but since logging operations and the burning and reburning that followed this logging, aspen and white birch have taken over the role as dominants.

Thirty-two islands and 5 mainland areas were chosen for the study. Three species of hibernators, *Eutamias minimus, Tamias striatus* and *Zapus hudsonias* were taken on the mainland with only the *Zapus* being found on the islands. All six of the non-hibernating species taken were found on the islands. These include *Microtus pennsylvanicus, Synaptomys cooperi, Clethrionomys gapperi, Peromyscus maniculatus, Blarina brevicauda* and *Sorex cinereus*.

The degree of occupancy changed from 4 out of 21 islands in 1950 to 21 of 24 islands in 1952. Along with this increase in the proportion of islands occupied there was an increase in the number of species found on most of the islands although the average number of species found on each occupied island did not change appreciably. This increase in islands occupied was correlated with an increase in the density of the mainland populations.

The water seemed to be an efficient barrier to travel during the summer but the ice of winter became a highway for dispersal of the various non-hibernating species in the winter.

Considerable difference was found in the vulnerability of the several species to trapping. The *Peromyscus* were readily caught. The *Clethrionomys* were not as easy to capture as the former but were still readily taken. The *Microtus*, however, posed a different problem. They were difficult to capture and on two islands of about 2.2 and 4 acres respectively we were not able to eliminate them after 16 days of trapping on a 15-foot grid.

The populations of small mammals fluctuate quite widely and the several populations appear to be somewhat independent of each other. The islands appeared to lag behind the mainland in the development of their populations.

References

Errington, P. L. 1946. Predation and vertebrate populations. Quart. Rev. Biol., **21**: 144-177, 221-245.

Manville, R. H. 1950. A comparison of trapping methods. Jour. Mamm., **31**: 377-383.

McCabe, T. T. and I. McT. Cowan. 1945. *Peromyscus maniculatus macrorhini* and the problem of insularity. Trans. Royal Canadian Inst., **25**: 117-216.

Rosendahl, C. O. and F. K. Butters. 1928. Trees and shrubs of Minnesota. vii + 385 pp. (Ref. pp. 7-12). Univ. Minnesota Press, Minneapolis.

Thiel, G. A. 1947. The geology and underground waters of northeastern Minnesota. Minn. Geol. Survey Bull., **32**: xiii + 247 pp. (Ref. pp. 154-162).

U. S. Dept. of Agr. 1941. Climate and Man. U.S.D.A. Yearbook. 1-1248 pp. (Ref. pp. 925-934).

Zumberge, J. H. 1952. The lakes of Minnesota, their origin and classification. Minn. Geol. Survey Bull., 35. xiii + 99 pp. (Ref. pp. 1-9).

DISTANCE TO NEAREST NEIGHBOR AS A MEASURE OF SPATIAL RELATIONSHIPS IN POPULATIONS

Philip J. Clark and Francis C. Evans
Institute of Human Biology, University of Michigan, Ann Arbor, Michigan

Introduction

The pattern of distribution of a population of plants or of animals is a fundamental characteristic of that population, but it is a feature that is extremely difficult to describe in precise and meaningful terms. The distributions exhibited by populations of living organisms in their natural environments include an almost infinite variety of patterns. However, the lack of an adequate method of description has prevented the development of any system of pattern classification other than a most general one. The situation is further complicated by the fact that it has generally been considered necessary, for practical reasons, to use samples rather than entire populations as the source of distributional information, and this frequently introduces bias and inaccuracy into the estimates of population parameters. The whole problem of the measurement and description of patterns of distribution has recently been the subject of an excellent review by Goodall (1952) in reference to plant populations. It is evident from this review that many important concepts of phytosociology are based upon the assumption that the individuals of most plant populations are distributed at random. This assumption is no longer a tenable one, and it is probably even less applicable to animal populations. A number of methods of demonstrating the occurrence of non-random distribution are now available, but the degree of departure from random expectation is much more difficult to ascertain, and the significance of differences in the distribution pattern of two or more populations is therefore hard to evaluate. Improvement in the quantitative analysis of distribution is greatly to be desired and would surely facilitate the interpretation of dispersion patterns.

Gleason (1920) and Svedberg (1922) seem to have been the first ecologists to test natural distributions of organisms for conformity to random expectation, the former employing the binomial distribution and the latter making use of the Poisson series. Several measures have since been used to test the hypothesis of random dispersion. The data to which these measures have been applied are usually counts of individuals or records of presence or absence in sample plots or quadrats and are generally expressed in terms of frequency. Such data are strongly influenced by the size of quadrat used in their collection (Curtis and McIntosh 1950).

Recognition of the widespread existence of non-randomly distributed populations has led to the development of mathematical models based on various assumptions about the natural forces active in the formation of particular patterns. Some models (Ashby 1935) attribute primary importance to variations in environmental factors, others (Neyman 1939) to behavioral characteristics of the species concerned. Data from numerous sources have been fitted to these models, with varying degrees of success.

The distance from one individual to another provides a variable for the measurement of spacing that obviates the use of quadrats and therefore eliminates the effect of quadrat size. Viktorov (1947, *fide* Goodall 1952) measured the distance from a given plant to every other individual connected to it by a straight line not crossing another plant and used this information to estimate the variability of the distances. Cottam and Curtis (1949) attempted to ascertain the average distance between trees in a forested area by using "randomly selected" pairs of individuals. Dice (1952) seems to have been the first to use distance between nearest neighbors in measuring departure from randomness. Dice's procedure consists of measuring the distance from a randomly selected individual to its nearest neighbor in each of the six sextants of the circular area which surrounds the chosen "center of origin." The method is somewhat laborious, however, for it requires several measurements from each center of origin and makes use of third moment statistics in testing for the degree of departure from random expectation. One also experiences considerable difficulty in expressing the results of this measure in meaningful terms of spatial relations. Further study of the spacing problem has suggested the following measure, which requires but a single measurement from each center of origin and which we believe to be superior in its simplicity of computation and ease of interpretation. Use of the distance to nearest neighbor in the detection of non-randomness in spatial pattern has also recently been advocated by Skellam (1952), who has given a derivation of the probability distribution of this distance.

Lee R. Dice gave much advice and encouragement in the development of this measure. A letter from David G. Kendall, of Magdalen College, Oxford, offered a valuable suggestion for the derivation of a formula to express the mean expected distance between nearest neighbors in a random distribution of specified density. The measure of spacing described in this paper has been applied to the distributions of three species of grassland forbs charted by the junior author in collaboration with Stanley A. Cain and Fernando Segadas-Vianna, then associated with the Cranbrook Institute of Science, and also to a map of the trees of a woodlot prepared by Morrison Ismond, Bruce Hayward and Theodore Herman. Raymond H. Brand carried out many of the measurements and much of the statistical computation. To all of these sources of information and assistance we are very grateful. Both authors are jointly responsible for the ideas of this paper and have shared in the preparation of the manuscript. The senior author is primarily responsible for the mathematical derivations and statistical treatments.

The Measure

For the purposes of simplification, this measure of spacing is explained here in terms of two-dimensional space, *i.e.,* with reference to populations on plane surfaces. With suitable modification, however, it is equally applicable to populations distributed along a line or dispersed throughout a volume. A generalization of the measure for use with k dimensions has been worked out and will be published later.

The measure of spacing which we propose is a measure of the manner and degree to which the distribution of individuals in a population on a given area departs from that of a random distribution. Some clarification of what is meant by a "random distribution" is therefore desirable. In a random distribution of a set of points on a given area, it is assumed that any point has had the same chance of occurring on any sub-area as any other point, that any sub-area of specified size has had the same chance of receiving a point as any other sub-area of that size, and that the placement of each point has not been influenced by that of any other point. Thus, randomness as here employed is a spatial concept, intimately dependent upon the boundaries of the space chosen by the investigator. A set of points may be random with respect to a specified area but decidedly non-random with respect to a larger space which includes the specified area. For meaningful results, therefore, the areas selected for investigation should be chosen with care.

The distance from an individual to its nearest neighbor, irrespective of direction, provides the basis for this measure of spacing. A series of such distances is measured in a given population, using all of the individuals present or a randomly selected sample, and the value of the mean distance to nearest neighbor is obtained for this set of observations. The mean distance to nearest neighbor that would be expected if the individuals of that population were randomly distributed is also calculated. The ratio of the observed mean distance to the expected mean distance serves as the measure of departure from randomness. The ratios that have been calculated for two or more populations can be directly compared with one another, as a measure of their relative departure from random expectation. Significance tests of the difference between ratios can be made without great difficulty, and the results of comparison can be expressed in terms that are readily visualized. To facilitate presentation, the derivation of some of the formulas used in this paper are included in an Appendix rather than in the body of the text. A list of the symbols and definitions of concepts employed herein is given in Table I.

TABLE I. *A list of the symbols and definitions of concepts employed in a measure of spacing based on the mean distance between nearest neighbors*

N	the number of measurements of distance taken in the observed population or sample. When a single sector is employed, N is also equal to the number of individuals used as centers of measurement.
r	the distance in any specified units from a given individual to its nearest neighbor.
ρ	the density of the observed distribution expressed as the number of individuals per unit of area. (The unit of measurement used in the calculation of rho must be the same as that used in measuring r.)*
Σr	the summation of the measurements of distance to nearest neighbor.
Σr^2	the summation of the squares of the measurements of distance.
$\bar{r}_A = \frac{\Sigma r}{N}$	the mean of the series of distances to nearest neighbor.
$\bar{r}_E = \frac{1}{2\sqrt{\rho}}$	the mean distance to nearest neighbor expected in an infinitely large random distribution of density rho.
$R = \frac{\bar{r}_A}{\bar{r}_E}$	the measure of the degree to which the observed distribution departs from random expectation with respect to the distance to nearest neighbor.
$c = \frac{\bar{r}_A - \bar{r}_E}{\sigma_{\bar{r}_E}}$	the standard variate of the normal curve.
$\sigma_{\bar{r}_E} = \frac{0.26136}{\sqrt{N\rho}}$	the standard error of the mean distance to nearest neighbor in a randomly distributed population of density rho.
F	the ratio of between-group variance to within-group variance in a test of significance of the differences between two or more populations.
p	the number of populations being compared.
k	the number of sectors in a circle of infinite radius surrounding the individual from which measurements of distance are taken.

*Computation of rho on the basis of N - 1 rather than N is theoretically proper, but with large samples the difference in results is negligible.

If, in a population of N individuals having a specified density ρ, the distance r from each individual to its nearest neighbor is measured, the mean observed distance may be represented as $\bar{r}_A = \frac{\Sigma r}{N}$. The mean distance which would be expected if this population were distributed at random, $\bar{r}_E$, can be shown to have a value equal to $\frac{1}{2\sqrt{\rho}}$ (see Appendix). The ratio $R = \frac{\bar{r}_A}{\bar{r}_E}$ can then be used as a measure of the degree to which the observed distribution approaches or departs from random expectation. In a random distribution, $R = 1$. Under conditions of maximum aggregation, $R = 0$, since all of the individuals occupy the same locus and the distance to nearest neighbor is therefore 0. Under conditions of maximum spacing, individuals will be distributed in an even, hexagonal pattern, and every individual (except those at the periphery of the population) will be equidistant from six other individuals. In such a distribution, the mean distance to nearest neighbor will be maximized and will have the value $\frac{1.0746}{\sqrt{\rho}}$ (see Appendix). When this is the case, $R = 2.1491$. Thus, R has a limited range, with values indicative of perfectly uniform, random, and completely aggregated patterns of distribution. In any given distribution, the mean observed distance to nearest neighbor is R times as great as would be expected in a random distribution of the same density. Thus, an R value of 0.5 would indicate that nearest neighbors are, on the average, half as far apart as expected under conditions of randomness. This measure can therefore be readily interpreted in simple terms. Since it is also easily computed, it should be of practical use in describing spatial relations.

Tests of Significance

The usefulness of any measure of spacing will be increased if its reliability can be ascertained. If the value of R indicates that a given population is not randomly distributed, the significance of the departure of $\bar{r}_A$ from $\bar{r}_E$ can be tested by the normal curve. The formula used in this test of significance is

$$c = \frac{\bar{r}_A - \bar{r}_E}{\sigma_{\bar{r}_E}}.$$

where c is the standard variate of the normal curve (Mather 1947) and $\sigma_{\bar{r}_E}$ is the standard error of the mean distance to nearest neighbor in a randomly distributed population of the same density as that of the observed population. The value of $\sigma_{\bar{r}_E}$ for a population of density ρ is $\frac{0.26136}{\sqrt{N\rho}}$, where N is the number of measurements of distance made (see Appendix). The c values 1.96 and 2.58 represent respectively the 5 per cent and the 1 per cent levels of significance (for a two-tailed test); for other values one may consult a table of the normal distribution.[1]

When two populations are being compared, it may not be sufficient merely to ascertain whether each of them departs significantly from randomness. One may also want to know whether the populations differ significantly from one another with respect to the direction and magnitude of their departures from random expectation. The significance of the difference in the values of R for two populations can be tested by the Student-Fisher t distribution, or by the F distribution. We have employed the latter method because it can be used with more than two populations (see Appendix for analysis of variance).

The tests of significance proposed above are based on the difference between $\bar{r}_A$ and $\bar{r}_E$. It is theoretically possible, however, for a non-randomly distributed population to exist in which $\bar{r}_A$ and $\bar{r}_E$ are equal. The distribution of such a population cannot be shown to be non-random by these tests. In this case, departure from randomness may be ascertained by comparing the frequency distributions of observed and expected distances to nearest neighbor by means of a χ^2 test (for procedure, see Cochran 1952). The majority of investigations of natural populations, however, will not require so sensitive a test.

Application of the Measure to Actual Data

The measure described above has been tested experimentally by applying it to a synthetically constructed random distribution, to the distributions of three species of grassland plants whose patterns exhibit various degrees of aggregation, and to the distribution of trees participating in the canopy of a woodlot, where competition for light might be expected to bring about a comparatively uniform spacing. The results of these analyses are shown in Table II.

Table II. *Comparison of certain statistics obtained in the application of the measure of spacing to various distributions. Measurements for the synthetic random distribution are in units of millimeters and square millimeters, those for the other distributions in units of feet and square feet*

Statistic	Synthetic random	*Solidago*	*Liatris*	*Lespedeza*	Forest trees
Size of area	25,781	23,936	23,936	23,936	69,696
N	116	89	197	184	174
ρ	.00449944	.00371825	.00823028	.00768717	.00249656
$\sqrt{\rho}$	.0670779	.0609774	.0907209	.0876765	.0499656
Σr	833.12	530.24	601.57	512.60	1979.99
Σr^2	7909.6616	4751.5652	4587.6121	3385.5290	30382.6942
$\bar{r}_A$	7.1821	5.9578	3.0537	2.7859	11.3793
$\bar{r}_E$	7.4540	8.1998	5.5114	5.7028	10.0069
R	0.9635	0.7266	0.5541	0.4885	1.1371
$\sigma_{\bar{r}_E}$	0.3618	0.4543	0.2053	0.2198	0.3965
c	0.75	4.93	11.97	13.27	3.46
Probability of a greater difference between $\bar{r}_E$ and $\bar{r}_A$	.453254	.000002	.000002	.000002	.000540

A synthetic random distribution

An artificial distribution was constructed on cross-section paper of 20 lines per inch by random placement of 116 points on a rectangle of 203 x 116 mm. The distance to the nearest neighbor was measured for each point, and the mean observed distance to nearest neighbor proved to be 7.1821 mm. while the mean expected distance was calculated to be 7.4540 (Table II). For this synthetic distribution, $R = 0.96$, not far from the 1.00 expected of a perfectly random distribution. A test of the significance of the difference between observed and expected mean distances gave a c value of 0.75, indicating that greater departure from expectation might occur 45 per cent of the time purely by chance. The synthetic model was thus shown to be a satisfactory representation of random dispersion.

Distributions of grassland plants

The locations of all individuals of *Lespedeza capitata* Michx., *Liatris aspera* Michx., and *Solidago rigida* L. occurring on several acres of an abandoned field on the University of Michigan's Edwin S. George Reserve were recently mapped by Cain and Evans (1952). Inspection of the

[1] When N is small, somewhat greater accuracy in this test of significance may be obtained by use of the Pearson type III distribution than by use of the normal curve. It can be shown that the skewness of the distances to nearest neighbor in a randomly dispersed set of points on a plane surface is $\alpha_3 = .631$. The probability of a given difference between $\bar{r}_A$ and $\bar{r}_E$ should therefore be found, with the type III distribution, under $\alpha_3 = \frac{.631}{\sqrt{N}}$. When N is large, say over 100, the difference in results is negligible.

map and quantitative studies of the distributions by Thomson (1952) and Dice (1952) indicated that *Lespedeza* was the most strongly aggregated of the three species, whereas *Solidago* was the least aggregated. The measure of spacing described above has been applied to the distribution patterns of these plans on a portion of the map having field dimensions of 136 x 176 feet. The results of analysis are shown in Table II. For *Lespedeza,* R = 0.49, for *Liatris,* R = 0.55, and for *Solidago,* R = 0.73, thus confirming the relative degrees of aggregation shown by previous studies. None of these R values is close to 1, and in each case the probability of obtaining, in a random distribution of the same density, a c value as large as that observed is less than .000002. All three populations clearly depart from random expectation with a high degree of significance, so far as the distance to nearest neighbor is concerned. An analysis of variance was made for an over-all comparison of the three species and yielded an F value of 4.57, significant at the 5 per cent level, indicating differences in the degree of aggregation. Analysis of individual differences, by the methods of Tukey (1949), then showed that *Solidago* differed from both *Liatris* and *Lespedeza,* but the latter two populations were not shown to differ from each other.

A distribution of forest trees

The measure has also been used to evaluate the spacing between individual trees which formed the canopy of an oak-hickory woodlot on the North Campus of the University of Michigan. The distribution pattern of these trees was not readily discernible, although it was suspected of being more uniform than random. Analysis of the trees which occurred on a square plot of this woodlot measuring 264 feet on a side gave the results indicated in Table II. For this distribution, R = 1.14, indicating a departure from random expectation in the direction of uniformity. This difference was shown to be highly significant by the c test.

Application of the Measure to Large Populations

This measure of spacing requires that the true density of the population under investigation be known. Such knowledge permits the calculation of the exact values of the mean expected distance to nearest neighbor and of its standard error. If an estimate of mean density is substituted for its true value, the rigor of the proposed tests of significance will be lost. We recommend, therefore, that the measure be applied to populations which are small enough to permit ascertainment of the true density. This will involve a complete count or census of the population.

If it is considered impractical to measure the distance to nearest neighbor for all of the individuals in the area selected for study, an estimate of the mean distance can be obtained from a random sample and can be used in the formulas already presented. To obtain a truly random sample of n individuals, it is necessary that every possible combination of n individuals have an equal chance of being drawn. This requires that every individual be identified, *i.e.,* marked in some manner, and that n of these marked individuals be drawn at random. In some situations, this may be impractical and the investigator may wish to obtain his measurements of distance to nearest neighbor from randomly selected quadrats. Some rigor will be lost by this procedure, but the saving in time may be considerable. The method of selecting quadrats at random will differ according to circumstances. Time and labor will be saved if the locations of all quadrats to be examined are established prior to collection of data, for the examination of quadrats can then proceed according to location rather than order of draw. The measurements of distance to nearest neighbor obtained in the series of quadrats should be pooled to provide an estimate of $\bar{r}_A$. This estimate may be used in the estimate of R and in the tests of significance given above but is less satisfactory than an estimate based upon a truly random sample.

The problem of the size of sample needed for a satisfactory estimate is discussed in many statistical textbooks and additional space need hardly be devoted to it here. Briefly, however, the size of sample (*i.e.,* number of individuals) required depends upon the degree of accuracy with which the investigator will be satisfied and upon the variability in the distance to nearest neighbor. The more variable the distance, the larger the size of the sample needed to provide a specified degree of accuracy. In general, it may be said that the less uniform the spacing of the individuals in the population, the greater the variability in distance to nearest neighbor will be and the larger the sample required. In any given case, the number of quadrats needed to secure a sufficient sample will depend upon the size of quadrat used. If many small quadrats are used, the sample will more closely approximate a random sample as defined above, and will therefore be more satisfactory, than if a few large quadrats are employed.

Problems of Procedure

When measurements are being taken, one may find that the nearest neighbor of a given individual lies outside of the specified area. Such distances should be measured and included in the computations. However, no individual lying outside the specified area should be used as a center of measurement.

It may also happen that two individuals selected as centers of measurement are closer to one another than to any other individuals. In this case, the same distance will be measured twice. Such double distances introduce no bias and both measurements should be used in the calculations.

The derivation of the mean expected distance between nearest neighbors in a randomly distributed population of specified density is based on the assumption that the area occupied by the population is infinite. In practice, however, we apply the measure to finite populations occupying definite areas. The presence of a boundary beyond which measurements cannot be made will tend to make the value of $\bar{r}_A$ greater than would be obtained if an infinite area were involved. For this reason it will be desirable, whenever possible, to select an area for investigation that lies well within the total area covered by the entire population.

In the foregoing description of this measure of spacing, the individual components of a population have been treated as dimensionless points, whereas in reality they will occupy definite areas. In many situations, the area occupied by an individual is so small in relation to the total area under examination that the individual can legitimately be treated as a point. However, if the individual occupies an area of some size it can no longer be treated as a point, and its spatial requirements become important in determining its relationship to other individuals. In such cases, this measure of spacing is applicable only to the centers of individuals, and measurements of distance should be taken accordingly. It is possible for individuals to be as closely spaced as their size permits and at the same time to have their centers distributed more uniformly than random. Consequently, R values greater than unity may sometimes result from the spatial requirements of the individuals alone.

Extension of the Method

A description of distribution pattern solely in terms of the distance to nearest neighbor is not entirely complete, for it disregards all of the other spatial relations which exist in a population. The measure as described above may therefore not distinguish between certain types of distribution patterns. For example, a population whose individuals are congregated in one spot will not be distinguished from a population consisting of scattered pairs of individuals, such as conjugating *Paramecium*, for in each case the distance to nearest neighbor will be 0. In such situations, the measure can be extended to make use of additional space relations.

Consider each individual selected as a center of measurement to be surrounded by a circle of infinite radius, which can be divided into equal sectors. The distance from the individual at the center to the nearest individual in each of the sectors can then be measured. These distances become the basis of calculations. In this case, the formulas (1) for the expected mean distance to nearest neighbor in a randomly distributed population of density ρ and (2) for its standard error are respectively $\bar{r}_E = \frac{\sqrt{k}}{2\sqrt{\rho}}$ and $\sigma_{\bar{r}_E} = \frac{0.26136\sqrt{k}}{\sqrt{N\rho}}$ where N is the number of measurements and k is the number of sectors about each individual used as a center of measurement. The ratio of the observed mean distance to nearest neighbor to the mean distance expected if the population were randomly distributed will again indicate deviation from randomness whenever its value is greater or less than 1. For completely aggregated populations, its value is, of course, 0. The value indicative of perfectly even, hexagonal spacing depends upon k and is equal to $\frac{2.1491}{\sqrt{k}}$ when k is not greater than 6.

It should be pointed out that an increase in the number of sectors may not always result in greater knowledge about the population under investigation. For example, the *Liatris* and *Lespedeza* populations previously described were re-examined using 2, 3 and 6 sectors. In no case was a significant F value obtained, and the conclusions as to spacing were identical with those based on a single sector. The number of sectors to be used in any given situation, therefore, will depend upon the nature of the distribution patterns involved. Caution should be observed when increasing the number of sectors, for the assumptions underlying the derivations of $\bar{r}_E$ and $\sigma_{\bar{r}_E}$ preclude the existence of empty sectors. We believe that the use of a single sector, *i.e.,* the entire circular area surrounding an individual chosen as a center of measurement, will be sufficient for the description and comparison of spatial relations in most natural populations.

An alternative to the use of more than one sector might be developed by employing successively the second, third, etc. nearest neighbors. The formulas for such a procedure would, however, be more complex than those used in this paper and we have not attempted to explore this possibility.

Summary

To obtain a measure of the spacing of individuals in a population of known density, the distance, r, from each individual to its nearest neighbor may be measured and the mean value of r, here designated $\bar{r}_A$, computed. The mean value of r which would obtain in an infinitely large random distribution of density ρ is found by the formula $\bar{r}_E = \frac{1}{2\sqrt{\rho}}$. The ratio $R = \frac{\bar{r}_A}{\bar{r}_E}$ of observed to expected mean distance provides a measure of the degree to which the distribution pattern of the observed population deviates from random expectation. This ratio is less than, equal to, or greater than 1 according to whether the distribution pattern of the individuals in the population is more aggregated, the same as, or more uniform than would be expected in an infinitely large random distribution of the same density. R ranges in value from 0 for a distribution with maximum aggregation to 2.1491 for a distribution which is as evenly and widely spaced as possible. The significance of departure from random expectation on the part of a given population may be tested by the normal variate

$$c = \frac{\bar{r}_E - \bar{r}_A}{\sigma_{\bar{r}_E}}$$

where $\sigma_{\bar{r}_E}$, the standard error of $\bar{r}_E$, is $\frac{0.26136}{\sqrt{N\rho}}$. The significance of the differences between values of R from various populations may be tested by a one-way analysis of variance.

To demonstrate the simplicity and ease of interpretation of this measure, it has been applied to a synthetically constructed random distribution, to the distribution of three species of grassland plants whose patterns showed various degrees of aggregation, and to the distribution of trees in an oak-hickory woodlot where the pattern was suspected of being more uniform than random. For the synthetic distribution $R = 0.96$, a non-significant deviation from unity. The three grassland populations yielded the following R values: 0.73 for *Solidago,* 0.55 for *Liatris,* and 0.49 for *Lespedeza,* all of which are significant departures from randomness in the direction of aggregated spacing. For the trees of the woodlot, $R = 1.14$, a significant deviation from randomness in the direction of uniform spacing.

This measure requires a knowledge of the true density of the population. If this is known, it can be applied to populations in which it is considered impractical to measure the distance to nearest neighbor for all of the individuals. In such cases an estimate of the mean distance to nearest neighbor may be substituted for its true value. This estimate will give more satisfactory results if it is based upon a truly random sample of individuals but it can be obtained from randomly selected quadrats. The size of sample and the number and size of quadrats required are discussed briefly.

Since this measure involves only the relationship between a given individual and its nearest neighbor, the majority of spatial relations in the population are ignored. An extension of the method may be achieved by constructing a circle of infinite radius about each individual from which distances are to be measured, dividing this circle into equal sectors, and measuring the distance from the individual at the center to the nearest neighbor in each of the sectors. The occurrence of empty sectors should be avoided. It is believed that for most purposes a single sector, *i.e.,* the entire circle surrounding the individual chosen as a center of measurement, will be adequate.

An appendix to the paper gives the derivation of certain formulas employed in the development of the measure.

Appendix

Derivation of formulas for $\bar{r}_E$ and $\sigma_{\bar{r}_E}$

The formula for the mean distance to nearest neighbor expected in a randomly dispersed population of specified density seems first to have been derived by Hertz (1909). This paper is virtually unknown to biologists and was brought to our attention by Dr. F. D. Miller, Department of Astronomy, University of Michigan, after the senior author had independently obtained the derivations given below. These are presented here because of the inaccessibility of Hertz's paper and because they are considerably simpler than those given by Hertz.

For a random distribution of points in two dimensions the probability that a randomly chosen area of specified size will contain exactly x points is, by Poisson's exponential function, $\frac{m^x e^{-m}}{x!}$, where m is the mean number of points per area. Let the specified area be a sector of a circle of radius r, formed by dividing the circle into k equal sectors. If ρ is the mean density of the distribution, then $m = \rho k^{-1} r^2$ is the mean number of points per area. In this case

$$\frac{(\rho\pi k^{-1}r^2)^x\, e^{-\rho\pi k^{-1}r^2}}{x!}$$

is the probability of finding exactly x points in an arbitrary area of $\pi k^{-1}r^2$ units. Consequently, $e^{-\rho\pi k^{-1}r^2}$ is the probability that a randomly chosen area of $\pi k^{-1}r^2$ units will contain no points. If our area is a sector of a circle about a randomly chosen point, the probability that the sector will contain no other point within a distance r of the chosen point is also $e^{-\rho\pi k^{-1}r^2}$. Considered as a function of r, this is the proportion of distances to nearest neighbor (within sectors) $\geq$ r. Consequently $1 - e^{-\rho\pi k^{-1}r^2}$ is the proportion of distances to nearest neighbor $\leq$ r. Differentiating the last expression with respect to r we obtain $2\rho\pi k^{-1} r e^{-\rho\pi k^{-1}r^2} dr$ as the probability distribution of r. The mean of r, called $\bar{r}_E$, can be obtained by multiplying the above expression by r and integrating over the interval from 0 to ∞. Thus $\bar{r}_E = \int_0^\infty 2\rho\pi k^{-1} r^2 e^{-\rho\pi k^{-1}r^2} dr$, which can be shown to be $\frac{\sqrt{k}}{2\sqrt{\rho}}$.

The second moment of r, $E(r^2)$, is obtained by multiplying the probability distribution function of r by r^2 and integrating over the interval from 0 to ∞. Thus

$$E(r^2) = \int_0^\infty 2\rho\pi k^{-1} r^3 e^{-\rho\pi k^{-1}r^2} dr.$$

This integral has the value $\frac{k}{\rho\pi}$. The variance of r is $E(r^2) - (\bar{r}_E)^2$, which proves to be $\frac{(4-\pi)k}{4\pi\rho}$. The standard error of $\bar{r}_E$, $\sigma_{\bar{r}_E}$, is thus

$$\sqrt{\frac{(4-\pi)k}{4\pi\rho N}} = \frac{0.26136\sqrt{k}}{\sqrt{N\rho}},$$

where N is the number of measurements made.

The upper limit of $\frac{\bar{r}_A}{\bar{r}_E}$

The mean distance between nearest neighbors is maximized in a hexagonal distribution, where each point has 6 equidistant nearest neighbors. Let r_u denote the constant distance between nearest neighbors in this distribution. In such a uniform distribution each point can be shown to occupy an area of $\frac{r_u^2 3^{1/2}}{2}$. The density, ρ, of the population is thus $\rho = \frac{2}{r_u^2 3^{1/2}}$. Solving for r_u we obtain $r_u = \frac{2^{1/2}}{3^{1/4}\rho^{1/2}}$. Since r_u is the greatest possible value of $\bar{r}_A$, the maximum value of the ratio $R = \frac{\bar{r}_A}{\bar{r}_E}$ is $\frac{r_u}{\bar{r}_E}$, or $\frac{2^{3/2}}{3^{1/4}k^{1/2}}$, which is approximately equal to $\frac{2.1491}{\sqrt{k}}$.

The application of the analysis of variance to the ratio R

In applying the analysis of variance, we desire to test for differences in R rather than $\bar{r}_A$. Since $R = \frac{\bar{r}_A}{\bar{r}_E} = 2\sqrt{\rho}\ \bar{r}_A$ it is necessary to multiply each value of r by the square root of the density of the population to which it belongs. Such a transformation is most readily effected by multiplying each Σr by its corresponding $\sqrt{\rho}$ and each Σr^2 by its ρ. If p populations are being compared the following computational scheme may be used:

$$a = \rho_1\Sigma r_1^2 + \rho_2\Sigma r_2^2 + \ldots + \rho_p\Sigma r_p^2$$

$$b = \frac{(\sqrt{\rho_1}\Sigma r_1 + \sqrt{\rho_2}\Sigma r_2 + \ldots + \sqrt{\rho_p}\Sigma r_p)^2}{N_1 + N_2 + \ldots + N_p}$$

$$c = \frac{\rho_1(\Sigma r_1)^2}{N_1} + \frac{\rho_2(\Sigma r_2)^2}{N_2} + \ldots + \frac{\rho_p(\Sigma r_p)^2}{N_p}$$

where the subscripts 1, 2, . . . , p which follow ρ, r, and N represent the populations to which these data pertain. The variance ratio, F, for testing the significance of the difference in the values of R for the p populations is

$$F = \frac{(c-b)(N_1 + N_2 + \ldots + N_p - p)}{(a-c)(p-1)},$$

there being $p - 1$ and $N_1 + N_2 + \ldots + N_p - p$ degrees of freedom for the between-group and within-group variances respectively. If a significant F is obtained from this over-all analysis, we know that not all of the populations are alike in their degrees of clumping. The problem of ascertaining which populations differ from each other, however, is more complicated and will not be discussed here. Tukey (1949) has suggested ways of dealing with this problem.

References

Ashby, E. 1935. The quantitative analysis of vegetation. Ann. Bot. **49**: 779-802.

Cain, S. A., and F. C. Evans. 1952. The distribution patterns of three plant species in an old-field community in southeastern Michigan. Contrib. Lab. Vert. Biol. Univ. Mich. **52**: 1-11.

Cochran, W. D. 1952. The χ^2 test of goodness of fit. Ann. Math. Statist. **23**: 315-345.

Cottam, G., and J. T. Curtis. 1949. A method for making rapid surveys of woodlands by means of pairs of randomly selected trees. Ecology **30**: 101-104.

Curtis, J. T., and R. P. McIntosh. 1950. The interrelations of certain analytic and synthetic phytosociological characters. Ecology **31**: 434-455.

Dice, L. R. 1952. Measure of the spacing between individuals within a population. Contrib. Lab. Vert. Biol. Univ. Mich. **55**: 1-23.

Gleason, H. A. 1920. Some applications of the quadrat method. Bull. Torrey Bot. Club **47**: 21-33.

Goodall, D. W. 1952. Quantitative aspects of plant distribution. Biol. Rev. **27**: 194-245.

Hertz, P. 1909. Über den gegenseitigen durchschnittlichen Abstand von Punkten, die mit bekannter mittlerer Dichte im Raume angeordnet sind. Mathematische Annalen **67**: 387-398.

Mather, K. 1947. Statistical analysis in biology. Interscience Publishers, Inc., New York.

Neyman, J. 1939. On a new class of "contagious" distributions, applicable in entomology and bacteriology. Ann. Math. Statist. **10**: 35-37.

Skellam, J. G. 1952. Studies in statistical ecology. I. Spatial pattern. Biometrika **39**: 346-362.

Svedberg, T. 1922. Ett bidrag till de statistiska metodernas användning inom växtbiologien. Svensk Bot. Tidskr. **16**: 1-8.

Thomson, G. W. 1952. Measures of plant aggregation based on contagious distributions. Contrib. Lab. Vert. Biol. Univ. Mich. **53**: 1-16.

Tukey, J. W. 1949. Comparing individual means in the analysis of variance. Biometrics **5**: 99-114.

Viktorov, S. V. 1947. A study of the distribution and dispersion of plants by aerial photographs. (in Russian) Bull. Soc. Nat. Moscou (sect. biol.), N.S. **52** (4): 71-78.

ECOLOGY OF SPECIES OF *MEGACHILE* LATREILLE IN THE MIXED PRAIRIE REGION OF SOUTHERN ALBERTA WITH SPECIAL REFERENCE TO POLLINATION OF ALFALFA[1]

G. A. Hobbs and C. E. Lilly[2]
Field Crop Insect Section, Science Service Laboratory, Lethbridge, Alberta, Canada

Introduction

In the semi-arid regions of western North America, bees of the genera *Megachile* Latr., *Bombus* Latr., and *Nomia* Latr. are considered the most important native bee pollinators of alfalfa (Tysdal 1940). In Utah, Bohart (1951) and Carlson *et al.* (1950) consider *Nomia melandri* Ckll. the most important in areas where it is prevalent. Species of *Nomia* are not found in southern Alberta. Sladen (1918), Gray (1925), and Salt (1940) gave some credit to bumble bees as pollinators of alfalfa in southern Alberta.

In 1950 Hobbs (1950) observed only one bumble bee tripping alfalfa, this record having been made prior to August 5. In 1951, 11 queens and workers of *Bombus nevadensis* Cress., *B. rufocinctus* Cress., *B. borealis* Kby., *B. occidentalis* Greene, and *B. huntii* Greene were taken on alfalfa but none were recorded before August 5. In 1952, 21 members of these species and *B. fervidus* (F.) were captured on alfalfa before August 15. (Experiments on pollination of alfalfa by honey bees indicated that the seed produced by flowers tripped after August 5 in 1950 and 1951, and after August 15, 1952, was destroyed by frost before it had time to mature.) Hence, only a few bumble bees were seen to trip alfalfa flowers, and still fewer did this at a time when tripping would have resulted in the production of mature seed. In 1951 and 1952 many workers of the common species *B. nevadensis, B. rufocinctus, B. fervidus,* and *B. huntii* were found to be gathering pollen from sweet clover that was growing near alfalfa. It is realized that certain species of bumble bees may be an important factor in certain isolated, mixed prairie areas where they are abundant and/or where competing plants, especially sweet clover, are few.

Osmia texana Cress. and *Anthophora occidentalis* Cress. have been found nesting gregariously in large colonies in cutbanks near alfalfa seed fields. Neither species has been seen to gather pollen or nectar from alfalfa by the writers.

Alfalfa is grown for seed in two ecological areas in Western Canada. Most of the seed is now produced in that area, referred to by Coupland (1950) as aspen "parkland" ecotone, which lies between mixed prairie and boreal forest and is associated with fescue grassland. The mixed prairie (*Stipa-Bouteloua* association), which is also an important alfalfa seed-growing area, is clearly defined by Coupland (1950). The borders of both fescue grassland and aspen parkland cannot be as clearly defined because they are associated with each other to some extent. Aspen poplar, *Populus tremuloides* Michx., is the principal, and often the only, tree species in the groves, or "bluffs," of the aspen parkland. The grassland between the groves is characterized by fescue, *Festuca scabrella* Torr. (This description was furnished by Dr. W. G. Dore, Senior Botanist, Botany and Plant Pathology Division, Ottawa, Canada.)

Most of the alfalfa grown for seed is grown on cleared land formerly covered with deciduous

[1] Contribution No. 3067, Division of Entomology, Science Service, Department of Agriculture, Ottawa, Canada. Part of a thesis submitted by the senior author for the degree of Doctor of Philosophy, Oregon State College, 1952.

[2] Associate Entomologist and Assistant Entomologist.