

## **Copyright Notice**

This electronic reprint is provided by the author(s) to be consulted by fellow scientists. It is not to be used for any purpose other than private study, scholarship, or research.

Further reproduction or distribution of this reprint is restricted by copyright laws. If in doubt about fair use of reprints for research purposes, the user should review the copyright notice contained in the original journal from which this electronic reprint was made.

# UNA NOTA ACERCA DE LA DIVERSIDAD

Exequiel EZCURRA (1)

En Ecología 3 (pp. 125-131), Schlichter, León y Soriano comparan dos índices de diversidad ( $H$  y  $D$ ) sobre muestras de pastizales de Chubut, y concluyen que "el que mejor representa el estado del pastizal es el índice de diversidad derivado de teoría de la información ( $H$ )". Agregan que "el índice de Simpson pondera excesivamente la dominancia, siendo muy poco sensible a variaciones no muy marcadas en el número de especies". En esta nota pretendo demostrar que:

a. Ambos índices están relacionados con la teoría de la información (si bien este hecho es irrelevante en ecología).

b. La forma de comparación que utilizan Schlichter *et al.* no es matemáticamente válida.

c. En general los índices de diversidad derivados de la fórmula de Simpson poseen mejores propiedades que el llamado índice de Shannon.

Para ello, partiré de los postulados básicos de la teoría de información y desarrollaré brevemente las generalizaciones necesarias.

En un sistema compuesto por  $s$  categorías (v.g. especies) de igual frecuencia de aparición, la frecuencia ( $p$ ) (abundancia relativa o probabilidad) de cada categoría es:

$$p = \frac{1}{s} \quad (1),$$

$$y \therefore s = \frac{1}{p} \quad (2)$$

Si se desean codificar los estados del sistema en base a información binaria, el número de dígitos binarios necesarios para describir cada categoría es:

$$H = \log_2 s \quad (3) \quad \text{ó} \quad H = \log_2 \frac{1}{p} \quad (4)$$

Si las categorías del sistema no son equiprobables, la frecuencia relativa ( $p_i$ ) de cada categoría  $i$  es:

$$p_i = \frac{n_i}{n} \quad (5),$$

donde  $n_i$  es el número de elementos en la categoría  $i$ , y  $n$  es el número total de elementos en el sistema ( $n = \sum_i n_i$ ). Dado que las frecuencias relativas ( $p_i$ ) no son iguales para todas las categorías, la fórmula (4) es reemplazada por:

$$H = \log_2 \frac{1}{\bar{p}} \quad (6)$$

donde  $\bar{p}$  es la media de las frecuencias de las distintas categorías.

Esta fórmula (6) es de central importancia en la medición de diversidades, o más ampliamente, de heterogeneidad de distribuciones de frecuencias. Es importante notar que: (a) la función  $N = 1/\bar{p}$  es una buena medida de diversidad, e indica el número "aparente" o "efectivo" de categorías; (b) la transformación logarítmica ( $H = \log N$ ) no tiene demasiado sentido cuando lo que se desea es medir la diversidad ( $N$ ) del sistema, y no el número de dígitos necesarios para describirla; y (c) si se usa la transformación logarítmica, el logaritmo en base 2 no es en absoluto necesario, ya que la codificación de las categorías (especies) en ecología no es en general de tipo binario.

La probabilidad media ( $\bar{p}$ ) de la ecuación (6) se puede calcular, en series de frecuencias, de dos formas igualmente válidas: como media geométrica ( $\bar{p}_g$ ), o como media aritmética ( $\bar{p}_a$ ). Entonces

$$\bar{p}_g = \left( \prod_i p_i^{n_i} \right)^{1/n} = \prod_i p_i^{p_i} \quad (7)$$

$$\bar{p}_a = \frac{\sum_i n_i p_i}{\sum_i n_i} = \sum_i p_i^2 \quad (8)$$

donde  $\bar{p}_a$  es el llamado "índice de Simpson". Reemplazando (7) y (8) en (6), definiremos dos medidas posibles de diversidad ( $N_1$  y  $N_2$ ), y dos medidas de información asociadas ( $H_1 = \log N_1$ ; y  $H_2 = \log N_2$ ); tal que:

$$N_1 = \frac{1}{\prod_i p_i^{p_i}} \quad (9)$$

$$H_1 = -\log \prod_i p_i^{p_i} = -\sum_i p_i \log p_i \quad (10)$$

(1) Instituto de Ecología  
Apartado 18-845  
México 20, D.F.

$$N_2 = \frac{1}{\sum_i p_i^2} \quad (11) \quad H_2 = -\log \sum_i p_i^2 \quad (12)$$

Los cuatro índices son estimadores eficientes de la diversidad, pero en ecología los números de diversidad ( $N$ ) son preferibles a las informaciones ( $H$ ) por las siguientes razones: (a) las variaciones en su escala son linealmente proporcionales a cambios en la riqueza específica para una misma distribución especie-abundancia (*v.g.* log-normal); y (b) indican al ecólogo el número de especies "aparentes", es decir, fácilmente visibles en la comunidad.

A. Rényi ha propuesto un marco teórico adecuado para comprender el problema (Hill, 1973; Rényi, 1961; Rényi, 1973), definiendo

$$N = \left( \sum_i p_i^\alpha \right)^{1/1-\alpha} \quad (13),$$

$$0 \leq \alpha \leq \infty,$$

donde  $N_\alpha$  se conoce como el "número de diversidad de orden  $\alpha$ ". Se observa fácilmente que  $N_0$  es igual a  $s$ , o riqueza específica de la comunidad; que  $N_1$  (que se resuelve por límites) corresponde con la fórmula (9) o antilogaritmo del índice de Shannon; y que  $N_2$  es la inversa del índice de Simpson (Hill, 1973). Se observa también que  $N_0$  es la inversa de la media armónica ponderada de las frecuencias específicas; y que  $H_\alpha = \log N_\alpha$  es la "información de orden  $\alpha$ ", de las cuáles el índice de Shannon es el caso particular de orden 1.

Si Schlichter *et al.* hubieran comparado  $N_1$  y  $N_2$  (o alternativamente  $H_1$  y  $H_2$ ) hubieran encontrado que las diversidades de ambos órdenes ajustan significativamente el puntaje de Soriano y Brun (si bien esto no es prueba suficiente de la bondad del índice), y que la correlación entre ambas diversidades es estrecha (admitiendo la dispersión lógica entre media aritmética y geométrica). Su apreciación sobre la inadecuación del índice de Simpson proviene de las transformaciones elegidas más que de las propiedades del índice en sí: al comparar el logaritmo de una función ( $H = \log N_1$ ) con la inversa de una función muy similar ( $D = 1/N_2$ ), encuentran que la segunda ( $D$ ) exagera los valores bajos de  $N_2$ , y concluyen que  $D$  "pondera excesivamente la dominancia".

Sin embargo, en ecología parece más apropiado usar  $N_2$  como medida de diversidad, por las siguientes razones:

a. A medida que  $\alpha$  aumenta, el sesgo muestral de  $N_\alpha$  y  $H_\alpha$  disminuye, y los índices son más consistentes (Hill, 1973).

b.  $N_2$  se puede calcular con la fórmula abreviada

$$N_2 = \frac{\left( \sum_i n_i \right)^2}{\sum_i n_i^2} \quad (14)$$

que facilita enormemente los cálculos.

c. Utilizando  $N_2$  es posible relacionar la diversidad de especies con la heterogeneidad ambiental, la diversidad de utilidades y la matriz de competencia. Para un desarrollo matemático detallado de este problema, el lector es referido a Mac Arthur (1972).

d. La uniformidad de la distribución o equitabilidad de las especies ( $E$ ) se puede medir como

$$E = N_\alpha / N_\beta \quad (15)$$

donde  $\alpha > \beta$  (Hill, 1973). Por la monotonicidad de su comportamiento y por su facilidad de cálculo, es de utilidad en ecología  $E = N_2/N_0$ .

e. El uso de  $N_2$  reduce el cálculo de diversidades al nivel de un estadístico descriptivo más, y evita la proliferación de conceptos confusos como la "información" o la "entropía" de una comunidad, o la "naturalidad cibernética" del ecosistema. Ideas que poseen poco sentido ecológico (May, 1976, pp. 157-158; Engleberg y Boyarsky, 1979) y evitan enfrentar en forma seria la compleja interacción evolutiva de los seres vivos.

## BIBLIOGRAFÍA

- HILL, M.O. 1973. Diversity and evenness: a unifying notation and its consequences. *Ecology*, 54: 427-432.
- ENGLEBERG, J. y BOYARSKY, L.L. 1979. The noncybernetic nature of ecosystems. *Am. Nat.* 114: 317-324.
- MAY, R.M. (Ed.). 1976. *Theoretical Ecology, principles and applications*. Blackwell, London.
- McARTHUR, R.H. 1972. *Geographical Ecology*. Harper and Row, N. York.
- RÉNYI, A. 1961. On measures of entropy and information. En J. Neyman (Ed.) *4th. Berkeley Symposium on mathematical statistics and probability*. Berkeley.
- 1976. *Cálculo de probabilidades*. Editorial Reverté, Barcelona.
- SCHLICHTER, T., LEÓN R.J.C. y SORIANO, A. 1978. Utilización de índices de diversidad en la evaluación de pastizales naturales en el centro-oeste del Chubut, *Ecología*, 3: 125-131.