the directing gaussian random function. However the coding processes are different. Each of them has four states  $(0, 1, 2, 3)$  depicted in red, yellow, green and black respectively. The top left coding process tends to favour periodic sequences of states. The top right one creates symmetric communications between adjacent states. The bottom left coding process is a chain with two extremal states  $(0$  and  $3)$  and two intermediary ones  $(1$  and  $2)$ . Finally, the bottom right coding process presents a hierarchy: communications between states  $0, 1$  or  $2$  require a transition via state  $3$ . Here are the four transition matrices:

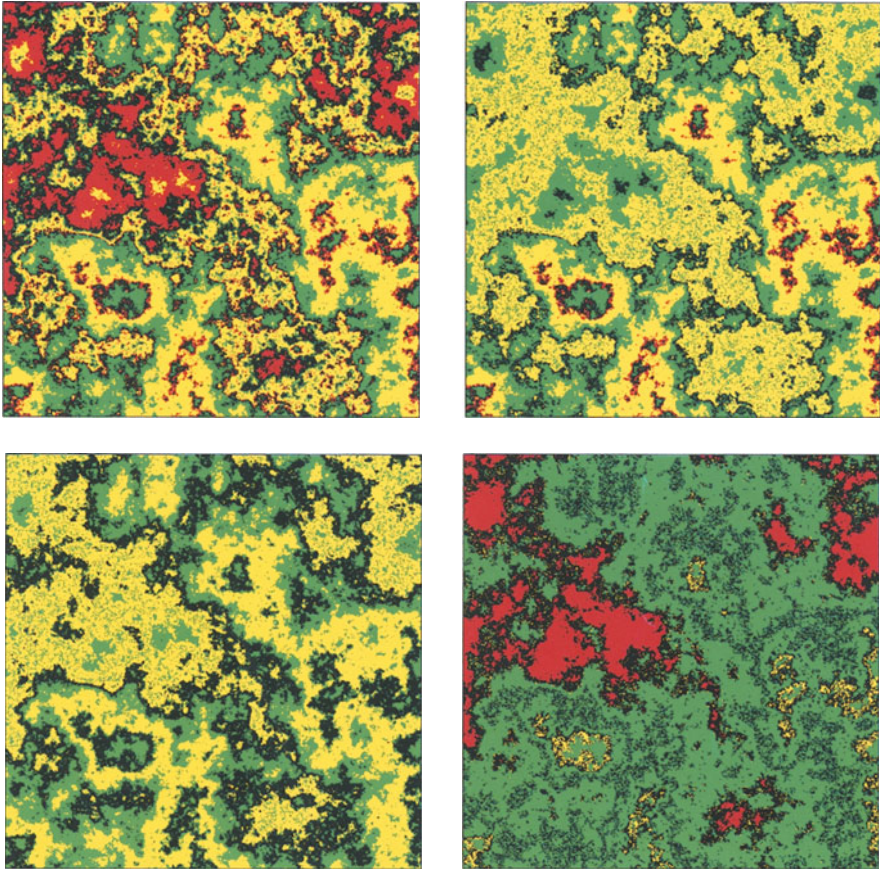
$$P_{tl} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \qquad P_{tr} = \frac{1}{2} \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

---

<sup>2</sup> This procedure is not significantly different from that used by Chilès (1978) for generating gaussian random functions with linear variograms (turning bands method). The only difference is that the number of hyperplanes is random instead of being fixed.

$$P_{bl} = \frac{1}{2} \begin{pmatrix} 0 & 2 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{pmatrix} \quad P_{br} = \frac{1}{3} \begin{pmatrix} 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 3 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

The sojourn times in each state are exponentially distributed. Their parameters have been chosen so that the point distribution of each of the four substitution random functions is uniform on  $\{0, 1, 2, 3\}$ . These are  $a_0 = a_1 = a_2 = a_3 = 0.125$  for both top coding processes,  $2a_0 = a_1 = a_2 = 2a_3 = 0.0625$  for the bottom left one and  $a_0 = a_1 = a_2 = a_3/3 = 0.125$  for the bottom right one.



**Fig. 17.5.** Simulations of four different substitution random functions. All have been built using the same realization of the directing random function, which conveys them some similarity. The coding processes are different so that the four simulations have different topological properties

### 17.3 Conditional simulation

In this section we are concerned with simulating the substitution random function  $Z$  in the field  $D \subset \mathbb{R}^d$  given the finite set of conditions  $(Z(x_a) = z_a, a \in A)$ .

To simplify the presentation, we are going to assume that the directing random function  $T$  and the coding process  $Y$  are discrete. Moreover, it is also convenient to resort to the vector notation used in the previous chapter. Using this notation, the conditions can be concisely written as  $Z(x_A) = z_A$  with  $x_A = (x_a, a \in A)$  and  $z_A = (z_a, a \in A)$ .

Let  $I$  be any finite population of indices. Because  $T$  and  $Y$  are independent, the conditional distribution of  $Z(x_I)$  can be expressed as follows

$$P\{Z(x_I) = z_I \mid Z(x_A) = z_A\} = \sum_{t_A} P\{T(x_A) = t_A \mid Z(x_A) = z_A\} \times \\ \sum_{t_I} P\{T(x_I) = t_I \mid T(x_A) = t_A\} P\{Y(t_I) = z_I \mid Y(t_A) = z_A\}$$

This formula shows that the conditional simulation of  $Z$  can be achieved by first simulating the values of the directing random function at the conditioning data points, and then by simulating both the directing random function and the coding process conditionally. Here is the algorithmic transcription of this formula:

**Algorithm 17.3.1.** (*Conditional simulation of a substitution random function*)

1. Generate  $t_A \sim \mathcal{D}(T(x_A) \mid Z(x_A) = z_A)$ .
2. Generate  $t_D \sim \mathcal{D}(T(x_D) \mid T(x_A) = t_A)$ .
3. Generate  $y_D \sim \mathcal{D}(Y(t_D) \mid Y(t_A) = z_A)$ .
4. Deliver  $y_D$ .

Assuming that we can simulate  $T$  and  $Y$  conditionally, the only difficulty is running step 1. Writing the conditional distribution of  $T(x_A)$  as

$$P\{T(x_A) = t_A \mid Z(x_A) = z_A\} \propto P\{T(x_A) = t_A\} P\{Y(t_A) = z_A\}$$

suggests the following rejection procedure

**Algorithm 17.3.2.** (*Step 1 of algorithm 17.3.1. Rejection approach*)

1. Generate  $t_A \sim \mathcal{D}(T(x_A))$ .
2. Generate  $y_A \sim \mathcal{D}(Y(t_A))$ .
3. If  $y_A \neq z_A$ , then goto 1.
4. Deliver  $t_A$ .



This procedure is possible but rather slow because the rejection rate is usually very high. For this reason, instead of generating all components of  $t_A$  in one step, an iterative procedure based on Gibbs sampler that updates one component of  $t_A$  at each iteration is more appropriate.

The initialization of the vector  $t_A$  is achieved by a nonconditional simulation of  $T(x_A)$ .

Regarding the current stage, suppose that the component  $t_a$  of  $t_A$  has to be updated. Its conditional distribution is

$$P\{T(x_a) = t_a \mid T(x_A^a) = t_A^a, Z(x_A) = z_A\} \propto \\ P\{T(x_a) = t_a \mid T(x_A^a) = t_A^a\} P\{Y(t_a) = z_a \mid Y(t_A^a) = z_A^a\}$$

This suggests generating a candidate value  $t'_a$  from the conditional distribution of  $T(x_a)$  given  $T(x_A^a) = t_A^a$  and accepting it with probability

$$\alpha(t_a, t'_a) = \frac{p'}{p + p'}$$

where

$$p = P\{Y(t_a) = z_a \mid Y(t_A^a) = z_A^a\} \quad p' = P\{Y(t'_a) = z_a \mid Y(t_A^a) = z_A^a\}$$

in accordance with Metropolis-Hastings algorithm (Barker's variation). We thus arrive to the following algorithm.

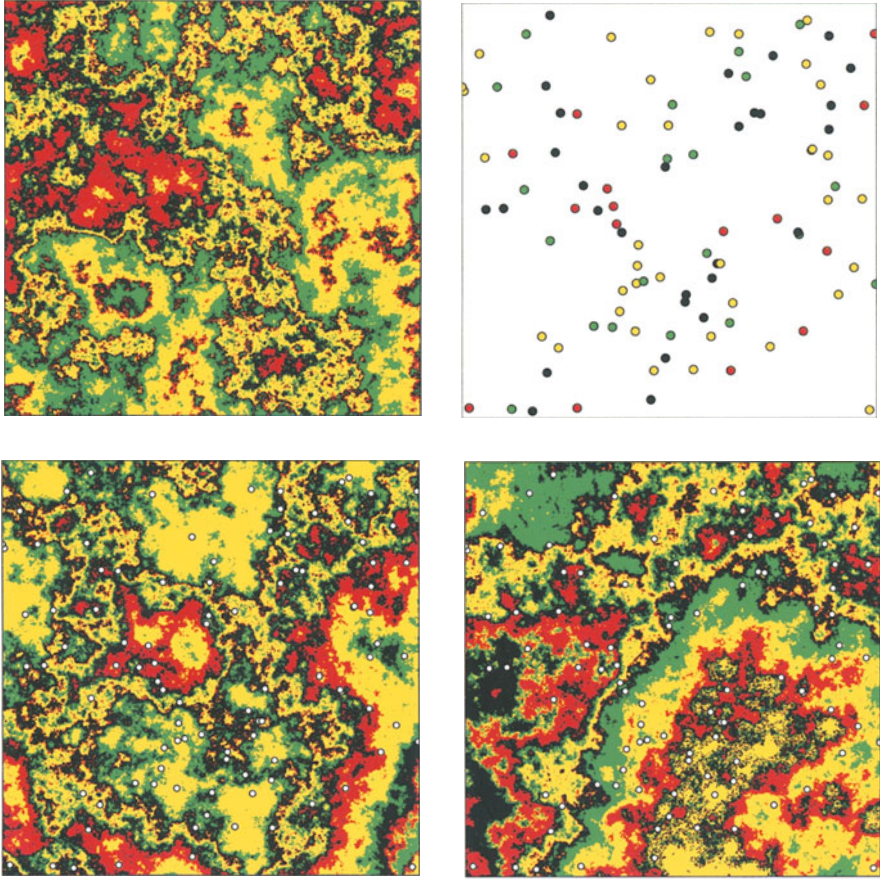
**Algorithm 17.3.3.** (Step 1 of algorithm 17.3.1)

1. Generate  $t_A \sim \mathcal{D}(T(x_A))$ .
2. Select  $a \sim \text{Unif}(A)$ .
3. Generate  $t'_a \sim \mathcal{D}(T(x_a) \mid T(x_A^a) = t_A^a)$  and  $u \sim \text{Unif}$ .
4. If  $u \leq \alpha(t_a, t'_a)$ , then put  $t_a = t'_a$ .
5. Goto 2.

**Remark 17.3.1.** Running step 4 of this algorithm presents no difficulty if the spatial distribution of the coding process is numerically tractable. This case is frequently encountered (e.g. if  $Y$  is a Markov chain), but not always. An alternative is to replace step 4 by the following procedure. Accepting the new value  $t'_a$  or conversely keeping the old value  $t_a$  is equivalent to simulating the distribution  $(\alpha(t'_a, t_a), \alpha(t_a, t'_a))$  on  $(t_a, t'_a)$ . This distribution can be simulated exactly (in the sense of Propp and Wilson) because it is invariant under the transition kernel

$$P = \begin{pmatrix} 1 - p' & p' \\ p & 1 - p \end{pmatrix}$$

and because simulating  $P$  is equivalent to simulating the conditional distributions of  $Y(t_a)$  and  $Y(t'_a)$ .



**Fig. 17.6.** Conditional simulation of a substitution random function. Top left, a non conditional simulation. Top right, a set of 100 conditioning data points. Bottom, two conditional simulation

Figure 17.6 shows an example of a conditional simulation of a substitution random function. The model used is as in the top left of Figure 17.5. The directing function is gaussian with the linear variogram  $\gamma(h) = |h|$ . The transition matrix is  $P = P_{tr}$  and the four states 0, 1, 2 and 3 have their mean sojourn time equal to  $1/a = 8$  in the coding process. The simulation field is  $400 \times 400$ . One hundred uniform points have been independently selected from the non conditional simulation (top left) to act as conditioning data points (top right). In the present case, explicit formulae exist for the transitions of the coding process

$$P^{(t)}(i, j) = \frac{1}{2} \exp(-at) \frac{d^n}{du^n} \left( \cosh u + \cos u \right)_{u=at}$$

with  $n = 4 - j + i \equiv (4)$ . So it was possible to run step 4 of algorithm 17.3.3. The two conditional simulations shown at the bottom of Figure 17.6 were generated this way using 30000 iterations. Despite the relatively large number of conditioning data points and the infinite integral range of the substitution random function, the conditional simulations still offer a wide range of variability. This can be attributed to the behaviour of the covariance function at the origin (finite, but steep slope), which is responsible for the amount of details that can be seen on each simulation. This also implies that the range of influence of a data point decreases rapidly. From this point of view, the conditioning data set cannot be considered as very informative.

**Appendix 1:** Calculation of the covariance function of a substitution random function.

Let us consider a model where the directing random function is based on Chentsov’s construction and where the coding process is a Markov chain on  $(0, 1)$  with transition kernel

$$P = \begin{pmatrix} p & q \\ q & p \end{pmatrix}$$

This Markov chain is reversible and its limit distribution is uniform. Direct calculations show that the  $n^{th}$  iterate of the transition kernel can be written as

$$P^{(n)} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{(p - q)^n}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

This gives

$$P\{Y(t) = i, Y(u) = j\} = \frac{1}{4} [1 + (-1)^{i-j} (p - q)^{|t-u|}]$$

for any  $t, u \in \mathbb{Z}$  and any  $i, j \in \{0, 1\}$ . It follows that the bivariate distribution of the substitution random function is

$$P\{Z(x) = i, Z(y) = j\} = \frac{1}{4} [1 + (-1)^{i-j} E\{(p - q)^{|T(x-y)|}\}]$$

for any  $x, y \in \mathbb{R}^d$  and any  $i, j \in \{0, 1\}$ . From this, we derive its covariance function

$$Cov\{Z(x), Z(y)\} = \frac{1}{4} E\{(p - q)^{|T(x-y)|}\}$$

Let  $\lambda = \theta \omega_{d-1} |x - y|$  be the mean number of Poisson hyperplanes hitting  $[x, y]$ . Two cases are now considered:

- If  $T$  is isotropic (this is the standard Chentsov’s construction of section 17.2.1, then  $T(x - y)$  is the difference of two independent Poisson variables

with common mean  $\lambda/2$ . A direct calculation (see also the next case) shows that its distribution is

$$P\{T(x - y) = n\} = e^{-\lambda} I_n(\lambda) \quad n \in \mathbb{Z}$$

where the  $I_n$ 's are the modified Bessel functions<sup>3</sup>. Therefore

$$Cov\{Z(x), Z(y)\} = \frac{1}{4} e^{-\lambda} \sum_{n \in \mathbb{Z}} I_n(\lambda) (p - q)^{|n|}$$

– Suppose now  $T$  anisotropic as in the first example of section 17.2.2. Let  $H$  be a uniform hyperplane hitting  $[x, y]$ . The point  $x$  belongs to the "positive" half-space delimited by  $H$  with probability

$$r = \frac{1}{\omega_{d-1}} \int_{S(u_0) \cap S(\alpha_0)} \langle u_0, \alpha \rangle d\alpha$$

where  $\alpha_0 = (0, \dots, 0, 1)$ ,  $u_0 = (x - y)/|x - y|$  is the unit vector in the direction of  $x - y$  and

$$S(u_0) = \{\alpha \in S_d : \langle u_0, \alpha \rangle \geq 0\} \quad S(\alpha_0) = \{\alpha \in S_d : \langle \alpha_0, \alpha \rangle \geq 0\}$$

Then  $T(x - y)$  is the difference of two independent Poisson variables with mean  $\lambda r$  and  $\lambda(1 - r)$ . It follows that if  $n \geq 0$

$$\begin{aligned} P\{T(x - y) = n\} &= \sum_{j=0}^{\infty} e^{-\lambda r} \frac{(\lambda r)^{n+j}}{(n+j)!} e^{-\lambda(1-r)} \frac{(\lambda(1-r))^j}{j!} \\ &= e^{-\lambda} (\lambda r)^n \sum_{j=0}^{\infty} \frac{(\lambda^2 r(1-r))^j}{j!(n+j)!} \\ &= e^{-\lambda} \left(\frac{r}{1-r}\right)^{n/2} I_n(2\lambda\sqrt{r(1-r)}) \end{aligned}$$

If  $n < 0$ , it suffices to permute  $r$  and  $1 - r$ , and replace  $n$  by  $-n$ :

$$\begin{aligned} P\{T(x - y) = n\} &= P\{-T(x - y) = -n\} \\ &= e^{-\lambda} \left(\frac{1-r}{r}\right)^{-n/2} I_{-n}(2\lambda\sqrt{r(1-r)}) \end{aligned}$$

<sup>3</sup> The modified Bessel function of order  $\nu$  is defined as

$$I_\nu(z) = (z/2)^\nu \sum_{k=0}^{\infty} \frac{(z^2/4)^k}{k! \Gamma(\nu + k + 1)}$$

It satisfies  $I_\nu = I_{-\nu}$  if  $\nu \in \mathbb{Z}$  (Abramovitz and Stegun, 1965).

But  $I_{-n} = I_n$ , so we finally have

$$P\{T(x - y) = n\} = e^{-\lambda} \left(\frac{r}{1-r}\right)^{n/2} I_n(2\lambda\sqrt{r(1-r)})$$

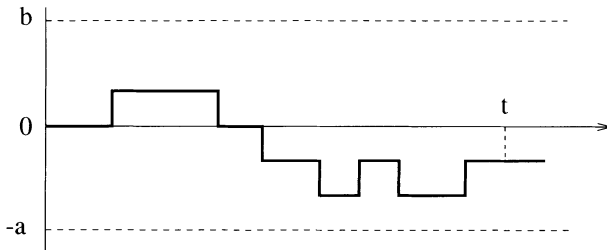
for any  $n \in \mathbb{Z}$ . Consequently

$$Cov\{Z(x), Z(y)\} = \frac{1}{4} e^{-\lambda} \sum_{n \in \mathbb{Z}} \left(\frac{r}{1-r}\right)^{n/2} I_n(2\lambda\sqrt{r(1-r)}) (p - q)^{|n|}$$

**Appendix 2:** Distribution of the extrema of a Poisson random walk.

Let  $X$  be a Poisson random walk on  $\mathbb{R}_+$ . Jumps with amplitude  $\pm 1$  occur at the points of a Poisson process with intensity  $\theta$ . Let  $a$  and  $b$  two integers. We are interested in the probability  $F_{a,b}$  that the trajectory of the process is bounded by the values  $-a$  and  $b$  in the interval  $[0, t]$  (see Figure 17.8).

$$F_{a,b}(t) = P\{-a \leq \min_{0 \leq u \leq t} X(u) \leq \max_{0 \leq u \leq t} X(u) \leq b\}$$



**Fig. 17.8.** Example of a Poisson random walk. What is the probability that the trajectory remains between  $-a$  and  $b$  in the interval  $[0, t]$ ?

Using a standard renewal argument, we can write

$$F_{a,b}(t) = e^{-\theta t} + \frac{1}{2} \int_0^t \theta e^{-\theta s} [F_{a+1,b-1}(t-s) + F_{a-1,b+1}(t-s)] ds$$

with the convention  $F_{a,b} = 0$  when  $a$  or  $b$  are negative. Taking the Laplace transform of  $F_{a,b}$

$$\phi_{a,b}(\lambda) = \int_0^\infty e^{-\lambda t} F_{a,b}(t) dt$$

we obtain

$$\phi_{a,b}(\lambda) = \frac{1}{\lambda + \theta} + \frac{1}{2} \frac{\theta}{\lambda + \theta} \left[ \phi_{a+1,b-1}(\lambda) + \phi_{a-1,b+1}(\lambda) \right]$$

or equivalently

$$\frac{2(\lambda + \theta)}{\theta} \phi_{a,b}(\lambda) - \phi_{a+1,b-1}(\lambda) - \phi_{a-1,b+1}(\lambda) = \frac{2}{\theta}$$

This shows that all the  $\phi_{a,b}(\lambda)$  associated with  $a + b = c$  fixed are the solution of a system of  $c + 1$  linear equations that can be written in matrix form as

$$\begin{pmatrix} \alpha & -1 & \cdots & 0 \\ -1 & \alpha & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \alpha & -1 \\ 0 & \cdots & -1 & \alpha \end{pmatrix} \begin{pmatrix} \phi_{c,0}(\lambda) \\ \phi_{c-1,1}(\lambda) \\ \vdots \\ \phi_{1,c-1}(\lambda) \\ \phi_{0,c}(\lambda) \end{pmatrix} = \frac{2}{\theta} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}$$

with  $\alpha = \frac{2(\lambda + \theta)}{\theta}$ . The matrix has the eigenvalues

$$\lambda_k = \frac{2\lambda}{\theta} + 4 \sin^2 \frac{k\pi}{2(c+2)} \quad k = 1, \dots, c+1$$

and a (normed) eigenvector associated with the eigenvalue  $\lambda_k$  is

$$u_k(l) = \sqrt{\frac{2}{c+2}} \sin \frac{kl\pi}{c+2} \quad l = 1, \dots, c+1$$

The solution of the system can be expressed very simply in terms of the  $\lambda_k$ 's and the  $u_k$ 's. Indeed, if  $\sum_k \beta_k u_k$  is the expansion of the right hand side of the system, its solution is  $\sum_k \beta_k / \lambda_k u_k$ . Explicitly, we find

$$\beta_k = \frac{2}{\theta} \sqrt{\frac{2}{c+2}} \sin^2 \frac{k\pi}{2} \cot \frac{k\pi}{2(c+2)}$$

from which we derive

$$\phi_{a,b}(\lambda) = \frac{2}{c+2} \sum_{k=1}^{c+1} \frac{\sin^2 \frac{k\pi}{2} \cot \frac{k\pi}{2(c+2)} \sin \frac{(b+1)k\pi}{c+2}}{\lambda + 2\theta t \sin^2 \frac{k\pi}{2(c+2)}}$$

Hence

$$F_{a,b}(t) = \frac{2}{c+2} \sum_{k=1}^{c+1} \sin^2 \frac{k\pi}{2} \cot \frac{k\pi}{2(c+2)} \sin \frac{(b+1)k\pi}{c+2} \times \exp \left( -2\theta t \sin^2 \frac{k\pi}{2(c+2)} \right)$$

Observe in particular that the property  $F_{a,b}(t) = F_{b,a}(t)$  is satisfied as a consequence of  $\sin \frac{(b+1)k\pi}{c+2} = (-1)^{k+1} \sin \frac{(a+1)k\pi}{c+2}$  and because only the odd values of  $k$  are involved in the summation.

Suppose now that  $a$  and  $b$  are independent random variables whose distribution is geometric with parameter  $p$ . Then we have

$$F(t) = \sum_{a,b=0}^{\infty} p^{a+b} (1-p)^2 F_{a,b}(t)$$

Explicitly we arrive at

$$F(t) = \sum_{c=0}^{\infty} \frac{2p^c(1-p)^2}{c+2} \sum_{k=1}^{c+1} \sin^2 \frac{k\pi}{2} \cot^2 \frac{k\pi}{2(c+2)} \exp\left(-2\theta t \sin^2 \frac{k\pi}{2(c+2)}\right)$$

## Appendix

This appendix gives the spectral measure and the integral range of some covariance functions commonly used.

Table 1 gives the list of the covariance functions. They are normalized, in the sense that their value at 0 is 1, their scale factor is 1 and that they are isotropic. Owing to their isotropy, their integral ranges can be calculated by changing to polar coordinates

$$A_d = \int_{\mathbb{R}^d} C(h) dh = d\omega_d \int_0^\infty C_d(r) dr$$

where  $C_d$  denotes the radial form of the covariance  $C$  (i.e.  $C(h) = C_d(|h|)$ ), and  $\omega_d$  is the volume of the  $d$ -dimensional unit ball. Table 2 gives the integral ranges in dimensions 1, 2 or 3. Table 3 gives the general expression for the integral range of the covariance models that are valid whatever the space dimension.

Table 4 gives the spectral measure of the covariance functions. It is easier to calculate a spectral measure when the covariance function is square integrable. In that case, the spectral measure has a density that is equal to the inverse Fourier transform of the covariance function

$$f(u) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i \langle u, h \rangle} C(h) dh.$$

In the isotropic case, this integral simplifies into

$$f(u) = \frac{1}{(2\pi)^{d/2}} \frac{1}{|u|^\nu} \int_{\mathbb{R}^d} J_\nu(|u|r) C_d(r) r^{d/2} dr$$

where  $\nu = d/2 - 1$  and  $J_\nu$  denotes the Bessel function of order  $\nu$ . This formula has been used to obtain the spectral densities in Table 4.

Unfortunately, we have been unable to calculate the spectral measure of the hyperbolic and of the stable covariance functions. This has been indicated with a question mark in Table 4.



**Table 1.** Covariance models

Model	Formula	Remark
Triangular	$(1 -  h ) 1_{ h  \leq 1}$	$d = 1$
Cosine	$\cos( h )$	$d = 1$
Circular	$\frac{2}{\pi} \left( \arccos  h  -  h  \sqrt{1 -  h ^2} \right) 1_{ h  \leq 1}$	$d \leq 2$
Spherical	$\left( 1 - \frac{3}{2} h  + \frac{1}{2} h ^3 \right) 1_{ h  \leq 1}$	$d \leq 3$
Hyperbolic	$\frac{1}{1 +  h }$	
Exponential	$\exp\{- h \}$	
Gaussian	$\exp\{- h ^2\}$	
Stable	$\exp\{- h ^\alpha\}$	$0 < \alpha \leq 2$
Cardinal sine	$\frac{\sin( h )}{ h }$	$d \leq 3$
$J$ -Bessel	$2^\mu \Gamma(\mu + 1) \frac{J_\mu( h )}{ h ^\mu}$	$\mu \geq \frac{d}{2} - 1$
$K$ -Bessel	$\frac{ h ^\mu}{2^{\mu-1} \Gamma(\mu)} K_{-\mu}( h )$	$\mu > 0$

**Table 2.** Integral ranges in dimensions 1, 2 and 3

Model	$A_1$	$A_2$	$A_3$
Triangular	1	—	—
Cosine	0	—	—
Circular	$\frac{8}{3\pi}$	$\frac{\pi}{4}$	—
Spherical	$\frac{3}{4}$	$\frac{\pi}{5}$	$\frac{\pi}{6}$
Hyperbolic	$\infty$	$\infty$	$\infty$
Exponential	2	$2\pi$	$8\pi$
Gaussian	$\sqrt{\pi}$	$\pi$	$\pi\sqrt{\pi}$
Stable	$2\Gamma\left(\frac{\alpha+1}{\alpha}\right)$	$\pi\Gamma\left(\frac{\alpha+2}{\alpha}\right)$	$\frac{4}{3}\pi\Gamma\left(\frac{\alpha+3}{\alpha}\right)$
Cardinal sine	$\pi$	$\infty$	$\infty$
$J$ -Bessel	$\infty$ if $\mu \leq \frac{1}{2}$ $2\sqrt{\pi}\frac{\Gamma(\mu+1)}{\Gamma(\mu+\frac{1}{2})}$	$\infty$ if $\mu \leq \frac{3}{2}$ $4\pi\mu$	$\infty$ if $\mu \leq \frac{5}{2}$ $(2\sqrt{\pi})^3\frac{\Gamma(\mu+1)}{\Gamma(\mu-\frac{1}{2})}$
$K$ -Bessel	$2\sqrt{\pi}\frac{\Gamma(\mu+\frac{1}{2})}{\Gamma(\mu)}$	$4\pi\mu$	$(2\sqrt{\pi})^3\frac{\Gamma(\mu+\frac{3}{2})}{\Gamma(\mu)}$

**Table 3.** Integral ranges in arbitrary dimensions

Model	$A_d$
Triangular	–
Cosine	–
Circular	–
Spherical	–
Hyperbolic	$\infty$
Exponential	$\omega_d d!$
Gaussian	$(\sqrt{\pi})^d$
Stable	$\omega_d \Gamma\left(\frac{\alpha + d}{\alpha}\right)$
Cardinal sine	–
$J$ -Bessel	$\begin{cases} \infty & \text{if } \mu \leq d - \frac{1}{2} \\ (2\sqrt{\pi})^d \frac{\Gamma(\mu + 1)}{\Gamma(\mu + 1 - \frac{d}{2})} & \text{if } \mu > d - \frac{1}{2} \end{cases}$
$K$ -Bessel	$(2\sqrt{\pi})^d \frac{\Gamma(\mu + \frac{d}{2})}{\Gamma(\mu)}$

**Table 4.** Spectral measures

Model	Spectral measure
Triangular	$\frac{1}{\pi} \frac{1 - \cos u}{u^2}$
Cosine	$\frac{1}{2} (\delta_{-1}(du) + \delta_1(du))$
Circular	$\frac{1}{\pi u ^2} J_1^2\left(\frac{ u }{2}\right)$
Spherical	$\frac{3}{4\pi u ^3} J_{\frac{3}{2}}^2\left(\frac{ u }{2}\right)$
Hyperbolic	?
Exponential	$\frac{\Gamma\left(\frac{d+1}{2}\right)}{\pi^{\frac{d+1}{2}}} \frac{1}{(1 +  u ^2)^{\frac{d+1}{2}}}$
Gaussian	$\frac{1}{(2\sqrt{\pi})^d} \exp\left(-\frac{ u ^2}{4}\right)$
Stable	?
Cardinal sine	$\frac{1}{4\pi} \delta_{ u =1}(du)$
$J$ -Bessel	$\frac{\Gamma(\mu + 1)}{\Gamma\left(\mu + 1 - \frac{d}{2}\right) \pi^{\frac{d}{2}}} \frac{1}{(1 -  u ^2)^{\mu - \frac{d}{2}}}$
$K$ -Bessel	$\frac{\Gamma\left(\mu + \frac{d}{2}\right)}{\Gamma(\mu) \pi^{\frac{d}{2}}} \frac{1}{(1 +  u ^2)^{\mu + \frac{d}{2}}}$

## References

- Abramowitz, M. and Stegun, I. (1970) Handbook of Mathematical Functions, Dover, New York (9<sup>th</sup> edition).
- Adler, R.J. (1981) The Geometry of Random Fields, Wiley, New York.
- Ahrens, J.H. and Dieter, U. (1982) Computer generation of Poisson deviates from modified normal distributions", ACM Trans. Math. Soft., 8, pp. 163-179.
- Akhiezer, N.I. and Glazman, I.M. (1966) Theory of Linear Operators in Hilbert Space. Frederick Ungar Pub. Co., New York.
- Alabert, F.G. (1987) Stochastic imaging of spatial distributions using hard and soft data, Master's thesis, Stanford University.
- Alfaro Sironvalle, M. (1979) Etude de la robustesse des simulations de fonctions aléatoires, Geostatistics Doctoral Thesis, School of Mines of Paris.
- Allard D. (1993) Connexité des ensembles aléatoires: applications à la simulation de réservoirs pétroliers hétérogènes, Geostatistics Doctoral Thesis, School of Mines of Paris.
- Allard, D. (1994) "Simulating A Geological Lithofacies with Respect to Connectivity Information using the Truncated Gaussian Model", in Armstrong, M. and Dowd P.A. (eds.) Geostatistical Simulations, Kluwer, Dordrecht, pp. 197-211.
- Arak, T. and Surgailis, D. (1989) "Markov fields with polygonal realizations", Probability Theory and Related Fields, Vol. 80, pp. 543-579.
- Armstrong, M., Galli, A., Le Loch, G., Geffroy, F. and Eschard, R. (2001) Pluri-gaussian simulations, Kluwer, Dordrecht.
- Barker, A.A. (1965) "Monte Carlo calculations of radial distribution functions for a proton-electron plasma", Aust. J. Phys., Vol. 18, pp. 119-133.
- Baxter, J.R. and Rosenthal, J.S. (1995) "Rates of convergence for everywhere-positive Markov chains", Stat. Prob. Letters, Vol. 22, pp. 333-338.
- Bélisle, C.J.P., Romeijn, H.E. and Smith, R.L. (1993) "Hit-and-run algorithms for generating multivariate distributions", Math. of Operation Research, Vol. 18-2, pp. 255-266.
- Bélisle, C.J.P. (1997) "Slow convergence of the Gibbs sampler", Technical report 172, DEpt. of Statistics, Univ. of British Columbia.
- Besag, J.E. (1974) "Spatial interaction and the statistical analysis of lattice systems", Journal of the Royal Statistical Society, Vol. B36, pp. 192-225.
- Boissonnat, J.D. and Yvinec, M. (1998) Algorithmic Geometry. University Press, Cambridge.
- Borovkov, A.A. and Foss, S.G. (1994) "Two ergodicity criteria for stochastically recursive sequences", Acta Applic. Math, Vol. 34, pp. 125-134.
- Bouleau, N., (1986) Probabilités de l'Ingénieur, Hermann, Paris.
- Box, G.E.P. and Muller, M.E. (1958) "A note on the generation of random normal deviates", Ann. Math. Stat., Vol. 29, pp. 610-611.
- Brézis, H. (1992) Analyse fonctionnelle. Théorie et applications. Masson, Paris.

- Brooker, P.I. (1985) "Two-dimensional simulation by turning bands", *J. of Math. Geol.*, Vol. 17-1, pp. 81-90.
- Carter, D.S. and Prenter, P.M. (1972) "Exponential spaces and counting processes", *Zeit. für Wahr. und ver. Gebiete*, 21, pp. 1-19.
- Chambers, J.M., Mallows, C.L. and Stuck B.W. "A method for simulating stable random variables", *JASA*, Vol. 71, pp. 340-344.
- Chauveau, D., Diebolt, J. and Robert, Ch. (1998) "Control by the Central Limit Theorem", in Robert, Ch. (ed.) *Discretization and MCMC convergence assessment*, *Lectures Notes in Statistics* 135, Springer, New York, pp. 99-126.
- Chauvet, P. (1994) *Aide-Mémoire de Géostatistique Linéaire*. Cahiers du Centre de Géostatistique, Fascicule 2. School of Mines of Paris.
- Cheng, R.C.H. (1977) "The generation of gamma variables", *Appl. Stat.*, Vol. 26, pp. 71-75.
- Chentsov, N.N. (1957) "Lévy brownian motion for several parameters and generalized white noise", *Th. of Prob. and its Appl.*, Vol. 2, pp. 265-266.
- Chessa, A. (1995) *Conditional simulation of Spatial Stochastic Models for Reservoir Heterogeneity*, Ph. D. thesis, Technical University of Delft.
- Chilès, J.P. (1977) *Géostatistique des phénomènes non stationnaires (dans le plan)*, Doctoral Thesis, University of Nancy 1, France.
- Chilès, J.P. (1995) "Quelques méthodes de simulation de fonctions aléatoires intrinsèques", *Cahiers de Géostatistique*, Ecole des Mines de Paris, pp. 97-112.
- Chilès, J.P. and Delfiner, P. (1999) *Geostatistics: Modeling Spatial Uncertainty*, Wiley, New York.
- Comtet, L. (1970) *Analyse combinatoire*, Presses Universitaires de France, Paris.
- Cox D.R. (1955) "Some statistical models connected with series of events", *Journ. Royal Stat. Soc.*, B17, pp. 129,164.
- Cramér H. (1945) *Mathematical methods of statistics*, Princeton University Press, New Jersey.
- Daley D.J. and Vere-Jones, D. (1988) *An introduction to the theory of point processes*, Springer, New York.
- Daly C. (1991) *Applications de la Géostatistique à quelques problèmes de filtrage*, *Geostatistics Doctoral Thesis*, School of Mines of Paris.
- Daly C. (1993) "Filtering non-linear functions", in A. Soares (ed.) *Geostatistics Tróia'92*, Kluwer, Dordrecht, pp. 61-72.
- De Fouquet C., Beucher H., Galli A. and Ravenne C. (1989) "Conditional simulation of random sets. Application to an argileous sandstone reservoir", in Armstrong M. (ed.) *Geostatistics*, Kluwer Academic Publishers, Dordrecht, pp. 517-530.
- Dellacherie, C. and Meyer, P.A. (1983) *Probabilités et potentiel. Théorie discrète du potentiel*, Hermann, Paris.
- Delfiner, P. (1970) "Le schéma booléen-poissonien", Internal Report N-212, Centre de Morphologie Mathématique, Fontainebleau.
- Delfiner, P. (1972) "A generalization of the concept of size", *Journal of Microscopy*, Vol. 95-2, pp. 203-216.
- Devroye, L. (1986) *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.
- Diaconis, P. and Freedman, D. (1999) "Iterated random functions", *SIAM Review*, Vol. 41-1, pp. 45-76.
- Dietrich, C.R. (1995) "A simple and efficient domain implementation of the turning bands method", *Water Res. Res.*, Vol. 31-1, pp. 147-156.
- Doebelin, W. (1937) "Sur les propriétés asymptotiques de mouvements régis par certains types de chaînes simples", *Bull. Math. Soc. Roum. Sci.*, Vol. 39-1, pp. 57-115; Vol. 39-2, pp. 3-61.

- Dunford, N. and Schwartz, J.T. (1967) *Linear operators*, Interscience Publishers, New York.
- Feller, W. (1971) *An introduction to Probability Theory and its Applications*, Vol. 2, Wiley, New York.
- Fill, J. (1998) "An interruptible algorithm for exact sampling via Markov chains", *Annals of Applied Probability*, Vol. 8, pp. 131-162.
- Fill, J., Machida, M., Murdoch, D.J. and Rosenthal, J.S. (2000) "Extensions of Fill's perfect rejection sampling algorithm to general chains", Technical Report, Department of Mathematical Sciences, the Johns Hopkins University.
- Freulon, X. (1992) *Conditionnement du modèle gaussien par des inégalités ou des randomisées*, Geostatistics Doctoral Thesis, School of Mines of Paris.
- Freulon X. (1994) "Conditional Simulation of a Gaussian Random Vector with Non Linear and/or Noisy Observations", in M. Armstrong and P.A. Dowd (eds.) *Geostatistical Simulations*, Kluwer, Dordrecht, 57-71.
- Gaetan, C. and Guyon, X. (1997) "Simulation des modèles de Gibbs spatiaux par chaîne de Markov", in Jeulin, D. (ed.) *Advances in Theory and Applications of Random Sets*, World Scientific Publishing Company, pp. 289-318.
- Galli A., Beucher H., Le Loc'h G. and Doligez B. (1994) "The Pros and Cons of the Truncated Gaussian Method", in Armstrong, M. and Dowd P.A. (eds.) *Geostatistical Simulations*, Kluwer, Dordrecht, pp. 217- 233.
- Galli A. and Gao H. (2001) "Rate of convergence of the Gibbs sampler in the Gaussian case", accepted for publication in the *J. of Math. Geol.*
- Gedler, G. (1991) *Algorithme de simulation du schéma booléen conditionnel*, Doctoral Qualifying Project, Centre de Géostatistique.
- Geman, S. and Geman, D. (1984) "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images", *I.E.E.E. Trans. Pattern Analysis and Machine Intelligence*, 6, pp. 721-741.
- Gneiting, T. (1996) "Comment on "A simple and efficient space domain implementation of the turning bands method" by C.R. Dietrich", *Water Res. Res.*, Vol. 32-11, pp. 3391-3396.
- Gneiting, T. (1998) "Closed form solutions of the two-dimensional turning bands equation", *J. of Math. Geol.*, Vol. 30, pp. 379-390.
- Guibal, D. (1972) "Simulation de schémas intrinsèques", Internal Report N-291, Centre de Morphologie Mathématique, Fontainebleau.
- Guyon, X. (1995) *Random fields on a network - Modelling*, Statistics and Applications, Springer, New York.
- Haas, A., Matheron, G. and Serra, J. (1967) "Morphologie Mathématique et granulométrie en place", *Annales des Mines*, Vol. 11, pp. 736-753 and Vol. 12, pp. 767-782.
- Hadwiger, H. (1957) *Vorlesungen über Inhalt, Oberfläche und Isoperimetrie*. Springer, Berlin.
- Häggstöm, O., van Lieshout, M.N.M. and Møller, J. (1999) "Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes", *Bernoulli*, Vol. 5, pp. 641-659.
- Hajek, B. (1988) "Cooling schedules for optimal annealing", *Math. Op. Res.*, Vol. 13-2, pp. 311-329.
- Haldorsen, H.H. (1983) *Reservoir characterization Procedures for Numerical Simulation*, Ph. D. thesis, University of Texas, Austin.
- Hall, P. (1988) *Introduction to the Theory of Coverage Processes*, Wiley, New York.
- Halmos, P.R. (1969) *Measure theory*, Van Nostrand, New York (13<sup>th</sup> edition).
- Hammersley, J.M. and Nelder, J.A. (1955) "Sampling from an isotropic Gaussian process", *Proc. Cambridge Phil. Soc.*, Vol. 51, pp. 652-662.

- Hardy, G.H. and Wright, E.M. (1979) *An Introduction to the Theory of Numbers*, Oxford University Press, London (5<sup>th</sup> edition).
- Hastings, W.K. (1970) "Monte-Carlo sampling methods using Markov chains and their applications", *Biometrika*, Vol. 57, pp. 97-109.
- Hegstad, B.K., Omre, H. Tjelmeland, H. and Tyler, K. (1994) "Stochastic simulation and conditioning by annealing in reservoir description", in Armstrong, M. and Dowd P.A. (eds.) *Geostatistical Simulations*, Kluwer, Dordrecht, pp. 43-55.
- Huber, P.J. (1981) *Robust statistics*, Wiley, New York.
- Hwang, C. (1980) "Laplace's method revisited: weak convergence of probability measures", *The Annals of Stat.*, Vol. 8-6, pp. 1177-1182.
- Jeulin, D. (1991) *Modèles morphologiques de structures aléatoires et de changement d'échelle*, Doctoral Thesis, University of Caen.
- Jeulin, D. and Jeulin, P. (1981) "Synthesis of rough surfaces of random morphological functions", *Stereo. Iugosl.*, Vol. 3-1, pp. 239-246.
- Jørgensen B. (1982) *Statistical Properties of the Generalized Inverse Gaussian Distribution*, Lecture Notes in Statistics, Springer-Verlag, New York.
- Journel, A.G. and Huijbregts, C.J. (1978) *Mining Geostatistics*, Academic Press, London.
- Kendall, D.G. (1974) "Foundations of a Theory of Random sets", in Harding, E.F. and Kendall, D.G. (ed.) *Stochastic Geometry*, Wiley, London, pp. 322-376.
- Kendall, M.G. and Stuart, A. (1977) *The Advanced Theory of Statistics*, Charles Griffin Ltd., London.
- Kendall, W.S. and Thönnnes, E. (1999) "Perfect simulation in Stochastic Geometry", *Pattern Recognition*, Vol. 32, pp. 1569-1586.
- Kendall, W.S. and Møller, J. (1999) "Perfect Metropolis-Hastings simulation of locally stable point processes" Technical Report R-99-2001, Department of Mathematical Sciences, Aalborg University.
- Kiêu, K. (1997) *Three Lectures on Systematic Geometric Sampling*. Memoir No. 13, Department of Theoretical Statistics, University of Aarhus.
- Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. (1983) "Optimization by simulated annealing", *Science*, Vol. 220, pp. 671-680.
- Kleingeld, W.J. (1987) *La Géostatistique pour des variables discrètes*, Geostatistics Doctoral Thesis, School of Mines of Paris.
- Kleingeld, W.J. and Lantuéjoul, C. (1993) "Some comments on the sampling of highly dispersed type orebodies", in A. Soares (ed.) *Geostatistics Tróia'92*, Kluwer, Dordrecht, pp. 953-964.
- Kleingeld, W.J., Lantuéjoul, C., Prins, C.F. and Thurston, M.L. (1997) "The conditional simulation of a Cox process with application to deposits with discrete particles", in E.Y. Baafi and N.A. Schofield (eds.) *Geostatistics Wollongong'96*, Kluwer, Dordrecht, pp. 683-694.
- Kolmogorov, A. (1933) *Grundbegriffe der Wahrscheinlichkeitrechnung*, Springer, Berlin.
- Krige, D.K. (1951) *A Statistical Approach to Some Mine Valuations and Allied Problems in the Witwatersrand*, Thesis, University of Witwatersrand, Johannesburg.
- Knuth, D.E. (1969) *The Art of Computing Programming*, Vol. 2 (Seminumerical Algorithms), Addison-Wesley, Reading.
- Lajaunie, C. (1998) Personal communication.
- Lajaunie, Ch. et Lantuéjoul, Ch. (1989) "Setting up the general methodology for discrete isofactorial models", in Armstrong, M. (ed.) *Geostatistics*, Kluwer, Dordrecht, pp. 323-334.



- Lantuéjoul, C. (1988) "Some Stereological and statistical consequences derived from Cartier's formula", *Journal of Microscopy*, Vol. 151.3, pp. 265-276.
- Lantuéjoul, C. (1991) "Ergodicity and Integral Range", *Journal of Microscopy*, Vol. 161-3, pp. 387-404.
- Lantuéjoul, C. (1993) "Substitution Random Functions", in Soares A. (ed.) *Geostatistics Tróia'92*, Kluwer, Dordrecht, pp. 37-48.
- Lantuéjoul C. (1994) "Non conditional Simulation of Stationary Isotropic Multi-gaussian Random Functions", in Armstrong, M. and Dowd P.A. (eds.) *Geostatistical Simulations*, Kluwer, Dordrecht, pp. 147-177.
- Lantuéjoul, C. (1996) "A stereological algorithm to generate uniform points over a domain in a high dimensional space", *Acta Stereologica*, Vol. 15-1, pp. 71-76.
- Lantuéjoul C. (1997) "Conditional simulation of object-based models", in Jeulin, D. (ed.) *Advances in Theory and Applications of Random Sets*, World Scientific Publishing Company, pp. 289-318.
- Lehmer, D.H. (1951) "Mathematical methods in large-scale computing units", *Ann. Comp. Lab. Harvard Univ.*, 6, pp. 141-146.
- Le Loc'h, G., Beucher, H., Galli, A. and Doligez, B. (1994) "Improvement in the truncated gaussian method: combining several gaussian functions", *Proc. ECMOR IV, 4th European Conference on the mathematics of oil recovery*, Røros, Norway, 13 pp.
- Le Loc'h, G. and Galli, A. (1997) "Truncated plurigaussian method: theoretical and practical points of view", in Baafi, E.Y. and Schofield, N.A. (eds.) *Geostatistics Wollongong'96*, Kluwer, Dordrecht, pp. 211-222.
- Lewis, P.A.W. (1972) *Stochastic Point Processes*. Wiley-Intersciences, New York.
- Machida, M. (1999) *Stochastic monotonicity and realizable monotonicity*. Ph.D. Dissertation, Department of Mathematical Sciences, The Johns Hopkins University.
- Mandelbrot, B. B. (1975) *Fonctions aléatoires pluri-temporelles: approximation Poissonienne et généralisations*. C.R.A.S. Série A, 280, 1075-1078.
- Mantoglou, A. and Wilson, J.L. (1982) "The Turning Bands Method for Simulation of Random Fields Using Line Generation by A Spectral Method", *Water Res. Res.*, Vol. 18-5, pp. 1379-1394.
- Matérn, B. (1986) *Spatial Variation*, Lecture Notes in Statistics Vol. 36, Springer-Verlag, Berlin (2<sup>nd</sup> edition).
- Matheron G. (1965) *Les variables régionalisées et leur estimation*. Masson, Paris.
- Matheron, G. (1967) *Éléments pour une théorie des milieux poreux*, Masson, Paris.
- Matheron, G. (1968) "Schéma Booléen séquentiel de partition aléatoire", Internal Report N-83, Centre de Morphologie Mathématique, Fontainebleau.
- Matheron, G. (1969) *Théorie des Ensembles Aléatoires*. Cahiers du Centre de Morphologie Mathématique, Fasc. 4. School of Mines of Paris.
- Matheron, G. (1970) *La Théorie des Variables Régionalisées et ses Applications*. Cahiers du Centre de Morphologie Mathématique, Fascicule 5. School of Mines of Paris. Translation (1971) *The Theory of Regionalized variables and its Applications*.
- Matheron, G. (1971) "Les polyèdres Poissonniens isotropes". 3<sup>ème</sup> European Symposium on Fragmentation, Cannes, 5-8 octobre.
- Matheron, G. (1972) *Le/c cons sur les fonctions aléatoires d'ordre 2*. Course C53, School of Mines of Paris.
- Matheron, G. (1972a) "Quelques aspects de la mont'ee", Internal Report N-271, Centre de Morphologie Mathématique, Fontainebleau.
- Matheron, G. (1973) "The intrinsic random functions and their applications", *Adv. Appl. Prob.*, Vol. 5, pp. 439-468.
- Matheron, G. (1975) *Random sets and Integral Geometry*. Wiley, New York.

- Matheron G. (1975a) "Compléments sur les modèles isofactoriels", Internal Report N-432, Centre de Géostatistique.
- Matheron, G. (1978) "La formule de Steiner pour les érosions", J. Appl. Prob., Vol. 15, pp. 126-135.
- Matheron G. (1981) "Quatre familles discrètes", Internal Report N-703, Centre de Géostatistique.
- Matheron G. (1981) "La déstructuration des hautes teneurs et le krigeage des indicatrices", Internal Report N-761, Centre de Géostatistique.
- Matheron G. (1984) "The Selectivity of the distributions and the Second principle of Geostatistics", in Verly G. and al. (eds.) Geostatistics for Natural Ressources Characterization, NATO ASI Series C, Vol. 122, Reidel, Dordrecht, pp. 421-433.
- Matheron G. (1985) "Comparaison de quelques distributions du point de vue de la sélectivité", Internal Report N-964, Centre de Géostatistique.
- Matheron, G., Beucher, H., de Fouquet, C., Galli, A., Guérillot, D. and Ravenne, C. (1987) "Conditional simulation of a fluvio-deltaic reservoir", SPE paper 16753, 62nd Annual Technical Conference and Exhibition of SPE, , Dallas, pp. 591-599.
- Matheron, G. (1989) Estimating and Choosing, Springer-Verlag, Berlin.
- Matheron G. (1989a) "Two classes of isofactorial models", in Armstrong M. (ed.) Geostatistics, Kluwer Academic Publishers, Dordrecht, pp. 309-322.
- Matheron, G. (1993) "Une conjecture sur la covariance d'un ensemble aléatoire", Cahiers de Géostatistique, Fasc. 3, Ecole des Mines de Paris, pp. 107-113.
- Mecke, J. (1983) "Zufällige Mosaik", Stochastische Geometrie, Akademie-Verlag, Berlin.
- Meijering, J.L. (1953) "Interface area, edge length, and number of vertices in crystal aggregates with random nucleation", Philips Res. Rep., Vol. 8, pp. 270-290.
- Metropolis N. and al. (1953) "Equations of State Calculations by fast Computing Machines", the Journal Of Chemical Physics, Vol 21-6, pp. 1087-1092.
- Meyn, S.P. and Tweedie, R.L. (1993) Markov chains and Stochastic Stability, Springer-Verlag, London.
- Miles, R.E. (1961) Random polytopes: the generalization to  $n$  dimensions of the intervals of a Poisson point process, Ph.D. Thesis, University of Cambridge.
- Miles, R.E. (1972) "Multi-dimensional perspectives on stereology", Journal of Microscopy, Vol. 95-2, pp. 181-195.
- Miles, R.E. (1972a) "The random division of space", special supplement to Adv. in Applied Prob., pp. 243-266.
- Miles, R.E. (1974) "A synopsis of Poisson flats in Euclidean spaces". Stochastic Geometry, publié par Harding, E.F. et Kendall, D.G. Wiley, London, pp. 202-227.
- Miles, R.E. and Maillardet, R.J. (1982) "The basic structures of Voronoi and generalized Voronoi polygons", Journal of Appl. Prob., Vol. 19-1, pp. 97-111.
- Mira, A., Møller, J. and Roberts, G.O. (1999) "Perfect slice samplers", Technical Report R-99-2020, Department of Mathematical Sciences, Aalborg University.
- Molchanov, I.S. (1993) Limit Theorems for Unions of Random Closed Sets, Lecture Notes in Mathematics 1561, Springer-Verlag, Berlin.
- Molchanov, I.S. (1997) Statistics of the Boolean Model for Practitioners and Mathematicians, John Wiley and Sons, Chichester.
- Møller, J. (1989) "Random tessellations in  $\mathbb{R}^d$ ", Adv. Appl. Prob., Vol. 21, pp. 37-73.

- Møller, J. (2000) "A review on perfect simulation in stochastic geometry", Technical Report R-00-2016, Department of Mathematical Sciences, Aalborg University.
- Monestiez, P., Meiring, W., Sampson, P.D. and Guttorp, P. (1997) "Modelling non-stationary spatial covariance structure from space-time monitoring data", In Precision agriculture: spatial and temporal variability of environmental quality, Ed. Ciba Foundation, Wiley, Chichester, pp. 38-51.
- Murdoch, D.J. and Green, P.J. (1998) "Exact sampling from a continuous state space", *Scand. J. Stat.*, Vol. 25, pp. 483-502.
- Neveu, J. (1965) *Mathematical Foundations of the Calculus of Probability*, Holden-Day, San Francisco.
- Preparata, F.P. and Shamos, M.I. (1985) *Computational Geometry: An introduction*. Springer-Verlag, New York.
- Preston, C. (1977) "Spatial Birth-and-Death Processes", *Bull. Int. Stat. Inst.*, Vol. 46, pp. 371-391.
- Préteux, F. and Schmitt, M. (1988) "Boolean texture analysis and synthesis", in Serra, J. (ed.) *Image Analysis and Mathematical Morphology, Theoretical Advances*, Academic Press, London, pp. 377-400.
- Propp, J.G. and Wilson, D.B. (1996) "Exact sampling with coupled Markov chains and applications to statistical mechanics", *Random structures and Algorithms*, Vol. 9, pp. 223-252.
- Rao, M.M. (1984) *Probability Theory with Applications*, Academic Press, Orlando.
- Rivoirard, J. (1994) *Introduction to Disjunctive Kriging and non-linear Geostatistics*, Clarendon Press, Oxford.
- Roberts, G.O. and Rosenthal, J.S. (1998) "On convergence rates of Gibbs samplers for uniform distributions", *The Annals of Appl. Prob.*, Vol. 8-4, pp. 1291-1302.
- Rockafellar, R.T. (1972) *Convex Analysis*, Princeton University Press, Princeton.
- Rohatgi, V.K. (1976) *An Introduction to Probability Theory and Mathematical Statistics*, Wiley, New York.
- Rubinstein, R.Y. (1981) *Simulation and the Monte Carlo Method*, Wiley, New York.
- Sampson, P.D. and Guttorp, P. (1992) "Nonparametric estimation of nonstationary spatial covariance structure", *J. Amer. Stat. Assoc.*, Vol. 87, pp. 108-119.
- Santaló, L. (1976) *Integral Geometry and Geometric Probability*. Addison-Wesley, Reading.
- Schmitt M. (1991). "Estimation of the density in a stationary Boolean model". *J. of Appl. Prob.*, Vol. 28, pp. 702-708.
- Schmitt, M. (1997) "Estimation of intensity and shape in a non-stationary Boolean model", in Jeulin, D. (ed.) *Advances in Theory and Applications of Random sets*, World Scientific, Singapore, pp. 251-267.
- Schmitt, M. (2000) "Random sets, sampling and connectivity", presented at the 6<sup>th</sup> Geostatistics Congress, Somerset West (South Africa).
- Shoenberg, I.J. (1938) "Metric spaces and completely monotone functions", *Annals of Mathematics*, Vol. 39., pp. 811-841.
- Senoussi, R., Chadoeuf, J. and Allard, D. (2000) "Weak homogenization of point processes by space deformation", *J. of Appl. Prob.*, Vol. 32-4, pp. 948-959.
- Serra, J. (1968) "Fonctions aléatoires de Dilution", Internal Report C-12, Centre de Morphologie Mathématique, Fontainebleau.
- Serra, J. (1982) *Image Analysis and Mathematical Morphology*. Academic Press, London.

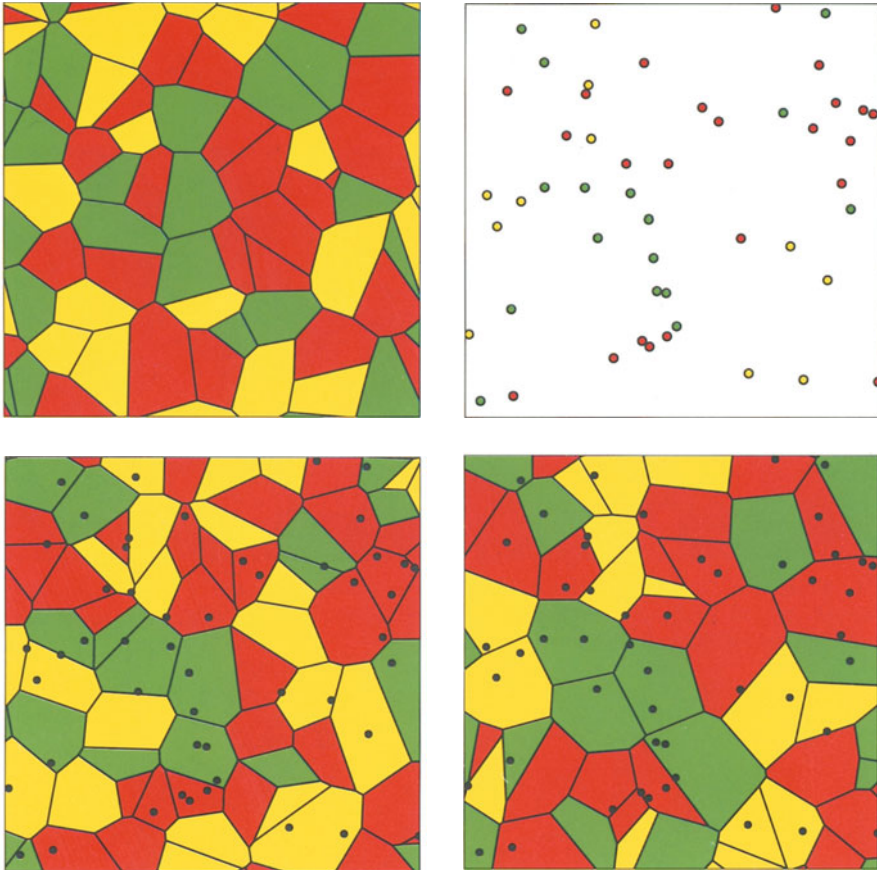
- Serra, J. (1988) "Boolean random functions", in Serra, J. (ed.) *Image Analysis and Mathematical Morphology, Theoretical Advances*, Academic Press, London, pp. 317-342.
- Shapiro, S.S. and Wilk, M.B. (1965) "An analysis of variance for normality (complete samples)", *Biometrika*, Vol. 52-3, 591-611.
- Shinozuka, M. and Jan, C.M. (1972) "Digital simulation of random processes and its applications", *J. of Sound and Vib.*, Vol. 25-1, pp. 111-128.
- Sichel H.S. (1973) Statistical valuation of diamondiferous deposits, in Salamon M.D.G. and Lancaster F.H. (eds.), *Application of Computer methods in the Mineral Industry*, The South African Institute of Mining and Metallurgy, Johannesburg, 17-25.
- Sichel H.S. (1974) On a distribution representing sentence-length in written prose, *J. Royal Stat. Soc., Serie A*, Vol. 137, pp. 25-34.
- Smith, R.L. (1983) "Efficient Monte Carlo procedures for Generating Points Uniformly Distributed over Bounded Regions", *Operations Research*, Vol. 32-6, pp. 1296-1308.
- Stoyan, D., Kendall, W.S. and Mecke, J. (1985) ( $2^{nd}$  ed., 1995) *Stochastic Geometry and Its Applications*, Wiley, New York.
- Sylverseen, A.R. and Omre, H. (1994) "Marked point models with complex conditioning used for the modelling of shales", *Proceedings ECMOR IV*, Røros, Norway.
- Sylverseen, A.R. and Omre, H. (1997) "Conditioning of Marked point Processes within a Bayesian framework", *Scand. Journ. of Stat.*, Vol. 24-3, pp. 341-352.
- Tierney, L. (1994) "Markov chains for exploring posterior distributions", *The Annals of statistics*, Vol. 22-4, pp. 1701-1762.
- Turchin, V.F. (1971) "On the computation of multidimensional integrals by the Monte-Carlo method", *Theory of Prob. and Appl.*, Vol. 16, pp. 720-724.
- Vedel-Jensen, E.B. and Stougaard-Nielsen, L. (1998) "Inhomogeneous Markov point processes by transformation", *Research Report 2, Lab. Comp. Stoch.*, pp. 1-24.
- Vitale R.A. (1983) "Some developments in the theory of random sets", *Bull. Int. Stat. Inst.*, Vol. 50, pp. 863-871.
- von Neumann J. (1951) "various techniques used in connection with random digits", *U.S. Nat. Bur. Stand. Appl. Math. Ser.*, No. 12, pp. 36-38.
- Weil W. and Wieacker J.A. (1984). "Densities for stationary random sets and point processes". *Adv. in Applied Prob.*, Vol 16-2, pp. 324-346.
- Winkler, G. (1995) *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer, Berlin.
- Yaglom, A.M. (1987) *Correlation Theory of Stationary and related Random Functions*, Springer Series in Statistics, New York.

# Index

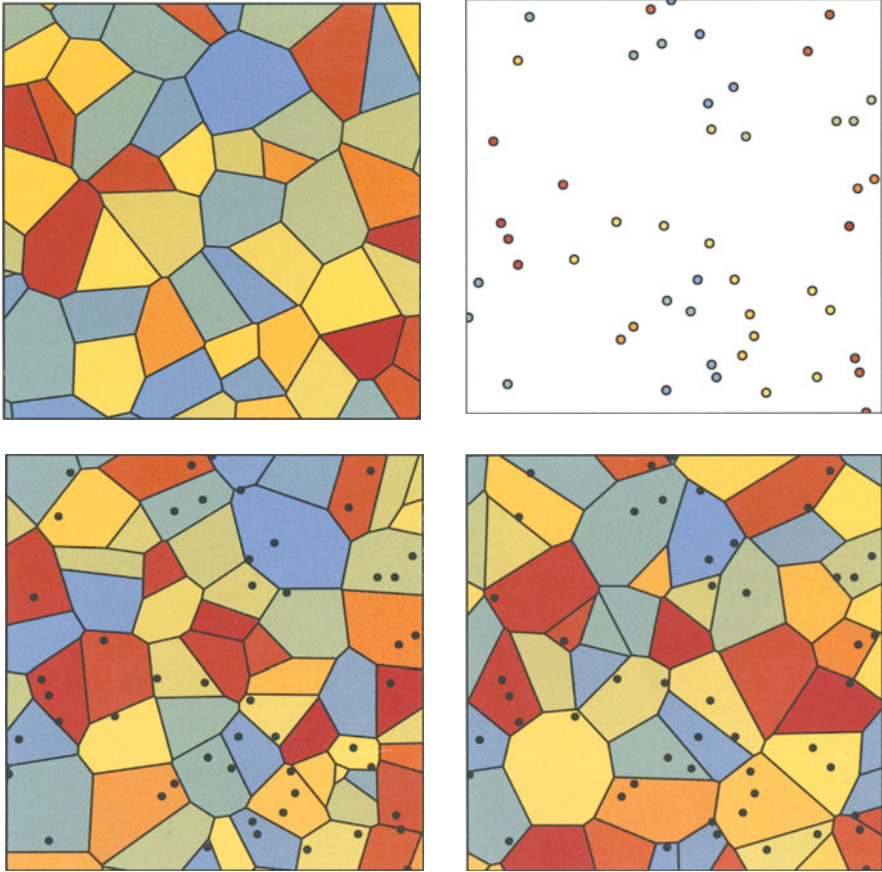
- $\sigma$ -algebra, 9
  
- Acceptance-rejection method, 60, 112
- Adler's criterion, 187, 230
- Aperiodicity, 69
- Autodual set, 19
- Avoiding functional, 14
  
- Backward iteration, 101
- Berry-Esséen theorem, 183
- Bochner's theorem, 24, 191
- Boolean model, 153
- Boolean random function, 172
- Borelian, 16
  
- Cauchy's formula, 49
- Central limit theorem, 183
- Chentsov's model, 224
- Closing, 42
- Coding process, 221
- Conditionally negative definiteness, 25
- Convex ring, 52
- Convexity, 47
- Correlation function, 24
- Coupling from the past, 101
- Covariance function, 24
- Cox process, 129
- Crofton's formula, 50
  
- Dead leaves model, 133, 176
- Dilation, 39
- Dilution random function, 167
- Directing random function, 221
- Dispersion variance, 32
- Down-set, 107
  
- Ergodicity, 30
- Erosion, 40
- Euler-Poincaré characteristic, 52
- Event, 9
- Exact simulation, 101, 178, 233
  
- Excursion set, 205
  
- Fatou's lemma, 106
- Fermat's theorem, 58
- Finite order population, 18, 155
- Fundamental cell, 136
  
- Gamma function, 48
- Gaussian random function, 185
- Generating function, 127
- Genus, 52
- Geometric covariogram, 22, 136, 165, 168, 177
- Geometric ergodicity, 87
- Gibbs sampler, 76, 210, 233
  
- Hausdorff distance, 18, 42
- Hermite polynomials, 214
- Hitting functional, 14
  
- Increasing mapping, 107
- Indicator function, 11
- Indicator variogram, 26
- Integral range, 30, 95, 186, 227
- Interruptibility, 113
- Invariance, 70
- Inversion method, 61
- Irreducibility, 69
- Is factorial expansion, 91, 224
- Isopartition, 228, 229
  
- Jacobi transition kernel, 94
- Jensen's inequality, 90
  
- Kolmogorov distance, 197
  
- Lévy distance, 197
- Law of large numbers, 183
- Law of quadratic reciprocity, 59
- Legendre coefficient, 58
- Locally finiteness, 16

- Markov chain, 67
- Matheron distance, 198
- Metropolis-Hastings algorithm, 233
- Migration process, 196
- Minkowski functionals, 48
- Minkowski operations, 39
- Minorization, 89
- Multiplicative generator, 57
  
- Object detection, 43
- Object population, 17
- Objective function, 82
- Opening, 41
  
- Plurigaussian random function, 211
- Point process, 16, 119
- Poisson cells, 144
- Poisson hyperplane, 144
- Poisson point process, 120, 122
- Polyconvexity, 51
- Poset, 107
- Positive definiteness, 24, 186
- Probability, 9
- Probability space, 9
  
- Random set, 11
- Random coin model, 167
- Random function, 10
- Random mapping, 101
- Random token model, 167
- Rate of convergence, 87
- Realization, 9
- Recurrent state, 105
- Regular set, 15
- Replacement, 171
- Reversibility, 71, 72
  
- Second-order stationarity, 24
- Self similarity, 225
- Separability, 15
- Simple kriging, 199
- Simulated annealing, 80
- Spatial distribution, 10, 17, 18
- Specific volume, 26
- Spectral measure, 25, 191
- Star of a point in a set, 79
- State, 9
- Stationarity, 134
- Steiner's formula, 51
- Stochastic monotonicity, 109
- Stochastic process, 10
- Substitution random function, 221
  
- Tessellation, 133
  
- Time-reversal transition kernel, 111
- Total variation norm, 68, 197
- Transient state, 105
- Transition kernel, 68
- Transitive covariogram, 21, 169
- Typical cell, 136
- Typical object, 158, 173
  
- Uniform  $i$ -flat, 50
- Uniform distribution, 57
- Uniform ergodicity, 88
- Up-set, 107
  
- Variogram, 25
  
- Weak discrepancy, 194
- Weight in number, 135
- Weight in volume, 135
- Weighted point process, 17

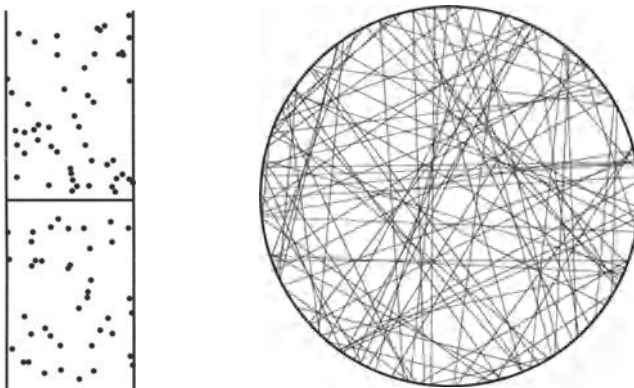
In the following, please find the (corrected) colour version of 5 figures (12.4, 12.5, 14.7, 14.8, 16.2), and Fig. 12.6 which is incomplete in the Text.



**Fig. 12.4.** Conditional simulation of a Voronoi tessellation with discrete values. Top left, a non-conditional simulation. Top right, the conditioning data set. Second row, two conditional simulations

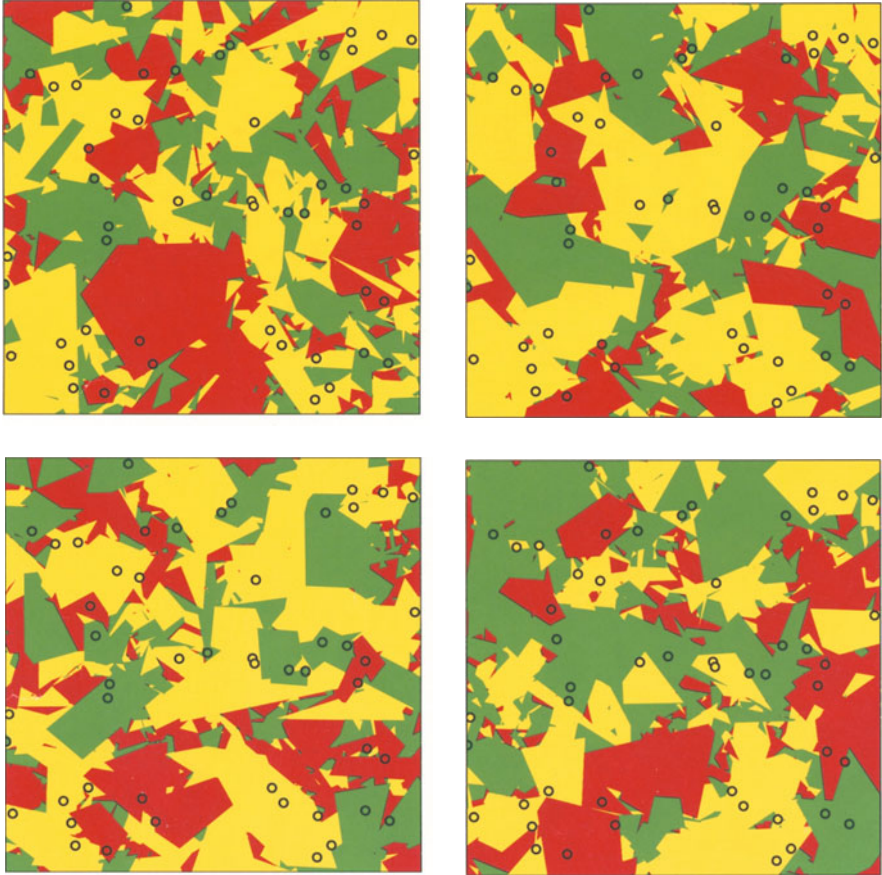


**Fig. 12.5.** Conditional simulation of a Voronoi tessellation with continuous values. Top left, a non-conditional simulation. Top right, the conditioning data set. Second row, two conditional simulations

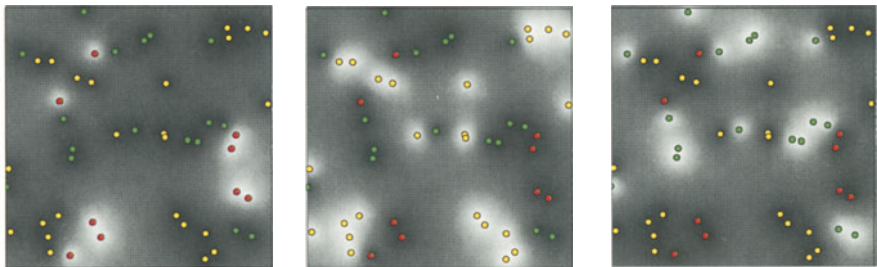


**Fig. 12.6.** Realization of a Poisson line network represented in the parameter space  $[0, \pi] \times \mathbb{R}$  (left) and in Euclidean space (right). The polygons that the lines delineate are the Poisson polygons

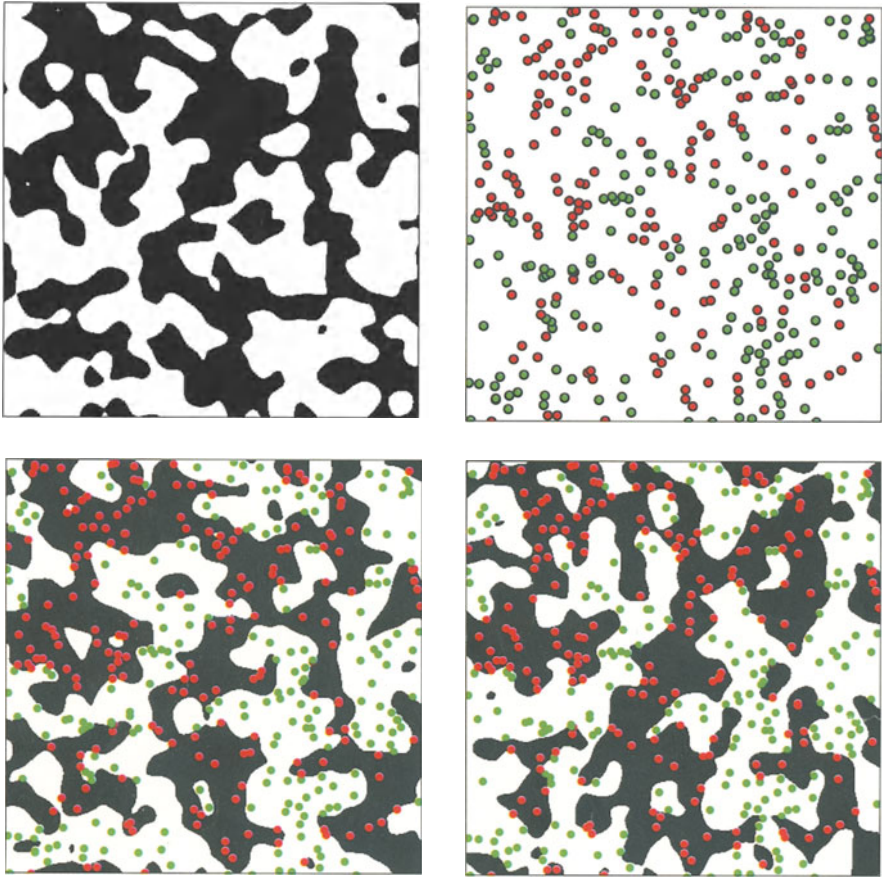




**Fig. 14.7.** Conditional simulation of a dead leaves model. The leaves are typical Poisson polygons (same distribution and same proportion for the three colours). Top left, a non conditional simulation. Top right and bottom, three conditional simulations



**Fig. 14.8.** Conditional distribution of the colour of the points of the simulation field. From left to right, this gives the probability that a point is red, yellow or green



**Fig. 16.2.** Conditional simulation of an excursion set of a gaussian random function. Top left, a non conditional simulation. Top right, the conditioning data points (the red points belong to the black phase, the green ones to the white) . Bottom, two conditional simulations