Diversity index

From Wikipedia, the free encyclopedia
(Redirected from Diversity Index)

A **diversity index** is a *statistic* which is intended to measure the differences among members of a set consisting of various types of objects. Diversity indices can be used in many fields of study to assess the diversity of any population in which each member belongs to a unique group, type or species. For instance, it is used in ecology to measure biodiversity in an ecosystem, in demography to measure the distribution of population of various demographic groups, in economics to measure the distribution over sectors of economic activity in a region, and in information science to describe the complexity of a set of information.

In measuring human diversity, the diversity index measures the probability that any two residents, chosen at random, would be of different ethnicities. If all residents are of the same ethnic group it's zero. If half are from one group and half from another it's .50.[1]

Below, a series of diversity indices is discussed.

Contents

Species richness

The species richness S is simply the number of species present in an ecosystem. This index makes no use of relative abundances. In practice, measuring the total species richness in an ecosystem is impossible, except in very depauperate systems. The observed number of species in the system is a **biased estimator** of the true species richness in the system, and the observed species number increases non-linearly with sampling effort. Thus S, if indicating the *observed* species richness in an ecosystem, is usually referred to as **species density**.

Species Evenness

The species evenness is the relative abundance or proportion of individuals among the species.

Concentration ratio

Concentration ratio is a crude indicator of the extent to which a few groups such as species, demographic groups or companies dominate an environment, the total share taken by the top *n* species or firms. However by itself the concentration ratio does not indicate how much that share is divided between those top *n* firms or species.

Simpson's diversity index

If p_i is the fraction of all organisms which belong to the i-th species, then Simpson's diversity index is most commonly defined as the statistic

$$D = \sum_{i=1}^{S} p_i^2.$$

This quantity was introduced by Edward Hugh Simpson.

If n_i is the number of individuals of species i which are counted, and N is the total number of all individuals counted, then

$$\frac{\sum_{i=1}^{S} n_i(n_i - 1)}{N(N-1)}$$

is an estimator for Simpson's index for sampling without replacement.

Note that $0 \leq D \leq 1$, with values near zero corresponding to highly diverse or heterogeneous ecosystems and values near one corresponding to more homogeneous ecosystems. Biologists who find this confusing sometimes use $1 / D$ instead; confusingly, this reciprocal quantity is also called Simpson's index. Another response is to redefine Simpson's index as

$$\tilde{D} = 1 - D = 1 - \sum_{i=1}^{S} p_i^2,$$

This quantity is called by statisticians the index of diversity.

In sociology, psychology and management studies the index is often known as Blau's Index, as it was introduced into the literature by the sociologist Peter Blau.

In economics essentially the same quantity is called the Hirschman-Herfindahl index (HHI), defined as the sum of the squares of the shares in the population across groups (with E as the group size, that is, the number of employees or the number of specimina):

$$D = \sum_{i=1}(\frac{E_i}{E})^2.$$

Note that a HHI is also used within sectors, to measure competition.

The **index of diversity** (also referred to as the *Index of Variability*) is a commonly used measure, in demographic research, to determine the variation in categorical data.

The most common index of diversity measure was created by Gibbs and Martin (Gibbs, Jack P., and William T. Martin, 1962. "Urbanization, technology and the division of labor." American Sociological Review 27: 667–77); also referred to by Judith Blau (Group Enmity and Accord: The Commercial Press in Three American Cities, Social Science History 24.2, 2000: 395-413) -

$$D = 1 - \sum_{i=1}^{N} p_i^2$$

where

> *p* = proportion of individuals or objects in a category
> *N* = number of categories.

A perfectly homogeneous population would have a diversity index score of 0. A perfectly heterogeneous population would have a diversity index score of 1 (assuming infinite categories with equal representation in each category). As the number of categories increases, the maximum value of the diversity index score also increases (e.g., 4 categories at 25% = .75, 5 categories with 20% = .8, etc.)

An example of the use of the index of diversity would be a measure of racial diversity in a city. Thus, if Sunflower City was 85% white and 15% black, the index of diversity would be: .255.

The interpretation of the diversity index score would be that the population of Sunflower City is not very heterogeneous but is also not homogeneous.

Shannon's diversity index

Shannon's diversity index is simply the ecologist's name for the *communication entropy* introduced by Claude Shannon:

$$H' = -\sum_{i=1}^{S} p_i \ln p_i$$

where p_i is the fraction of individuals belonging to the i-th species. This is by far the most widely used diversity index. The intuitive significance of this index can be described as follows. Suppose we devise binary codewords for each species in our ecosystem, with short codewords used for the most abundant species, and longer codewords for rare species. As we walk around and observe individual organisms, we call out the corresponding codeword. This gives a binary sequence. If we have used an efficient code, we will be able to save some breath by calling out a shorter sequence than would otherwise be the case. If so, the average codeword length we call out as we wander around will be close to the Shannon diversity index.

It is possible to write down estimators which attempt to correct for bias in finite sample sizes, but this would be misleading since communication entropy does not really fit expectations based upon parametric statistics. Differences arising from using two different estimators are likely to be overwhelmed by errors arising from other sources. Current best practice tends to use bootstrapping procedures to estimate communication entropy.

Shannon himself showed that his communication entropy enjoys some powerful formal properties, and furthermore, it is the unique quantity which does so. These observations are the foundation of its interpretation as a measure of statistical diversity (or "surprise", in the arena of communications). The applications of this quantity go far beyond the one discussed here; see the textbook cited below for an elementary survey of the extraordinary richness of modern information theory.

Berger-Parker index

The **Berger-Parker diversity index** is simply

$$\max_{1 \le i \le s} p_i$$

This is an example of an index which uses only partial information about the relative abundances of the various species in its definition.

Gini index

Gini coefficients are used to measure diversity.

Renyi entropy

The Species richness, the Shannon index, Simpson's index, and the Berger-Parker index can all be identified as particular examples of quantities bearing a simple relation to the Renyi entropy,

$$H_\alpha = \frac{1}{1-\alpha} \log \sum_{i=1}^{S} p_i^\alpha$$

for α approaching $0,\ 1,\ 2,\ \infty$ respectively.

Unfortunately, the powerful formal properties of communication entropy do not generalize to Renyi's entropy, which largely explains the much greater power and popularity of Shannon's index with respect to its competitors.

See also

- Alpha diversity
- Qualitative variation
- Shannon index

External links

- Simpson's Diversity index (http://www.countrysideinfo.co.uk/simpsons.htm)
- Diversity indices (http://www.tiem.utk.edu/~gross/bioed/bealsmodules/simpsonDI.html) gives some examples of estimates of Simpson's index for real ecosystems.
- [1] (http://chao.stat.nthu.edu.tw/paper/2003_EEST_10_P429.pdf)

References

1. ^ "Mapping L.A..," *Los Angeles Times* website (http://projects.latimes.com/mapping-la/neighborhoods/diversity/neighborhood/list/)

Further reading

- Colinvaux, Paul A. (1973). *Introduction to Ecology*. Wiley. ISBN 0-471-16498-4.
- Cover, Thomas M.; and Thomas, Joy A. (1991). *Elements of Information Theory*. Wiley. See *chapter 5* for an elaboration of coding procedures described informally above.

Categories: Measurement of biodiversity | Index numbers | Summary statistics for categorical data

---