



## FORUM

### On parametric diversity indices in ecology: A historical note

C. Ricotta

*Department of Plant Biology, University of Rome "La Sapienza", Piazzale Aldo Moro 5, I-00185 Rome.  
Tel: +39 06 49912408, Fax: +39 06 4457540, E-mail: carlo.ricotta@uniroma1.it*

In a recent paper, Keylock (2005) suggested the use of Tsallis (1988) generalized entropy as a parametric measure of biological diversity. Keylock (2005) notes that since the Shannon and the Simpson diversity indices are special cases of Tsallis generalized entropy, the proposed function represents the cornerstone for a continuum of possible diversity measures.

Also, beyond the practical aspects of diversity measurement and comparison, the author showed that Tsallis generalized entropy may be useful for modeling various properties of ecological systems within the broader context of statistical mechanics and generalized information theory. As shown by Keylock (2005), a generalized version of the Zipf-Mandelbrot distribution sometimes used to characterize rank-abundance relationships may be obtained from the Tsallis entropy, thus providing an information-theoretical framework for the macro-ecological study of the energetics of communities. In this view, as well as by providing a useful practical tool for field ecology, the Tsallis entropy has more general implications in ecological theory.

Incidentally, the proposed measure of parametric diversity has a long history in statistical ecology. It has been well-known for more than 25 years as the 'diversity index of degree  $\beta$ ' (Patil and Taillie 1979, 1982). Unfortunately, neither the author nor the referees of Keylock's (2005) paper were able to detect the analogy between the Tsallis generalized entropy and the Patil and Taillie diversity index of degree  $\beta$ . Therefore, I came to the idea of shortly revisiting some historical aspects on the development of parametric diversity in statistical ecology in order to understand why most of this work was ignored by the ecological community.

#### A brief history of parametric diversity indices

Since the work of Hurlbert (1971), ecologists have developed a number of parametric measures for summarizing community diversity (Hill 1973, Patil and Taillie 1979, 1982, Tóthmérész 1998, Baczkowski et al. 2000, Southwood and Henderson 2000, Ricotta 2004), the most celebrated of which is probably the Hill family of indices.

For a set composed of  $S$  elements with the relative abundance vector  $p = (p_1, p_2, \dots, p_S)$ , Rényi (1961) extended the concept of Shannon's entropy by defining a parametric entropy of order  $\alpha$  as:

$$H_\alpha = \frac{1}{1-\alpha} \ln \sum_{j=1}^S p_j^\alpha \quad (1)$$

where  $p_j$  denotes the relative abundance of the  $j$ -th element ( $j = 1, 2, \dots, S$ ) within the set such that  $0 \leq p_j \leq 1$ , and  $\sum_{j=1}^S p_j = 1$ . In principle,  $\alpha$  makes mathematical sense for  $-\infty \leq \alpha \leq \infty$ . However, a parameter restriction ( $\alpha \geq 0$ ) has to be imposed if  $H_\alpha$  should possess certain desirable properties, such as the Pigou-Dalton principle of transfers that renders it adequate in ecological research (for details, see Patil and Taillie 1982).

In its very essence, the entropy  $H$  of a given set is a measure of uncertainty in predicting the relative abundance of elements. High entropy thus implies high unpredictability and is maximum when all elements occur in equal abundance (an even distribution). The amount of information obtained from observing the result of an experiment depending on chance can be taken to be numerically equal to the amount of uncertainty in the outcome of the experiment before carrying it out. Therefore, entropy can also be viewed as a measure of information (Rényi 1970).

Expanding on Rényi's work, Hill (1973) showed that the generalized entropy  $H_\alpha$  has many desirable properties for summarizing the diversity of a biological community. One particularly convenient property is that a number of traditional diversity indices are special cases of  $H_\alpha$ . For instance, for  $\alpha = 0$ ,  $H_0 = \ln S$ , where  $S$  is species richness; for  $\alpha = 2$ ,  $H_2 = \ln 1/D$ , where  $D$  is the Simpson concentration or dominance  $D = \sum_{j=1}^S p_j^2$ , and  $p_j$  denotes the relative abundance of the  $j$ -th species. Finally, for  $\alpha = \infty$ ,  $H_\alpha = \ln 1/p_{\max}$ , where  $p_{\max}$  is the relative abundance of the most frequent species in the assemblage. Notice that for  $\alpha = 1$ , Eq. (1) is defined in the limiting sense, and  $H_1$  equals the Shannon entropy,  $-\sum_{j=1}^S p_j \ln p_j$ .

Traditional diversity indices such as the Simpson concentration or the Shannon index supply point descriptions of community structure. On the other hand, according to Hill's generalization, there is a continuum of possible diversity measures, which differ in their sensitivity to the rare and abundant species in the assemblage, becoming increasingly dominated by the commonest species for increasing values of the parameter  $\alpha$  (Patil and Taillie 1982, Ricotta 2000). In this view, changing  $\alpha$  can be considered a scaling operation that takes place not in the real, but in the data space (Podani 1992)<sup>1</sup>.

Besides Rényi's generalized entropy of order  $\alpha$ , several additional generalized information functions were discovered and rediscovered during the years in fields as different as mathematics, physics and engineering. For a short overview on "the labyrinthic history of entropies", see Tsallis (2002).

For instance, in the physics community, Tsallis (1988) developed another generalized information function given by

$$H^\alpha = (1 - \sum_{j=1}^S p_j^\alpha) / (\alpha - 1) \quad (2)$$

Here, for  $\alpha = 0$ ,  $H^0 = S - 1$  (i.e., a monotone function of species richness for which the diversity of monospecific assemblages equals zero); for  $\alpha = 1$ ,  $H^1 = -\sum_{j=1}^S p_j \ln p_j$ , for  $\alpha = 2$ ,  $H^2 = \sum_{j=1}^S p_j^2$  (i.e., the Simpson diversity), and for  $\alpha \rightarrow \infty$ ,  $H^\infty$  converges to zero. Since the Rényi and the Tsallis entropy both depend on the same expression  $\sum_{j=1}^S p_j^2$ , the relation between  $H_\alpha$  and  $H^\infty$  follows from a simple comparison of Equation (1) with Equation (2):

$$H_\alpha = \frac{1}{1 - \alpha} \ln[1 + (1 - \alpha)H^\alpha] \quad (3)$$

From a mathematical viewpoint, the Rényi entropy and the Tsallis entropy have slightly different properties. While the Shannon entropy  $-\sum_{j=1}^S p_j \ln p_j$  is both concave and extensive, the Rényi entropy  $H_\alpha$  is extensive but concave only for  $0 < \alpha \leq 1$ . By contrast, the Tsallis entropy  $H^\alpha$  is concave and nonextensive for all  $\alpha > 0$ . Therefore, for practical purposes one uses an entropy most adequate to solve the given problem (Aczél and Daróczy 1975, Tsallis 2002).

In statistical ecology, the same measure of parametric diversity was independently proposed more than 25 years ago by Patil and Taillie (1979, 1982) as the 'diversity index of degree  $\beta$ '

$$\Delta_\beta = (1 - \sum_{j=1}^S p_j^{\beta+1}) / \beta. \quad (4)$$

Interestingly,  $\Delta_\beta$  was proposed in the ecological literature within the context of intraspecific encounter theory and without any information-theoretical interpretation (see Patil and Taillie 1979, 1982). Nonetheless, putting  $\alpha = \beta + 1$ , the diversity index of degree  $\beta$  is identical to Tsallis (2002) nonextensive entropy (see Eq. 2). This identity was first noticed by Ricotta (2003).

While Keylock (2005) remarks that: "From the perspective of [the Tsallis] generalized measure of entropy [the Shannon entropy] is no more justifiable than [the Simpson diversity] as a diversity measure because both can be linked to a generalized formulation of entropy", Patil and Taillie (1979), reversing the same argument, observe: "Particularly, we note that the Shannon index has an encounter theoretic interpretation and thus should not be singled out or criticized simply because its continuing use in information theory".

Also, within the context of intraspecific encounter theory (that is certainly rooted deeper in the ecological practice than the abstract concept of information theory) Patil and Taillie (1979) show the close relationship of the diversity index of degree  $\beta$  with the parametric diversity index proposed by Hurlbert (1971):

$$S_\alpha = \sum_{j=1}^S (1 - (1 - p_j)^\alpha) \quad (5)$$

where  $S_\alpha$  represents the expected number of species present when a random sample of size  $\alpha$  ( $\alpha \geq 1$ ) is drawn with replacement from a given species assemblage (Hurlbert 1971, Tóthmérész 1994, Ricotta 2004). Note that, in the ecological community, the fundamentals of Hurlbert's

1 Here, I consider extremely important László Orlóci's influence who, as a referee of the first version of Hill's diversity paper submitted to the *American Naturalist* (Bob Sokal editor), brought Rényi's work to the attention of the author. The paper was published with the Rényi content later in *Ecology*, after rewriting (Orlóci, personal communication).

scaling function have been recognized in an elementary form since the 1930's (Tóthmérész, personal communication).

To complete the picture, it is also worth mentioning that another generalized entropy function of the form

$$H^\alpha = \frac{\sum_{j=1}^S p_j^\alpha - 1}{2^{1-\alpha} - 1} \quad (6)$$

was extensively studied by Aczél and Daróczy (1975). As emphasized by Tsallis (2002), this function is very similar to the generalized entropy of Eq. (2), though with a different prefactor fitted for binary variables. In the ecological literature, its use as a parametric diversity measure was first proposed by Tóthmérész (1993, 1995).

### Where next?

Why were the many pioneering contributions on parametric diversity (such as the papers of Patil and Taillie 1979, 1982) largely overlooked by later authors? One might argue that, since the 1979 paper was published in an edited book, while the 1982 paper was published in the *Journal of the American Statistical Association*, both papers experienced a rather limited circulation among ecologists. However, this is probably not the case, at least for the 1982 paper, which was reviewed in the same journal issue containing the Patil and Taillie article by George Sugihara, a central figure of quantitative ecology.

By contrast, a more plausible explanation emerges by considering the different approaches to biodiversity research in the 1970's and 1980's. Many statistical ecologists considered (and still consider) the problem of measuring biodiversity from a statistical perspective with little or no reference to the biological factors that affect the structure of species assemblages. As clearly illustrated by Colwell (1979): "The statistical approach to diversity has focused on the goals of relating empirical distributions (principally species abundance data) to theoretical ones, and on finding parameters or indices appropriate for comparing the species diversity of collections or communities. These goals are usually seen as preliminary to the inference of mechanisms and processes from differences or universals among empirical patterns."

Hence, in the view of these authors, the statistical aspects of diversity metrics have a primary role that needs to be examined first: "Diversity as an unequivocal and inherent property of a biological system does not exist. Instead, as it is the case with any other statistic, diversity values are merely numbers and their relevance to ecologi-

cal problems must be judged on the basis of observed correlations with other environmental variables" (Molinari 1989; see also Ricotta 2005).

By contrast, from the perspective of a field ecologist, there was a strong belief that diversity as an ecological concept is different from diversity as a statistical index. For instance, in his comment on Patil and Taillie (1982), Sugihara (1982) noted: "To the ecologist, diversity is interesting as a property of state in so far as it has the potential to reflect the nature of the underlying processes and organization that structure the community. Therefore, beyond arbitrary or weakly motivated definitions the scientific interest in and importance of ecological diversity hinges directly on its possible connection with the functioning and organization of communities..."

Similarly, in reviewing the book edited by Grassle et al. (1979), Pielou (1980) wrote: "I was distressed to find that so many people treat 'diversity' and 'diversity index' as synonyms. It will be evidently new to many statisticians that an ecologist studying diversity is not merely engaged in devising, and estimating, an index that is the qualitative analogue to variance. Ecological diversity is a biological phenomenon..."

In a sense, the pioneering work of Hurlbert, Hill, Patil and Taillie and many others was not relevant for the majority of the ecologists of the 1970's and 1980's. For instance, they simply ignored how to use the proposed diversity measures for relating community structure to ecological functioning, going beyond the purely statistical aspect of diversity measurement and comparison.

It is just in the latest few years, under the impulse of modern biocomplexity research, that some theoretical ecologists rediscovered the more general utility of parametric entropy and diversity for modeling various properties of ecological systems, as elegantly illustrated by Keylock (2005).

To conclude, an up to date re-examination of the pioneering papers on parametric diversity of the 1970's and 1980's might promote deeper understanding of new aspects of ecological theory and its modeling implications within the broader context of statistical mechanics and generalized information theory. On the other hand, by ignoring more than 25 years of work on parametric diversity we risk throwing out the baby with the bathwater.

**Acknowledgement.** I am grateful to Béla Tóthmérész for fruitful discussion on parametric diversity indices.

### References

- Aczél, J. and Daróczy, Z. 1975. *On Measures of Information and their Characterizations*. Academic Press, London.

- Baczkowski, A., Joanes D.N. and Shamia, G. 2000. The distribution of a generalized diversity index due to Good. *Environ. Ecol. Stat.* 7: 329-342.
- Colwell, R.K. 1979. Toward an unified approach to the study of species diversity. In: Grassle, J.F., Patil, G.P., Smith, W. and Taillie, C. (eds.), *Ecological Diversity in Theory and Practice*. International Cooperative Publishing House, Fairland, Maryland. pp. 75-91.
- Grassle J.F., Patil, G.P., Smith, W. and Taillie, C. (ed.). 1979. *Ecological Diversity in Theory and Practice*. International Cooperative Publishing House, Fairland, Maryland.
- Hill, M.O. 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology* 54: 427-431.
- Hurlbert, S.H. 1971. The nonconcept of species diversity: a critique and alternative parameters. *Ecology* 52: 577-586.
- Keylock, C.J. 2005. Simpson diversity and the Shannon-Wiener index as special cases of a generalized entropy. *Oikos* 109: 203-207.
- Molinari, J. 1989. A calibrated index for the measurement of evenness. *Oikos* 56: 319-326.
- Patil, G.P. and Taillie, C. 1979. An overview of diversity. In: Grassle, J.F., Patil, G.P., Smith, W. and Taillie, C. (eds.), *Ecological Diversity in Theory and Practice*. International Cooperative Publishing House, Fairland, Maryland, pp. 3-27.
- Patil, G.P. and Taillie, C. 1982. Diversity as a concept and its measurement. *J. Am. Stat. Ass.* 77: 548-567.
- Pielou, E.C. 1980. Review on Grassle et al. (1979). *Biometrics* 36: 742-743.
- Podani, J. 1992. Space series analysis: processes reconsidered. *Abstr. Bot.* 16: 25-29.
- Rényi, A. 1961. On measures of entropy and information. In: Neyman, J. (ed.), *Proceedings of the 4<sup>th</sup> Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I. University of California Press, Berkeley, pp. 547-561.
- Rényi, A. 1970. *Probability Theory*. North-Holland, Amsterdam.
- Ricotta, C. 2000. From theoretical ecology to statistical physics and back: self-similar landscape metrics as a synthesis of ecological diversity and geometrical complexity. *Ecol. Model.* 125: 245-253.
- Ricotta, C. 2003. Additive partition of parametric information and its associated  $h'$ -diversity measure. *Acta Biotheor.* 51: 91-100.
- Ricotta, C. 2004. A parametric diversity measure combining the relative abundances and taxonomic distinctiveness of species. *Divers. Distrib.* 10: 143-146.
- Ricotta, C. 2005. Through the jungle of biological diversity. *Acta Biotheor.* 53: 29-38.
- Southwood, T.R.E. and Henderson, P.A. 2000. *Ecological Methods*, 3<sup>rd</sup> ed. Blackwell Science, Oxford.
- Sugihara, G. 1982. Comment to Patil and Taillie (1982). *J. Am. Stat. Ass.* 77: 564-565.
- Tóthmérész, B. 1993. DivOrd 1.50: A program for diversity ordering. *Tiscia* 27: 33-44.
- Tóthmérész, B. 1994. Statistical analysis of spatial pattern in plant communities. *Coenoses* 9: 33-41.
- Tóthmérész, B. 1995. Comparison of different methods for diversity ordering. *J. Veg. Sci.* 6: 283-290.
- Tóthmérész, B. 1998. On the characterization of scale-dependent diversity. *Abstr. Bot.* 22: 149-156.
- Tsallis, C. 1988. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* 52: 479-487.
- Tsallis, C. 2002. Entropic nonextensivity: a possible measure of complexity. *Chaos, Solitons and Fractals* 13: 371-391.