BVStep stepwise selection

There is one fundamental problem with applying BEST (Bio-Env) in many of the above scenarios: the number of variable *combinations* from the active matrix that must be considered in a full search increases exponentially with the number of variables. For *p* variables, there are (2^p – 1) combinations, and this is prohibitive for *p* more than about 16 (c. 65,000 combinations). Searching across all subsets of species from a typical community matrix will therefore usually prove impossible. The (•BVSTEP) option under **Analyse>BEST** instead carries out a stepwise search: the best single variable is selected (maximising the matching coefficient, ρ); this is retained and the best variable to add to this is selected (maximising ρ); these two are retained and a third variable is added, and so on, resulting in a declining number of combinations to be considered at each step. This is called *forward selection*. BVStep also carries out *backward elimination*: starting with all the variables included, the one that decreases ρ least, when omitted, is dropped from the set, and this elimination process repeated. In fact, as is common with stepwise procedures elsewhere (e.g. in multiple linear regression), BVStep implements both forward and backward steps successively, so that after each addition of a variable by forward selection, the current set of variables is scanned to see if any of the other variables can now be eliminated. (The analogy with stepwise multiple regression is not perfect, note, because there the residual sums of squares always decreases as more variables are added – here the ρ value may go up or *down*, giving a natural optimisation). It follows, however, from the fact that only a small fraction of the possible combinations are considered, that the routine can become trapped in a non-optimal maximum, just as *n*MDS can get trapped in a local minimum of the stress function (Section 8). The answer is the same as for MDS – repeat the search from a different starting position. So, the BVSTEP dialog lets the user specify how many random restarts are required (choose as many as are computationally feasible). Each restart is from a different, randomly chosen, combination of the variables – experience suggests that it is better not to start with too large a number because it can be difficult to shed extremely sparse variables that neither help nor harm the best solution, so the default is set at 6. Chapter 16 of CiMC gives more detail on the operation of the forward/backward stepping algorithm and the application below.

Revision #2
Created 26 October 2024 03:49:02 by Arden
Updated 25 February 2025 22:58:43 by Abby Miller