5.7 MDS strengths and weaknesses

MDS strengths

1. *MDS is simple in concept*. The numerical algorithm is undeniably complex, but it is always clear what non-metric MDS is setting out to achieve: the construction of a sample map whose inter-point distances have the *same rank order* as the corresponding dissimilarities between samples.

2. *It is based on the relevant sample information*. MDS works on the sample dissimilarity matrix not on the original data array, so that there is complete freedom of choice to define similarity of community composition in whatever terms are biologically most meaningful.

3. *Species deletions are unnecessary*. Another advantage of starting from the sample dissimilarity matrix is that the number of species on which it was based is largely irrelevant to the amount of calculation required. Of course, if the original matrix contained many species whose patterns of abundance across samples varied widely, and prior transformation (or choice of similarity coefficient) dictated that all species were given rather equal weight, then the structure in the sample dissimilarities is likely to be more difficult to represent in a low number of dimensions. More usually, the similarity measure will automatically down-weight the contribution of species that are rarer (and thus more prone to random and uninterpretable fluctuations). There is then no necessity to delete species, either to obtain reliable low-dimensional ordinations or to make the calculations viable; the computational scale is determined solely by the number of samples.

4. *MDS is generally applicable*. MDS can validly be used in a wide variety of situations; fewer assumptions are made about the nature and quality of the data when using non-metric MDS than (arguably) for any other ordination method. It seems difficult to imagine a more parsimonious position than stating that all that should be relied on is the rank order of similarities (though of course this still depends on the data transformation and similarity coefficient chosen). The step to considering only rank order of similarities, rather than their actual values, is not as potentially inefficient as it might at first appear, in cases where the resemblances are genuine Euclidean distances. Provided the number of points in the ordination is not too small (nMDS struggles when there are only 4 or 5, thus few dissimilarities to rank), nMDS will effectively reconstruct those Euclidean distances solely from their rank orders so that metric MDS (mMDS) and nMDS solutions will appear identical. The great advantage of nMDS, of course, is that it can cope equally well with very non-Euclidean resemblance matrices, commonplace in biological contexts.

5. *The algorithm is able to cope with a certain level of 'missing' similarities.* This is not a point of great practical importance because resemblances are generally calculated from a data matrix. If that has a missing sample then this results in missing values for all the similarities involving that sample, and MDS could not be expected to 'make up' a sensible

place to locate that point in the ordination! Occasionally, however, data arrives directly as a similarity matrix and then MDS can cleverly stitch together an ordination from incomplete sets of similarities, e.g. knowing the similarities A to (B, C, D) and B to (C, D) tells you quite a lot about the missing similarity of C to D. And if, as noted above, there are a reasonable number of points, so a fairly rich set of ranks, even nMDS (as found in PRIMER) would handle such missing similarities.

MDS weaknesses

1. *MDS can be computationally demanding.* The vastly improved computing power of the last two decades has made it comfortable to produce MDS plots for several hundred samples, with numerous random restarts (by default PRIMER now does 50), in a matter of a few seconds. However, for n in the thousands, it is still a challenging computation (processor time increases roughly proportional to n^2). It should be appreciated, though, that larger sample sizes generally bring increasing complexity of the sample relationships, and a 2 or 3-dimensional representation is unlikely to be adequate in any case. (Of course this last point is just as true, if not more true, for other ordination methods). Even where it is of reasonably low stress, it becomes extremely difficult to label or make sense of an MDS plot containing thousands of points. This scenario was touched on in Chapter 4 and in the discussion of Fig. 5.7, where it was suggested that data sets will often benefit by being sub-divided by the levels of a factor, or on the basis of subsets from a cluster analysis, and the groups analysed separately by MDS (agglomerative clustering is very fast, for large numbers of samples[¶]). Averages for each level might then be input to another MDS to display the large-scale structure across groups. It is the authors' experience that, far too often, users produce ordination plots from all their (replicate) samples and are then surprised that the ordination, containing many points, has high stress and little apparent pattern. Not enough use is made of averaging, whether of the transformed data matrix, the similarities, or the centroids from PCO (Anderson, Gorley & Clarke (2008)), taken over replicates, over sites for each time, over times for each site etc, and entering those averages into MDS ordinations. In univariate analysis, it is rare to produce a scatter plot of the replicates themselves: we are much more likely to plot the means for each group, or the main effects of times and sites etc (for each factor, averaging over the other factors), and the situation should be no different for multivariate data.

2. *Convergence to the global minimum of stress is not guaranteed*. As we have seen, the iterative nature of the MDS algorithm makes it necessary to repeat each analysis many times, from different starting configurations, to be fairly confident that a solution that re-appears several times (with the lowest observed stress) is indeed the global minimum of the stress function. Generally, the higher the stress, the greater the likelihood of non-optimal solutions, so a larger number of repeats is required, adding to the computational burden. However, the necessity for a search algorithm with no guarantee of the optimal solution (by comparison with the more deterministic algorithm of a PCA) should not be seen, as it sometimes has, as a defect of MDS vis-à-vis PCA. Remember that an ordination is only ever an approximation to the high-dimensional truth (the resemblance matrix) and it is much better to seek an approximate answer to the right problem (MDS on Bray-Curtis

similarity, say) rather than attack the wrong problem altogether (PCA on Euclidean distance), however deterministic the computation is for the latter.

3. *The algorithm places most weight on the large distances*. A common feature of most ordination methods (including MDS and PCA) is that more attention is given to correct representation of the *overall* structure of the samples than their *local* structure. For MDS, it is clear from the form of equation (5.1) that the largest contributions to stress will come from incorrect placement of samples which are very distant from each other. Where distances are small, the sum of squared difference terms will also be relatively small and the minimisation process will not be as sensitive to incorrect positioning. This is another reason therefore for repeating the ordination *within* each large cluster: it will lead to a more accurate display of the fine structure, if this is important to interpretation. An example is given later in Figs. 6.2a and 6.3, and is typical of the generally minor differences that result: the subset of points are given more freedom to expand in a particular direction but their relative positions are usually only marginally changed.

¶ *PRIMER has no explicit constraint on the size of matrices that it can handle; the constraints are mainly those of available RAM. On a typical laptop PC it is possible to perform sample analyses on matrices with tens of thousands of variables (species or OTUs) and hundreds of samples without difficulty; once the resemblance matrix is computed most calculations are then a function of the number of samples (n), and cluster analysis on hundreds of samples is virtually instantaneous. (The same is not true of the SIMPROF procedure, note, since it works by permuting the data matrix and is highly compute-intensive; v7 does however make good use of multi-core processors where these are available).*

Revision #10
Created 22 February 2022 10:27:01 by Arden
Updated 23 October 2024 04:15:08 by Marti