3   Statistical image modelling

In Figure 32 we see two examples of images obtained by simulation from simple models with independent pixel values. To the left we have a 'pepper-and-salt' pattern corresponding to equal probabilities for black and white. To the right we have a grey-level image from a normal distribution (μ, σ^2) with $\mu = 0.5$, $\sigma = 0.2$ and truncated to the interval $[0, 1]$, that is, if a value less than 0 was generated it was replace by 0 and if a value larger than 1 was generated it was replaced by 1.
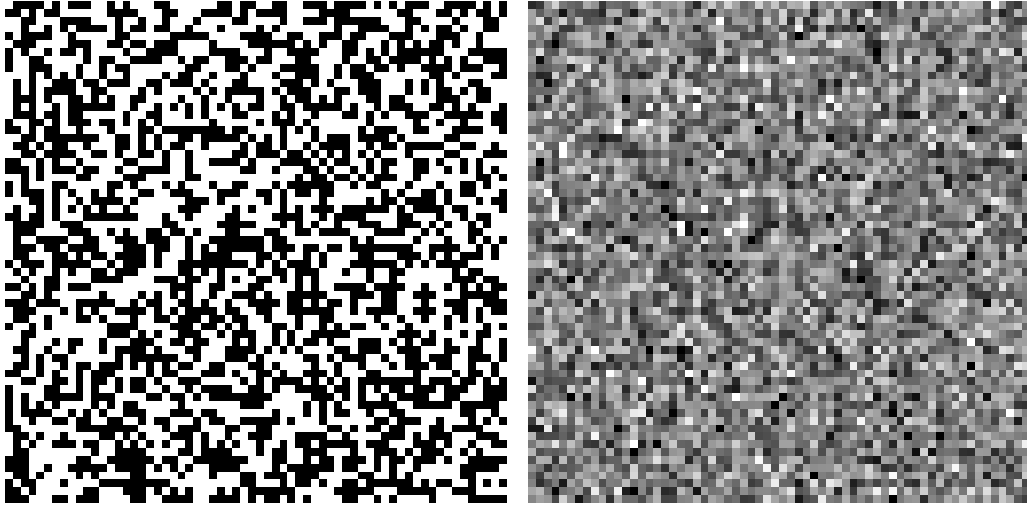

Figure 32: Images of size 64×64 obtained by simulation from models with independent pixel values: to the left a black-and-white image with equal probabilities for the two colours, and to the right a grey-level image with values from a normal distribution with expectation $\mu = 0.5$, a standard deviation $\sigma = 0.2$ and truncated to the interval $[0, 1]$.

In the following sections we will generalize to models with dependence between pixel values. We start by considering Markov random field models defined by a neighbourhood for each pixel and a corresponding conditional distribution for the pixel value given the pixel values in the neighbourhood.

3.1   Markov random field models

Let us regard a random image $X = (X_s, s \in S)$, where S denotes the set of sites (pixel locations). We suppose that to each site $s \in S$ there is defined a set $N_s \subset S$ of neighbour sites such that the following two conditions are satisfied:

(i) $s \notin N_s$,
(ii) $t \in N_s$ if and only if $s \in N_t$.

Two often used neighbourhood systems are shown in Figure 33. To the left we see the system where the site $s = (i, j)$ has the neighbourhood

$$N_s = \{(i-1, j), (i+1, j), (i, j-1), (i, j+1)\}. \tag{47}$$

In the system shown in the right part of the figure there are four additional neighbours so that N_s consists of eight sites.

Figure 33: Two often used neighbourhood systems: to the left the site s has four neighbours and to the right it has eight neighbours.

Suppose that $X = (X_s, s \in S)$ is a set of discrete random variables taking values in the set V. We say that X is a *Markov random field* with respect to the system $(N_s, s \in S)$ of neighbourhoods if

$$\Pr(X_s = x | X_t, t \neq s) = \Pr(X_s = x | X_t, t \in N_s), \quad x \in V, s \in S. \tag{48}$$

This means that if we want to predict the pixel value X_s at s knowing all other pixel values we get the same prediction as when we only know the pixel values in the neighbourhood N_s. This will turn out to be highly useful in an iterative sampling method called Gibbs sampling, which may be used for simulation of a Markov random field.

Neighbourhoods of border sites have to be considered separately. Suppose that the set of sites is
$$S = \{(i, j) : i = 1, \ldots, m, j = 1, \ldots, n\}. \tag{49}$$

One possibility is to use *periodic boundary conditions* which means that sites in the leftmost column are considered as neighbours of sites in the rightmost column, and, similarly, that sites in the top row are considered as neighbours of the bottom row. Specifically, if (47) gives neighbourhoods for non-border sites, we define for $s = (i, n)$ with $1 < i < m$
$$N_s = \{(i - 1, n), (i + 1, n), (i, n - 1), (i, 1)\}, \tag{50}$$

with similar definitions for other border sites. We can think of periodic boundary conditions as corresponding to a folding of S like a torus (a doughnut).

Example 3.1. The Ising model. Let S be given by (49) with periodic boundary conditions. In physical applications to be discussed below we are interested in large values of m and n. Suppose that X_s can take two possible values, -1 and $+1$. Let X_s^+ and X_s^- denote the number of neighbours of s that take positive and negative values, respectively. Thus $X_s^+ + X_s^- = 4$. In the basic two-dimensional model we assume that

$$\Pr(X_s = +1 | X_t, t \in N_s) = \frac{\exp(2\beta(X_s^+ - X_s^-))}{1 + \exp(2\beta(X_s^+ - X_s^-))}. \tag{51}$$

We assume that $\beta > 0$. Note that if $X_s^+ > X_s^-$, that is, if the number of neighbours of s with positive values is larger than the number of neighbours with negative values, then the probability that s shall also have a positive value is greater than $1/2$.

An alternative way of specifying the probability distribution of X is as a Gibbs distribution,

$$\Pr(X = x) = \frac{1}{Z} \exp(\beta \sum_{s \sim t} x_s x_t), \tag{52}$$

where Z is a normalizing constant, which is notoriously difficult to compute in models of this type, and where $s \sim t$ denotes that s and t are neighbours. Thus we sum in the right member of (52) over all pairs (s, t) of sites that are neighbours. In physics the Ising model is used as a model for ferromagnetism and β may be interpreted as inverse temperature. It turns out that for temperature below a critical value, that is for $\beta > \beta_c$, there are long range dependencies and possible phase transitions, that is a clear majority of the X_s-values will either be equal to $+1$ or a clear majority will be equal to -1. But for $\beta < \beta_c$ there are no phase transitions and the value of X_s averaged over large sets of sites is close to zero. A famous computation by Onsager (1944) gives

$$\beta_c = \frac{1}{2} \log(1 + \sqrt{2}) = 0.44069 \tag{53}$$

An up-to-date review of Gibbs distributions and their use in mathematical physics may be found in Georgii, Häggström and Maes (2000). □

3.2 Autonormal random field models

Let us now also regard Markov random field models, where $X_s, s \in S$ are continuous real-valued random variables. The condition (48) needs then a modification to

$$\Pr(X_s \in A | X_t, t \neq s) = \Pr(X_s \in A | X_t, t \in N_s), \quad A \subseteq \mathbb{R}, s \in S, \tag{54}$$

for all considered subsets A of $\mathbb{R}$. We here only consider some simple *autonormal* models where we assume that the conditional distribution of X_s given its neighbours is normal with a constant variance σ^2 and an expectation that is a linear combination of the neighbour values. Specifically, let us consider the neighbourhood system given by the left part of Figure 33 and denote the neighbours of s in the West, North, East and South directions $W(s)$, $N(s)$, $E(s)$, and $S(s)$, and assume that

$$\mathbf{E}(X_s | X_t, t \in N_s) = \mu + \beta_W(X_{W(s)} - \mu) + \beta_N(X_{N(s)} - \mu) + \beta_E(X_{E(s)} - \mu) + \beta_S(X_{S(s)} - \mu). \tag{55}$$

3.3 Simulation of Markov random fields

There are several ways of simulating images from Markov random field models. We will describe one of the most used methods, Gibbs sampling.

In *Gibbs sampling* we visit the sites $s \in S$ in a specified way which may be random or deterministic. An often used random method is to choose successive sites to be visited independently and in a purely random way from the set of all sites. And an often used deterministic visiting scheme for a set of sites such as (49) is to choose sites to be visited row-wise from left to right starting with the first row and proceeding until all sites have

been visited. Such a set of visits is called a sweep. The procedure is iterated a given number of of sweeps.

Example 3.2. The Ising model. Continuation. Consider Gibbs sampling for the Ising model by use of (51). As start configuration we use a purely random configuration as in the left part of Figure 32. For a set of β-values we see in Figure 34 binary images obtained by deterministic row-wise sweeps as described above. The upper two rows correspond to β values under the critical value (53), that is to high temperature, while the two lower rows correspond to low temperature. In the middle row we have β very close to the critical value, actually slightly above.

It may be noted that for large β-values (the two lower rows) the number of iterations used in Figure 34 is far too small to arrive at a stationary distribution for the Markov chain formed by the successive iterations. $\square$
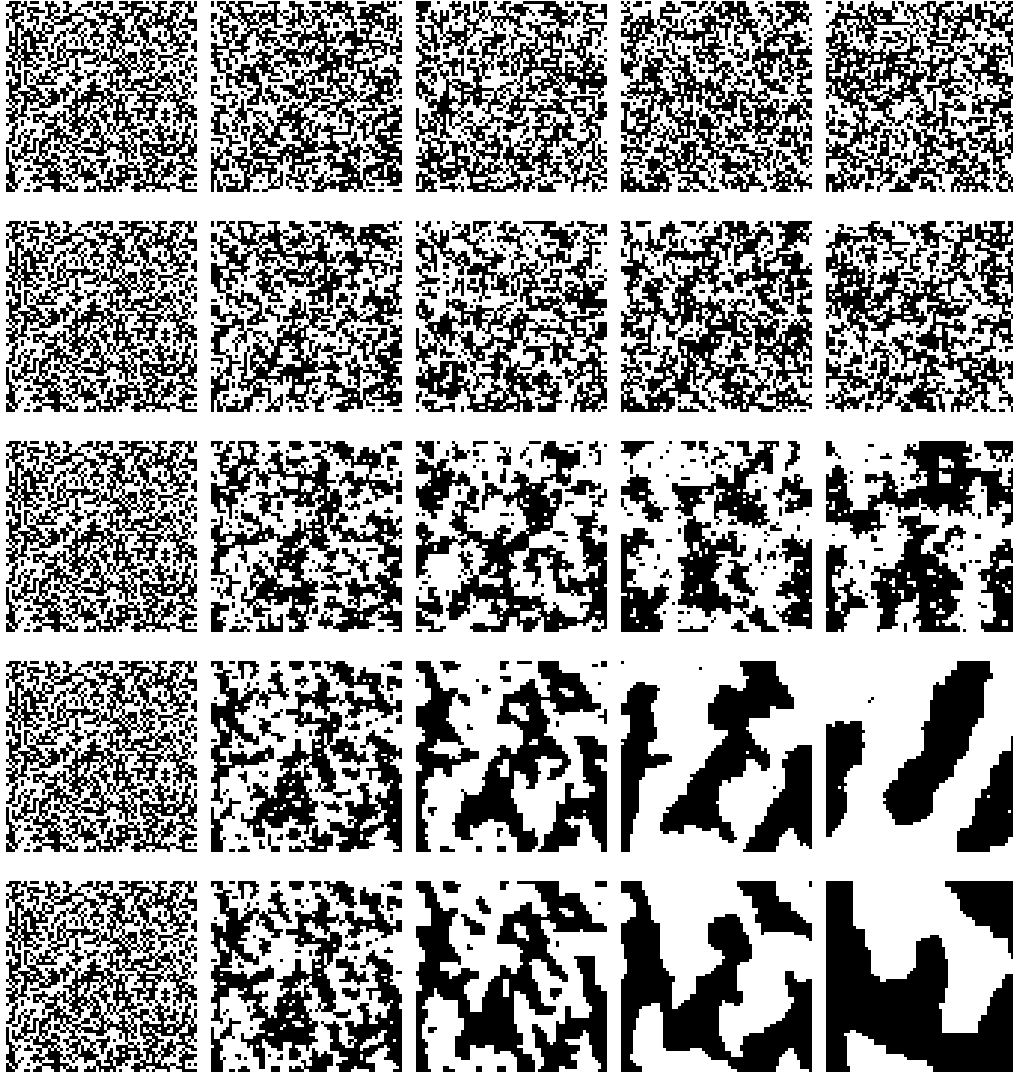
Figure 34: Binary images obtained by simulation for the Ising model with $\beta = 0.11$, 0.22, 0.4407, 0.88 and 1.76 in rows 1 to 5, respectively. In the columns we have to the left a purely random start configuration and then the result after 1 sweep, after 4 sweeps, after 16 sweeps and after 64 sweeps, respectively.
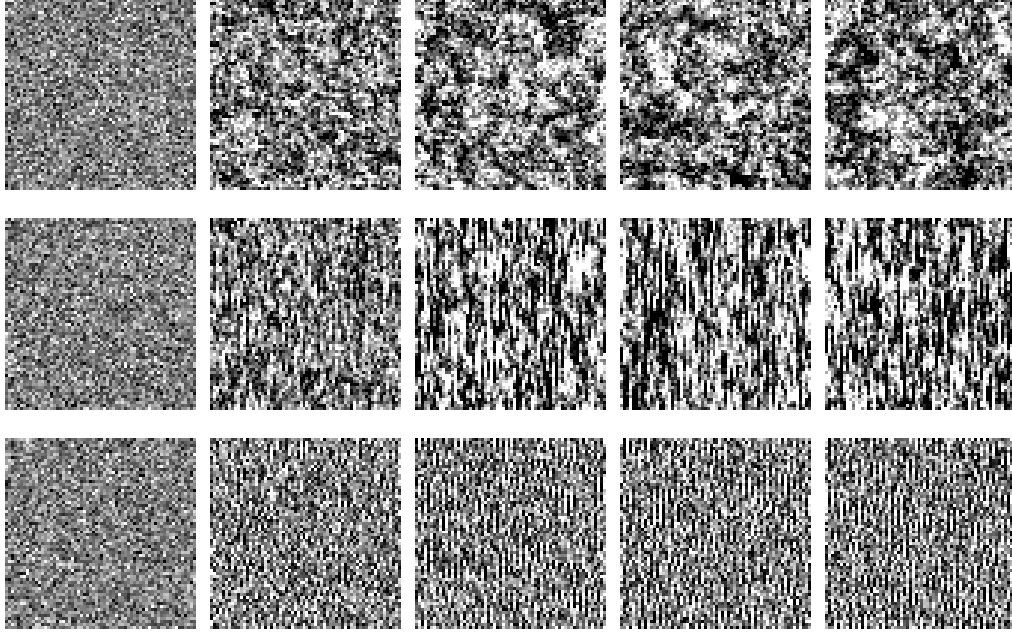
Figure 35: Grey-scale images obtained by simulation for autonormal models. In the columns we have to the left a purely random start configuration and then the result after 1 sweep, after 16 sweeps, after 128 sweeps and after 256 sweeps, respectively. The parameters in (55) are in the upper row $\beta_W = \beta_E = \beta_N = \beta_S = 0.24$, in the second row $\beta_W = \beta_E = 0$ and $\beta_N = \beta_S = 0.48$, and in the third row $\beta_W = \beta_E = -0.24$ and $\beta_N = \beta_S = 0.24$. In all three rows we have $\mu = 0.5$ and the residual standard deviation $\sigma = 0.3$.

Example 3.3. Simulation of an autonormal model. Consider Gibbs sampling for the autonormal model with conditional expectations (55) and constant conditonal variance given the neighbour values. For three sets of parameters we obtain results shown in Figure 35. □

3.4   Bayesian analysis of images

A common approach in Bayesian image analysis, is to assume that we start with a random image X given by a Markov random field. Then we observe a distorted image Y and one basic problem is to reconstruct X from Y. A simple model for the observed image $Y = (Y_s, s \in S)$ is to assume that given X the Y_s-variables are independent and furthermore that the distribution of Y_s only depends on X_s, that is we assume that

$$\Pr(Y = y | X) = \prod_{s \in S} \Pr(Y_s = y_s | X_s). \tag{56}$$

The reconstruction of X from Y is a difficult computional problem, and a series of iterative algorithms have been developed for this type of problems, most of them based on Markov chain Monte Carlo algorithms.

The use of Bayesian models for image reconstruction by use of Markov random field models as priors for the unobserved image X has generally suffered from the problem

that it seems difficult to specify realistic priors for images typically found in applications. A recent interesting approach developed in particular by David Mumford and Song Chun Zhu is based on the following type of models, see for instance Zhu & Mumford (1997) for details and examples of which images that might be generated. Briefly the structure of the model for the prior is a Gibbs distribution, cf. (52) above, with

$$\Pr(X = x) = \frac{1}{Z} \exp(-U(x; \Lambda, F)), \tag{57}$$

where

$$U(x; \Lambda, F) = \sum_{\alpha=1}^{K} \sum_{s \in S} \lambda^{(\alpha)}((F^{(\alpha)} * x)(s)). \tag{58}$$

Here $F = \{F^{(1)}, \ldots, F^{(K)}\}$ is a set of linear filters and $\Lambda = \{\lambda^{(1)}, \ldots, \lambda^{(K)}\}$ is a set of functions, called potential functions, acting on the features extracted by the filter bank F.

3.5 Exercises

Exercise 3.1. Simulate images with independent pixel values as in Figure 32 but with k equi-distributed levels. Choose $k = 3$ and $k = 256$. (Note that the left image in Figure 32 corresponds to $k = 2$.)

Exercise 3.2. Regard the Ising model with negative β-values. (In physics this model is used as a model for anti-ferromagnetism.) Use Gibbs sampling to simulate images as in Figure 34 with β = -0.11, -0.22, -0.44, -0.88 and -1.76. Try also to guess what the images will look like before making the simulations.

Exercise 3.3. Regard an autonormal model with a neighbourhood system as in the right part of Figure 32. Choose suitable notation and write a model corresponding to (55). Use Gibbs sampling to simulate images as in Figure 35 and suggest parameter combinations to obtain different types of random textures.

Exercise 3.4. Show that if the distribution of X is given by (52), then (51) holds. Hint: one can use that

$$\Pr(X_s = +1 | X_t = x_t, t \in N_s) = \frac{\Pr(X_s = +1, X_t = x_t, t \in N_s)}{\Pr(X_s = +1, X_t = x_t, t \in N_s) + \Pr(X_s = -1, X_t = x_t, t \in N_s)}.$$

3.6 Literature on statistical image modelling

Bayesian models for images became popular in the eighties following work by Grenander (1983) and Geman & Geman (1984). Markov chain Monte Carlo methods play an important role in reconstruction of images observed with noise. Important algorithms are simulated annealing, the Metropolis algorithm and Gibbs sampling, which all are examples of randomized algorithms. A simple iterative method, iterated conditional modes, was introduced by Besag (1986). Winkler (1995) gives a thorough treatment of these methods from a mathematical point of view. For an up-to-date introduction to randomized algorithms viewed as Markov chains, see Häggström (2001), including a description of the recently introduced *exact* or *perfect* simulation algorithms.

PART 2 SPATIAL STATISTICS

4 Spatial random processes

Let $X = (X_s, s \in S)$ be a spatial random process, where s is a spatial coordinate. In this chapter S may either be a discrete set, as when X is a digital image, or a continuous set, e.g. a rectangle $S = \{(s_1, s_2) \in \mathbb{R}^2 : a_1 \leq s_1 \leq b_1, a_2 \leq s_2 \leq b_2\}$. In these notes we limit ourselves to spatial processes in two dimensions, but generalizations to d dimensions are fairly straightforward.

A spatial random process may be characterized by its mean value function,

$$m_s = \mathbf{E}X_s \tag{59}$$

and its covariance function

$$C(s, t) = \mathbf{E}(X_s - m_s)(X_t - m_t). \tag{60}$$

A Gaussian random process is completely specified by its mean value and covariance functions. It should, however, be noted that not all functions of two variables are possible covariance functions. In fact, a necessary and sufficient condition that C is a valid covariance functions is that C is symmetric, that is $C(s, t) = C(t, s)$, and that it is positive-definite, that is satisfies

$$\sum_i \sum_j a_i a_j C(s_i, s_j) \geq 0 \tag{61}$$

for all n, $a_1, \ldots, a_n$, and $s_1, \ldots, s_n$. Note that the necessity of the condition (61) follows directly from the fact that

$$\mathbf{E}(\sum_{i=1}^n a_i (X_{s_i} - m_{s_i}))^2 = \sum_i \sum_j a_i a_j C(s_i, s_j). \tag{62}$$

A covariance function $C(s, t)$ is called *stationary* if $C(s, s + t)$ only depends on t, and it is called *isotropic* if it can be written on the form

$$C(s, t) = \sigma^2 \rho(|s - t|), \tag{63}$$

where $|s - t|$ is the Euclidean distance between s and t. Examples of ρ-functions that give valid (positive-definite) covariance functions are

$$\rho(r) = \exp(-ar), \tag{64}$$

$$\rho(r) = \exp(-ar^2) \tag{65}$$

with a positive constant a, and

$$\rho(r) = (1 + r^2/b^2)^{-\beta} \tag{66}$$

with positive constants b and β.

Suppose now that we have a valid covariance function $C(s, t)$, and that $\sigma_0^2 > 0$. Then we can construct a new valid covariance function $C_0(s, t)$ by putting

$$C_0(s, t) = \begin{cases} \sigma_0^2 + C(s, t) & \text{if } s = t \\ C(s, t) & \text{if } s \neq t. \end{cases} \tag{67}$$

The constant σ_0^2 in (67) is sometimes called a *nugget* effect with regard to applications in mining. Another interpretation of the added quantity σ_0^2 in (67) is that it just corresponds to adding independent noise with variance σ_0^2 to all our original observations.
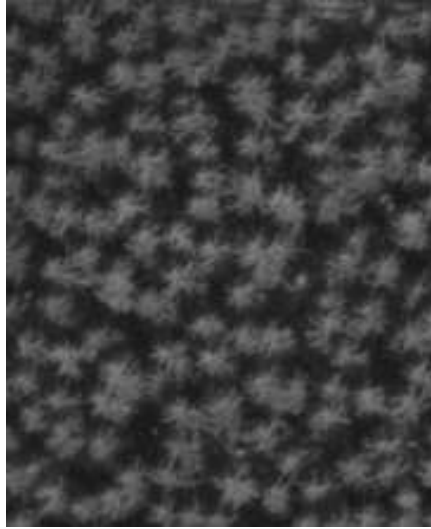

Figure 36: Aerial photograph of Norway spruce trees.

4.1 Prediction (kriging)

Suppose that

$$X_s = m_s + \epsilon_s, \tag{68}$$

where m_s is a slowly varying *trend* function, known or with a known parametric form, and that ϵ_s is a zero-mean random process with a covariance function, also assumed to be either known or of a known parametric form.

Suppose that we have observed X_{s_i}, $i = 1, \ldots, n$, and that we want to predict X_t. In mining this problem is often called kriging after the South African mining engineer D. G. Krige.

Assume first that the functions m and C are known. By regarding $X_s - m_s$ instead of X_s we can transform the problem into one where $m_s = 0$, which we now assume.

Consider a linear predictor

$$\hat{X}_t = \sum_{i=1}^n a_i X_{s_i} = a^T X_{(n)}, \tag{69}$$

where $a = [a_1 \ldots a_n]^T$ and $X_{(n)} = [X_{s_1} \ldots X_{s_n}]^T$ denotes the observations. We choose a to minimize the expected squared error

$$\mathbf{E}(\hat{X}_t - X_t)^2 = a^T G a - 2a^T g_t + \sigma^2(t), \qquad (70)$$

where G is the $n \times n$-matrix with elements $G_{ij} = C(s_i, s_j)$, $g_t^T = [C(s_1, t) \ldots C(s_n, t)]$, and $\sigma^2(t) = C(t, t)$. It is straightforward to show that (70) is minimized for $a = G^{-1} g_t$, and the optimal predictor thus becomes

$$\hat{X}_t = X_{(n)}^T G^{-1} g_t. \qquad (71)$$

The corresponding expected squared error becomes

$$\sigma_{\text{opt}}^2(t) = \sigma^2(t) - g_t^T G^{-1} g_t. \qquad (72)$$

It should be noted that in practice we often only assume that m and C are of known parametric forms but with unknown parameters, and our observations $X_{(n)}$ have to be used to estimate these parameters.

4.2 Exercises

Exercise 4.1. Regard the image in Figure 36. The image $X_s, s \in S$ with $S = \{1, \ldots, 223\} \times \{1, \ldots, 183\}$ is available as `ku94-148Dpart.tif`
(*a*). Assume first that the random function $X_s, s \in S$, has a stationary covariance function that can be written on the form $C(s, s + t) = R(t_1, t_2)$ for $t = (t_1, t_2)$. Estimate the covariance function $R_1(t_1) = R(t_1, 0)$ and the covariance function $R_2(t_2) = R(0, t_2)$ in two orthogonal directions and plot the estimated functions R_1 and R_2 in the same diagram. Does it seem as the covariance function C is identical in the two directions studied?
(*b*). Assume now that the random function $X_s, s \in S$, has an isotropic covariance function. Try to estimate the corresponding ρ-function in (63).
(*c*). Assume that the random function $X_s, s \in S$, is stationary such that the distribution of X_s is the same for all $s \in S$. Try to estimate this distribution, often called the marginal distribution of X.

Exercise 4.2. Delete, say, the three bottom rows in the image in Figure 36. See how well you can reconstruct these three rows by use of prediction according to (71). Assume that the mean value function is a constant, which you estimate from the data. Use an isotropic covariance function with one of the three forms (64) – (66) with parameter(s) adapted to the result of Exercise 4.1(b). To limit computations in the prediction, use as $X_{s_1}, \ldots, X_{s_n}$ a limited set of observations from, say, the last two remaining rows. Note that if you want to use (71) for several g_t it is computationally advantageous to multiply together $X_{(n)}^T$ and G^{-1} before starting to vary g_t.

Exercise 4.3. Consider the three images in the rightmost column of Figure 35. Estimate the covariance function in two orthogonal directions (horisontal and vertical in the figure) as in Exercise 4.1 above. Can any of the three covariance functions be assumed to be isotropic?

Exercise 4.4. Show that if C is a valid covariance function, that is satisfies the inequality (61), then C_0 in (67) is also a valid covariance function.

Exercise 4.5. Verify that $\hat{X}_t$ in (71) minimizes (70) and that (72) gives the corresponding expected squared error.

4.3 Literature on spatial random processes

See Ripley (1981) and Cressie (1991).

5 Point processes. Poisson processes.

Let A be a subset of $\mathbb{R}^2$ with finite and positive area $|A|$. We will consider a random subsets X of A consisting of finitely many points, and call X a point process on A. If $B \subseteq A$ we let $X(B)$ denote the number of points in X that belong to B.

The point process X is said to be *stationary* if the probability distribution of X is invariant under any translation of the sets B where we regard the point process, and we say that X is *isotropic* if the process is stationary and if, additionally, the distribution of X is invariant under any rotation of such sets B.

Consider a stationary point process X on A such that $X(A)$ has finite expectation. One can then show that

$$\mathbf{E}(X(B)) = \lambda|B| \tag{73}$$

for some constant λ which we call the intensity of the point process.

Example 6.1. A Poisson process with constant intensity. A point process X is called a Poisson process with constant intensity $\lambda \geq 0$ on A if $X(B_1)$ and $X(B_2)$ are independent for disjoint subsets B_1 and B_2 of A and if $X(B)$ is Poisson distributed with expectation $\lambda|B|$ for a subset $B \subseteq A$ with area $|B|$, that is

$$\Pr(X(B) = n) = \frac{(\lambda|B|)^n}{n!} \exp(-\lambda|B|). \tag{74}$$

A Poisson process with constant intensity is stationary and isotropic.

A Poisson process on A with intensity λ can be generated in the following way. Let first N be Poisson distributed with expectation $\lambda|A|$. Given that $N = n$, generate $X_1, \ldots, X_n$ as independent and identically distributed variables, each with a uniform distribution over A. Then we let X consist of the points $X_1, \ldots, X_n$, that is $X = \{X_1, \ldots, X_n\}$.

In Figure 37 we see two examples of such generation of a Poisson process in the unit square with the constant intensity $\lambda = 50$. □

Example 6.2. A Poisson process with varying intensity. A point process X is called a Poisson process with intensity function $\lambda(s), s \in A$, if $X(B_1)$ and $X(B_2)$ are independent for disjoint subsets B_1 and B_2 of A and if $X(B)$ is Poisson distributed with expectation $\int_B \lambda(s)\,ds$ for $B \subseteq A$.

A Poisson process with intensity function $\lambda(s), s \in A$, can be generated in the following way. Let first N be Poisson distributed with expectation $\int_A \lambda(s)\,ds$. Given that $N = n$, generate $X_1, \ldots, X_n$ as independent and identically distributed variables, each with a distribution specified by

$$\Pr(X_i \in B) = \frac{\int_B \lambda(s)\,ds}{\int_A \lambda(s)\,ds} \quad \text{for} \quad B \subseteq A. \tag{75}$$
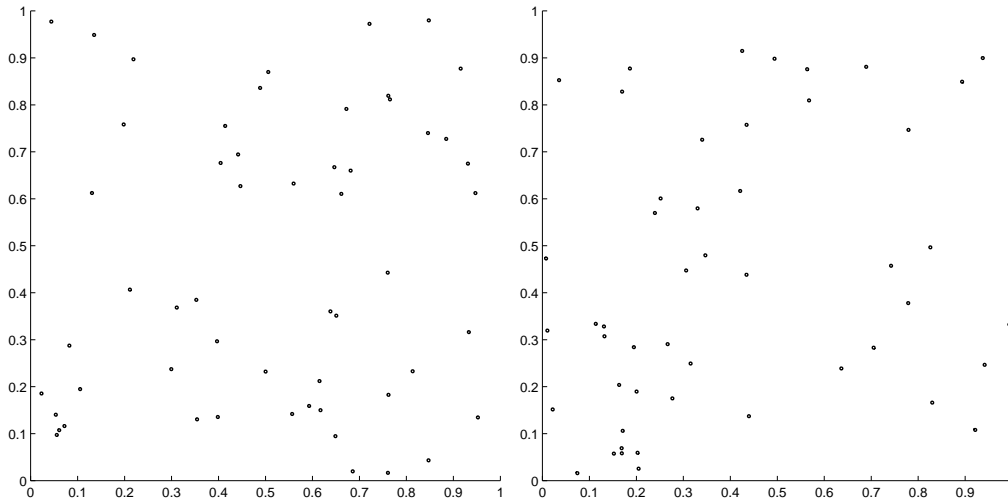
Then we put $X = \{X_1, \ldots, X_n\}$. □

Figure 37: Two examples of Poisson point processes generated in the unit square with $\lambda = 50$. The generated number of points is to the left $N = 55$ and to the right $N = 49$.

5.1 The Neyman-Scott process, a point processes with clustering

Consider a Poisson process with constant intensity λ, and regard the points of this process as mother points. From a mother point we generate daughter points such that the number of daughter points from a mother point are all independent and identically distributed. Further, the two-dimensional vectors from a mother point to a daughter point are all independent and identically distributed. This distribution we call the scattering distribution. The process of daughter points is called a Neyman-Scott process.

Suppose that we want to generate a Neyman-Scott process. If the daughter process is regarded on a set A we need to start by generating the mother point process on a set larger than A, in fact so large that all points from which daughters can get scattered into A are included. With this observation it is straightforward to generate a Neyman-Scott process from the definition above.

5.2 A hard-core inhibition point process

In the cluster point process in the previous section the occurrence of a point typically increases the intensity of points in a neighborhood of this point. We will now describe a point processes with inhibition, suggested 1960 by Matérn, see Matérn (1986), which has the opposite property: the occurrence of a point inhibits other points within a certain distance.

Start by generating a Poisson point process with intensity λ on a bounded set A. To each point $X_i, i = 1, \ldots, N$, we associate a random mark consisting of random variable U_i, which is uniformly distributed on the interval $(0, 1)$ and such that the U_i's are indendent, mutually and of the X_i's. We can think of U_i as the birth time of the point X_i.

Then we thin the X-process by deleting each point X_i for which there exists an older point X_j of the original point process closer than a distance d, that is a point X_j satisfying

$|X_i - X_j| < d$ and $U_j < U_i$. The distance d is called the hard core distance.

5.3 The K-function, a diagnostic tool for detecting clustering and inhibition

Consider an isotropic point process with intensity λ and suppose that x is a point of the point process X. Let $\|y - z\|$ denote the distance between two points y and z in $\mathbb{R}^2$, and define the K-function of X as follows,

$$K(r) = \frac{1}{\lambda}\mathbf{E}(\text{number of further points of } X \text{ within distance } r \text{ from } x) \tag{76}$$

or more precisely

$$K(r) = \frac{1}{\lambda}\mathbf{E}(X(C_x(r))), \tag{77}$$

where $C_x(r) = \{y : 0 < \|y - x\|) \leq r\}$ denotes a circular disk with radius r around x with the point x excluded.

For a stationary Poisson process it follows that

$$K(r) = \pi r^2. \tag{78}$$

Sometimes one chooses to regard $L(r) = (K(r))^{1/2}$ as this function is linear in r for a Poisson process, for which

$$L(r) = \sqrt{\pi}r. \tag{79}$$

5.4 Estimation of characteristics for point processes

Suppose that we have observed a stationary point process X on a set $A \subset \mathbb{R}^2$. The intensity of X we estimate by

$$\hat{\lambda} = \frac{X(A)}{|A|}. \tag{80}$$

It follows generally that for a stationary point process with finite intensity λ the estimator (80) is an unbiased estimator of the intensity, that is, $\mathbf{E}(\hat{\lambda}) = \lambda$.

For a Poisson process we can also compute the variance of the estimator (80). We find

$$\text{var}(\hat{\lambda}) = \frac{\lambda}{|A|}. \tag{81}$$

Let us now regard estimation of the K-function of a point process X observed in the region A. The basic problem in estimating $K(r)$ is that for a point $x \in X$ we want to consider all neighbouring X-points within distance r. But some of these neighbours may be located outside A.

For our first estimator of $K(r)$ we consider pairs of X-points x and y such that $x \in A_r^-$, where A_r^- denotes the subset of A of points with a distance at least r to the border of

A. Let $1\{P\}$ denote the function which is 1 when P is true and zero else. From the definition (76) it follows

$$\sum_{x \in X \cap A_r^-} \sum_{y \in X} 1\{0 < \|y - x\| < r\} \tag{82}$$

is an unbiased estimator of $\lambda^2 |A_r^-| K(r)$. The procedure of restricting to points within a certain distance to the border is sometimes called minus-sampling, and the corresponding estimator of $K(r)$ is therefore called $\hat{K}_{\text{minus}}(r)$, and it is obtained from the unbiased estimator (82) of $\lambda^2 |A_r^-| K(r)$ by replacing λ with its estimator (81). We get

$$\hat{K}_{\text{minus}}(r) = \frac{1}{\hat{\lambda}^2 |A_r^-|} \sum_{x \in X \cap A_r^-} \sum_{y \in X} 1\{0 < \|y - x\| < r\}. \tag{83}$$

Let us now give another estimator of the K-function which utilizes our observations more effectively. Regard two points x and y in the region A and a circle with centre at x and radius $\|y - x\|$. Let $w(x, y)$ denote the proportion of the perimeter of this circle that lies within A. If, for instance A is the unit square, $x = (1/2, 1/2)$ and $y = (1/2, -1/2 + 1/\sqrt(2))$, then a straightforward compution shows that $w(x, y) = 1$ and $w(y, x) = 3/4$. One can now show that

$$\sum_{x \in X} \sum_{y \in X} \frac{1\{0 < \|y - x\| < r\}}{w(x, y)} \tag{84}$$

is an unbiased estimator of $\lambda^2 |A| K(r)$. The corresponding estimator of the K-function is

$$\hat{K}(r) = \frac{1}{\hat{\lambda}^2 |A|} \sum_{x \in X} \sum_{y \in X} \frac{1\{0 < \|y - x\| < r\}}{w(x, y)}. \tag{85}$$

There is one minor restriction in the use of (85) which means that we cannot consider r so large that $w(x, y)$ become close to zero. In practice this is not important as we are usually interested in reasonably small r-values. Thus, for observations in the unit square an upper limit for r is $\sqrt(1/2)$.

5.5 Exercises

Exercise 6.1. Generate a Poisson process on the unit square $[0, 1] \times [0, 1] \subset \mathbb{R}^2$ with constant intensity 100. Show the result in a figure.

Exercise 6.2. Generate a Poisson process on the unit square $A = [0, 1] \times [0, 1]$ with varying intensity $\lambda(s) = 200 s_1$, $s = (s_1, s_2) \in A$. Show the result in a figure.

Exercise 6.3. Generate a Neyman-Scott process on the unit square $A = [0, 1] \times [0, 1] \subset \mathbb{R}^2$ in the following way. Assume that (i) the mother process is a Poisson process with constant intensity 50, (ii) each mother point generates two daughter points, and (iii) the scattering distribution (from mother to daughter) is an isotropic two-dimensional

normal distribution with zero means and standard deviation 0.01 in two orthogonal directions. (Truncate here the normal distributions at, say, plus and minus three standard deviations.) Show the result in a figure.

Exercise 6.4. Compute the expected distance from one mother point to its nearest neighbour mother point for the point process of the previous exercise, and also the expected distance between the two daughter points from one mother point (disregard in these computations edge effects, that is the limited size of the set A). Instead of the two expected distances you may choose to compute root-mean square distances, that is the square root of the expected squared distances, which are a bit easier to compute.

Exercise 6.5. Generate a hard core Matérn point process on the unit square $[0, 1] \times [0, 1] \subset \mathbb{R}^2$ with $\lambda = 100$ and $d = 0.1$. Show the result in a figure.

Exercise 6.6. Estimate the intensity and the K-function for the point processes considered in (a) Exercise 6.1, (b) Exercise 6.3, and (c) Exercise 6.5. Compare the three K-function estimates.

5.6 Extensions and literature on point processes

A highly readable introduction to spatial point processes is given in Diggle (1983). In particular, one finds here methods for testing complete spatial randomness, which means testing whether an observed point pattern is consistent with an hypothesis that it is generated from a Poisson process with constant intensity. One of these test methods consists of first estimating the K-function for the observations and then to find confidence limits for this function by simulating a number of Poisson processes with the same intensity and computing upper and lower envelopes for the corresponding K-function estimates.

Stoyan et al. (1995) gives a thorough coverage of stochastic geometry, including the theory of point processes. Many different methods for taking care of edge effects for estimates of functions like the K-function are described and compared.

6 Marked point processes and patterns of randomly placed objects

Point processes are natural building blocks for more complicated spatial processes such as patterns of random objects, for instance disks of random sizes. Let us consider a point process X and associate with each point X_i of X a random mark M_i, which could be the radius of a disk centered at X_i. By letting the mark be a vector with several components we could model more complex objects.

For the 2D gel electrophoresis images in Figures 9 and 10 we could associate with a protein at position $X_i = [X_{1i} X_{2i}]^T$ the mark $M_i = (S_i, C_i)$, where S_i is the expression level of the corresponding protein and C_i could describe the shape of the spot at X_i. A simple model would be to assume that the spot shape is a two-dimensional normal distribution with $2{\times}2$ covariance matrix C_i. The observed pixel gray level Y_x at a pixel with location x could then modeled by

$$Y_x = \sum_i S_i f(x, X_i, C_i) + \epsilon_x, \qquad (86)$$

where ϵ_x is the observation noise at pixel x and

$$f(x, X_i, C_i) = \frac{1}{2\pi (\det C_i)^{1/2}} \exp(-\frac{1}{2}(x - X_i)^T C_i^{-1}(x - X_i)). \qquad (87)$$

For the diffusing particles in Figures 13 and 14 we could consider a model

$$Y_x = \sum_i f(x, X_i, z_i) + \epsilon_x, \qquad (88)$$

where again ϵ_x is the observation noise at pixel x, but the mark consists of the scalar z_i representing the vertical position of a particle relative to the focal plain. The function f could be assumed to be the same for all particles but needs to be estimated from data or by applying optical theory for the light scattering of the diffusing objects.

Similar models could be considered for the aerial photographs in Figures 2 and 4 where we could assume a similar shape for trees in a given view. This shape function could then be estimated from data combined with a simulation model based on the geometry and illumination of the trees from the sun (Larsen & Rudemo, 1998).

A specific problem is interaction between objects that overlap partly. In 2D gel electrophoresis it is natural to assume an additive model as in (86), but in the aerial photographs, and particularly for the diffusing particles, objects may occlude each other and then an additive model may be an untenable approximation. In some applications such as the one shown in Figure 38 objects do not overlap.

Let us regard models for random placed disks. For disks of constant size we can then use the inhibition point process of Section 5.2 by placing disks of diameter d centered at the points of the thinned point process. In the following section we shall regard two modifications of this model.
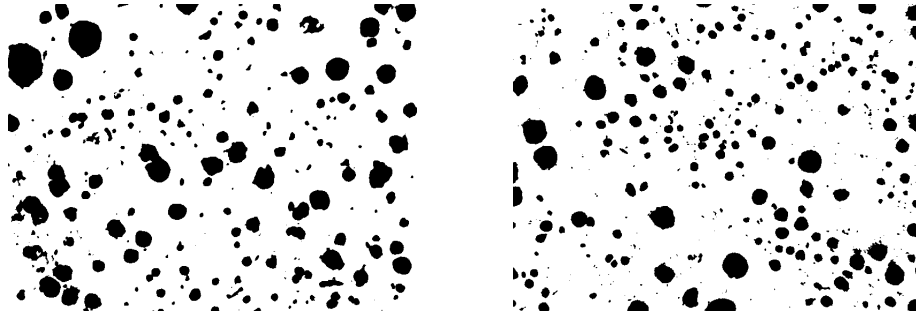
Figure 38: Binary images of two cuts in cast iron showing approximately disk-shaped defects. Data from Beretta (2000) and Månsson and Rudemo (2002).

6.1 Two processes of varying-sized disks

Let us regard marked point processes constructed in two steps as follows.

In the first step we generate a Poisson point process with constant intensity λ in the plane, and to each point in this point process we generate identically distributed radii with a *proposal* distribution function F_{pr}. The radii are independent mutually and of the point process.

In the second step we thin the generated point process by letting all pairs of points whose associated disks intersect 'compete'. A point is kept if it has higher weight in all pairwise comparisons, where the, possibly random, weights are assigned to the points according to two different approaches:

1) *Pairwise assignment of weights*: For each comparison, weights are assigned to the involved pair of points, and assignments are independent both within and between pairs.

2) *Global assignment of weights*: Weights are assigned once and for all to all points, and assignments to different points are independent. These weights are then used in all comparisons.

In both cases the weight of a point may depend on the associated radius. (When the weights are constant or deterministic functions of the radii, the two approaches coincide.)

It is possible to compute both the intensity of the point process after thinning and the radius distribution function after thinning. Details are given in Månsson and Rudemo (2002). Let us here only show a simulation example of disks before and after thinning with three different thinning procedure, see Figure 39.
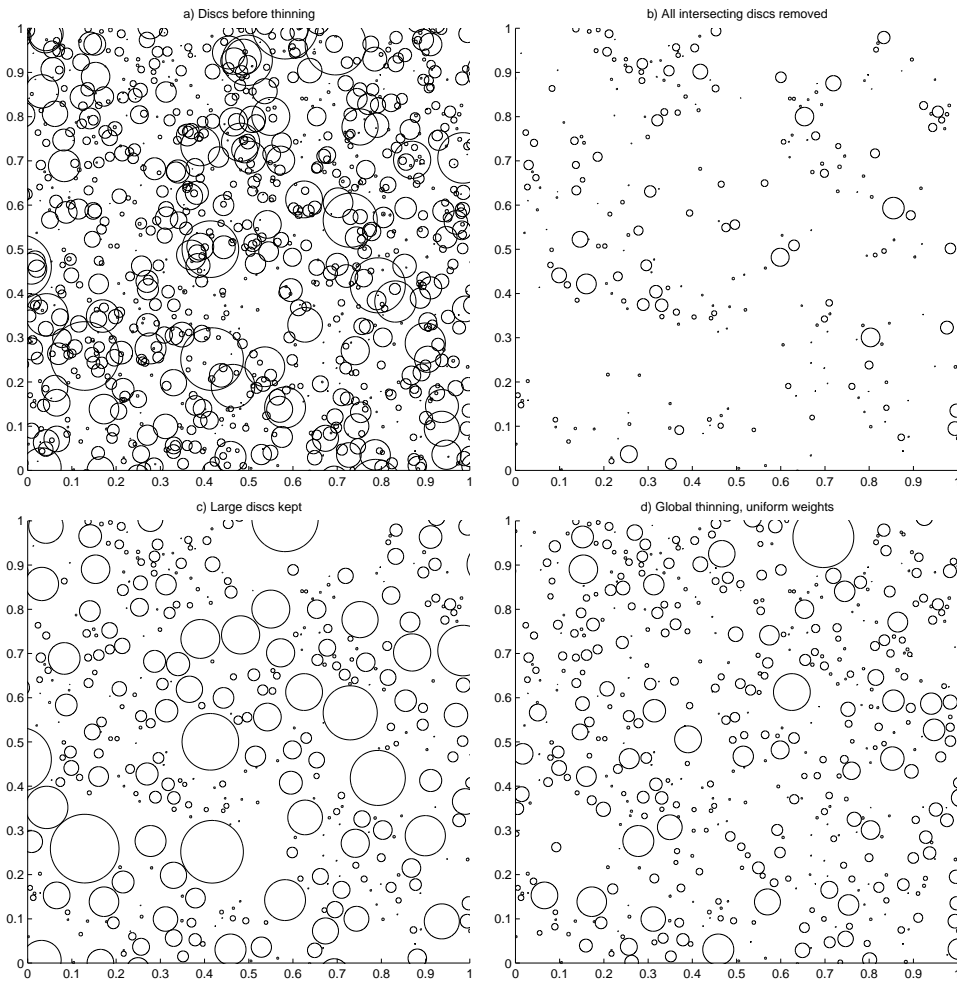
Figure 39: Simulation of a disk process before and after three different thinning procedures. In the first step a Poisson process with intensity 1000 in the unit square is generated with exponentially distributed disk radii with expectation 0.01.