

TESTING WHETHER ISLAND COMMUNITIES ARE RANDOM

By Richard F. Green

Technical Report No. 63

Department of Statistics
University of California
Riverside, California 92521

December, 1979

TESTING WHETHER ISLAND COMMUNITIES ARE RANDOM

by

Richard F. Green
Department of Statistics
University of California
Riverside, CA 92521

Strong, Szyska and Simberloff (1979) recently tested for community-wide character displacement by comparing morphological differences between members of island bird communities with average differences found in simulated communities constructed at random using mainland source species. Their tests do not yield any evidence of character displacement, in contrast to the frequent claims of character displacement and resource partitioning found in the recent ecological literature. Strong et al. suggest that this difference may be due to the fact that few authors have compared observed patterns to those generated under appropriate null hypotheses (but see Pleasants 1977 and Poole and Rathcke 1979).

The general idea of testing rigorously whether patterns in real communities are consistent with a priori null hypotheses, especially null hypotheses of randomness, is an important one. However, the tests proposed by Strong et al. do not serve their stated purpose. That the tests are not able to reject the null hypothesis and detect character displacement simply may be due to the fact that they are virtually powerless to detect character displacement if it does exist. Other, more powerful tests are easy to find and one such test is proposed in this note.

Ecological theory suggests that competing species should not overlap too much in resource use, either because of extinction of species too similar to others (MacArthur 1972), or because of selection to avoid competition and consequent divergence of species in resource use (Roughgarden

1976). Schoener (1974) has pointed out that competition might be expected to produce regularly spaced niches. He says (p. 28):

Although if species had no influence on each other's resource utilization their niches would still differ, competition should result in an overdispersion of niches in niche-space. Where niches are regularly and widely spaced over one or more dimensions, the alternative or "null" hypothesis of randomly generated differences must be rejected.

Unfortunately, Schoener does not explicitly describe or perform any tests of such "null" hypotheses. Strong et al. (1979) are attempting to remedy this situation.

Niche dimensions are difficult to measure directly and Strong et al. use size of birds itself as an indicator of resource use. Culmen length is taken as an indicator of food size use and wing length as an indicator of overall body size. Birds in a given island fauna are grouped into families and within each family are ranked in order of culmen and wing length. Then ratios of culmen length and wing length are calculated for each contiguous pair of ordered values. These values are compared with ratios obtained in the same way from simulated island communities, or "null faunas".

Null faunas are found by choosing a sample of species from the source fauna at random and treating these species as if they were the actual island fauna. For example, if the island fauna contains 10 species from a certain family and the mainland source fauna contains 20 species from that family which could conceivably be found on the island, then 10 species are drawn at random from the 20 source species. The species drawn are ordered according to culmen length and wing length and the ratios of lengths are calculated for contiguous pairs. The simulation is repeated 100 times and the average ratio calculated for culmen length and wing

length for each family. The ratios found in the actual fauna are compared with the average of those found in the null faunas. A test of the null hypothesis is described by Strong et al. (p. 898):

The ratios of culmen and wing length between adjacent species in rankings by size are termed contiguous ratios. The first specific null hypothesis is that contiguous ratios among island species are generally no different from those in null faunas. There are two alternative possibilities, that island ratios are either larger or smaller than null ratios. Character displacement is consistent only with larger ratios.

The tests proposed by Strong et al. are inappropriate for several reasons. I will criticize only the test based on the binomial distribution mentioned above, but similar criticisms may be made against the other tests as well.

Strong et al. propose comparing the observed ratios with the average ratio calculated from the simulations. They assume that under the null hypothesis the number of actual ratios that are larger than the average value found from the simulations will have a binomial distribution with n equal to the number of ratios considered and $p \doteq .5$. This assumption is wrong on two counts.

1) The probability, p , of an actual ratio exceeding the mean of the simulated ratios is not .5 under the null hypothesis because the distribution of ratios is skewed. For the extreme case where a large number of species in a family are drawn from a still larger source fauna whose culmen or wing lengths are roughly uniformly distributed, the ratios will be approximately exponentially distributed and the probability that a given ratio exceeds the mean will be approximately $e^{-1} \doteq .368$. If this were the case then for 30 ratios the expected number of times an observed ratio would exceed the mean is about 11 and not 15. For birds

of the Tres Marias Islands off Mexico, 18 of 30 observed ratios of culmen lengths exceed the mean. This is not significant if the expected number of large ratios is 15, but it is significant if the expected number is 11. It is unlikely that the distribution of ratios would be as skewed as the exponential distribution, so the true probability that a given ratio will exceed the mean under the null hypothesis is probably between .368 and .5. If the median were used the probability that a given observed ratio exceeds the median would be about .5 under the null hypothesis.

2) To use the binomial distribution the ratios should be independent. This is certainly not the case; if one ratio is large this will make it more likely that other ratios are small since the total ranges of culmen and wing length are fixed for assemblages of species either on islands or in source pools. Because of this negative correlation between the different observed ratios the variance of the number of ratios that exceed the mean will be smaller than it would be in the binomial case where the ratios would have to be independent.

These criticisms, however, are merely technical. The most serious mistake that Strong et al. make is that they use a test which is virtually powerless. They test whether observed ratios exceed the mean found from null faunas. If community-wide character displacement exists one would expect a regular spacing of sizes and the contiguous ratios should not be either large or small but should be relatively constant. One would expect some ratios above the mean and some below the mean whether character displacement tends to spread out the island community or not.

What should be done is to use a test which is sensitive to the variance in the contiguous ratios rather than the mean. Such a test should be based

on the sample variance. The method suggested here is to perform a Monte Carlo test using the sample variance of the contiguous ratios as the test statistic.

To perform a Monte Carlo test one chooses a test statistic, such as the variance in this case, and calculates the value of that statistic using the observed data. Then a number of simulated samples, say 99 of them, are produced and the same test statistic is calculated for each of the simulated samples. In this case 99 "null faunas" are chosen and the variance of the ratios calculated for each set. Thus, the test statistic is calculated a total of 100 times and if the null hypothesis is true then the test statistic for the observed island fauna is equally likely to be any of the 100 values. In this case we are interested in low values of the variance. If the variance for the observed ratios is one of the five smallest values from the 100 values then we would reject the null hypothesis that the island communities are random at the 5% level. In general, if the observed variance is the k^{th} smallest value the p-value will be $k\%$.

If tests are based on too few data they may not be significant even if the effect they are designed to detect is present. However, if a number of independent tests are performed on small amounts of data, for example when different families of birds are considered separately, the results may be combined using the method suggested by Fisher (see Sokal and Rohlf 1969 pp. 621-624). The significance levels (p-values) are calculated for each test and combined in the summary statistic

$$-2 \sum_{i=1}^n \ln p_i$$

which will have a Chi-squared distribution with $2n$ degrees of freedom if the null hypothesis is true. Here n is the number of tests performed and $\ln p_i$ is the natural logarithm of the significance level for the i^{th} test.

The idea of the Monte Carlo test was suggested by Barnard during the discussion of a paper by Bartlett (1963) read before the Royal Statistical Society. Monte Carlo tests are conceptually simple and are "exact" in the sense that the probability of falsely rejecting the null hypothesis is exactly equal to the nominal value, α . Monte Carlo tests are useful in cases where it is difficult or impossible to calculate the distribution of the test statistic. With Monte Carlo tests it is not necessary to find the distribution of the test statistic, either by exact calculation or by simulation. All that is necessary is to be able to calculate the test statistic, such as the sample variance, from the data, and to be able to simulate samples.

Monte Carlo tests are performed by comparing the observed test statistic with simulated values rather than with a theoretical distribution as is done in classical statistics. Hope (1968) has shown that if the test statistic used to perform a Monte Carlo test is uniformly most powerful when the sampling distribution is known then the Monte Carlo test, which does not require the sampling distribution to be known, will be almost as powerful. Of course a Monte Carlo test will not be powerful unless the test statistic used is sensitive to the difference between the null hypothesis and the alternative.

A test almost identical to that proposed here was proposed for a similar problem by Poole and Rathcke (1979). They were interested in determining whether the flowering times given by Stiles (1977) for 11

hummingbird-pollinated plants are more regular than they would be if they were arranged at random throughout the growing season. If the flowers are considered to compete for the pollinators and if they tend to flower at different times, then the median flowering times for the different species might be regularly spaced over the growing season.

Poole and Rathcke consider the sample variance of the differences between the median flowering times for the plants. This variance should be small if the median flowering times are regularly spaced. In fact, for each of the four years of Stiles' study, the observed values of the sample variance are not smaller but Poole and Rathcke claim they are significantly larger than one would expect if median flowering times were chosen at random.

The difference between Poole and Rathcke's procedure and that proposed here is that their null hypothesis is equivalent to having points distributed at random on a line segment and the theoretical distribution of the interval between successive points is known. Poole and Rathcke do not have to use simulation because the distribution of the sample variance in their case does not depend on the particular values in some source population.

Poole and Rathcke standardize the sample variance and test the null hypothesis assuming that their test statistic has a Chi-squared distribution. One trouble with their procedure is that their test statistic does not actually have a Chi-squared distribution and its true distribution looks less and less like a Chi-squared as the sample size increases. The probability of falsely rejecting the null hypothesis is greater than they think it is.

Stiles (1979) objects that the null hypothesis of random flowering times considered by Poole and Rathcke is biologically unreasonable because the season most suitable for flowering is divided into two separate parts. This may well be so. However, if one expects to find flowering times clumped because of separate growing seasons it is difficult to see how the distribution of flowering times could be used to test the hypothesis that competition for pollinators tends to make the intervals between flowering times regular.

In testing biological hypotheses it is important that different observations be predicted by the different hypotheses. If different hypotheses predict the same observations then the observations will not help distinguish between the hypotheses. One must be able to determine what observations are expected under each biological hypothesis. Then one must find a good procedure to test the statistical hypotheses about the observations.

Strong et al. (1979) have proposed statistical hypotheses which may be tested to detect whether community-wide character displacement exists. They have proposed and applied a test of the null hypothesis that island communities are constructed at random from mainland source faunas. The hypothesis tested is reasonable but the test used is inappropriate. The fact that Strong et al. fail to detect any effect of character displacement may be due to the fact that the test they use is virtually powerless to detect such an effect if it does exist.

I would like to thank Nick Waser and Mary Price for reading and commenting on the manuscript. This paper is Technical Report No. 63 of the Department of Statistics, University of California, Riverside.

LITERATURE CITED

- BARTLETT, M. S. 1963. The spectral analysis of point processes (with Discussion). *J. R. Statist. Soc. B.* 25:264-296.
- HOPE, A. C. A. 1968. A simplified Monte Carlo significance testing procedure. *J. R. Statist. Soc. B.* 30:582-598.
- MACARTHUR, R. H. 1972. *Geographical Ecology*. Harper & Row, New York.
- PLEASANTS, J. M. 1977. Competition in plant-pollinator systems: an analysis of meadow communities in the Colorado Rocky Mountains. Ph.D. dissertation. University of California, Los Angeles.
- POOLE, R. W., and B. J. RATHCKE. 1979. Regularity, randomness, and aggregation in flowering phenologies. *Science*. 203:470-471.
- ROUGHGARDEN, J. 1976. Resource partitioning among competing species -- a coevolutionary approach. *Theor. Popul. Biol.* 9:388-424.
- SCHOENER, T. W. 1974. Resource partitioning in ecological communities. *Science*. 185:27-39.
- SOKAL, R. R. and F. J. ROHLF. 1969. *Biometry*. Freeman, San Francisco.
- STILES, F. G. 1977. Coadapted competitors: the flowering seasons of hummingbird-pollinated plants in a tropical forest. *Science*. 198:1177-1178.
- STILES, F. G. 1979. Reply to Poole and Rathcke. *Science*. 203:471.
- STRONG, D. R., L. A. SZYSKA, and D. S. SIMBERLOFF. 1979. Tests of community-wide character displacement against null hypotheses. *Evolution*. 33:897-913.