Algunos índices de diversidad comúnmente utilizados

Consideremos un vector de probabilidades P, $P=(p_1,\ldots,p_k)$, $p_i \geq 0$, $\sum_{i=1}^{k} p_i = 1$ y una muestra multinomial (frecuencias absolutas) $\mathbf{n} = (n_1,\ldots,n_k)$, $n_i \geq 0$, $n_1+\ldots+n_k = n$, obtenida a partir de P al haber repetido en condiciones de independencia una experiencia n veces y haber contado cuantas veces se producía el suceso número i, i =1,…,k, que se podía producir con probabilidad p_i. Indicaremos como $\hat{p}_i = \frac{n_i}{n}$ las correspondientes frecuencias relativas, estimaciones de las probabilidades p_i.

Entropía o diversidad de Shannon:

$$H^S(P) = -\sum_{i=1}^{k} p_i \log_2 p_i .$$

Se propuso primero en un contexto de teoría de la comunicación pero se ha empleado en numerosas disciplinas. En particular Ramón Margalef propuso su utilización en un contexto de biodiversidad, siendo por ejemplo p_i las proporciones reales de individuos de distintas especies en un ecosistema y $\hat{p}_i$ sus estimaciones a partir de una muestra de tamaño n.
Si la base del logaritmo empleado en su definición es 2, la diversidad se mide en “bits”, interpretable como número de decisiones binarias equiprobables que se tendrían que realizar para determinar el valor de i. Como toda medida razonable de diversidad, tiene su máximo valor cuando todas las opciones se presentan con la misma probabilidad, $p_i = 1/k$ para toda i. En particular si P = (0.5, 0.5) y tomamos logaritmos en base 2, tenemos que $H^S(P) = 1\text{bit}$.
A pesar de lo dicho anteriormente, este índice tiene sentido para cualquier base logarítmica, en particular para el caso de logaritmos naturales, en base e.
A partir de una muestra $\mathbf{n} = (n_1,\ldots,n_k)$, $n_i \geq 0$, $n_1+\ldots+n_k = n$ obtenida de P, la diversidad muestral de Shannon:

$$\hat{H}^S = H^S(\hat{P}) = -\sum_{i=1}^{k} \hat{p}_i \log_2 \hat{p}_i \qquad \hat{p}_i = \frac{n_i}{n}$$

es un estimador razonable de $H^S(P)$ cuyo error estándar se puede estimar mediante:

$$\widehat{\text{se}}_{\hat{H}^S}(\hat{P}) = \frac{1}{\sqrt{n}} \sqrt{\sum_{i=1}^{k} \hat{p}_i \left(\log_2 \hat{p}_i\right)^2 - \left(H^S(\hat{P})\right)^2}$$

Entropía o diversidad de Rényi de orden α:

$$H_\alpha^R(P) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{k} p_i^\alpha\right), \text{ para } P=(p_1,\ldots,p_k),\ p_i \geq 0,\ \sum_{i=1}^{k} p_i = 1,$$

es una generalización de la entropía de Shannon bastante utilizada en distintos campos, como Física, Ecología, Bioinformática o Economía, como medida de diversidad o de información. Depende de un parámetro α positivo arbitrario (a fijar por quien utiliza este índice), distinto de 1. Valores de este parámetro

cercanos a cero confieren peso similar a todas las probabilidades p_i, valores progresivamente grandes de α dan cada vez más peso a los valores de p_i mayores. En el valor intermedio $\alpha = 1$ se reduce a la entropía de Shannon, en el sentido de que tiende a la entropía de Shannon cuando $\alpha \rightarrow 1$ (y cuando se toman logaritmos en base 2).
A partir de una muestra multinomial **n** = (n_1,…,n_k), $n_i \geq 0$, n_1+…+n_k = n obtenida de P, la entropía muestral de Rényi:

$$\hat{H}_\alpha^R = H_\alpha^R\left(\hat{P}\right) = \frac{1}{1-\alpha}\log\left(\sum_{i=1}^{k}\hat{p}_i^{\alpha}\right),\ \hat{p}_i = \frac{n_i}{n}$$

es un estimador razonable de $H_\alpha^R(P)$ cuyo error estándar se puede estimar mediante:

$$\widehat{\text{se}}_{\hat{H}_\alpha^R}\left(\hat{P}\right) = \frac{1}{\sqrt{n}}\left|\frac{\alpha}{1-\alpha}\right|\sqrt{\frac{\sum_{i=1}^{k}\hat{p}_i^{2\alpha-1}}{\left(\sum_{i=1}^{k}\hat{p}_i^{\alpha}\right)^2} - 1}.$$

´

Entropía o diversidad de Havrda-Charvat de orden α:

$$H_\alpha^{HC}\left(P\right) = \frac{1}{\alpha-1}\left[1-\sum_{i=1}^{k}p_i^{\alpha}\right]$$

También depende de un parámetro α positivo arbitrario (a fijar por quien utiliza este índice), distinto de 1. El caso $\alpha = 2$ es especialmente interesante en genética ya que correspondería al grado máximo de heterocigosis alcanzable para un gen con k alelos que se presentan en proporciones $P = (p_1,\ldots,p_k)$.

Substituyendo P por su estimación $\hat{P}$ en la expresión anterior tendríamos la diversidad muestral de Havrda-Charvat de orden α.

$$\hat{H}_\alpha^{HC} = H_\alpha^{HC}\left(\hat{P}\right) = \frac{1}{\alpha-1}\left[1-\sum_{i=1}^{k}\hat{p}_i^{\alpha}\right]$$

Su error estándar se puede estimar mediante:

$$\widehat{\text{se}}_{\hat{H}_\alpha^{HC}}\left(\hat{P}\right) = \frac{1}{\sqrt{n}}\left|\frac{\alpha}{\alpha-1}\right|\sqrt{\sum_{i=1}^{k}\hat{p}_i^{2\alpha-1} - \left(\sum_{i=1}^{k}\hat{p}_i^{\alpha}\right)^2}.$$