Spatial Statistics

This Page Intentionally Left Blank

Spatial Statistics

This Page Intentionally Left Blank

Spatial Statistics

BRIAN D. RIPLEY
University of London

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data is available.

ISBN 0-471-69116-X

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Preface

This is a guide to the analysis of spatial data. Spatially arranged measurements and spatial patterns occur in a surprisingly wide variety of scientific disciplines. The origins of human life link studies of the evolution of galaxies, the structure of biological cells, and settlement patterns in archeology. Ecologists study the interactions among plants and animals. Foresters and agriculturalists need to investigate plant competition and account for soil variations in their experiments. The estimation of rainfall and of ore and petroleum reserves is of prime economic importance. Rocks, metals, and tissue and blood cells are all studied at a microscopic level. The aim of this book is to bring together the abundance of recent research in many fields into the analysis of spatial data and to make practically available the methods made possible by the computer revolution.

The emphasis throughout is on looking at data. Each chapter is devoted to a particular class of problems and a data format. The two longest and most important are on *smoothing and interpolation* (producing contour maps, estimating rainfall or petroleum reserves) and on *mapped point patterns* (trees, towns, galaxies, birds' nests). Shorter chapters cover:

The regional variables of economic and human geography.

Spatially arranged experiments.

Quadrat counts.

Sampling a spatially correlated variable.

Sampling plants and animals and testing their patterns.

The final chapter looks briefly at the use of image analyzers to investigate complex spatial patterns, and stereology: how to gain information on three-dimensional structures from linear or planar sections. Some emphasis is placed on going beyond simple tests to detect "nonrandom" patterns as well as on fitting explanatory models to data. Some general families of models are discussed, but the reader is urged to find or invent models that

reflect the theories of his or her own discipline, such as central place theory for town locations. The techniques presented are designed for both of John Tukey's divisions of exploratory and confirmatory data analysis.

The level of mathematical difficulty varies considerably. The formal prerequisites are few: matrix algebra, some probability and statistics, and basic topology in parts of Chapter 9. An acquaintance with time series analysis would be helpful, especially for Chapter 5. I have tried to confine the formal mathematics to the essential minimum. Mathematically able readers will be able to find their fill in the references. It is perhaps inevitable that some of the mathematical justifications are far deeper than is the practical import of the results. But beware. There is much appealing but incorrect mathematics in the spatial literature, and some of the subtlest arguments are used to discover undesirable properties of simple procedures. I recommend readers who find the going tough to skip ahead to the examples before seriously tackling the theory.

Computers, especially computer graphics, are an essential tool in spatial statistics. Useful data sets are too large and most of the methods too tedious for hand calculation to be contemplated. Even data collection is being increasingly automated. The worked examples were analyzed at an interactive graphics terminal by FORTRAN programs running on Imperial College's CDC 6500/Cyber 174 system. Unfortunately, the reader cannot follow my decisions as I rotated plots, investigated contour levels, and altered smoothing parameters. There is no substitute for experience at a computer terminal using one's own data. Therefore, it was a difficult decision not to include programs. There was at the time of writing no agreed-upon standard for computer graphics, and the availability of plotting and other utility operations varied widely. The choice of language was also debatable. I could only use interactive graphics from FORTRAN, whereas microcomputers were becoming available with BASIC or PASCAL. Hints on algorithms and computation are included.

The bibliography is the only example I know of that attempts a comprehensive coverage of the spatial literature. It contains not only references to the theory and methods, but a large number of accounts of applications in many disciplines as well. Guides to the literature are given at the end of several chapters and sections.

B. D. RIPLEY

London
March 1981

Acknowledgments

This book was written during two periods of leave visiting the Department of Statistics, Princeton University and the Afdeling for Teoretisk Statistik, Aarhus University, Denmark. My stay at Princeton was supported by contract EI-78-5-01-6540 with the U.S. Energy Information Administration. I am grateful to Geof Watson and Ole Barndorff-Nielsen for their interest and encouragement.

Most of the figures were computer-drawn on 35-mm microfilm at the University of London Computer Centre, using procedures set up by Imperial College Computer Centre. The perspective plotting routines were developed jointly with Dan Moore. The Dirichlet tessellations and Delaunay triangulation were drawn by the program TILE of Peter Green and Robin Sibson.

Karen Byth read through the manuscript and removed many errors. I would appreciate being informed of any remaining errors and of work I have missed.

B. D. R.

This Page Intentionally Left Blank

Contents

Spatial Statistics

This Page Intentionally Left Blank

CHAPTER 1

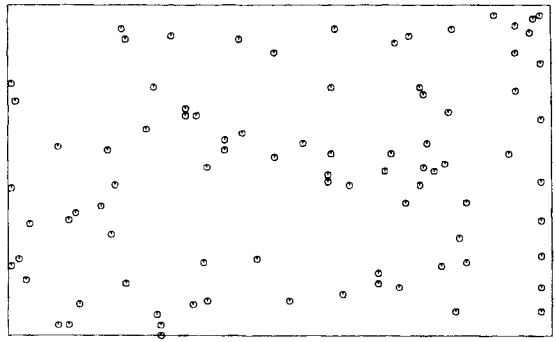
Introduction

1.1 WHY SPATIAL STATISTICS?

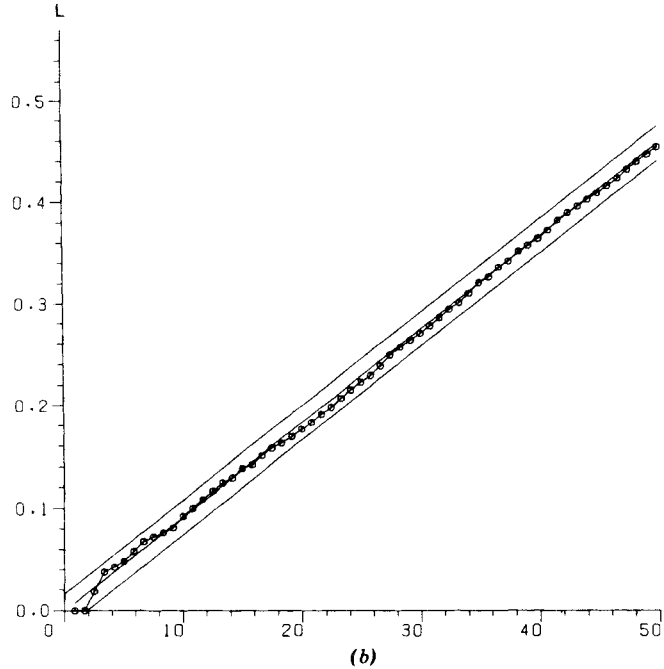
Men have been drawing maps and so studying spatial patterns for millenia, yet the need to reduce such information to numbers is rather recent. The human eye and brain form a marvelous mechanism with which to analyze and recognize patterns, yet they are subjective, likely to tire, and so to err. The explosion in computing power available to the average researcher now makes it possible to do routinely the intricate computations needed to explore complex spatial patterns.

One sense of the word "statistics" is a collection of numbers, and spatial statistics includes "spatial data analysis," the reduction of spatial patterns to a few clear and useful summaries. But statistics goes beyond this into what John Tukey has called "confirmatory data analysis," in which these summaries are compared with what might be expected from theories of how the pattern might have originated and developed. Consider, for example, Figure 1.1*a*, which is a map of trees in a rectangular plot. Figure 1.1*b* shows a summary of these data as a graph, together with confidence limits for the sort of graph we would get if each tree had been placed at random in the plot, without any reference to the positions of the rest of the trees. This example shows one of the characteristic features of the subject. There are so many different types of spatial patterns that we need to summarize the data in one or more graphs rather than by single numbers, such as the mean and standard deviation of classical statistics.

Almost invariably we will have only a single example of a particular pattern rather than the many replications of measurements found in the experimental sciences. To get some idea of the variability of such data, we are forced to make some assumption of stationarity of the underlying mechanism that generated the pattern. Such an assumption has often been disputed, particularly in the geographic literature. Its validity may depend on the questions being asked. For instance, if we are looking at



(a)



(b)

Fig. 1.1 (*a*) Point patterns of trees. (*b*) Summary of data with a 95% confidence band. See Figure 8.6 for further details.

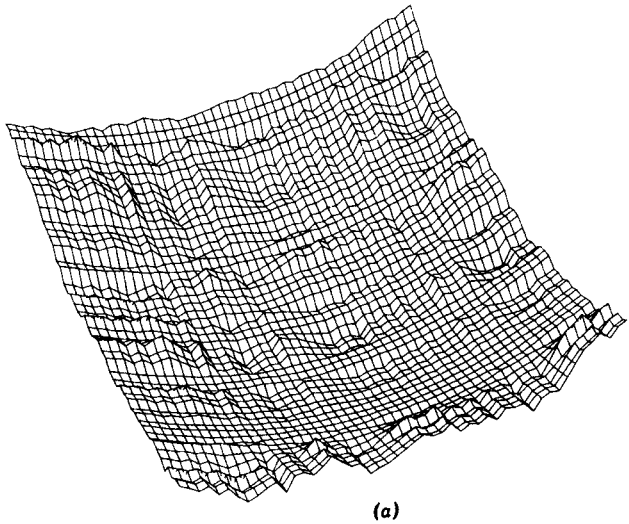
population density we may wish to know whether we need to invoke the topography (which might suggest nonstationarity) to explain the observed variations in density. Patterns that vary in a systematic way from place to place are called *heterogeneous* (opposite *homogeneous*). But we might be studying the grouping of houses that we might expect to *interact*, either *clustering* together because of human gregariousness or *inhibiting* where houses need to be close to sufficient land. Patterns can also exhibit preferred directions, called *anisotropy* (opposite *isotropy*). For example, forests that were originally planted in rows may show directionality in the crowns of the trees (Ford, 1976). We will assume that the data have been subdivided into sufficiently small units or that they have had obvious trends removed to permit us, where necessary, to invoke homogeneity or isotropy.

1.2 TYPES OF DATA

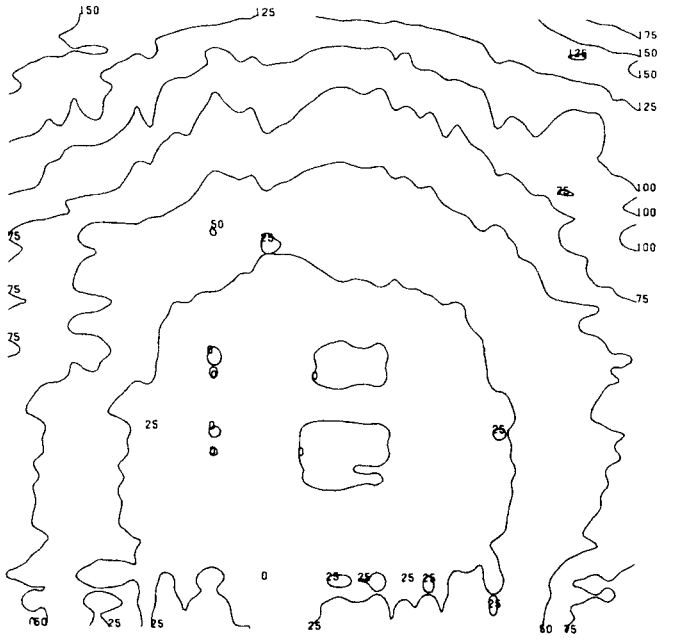
The basic subdivision of this volume is by the type of data to be analyzed. The tree positions given in Figure 1.1*a* are an example of a *point pattern*. Other examples are the locations of birds' nests, of imperfections in metals or rocks, galaxies, towns, and earthquakes. Of course, none of these is actually a point, but in each case the sizes of the objects are so small compared with the distances between them that their size may be ignored. (Sometimes size is an important explanatory variable associated with a point. For example, we might expect the area of the hinterland of a town to depend on its population size.) Maps of point patterns are discussed in Chapter 8.

Sometimes points are so numerous that complete mapping would be an unjustified effort (consider clover plants in a grassland). Two methods of sampling such point patterns are discussed in Chapters 6 and 7. In Chapter 6 we consider methods based on taking sample areas, called *quadrats*, and counting objects within each, whereas in Chapter 7 the methods are based on measuring distances to or between objects. Chapter 7 also deals with two cases in which complete mapping is either uneconomical or impossible; trees in a dense forest and animal populations such as deer and moorland grouse (game birds).

Many variables that were originally point patterns are recorded as regional totals, such as census information. If these regions are genuinely distinct, we may wish to test for correlation between the regional statistics, taking account of the connections between the regions measured by, say, the lengths of the common borders (if any) or freight costs between them.

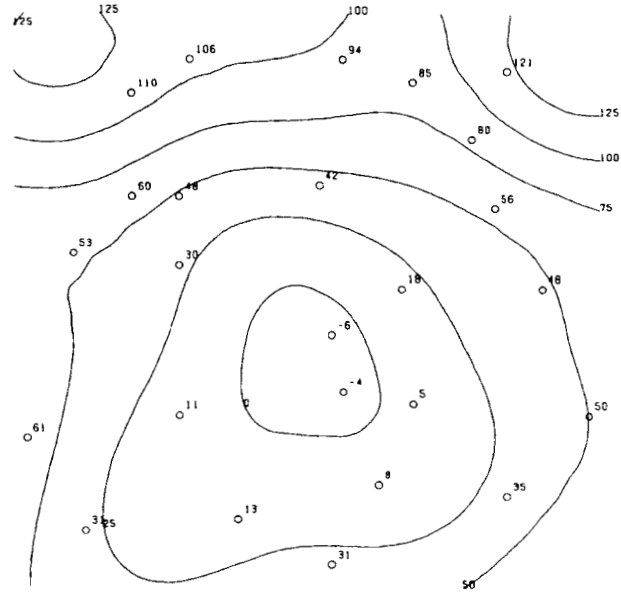


(a)

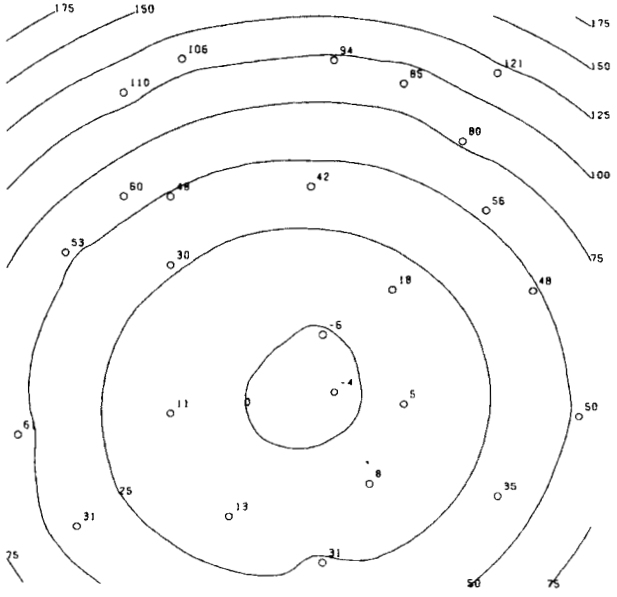


(b)

Flg. 1.2 (*a*), (*b*) A simulated surface. (*c*), (*d*) Two reconstructions from the sample points indicated as circles.



(c)



(d)

Fig. 1.2 (*continued*)

Summary measures for what is known as "spatial autocorrelation" are discussed in Section 5.4. They are particularly useful when applied to the residuals from the regression of one regional statistic on others.

Where the regions are small administrative units we might wish to smooth the data to produce a map of population density, average income, or similar variable. This problem of reconstructing a surface from irregularly spaced sample points is common; all topographical maps are prepared from such data, as is rainfall information. Geologists, oil prospectors, and mining engineers all have to reconstruct facets of an underground pattern such as the volume and average grade of ore in various parts of a mine, using spatially arranged samples. Such problems are considered in Chapter 4. Figure 1.2 illustrates a surface and two reconstructions.

Usually the locations of the sample points are fixed from other considerations, but in Chapter 3 we consider how sample points should be chosen to give the best estimate of the average level of a surface.

Data arranged on a rectangular grid are not as common as might be expected by analogy with time-series theory. They seem to arise only from man's experiments, either where he has deliberately sampled systematically or from agricultural field trials in which a field has been divided into rectangular parcels. Clearly, we would expect neighboring plots to have similar fertility and hence that the yields would be spatially autocorrelated. We show in Chapter 5 how such data might be analyzed.

The least explored class of patterns are those of two or more phases forming a mosaic. Patterns of vegetation provide two-dimensional examples, but most of the interest is in three dimensions, in bone and tissue and rock grains and pores. Descriptions of patterns such as that shown in Figure 1.3 were facilitated by the invention of image analyzers during the 1960s, these being scanning microscopes connected to computers to analyze the vast amounts of output. Stereology is the theory of reconstructing information on three-dimensional patterns from planar sections (see, for example, Figure 1.3) or linear probes. This area is the subject of Chapter 9.

All the models of the mechanisms that might generate patterns described in the chapters for each type of data are stochastic processes. Chapter 2, on "basic stochastic processes," gives an introduction to what is needed of the mathematical theory, to generic families of models, and to ways in which the computer can be used to experiment with models.

Most of the theory and methods apply equally in two or three dimensions. Where formulas depend on the dimension, only the two-dimensional case is given unless otherwise stated. Planar data are by far the most common; all the examples are planar.

Fig. 1.3 Simplified pore space (black) in a section of smackover carbonate rock.

Spatial Topics Omitted

This volume concentrates on information on location, ignoring the concepts of shape and form reflected in the monographs of Grenander (1976, 1978), Mandelbrot (1977), and Bookstein (1978). Some specialized topics omitted are on the spread of epidemics (Bartholomew, 1973; Mollison, 1977) and percolation theory (Shante and Kilpatrick, 1971; Welsh, 1977; Smythe and Wierman, 1978). Each of these references is more concerned with mathematical modeling than with analyzing data.

Little attention is given here to space–time problems. Many of the same methods can be used, but adequate data seem rare (earthquake occurrences being an exception). Often the best way to deal with space–time data is to compare the maps in successive time periods. Another generalization is to multitype problems in which the objects are of different types or where two or more patterns or surfaces are to be related. Again, the extension of many of the methods is simple. Whenever a pair of points is considered, take one from each of the two patterns or surfaces. If three or more surfaces or patterns are considered, take them in pairs. In general, the theory of multitype procedures is not satisfactory and there are few examples of its use. Pielou (1977) gives examples of some of the methods of "classical" statistics used on these problems.

More information on applications in specific disciplines may be found in:

Animal ecology	Southwood (1978)
Archeology	Hodder and Orton (1976)
Geography	Bartels and Ketellapper (1979)
	Bennett (1979)
	Berry and Marble (1968)
	Cliff and Ord (1973)
	Getis and Boots (1978)
	Haggett et al. (1977)
	Rayner (1971)
	Rogers (1974)
Geology	Davis (1973)
Mining	David (1977)
	Guarascio et al. (1976)
	Journel and Huijbregts (1978)
	Matheron (1965, 1967a)
Plant ecology	Greig-Smith (1964)
	Kershaw (1973)
	Patil et al. (1971)
	Pielou (1977)

CHAPTER 2

Basic Stochastic Processes

This chapter assumes a basic knowledge of probability theory and sets up some of the background of the models and methods used in later chapters. Section 2.4 is more mathematical and is not necessary for an understanding of the rest of the material (although its ideas are used in Sections 5.2 and 8.4).

2.1 DEFINITIONS

A *stochastic process* is a collection of random variables $\{Z(t)|t\in T\}$ indexed by a set T. It has been usual to take T to be a subset of the real numbers, say $\{1,2,3\cdots\}$ or $[0,\infty)$. However, we need more general index sets such as pairs of integers (labeling the plots in a field trial), the plane (labeling topographic heights), and rectangles in the plane (labeling counts of plants). The great distinction between these indices and those representing time is that the latter have an ordering.

The Daniell–Kolmogorov theorem states that to specify a stochastic process all we have to do is to give the joint distributions of any finite subset $\{Z(t_1),\ldots,Z(t_n)\}$ in a consistent way, requiring

$$P(Z(t_i)\in A_i, i=1,\ldots,m, Z(s)\in\mathbb{R})=P(Z(t_i)\in A_i, i=1,\ldots,m)$$

Such a specification is called the *distribution* of the process. We avoid subtle mathematics by only considering a finite number of observations on a stochastic process (except for the differentiability properties in Section 4.4).

We say that the stochastic process is stationary under translations or *homogeneous* if the distribution is unchanged when the origin of the index set is translated. For this to make sense the index set has to be unbounded; it has to be either all pairs of integers or the whole plane. If T

is the whole of the plane or three-dimensional space, we can also consider processes that are stationary under rotations about the origin, called *isotropic*. Homogeneous and isotropic processes are stationary under rigid motions. The philosophy behind these definitions is discussed in Chapter 1. Note that they can, at most, be partially checked by, for example, splitting the study region into disjoint parts and checking their similarity.

2.2 COVARIANCES AND SPECTRA

The covariance C and correlation R between $Z(s)$ and $Z(t)$ for two points in T are defined by

$$C(s,t)=E\left[\{Z(s)-E(Z(s))\}\{Z(t)-E(Z(t))\}\right]$$

$$R(s,t)=C(s,t)/\sqrt{\{C(s,s)C(t,t)\}}$$

Homogeneity implies that C and R depend only on the vector $\mathbf{h}$ from s to t, whereas with isotropy they depend only on $d(s,t)$. We will use the notation $C(\mathbf{h})$ or $C(r)$ for these reductions. Note that by symmetry $C(\mathbf{h})=C(-\mathbf{h})$, but $C((-h_1,h_2))$ may differ from $C((h_1,h_2))$. We will usually plot C in the right half-plane; the other half-plane is found by a half-turn rotation.

In general the distribution of a stochastic process is not completely determined by the mean $m(s)=E[Z(s)]$ and covariance $C(s,t)$. This *is* the case for an important class of processes, the *Gaussian processes* defined by the property that all finite collections $\{Z(t_1),\ldots,Z(t_n)\}$ are joint Normal (that is, every linear combination has a Normal distribution). It is important to know which covariance functions can occur, for given m and C we can construct a Gaussian process via the Daniell–Kolmogorov theorem with that mean and covariance. The necessary and sufficient condition is that C should be *nonnegative definite* and symmetric, that is, that $C(t,s)=C(s,t)$ and

$$\left[\sum_i \alpha_i Z(t_i)\right]=\sum_i\sum_j \alpha_i\alpha_j C(t_i,t_j)\geqslant 0 \tag{2.1}$$

for all n, $\alpha_1,\ldots,\alpha_n$, $t_1,\ldots,t_n$ (Breiman, 1968, Chapter 11). We often ask that C be strictly positive definite when (2.1) must be nonzero unless all α_i are zero.

This condition of nonnegative definiteness occurs elsewhere, thus enabling us to give examples of valid covariance functions. The characteristic function of a d-dimensional symmetric random vector $\mathbf{X}$ is a nonnegative definite continuous symmetric function on $\mathbb{R}^d$. A continuous homogeneous covariance function of a stochastic process on $\mathbb{R}^d$ will be proportional to such a characteristic function. If $\mathbf{X}$ is rotationally symmetric, the covariance is isotropic. Taking the d-dimensional Cauchy and Normal distributions (with densities proportional to $1/(1+\alpha\|\mathbf{x}\|^2)$ and $\exp -\alpha\|\mathbf{x}\|^2$) shows that e^{-ar} and e^{-ar^2} are both isotropic covariances in any number of dimensions. The *spectral density* f is defined by

$$f(\omega)=\frac{1}{(2\pi)^d}\int \exp\{-i\omega^T\mathbf{h}\}C(\mathbf{h})\,d\mathbf{h} \tag{2.2}$$

when this integral exists. Then

$$C(\mathbf{h})=\int \exp\{+i\omega^T\mathbf{h}\}f(\omega)\,d\omega \tag{2.3}$$

Thus $f/C(\mathbf{0})$ is the *pdf* of a random vector with characteristic function $C/C(\mathbf{0})$. For processes on a lattice (2.2) is replaced by a sum and only frequencies for which each component is in the range $[-\pi,\pi]$ are considered, so the integration in (2.3) is restricted to $[-\pi,\pi]^d$. Any nonnegative function that gives a finite value of $C(\mathbf{0})$ in (2.3) is a spectral density.

The spectral density inherits the symmetry condition $f(-\omega)=f(\omega)$ from the covariance function. If the covariance is isotropic $f(\omega)$ becomes a function of $\tau=\|\omega\|$ only, and we have

$$f(\tau)=\frac{1}{2\pi}\int_0^\infty J_0(r\tau)C(r)r\,dr$$

$$C(r)=2\pi\int_0^\infty J_0(r\tau)f(\tau)r\,d\tau$$

in $\mathbb{R}^2$, where J_0 is the Bessel function (Quenouille, 1949).

The requirement of isotropy on a covariance function is quite restrictive. Schoenberg (1938) conjectured that such a function was continuous, except possibly at the origin. Furthermore, Matérn (1960, pp. 13–19) shows that

$$C(r)\geqslant \inf_u \{k!(2/u)^k J_k(u)\}C(0) \qquad k=(d-2)/2$$

so that isotropic correlations are bounded below by -0.403 in $\mathbb{R}^2$ and

-0.218 in $\mathbb{R}^3$. Isotropic correlation functions are usually specified either by giving an isotropic spectral density or by "mixing" the family e^{-ar}. Suppose we choose a from some distribution, then use a process with correlation function e^{-ar}. The correlation function of the mixed process is $E(e^{-ar})$. This argument shows that any Laplace transform can be a covariance function. (This is the class of functions with $(-1)^n C^{(n)}(r) \geqslant 0$ for $n=0,1,2,\ldots$, and all $r>0$). One such family of functions are those proportional to $r^\nu K_\nu(ar)$ for $\nu>0$, with spectral densities proportional to $1/(b+\|\omega\|^2)^{\nu+d/2}$. Here K_ν is a Bessel function. The exponential correlation function is the special case $\nu=\frac{1}{2}$ (Whittle, 1954, 1956, 1963a).

The class of known examples of isotropic correlation functions is not totally satisfactory, for in practice one often finds

$$\operatorname{var}\left\{\int_A Z(\mathbf{x})\,d\mathbf{x}\right\} \sim \text{const}\{\text{area}(A)\}^{(2-\lambda/d)} \tag{2.4}$$

A famous example is given by Fairfield Smith (1938), who found $\lambda \approx 3/2$ for yields from wheat trials. Whittle (1956) showed that (2.4) is equivalent to $C(r)$ and $f(\tau)$ behaving as $r^{-\lambda}$ and $\tau^{(\lambda-d)}$ for large r and small τ, whereas all the standard examples of isotropic covariances decay exponentially at large distances.

One way to form an isotropic process in $\mathbb{R}^d$ is to take a homogeneous process Z_1 with covariance function C_1 on $\mathbb{R}$, to let $Z(\mathbf{x})=Z_1(x_1)$ and then give the whole of each realization an independent uniformly distributed rotation about the origin in $\mathbb{R}^d$. Then the covariance function of Z is

$$C(r)=\frac{2\Gamma(d/2)}{\sqrt{\pi}\,\Gamma\{\frac{1}{2}(d-1)\}}\int_0^1 C_1(vr)(1-v^2)^{(d-3)/2}\,dv \tag{2.5}$$

[Matheron, 1973, equation (4.1)]. For $d=3$ we have the simple results

$$C(r)=\int_0^1 C_1(vr)\,dv, \qquad C_1(r)=\frac{d}{dr}[rC(r)] \tag{2.6}$$

In fact (2.5) is the general form of an isotropic covariance in $\mathbb{R}^d$. It can be re-expressed as

$$C(r)=E\{C_1(rV)\} \tag{2.7}$$

where V is the first coordinate of an independent uniformly distributed point on the surface of the unit ball in $\mathbb{R}^d$. We have noted that $C/C(0)$ is the characteristic function of a random vector $\mathbf{X}$. Because C is isotropic

$\mathbf{X}$ has a rotationally symmetric distribution and

$$C(r) = C(0)\, E(\exp\{i \|\mathbf{X}\| V\}) \tag{2.8}$$

Comparison of (2.7) and (2.8) shows that we can take $C_1(t) = C(0)\, E(\exp\{it\|\mathbf{X}\|\})$, which is nonnegative definite and symmetric and so a covariance function in $\mathbb{R}^1$.

The inversion of (2.5) to find C_1 from C provides a way to simulate these processes, as discussed in Section 2.5.

Whittle (1954, 1956), Heine (1955), and Bartlett (1975) discuss the definition of continuous stochastic processes via differential equations. Stochastic differential equations theory is needed to justify their manipulations, which lead to explanatory models for some of the covariances studied here.

2.3 POISSON AND POINT PROCESSES

Point patterns convey a different sort of spatial information from those processes considered so far. They can be included by defining $Z(\mathbf{x}) = 1$ if there is a point at $\mathbf{x}, 0$ otherwise. This representation is useless, however, for $P(Z(\mathbf{x}) = 1)$ is usually zero and the distribution of the process then contains no information at all. We overcome this problem by indexing the stochastic process not by points but by sets, so $Z(A)$ is the number of points in set A. Every realization of a point process is then a countable set of points, of which a finite number fall within any bounded set. The points can certainly be located by knowing the counts in each rectangle. In fact, it is sufficient to know which rectangles are nonempty.

The basic point process is a Poisson process, defined by either or both of the following properties:

1. The number of points in any set A has a Poisson distribution mean $\Lambda(A)$.

2. Counts in disjoint sets are independent.

Here Λ is a measure giving finite mass to bounded sets, called the *mean measure*. It is often defined by $\Lambda(A) = \int_A \lambda(\mathbf{x})\, d\mathbf{x}$ for some nonnegative bounded function $\lambda(\mathbf{x})$. A homogeneous Poisson process has mean measure $\Lambda(A) = \lambda$ area (A), where λ is a constant known as the *intensity*, the expected number of points per unit area. Note that a homogeneous Poisson process is automatically isotropic.

The Poisson process can also be defined in a more constructive way. Consider the process on a bounded set E. By property 2 it is sufficient to