ESTIMATING DENSITY AND DISPERSION FROM TRUNCATED OR UNRESTRICTED JOINT POINT-DISTANCE NEAREST-NEIGHBOUR DISTANCES

C. L. BATCHELER

Forest and Range Experiment Station, Rangiora

SUMMARY: Forty computer-simulated populations were analysed to derive formulae for estimating density of populations from a set of distances from sample points to the nearest member, from that member to its neighbour and from that neighbour to its nearest neighbour. Distances may be truncated or unlimited.

The formulae were applied to data from 11 paper-dot and 18 field populations in a total of 37 experiments. Fifty-nine percent of the corrected estimates were within 10 percent of "true density", 76 percent were within 20 percent, and 93 percent were within 30 percent. Most of the unacceptable estimates were attributable to sampling difficulties (particularly in paper dot populations) or sampling errors of the "true density" values.

An index of non-randomness is an intrinsic part of the density estimating formula. This index is described, and values for the experimental data are given to illustrate the spectrum which can be expected in biological populations.

INTRODUCTION

An estimate of a population's density which is based on distances from sample points to the nearest member is biased if the population is not randomly distributed (Kendal and Moran 1963, Pielou 1969). If population members tend to be uniformly spaced, the estimate will be high, whereas if they are distributed in clumps the estimate will be low (Cottam and Curtis 1956 Pielou 1959, Lyon 1968).

Recent work (Batcheler and Bell 1970, Batcheler 1971, James 1971) has shown that these biases are inversely related to ratios which can be derived from measurements from the sample point to the nearest member, and from that member to its nearest neighbours. This function is easily visualised by comparing a perfectly uniform population with one which is extremely aggregated (clumped). In the former case, which is analogous to the position occupied by a person in a square room, the distance from any position in the room to any corner must be less than the distance from that corner to another. Consequently, the average of any set of distances from random positions in the room, divided by the distance between neighbouring corners, will produce a quotient which is less than 1. The small quotient can be used to "knock down", and reduce, the high biased estimate derived from position-to-corner distances. Conversely, if it is supposed that compact clumps of population members occur at each corner of the room (i.e. the population is aggregated), the average of a set of distances from random positions to the nearest member of any clump (r_p) will be larger than the average distance from one member of a clump to its nearest neighbour (r_n). This quotient of r_p/r_n, greater than 1, can be used to raise the low biased estimate which is characteristic of aggregated populations.

The idea of using these ratios to correct bias was developed earlier (Batcheler and Bell 1970, Batcheler 1971) to give one of many possible equations for estimating density. The equation was called the 50 percent corrected point distance estimator:

$$d/D = 1.386 \times 1.450^{-b}$$

Where d is the biased estimate calculated from

$$p/\pi \left(\sum_{o}^{R} r_p^{\,2} + (N-p)R^2\right),$$

R is a distance large enough to include a population member at 50 percent of the sample points, N is the total number of sample points, p is the number of distances measured within R, r_p is a distance less than R, and b is the ratio of the sum of all r_p's divided by the sum of all r_n's).

The equation was tested on only 13 populations at the time the earlier papers were published. Despite this rather superficial treatment, some features, particularly the use of a truncated distance formula for d, and the use of joint point and nearest-neighbour distances as an index of dispersion (rather than independent point and neighbour-distances as employed by Hopkins (1954) opened up some useful possibilities in the quest to derive a simple yet reliable method since Cottam (1947) drew attention to the use of distance techniques for woodland assessment.

No new theoretical papers have come to notice in the intervening three years to push the issue any further. Indeed, it seems clear enough from the forbidding assumptions which attend formal mathematical treatment of distance functions (e.g. Clark and Evans 1954, Morisita 1957, Perrson 1962, Holgate 1965, Thompson 1956, Pielou 1969, Eberhardt 1967), that formal methods applicable to populations of which the dispersion parameters cannot be easily defined, will be difficult if at all possible.

Meanwhile, James (1971) has further developed several aspects of distance sampling by study of 40 computer-simulated populations. He chose 26 measures of dispersion based on distances from point-to-member and member-to-neighbours, and tested their capacity to correct the dispersion-dependent bias of nine estimates of density based on first and second moments of the nearest member, second nearest member, and joint nearest neighbour. Using polynomial regression methods he found three equations for correcting bias for which the co-efficient of multiple correlations, R, exceeded -0.98. In these, the ratios used were:

$$\bar{r}_p^2/\bar{r}_n^2, \quad \Sigma\, r_p^2/\Sigma r_n^2 \quad \text{and} \quad \Sigma\, r_p^2/\Sigma\, r_{p2}^2$$

(r_{p2} is the second nearest member to the sample point).

That is to say, of 26 x 9 combinations of density estimators and corrections which were tested, the best two were based on transformations of the joint point and nearest-neighbour distance technique published earlier.

These results, and the daunting quantity of statistical and computer work they represent, are in three senses the keystone of this paper: they provide very much stronger evidence of the utility of the joint point-distance — nearest-neighbour distance function than was formerly available; they provide experimental data which enabled me to identify a serious lack of sensitivity of the earlier model to certain patterns of aggregation; lastly, they facilitated development of a density estimating formula for the field situation where, for practical reasons, a limit has to be imposed on the distance searched from the sample point to the nearest member and its sequential neighbours.

SAMPLING AND SAMPLE STATISTICS USED

Before going further, it is useful to outline the strategy of the sampling technique and to summarise the statistics to be used in remaining sections of the paper.

A number of sample points, N, are located at systematic intervals along random lines, or completely at random, and the observer measures the following distances within R, a chosen maximum distance:-

(a) r_p, if the nearest member is distance R or less from the point, r_p is measured, and p is the number of such measurements made;

(b) r_n, if the member nearest to the sample point is found at R or less, the distance from that member to its nearest neighbour is also measured provided it is R or less, and n is the number of such measurements;

(c) r_m, if the nearest neighbour is closer than R to the point member, the distance to its nearest neighbour (exclusive of the member nearest to the sampling point) is measured up to the limit R, and m is the number of such measurements.

These quantities are used to calculate the following statistics, most of which are described in detail in Batcheler and Bell (1970) and Batcheler (1971):

f=p/N,
point distance frequency within R;

$$d=p/\pi\ (\sum_{o}^{R} r_p^2 + (N-p)R^2),$$

the maximum likelihood estimate of density (in Batcheler, 1971), which tends to
$N/\pi\ \Sigma\ r_p^2$
as R increases;

$$b_1 = \sum_{o}^{R} r_p\ n^2N/\sum_{o}^{R} r_n p^3$$

an empirical measure of dispersion which is sensitive to simple clustering in a population, and which tends to
$\Sigma\ r_p/\Sigma r_n$
as R increases;

$$b_2=\sum_{o}^{R} r_p m^2N/\sum_{o}^{R} r_m p^2 n,$$

an empirical measure of dispersion sensitive to clustering within clusters, which tends towards
$\Sigma r_p/\Sigma r_m$
as R increases; and

$$CV = \sqrt{p(\Sigma r_p^2 - (\Sigma r_p)^2/p)/\Sigma r_p},$$

the coefficient of variation of the point distance (syn. Shortest distance, Morisita, 1954).

CHARACTERISTICS OF JAMES' POPULATION DATA

As mentioned earlier, 40 populations were analysed to derive empirical formulae for estimating density. They were of equal density and each was comprised of 1,000 members at plotted co-ordinates within a square area of 10,000 x 10,000 units. One of the 40 was a natural forest tree population in the Manawatu District, North Island. The remaining 39 were simulated in a computer by generating sets of 1,000 co-ordinates by programmes designed to induce uniformity, randomness or aggregation in the distribution of members. Of these, 12 were designed to span the range from uniform to random whereas the remaining 27 tended from approximately random to strongly aggregated. Two hundred samples of r_p, r_n and r_m were measured from random co-ordinates in each of these populations and at each of the 200 sample points the members were counted within a circular plot which was made large enough to include an average of four per plot. The sample plots were established at the sample points primarily to check the density of population members in the vicinity of each sample point.

Besides the check on density, the sample plot counts provide a statistic of dispersion which serve to rank the degree of uniformity or aggregation in the populations. Since in this instance the mean was expected to be four per plot, the variance under Poisson assumptions is four, and 100 $s/\bar{x}$, the coefficient of variation, is 50. A lower value indicates uniformity, whereas a higher value indicates aggregation. Sample coefficients of the 40 populations are listed in column 5, Table 1.

The programme for controlling the degree of uniformity allocated a proportion of members to the grid intercepts of either a square or triangular lattice, and the remainder to random co-ordinates. In the 12 such experiments, the degree of uniformity was graduated by lowering the proportion of members at lattice positions from 961 (of the 1,000) to 289, and correspondingly raising the random proportion. The evident correlation between the proportion of random members in these populations and the coefficient of variation calculated from the 200 plot samples (columns 3 and 5, Table 1) shows that a comprehensive gradient of distributions was generated by this technique.

Random-to-aggregated populations, with specified degrees of clumping, proved relatively unpredictable to compile. Uniform, random, and aggregated clump centres were chosen, and with square sub-units of three different sizes about these points 10, 50 or 250 members were placed in uniform, random or aggregated groups. This technique was used to obtain the 27 combinations of between-clump distribution (Uni, Ran, Agg), within-clump distribution (Uni, Ran, Agg) and clump size (10, 50, 250) (Table 1). Generally, a large number of clumps endowed a population with only a mild degree of aggregation. Higher density within clumps, and few clumps with many individuals, created a higher degree of aggregation. The most severely aggregated populations were made up of only four large clumps, some of which were subsequently found by mapping to overlap.

In early runs with the programme it was found that many sampling points, particularly in aggregated populations, fell closer to the edge of the population than to the nearest member. These sample points were rejected, and new ones generated.

TABLE 1. *Characteristics of Computer Populations.*

1*A Uniform populations*

Lattice	No. of uniform members	No. of random members	Plot count E = 800	CV% (plot count)
Square	961	39	797	12
Triangular	900	100	813	27
Square	900	100	824	27
Triangular	784	216	819	29
Triangular	961	39	824	30
Square	754	246	780	30
Triangular	529	471	828	35
Triangular	625	375	820	36
Square	625	375	784	40
Square	529	471	824	41
Triangular	289	711	810	43
Square	289	711	786	44

1*B Clumped populations*

Clump distribution	No. of clumps	Distribution within clumps	No. within clumps	Clump density per unit area	Plot count E = 800	CV% (plot count)
Uni	20	Ran	50	.08	807	48
Uni	20	Agg	50	.08	923	53
Natural forest population					933	53
Uni	20	Uni	50	.08	924	58
Ran	20	Ran	50	2	859	59
Uni	100	Agg	10	40	839	62
Agg	20	Agg	50	.08	1027	66
Ran	20	Agg	50	.08	953	67
Ran	20	Uni	50	.08	869	79
Agg	20	Uni	50	.08	1022	80
Uni	100	Ran	10	40	789	80
Uni	4	Ran	250	.4	1094	88
Agg	20	Ran	50	.08	1055	88
Ran	20	Ran	50	.08	825	95
Agg	4	Agg	250	.4	845	101
Ran	4	Agg	250	.4	944	105
Agg	100	Agg	10	40	1009	105
Ran	100	Agg	10	40	877	109
Ran	100	Uni	10	40	837	111
Agg	100	Uni	10	40	1190	122
Agg	100	Ran	10	40	1180	124
Ran	100	Uni	10	40	859	127
Agg	4	Uni	250	.4	1073	132
Agg	4	Ran	250	.4	1155	135
Uni	4	Uni	250	.4	1014	136
Ran	4	Uni	250	.4	1003	146
Ran	4	Ran	250	.4	883	171
Uni	4	Agg	250	.4	706	178

This had the insidious effect of tending to locate "valid" sample points within or near clumps, and cause sampling to be concentrated in regions of the plane which contained higher than average density. The severity of this effect was appreciated when the plot counts were found to range from about 800 (the figure expected from 200 samples with a mean count of four per plot) to 1.200, and when this trend was found to be correlated with the degree of aggregation as measured by coeffi-

cient of variation (plot count = 837 + 16.79 (CV_p); r = 0.516; P 0.01, where CV_p = coefficient of variation of plot samples). In other words, relative to actual density in the vicinity of the sample points, the population parameter was wrong by up to 50 percent because of edge effects. Consequently, estimates of density in the vicinity of the sample points, calculated from (1,000 x count)/800, are used as an appropriate estimate of true density of each population.

TESTS OF THE 50 PERCENT CORRECTED POINT DISTANCE TECHNIQUE

The estimating formula—

$$d/D = 1.386 \times 1.450^{-b_{max}},$$

where b_{max} is the maximum of b_1 and b_2 (Batcheler 1971) was used to estimate density of the 40 computer populations and yielded 33 estimates within ± 15 per cent of the parameter value (Fig. 1). Although these are superficially reasonable (they are nearly all within the range of error to be expected on shortest distance theory if the populations had actually all been random (Kendal and Moran 1963)), two obvious problems are displayed by the figure. Firstly, there is a clear residual tendency for the corrected estimate to regress across the expected value (1.0) as a function of $N/\pi \Sigma r^2$. This suggests that, even in relatively uniform populations, the corrected 50 percent PDE did not fully cope with dispersion-dependent bias. Secondly, for seven of the populations, in which aggregation was so severe that $N/\pi \Sigma r^2$ gave estimates of only three to five percent of true density, the "corrected" estimates are still 50-75 percent biased. A new regression formula calculated from the computer data itself gives—

$$(50\%\ PDE)/D = 1.413 \times 1.627^{-b_{max}}$$

(r = 0.9459). This virtually eliminated perceptible drift in the estimates (compared with that in Fig. 1), but made negligible improvement to the estimates for the seven severely aggregated populations.

These problems, and the over-riding objective of finding a formula applicable to truncated distances necessitated further work. It was decided to investigate initially the causes of aberration in distances not subject to any constraint and then to test whether any solution could be applied to truncated data.

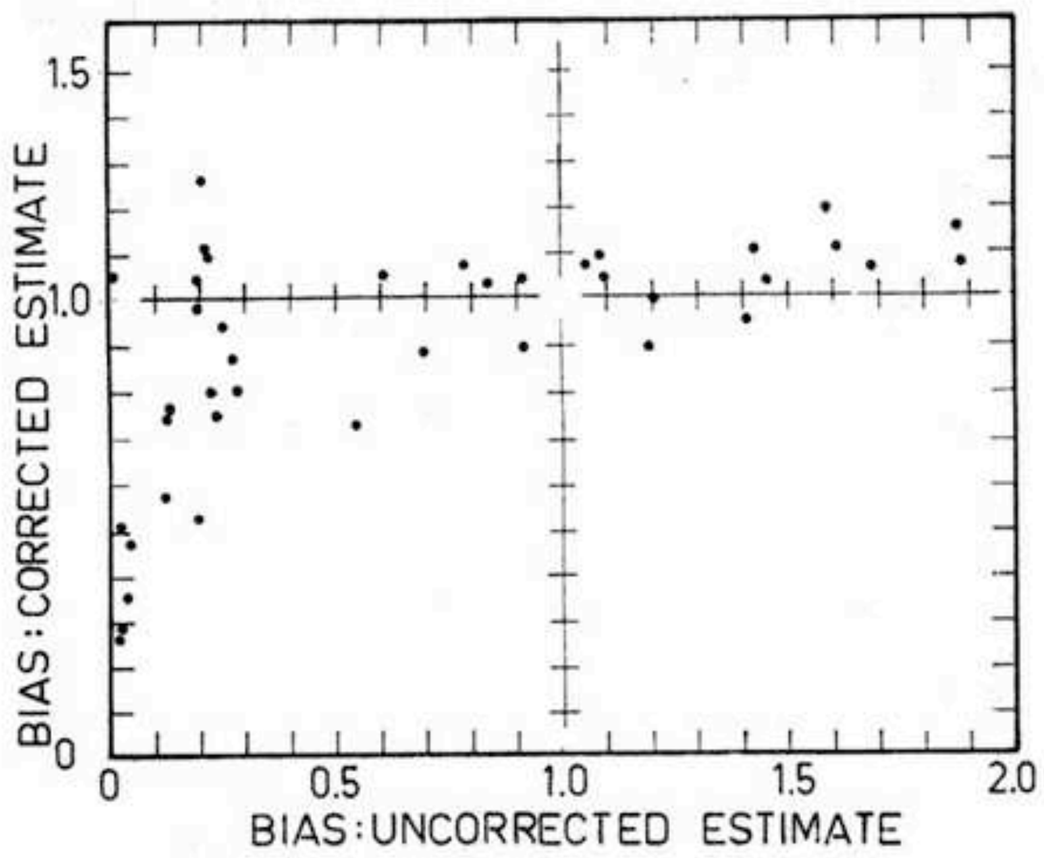

FIGURE 1. *Values of density calculated by the shortest distance method on the x-axis* ($d = N/\pi \Sigma r_p^2$) *plotted against estimates from the corrected 50 percent point-distance method, y-axis. Uncorrected estimates, ranging from 2 percent to 188 percent of the expected value, are a measure of non-randomness in the populations. 19 of the 40 "corrected" estimates are within ± 10 percent, 28 are within ± 20 percent, 33 are within ± 30 percent.*

ESTIMATING WITHOUT CONSTRAINT ON THE DISTANCE SEARCHED

Here, the maximum likelihood estimate of density is—

$$N/\pi\ \Sigma r^2,$$

while b_1 ($\Sigma r_p/\Sigma r_n$), b_2 ($\Sigma r_p/\Sigma r_m$) and b_{max} (the largest value of b_1 and b_2) are under evaluation as point-neighbour ratio corrections.

A semi-log plot of the bias of $N/\pi\Sigma r^2$ against b_{max} (Fig. 2) shows a linear trend over the uniform-to-random range. With increasing aggregation, however, the fit of points drops away in a sickle shaped arc towards the bottom right of the figure. Clearly, within this range, b_{max} is of

little use as a correction to the bias of the shortest distance estimate.

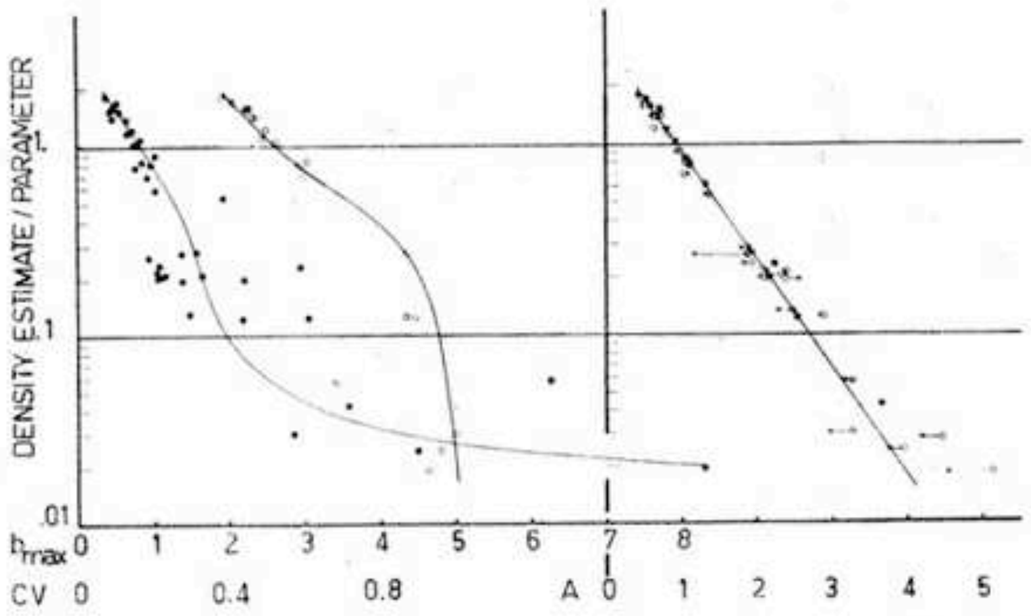

FIGURE 2. *Three scattergrams and a regression of log bias of* $N/\pi \Sigma r_p^2$ on b_{max} *(i), co-efficient of variation (ii), and the composite correction based on* $\frac{CV}{E(CV)} \times \sqrt{b_1}$ *(crosses) or* $\sqrt{b_2}$ *(circles), giving* A_1 *and* A_2 *respectively (iii). The opposing sickle-shaped arcs of* b_{max} *and* CV *against log bias (free-hand lines) are evident in (i) and (ii). The line in (iii) is calculated as mean regression of log bias against* A_{max}, *for 28 selected population samples (see text).*

The second scattergram in Figure 2 shows the semi-log plot of bias against the coefficient of variation of r_p. It is easily seen that this scattergram also displays a sickle-shaped arc of points, and that they have the interesting property of mirror-imaging the arc of the b_{max} correction. Indeed for each of the 40 sample statistics, a "left-hand" cast of b_{max} is faithfully mirrored by an extreme "right-hand" cast of CV, and *vice-versa*, or alternatively, both are neatly centred on their respective correlation lines. Therefore, it seemed certain that these two indices of dispersion could be combined into a single powerful index, correlated with bias of the shortest distance estimate.

The scattergram of bias against CV passed through 1 on the Y axis at about 0.5 (Fig. 2). When Dr J. H. Darwin showed me (pers. comm.) that this is nearly the expected value of CV for a random population ($\sqrt{(4-\pi)/\pi} = 0.5227$), it was realised that both CV/E(CV) and b were approximately 1 for the random case, that their product would be approximately 1, that smaller values would be indicative of uniformity, and that larger values indicative of aggregation. It was subsequently found by iterative and least squares tests that $CV/E(CV) \times \sqrt{b_{max}}$ formed a very good straight-line scattergram against log of bias of the point distance estimate. Defining these as A_1, A_2 and A_{max} when b_1, b_2 and b_{max} point and neighbour ratios respectively are used, the dependent regression of log bias on A_{max} for all 40 populations samples is

$$d/D = 2.736 \times 3.04_1{}^{-A_{max}} \quad (r = -0.983)$$

as shown in scattergram (iii), Figure 2.

However, this statement does not discriminate between A_1, A_2, and A_{max} as the best choice for correction of the bias of the point distance estimate. There is virtually no difference between them in 28 of the samples, and, for these, A_1 would be chosen on grounds of relative ease of field sampling. But in five of the 40 samples log bias against A_1 lies closer to the general correlation set, while in another three A_2 lies closer. At this preliminary stage of screening the behaviour of A against bias it therefore seemed reasonable to reject samples where A_1 differs markedly (say 10%) from A_2.

Similar consideration is required of the need to reject some samples in which the density estimates are extremely biased. James (1971) has pointed out that random sampling error in any corrected estimate is likely to be positively related to variance. Therefore, since point distance estimates as low as only two to five percent of true density were obtained from some aggregated populations, the problem in evaluating A_1 and A_2 is one of evaluating the fidelity of a correction term which is required to raise the point distance estimate by up to 50 times the yield a reasonable value. Quite small errors in sampling can conceivably be compounded into an estimate which is grossly wrong. This argument favours rejection of very biased samples from the analysis.

Accordingly, twelve populations in which A_1 differed from A_2 by more than 10 percent, or in which $N/\pi\Sigma r_p^2$ is biased by more than 80 percent, were set aside from the list, and the mean regres-

sion of log bias against A_{max} was calculated for the remaining 28. Mean regression, where

$$\bar{b} = b_{yx} + 1/b_{xy},$$
$$\text{and } \bar{a} = \bar{y} - \bar{b}.\bar{x}$$

(Simpson, Roe and Lewontin 1960) is taken as a more appropriate regression than the dependent form for log bias on A_{max}, because the variables are mutually dependent. These give—

$$d/D = 3.473 \times 3.717^{-A_{max}} \ (r = -0.997).$$

THE GENERAL FORMULA WHEN A LIMIT DISTANCE (R) IS IMPOSED ON THE SEARCH FOR THE NEAREST MEMBER AND ITS NEIGHBOURS

When used for either truncated or unconstrained measurements it is obviously desirable that the estimating formula be basically of the same form, that it should give a reasonable estimate of density when even a small proportion of successful searches are made within any given distance R, that it be subject to few if any "special case" rules, and it should be reasonably easy to calculate.

Besides the semi-log approach introduced in the previous section, alternative methods based on polynomial regressions (James 1971) were explored, but ran into difficulties. When no limit was imposed on R, an extremely good regression estimator was obtained from the 40 population samples by his method. Taking $\sqrt{A_{max}}$ as A′, it is—

$$d/D = 2.4847 - 3.3701(A'/(1 + A')) + 1.2052 (A'/(1 + A'))^2,$$

with R = -0.9948 (Fig. 3). However, when progressively smaller R limits were imposed, third degree polynomials were required to fit a regression, and the constants changed almost unpredictably. I could find no way to reduce them to a simple estimator which could be used at any frequency value which might be derived from field data. Further, although the very high coefficient of multiple correlation suggests the quadratic fit is adequate for most samples in which no limit is set upon R, the transformation used in the regression engenders false confidence in its power to correct estimates from aggregated populations. Hopkins' transformation of A′ (i.e. $A'/(1 + A')$) compresses the A′ values for extremely aggregated populations into a small region at the tail of the curve (Fig. 3). Second, the curve is intuitively undesirable in that the line cannot produce negative estimates, whereas these are in fact fitted when A′ exceeds 0.96 — corresponding to A of 4.9. Polynomial methods were rejected on these grounds.

The semi-log regression model leads to a relatively simple estimator. Only one special case was found to arise (that of very uniform populations), and the only statistic which is ponderous to calculate is the E(CV) for data in which f is less than 1.

J. H. Darwin (pers. comm.) helped me to dispose of this problem by calculating a set of ap-

TABLE 2. *Expected Coefficient of Variation of Point Distances, Calculated from an Empirical Cubic Curve.*

		B									
	A*	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0			.3564	.3564	.3565	.3566	.3568	.3570	.3571	.3575	.3577
.1	.3579	.3581	.3584	.3588	.3592	.3597	.3602	.3607	.3612	.3618	.3624
.2	.3640	.3630	.3636	.3643	.3650	.3657	.3664	.3672	.3680	.3688	.3696
.3	.3706	.3704	.3712	.3721	.3730	.3738	.3747	.3756	.3765	.3774	.3784
.4	.3782	.3793	.3802	.3812	.3821	.3830	.3840	.3849	.3859	.3868	.3878
.5	.3875	.3887	.3897	.3906	.3916	.3925	.3935	.3945	.3954	.3964	.3974
.6	.3980	.3984	.3994	.4004	.4014	.4025	.4036	.4047	.4058	.4070	.4082
.7	.4113	.4095	.4108	.4122	.4136	.4151	.4167	.4184	.4201	.4220	.4240
.8	.4283	.4261	.4283	.4307	.4333	.4360	.4389	.4421	.4454	.4490	.4529
.9	.4533	.4571	.4616	.4664	.4716	.4772	.4832	.4897	.4968	.5043	.5125
1.0	.5227*	.5214									

$$\log_e E_{cv} = -1.03187 + .48924f^2 - .71817f^4 + .60946f^6$$

*Column A, .10 steps of E_{cv} calculated from measurements of a simulated random population. That for f = 1.0 is $\sqrt{(4-\pi)/\pi}$. B are .01 steps of E_{cv} calculated from the cubic regression equation based on the values of column A.

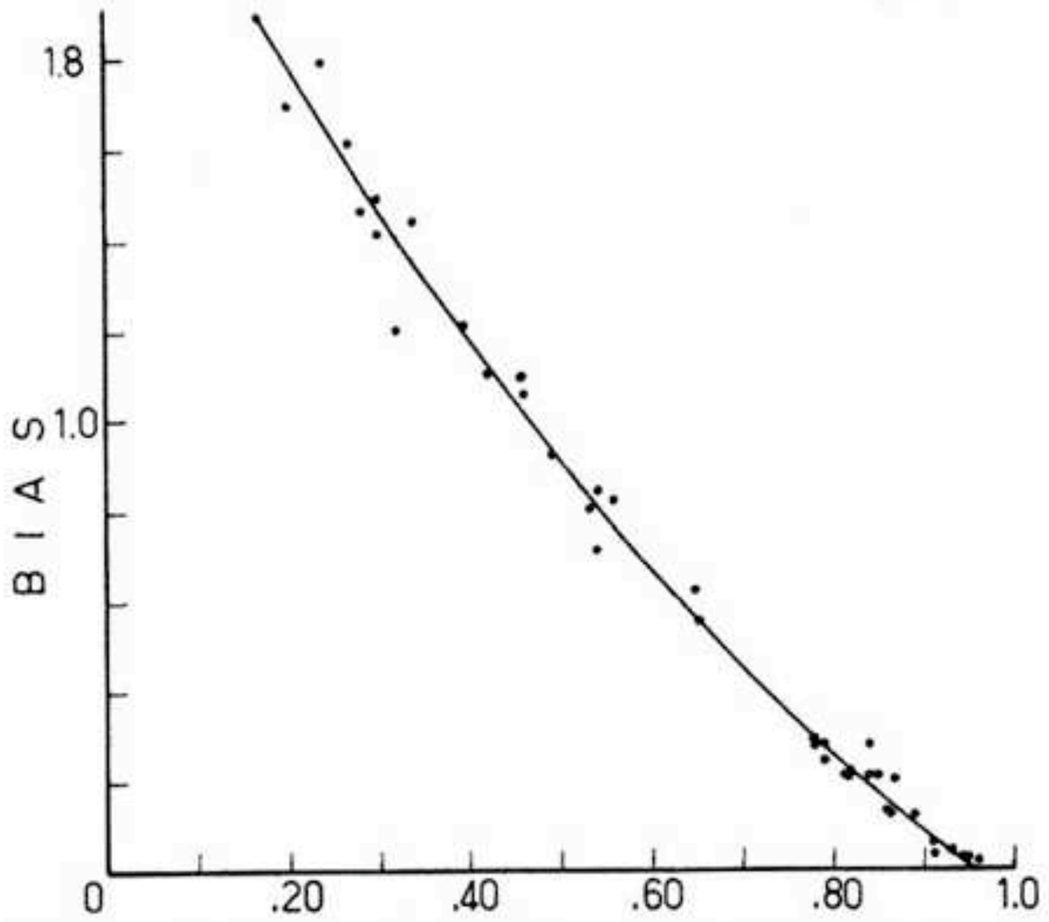

FIGURE 3. *Regression of bias of the* 40 *experimental populations against* $A'/(1 + A')$, *where* A' *is the square root of* $(CV/E(CV)) \sqrt{b_{max}}$. *The regression line is the quadratic curve of*

$$\text{bias} = 2.4847 - 3.7401\ (A'/(1 + A') + 1.2052\ (A'/(1 + A'))^2$$

for which coefficient of multiple correlation is —0.9948.

proximate E(CV) values by sampling a simulated random population in a computer. These are shown in 10 percent steps of frequency up to 90 percent in the second column of Table 2, and the value for 100 per cent frequency is taken as $\sqrt{(4-\pi)/\pi}$. Approximate values for one percent frequency steps were subsequently calculated empirically by least squares, giving—

$$\log_e E(CV) = -1.0319 + 0.4892f^2 - 0.7182f^4 + 0.6095f^6,$$

as given in the body of Table 2.

Mean regressions of log bias against A_{max} were fitted to sample data arbitrarily restricted to include 20, 30, 50, 70 and 90 percent of the point distances. Population samples in which b_1 differed by more than 10 percent from b_2, or the bias exceeded 80 percent, or A_1 was less than 0.5 (see below and caption, Fig. 4), were excluded from the analysis. Semi-log regressions described these data very adequately, and it is clear that as f is reduced towards zero, both $\bar{a}$ and $\bar{b}$ tend from 100 percent values towards 1 (Fig. 4, Table 3).

An objective measure of the rate of change of these constants with change of frequency appears difficult to establish. Fitting the series as dependent variables on f is not valid, because the successive estimates are not independent, and because variance of the constants must change with f. However, since it is known that d/D is unbiased when f is infinitely small (but variance is very large, $a \cdot b^{-A}$ must also tend towards 1, and all three terms must equal 1 when f is zero. It there-

TABLE 3. *Calculated Regression Constants of Bias Against Correction Factors for Unconstrained and Truncated Measurements of Distance**.

			Dependent regression of bias on A		Mean regression		Coefficient of Correction
A. No Constraint on R	N	f	a	b	a	b	
All populations, bias on A_1	40	100	2.814	3.281	2.956	3.376	0.976
All populations, bias on A_2	40	100	2.706	3.059	2.801	3.119	0.983
All populations, bias on A_{max}	40	100	2.736	3.041	2.836	3.102	0.983
28 selected populations bias on A_{max}	28	100	3.459	3.705	3.743	3.717	0.997
B. Constraint on R, giving specified	16	90	2.723	2.963	2.741	2.980	0.995
frequencies, and rejecting samples	27	70	2.752	3.104	2.846	3.189	0.976
specified in text	26	50	2.427	2.638	2.474	2.681	0.984
	30	30	1.631	1.660	1.808	1.805	0.867
	30	20	1.398	1.385	1.501	1.467	0.859

*The constants ultimately used from these iterations were based on mean regression for population samples in which A_1 and A_2 were nearly equal, and bias of the shortest distance estimate was less than —80%.

fore seemed reasonable to propose an estimator of the form—

$$d/D = (1 + 2.473f)\ (1 + 2.717f)^{-A_{max}}$$

as a general estimator when $A_1 \eqsim A_2$. As shown by the regression lines in Figure 4, this turned out to be a good empirical proposition.

When $A_2 > A_1$, use of A_2 as the exponent overestimated true density of the aggregated populations, and some natural aggregated populations yet to be described, by as much as two to three times. Conversely, A_1 as the exponent underestimated true density. Subsequently, in attempting to bring these difficult cases into line, it was found by trial and error that the mean of the two estimates was consistently closer to true density, so—

$$\overline{(d/D)} = \frac{(1 + 2.473f)}{2}\left[\ (1 + 2.717f)^{-A_1} + (1 + 2.717f)^{-A_2}\ \right]$$

becomes the basic general estimator wherever A_1 exceeds 0.5.

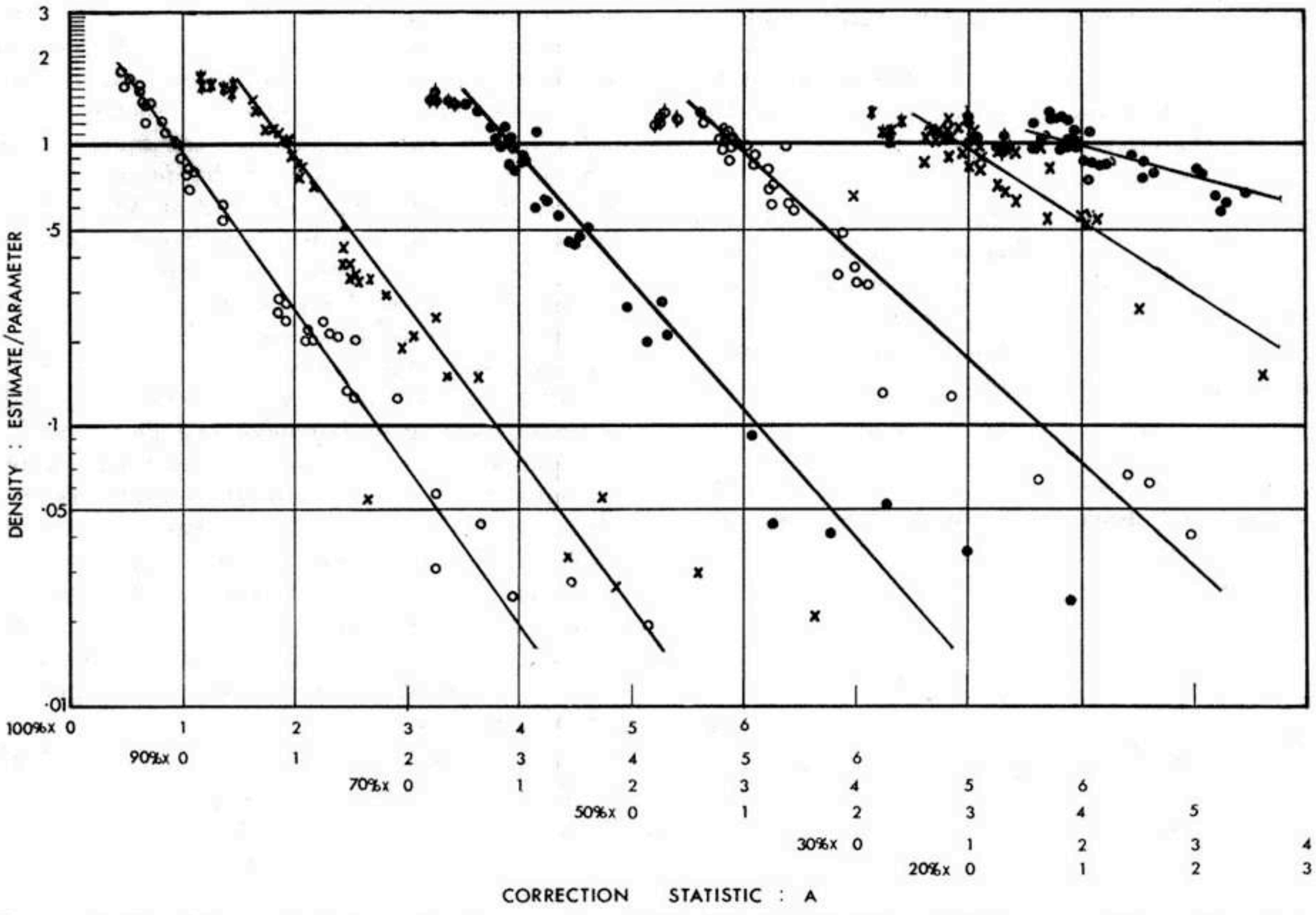

FIGURE 4. *Log bias against* A_{max}, *for frequencies of* 100, 90, 70, 50, 30 *and* 20 *percent. The scales of the abscissae are moved in steps from left to right to separate the points for each frequency. The calculated regression lines are from equation* 3 (*assuming* $A_1 = A_2$), *and clearly show the pivoting of the regression lines from maximum slope b and highest a intercept at* 100 *percent f, towards zero slope and a intercept at* 1, *as frequency drops towards zero* (*right side*). *Displacement of values* (*slashed symbols*) *at the top left of each set* (*less than* 100 *percent f*) *show the asymptotic nature of* A_{max} *in uniform populations* ($A < 0.5$). These *values were extracted and used to calculate the quadratic equation for uniform populations.*

As shown in Figure 4, six of the point distance estimates, drawn from uniform populations in which $A_1 < 0.5$, give log bias $-A_{max}$ intercepts which depart abruptly from the regression line indicated by the other 34 values. These represent the behaviour of uniform or nearly uniform populations and appear to form the only special case of the method for which a basically different calculating formula is necessary. This arises in uniform populations because r_p must always be less than the distance between neighbours. In the case of the grid square lattice distribution, r_p cannot exceed 0.707 of the grid distance; within a triangular lattice it cannot exceed 0.625 of the grid distance. Consequently, imposition of any limit R on the search for r_p and r_n effectively makes all r_n's greater than any r_p. Therefore

$$\Sigma r_p n^2 N / \Sigma r_n p^3$$

(i.e. b_1) tends rapidly towards zero, and is a measure of asymptotic behaviour of the index of dispersion. The data for these populations were therefore extracted and an independent regression of the bias of the estimate was fitted as the dependent variable against frequency, giving—

$$d/D = 1.0065 + 0.3401f + 0.4235f^2.$$

It provides a simple means of correcting the bias of the point distance estimate in these cases, and the need to employ it is invariably identified by $A_1 < 0.5$.

SUMMARY OF THE CALCULATING FORMULAE

The following brief summary draws together the formulae used in succeeding sections of this paper.

(1) The shortest distance estimate is calculated from—

$$d = p/\pi \left(\sum_{o}^{R} r_p^2 + (N - p)R^2\right).$$

(2) The correction terms A_1 and A_2 are calculated from—

$$A_1 = CV/E(CV) \sqrt{\Sigma r_p n^2 N / \Sigma r_n p^3}$$

$$A_2 = CV/E(CV) \sqrt{\Sigma r_p m^2 N / \Sigma r_m p^2 n}.$$

(3) If $A_1 > 0.5$—

$$\overline{(d/D)} = \frac{(1 + 2.473f)}{2} \left[(1 + 2.717f)^{-A_1} + (1 + 2.717f)^{-A_2} \right]$$

or, when $A_1 = A_2$,

$$d/D = (1 + 2.473f)\ (1 + 2.717f)^{-A_1}.$$

(4) If $A_1 < 0.5$,

$$d/D = 1.0065 + 0.3401f + 0.4235f^2.$$

TESTS WITH PAPER DOT AND FIELD POPULATIONS

So far, the formulae given in the above summary have been tested by the writer, colleagues at this station and students of Christchurch Teachers Training College on 29 populations for which density was also determined by plot samples or total counts. Of the 29, 11 were simulated on paper and 18 are field populations. Thirty-seven independent distance measurement experiments have been undertaken in them. As shown by the following list, two populations — a square chain area of the mountain daisy (*Celmisia spectabilis*), and the pines of Compartment 2 at Ashley State Forest — became favoured experimental subjects. Efforts were concentrated on the daisy population because its intense clumping posed the first encounter with extremely biased shortest distance estimates, and ultimate recognition of clumps within clumps and the use of A_2 (described as second-order aggregation by Batchelor and Bell (1970)). The pine plantation became favoured because it represented a typical production forest inventory problem, and because estimates of density were available from forest inventory records. The compartment lies across ridges, slopes and shallow gullies, with concommittent gradients of tree density and size. In addition, it contained thinning trials, extensive gaps caused by storms and other gaps in scrub-choked gullies.

The characteristics of these populations and the number of distance sample points used to sample them are summarised in the following list.

Paper dot populations

(1-3) Three uniform populations simulated by a grid square of four points within which point distances were measured from random co-ordinates. The grid distance is the neighbour distance (sample sizes 100, 200, 200).

(4-5) Two random paper dot populations of 400 and 500 members in which three distance samples were taken from systematically spaced

points on randomly chosen lines (sample sizes 153, 159, 170).

(6-11) Six aggregated populations of 360 to 600 members, compiled by arbitrary clustering procedures (Batcheler 1971) and sampled from systematic points on random lines (sample sizes 79, 97, 112, 119, 147, 152).

Natural populations

(12) Three seperate experiments on random lines in a final crop plantation of *Pinus radiata* (Compartment 2, Ashley State Forest), in which an independent estimate of 346 ± 20/ha (P = 0.95) was obtained by counts within circular sample plots (sample sizes 100, 253, 348).

(13) One experiment using systematically spaced points on random lines in thinned *P. radiata* on a dissected slope (Compartment 4, Ashley State Forest), which was estimated (692 ± 59/ha, P = 0.95) from counts within 56, 0.04 ha plots (278 distance samples).

(14) An unthinned plantation of *P. nigra* which had suffered extensive natural mortality (Compartment 66, Eyrewell State Forest), which was estimated at 1132 ± 84/ha (P = 0.95) from counts on 38, 0.07 ha plots (sample size 129).

(15-16) Two experiments in a naturally regenerated stand of *P. radiata* (Compartment 68, Balmoral State Forest) which had been thinned twice, and marked for final crop thinning. The total stand was estimated at 677 ± 47/ha (P = 0.95) on 56, 0.08 ha plots. The entire marked crop had been tallied at 326 trees/ha as a check against calculated specifications for the crop. Fifty-six sample plots gave an estimate of 341 ± 32/ha (P = 0.95). Both total stand and marked crop were estimated by distances, on 10 random sample lines (292 distance samples for both total stand and crop).

(17) A thinned stand of *P. radiata* at Tiritea Reserve, Manawatu, in which all trees of the sampled area were counted (200 distance samples).

(18) A natural population of *Beilschmedia tawa* at Tiritea in which all trees of the sampled area were counted (200 distance samples).

(19-21) Three populations of indigenous beech (*Nothofagus* spp) at Makahu, Kaweka State Forest, Wellington, and Craigieburn State Forest, Canterbury, in which all trees were counted (sample sizes 134, 200, 234).

(22) A single experiment in a 405 m² area of dense hard fescue tussocks (*Festuca novae-zelandiae*) at Craigieburn in which all tussocks were counted. Distances were sampled on a systematic grid (sample size 100).

(23) A single experiment in a 405 m² area of hard fescue tussocks, patchily distributed across an area dissected by erosion pavement. All tussocks were counted, and distances were sampled on a systematic grid (100 samples).

(24) Six different experiments in a 405 m² area of *C. spectabilis*. All rosettes were counted, and distances were measured from both the centres and intercepts of a systematic grid, and from random points (sample sizes 100, 100, 100, 121, 200 and 200).

(25-26) Two experiments to estimate density of rabbit faecal pellets at Cairnhill, Central Otago. Each consisted of a single line of 100 systematically spaced points on a line, at each of which pellets were counted in a 0.09 m² plot (1.98 ± 0.85 and 2.52 ± 1.08/m², P = 0.95) and distances were measured (sample sizes 100, 100).

(27-29) Three experiments with populations of hare faecal pellets at Harper-Avoca field station, Canterbury. Independent counts were made in 100 0.09 m² plots giving 7.65 ± 1.98 m², 5.94 ± 1.62 m² and 10.17 ± 2.79 m² (P = 0.95) (distance samples 100, 100, 100).

Overall, distance sample size averaged 156 points, and ranged from 79 to 348.

Results of population experiments

Uncorrected and corrected estimates were calculated for approximately 15 percent steps of the frequency range 25-100 percent. Eight others, representing lower frequency values, were calculated from the data of the three biggest experiments. All are summarised as a fraction of the parameter or sample plot estimate (i.e. the expected value is 1.0) in Figures 5 and 6, and the accumulated departures from the expected values are shown in Table 4.

Estimates obtained by equation (1), the shortest distance estimator, effectively indicate the general

distribution of the populations (Fig. 5). Only five of the 36 results at 100 percent frequency lie within ± 10 percent of the expected value, and could accordingly be considered random or nearly so. Thirteen of the remainder were uniform, giving estimates up to almost twice the expected value, and 18 were aggregated, giving values ranging down to five percent of the expected value. As expected on theoretical grounds, the biases in these estimates are progressively smaller as f diminishes towards zero (left side of Fig. 4).

TABLE 4. 135 *Uncorrected (left) and Corrected (right) Estimates Derived from 37 Experiments with 29 Populations.*

Difference between parameter and estimate (±%)	Uncorrected estimates: No. within range	Uncorrected estimates: Accumulated %	Corrected estimates: No. within range	Corrected estimates: Accumulated %
1-10	24	17.8	80	59.3
11-20	14	28.1	23	76.3
21-30	18	41.5	23	93.6
31-40	12	50.4	6	97.8
41-50	8	56.3	2	99.3
51-60	7	61.5	0	99.3
61-70	8	67.4	0	99.3
71-80	16	79.3	1	100.0
81-90	19	93.3	0	100.0
91-100	9	100.0	0	100.0
Totals	135		135	

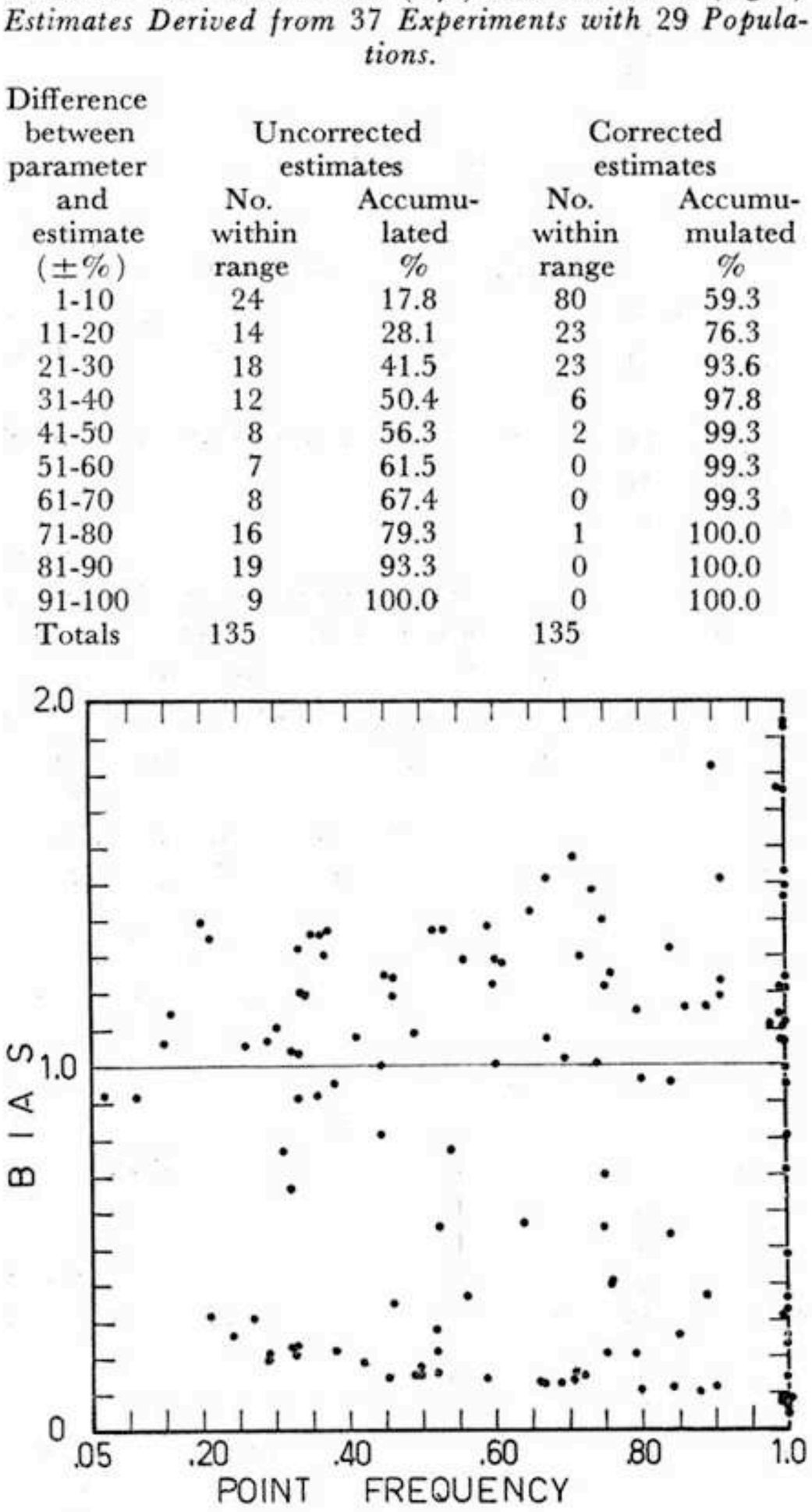

FIGURE 5. *Uncorrected estimates of density of paper dot and natural populations (from equation 1), plotted against frequency.*

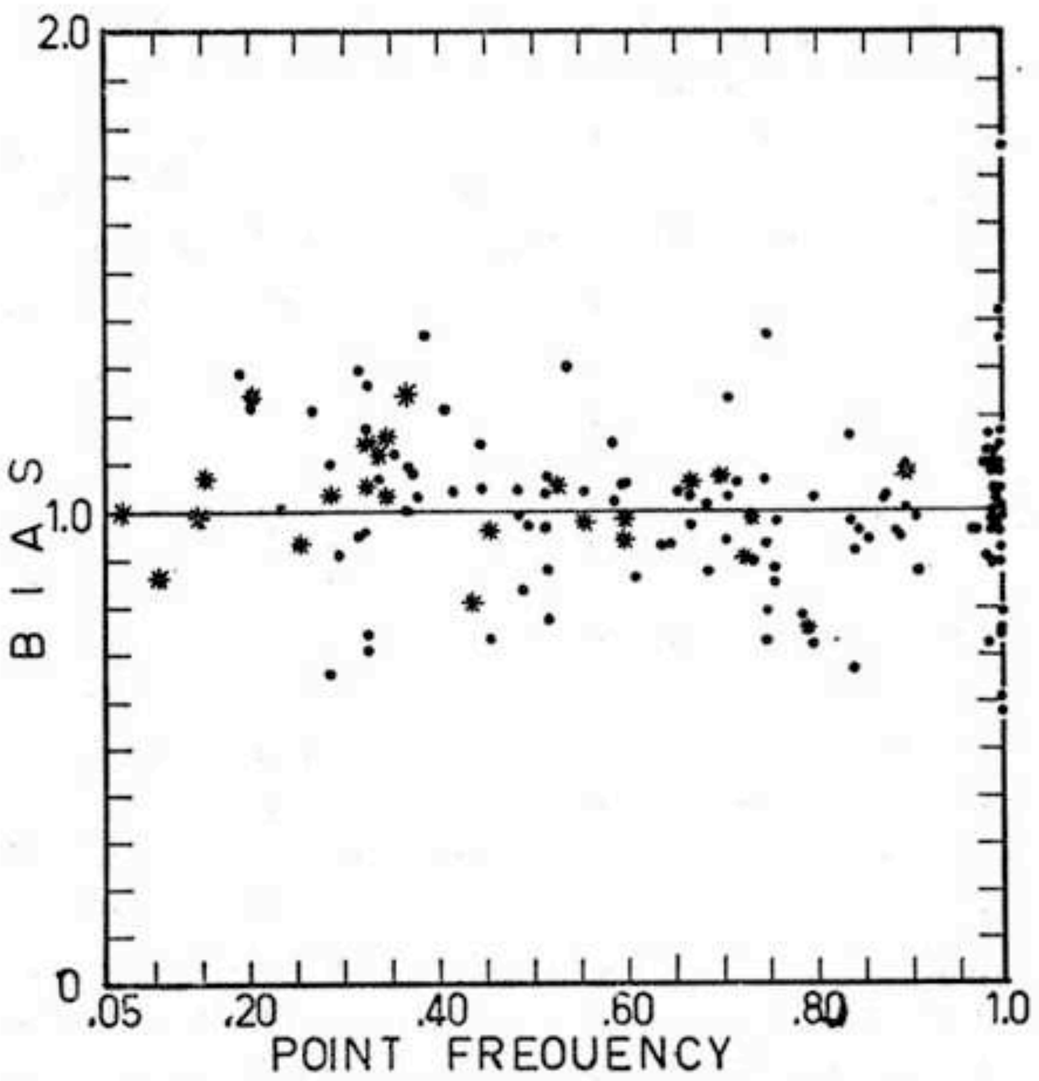

FIGURE 6. *Corrected estimates of density of paper dot and natural populations, from equations 2, 3 and 4 as appropriate. Those calculated from the quadratic equation are shown as asterisks.*

In contrast, the corrected estimates obtained from equations (2) and (3) are strongly grouped, and with only three exceptions among the 135 calculated values all are within ±40 percent of the expected value (Fig. 6). By accumulating these as departures from one, 59 percent lie within ± 10 percent, 76 percent within ± 20 percent, and 93 percent lie within ±30 percent of the expected value (Table 4). This empirically infers that there is about a 93 percent chance of an estimate being within ± 30 percent of the true population value, provided an estimate is based on at least 30 positive records from 100 or more sample points.

DISCUSSION

Density

In pursuing this study there has been no point in lamenting that the sampling distributions of the distance measurements have not been derived from

theoretical considerations. Rather, given the formidable nature of the theoretical problem, the evidence that many different distance functions yield a meausre of non-randomness (Hopkins 1954, Moore 1954, Holgate 1965a and 1965b, Eberhardt 1967), and given evidence that distance ratios can be used in non-random populations to correct dispersion-dependent bias of a density estimate (Morisita 1957, Batcheler and Bell 1970, Batcheler 1971, James 1971), there has been ample justification for digging into the problem from every imaginable angle.

Considerable licence was employed in accepting or rejecting samples simply because the objective was to develop an empirical function which is free from gross sampling problems. The coefficient of variation, the r_m distance and the quadratic regression approach to estimating uniform populations from truncated distance measurements, were progressively brought into the picture as a tedious string of analyses whittled away at the obvious remaining errors. The only completely subjective element in this process has been the decisions taken as to whether occasional "maverick" estimates indicated some fundamental cause worth pursuing, or whether they could be written off to chance. Nonetheless, like any empirical model, the distance formulae presented must be presumed to carry some unmeasured taint of the data from which they have been derived.

At present, the only real test of their worth is the evidence from the 29 paper dot and natural population experiments. This is summarised graphically in Figures 5 and 6 and numerically by the statement that 93 percent of the corrected estimates lie within 30 percent of "true density" (Table 4) — a satisfactory result for most practical purposes. Beyond this, any statement about the reasons why seven percent of the estimates remained intractably bad becomes inseparable from an object lesson in the pitfalls of experimental investigation.

Among the estimates for paper dot populations, where there is of course no error in the density parameter, 13 percent of the corrected distance estimates lay outside the range which would be expected by chance from point distance theory, assuming the populations to actually be random.* These unacceptable values came from intensely clumped populations in which, as with the computer-simulated populations, considerable numbers of sampling points had to be declared invalid because they were closer to the edge of the population map than to the nearest member. However, if the populations were assumed to lie within regions of the respective maps such that no sampling point was nearer to the edge than to the nearest member, the estimates in all except one instance were within the limits for a random population.

The field populations pose a much more mixed bag.

In one pine plantation (Eyrewell, Cpt. 66), where distances gave an estimate of only 74 percent of the plot-estimated density, it was subsequently learned that double-leader trees, forking below 4 ft. 6 in., are tallied as two trees in routine inventories.

Among the other natural populations, particularly aggregated ones, three common problems were encountered. Some study populations were so small, or population density was so low, that several measurements were made to particular members and their neighbours or, particularly in two experiments (one beech forest, one eroded *Festuca* area), several measurements were made across the delineated population boundary to reach the nearest member. In experiments with hare pellets, which yielded distance estimates of 58-69 percent of the plot values, the possibility of a significant degree of clustering within primary clusters (second order aggregation) was overlooked. This was later recognised to be a normal attribute of pellets which are mostly voided in groups in preferred feeding areas of the range.

The most common problem, however, was large probable limits of error of the sample plot counts.

---

*$\chi^2_{2Ndf} = 2d_{(\text{upper or lower})} \pi \Sigma r_p^2$ (Kendall and Moran 1963) where d_{upper} and d_{lower} are the upper and lower confidence limits of d, and χ^2_{2Ndf} is Chi-square for 2N degrees of freedom.

These were ± 14, 26, and 71 percent of the means in the hare pellet experiments and ± 43 percent in both of the rabbit pellet counts ($P = 0.95$). Such broad confidence limits for the estimates clearly inhibit any precise statement about relative or absolute accuracy of the distance estimate compared with plot counts: both may be right; one may be right and the other wrong; both may be equally wrong. The only definite proof is yet more testing.

Note on the index of non-randomness

If bias of the point distance estimate is a function of dispersion as measured by departure from the Poisson case, and if, as concluded above, the corrected estimate is unbiased, then A_1 and A_2 must be measures of non-randomness. This hypothesis cannot be tested with the data from field populations because in non-random populations $S^2/\bar{x}$ may be unpredictably influenced by the choice of plot size (Greig-Smith 1964, Kershaw 1957, Southwood 1968, Pielou 1969, Iwao 1972). On the other hand, since the computer-simulated populations were nominally of equal density (ignoring edge effects) and since density was sampled by plots of equal size, it was expected that A_1 and A_2 would be correlated with variance, or its dimensionless statistic, coefficient of variation. For the 28 populations in which A_2 differed from A_1 by less than 10 percent, their geometric mean, —

$$\bar{A}g = -.0075 + .0203\ CV_p,$$

or, within the error limits of regression, $CV_p = 50Ag$ i.e. $\bar{A}g$ is a linear function of the square root of variance.

This attribute of $\bar{A}$ is a particularly useful measure of non-randomness, because it can be estimated in unknown populations by the single-stage distance sampling method without need for choice of the sampling unit. The size of the sampling unit (r_p, r_n, r_m) is automatically related to density of the population. It is therefore more versatile than dispersion indices derived from bounded plot sampling because, as shown by Pielou (1969) and Iwao (1972), even powerful indices such as Lloyd's (1967) index of mean crowding and Morisita's (1959) $I\delta$ index cannot be presumed unique to the population without analysis of a series of plot sizes sufficient to show that all or nearly all of at least the smallest plots lie completely within any density phase. Furthermore, these indices are known to be valid only if populations occur in randomly distributed patches of different density phases, within which the members of a patch are randomly distributed.

The values for A_1 and A_2 calculated for the paper dot and natural populations are summarised in Table 5 with their corresponding uncorrected and corrected density values, to illustrate the spectrum which can be expected from a wide range of biological populations. Uniform grid square populations are characterised by $A < 0.45$. Pine plantations, in which the original grid distribution is broken by silvicultural thinning, natural mortality and such irregularities as scrub-choked gullies, give A in the order of 0.6-0.8, tending to approximately 0.9 in relatively uneven stands.*

A_1 exceeds 0.95 (the average value for random populations from equation (3)) in aggregated populations. As aggregation intensifies to the degree exemplified by large mats of *Celmisia* rosettes, and by rabbit pellets clumped into "pill heaps" related to territorial display, A_1 is in the order of 2.5, and A_2 may be as high as 3.5-4.

When R is set such that f is less than 100 percent, A_1 and A_2 differ from the 100 percent value according to the pattern of the population. In uniform populations, it rises rather abruptly from zero as R is raised to the extent necessary to take in the uniformly spaced neighbours. In aggregated populations however, A may rise or fall as f (and R) increases, in opposition to the effect of pattern on the point distance estimate of density. As shown in Figure 7 where six characteristic patterns of the distance measures are shown against f, it appears to be impossible to derive a unique index of non-randomness from truncated distance measurements. Nevertheless, the broad pattern of deviations from randomness are usually evident by $A >$ or < 1 when f is 30 percent or more, particularly if populations are compared by the data for a common frequency.

*One traverse line through a recently studied compartment of *P. nigra* which had been badly influenced by frost and scrub-competition gave $A_1 = 1.01$, suggesting a slightly aggregated distribution.

Field practice and sampling requirements

There are several "tricks of the trade" in estimating density by this distance technique. Sampling at systematic intervals on random lines is advocated because of the speed with which sample points can be objectively located. Each line of points is then considered as a single replicate and variance and confidence limits of the mean can validly be calculated by normal distribution techniques (Jowett 1967). However, if density follows some discernible gradient along the line, or the line traverses a patchwork of density and dispersion phases, a single estimate of density for the whole line is rarely unbiased. So far as can be judged from present experience, this problem is best countered by subdividing the line into relatively homogenous units (strata?), calculating the density for each unit, and combining them into a single estimate for the line by weighting according to density and the number of sample points in each unit. The weighted mean is then used as the replicate. The need to subdivide in this way is invariably indicated by ordering the point-distances from smallest to largest and calculating about four estimates at uniform intervals of f. A substantial decline of corrected estimates associated with increasing f usually suggests a polymodal distribution of distances, and the advisability of subdivision.

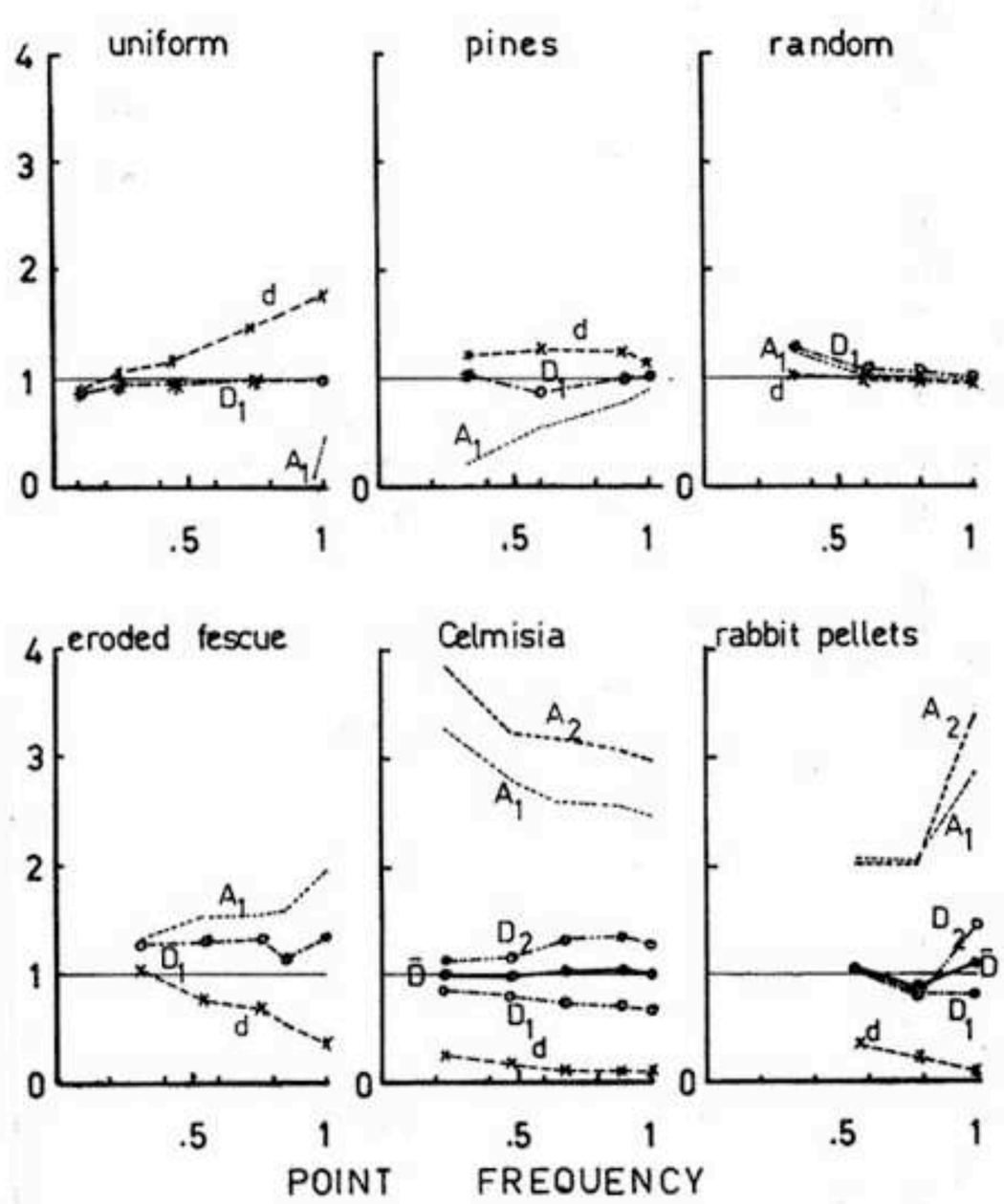

FIGURE 7. *Six examples of the magnitude and trend of* A_1 *and* A_2*, uncorrected estimates and corrected estimates, illustrating patterns which can be expected in biological populations.*

Of course, it is relatively difficult to estimate density of sparse populations, particularly if members cannot easily be seen from the sample point. For a given effort, however, it is usually advantageous to keep R relatively small (say, large enough to give f = 30 percent or more) and concentrate the available sampling effort on establishing a large number of sample points. This recommendation arises from several interwoven aspects of behaviour of the distance equation. First, since the distribution of point distances is root-skewed, a very large increase of R is usually required beyond the modal distance to measure the large distances at the "tail-end" of the distribution. The effort is rarely worthwhile; because "tail-end" distances, of large value, often unduly disturb the estimate of density, and, when large R is employed, the resulting estimate is more severely biased and requires a formidable order of correction (see, for example, Fig. 7E and 7F).

The required number of sample points therefore differs according to the pattern of dispersion, and ease of sighting members of the population.

The estimate for any replicate should be based on at least 30 measurements of r_p, so that N would normally be designed to range from about 30 for a replicate of a uniform population of large entities such as trees (and no R limit would be set) to about 100 for a sample of aggregated populations where A_1 exceeds about 2, or members are too difficult to find at distance from the sample point (so f may be 0.3 or even less). When variance between the replicates is not large, it will turn out that about five replicates will be required to yield an estimate with less than ± 10 percent probable error at 95 percent probability. Accordingly, 150-500 p distances, distributed over about five replicates constitute an adequate sample under average conditions. But it is imperative that these samples

TABLE 5. *Summary of data at f = 100% for the Paper Dot and Natural Populations to Show A values for Different Kinds of Populations. Bias of Uncorrected Estimates, the Corrected Estimates and Coefficient of Variation of the True Values are also Tabulated.*

Type of population	Samples	Uncorrected estimate	Corrected estimate	A_1	A_2	C V (plots)
Paper dot						
Uniform grid	100	1.930	1.040	0.472	—	—
Uniform grid	200	1.940	1.020	0.468	—	—
Uniform grid	200	1.760	0.994	0.436	—	—
Random 1st sample	153	1.212	1.098	0.882	—	—
2nd sample	159	0.950	0.887	0.905	—	—
Random	170	0.994	0.953	0.926	—	—
Aggregated	79	0.258	0.956	2.056	—	—
Aggregated	97	0.243	1.146	2.142	—	—
Aggregated	112	0.712	1.309	1.127	—	—
Aggregated	119	0.346	1.089	1.787	1.876	—
Aggregated	152	0.331	1.167	1.920	—	—
Natural populations						
Pine plantation (Ctp. 2)	100	1.239	1.088	0.858	—	
Pine plantation (Cpt. 2)	253	1.538	1.040	0.659	—	31
Pine plantation (Cpt. 2)	348	1.064	1.043	0.942	—	
Pine plantation (Cpt. 4)	278	1.109	1.055	0.920	—	36
Pine plantation (Cpt. 66)	129	1.145	0.735	0.619	—	22
Pine plantation (Cpt. 68)	292	1.219	0.920	0.742	—	26
Pine plantation (Cpt. 68)	292	1.333	1.014	0.749	—	35
Tiritea pine plantation	200	1.467	1.174	0.750	0.823	—
Tiritea tawa	200	0.775	0.884	0.884	—	—
Makahu beech forest	200	1.084	1.117	0.981	—	—
Craigieburn beech forest	200	0.630	0.634	0.952	0.973	—
Craigieburn beech forest	200	0.800	0.956	1.089	1.099	—
Eroded *Festuca*	100	0.369	1.363	1.956	—	—
Dense *Festuca*	100	1.125	1.015	0.879	—	—
Celmisia spec.	100	0.094	0.983	2.465	2.958	
Celmisia spec.	100	0.117	1.013	2.252	2.997	
Celmisia spec.	100	0.088	0.743	2.609	2.569	143
Celmisia spec.	121	0.083	0.871	2.402	2.991	
Celmisia spec.	200	0.096	0.796	2.468	2.665	
Celmisia spec.	200	0.090	0.609	2.225	2.578	
Rabbit pellets	100	0.062	1.121	2.903	3.375	104
Rabbit pellets	100	0.046	1.761	3.592	3.870	104
Hare pellets	100	0.296	0.694‡	1.609	—	132
Hare pellets	100	0.193	0.710‡	1.953	—	132
Hare pellets	100	0.476	0.581‡	1.110	—	136

‡ r_m should have been measured, see text.

be well spread throughout the population. As with the traditional pilot trial practice of step-wise analysis of counts within bounded plots to determine what constitutes an adequate sample, there is no surer guide to distance sampling requirements, than circular reasoning from results already in hand.

Basically, a large though undefinable number of sample points must be advocated for surveys of unknown populations because determination of the proportions of different density phases which makes up the total is invariably the paramount task of any density assessment.

ACKNOWLEDGMENTS

I am particularly indebted to Mr I. L. James for allowing me to use his simulated population data, for writing programmes to calculate some of the sampling statistics used, and for permission to use data from field experiments; to Dr W. A. Laycock, U.S. Forest Service, for use of some data; to Dr J. H. Darwin, DSIR, for

guidance with some problems; to many people, particularly Messrs D. J. Bell, C. W. Barr, P. Crozier, R. Hodder, M. Harvie, R. R. McPherson and the Mathematics students of Christchurch Teachers Training College, and Ms K. P. Smith, H. Arnold and H. M. Brasell, for their help in the field and laboratory; also, to J. H. Darwin, I. L. James, I. G. Crook, and A. B. Robson, for their criticisms of an earlier draft of the paper.

REFERENCES

BATCHELER, C. L. 1971. Estimation of density from a sample of joint point and nearest-neighbour distances. *Ecology* 52: 703-709.

BATCHELER, C. L.; BELL, D. J. 1971. Experiments in estimating density from joint point- and nearest-neighbour distance samples. *Proceedings of the New Zealand Ecological Society* 17: 111-117.

CLARK, P. J.; EVANS, F. C. 1954. Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology* 35: 445-453.

COTTAM, G. 1947. A point method for making rapid surveys of woodlands. *Bulletin of the Ecological Society of America* 28: 60.

COTTAM, G.; CURTIS, J. T. 1956. The use of distance measures in phytosociological sampling. *Ecology* 37: 451-460.

EBERHARDT, L. L., 1967. Some developments in "distance sampling". *Biometrics* 23: 207-215.

GREIG-SMITH, P. 1964. *Quantitative Plant Ecology*. Butterworths, London.

HOLGATE, P. 1965a. Tests of randomness based on distance methods. *Biometrika* 52: 345-353.

HOLGATE, P. 1965b. Some new tests of randomness. *The Journal of Ecology* 52: 261-266.

HOPKINS, B. 1954. A new method for determining the type of distribution of plant individuals. *Annals of Botany*. New Series 18: 213-227.

IWAO, S. 1972. Application of the M*-M method to the analysis of spatial patterns by changing the quadrat size. *Researches in Population Ecology* 14: 97-128.

JAMES, I. L. 1971. A computer study of corrected density estimators for distance sampling of non-random populations. *Diploma of Agricultural Science, Thesis,* Massey University.

JOWETT, G .H. 1967. An extension of the application of the statistical formula $S/\sqrt{n}$. *New Zealand Mathematics Magazine* 4: 61-63.

KENDAL, M. G.; MORAN, P. A. P. 1963. *Geometrical probability*. Griffin, London.

KERSHAW, K. A. 1957. The use of cover and frequency in the detection of patterns in plant communities. *Ecology* 38: 291-299.

LLOYD, M. 1967. Mean crowding. *Journal of Animal Ecology* 36: 1-30.

LYON, J. L. 1968. An evaluation of density sampling methods in a shrub community. *Journal of Range Management* 21: 16-20.

MOORE, P. G. 1954. Spacing in plant populations. *Ecology* 35: 222-227.

MORISITA, M. 1954. Estimation of population density by spacing method. *Memoirs of the Faculty of Science,* Kyushu University, Series E1: 187-197.

MORISITA, M. 1957. A new method for the estimation of density by the spacing method applicable to non-randomly distributed populations. *Physiology and Ecology* 7: 134-144.

PIELOU, E. C. 1959. The use of point-to-plant distances in the study of the pattern of plant populations. *Journal of Ecology* 47: 607-613.

PIELOU, E. C. 1969. *An Introduction to Mathematical Ecology*. Wiley, New York.

SIMPSON, G. G.; ROE, A.; LEWONTON, R. C. 1960. *Quantitative Zoology*. Harcourt, Brace, New York.

SOUTHWOOD, T. R. E., 1968. *Ecological Methods*. Methuen, London.

THOMPSON, H. R. 1956. Distribution of distance to nth nearest neighbour in a population of randomly distributed individuals. *Ecology* 37: 391-394.

ERRATA: BATCHELER, C.L. (1973). ESTIMATING DENSITY AND DISPERSION FROM TRUNCATED OR UNRESTRICTED JOINT POINT-DISTANCE NEAREST-NEIGHBOUR DISTANCES. Proc. N.Z. Ecol. Soc. 20: 131-147.

In para. 2 p. 139 it is stated that the average of two estimates of density, derived from

$$D_1 = d/ab^{-A_1},$$

and

$$D_2 = d/ab^{-A_2},$$

consistently gave the most accurate estimate for the population;

i.e., $$\bar{D} = (D_1 + D_2)/2.$$

However, the formula given on p. 139, and in the summary of calculating formulae (No. 3, p. 140) for calculating this average, is wrong. It should be

$$\bar{D} = \tfrac{1}{2}\left(\frac{d(b^{-A_1} + b^{-A_2})}{ab^{-(A_1 + A_2)}}\right).$$

which simplifies to

$$\bar{D} = \frac{d}{2a}\ (b^{A1} + b^{A2})$$

Perpetration of this blunder in print is regretted.