On Sampling Spatial Patterns by Distance Methods

Author(s): Karen Byth and B. D. Ripley

Source: *Biometrics*, Vol. 36, No. 2 (Jun., 1980), pp. 279-284
Published by: International Biometric Society
Stable URL: http://www.jstor.org/stable/2529979

BIOMETRICS 36, 279–284
June 1980

SHORTER COMMUNICATIONS

EDITOR:

JOHN J. GART

On Sampling Spatial Patterns by Distance Methods

Karen Byth and B. D. Ripley
Department of Mathematics, Imperial College, London SW7 2BZ, England

SUMMARY

Hopkins' test for randomness involves choosing random trees and has been considered impracticable. We introduce a semisystematic sampling scheme which overcomes this objection and allows more intensive sampling. A related test statistic is introduced and a Monte Carlo study is presented which shows that the power of these tests compares favourably with that of other distance methods.

1. Introduction

'Distance' or 'nearest neighbour' methods are often used as alternatives to counting plants within squares ('quadrat' methods) either to estimate the number of plants in a study region or to test the 'randomness' of their pattern.To be specific we will consider trees; Figure 1 illustrates a 10 metre square plot of pines (from Strand, 1972).

The suggested procedures for density estimation and for testing have been compared by Diggle (1975, 1977), Diggle, Besag and Gleaves (1976), and Hines and Hines (1979). The method suggested by Hopkins (1954) involves measuring the following squared distances: u from a random point to the nearest tree and v from a randomly chosen tree to its nearest neighbour. Edge effects should be made negligible by placing the study region from which random points and plants are to be selected well within the region of interest. Hopkins' method has generally been preferred in comparison studies but is usually regarded as impracticable since, to find a random tree, all trees in the study region should be counted and a randomly numbered tree chosen. Squared distances arise because the areas swept out in searches for the nearest trees are πu and πv, if the search is thought of in concentric circles about the chosen point or tree. The distribution theory assumes that for a Poisson process the areas πu and πv will be independent. For this to hold it should be checked that the areas searched do not overlap, which constrains the number of samples that can be taken.

We consider bounds on the sampling intensity and we introduce a semisystematic sampling scheme which allows a higher sampling intensity and permits Hopkins' method to be used without complete enumeration of the study region. A new test for 'randomness' related to Hopkins' test is introduced; a Monte Carlo study shows that this test and

Key words: Distance methods; Power; Semisystematic sampling; Spatial randomness.

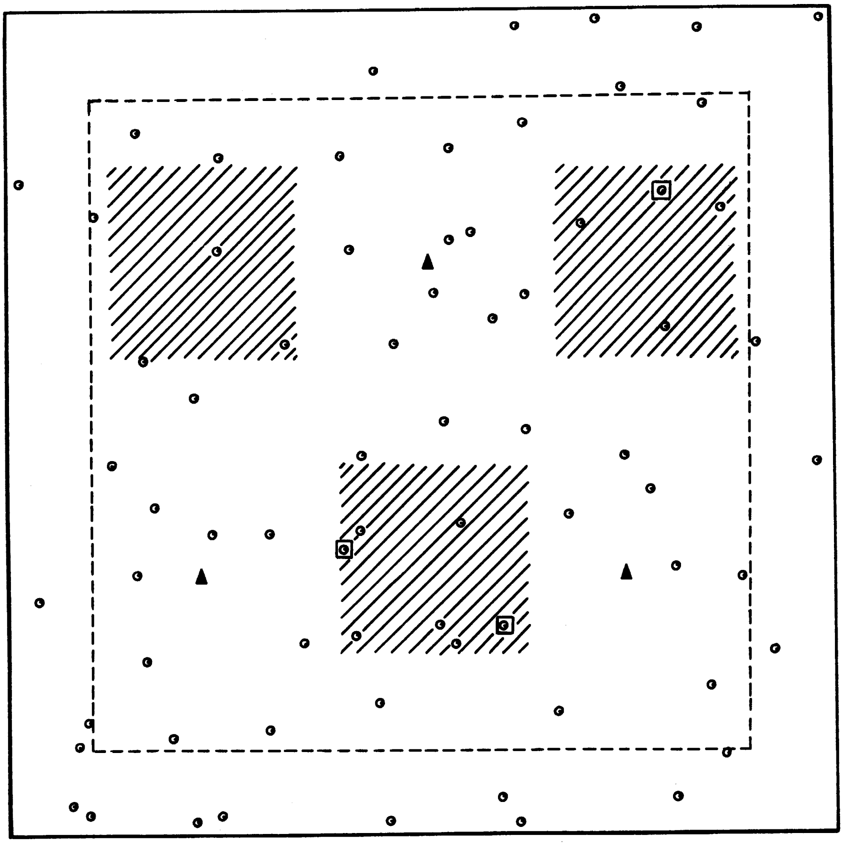


Figure 1. Semisystematic sampling of $m = 3$ sample points and $m = 3$ plants from a 10 metre square stand containing 71 pines ○ (Strand, 1972). ---- boundary of study area, ▨ enumerated area, ▲ sample point, ▣ selected plant.

Hopkins' original test are leading contenders. Diggle (1977) showed that density estimates based on Hopkins' sampling were usually best and we think that they should be at least as good with our sampling scheme.

In Fig. 1 we illustrate our recommended strategy to test for significant pattern in a region. We set up a fairly regular grid of $2m$ points within the study region (here chosen to be an 8 metre square lying 1 metre inside the region of interest). From half of these points, which we will refer to as the m sample points, we measure the squared distance to the nearest tree, obtaining measurements $u_1, \ldots, u_m$. Around each remaining point we lay out a small plot of a size which would contain about five trees on average and we count the trees in each plot. We then select m trees at random from those enumerated and measure the squared distances $v_1, \ldots, v_m$ from each to its nearest neighbour. Calculate

$$\mathrm{Hop}_F = \sum u_i / \sum v_i \quad \text{and} \quad \mathrm{Hop}_N = (1/m) \sum \{u_i/(u_i + v_i)\},$$

and use these as tests of 'randomness', employing Hop_F when clustering is suspected and Hop_N for regular alternatives. For Hop_N the u_i and v_i are paired at random.

The pattern of Fig. 1 shows slight regularity when tested by the methods of Ripley (1977, 1979b). If a complete map is available, distance methods of the type considered here should not be used.

2. Distribution Theory

Suppose the trees were generated by a Poisson process giving an intensity of λ trees per unit area. Then the area swept out in searching for the nearest tree from an arbitrary point has an exponential distribution, mean $1/\lambda$. In particular, this will be so if the distance is measured from a position which is uniformly distributed over the study region. A point or a tree chosen at random or by our semisystematic scheme is such a position. We shall be justified in regarding the measurements as independent if there is a negligible probability that two swept-out areas overlap. This probability is about 5% if the minimum distance between the sample positions is at least $3/\sqrt{(\pi\lambda)}$. This can be checked in the field as the usual estimate of $1/\sqrt{(\pi\lambda)}$ is the root mean square distance from a point to the nearest tree. For m sites (either points or trees) selected by random sampling, the average distance between sites is about $\sqrt{(A/\pi m)}$ if the study region has area A. Thus, for random sampling, the number of sample sites certainly should not exceed 10% of the number of trees in the study region. This conclusion is confirmed by our simulation study reported below. For semisystematic sampling the average distance between the sites becomes $\sqrt{(A/m)}$, so the sampling intensity can be increased by a factor of about three.

The tests Hop_F and Hop_N were defined in §1. Hop_F was introduced by Hopkins (1954) with Skellam showing that for a Poisson forest Hop_F has an F distribution with $(2m, 2m)$ degrees of freedom; Hop_N is a new test, its distribution is the average of m independent variables each uniformly distributed on $[0, 1]$ and so it is well approximated by the Normal distribution with mean $\frac{1}{2}$ and variance $m/12$. We considered a variant of this test in which $u_i/(u_i + v_j)$ was summed over all pairs i and j. This had almost identical power and so was discarded.

Various test statistics have been based on T-square sampling. A random point is chosen, the squared distance u to the nearest tree is measured, then the squared distance t from this tree to its nearest neighbour in a direction away from the initial point is measured. For a Poisson forest, u and $\frac{1}{2}t$ are independent, each having an exponential distribution with mean $1/(\pi\lambda)$. We denote samples obtained by this procedure by $u_1, \ldots, u_m$ and $t_1, \ldots, t_m$.

Besag and Gleaves (1974) introduced

$$T_F = \sum u_i / \sum (\tfrac{1}{2}t_i) \quad \text{and} \quad T_N = (1/m) \sum \{u_i/(u_i + \tfrac{1}{2}t_i)\},$$

which have the same distributions as their Hop analogues. Hines and Hines (1979) recommend using

$$T_E = 2m \sum (2u_i + t_i) \Big/ \Big[\sum\{\sqrt{(2u_i)} + \sqrt{t_i}\}\Big]^2,$$

for which they give a table of percentage points.

The tests proposed by Holgate (1965) were also considered, but the results were poor and are not presented. Hines and Hines show that the statistic $\bar{r}$ of Cox and Lewis (1976) is very similar to T_N.

3. Monte Carlo Study

We tested the five statistics Hop_F, Hop_N, T_F, T_N and T_E for Poisson processes and for the clustered and regular alternatives with the usual random sampling scheme and semisystematic sampling.

For m sample points and n trees in the study area, we define the sampling intensity $\rho = m/n$. Diggle *et al.* (1976) reported consistency with the null distributions of Hop_F, T_F and T_N for $\rho \leq 10\%$ on the basis of 57 realizations with $m = 25$. Our results based on a variety of combinations of ρ and λ suggest that 10% is an acceptable upper bound for T-type statistics but is too high for Hop-type. With $m=9$ we observed 71 and 68 values out of 1000 Hop_F and Hop_N, respectively, beyond the lower 5% point of the appropriate proposed theoretical distribution; the bound $\rho \leqslant 5\%$ seems adequate. Since hop-type statistics need two separate sets of m locations we should expect a smaller bound. When semisystematic sampling was used ρ could be at least as high as 25% for T-type statistics. For Hop-type statistics $\rho \leqslant 10\%$ sufficed if half the area was enumerated and $\rho \leqslant 5\%$ if one quarter was enumerated. These bounds correspond to enumerating five trees per sample point.

For the rest of the simulation study we took $m=9$, $\rho \leqslant 5\%$ and considered 5% equal-tailed tests. The results are presented in Table 1, each power estimate being based on 100 or 200 realizations.

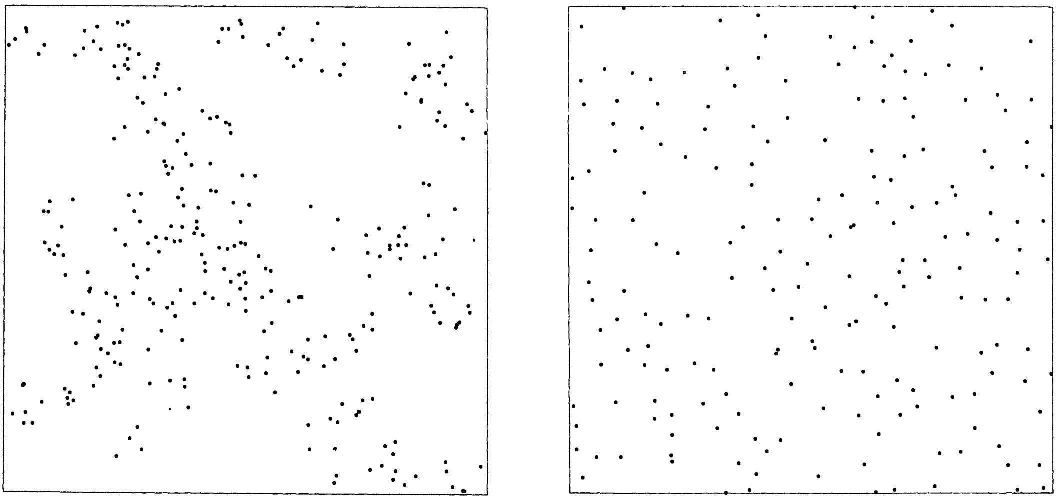
Matérn cluster processes (Matérn, 1971), modified to contain a fixed number of objects, N, were used as clustered alternatives. Realizations were simulated by generating a Poisson number of cluster centres, N_C, mean μ, and distributing these independently and uniformly within the region of interest. The remaining $N-N_C$ objects were then distributed uniformly within a disc of diameter R centred on a cluster centre chosen at random, independently for each object. If the object so added fell outside the region of interest, the process was repeated for that object. Figure 2a illustrates a realization.

Strauss processes (Kelly and Ripley, 1976; Ripley, 1977, 1979a) were used as regular alternatives, with R measuring the range of inhibition and C the strength.The case $C=0$ corresponds to nonoverlapping discs of diameter R about each object. Figure 2b illustrates a realization of a Strauss process.

Table 1
Estimated power in percent against clustered and regular alternatives

Clustered alternatives								
Mean size of cluster	Diameter R	Hop_F		Hop_N		T_F	T_N	T_E
		Random	Semi-systematic	Random	Semi-systematic			
2	0.05	58	35	42	26	43	36	36
	0.02	80	55	92	67	50	79	85
3	0.05	85	71	67	65	70	65	68
	0.03	97	87	95	92	86	92	96
6	0.15	40	37	19	16	23	11	29
	0.10	69	75	45	48	65	36	59
	0.07	96	98	69	74	91	76	88
Regular alternatives								
Strength C	Range R							
0	0.03	19	20	50	47	11	20	22
	0.04	49	48	84	80	18	33	45
0.1	0.04	33	38	48	46	10	24	30
0.25	0.04	16	23	21	31	10	18	15

The results under both semisystematic and random sampling are given for Hop_F and Hop_N but only the average power is tabulated for T_F, T_N and T_E.



(a) (b)

Figure 2. A realization on the unit square of (a) a modified Matérn process having mean cluster size 6 and cluster diameter 0.15, and (b) a Strauss process having 200 points, strength of inhibition 0.1 and range of inhibition 0.035.

The sampling scheme used did not appreciably affect the power for T-type statistics, and the average of the results for semisystematic and random sampling is presented. We recommend the semisystematic scheme as a precaution against oversampling. For clustered alternatives, Hopkins' original test is best with both Hop_F and T_E doing well under semisystematic sampling. For regular alternatives, Hop_N is clearly best and does not appear to lose power when semisystematic sampling is used. It is interesting to note the substantial decrease in power against regular alternatives for all the statistics when the strength of inhibition is slightly weakened from the case of nonoverlapping discs, $C=0$. For example, compare $C=0$ with $C=0.1$ for $R=0.04$. Both Diggle *et al.* (1976) and Hines and Hines (1979) considered as regular alternatives only simple sequential inhibition processes which are very similar to our case $C=0$.

This study shows that Hop-type statistics are worthy of consideration. The semisystematic scheme allows Hop-type tests to be used without complete enumeration and allows more intensive sampling if complete enumeration or T-type tests are used.

Résumé

Le test de Hopkins pour étudier le caractère aléatoire implique un choix aléatoire des arbres et on l'a considéré comme impossible à appliquer. Nous introduisons un plan d'échantillonnage semisystématique qui permet de surmonter cette objection et qui autorise un échantillonnage plus intensif. On introduit un test statistique associé et on présente une étude de Monte Carlo qui montre la puissance de ces tests dont la comparaison par rapport aux autres méthodes de distances est avantageuse.

References

Besag, J. E. and Gleaves, J. T. (1974). On the detection of spatial pattern in plant communities. *Bulletin of the International Statistical Institute* **45** (1), 153–158.

Cox, T. F. and Lewis, T. (1976). A conditioned distance ratio method for analyzing spatial patterns. *Biometrika* **63,** 483–491.

Diggle, P. J. (1975). Robust density estimation using distance methods. *Biometrika* **62,** 39–48.
Diggle, P. J. (1977). A note on robust density estimation for spatial point patterns. *Biometrika* **64,** 91–95.
Diggle, P. J., Besag, J. E. and Gleaves, J. T. (1976). Statistical analysis of spatial point patterns by means of distance methods. *Biometrics* **32,** 659–667.
Hines, W. G. S. and Hines, R. J. O'H. (1979). The Eberhardt statistic, and the detection of non randomness of spatial point distributions. *Biometrika* **66,** 73–79.
Holgate, P. (1965). Some new tests of randomness. *Journal of Ecology* **53,** 261–266.
Hopkins, B. (1954). A new method of determining the type of distribution of plant individuals. *Annals of Botany* **18,** 213–227.
Kelly, F. P. and Ripley, B. D. (1976). A note on Strauss's model for clustering. *Biometrika* **63,** 357–360.
Matérn, B. (1971). Doubly stochastic Poisson processes in the plane. In *Statistical Ecology, Vol.* 1. G. P. Patil, E. C. Pielou and W. E. Waters (eds), 195–206. University Park: Pennsylvania State University Press.
Ripley, B. D. (1977). Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society, Series B* **39,** 172–212.
Ripley, B. D. (1979a). Simulating spatial patterns: dependent samples from a multivariate density. *Applied Statistics* **28,** 109–112.
Ripley, B. D. (1979b). Tests of 'randomness' for spatial point patterns. *Journal of the Royal Statistical Society, Series B* **41,** 368–374.
Strand, L. (1972). A model for stand growth. *IUFRO Third Conference Advisory Group of Forest Statisticians,* 207–216.

Received July 1979; *revised October* 1979