Taxonomic dissimilarity measures

This concept of taxonomic distinctness can be carried over from a diversity index to a dissimilarity coefficient. Two measures are given under **Analyse>Resemblance**>(Measure•Other: ✓Taxonomic P/A). Both are presence/absence measures only, indicated by the plus sign superscript: Γ^+ (upper case Greek gamma) is a natural extension of Bray-Curtis dissimilarity on P/A data (the latter is just the complement of Sørensen S_8), and Θ^+ (upper case Greek theta) similarly extends Kulczynski P/A dissimilarity, the complement of S_{13}. They are formally defined as:

$\Gamma^+ = \frac{ \left(\sum_{i=1}^{s_1} \min_j \left\{ \omega_{ij} \right\} + \sum_{j=1}^{s_2} \min_i \left\{ \omega_{ij} \right\} \right) }{ \left(s_1 +s_2 \right) } \text{, \hspace{8mm}} \Theta^+ = \frac{1}{2} \left(\frac{ \sum_{i=1}^{s_1} \min_j \left\{ \omega_{ij} \right\} }{ s_1} + \frac{ \sum_{j=1}^{s_2} \min_i \left\{ \omega_{ij} \right\} }{ s_2} \right) $

where there are s_1 species present in sample 1 and s_2 in sample 2, and ω_{ij} is the distance through the tree from species i of sample 1 ($i =$ 1, 2, …, s_1) to species j of sample 2 ($j =$ 1, 2, …, s_2). This is almost simpler to express in words: for each species one finds the most closely related species in the opposite sample, then averages these minimum path lengths over all ($s_1 + s_2$) species, to obtain Γ^+. (If the nearest relation in the opposite sample is the same species, the path length is defined to be zero, of course). For Θ^+, these averages are calculated separately, i.e. the average path length for all species in sample 1 to their nearest neighbours in sample 2, then for all species in sample 2 to their nearest neighbour in sample 1, with these two averages then themselves being averaged.

As noted, these constructions result in Γ^+ and Θ^+ reducing to the dissimilarity forms of Sørensen and Kulczynski (P/A) when the hierarchy collapses, i.e. when all species are in one higher-order group and the path lengths are 0 or 100 (species do or do not have a match in the opposite sample).

Θ^+ was defined (and referred to as an ‘optimal mapping statistic’, denoted M) by Clarke KR & Warwick RM 1998, *Oecologia* 113: 278-289, and Γ^+ is (to within a constant) the TD of Izsak C & Price ARG 2001, *Mar Ecol Prog Ser* 215: 69-77. They are clearly closely related, and will be identical when $s_1=s_2$. Their use is in ordinating samples from widely-spread biogeographic regions with few, if any, shared species, but which will always have higher-order taxa in common. They also provide a certain amount of robustness in dissimilarity value to mistakes or inconsistent identification at the finest taxonomic levels (see CiMC, end of Chapter 17, for two applications from Clarke KR, Somerfield PJ, Chapman MG 2006, *J Exp Mar Biol Ecol* 330: 55-80).

Revision #20
Created 11 June 2024 20:41:57 by Arden
Updated 11 June 2024 21:16:22 by Arden