Other coefficients

Returning to the quantitative resemblance coefficients in the •Others list, five further measures given under the ✓Distance/dissimilarity heading are (loosely) based on likelihood-ratio tests. All are motivated by the (usually unrealistic) model in which the individuals of a species are randomly distributed in space or time (i.e. the data are strict counts, Poisson distributed), independently of other species, and with the mean count differing over species. A generalised likelihood ratio (GLR) test that two samples come from the same assemblage then produces the test statistic:

$D^{BinD} = 2\sum_i \left[y_{i1} \log \left(\frac{y_{i1}}{ y_{i1}+y_{i2}} \right) + y_{i2} \log \left(\frac{y_{i2}}{ y_{i1}+y_{i2}} \right) + \left(y_{i1} + y_{i2}\right) \log 2 \right] \text{\hspace{10mm} Binomial deviance,} $

where the sum is over all p species as usual (note the first two terms do go to zero, unambiguously, when y_{i1} and y_{i2} are zero, respectively). In fact, the coefficient is of the form $2 \sum \left[O \log(O/E) \right]$, where $O=y_{i1}$ or y_{i2} and $E=(y_{i1}+y_{i2})/2$ are the observed and expected values in a chi-squared type test of equality of counts for species i, then summed over the (supposedly independent) species, $i = 1,\ldots, p$. The more familiar Wald test statistic for this situation is $\sum \left[(O – E)^2 /E \right]$, but the two measures are likely to behave very similarly in practice (both having large-sample distributions of χ^2 on p df). A more useful variant of the latter is therefore given under Measure•Others, by simply dividing the chi-squared by the number of non jointly-absent species (p_{12}) for these two samples:

$D^{Wald} = \frac{1}{p_{12}} \sum_i \left[\frac{ \left(y_{i1}-y_{i2} \right)^2}{ \left(y_{i1}+y_{i2} \right)} \right] \text{\hspace{30mm} Wald (chi-squared) coefficient,}$

thus making this form of the coefficient independent of joint absences. This could be further modified in a natural way, to make it more robust to large y_{ij} (outliers) whilst preserving similar behaviour, by replacing a sum of squares with a sum of absolute values:

$D^{Chi} = \frac{1}{p_{12}} \sum_i \left[\frac{ \left| y_{i1}-y_{i2} \right|}{ \sqrt {y_{i1}+y_{i2}}} \right] \text{\hspace{30mm} ‘Chi’ statistic.}$

All three coefficients above are not dimensionless, i.e. they make sense only when applied to real counts and not densities, biomass, area cover etc. Millar RB & Anderson MJ 2004, *J Exp Mar Biol Ecol* 305: 191-221 therefore suggest a scale-invariant form of the first one:

$ D^{SBinD} = \sum_i \frac{1}{ \left(y_{i1}+y_{i2} \right)} \left[y_{i1} \log \left(\frac{y_{i1}}{ y_{i1}+y_{i2}} \right) + y_{i2} \log \left(\frac{y_{i2}}{ y_{i1}+y_{i2}} \right) + \left(y_{i1} + y_{i2}\right) \log 2 \right] $ $\text{\hspace{95mm} Binomial deviance (scaled).} $

(They choose to drop the 2 outside the sum and work in logs to the base 10, so for consistency with that paper, PRIMER does the same. Resulting analyses would be unchanged either way, since the difference is just the same constant multiplier for all pairs of samples). Because of the close link between likelihood ratio and Wald statistics, D^{SBinD} is seen to be a form of Clark’s

divergence, D_{11}, though without the adjustment for double zeros that comes through the p_{12} divisor.

Cao Y, Bark AW, Williams WP 1997, *Hydrobiologia* 347: 25-40 suggested a coefficient which has been advocated or used in subsequent studies. It looks very reminiscent of the (scaled) likelihood ratio statistic, but with an important switch of the y_{i1} and y_{i2} inside the logs:

$ D^{CY} = -\frac{1}{p_{12}} \sum_i \frac{1}{ \left(y_{i1}+y_{i2} \right)} \left[y_{i1} \log \left(\frac{y_{i2}}{ y_{i1}+y_{i2}} \right) + y_{i2} \log \left(\frac{y_{i1}}{ y_{i1}+y_{i2}} \right) + \left(y_{i1} + y_{i2}\right) \log 2 \right] \text{\hspace{5mm} CY.} $

(It does take positive values in spite of the negative sign outside the sum!). Like D^{Wald} and D^{Chi}, it too contains the important p_{12} denominator adjustment to ignore joint absences, which the binomial deviance measures omit, but like D^{SBinD} it adds a denominator scaling to make the measure scale-invariant. However, it is now undefined when either $y_{i1} = 0$ (and $y_{i2} \ne 0$) or vice-versa, which could be much of the time, in fact! Zeros have to be replaced with a small positive number therefore, and the outcome is sensitive to this choice. No theoretical basis has been advanced for this coefficient, and it does not have an intuitively simple form, so any good operational properties it may possess must be somewhat fortuitous, and it is probably best avoided by the novice user.

Revision #48
Created 9 June 2024 21:47:17 by Arden
Updated 15 January 2025 02:35:54 by Abby Miller