LOCAL DENSITY ANALYSIS:
A NEW METHOD FOR QUANTITATIVE SPATIAL ANALYSIS

Ian Johnson

Department of Prehistory
Research School of Pacific Studies
Australian National University

Abstract

This paper presents a new method for the quantitative spatial analysis of occupation floors, or more generally of archaeological horizons. The method is directly applicable to 'precise' coordinate data, counts for quadrats or irregular areas, and to an intermediate form of data involving collection by a sub-grid which effectively yields 'approximate' coordinates (typically ± 15 cm). It is suggested that the latter method is often ideal for archaeological applications where 'precise' coordinates may merely resolve post-depositional noise or juxtapositions which should be noted during excavation. Local Density Analysis is believed to be the only method capable of treating such data, and the datasets necessary and programs available are discussed.

Introduction

Over the last few years there has been a growing interest, amongst prehistorians, in the application of quantitative methods for analysing the distributions of archaeological remains on occupation floors, or more generally in archaeological horizons. This interest has been stimulated by the increasing availability of data relating to horizontal patterning, based on an increasing awareness of the relevance of this data and of the sort of information which can be obtained from it. Many people have realised that subjective judgments of distributions can be very different from one person to the next and that some form of quantification would be desirable to back up such judgments. Quantification would provide a mathematical resume of the raw data on which to base either direct interpretations or multivariate analyses destined to isolate factors of covarying artefact types. The result of this interest has been the borrowing of analytical methods from other fields, particularly from human geography and plant ecology. These methods have not proved very successful, mainly due to their inability to cope with the irregular patterning and often numerically sparse data available in most archaeological situations.

In general, what is needed by prehistorians is some measure of association between the distributions of different sorts of artefact,[1] on which to base interpretations bearing on the layout of occupation areas or on the existence of spatially segregated activities identified by the existence of spatially covarying artefact types.[2] This requirement, coupled with the lack of

---

[1] I use 'artefact' in the broad sense to include waste products and utilised objects as well as formalised tools.

[2] It is very dangerous to make the uncritical assumption that spatial covariation equates with spatial segregation of activities, as this overlooks the fundamental effect that site maintenance activities (particularly displacement of waste products) and curation may have on spatial patterning.

coordinate data for all or a proportion of artefacts excavated (these are generally collected by grid units such as metre squares) has focussed attention on a broad class of methods termed 'quadrat methods'.

Quadrat Methods

The site is divided into a series of 'quadrats' - these are generally square or rectangular (l=2b) areas, non-overlapping and contiguous. For most methods the basic data consists of a matrix of frequencies for each category of artefact in each quadrat. In most applications of these methods a number of such matrices are constructed for different quadrat sizes[1] and some form of test is applied to find the grouping giving rise to the matrix with the greatest variability. This grouping is assumed to be that most nearly approximating the scale of patterning present (note that more than one grouping may be selected if patterning peaks at more than one scale, but one cannot hope to detect several or closely spaced scales of patterning, as is often the case in archaeological sites).

The association between any pair of artefact categories can now be represented by any index or coefficient expressing the covariation between quadrats of the artefact frequencies for those categories. This calculation takes no account of the spatial relationship between the quadrats.

Weaknesses of quadrat methods

The basic weakness of quadrat methods is associated with the *a priori* imposition of a geometrical grid of quadrats on distributions of objects which are irregular and often with diffuse boundaries.[2]

It has been shown (Kershaw 1957) that patterning cannot be reliably detected if the quadrats are less than twice the size of that patterning, owing to the fact that clusters of artefacts are generally split up between more than one quadrat, and no account is taken of the adjacency of quadrats (cf. above). Equally, small quadrat sizes run into the problems associated with small artefact frequencies, particularly serious in many archaeological situations where the material is sparse (if the material is plentiful, repeated occupations over a long period will often have given rise to too much archaeological noise for distribution analysis to be useful). In particular the Pearson Correlation Coefficient, probably the most used for quadrat analyses, is

---

1 Generally by grouping adjacent quadrats, going from a grid of squares to 2x1 rectangles and back to squares etc.

2 Herein lies an important difference between archaeological distributions and those treated in plant ecology, where patterning is often repetitive and the boundaries of the study are generally defined by the researcher.

totally unsuitable.

For larger quadrat sizes one runs into the twin problems of poor resolution of all but the largest scales of patterning, and a reduction of the number of quadrats, and hence of artefact frequency pairs for the calculation of the relevant indices or coefficients.

Thus quadrat methods can only be applied to a limited range of quadrat sizes, which will not generally span the range of scales of patterning present in an archaeological site. Even under optimum conditions their resolution of patterning will be poor owing to the disparity between the shape of the sampling units (square/rectangular with fixed orientation) and the shape of the clusters to be detected (often sub-circular or elongated with varied orientations). Clearly what is required is a method which can cope with low artefact frequencies, and therefore use small quadrats giving high resolution of the distributions present, while at the same time exploiting the spatial relationship between those quadrats so that larger scales of patterning can be detected despite their being split up between many quadrats. It was for these reasons that Local Density Analysis was developed.[1] As the quadrats used by this method will typically be an order of magnitude smaller than the smallest quadrats that a conventional quadrat method can cope with, I will use the term 'cells' to refer to them.

Artefact collection by cells

There are two reasons why one cannot measure 'precise' coordinates for every artefact found in a site - firstly the time required and secondly the fact that a certain proportion of objects will be missed during excavation and recovered in the sieves. The latter are often collected by large, low resolution units, such as metre squares, in contrast with the high resolution coordinate data for a proportion of the objects found in *situ*. The data collected is therefore of very uneven quality. A compromise, economical in time but giving good resolution of the distributions present, is the collection of all objects by small grid units or cells,[2] typically of 20 or 25 cms. This effectively gives rise to 'approximate' coordinates. The loss of resolution relative to 'precise' coordinate data will be negligible except for very small scales of patterning (which will generally be masked by 'noise' due to post-depositional disturbance). The technique is thus ideal provided there is a well defined archaeological horizon not necessitating tight vertical control.

---

1 This method can also be applied directly to coordinate data and as an ordinary quadrat method, an added bonus being that the area of the site can be eliminated from the calculations, obviating the need to define the boundaries of the site.

2 This method has been used for the collection of small waste flakes etc. at the site of Pincevent, near Paris, excavated by Professor A. Leroi-Gourhan.

The only way that cell-count data can be treated by conventional methods is by grouping cells into quadrats, thus introducing all the disadvantages previously mentioned. Local Density Analysis is the only method that, to my knowledge, is available to treat such data directly.

Local Density Analysis

The method was initially conceived as a coordinate method in order to eliminate the problems, noted above, associated with the existence of quadrats. These problems were eliminated by replacing the imposed grid of quadrats by a series of circular sampling units centred on the individual objects making up the distribution. The size of these circles determines the sensitivity of the method to different scales of patterning, and the determination of an optimum size will be treated below. The jump to treating cell-count data is made by the approximation that all the objects within a cell are located at its centre, and the sampling units can therefore be formed by all cells whose centres lie within a specified radius of each cell in turn.

The use of sampling units centred on the objects themselves means that the number of sampling units is independent of their size. On the other hand, the fact that they frequently overlap means that many of the mathematical techniques used for analysing quadrat data are no longer applicable.

The algorithm I have chosen is to compare the *local* density of one category in the vicinity of objects of another category[1] with the *global* density of the first category, i.e. the density determined over the whole area of the site. This clearly raises the question of determining the area of the site. Though an approximate value can generally be determined by examination of site plans, an accurate value is generally impossible to obtain, and it is for this reason that the area of the site will be eliminated from the calculations at a later stage.

Index of co-clustering (association) of j with i is termed $C_{ij(r)}$ for sampling unit radius r.

$$C_{ij(r)} = \sum_{k=1}^{N_i} \frac{(M_{ij})_k / N_i \pi r^2}{N_j / A}$$

$(M_{ij})_k$ = Number of objects of category j within radius r of the kth object of category i

N_i, N_j = Total number of objects of category i and j respectively

A = Estimated area of the site

[1] Defined as the mean density of the first category measured over the sampling units centred on the objects of the second category.

It will be noted that $C_{ii(r)}$ is a measure of clustering of the objects of category i for a sampling unit radius of r, and that $C_{ij(r)} = C_{ji(r)}$.

If the two categories i and j are unassociated, the *global* density and the *local* density should be equal, and the value of $C_{ij(r)}$ will be unity (within the limits of statistical error). This value can be shown to be obtained when either distribution is random with respect to the site, whatever the nature of the other distribution (this is a valuable property, as this is the only *true* form of lack of association; any lack of association involving two non-random distributions is method-dependant. Note however, that this result will only be obtained with infinite distributions).

If the two categories i and j are associated the *local* density will be higher than the *global* density, and the index will therefore be greater than unity. Conversely a dissociation or mutual exclusion will be indicated by an index less than unity. The limiting values are 0 for dissociation (no objects of one category within a distance of r from objects of the other) to $A/\pi r^2$ for association (all the objects of both categories lie within a distance r from one another). It should be noted that the index is assymetrical around unity, and for this reason it will generally be subjected to a log transformation, making it symmetrical about zero.

$$CL_{ij(r)} = Ln(C_{ij(r)})$$

Elimination of the site area

Since the estimate of the area of the site occurs as a constant factor throughout the matrix of association indices, it can be eliminated by the calculation of a new matrix based on the Pearson Correlation Coefficient between pairs of rows (or columns) of the association index matrix. This leads directly to the possibility of a multivariate analysis such as Principal Components Analysis. The correlation coefficient matrix is more stable than the association index matrix, unless the latter is small, as the correlation coefficients are based on the relationships between the distribution of one category and those of all the other categories, thus smoothing out chance variation which may occur for a single association index. Equally the correlation coefficients are easier to interpret, being bounded by fixed limits of + and - 1 and being familiar to most archaeologists.

The Scale of the Analysis

In order to resolve the detail of the relationships between the distributions of different categories of remains, the radius of the sampling units, r, should be as small as possible. However there is a limit to how small the sampling units can be due to the reduction of $\sum_{k=1}^{N_i} (M_{ij})_k$ which results in increasing chance variation or instability of the association index. At the other end of the scale, as the radius of the sampling units is increased, they increasingly overlap the boundaries of the site and the boundaries of any concentrations within the site. As a result $\sum_{k=1}^{N_i} (M_{ij})_k/\pi r^2$ drops and $C_{ij(r)}$ is therefore reduced. I have called this the 'edge-effect'. It can be mathematically predicted and simulation studies have so far confirmed the predictions, so that the change in the edge effect as r is increased can actually be used as a means of determining the scale of patterning present (Johnson 1976). A detailed discussion is beyond the scope of this paper, but in effect variations in the indices due to small scale patterning are successively smoothed out as the radius of the sampling units is increased, i.e. the scale of the analysis increases.

In practice an optimum value for r is chosen by examination of the association and correlation matrices and the contributions of each component in the multivariate analysis. Unstable indices (r too small) are indicated by a lack of strong positive or negative correlations, inconsistencies in the association index matrix (two categories mutually associated but showing widely divergent associations with a third) and the need for several components of similar strength to provide a reasonable summary of the correlation matrix. As r is increased, the range of the correlation coefficients increases and the main components of the multivariate analysis become stronger. Finally, as r starts to approach the same order of magnitude as the size of the site, the association coefficients drop due to the edge-effect.

Application to cell-count data

If we make the approximation that all objects in a cell are concentrated at its centre, we can write:

$$C_{ij(r)} = \frac{\sum_{k=1}^{N_c} [(M_j)_k . (L_i)_k] \; / \; N_i \; l^2}{N_j/A}$$

'Equivalent radius'[1] of the analysis = r =[√(Total area of cells whose centres lie at less than R from the centre of the kth cell)/π]

$(M_j)_k$ = Number of objects of category j in all cells whose centres lie at less than R from the centre of the kth cell

N_c = Total number of cells

$(L_i)_k$ = Number of category i objects in the kth cell

l^2 = Area of cells

[1] The 'equivalent radius', r, is the simplest measure of the scale of the analysis for comparison with analyses made using coordinate data.

Local Density Analysis as a Quadrat Method

If we apply the equation above with R = 0, the method becomes a quadrat method. Note that the distinguishing feature of a quadrat method is that it takes no account of the relative positions of the sampling units (the values summed in the numerator are each based on the artefact frequencies for a single quadrat) and that any given quadrat occurs in only one sampling unit, unlike cells which may contribute to several sampling units. The quadrats can therefore be non-contiguous, either as a random or patterned sampling procedure or occurring in two or more excavated areas. Note that the method can accommodate 'quadrats' of unequal sizes and shapes. In this case several different formulae can be applied, depending on the weighting one applies to different sized quadrats, but the simplest, and probably the best is:

$$C_{ij(r)} = \frac{\sum_{k=1}^{N_c} [(L_j)_k . (L_i)_k / S_k]}{N_i . N_j / A}$$

where S_k is the area of the kth 'quadrat'

Applications

So far Local Density Analysis has been tested on four sites, with good results. The main test was on a part of the Magdalenian open-air site of Pincevent (Section 36:V105), an area of 10x11 m. This test confirmed the validity of the cell-count approximation (results of cell and coordinate analyses were compared) and no obvious discrepancies were found between the results of Local Density Analysis and the detailed subjective study made by Professor Leroi-Gourhan (Leroi-Gourhan & Brezillon 1972). In fact it proved possible to suggest one or two further interpretations, despite the methodological-test nature of the analysis.

A second test on the site of Les Tarterets II (Brézillon 1971) was intended to compare Local Density Analysis with the quadrat correlation method mentioned earlier. The latter analysis was carried out by A. Hesse and yielded very inconsistent results clearly at variance with strongly visible patterning (Hesse 1971). The results of Local Density Analysis were consistent and corresponded well with subjective judgement of the patterning present. Furthermore the quadrat analysis version of the method gave results remarkably close to those of the coordinate analysis, thus confirming the stability of the association indices available from this method.

Computer Implementation

Two programs are available for Local Density Analysis, one for coordinate data (LOCD), the other for cell-count of quadrat data (CELLS) (note that a quadrat analysis is run simply by specifying a zero radius for the analysis, see above).

Both programs are available in Fortran off-line or interactive versions, the first card(s) or question specifying the format of the dataset. The second card or the following questions establish the parameters for the analysis, input and output unit numbers etc., and the values are checked to see that they lie within reasonable limits.

The dataset can be on cards or file and the first record is taken as a title. Subsequent records should specify X and Y coordinates for one or more objects, together with a numerical code indicating the object's category (LOCD) or the numbering of the cell outwards from the origin along the X and Y axes, followed by the counts for each category in turn within that cell (CELLS). An interactive program (CONVT) is available for converting coordinates measured within metre squares identified by a letter/number combination, to cartesian coordinates.

The output of LOCD and CELLS consists of tables of summed raw counts, mean counts per sampling unit, association indices before and after log transformation and correlation coefficients. Any of these can be filed for further computations, such as multivariate analysis. It is hoped in the near future to add a means of combining categories, at present a separate program. Other problems being worked on are the plotting back of the factors extracted from a multivariate analysis, in the form of factor density plans of the site, quadrat analysis with irregular sized quadrats and the creation of simulated distributions with known characteristics.

Conclusion

The development of Local Density Analysis is by no means complete, but it is now a perfectly usable tool which test applications show can give results of archaeological use, rather than simply being a methodologist's plaything. The programs available are very straightforward and require no special knowledge to be applied, and the extreme versatility of the method makes it suitable for application in most circumstances where horizontal patterning may be expected. It would be useful to get further experience of its application over and above my own work, so I would be very willing to supply programs and advise anyone interested in applying it.

The work described in this paper was largely carried out at the Institut du Quaternaire, Bordeaux University. I received much encouragement and advice from Paul Callow, University of Cambridge, and J-Ph.Rigaud, Director of Prehistoric Antiquities of the Aquitaine Region, but errors are my own.

Brézillon, M. 1971 — 'Les Tarterets II, site Paléolithique de plein-air à Corbeil-Essonnes (Essonne)'. GALLIA PREHISTOIRE XIV, 3-40

Hesse, A. 1971 — 'Les Tarterets II, site paléolithique de plein-air à Corbeil-Essonnes (Essonne). II. Comparaison par le calcul des distributions horizontales des vestiges lithiques'. GALLIA PREHISTOIRE XIV, 41-46

Johnson, I. 1976 'Contribution méthodologique a l'étude de la répartition des vestiges dans des niveaux archéologiques. D.E.S. Thesis, Universite de Bordeaux I'

Kershaw, K.A. 1957 'The Use of Cover and Frequency in the Detection of Pattern in Plant Communities'. ECOLOGY 38:291-299

Leroi-Gourhan, A. Brézillon, M. 1972 'Fouilles de Pincevent. Essai d'analyse ethnographique d'un habitat magdalénien (la section 36)'. VII supplement to GALLA PREHISTOIRE, C.N.R.S., Paris