Procrustes mean shape estimation

Ben Flood

University of Limerick

Analysis of the shape of a sample of biological organisms often begins with the selection of landmarks. Landmarks are features that can consistently be identified as points on each member of the population. There are a number of different types of landmarks in use, the most common being anatomical and mathematics landmarks. The number of landmarks available depends on the structure of the organism, and there may be differences in opinion between scientists as to which landmarks were consistent and locatable across a population.

Geometric morphometrics is the most common approach to analysing landmark configurations in modern statistics. The aspect that distinguishes geometric morphometrics from traditional morphometric methods is the Procrustes registration method. The Procrustes registration method is used to superimpose landmarks configurations onto a common mean shape, according to an optimality criterion. The Procrustes algorithm uses the Procrustes registration method in an iterative approach to estimate the mean shape for a population of landmark configurations.

The Procrustes estimate of the mean shape asymptotically converges to the population mean under the right conditions (Dryden and Mardia, 1998). However, in a small sample the results may be affected by the number of landmarks identified on the organism. It seems intuitive that by finding more landmarks the estimate of the population mean could be improved. Nevertheless, the effect of increasing the number of landmarks used is not necessarily desirable.

Suppose two estimates of the mean of a \( k \)-landmark configuration are produced, the first using the Procrustes algorithm on a sample of the \( k \)-landmark configurations. The second estimate uses the Procrustes algorithm on the same sample but with an extra \( d \) identified landmarks, which are then discarded. Both estimates of the mean shape of \( k \) landmarks will be convergent towards the population mean. The uncertainty about the estimate of the mean will change.

The first estimate of the mean shape of the \( k \)-landmark configuration will in fact have uncertainty attached that is less than or equal to the uncertainty of the second estimate. The Procrustes registration of a configuration is calculated by finding a Euclidean transformation of the position of each configuration that minimises the Procrustes distances to some common mean shape. In the implementation of the Procrustes algorithm on the \( k \)-landmark configurations, the Procrustes distances are minimised. Thus, using extra landmarks can either leave the minimisation unaffected, or move the configuration such that its position is no longer optimised for the subset of \( k \) landmarks.

This poster presents the issue and uses the mice thoracic vertebrae set from the R (R Development Core Team, 2005) package shapes (Dryden, 2004) as an example.

References

