Spatial Tessellations: Concepts and Applications of Voronoi Diagrams
Edition by ATSUYUKI OKABE & BARRY BOOTS and SUNG NOK CHIU

CHAPTER 8

Point Pattern Analysis

A common concern that runs through a wide range of disciplines is the examination of the spatial occurrence of a particular phenomenon. Illustrative examples, which also reflect the extensive range of scales covered by such phenomena, include the location of sulphide inclusions in a steel (metallurgy), artefacts over a site (archaeology), nest sites of bird species in a habitat (ecology), towns in a state (geography), seismic events in a continent (geology) and galaxies of stars in the universe (astronomy). In each of these instances it is possible to represent individual incidences of the phenomenon as a set of n distinct points, $P = \{p_1, p_2, \ldots, p_n\}$, in a bounded region, B, in either $\mathbb{R}^2$ or $\mathbb{R}^3$. We call such representations, empirical point patterns. Note that while the individual objects themselves are not points, such representations are possible because the physical sizes of the objects are very small relative to both the distances between them and the extent of the region in which they occur.

Empirical point patterns are examined for a variety of reasons. A major one is the expectation that a pattern represents one source of evidence that may be useful in learning more about the process responsible for creating it. If our ideas concerning a phenomenon are sufficiently developed, we may be able to build an explanatory model of it. We can use such models to derive hypotheses concerning the locational behaviour of the phenomenon. Support for such hypotheses, and thus the models from which they are derived, can be obtained by analysing empirical point patterns of the phenomenon. Sometimes such analysis focuses on the relative locations of individual points of P or on the spatial distribution of P over B. Such situations are dealt with in Sections 8.1 and 8.2. At other times we may be interested in how the members of P are located with respect to other objects, not members of P , also located in B. These objects may be represented as points, lines or areas, or some combination of all three. We examine such circumstances in Section 8.3 which is further subdivided in terms of the dimensionality of the other objects.

When our knowledge of the phenomenon represented by the point pattern is more rudimentary, modelling may be less appropriate. Then analysis takes on a more exploratory form as we seek to identify fundamental characteristics of the point pattern. Sections 8.4–8.6 discuss some exploratory analyses involving shape, spatial intensity, and segmentation of point patterns which have been pursued using concepts of Voronoi diagrams and related

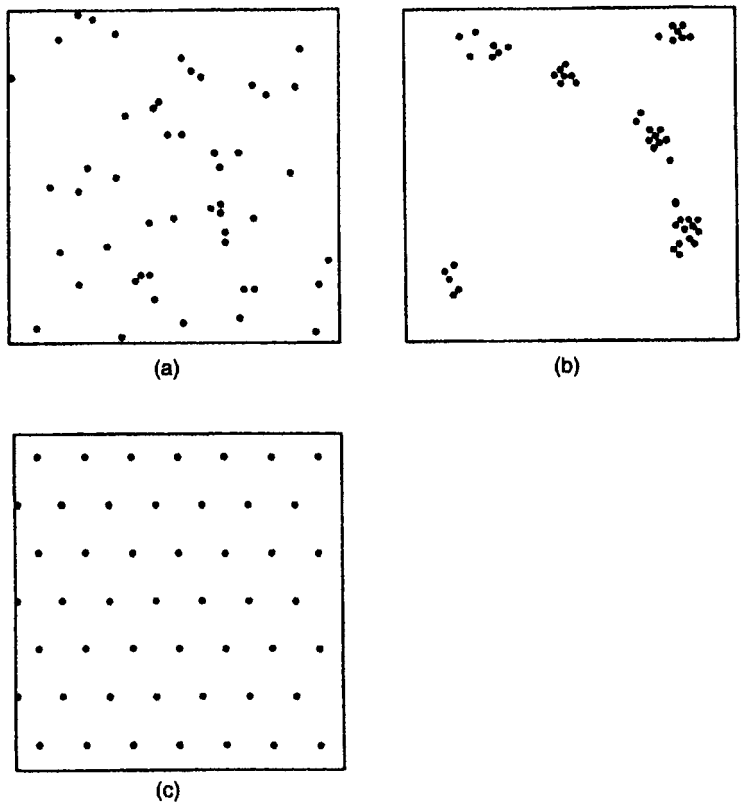

Figure 8.0.1 Classes of point patterns: (a) complete spatial randomness; (b) clustered; (c) regular.

structures. Finally, in Section 8.7 we show how such concepts can also be used to generate point processes with specific properties.

In those circumstances where our primary concern is with modelling and our analysis is limited to a consideration of only the members of P, it is usual to adopt an approach that involves establishing a theoretical point pattern with respect to which other patterns are compared. The theoretical point pattern chosen is that which results from the operation of the homogeneous Poisson point process, Θ_P, described in Section 1.3.3. A pattern that results from such a process can be considered one that would arise by chance in a completely undifferentiated environment. Following Diggle (1983) we call such a point pattern *complete spatial randomness* (CSR). Figure 8.0.1(a) shows such an example of CSR. In view of the conditions involved in generating such a pattern, it is unlikely that true CSR occurs in any empirical situations. Instead, our concern with CSR is primarily for its role as an idealized standard. The two conditions that are assumed to exist when CSR results provide a simple model that can be useful in several circumstances. For

instance, if we know little about the process responsible for a particular empirical point pattern, we can begin by testing a hypothesis that the pattern is CSR. This allows for exploration of the empirical point pattern. Whether or not the initial hypothesis relating to CSR is rejected, a description of the empirical point pattern remains. If CSR cannot be rejected, further analysis is unwarranted, otherwise the results of the test can be used as an aid in formulating new hypotheses regarding the empirical point pattern.

Classes of point patterns can be recognized using CSR as a benchmark. *Clustered* point patterns are those in which the points are significantly more grouped in B than they are in CSR (see the example in Figure 8.0.1(b)) whereas *regular* patterns (sometimes also called uniform or dispersed point patterns) are those in which the points are more spread out over B than they would be in CSR (see Figure 8.0.1(c)). Clustered and regular patterns can arise as the result of changing one or more of the postulates of Θ_P identified in Section 1.3.3.

Homogeneity may be relaxed by differentiating subareas of B in some way. Environmental inhomogeneity implies that some subareas of B are less likely to receive a point than other subareas, or might even be prohibited from receiving a point. In such circumstances we would expect to find more points in the favoured subareas of B than elsewhere, thereby producing a clustered pattern.

One way of relaxing independent scattering is to permit interaction among points. Points may either attract or repulse one another. Attraction may result from processes such as agglomeration, association, and some types of diffusion and competition. In each instance the result is a clustered pattern. Such situations, where different processes result in similar types of point patterns, are not unusual in point pattern analysis. When they occur, if the only evidence we have is the pattern itself, further analysis is necessary to determine the conditions responsible for the observed clustering. Instead of points attracting each other, in some instances, such as diffusion or competition, points may repel each other. Repulsion will likely produce regular patterns. The types of models that produce non-CSR outcomes are reviewed by Haggett *et al.* (1977, Chapter 13), Getis and Boots (1978), Ripley (1981), Cliff and Ord (1981), Diggle (1983), Upton and Fingleton (1985), Cressie (1991), Stoyan and Stoyan (1994) and Stoyan *et al.* (1995) (see also Sections 1.3.3 and 5.12).

In general, the hypothesis of CSR (or any other hypothesis concerning the spatial nature of the empirical point pattern) is tested by comparing measures of selected characteristics of the empirical point pattern with those of the hypothesized pattern. The most frequently used approaches are quadrat analysis, nearest neighbour analysis, and second-order analysis (see Boots and Getis, 1988, for details). Quadrat analysis involves defining a set of spatial sampling units (quadrats) of consistent size and shape which, depending on the specific approach adopted, may or may not overlap and which, collectively, may or may not completely cover B. A frequency array, $f(x)$, is generated by counting the number of quadrats containing $x = 0, 1, 2, \ldots$

points. $f(x)$ can then be compared with the corresponding distribution, $p(x)$, expected under CSR with the same intensity as the empirical pattern, using a χ^2 goodness-of-fit test.

Nearest neighbour analysis involves measuring the distance between each point in P and the nearest other point. The nearest neighbour distances so obtained are compared with those expected under CSR with the same intensity as the empirical pattern using a variety of tests. Second-order analysis extends this approach by examining the distances between all pairs of points in P. These interpoint distances are compared with the CSR expectations, again taking into consideration the intensity of the points in the empirical pattern.

8.1 POLYGON-BASED METHODS

8.1.1 Direct approach

If we represent the empirical point pattern as a set, $P = \{p_1, p_2, \ldots, p_n\}$ of n distinct points in a bounded region, B, of $\mathbb{R}^m$, we can define the bounded Voronoi diagram, $\mathcal{V}(P)$, of P (see Section 2.1). For each polygon of $\mathcal{V}(P)$ we can measure the values of selected characteristics. For $\mathcal{V}(P)$ in $\mathbb{R}^2$ the most obvious characteristics in terms of dimensionality from zero to two are the number of sides (vertices) N of a polygon, the perimeter P, and the area A, although several measures of shape have also been considered (Burger *et al.*, 1990; Marcelpoil and Usson, 1992). We can compare the values of these characteristics with those of polygons of the Voronoi diagram defined for a hypothesized point process. Since edge effects influence the values for polygons around the boundary of B, some allowance must be made for this. One approach (already considered in Section 5.4) is to exclude values for any polygon for which a circle, centred at any vertex, v, of the polygon and passing through the three points of P which are equidistant from v, intersects the boundary of B.

The polygonal approach was originally proposed by Evans (1967), but despite being advocated by a number of researchers (Haggett *et al.*, 1977, pp. 436–439; Matérn, 1979; Cormack, 1979, pp. 171–175; Ripley, 1981, p. 149; Cliff and Ord, 1981, p. 110; Thomas, 1981), it was not widely developed until later. One reason for this delay is that the properties of N, by far the most studied and most easily measured characteristic initially, do not always vary strongly amongst Voronoi diagrams defined for patterns generated by different processes. Recall that in Section 5.5 (equation (5.5.6)) we noted that the average value of N, $\mathrm{E}(N)$, is always 6 for any normal, planar tessellation. Furthermore, Boots (1977) demonstrates that the frequency distribution of N for at least one set of clustered patterns is not significantly different from that of CSR. Vincent and Howarth (1982) reach a similar conclusion for a set of regular patterns, although Burger *et al.* (1990) suggest that the standard deviation of N is a strong indicator of the level of regu-

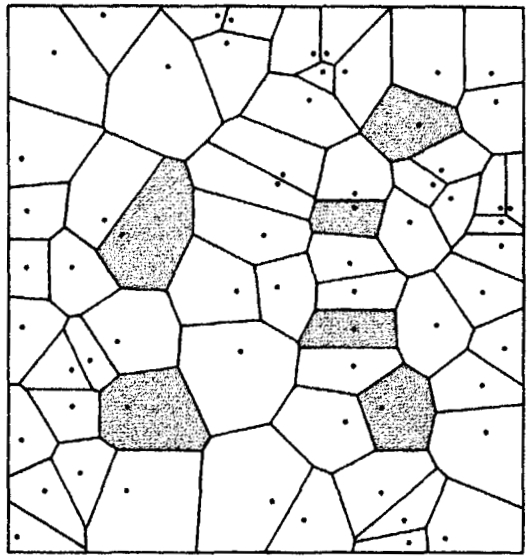

Figure 8.1.1 Voronoi diagram of Japanese black pine saplings (shaded polygons are those used in calculating U_k in equation (8.1.1)). (Redrawn from data in Upton and Fingleton, 1985, Figures 1.39a and 1.42a.)

larity in some point patterns encountered in metallography. Krawietz and Lorz (1991) also explore tests using moments of N and find that a variance test is the most powerful of these in general, although a skewness test is more sensitive in detecting regularity.

When, as is usually the case initially, the theoretical pattern we are comparing with the empirical one is CSR, the values for the polygons are those reported for the Poisson Voronoi cells (PVCs) in Section 5.5. Recall that these polygons comprise the Voronoi tessellation defined for points located according to a homogeneous Poisson point process, Θ_P. For PVCs the distribution of P, the perimeter of a polygon, was estimated by Hinde and Miles (1980) to be approximately normal with mean $4\lambda^{-1/2}$ and variance $0.9472\lambda^{-1}$, where λ is the intensity of Θ_P (see Table 5.5.1). This led Upton and Fingleton (1985, p. 99) to propose the statistic U_k as a test of CSR where

$$U_k = \frac{1}{0.9472}\hat{\lambda}\sum_{i=1}^{k}\left(P_i - 4\hat{\lambda}^{-1/2}\right)^2 \tag{8.1.1}$$

and $P_1, P_2, \ldots, P_k$, are the perimeters of a sample of k of the n polygons of $\mathcal{V}(P)$ and $\hat{\lambda} = n/|B|$ (where $|B|$ is the area of B) is the sample estimate of λ. When the k polygons are selected following certain conditions, the distribution of U_k for CSR will be approximately χ^2 with k degrees of freedom. These conditions involve avoiding both edge effects as discussed above and dependence between the k polygons. As noted by Boots and Murdoch (1983) (see equation (5.5.19)) the lengths of the perimeters of neighbouring PVCs

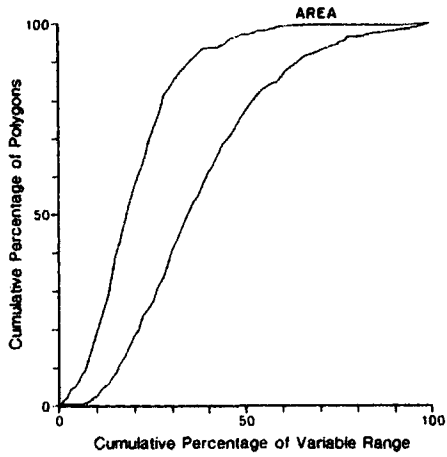

Figure 8.1.2 Standardized cumulative frequency histogram for the areas of Voronoi polygons of points in a CSR pattern. The curves are the limits to an envelope describing the standardized cumulative frequency histograms for a family of simulations of 500 points. (*Source*: Hutchings and Discombe, 1986, Figure 2.)

are not independent and so the k polygons must be selected so that no two are adjacent.

To illustrate this approach consider Figure 8.1.1, which shows the Voronoi diagram for a set of 65 Japanese black pine saplings (Upton and Fingleton, 1985, Figures 1.39a and 1.42a). In this example $k = 6$ (the shaded polygons of Figure 8.1.1) and U_k is calculated to be 4.17. Since $\chi^2_{0.05,6} = 12.59 > 4.17$, the pattern of pines is not considered significantly different from CSR.

Because the dependence between the lengths of the perimeters of neighbouring PVCs also applies to other polygon characteristics such as area and number of sides (see equations (5.5.18) and (5.5.19), respectively), Hutchings and Discombe (1986) propose an analytical approach using Monte Carlo simulation tests. This involves generating a number of patterns, typically 100, using a known spatial point process (e.g. the homogeneous Poisson point process which produces CSR) containing the same number of points as the empirical pattern being investigated. The Voronoi diagram is generated for each simulated pattern and the values of the desired characteristics of all polygons not subject to edge effects (as defined above) are measured. These values are then formed into a cumulative percentage frequency histogram in which the class intervals are set equal to 1% of the range between the smallest and largest values in the set of measurements. The use of percentages avoids the problem of different numbers of polygons in each of the simulations. A curve is produced by plotting the histogram class interval against the cumulative percentage frequency. If the curves for all simulations are superimposed, an envelope of values is created which contains all the values from the simulated patterns (see Figure 8.1.2). A similar frequency curve can be generated for the values of the polygons of the Voronoi diagram

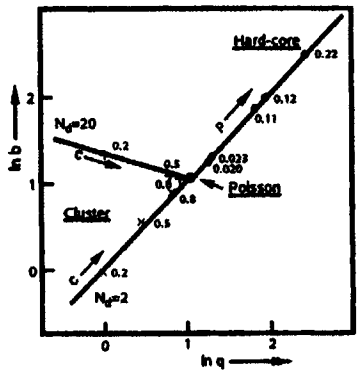

Figure 8.1.3 Values of b and q of gamma distributions fitted to the density functions of areas of Voronoi polygons for sets of points generated by different point processes. For cluster processes, N_{cl} is the mean number of points per cluster and c characterizes the distance of the points from the cluster centre, which is inversely proportional to c. For hard core proceses, p characterizes the degree of order which increases with increasing p. (*Source*: Hermann *et al.*, 1989, Figure 4(c).)

of the empirical pattern. If this empirical curve lies entirely within the envelope of simulated values, the hypothesis that the process which generated the simulated patterns is also responsible for the empirical one is not rejected.

Using this approach Hutchings and Discombe (1986) suggest that the best single polygon characteristic for distinguishing both regular and clustered empirical patterns from a hypothesized pattern of CSR is area, while perimeter is sensitive to clustering only. In view of this conjecture it is not surprising that recent work has emphasized examining distributions of polygon areas (Shehata and Boyd, 1988; Hermann *et al.*, 1989; Burger *et al.*, 1990; Byers, 1992, 1996; Duyckaerts *et al.*, 1994; Horizoe *et al.*, 1995: Itoh *et al.*, 1995; Stone and Tsakiropoulos, 1995; Bertram and Wendrock, 1996).

Recall from Section 5.5.4 that the distribution of areas for PVCs follows a two parameter gamma distribution (see equation (5.5.33)). The same distribution has also been found to fit Voronoi polygon area distributions in a number of empirical situations (e.g. Zaninetti, 1991b, 1992; Gotoh, 1993; Lemaître *et al.*, 1993; Le Caër and Delannay, 1993b; Itoh *et al.*, 1995; Mulheran and Blackman, 1996). This finding led Hermann *et al.* (1989) to suggest that the estimates of b and q in equation (5.5.33) can be used to define a two dimensional parameter space which can be used to classify patterns. As Figure 8.1.3 shows, clustered and regular (hard core) patterns each occupy distinct regions of the parameter space. Since both Maxwell and lognormal distributions have also been shown to provide good fits to Voronoi polygon areas, Hermann *et al.* (1989) suggest that other two-dimensional parameter spaces created by using the estimated parameters of these distributions or the values of the extremes of the distributions (the 5th and 95th quantiles) may also prove useful.

Marcelpoil and Usson (1992) adopt a similar approach but instead create a three-dimensional graph by using a measure of the variation in cell areas

$$AD = 1 - (1 + \sigma_A / \mu_A)^{-1}, \tag{8.1.2}$$

where μ_A and σ_A are the mean and standard deviation of the polygon areas, respectively, together with two measures which use the 'roundness factor',

$RF_i = 4\pi A_i / L_i^2$, where L_i and A_i are the perimeter and area of polygon i, respectively. These measures are the average roundness factor

$$RF_{AV} = \frac{1}{N}\sum_{i=1}^{N} RF_i \tag{8.1.3}$$

and the roundness factor homogeneity

$$RFH = \left(1 + \sigma_{RF} / RF_{AV}\right)^{-1}, \tag{8.1.4}$$

where σ_{RF} is the standard deviation of RF. All three of these measures are defined on the interval [0, 1]. Bertin *et al.* (1992) propose parallel measures for points located in $\mathbb{R}^3$ using cell volume and 'sphericity'.

Despite the number of analyses using Voronoi polygon areas, there has been little investigation of the statistical power of the procedures involved relative to each other or to other analytical procedures. Myles *et al.* (1995) examine two statistics which use empirical distribution functions. These are

$$H_M = \max_{0 \le r \le r_0} |\hat{H}(r) - \hat{E}[\hat{H}(r)]| \tag{8.1.5}$$

and

$$H_I = \int_0^r \{\hat{H}(r) - \hat{\mathrm{E}}[\hat{H}(r)]\}^2 \, \mathrm{d}r, \tag{8.1.6}$$

where $\hat{H}(r) = U(r)/n$, $U(r)$ is the total number of Voronoi polygons with areas less than r and $\hat{\mathrm{E}}[\hat{H}(\mathrm{r})]$ is the estimated value for a CSR pattern. Their results suggest that H_I is almost always the more powerful of the two statistics against clustering while neither test is better than the other with respect to regularity. When other analytical procedures are considered, they find that against clustering quadrat tests perform best, Voronoi and second order tests are essentially equal, and all three are considerably better than nearest neighbour. However, against regularity it is the second order test which is best, followed in order by the nearest neighbour, Voronoi, and quadrat tests.

Wallet and Dussert (1997) examined the three parameters AD, RF_{AV} and RFH described in equations (8.1.2)–(8.1.4) and found that AD performed best against both clustering and regularity. Relative to other analytical procedures considered (nearest analysis, second-order, quadrat, and edge lengths of minimum spanning trees), the Voronoi measures were again found to be more successful against clustering than regularity.

8.1.2 Indirect approaches

In addition to their direct use, the characteristics of $\mathcal{V}(P)$ may be exploited to implement other procedures for analysing point patterns. This subsection describes two such procedures, one based on information theory and the other on rearranging the points.

Information theory attempts to measure the degree of organization in a given system. Shannon (1948a, b) applied this idea to a set of probabilities by measuring organization using the entropy statistic, H, given by

$$H = \sum_{i=1}^{n} p_i \ln \frac{1}{p_i}, \tag{8.1.7}$$

where p_i is the probability of a random variable, x, taking the value x_i ($i = 1, \ldots, n$) and

$$\sum_{i=1}^{n} p_i = 1 \quad (p_i \geq 0). \tag{8.1.8}$$

Originally, in communications theory, logarithms to the base 2 were used, but in applications elsewhere these have been replaced by natural logarithms to the base e, as in equation (8.1.7).

If the system being investigated is an empirical point pattern consisting of a set $P = \{p_1, p_2, \ldots, p_n\}$ of n distinct points in a bounded region, B, of $\mathbb{R}^m$, one way the level of organization will be reflected is by the variation in the frequency with which points occur in subregions of B. To operationalize this concept in $\mathbb{R}^2$, Chapman (1970) defines the Voronoi diagram of P, $\mathcal{V}(P)$, and sets $p_i = A_i / \sum_{i=1}^{n} A_i$ ($i = 1, \ldots, n$), where A_i is the area of the polygon associated with p_i. Edge effects are avoided by disregarding any member of P which does not meet the criterion described at the beginning of Section 8.1.1.

In fact, H is more properly considered as a measure of disorder since its maximum value, $H_{\max}$, occurs when there is no information other than n and the area of B, $|B|$. In this case, $\hat{A}_i$, the best estimate of A_i for any member p_i of P, would be $\hat{A}_i = |B|/n$. This would mean that $\hat{A}_i$ would equal $1/n$ for each of the n points so that

$$H_{\max} = n\left(\frac{1}{n} \ln \frac{1}{1/n}\right) = \ln n. \tag{8.1.9}$$

However, there are problems in using H directly to evaluate empirical point patterns (Chapman, 1970, pp. 320–321; Lenz, 1979, pp. 376–377). In particular, although H decreases as the extent of clustering in the pattern increases, maximum regularity (members of P located on a triangular grid, see Figure 8.2.2) also yields $H_{\max}$. Also as n gets very large the value of H for CSR also approaches $H_{\max}$ thus making the interpretation of H for an empirical point pattern almost impossible. In addition, since $\ln n$ increases as n increases, patterns generated by the same process but with different intensities will not produce the same value of H.

These problems with H can be overcome by using Thiel's redundancy measure, R^*, given by

$$R^* = \sum_{i=1}^{n} p_i \ln \frac{p_i}{1/n} = H_{\max} - H. \tag{8.1.10}$$

For a regular pattern, $R^* = 0$, while R^* increases with increasing clustering of points. In addition, R^* is less sensitive to variations in n.

If we wish to test an empirical point pattern against a hypothesis of CSR we need both the expected value, $E(R^*)$, and the variance, $\mathrm{Var}(R^*)$, of R^* for CSR. Recall that the area A of a Poisson Voronoi cell (PVC) has a gamma distribution,

Table 8.1.1 Estimates of the expected value and variance of Thiel's redundancy measure R^* for CSR.

n	$E(R^*)$	$Var(R^*)$
15	0.124	0.001875
30	0.129	0.001225
60	0.131	0.000660
90	0.132	0.000412
180	0.132	0.000204
360	0.132	0.000090
∞	0.134	0

Source: Lenz (1979).

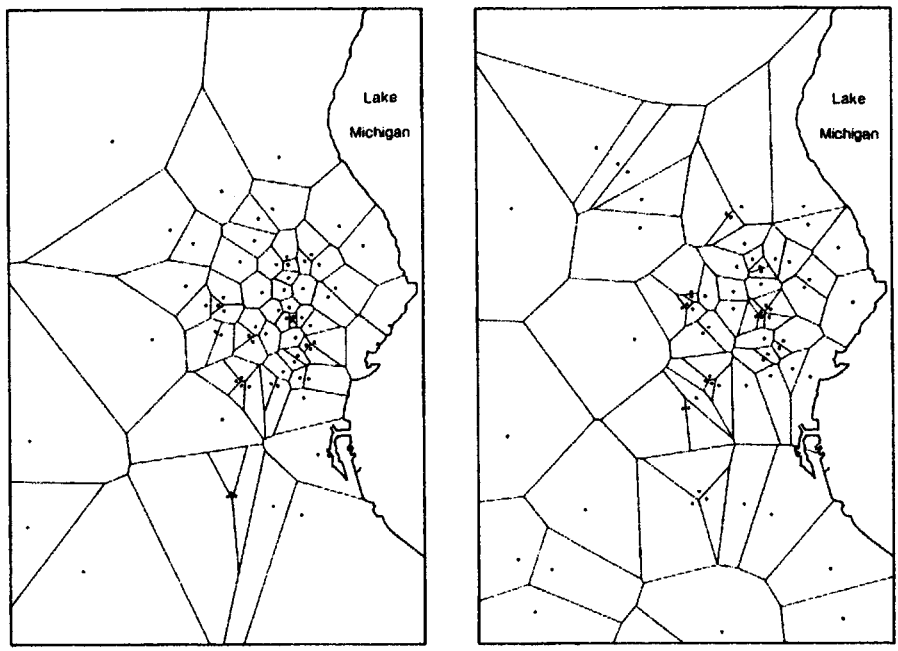

Figure 8.1.4 City of Milwaukee, 1975: (a)Voronoi diagram of locations of personal robberies; (b)Voronoi diagram of locations of victims' addresses (*Source*: Lenz, 1979, Figures 3 and 4.)

$$f(A,\alpha) = [\alpha(\alpha A)^{\alpha-1} e^{-\alpha A}] / \Gamma(\alpha), \tag{8.1.11}$$

for which $\alpha = 3.57$ if the variance of $A = 0.280\lambda^{-2}$ The best fit to the data in Table 8.1.1 yields

$$\text{Var}(R^*) = 0.2157 / \sqrt[1.9]{n}, \tag{8.1.12}$$

which can be used to estimate $\text{Var}(R^*)$ for values of n not given in Table 8.1.1.

Lenz (1979) shows that when $n \geq 15$, R^* is approximately normally distributed for CSR so that we may use the variate of the standard normal curve

$$z = [R^*_{\text{obs}} - \text{E}(R^*)] / \text{Var}(R^*)^{1/2}, \tag{8.1.13}$$

where R^*_{obs} is the value of R^* for the empirical point pattern, as a test statistic.

Since R^* is independent of n we can also use it to compare two empirical point patterns, for example of the same phenomenon at different times or different phenomena in the same region. An example is provided by Lenz who compares the pattern of personal robbery locations in the city of Milwaukee in 1975 with the pattern of victims' addresses (see Figure 8.1.4). Values of R^*_{obs} indicate that the former pattern shows more clustering than the latter, which leads Lenz to conclude that victims contribute to the occurrence of personal robberies by exposing themselves to high crime risk districts of the city.

Another set of procedures, based on rearranging the points, has been developed by Perry (1995). In summary, this involves moving the points of the pattern simultaneously until the pattern converges on a regular one and then comparing the locations of the points in the initial and final patterns. The procedure involves four steps. The first step is to construct $\mathcal{V}(P)$. In step two, each point whose Voronoi polygon does not include part of the boundary ∂B of the study area B is relocated to a new position which is the centroid of the locations of its Voronoi neighbours weighted by the lengths of the shared Voronoi polygon edges. Thus, if the initial coordinates of the q neighbours of a point p_i are (x_j, y_j), $j = 1, \ldots, q$, and the length of the common edge between $V(p_i)$ and $V(p_j)$ is l_j, the new coordinates for p_i are given by

$$\left(\frac{\sum_{j=1}^{q} l_j x_j}{\sum_{j=1}^{q} l_j}, \frac{\sum_{j=1}^{q} l_j y_j}{\sum_{j=1}^{q} l_j} \right). \tag{8.1.14}$$

When one or more of the edges of the Voronoi polygon of a point are parts of ∂B, a temporary, imaginary point is located at the mid-point of each boundary edge and assigned a weight of $4/\sqrt{3}$ times the edge length. This value is derived from the situation where the points are located on a triangular grid (see Figure 8.2.2). The imaginary points are removed once the point has been re-located. Step three is to construct a new Voronoi diagram using the new set of point locations. The final step is to assess if the new

pattern of points has achieved a sufficient degree of regularity. If the variation in the areas of the Voronoi polygons for the current pattern is smaller than that of the previous pattern and differs from it by less than some specified value, the process ends; if not, steps two through four are repeated.

If d_i is the Euclidean distance between the initial and final locations of p_i, Perry proposes Monte Carlo test procedures involving either the empirical distribution function of d_i or

$$D = \sum_{i=1}^{n} d_i. \tag{8.1.15}$$

By applying these tests to some of the same data sets analysed by Diggle (1983), Perry suggests that they have power which is comparable with that of quadrat, nearest neighbour and second-order tests, especially for detecting clustering.

8.2 TRIANGLE-BASED METHODS

If an empirical point pattern consists of a set $P = \{p_1, p_2, \ldots, p_n\}$ of n distinct points in a bounded region, B, of $\mathbb{R}^m$, this approach involves defining the Delaunay tessellation of P, $\mathcal{D}(P)$, measuring properties of the polytopes of $\mathcal{D}(P)$ and comparing these with the properties of the polytopes of a Delaunay tessellation defined for a hypothesized point pattern. As with all areas of point pattern analysis, the hypothesized pattern which has been emphasized is CSR. Such patterns correspond to those produced when P is generated according to a homogeneous Poisson point process (see Section 1.3.3). Known results relating to this model are reported in Section 5.5 and several of them are used here.

In $\mathbb{R}^2$ the probability density function (pdf), $f(\alpha)$, of α, a random angle of a randomly selected triangle of $\mathcal{D}(P)$ when P is CSR, was derived by Miles (1970a) (see equation (5.11.12)). Boots (1974, 1975b) and later Vincent *et al.* (1976, 1977, 1983) argued that a test for CSR can be obtained by integrating equation (5.11.12) for chosen class intervals to obtain a set of expected frequencies which can be compared with the empirical frequencies using a Pearson goodness-of-fit test. The probability $F_\alpha(x)$ that $\alpha \leq x$ derived by integrating equation (5.11.12) is

$$\begin{aligned} F_\alpha(x) &= \int_0^x f(\alpha)\, d\alpha \\ &= (1/3)\left[2\sin^2\alpha + \{\alpha\cos 2\alpha - (3\sin 2\alpha)/2 + 2\alpha\}\pi^{-1}\right]_0^x. \end{aligned} \tag{8.2.1}$$

Both Boots (1975b) and Vincent *et al.* (1977) use this procedure to analyse patterns of urban settlements in various parts of the United States and demonstrate that, in this context, the procedure is more effective than the nearest neighbour approach (see Section 8.1) in correctly rejecting a null hypothesis of CSR.

However, since this approach requires that we take a random sample of triangles from $\mathcal{D}(P)$, Mardia *et al.* (1977) have argued that it is wasteful of the available information. They propose instead using the marginal density of the minimum angle, α_{min}, of all triangles in $\mathcal{D}(P)$. The pdf of α_{min} for CSR is given in equation (5.11.13). Mardia *et al.* assume that the values of α_{min} are independent for the triangles of $\mathcal{D}(P)$. Integrating equation (5.11.13) we get

$$F_{\alpha_{min}}(x) = \left[1 + \tfrac{1}{2}\pi\left\{(6\alpha_{min} - 2\pi)\cos 2\alpha_{min} - \sin 2\alpha_{min} - \sin 4\alpha_{min}\right\}\right] \qquad (0 < x \leq \pi/3) \tag{8.2.2}$$

as the probability that $\alpha_{min} \leq x$.

Boots (1986) has proposed that a similar argument may also be made in favour of the use of the maximum angle, α_{max}, of the triangles of $\mathcal{D}(P)$. When P is CSR the marginal density of α_{max} is given by equation (5.11.14) which yields

$$F_{\alpha_{max}}(x) = \tfrac{1}{2}\pi\left[\sin 4x + \sin 2x + 2\pi\cos x - 6x\cos 2x\right] \qquad (\pi/3 \leq x \leq \pi/2)$$

and

$$F_{\alpha_{max}}(x) = \tfrac{1}{2}\pi\left[2x\cos 2x - 2\pi\cos 2x - 3\sin 2x + 4x - 2\pi\right] \qquad (\pi/2 \leq x \leq \pi) \tag{8.2.3}$$

as the probability that $\alpha_{max} \leq x$.

Preliminary investigations by Boots (1986) using both empirical and simulated patterns containing different degrees of regularity or clustering suggest that α_{min} has the best overall ability to reject a null hypothesis of CSR against both regular and clustered alternatives, although there is some indication that α_{max} is able to detect instances of clustering not discernible by α_{min}. In view of these findings we illustrate this procedure using α_{min} for a pattern of settlements in an area of southeastern Montana, USA, in 1973 (see Figure 8.2.1).

For each triangle in $\mathcal{D}(P)$ in Figure 8.2.1 we identify and measure α_{min}. These values are used to determine the number of values, $G(\alpha_{min})$, of α_{min} in a specified interval. In calculating $G(\alpha_{min})$ we must account for edge effects produced by the bounded nature of $\mathcal{D}(P)$. Mardia (1989) (see also the discussion in Kendall, 1989, p. 110) proposes doing this by discounting any triangle for which the circumcentre does not lie in B (recall from Section 2.3 (Property V7) that the circumcentres are vertices of the Voronoi diagram of P). This procedure was followed in this example and such discounted triangles are shaded in Figure 8.2.1. The values of $G(\alpha_{min})$ so obtained are compared with the corresponding values, $F(\alpha_{min})$, for CSR obtained from equation (8.2.2) using a χ^2 goodness-of-fit test (see Table 8.2.1). Following usual statistical convention, categories for which the expected values are less than five are combined with adjacent categories. Since the value of $\chi^2 = 11.198$ which results has a probability of less than 0.05 of occurring for a χ^2_4 distribution, we reject the hypothesis of CSR for this settlement pattern.

Table 8.2.1 Analysis of locations of settlements in southeastern Montana using the minimum angle technique.

Minimum angle, α_{min} (degrees)	Observed frequency, $G(\alpha_{min})$		Expected frequency, $F(\alpha_{min})$	
0–10	1	2	2.468	9.441
11–20	1		6.973	
21–30	15		10.007	
31–40	8		10.521	
41–50	15	16	8.023	11.031
51–60	1		3.008	

$\chi^2_4 = 11.198$.

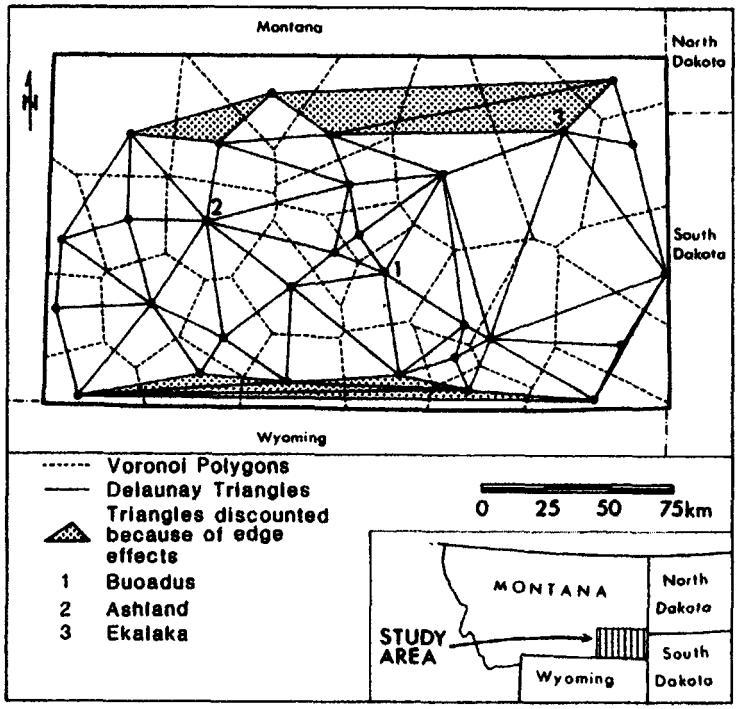

Figure 8.2.1 Voronoi diagram and Delaunay triangulation for a pattern of settlements in an area of Montana, USA, in 1973. Re-drawn from data in Boots and Getis (1988, Figure 4.3).

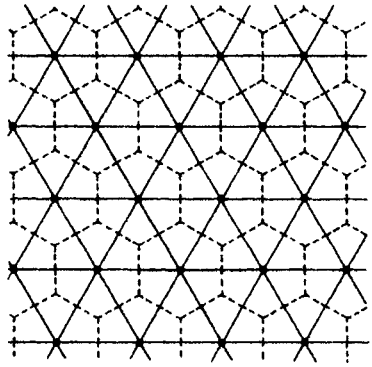

Figure 8.2.2 Voronoi polygons (indicated by dashed lines) and Delaunay triangles (indicated by solid lines) for points located on a triangular grid.

However, can we proceed further beyond this result? Currently, the answer is a tentative yes. If the points of P are arranged in a perfectly regular way on a triangular grid such as that shown in Figure 8.2.2, all the Voronoi polygons (indicated by dashed lines) will be regular hexagons and all the Delaunay triangles in $\mathcal{D}(P)$ (indicated by solid lines) will be equilateral so that all angles would be equal to $\pi/3$. Thus, if for an empirical point pattern there is an excess of values (relative to the number expected for CSR) at the upper tail of $G(\alpha_{min})$, a regular pattern is indicated. However, in Section 5.11 we noted that there will also be a high proportion of nearly equilateral triangles in $\mathcal{D}(P)$ under CSR so that the excess referred to might prove difficult to detect. If the points are located on a square grid (see Figure 8.2.3) the resulting Voronoi polygons (indicated by dashed lines) will be squares and the Delaunay triangles of $\mathcal{D}(P)$ (indicated by the solid lines) will be right-angled so that $\alpha_{min} = \pi/4$. A significant excess of such angles would indicate a point pattern with affinities to a square grid. In contrast, a pattern in which points are clustered in distinct groups will have some edges of $\mathcal{D}(P)$ which correspond to links between points on the peripheries of different clusters (see Figure 8.2.4). Many of the triangles formed by such edges will be elongated so that their minimum angles will be small. Thus, a distribution of empirical values of $\alpha_{min}$with a significant excess of small angles is suggestive of a pattern of clustered points.

On the basis of these speculations and noting that the greatest differences between the frequencies for the observed angles and the CSR expectations occur in the ranges of 0–20 degrees (where there is a deficit of minimum angles relative to CSR) and 41–50 degrees (where there is an excess) (see Table 8.2.1), we may conclude that the pattern of settlements in Figure 8.2.1 shows a significant degree of regularity. Furthermore, it appears that this regularity is more consistent with the expectations of a regular square grid than with a triangular one.

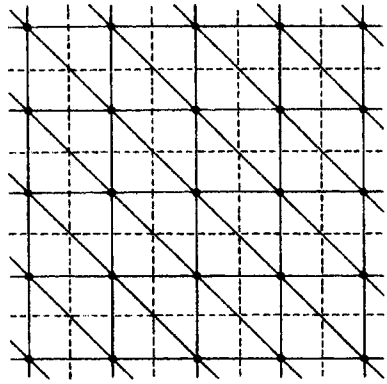

Figure 8.2.3 Voronoi polygons (indicated by dashed lines) and Delaunay triangles (indicated by solid lines) for points located on a square grid.

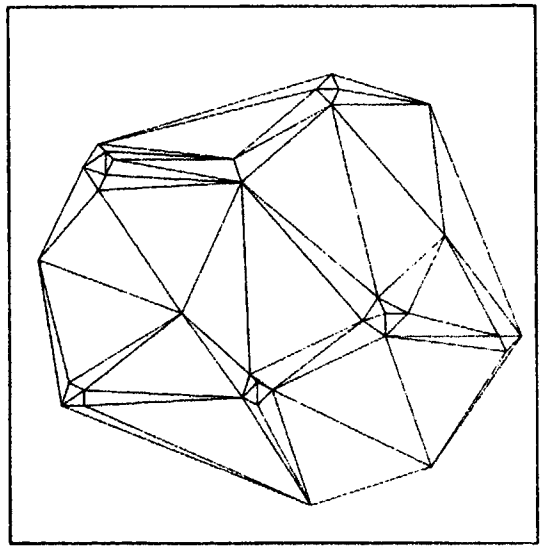

Figure 8.2.4 Delaunay triangulation for a clustered pattern of points.

An alternative to considering individual angles of triangles in $\mathfrak{D}(P)$ is to focus on triangle shape. Mardia (1989) has proposed a test which involves comparing the positions of the triangles of $\mathfrak{D}(P)$ in Kendall's shape space (see Section 5.11 and Figure 5.11.12) with the positions of those expected under CSR. When testing if an empirical point pattern is significantly different from a regular one, Taylor *et al.* (1995) and Dryden *et al.* (1995, 1997) define test statistics that incorporate the Procrustes shape distance from each triangle in $\mathfrak{D}(P)$ to an equilateral triangle. An example of such a

measure is the Riemannian shape distance $(\cos\rho)^2$ between two triangles $((x_i, y_i)$ and $(\alpha_i, \beta_i), i = 1, 2, 3)$, where

$$(\cos\rho)^2 = \frac{(S_{x\alpha} + S_{y\beta})^2 + (S_{x\beta} - S_{y\alpha})^2}{(S_{xx} + S_{yy})(S_{\alpha\alpha} + S_{\beta\beta})}, \tag{8.2.4}$$

where $S_{xy} = \sum_{i=1}^{3}(x_i - \bar{x})(y_i - \bar{y}), \bar{x} = (\sum_{i=1}^{3} x_i)/3$ and $\bar{y} = (\sum_{i=1}^{3} y_i)/3$.

Other characteristics of triangles can also be examined. So far most attention has been paid to N the number of edges incident at a vertex in $\mathcal{D}(P)$ and L, the length of an edge of $\mathcal{D}(P)$ (Vincent *et al.*, 1976, 1977, 1983; Pryz, 1994; Bertram and Wendrock, 1996). Because of the duality between $\mathcal{D}(P)$ and the ordinary Voronoi diagram of P, N is also the number of edges (vertices) of the Voronoi polygon associated with the vertex in $\mathcal{D}(P)$ and L is the distance between a pair of Voronoi neighbours.

For CSR the pdf of L, $f(L)$, is given by equation (5.11.3), while pdfs for other processes whose realizations are either clustered (Pyrz, 1994; Bertram and Wendrock, 1996) or regular patterns (Vincent and Howarth, 1982; Pyrz, 1994; Bertram and Wendrock, 1996) have been approximated using Monte Carlo procedures. Using the Monte Carlo simulation test procedure described in Section 8.1.1, Hutchings and Discombe (1986) suggest that L is effective in distinguishing both regular and clustered patterns from those of CSR. Bertram and Wendrock (1996) reach a similar conclusion. They also consider a local average of L for each point p_i given by

$$L_i = \frac{1}{N_i}\sum_{j=1}^{N_i} L_{ij}, \tag{8.2.5}$$

where L_{ij} is the length of the edge between point i and point j, and N_i is the number of edges incident at point p_i, and demonstrate that the coefficient of variation (cv) of L_i, $\mathrm{cv}(L_i)$, can be used together with $\mathrm{cv}(N)$ to create a two-dimensional parameter space in which different patterns are shown to occupy different locations. L_i has also been used in empirical investigations of metals (Wray *et al.*, 1983; Spitzig *et al.*, 1985; Shehata and Boyd, 1988; Burger *et al.*, 1990; Stone and Tsakiropoulos, 1995) and unidirectional composites (Pyrz, 1994) where it is usually referred to as the *near neighbour distance*.

Finney (1991, 1993) also uses L to construct a classification for individual cells in three-dimensional simplexes. Depending on its length relative to a specified threshold, each edge is designated as either short or long. Since each simplicial cell has six edges, it can be assigned to one of seven classes depending on the number of its short and long edges. The complete structure is summarized by the frequency distribution of the classes and by a table showing the adjacencies between cells of a given class. Such summary measures can then be compared with model expectations.

Rather than considering individual characteristics of the triangles of $\mathcal{D}(P)$ independently, both Mardia *et al.* (1977) and Kendall (1989) propose pursuing them jointly. For CSR a lemma due to Miles (1970a) (see equation (5.11.2)) demonstrates that the size (measured by the circumradius, r) and the shape

(measured by the joint orientation of the edges) of an individual triangle are independent with the distribution of both characteristics being known. Using this information Kendall (1989) suggests the following test. First, sort the triangles of $\mathcal{D}(P)$ in terms of their shapes into three groups: those that are 'nearly equilateral', those that are highly 'splinter shaped', and the remainder. Then for the first and last of these groups, examine the departures from independence and from the known distribution of r.

8.3 NEAREST NEIGHBOUR DISTANCE METHODS

In this section we show the nearest neighbour distance method, abbreviated to the NND method. The NND method is a statistical method which examines the effect of physical objects located in a region upon the distribution of point-like elements over the region. The effect of nuclear power stations on the distribution of abnormal plants, the locational dependency of fast food stores on arterial streets, the spatial association of firms producing integrated circuits around airports, etc. can be investigated with the NND method. We should note that the NND method discussed here is different from the ordinary nearest neighbour distance method referred to in biology, geography and ecology (Bartlett, 1975; Pielou, 1977; Getis and Boots, 1978). The ordinary nearest neighbour method also deals with the distribution of point-like elements but it does not deal directly with physical objects affecting the distribution of the point-like elements. In this section we focus on the NND method because this method is an application of the Voronoi diagram which is the major concern of this text. The reader who wishes to study the ordinary nearest neighbour distance method in depth should consult Diggle (1983), Upton and Fingleton (1985) and Cressie (1991).

To fix a general setting for the analysis, let us consider a region in which m physical objects (such as power stations, arterial streets and airports referred to above) are located in a region, and n point-like elements (such as plants, stores and firms) are distributed over the region excluding the area occupied by the objects. A set of the m objects in the region is represented by a set of points, lines or areas in a bounded region S in $\mathbb{R}^2$, which is denoted by $O = \{O_1, \ldots, O_m\}$ $(1 \leq m < \infty)$, where $O_i \cap O_j = \emptyset, i \neq j, i, j \in I_n$. A set of the n point-like elements is represented by a set of points distributed over $S_0 = S \backslash \cup_{i=1}^{m} O_i$, which is denoted by $P = \{p_1, \ldots, p_n\}$ $(1 \leq n < \infty)$ with the set of their location vectors $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$.

Given O and P, we define the distance from a point p_i in $\mathbb{R}^2$ to an object O_j by

$$d(p_i, O_j) = \min_{\boldsymbol{u}_j} \{ \|\boldsymbol{x}_i - \boldsymbol{u}_j\| \mid \boldsymbol{u}_j \in O_j \}, \tag{8.3.1}$$

where $\|\boldsymbol{x}_i - \boldsymbol{u}_j\|$ is the Euclidean distance between $\boldsymbol{x}_i \in S_0$ and $\boldsymbol{u}_j \in O_j$. With $d(p_i, O_j)$, we next define the distance from a point p_i to the nearest object in O by

$$d(p_i, O) = \min_j \{d(p_i, O_j) \mid j \in I_n\}. \tag{8.3.2}$$

We call this distance the *nearest neighbour distance*, or briefly the *NN-distance* from p_i to O. Third, we define the NN-distance averaged across the n points, i.e.

$$\bar{t} = \frac{1}{n} \sum_{i=1}^{n} d(p_i, O). \tag{8.3.3}$$

We call $\bar{t}$ the *average nearest neighbour distance* or the *average NN-distance* for short.

In the NND method, several indices are proposed to measure or to test the effect of objects O on the distribution of points P. To formulate one of the simplest indices, let us consider the hypothesis that points P are independently and randomly distributed according to the uniform distribution over S_0. This hypothesis implies that objects O do not affect the distribution of points P. Under this hypothesis, the NN-distance $t_i = d(p_i, O)$ is a random variable; consequently $\bar{t}$ of equation (8.3.3) is also a random variable. We define an index by

$$R = \frac{\bar{t}}{\mathrm{E}(\bar{t})}, \tag{8.3.4}$$

where $\mathrm{E}(\bar{t})$ is the expected value of $\bar{t}$ (cf. Pielou's index, 1977, p.155). Since $\bar{t}$ is the average of independent random variables having the same distribution with a finite mean, the central limit theorem guarantees that the distribution of $\bar{t}$ asymptotically approaches the normal distribution with mean $\mathrm{E}(\bar{t}) = \mathrm{E}(t)$ and variance $\mathrm{Var}(\bar{t}) = \mathrm{Var}(t)/n$, where $t = d(p, O)$ is the NN-distance from a random point p to O. It follows from this property that the random variable R asymptotically follows the normal distribution with $\mathrm{E}(R) = 1$ and $\mathrm{Var}(R) = \mathrm{Var}(t)/(n\,\mathrm{E}(t)^2)$ as n increases. With this test statistic we may conclude that if the observed value $\hat{R}$ is less than a lower critical value of R, then the points P are more closely distributed around objects O than they would be in the random distribution. If $\hat{R}$ is greater than an upper critical value of R, we may infer that the points P are more sparsely distributed around objects O than they would be in the random distribution.

We should make one remark on the boundary effect often discussed in the related literature. The nature of the boundary effect of the NND method is different from that of the ordinary nearest neighbour method. Since the NND method is based upon a uniform distribution over a bounded region and a random NN-distance t_i is a distance from a random point to a fixed object (not a random object), in theory the NND method is not bothered by the boundary effect. In practice, the boundary effect may appear if we apply the NND method to a subregion of a region in which points P are distributed, because we usually ignore objects outside the subregion that affect the distribution of P. In this case we can overcome the boundary effect by considering the objects outside the subregion whose Voronoi regions (generated by O) overlap the subregion.

When we use the index R in practice, we have to compute $d(p_i, O)$, $E(\bar{t}) = E(t)$ and $Var(\bar{t}) = Var(t)/n$. The latter two values are obtained from the probability distribution function, $F(t)$, of t. Since a point is placed according to the two-dimensional uniform distribution over S_0, the probability of the point being placed in a subregion in S_0 is given by the ratio of the area of the subregion to that of the whole region S_0 (recall Section 1.3.3). As a subregion, we consider the region $S(t)$ in which the NN-distance from any point p in $S(t)$ to O is less than or equal to t, i.e.

$$S(t) = \{p \mid d(p, O) \leq t,\ p \in S_0\}. \tag{8.3.5}$$

We call this region the *buffer zone* of objects O with distance t. Since the value of $F(t)$ is the probability that the NN-distance from a random point to O is less than or equal to t, $F(t)$ is given by

$$F(t) = \frac{|S(t)|}{|S_0|}, \tag{8.3.6}$$

where $|S(t)|$ indicates the area of $S(t)$. At first glance, the exact form of the function $|S(t)|$ or $F(t)$ is difficult to obtain. Fortunately, however, if we use a Voronoi diagram, we can exactly obtain the function $|S(t)|$. Since the computational method differs according to the shape of objects O, we deal with the NND methods for point-like objects, line-like objects, and area-like objects, separately.

8.3.1 Nearest neighbour distance method for point-like objects

When the area occupied by an object O_i is very small relative to the whole region S_0, we may regard O_i as a point, and O as a set of points. We call the NND method for this set the *NND method for point-like objects*. Using this method, we may examine the effect of point-like objects O on the distribution of point-like elements P. For example, we may examine the effect of firms (O) polluting the air with smoke on the distribution of patients (P) suffering from asthma, and the association of news stands (P) around stations (O). An actual application was carried out by Okabe and Miki (1984) who examined the effect of railway stations (O) on the distribution of several kinds of retail stores (P) in Toshima, Tokyo. Figure 8.3.1 shows the distribution of book stores (the filled circles) with respect to railway stations (the unfilled circles).

As we mentioned above, when we use the index R in practice, we have to calculate the value of $d(p_i, O)$ and the moments of $t = d(p, O)$, i.e. $E(t)$ and $Var(t) = E(t^2) - E(t)^2$. Recalling Problem P3 in Section 2.3, we notice that $d(p_i, O)$ can be efficiently computed using the ordinary Voronoi diagram. First, we generate the Voronoi diagram $\mathcal{V} = \{V(O_1), \ldots, V(O_n)\}$ with point-like objects O (indicated by continuous lines in Figure 8.3.1). Second, we find in which Voronoi polygon a point p_i is located. This search problem is known as the *point-location problem*, and several efficient computational methods are proposed (see Chapter 4). Once $p_i \in V(O_j)$ is known, $d(p_i, O)$

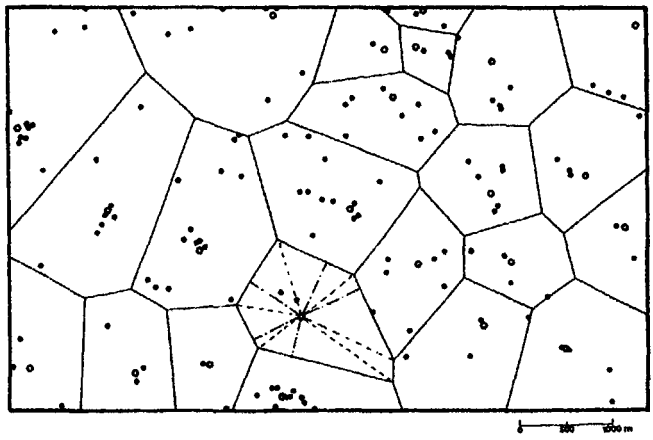

Figure 8.3.1 The location of the railway stations (the unfilled circles), the distribution of book stores (the filled circles) in Toshima and the Voronoi diagram generated by the railway stations. (*Source*: Miki, 1983, Figure 3.7.)

is readily given by $d(p_i, O_j)$. Substituting this value into equation (8.3.3), we obtain the value of $\bar{t}$ in equation (8.3.4).

The moments of the NN-distance to point-like objects (the moments of the distance from a random point to the nearest generator point) We next show the procedure for calculating the moments of t. First, we generate the Voronoi diagram $\mathcal{V}$ with point-like objects O. We next triangulate each Voronoi polygon by the line segments connecting generator O_i and Voronoi vertices (indicated by the broken lines in Figure 8.3.1). We further partition the resulting triangle by the line passing through O_i perpendicular to the facing edge if the line crosses the edge (indicated by the dash-dotted lines in Figure 8.3.1). As a result, the whole region S is partitioned into a set of obtuse or right angle triangles $S = \{S_1, \ldots, S_k\}$, where $S = S_1 \cup \ldots \cup S_k$.

Figure 8.3.2 depicts a triangle S_i in S, where the vertex O_i is placed on the v-axis and the other two vertices are placed on the u-axis. Let θ_{i1}, θ_{i2}, u_{i1}, u_{i2}, v_{io}, t_{i1} and t_{i2} be those indicated in Figure 8.3.2, respectively, and $S_i(t) = S(t) \cap S_i$. Then the probability distribution function of equation (8.3.6) is written as

$$F(t) = \frac{1}{|S_0|} \sum_{i=1}^{k} |S_i(t)|. \tag{8.3.7}$$

With the help of Figure 8.3.2, we obtain $|S_i(t)|$ as

$$|S_i(t)| = \begin{cases} \frac{1}{2} t^2 (\theta_{i2} - \theta_{i1}) & \text{for } 0 \le t \le t_{i1}, \\ \frac{1}{2} v_{i0} \left(\sqrt{t^2 - v_{i0}^2} - u_{i1} \right) + \frac{1}{2} t^2 \left(\theta_{i2} - \arccos \frac{v_{i0}}{t} \right) & \text{for } t_{i1} \le t \le t_{i2}, \\ \frac{1}{2} v_{i0} (u_{i2} - u_{i1}) & \text{for } t_{i2} \le t. \end{cases} \tag{8.3.8}$$

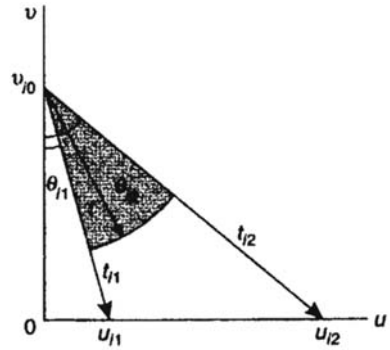

Figure 8.3.2 A triangle in the triangulated Voronoi diagram shown in Figure 8.3.1.

Thus we can explicitly obtain the probability distribution function $F(t)$ by substituting equation (8.3.8) into equation (8.3.7).

The probability density function $f(t)$ of $F(t)$ is obtained from

$$f(t)=\frac{\mathrm{d}F(t)}{\mathrm{d}t}=\frac{1}{|S_0|}\frac{\mathrm{d}|S(t)|}{\mathrm{d}t}=\frac{1}{|S_0|}\sum_{i=1}^{k}\frac{\mathrm{d}|S_i(t)|}{\mathrm{d}t}=\frac{1}{|S_0|}\sum_{i=1}^{k}s_i(t). \tag{8.3.9}$$

The rth moment of t is given by

$$\mathrm{E}(t^r)=\int_0^\infty t^r f(t)\,\mathrm{d}t=\frac{1}{|S_0|}\sum_{i=1}^{k}\int_0^\infty t^r s_i(t)\,\mathrm{d}t. \tag{8.3.10}$$

After several steps of calculation, we obtain the second integral in equation (8.3.10) as:

if r is even ($r = 2q - 2, q \geq 1$),

$$\int_0^\infty t^r s_i(t)\mathrm{d}t=\frac{(2q-2)!!}{2q\,(2q-1)!!}\sum_{j=0}^{q-1}\frac{(2j-1)!!}{(2j)!!}\times\left[u_{i2}(u_{i2}^2+v_{i0}^2)^j-u_{i1}(u_{i1}^2+v_{i0}^2)^j\right]v_{i0}^{2q-2j-1}; \tag{8.3.11}$$

if r is odd ($r = 2q - 1, q \geq 1$),

$$\int_0^\infty t^r s_i(t)\mathrm{d}t=\frac{(2q-1)!!}{(2q+1)(2q)!!}\times\left\{\sum_{j=0}^{q-1}\frac{(2j)!!}{(2j+1)!!}\left[u_{i2}(u_{i2}^2+v_{i0}^2)^{j+1/2}-u_{i1}(u_{i1}^2+v_{i0}^2)^{j+1/2}\right]v_{i0}^{2q-2j-1}+v_{i0}^{2q+1}\ln\left(\frac{\sqrt{v_{i0}^2+u_{i2}^2}+u_{i2}}{\sqrt{v_{i0}^2+u_{i1}^2}+u_{i1}}\right)\right\} \tag{8.3.12}$$

(Okabe, 1987; Yoshikawa, 1989). Substituting this equation into equation (8.3.10), we can explicitly obtain the expected value $\mathrm{E}(t)$ and the variance

$\mathrm{Var}(t) = \mathrm{E}(t^2) - \mathrm{E}(t)^2$. Okabe and Miki (1984) calculated the value of R for book stores in Figure 8.3.1 with these equations, and obtained $R = 0.727$. Since this value was smaller than unity with significance level 0.05, they rejected the null hypothesis that the book stores were independently and uniformly distributed over Toshima. The book stores tended to gather around railway stations.

Besides the NND method, the moments of the NN-distance are also used in the quantizing problem encountered in applications to analog-to-digital conversion (Conway and Sloane, 1993, Chapter 22). This problem asks how to place points in space so that the average second moment of their Voronoi polygons is as small as possible. This problem is discussed with the second moments of the the NN-distance of m-dimensional Voronoi diagrams generated by several types of lattice points. These moments are shown in Conway and Sloane (1993, Chapter 22) (see also a related discussion in Section 7.1).

8.3.2 Nearest neighbour distance method for line-like objects

When the area occupied by an object O_i is very narrow and long relative to the whole region S_0, we may regard the object as a line segment, and O as a set of line segments. We call the NND method for this set the *NND method for line-like objects*.

Applications of the NND method for line-like objects are shown by Okabe and Fujii (1984) and Okabe and Yoshikawa (1989). Figure 8.3.3 illustrates the distribution of high-class apartment buildings (the dots) and arterial streets (the continuous lines) in Kohto-Sumida, Tokyo. Okabe and Yoshikawa

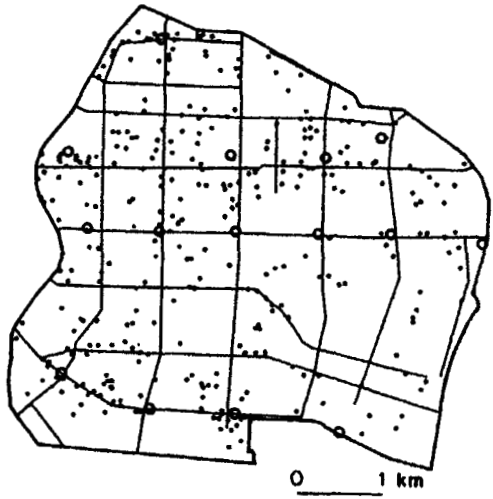

Figure 8.3.3 The distribution of high-class apartment buildings (dots) and the arterial streets (continuous lines) in Kohto-Sumida, Tokyo. (*Source*: Okabe and Yoshikawa, 1989, Figure 9.)

(1989) asked whether or not the distribution of high-class apartment buildings was affected by arterial streets. This question can be examined by the NND method for line-like objects.

To carry out this statistical examination, we have to calculate $d(p_i, O)$ and the moments of $t = d(p, O)$. First, to gain computational tractability, we assume that a set of line-like objects O is represented by a set of chains of straight line segments. A curved line is approximated by the chain of connected small straight line segments. If this modification is made, the procedure is essentially the same as that of the NND method for point-like objects. We construct the Voronoi diagram $\mathcal{V} = \{V(O_1), \ldots, V(O_n)\}$ generated by the decomposed set of the line generator set O (see Section 3.5). We next find in which Voronoi region a point p_i is placed. Once $p_i \in V(O_j)$ is known, the NN-distance $d(p_i, O)$ is given by $d(p_i, O_j)$.

The moments of the NN-distance to line-like objects (the moments of the distance from a random point to the nearest generator line) We next show the procedure for calculating the moments of t. This procedure is similar to that in Section 8.3.1. We first generate the Voronoi diagram $\mathcal{V}$. Second, when a generator is a point O_i, we join the vertices of $V(O_i)$ and O_i by line segments (the broken lines in Figure 8.3.4). Third, when a generator is a line segment O_i, we consider the line that passes through a vertex of $V(O_i)$ and is perpendicular to the generator line O_i. If this line crosses the generator line O_i other than its end points (the dash-dotted lines), we partition $V(O_i)$ by that line. As a result, S_0 is partitioned into subregions and these subregions are classified into four types, three of which are shown in Figure 8.3.5 and the other one is shown in Figure 8.3.2. Let S_{ij} be the jth subregion of type $i, j = 1, \ldots, k_i, i = 1, \ldots, 4$, and $S_{ij}(t) = S_{ij} \cap S(t)$, i.e. the buffer zone in S_{ij}. Then the probability distribution function $F(t)$ of equation (8.3.6) is written as

$$F(t) = \frac{|S(t)|}{|S_0|} = \frac{1}{|S_0|} \sum_{i=1}^{4} \sum_{j=1}^{k_i} |S_{ij}(t)|. \tag{8.3.13}$$

To give the explicit form of $S_{ij}(t)$, let u_{j1}, u_{j2}, t_{j1}, t_{j2}, a, θ_{j1} and θ_{j2} be those indicated in Figure 8.3.5, respectively. Then, $S_{ij}(t)$ is given by

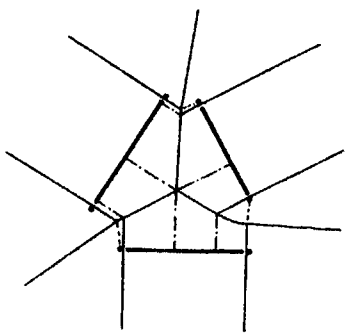

Figure 8.3.4 The partition of a Voronoi diagram generated by a set of lines for calculating the moments of t.

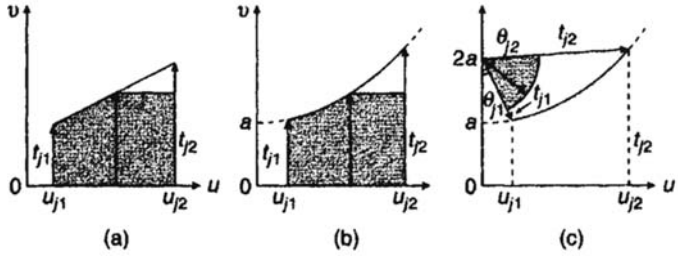

Figure 8.3.5 Four types of subregions constituting a Voronoi diagram generated by a set of line segments (three types are shown here; the fourth type is shown in Figure 8.3.2).

$$|S_{1j}(t)| = \begin{cases} (u_{j2}-u_{j1})t & \text{for } 0 \le t \le t_{j1}, \\ -\dfrac{(u_{j2}-u_{j1})(t-t_{j1})^2}{2(t_{j2}-t_{j1})} + (u_{j2}-u_{j1})t, & \text{for } t_{j1} \le t \le t_{j2}, \\ \frac{1}{2}(u_{j2}-u_{j1})(t_{j2}+t_{j1}) & \text{for } t_{j2} \le t; \end{cases} \tag{8.3.14}$$

$$|S_{2j}(t)| = \begin{cases} (u_{j2}-u_{j1})\,t & \text{for } 0 \le t \le t_{j1}, \\ -\frac{4}{3}\sqrt{a}\,(t-a)^{3/2} - \dfrac{u_{j1}^3}{12a} - au_{j1} + u_{j2}t & \text{for } t_{j1} \le t \le t_{j2}, \\ a(u_{j2}-u_{j1}) + \dfrac{u_{j2}^3 - u_{j1}^3}{12a} & \text{for } t_{j2} \le t; \end{cases} \tag{8.3.15}$$

$$|S_{3j}(t)| = \begin{cases} \frac{1}{2}(\theta_{j2}-\theta_{j1})t^2 & \text{for } 0 \le t \le t_{j1}, \\ \frac{1}{3}\sqrt{a}\,(t+2a)\sqrt{t-a} - \frac{1}{2}au_{j1} - \dfrac{u_{j1}^3}{24a} + \frac{1}{2}t^2\left[\theta_{j2} - \arccos\left(\dfrac{2a}{t}-1\right)\right] & \text{for } t_{j1} \le t \le t_{j2}, \\ \dfrac{u_{j2}^3 - u_{j1}^3}{24a} + \frac{1}{2}a(u_{j2}-u_{j1}) & \text{for } t_{j2} \le t. \end{cases} \tag{8.3.16}$$

Notice that $S_{4j}(t)$ is given by $S_j(t)$ of equation (8.3.8).

The probability density function $f(t)$ of $F(t)$ is obtained from

$$f(t) = \frac{dF(t)}{dt} = \frac{1}{|S_0|}\frac{d|S(t)|}{dt} = \frac{1}{|S_0|}\sum_{i=1}^{4}\sum_{j=1}^{k_i}\frac{d|S_{ij}(t)|}{dt} = \frac{1}{|S_0|}\sum_{i=1}^{4}\sum_{j=1}^{k_i}s_{ij}(t). \tag{8.3.17}$$

The rth moment of t is given by

$$E(t^r) = \int_0^\infty t^r f(t)\,dt = \frac{1}{|S_0|}\sum_{i=1}^{4}\sum_{j=1}^{k_i}\int_0^\infty t^r s_{ij}(t)\,dt. \tag{8.3.18}$$

Differentiating equations (8.3.8) and (8.3.14)–(8.3.16) with respect to t and evaluating the integral in equation (8.3.18), we obtain

$$\int_0^\infty t^r s_{1j}(t)\,\mathrm{d}t = \frac{(u_{j2}-u_{j1})\,(t_{j2}^{r+2}-t_{j1}^{r+2})}{(t_{j2}-t_{j1})\,(r+1)\,(r+2)}, \tag{8.3.19}$$

$$\int_0^\infty t^r s_{2j}(t)\,\mathrm{d}t = \frac{1}{r+1}\sum_{i=1}^{r+1}\binom{r+1}{i}\frac{a^{r+1-2i}(u_{j2}^{2i+1}-u_{j1}^{2i+1})}{2^{2i}(2i+1)}, \tag{8.3.20}$$

$$\int_0^\infty t^r s_{3j}(t)\,\mathrm{d}t = \frac{1}{r+2}\sum_{i=0}^{r+1}\binom{r+1}{i}\frac{a^{r+1-2i}(u_{j2}^{2i+1}-u_{j1}^{2i+1})}{2^{i}(2i+1)} \tag{8.3.21}$$

(Yoshikawa, 1989). Notice that $\int_0^\infty t^r s_{4j}(t)\,\mathrm{d}t$ is given by equations (8.3.11) and (8.3.12). Substituting equations (8.3.11), (8.3.12) and (8.3.19)–(8.3.21) into equation (8.3.18), we can explicitly obtain the values of $\mathrm{E}(t)$ and $\mathrm{Var}(t)$.

Through the above procedure, Okabe and Yoshikawa (1989) calculated the statistic R and showed that high-class apartment buildings tended to locate along arterial streets in Kohto-Sumida, Tokyo (Figure 8.3.3).

8.3.3 Nearest neighbour distance method for area-like objects

When we cannot ignore the area of object O_j, we should apply the NND method for area-like objects. Figure 8.3.6 shows its typical application provided by Okabe *et al.* (1988). In the figure, the area-like objects O indicated by shaded polygons are big parks in Setagaya, Tokyo, and the point-like elements P indicated by the unfilled circles are high-class apartment buildings. They examined whether or not the distribution of high-class apartment buildings was affected by big parks.

The procedure of the the NND method for area-like objects is just the same as that for line-like objects, and so we do not repeat it here. Through that procedure, Okabe *et al.* (1988) calculated the statistic R and came to the conclusion that the distribution of high-class apartment buildings being independent of the location of big parks could not be rejected with significance level 0.05 in Setagaya, Tokyo.

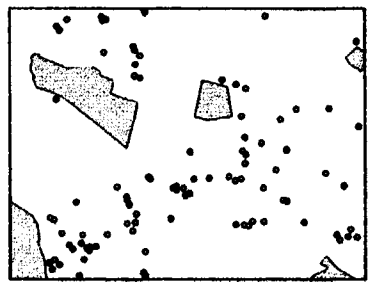

Figure 8.3.6 The distribution of high-class apartment buildings (the unfilled circles) and big parks (the shaded polygons) in Setagaya, Tokyo. (*Source*: Okabe *et al.*, 1988, Figure 1.)

8.3.4 Multi nearest neighbour distance method

In the above three subsections we showed the methods that examined the effect of only one type of objects O on the distribution of point-like elements P. In the real world, the effect is likely to be compounded; more than one type of objects, $O_{(j)}, j \in I_m$, compoundly affect the distribution of point-like objects P. For instance, as is shown in Figure 8.3.3, it is likely that the distribution of high-class apartment buildings is compoundly affected by subway stations (the unfilled circles) and arterial streets (the solid lines).

To examine this compound effect, Okabe and Yoshikawa (1989) generalize the above univariate index R into the multivariate index

$$R = \left(\frac{\bar{t}_{(1)}}{E(\bar{t}_{(1)})}, \ldots, \frac{\bar{t}_{(m)}}{E(\bar{t}_{(m)})} \right), \tag{8.3.22}$$

where $\bar{t}_{(i)}$ is the average NN-distance defined by equation (8.3.3) for objects $O_{(i)}, i \in I_m$. With this multivariate index, we can examine the compound effect of several kinds of objects, $O_{(1)}, \ldots, O_{(m)}$, on the distribution of point-like elements P. Using this index, Okabe and Yoshikawa (1989) found that there was a compound effect of arterial streets and subway stations on the distribution of high-class apartment buildings in Kohto-Sumida, Tokyo (see Okabe and Yoshikawa, 1989, for a detailed discussion).

8.4 THE SHAPE OF A POINT PATTERN

There is no single precise definition of the term shape. Instead, exact meanings are context dependent (for example, see Section 5.11). The shape of a point pattern is the geometric form produced by identifying and linking 'essential' members of a point set $P = \{p_1, p_2, \ldots, p_n\}$ in $\mathbb{R}^m$ (Kirkpatrick and Radke, 1985; Radke, 1988). Essential points are defined in terms of their spatial relationships with other members of P. Since there is no unique way of either selecting or joining essential points, more than one shape can be defined for P. However, most of the resulting shapes are either linear or polygonal structures. Following Kirkpatrick and Radke (1985) and Jaromczyk and Toussaint (1992), we refer to these as internal and external shapes, respectively, although the terms endo-skeleton and exo-skeleton (Radke, 1988) and skeleton and shape hull (Toussaint, 1980b) are also used elsewhere. In this section we describe measures that use features of Voronoi diagrams or related structures, with the emphasis placed on point sets in $\mathbb{R}^2$.

8.4.1 Internal shape

We begin by noting that each shape defined in the above way is equivalent to a graph whose vertices are a subset of P and whose edges join points that are related in some sense. Since the Delaunay triangulation of P, $\mathcal{D}(P)$, can be considered a maximal planar description of the internal structure of P,

attention has focused on using subgraphs of $\mathcal{D}(P)$ described in Section 2.5 as measures of internal shape, in particular the Euclidean minimum spanning tree EMST(P), the relative neighbourhood graph RNG(P), and the Gabriel graph GG(P) which provide, respectively, increasingly more detailed descriptions of the internal shape of P.

Fairfield (1983) also proposes using subgraphs of $\mathcal{D}(P)$ generated using a threshold value t. His method involves examining paths of $\mathcal{V}(P)$. A path is a connected series of segments of $\mathcal{V}(P)$. For any location s on the path, the spread angle of the path is the angle defined by p_i and p_j where p_i and p_j are the two closest points to s on either side of the path. Then define the internal concavity of the path as the difference between the spread angle of the first and last points on the path as it is traversed in a given direction. A path is said to be *t-increasing* if it contains a subpath of concavity $\geq t$ and it contains no subpath of concavity $\leq -t$. A maximal t-increasing path is a t-increasing path which is contained within a no longer t-increasing path. Once all of the maximal t-increasing paths of $\mathcal{V}(P)$ have been found, they are 'cut' into pieces of internal concavity less than t, using the minimal number of cuts. When a path is cut on a Voronoi edge e, the edge of $\mathcal{D}(P)$ which corresponds to e is drawn. The set of all such Delaunay edges so drawn constitutes the internal shape of P. Note that $t = 0$ gives $\mathcal{D}(P)$ and as t increases, smaller subsets of $\mathcal{D}(P)$ are produced (see Figure 8.4.1).

Recall from Section 2.5 that the links in RNG(P) and GG(P) can be defined in terms of empty spindles and empty circles, respectively. Kirkpatrick and Radke (1985) generalize this idea to produce a spectrum of internal shapes. They do this by defining two points p_i and p_j in P, distance $d(p_i, p_j)$ apart, as *β-neighbours* if their neighbourhood or region of influence $N(p_i, p_j, \beta)$ contains no other points of P, where β is a positive real valued parameter whose magnitude reflects the size of the neighbourhood. The *β-skeleton* of P is the set of edges joining the β-neighbours of P. They consider two different ways of operationalizing the neighbourhoods. In the first, for

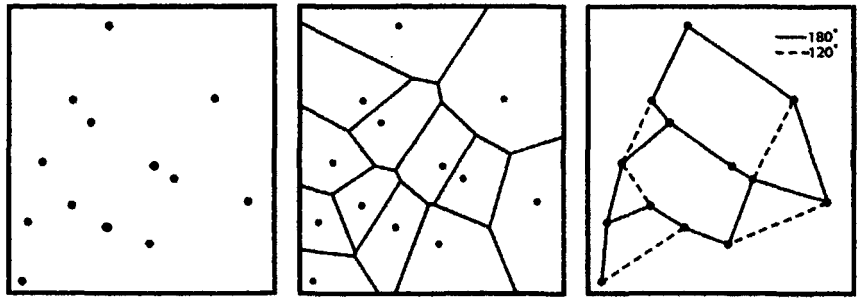

Figure 8.4.1 Defining an internal shape of a point pattern using Voronoi diagram concavity: (a) point pattern; (b) Voronoi diagram; (c) internal shape for two different values of t. (*Source*: Fairfield, 1983, Figure 4.)

$\beta \geq 1$, $N(p_i, p_j, \beta)$ is the intersection of the two circles of radius $\beta d(p_i, p_j)/2$ centred at locations $[(1 - \beta/2)p_i + (\beta/2)p_j]$ and $[(\beta/2)p_i + (1 - \beta/2)p_j]$, respectively, while for $0 < \beta < 1$, $N(p_i, p_j, \beta)$ is the intersection of two circles of radius $d(p_i, p_j)/(2\beta)$ passing through both p_i and p_j. When $\beta = 1, 2$, the β-skeleton is identical to GG(P) and RNG(P), respectively. An alternative is to define $N(p_i, p_j, \beta)$ for $\beta \geq 1$ as the union of the two circles of radius $\beta d(p_i, p_j)/2$ which pass through both p_i and p_j and for $0 < \beta < 1$ as the intersection of the two circles of radius $d(p_i, p_j)/2\beta$ that pass through p_i and p_j. In this case, when $\beta = 1$ the β-skeleton again corresponds to GG(P).

8.4.2 External shape

One of the simplest measures of this type is provided by the convex hull CH(P) of P (see Section 2.5 and Figure 8.4.2(b)). Simple extensions of this measure have been proposed to accommodate specific features of the spatial distribution of P such as concavities. For example, Jarvis (1977) suggests defining n subsets of P consisting of each point of P and its k ($1 \leq k \leq n$) nearest neighbours, generating the convex hull of each subset, and overlaying them. The external shape is defined by the union of the free edges of the union of convex hulls. Thus, as $k \to n$, this shape approaches CH(P). Figure 8.4.2(c) shows this shape hull for $k = 3$. Toussaint (1980) proposes the *Gabriel hull* which involves removing from CH(P) any edge which does not intersect the dual Voronoi diagram edge and replacing it with two new edges, one of which extends from one of the vertices of the removed edge to the generator point of the first Voronoi polygon traversed by the removed edge (see Figure 8.4.2(d)).

Edelsbrunner *et al.* (1983) generalize the notion of a convex hull to produce a spectrum of shapes which they call α-*hulls*. For an arbitrary real number α, they define a generalized disc of radius $1/\alpha$ as a disc of radius $1/\alpha$ if $\alpha > 0$, the complement of a disc of radius $|1/\alpha|$ if $\alpha < 0$, and a half plane if $\alpha = 0$. The α-*hull* of P is the intersection of all closed generalized discs of radius $1/\alpha$ that contain all points of P (see Figures 8.4.2(e) and (f)). If α is sufficiently large, the α-hull is the smallest enclosing circle of P; if $\alpha = 0$ it is CH(P); and if α is sufficiently small it is P itself. Furthermore, they define a point p_i in P as α-*extreme* if there exists a closed generalized disc of radius $|1/\alpha|$ such that p_i lies on its boundary and it contains all points of P. If for two α-extreme points, p_i and p_j, there exists a closed generalized disc of radius $|1/\alpha|$ with both points on its boundary and which contains all other points of P, p_i and p_j are said to be α-*neighbours*. Then the α-*shape* of P is a straight line graph whose vertices are the α-extreme points and whose edges connect the respective α-neighbours. As the value of α decreases, the α-shapes provide increasingly more detailed descriptions of the external shape of P. Edelsbrunner *et al.* (1983) show that when $\alpha \geq 0$, any α-shape is a subgraph of the Delaunay triangulation of the farthest point Voronoi diagram of P (see Section 3.3.1), while for $\alpha \leq 0$ it is a subgraph of $\mathfrak{D}(P)$. Edelsbrunner and Mücke (1994) extend the concepts of α-hulls and

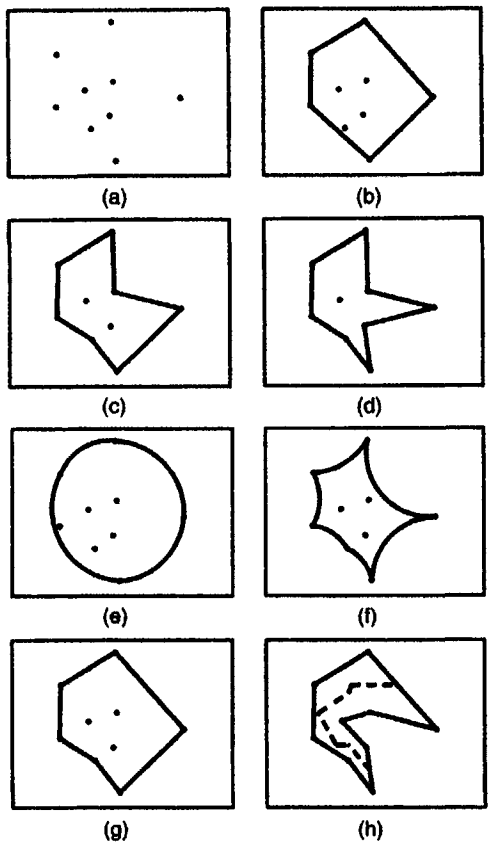

Figure 8.4.2 Different external shapes of a point pattern: (a) point pattern; (b) convex hull; (c) union of convex hulls of subsets of points; (d) Gabriel hull; (e) α-hull ($\alpha > 0$); (f) α-hull ($\alpha < 0$); (g) Pernus's circles of influence; (h) 'natural' simple polygon (the dashed line is the minimal spanning Voronoi tree).

α-shapes to finite point sets in $\mathbb{R}^3$, while Saito *et al.* (1991) extend the concept of the α-hull to sets of figures of different shapes and sizes located in a digitized picture plane, labelling the resulting structure the *extended digital α-hull*, ED α-hull. They also consider different distance metrics in addition to Euclidean with the disc being replaced by a rhombus or a square for the Manhattan metric (L_1) and the supremum metric (L_∞) (see Section 3.7.1), respectively.

Pernus (1988) proposes a method combining aspects of both the Gabriel hull and the α-shape which involves circles of influence similar to those used by Kirkpatrick and Radke (1985) to develop the β-skeletons discussed in the previous section. For two points p_i and p_j located at (x_i, y_i) and (x_j, y_j), respectively, a distance $d(p_i, p_j)$ apart, and a parameter $t\,(0 \leq t \leq \infty)$, the circles of influence are the set of circles with centres

at $([x_i + x_j]/2 - [y_j - y_i]t,\ [y_i + y_j]/2 + [x_j - x_i]t)$ and radius $d(p_i, p_j)$ $(1 + 4t^2)^{1/2}/2$. The method begins by generating $\mathcal{D}(P)$. Then for each pair of points p_i and p_j which are the end points of an edge of CH(P) we examine their circle of influence for the specified value of t. If that circle is empty, the edge between p_i and p_j is retained; if not, the edge is removed. This process is repeated until no further edges are removed (see Figure 8.4.2(g)). The resulting shapes decrease in detail from the Gabriel hull ($t = 0$) to CH(P) ($t = \infty$).

Another way to summarize the shape of P is with a simple polygon. Since the complete graph of P consists of $\binom{n}{2}$ edges, to create a simple polygon we need to choose $\binom{\binom{n}{2}}{n}$ edges. The magnitude of this search can be reduced by considering $\mathcal{D}(P)$ which has at most $(3n-6)$ edges (Property V11). The problem then becomes a search for a subgraph of $\mathcal{D}(P)$ which is a simple polygon. Furthermore, since the polygonal path must touch each and every vertex just once, it must be a Hamiltonian cycle. O'Rourke *et al.* (1987) note that the set of Delaunay triangles forming a simple polygon is also associated with a tree in $\mathcal{V}(P)$. In light of this, their choice for the most 'natural' simple polygon is that associated with the minimal spanning Voronoi tree of P which they define as the shortest subgraph of $\mathcal{V}(P)$ whose corresponding collection of Delaunay triangles forms the triangulation of a simple polygon which spans every point of P (see Figure 8.4.2(h)).

Boissonnat (1984) generates a simple polygon by sculpting CH(P). He begins by defining $\mathcal{D}(P)$ and then eliminating sequentially triangles that have one or more edges forming part of CH(P) or its subsequently sculpted form. The procedure continues until all members of P are located on the sculptured CH(P). However, these simple polygon approaches will only be successful if $\mathcal{D}(P)$ contains a Hamiltonian cycle, which, although highly probable, is not universally true (Kantrabutra, 1983; Dillencourt, 1987a,b, 1990c, 1996a,b). Unfortunately, Dillencourt (1996a) shows that it is an NP-complete (Section 1.3.4) problem to determine if there is a Hamiltonian cycle in a non-degenerate Delaunay triangulation.

8.5 SPATIAL INTENSITY

When the point pattern under investigation is not homogeneous, we may wish to explore how the intensity of the phenomenon represented by the points varies over the study area. If we generate the Voronoi diagram of the point set $P = \{p_1, p_2, \ldots, p_n\}$ and measure the area A_i of each polygon p_i, we can consider $1/A_i$ as an indicator of the local intensity of the point pattern at p_i (Duyckaerts *et al.*, 1994). In order to visualize variations in local intensity, Dahlberg (1967) suggests that we may represent each polygon as a prism whose height is inversely proportional to the area of the polygon (see Section 6.1.1 for a similar procedure). However, other characteristics of the Voronoi polygons besides area also reflect changes in spatial intensity. For example,

Voronoi polygons are elongated in a direction perpendicular to that of increasing intensity, while their generator points are increasingly displaced from their centres of gravity in the direction of increasing density (Ahuja and Tuceryan, 1989).

In some situations, instead of being concerned with local variations in intensity, the focus of our investigation is on estimating the intensity for the entire pattern. Such instances arise in forestry where estimates of the intensity of a particular tree species may be used in deriving expected timber yields, or in metallurgy where intensities of impurities in a given material above a specified level may invalidate its use for a particular purpose.

The obvious way of obtaining an estimate $\hat{\rho}$ of ρ, the spatial intensity of a point process, is to use

$$\hat{\rho} = n / |B|, \tag{8.5.1}$$

where $|B|$ is the area of B, or alternatively, to count the number, n_b, of the members of P in some bounded subregion, b of B of area $|b|$ and use

$$\hat{\rho} = n_b / |b|. \tag{8.5.2}$$

However, ρ may also be estimated using any of the techniques mentioned above such as quadrat, nearest neighbour and second-order methods. (See Upton and Fingleton, 1985, Chapter 2, for a review of the use of such methods in estimating intensity.) Voronoi polygons may also be used in this role.

If ρ is the intensity or number of points per unit of area, then the area per point is ρ^{-1}. Recognition of this provides an alternative way of estimating ρ. If we select a sample of m polygons from $\mathcal{V}(P)$ which satisfy the avoidance of edge effects and the independence conditions specified in Section 8.1.1, then the mean area $\bar{A}$ of the sample polygons is given by

$$\bar{A} = \frac{1}{m} \sum_{i=1}^{m} A_i. \tag{8.5.3}$$

Brown (1965) suggests that the intensity may be estimated from

$$\hat{\rho} = 1 / \bar{A}. \tag{8.5.4}$$

However, this estimate is only asymptotically unbiased. To provide an unbiased estimate of ρ Ord (1978) uses the reciprocal of A_i so that the intensity estimate is given by

$$\hat{\rho} = \frac{1}{m} \sum_{i=1}^{m} (A_i^{-1}). \tag{8.5.5}$$

If we wish to provide a range for the above estimates, for equation (8.5.4) we need the variance $\mathrm{Var}(A_i)$ of the sample areas given by

$$\mathrm{Var}(A_i) = \frac{1}{m-1} \left[\sum_{i=1}^{m} (A_i^2) - \frac{1}{m} \left(\sum_{i=1}^{m} A_i \right)^2 \right], \tag{8.5.6}$$

and for equation (8.5.5) the variance $\mathrm{Var}(A_i^{-1})$ of the reciprocals given by

$$\text{Var}(A_i^{-1}) = \frac{1}{m-1}\left[\sum_{i=1}^{m}(A_i^{-2}) - \frac{1}{m}\left(\sum_{i=1}^{m}A_i^{-1}\right)^2\right]. \quad (8.5.7)$$

Thus, a two-standard-deviation range for $\bar{A}$ in equation (8.5.4) is given by $\{\bar{A} \pm [\text{Var}(A_i)/m]^{1/2}\}$ while that for the mean of the reciprocals is given by $\{(\sum_{i=1}^{m}A^{-1})/m \pm [\text{Var}(A_i^{-1})/m]^{1/2}\}$. In general, the size of the range reflects the nature of the point pattern. If it is regular, there is little variation in the areas of the Voronoi polygons and so the variances in equations (8.5.6) and (8.5.7) are small leading to a small range. Conversely, clustered patterns produce Voronoi polygons of greatly varying sizes, thus producing large variances in equations (8.5.6) and (8.5.7) and large ranges.

When intensities are estimated in the field, as often arises in forest inventories, it is more practical to use the areas of Delaunay triangles defined by trees enclosing sampling sites (Ward, 1991), in which case a factor of 2 is included in the denominator and numerator of equations (8.5.4) and (8.5.5), respectively. Simulation studies undertaken by Ward suggest that using triangle areas produces more accurate estimates of intensity than either the quadrat or nearest neighbour methods, regardless of whether the distribution of trees is random, regular or clustered.

8.6 SEGMENTING POINT PATTERNS

For those patterns where there is marked spatial variation in the local intensity of points, we may wish to attempt to segment the pattern into component parts. Ahuja (1982) and Ahuja and Tuceryan (1989) suggest that individual points may be considered to belong to one of five types; isolated points, members of a curvilinear structure, members of a cluster with an empty interior, or members of either the boundary or the interior of a cluster with a non-empty interior (see Figure 8.6.1). They suggest that various geometric properties of the Voronoi polygons of a point and its neighbours (compactness, area, elongation, and eccentricity) and ratio measures involving the associated Delaunay edges can be used to determine to which group a point belongs. For example, the Voronoi polygons of points in the interior of a homogeneous cluster will all be compact, have approximately equal areas, small eccentricities, and little elongation. In contrast, in the interior of a non-homogeneous cluster, the areas, eccentricities and elongation of the Voronoi polygons all change with respect to the direction of density changes. David (1988a) illustrates the utility of such an approach by using volumes and adjacencies of Voronoi polyhedra of atoms to identify channels in macromolecular assemblies.

Tüceryan and Jain (1990) extend this approach to texture segmentation in images. This is the process of identifying regions in an image with similar texture and separating regions with different textures. Textures are defined in terms of tokens, which may take the form of either points or more complex primitives such as triangles or arrows, extracted from grey level images. They

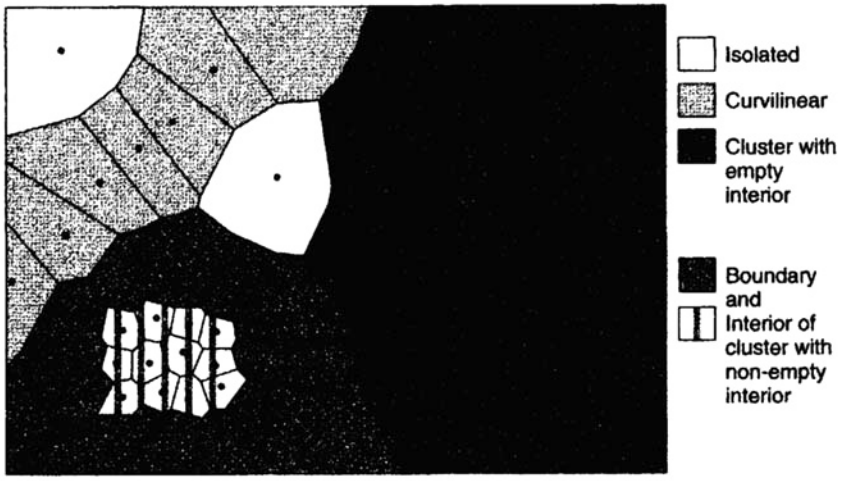

Figure 8.6.1 Segmenting a point pattern into component parts.

generate both the Voronoi diagram and the Delaunay tessellation of the set of image tokens and measure various order moments of each Voronoi region. A neighbour of token t_i is defined as any token whose Voronoi region is contiguous to the Voronoi region of t_i, $V(t_i)$, or contiguous to a Voronoi region which is contiguous to $V(t_i)$. If e_{ij} is the Delaunay edge linking tokens t_i and t_j, and N_i and N_j are the sets of neighbouring tokens for t_i and t_j, respectively, when e_{ij} lies in a single texture region the properties of N_i and N_j should be similar. Thus a 'break probability' can be assigned to e_{ij} in terms of the similarity of the order moments of the Voronoi regions of N_i and N_j.

If we are only interested in identifying clusters within a point pattern it may be possible to focus on just one characteristic. For instance, Yokoi and Toriwaki (1986) suggest defining a neighbourhood graph such as one of those described in Section 2.5 and then removing any point for which the average length of edges incident at the point exceeds some threshold value. The remaining connected points then constitute the clusters. A specific example of this approach is provided by Zahn (1971) who uses the Euclidean minimum spanning tree (see Section 2.5) and deletes any edge whose length is significantly longer than the average of those edges on both sides of the edge. Duyckaerts *et al.* (1994) propose a three-step iterative method for cluster identification which uses the areas of the Voronoi polygons of the points. The first step involves identifying the smallest polygon in the Voronoi diagram of the points which is designated the reference polygon. Step two examines the polygons which are contiguous to the reference polygon and adds to it those polygons whose areas do not exceed that of the reference polygon by more than some pre-specified proportion. This step is repeated until all newly examined polygons exceed the threshold. In the third step, the cluster so defined is removed from the pattern and the three steps are

repeated for the remaining polygons until no further clusters result. Isaac and Wyatt (1997) adopt a similar strategy to automatically segment multimodal wave spectra into the individual modes present when the spectra are sampled by means of a set of sampling points. They show that the Voronoi diagram approach is particularly effective when the sampling points are non-uniformly distributed as they are for wave spectra extracted by the inversion of high-frequency radar backscatter.

8.7 MODELLING POINT PROCESSES

In this section we turn our attention away from the ways in which characteristics of Voronoi diagrams, Delaunay tessellations and related structures can be used in both exploring and modelling of point patterns to considerations of their use in constructing point processes. Most of this work focuses on *Markov point processes.*

A Markov point process on B is one in which the conditional distribution on A (a subset of B), given its behaviour in $B \backslash A$, depends only on its behaviour in $E(A) \backslash A$, where $E(A)$ is the environment of A given by means of a neighbourhood relation $\sim$ on B (Stoyan *et al.*, 1995, p. 175). $E(A)$ is defined by

$$E(A) = \{x : x \sim y \text{ for some } y \in A\} \tag{8.7.1}$$

One possible definition of $\sim$ is

$$x \sim y \text{ if and only if } \| x - y \| < r, \tag{8.7.2}$$

so that interaction occurs between all pairs of points within a fixed, predetermined range. Building on a suggestion by Ord (1977), Baddeley and Møller (1989) offer an alternative approach in which the neighbourhood relationships are defined in terms of the configuration of the points. One possibility is that a pairwise interaction occurs between points which are Voronoi neighbours. Another is to consider the interaction between triples of points whose Voronoi polygons share a common vertex. As an example, consider the simple sequential inhibition process mentioned in Section 5.12. In this process points are located sequentially at random in B subject to the constraint that a new point is at least some fixed distance r from any existing point. Weaker inhibitory effects can be produced by replacing this constraint with one in which a new point is realized if the area of its Voronoi polygon exceeds some pre-specified threshold size (Ord, 1977) or if the distance of a point to each of the vertices of its Voronoi polygon exceeds some pre-specified value (Baddeley and Møller, 1989).

Markov point processes may also be applicable in situations where the points in the pattern consist of more than one type so that a mark can be attached to each point indicating its type. Brown *et al.* (1981) describe a class of *two-type point processes* where the marginal process of both types of point is Poisson and an interaction occurs between points of different types.

In particular, they focus on the situation of inhibition between the two types. Such a process may be constructed in the following way. First, generate a Poisson Voronoi diagram. Then, for each Poisson Voronoi cell generate the number of events of each type, e_1 and e_2, based on the marginal distributions of the processes and level of their correlation, and the area of the cell. Finally, locate the e_1 and e_2 events uniformly in the cell. The resulting process is stationary under rigid motions, ergodic, and mixing. Since this approach can be extended to any random partition, the Poisson Delaunay tessellation may be used instead of the Poisson Voronoi diagram.

In those situations where we wish to consider the phenomenon represented by the points at more than one scale, *fractal set models* may be appropriate, especially if the scales span several orders of magnitude. A set A has a fractal dimension d if the number of points inside a sphere of radius r is given by $N(A, r) \approx cr^d$ for some positive constant c (Zaninetti, 1991b). In an application in astronomy, Martinez *et al.* (1990) describe a stochastic point process in $\mathbb{R}^3$ which results in clustered patterns with multifractal properties very similar to those of real galaxies. This process involves generating a Poisson Voronoi diagram, then using the vertices of this diagram to generate a second Voronoi diagram whose vertices constitute the realization of the process.