

MULTINOMIAL EXPECTATION: AN ALTERNATIVE TO 2 X 2
PRESENCE-ABSENCE TESTS OF SPECIES ASSOCIATION

Douglas W. Morris*
Department of Zoology
University of Western Ontario
London, Ontario, N6A 5B7

Received June 2, 1982; October 12, 1982; April 21, 1983

ABSTRACT: Expected frequencies of single and mixed species occurrence can be estimated by multinomial expansion of species' frequencies and the fit of expected to observed values analyzed by single classification goodness of fit tests. A worked example on the co-distribution of two rodent species demonstrates the superiority of this technique over classical 2 X 2 presence-absence analyses when species occur in low numbers. This procedure will be most useful for association analyses of rare species and for the simultaneous analysis of joint and separate distribution in multi-species samples.

* * *

INTRODUCTION

Ecologists interested in patterns of species association often analyze the joint distribution of species pairs with 2 X 2 presence-absence tables (Pielou, 1977; Poole, 1974). In addition to the initial test of independence, these analyses can be extended to include various statistics of association. A serious shortcoming of 2 X 2 presence-absence data is that the association between rare species is often biased by an abundance of sampling points containing neither. These 2 X 2 tables (dominated by the absence-absence cell) will frequently indicate independence even if the constituent species interact. One solution is to increase the size of the sampling unit so that more samples include one or both species. This is impractical because the size of the sampling unit is usually determined a priori as a unit within which association is meaningful. More commonly, rare species are simply excluded from the analysis. If we could determine the expected frequencies of isolation and co-occurrence between these species based on the assumption of independence, we could still proceed with a suitable goodness of fit test of observed and expected distributions.

THE MODEL

If individuals within two species (A and B in proportions p_1 and p_2 respectively, $p_1 + p_2 = 1$) are independently distributed, and possess similar capabilities of movement, the expected proportions of individuals of each species occurring with conspecifics or with the other species are obtained by the binomial expansion. Separate and joint occurrence of two individuals per sample is given by $(p_1 + p_2)^2 = p_1^2 + 2p_1p_2 + p_2^2$; for three individuals by $(p_1 + p_2)^3 = p_1^3 + 3p_1^2p_2 + 3p_1p_2^2 + p_2^3$; and for n individuals as $(p_1 + p_2)^n$. For the degenerate case of one individual per sampling point, expected frequencies are simply the initial proportions of each species.

* * *

*Present address: Douglas W. Morris
Department of Biology
Memorial University of Newfoundland
St. John's, Newfoundland
A1B 3X9, Canada

Evolutionary Theory 6: 275-279 (August, 1983)

The editors thank C. Goodnight and two other referees for help in evaluating this paper.

© 1983, The University of Chicago; rights owned by author.

For three species, the expected frequencies of single and mixed species occurrence are given by collecting terms in the expanded trinomial $(p_1 + p_2 + p_3)^n$ where p_3 is the proportion of the total numbers of the three species contributed by species C. Thus the distribution of exactly two individuals per sample of three species is given by $(p_1 + p_2 + p_3)^2 = p_1^2 + p_2^2 + p_3^2 + 2p_1p_2 + 2p_1p_3 + 2p_2p_3$ (six classes - 2A, 2B, 2C, A+B, A+C, B+C) and for three individuals as $(p_1 + p_2 + p_3)^3 = p_1^3 + p_2^3 + p_3^3 + (3p_1^2p_2 + 3p_1p_2^2) + (3p_1^2p_3 + 3p_1p_3^2) + (3p_2^2p_3 + 3p_2p_3^2) + 6p_1p_2p_3$ (seven classes - 3A, 3B, 3C, 2A+B or A+2B, 2A+C or A+2C, 2B+C or B+2C, A+B+C). The general case of n individuals per sampling point of s species is obtained by expanding the multinomial $(p_1 + p_2 + \dots + p_s)^n$.

STRATEGIES OF ANALYSIS

There are two statistical strategies appropriate for comparing multinomial expectations with observed field data. One is simply to analyze each combination of individuals separately by an appropriate single classification goodness of fit test, or for small samples, exact tests such as those proposed by Ghent (1972). Thus, for two species which occur at sampling points as doublets and triplets, we would perform one analysis of the expected versus observed frequencies of joint and separate distribution for doublets, and another for triplets. For large samples one could use the normal approximation to the multinomial distribution and analyze the fit of expected and observed frequencies by the Kolmogorov-Smirnov test. The separate analysis approach has the advantage of allowing the investigator to analyze the orderings of the various co-occurrence classes. It may occasionally be important to distinguish between the triplet 2A+B versus A+2B for example. The disadvantage of this strategy is that the frequency of empty or small sample cells in the goodness of fit test increases rapidly with slight increases in n . This may frequently be a problem because the multinomial model will often be used to test the association of rare species.

An alternative strategy is appropriate for instances where the investigator is primarily concerned about joint or separate distribution, and is not interested in the relative orderings of species in groups of more than two individuals. It will also be valuable for the analysis of species sets where groupings of $n \geq 3$ are too few for independent analysis. This strategy groups the various possible classes of multiple occurrence for a particular species set by weighting each multinomial expansion of the series $n = 2, 3, 4, \dots, \infty$ by the number of N_n observations of groupings of size n . In this way, all of the data are made use of in the analysis, including infrequent but important records of multiple occurrence of $n > m$, where m is the smallest grouping of interest.

For example, in the two species case occurring with observed frequencies of N_2 doublets and N_3 triplets ($N_2 + N_3 = N$), expected proportions of the three possible classes (2A or 3A, 2B or 3B, A+B or 2A+B or A+2B) are given by

$$N_2/N((p_1 + p_2)^2) + N_3/N((p_1 + p_2)^3).$$

After collecting terms with respect to species composition classes and multiplying by N , the expected frequencies are

$$2A \text{ or } 3A: p_1^2N_2 + p_1^3N_3$$

$$2B \text{ or } 3B: p_2^2N_2 + p_2^3N_3$$

$$A \text{ and } B: 2(p_1p_2)N_2 + 3(p_1^2p_2)N_3 + 3(p_1p_2^2)N_3.$$

The degrees of freedom for all goodness of fit tests are $M - 1 - (s-1)$ where M is the total number of terms in the expanded multinomial (strategy 1), or the total number of combinations of species taken from one through s , i.e.,

ALTERNATE TESTS OF SPECIES ASSOCIATION

$\binom{s}{1} + \binom{s}{2} + \dots + \binom{s}{s} = s!/(s-1)! + s!/(s-2)! + \dots + s!/s! = s + s!/(s-2)! + \dots + 1$ (strategy 2). One degree of freedom is lost because the total sample is fixed; the remainder are lost because $s - 1$ parameters (species' proportions) are generated from an intrinsic hypothesis in the same data set as the observed outcomes (Sokal and Rohlf, 1981). In cases where expected relative abundances are generated from extrinsic hypotheses such as independent estimates of relative densities, the degrees of freedom equal $M - 1$.

One drawback of this test procedure is that it requires both presence-absence data and relative abundance estimates of the species being studied. This inconvenience should not be serious because the test is oriented toward association of rare species where standard presence-absence analyses are inappropriate because of the high frequency of absence-absence cells.

The method assumes that the observed frequencies of groupings of size n reflect the actual underlying distribution of those groupings. Where the sampling regime leaves some doubt about this assumption, then the appropriate degrees of freedom for the analysis becomes $M-1-(s-1)-(k-1)$ where k is the number of different sized groupings. One degree of freedom is lost for each intrinsic estimate of the frequency of groupings of $n > m$. Where previous work on the spatial distribution of individuals has identified the theoretical distribution of groupings (eg. Poisson, Neyman type A, negative binomial, see for example, Pielou (1977)) then the expected frequencies of groupings can be used in place of the observed groupings with no loss of additional degrees of freedom.

A FIELD STUDY

The following worked example demonstrates that 2 X 2 presence-absence analyses may lead to mis-interpretations of species distribution when those species are rare. The co-distribution of two rodent species, Clethrionomys gapperi (red-backed vole) and Peromyscus maniculatus (deer mouse) was investigated in several habitats in the Rocky Mountains of Alberta (Morris, In Press). These animals have similar sized home ranges and in western Alberta, both prefer forested habitats. Each sampling plot consisted of 49 trap stations at which small mammals were live-trapped, individually marked and released during three sampling sessions in the summer of 1977. Co-distribution of Clethrionomys and Peromyscus at individual trap sites was analyzed by 2 X 2 presence-absence tables. Only three out of twelve plots met the selection criterion of no expected frequencies less than 1.0, because both Clethrionomys and Peromyscus numbers were low. In all goodness of fit analyses of these data, either for the summer as a whole, or for individual 36 hr trap sessions, Clethrionomys and Peromyscus were always independently distributed. In all cases however, the values of the absence-absence cells were two or more times greater than the value of any other cell in the same 2 X 2 table. Clearly low populations and the absence-absence cells dominated the analysis and could have concealed significant patterns of species association.

To overcome these problems, the minimum number of each species known alive was used to determine relative proportions of the two species in forested habitats (Table 1). The binomial expansion of these proportions estimated frequencies of single and mixed species capture points (Table 2). Observed values are actual records of separate and joint capture at individual trap stations within each 36 hr capture period. The fit of expected to observed frequencies by the log likelihood ratio test was marginally significant ($G = 2.72$; $0.05 < p < 0.10$, Table 3). This changed the interpretation of

Clethrionomys-Peromyscus distribution from one of species independence to one indicating moderate interaction between individuals.

* * *

Table 1. Density estimates of two rodent species in the Rocky Mountains of Alberta. Animals known alive were summed across plots in which at least two different individuals of each species were known to have been active during the same 36 hr sampling period.

| SPECIES | NUMBER KNOWN ALIVE | PROPORTION |
|----------------------|--------------------|------------|
| <u>Clethrionomys</u> | 45 | 0.61 |
| <u>Peromyscus</u> | 29 | 0.39 |

* * *

Table 2. Expected frequencies of joint and separate occurrence of two rodent species obtained by binomial expansion of the observed proportions in Table 1, and weighted by observed values of 26 doublets and 2 triplets.

| OCCURRENCE CLASS | ACCUMULATED FREQUENCIES |
|---|--|
| <u>Clethrionomys</u> alone | $(.61)^2(26) + (.61)^3(2) = 10.1$ |
| <u>Peromyscus</u> alone | $(.39)^2(26) + (.39)^3(2) = 4.1$ |
| <u>Clethrionomys</u> + <u>Peromyscus</u> | $2((.61)(.39))(26) + 3((.61)^2(.39))(2) + 3((.61)(.39)^2)(2) = 13.8$ |

* * *

Table 3. Expected versus observed frequencies of separate or joint capture of two rodent species in the Rocky Mountains of Alberta ($G = 2.72$ with Yates correction for continuity; $0.05 < p < 0.10$) (data from Morris (In Press)).

| SPECIES COMPARISON | EXPECTED FREQUENCY | OBSERVED FREQUENCY |
|------------------------------------|--------------------|--------------------|
| <u>Clethrionomys-Clethrionomys</u> | 10.1 | 11 |
| <u>Peromyscus-Peromyscus</u> | 4.1 | 7 |
| <u>Clethrionomys-Peromyscus</u> | 13.8 | 10 |
| Total | 28 | 28 |

* * *

THREE CAVEATS

A few additional points about multinomial expectation should be stressed. First, this test makes all of the assumptions common to any contingency analysis. Second, it is susceptible to bias when comparing organisms of unequal mobility because only relative densities are used to estimate expected frequencies of conspecific and mixed species occurrence. This can be corrected through the weighting of expected frequencies by the ratios of encounter radii. Third, the assumption of independence of individuals within species can lead to ambiguities. Contagion within species is always reflected by greater than expected frequencies of conspecific co-occurrence, but the same pattern could also be caused by negative association between species. Similarly, observed

contagion of mixed species samples could imply either positive association between species, or consistent intraspecific spacing.

ACKNOWLEDGEMENTS

I thank C. Goodnight, Roy Knoechel, John S. Millar, Leigh M. Van Valen and anonymous reviewers for several helpful suggestions which improved this paper. This report was conceived while the author held a Natural Sciences and Engineering Research Council of Canada, Postdoctoral Fellowship.

LITERATURE CITED

- Ghent, A.W. 1972. A method for exact testing of 2X2, 2X3, 3X3, and other contingency tables, employing binomial coefficients. *Amer. Midl. Nat.* 88: 15-27.
- Morris, D.W. In press. Field tests of competitive interference for space among temperate zone rodents. *Can. J. Zool.*
- Pielou, E.C. 1977. *Mathematical ecology*. John Wiley and Sons, New York. 385 p.
- Poole, R.W. 1974. *An introduction to quantitative ecology*. McGraw-Hill, New York. 532 p.
- Sokal, R.R., and F.J. Rohlf. 1981. *Biometry* (2nd edition). W.H. Freeman and Company, San Francisco. 859 p.

