Similarity Indices II: The Power of Goodall's Significance Test for the Simple Matching Coefficient

by
J. W. Johnston

December 1976

**Prepared for The Nuclear
Regulatory Commission**

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Nuclear Regulatory Commission, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

PACIFIC NORTHWEST LABORATORY
operated by
BATTELLE
for the
ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION
Under Contract EY-76-C-06-7830

Printed in the United States of America
Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, Virginia 22151
Price: Printed Copy \$____*; Microfiche \$3.00

Pages	NTIS Selling Price
001-025	\$4.50
026-050	\$5.00
051-075	\$5.50
076-100	\$6.00
101-125	\$6.50
126-150	\$7.00
151-175	\$7.75
176-200	\$8.50
201-225	\$8.75
226-250	\$9.00
251-275	\$10.00
276-300	\$10.25

3 3679 00062 7002

**Similarity Indices II:
The Power of Goodall's
Significance Test for the
Simple Matching Coefficient**

**by
J. W. Johnston**

December 1976

**Prepared for The Nuclear
Regulatory Commission**

**Battelle
Pacific Northwest Laboratories
Richland, Washington 99352**

SUMMARY

The Simple Matching Coefficient, K_{SM} , of Sokal and Michener (1958) is representative of a class of coefficients used to indicate the degree of similarity between two multivariate samples from ecological communities. Goodall (1967) proposed a statistical decision rule based on K_{SM} for deciding whether or not the degree of similarity was small enough to allow the conclusion that the two samples come from different populations. This paper shows that the statistical power of Goodall's (1967) test was only 0.42 when a "95 percent confidence level" test and a "reasonable" alternative was used. That is, when a 0.05 chance was allowed (providing 95 percent confidence) for making the error of concluding that the two communities were different, when in fact they were the same, then there was less than a 50-50 chance (0.42-0.58) of concluding that the two samples were different when the samples were constructed to be different. No general power function for Goodall's test was developed, the power calculation for two specific alternatives being considered adequate to demonstrate the poor behavior of K_{SM} as a test statistic.

CONTENTS

SUMMARY	1
LIST OF FIGURES AND TABLES	i v
INTRODUCTION	1
SIGNIFICANCE TESTS	3
GOODALL'S SIGNIFICANCE TEST	7
EXAMPLE OF APPLICATION OF THE DECISION RULE	13
THE POWER OF THE DECISION RULE	17
CONCLUSION	35
REFERENCES	37

LIST OF FIGURES

1	Cumulative Distributions of m for $N = 8$ under H_0 : No Difference	14
2	Cumulative Distributions of m for $N = 8$ under H_0 and H_A	25
3	Normal Approximations to the Frequency Distribution of m under H_0 and H_A	27
4	Cumulative Distributions of m for $N = 8$ under H_0 and $H_{\max}$	31

LIST OF TABLES

1	Comparison of the Distribution of m under H_0 and H_A	24
2	The Distribution of m for Samples from Populations A and C (under $H_{\max}$)	30

SIMILARITY INDICES II:
THE POWER OF GOODALL'S SIGNIFICANCE
TEST FOR THE SIMPLE MATCHING COEFFICIENT

INTRODUCTION

A previous report by Johnston (1976) investigated the algebraic meaning of 25 similarity indices. One of these was the Simple Matching Coefficient introduced to taxonomic studies by Sokal and Michener (1958). Goodall (1967) attempted to determine the distribution of this coefficient under the null hypothesis that two sample vectors of N binary coded attributes come from the same statistical population. In the fifth section of his 1967 paper, Goodall suggests significance tests for the above null hypothesis. Despite some reservations about the distributional theory Goodall uses to develop his test, we will accept his suggested test as indicative of the state of the art of testing similarity coefficients and calculate the power of his test against a specific alternative. The goal of this paper is to demonstrate that the power of Goodall's test is so poor that it should not be used in studies attempting to discern a change in an ecological population. This will be accomplished by using two specific and reasonable alternatives to α -level tests of the null hypothesis. Finally, it is pointed out that improvements in discriminating between two ecological communities will require basic statistical research using statistical decision theory.

SIGNIFICANCE TESTS

A statistical significance test is an objective method for making a decision. The decision of concern here is whether or not a change has occurred in an ecological community, or whether or not two ecological communities are different. The basic data for making such a decision are two samples, one from each community. A statistic is calculated based on these samples and compared with the expected value of the statistic under certain theoretical assumptions. If the statistic is far from the expected value under the hypothesis that the ecological communities are the same, the hypothesis is rejected and the conclusion made that the communities are statistically different. This basic decision theory for significance tests implies the following adjuncts of the data in making the decision.

1. A sample space specifying the set of possible outcomes of a sampling operation, $S = \{x: x \text{ is a possible outcome}\}$.
2. A class of statistical distribution functions, $P = \{P_\theta: \theta \in \Omega\}$ depending on a parameter, θ , which may be a vector and which assigns probabilities to the outcomes in 1. Ω is the parameter space.
3. A decision space, $D = \{d: d \in \delta(x)\}$ which assigns a decision, d , based on the decision rule $\delta(x)$, to each outcome, x . (Frequently, the decision rule is mediated by a statistic which is a function of x and which estimates the parameter θ .)
4. Usually, a "Risk Function" (see E. L. Lehmann, 1959, Cpt. 1), which quantifies the overall risk in using the decision rule δ when θ is the true parameter, is also specified. To simplify matters, we will simply assign a risk value of zero to making the correct decision and a value of unity to making an error of either kind. (See Lehmann, 1959, p. 63.)

For our discrimination problem, the set of possible outcomes consists of the possible binary vectors of attribute scores, that is, presence-absence scores (1 or 0) for each of the N species which could be sampled from a space-time community. The appropriate family of distribution functions has received little attention (except for Goodall, 1967; Preston, 1948, and Fisher, et al., 1943) with regard to species presence-absence data. The class of distribution functions for our problem consists of the multivariate binomial distributions specified by the vector of probabilities.

$$F = (f_{11}, f_{12}, \dots, f_{1N})$$

where f_{1k} is the probability of observing the k^{th} attribute (species) in a random sample from the community (f_{1k} is between zero and unity).

The decision space contains two points: the two communities are different or they are not different.* The decision maker can make two kinds of errors as indicated in the following decision table.

<u>Decision</u>	<u>State of Environment</u>	
	<u>Not Different</u>	<u>Different</u>
Not different	No error (1- α) = Confidence	Type II error β
Different	Type I error α	No error (1- β) = power

It appears to be fairly common practice to use a decision rule based on allowing a 5 or 10 percent (100α) chance of making an error of Type I, that is, concluding that the communities are different based on a statistical test when in fact they are not different. The decision maker can be $100(1-\alpha)\%$ confident that his decision is correct. This level of confidence

*Note that only two alternatives are allowed. The noncommittal assertion that the data were not adequate to detect a difference is not a possible conclusion, i.e., it is not included in the "decision space."

in a correct decision pertains only to the case where the true, but unknown, state is one of no difference. The "a-level" decision rule described above has a calculatable "power" depending on the extent of the difference to be detected, the variability of the statistics compared and the statistical distribution appropriate for the data and statistics used. This "power" gives the odds of correctly concluding that there is a difference when there truly is and is the complement of the odds of making the Type II error of concluding there is no difference when in fact there is. The chance of failing to detect a true difference (and by inferential extrapolation, an impact) should be an important consideration in specifying a decision rule. **It** forces quantification in place of the noncommittal alternative excluded in the last footnote by assigning probabilities to both kinds of errors. For example, a decision maker using a "90% confidence" decision rule would appear much less magnanimous **if it** were realized that this rule could only detect a difference in half the cases of true differences of a reasonable magnitude. Such a situation would be analogous to an even money bet in which a potentially polluting operator had 9 to 10 odds of winning, but the "environmentalists" only 1 to 2 odds.*

*Before such hypothetical odds for the parties involved can be quoted, other than "out of context," estimates of power in actual studies are needed.

GOODALL'S SIGNIFICANCE TEST

Goodall proposed the Simple Matching Coefficient as a statistic which could be used in this discrimination problem. The basic assumption used by Goodall (1967) to get started on a distributional theory for the simple matching coefficient, K_{SM} , is that each of the N attributes has a probability, f_{1k} , of being observed and $(1-f_{1k})$ of not being observed, that is, each of the N attributes has a binomial distribution with parameter f_{1k} . Generally f_{1k} for attribute k is not the same as $f_{1k'}$ for another attribute k' . Of course, the specific values for the f_{1k} are unknown. Goodall advised estimating the f_{1k} , for $k = 1, \dots, N$ from the entire set of L samples in the study by

$$\hat{f}_{1k} = \sum_{\ell=1}^L x_{\ell k} / L$$

where $x_{\ell k}$ is 0 or 1. Then the estimated probability of a match on attribute k is

$$\hat{p}_k = \hat{f}_{1k}^2 + (1-\hat{f}_{1k})^2.$$

Returning to the unknown population parameters, we can write

$$p_k = f_{1k}^2 + (1-f_{1k})^2$$

as the probability of a match, defined as the outcome (0,0) or (1,1) for attribute k . Again, the binomial (Bernoulli trials) model is appropriate with parameters p_k for a "success", i.e., a match, and $(1-p_k)$ for no match. This defines a match vector of 1 and 0's of dimension N . If m is the number of matches (1's in the match vector), the probability of observing m matches is the product of the individual attribute match probabilities, under the assumption of independence of attributes, summed over all possible match vectors with m one's. The number of terms in this summation

is the number of ways m one's and $(N-m)$ 0's can be arranged in the N dimensional match vector, which is the combinatorial

$$\binom{N}{m} = \frac{N!}{m!(N-m)!} = C$$

Algebraically, the exact probability of m matches is

$$P_m = \sum_{v=1}^C \left[\prod_{k=1 \text{ in } v} (p_k) \right] \left[\prod_{k=0 \text{ in } v} (1-p_k) \right] ; m = 0, 1, \dots, N \quad (1)$$

where the notation $\prod_{k=1 \text{ in } v}$ specifies the operation of multiplying together all p_k for which a match occurs on the k^{th} attribute in the match vector v with a similar definition for the product of the no-match probabilities.

Equation 1 satisfies the requirement of specifying a class of distribution functions defined on the outcome space.* For any possible pair of attribute vectors we can determine the match vector, and from this, the value of the simple matching coefficient

$$K_{SM} = m/N$$

The parameter space is defined by the vector

$$\theta = (p_1, p_2, \dots, p_N)$$

for arbitrary p_k .

Goodall suggests using "as the first approximation to a significance test" the decision rule (in our notation)

$$\delta(K_{SM}) = \begin{cases} \text{Not Different, if } m/N \geq Bi(\alpha) \\ \text{Different, if } m/N < Bi(\alpha) \end{cases}$$

*Although this is not the basic outcome space consisting of binary, N -dimensional vectors, it is generated by the cross product of the basic space into itself which gives the same space again, but with different probabilities.

where $Bi(\alpha)$ is the α -fractile of the binomial distribution with parameter μ and

$$\hat{\mu} = \sum_{k=1}^N \hat{p}_k / N$$

He points out that the only case in which this binomial approximation is exact is when the p_k are all the same, say p_0 , then

$$p_m = \binom{N}{m} p_0^m (1-p_0)^{N-m}, \quad (2)$$

the probability density for m successes in N binomial trials, for which the mean and variance of m are given by

$$E(m) = Np_0$$

and

$$V'(m) = Np_0(1-p_0) = Np_0 - Np_0^2.$$

For m distributed according to Eq. 1

$$E(m) = \sum_{k=1}^N p_k = p.$$

as for Eq. 2, but

$$\begin{aligned} V(m) &= \sum_{k=1}^N p_k(1-p_k) \\ &= \sum_{k=1}^N p_k - \sum_{k=1}^N p_k^2 \\ &= N\bar{p} - \sum_{k=1}^N p_k^2, \end{aligned}$$

where $\bar{p} = \sum_{k=1}^N p_k / N$. Adding and subtracting $N\bar{p}^2$ from $V(m)$ gives

$$\begin{aligned} V(m) &= (N\bar{p} - N\bar{p}^2) + (N\bar{p}^2 - \sum_{k=1}^N p_k^2) \\ &= N\bar{p}(1 - \bar{p}) - (\sum_{k=1}^N p_k^2 - N\bar{p}^2) \\ &= V'(m) - N[V(p_k)] \end{aligned}$$

by the definition of the population variance. When the p_k are all equal $V(p_k)$ is zero and $V(m) = V'(m)$. When the p_k are not equal, the variance of m is reduced by N times the variance of the p_k from what it would be in N independent "trials" each with the same probability of success. Since

$$K_{SM} = m/N$$

and

$$E(K_{SM}) = \bar{p},$$

$$\begin{aligned} V(K_{SM}) &= V(m)/N^2 \\ &= \frac{\bar{p}(1 - \bar{p})}{N} - \frac{V(p_k)}{N} \end{aligned}$$

K_{SM} cannot be distributed as the estimate $\bar{p}$ from N binomial trials, because this assumption leads to overestimation of the variance (as recognized by Goodall). Goodall suggests a normal approximation using the "correct" variance as being a better approximation. This *may* be true. Goodall assumes it is true by "analogy with the binomial case."

The decision rule can be stated in terms of m rather than K_{SM} . Usually the number of attributes, N , will be sufficiently large to assure that at least the equal probability case will be adequately approximated by the normal distribution.

The normal theory decision rule is

$$\delta(m) = \begin{cases} \text{Not Different if } \frac{m - E(m)}{sd(m)} \geq z(\alpha) \\ \text{Different if } \frac{m - E(m)}{sd(m)} < z(\alpha) \end{cases}$$

where

$$sd(m) = [V(m)]^{1/2}$$

and $z(\alpha)$ is the α -fractile of the standard normal distribution. Both $E(m)$ and $sd(m)$ depend on the unknown p_k so their sample estimates based on $\hat{p}_k$ must be used. (In all rigor we should devise some kind of t-statistic but this leads to complication of degrees of freedom and non-centrality parameters. The normal approximation provides an upper limit to the power of the decision rule, so will be adequate for expository purposes.) The value $z(\alpha)$ is called the critical value, the "cut point," of the decision rule, and the part of the real line less than $z(\alpha)$ is called the "critical region." The standardized m values are treated as though they were distributed as $N(0,1)$ variables, thus assigning probabilities to any observed m value.

The decision rule can be stated: Decide the samples are different whenever

$$m < \sum_{k=1}^N \hat{p}_k + z(\alpha) \hat{sd}(m)$$

and not different otherwise. Since m , the $\hat{p}_k$ and $\hat{sd}(m)$ are all calculated from the data, all that remains to completely specify the decision rule is to set α and look up $z(\alpha)$ in a table of the standard normal distribution (e.g., Pearson and Hartley, 1962). Once α is fixed, we have an " α -level decision rule." Suppose we want "95% confidence" of making the correct decision when there is no difference. This means we are willing to allow

an $\alpha = 0.05$ probability of deciding there is a difference when, in fact, there is not. Now $z(0.05)$ is -1.645 so that we decide the samples are different whenever

$$m < \sum_{k=1}^N \hat{p}_k - 1.645 \hat{sd}(m).$$

The null hypothesis tested is,

$$H_0: E(m) \geq E_0(m)$$

where $E_0(m)$ is the expectation of m under the assumption that the attribute population probabilities are $\hat{f}_{1k}$, i.e., assuming $f_{1k} = \hat{f}_{1k}$ and assuming that the normal approximation is appropriate.

EXAMPLE OF APPLICATION OF THE DECISION RULE

To help clarify the ideas used in arriving at this decision rule, consider the following example. Suppose we are sampling 8 binary attributes (an order of magnitude less than usually encountered in situations where the matching coefficient has been applied). Suppose further that it has been observed that the $\hat{f}_{1k}$ are as follows and $\hat{p}_k = \hat{f}_{1k}^2 + (1-\hat{f}_{1k})^2$.

Attribute Seen: Attribute (k):	Frequently <u>1 2 3 4</u>	Rarely <u>5 6</u>	Predominantly <u>7 8</u>
$\hat{f}_{1k}$	.5	.05	.95
$\hat{p}_k$	.5	.905	.905
$1-\hat{p}_k$	.5	.095	.095

Then, the statistics of interest are:

$$\sum_{k=1}^N \hat{p}_k = 4(.5) + 4(.905) = 5.620 = \widehat{E(m)}$$

$$\sum_{k=1}^N \hat{p}_k (1-\hat{p}_k) = 4(.5)^2 + 4(.905)(.095) = 1.3439 = \widehat{V(m)}$$

$$\widehat{sd(m)} = (1.3439)^{1/2} = 1.1593$$

$$\widehat{\bar{p}} = 5.620/8 = .7025.$$

The associated cumulative probabilities for the possible values of m are given in Figure 1.

The normal theory 0.05 critical region for this example is

$$m < 5.620 - 1.645(1.1593) = 3.71$$

FIGURE 1. Cumulative Distributions of m for $N = 8$ under H_0 : No Difference.

and the 0.1 level region is

$$m < 5.620 - 1.2816(1.1593) = 4.13.$$

This means that the samples would be judged different if m is 0, 1, 2, or 3 with 95% confidence and 0, 1, 2, 3, or 4 with 90% confidence. These "cut points" are associated with the confidence levels in Figure 1. The cumulative distributions for both the exact distribution, using Eq. 1, and the normal approximation are graphed in Figure 1 as a step function and a straight line respectively. The binomial approximation is indicated by a big dot for each m value. It will be noted the binomial test would not lead to the decision that the samples are different for $m = 3$ for an $\alpha = 0.05$ test, nor for $m = 4$ for the $\alpha = 0.1$ test. At the 95% confidence level, the exact distribution results in the same rejection values for m as for the normal case, but an m of 4 would not indicate a difference at the 90% level.

THE POWER OF THE DECISION RULE

We now ask the question, what is the power, i.e., the chance of detecting a true difference, using the $\alpha = 0.05$ level decision rule? In order to answer this question, we need to clarify what we mean by a "true difference." In terms of the normal theory decision rule, a difference is specified by different values for the parameter vector, $[E(m), V(m)]$. These, in turn, are dependent upon different values for the unknown f_{1k} in the two samples from which the match vector is determined and m calculated. Goodall's theory of the distribution of the matching coefficient depends completely on the truth of the "null hypothesis" of no difference so that the f_{1k} are the appropriate population probabilities for attribute k from whatever sample. This assumption allows study of the matching coefficient under a "suitable null hypothesis" (Goodall, 1967)*, and is, of course, standard practice. Also standard practice is to specify what happens or at least indicate there may be theoretical problems, when the null hypothesis fails to hold, that is, under an alternative hypothesis. The only reasonable alternative to Goodall's assertion of the null hypothesis by specifying that each f_{1k} is constant in the population (equivalently, all samples come from the same ecological community) is that at least some f_{1k} are different in some samples. This alternative hypothesis implies at least two different ecological communities, the second state of nature the decision rule must discriminate against. Suppose, under the alternative hypothesis, that two samples used to calculate m are from different ecological communities with respective attribute probabilities f_{A1k} and f_{B1k} . Suppose further that the two communities are measured on the 8 attributes of the example and, as in the example

$$f_{A1} = (.5, .5, .5, .5, .05, .05, .95, .95)$$

but let

$$f_{B1} = (.05, .05, .95, .95, .5, .5, .5, .5)$$

*It also allows using the sample estimates of the f_{1k} to be unquestioned (by Goodall) in application of the theory to the discrimination problem.

By the hypotheses, f_{A1} and f_{B1} are population parameters, not empirical estimates. One would judge these two communities to be quite different with respect to attribute probabilities; the commonly, predominantly and rarely observed species being reversed for the same set of probabilities. The probability distribution of m is the same in both populations, since m is invariant under permutation of the attributes, that is, m only counts the number of matches irrespective of the attributes for which they occur. This completely confounds the usual procedure for determining the power of a decision rule by comparing the distributions of the test statistic (in this case m) under the null hypothesis and under the alternative hypothesis. Only one of a potentially infinite number of alternatives has been specified. Although it appears to be a reasonable alternative and reflects a quite extreme difference in the two communities, Goodall's "Poisson binomial" model does not allow for calculating the power of his test.

Of course the fallacy in our argument is the suggestion to compare the m 's from two different communities, not the attribute matches for a sample from each of the two communities. But this is the quandry that extending Goodall's logic leads us into. If all samples are from the same community, as specified by its f_{1k} , then Goodall's theory follows. But this theory leaves no room for specifying a meaningful alternative community under different specifications of the f_{1k} . The problem is that his null hypothesis is not "suitable." Goodall uses a poorly chosen model. The discussion in Kendall and Stuart (1958, Sec. 5.10) referenced by Goodall in support of his assertion that the "distribution of N is in fact a particular case of the "Poisson binomial" (sic), has to do with sampling from N different statistical populations. We are attempting to distinguish between two statistical populations, each characterized on N different attributes. The key to a more appropriate model is given by Kenaall and Stuart (1958) in Section 5.31, where the bivariate binomial and a potential extension to the multivariate binomial distribution are discussed. Since K_{SM} is defined on pairs of vectors, the most general "suitable" model would involve use of the multivariate multinomial

distribution (Wishart, 1949) in which the "multivariate" applies to the N dimension (attributes or species) and the "multinomial" to the four possible outcomes of a pairwise attribute comparison. The effect of assumptions of sample independence, of attribute independence, and of population identity or diversity could then be explicitly stated in terms of the model. Since K_{SM} groups matches of either kind and mismatches of either kind, it is conjectured that the distribution of m under the null hypothesis would reduce to the multivariate binomial mentioned above. It is also conjectured that the distribution under the alternative hypothesis could be handled more generally. These topics require more basic research than lies within the scope of this paper.

Some clarification of the nature of the decision problem posed, and specification of Goodall's hypotheses are required. Goodall does not state his hypotheses explicitly. His implied null hypothesis is that K_{SM} (his M) is not different from "its expectation under random occurrence of attributes." Taken in the context of our discrimination problem, "random occurrence" means following the attribute relative frequencies, $\hat{f}_{1k}$, determined from the entire set of samples in the study. These are used to calculate the match probabilities, $\hat{p}_k$, which in turn provide the expected value and variance of K_{SM} (and m) used in the normal approximation. The null hypothesis tested is thus

$$H_0: \mu = \mu_0$$

where μ_0 is the mean of a normal distribution with population mean

$$\mu_0 = \sum_{k=1}^N \hat{p}_k$$

and population variance

$$\sigma^2 = V(m).$$

(The equating of population parameters and sample statistics is necessary for the statement of Goodall's significance test.) This null hypothesis is the same as stated previously except that equality is hypothesized rather than $\mu \geq \mu_0$. Goodall's procedure has a two-sided alternative in which

$$m > \mu_0 + z_{(1-\alpha/2)} \widehat{sd}(m)$$

implies the two samples "are significantly more similar than could be expected on random assortment of attributes," and

$$m < \mu_0 + z_{\alpha/2} \widehat{sd}(m)$$

implies "they are significantly dissimilar." Goodall's decision space has three alternatives, namely; too similar, no difference, and dissimilar. The "too similar" decision has no practical meaning. If $K_{SM} = 1$, or equivalently $m = N$, the two samples would be identical. Any K_{SM} less than unity is indicative of some shortfall from identity or perfect similarity. One might ask what action an investigator would take if informed that the samples indicated the populations were "too similar." The two alternatives, "no difference" and "dissimilar" exhaust the alternatives of practical interest. Combining the two alternatives "too similar" and "no difference" does not detract from the statement of the decision problem. It also simplifies the calculation of the power.

The remarks in the above paragraph are applicable when the significance test is intended to discriminate between two ecological communities. Most probably Goodall did not intend that his significance test be used for such a purpose. Since the distribution of m is empirically determined from the set of all samples in the study, and since each sample could be from a different community, the attribute relative frequencies ($\hat{f}_{1k}$) are averages over all the different communities which might be in the sample. Assuming that these averages characterize a single community in terms of its binomial probabilities for each attribute (f_{1k} , without the carat) is usually an

obviously fallacious assumption. We are not accusing Goodall of using this assumption. In addition, given a set of N binomial attribute probabilities specifying the population under H_0 , any permutation of the species list can be made without changing the value of $E_0(m)$. This means that it makes no difference which particular species are predominant and which rare. In effect Goodall's null hypothesis only specifies that the distribution of the f_{1k} is the same from sample to sample. He does not put the same interpretation on "random occurrence" as we do. This is evident from his use of the more descriptive words "random assortment of attributes" in his statement of the alternatives to his null hypothesis. We must (and validly may) interpret "random occurrence" as implying that f_{1k} is associated with a particular attribute, k , in order to adequately characterize the decision problem. When Goodall uses the words "random assortment" his results indicate that he means that any permutation of the species list can be associated with the f_{1k} in their fixed, but arbitrary, order. In effect Goodall provides a test of the null hypothesis

H_0 : the two populations have the same index of diversity,
against the two-tailed alternative

H_A : the two populations have different indices of diversity.

The index of diversity referred to is that discussed by Mountford (1962), Williams (1949) and Fisher, et al. (1943), and is the parameter, called a (not to be confused with the probability allowed for a Type I error), of the logarithmic series. This follows from the interpretation of the terms of the logarithmic series

$$\frac{\alpha x^n}{n}, \quad n = 1, 2, \dots$$

as representing the number of species with n individuals in the population, (Williams, 1949). Then the probability of observing a species with n individuals is greater than the probability of observing a species with, say, $n/2$ individuals when a random sample is taken. The particular species

identification is irrelevant to the estimation of a , or the calculation of $E_0(m)$, as long as the species are differentiated into classes with the same expected number of individuals per species.

Goodall's significance test is thus not appropriate for our discrimination problem. No statistic based on a comparison with $E_0(m)$ will have the desirable property of being sufficient (in the sense of Lehman, 1959) for the family $F = (f_{11}, \dots, f_{1N})$ of multivariate binomial distribution functions. This is because $E_0(m)$ is invariant under permutation of the vector

$$F_0 = (\hat{f}_{11}, \dots, \hat{f}_{1N})$$

specifying the distribution under the null hypothesis. In this context the "too similar" alternative is meaningful. If a pair of sample vectors have significantly more matches than expected, then there is an indication of relatively more species in the rare and/or predominant range of probabilities, indicating a different index of diversity.

The above remarks do not void the applicability of Goodall's test to the discrimination problem. They merely indicate that not much power can be expected to be associated with his test. His test has the same null hypothesis as ours, but different alternative hypotheses. Whereas his alternative hypotheses do not require associating a particular f_{1k} with a particular species, k , ours does. Our alternative distribution, specified by the vector, f_{B1} , specifies a population with the same index of diversity as specified by f_{A1} . The f_{B1k} are associated with particular species. This causes the match probabilities, p_{k1} to change and, consequently, the expectation and variance to change.

In order to carry through our example of a determination of the power of the decision rule, we will assume that the match probabilities under the alternative hypothesis can be calculated as

$$p_k^i = f_{A1k}f_{B1k} + (1-f_{A1k})(1-f_{B1k}) \quad (3)$$

We do this only to show what happens when Goodall's theory is applied to a reasonable alternative. We do not advise it as a good approach to solving the discrimination problem. As it turns out, for the two specific populations p_k^i is 0.5 for all k . All p_k^i are the sum of the products $(.5)(.05) + (.5)(.95) = .025 + .475 = .5$. This means that a match or a mismatch is equally likely for each attribute when a sample from community A is compared with a sample from community B.*

Table 1 gives the probabilities of observing each attribute, f_{A1} and f_{B1} , the match probabilities when both samples are from community A (Under H_0) and the match probabilities when two samples come, one each, from both communities (Under H_A). The middle part of the table gives the associated probability distributions of m under Goodall's formulation. Corresponding to each possible value of m , called m' , are the exact probabilities ($m = m'$), the probability that m is less than or equal to m' (the cumulative distribution) and the cumulative distribution under the normal approximation ($z < z'$). The constructed population parameters (not statistics) are given at the bottom of the table.

We now have the requisites for calculating the power of discrimination of Goodall's suggested decision rule against the single alternative specified by the f_{B1} and Eq. 3. (Under the questionable assumption that Goodall's theory is correct.) The $\alpha = 0.5$ decision rule defined and exemplified above rejects the null hypothesis of no difference whenever a small number of matches ($m = 0, 1, 2$, or 3) is observed, no matter which distribution is the correct one. The two exact cumulative distributions are graphed in Figure 2 (as step functions) along with their normal theory approximations (as continuous straight line functions). Under the alternative hypothesis distribution, the probability that $m' = 0, 1, 2$, or 3 is

$$P[z \leq \frac{3.71 - 4}{1.4142}] = P[Z \leq -.2051] = .419$$

*This provides a counter example to Goodall's assertion "that 0.5 is in fact the lower limit of the expectation (of the matching coefficient), only realized when $f_{ij} = 0.5$ for all values of i and j ."

TABLE 1. Comparison of the Distribution
of m under H_0 and H_A

Attribute (k)	Probabilities of Observing Attributes and Scoring a Match			
	Under H_0		Under H_A	
	f_{A1}	p_k	f_{B1}	p'_k
1	.5	.5	.05	.5
2	.5	.5	.05	.5
3	.5	.5	.95	.5
4	.5	.5	.95	.5
5	.05	.905	.5	.5
6	.05	.905	.5	.5
7	.95	.905	.5	.5
8	.95	.905	.5	.5

m'	Probability			Probability		
	$m=m'$	$m \leq m'$	$z \leq z'$	$m=m'$	$m \leq m'$	$z \leq z'$
0	5.09×10^{-6}	5.09×10^{-6}	6.9×10^{-7}	.00391	.00391	.0023
1	.00021	.00022	.00003	.03125	.03516	.0169
2	.00358	.00380	.00090	.10938	.14453	.0786
3	.02988	.03368	.01191	.21875	.36328	.2397
4	.12975	.16342	.0812	.27344	.63672	.5000
5	.28461	.44804	.2963	.21875	.85547	.7603
6	.32474	.77278	.6473	.10938	.96484	.9214
7	.18530	.95808	.8925	.03125	.99609	.9831
8	.04192	1.00000	.9823	.00391	1.0000	.9977

$E(m)$	5.620	4.0000
$V(m)$	1.3439	2.0000
$sd(m)$	1.1593	1.4142
$\bar{p}$	0.7025	0.5
$V'(m)$	1.67195	2.0
$sd'(m)$	1.29303	1.4142

FIGURE 2. Cumulative Distributions of m for $N = 8$ under H_0 and H_A .

using the normal theory decision rule. The probability of concluding that the two samples come from the specified different populations is thus 0.419 or odds of less than 50-50. The chance of failing to detect the difference is the complement of this or 0.581. The power of this 0.05 level decision rule is

$$(1 - \beta) = P(\text{rejecting } H_0 \text{ when } H_A \text{ is true}) = .42$$

and the probability of making a Type II error is

$$\beta = P(\text{not rejecting } H_0 \text{ when } H_A \text{ is true}) = .58.$$

Another view of the solution to the problem of calculating power is given in Figure 3. The two normal frequency functions are graphed for the two distributions of Table 1, the normal distribution with mean 5.62 and variance 1.34, the distribution under the hypothesis (H_0) of no difference, and the distribution with mean 4.0 and variance 2.0, the distribution under the specified alternative (H_A). The normal theory cut point is 3.71 and 5% of the area under the distribution for H_0 is in the rejection region of the decision rule, that is, a decision of "different" will be made 5% of the time when the decision should be "not different." However, a decision of "not different" will be made in 58% of the cases when the alternative distribution is the correct one. Using this rule, there is less than a 50-50 chance of making the right decision when the two samples are in fact from the different populations specified. Using an $\alpha = 0.1$ decision rule, the cut point moves up to 4.13 and the chance of making the right decision when the communities are different increases to about 54 in 100. In order to attain a power of 0.9 the cut point needs to be moved clear up to 5.81, implying an α -level of 0.56, or a better than 50-50 chance of making the wrong decision when the samples are from the same community. The decision rule is not very satisfactory against this alternative.

FIGURE 3. Normal Approximations to the Frequency Distribution of m under H_0 and H_A .

The alternative specified by the f_{B1} led to match probabilities of 0.5 for all 8 attributes. This result occurred since either f_{A1k} or f_{B1k} was 0.5 for all species, so that by Eq. 3,

$$\begin{aligned} p'_k &= f_{1k}f'_{1k} + (1 - f_{1k})(1 - f'_{1k}) \\ &= 0.5[f'_{1k} + (1 - f'_{1k})] = 0.5 \end{aligned}$$

whatever the value of f'_{1k} , the observation probability (f_{A1k} or f_{B1k}) which is not equal to 0.5. As shown above (see also Feller, 1957, p 216)

$$V(m) = V'(m) - N[V(p_k)]$$

so that $V(p_k) = 0$ and

$$V(m) = V'(m) = 8(0.5)^2 = 2$$

for the p'_k . As pointed out by Feller (1957, p 217), $V(m)$ assumes its maximum when all the p_k are equal under the condition that the average of the p_k remains the same. It can also be shown, by differentiating $V'(m) = Np_0(1 - p_0)$, that $V'(m)$ attains its maximum when $p_0 = 0.5$. It then follows that p'_k produces the maximum variability of m for the 8 species example under any alternative based on permutations of the f_{A1k} . The alternative specified by the f_{1Bk} is the least discriminable against the null hypothesis that the f_{1Ak} are appropriate for both samples. The power against this alternative thus provides the lower limit to the power of the significance test.

The most extreme alternative hypothesis using permutations of the 4 frequently observed ($f_{1k} = .50$), 2 rarely observed ($f_{1k} = .05$), and 2 predominantly observed ($f_{1k} = .95$) species is characterized by population C as specified in Table 2, under f_{C1} .* If 0.5 is paired with 0.05 or 0.95,

*Note that under this alternative the expected value of m is 2.38 implying $E(K_{SM}) = 2.38/8 = 0.2975$ and controverting Goodall's assertion that "0.5 is in fact the lower limit of the expectation." This assertion only holds under the assumption that the null hypothesis is true.

TABLE 2. The Distribution of m for Samples
from Populations A and C (under $H_{\max}$)

Attribute	f_{C1}	f_{A1}	p_k''
1	.5	.5	.5
2	.5	.5	.5
3	.5	.5	.5
4	.5	.5	.5
5	.95	.05	.095
6	.95	.05	.095
7	.05	.95	.095
8	.05	.95	.095

Probability Distributions

m'	Exact		Binomial		Normal	
	$m=m'$	$m \leq m'$	$m=m'$	$m \leq m'$	$m=m'$	$m \leq m'$
0	.04192	.04192	.05932	.05932	.08315	.03828
1	.18530	.22723	.20096	.26027	.23547	.15224
2	.32474	.55197	.29785	.55813	.38330	.38868
3	.28461	.83657	.25228	.81041	.35887	.67772
4	.12975	.96633	.13355	.94396	.19292	.88599
5	.02988	.99620	.04524	.98920	.05965	.97439
6	.00358	.99978	.00958	.99878	.01060	.99647
7	.00021	.99999	.00116	.99994	.00108	.99971
8	5.09×10^{-6}	1.00000	.00006	1.00000	.00006	.99998

$$\begin{aligned}
 E(m) &= 2.38 & \bar{p}_* &= 0.2975 \\
 V(m) &= 1.3439 & V'(m) &= 1.67195 \\
 sd(m) &= 1.1593 & sd'(m) &= 1.29303
 \end{aligned}$$

as for population B above, then the difference between attribute probabilities is 10.451 for all attributes, giving 1.62 as the sum of squared differences. When attribute probabilities are paired as in Table 2, four of the differences are zero but four are 10.901 giving 3.24 as the sum of squared differences. The population maximally different from population A by this criterion is population C. Calculation of the power against population C will provide the upper limit to the power of the decision rule under the constraints imposed (8 attributes and permutations of the given probabilities). The calculation of the power against population B provided the lower limit to the power.

Using the same calculational methods as before we have

$$p_k'' = f_{A1k}f_{C1k} + (1 - f_{A1k})(1 - f_{C1k})$$

$$= \begin{cases} (.5)(.5) + (.5)(.5) & = .5 & \text{for } k = 1, 2, 3, 4 \\ (.95)(.05) + (.05)(.95) & = .095 & \text{for } k = 5, 6, 7, 8, \end{cases}$$

as listed in Table 2. The exact distribution of m was calculated using Eq. 1, and is given in Table 2 along with the binomial approximation using $\bar{p}$ for the p_k'' (.2975) and the normal approximation using $E(m)$ and $V(m)$, (2.38 and 1.3439 respectively). The cumulative distributions are given in each case.

Figure 4 gives the cumulative probability plots for these 3 distributions and, for comparison, for the normal approximation for population A. The $\alpha = 0.05$ decision rule specified that the populations should be judged differently wherever $m = 0, 1, 2$ or 3. The power of this decision rule against population C is thus,

$$p(m = 0, 1, 2, 3 \text{ given } p_k'') = \begin{cases} 0.839 & \text{for the normal} \\ 0.810 & \text{for the binomial} \\ 0.837 & \text{for the exact} \end{cases}$$

FIGURE 4. Cumulative Distribution of m for $N = 8$ under H_0 and $H_{\max}$.

The power against population B was

$$p(m = 0,1,2,3 \text{ given } p_k^t) = \begin{cases} 0.419 & \text{for the normal} \\ 0.363 & \text{for the binomial} \\ 0.363 & \text{for the exact} \end{cases}$$

Thus the chance of failing to detect a true difference varies from about 2 out of 3 ($1 - 0.363 = 0.637 \cong 2/3$) to 1 out of 6 ($1 - 0.837 = 0.163 \cong 1/6$). The power of the test for the maximally discriminable alternative is not too bad considering that only $N=8$ attributes are used. However, the population specified by the f_{1Bk} , which intuitively looked quite different from the f_{1Ak} , turned out to be the minimally discriminable population. The significance test had unacceptably low power, less than 0.50, against population B.

As pointed out above, the normal theory test provides an upper limit to the power when a t-test should be used. It also provides an upper limit when used as an approximation to a discrete distribution, as is the case here. The situation is thus even worse than the normal theory test indicated. The true distributional form is a step function, not continuous. In practical situations where 100 or more attributes are compared, the normal approximation may be almost indistinguishable from the step function when graphed as in Figure 2. The exact distribution is still calculatable but requires calculating 2^N probabilities and summing the right ones together to get the probability of an observed m . In the example with only 8 attributes, reference to Table 1 indicates that under H_0 (community A)

$$P(m \leq 3) = .03368 < 0.05$$

and

$$P(m \leq 4) = .16342 > 0.1.$$

If our desire is for an $\alpha = 0.05$ level test, the closest we can come is to use a rejection region of $m = 0, 1, 2, 3$ which has an exact α -level of 0.03368. The corresponding probability, in Table 1, under H_A is 0.36328,

so that the chance of concluding the samples come from the two different communities when in fact they do is only slightly better than 1 in 3 and we have a 2 out of 3 chance of failing to detect the difference. For larger N , where the normal approximation is more appropriate, the difference in power will not be so dramatic, but the normal theory decision rule will still overestimate the power.

We have only considered two specific alternatives and even for these cases the calculations were not few, although they were fairly simple. A better formulation of the underlying probability theory for m would make the calculations more tractable and not subject to such detailed specification of each case and alternative. The important point is that the power calculated (under somewhat shaky assumptions) was so poor that use of K_{SM} in an ecological context as a discrimination rule is futile.

We have also only considered one similarity measure of those discussed in the previous report. After reading this discussion of the power calculation for the comparatively simple K_{SM} , it is hoped that the reader will understand why the other coefficients were not investigated in the same detail. For most of the coefficients mentioned in Similarity Indices I, the calculations would be more complicated supposing that probabilistic models could be developed for them.

CONCLUSION

There is one major point the author hopes has come through in this and its companion report. This is: the use of similarity measures in a discrimination problem may be possible, but **it** is an extremely laborious approach providing little hope of success and dependent upon a multitude of decision points over which the final formulator of the decision rule (the statistician) has no control. The statistician is forced to use a statistic designed for the specific purpose of describing some specific aspect of an ecological community (or sets of ecological communities) as a statistic to test whether or not the difference in two (or more) communities is statistically significant. The problem could be approached from the other end by designing a sampling scheme to provide a statistic which would discriminate between two populations and have a more tractable probability theory. **If** an ecologist can "ad hoc" descriptive statistics, a statistician should be able to "ad hoc" discrimination ecology--with, of course, the assistance of ecologists. Statistical decision theory has a well developed literature which could provide valuable contributions to the discrimination problem.

REFERENCES

Feller, W. An Introduction to Probability Theory and Its Applications. John Wiley and Sons, New York, 1957.

Fisher, R. A., A. S. Corbet and C. B. Williams. The relation between the number of species and the number of individuals in a random sample of an animal population. J. Animal Ecol. 12:42-58, 1943.

Goodall, D. W. The distribution of the simple matching coefficient. Biometrics 23:647-656, 1967.

Johnston, J. W. Similarity Indices I: What Do They Measure? Addendum to BNWL-2151, Battelle, Pacific Northwest Laboratories, Richland, Washington, 1976.

Kendall, M. G. and A. Stuart. The Advanced Theory of Statistics, Vol. 1. Distribution Theory. Charles Griffin & Co., London, 1958.

Lehmann, E. L. Testing Statistical Hypothesis, John Wiley and Sons, New York, 1959.

Mountford, M. D. An index of similarity and its application to classificatory problems, pp. 43-50. In: P. W. Murphy (ed.) Progress in Soil Zoology, Butterworth, London, 1962.

Pearson, E. S. and H. O. Hartley. Biometrika Tables for Statisticians, Vol. 1. Cambridge U. Press, London, 1962.

Preston, F. W. The commonness and variety of species. Ecology 29: 254-283, 1948.

Sokal, R. R. and C. D. Michener. A statistical method for evaluating systematic relationships. U. Kansas Sci. Bull. 38:1409-1438, 1958.

Williams, C. B. Jaccard's generic coefficient and coefficient of floral community in relation to the index of diversity. Annals of Botany 13:53-58, 1949.

Wishart, J. Cumulants of multivariate, multinomial distributions. Biometrika 36:47, 1949.

DISTRIBUTION

No. of
Copies

No. of
Copies

OFFSITE

ONSITE

A. A. Churm
ERDA Chicago Patent Group
9800 South Cass Avenue
Argonne, IL 60439

3 M. Jinks, Chief Mail and
Files
USNRC Central Files
Washington, DC 20555

245 ERDA Technical Information
Center
For Basic Distribution
Under NRC-1

P. G. Voilleque
ERDA Health Services
Laboratory
Idaho Falls, ID 83401

H. T. Peterson
USNRC Office of Standards
Development
Washington, DC 20555

ERDA Richland Operations Office

P. G. Holsted

Atlantic Richfield Hanford Company

G. E. Backman

United Nuclear Industries, Inc.

A. E. Engler

Hanford Engineering Development
Laboratory

G. D. Carpenter

40 Battelle-Northwest

K. L. Gore (5)
J. L. Helbling
C. Huges
J. Johnston (14)
L. D. Kannberg (5)
J. Mahaffey
J. A. Strand
J. M. Thomas
D. G. Watson (2)
B. E. Vaughan
Technical Information Files (5)
Technical Publications (3)