PRODUCT MOMENT

The expected value of a product of random variables \((X_1, X_2, \ldots, X_m)\) possibly raised to different powers. The usual notation is

\[
\mu_{r_1'\ldots r_m'}(X_1, X_2, \ldots, X_m) = E\left[ \prod_{j=1}^{m} X_j^{r_j} \right].
\]

(For brevity, the symbol \(\mu_{r_1'\ldots r_m'}\) is often used.)

The corresponding central product moment is

\[
\nu_{r_1'\ldots r_m'} = E\left[ \prod_{j=1}^{m} (X_j - E[X_j])^{r_j} \right].
\]

To emphasize the distinction between \(\mu_{r_1'\ldots r_m'}\) and \(\nu_{r_1'\ldots r_m'}\), the former is sometimes called a crude product moment.

Some of the \(r_j's\) may be zero, but at least two should differ from zero.

Another name is mixed moment.

Absolute product moments are defined by

\[
\psi_{r_1'\ldots r_m'} = E\left[ \prod_{j=1}^{m} |X_j|^{r_j} \right],
\]

\[
\nu_{r_1'\ldots r_m'} = E\left[ \prod_{j=1}^{m} (X_j - E[X_j])^{r_j} \right].
\]

(PRODUCT MULTINOMIAL)

PROFILE OF DIVERSITY

Diversity is an important concept in ecology and, under various names, it appears in several biological, physical, social, and management sciences. The common theme is the apportionment of some quantity into a number of well-defined categories. The quantity
defined by

\[ J(t) = \sum_{i=0}^{n} K_i(t)(1 + \lambda)^i \]  \hspace{1cm} (12)

\( K_i(t) \) is the gross amount of capital produced in year \( i \) and in use in year \( t \), and \( \lambda \) is the rate of improvement in the productivity of new machines over those produced in the previous year. In this model, the rate of increase in the TFP index is given by

\[ \frac{\dot{A}(t)}{A(t)} = \frac{Q(t)}{Q(t)} - \frac{\dot{L}(t)}{L(t)} \]

\[ = (1 - \alpha) \left[ \frac{K(t)}{K(t)} + \lambda - \lambda \dot{\bar{a}} \right] \]  \hspace{1cm} (13)

where \( \bar{a} \) is the average age of capital. Maximum likelihood estimation* of the parameters of this model is discussed in ref. 41.

**Best Practice Technique**

Salter [30] attributes productivity growth to best practice technique which yields minimum production costs. He decomposes the TFP index into advances in technical knowledge, scale economies, and factor substitution. Salter examines proportionate changes in labor and capital, so as to attribute them to both neutral and nonneutral technical progress as well as to substitution between labor and capital. The discussion of Salter's measures and their application to U.K. and U.S. industries is also contained in ref. 30.

**NOTE**

1. A survey of partial productivity comparisons between countries can be found in [17]. The weighted mean of the ratio of labor hours worked at two points of time is called the man-hour index. The weights are usually the hourly earnings.

**References**

may be in such a form as abundance, time, resource, investment, or energy. The concern in diversity analysis is about the degree of the apportionment being more diverse or less diverse, expressed variously as diversification or concentration, spread or specialization, monopoly or lack of it, and so on.

As a simple example, consider the apportionment of time spent on two activities, music and mathematics. Suppose that John and Jane apportion their time as follows:

<table>
<thead>
<tr>
<th>Category</th>
<th>Music</th>
<th>Mathematics</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>1/3</td>
<td>1/3</td>
</tr>
<tr>
<td>Jane</td>
<td>1/3</td>
<td>1/3</td>
</tr>
</tbody>
</table>

If we ask whether John has a different kind of specialization/ diversification from Jane, the answer is yes: The subject identity matters. Instead, if we ask whether John has a different degree of specialization/diversification from Jane, the answer is no: The subject identity does not matter. The degree of specialization/diversification, or diversity, is permutation invariant; the identities of the categories are disregarded.

Consider a quantity distributed among a finite set of categories labeled \( i = 1, 2, 3, \ldots, s \). Examples of such a quantity are (a) total abundance of an ecological community distributed among \( s \) species, and (b) total market sales distributed among \( s \) competing firms. Let \( \pi_i \) be the proportion of the quantity found in the \( i \)th category, as, for example, the relative abundance of the \( i \)th species in an ecological community. We assume for this discussion that \( \pi_1 > \pi_2 > \cdots > \pi_s > 0 \), and \( \pi_1 + \pi_2 + \cdots + \pi_s = 1 \). The relative abundance vector, \( \pi \), is given by \( \pi = (\pi_1, \pi_2, \ldots, \pi_s) \). Since the vast bulk of applications of diversity analysis occurs in the field of ecology (see ref. 1), we will here refer to the categories as species and define a community \( C \) as the pair \( C = (s, \pi) \).

The concept of diversity of a community has been defined by Patil and Taillie [9–12] independent of any particular diversity indices. In a diverse community, the typical species is relatively rare. Intuitively, a transfer of abundance either from a common to a rare species or from a common to a new, additional species would increase an observer's assessment of the rarity of a typical species. Accordingly, a community \( C' = (s', \pi') \) is defined in [10] to be intrinsically more diverse than a community \( C = (s, \pi) \), written \( C' > C \), if \( C' \) can be constructed from \( C \) by a finite sequence of the following two operations:

1. Introducing a species in this manner: (a) subtract an amount \( h \) from the relative abundance \( \pi_i \) of an existing species, where \( 0 < h < \pi_i \), and (b) add a new species to the community of relative abundance \( h \).

2. Transferring abundance between two existing species in this manner: (a) subtract an amount \( h \) from the relative abundance \( \pi_i \) of a species, and (b) add \( h \) to the relative abundance \( \pi_j \) of another species, where \( 0 < h < \pi_i - \pi_j \).

Note that the species' actual labels are ignored in transforming the vector \( \pi \) into the vector \( \pi' \). It is shown in ref. 12 that \( C' \) is intrinsically more diverse than \( C \) if and only if every right tail-sum of \( \pi' \) is greater than or equal to the corresponding right tail-sum of \( \pi \), that is,

\[
\sum_{i=k}^{s} \pi'_i \geq \sum_{i=k}^{s} \pi_i, \quad k = 1, 2, 3, \ldots, s. \tag{1}
\]

Since the \( k \)th right tail-sum is the combined abundance of those species rarer than the \( k \)th ranked species; greater diversity in this sense means a greater amount of rarity.

A diversity measure of a community \( C \) is defined in [10, 12] as the average rarity of the community, given by

\[
\delta(C) = \frac{1}{s} \sum_{i=1}^{s} \pi_i R(i, \pi). \tag{2}
\]

Here \( R(i, \pi) \) is a numerical measure of rarity associated with the \( i \)th species. Different forms for the function \( R \) produce different
diversity measures. A natural requirement is to use only those forms of $R$ which yield diversity measures with the following properties: (1) $\Delta(C') > \Delta(C)$ whenever $C' > C$, and (2) $\Delta(C) = 0$ if $C$ has only one species.

Two general categories of diversity measures are used frequently, based on two types of rarity functions: (1) rarity functions of the dichotomous type depend only on the relative abundance of the $i$th species: $R(i, \pi) = R(i);$ and (2) rarity functions of the ranking type depend only on the (descending) rank of the $i$th species: $R(i, \pi) = R(i)$.

One dichotomous rarity function suggested in refs. 10 and 12 is

$$R(i, \pi) = (1 - \pi_i^\beta) / \beta,$$

where $\beta$ is a constant chosen by the investigator such that $\beta > 1$. The limiting form $R(i, \pi) = -\log \pi_i$ is used for $\beta = 0$. The result is a parametric family of diversity measures:

$$\Delta_{\beta}(C) = \left(1 - \sum_{i=1}^{r} \pi_i^{\beta+1}\right) / \beta.
$$

Three diversity indices popular in ecological studies are special cases of $\Delta_{\beta}$: (a) $\Delta_{-1} = s - 1$, the species richness index, (b) $\Delta_0 = -\sum \pi_i \log \pi_i$, the Shannon index, and (c) $\Delta_1 = 1 - \sum \pi_i^2$, the Simpson index. However, any other values of $\beta > 1$ also yield diversity indices with the required properties mentioned above. Monotone transformation of $\Delta_{\beta}$ would preserve the ordering properties required of a diversity index; some have been proposed as indices. The transformation $S_\beta = 1/(1 - \beta \Delta_\beta)^{\beta}$ is Hill's [3] family of indices. It has the interpretation of being the numbers equivalent of $\Delta_{\beta}$; that is, $S_\beta$ is the number of species a totally even community must have in order for its value of $\Delta_{\beta}$ to equal that of the given community. The function $\log S_\beta$ is Rényi's [15] entropy* of order $\beta + 1$, proposed by Pielou [13] as a diversity index. (See refs. 3, 13, and 15 and DIVERSITY INDICES for additional information.)

Another dichotomous rarity function is

$$R(i, \pi) = (1 - \pi_i^m) \left[1 - (1 - \pi_i)^m\right] / \pi_i,$$

where $m$ is a nonnegative integer. It leads to the family of diversity indices studied in refs. 4 and 16,

$$\Delta_{\infty}^{\pi}(C) = \sum_{i=1}^{r} \left(1 - \pi_i\right) \left[1 - (1 - \pi_i)^m\right].$$

The quantity $\Delta_{\infty}^{\pi} + 1$ has the interpretation of being the expected number of species obtained when $m + 1$ individuals are randomly selected from the community $C$.

One ranking-type rarity function takes the $k$th ranked species to be the standard against which a species' rarity is assigned the value 1 or 0:

$$R(i) = \begin{cases} 1, & i > k \\ 0, & i < k. \end{cases}$$

The diversity measure produced, denoted $T_k(C)$, is simply the right tail-sum of the relative abundances:

$$T_k(C) = \sum_{i=k}^{r} \pi_i.
$$

Other ranking-type indices are discussed in ref. 12.

Diversity is a partial ordering in the sense that (a) $C > C$, (b) $C'' > C$ whenever $C'' > C'$ and $C' > C$, and (c) a given pair of communities need not be comparable. This explains the complaint frequently heard among ecologists (e.g., in ref. 4) that different diversity indices may give different orderings. In ref. 12, for example, a simple illustration involving two communities is given.

The intrinsic diversity profile of a community $C$ is a plot of the diversity measure $T_k(C)$ (4) against the ranks $k = 1, 2, \ldots, r$. By plotting the profiles of two communities on one graph, one can tell at a glance whether the communities can be ordered by diversity. By (1), $C' > C \Rightarrow T_k(C') > T_k(C)$ for all $k = 1, 2, \ldots, r$. Thus the profile of an intrinsically more diverse community is everywhere above or equal to that of a less diverse community (Fig. 1). The profiles will cross if the communities are not intrinsically comparable.

The measure $T_k(C)$ is a family of diversity indices parameterized by $k$. For larger values of $k$, $T_k$ is in some sense more sensitive
to changes in numbers and abundances of rare species in the community. In ref. 19, a formal definition of sensitivity of an index to rare species is provided. In particular, it is shown that the sensitivity of $\Delta_0$ to rare species is a strictly decreasing function of $\beta$, while that of $\Delta_{\alpha}$ is a strictly increasing function of $m$.

Other profiles of diversity can be constructed by plotting a parametric diversity index against a parameter which measures sensitivity of the index to rare species. For example, $\Delta_0(C)$ can be computed for a given community $C$ using a variety of values of $\beta > -1$. The plot of $\Delta_0$ against $\beta$ is a type of diversity profile. Similarly, $\Delta_{\alpha}$ may be computed and plotted for a variety of $m$ values. If the diversity profiles of two communities cross, the communities cannot be intrinsically ordered according to diversity.

Additional information on diversity analysis may be found in the references following. Ecological studies using diversity profiles include refs. 2, 11, and 8 (this last study refers to the intrinsic diversity profile as a "$k$-dominance curve"). In ref. 21 the $\Delta_0$, $S_0$, and $\Delta_{\alpha}$ families are used to explore relationships between genetic variation and species diversity. In ref. 19 the $\Delta_0$ family is applied to forestry. In ref. 17 estimation for diversity index families is studied. In ref. 18 a diversity ordering based on the concept of majorization is proposed. It turns out to be equivalent to the intrinsic diversity ordering (see ref. 12). In ref. 5 the log series distribution is employed as a model of species abundances, and the estimate of the parameter $\alpha$ is used as a diversity index (see also refs. 6 and 7). In ref. 20 it is shown that the parameter $\alpha$ possesses the intrinsic diversity ordering property. Finally, we mention a different approach to diversity analysis taken in ref. 14. See Rao's axiomatization of diversity.

References


PROGRESSIVE CENSORING SCHEMES

In clinical trials*, survival analysis*, follow-up* studies, and in some problems in reliability*, some incompleteness and/or incompatibility of data may arise due to nonsimultaneous entry of the units into the (experimental) scheme, possible withdrawal of the units from the scheme, and more typically, due to intended termination of the study before all the responses become available. There may be additional complications due to incorporation of interim or repeated analysis schemes for the accumulating data (over the tenure of the study period). For example, in a clinical trial [designed to study the comparative effectiveness or performance of two (or more) drugs (or treatments) relating to a common response] yielding data on the times to failure (response) for the different units (subjects) under study, we may encounter a variety of complications due to one or more of the following factors (these call for special attention for a valid and efficient statistical analysis of the clinical trial data):

1. Censoring/truncation. Due to limitations of time, cost of experimentation, and other considerations, the clinical trial may be planned so that after a prefixed period of time, the trial may be terminated and a statistical analysis be performed on the (partial) data accumulated during this tenure of the experiment. This may be termed a truncated or type I censored experiment. In some other situations, it may be decided to curtail experimentation after a prefixed number (or proportion) of responses becomes available. This may be termed a type II censoring scheme. Note that in a type I censoring scheme, the duration of the trial is prefixed, but the number of responses occurring within this period is random. In a type II censoring scheme, this number is fixed in advance, but the duration of the trial is a random variable. In either case, at the termination of the experimentation, there may still be some surviving units whose failure times are not totally known (these are, however, known to be larger than the censoring times). These censored observations introduce some incompleteness in the response data.

2. Staggering entry. The subjects (or units) may not all be available at the begin-