ANOTHER LOOK AT CONFIDENCE LIMITS FOR SPECIES PROPORTIONS

MARTIN A. BUZAS
Department of Paleobiology, Smithsonian Institution, Washington, D.C. 20560

In palaeontological investigations the number of individuals of a particular species, \( n \), is often expressed as a proportion of the total number of individuals, \( \sum n_i = n \), for all, S species.

This proportion, \( p = \frac{n}{n} \), is referred to by various authors as percent species, relative abundance, percentage abundance, species frequency, fractional abundance, and so on. Naturally, researchers are interested in the confidence limits that can be placed on these estimates, and in the number of individuals required to obtain them. The binomial distribution was used by Dryden (1931), Dennison and Hay (1967), Wright and Hay (1971), and Patterson and Fishbein (1989) for this purpose.

While the above authors realized that a single sample may not be representative of an entire locality, no replicates were utilized to ascertain how accurate the confidence limits obtained from a single sample were for a locality. In this study, confidence limits are compared for foraminifera samples using binomial and cluster estimates to help resolve this problem.

**BINOMIAL CONFIDENCE LIMITS**

The binomial standard error of the proportion, \( p = \frac{n}{n} \), is given by

\[
\hat{\sigma}_{\text{bin}} = \sqrt{\frac{pq}{n^2}}
\]

(1)

where \( p \) is the proportion of the ith species, \( n_i \) is the number of individuals of the ith species, \( n \) is the total number of individuals, and \( q = 1 - p \). The confidence limits are \( p - \pm \hat{\sigma}_{\text{bin}} \times t \).

Thus, the confidence limits are \( t = 1.96 \) (Cochran, 1963). Various values of \( t \) in percent at the 95 percent confidence level, for \( p \) in percent, and \( n \) are tabulated by Patterson and Fishbein (1989). Clearly, from (1) the largest value of \( \hat{\sigma}_{\text{bin}} \) occurs when \( p = 0.50 \).

In calculating the confidence limits using the binomial, we assume that the n individuals are chosen at random. When sampling large organisms in the field, this is possible. Often, however, we must sample some volume of sediment containing the individuals that we wish to enumerate. Using standard laboratory procedures, we can remove \( n \) individuals at random from the large number of individuals in a homogenized sediment sample. Providing the number of individuals in the sediment sample is large, (a few hundred), (1) will provide a reasonable estimate for the standard error of the species proportions in our laboratory sample.

**CLUSTER CONFIDENCE LIMITS**

The spatial distribution of individuals, whether living, dead, or fossil, is usually not random, but is instead aggregated (Buzas, 1968; Buzas and Gibson, 1990). The sediment may be sampled at random, but we must accept the particular individuals contained in the sediment sample. Consequently, more than one sediment sample must be taken to measure the variability between samples to ascertain the confidence limits for a locality.

If \( m \) sediment samples are chosen at random and the \( j \)th sample contains \( a_i \) individuals of a particular species and \( n_i \) total individuals, then the cluster standard error of the species proportion is given by

\[
\hat{\sigma}_{\text{clus}} = \sqrt{\frac{a_i q_i}{n_i^2}} + \frac{q_i}{m} \sqrt{\frac{t^2}{m-1}}
\]

(2)

where \( \sigma_{\text{clus}} = a_i/n_i, p = a_i/n_i, \) and \( n = n_i/m \). As before, the confidence limit, \( d_i \), is obtained by multiplying \( \hat{\sigma}_{\text{clus}} \) by \( t \).

To illustrate the different results obtained using binomial versus cluster estimations, foraminifera from \( m = 4 \) sediment samples were collected at a single locality in May 1977 at Link Port, Florida. The locality is about 1 m\(^2\) and at a depth of about 1 m. Each sample consisted of 5 ml of sediment, which was stained with rose Bengal and washed over a 63 \( \mu \) sieve. The number of living (stained) individuals of Quinquelocula, Elphidium, Ammonia, and Bolivina and total living individuals were counted. These counts are shown in Table 1. Using (1) and (2), the standard errors for the species proportions were calculated for binomial and cluster sampling (note, \( n \) in (1) now becomes \( 2n \), or \( m \)). By multiplying the standard errors by 1.96, 95 percent confidence limits, \( d_i \), were obtained and are shown in Table 2.

**DISCUSSION**

Table 2 shows the confidence limits for the proportion of Quinquelocula (\( p = 0.50 \)) using the binomial is \( \pm 0.02 \), while for the cluster method it is \( \pm 0.09 \). For Elphidium (\( p = 0.36 \)) the binomial confidence is also \( \pm 0.02 \), while the cluster confidence is \( \pm 0.07 \). Using the binomial confidence limits, we might conclude that the two taxa have different proportions. The cluster confidence limits, however, overlap and our confidence is shaken. When comparing species proportions between areas, the use of binomial confidence limits will lead the investigator to become more confident of differences than is warranted. As \( p \) decreases from 0.50, and becomes very small, the confidence limits decrease and the binomial and cluster estimates converge. Consequently, for Bolivina (\( p = 0.07 \)) little is gained by using (2), providing \( m \) is large.

Dryden (1931) used the binomial formula in a slightly different way to assess the reliability of mineral grain counts. He suggested a few hundred counts, perhaps 300, would suffice for most purposes. On the basis of Dryden's study, Phleger (1960) also suggested a count of 300 was sufficient for foraminiferal analysis. Because for the binomial distribution, \( 1 - q^2 \) is the probability of at least one success, Dennison and Hay (1967) and Wright and Hay (1971) pointed out that 300 individuals would be required to record with 95 probability the occurrence (at least one individual) of a species with \( p = 0.01 \). Over the years the number 300 has taken on an almost magical significance, and is used by most micropaleontologists in their analyses (Patterson and Fishbein, 1989).
TABLE 1—Numbers of individuals for each species, \( n \), and total number of individuals for all species, \( n \), in four sediment samples of 5 ml each from Link Port, Florida.

<table>
<thead>
<tr>
<th>Species</th>
<th>Quinqueloculina</th>
<th>Elphidium</th>
<th>Ammonia</th>
<th>Bolivina</th>
<th>Total individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>181</td>
<td>229</td>
<td>344</td>
<td>299</td>
<td>1,323</td>
</tr>
<tr>
<td>( n )</td>
<td>134</td>
<td>211</td>
<td>60</td>
<td>56</td>
<td>330</td>
</tr>
<tr>
<td>( n )</td>
<td>56</td>
<td>48</td>
<td>46</td>
<td>7</td>
<td>175</td>
</tr>
<tr>
<td>( n )</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Totals</td>
<td>1,323</td>
<td>1,303</td>
<td>215</td>
<td>194</td>
<td>3,045</td>
</tr>
</tbody>
</table>

In the present study, about 2,000 individuals were counted. Table 2 shows that at \( p = 0.11 \) (Ammonia) the binomial confidence is \( 0.0294 \), and the cluster \( 0.03 \). A random sample of 300 individuals having a species with \( p = 0.10 \) would have a binomial confidence limit of \( 0.0339 \), for 400 individuals the confidence would be \( 0.0294 \). Considering the variation documented in the field (Table 2), little appears to be gained by counting more when \( p > 0.10 \). Table 2 shows the binomial and cluster confidence limits are almost identical for Bolivina (\( p = 0.007 \)). Now from (1), as \( p \) decreases and \( n \) increases, \( p \) approaches \( \sigma_{\text{max}} \). At \( p = 0.10 \) and \( n = 300 \), \( 0.0122 \), and for \( n = 400 \), \( 0.0098 \). Therefore, at \( p = 0.10 \), our confidence limit equals the species proportion. For \( p = 0.001 \) and \( n = 1,000 \), \( p = \sigma_{\text{max}} \). For practical purposes, then, when \( p < 0.01 \), confidence limits are meaningless.

When species proportions are less than \( 0.01 \), confidence limits are no longer meaningful.

ACKNOWLEDGMENTS

I thank Karen Carle for counting the foraminifera and Lee-Ann Hayek, W. D. Liddell, and R. C. Wright for helpful comments on the manuscript. This is contribution number 254 from the Smithsonian Marine Station at Link Port, Florida.

REFERENCES


Accepted 8 March 1990

The Smithsonian Institution provided $300 in support of this article.