A Note on Robust Density Estimation for Spatial Point Patterns
Author(s): Peter J. Diggle
Source: *Biometrika*, Vol. 64, No. 1 (Apr., 1977), pp. 91-95
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: http://www.jstor.org/stable/2335776
Accessed: 03-11-2015 23:42 UTC

Biometrika (1977), **64**, 1, *pp.* 91–5
With 3 *text-figures*
Printed in Great Britain

A note on robust density estimation for spatial point patterns

By PETER J. DIGGLE
Department of Statistics, University of Newcastle upon Tyne

Summary

The author's earlier analytical results on the robustness of distance-based estimators of density are supplemented by a simulation study, with particular attention being given to various types of aggregated spatial point patterns. In addition, some remarks on heterogeneous patterns are made. Finally, the results of an application to data on tree locations are described.

Some key words: Density estimation; Distance method; Ecology; Robustness; Spatial distribution.

1. Introduction

Let γ denote the mean area per plant in a population, to be regarded as a partial realization of a two-dimensional point process, and let X and Y represent the random point-to-plant and plant-to-plant nearest neighbour distances, respectively. In a previous paper (Diggle, 1975), the author has shown the estimator

$$\gamma^* = (\pi/m)\left(\sum_{i=1}^{m} x_i^2 \sum_{i=1}^{m} y_i^2\right)^{\frac{1}{2}}$$

to be comparatively robust against two alternative classes of point process which together embrace a continuous range of variation in pattern, from extreme regularity as exemplified by deterministic lattice patterns, through complete spatial randomness to the aggregated extreme of randomly distributed point clusters. In practice, random plant-to-plant measurements are, of course, unobservable until a complete enumeration of the population has been made. However, let P be a random point and Q the nearest plant, so that $PQ = X$; the T-square nearest neighbour distance, Z say, is defined to be the distance from Q to the nearest plant, within the half-plane which is defined by the line through Q perpendicular to PQ and which excludes the point P (Besag & Gleaves, 1973). If we now replace y_i^2 by $\frac{1}{2}z_i^2$, we obtain a second estimator

$$\gamma_T^* = (\pi/m)\left(\sum_{i=1}^{m} x_i^2 \sum_{i=1}^{m} \tfrac{1}{2}z_i^2\right)^{\frac{1}{2}},$$

which will be stochastically equivalent to γ^* in the completely random case.

We now assess the robustness of γ^* and γ_T^* by a simulation study, which extends the results obtained previously by analytical methods. In this assessment, we measure robustness by the standardized root mean squared error, $R(\hat{\gamma}) = (E[\{(\hat{\gamma}-\gamma)/\gamma\}^2])^{\frac{1}{2}}$.

2. A simulation study

As always, considerable care must be taken to ensure that the simulation procedures used give a reliable representation of the underlying processes. Details will be omitted here, but may be obtained from the author. Throughout, we consider a sample size $m = 25$ and use 100 realizations of each process to estimate $R(\hat{\gamma})$. The author's earlier analytical results for γ^* provide a standard against which the robustness of γ^* and γ_T^* may be assessed, and are represented in Figs. 1 and 2 by a solid, smooth curve. Note in particular that the smooth curve in Figs. 1(*b*) and 2(*b*) refers to the estimator γ^*, and not to γ_T^*.

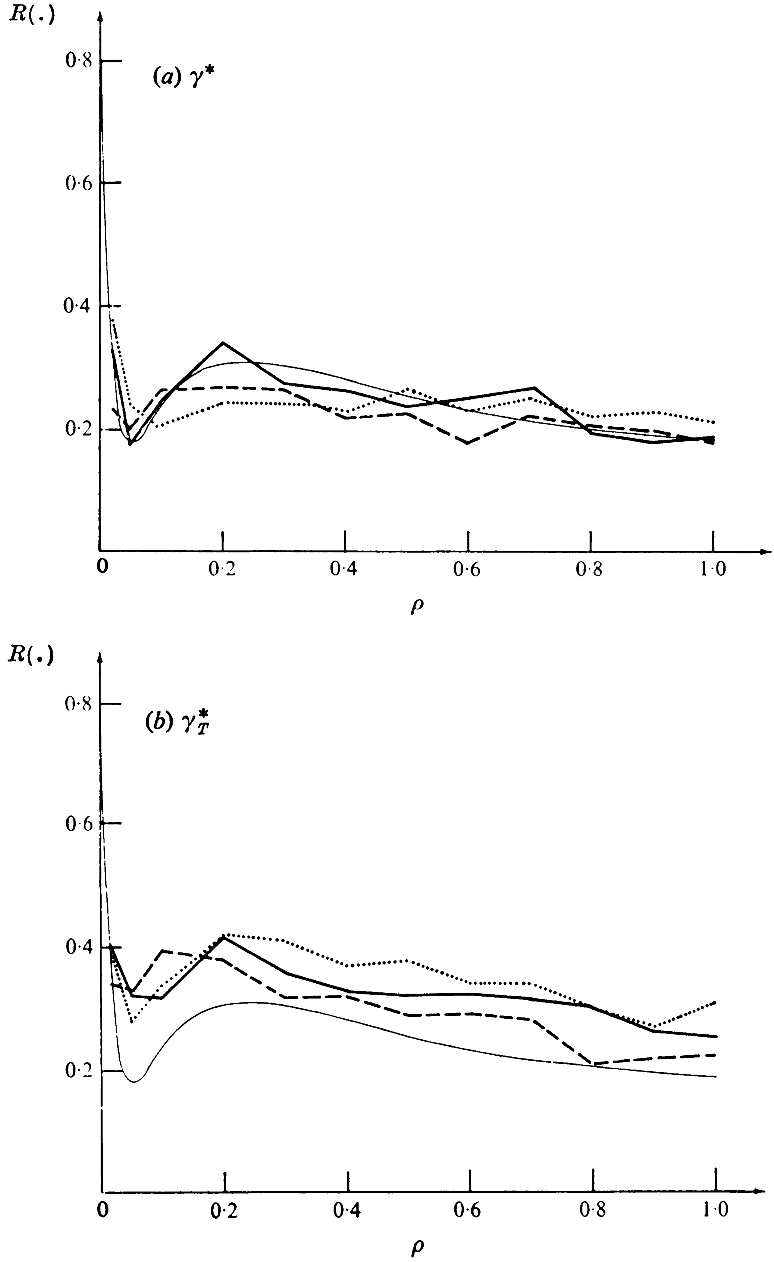

Fig. 1. Root mean squared error for alternative spatial dispersion mechanisms: ——, semi-deterministic; -----, uniform; ····· normal; solid, smooth curve relates to analytical results for γ^*.

For regular patterns, there is an essentially stable situation and we contend that further investigation of the robustness of γ^* is unnecessary. With regard to γ_T^*, we note that simulations of a regular lattice with superimposed Poisson process suggest that γ_T^* is, in fact, superior to γ^*. Again, further details may be obtained on request.

Aggregated patterns may be generated by considering a Poisson process of parents each of which, independently, gives rise to a random number of offspring; the offspring are, again independently, spatially distributed relative to the parent. The final pattern consists of the

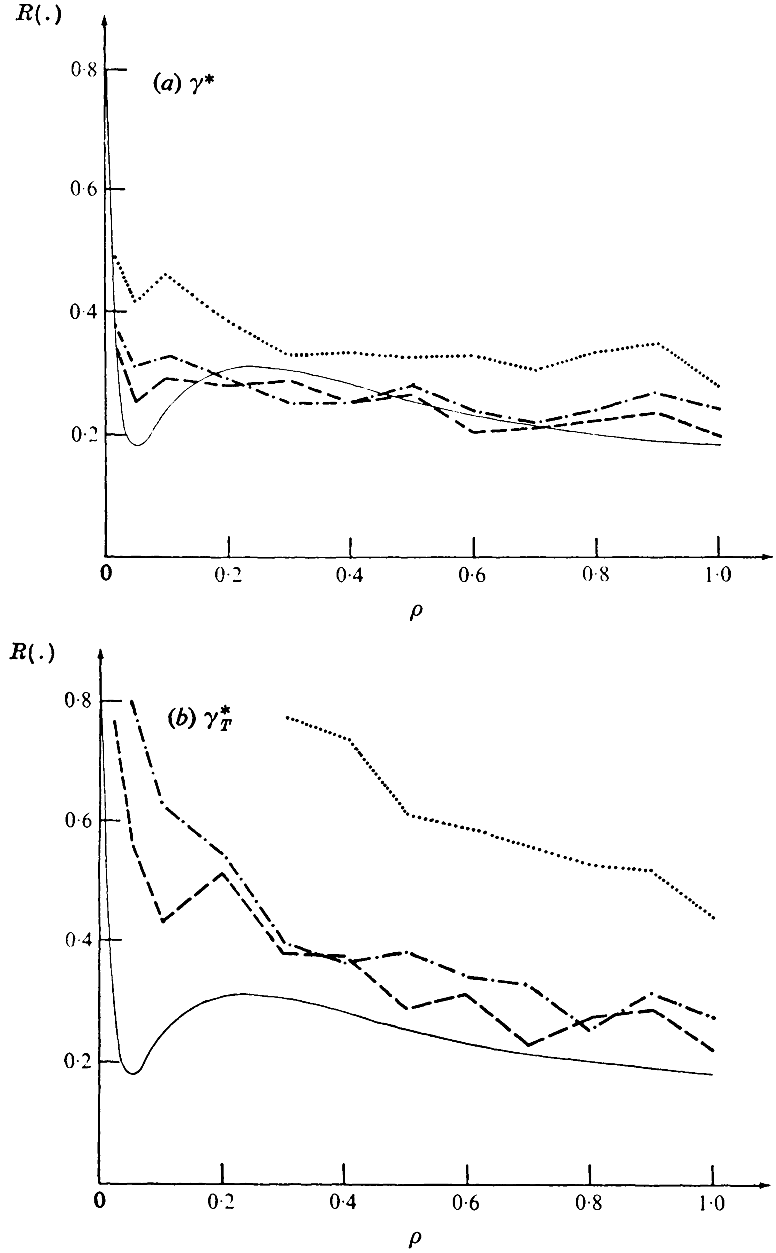

Fig. 2. Root mean squared error for zero-modified Poisson cluster size distribution: - - - - -, $p = 0{\cdot}3$; —·—·—·—, $p = 0{\cdot}5$; ·····, $p = 0{\cdot}8$; solid, smooth curve relates to analytical results for γ^*.

parents and offspring, which are assumed to be indistinguishable. Clearly, we should investigate the effects of modifications to two fundamentally different distributions, those of cluster size and spatial dispersion within a cluster.

The analytical results correspond to a Poisson distribution, with mean μ say, for the number of offspring per parent and an extreme bowl-shaped, or 'semideterministic' spatial dispersion mechanism whereby the radial dispersion is unity and the angular dispersion uniform on $(0, 2\pi)$. A natural counterpart is the bell-shaped normal, with a flat-topped uniform as an intermediary; radial symmetry and a mean squared radial dispersion equal to unity are imposed

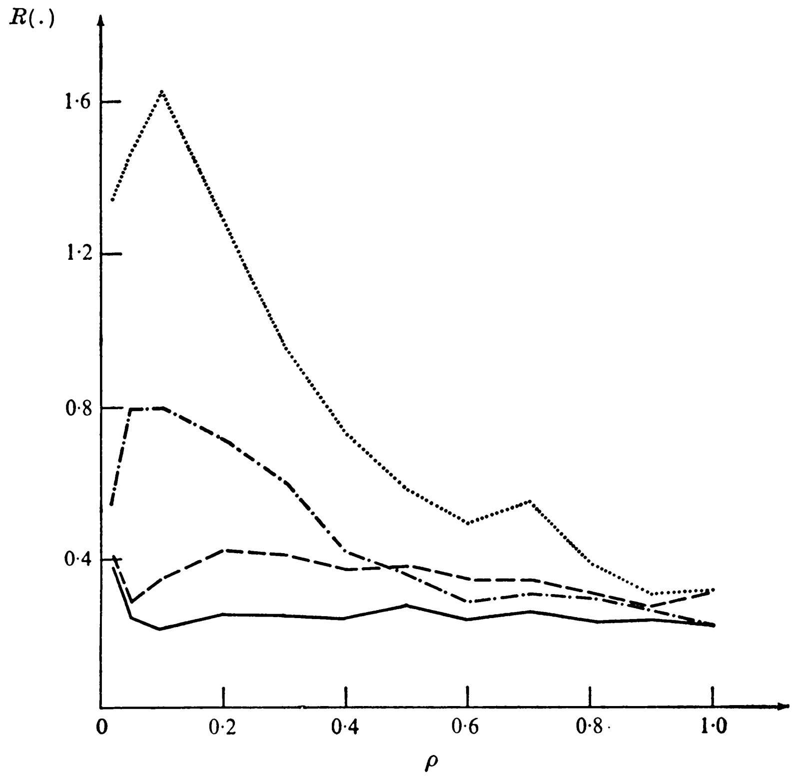

Fig. 3. Root mean squared error for normal spatial dispersion; number of offspring per parent Poisson distributed with mean μ: ——, γ^*, $\mu = 3$; -----, γ_T^*, $u = 3$; -·-·-·-, γ^*, $\mu = 20$; ·····, γ_T^*, $\mu = 20$

in both cases, to achieve direct comparability. Figure 1, which relates to $\mu = 3$, suggests that neither estimator is particularly sensitive to changes in the spatial dispersion mechanism, although there is an indication that γ_T^* is now rather less robust than is γ^*. The parameter ρ denotes the mean number of parents per unit area; complete spatial randomness results in the limit $\rho \to \infty$. Note, in Fig. 1(*a*), that the simulation results for γ^* applied to the semi-deterministic model are included for comparison with the analytical curve.

We now consider the number of offspring per parent to be distributed according to

$$p_n = \begin{cases} p + (1-p)\, e^{-\nu} & (n = 0), \\ (1-p)\, e^{-\nu} \nu^n / n! & (n = 1, 2, \ldots), \end{cases}$$

for various values of p and ν such that the mean assumes a constant value 3. This provides a severe test of γ_T^*, whose potential weakness lies in its tendency to focus on the more isolated plants in the population. Figure 2 confirms this weakness when $p = 0{\cdot}8$, although the performance of γ_T^* for p as large as 0·5 is encouraging. On the other hand, γ^* is markedly more robust than is γ_T^* and shows only slight deterioration in performance with increasing p.

Finally, in Fig. 3, we consider the effect of large mean cluster size together with a relatively diffuse spatial dispersion mechanism, the symmetric radial normal. Again, γ_T^* fares ill in comparison with γ^* as it tends to sample from the relatively less dense peripheries of large clusters.

3. Heterogeneous patterns

Notwithstanding that a statistical distinction between clustering and heterogeneity may be drawn only with some difficulty, or possibly not at all (Bartlett, 1964), further insight into the robustness of γ^* and γ_T^* may be gained by considering a population divided into k subareas A_i containing numbers n_i of plants, within each of which the underlying process is completely random. Writing $A = \Sigma A_i$ and $n = \Sigma n_i$, we see that $\gamma = A/n = \Sigma(n_i/n)(A_i/n_i) = \Sigma(n_i/n)\gamma_i$, say; all summations are over $i = 1, ..., k$. Thus, any estimator for γ_i which is unbiased for a completely random process will provide an unbiased estimator for γ, if it can be applied in such a way as to ensure that sampling takes place within the ith subarea with probability $p_i = n_i/n$. This is, of course, precisely what would be achieved by random plant-to-plant nearest neighbour sampling. On the other hand, random point-to-plant and T-square nearest neighbour sampling correspond to $p_i = A_i/A$, which induces an element of positive bias into the estimator for γ. Relatively diffuse clustering processes with large mean cluster size may, as a first approximation, be viewed in this alternative light, and comparison may be made with Fig. 3.

4. Discussion

The results of the present study suggest that, with the exception of regular departures from complete spatial randomness, γ_T^* is generally less robust than is γ^*. Although this is disappointing, the cases in which γ_T^* performs particularly poorly are relatively extreme and may furthermore be detected by carrying out associated tests of spatial randomness based on the same data as are used to calculate γ_T^*. In particular, the tests proposed by Besag & Gleaves (1973) are designed to detect regularity or aggregation in the underlying spatial point pattern, while Diggle (1977) suggests a method for detecting the type of heterogeneity described in § 3 above. We therefore suggest that T-square sampling can be a useful tool in the preliminary analysis of field data to provide both an estimate of density and a characterization of spatial pattern.

To illustrate this procedure we have applied the T-square sampling method to a map of 2251 trees divided by species into six smaller populations. Notice that the estimator γ_T^* is a function of two dependent sample means; thus, the delta technique may be used to calculate an approximate standard error for the observed value of γ_T^* and the range $\gamma_T^* \pm 2$ std err. (γ_T^*) used as an interval estimate for γ in each case. In the event, these intervals were found to exclude the known value of γ only in conjunction with the detection of heterogeneity at the 5% level of significance, and we would claim that such interval estimates will generally be reasonable unless the associated tests of spatial randomness suggest heterogeneity, when the corresponding point estimate will tend to be too large, or extreme aggregation, in which case the concept of mean area per plant for the population as a whole is of limited relevance. Again, further details may be obtained from the author.

I am grateful to Professor D. G. Gerrard for providing data on tree locations.

References

Bartlett, M. S. (1964). Spectral analysis of two-dimensional point processes. *Biometrika* **51**, 299–311.

Besag, J. E. & Gleaves, J. T. (1973). On the detection of spatial pattern in plant communities. *Bull. Inst. Int. Statist.* **45**, 153–8.

Diggle, P. J. (1975). Robust density estimation using distance methods. *Biometrika* **62**, 39–48.

Diggle, P. J. (1977). The detection of random heterogeneity in plant populations. *Biometrics* **33**. To appear.

[*Received April* 1976. *Revised July* 1976]