Robust Density Estimation Using Distance Methods
Author(s): Peter J. Diggle
Source: *Biometrika*, Vol. 62, No. 1 (Apr., 1975), pp. 39-48
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: http://www.jstor.org/stable/2334485
Accessed: 03-11-2015 23:42 UTC

Biometrika (1975), **62**, 1, *p.* 39
With 4 text-figures
Printed in Great Britain

Robust density estimation using distance methods

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

SUMMARY

Distance estimators of density may exhibit serious bias unless the population under consideration forms a completely random spatial pattern, i.e. the estimators are not robust. In this paper some new estimators are proposed, and their robustness is assessed analytically against two stochastic models, which together embrace a continuous range of spatial pattern, from extreme regularity, through randomness, to extreme aggregation.

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

1. INTRODUCTION

The use of distance methods for estimating the density of plant populations, envisaged as spatial point patterns, has been widely discussed (see, for example, Persson, 1964; Holgate 1972), but such estimators have been found lacking in robustness (Persson, 1971).

In particular, let X and Y represent the random point-to-plant and plant-to-plant nearest neighbour distances, respectively. 'Simple' estimators for the mean area, γ say, per plant are defined as

$$\hat{\gamma}_x = \pi \sum_{i=1}^{n} x_i^2/n, \quad \hat{\gamma}_y = \pi \sum_{i=1}^{n} y_i^2/n,$$

each of which will be unbiased, and fully efficient, when the plants are distributed completely at random, i.e. they constitute a realization of a two-dimensional, homogeneous, Poisson point process, or 'Poisson forest'. When the underlying spatial pattern is not completely random, $\hat{\gamma}_x$ and $\hat{\gamma}_y$ will tend to be biased in opposite directions, since squares of distances from random points to nearest plants tend to be increased by aggregation and decreased by regularity in the underlying pattern, while the reverse is true of squares of distances from random plants to their nearest neighbours. Thus, some form of average of $\hat{\gamma}_x$ and $\hat{\gamma}_y$ should result in improved robustness. For example, we could define a 'compound' estimator

$$\hat{\gamma}(r) = r\hat{\gamma}_x + (1-r)\hat{\gamma}_y \quad (0 < r < 1).$$

In particular, the estimator $\tilde{\gamma} = \frac{1}{2}(\hat{\gamma}_x + \hat{\gamma}_y)$ will be optimal among all $\hat{\gamma}(r)$ for a completely random pattern, when $\hat{\gamma}_x$ and $\hat{\gamma}_y$ are based on samples of equal size. Other values of r may, however, result in greater robustness. For example, in a strongly aggregated pattern, $\hat{\gamma}_x$ may exhibit arbitrarily large positive bias, whereas $\hat{\gamma}_y$ must at least be non-negative. Because of this asymmetry, a linear combination of $\hat{\gamma}_x$ and $\hat{\gamma}_y$ may not always be the most efficient form of average, and we shall also consider the compound estimator

$$\gamma^* = \sqrt{(\hat{\gamma}_x \hat{\gamma}_y)}.$$

The estimators described above will be termed estimators based on Hopkins sampling, after Hopkins (1954) who suggested the combined use of point-to-plant and plant-to-plant

nearest neighbour distances to provide a test of randomness. While the truly random selection of a plant from the population is impossible in the present context, the properties of such estimators will provide a useful pointer to the performance of analogous estimators, based on a practical, but comparatively intractable, alternative sampling procedure, which will be considered in § 4.

Our discussion of robustness will be concerned primarily with the mean standardized bias,

$$B(\hat{\gamma}) = E\{(\hat{\gamma}-\gamma)/\gamma\},$$

which should remain small in absolute value for a wide variety of underlying spatial patterns. A secondary requirement will be that the variance of $\hat{\gamma}$ should not be too large, for a grossly inefficient estimator would be rejected as unsatisfactory on this basis alone.

2. DENSITY ESTIMATION FOR REGULAR PATTERNS

Extreme examples of regular spatial patterns are furnished by populations in which a plant occurs at each vertex of a regular lattice of unit side. A more flexible model would hold greater intuitive appeal, and is obtained if we superimpose upon the lattice structure a realization of a Poisson forest with mean number of plants per unit area equal to ρ, thus providing a continuous range of patterns, from extreme regularity when ρ is zero, towards complete randomness as ρ tends to infinity, against which the robustness of the estimators based on Hopkins sampling may be investigated analytically.

Consider first the distribution of the squared point-to-plant nearest neighbour distance, U say. For the square lattice of unit side, a geometrical argument readily provides the distribution function of U, given by Persson (1964) as

$$G(u) = \begin{cases} \pi u & (0 \leqslant u < \tfrac{1}{4}), \\ 2u\sin^{-1}\{(1-2u)/(2u)\} + \sqrt{(4u-1)} & (\tfrac{1}{4} \leqslant u < \tfrac{1}{2}), \\ 1 & (u \geqslant \tfrac{1}{2}). \end{cases} \tag{2.1}$$

The distribution function of U in a realization of the Poisson forest is, for $u \geqslant 0$, $F(u) = 1 - e^{-\pi\rho u}$. When the Poisson forest is superimposed upon the lattice, U is the minimum of two random variables, with distribution functions F and G, and must therefore have distribution function

$$\begin{aligned} H(u) &= F(u) + G(u) - F(u)\,G(u) \\ &= 1 - e^{-\pi\rho u}\{1 - G(u)\}, \end{aligned}$$

an explicit expression for which may be obtained by substitution from (2·1).

The squared plant-to-plant nearest neighbour distance, V say, will also have distribution function H, conditional on the randomly selected plant being in the realization of the Poisson forest, as will occur with probability $p = \rho/(1+\rho)$. Otherwise, the nearest 'lattice-plant' must be unit distance away, and the distribution function of V will be

$$J(v) = \begin{cases} 1 - e^{-\pi\rho v} & (v < 1), \\ 1 & (v \geqslant 1). \end{cases}$$

The distribution function of V in the superimposed process is therefore

$$K(v) = pH(v) + (1-p)\,J(v).$$

The moments of U and V may now be obtained from their respective distribution functions, using a repeated Simpson's rule for the necessary numerical integrations. Figure 1 gives the mean standardized biases of the various estimators based on Hopkins sampling, as functions of ρ. The improved robustness of the averaging estimators in comparison with the simple estimators is clear, as is the superiority of γ^* over $\tilde{\gamma}$. Also γ^* is more efficient than $\tilde{\gamma}$ for all ρ greater than about 0·04; indeed, its variance is never greater than the limiting value of $\gamma^2/(2n)$, approached as ρ tends to infinity, the completely random case. For the general linear averaging estimator $\hat{\gamma}(r)$, the value $r = 0{\cdot}8$ represents a compromise over the whole range of ρ. A slightly smaller value for r would reduce the negative bias which prevails beyond $\rho \simeq 0{\cdot}09$, but at the expense of a proportionately greater increase in the positive bias for smaller ρ.

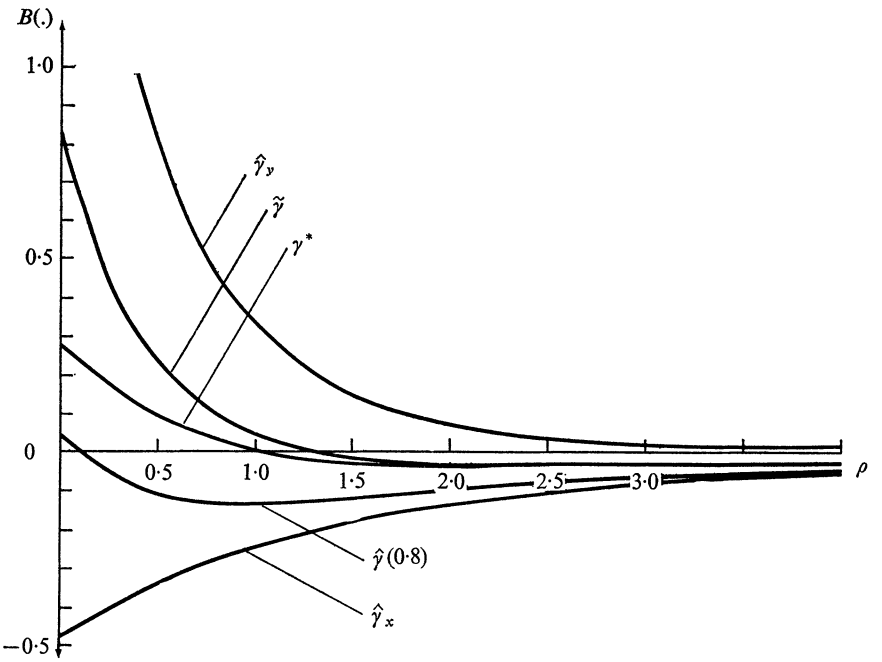

Fig. 1. Mean standardized bias, for density estimators based on distance methods, applied to the square lattice plus Poisson forest superimposed process.

The corresponding analysis for a superimposed process of Poisson forest plus triangular lattice proceeds along similar lines, and a closely comparable set of results may be obtained, details of which are available on request. Further discussion of regular spatial patterns, in a geographical context, is given by Dacey (1971).

3. Density estimation for aggregated spatial patterns

The case of aggregation embraces a much wider variety of patterns than does regularity, but analytical progress is possible in certain cases. Consider first the extreme case in which plants occur in tight clusters, each plant within a cluster occupying the same physical point. For example, Thomas (1949) postulated a completely random distribution of parents, each producing, independently, a Poisson number, with mean μ, of offspring. Each offspring occupies the same point as its parent, and is indistinguishable from it. Thus, cluster size is

one plus a Poisson variate. The point-to-plant nearest neighbour distance has the same distribution, apart from a change in the scale parameter, as in the completely random case, but the plant-to-plant nearest neighbour distance will be zero with probability $1-e^{-\mu}/(1+\mu)$, and will otherwise have the same distribution as in the random case, again apart from the value of the scale parameter. While, inevitably, all the estimators perform poorly in this situation when μ is large, it will be seen later that a reasonable degree of robustness may be achieved for moderate values of μ.

A more realistic model for aggregated spatial patterns would allow each offspring to be spatially distributed around its parent, as considered by Neyman & Scott (1958), Bartlett (1964) and Warren (1971). Bartlett (1974) gave general expressions for the distribution functions of both the point-to-plant and plant-to-plant nearest neighbour distances, when the angular displacement of each offspring relative to its parent is, independently, uniformly distributed on $(0, 2\pi)$, the radial displacement being independent of the angular displacement, but otherwise arbitrary. The point-to-plant distance, X say, has distribution function

$$F(x) = 1-\exp\left(-\pi\rho\left[x^2+2\int_x^\infty \{1-P(y;x)\}\,y\,dy\right]\right),$$

where ρ represents the mean number of parents per unit area, and $P(y;x) = \text{pr}$ (no offspring are found in the circle with centre the origin and radius x, from a parent a distance $y > x$ from the origin). The plant-to-plant distance, Y say, then has distribution function $G(y)$ given by

$$1-G(y) = \{1-F(y)\}\,P_2(y), \tag{3·1}$$

where $P_2(y) = \text{pr}$ (no parent or offspring from the same cluster as the random sample plant is found less than a distance y from the random sample plant), since the density of the process is unaltered by the removal of a single cluster from a conceptually infinite population.

The probabilities $P(y;x)$ and $P_2(y)$ will usually each involve an extremely awkward integral, but may be evaluated explicitly if we consider a semi-deterministic clustering process in which the radial displacement of each offspring relative to its parent is deterministically equal to unity. For each value of μ this process affords a continuous range of variation, from the Thomas process to complete randomness, as ρ increases from zero to infinity.

Considering first the point-to-plant distance X, we note that now $P(y;x) = 1$ for all $y \geqslant x+1$, while for $x \leqslant y \leqslant x+1$ we must distinguish between two cases; see Fig. 2*a*.

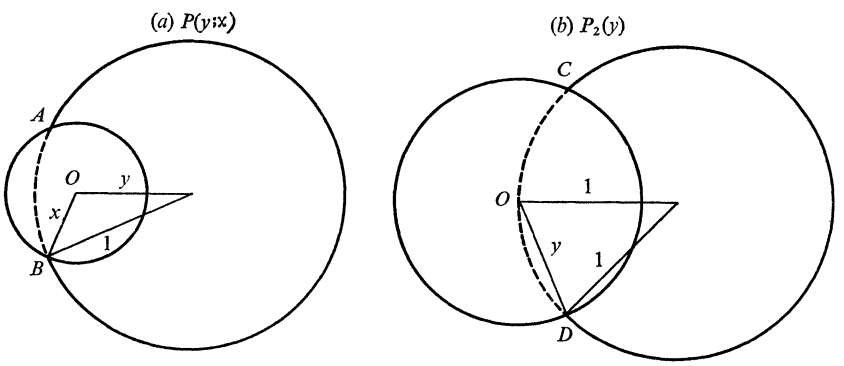

Fig. 2. Evaluation of probabilities for the semi-deterministic clustering process. (*a*) $P(y;x)$, (*b*) $P_2(y)$.

(*a*) If $x \leqslant \frac{1}{2}$, then $P(y;x) = 1$, for all y in the range $(x, 1-x)$, while, for $1-x \leqslant y \leqslant 1+x$,

$$\begin{aligned} P(y;x) &= \text{pr (no offspring on minor arc } AB \text{ of circle of unit radius)} \\ &= \exp\{-\mu l(y;x)/(2\pi)\}, \end{aligned}$$

where

$$l(y;x) = 2\cos^{-1}\{(y^2+1-x^2)/(2y)\}, \tag{3·2}$$

since the number of offspring on an arc of length l is a Poisson variate, with mean $\mu l/(2\pi)$.

(*b*) If $x > \frac{1}{2}$, then $P(y;x) = \exp\{-\mu l(y;x)/(2\pi)\}$ for all $x \leqslant y \leqslant x+1$, where $l(y;x)$ is again defined by (3·2). If we now define

$$q(y;x) = 1-\exp\{-\mu l(y;x)/(2\pi)\}, \quad Q(x) = \int_{\max(x,1-x)}^{1+x} q(y;x)\, y\, dy,$$

the distribution function of X may be written

$$F(x) = 1-\exp[-\pi\rho\{x^2+2Q(x)\}].$$

Turning now to the plant-to-plant distance Y, we may also calculate $P_2(y)$ explicitly by distinguishing between two cases.

(*a*) With probability $\mu/(1+\mu)$, the random sample plant will be an offspring, whose parent will be unit distance away, so that $P_2(y) = 0$, for all $y \geqslant 1$, while for $y < 1$ (see Fig. 2*b*),

$$\begin{aligned} P_2(y) &= \text{pr (no other offspring on minor arc } CD \text{ of circle of unit radius)} \\ &= \exp\{-\mu d(y)/2\pi)\}, \end{aligned}$$

where $d(y) = 2\cos^{-1}(1-\frac{1}{2}y^2)$.

(*b*) With probability $1/(1+\mu)$, the random sample plant will be a parent and

$$P_2(y) = \begin{cases} 1 & (y < 1), \\ e^{-\mu} & (y \geqslant 1). \end{cases}$$

Combining the two cases, we have

$$P_2(y) = \begin{cases} [1+\mu\exp\{-\mu\cos^{-1}(1-\frac{1}{2}y^2)/\pi\}]/(1+\mu) & (y < 1), \\ e^{-\mu}/(1+\mu) & (y \geqslant 1). \end{cases} \tag{3·3}$$

The explicit form for $G(y)$ may be obtained by substitution of (3·3) into (3·1).

The moments of $U = X^2$ and $V = Y^2$ may now be evaluated numerically. The function $Q(x)$ was evaluated by Gauss integration, and the mean and variance of U and V were then obtained from the distribution functions of X and Y, respectively, using a repeated Simpson's rule with a conservative upper truncation point for the range of integration.

Figure 3 gives some indication of the rate at which the bias of each estimator damps down to zero as ρ increases. Again, γ^* is moderately robust in terms of both bias and efficiency, except against the combination of large μ and small ρ. The optimal choice of r for the linear estimator $\hat{\gamma}(r)$ will depend on the value of μ; for the three values considered here, $\gamma(0{\cdot}2)$ represents a good compromise, although a larger value of r would be preferable when μ is small, and a smaller value when μ is large. Detailed results are available from the author on request.

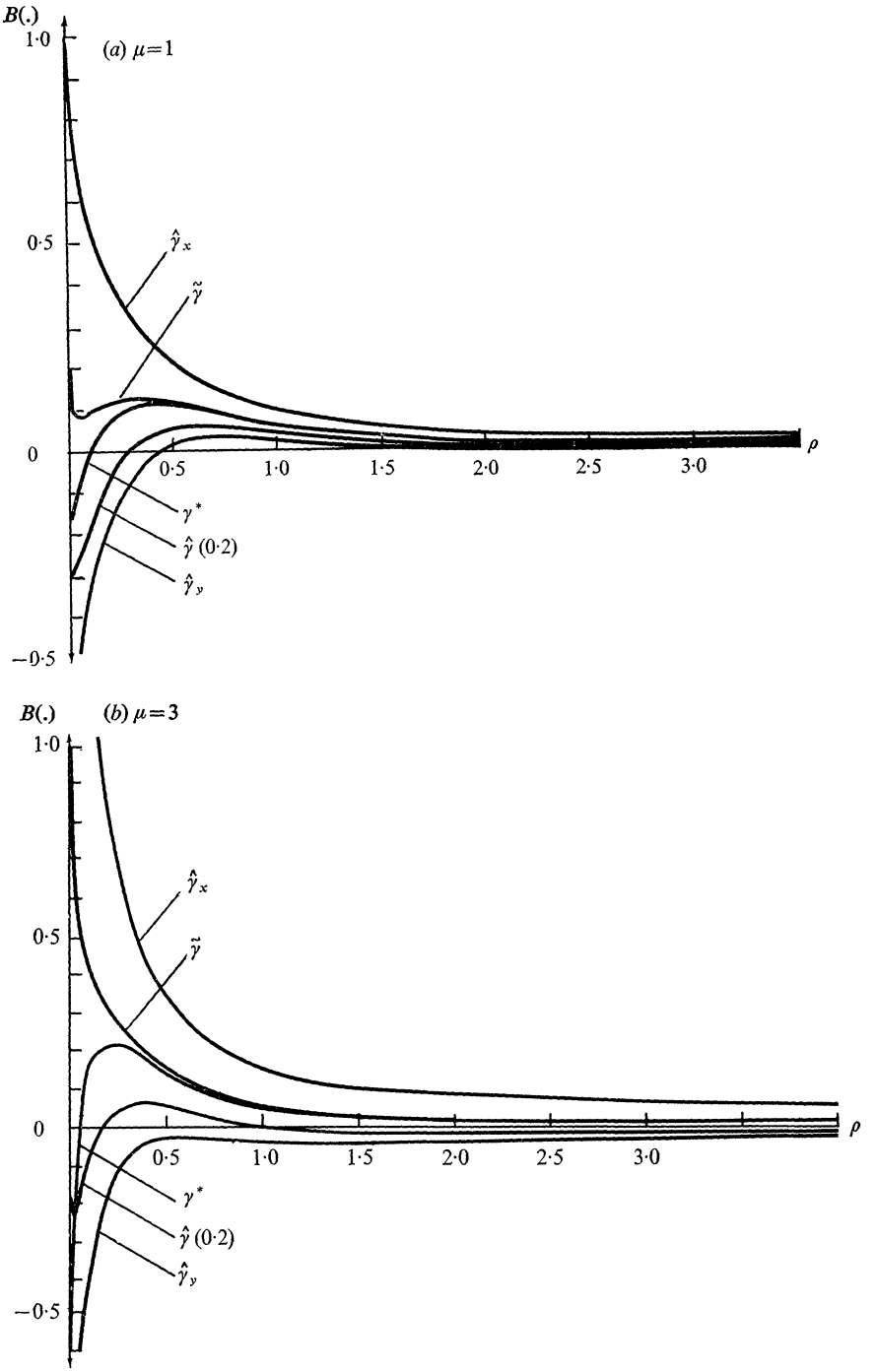

Fig. 3. For legend see facing page.

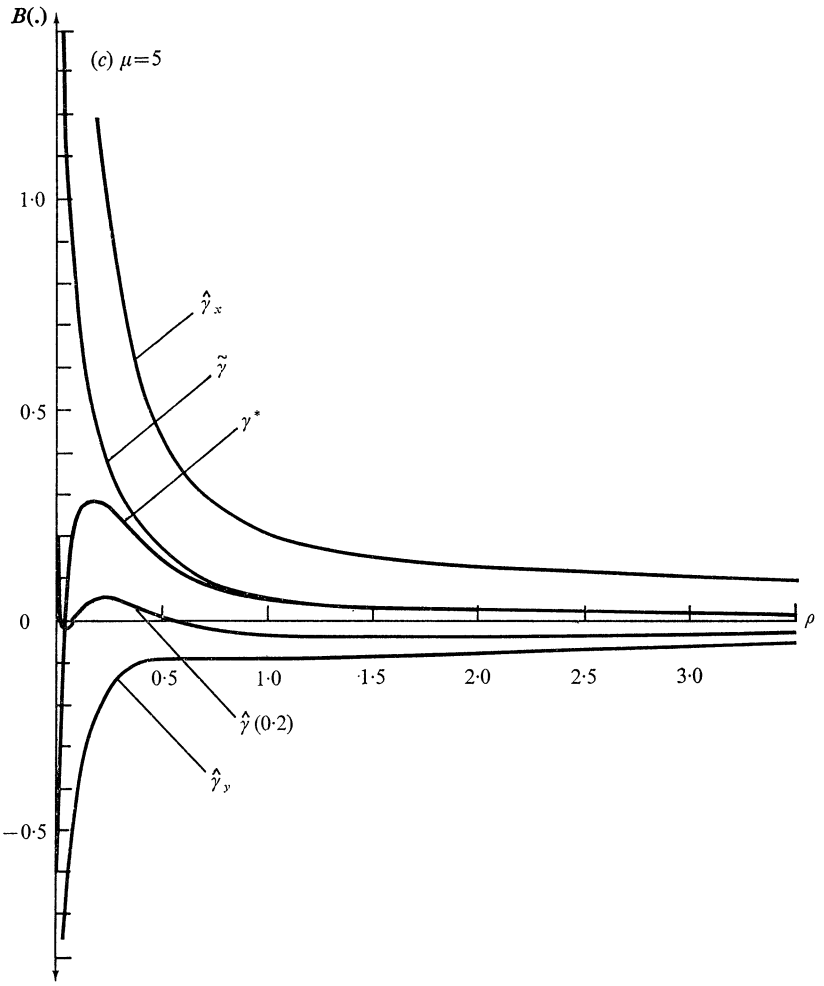

Fig. 3. Mean standardized bias, for density estimators based on Hopkins sampling, applied to the semi-deterministic clustering process. (*a*) $\mu = 1$, (*b*) $\mu = 3$, (*c*) $\mu = 5$.

4. Estimators based on T-square sampling

As previously noted, the truly random selection of a plant from the population is impossible in the present context. However, Besag & Gleaves (1973) introduced, as a practicable alternative to Hopkins's test of randomness, a 'T-square' sampling procedure which may readily be applied to the problem of density estimation. Let P be a random point, and Q the nearest plant. 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. Then the simple estimator

$$\hat{\gamma}_T = \pi \sum_{i=1}^{n} z_i^2/(2n)$$

will be stochastically equivalent to $\hat{\gamma}_y$ in the completely random case, and we may define compound estimators based on T-square sampling,

$$\hat{\gamma}_T(r) = r\hat{\gamma}_x + (1-r)\hat{\gamma}_T, \quad \tilde{\gamma}_T = \tfrac{1}{2}(\hat{\gamma}_x + \hat{\gamma}_T), \quad \gamma_T^* = \sqrt{(\hat{\gamma}_x \hat{\gamma}_T)},$$

analogous to the corresponding estimators $\hat{\gamma}(r)$, $\tilde{\gamma}$ and γ^* defined in § 1. An analytical assessment of the robustness of the T-square estimators is unfortunately not available for the lattice plus Poisson forest superimposed processes, or for the semi-deterministic clustering process, but may be made in the extreme cases of regularity and aggregation provided by, on the one hand, lattice structures, and on the other, the Thomas process of point clusters. In either case, the T-square estimators exhibit greater robustness than do their Hopkins counterparts. For example, the mean standardized bias of γ_T^* for the square lattice structure is $-0{\cdot}093$, compared to a figure for γ^* of $0{\cdot}283$, while for the triangular lattice structure the corresponding figures are $-0{\cdot}004$ and $0{\cdot}352$. For the Thomas process,

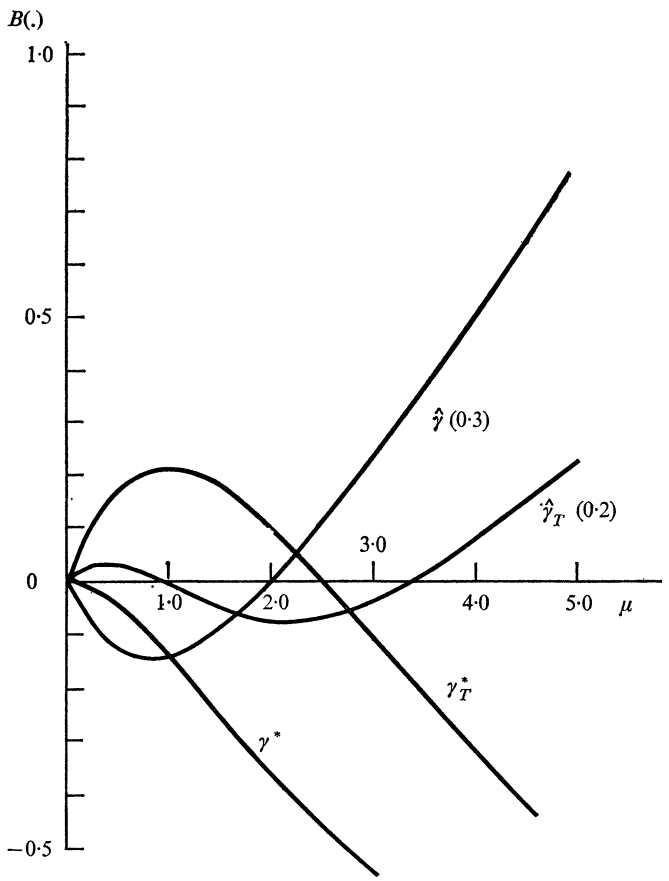

Fig. 4. Mean standardized bias for density estimators applied to the Thomas process.

Fig. 4 gives the mean standardized bias for each of γ^*, γ_T^*, $\hat{\gamma}(0{\cdot}3)$ and $\hat{\gamma}_T(0{\cdot}2)$ as a function of μ; as before the values of r chosen for illustration represent a compromise over the whole range of μ.

5. CONCLUSIONS AND RESERVATIONS

On considering the results of the preceding sections as a whole we see clearly that the compound estimators exhibit greatly improved robustness in comparison to the simple estimators $\hat{\gamma}_x$, $\hat{\gamma}_y$ and $\hat{\gamma}_T$. Let us first restrict our attention to the estimators based on

Hopkins sampling. The estimator γ^* is moderately robust over a wide range of regular and aggregated spatial patterns and, in the terminology of § 1, is the best available compound estimator for γ. We should, however, consider how sensitive the results will be to changes in the stochastic models which generate the spatial patterns. This is particularly important in the case of aggregation where, in addition to changes in the dispersion mechanism of each offspring relative to its parent, we may investigate the effect of changes in the distribution of cluster size.

Turning to the estimators based on T-square sampling, we may recommend the analogous compound estimator γ_T^*, but the evidence to support this is now rather sparse, and the lack of results for any other than the extreme cases of regularity and aggregation is unsatisfactory. Also, we should investigate whether the apparent superiority of the T-square estimators over their Hopkins counterparts in the extreme cases would extend to a more general situation.

The results also point to the possibility of achieving improved robustness by a 'two-stage' estimation procedure, in which the result of a test of randomness is incorporated into the estimator, as mentioned by D. J. Anderson in a comment on Persson (1971), and in lectures by J. K. Ord. For example, the results for the lattice plus Poisson forest superimposed process suggest that the optimum value of r in the estimator $\hat{\gamma}(r)$ tends to decrease with the transition from extreme regularity to randomness, while those for the semi-deterministic clustering process suggest further decrease in the optimum value with increasing aggregation. Thus, a modified version of $\hat{\gamma}(r)$ may be considered in which r is made a function of Hopkins's test statistic, and it would be instructive to discover an appropriate functional form of r. Similarly, in the estimator $\hat{\gamma}_T(r)$, r could be made a function of Besag & Gleaves's T-square statistic.

Some further analytical progress is possible, for example, with the 'thinned plantation' model for regular spatial patterns, discussed by Brown & Holgate (1974), but many of the questions raised above can only be answered by simulation studies and applications to field data, work which is currently in progress and which will be reported in the near future.

I am particularly pleased to acknowledge the help of Professor M. S. Bartlett, under whose supervision I began this work. I am also grateful to Mr J. E. Besag, who has been a frequent source of encouragement and advice, and to Professor R. L. Plackett for his comments on an early draft of the paper.

References

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

Bartlett, M. S. (1974). The statistical analysis of spatial pattern. *Adv. Appl. Prob.* **6**, 336–58.

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

Brown, S. & Holgate, P. (1974). The thinned plantation. *Biometrika* **61**, 253–62.

Dacey, M. F. (1971). Regularity in spatial distributions, a stochastic model of the imperfect central place plane. In *Statistical Ecology*, Vol. 1, Ed. G. P. Patil, E. C. Pielou and W. E. Waters, pp. 287–309. Pennsylvania State University Press.

Holgate, P. (1972). The use of distance methods for the analysis of spatial distributions of points. In *Stochastic Point Processes*, Ed. P. A. W. Lewis, pp. 122–35. New York: Wiley.

Hopkins, B. (1954). A new method of determining the type of distribution of plant individuals. *Ann. Bot.* **18**, 213–26.

NEYMAN, J. & SCOTT, E. L. (1958). Statistical approach to problems of cosmology. *J.R. Statist. Soc.* B **20**, 1–43.

PERSSON, O. (1964). Distance methods. *Studia Forestalia Suecica*, **15**, 68.

PERSSON, O. (1971). The robustness of estimating density by distance measurements. In *Statistical Ecology*, Vol. 2, Ed. G. P. Patil, E. C. Pielou and W. E. Waters, pp. 175–90. Pennsylvania State University Press.

THOMAS, M. (1949). A generalisation of Poisson's binomial limit, for use in ecolog y. *Biometrika* **36**, 18–25.

WARREN, W. G. (1971). The centre-satellite concept as a basis for ecological sampling. In *Statistical Ecology*, Vol. 2, Ed. G. P. Patil, E. C. Pielou and W. E. Waters, pp. 87–118, Pennsylvania State University Press.

[*Received May* 1974. *Revised July* 1974]