CONCENTRATION CURVES AND HAVE-STATISTICS FOR ECOLOGICAL ANALYSIS OF DIVERSITY:

PART III: COMPARISON OF MEASURES OF DIVERSITY

Dianne G. Goodwin
James W. Vaupel

December 1985
WP-85-91

NOTE: This is Part III of a series of three working papers. For a preface, forward, acknowledgements, and a note about the authors, please see Part I of the series, which is subtitled "Dominance and Evenness in Reproductive Success", IIASA WP-85-72.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

Contents

CONCENTRATION CURVES AND HAVE-STATISTICS FOR ECOLOGICAL ANALYSIS OF DIVERSITY:

PART III: COMPARISON OF MEASURES OF DIVERSITY

Dianne G. Goodwin and James W. Vaupel

ORIENTATION

Given the central importance of diversity in ecology and the life sciences more generally, it is not surprising that a variety of methods and measures have been developed to describe and summarize diversity. In the two previous parts of this series of papers, comparisons were drawn between concentration curves and frequency distributions, the most widely used graphical display of variation, and between concentration curves and dominance-diversity curves. This final part of the three paper series compares various statistics that might be used to summarize diversity, with a focus on the usefulness of have-statistics as a supplement to more traditional measures. The first section of our discussion lays out some reasonable criteria and principles that good measures of diversity should satisfy: some traditional measures violate at least one of the criteria; the have-statistics pass the hurdles and have some desirable properties in addition. We then illustrate the use of different measures by way of examples drawn from Howard's studies of bullfrogs (discussed in Part I), the study of species diversity among diatoms (discussed in Part II), an analysis of mating systems of various birds, and a survey of human fertility in 41 countries.

PRINCIPLES FOR JUDGING MEASURES OF EVENNESS

The literature on measures of diversity is so vast and chaotic (see, e.g., Hurlbert 1971, Patil and Taillie 1982, Peet 1974, and Rao 1982b for overviews) that it is impossible to make headway without a clear goal and some principles of navigation. Our goal is to try to gain some understanding of the uses and limitations of have-statistics and concentration curves by comparing them with other kinds of measures of diversity. We will base this comparison on some principles and desirable properties of diversity measures.

For our purposes, it is convenient to begin with the aspect of diversity known as evenness. There is widespread agreement among researchers who have thought about the principles that a measure of evenness should satisfy (e.g., Marshall and Olkin 1979, Foster 1985) that the following four principles are reasonable. For expository simplicity, we use x to mean individual, species, or any other "have" and we use y to mean offspring, zygotes, mates, members of a species, or any other "had".

1. The Anonymity Principle

Consider any two x's in a population. Suppose they can be identified: for convenience, call one Harry and the other Larry. Suppose one has y and the other has y'. An evenness measure should not change if Harry is the one with y rather than the one with y'.

2. The Relativity Principle

Evenness should depend only on the relative amount, i.e., the proportion of the total, each x has and not on the absolute amount. Consider, for instance, a population with two x's, one having 70% of the y's and the other having 30%. Evenness should be the same regardless of whether the total number of y's is ten, a thousand, or a million.

3. The Replication Principle

Suppose a population is replicated so that there are now m identical populations. Suppose the m populations are combined into a new population with m times as many x's. The evenness of this new population should be the same as the evenness of the original population. For instance, suppose the original population consists of two x's, one having 70% of the y's and the other having 30%. After a single replication and combination, the new population will consist of four x's, the top half (i.e., top two) having 70% of the y's and the bottom half having 30%. Evenness should remain the same.

4. The Transfer Principle

Consider any two x's in a population such that one has y and the other has somewhat fewer, $y-d$. Suppose a transfer is made such that the first x now has even more, $y+c$, and the second x has even fewer, $y-d-c$. According to the Fully-Responsive Transfer Principle, an evenness measure should decrease. According to the Partially-Responsive Transfer Principle, an evenness measure should not increase and there should be at least one pair of x's such that such a transfer would result in a decrease in the evenness measure. Note that the Fully-Responsive Transfer Principle implies a transfer from a "poor" x to a very rich x should decrease evenness by more than an equal transfer to a not-so-rich x, and the Partially-Responsive Transfer Principle implies that such a transfer from the poor to the very rich should decrease evenness by at least as much as a transfer to the less rich.

A measure of evenness that is consistent with principles 1, 2, and 3 and with the partially-responsive version of principle 4 might be called a consistent measure. A measure consistent with 1, 2, and 3 and the fully-responsive version of 4 might be called a strictly-consistent measure. As documented below, three of the most commonly used "measures of evenness" are neither consistent nor strictly-consistent.

It turns out that the four principles have a close relationship with concentration curves, as follows. If and only if the concentration curve for one population lies between the curve for another population and the diagonal line at all points between 0 and 1, will any consistent measure of evenness indicate that the evenness of the first population is greater than the evenness of the second population. Hence, if one concentration curve lies under another it can be said that the first population is definitely more even (regardless of the measure of evenness used) than the second population. This is one of the key reasons that concentration curves are so useful and why the central role of concentration curves in the analysis of evenness is, as Allison (1978) put it, "virtually unquestioned" by economists, other social scientists, and mathematicians who have studied inequality.

If one concentration curve is lower than a second curve at some points but at other points the curves either touch or run along together, then all strictly-consistent measures of evenness will indicate that the first population is more even than the second. Depending on the measure and on where the curves touch, a measure that is consistent but not strictly-consistent may indicate either that the two populations are equally even or that the first population is more even than the second. Consequently, it is possible to reformulate the criteria for a measure of evenness as follows:

- a measure of evenness is strictly-consistent if and only if the measure gives a lower value of evenness to a concentration curve that lies outside another concentration curve at at least some points and never lies inside the other concentration curve.
- a measure of evenness is consistent if and only if the measure gives a lower value of evenness to a concentration curve that lies outside of another concentration curve at all points between 0 and 1 and gives a lower or equal value of evenness to a concentration curve that either lies outside or touches another concentration curve at all points.

These two criteria might be called the concentration-curve criteria.

In our empirical analyses, on occasion the concentration curves crossed over. In these cases two different summary measures may give the two populations a different ordering: according to some measures, the first population may be more concentrated and according to other measures, the second population may be more concentrated or, at least, equally concentrated. The differences in the measures were small and when a number of populations were considered the rankings according to different measures tended to be more or less the same. This highlights the importance of concentration curves themselves as compared with any particular summary measure.

WHICH MEASURES ARE CONSISTENT?

Many measures have been used to capture the evenness of a population; we consider only the most commonly used measures, as well as various have-statistics. Among these measures the following distinctions can be made:

- The havehalf, haveall, and other have-*x* measures are consistent measures of evenness.
- The halfhave and other *y*-have measures are consistent measures of unevenness. That is, these measures are consistent with the four principles of evenness, except that the measures decrease as evenness increases.

- -- The havenone is a consistent measure of unevenness.
- -- The Gini coefficient, which is usually defined as the proportion of the area above the diagonal line that lies between the diagonal line and a concentration curve, is a strictly-consistent measure of unevenness. An alternative expression for the Gini coefficient is

$$\frac{\sum_i \sum_j |p_i - p_j|}{2N} ,$$

where p_i is the proportion of total y's attributable to the i'th x and N is the number of x's.

The coefficient of variation and Crow's I, which equals the square of the coefficient of variation, are both strictly-consistent measures of unevenness. Crow's I is usually defined as the ratio of the variance in number of offspring divided by the square of the mean number of offspring. This ratio reduces to:

$$N \sum_i p_i^2 - 1 .$$

The core of this expression, it might be noted, is Simpson's well-known index of dominance:

$$\sum_i p_i^2 .$$

Another expression for Crow's I is:

$$\frac{\sum_i \sum_j (p_i - p_j)^2}{2} .$$

This expression is intriguingly analogous to the formula for the Gini coefficient.

- -- One of the entropy measures proposed by Thiel, namely

$$\ln N + \sum_i p_i \ln p_i ,$$

(where ln denotes the natural logarithm) is a strictly-consistent measure of unevenness.

- -- The most commonly used "measure of evenness", Pielou's J', is not a consistent measure of evenness. The measure J', given by

$$J' = \frac{H'}{H'_{\max}} ,$$

where H' is Shannon's measure of information (or entropy),

$$H' = -\sum_{i} p_i \ln p_i$$

and

$$H'_{\max} = \ln N \ ,$$

violates the replication principle. Consider, for instance, the following example. A population consists of two x's with 80% and 20% of the y's, respectively. Another population consists of twenty x's, the first ten having 8% of the y's each and the second ten having 2% of the y's each. The second population clearly can be created by replicating the first population ten times. In both populations the top half have 80% and the bottom half have 20% of the y's, and the concentration curves for the two populations are identical. However, J' for the first population is 0.72, whereas it is 0.94 for the second population. In general, for any distribution of y's among the x's, if the number of x's increases but the concentration curve remains the same (i.e., as a population is increasingly replicated), J' will asymptotically approach 1. It might be noted that although Thiel's entropy and J' are both simple transforms of Shannon's measure of entropy, Thiel's entropy is a strictly-consistent measure of evenness whereas J' violates the replication principle.

-- Pielou's J is also not a consistent measure of evenness. J is defined by

$$J = \frac{H}{H_{\max}} \ ,$$

where H is Brillouin's measure of information (or entropy),

$$H = \frac{\ln\left(\dfrac{K!}{k_1!k_2! \cdots k_N!}\right)}{K}$$

where K is the total number of y's and k_i is the number of y's of the i'th x. The formula for $H_{\max}$, which is the maximum value H can attain, can be found in Pielou (1969, p. 233). It is not difficult to show that J violates both the relativity principle and the replication principle. Consider, for instance, the following three populations:

1) a population with two x's, one with 5 y's and the other with 1 y.

2) a population with two x's, one with 50 y's and the other with 10 y's.

3) a population with twenty x's, half with 5 y's each and half with 1 y each.

The concentration curves for these three populations are identical and any consistent measure of evenness should be the same for all three. The value of J, however, is 0.60 for the first population, 0.64 for the second, and 0.92 for the third. Peet (1974) provides another example demonstrating that J violates the relativity principle.

— McIntosh's "index of evenness" (McIntosh 1967; Pielou 1969), which we will denote by Mc, is not a consistent measure of evenness because it violates the replication principle. The index can be expressed as:

$$Mc = \frac{1 - \sqrt{\sum_i p_i^2}}{1 - \frac{1}{\sqrt{N}}} .$$

As an example, consider a population which consists of two x's, with 90% and 10% of the y's respectively. Suppose this population is replicated ten times to produce a population in which the top ten x's each have 9% of the y's and the bottom ten each have 1%. Evenness should be the same in both cases, but McIntosh's index is 0.32 for the first population and 0.92 for the second.

SOURCES OF CONFUSION

The three measures that are not consistent measures of evenness, J', J, and Mc, were all derived by Pielou by standardizing a measure of diversity—Shannon's entropy, Brillouin's entropy, and McIntosh's index of diversity, respectively--so that the standardized measure ranges from zero, when one x has all the y's, to one, when all x's have the same number of y's. This approach is unsatisfactory on three counts.

First, standardization of a measure of diversity does not guarantee that the resulting measure will be a consistent measure of evenness. If standardization is desired, the correct approach is to appropriately standardize a consistent measure of evenness: the resulting measure will then also be a consistent measure of evenness.

Second, standardizing a measure so that its range is from zero to one does not imply that the measure itself is independent of N (i.e., the number of x's). Such standardization merely implies that the ***range*** of the measure is independent of N. Pielou argues that a measure of evenness should be independent of N, but her measures are not. The replication principle is a way of defining and operationalizing the idea that evenness should not vary across populations of different sizes N that are identical in their distribution of the y's. The measures J', J, and *Mc* all violate the replication principle.

Third, a measure that is standardized so that its range stretches from zero to one (or any other interval that does not depend on N) will necessarily be inconsistent with the replication principle, i.e., with the idea that evenness should not depend on population size. A population in which one x out of two has all the y's is more even than a population in which one x out of 20 has all the y's because:

1. according to the replication principle, a population in which one x out of two has all the y's is just as even as a population in which 10 x's out of 20 have all the y's, and
2. according to the transfer principle, a population in which 10 x's out of 20 have all the y's is more even that a population in which one x out of 20 has all the y's.

Thus, a measure that gives all populations the same value when one x dominates cannot be a consistent measure of evenness. On the other hand, a consistent measure must always give the same value to populations that are perfectly even (because of the replication principle). A consistent, reasonable and intelligible way to standardize a measure of evenness is to set the measure equal to 1/N when one x out of N has all the y's and set the measure equal to 1 when all x's have equal numbers of y's.

The measure J, which is a standardized version of Brillouin's entropy H, violates both the replication principle and the relativity principle. It violates the relatively principle because it depends on the total number of y's. Pielou argues that J should be used for fully-censused collections whereas J', which is a standardized version of Shannon's entropy H', should be used for samples from very large communities. She bases this position on three notions:

1. In information theory, Shannon's entropy "is strictly defined only for an infinite population", whereas Brillouin's entropy is appropriate for messages of finite length (Pielou 1969, p. 231). However, as Pielou notes, "analogies with

information theory ... do not, of course, provide a compelling reason for using H' and H in the way just outlined" (Pielou 1975, p. 10). Indeed, why should the diversity or evenness of a population be measured the same way as the information content of a message or code?

2. "A value of H is determined from a complete census and hence is free of statistical error whereas a value of H' is estimated ... and thus has sampling error; estimates of H' should always be accompanied by estimates of their standard errors" (Pielou 1975, p. 11). Any measure of evenness that is estimated has a sampling error and could be accompanied by an estimate of its standard error. Thus, Pielou's argument here is simply an argument for calculating standard errors and not an argument in favor any particular kind of measure.

3. "If, from an indefinitely large community, we take two samples, one small and one large, and treat both as collections, the small collection would be expected to have a lower value of H than the large collection. This result accords with what we intuitively require of a diversity index..." (Pielou 1975, p. 11). The underlying idea here, as we understand it, is as follows. Consider a community that consists of a large number of different species (the x's), some of which have large populations of individuals (the y's) and some of which have small populations. If a small sample is taken, many of the rare species are likely to be missed. Hence the diversity of the sample will tend to be less than the diversity of the entire community. This, however, does not imply that an index of diversity—or an index of evenness—should tend to decrease with the size of the sample: indeed, such variation would violate the relativity principle. Rather, the diversity (or evenness) of the sample should be summarized by a measure that is consistent with underlying principles. If it is desired, an estimate of the diversity of the entire community might then be made, using appropriate statistical methods for drawing inferences about a universe from a sample. It might be possible to develop a short-cut approach to estimating the diversity of the entire community: in such an approach, a measure of estimated diversity would have to tend to *increase* as sample size decreased, in order to counterbalance the tendency for small samples to be less diverse than the entire community.

DESIRABLE PROPERTIES OF MEASURES OF EVENNESS

Although the "measures of evenness" most commonly used by ecologists are not consistent with the four principles of evenness or the concentration-curve criterion, a variety of consistent measures of evenness exist. In choosing amongst the alternatives, an analyst might want to consider how the measures compare according to some desirable properties. We consider five such properties below: standardization, intelligibility, decomposibility, sensitivity, and robustness.

1. Standardization

As discussed above, it is often desirable to use standardized measures of evenness that range in value from $1/N$ when one x out of N has all the y's to 1 when all x's have the same number of y's. Only one of the measures discussed above has this property, the haveall statistic. It is not difficult, however, to standardize other measures.

-- The havehalf ranges from $1/2N$ to $1/2$. Thus, twice the havehalf (a measure which we will refer to as the double-havehalf) is a standardized measure of evenness as is, more generally, x times any have-x statistic.

-- The Gini coefficient ranges from $1-1/N$ to 0. Hence the complement of the Gini Coefficient (i.e., one minus the Gini coefficient) is a standardized measure of evenness. By analogy to terms such as cosine and colog in which co indicates complement, we will call this measure the co-Gini index of evenness.

-- Crow's I ranges from $N-1$ (for a population in which one x has all the y's) to 0 for a perfectly even population. Hence,

$$\frac{1}{I+1} = \frac{1}{N \sum_i p_i^2}$$

is a standardized measure of evenness. This measure can also be interpreted as the inverse of N times Simpson's index of concentration. We will refer to it as the reciprocal-Simpson index of evenness.

-- Thiel's entropy varies from $\ln N$ to zero. One transformation that might be used to convert this measure into a standardized measure of evenness is:

$$e^{-(\ln N + \sum_i p_i \ln p_i)} = \frac{e^{-\sum_i p_i \ln p_i}}{N}$$

$$= \frac{1}{N \prod_i p_i^{p_i}} ,$$

(where, by convention, zero times the log of zero is taken as zero and zero raised to the zero power is taken as one.) Buzas and Gibson (1969) proposed this measure as a measure of evenness. It can also be derived by raising e to the Shannon index and then dividing by N. We will call this measure the exponential-Shannon index of evenness.

-- It might be noted that concentration curves are standardized in that the vertical and horizontal axes both run from 0 to 1. Thus, it is easy to compare the evenness of two populations merely by examining their concentration curves.

2. Intelligibility

A second desirable property of measures of evenness is intelligibility. Ideally, a measure should be easy to comprehend, intuitively meaningful, simple to explain to others, and naturally relevant to the problems being addressed. Although there is no disputing taste, and intelligibility is clearly a matter of taste and personal opinion, have-statistics, especially the havehalf and the haveall (or havenone), achieve these goals for us better than any other measures of evenness we are familiar with.

Gini's coefficient has a simple geometric interpretation on a concentration graph as the proportion of the area above the diagonal line that lies between the concentration curve and the diagonal line. Yet its biological interpretation is not directly clear. What does it *mean* if the Gini coefficient is .3 as opposed to .4?

Simpson's index of dominance,

$$\sum_i p_i^2 \quad ,$$

forms the core of two of the indices discussed above, Crow's I,

$$N \sum_i p_i^2 - 1 \quad ,$$

which is a measure of unevenness, and the "reciprocal-Simpson index",

$$\frac{1}{N \sum_i p_i^2} \quad ,$$

which is a standardized measure of evenness. Simpson's index can be interpreted as the probability that two randomly selected y's belong to the same x, e.g., the probability that two individuals in a population belong to the same species. This is a helpful, ecologically-relevant interpretation, but unfortunately the interpretation pertains to Simpson's index rather than to the measures of evenness themselves. Suppose, for instance, that the value of Crow's I was 9.26 and, correspondingly, that the value of the reciprocal-Simpson measure was 0.097. Without knowledge of N, it is impossible to convert these values into their Simpson equivalent and even if it was known that N was thirty-eight, say, the calculation of the value of 0.27 of Simpson's index takes a bit of effort.

Crow's I has a direct interpretation that has some ecological meaning. Define the "importance" of each x as the amount of y's that x has and, similarly, define the importance of each y as the total amount of y's the x that has that y has. Let X be the average of the first of these importance variables and let Y be the average of the second. In the case, for example, of a population of females having broods of children, X would be the average brood size per female and Y would be the average brood size per child. Then it can be shown that

$$Y = X(I+1) \quad .$$

If, as above, I is 9.26, then this implies that the average child has 10.26 times as many siblings (including itself) as the average mother has children. Such a situation could arise if most females have no children and if almost all children come from large families. The relationship between X and Y implies that in a stationary population females on average only have $1/(I+1)$ as many offspring as their own average brood size. Hence, in the example given, the average child would have less than a tenth the offspring her mother had—perhaps because more than nine-tenths of each birth cohort leaves no offspring. Preston (1976) provides an interesting discussion of the relationship, for humans, between family sizes of children and family sizes of women.

Latter (1980) argues that entropy "has many convenient properties from a mathematical point of view, but is extremely difficult to interpret genetically". We have not been able to find any helpful biological interpretations of any of the various entropy measures and have not been able to develop much of a feeling for what a Theil entropy of, say, 0.52 means.

3. Decomposibility

Theil's entropy, like various other measures of entropy, does, however, have the desirable property of decomposibility. Foster (1985) calls decomposition "the most useful property of the Theil entropy measure". Suppose a population is comprised of several groups. Then, as explained by Foster and by Theil (1972), it is possible to calculate a "within-group" entropy and a "between-group" entropy. The within-group entropy measures the average unevenness within the various groups; the between-group entropy measures the unevenness of the distribution where the group mean replaces each group member's y value. The desirable feature of Theil's entropy is that the value of Theil's measure for the entire population is simply the sum of the within-group and between-group measures.

As discussed by Foster (1985), it is also possible to decompose Crow's I (and some other measures described by Foster) into within-group and between-group components, although the decomposition is somewhat complicated. Patil and Taille (1982) also discuss a number of measures that can be decomposed. We have not yet investigated whether it is possible to find some useful decompositions based on various have-statistics nor have we explored the uses of decomposition in our studies.

4 and 5. Sensitivity and Robustness

A measure is sensitive if it responds to changes in the underlying data. If the data are known to be accurate, this is a desirable property. If, however, some of the data may be in error, a robust measure that is insensitive to errors is desirable. Hence, for some applications sensitive measures are preferable and for other applications robust measures are indicated. Some measures are sensitive to data in certain ranges--say in the middle of the overall range--and robust to data in other ranges--say at the extremes. In investigating some biological questions, it may be desirable to use a measure that is sensitive to prolific or dominant x's but robust to changes in x's that have little or no y's, but in other analyses the opposite may be the case--e.g., in studies where the rare species with small populations are of great interest. A good introduction to the concepts of sensitivity and robustness, illustrated by a comparison of the mean (which is a sensitive measure), the median (which is robust) and the mid-mean, the mean of the middle half of the data values, (which is sensitive to the middle range and robust to the extreme ends of the range), can be found in Tukey (1978).

Strictly-consistent measures of evenness or unevenness, like Gini's coefficient, Crow's I, or Thiel's entropy, are more sensitive to transfers of y's among the x's than are consistent measures like the havehalf, haveall, or halfhave. The haveall is an extreme case because it only depends on the proportion of x that have all the y's: the distribution of the y's among these x's is irrelevant. Similarly, other have-x and y-have statistics are insensitive to certain kinds of transfers among the x's. When field data in ecological studies may be subject to substantial error, this robustness of have-statistics may be a valuable property.

Although robust to certain kinds of transfers, have-statistics are sensitive to changes in the amount of y's any particular x has. A statistician might call the havehalf the ".5 fractile of the inverse right-hand concentration curve". The median is also a .5 fractile (of a frequency distribution), but the havehalf differs from the median in a key respect. The median is a robust statistic that will not change in value if any of the values of the frequency distribution are changed, except for the one or two middle values of the distribution. The value of the havehalf, on the other hand, changes if any single value of the underlying frequency distribution is altered. This sensitivity to changes in any of the values of the underlying frequency distribution holds for all the have-statistics except the haveall and its complement the havenone.

The sensitivity and robustness of different measures of evenness deserves further attention. A useful exercise would be to do some empirical calculations based on plausible changes and errors in a data set, to determine how different measures respond.

DIVERSITY

Ecologists use the word diversity in two different senses, one broad and the other narrower. In the broader sense, diversity refers to the differences among individuals in a population. In its narrower, more technical meaning, which is typically used in studies of species diversity, diversity is defined as a measure that captures both "evenness" and "richness", where richness is a measure of how many x's (e.g., species) there are in the population or community, that is, the variable we have been calling N. For example, Pielou (1975, p. 14) explains that "the diversity of a community depends on two things: the number of species and the evenness with which the individuals are apportioned among them." It seems reasonable that such a measure of diversity should satisfy three of the principles laid out at the beginning of this paper, namely the anonymity, relativity, and transfer principles. Instead of the replication principle used for a measure of evenness, the following replication principle might be used for a measure of diversity.

The Replication Principle for a Measure of Diversity:

Suppose a population is replicated so that there are now m identical populations. Suppose the m populations are combined into a new population with m times as many x's. The diversity of this new population should be greater than the diversity of the original population. Furthermore, the bigger m is, the larger the diversity should be.

This principle, plus the other three principles, implies that a diversity measure D can be considered a function of N (the number of x's) and E, where E is some consistent or strictly-consistent measure of evenness. (D, in these two cases, might be called a consistent or a strictly-consistent measure of diversity.) The functional relationship is defined by:

$$D = f(N,E) \quad ,$$

where

$$f(N+1,E) > f(N,E) \quad , \text{ any } N \text{ , and any } E \quad ,$$

and

$$f(N,E+a) > f(N,E) \quad , \text{ any } N \text{ , any } E \text{ , and any } a > 0 \quad .$$

That is, diversity should increase if either richness N or evenness E increases.

Numerous functions f satisfy these criteria, but one seems particularly appropriate, at least for ecological applications:

$$D = N E \quad .$$

Diversity in this case is simply richness times evenness, where richness is measured by N and evenness is measured by any consistent or strictly-consistent measure. If evenness is standardized to vary from $1/N$ when one x dominates to 1 when all x's have equal y's, then this measure of diversity varies from 1 to N. Such a measure of diversity can be interpreted as the "equivalent number" (MacArthur

1965), "effective number", or "even number" of x's, i.e., the number of x's that in a situation of complete evenness would produce the same diversity as the actual diversity. For instance, consider a community of different species with differing populations. If N, the number of species, is 120 and E, the evenness, is 0.5, then the diversity would be 60 and it might be said that the community has a diversity equivalent to the diversity of a community with 60 equally numerous species.

The concept of diversity as the product of N and E is so natural that the following principle may seem appropriate:

The Proportional Replication Principle:

Suppose a population is replicated so that there are now m identical populations. Suppose the m populations are combined into a new population with m times as many x's. The diversity of this new population should be m times the diversity of the original population.

This principle, together with the other three principles, implies that diversity should be measured as the product of N and some measure of evenness. Such a diversity measure might be called a "proportional" measure.

MEASURES OF DIVERSITY

Numerous measures of diversity have been proposed and it is beyond our scope to review more than a few of the most widely known measures as well as some have-statistic measures. Among these measures the following distinctions can be made:

— N times a have-y measure is a consistent, proportional measure of diversity. Such a measure can be interpreted as the number of x's that account for the specified proportion of the y's. Thus the $N \cdot$ havehalf is the number of x's that have half the y's. Twice the $N \cdot$ havehalf is a standardized measure of diversity that varies from one to N, and may be interpreted as the number of x's in an even population with the same diversity as the actual population. We will refer to this measure as the double-halfhave measure of diversity.

- N times the haveall is a standardized, consistent, proportional measure of diversity that gives the number of x's that have any y's. For instance, in the case of females having offspring, if the haveall is .80 and N is 50, then 40 females had offspring. In studies of the diversity of species in a community, where every species included has to have, by definition, at least one member, the $N \cdot$ haveall is simply equal to N: in this special case, this measure of diversity coincides with the measure of richness.
- Simpson's index of concentration is a strictly-consistent, proportional measure of concentration and its inverse is a standardized, strictly-consistent, proportional measure of diversity, which might be called the reciprocal-Simpson index of diversity.
- N times the complement of Gini's coefficient is a standardized, strictly-consistent, proportional measure of diversity, which might be called the co-Gini index of diversity.
- Shannon's entropy,

$$-\sum_i p_i \ln p_i \quad ,$$

is a strictly-consistent measure of concentration, but it is not standardized or proportional.

- By multiplying the exponential-Shannon index of evenness by N, the following standardized, strictly-consistent, proportional measure of diversity, which might be called the exponential-Shannon index of diversity, can be derived:

$$\prod_i p_i^{-p_i} \quad .$$

- Brillouin's entropy is not a consistent measure of diversity because it violates the relativity principle. Peet (1974) has noted that "... the Brillouin formula does not provide an acceptable index of heterogeneity".
- The variance is not a consistent measure of diversity because it violates the relativity principle.

APPLICATIONS

To illustrate some of the points made above about alternative measures of evenness and diversity, we provide three examples.

1. Lifetime Reproductive Success of Male vs. Female Bullfrogs

Table 1 presents various summary statistics pertaining to lifetime reproductive success of male vs. female bullfrogs. The original data are from Howard (1983); in Part I of this series of working papers, in Figure 9, concentration curves were drawn based on these data. Scrutiny of the two curves and the various summary statistics might help the interested reader form his or her own judgments of the merits of concentration curves and of different summary statistics.

2. Diversity in a Community of Herbaceous Plants

Table 2 is similar to Table 1 and has a similar purpose. It presents various summary statistics pertaining to the diversity of herbaceous plants in a deciduous woodlot, as described in Part 2 of this series of working papers, in conjunction with Figure 1. Note that whereas Table 1 includes various measures of evenness, Table 2 presents alternative measures of diversity. Only those measures of diversity that were discussed above and that seem particularly useful are included. All of the standardized measures of diversity given in the table correspond to N times a standardized measure of evenness.

3. Birds of a Feather

An example of the use of summary measures of evenness for ecological analysis is Payne and Payne's (1977) comparison of the distribution in mating success of male birds in different mating systems. Payne and Payne argue that "mating systems and the statistics of mating success among males are closely related" and that measures that summarize the evenness in the distribution of mating success among individual birds are useful tools in describing and comparing the mating systems of populations (Payne and Payne 1977, p. 165).

Payne and Payne use three particular measures of "evenness" in their study, namely the coefficient of variation, Pielou's evenness index and the coefficient of skewness. As we have noted, the coefficient of variation (often abbreviated *CV*), which is the square root of Crow's I, is a strictly-consistent measure of uneven-

Table 1. Summary statistics for the predicted lifetime reproductive success of male and female bullfrogs.

Standardized measures of evenness:	Males	Females
Haveall	.31	.46
Double-Havehalf (2 · Havehalf)	.13	.20
Co-Gini (1-Gini)	.18	.26
Reciprocal (1/N · Simpson)	.18	.28
Exponential-Shannon ($e^{\text{Shannon Index}}/N$)	.23	.34
Have-statistics:		
Havequarter	2.28%	4.71%
Havehalf	6.50%	10.17%
Quarterhave	92.68%	85.42%
Halfhave	100.00%	100.00%
Havenone	69.00%	54.00%
Measures of unevenness:		
Crow's Index	4.68	2.56
Gini's Coefficient	.82	.74
Thiel's Entropy	1.47	1.08

ness. Payne and Payne (1977, p. 167) note "monogamous and polygynous species overlap perhaps less in *CVs*, and this may be the single statistic best describing the distribution of mating success in different mating systems." Pielou's evenness index has already been discussed; it is not a consistent measure of evenness. The coefficient of skewness is also a defective measure of evenness because it does not satisfy the transfer principle. For example, take three populations of three x's each. In the first population the distribution of y among the x's is 1,7,7. In the second population it is 4,5,6 and in the third it is 1,1,13. The first population has a skewness of -.707, the second a skewness of 0 and the third a skewness of .707. But the transfer principle implies that it is the middle population that is most even.

Table 2. Summary statistics of diversity of herbaceous plant species in a deciduous woodlot.

Standardized measures of diversity (i.e. Equivalent numbers of species)	Value
Haveall (richness; number of species)	62.
Double-Havehalf ($2 \cdot N \cdot$ Havehalf)	14.38
Co-Gini ($N \cdot$ (1-Gini))	21.18
Reciprocal-Simpson (1/Simpson)	21.10
Exponential-Shannon ($e^{\text{Shannon Index}}$)	27.92
Other measures of diversity:	
Shannon Index (Entropy)	3.33
Simpson's Dominance Index (of lack of diversity)	.05

Table 3 is styled after Payne and Payne's presentation and uses an accessable subset of their data sources. However, we relied on our own statistical calculations. For details of the specific data sets the reader should refer to Payne and Payne's article and the source material.[1] We present a variety of measures of evenness for the birds of different species, but omit Pielou's index and the coefficient of skewness because they are not consistent measures of evenness. We have not included the havehalf as it is simply one half of the double-havehalf.

The species of male birds are placed in three categories:

A. Those which generally form lek or dispersed lek mating systems in which males display, but form no pair bond and provide no parental care;

B. Those which form mating systems in which males are sometimes polygynous, form pair bonds and may provide some parental care, and

---

[1] A cautionary note is in order. As was demonstrated in Part I, the concentration of reproduction changes depending on the stage of reproduction considered and whether all, only the reproductively viable or only the reproductively successful animals are included. Also, concentration is reduced with the use of averaged data. The nature of the data must obviously be considered before any substantive conclusions can be drawn. The purpose of this example is illustrative.

Table 3. Distribution of mating success among male birds from species with different mating systems.

				STANDARDIZED MEASURES OF EVENNESS					MEASURES OF UNEVENNESS			HAVESTATISTICS				
	Type	Number of adult males	Mean no. matings, mates or fledglings/adult	Double-Havehalf	Havaall	Co-Gini	Reciprocal Simpson	Exponential Shannon	Crow's Index	Gini's Coeff.	Thiel's Entropy	Have quarter	Have none	Half have	Quarter have	Reference
Vidua chalybeata (Indigo)	A	14	1.0 matings observed	0.100	0.286	0.143	0.132	0.175	6.571	0.857	1.744	0.025	0.714	1.000	0.964	Payne&Payne 1977 (Cowpie,1976)
Vidua chalybeata (Indigo)	A	14	2.2 "	0.117	0.429	0.177	0.165	0.225	5.075	0.823	1.494	0.029	0.571	1.000	0.919	Payne&Payne 1977 (Junction,1973)
Manacus manacus (White-bearded manakin)	A	12	2.3 "	0.130	0.250	0.161	0.171	0.199	4.847	0.839	1.815	0.032	0.750	1.000	1.000	Lill 1974 (LekB,Group1,1967)
Manacus manacus (White-bearded manakin)	A	10	43.8 "	0.136	0.800	0.199	0.178	0.249	4.829	0.801	1.391	0.034	0.200	0.989	0.908	Lill 1974 (LekB,1968)
Tympanuchus cupido (Prairie chicken)	A	9	3.3 "	0.159	0.333	0.193	0.202	0.240	3.940	0.607	1.427	0.040	0.667	1.000	0.950	Robel 1966
Lyrurus tetrix (Black grouse)	A	6	4.0 "	0.235	0.500	0.278	0.294	0.344	2.396	0.722	1.069	0.059	0.500	1.000	0.833	Koivisto 1965
Pipra erythrocephala (Golden-headed manakin)	A	16	4.9 "	0.325	0.625	0.367	0.406	0.470	1.462	0.633	0.755	0.069	0.375	0.962	0.718	Lill 1976 (Lek4,1970)
Pipra erythrocephala (Golden-headed manakin)	A	13	6.7 "	0.390	0.769	0.438	0.468	0.546	1.135	0.562	0.605	0.076	0.231	0.925	0.615	Lill 1976 (Lek2,1969)
Telmatodytes palustris (Marsh wren)	B	25	3.8 fledglings	0.432	0.760	0.527	0.566	0.645	0.766	0.473	0.439	0.087	0.240	0.830	0.545	Verner 1965 per season
Agelaius phoeniceus (Red-winged blackbird)	B	51	2.5 "*	0.461	0.961	0.594	0.594	0.742	0.683	0.406	0.299	0.064	0.039	0.779	0.529	Holm 1973 *(Harem averages)
Agelaius phoeniceus (Red-winged blackbird)	B	53	3.6 "*	0.548	0.981	0.635	0.689	0.792	0.452	0.365	0.233	0.114	0.019	0.769	0.472	Holm 1973 *(Harem averages)
Melospiza melodia (Song sparrow)	C	15	4.3 "	0.587	0.933	0.698	0.763	0.841	0.311	0.302	0.173	0.124	0.067	0.703	0.445	Nice 1937
Agelaius phoeniceus (Red-winged blackbird)	B	53	3.0 female mates	0.613	0.981	0.714	0.794	0.670	0.259	0.286	0.140	0.140	0.019	0.709	0.425	Holm 1973
Lagopus lagopus (Red grouse)	C	72	5.2 size of fledged broods	0.639	0.917	0.703	0.784	0.834	0.276	0.297	0.181	0.144	0.083	0.712	0.406	Jenkins et al 1963 (1960,highlands)
Agelaius phoeniceus (Red-winged blackbird)	B	51	2.7 female mates	0.641	0.961	0.725	0.801	0.869	0.248	0.275	0.140	0.135	0.039	0.699	0.420	Holm 1973
Lagopus lagopus (Red grouse)	C	74	5.0 size of fledged broods	0.667	0.986	0.742	0.828	0.889	0.207	0.258	0.117	0.149	0.014	0.684	0.394	Jenkins et al 1963 (1960,lowlands)

MATING SYSTEMS:

A. Leks and dispersed leks – males display, but form no pair bonds and provide no parental care.
B. Polygynous – male form pair bonds and may provide some parental care.
C. Monogamous – males have well developed pair bonds, males and females provide parental care.

C. Those in which males form pair bonds, are generally monogamous and both males and females care for the young.

These classifications are generally consistent with a trend from extreme polygamy to monogamy. We have ordered the birds, according to their double-havehalf measures, from those populations in which individual mating success is most concentrated to those in which it is least concentrated. As suggested by Payne and Payne, there is a tendency for mating success to be progressively more evenly distributed in systems moving from those which are extremely polygamous to those which are monogamous.

Note that all of the various measures of evenness rank the species in more or less the same order. Consequently, any one of the measures could be used to draw Payne and Payne's conclusion. If a single measure were to be used, it seems to us that the havehalf offers the advantage of being simple to explain and easy to comprehend. If two measures were to be used, the havehalf and the haveall provide, at least for us, more readily intelligible information that any other pair of measures.

CORRELATIONS BETWEEN MEASURES OF EVENNESS

The comparison of mating success of birds presented above suggests that the various measures of evenness and unevenness are highly intercorrelated. To check this conjecture, we calculated the correlation (as measured by Pearson's r^2) between each possible pair of the measures. Table 4 shows the results. Note that we grouped the double-havehalf and the havehalf together and we grouped the co-Gini and Gini measures together because each of these twin measures has identical correlations with the other measures.

All the measures are highly correlated with all the other measures. This suggests that they all are providing more or less the same information. Indeed, the set of alternative measures can be reduced even further than the high correlation coefficients imply. Although the reciprocal-Simpson measure and Crow's I are not perfectly correlated with each other, these two measures are simply transformations of each other and each one is completely determined by the other. It is only because the function linking the two measures is not a linear function that the correlation between the measures is not one. Similarly, the exponential Shannon measure and Thiel's entropy are deterministic transformations of each other.

Table 4. Correlation coefficients (for Pearson's r^2) for evenness measures in Table 3.

	Haveall	Co-Gini/Gini	Reciprocal Simpson	Exponential Shannon	Crow's I	Thiel's Entropy	Have quarter	Half have	Quarter have
Double-Havehalf/Havehalf	.784	.991	.997	.986	.862	.952	.979	.939	.976
Haveall		.820	.788	.831	.738	.844	.729	.762	.842
Co-Gini/Gini			.996	.998	.859	.963	.956	.960	.985
Reciprocal Simpson				.992	.859	.955	.975	.953	.978
Exponential Shannon					.873	.973	.944	.949	.987
Crow's I						.953	.803	.707	.884
Thiel's Entropy							.893	.859	.979
Havequarter								.921	.928
Halfhave									.914

When two measures are highly correlated or are perfectly determined by each other, a choice between the measures can be based on considerations of convenience, intuitiveness, comprehensibility, explainability, and the like. Just as saying a glass is half-full may convey a different vector of meaning than saying the glass is half-empty, use of the haveall measure rather than the havenone measure may highlight a different aspect of evenness in a population. Thus, even in this simple case of two complementary measures, we think that a careful analyst should devote some attention to considering the most appropriate measure to use to present information--and perhaps decide to present both measures. Tversky and Kahneman (1981) and Vaupel (1982) provide further discussion of statistical insinuation and implicational honesty in the use of alternative measures.

We also used data from Lutz (1985) to investigate another data set, pertaining to the concentration of reproduction among human females in 41 different countries. The correlations between the various pairs of evenness and unevenness measures are displayed in Table 5. It is interesting to note that all the measures, with the exception of the haveall statistic, are highly correlated. The haveall measure appears to provide another dimension of information. In Part I of this series of papers, we frequently found the haveall statistic (or its complement, the havenone) be useful in addition to the havehalf measure; the havehalf plus the haveall generally seemed to be the most informative pair of statistics.

CONCLUSION

In the three parts of this series of papers we have strived to persuasively make a single, simply-stated point: concentration curves and various associated have-statistics are useful in ecological analyses of diversity. In Part I, we provided several examples of how concentration curves and have-statistics could be used to analyze evenness in reproductive success. In Part II, we extended the approach to species diversity and some related topics. Finally, here in Part III, we compared have-statistics with other measures of evenness and diversity, both according to some general principles and as applied to some specific examples.

Diversity, heterogeneity, variety, and inequality in populations is a vast subject, at the heart of ecology, demography, and the life sciences more generally. One approach to this subject is to study population concentrations. What proportion of the population has all the offspring, what proportion of the species accounts for half the total biomass, what proportion of the matings are attributable to the most successful quarter of the males? These are important questions that are directly addressed by concentration curves and have-statistics.

Table 5. Correlation coefficients (for Pearson's r^2) for evenness measures, World Fertility Survey.

	Haveall	Co-Gini/Gini	Reciprocal Simpson	Exponential Shannon	Crow's I	Thiel's Entropy	Have quarter	Half have	Quarter have
Double-Havehalf/Havehalf	.816	.959	.967	.864	.926	.849	.971	.918	.993
Haveall		.485	.462	.669	.483	.689	.251	.498	.222
Co-Gini/Gini			.976	.959	.942	.948	.888	.985	.947
Reciprocal Simpson				.948	.989	.940	.935	.929	.972
Exponential Shannon					.938	.999	.788	.943	.862
Crow's I						.936	.918	.886	.953
Thiel's Entropy							.775	.932	.851
Havequarter								.819	.982
Halfhave									.897

REFERENCES

Allison, P.D. (1978) Measures of inequality. *American Sociological Review* 43:865-880.

Brillouin, L. (1962) *Science and Information Theory*. 2nd edition. New York: Academic Press.

Crow, J.F. (1958) Some possibilities for measuring selection intensities in man. *Hum. Biol.* 30:1-13.

Foster, J.E. (1985) Inequality measurement. In *Fair Allocation*, edited by H.P. Young. *Proceedings of Symposia in Applied Mathematics*, vol. 33, American Math. Society.

Holm, C.H. (1973) Breeding sex ratios, territoriality, and reproductive success in the red-winged blackbird. Agelaius phoeniceus. *Ecology* 54:356-365.

Howard, R.D. (1983) Sexual selection and variation in reproductive success in a long-lived organism. *Amer. Natur.* 122:301-325.

Hurlbert, S.H. (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52: 577-586.

Jenkins, D., A. Watson, and G.R. Miller (1963) Population studies on red grouse. Lagopus lagopus scoticus (Lath.) in north-east Scotland. *Journal of Anim. Ecol.* 32: 317-376.

Koivisto, I. (1965) Behavior of the black grouse, Lyrurus tetrix (L.), during the spring display. *Finnish Game Res.* 26: 1-60.

Latter, B.D.H. (1980) Genetic differences within and between populations of the major human subgroups. *Amer. Natur.* 16:220-237.

Lill, A. (1974) Sexual behavior of the lek-forming white-bearded manakin. *Manacus manacus trinitatis hartert. Z. Tierpsychol.* 36: 1-36.

Lill, A. (1976) Lek behavior in the golden-headed manakin. Pipra erythrocephala in Trinidad (West Indies). *Z. Tierpsychol. Suppl.* 18:1-84.

Lutz, W. (1985) *A Comparative Study on 41 Countries Participating in the World Fertility Survey*. Demographic Institute of the Austrian Academy of Sciences, Vienna, Austria.

MacArthur, R.H. (1965) Patterns of species diversity. *Biol. Rev.* 40:510-533.

McIntosh, R.P. (1967) An index of diversity and the relation of certain concepts to diversity. *Ecology* 48: 392-404.

Marshall, A.W. and I. Olkin (1979) *Inequalities: Theory of Majorization and Its Applications*. New York: Academic Press.

Nice, M.M. (1937) *Studies in the Life History of the Song Sparrow*. Vol. 1. New York: Dover.

Patil, G.P. and C. Taillie (1979) An overview of diversity. In *Ecological Diversity in Theory and Practice*, Vol. 1.

Patil, G.P. and C. Taillie (1982) Diversity as a concept and its measurement. *Journal of the American Statistical Association* 77: 548-567.

Payne, R.B. and K. Payne (1977) Social organization and mating success in local song populations of village indigobirds. *Vidua chalybeata. Z. Tierpsychol.* 45: 113-173.

Peet, R.K. (1974) The measurement of species diversity. *Ann. Rev. Ecol. Syst.* 5:285-307.

Pielou, E.C. (1966) The measurement of diversity in different types of biological collection. *Journal of Theoretical Biology* 13: 131-144.

Pielou, E.C. (1969) *An Introduction to Mathematical Ecology*. New York: Wiley.

Pielou, E.C. (1975) *Ecological Diversity*. New York: Wiley.

Preston, S.H. (1976) Family sizes of children and family sizes of women. *Demography* 13:105-114.

Rao, C.R. (1982a) Gini-Simpson index of diversity: a characterization, generalization and applications. *Utilitas Mathematics* 21: 273-282.

Rao, C.R. (1982b) Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology* 21:24-43.

Robel, R.J. (1966) Booming territory size and mating success of the greater prairie chicken. Tympanuchus cupido pinnatus. *Anim. Behav.* 14:328-331. University of Illinois Press, Urbana.

Simpson, E.H. (1949) Measurement of Diversity. *Nature* 163:688.

Thiel, H. (1972) *Statistical Decomposition Analysis*. Amsterdam: North Holland.

Tversky, A. and D. Kahneman (1981) The framing of decisions and the psychology of choice. *Science* 211: 453.

Tukey, J. (1978) *Exploratory Data Analysis*. Reading: Addison-Wesley.

Vaupel, J.W. (1982) Statistical insinuation. *Journal of Policy Analysis and Management* 1(2):261-263.

Verner, J. (1965) Breeding biology of the long-billed marsh wren. *Condor* 67:6-30.