MEASURES OF DIVERSITY WITH UNBIASED ESTIMATES

WOOLLcott Smith, J. Frederick Grassle
AND David KRAVITZ

Woods Hole Oceanographic Institution
Woods Hole, MA 02543 USA

SUMMARY. The results of Smith and Grassle (1977) for the family of expected species diversity measures are reviewed. These results are applied to estimating variability in local diversity in the deep sea from small benthic cores. These results are extended to a wider class of diversity indices including a modification of Brillouin's index. Finally, unbiased estimation methods are applied to the estimation of species-area diversity.

KEY WORDS. species diversity, sampling variance, unbiased estimators, species-area diversity, Shannon index, Brillouin index, deep sea, benthic communities.

1. INTRODUCTION

Many measures have been designed to quantify the species diversity of a community (Fager, 1972; Hill, 1973; Hurlbert, 1971; Peet, 1974; Patil and Taillie, 1979). These diversity indices have a variety of theoretical, mathematical, and ecological justification. The debate surrounding the use and interpretation of diversity measures concerns more than one methodological problem. In this paper we address the statistical problem of determining the accuracy and precision of diversity estimates based on finite, sometimes very small, samples taken from a large community. With these statistical problems solved, the empirical investigation of species diversity can proceed on a firmer statistical basis.
This paper addresses the following general problem: although species diversity is a property of a community, in practice, diversity measures are estimated from relatively small samples taken from the community. For example, we might be interested in comparing two deep sea benthic communities. One community is sampled with a core 35 cm² with a mean of 10 individuals per core; the other community is sampled with a large box core, 2500 cm², with about 1,000 individuals per core. To compare diversity for these communities we need to know how accurately we can estimate diversity from each sampling device and whether the smaller samples will consistently underestimate the true community diversity. The same kind of problems arise when we compare small scale (within habitat) diversity with community diversity.

Bowman, Hutcheson, Odum, and Shenton (1971) have given an extensive review of the small sample properties of Shannon-Wiener information and Simpson's index. Blyth (1958) has shown that there is no unbiased estimator for Shannon's information, so there is no way to remove completely the sample size dependence of Shannon's information. On the other hand, Simpson's index has a simple unbiased estimator. We shall show in later sections that this property of Simpson's index will simplify much of the sampling theory. The basic objection to Simpson's index is ecological: it is heavily dependent on the dominant species in the community, and thus will only reflect the properties of the dominant species.

This paper describes a family of diversity measures that have unbiased estimators, and derives unbiased estimators for their sampling variances. Section 2 reviews the results of Smith and Grassle (1977) for the family of expected species diversity measures. In Section 3, these results are applied to estimating variability in local diversity in the deep sea from small benthic cores. In Section 4, the expected species index is compared with Shannon's information. Section 5 extends these results to a wider class of diversity indices including a modification of Brillouin's index. Finally, Section 6 applies unbiased estimation methods to estimation of species area diversity. The new results are contained in Sections 4 through 6.

The basic intent of this work is to apply a well-known and straightforward statistical theory, unbiased estimation, that is valid for small samples. This requires a large increase in computational complexity of the estimators, but modern computers can easily handle this problem. Some of the unbiased estimation results developed here are closely related to the more general work of Patil and Joshi (1970) and Joshi and Patil (1972) on sum-symmetric power series distributions.
2. EXPECTED SPECIES DIVERSITY

In this section some notation is introduced and the results in Smith and Grassle (1977) are summarized. Hurlbert (1971), Sanders (1968), and others have proposed a generalization of Simpson’s (1949) diversity measure. The proposed diversity measure is based on the expected number of species among \( m \) individuals drawn at random from a multinomial population.

Let \( \pi_i \) denote the probability that an individual drawn from the population will be from species \( i \). The vector \( \pi = (\pi_1, \pi_2, \ldots, \pi_k) \) defines the multinomial population. The random variable \( S_m \) denotes the number of species present when \( m \) individuals are drawn at random from the population. The expected species diversity measure at individual index \( m \) is defined as

\[
s(m) = E[S_m | \pi] = \sum_{i=1}^{k} \left[ 1 - (1-\pi_i)^m \right].
\]  

(2.1)

Expected species diversity is a family of diversity indices with varying rare species sensitivities. For large \( m \), \( s(m) \) is sensitive to the rare species in the population. For small \( m \), \( s(m) \) depends mainly on the dominant species. The diversity \( s(2) \) is linearly related to Simpson’s diversity measure.

An important property of expected species diversity is that an unbiased estimator for \( s(m) \) exists. Suppose that we have a random sample of size \( N \) from the population and let \( N_i \) denote the number of individuals of species \( i \). The vector \( N = (N_1, N_2, \ldots, N_k) \) is the minimal sufficient statistic for the multinomial population \( \pi \). A straightforward application of the Rao-Blackwell theorem shows that the minimum variance unbiased (MVU) estimator for \( s(m) \) is

\[
\hat{s}(m) = E[S_m | N] = \sum_{i=1}^{k} \frac{1}{m} \left[ 1 - \left( \frac{N_i - N}{m} \right)^m \right] \left( \frac{N}{m} \right)
\]  

(2.2)

when \( N > m \). The MVU estimator for the variance can also be found and is

\[
\hat{\text{Var}}[\hat{s}(m)] = - \text{Cov}[S'_m, S''_m | N]
\]  

(2.3)

for \( N \geq 2m \), where random variables \( S'_m \) and \( S''_m \) denote the number of species in two mutually exclusive subsamples of size \( m \) from the original sample, \( N \). The rather complicated numerical
formulas for the variance estimator (2.3) depends on the properties of the multivariate hypergeometric distribution and is given in Smith and Grassle (1977). In Section 5 we define a larger family of diversity measures for which unbiased estimation theory applies. And in Section 6 we will extend these ideas to species-area diversity measures.

3. SMALL SCALE VARIATION IN DIVERSITY

Engen (1977) and Peterson (1976) have both proposed measures of local variation in diversity within a community. The basic problem is to partition the variation in diversity between sampling error and true variation in small scale diversity. For expected species diversity one can write

\[ \text{Var}[s(m)] = \text{Var}[\hat{s}(m)] - E_{\pi} \left[ \text{Var}[\hat{s}(m)|\pi] \right]. \]  

(3.1)

This formula will hold for any unbiased estimator of diversity. Suppose we have a set of unbiased estimates of the diversity \( s(m) \) from \( n \) samples, \( N_1, N_2, \ldots, N_n \), then an unbiased estimate of \( \text{Var}[s(m)] \) is

\[ \hat{\text{Var}}[s(m)] = \frac{\hat{\sigma}^2}{\hat{\sigma}^2} - \frac{1}{n} \sum_{i=1}^{n} \text{Var}[\hat{s}(m)|N_i]. \]  

(3.2)

where \( \hat{\sigma}^2 = \frac{1}{n-1} \sum \left( \hat{s}(m) - \hat{s}(m) \right)^2 \).

Equation (3.2) will be useful only when we have unbiased estimators for both \( s(m) \) and \( \text{Var}[\hat{s}(m)|\pi] \).

To illustrate the method we have computed the expected species index at two individuals for polychaetes from 16 cores taken with the submersible Alvin over a three year period at the Woods Hole bottom station DSO1 at a depth of 1,750 to 1,830 meters (Table 1). The main feature of these cores is that there are very few individuals in each core and the species diversity is high. The size of each core is 35 cm². The variance estimators were computed using (2.3) and using the maximum likelihood estimator, MLE, from Simpson's (1949) formula for the variance

\[ \text{Var}[\hat{s}(m)|\pi] = \frac{4 \sum_{i=1}^{k} \pi_i^2 + 2 \sum_{i=1}^{k} \pi_i - 2(N-3)(\sum \pi_i^2)^2}{N(N-1)}. \]  

(3.3)

For small diverse samples it is clearly inappropriate to use the MLE estimator for the variance. In this case the MLE estimator

<table>
<thead>
<tr>
<th>No. of Core</th>
<th>Species</th>
<th>Individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>1A</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>2A</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>18</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>19</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>15</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

**Table 1: Variation in polychaetes.**

On average, estimates slightly underestimate the true variance computed from \( \hat{s}(m) \), but this is expected. We can conclude that if there is a significant relationship it is too small to be detected.
TABLE 1: Variation in polychaetes diversity from sixteen 35 cm² cores.

<table>
<thead>
<tr>
<th>Core</th>
<th>No. of Species</th>
<th>No. of Individuals</th>
<th>$\hat{s}(2)$</th>
<th>Estimate for $\text{Var}[\hat{s}(2)] \times 10^2$</th>
<th>MVU</th>
<th>MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR</td>
<td>4</td>
<td>5</td>
<td>1.900</td>
<td>1.000</td>
<td>2.592</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>9</td>
<td>1.888</td>
<td>.440</td>
<td>.828</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>14</td>
<td>1.945</td>
<td>.068</td>
<td>.203</td>
<td></td>
</tr>
<tr>
<td>1A</td>
<td>7</td>
<td>8</td>
<td>1.964</td>
<td>.127</td>
<td>.596</td>
<td></td>
</tr>
<tr>
<td>2A</td>
<td>8</td>
<td>9</td>
<td>1.972</td>
<td>.077</td>
<td>.408</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>13</td>
<td>16</td>
<td>1.975</td>
<td>.007</td>
<td>.009</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>4</td>
<td>6</td>
<td>1.800</td>
<td>4.000</td>
<td>2.96</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>5</td>
<td>2.000</td>
<td>.000</td>
<td>1.600</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>8</td>
<td>1.892</td>
<td>1.147</td>
<td>1.238</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>10</td>
<td>1.866</td>
<td>.349</td>
<td>.722</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>6</td>
<td>6</td>
<td>2.000</td>
<td>.000</td>
<td>.925</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>8</td>
<td>1.964</td>
<td>.127</td>
<td>.596</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>8</td>
<td>13</td>
<td>1.923</td>
<td>.032</td>
<td>.250</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>10</td>
<td>1.911</td>
<td>.313</td>
<td>.598</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2.000</td>
<td>.000</td>
<td>3.125</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>7</td>
<td>1.904</td>
<td>.045</td>
<td>1.067</td>
<td></td>
</tr>
<tr>
<td>Variance estimates</td>
<td>.313</td>
<td>.477</td>
<td>.112</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

On average overestimates the true sampling variation. The MVU estimate slightly underestimates the true sampling variation computed from $\hat{s}_1^2(m)$, but this is well within sampling error. We can conclude that if there is variation in dominant species diversity it is too small to be detected with these small cores.

4. RELATIONSHIP BETWEEN EXPECTED SPECIES AND SHANNON'S INFORMATION

In studies of natural communities, diversity measures with quite different functional forms often are statistically strongly correlated. Loya (1972) compared Shannon's information index with Simpson's index using coral communities and found a nearly linear relationship between the two indices. Similar results have been reported by Risser and Rice (1971) and Johnson and Raven (1970). It is not surprising, then, that there are strong statistical relationships between the expected species index and other diversity indices for natural communities. We compared Shannon's information index with the expected species index for 10 and 200 individuals, using a series of benthic samples from the West Falmouth oil spill study (Sanders, Grassle, and Hampson, 1971). Only the polychaetes were used for the diversity calculations. The results
show a strong linear relationship for the 10-individual diversity index as indicated in Figure 1. In Figure 2 we see that the correlation for the 200-individual diversity index is not as strong.

The nearly linear relationship between Shannon's information index and expected species for 10 individuals can be explained by comparing the contribution of an individual species to each diversity index (Peet, 1974). If the proportion of a species in the population is \( \pi_i \), then its contribution to the Shannon diversity index is \( \pi_i \log \pi_i \). Similarly its contribution to the expected species index with \( m \) individuals is \( 1-(1-\pi)^m \). In Figure 3 the contribution terms for the information index and for the expected species index for 2, 10, and 200 individuals is shown. For \( .05 < \pi < .5 \) the curves for Shannon's information and the expected species index for 10 individuals have similar shapes, indicating that within this range the relationship between the two diversity indices is approximately linear. This relationship will hold for most natural populations; however, for populations with many very rare species or one abundant species this linear approximation will not hold.

5. A GENERALIZED DIVERSITY MEASURE

The results developed in Sections 2 and 3 can be extended to any diversity measure which has an unbiased estimator. We will define in this section a larger family of diversity measures with unbiased estimators. Following the notation of Patil and Taillie (1979), a dichotomous diversity measure \( \Delta \) is defined as

\[
\Delta(\pi) = \sum \pi_i R(\pi_i),
\]

(5.1)

where \( R(\pi) \) satisfies the more restrictive conditions of Patil and Taillie: \( R(\pi) \) is a decreasing function on the interval \( (0, 1) \) and \( \pi R(\pi) \) is concave on the interval \( (0, 1) \).

Now suppose as in Section 2 we have a random sample \( N \) from a multinomial population \( \pi \). Let \( \hat{\pi} \) denote the maximum likelihood estimate for \( \pi \), \( \hat{\pi}_i = N_i / N \). If the family of dichotomous diversity measures, \( \Delta \), satisfies the concave property, then

\[
E[\Delta(\hat{\pi}) | \pi] < \Delta(E[\hat{\pi} | \pi]) = \Delta(\pi).
\]

(5.2)

This follows from the expectation of a convex function (Fraser, 1957, p. 54). In other words, the estimator \( \Delta(\hat{\pi}) \) will on average underestimate the true population diversity.


**FIG. 1:** The expected species index for 10 individuals plotted for 114 samples of polychaetes from the West Falmouth oil spill study.

**FIG. 2:** The expected species index for 200 individuals plotted against Shannon's information index for the polychaete samples from the West Falmouth oil spill study.
FIG. 3: The contribution of a single species to Shannon's information index and to the expected species index for 2, 10, and 200 individuals as a function of $\gamma$, the proportion of individuals in the population.

For large samples this does not present much of a problem because of the asymptotic properties of the ML estimators. However, for small samples or to account for variability in diversity between small quadrats (as in the previous sections), it is important to obtain unbiased estimators for both $\Delta$ and $\text{Var}(\Delta)$. However, not all diversity measures possess unbiased estimators. Blyth (1958) and Good (1953, 1969) have shown that only polynomials in $\pi_i$ of degree $m \leq N$, where $N$ is the sample size, have unbiased estimators. This in turn implies that the family of unbiased dichotomous diversity measures must have the form

$$\Delta(\pi) = \sum_{i=1}^{k} V(\pi_i),$$

where $V(\pi_i) = \pi_i R(\pi_i) = \sum_{j=0}^{m} a_j \pi_i^j$.  

Since the multinomial estimator for $\Delta$

$$\hat{\Delta} = \sum \hat{V}_i,$$

where $\hat{V}_i = \sum_{j=0}^{m} a_j s_{i,j}$

Rao (1965) and Good (1969) assume that these summations one can and that the unbiased estimator

The diversity

$$\Delta = \sum \hat{V}_i,$$

given in Patil and Rao (1967) and for $B = 1, 2, \ldots$ are Shannon's information indices which do not have unbiased estimators.

An unbiased estimator of the argument given in Patil and Rao (1967) is a polynomial of degree $m$ and $\hat{\Delta}_m$ denote the $m$-th sub-samples of size $n$.

$$\text{Var}(\hat{\Delta}) = \sum \hat{V}_i$$

Depending on the $n$ by (5.6) can be replaced by (5.7). Section 3, it is a variance when same methods can be used (1979). This is due to sub-samples from the $\hat{\Delta}_m$ and $\hat{\Delta}_n$; the other an estimate of equi-

Since Shannon's mator, a natural way to produce a diverse
Since the multinomial distribution family is complete the MVU
estimator for $\Delta$ is

$$\hat{\Delta} = \sum \hat{v}_i$$

where

$$\hat{v}_i = \sum_{j=0}^{m} a_j \binom{N}{j} \binom{i}{j},$$

(5.4)

Rao (1965) and Good (1953). After some manipulation of binomial
sums one can show that $s(m)$ has the form given in (5.3)
and that the unbiased estimator (2.2) is the same as (5.4).

The diversity measure

$$\Delta_B = \sum \pi (1-\pi^\beta) / \beta$$

(5.5)

given in Patil and Taillie (1979) satisfies the unbiased condition
for $\beta = 1, 2, \cdots m < N$. However, for $\beta = 0$ and $\beta = -1$, which
are Shannon's information and the species index, respectively,
$\Delta_B$ does not have an unbiased estimator.

An unbiased estimator for the variance of $\hat{\Delta}$ is obtained by
the argument given in Smith and Grassie (1977). If $V(\pi)$ is a
polynomial of degree $m$, $2m < N$, and the random variables $\hat{\Delta}_m$
and $\hat{\Delta}_m'$ denote the values of the index $\Delta$ in two non-overlapping
subsamples of size $m$ from the original sample $N$, then

$$\text{Var}(\hat{\Delta}) = \text{Cov}(\hat{\Delta}_m, \hat{\Delta}_m') |_N.$$  

(5.6)

Depending on the form of the polynomial, the calculation implied
by (5.6) can be rather complicated; however, as illustrated in
Section 3, it is important to obtain the unbiased estimator of
variance when sample size is small. Monte Carlo simulation
methods can be used to obtain good estimates of (5.6) (Simberloff,
1979). This is done by repeatedly drawing two nonoverlapping
subsamples from the original sample, $N$, and computing
$\hat{\Delta}_m$ and $\hat{\Delta}_m'$; their sample covariance over the repeated trials is
an estimate of equation (5.6).

Since Shannon's information does not have an unbiased esti-
mator, a natural step would be to modify Shannon's information to
produce a diversity measure that has an unbiased estimator. Blyth
(1958) substituted a Taylor series expansion of $(x+b) \log(x+b)$,
for $\prod_{i=1}^{k} \log \pi_i$; however, this expansion includes a term with $k$, the number of species in the population. If $k$ is unknown, then one cannot obtain an unbiased estimator for the series expansion.

Another approach is to use Brillouin's (1962) information for a finite population of size $m$:

$$B_m = \frac{\prod_{i=1}^{k} (m_i!)}{m^{k}} \log \left( \frac{1}{m} \right),$$

where $m_i$ is the number of individuals of species $i$ in the finite sample of size $m$. For a fixed number of individuals, $m$, we can define the expected Brillouin information when $m$ individuals are sampled at random from the multinomial population,

$$b(m) = E[B_m | \Pi] = (\log m! - \sum_{i=1}^{k} E[\log(m_i! | \Pi)])/m. \quad (5.7)$$

This can be written in the general form

$$b(m) = \sum_{i=1}^{k} V_b(\pi_i),$$

where

$$V_b(\pi_i) = \frac{\pi_i \log m! - \sum_{j=2}^{m} \log(j!)(m/j)^{\frac{j}{j}(1-\pi)^{m-j}}}{m}. \quad (5.8)$$

After some manipulation one can show that $V_b(\pi)$ is concave on the interval $(0,1)$ and that $V_b(\pi)/\pi$ is a monotone decreasing function on $(0,1)$. (5.6) and (5.4) can then be applied to the general results to obtain the estimators for $b(m)$ and $\text{Var}[b(m)]$.

The class of diversity measures with unbiased estimators is still large enough to provide a considerable freedom in choosing a diversity measure. It is clear, however, that we must drop or modify the additivity principle for information (Pielou, 1975) if one is to obtain a measure that has an unbiased estimator.

6. SPECIES ARE

In Sections 2 through 5 we from a multinomial population, nominal populations $\pi_1, \pi_2, ..., \pi_k$, for the sample statistic $N$ to that the species identification of every other individual in the communities this assumption never, in other communities, par competition for space within a that individuals within a sample to be independent. Under these Simpson's index or Shannon's diversity is unclear since the definition on the multinomial model.

One solution to the problem in terms of the number selected quadrats from the community. Engen (1976), Holthe (have discussed species area cur the basic estimation techniques used to find estimators for the.

Suppose one samples a comm $N$ random quadrats of a fixed a denote the presence or absence $D_{ij}$ is equal to zero if species The vector $D_j = (D_{1j}, D_{2j}, ..., D_j)$ of all $K$ species in quadrat combinations of species within a distribution of $D_j$ is then defined $\Pr(D_j = d) = \pi(d)$. From these probabilities we can species $i$ in a random quad species $i$ and $j$ are in a random number of species in a random denote its mean.

$$s_q(m) = E[S_m | \Pi] = \sum_{i=1}^{K} \pi_i$$
6. SPECIES AREA RELATIONSHIPS

In Sections 2 through 5 we assumed that samples were drawn from a multinomial population, \( \pi \), or from a sequence of multinomial populations \( \pi_1, \pi_2, \ldots, \pi_n \). The basic assumption needed for the sample statistic \( N \) to have a multinomial distribution is that the species identification of each individual is independent of every other individual in the sample. For deep sea benthic communities this assumption seems to hold for most species. However, in other communities, particularly plant communities, competition for space within a sampling unit, may be such that individuals within a sampling unit cannot be assumed to be independent. Under these conditions the usefulness of Simpson's index or Shannon's index in describing community diversity is unclear since the definition of both these measures depends on the multinomial model.

One solution to the problem described above is to define diversity in terms of the number of species in \( m \) randomly selected quadrats from the community to produce species area curves. Engen (1976), Holthe (1975), Krylov (1971) and others have discussed species area curves. In this section we show that the basic estimation techniques developed in Section 2 can be used to find estimators for the species area curve.

Suppose one samples a community containing \( K \) species with \( N \) random quadrats of a fixed area. Let the random variable \( D_{ij} \) denote the presence or absence of the \( i \)th species in quadrat \( j \). \( D_{ij} \) is equal to zero if species \( i \) is absent, one if present.

The vector \( D_j = (D_{i1}, D_{i2}, \ldots, D_{iK}) \) denotes the presence or absence of all \( K \) species in quadrat \( j \). There are \( 2^K \) possible combinations of species within a single quadrat. The sampling distribution of \( D_j \) is then defined by the \( 2^K \) probabilities:

\[
\Pr(D_j = d) = \pi(d). \tag{6.1}
\]

From these probabilities we can find \( \pi_i \), the probability that species \( i \) is in a random quadrat and \( \pi_{ij} \), the probability that species \( i \) and \( j \) are in a random quadrat. Let \( S_m \) denote the number of species in \( m \) randomly selected quadrats and let \( s_q(m) \) denote its mean.

\[
s_q(m) = E[S_m | \pi] = \sum_{i=1}^{K} \left( 1 - (1 - \pi_i)^m \right). \tag{6.2}
\]
Suppose we have a random sample of $N$ quadrats from a community, $D_1, D_2, \ldots, D_N$; these denote the species composition of the $N$ quadrats. Let the sample statistic $L_i$ denote the number of stations with species $i$ present and let $L_{ij}$ denote the number of stations with both species $i$ and $j$ present. The unbiased estimators can be defined in terms of the sample statistics $L_i = (L_1, L_2, \ldots, L_i)$ and the $K \times K$ matrix $L = [L_{ij}]$.

\[
\hat{s}_q(m) = E[S_m | L_1^{(1)}, L_2^{(2)}] = \sum_{i=1}^{K} \left( 1 - \frac{\binom{N-L_i}{m}}{\binom{N}{m}} \right). \tag{6.3}
\]

Using the methods developed in expected species diversity, we can compute the estimator for the sampling variance of $\hat{s}_q(m)$.

Let the random variable $S'_m$ and $S''_m$ denote the number of species in two non-overlapping samples of size $m$ from the original sample of $N$ quadrats, $N \geq 2m$. The estimator for the variance is:

\[
\hat{\text{Var}}[\hat{s}_q(m)] = -\text{Cov}[S'_m, S''_m | L_1^{(1)}, L_2^{(2)}]. \tag{6.4}
\]

The computational form for the variance estimate is rather different from the one for Hurlbert's expected species. Details are given in Kravitz (in preparation). Simulations methods similar to Simberloff's (1979) methods could be used here to estimate equation (6.4).

The species quadrat method can be applied to the small deep sea benthic core data presented in Section 3. Here we want to know how well we can estimate the expected number of species in $m \text{ cm}^2$ deep sea cores and the given $N$ cores. The results are presented in Table 2.

### 7. CONCLUSIONS

Many diversity measures meet the general theoretical requirements for measuring diversity. In this paper we have suggested an additional statistical condition: that the estimator of diversity have well defined statistical properties. In particular we have suggested a more restricted class of diversity measures which

<table>
<thead>
<tr>
<th>Number of Cores (m)</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>have unbiased estimators been used for the diversification is the simplest and simplest for us to obtain a simple est</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This paper has dealt with a problem of Engen (1977) and Taylor, diversity measures based on the same question asked in a different context applied to the parametric log series model have unbiased estimators. Developed in Section 2, the difficult computations for the log series distribution difficult to compute.

This research was supported by grant OCE76-21968 and NOA.


TABLE 2: Species-core estimates for 16 deep sea benthic cores.

<table>
<thead>
<tr>
<th>Number of Cores (m)</th>
<th>Expected Number of Species ( s^q(m) )</th>
<th>Standard Error of ( s^q(m) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.62</td>
<td>.59</td>
</tr>
<tr>
<td>2</td>
<td>11.47</td>
<td>.94</td>
</tr>
<tr>
<td>4</td>
<td>18.54</td>
<td>1.47</td>
</tr>
<tr>
<td>8</td>
<td>28.22</td>
<td>2.18</td>
</tr>
</tbody>
</table>

have unbiased estimators. Other statistical criteria might have been used for the diversity problem; however, the unbiased criterion is the simplest and most straightforward to use and allows us to obtain a simple estimator for the sampling variance.

This paper has dealt with the non-parametric model, the sample statistic \( N \) has the general multinomial distribution. Engen (1977) and Taylor, Kempton, and Woiwod (1976) have suggested diversity measures based on the parametric log series model. The same question asked in the non-parametric case can of course be applied to the parametric case: namely, what properties of the log series model have unbiased estimators? The class of diversity measures developed in Section 5 has unbiased estimators with respect to the log series model; however, in practice they may be difficult to compute.

ACKNOWLEDGMENTS

This research was supported by National Science Foundation grant OCE76-21968 and NOAA Sea Grant Contract 04-6-159-44016.

REFERENCES


[Received June 1]


[Received June 1977. Revised April 1979]