EXPERIMENTS IN ESTIMATING DENSITY FROM JOINT POINT- AND NEAREST-NEIGHBOUR DISTANCE SAMPLES

C. L. BATCHELER and D. J. BELL

Forest and Range Experiment Station, Rangiora

SUMMARY: A new model has been developed for estimating density of plant or animal populations. For uniform, randomly-distributed or contagious populations which are distributed with one or two orders of aggregation, it employs the distance from each sample point to the nearest population member, from that individual to the nearest neighbour and from that neighbour to its nearest neighbour. The principle of the model is that an estimate of density is first obtained from the sample point-nearest population member distances. This point-distance estimate is then corrected for bias arising from non-randomness by using the sums and frequencies of first and second neighbour distances. Procedures are given for extending the model to populations distributed with three or more orders of contagion.

The model was tested on five populations of dots, of which four were contagious, two pine plantations of uniform distribution and a population of the contagiously-distributed Composite, *Celmisia spectabilis*. Estimates were all within 25% of the true values. The model is considered to be suitable for estimating a wide variety of populations.

INTRODUCTION

In another paper (Batcheler, in press), J. H. Darwin (in appendix) has shown that

$$\lambda = a/\pi(r_1^2 + \ldots + r_a^2 + (N-a)R^2) \qquad (1)$$

is a useful equation for estimating the density of randomly-distributed populations from a sample of point-distances when it is not feasible to find population members which lie at distances exceeding R. Here, λ is density; N is the total number of sample points; R is the maximum distance over which the search is made at any point; a is the number of N points at which a member is found at a distance equal to or less than R; and $r_1 \ldots r_a$ are the distances recorded at the a points.

The effect of varying R in (1) was studied by using data from synthetic populations of dots, pine plantations and simulated computor populations. As was shown in detail in the other paper, the effect on density caused by changing R is characteristic of the type of distribution of the population. It gives biased estimates which are, (i) progressively larger for uniform populations, and (ii), progressively smaller for contagious populations. These properties are similar to the biases of the well-known equation, $N/\pi\Sigma R^2$ when it is applied to uniform or contagious populations; and this arises because, of course, equation (1) assumes this form when R is large enough to include distance measurements from all the sample points.

J. H. Darwin found, from theoretical considerations of equation (1), that as R diminishes and approaches zero, λ approaches true density but, because this allows use of progressively fewer measured distances, variance of the estimate increases enormously. Therefore, since large variance or large bias is an undesirable attribute of λ, it was evident that an estimate of λ from (1), using an intermediate value of r, was likely to be more reliable. Equation (1) was therefore used with R being set large enough to include half the distances, and its result was termed the 50% point-distance estimate (50% PDE).

Biases in the expected directions persisted in the 50% PDE. Therefore, the estimates obviously were still suffering from the intrinsic distribution characteristics of non-randomly distributed populations. In uniform populations, sampling points are always close to a population member and the sum of the squares of distances is therefore small and λ is high. Conversely, in contagious populations, sample points fall with as equal likelihood in sparsely populated areas as in those containing clusters. Therefore, the distances tend to measure the space between clusters, and λ is low.

A correction was found for these residual biases by making use of the sum of the distances from sample points to the nearest population members and dividing by the sum of the distances from those to their nearest neighbours. Formally, we denote this $\Sigma r_p/\Sigma r_n$ (Fig. 1). In a perfectly uniform
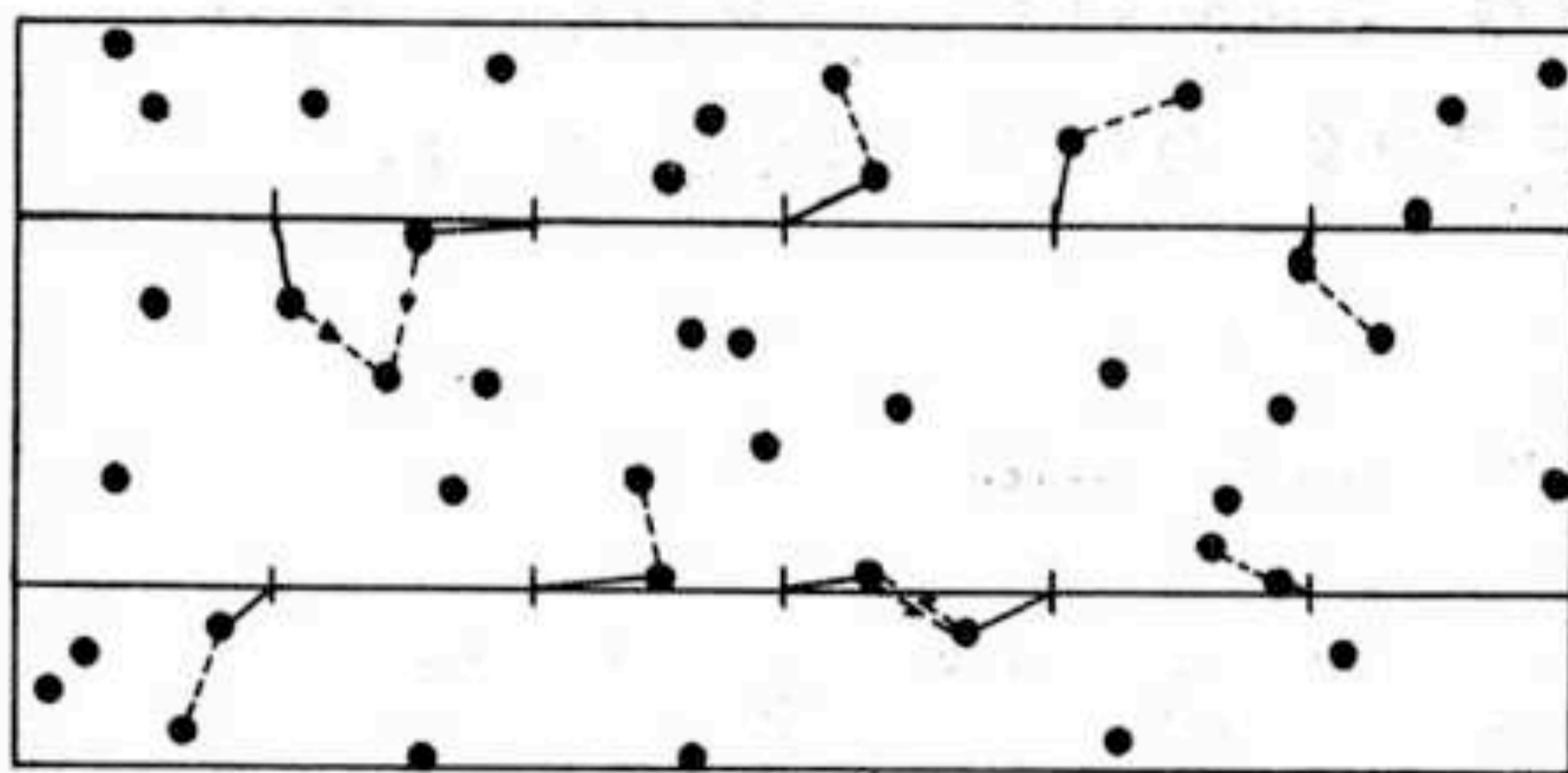
FIGURE 1. *Diagrammatic layout of the procedure for sampling point- and nearest-neighbour distances. The same population member may be nearest to two or more points or be the nearest neighbour to more than one other member.*

population, where members are located on a square grid, sample points within the grid always give distances less than the distance between members; so $\Sigma r_p/\Sigma r_n$ is less than 1, and the 50% PDE is shifted in the required direction. Conversely, in a highly contagious population (e.g. Fig. 2D), the point-distances tend to measure the average gap between clusters, whereas the distances from the nearest individuals to the nearest neighbours measure the distance within the clusters; so $\Sigma r_p/\Sigma r_n$ is greater than 1. Again, the 50% PDE is moved in the required direction. These results were combined to give the single equation for density,

$$\lambda = a/\pi(r_1^2 + \ldots + r_a^2 + (N-a)R^2)\Sigma r_p/\Sigma r_n$$
$$\ldots\ldots(2)$$

This estimator was tested on data from five synthetic populations of dots and two thinned plantations of *Pinus radiata* at Ashley State Forest.

METHOD

Four of the five dot populations were made by placing 400 dots on 100 sq. in. surfaces, as shown in Figure 2. The method is explained in detail by Batcheler (in press). Population A is random and B, C and D are contagious. Population E was made by I. L. James of Forest and Range Experiment Station, Napier, by placing clusters of 10 dots around 36 random points on a 10,000 sq. unit surface. Two *P. radiata* compartments at Ashley State Forest were chosen for measurements to represent a final crop stand (Compartment 2), and one had been thinned once from its original 6 ft. × 6 ft. spacing (Compartment 31).

Density and basal area of both compartments had been estimated by enumeration within sample plots.

Measurements of point- and nearest-neighbour distances were made from systematic points on sample lines. Distances were recorded from the sample point to the nearest population member (r_p) and from that member to its nearest neighbour (r_n). Distances are measured parallel to the ground, and are corrected by transformation to the mean radius of an ellipse of great dimension r on slope α, $\sqrt{r^2 \cos \alpha}$.

RESULTS

The parameter values of quadrat estimates and estimates of density from equation (2) are given in Table 1. These show that the corrected 50% PDEs are all within 25% of the expected population size.

TABLE 1. *Estimates of density of dot and pine plantations obtained by correcting the 50% PDE with the sums of point- and nearest-neighbour distances.*

Population	Para- meter	Estimate (p=.95)	Equa- tion 2	Differ- ence (%)
Random 1	400		385	−3.75
Contagious 2	400		499	+24.75
Contagious 3	400		352	−12.00
Contagious 4	400		410	+2.50
Contagious 5	360		397	+9.97
Compartment 2 (uniform) 6		116±41	101	−12.93
Compartment 31 (uniform) 7		334±73	276	−17.37

CORRECTION OF (1) USING TRUNCATED DISTANCES

The correction factor for non-randomness, $\Sigma r_p/\Sigma r_n$, requires measurement of the point- and nearest-neighbour distances from all sample points. Therefore, since the 50% PDE uses only half the measurements, whereas the correction factor uses all, the field technique is a hybrid which is not applicable to the problem originally studied — estimation of the density of deer pellet groups, because the nearest pellet group is often so far from a sample point that it is impossible
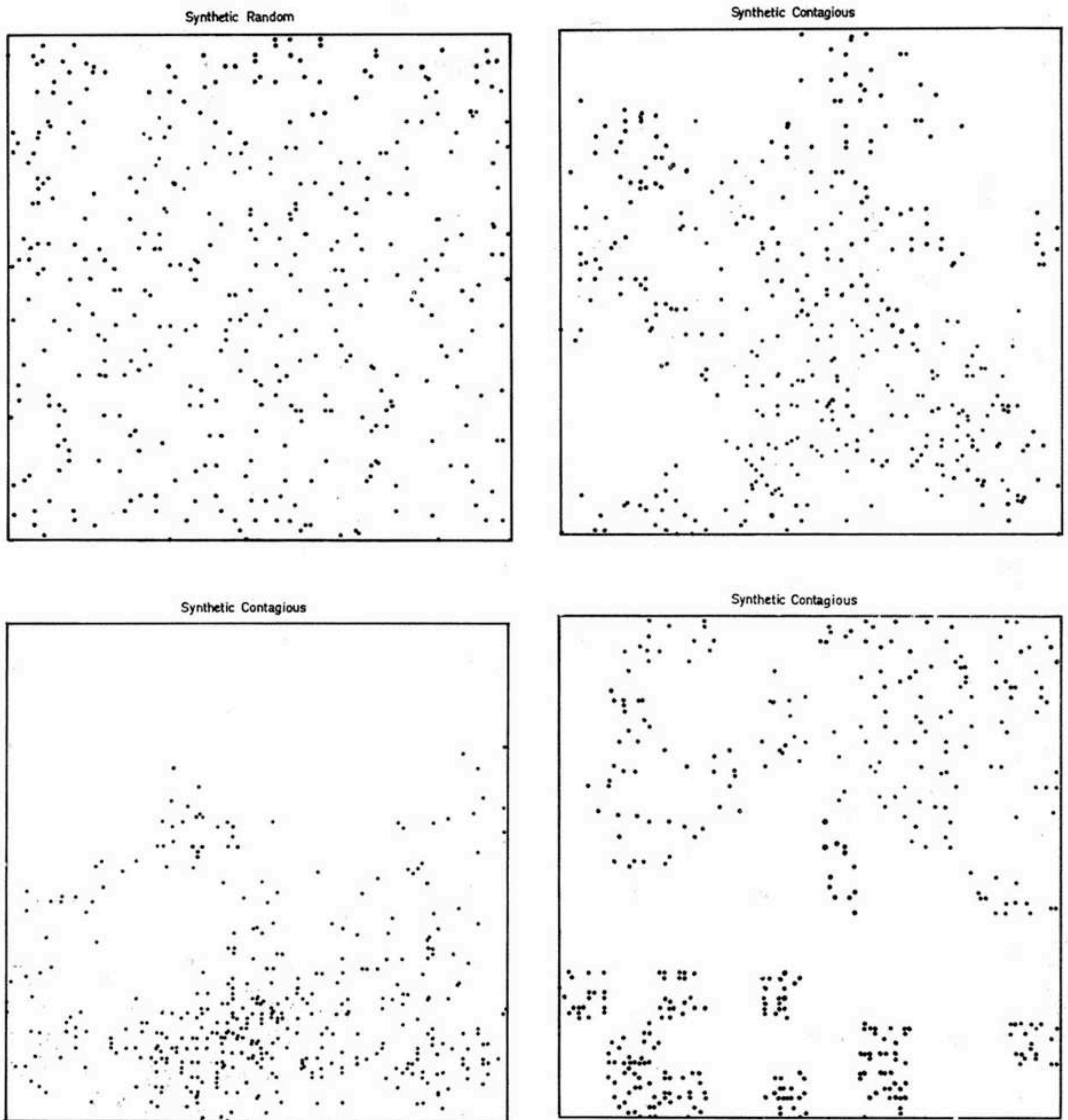
FIGURE 2. *Layout of four of the dot populations studied. They were sampled from systematic points on systematic lines across the surfaces. Each includes 400 members but distribution varies from random (A), through "mildly" contagious (B), to highly contagious with one order of contagion (C and D).*

to find. We therefore require an extension of the model which gives an unbiased estimate from point- and nearest-neighbour distances made within a practicable searching limit, R.

(a) Observed and expected frequencies of nearest neighbours within R

If the search for a member of a randomly-distributed population is limited within R, and one is found at p of N sample points, the probability of occurrence is p/N. If we then suppose that plots of radius R are set together in pairs, each will have probability p/N of being occupied, and the probability that both will be occupied is $(p/N)^2$. So, if an individual is observed within R from p/N points and a search is made around these discovered individuals, their nearest neighbours will be found within R at $(p/N)^2$ points.

This will not hold for non-random populations. In a contagious one, individuals tend to be close together; so when one occurs within R from a sample point, its nearest neighbour will lie within R at a frequency greater than $(p/N)^2$. Conversely, in a uniform population, the nearest neighbour will occur within R less frequently than $(p/N)^2$. Since densities of contagious populations are underestimated and those of uniform populations are overestimated by the 50% PDE, it follows that the observed frequency of nearest neighbours over the expected frequency, $(n/N)/(p/N)^2$, will tend to shift an estimate in the required direction.

This function has, however, two limitations: First, as p/N approaches 1, so also does $(p/N)^2$, the difference between them disappearing. The maximum potential difference occurs when p/N is 0.5. Second, in measurements of contagious populations, neighbour frequency is the same as the point distance frequency when R is just large enough to embrace all the nearest-neighbour distances but not all the point- measurements. This limits the degree of contagion which can be sensed by the difference between observed and expected frequencies and so confines use of the function to small values of R.

(b) Observed and expected means of point- and nearest-neighbour distances within common R

As is well-known, and as will be seen from the expected behaviour of the correction factor. $\Sigma r_p/\Sigma r_n$, in equation (2), the average distance of members of a randomly-distributed population to their nearest neighbours equals the average dis-

tance from random sample points to the nearest individuals. The expectation is thus 1.0, and is $\Sigma r_p n/\Sigma r_n p$, where these symbols carry the notations used earlier. However, when $\bar{r}_p$ is greater than $\bar{r}_n$, there is an element of contagion in the population. Conversely, when $\bar{r}_n$ is greater than $\bar{r}_p$, the population tends to uniformity. These relationships can be expected to hold to some degree whether a limit R is imposed on the measurements or not; but it will be clear that, as R diminishes, the difference in the means must tend to be suppressed by the process of incorporating only measurements within the same limit. Therefore the departure of $\bar{r}_p/\bar{r}_n$ must tend to give a better measure of non-randomness as R increases, ultimately, to the point (with p and n equal) where $\bar{r}_p/\bar{r}_n = \Sigma r_p/\Sigma r_n$. The fraction therefore functions in the opposite way to the departure of frequencies from expectation as a correction for the 50% PDE. That is, as the frequency correction loses sensitivity to non-randomness with increasing R, the sensitivity of $\bar{r}_p/\bar{r}_n$ increases.

(c) The joint correction for contagion of a single order

Taking the two corrections together, we have $(n/N)/(p/N)^2.\Sigma r_p n/\Sigma r_n p$ as a joint correction with an expected value of 1 for a random population. This simplifies to

$$n^2 N \Sigma r_p / p^3 \Sigma r_n \qquad \text{...........(3)}$$

and is the correction factor for the 50% PDE.

RESULTS OF COMPUTATIONS WITH (3)

As before, the data from the experimental populations were given progressively larger values of R, to give correspondingly larger values of p/N. and these values were applied as a correction to the 50% PDEs. The estimates of density were then divided by the parameter value or quadrat estimate, so that the expected value equalled 1.0, and were plotted against p/N (Fig. 3). The figure shows that the estimates were all within 25% of their expected values when the correction was used for high frequency values of p/N.

The estimates for the contagious populations behaved differently from those for uniform populations as p/N changed. For the contagious populations, the estimates were all within 25% of the expected value when R included a half or more of the distances from the sample points. On the other hand, nearly all the measurements of the uniform plantations had to be included before the estimates became satisfactory.
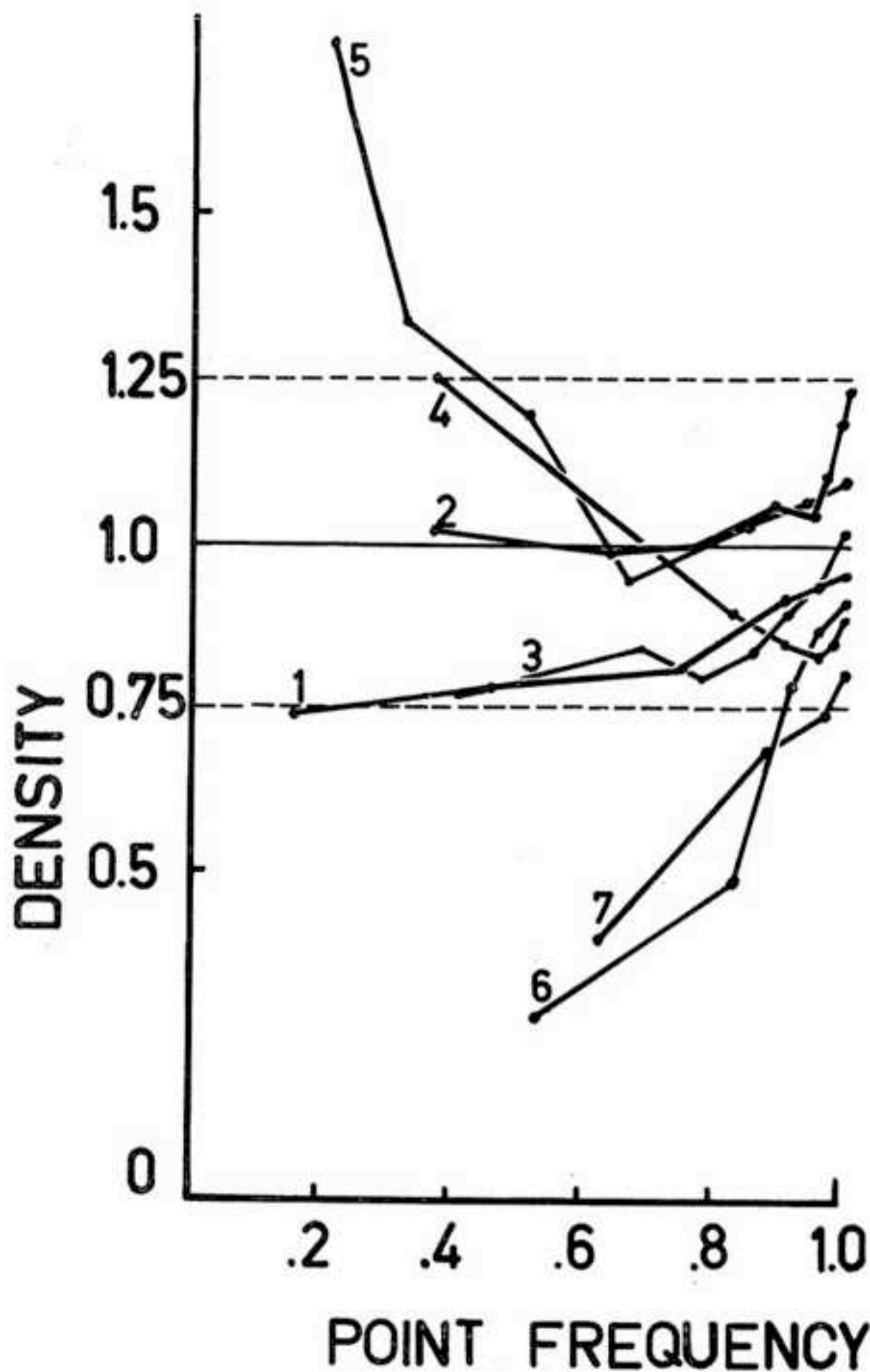
FIGURE 3. *The effect on density obtained by increasing R in the correction factor for the 50% PDE. The random and contagious populations are satisfactorily given when R includes half or more of the point- and their joint nearest-neighbour distances Nos. 1–5. Both distances from all points have to be included to correct the estimates of uniform populations (Nos. 6 and 7). Point frequency in the abscissa refers to the correction factor.*

This problem in the uniform populations has a very simple explanation. As a population tends towards the perfect grid-square, no measurements to nearest neighbours can be made within R until R is equal to the grid distance. The sample points, however, fall randomly within the grid and will always be $0.5\sqrt{2}$ of the grid distance or less.

Therefore, no nearest neighbours can occur within an R which is just large enough to embrace the point distances. For uniform populations, therefore, the correction and estimate will both be biased unless taken without a restriction upon R; and the bias will be increasingly severe as the population tends towards perfectly uniform. Fortunately, this situation can be identified in practice because, within any reasonable R, $\bar{r}_n$ exceeds $\bar{r}_p$ and the observed frequency of nearest neighbours is less than $(p/N)^2$.

CORRECTION FOR TWO ORDERS OF CONTAGION

During preparation of this paper W. A. Laycock showed us distance data from a population of the daisy, *Celmisia spectabilis*. Considering each rosette as a plant, the population is distributed in clusters and the clusters in the study area were concentrated in one half of the area. In the other half, the clusters were fewer and a higher proportion of them contained only one rosette. The population, therefore, exemplifies an extremely complex kind of non-randomness, with at least two orders of contagion.

Using 95 sample points over a sq. chain area of this population, the 50% PDE estimate yielded only 19% of the parameter value which had been found by total count. With equation (2), correcting for first order contagion, the estimate yielded 53%.

The appropriate correction for this population involves recognition of the second order of contagion. Suppose that one measures distance from the sample point to the nearest individual (r_p), from that one to its nearest neighbour (r_n) and from that neighbour to its nearest neighbour (r_m). Within any limit R, the means $\bar{r}_p$, $\bar{r}_n$ and $\bar{r}_m$ will be equal in a randomly-distributed population. Further, there will be $(p/N)^2$ occasions in which p and n distances occur together within R, and $(n/p)^2$ in which p, n and m occur together within R. It was shown earlier that the correction for the first order of contagion is $(n^2N\Sigma r_p)/p^3\Sigma r_n$. Similarly, the measure of the second order is given by $(\bar{r}_n/\bar{r}_m) \cdot ((m/p)/(n/p)^2)$. This simplifies to

$$\Sigma r_n m^2 p/\Sigma r_m n^3 \qquad \text{..........(4)}$$

The joint correction to (1) is the product of (3) and (4), and simplifies to

$$\Sigma r_p m^2 N/\Sigma r_m n p^2 \qquad \text{..........(5)}$$

When R is large enough to include all the distances, then $N=p=n=m$, and (5) becomes

$$\Sigma r_p/\Sigma r_m.$$

RESULTS FOR THE *C. Spectabilis* POPULATION

Distances obtained for the *C. spectabilis* population gave progressively smaller means from p to n to m, as postulated for two levels of contagion: $\bar{r}_p = 26.9$ in., $\bar{r}_n = 8.9$ in., $\bar{r}_m = 6.2$ in. The population estimates for progressively larger values of R and for estimates which incorporate zero and one and two orders of contagion in the 50% PDE appear in Table 2 as "λ/parameter value". The estimate with no limit on R was 80% of the parameter value.

TABLE 2. *Estimates of the* C. spectabilis *population showing the improvements obtained by taking account of up to two orders of contagion and correcting the 50% PDE by Equation (5) at progressively larger point-distance frequencies. The most accurate estimate was obtained when the restriction imposed on the correction included half the data.*

Method	Point-distance frequency (per cent)	Estimate (per sq. ft.)	Estimate/ total count (per cent)
Total count		0.48	100
50% PDE without correction		0.09	19
50% PDE × (3) without restriction		0.25	53
50% PDE × (5) with R including	27	0.72	150
,,	53	0.47	98
,,	72	0.45	94
,,	82	0.40	84
,,	89	0.41	86
,,	95	0.35	73
,,	100	0.38	80

By total count the area was estimated to contain 0.48 plants per sq. ft. Laycock calculated (pers. comm.) that 210 sampling quadrats, each 6 ft. × 6 ft., would be necessary to estimate the density of a large area of this population within ±20% at 95% probability. Since this would involve an inventory of over 3,600 rosettes, it is clear that the corrected point distance method, involving enumeration of only 285 (95 × 3) distances to plants, achieved a very reasonable estimate of the density at an extremely low intensity of sampling.

DISCUSSION

Correction of the 50% PDE by the $\Sigma r_p / \Sigma r_n$, or equation (3), for uniform, randomly-distributed, and single-order contagious populations (noting
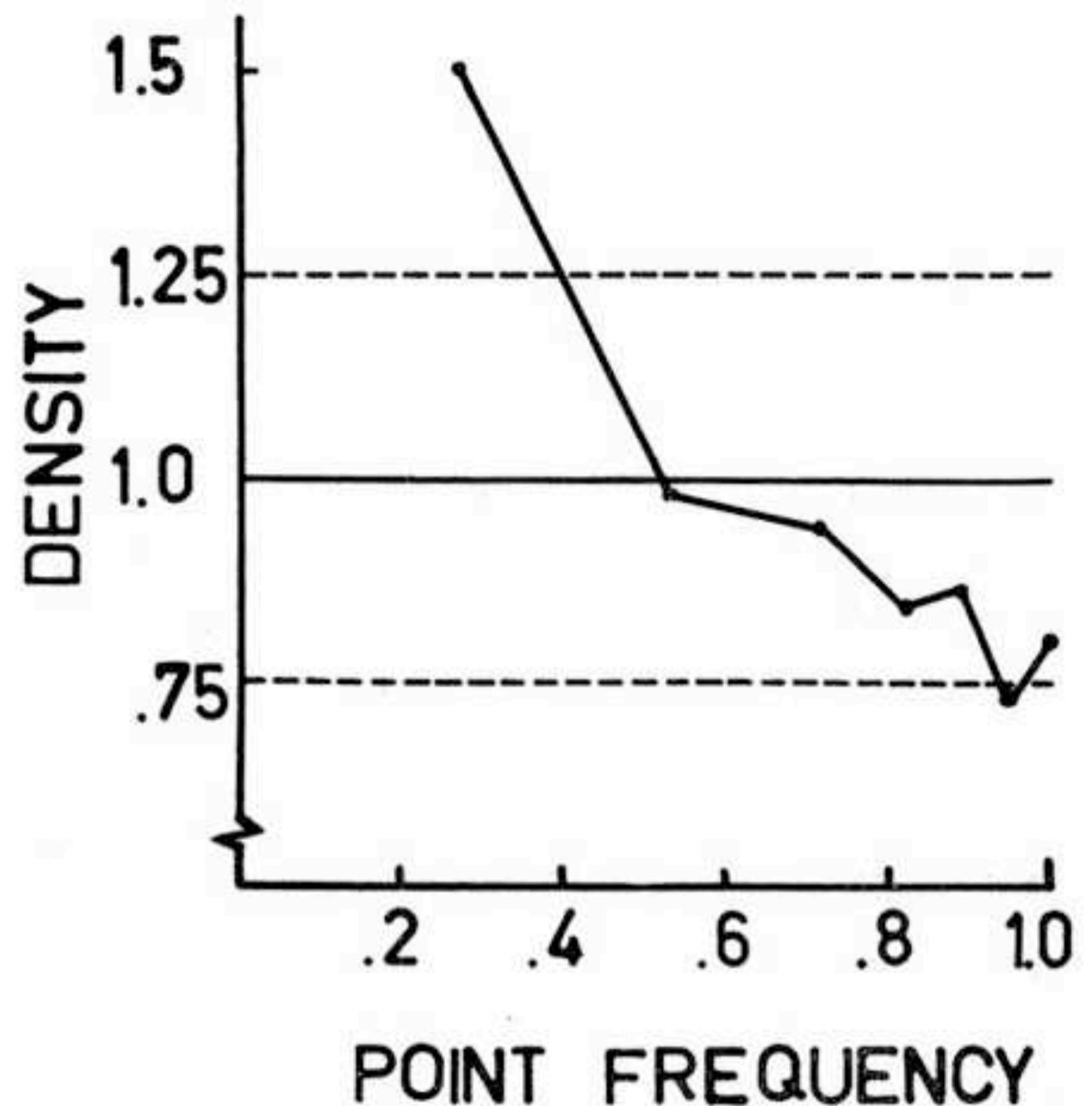

FIGURE 4. *The effect of increasing R in correcting the estimate of rosettes in a* Celmisia spectabilis *population which exhibits at least two orders of contagion. Point frequency in the abscissa refers to the correction factor.*

the special condition concerning truncated distances in uniform populations) and correction by equation (5) for second-order contagion, indicate that the general approach may be extended to more complex aggregations. For the third order (involving measurements to the third neighbour, denoted o) the observed and expected frequencies are o/n and $(m/n)^2$ and the means are $\Sigma r_m / m$ and $\Sigma r_o / o$. The correction is therefore $\Sigma r_m o^2 N / \Sigma r_o m^3$ and this, multiplied by equation (5), gives $\Sigma r_p o^2 N / \Sigma r_o m p^2$. In general, therefore, when no restriction is placed upon R, the 50% PDE is corrected by the sum of the point distances divided by the sum of joint neighbour distances for the most severely non-random component in the population. A restriction placed upon R greatly complicates computation but does not otherwise appear to confound the model.

Only one or two significant levels of contagion are likely to be encountered in most biological populations, so the necessary distances will usually not be more than p, n and m. A third or higher significant level is probably very rare. It may, perhaps, occur in populations of *Celmisia* such as

that described, if rosettes towards the centre of clumps are smaller, and consequently closer together, than rosettes near the fringes. Or, if the problem is to estimate the total number of leaves in the population of *Celmisia*, the third order would arise from the characteristic bunching of the leaves within each plant. Generally, the eyes and a sensitive appreciation of the concept of contagion will indicate to an investigator how many orders are likely to prevail. The significance of each order may be tested by comparing the roots of the distances p, n, m, etc., as normal variates.

Uniformly-distributed populations must be estimated from distances taken without a restriction on R; but random populations or contagious populations with at least one order of contagion are reasonably well estimated with R large enough to include a measurement at only half the points, provided N is over about 70. The same approximation will probably hold good for estimating populations with more than one order; because in these, this restriction will always include a fairly high proportion of first and second neighbour distances.

The estimates given in this paper cannot be expected to be very precise because experiments of the severe kind described contain fairly severe limitations. In common with Morisita (1954), Cottam and Curtis (1956) and Catana (1963), we found it necessary to restrict sampling to small dot populations. In highly contagious situations this resulted in the distance between sample points and the nearest individual often being greater than the distance to the edge of the population. Following Morisita's precedent such samples were ignored, with the inevitable result that sampling was biased towards areas containing more members. This may explain why the estimate for contagious populations tended to be high (Table 1 and Fig. 3). More than one measurement between certain neighbouring individuals, or to one individual from different sample points may cause other problems which have not been evaluated (see Fig. 1). Experiments with field populations are subject to the risks of missing individuals during sampling of distances, or of working against an inaccurate figure for the total number of individuals present. Differences of ±25% between the distance estimates and 'parameter' values are therefore considered to be reasonable and likely to arise as much from chance as from bias in the

estimator. More exhaustive evaluation of the model and its development into point-centred-quarter and other distance techniques will probably depend upon simulation of populations and sampling in a computor.

We consider that the corrected point-distance method is simpler to use and the result predictably more accurate than can be obtained for equal amounts of field and computing work involving quadrat methods of which we have experience. Quadrat boundaries do not have to be laid out, more sites can be sampled than is feasible with quadrats and, regardless of density, the number of population members which occur in any sample is directly proportional to the number of sample points. Even when a limit has to be imposed on R, because it is not feasible to measure all the joint distances, arbitrary decisions are surprisingly robust. This is because of the root skew distribution of distances. In surveys of the density of deer pellet groups, this Institute has used 8 ft. as the limit in areas where high densities are suspected (C. N. Challies, pers. comm.) and 10–12 ft. where lower values are suspected (Batcheler, unpubl.).

ADDENDUM
Further analyses have shown that
$$\lambda/D = 1.386 \times 1.450^{-(\Sigma r_p / \Sigma r_n)}$$
is a significantly better estimator of density (D) than is given in equation (2). λ is the 50% PDE. The effect of this result on corrections calculated from measurements within common R has not yet been determined.

REFERENCES
BATCHELER, C. L. (in press). A new model for estimating density from a joint sample of point- and nearest-neighbour distances. With Appendix on derivation, by J. H. Darwin.

CATANA, A. J., JR. 1963. The wandering quarter method of estimating population density. *Ecology* 44: 349–360.

COTTAM, J. T., and CURTIS, J. T. 1956. The use of distances in phytosociological sampling. *Ecology* 37: 451–460.

MORISITA, M. 1954. Estimation of population density by spacing method. *Mem. Fac. Sci. Kyushu Univ. Ser. E* 1: 187–197.