Once a point pattern has been successfully analysed by the exploratory methods and the analysis has suggested a class of suitable models, in the last step these results are synthesised, i.e. a model is fitted to the data. This includes ‘model criticism’, consisting of various forms of model evaluation – in particular, formal goodness-of-fit testing. This final chapter presents methodology that may be used for this purpose. First, fundamental methods for parameter estimation are covered, the maximum likelihood method in an approximate form and the method of moments. Second, the use of simulation tests for testing distributional hypotheses is discussed, which includes tests of the hypotheses of independent and geostatistical marking.

7.1 Choice of model

One of the most important steps in the statistical analysis of a point pattern or a series of patterns is constructing and fitting a model. In some ways this is the synthesis of the knowledge gained from a thorough exploratory analysis. Due to the strong link between the methods for model fitting and the material covered
This chapter assumes that the reader is familiar with such methods, in particular with exploratory data analysis and point process models.

This section briefly discusses those properties that make a good model and then provides some advice on a suitable strategy for identifying a suitable model for a data set.

In most cases, the aim of fitting a model to empirical data is to gain an understanding of the pattern(s). Of particular interest are the nature and spatial extent of the interaction or correlation between the points, the influence of the marks and the relation of covariates to the (marked) point pattern. A good statistical model is one that is easy to understand and interpret. In other words, a good model depends on only a small number of parameters but is still complex enough to suitably reflect those aspects in the data that are of scientific interest. It is important that the model is simple enough that both estimation and simulation are still feasible – models that are too complex easily result in algorithms that are prohibitively expensive to run.

However, it is not obvious how to identify a suitable model for a specific data set. Usually the data will originally have been collected with the aim of proving or disproving some hypotheses about the spatial distributions of the objects that are represented by the points, i.e. in the analysis of a spatial pattern the model is directly linked to some scientific question. In other cases, however, the aim is simply to describe the spatial behaviour in the pattern without reference to underlying mechanisms.

In most applications, some a priori knowledge is available, which often supports the choice of model. For example, seed dispersal may suggest that a cluster process model is suitable for a data set describing a plant pattern. However, further analysis may indicate that the modelling approach has to be refined. For example, airborne seeds disperse over large distances. Seeds within the observation window might stem from mother plants that are far away from the observation window but also from mothers close to or within the window. Hence, a superposition with a Poisson process to model global dispersal in addition to local dispersal may be more suitable.

Sometimes the structure of the model and the structure of the underlying process that formed the pattern are not directly linked. These models are fitted to merely describe or explain the statistical fluctuations and to use this information for the simulation of similar point patterns or to determine the precision of statistical estimates – in particular, of the intensity. However, these models do not explain any true natural mechanisms underlying the data. An example of this is the Matérn cluster process, which may be used to model clustering (around randomly located centres) or, alternatively, environmental heterogeneity. The variation of local point density can be suitably modelled with this process, but the structure of the model deviates strongly from the true structure in any realistic pattern.

As indicated in Section 2.7, the first step in the analysis of a spatial point pattern should be a test of the CSR hypothesis. Accepting the CSR hypothesis may be a disappointing result since it rules out any interesting correlations in the
pattern. On the other hand, modelling and further scientific calculations are clearly simplified and can be based on the elegant Poisson process model.

Only once the CSR hypothesis has been rejected, a suitable class of models should be sought. Often, visual inspection and experience from the CSR test aid decision-making on more fundamental questions such as whether the pattern is finite or not finite, stationary or non-stationary, and clustered or regular.

In the exploratory analysis below, the methods discussed in Chapters 3–5 are applied using some of the summary characteristics described there. Quite often, this may be sufficient since these characteristics reveal details on the range and strength of correlation in the pattern. If one wishes to gain a deeper understanding of the pattern, one should look for a suitable model based on the results of the exploratory analysis.

If the data can be considered stationary, the pair correlation function $g(r)$ is probably the best exploratory tool. The detailed discussion of this function and its interpretation in Section 4.3.4, along with the description of the models in Chapter 6, can help the reader to identify appropriate classes of models. Recall, for example, that large values of $g(r)$ for small $r$ indicate clustering and small values of $g(r)$ for small $r$ indicate regularity. This simple information on clustering or regularity at different scales, combined with a priori knowledge, may already indicate which classes of models are more suitable than others. As a general rule, in the first instance every modelling approach should start with basic classical models such as Cox, Neyman–Scott or Gibbs processes, since these can easily be fitted and simulated. Cox and Neyman–Scott processes are suitable for modelling fluctuating point density, while Gibbs processes may be used to describe interesting repulsive interactions between the points. If these models turn out not to be suitable, they can be modified by relaxing some of the underlying Poisson process assumptions, such as by choosing a non-Poisson parent process for a cluster process or by including hard cores, i.e. minimum inter-point distances. Any of these modifications should still be simple enough that the model can be simulated.

Once a suitable model has been identified, the next step involves the estimation of model parameters. These model parameters, such as $\lambda$ (for the Poisson process), $\lambda$, $\mu$ and $R$ (for the Matérn cluster process) or $\beta$ and $r_{\text{max}}$ (for the Strauss process), are of course of a different nature for different models. However, to simplify the notation and language, the general symbol $\theta$ is used throughout this chapter to represent any of these parameters. Note that parameter estimation is based on the same principles as the estimation of $\mu$ and $\sigma^2$ in classical statistics, even if it appears to be more complicated in the context of point processes.

Researchers often consider the analysis as finished once a model has been fitted to the data, supported by good agreement of some graphics of theoretical and empirical summary characteristics. However, a goodness-of-fit test should be performed to formally assess the suitability of a model. It is not difficult to do so if it is possible to simulate from the fitted model. The goodness-of-fit test may on the one hand confirm that the model is suitable, but on the other hand it may help to identify and eliminate any weaknesses of the model.
The following sections describe approaches to parameter estimation and discuss model tests, assuming that a model has been identified for a particular data set. All model tests considered here are simulation tests, which are very common in point processes statistics. Note that, unfortunately, classical tests such as the $t$-test or $\chi^2$ goodness-of-fit test cannot be applied here. Other approaches to assessing the model fit use, for example, residuals; see Section 4.6.5 and Baddeley et al. (2005). Mathematically more sophisticated tests (e.g. likelihood-ratio tests) can be found in Geyer (1999) and Møller and Waagepetersen (2004).

### 7.2 Parameter estimation

A number of different approaches to parameter estimation have been used in the context of spatial point processes. These are based on the same ideas as in classical statistics. Which estimation method is used for a specific data set depends on the model and the nature of the parameters and is to a certain extent also a matter of taste.

As far as the performance of the estimators is concerned, these are usually required to be unbiased, and to have a small mean squared error (mse). Another requirement is that the estimators are consistent, i.e. that their increases with increasing window size. This is the case for many parameter estimators of stationary point processes if these are ergodic.

#### 7.2.1 Maximum likelihood method

Maximum likelihood methods are widely used in classical statistics, and many statisticians believe that they should also be preferred in point process statistics. Indeed, famous theorems by Fisher, Rao and Cramér show that in classical statistics maximum likelihood estimators represent the ‘hard currency’ among the estimators as they are efficient, sufficient and consistent. Those readers familiar with estimation methods in classical statistics may know that maximum likelihood method techniques can only be applied if the likelihood function – describing the probability of observing the data given the model – is known. This probability is maximised (with fixed data and variable parameters), yielding parameter estimators that best fit the data. However, often and particularly for stationary point processes, it is extremely difficult, even impossible, to find the likelihood function. As a result, the maximum likelihood method can only be applied to specific classes of models. These are Poisson processes (pp. 80 and 121), Cox processes (with approximative likelihoods; see Møller and Waagepetersen, 2004, and below) and finite Gibbs processes (p. 161). In addition to the models’ flexibility and interpretability in applications, this might account partly for the popularity of finite Gibbs (and Markov) point processes in the statistical point process literature.

It is possible to apply the maximum likelihood method to spatial point patterns where the likelihood function is not known explicitly by (heuristically) approximating the likelihood function. This can be done in many ways.
An interesting example of this approach is a method developed by Tanaka et al. (2008) for stationary point processes. It is based on the model pair correlation function $g(r)$ and works well if $g(r)$ is known explicitly and contains the parameters of interest. The idea is to analyse the finite point process $N_{\delta}$ in addition to the original point process $N$. $N_{\delta}$ consists of all the difference points

$$\delta = x - y \quad \text{for } x \neq y,$$

where $x$ and $y$ are the points of the process $N$ in the window $W$, which is assumed to be convex. If the original pattern has $n$ points then the difference pattern $N_{\delta}$ has $n(n - 1)$ points. $N_{\delta}$ is central-symmetric in $W \oplus \bar{W} = \{ z : z = x - y, x \in W, y \in W \}$ since it contains $x - y$ as well as $y - x$. Figure 7.1 shows an example of a small point pattern and the corresponding pattern of difference points.

For the intensity function of the difference pattern a formula can be given in which $\lambda$ and $g(r)$ appear. Consider the mean number $\Lambda_{\delta}(r)$ of points of $N_{\delta}$ in the disc $b(o, r)$. The derivative $\Lambda_{\delta}'(r) = \lambda_{\delta}(r)$ can be calculated as

$$\lambda_{\delta}(r) = \lambda^2 \overline{\gamma}_W(r) g(r) \quad \text{for } r \geq 0,$$

where $\overline{\gamma}_W(r)$ is the set covariance of the window $W$.

Tanaka et al. (2008) then assume that $N_{\delta}$ can be approximated well by an inhomogeneous Poisson process with intensity function $\lambda_{\delta}(r, \theta)$. This function depends on $\theta$ through $g(r)$, and the authors use the maximum likelihood approach for inhomogeneous Poisson processes to estimate $\theta$. Using (3.4.4) and polar coordinates, the corresponding log-likelihood function becomes

$$\ln L(\theta) = \sum_{x, y \in N \cap W} \ln \left( \lambda_{\delta}(\|x - y\|, \theta) \right) - \int_0^R \lambda_{\delta}(r, \theta) db \Gamma(r^{d-1}) dr$$

Figure 7.1 A pattern of 10 points and the corresponding pattern of difference points.
for $R = \min\{r : \mathcal{F}_w(r) = 0\}$. Tanaka et al. (2008) simplify the likelihood function further and use numerical methods to compute the integral and find the maximum. They show that the method works well for Neyman–Scott cluster processes, for which the exact maximum likelihood method usually cannot be applied. The same approach can also be used for other Cox processes.

Other approximate likelihood methods include the pseudo-likelihood method (see Besag, 1975, 1978; Møller and Waagepetersen, 2004, p. 171) and its improvement by Huang and Ogata (1999); see also Section 3.6.

In classical statistics as well as in point process statistics often several potential models with different parameters and numbers of parameters may initially be considered appropriate for a specific data set. A common approach to the comparison of different models where parameter estimation has been done using maximum likelihood methods is the Akaike information criterion (Akaike, 1974). The $AIC$ is defined as

$$AIC = -2 \cdot \ln(L(\hat{\theta})) + 2k,$$

where $L(\hat{\theta})$ is the likelihood function evaluated at the maximum likelihood estimator $\hat{\theta}$, and $k$ is the number of independently fitted parameters. Tanaka et al. (2008) use the AIC for model comparison.

### 7.2.2 Method of moments

The method of moments has not been very popular in classical statistics in general as it lacks some of the desirable properties of the maximum likelihood method. Nevertheless, it often provides very good estimators that are unbiased or ratio-unbiased and consistent. It has many applications in the context of spatial point processes, especially when the likelihood function is not available.

Note that the term ‘method of moments’ is used here somewhat loosely since the approaches described here are all based on the same general idea but this idea is applied to moments or moment-measure-related characteristics as well as to other summary characteristics that are not moment-related. The general idea is to find parameters that minimise the difference between a ‘suitable’ summary characteristic $S$ that is known analytically (or from simulations) and the summary characteristic $\hat{S}$ as estimated from the data. It is important that $S$ depends on the unknown parameter $\theta$; to emphasise this dependence, the characteristic is denoted by $S_\theta$. The methods discussed in Chapter 3, 4 and 5 are then applied to the data to yield an empirical $\hat{S}$. The value $\theta$ for which $S_\theta$ and $\hat{S}$ are ‘as similar as possible’ is used as an estimator. The term ‘as similar as possible’ means here ‘similar in the sense of a specific approximation method’, such as the least-squares approach.

As indicated by the vague expression ‘suitable’, different summary characteristics $S$ may be used. Which of these is deemed suitable depends on the context. A first criterion for the choice of summary characteristics is often whether a formula for
$S_\theta$ is known. However, if this is not the case simulation approaches may be used instead. Another criterion should be that $S_\theta$ is sensitive to variation in $\theta$.

In its simplest form, the method of moments is applied to numerical summary characteristics. If, for example, the intensity $\lambda$ is a model parameter, $\hat{\lambda}$ is its estimator as in (4.2.10). Similarly, if the hard-core distance $r_0$ is a model parameter, as in Gibbs hard-core or Matérn hard-core processes, the estimator is simply the minimum inter-point distance in the sample pattern, see (4.2.46).

In many other examples the method of moments is based on a functional summary characteristic, i.e. the $S$ above is a function $S(r)$. Sometimes it is enough simply to plot $\hat{S}(r)$ and identify specific points, e.g. cusp points or points where $S_\theta(r)$ becomes constant. This approach may be used to find estimators of particular distances such as $r_{\text{cor}}$. In the context of Neyman–Scott processes the radius of the clusters may be found in this way. The parameters $r_1$ and $\beta$ of a Strauss process may be estimated using the cusp-point method.

Typically, however, the method of moments for functional summary characteristics usually applies a least-squares approach that is often referred to as the minimum contrast method. This is based on the simple idea of minimising

$$\Delta(\theta) = \int_{s_1}^{s_2} |\hat{S}(r) - S_\theta(r)|^\beta \, dr$$

with respect to $\theta$. The value of $\theta$ that yields the minimum is the estimator $\hat{\theta}$.

Here, the parameters $s_1$, $s_2$ and $\beta$ as well as the summary characteristic $S(r)$ can in principle be chosen arbitrarily; often $\beta = 2$ is used. In the literature most applications use either

$$S(r) = L(r) \quad \text{or} \quad S(r) = H_s(r).$$

For these summary characteristics, the integral limits can be chosen as $s_1 = 0$ and $s_2 = s$ where $s$ is some suitable maximum distance as on p. 95. In Stoyan and Stoyan (1996), Möller and Waagepetersen (2004, p. 183) and Taylor et al. (2001) the pair correlation function $g(r)$ is used as the summary characteristic $S(r)$. The reader is advised to follow this example in order to avoid dependence among the residuals in the sum of squares. However, because of the well-known inaccuracy of $\hat{g}(r)$ for small $r$, a positive value for $s_1$ should be chosen, somewhere in the region of $\hat{m}_{D}$, the estimated mean nearest-neighbour distance, or of $\hat{r}_0$, the estimated hard-core distance. For finite point processes $L_{\text{fin}}(r)$ (see p. 131) may be a good choice for $S(r)$.

In practice, the integral in (7.2.1) is replaced by a sum,

$$\int_{s_1}^{s_2} |\hat{S}(r) - S_\theta(r)|^\beta \, dr \approx \sum_{i=0}^{k} |\hat{S}(Q_i) - S_\theta(Q_i)|^\beta$$

(7.2.2)
with \( q_0 = s_1, q_i = r_1 + i\delta, q_k = s_2 \) and \( \delta = \frac{s_2 - s_1}{k} \) for some integer \( k \).

Jolivet (1986) and Heinrich (1992, 1993) investigate the statistical properties of \( \hat{\theta} \) for special cases. In reasonable cases the minimum contrast estimator is ‘consistent’ in the sense of probability theory, i.e. it converges to the true value as the size of the window \( W \) increases until it is the whole space.

### 7.2.3 Trial-and-error estimation

If all else fails, parameters can be estimated by trial and error, as demonstrated by Ripley (1977), even if there is a possibility that the results it yields are of dubious quality. The idea is to try several estimates \( \hat{\theta}_1, \hat{\theta}_2, \ldots \), and use a goodness-of-fit test from Section 7.4 with each of these. The first estimate \( \hat{\theta}_i \) for which the hypothesis that the model containing \( \hat{\theta}_i \) fits the data is accepted, is used as the estimate. An extension of this method may even be used to construct confidence intervals for \( \theta \), using the well-known duality between significance tests and confidence intervals: the set of all \( \hat{\theta}_i \) for which the test leads to acceptance determines such an interval.

Clearly, the dimension of \( \theta \), i.e. the number of parameters, should be small.

**Example 7.1. Phlebocarya pattern: fitting a cluster process model**

After a successful statistical analysis of the pattern of 207 *Phlebocarya filifolia* plants in a 22 \( \times \) 22 m square (see Figure 1.4) in Chapters 4 and 6, a cluster process model is now fitted to the data. Indeed, the pattern looks like a sample from a cluster process and the pair correlation function in Figure 4.19 also suggests this. The fact that in Example 6.2 a Cox process could be fitted to the pattern may trigger the idea of also trying a simpler cluster model. This model is rather primitive and likely to be too simplistic. It is discussed here mainly as an illustration. If such a model could really be fitted to the data this may lead to a clearer statement on the range of correlation in the pattern. Figure 4.19 does not give a clear answer on the question of the range of correlation: it may be that the range is smaller than 1 m and that the values of \( g(r) \) larger than 1 for \( r > 1 \) m have to be considered irrelevant or that the range is large, up to 10 m, but for distances between 1 and 10 m the correlations are very weak.

In this example, a Matérn cluster process is used as the cluster process model, which depends on the parameters cluster radius \( R \), mean number of points \( \bar{c} \) per cluster, and intensity \( \lambda \), as explained on p. 376. Two estimation methods are applied: the minimum contrast method with \( S(r) = g(r) \), \( s_1 = 0 \) and \( s_2 = 3.0 \), yielding the estimates \( \hat{R} = 0.12 \) m, \( \hat{c} = 0.078 \); and approximate maximum likelihood method, yielding \( \hat{R} = 0.19 \) m, \( \hat{c} = 0.134 \). For both models the intensity estimate is \( \hat{\lambda} = 0.428 \) m\(^{-2}\).

Although the methods yield similar results, in the following the minimum contrast estimates are preferred as their behaviour in goodness-of fit tests is a little better. Figure 7.2 shows the empirical pair correlation function and envelopes from 99 simulations of the fitted model, which show good agreement. This was to be expected since a second-order characteristic was used for parameter estimation. So
using a similar characteristic to assess the fit of the model is likely to indicate a (potentially spurious) fit and should not be considered a useful goodness-of-fit test, but see Example 7.2.

Note that for this example the estimates of $R$ deviate greatly from values which visual inspection of the empirical pair correlation function in Figure 4.19 may suggest.

In many applications, different estimation methods are combined: different components of the parameter vector $\theta$ are estimated by different methods, often in an iterative way. Consider, for example, a pair potential parameter $\theta = (\theta_1, \theta_2)$ where $\theta_1$ is the range of interaction and $\theta_2$ determines the strength of interaction. Then $\theta_1$ may be estimated by the cusp-point method and $\theta_2$ by the partial maximum likelihood method given $\theta_1$. Another example is the profile likelihood method explained on p. 167 in Section 3.6.

### 7.3 Variance estimation by bootstrap

In all areas of statistics, parameter estimates are difficult to interpret without additional information on their accuracy. Accuracy is usually characterised by mse, estimation variance or confidence interval, but in point process statistics it is rather difficult to calculate these. Simulation and resampling methods are often used in practice to overcome this problem. A very popular method is the bootstrap; see Davison and Hinkley (1997) and Manly (2004) for details.

The bootstrap approach is based on empirical data, but in some cases simulated data are also used for the same purpose. The method typically does not make any distributional assumptions. In classical statistics, a large number of new artificial
samples are resampled from the data $x_1, x_2, \ldots$ with replacement. These samples are analysed separately and the fluctuations of the estimates are considered and used in order to characterise the variability of the estimator of interest.

The analogue of the artificial data in point process statistics are independent point pattern samples, rather than samples that consist of points from one original pattern as this is a misuse of the bootstrap idea; see the discussion in Martínez and Saar (2002, p. 91) and Snethlage (1999). Methods discussed in the point process literature, where new patterns are constructed by ‘block resampling’, i.e. random subsamples of the pattern in $W$ are taken and combined to form new patterns within $W$, may be useful for data on the real axis, in time series analysis. However, for spatial patterns these are unsuitable as they ignore existing correlations along the edges of the subwindows and generate configurations that do not exist or may even be impossible in the original pattern.

Therefore, in point process statistics bootstrapping can only be successfully used in the context of replicated patterns, where a series of point patterns are given in windows $W_1, \ldots, W_k$, which are of equal size and shape and $k$ is not very small; see Diggle et al. (2000) and Schladitz et al. (2003). These windows may be subwindows of a large window or separate windows which observe the same spatial phenomenon. It is important that the patterns in the $W_i$ are (practically) independent.

This section provides a sketch of bootstrapping for the estimation of the variances of estimators of the intensity $\lambda$ and the pair correlation function $g(r)$ of replicated stationary point processes.

The intensity estimator (4.2.10) yields the values $\hat{\lambda}_i$ for $i = 1, 2, \ldots, k$ for the windows $W_i$. Equation (4.7.1) then yields the global value $\bar{\lambda}$,

$$\bar{\lambda} = \frac{1}{k} \sum_{i=1}^{k} \hat{\lambda}_i.$$  

The aim of bootstrapping is to evaluate the precision of $\bar{\lambda}$ based on the values $\hat{\lambda}_i$. A new sample of $k \hat{\lambda}_i$-values is generated by randomly resampling from the $\hat{\lambda}_i$ with replacement. If $k = 5$, it may for example consist of the values $\hat{\lambda}_2$, $\hat{\lambda}_3$, $\hat{\lambda}_5$, and $\hat{\lambda}_5$ or $\hat{\lambda}_1$, $\hat{\lambda}_2$, $\hat{\lambda}_3$, $\hat{\lambda}_3$ and $\hat{\lambda}_3$. These values are used to calculate a new global $\bar{\lambda}$. The whole procedure is repeated $m$ times to obtain $m$ global means $\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_m$. The corresponding sample variance $s^2_{\lambda}$ is used as an approximation of the estimation variance $\sigma^2_{\lambda}$ of the intensity.

A confidence interval for $\lambda$ results from rearranging the values $\bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_m$ in increasing order and using the values with indices closest to $\frac{\alpha}{2}m$ and $(1 - \frac{\alpha}{2})m$ as the bounds of a confidence interval for $\lambda$ of level $1 - \alpha$. (If $\alpha = 0.05$ and $m = 200$ the numbers are 5 and 195.)

Using a similar approach, the hypothesis $H_0 : \lambda_1 = \lambda_2$ i.e. that the intensities of two point processes are the same, can be tested by simulation. If the null hypothesis $H_0$ is rejected the data show evidence against the hypothesis of equal intensities.

Consider $k_1$ windows which yield the intensity estimates $\hat{\lambda}_{11}, \ldots, \hat{\lambda}_{1k_1}$ and $k_2$ with $\hat{\lambda}_{21}, \ldots, \hat{\lambda}_{2k_2}$ corresponding to $\lambda_1$ and $\lambda_2$ yielding $\bar{\lambda}_1$ and $\bar{\lambda}_2$ respectively as
well as $\Delta \lambda = \tilde{\lambda}_1 - \tilde{\lambda}_2$. This $\Delta \lambda$ is compared to $m$ values $\Delta \lambda_1, \ldots, \Delta \lambda_m$ which are obtained by resampling. Both samples of $\hat{\lambda}_{1i}$ and $\hat{\lambda}_{2j}$ are merged to form a unique sample of $k_1 + k_2$ values. Then $m$ new pairs of samples of $k_1$ and $k_2$ values are generated from this sample, by randomly resampling with replacement, and labelled as 1 and 2 even though they have been drawn from the same set of $\lambda$-values. Then the differences of $\lambda_1$- and $\lambda_2$-estimates are calculated, yielding $\Delta \lambda_1, \ldots, \Delta \lambda_m$. $H_0$ is rejected if $\Delta \lambda$ has an extreme position in the series of ordered values $\Delta \lambda_i$. If the error probability is $\alpha$, then values at positions smaller than $\alpha/2m$ and larger than $(1 - \alpha/2)m$ are considered extreme.

The same procedure as for $\lambda$ may be applied for all $r$ separately to the pair correlation function $g(r)$, i.e. instead of $\hat{\lambda}_i$ the $\hat{g}_i(r)$ are used now for all values of $r$ of interest.

Another common method is based on Monte Carlo simulation of models with estimated parameters. It is sometimes called the parametric bootstrap, although it is not a conventional bootstrap method. Assume that $\theta$ and $\hat{\theta}$ are the model parameter and its estimator, respectively. The model with parameter $\hat{\theta}$ is simulated $m$ times independently, and the estimators $\hat{\theta}_k$ are determined for $k = 1, 2, \ldots, m$ for the resulting data. The variance of these values is used as an approximation of the variance of the estimator $\hat{\theta}$; see Efron and Tibshirani (1993) and Givens and Hoeting (2005). This method can be applied regardless of the method that was used to derive $\hat{\theta}$. The parametric bootstrap is clearly computationally intensive, as it requires both simulation and parameter estimation for a large number of samples.

7.4 Goodness-of-fit tests

This section discusses goodness-of-fit tests for point processes, i.e. tests that are analogues of the Kolmogorov–Smirnov test familiar from classical statistics. For all these tests, the null hypothesis is simply $H_0$: ‘the model fits the data’. In point process statistics these tests are usually based on simulations, with the exception of some tests that are used in the context of the Poisson process, as discussed in Section 2.7. These tests are special cases of Monte Carlo tests as described in Chapter 4 of Davison and Hinkley (1997); see also Ripley (1988) and Robert and Casella (2005).

Two main approaches are discussed here. The first is a method that has been very popular since its introduction in Ripley (1977), even though it does not produce a formal significance test with known and predefined error probability $\alpha$. For this reason, Loosmore and Ford (2006) refer to the method as ‘incorrect’. However, this book warmly recommends the approach in applications.

7.4.1 Envelope test

This method is based on some functional summary characteristic $S(r)$ such as $g(r)$, $L(r)$, $L_{fin}(r)$, $H_s(r)$ or $D(r)$, as discussed in Chapters 3, 4 and 5. The idea of the
test is to compare the empirical summary characteristic estimated from a point pattern in the observation window $W$ to estimates of the summary characteristic for simulations from the model using the estimated parameters generated in the same window. The model is simulated $k$ times and the estimate of $S(r)$, $\hat{S}_i(r)$ for $i = 1, 2, \ldots, k$, is determined for each sample. Then the extreme values

$$ S_{\text{min}}(r) = \min_{(i)} \hat{S}_i(r) \quad \text{and} \quad S_{\text{max}}(r) = \max_{(i)} \hat{S}_i(r) $$

are determined. Finally, three curves showing $S_{\text{min}}(r)$, $\hat{S}(r)$ and $S_{\text{max}}(r)$ are plotted, as in Figure 7.3 as well as in many other figures in this chapter. Since $S_{\text{min}}(r)$ and $S_{\text{max}}(r)$ are envelopes of the $\hat{S}_i(r)$, the name ‘envelope method’ is often used, which leads to pointwise confidence bands.

The values $k = 19$ and $k = 99$ are often used for $k$, where $k = 19$ may appear to be rather small. On the other hand, the choice $k = 999$ might be considered as rather large. In the literature, values of $k$ satisfying $\alpha k \geq 5$ are recommended, where $\alpha$ is the error probability.

If the inequality

$$ S_{\text{min}}(r) \leq \hat{S}(r) \leq S_{\text{max}}(r) \quad (7.4.1) $$

holds for all $r$, the model is accepted, otherwise it is rejected. If the model is rejected, the values $r$ for which (7.4.1) is violated provide some information on the nature and reason for the deviations of the data from the model.

This test is often regarded and interpreted as a significance test. Indeed, if a fixed $r = r^*$ has been chosen prior to the simulation, the test which rejects the model if the inequality (7.4.1) is not satisfied for $r = r^*$ is a correct simulation test. Its error probability in one-sided testing is $1/(k + 1)$. Thus $k = 19$ corresponds to $\alpha = 0.05$ and $k = 99$ to $\alpha = 0.01$. In a two-sided test these values of $\alpha$ should be multiplied by 2. However, since ‘all’ $r$ are considered simultaneously, the probability of rejecting $H_0$ is increased and the true error probability is larger than 0.05 and 0.01, respectively. On the other hand, it is to be expected that a test based on single $r^*$ is rather conservative, i.e. the null hypothesis is rather unlikely to be rejected. This is because the model is simulated with parameters that have been estimated from the same data as those that were used for the test. Note that this problem has also been discussed in classical statistics in the context of the Kolmogorov–Smirnov test; see Conover (1999, pp. 443 and 448) and Armitage et al. (2001, p. 373).

To address this issue it is recommended to carefully choose the summary characteristic $S_{\text{test}}(r)$ used in the test. It should be of a different nature than $S_{\text{est}}(r)$, which is used to estimate the model parameters. A suitable approach, for example, is to estimate $\theta$ through $S_{\text{est}}(r) = g(r)$ (by the minimum contrast or approximate maximum likelihood methods) and to use $D(r)$ or $H_s(r)$ as $S_{\text{test}}(r)$ for the test rather than another second-order characteristic; see the discussion in Diggle (2003, p. 89). The rejection of a null hypothesis that is based on a test that applies the same
summary characteristic as for parameter estimation casts particular doubt on the null hypothesis.

This book suggests accepting that the classical choices of $\alpha = 0.05$ and 0.01 are only conventions and are not given by first principles. In other words, in the context of point processes a point process convention may be to work with $k = 19$ and $k = 99$ and to interpret these as ‘$\alpha = 0.05$’ and ‘$\alpha = 0.01$’, respectively.

Note that the test above can be modified and larger or smaller values can be used, rather than the minimum and maximum values. For $\alpha = 0.05$, $k = 999$ may be chosen and $S_{\min}$ is replaced by the 25th of the ordered $\hat{S}_i(r)$ values and $S_{\max}$ by the 975th.

### 7.4.2 Deviation test

Simulation tests with an exact error probability $\alpha$ can be constructed based on the general recipe described on p. 54.

Each simulated pattern $l$ is assigned a global deviation measure $\Delta_l$, by analogy with $D_l$ on p. 54, i.e.

$$
\Delta_l = \max_{0 \leq r \leq s} |S_{\hat{\theta}}(r) - \hat{S}_l(r)|
$$

or

$$
\Delta_l = \int_0^s |S_{\hat{\theta}}(r) - \hat{S}_l(r)|^\beta \, dr \quad \text{for } l = 1, 2, \ldots, k,
$$

where $S_{\hat{\theta}}(r)$ is the theoretical $S(r)$ with the estimated parameter $\hat{\theta}$ and $s$ a maximum $r$-value as on p. 96. In practice, the integrals are of course replaced by finite sums.

The values $\Delta_l$ and the $\Delta$-value for the original data,

$$
\Delta = \max_{0 \leq r \leq s} |S_{\hat{\theta}}(r) - \hat{S}(r)|
$$

or

$$
\Delta = \int_0^s |S_{\hat{\theta}}(r) - \hat{S}(r)|^\beta \, dr,
$$

are arranged in increasing order. If $\Delta$ has an extremely high position among these values, $H_0$ is rejected. For $\alpha = 0.05$ and $k = 19$, 99 and 999 the critical positions are those larger than 19, 95 and 950, respectively; for $\alpha = 0.01$ and $k = 99$ and 999 they are 99 and 990.

Summary characteristics commonly chosen as $S(r)$ are $L(r)$, $L_{\text{fin}}(r)$, $D(r)$ and $H_s(r)$. In accordance with common practice in classical statistics the density functions $g(r)$, $d(r)$ and $h_s(r)$ are not used here, because the estimation of $L(r)$, $D(r)$ and
$H_s(r)$ is more standardised than that of $g(r)$, $d(r)$ and $h_s(r)$. Again, as mentioned in the discussion of the envelope approach, a deviation test should be based on a different summary characteristic $S_{est}(r)$ than the one that was used as $S_{est}(r)$.

The observed $P$-value of the deviation test can be approximatively calculated as

$$\hat{p} = \frac{1 + \sum_{i=1}^{k} I(\Delta_i > \Delta)}{k + 1}.$$  \hspace{1cm} (7.4.2)

Loosmore and Ford (2006) discuss the variation of $\hat{p}$ as a function of $k$.

Practical experience has indicated that the deviation test is probably less powerful than the envelope test. To rectify this, the deviation test may be improved by using two summary characteristics, $S_1(r)$ and $S_2(r)$, of different nature (e.g. $L(r)$ and $D(r)$) and constructing a combined deviation measure, given here for the max case,

$$\Delta = \max_{0 \leq r \leq s} |S_{1,\hat{\theta}}(r) - \hat{S}_1(r)| + \max_{0 \leq r \leq s} |S_{2,\hat{\theta}}(r) - \hat{S}_2(r)|.$$ 

Deviation tests based on simulation approaches as before may be also applied if an explicit formula for $S_{\hat{\theta}}(r)$ is not known. Then $m$ additional independent patterns with the parameter $\hat{\theta}$ are generated, the corresponding $S_i(r)$ are calculated and averaged over $l$. Diggle (2003, p. 89), recommends using the $k$ samples above to also derive an estimate $S_{\hat{\theta}}(r)$.

**Example 7.2.** Phlebocarya pattern: testing a Matérn cluster process hypothesis

This example continues the analysis of the Phlebocarya pattern in Example 7.1 and tests the hypothesis that the Matérn cluster process with the minimum contrast estimates given in Example 7.1 fits the data.

![Figure 7.3 Empirical nearest-neighbour distance d.f. $\hat{D}(r)$ (solid line, left) and spherical contact d.f. $\hat{H}_s(r)$ (solid line, right) for the Phlebocarya pattern and envelopes from 99 simulations of the Matérn cluster process with minimum contrast estimates (dashed lines).](image_url)
Since the parameter estimation is based on the second-order characteristic $K(r)$, different summary characteristics are used for the test, $D(r)$ and $H_s(r)$. Figure 7.3 shows the empirical d.f. $\hat{D}(r)$ and the envelopes resulting from 99 simulations of the Matérn cluster process with the two parameter sets derived in Example 7.1. The empirical nearest-neighbour distance d.f. $\hat{D}(r)$ is completely within the envelopes for both sets and the test confirms that the models with the very small values of $\hat{R}$ fit the data. For the spherical contact d.f. $H_s(r)$ the situation is the same.

The result of the test is that the Matérn cluster model is also acceptable. However, the log-Gaussian Cox process performs better and seems to be a more realistic model. The Matérn cluster models suggest a very short range of correlation of 20 cm. This may show that the pattern is globally close to a Poisson process; however, an $L$-test of the CSR hypothesis leads to rejection.

**Example 7.3. Testing a Gibbs process hypothesis for the Spanish towns**

This example continues the analysis of the Spanish town pattern in Example 3.14 and tests the hypothesis that the hard-core Strauss process with the parameters obtained via the maximum likelihood method given in Example 3.14 fits the data. Since parameter estimation was not based on second-order characteristics the test applies the finite $L$-function. The parameters of the model are $\hat{r}_0 = 0.83$ miles, $\hat{r}_{\text{max}} = 3.5$ miles, $\hat{\alpha} = 2.08$ and $\hat{\beta} = 0.76$. Here the simulation is performed using the random variable $N(W)$ and that particular value of $\hat{\alpha}$ which corresponds to the mean value $E(N(W)) = 69$, the number of towns in $W$. (The value of $\hat{\alpha}$ was found by trial-and-error and simulation.) Figure 7.4 shows the empirical $L$-function and the envelopes resulting from each 99 simulations of the Strauss process. Since the empirical $L$-function is completely within the envelopes, the Strauss process hypothesis is accepted. By the way, for the distance summary characteristics the result is the same.

![Figure 7.4](image-url) The empirical finite $L$-function for the Spanish towns pattern (solid line) and the envelopes from 99 simulations of the hard-core Strauss process (dashed lines).
7.5 Testing mark hypotheses

7.5.1 Introduction

The discussion in Section 7.4 has not explicitly covered examples with marked point patterns, but in the context of marked point processes similar tests to those described above may also be used, based on suitable summary characteristics for marked point processes. Two fundamental hypotheses are particularly important:

1. the marks are independent;
2. the marks result from geostatistical marking, i.e. the random field model of Section 5.1.3 and 6.8 can be applied.

In the first case, the aim may be to find out whether the marks are positively or negatively correlated and in the second case, whether marks and points are correlated, e.g. whether marks depend on local point density. In some applications, the analysis may reveal that the marks can be considered independent. This might appear to be an unexciting result and be disappointing in a specific context where scientific theory would have suggested some correlations. Nevertheless, showing independence among the marks is also an important and useful result. If the marks turn out to be spatially correlated, a further question concerns the interaction among points and marks or whether the simple random field model without interaction is suitable.

The test procedures discussed in the following are all non-parametric approaches. They use functions of the type \( L_t(r) \) as defined in Section 5.4.1, e.g. \( L_{mm}(r) \), \( L_m(r) \) or \( L_{ij}(r) \).

7.5.2 Testing independent marking, test of association

All tests of independence hypotheses for marked point processes apply some form of resampling. This means that \( k \) new samples of marked point patterns with ‘independent’ marks are generated and the corresponding summary characteristics are compared to those of the original pattern. This is explained here in detail for \( L_t \)-functions as defined in Section 5.4.1.

Qualitative marks

Tests for qualitative marks are discussed here only for the most important bivariate case. This implies that, in a specific pattern, two marks \( i \) and \( j \) are selected from among the \( m \) different qualitative marks for an analysis of the correlation of the \( i \)- and \( j \)-marks, and points of type \( i \) a assigned the new mark 1 and points of type \( j \) the mark 2. Or, more generally, two groups of marks \( (i_1, \ldots, i_m) \) and \( (j_1, \ldots, j_m) \) are selected and all points of the first group get mark 1 and those of the second group...
mark 2. For example, in a forest conifers and deciduous trees may be considered, while the individual species are ignored.

The test considered here is based on the $L_{12}$-function as introduced in Section 5.3. Note that there are two different interpretations of the null hypothesis of ‘independent marking’. These need to be carefully distinguished and are discussed below: random labelling, and random superposition or population independence.

Random labelling means that the points in an originally non-marked point process are independently marked 1 or 2. Goreaud and Pélissier (2003) therefore refer to this as a posteriori marking. Typical examples of this may be forests in which the trees are infected by some disease or damaged by wind or frost (mark 1) or not (mark 2).

Random superposition means that a priori there are two independent point patterns in the same window $W$, one consisting of points of type 1, the other of type 2. These two patterns are combined, yielding the bivariate pattern. Typical examples include plant communities where the locations of plants from the different species within $W$ result from different dispersal mechanisms.

The summary characteristics for the two cases show a different but characteristic behaviour, as shown in Table 7.1. The quantities mentioned in the table are mainly defined in Section 5.3.2. $D_{12}(r)$ is the d.f. of the random distance from the typical type 1 point to its nearest type 2 neighbour and $H_{s,2}(r)$ is the spherical contact d.f. of the subprocess of type 2 points.

If the hypothesis of ‘independent marking’ is to be tested, the right null hypothesis has to be chosen first. In some cases, one of the two null hypotheses is completely inadequate (see Example 7.4) and thus can never be accepted. In many cases the choice is clearly determined by the study question and the origin and nature of the data. Below, a number of examples are used to illustrate this point. In more complicated cases where it is not clear which one of the two hypotheses seems more appropriate, statistical estimates of the pair correlation functions $g_{11}(r)$, $g_{22}(r)$ and $g_{12}(r)$ may be helpful. Certainly, $L_{11}(r)$, $L_{22}(r)$ and $L_{12}(r)$ may also be used,

### Table 7.1 Summary characteristics with index ‘12’ and simulation method for random labelling and random superposition. This table and the structure of the text referring to it were inspired by Goreaud and Pélissier (2003).

<table>
<thead>
<tr>
<th>Random labelling</th>
<th>Random superposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L(r) = L_{11}(r) = L_{22}(r) = L_{12}(r)$</td>
<td>$L_{12}(r) = r$</td>
</tr>
<tr>
<td>$g(r) = g_{11}(r) = g_{22}(r) = g_{12}(r)$</td>
<td>$g_{12}(r) = 1$</td>
</tr>
<tr>
<td>$p_{12}(r) = 2p_1p_2$</td>
<td>$p_{12}(r) = 2p_1p_2$</td>
</tr>
<tr>
<td>$p_{n}(r) = p_1^2$</td>
<td>$p_{n}(r)$ given by (5.3.18)</td>
</tr>
<tr>
<td>$D_{12}(r)$: (5.2.30)</td>
<td>$D_{12}(r) = H_{s,2}(r)$</td>
</tr>
<tr>
<td>reallocation</td>
<td>shifting</td>
</tr>
</tbody>
</table>
but due to the cumulative nature of the $L$-functions their use is not recommended
for exploration; see the discussion in Section 4.3. Further, it is important also to
assess $g_{11}(r)$ and $g_{22}(r)$, and not just $g_{12}(r)$.

Under the null hypothesis of random labelling, all three functions are equal and
their estimates will be similar. For random superposition, however, $g_{12}(r)\equiv 1$ but
$g_{11}(r)$ and/or $g_{22}(r)$ may differ substantially and this should be also reflected in
statistical estimates of the three functions. If it is not clear which of the hypotheses is
suitable for a given data set, both null hypotheses may be tested and the above func-
tions may be used for clarification. Note that the two cases cannot be distinguished
for Poisson processes.

Once the appropriate choice of null hypothesis has been determined, a test of this
null hypothesis may be carried out. This is again a simulation test, where $k$ marked
point patterns are generated with new marks for each of these. If the null hypothesis
is a random labelling hypothesis, random reallocation is used to generate these new
marks, i.e. the marks are permuted while the points are fixed. This implies that for
all simulated patterns the numbers of points of type 1 and type 2 are the same as
for the data.

If the null hypothesis is a random superposition hypothesis, random shifts are
used. Assume that $W$ is a rectangle or parallelepiped. The pattern of points of type
1 is fixed and the entire pattern of points of type 2 is shifted. If these points, which
are $x_i$ with $m(x_i) = 2$, are re-denoted by $y_1, y_2, \ldots, y_m$, a ‘shift’ implies

$$y_i \rightarrow y_i + u,$$

where $u$ is a random uniform location in $W$, the same for all $i$, but different
for different $l$. The operation $y_i \rightarrow y_i + u$ is based on an idea similar to periodic
edge-correction (see p. 184).

Envelopes $L_{12,\text{min}}(r)$ and $L_{12,\text{max}}(r)$ are constructed based on the $k$ simulated
patterns. If the empirical $L_{12}$-function derived from the data $\hat{L}_{12}(r)$ is not completely
within the envelopes or the corresponding confidence band the respective indepen-
dence null hypothesis is rejected.

If one or both null hypotheses have been rejected one should reconsider the
three partial pair correlation functions $g_{11}(r)$, $g_{22}(r)$ and $g_{12}(r)$ in order to find
the right explanation. Considering only $\hat{L}_{12}(r)$ and the envelopes may be risky, as
demonstrated in Goreaud and Pélissier (2003). It can lead to wrong conclusions as
to the reason for rejection, i.e. on the nature of correlations among the two types
of points.

Example 7.4. Testing the independence of amacrine on- and off-cells

Figure 1.2 shows the bivariate pattern of on- and off-cells (type 1 and type 2 points)
and Figure 5.5 the partial pair correlation functions $g_{ij}(r)$. The functions $g_{11}(r)$
and $g_{22}(r)$ are quite similar and seem to correspond to point processes with some
Figure 7.5 The envelopes $L_{12,\text{max}}(r) - \hat{L}_{12}(r)$ and $L_{12,\text{min}}(r) - \hat{L}_{12}(r)$ resulting from 99 simulations obtained by random reallocation (solid lines) and random shifts (dashed lines), for the amacrine cells. Since the null line is outside the confidence band corresponding to random reallocation for many values of $r$, the hypothesis of random marking is rejected. The fact that the null line is completely within the confidence band corresponding to random shifts suggests that the two subpatterns of type 1 and type 2 points are independent.

A tendency towards regularity. In contrast to this, $g_{12}(r)$ fluctuates randomly around 1. The statistical analysis in Example 5.5 suggests that the random superposition model is appropriate.

Here, the random labelling hypothesis is also considered, but only to point out the difference between the two hypotheses. In this particular context it is clear from the start that this hypothesis will not be accepted. The simulation tests yield the expected results. Figure 7.5 shows the differences of the empirical $L_{12}$-function and the $L_{12}$-envelopes from each 99 simulated patterns obtained by random allocations and shifts. Clearly, the random superposition hypothesis is accepted. In contrast, the random labelling hypothesis is rejected, since the null line is outside the confidence band for many values of $r$. Apparently, random allocations completely destroy the correlation structure in the pattern, in particular that in the patterns of type 1 and type 2 points. Consequently, the independence of the systems of type 1 and type 2 points appears to be proved and the random superposition model seems to be the appropriate model for the amacrine cells.

Example 7.5. **Frost shake in an oak forest**

This example reconsiders the distribution of trees damaged by frost shake in a young oak forest (*Quercus petraea*) at Allogny in France, as shown in Figure 7.6. ‘Frost shake is a split in a tree trunk, produced by the interaction of frost and sun, that leads to a lowering of timber quality’ (Goreaud and Pélissier, 2003). Visual inspection of Figure 7.6 shows that this example presents a difficult statistical problem as there is only some weak clustering of damaged trees. In the following, the question whether the pattern exhibits independent or dependent marking is addressed.

Figure 7.7 shows the three partial pair correlation functions $g_{11}(r)$, $g_{21}(r)$ and $g_{22}(r)$. There are no clear differences between the three functions. $g_{12}(r)$ is only
Figure 7.6  Positions of 392 oak trees in a 100 × 100 m square at Allogny, France. The pattern consists of 285 sound (○, type 1) and 107 damaged (●, type 2) trees. Courtesy of F. Goreaud and R. Pélissier. Reproduced by permission of Opulus Press.

Figure 7.7  The empirical partial pair correlation functions for the oak pattern: \(g_{11}(r)\) (solid curve), \(g_{12}(r)\) (dashed curve) and \(g_{22}(r)\) (dotted curve). The curves show that the pattern is close to a randomly labelled pattern with some tendency to clustering of trees of damaged type 2.

smaller than \(g_{11}(r)\) and \(g_{22}(r)\) for \(r\) between 2.0 m and 3.0 m, which indicates some weak repulsion among the trees of type 1 and 2. Between 2.5 and 3.5 m \(g_{22}(r)\) has larger values, which indicates some form of clustering of damaged trees.
In this example it is clear which null hypothesis is appropriate. The biological problem clearly dictates that the random labelling hypothesis is the right one. But even in the absence of biological background information it is clear that the random labelling hypothesis should be chosen, since the similarity of the $g_{ij}(r)$ indicates that there is some probability that the null hypothesis may be rejected and it is unlikely that the clusters have been generated by superposition. Hence, the random superposition hypothesis is tested here only to demonstrate the difference between the two hypotheses. Indeed, Figure 7.8 shows that it is rejected; for $r$-values smaller than 5.5 m the empirical $L_{12}$-function is too small. For the random labelling hypothesis the situation is not that clear; only for $r$-values around 5 m is $\hat{L}_{12}(r)$ smaller than $L_{12, \text{min}}(r)$. Note that, using data from a larger window ($125 \times 180$ m, containing the window considered here), Goreaud and Pélissier (2003) concluded that frost shake is a clustered phenomenon; damaged trees appear in clumps. This is a nice example that shows the strong impact of the choice of the observation window.

**Example 7.6. Testing the randomness of the distribution of oaks and beeches**

Figure 5.11 on p. 334 shows the mark connection functions $p_{ij}(r)$ which suggest some weak correlation between oaks (1) and beeches (2). Figure 5.6 shows the partial pair correlation functions, which do not indicate clearly which hypothesis for independence testing is appropriate here. Whereas $g_{11}(r)$ and $g_{22}(r)$ differ clearly ($g_{11}(r)$ looks like the pair correlation function of a hard-core process, $g_{22}(r)$ like that of a cluster process), $g_{12}(r)$ is smaller than 1 and does not indicate independence. Thus, both versions are considered in the following in an attempt to clarify the situation.

![Figure 7.8](image-url) The envelopes $L_{12, \text{max}}(r) - \hat{L}_{12}(r)$ and $L_{12, \text{min}}(r) - \hat{L}_{12}(r)$ resulting from 99 simulations of random allocations (solid lines) and random shifts (dashed lines) for the Allogny oak forest. The envelopes indicate varying degrees of deviation from the independence hypothesis: for the random labelling hypothesis it is not clear, whereas the random superposition hypothesis (which is nonsense for these data) is clearly rejected.
Figure 7.9 The envelopes $L_{12,\text{max}}(r) - \hat{L}_{12}(r)$ and $L_{12,\text{min}}(r) - \hat{L}_{12}(r)$ resulting from 99 simulations of random allocations (solid lines) and random shifts (dashed lines) for the pattern of oaks and beeches. The results of the two tests differ: whereas the null line is entirely within the confidence band for the random allocations, this is not true for the random shifts.

Figure 7.9 shows the differences in the empirical $L_{12}$-function and the $L_{12}$-envelopes resulting from 99 simulations, for random allocations and random shifts. For the random superposition test the null line is not completely within the confidence band. There are deviations for values around $r = 6.0$ m and hence the conclusion is, as expected, that the marks correlate weakly. For the random allocation method the result is different: the null line is completely within the confidence band. This might suggest the conclusion that there is no correlation between oaks and beeches; however, this has been proved wrong by the random shift test and the correlation functions. The discrepancy in the result based on the random allocation test may be due to the fact that relabelling does not destroy the global structure in the sample, which consists of clusters with many type 2 trees and few type 1 trees.

Note that for the example of palms and mounds on p. 335 the random labelling test also rejects the independence hypothesis, whereas the random shift test accepts it.

**Quantitative marks**

For quantitative marks, the test functions

$$t_2(m_1, m_2) = m_1 m_2 \quad \text{and} \quad t_4(m_1, m_2) = \frac{1}{2} (m_1 - m_2)^2$$

can be recommended for defining $L_{mm}(r)$ and $L_{\gamma}(r)$, respectively. If the marks are independent,

$$L(r) = L_{mm}(r) = L_{\gamma}(r).$$
For the $k$ $L_t$-functions $L_{t,i}(r)$ which result from simulated samples for $i = 1, 2, \ldots, k$, the envelopes based on $L_{t,\text{min}}(r)$ and $L_{t,\text{max}}(r)$ may be determined. An $\hat{L}_t(r)$ that is outside these envelopes indicates non-independence at specific distances. In graphical presentation it is useful to plot differences $L_{t,\text{max}}(r) - \hat{L}_t(r)$ and $L_{t,\text{min}}(r) - \hat{L}_t(r)$ and to assess whether the null line remains within the confidence band.

Marked point patterns with quantitative marks are resampled by random reallocation (or random marking or labelling), i.e. the points are fixed but are allocated new marks. There are two different approaches to this: permutation of the marks, and resampling the marks with replacement from the empirical mark distribution. Both approaches ignore existing mark correlations, but if the marks are independent these approaches do not change the mark correlation functions significantly. Note that permutation of the marks guarantees that all samples have the same empirical mark d.f.; this is not the case for independent marking using the empirical mark d.f.

The following examples discuss independence tests for two different data sets.

**Example 7.7. Gold particles: testing the independence of diameter marks**

Figure 5.18 on p. 342 shows the mark correlation function $k_{mm}(r)$ for the pattern of gold particles. For small $r$ its values are below 1, which seems to indicate that there is some inhibition among the particles. The aim of this example is to test whether the relationship is significant using the random reallocation method and the $L_{mm}$- and $L_{\gamma}$-functions.

Figure 7.10 shows the differences in the empirical $L_{mm}$-function and the $L_{mm}$-envelopes resulting from 99 simulations. The null line is not completely within the confidence band, indicating that the marks are indeed correlated. For $L_{\gamma}$ the situation is similar but much clearer, also shown in Figure 7.10. Considering the results of the tests here and of the exploratory analysis in Section 5.3.3, one may

![Figure 7.10](image-url)
conclude that there is a weak correlation among the diameter values with a tendency for diameters of particles at short distances to be similar.

**Example 7.8. Testing the independence of the marks in the spruce stand in the Tharandter Wald**

Figure 6.19 shows the mark correlation function $k_{mm}(r)$ and the mark variogram for the 134 spruce trees as discussed in Example 6.4. The two correlation functions lead to different conclusions: whereas $k_{mm}(r)$ indicates independence, $\gamma_m(r)$ indicates that close pairs of big and small trees appear slightly more frequently than would have been expected for independent marks.

The random allocation method is used and $L_{mm}$ and $L_{\gamma}$ are considered. Perhaps slightly surprisingly, the null hypothesis of independent marking is not rejected by the envelope test with $L_{\gamma}$. The results are different for $L_{mm}$. Figure 7.11 shows the differences in the empirical $L_{mm}$-function and the $L_{mm}$-envelopes resulting from 99 simulations. The null line is not completely within the confidence band and the null hypothesis of independence is rejected. In conclusion, a weak correlation among the diameter marks can indeed be assumed.

The contradictory results for the spruce stand might be a result of the small size of the point pattern, which consists of only 134 trees. Probably the most realistic assumption is that the marks are simply independent.

### 7.5.3 Testing geostatistical marking

Tests of geostatistical marking are based on the $L_m$-function and random allocation, and strongly resemble tests of independent marking. Under geostatistical marking,

$$L(r) = L_m(r).$$  \hspace{1cm} (7.5.5)

![Figure 7.11](image) The envelopes $L_{mm,\max}(r) - \hat{L}_{mm}(r)$ and $L_{mm,\min}(r) - \hat{L}_{mm}(r)$ from 99 simulations resulting from random reallocation for the stand of spruces. Since the null line is not completely inside the confidence band, the marks are considered correlated.
Figure 7.12 The envelopes $L_{m,\max}(r) - \hat{L}_m(r)$ and $L_{m,\min}(r) - \hat{L}_m(r)$ resulting from 99 simulations with random reallocation for the gold pattern. Since the null line is not completely within the confidence band, the hypothesis of geostatistical marking is rejected. But note that the upper envelope only slightly crosses the null line.

Example 7.9. Gold particles: testing for geostatistical marking
This example continues Example 7.7 and uses the $L_m$-function to test for geostatistical marking. Figure 7.12 shows the differences the empirical $L_m$-function and the $L_m$-envelopes from 99 simulations with random reallocation. The null line is not completely within the envelopes, but nevertheless the hypothesis may be accepted, as a simulation test with an estimated variogram shows (see p. 349).

Example 7.10. Testing for geostatistical marking in the pattern of spruce trees from the Tharandter Wald
This example continues Example 7.8 and tests for geostatistical marking, again using the $L_m$-function. Figure 7.13 shows the differences in the empirical $L_m$-function and the $L_m$-envelopes resulting from 99 simulations with random reallocation. The
null line is not completely within the confidence band, which indicates that the hypothesis of geostatistical marking might have to be rejected. Since independent marking is a special case of geostatistical marking, this result may suggest rejecting the independence hypothesis as well. However, for this example this result may be again due to the small size of the sample, and the authors tend to assume that the marking here is independent.

Note, by the way, that the pattern in this stand is the result of a forester’s work. However, since the aim was to grow trees of a similar size the result is certainly a pattern of trees with similar diameters and only random fluctuations around the mean, where any effect of competition among the trees can hardly be identified.

Table 7.2 lists the results of mark independence tests for a number of examples.

<table>
<thead>
<tr>
<th>Point pattern</th>
<th>$L_t$ used</th>
<th>Envelope test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold particles</td>
<td>$L_{mm}$</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>$L_m$</td>
<td>○</td>
</tr>
<tr>
<td></td>
<td>$L_\gamma$</td>
<td>–</td>
</tr>
<tr>
<td>Spruces in Tharandter Wald</td>
<td>$L_{mm}$</td>
<td>○</td>
</tr>
<tr>
<td></td>
<td>$L_m$</td>
<td>○</td>
</tr>
<tr>
<td></td>
<td>$L_\gamma$</td>
<td>+</td>
</tr>
<tr>
<td>Amacrine cells</td>
<td>$L_{12}$</td>
<td>+ tr</td>
</tr>
<tr>
<td></td>
<td>$L_{12}$</td>
<td>– pe</td>
</tr>
<tr>
<td>Oaks and beeches</td>
<td>$L_{12}$</td>
<td>○ tr</td>
</tr>
<tr>
<td></td>
<td>$L_{12}$</td>
<td>+ pe</td>
</tr>
<tr>
<td>Oak frost shake</td>
<td>$L_{12}$</td>
<td>– tr</td>
</tr>
<tr>
<td></td>
<td>$L_{12}$</td>
<td>○ pe</td>
</tr>
<tr>
<td>Concrete sample (planar section)</td>
<td>$L_{12}$</td>
<td>○ tr</td>
</tr>
<tr>
<td></td>
<td>$L_{12}$</td>
<td>+ pe</td>
</tr>
<tr>
<td>Palms and mounds</td>
<td>$L_{12}$</td>
<td>+ tr</td>
</tr>
<tr>
<td></td>
<td>$L_{12}$</td>
<td>– pe</td>
</tr>
<tr>
<td>Buttercup</td>
<td>$L_{12}$</td>
<td>+ tr</td>
</tr>
<tr>
<td>coverage</td>
<td>$L_{12}$</td>
<td>+ tr</td>
</tr>
<tr>
<td>stoniness</td>
<td>$L_{12}$</td>
<td>+ tr</td>
</tr>
<tr>
<td>Copper deposits</td>
<td>$L_{12}$</td>
<td>+ tr</td>
</tr>
</tbody>
</table>

– Rejection of independence hypothesis, empirical $L_t(r)$ for many $r$-values outside the envelope strip.
○ The same, but only for few ($\leq 3$) $r$-values.
+ No rejection.
tr Random shift.
pe Random allocation.
Note that all tests are based on 99 replications (simulations) and 20 equidistant $r$-values with maximum distance approximately equal to half the smaller side length of the window $W$.

### 7.6 Bayesian methods for point pattern analysis

The Bayesian approach to statistical inference has recently become very popular, especially in the analysis of complex data sets. This is mainly a result of the development of Markov chain Monte Carlo methods, which has made it possible to apply Bayesian methods, since the early 1990s, to a much wider range of situations than had previously been possible. Not surprisingly, the Bayesian approach, together with MCMC methodology, has also made its way into point process statistics. In Section 3.2.1 the Bayesian approach was briefly mentioned in the context of statistics for the binomial distribution, and in Example 5.7 in the context of identifying mother points in a cluster process, but it may also be applied in many other situations, including intensity estimation, reconstruction of tessellations from point process data, and estimation of pair potential functions for Gibbs processes.

The Bayesian approach follows a different philosophy than the classical frequentist approach to statistical inference; the parameters and other unknown quantities, such as unobserved variates (covariates) and missing data, are considered random variables that are assumed to follow some probability distribution. In other words, the analysis involves formulating a model for both the data and the parameters. All uncertainties are expressed in terms of probability distributions and probabilities.

This implies the following modelling steps:

- As in a non-Bayesian context, the observations $x$ are modelled by a probability distribution $f(x|\theta)$. However, in the Bayesian setting the distribution is interpreted as a conditional distribution of the data $x$ given the vector $\theta$ of all unknown and unobserved quantities in the model, including the parameters but possibly also unobserved variates and missing data.

- As mentioned above, the unknowns $\theta$ are further modelled with a prior distribution $\pi(\theta)$. This is a specific distribution that has to be chosen a priori and may be multivariate and may even have a complex dependence structure. This distribution reflects the investigator’s beliefs on the unknown quantities prior to the analysis.

- The investigator’s uncertainty on $\theta$ given the data $x$ is expressed by the posterior distribution

$$
\pi(\theta|x) = \frac{\pi(\theta) f(x|\theta)}{\int \pi(\theta) f(x|\theta) \, d\theta}.
$$

(7.6.6)
Equation (7.6.6) is an application of Bayes’ theorem. It relates unknown quantities given the data to the prior distribution and the distribution of the data.

- The posterior distribution is often difficult to handle and an analytical calculation of its characteristics is impossible in most cases due to the normalising integral in the denominator. However, MCMC methods can be applied to simulate from the posterior distribution. Refer to the literature (see below) for details on how this is done in practice.
- Once the posterior distribution has been determined, the analysis typically focuses very much on the characteristics of the marginal posterior distributions $\pi(\theta_i|x)$ for $\theta = (\theta_1, \ldots, \theta_p)$. This includes, in particular, the summary characteristics familiar from classical statistics, such as mean, standard deviation and quantiles.


To illustrate Bayesian methods, Example 3.4 is reconsidered here as this is based on a probability distribution and on a modelling approach that is familiar from classical statistics.

Example 7.11. Analysis of fruit dispersal around an ash tree
Consider the estimation of the intensity function $\lambda(r)$ for the data in Example 3.4 on p. 122. In this example, the locations of fruit dispersed by an ash tree are modelled using an inhomogeneous Poisson process. The data set $\{n_i, r_i\}$ consists of trap counts $n_i$ and of the distances $r_i$ of 66 traps from the mother tree, where the traps are located around the tree as shown in Table 3.1.

Again, as in Example 3.4, the observations are modelled by a Poisson distribution

$$f(n_i|\mu_i) = \prod_{i=1}^{n} \frac{\mu_i^{n_i}}{n_i!} e^{-\mu_i}. \quad (7.6.7)$$

The parameter $\{\mu_i\}$ of the Poisson distribution results from the intensity function $\lambda(r)$ on p. 122 and models the mean number of fruit in trap $i$. It depends on the total number $m$ of fruits and the random dispersal, which is assumed to follow a lognormal distribution with density function

$$d(r) = \frac{1}{\sigma r \sqrt{2\pi}} \exp \left( -\frac{(\ln r - \mu)^2}{2\sigma^2} \right) \quad \text{for} \quad r \geq 0,$$
as in Example 3.4. Given \( m \) and the dispersal parameters \( \mu \) and \( \sigma \), the mean number of fruits in trap \( i \) of area \( a \) is

\[
\mu_i = \frac{amd(r_i)}{2\pi r_i}.
\]

The analysis now applies the Bayesian philosophy and treats the unknowns \( m, \mu \) and \( \sigma \) as random variables, with \( \theta = (m, \mu, \sigma) \), requiring the specification of their joint prior distribution. It is natural to assume that the total number of fruit \( (m) \) and the dispersal parameters \( (\mu, \sigma) \) are independent of each other a priori. Hence,

\[
\pi(m, \mu, \sigma) = \pi(m) \pi(\mu, \sigma).
\]

In order to avoid complex notation \( \pi(\cdot) \) is used for multivariate distributions as well as for the corresponding marginal distributions. The specific distributions can be identified on the basis of the arguments.

Prior information on the number of fruit is not available here – it is not clear, for example, whether certain values are more likely than others. In many applications, however, researchers do have expert knowledge on this and would probably be able to specify a specific distribution. In the absence of prior information, a natural choice of a prior distribution for \( m \) is the uniform distribution on \( \{0, 1, \ldots \} \), which is well approximated by its continuous counterpart, the uniform distribution on \([0, \infty)\). For this prior \( \int_0^\infty \pi(m)dm \) is infinite and referred to as improper. This implies that \( \pi(m, \mu, \sigma) \) is also improper. However, such a prior may be used, provided that the posterior (7.6.6) is a proper distribution, i.e. that the integral of the posterior over the whole space is 1.

The parameters \( \mu \) and \( \sigma \) determine the distances \( r \) of the dispersion of fruits from the mother tree. Accordingly, the mean distance is

\[
E(r) = \exp \left( \mu + \frac{1}{2} \sigma^2 \right) \tag{7.6.8}
\]

and the mode

\[
r_{mode} = \exp(\mu - \sigma^2). \tag{7.6.9}
\]

The mode of the dispersion distribution is an intuitive characteristic for a priori reasoning. Assume now that some prior knowledge on the seed dispersal mechanism of the tree and hence prior information on the \( \mu \) and \( \sigma \) can be used to derive the prior distribution of these \( \pi(\mu, \sigma) \). In other words, the analysis now assumes that the maximum is between 0 m and 6 m with a high probability, i.e. \( \ln(r_{mode}) \) has a Gaussian distribution where the parameters are chosen as \( \mu = 3 \) and \( \sigma = 1.6 \) and \( \sigma \) can be modelled as

\[
1/\sigma^2 \sim \text{the gamma distribution with mean 1 and variance 10.}
\]
Note that in Bayesian statistics the precision parameter $1/\sigma^2$ is commonly modelled as a gamma distribution. A weaker prior such as the uniform distribution may be chosen if less information is available. Finally, the prior $\pi(\mu, \sigma)$ can be specified based on these two distributional assumptions and (7.6.9).

The posterior distribution $\pi(m, \mu, \sigma, \{\mu_i\} | \{n_i\})$ is rather complex but can be explored by simulation. Figure 7.14 gives a graphical description of the dependence structure. Note that complex hierarchical models are often displayed in this way or even constructed with reference to a similar graph.

The parameters of the marginal posterior distributions are shown in Table 7.3, and have been calculated using a large number of MCMC simulation runs, which were necessary to obtain stable results. This corresponds to the experience in Example 3.4 that the likelihood function was rather flat.

![Directed acyclic graph showing the hierarchical Bayesian model describing the dependence structure among the variables. Observed variables are represented by squares and unobserved variables by circles.](image)

Table 7.3 Description of the marginal posterior distributions for $m$, $\mu$ and $\sigma$ extracted from MCMC simulations: mean, standard deviation and 95% credible intervals (2.5% and 97.5% quantiles).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>$q_{0.025}$</th>
<th>$q_{0.975}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>219 800</td>
<td>84 760</td>
<td>138 700</td>
<td>440 700</td>
</tr>
<tr>
<td>$\mu$</td>
<td>4.11</td>
<td>0.35</td>
<td>3.65</td>
<td>5.04</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.95</td>
<td>0.13</td>
<td>0.77</td>
<td>1.28</td>
</tr>
</tbody>
</table>
The quantiles $q_{0.025}$ and $q_{0.975}$ are the bounds of the 95% credible intervals for the parameters. For example, the posterior probability that the mean number of fruits from the tree is between 138 700 and 440 300 is 0.95. This interval is wide, which means that the experiment provides little information on the number of fruit. However, this is more valuable than the simple $\hat{m} = 179 800$ obtained in Example 3.4.

Note that in this approach, which is often referred to as ‘fully Bayesian’, the estimation also provides information on the precision (in terms of standard deviation and credible interval) of the parameters. Unlike in the frequentist approach, which often depends on large sample size assumptions, these estimates are ‘exact’ in the sense that they do not rely on any such assumptions. Recall that the maximum likelihood estimates in Example 3.4 were: $\hat{m} = 179 800$, $\hat{\mu} = 3.93$ and $\hat{\sigma} = 0.94$. These are in good agreement with the results of the fully Bayesian analysis.

**Bayesian estimation of the intensity for a Poisson process**

The following discusses the Bayesian approach, now applied to a simple point process model, the inhomogeneous Poisson process. Bayesian approaches for more complex models are briefly discussed below.

Assume that $N$ is an inhomogeneous Poisson process with intensity function $\lambda(x)$ and a sample $x_1, \ldots, x_n$ of points in the observation window $W$ is given. In the Bayesian approach, $\lambda(x)$ is considered a random variable, i.e. a random field. Consequently, the conditional distribution of the data given $\{\lambda(x)\}$ is the likelihood of an inhomogeneous Poisson process, given by (3.4.4), which in the notation used here is

$$f(x_1, \ldots, x_n|\{\lambda(x)\}, \theta) = \prod_{i=1}^{n} \lambda(x_i; \theta) \exp \left( - \int_W \lambda(x; \theta) \, dx \right).$$

Several kinds of prior distributions for the intensity function, i.e. distributions of the random field $\{\lambda(x)\}$, have been suggested such as the log-Gaussian Cox process (Beneš et al., 2005), the gamma–Poisson process (Wolpert and Ickstadt, 1998), and the partition process (Heikkinen and Arjas, 1998; Ferreira et al., 2002).

Note that two of these processes have been discussed in this book in the context of Cox processes (refer to p. 380), but the partition process has not. The intensity of the partition process is

$$\lambda(x) = \sum_{k=1}^{K} \lambda_k \mathbf{1}_{E_k}(x),$$

where $\{E_k\}$ is a random partition of $W$, typically the Voronoi tessellation generated by some point process. Here, the $(\lambda_1, \ldots, \lambda_K)$ are the unknown parameters $\theta$. The function $\lambda(x)$ is not continuous, but the average of several simulated realisations from the posterior distribution of $\lambda(x)$ results in a smooth surface.
Any of the three processes listed above may be chosen as the prior for the random field and contains unknown hyperparameters $\theta$, i.e. for each of these a similar hierarchical model structure may be used and the posterior $\pi(\theta, \{\lambda(x)\}|\mathcal{N})$ is proportional to

$$\pi(\theta) \pi(\{\lambda(x)\}|\theta) f(x_1, \ldots, x_n|\{\lambda(x)\}).$$

Note that in this application the denominator in (7.6.6) may be ‘ignored’ in the MCMC simulation, i.e. it suffices to use the numerator to characterise the posterior distribution.

The estimation of the random intensity $\{\lambda(x)\}$ of point pattern data assumes a priori information on the smoothness of the result. There is some similarity to kernel smoothing of intensity functions as discussed on p. 115 and sometimes the term ‘Bayesian smoothing’ is used.

**Other applications**

Blackwell (2001), Blackwell and Møller (2003) and Skare et al. (2007) model badger territories. Badgers (*Meles meles*) of a clan create latrines near their territorial borders. These markings form a point pattern with increased intensity near the borders of territories. A natural tool in modelling territories is the Voronoi tessellation (see Section 1.8.4) generated as a secondary structure of a prior point pattern with one ‘central’ point for each territory. Bayesian modelling may be used to provide estimates of badger territories and their centre points, and posterior simulations produce measures of the uncertainty in these reconstructions. Similar approaches have been applied to image segmentation (Byers and Raftery, 2002), and delineation of homogeneous regions from categorical and quantitative soil measurements for precision farming purposes (Guillot et al., 2006).

Illian et al. (2008) apply both a non-Bayesian and a Bayesian approach to a multi-type pattern of 24 species of Australian plants (Armstrong, 1991; see also Section 4.9 above) and they notice that the Bayesian approach proves to be rather flexible in particular in the context of complex patterns and a large number of parameters.

For finite Gibbs processes (see Section 3.6) the likelihood is known and, in theory, a Bayesian approach could be applied to the estimation of the pair potential function. However, a big difficulty here is the fact that the normalising factor of the Gibbs likelihood (see (3.6.11)) remains in the MCMC algorithm for posterior simulation and also requires simulation, except in the Poisson case. Usually MCMC within MCMC is regarded as intractable. However, some progress has recently been made on this issue. It is possible to find the maximum of the posterior density; see Heikkinen and Penttinen (1999). Hence a Bayesian approach can be taken here, with the advantage that maximum a posteriori estimation is more useful than the maximum likelihood solution if regularisation of the estimates is important. This is the case, for example, if the pair potential is parametric but very flexible, such
as a multi-scale model with many steps. Møller et al. (2006) suggest an ingenious method based on an auxiliary variable for a fully Bayesian estimation of the pair potential.

Marked point process models have been used as *regularisation priors* in image analysis and computer vision; see Stoica et al. (2004).

For Bayesian methods to be applied to point pattern analysis the likelihood of the point process has to be known. However, many important point processes are defined by construction, and the likelihood is not readily available. For these models the Bayesian methods are, at least currently, not a competitive alternative to the frequentist approach. The inhomogeneous Poisson process and the finite Gibbs process are two of the few exceptions. Indeed, many Bayesian applications for point pattern analysis rely on the inhomogeneous Poisson process as the model for observations. This model has been frequently applied, for example, in epidemiology, most often to counts of points in small compartments or cells; see Elliott et al. (2000) and Diggle (2003). Currently, Bayes methods in the context of point process statistics are very topical and many new methods are being developed.