The Detection of Random Heterogeneity in Plant Populations

Author(s): Peter J. Diggle

Source: *Biometrics*, Vol. 33, No. 2 (Jun., 1977), pp. 390-394
Published by: International Biometric Society
Stable URL: http://www.jstor.org/stable/2529790

The Detection of Random Heterogeneity in Plant Populations

PETER J. DIGGLE

Department of Statistics, University of Newcastle upon Tyne, England

Summary

Most tests of spatial randomness distinguish between regular and aggregated departures from complete spatial randomness. In this paper, a two-stage procedure for the detection of random heterogeneity is proposed; a preliminary test of randomness is followed, in appropriate cases, by a heterogeneity test, thus effecting a four-way characterisation of spatial point patterns as regular, random-homogenous, random-heterogeneous or aggregated.

1. Introduction

A useful starting point in the analysis of the spatial pattern presented by a particular plant population is to test the null hypothesis, $\mathcal{H}_0$, that the individual plants, envisaged as point locations in the plane, are distributed completely at random or, more formally, that they constitute a partial realisation of a homogenous, two-dimensional, Poisson point process. Quadrat methods, as discussed for example by Greig-Smith (1964) were originally used in this context, but may be non-robust to changes in quadrat size (Holgate 1972) and a number of alternative tests have evolved, using plotless sampling techniques. Among such "distance methods," the "T-square" tests of Besag and Gleaves (1973) presuppose no knowledge of population density and aim to combine the intuitive appeal of Hopkins' (1954) test with the practicability of Holgate's (1965) ratio test.

Let X represent the distance from a randomly selected point, P, to the nearest plant in the population, at Q say, and define QQ^T to be the line through Q, perpendicular to PQ. Now let Y be the distance from Q to the nearest plant in the population, excluding all plants which lie on the same side of QQ^T as does P. For a sample of m randomly selected points, and vectors of observations $\mathbf{x}$ and $\mathbf{y}$, the T-square statistics may now be written as

$$t_b = \Sigma x_i^2 / \Sigma(x_i^2 + \tfrac{1}{2}y_i^2)$$

and

$$t_N = \frac{1}{m} \Sigma x_i^2/(x_i^2 + \tfrac{1}{2}y_i^2),$$

whose respective null distributions are exactly beta, with m and m degrees of freedom (d.f.), and approximately normal, with mean $\frac{1}{2}$ and variance $1/(12m)$, all summations being over the range $i = 1, \cdots, m$; details may be found in Besag and Gleaves (1973). The results obtained by Diggle, Besag and Gleaves (1976), using Monte Carlo methods, indicate that either test will be comparatively powerful against a wide variety of regular or aggregated alternatives, thus permitting an admittedly approximate three-way characterisation of spatial point patterns as regular, random or aggregated, corresponding respectively to a significantly small, non-significant or significantly large value for t_B or t_N.

In the case of aggregation, no attempt is made to distinguish between clustering of plants and random heterogeneity in the environment. Indeed, as shown by Bartlett (1964), it may be impossible to make any such formal distinction on the basis of *any* statistical test. However, we may encounter patterns which, at an intuitive level, exhibit heterogeneity, i.e., varying

Key Words: Spatial distributions; Tests of randomness; Distance methods; Heterogeneity.

local density, but could not accurately be described as aggregated. Consider, as a specific example, the two patterns in Figures 1a and b. Both have been generated by a doubly stochastic Poisson process described by Matérn (1971) in which circles of equal radius r_0 are positioned at random over some plane area, and Poisson-distributed numbers of plants are then placed, independently and uniformly at random, within the individual circles. Both patterns utilise the same pseudo-random number streams, and are identical in every respect save that the radius of each circle is 0.5 in Figure 1a, but has been increased to 2.0 in Figure 1b. The pattern in Figure 1a clearly exhibits aggregation, whereas that in Figure 1b might perhaps be more accurately described as "locally random but globally heterogeneous," or "random-heterogeneous."

To emphasise the ambiguity between clustering and heterogeneity, we remark that this process has a dual interpretation. It is doubly stochastic, or heterogeneous, in the sense that the local density at a point $\mathbf{x}$ is proportional to the number, $N(\mathbf{x})$, of circles which contain $\mathbf{x}$; but $N(\mathbf{x})$ is itself a Poisson variate, with mean proportional to πr^2_0. Alternatively, the centre of each circle may be regarded as a cluster centre, around which individual plants are distributed independently according to the bivariate p.d.f.

$$f(r, \theta) = r/(\pi r^2_0),\ 0 \leq r \leq r_0;\ 0 \leq \theta \leq 2\pi.$$

One objection to the use of distance methods as tests of randomness, voiced for example by Mead (1974), is that they "are concerned essentially with small-scale patterns." This need not, however, imply a weakness against heterogeneity; we have suggested above, and will demonstrate in Section 3 that patterns arising from extreme heterogeneity will tend to be classified, reasonably, as aggregated, using the T-square Normal test, whereas this test will seldom detect random-heterogeneous patterns. We require, therefore, a supplementary procedure, which will aim to distinguish between random-heterogeneous and completely random, or "random-homogeneous" patterns.

2. *A Test for Random Heterogeneity*

We now formalise the notion of random heterogeneity which was introduced in the preceding section. We shall make the following assumption, subsequently denoted by A: the

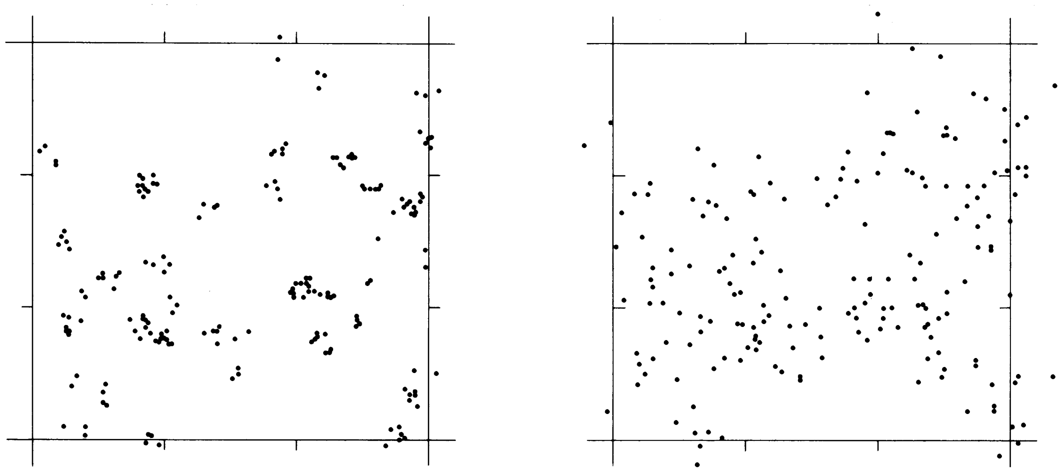


Figure 1a and 1b
A realisation of the Matérn process with individual circles of radius 0.5 (a) and 2.0 (b).

random variables $U_i = X^2_i + \frac{1}{2}Y^2_i$, obtained from each of m randomly selected points, are independent, with probability densities

$$f_i(u) = \lambda^2_i u\, e^{-\lambda_i u},\ u \geq 0.$$

Thus, assumption A corresponds to a gamma distribution for U_i, with shape and scale parameters 2 and λ_i respectively. The case $\lambda_i = \lambda$, $i = 1, \cdots, m$, corresponds to the random-homogeneous hypothesis, $\mathcal{H}_0$, whilst that of unequal λ_i represents the alternative hypothesis, $\mathcal{H}_1$ say, of random heterogeneity. Note that in either situation $X^2_i/(X^2_i + \frac{1}{2}Y^2_i)$ will be uniformly distributed on (0, 1) for each i, and the null distribution for t_N will apply, confirming the weakness of the T-square Normal test against random heterogeneity, as here defined.

We now consider the generalised likelihood ratio test of $\mathcal{H}_0$ against $\mathcal{H}_1$. Denoting by L_i the likelihood of the vector observation **u** under $\mathcal{H}_i$, and writing $M = -2 \log (\max L_0/\max L_1)$, we readily obtain

$$M = 4\ \{m \log (\sum U_i/m) - \sum \log U_i\}.$$

This is but a special case of Bartlett's (1937) statistic for testing the homogeneity of m estimated variances in Normal sampling, in which each variance estimate has four d.f. Accordingly, we follow Bartlett's recommendation that the statistic M should be divided by a factor C, equal to $(13m + 1)/(12m)$ in the present context, to improve the χ^2 approximation to the null distribution of the test statistic; thus we obtain a revised statistic,

$$M/C = 48m\ \{m \log (\sum u_i/m) - \sum \log u_i\}/(13m + 1),$$

whose null distribution is approximately χ^2 on $m - 1$ d.f.

Subsequent work on Bartlett's test by, among others, Hartley (1940) confirmed that the χ^2 approximation is quite adequate, albeit marginally conservative, in this situation. Let us now, however, consider the validity of assumption A, which was that the individual U_i are proportional to χ^2 variates with four d.f. Neither the independence nor the proportionality to χ^2 will often be satisfied exactly in practice, but the usefulness of A as an approximation may be checked by first conducting a test of randomness, using t_N. If this yields a non-significant result, the same data may be utilised for the M/C test for heterogeneity, which may therefore be regarded as the second half of a two-stage procedure in which the same data are used twice, and the quotation of "exact" significance levels would be misleading.

If we wish to detect non-randomness, we could omit the preliminary test, since any departure from A itself constitutes evidence of nonrandomness, and the question of spurious significance does not arise. However, we have indicated that our aim is rather to effect an approximate four-way characterisation of spatial point patterns as regular, random-homogeneous, random-heterogeneous or aggregated, for which the two-stage procedure becomes necessary. For this reason, the present procedure may be distinguished from those mentioned by Moore (1954) and Pollard (1971) which, although very similar in conception to the above, are based only on point-to-plant distance measurements, and make no reference to any preliminary test of randomness.

3. *Power and an Example*

An analytical investigation of the power of the two-stage procedure against plausible heterogeneous processes would present considerable difficulties and we shall therefore content ourselves with a simulation study of the Matérn (1971) process described in Section 1. We first distribute a prescribed number, n_0, of points uniformly and independently at random over a square of side x proportional to $\sqrt{n_0}$. Corresponding to each such point, we generate numbers of plants independently according to a Poisson distribution with mean μ. Each plant is then distributed uniformly and independently at random within the circle of radius r_0 centred on the corresponding random point. For each pseudo-random number stream, we produce parallel realisations for different values of r_0. The constraints $n_0 \geq 100$ and $n_0\mu \geq 400$ were imposed to produce populations of reasonable size. For each realisation, 25 sample points were positioned uniformly and independently at random within the centrally-placed interior square of side $4x/5$ to guard against, on the one hand over-intensive sampling and on the other, undesirable edge-effects (Diggle, Besag and Gleaves 1976). Figures 1a and b correspond, respectively, to $r_0 = 0.5$ and $r_0 = 2.0$, but with $n_0 = 40$ and $\mu = 5$; they give some indication of the change from "aggregation" to "heterogeneity" (and ultimately complete randomness) as r_0 increases.

The results for 100 sets of parallel realisations, summarised in Figures 2a, b and c, are consistent with the change in character of the patterns produced by the Matérn process as r_0 increases, and demonstrate the potential usefulness of the heterogeneity test as a supplement to the T-square Normal test of randomness.

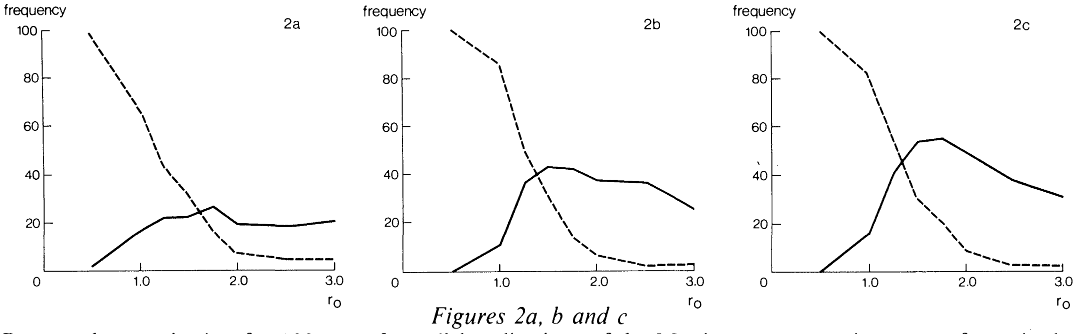
This potential usefulness is further demonstrated by the results which appear in Table 1. These relate to the application of the two-stage procedure, using a sample size $m = 25$ and tests of nominal significance level five percent, to Gerrard's Lansing Woods data, details of which may be found in Gerrard (1969). Separate analyses were performed on three sub-groups from the data—oaks, hickories and maples—which together embrace all but 105 of the 2251 trees included in Gerrard's listing.

4. Conclusions

The two-stage procedure proposed in this paper aims to remedy one defect of most tests of spatial randomness which use distance measurements, namely their weakness against heterogeneity. The results of Section 3 indicate that this aim is at least partially fulfilled. In contrast to the work of Mead (1974), the procedure cannot investigate pattern at different scales simultaneously, but will usually be a much less time-consuming operation in the field, and should provide a useful foundation for subsequent, more detailed analyses.

Acknowledgments

I would like to thank Mr. J. E. Besag, under whose supervision I began this work at the University of Liverpool, and Professor R. L. Plackett for his comments on an early draft of the paper. I am also extremely grateful to Dr. D. J. Gerrard, for giving me permission to use his data.



Figures 2a, b and c

Pattern characterisation for 100 sets of parallel realisations of the Matérn process, using tests of nominal size 5%---- Aggregated; —— Random heterogeneous; $a)\mu = 2$; $b)\mu = 5$; $c)\mu = 8$

Table 1
Analysis of Gerrard's Lansing Woods Data

Population	t_N	p-value (two-sided)	M/C	p-value (one-sided)	Characterisation
Oaks	0.398	0.078	29.25	0.211	random-homogeneous
Hickories	0.485	>0.5	40.24	0.020	random-heterogeneous
Maples	0.526	>0.5	64.93	<0.001	random-heterogeneous

References

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Statistical Society, A, 160,* 268–282.
Bartlett, M. S. (1964). Spectral analysis of two-dimensional point processes. *Biometrika 51,* 299–311.
Besag, J. E. and Gleaves, J. T. (1973). On the detection of spatial pattern in plant communities. *Bulletin of the International Statistical Institute 45(1),* 153–158.
Diggle, P. J., Besag, J. and Gleaves, J. T. (1976). Statistical analysis of spatial point patterns by means of distance methods. *Biometrics 32,* 659–667.
Gerrard, D. J. (1969). Competition quotient: a new measure of the competition affecting individual forest trees. Research Bulletin No. 20, Agricultural Experiment Station, Michigan State University.
Greig-Smith, P. (1964). *Quantitative Plant Ecology,* 2nd ed., Butterworth, London.
Hartley, H. O. (1940). Testing the homogeneity of a set of variances. *Biometrika 40,* 318–335.
Holgate, P. (1965). Some new tests of randomness. *Journal of Ecology 53,* 261–266.
Holgate, P. (1972). The use of distance methods for the analysis of spatial distributions of points. In *Stochastic Point Processes* P. A. W. Lewis, ed., Wiley, New York, 122–135.
Hopkins, B. (1954). A new method of determining the type of distribution of plant individuals. *Annals of Botany 18,* 213–226.
Matern, B. (1971). Doubly stochastic Poisson processes in the plane. In *Statistical Ecology, Vol. 1,* Patil, Pielou and Waters, ed., Pennsylvania State University Press, 195–213.
Mead, R. (1974) A test for spatial pattern at several scales using data from a grid of contiguous quadrats. *Biometrics 30,* 295–308.
Moore, P. G. (1954). Spacing in plant populations. *Ecology 35,* 295–308.
Pollard, J. H. (1971). On distance estimators of density in randomly distributed forests. *Biometrics 27,* 991–1002.

Received October 1976