

pfinder

Tool for detection of polyhedron shaped suites of variation

User manual

Version 1.0 (April 2012) available at:

www.weizmann.ac.il/mcb/UriAlon

Written by Hila Sheftel, Omer Ramote, and Oren Shoval¹

Uri Alon lab

Department of Molecular Cell Biology

Weizmann Institute of Science

Rehovot, Israel

Instructions

1. Download Pareto.zip
2. Unpack zip
3. Perform test run:
 - a. Open DOS window
 - b. Move to the software root directory ('..\pareto\')
 - c. Run the following command: 'pareto.exe ..\Finches.csv'
 - d. Since the data in this example includes 5 dimensions, the software performs principal component analysis (PCA), and plots the first two components, showing the relative variance of each component. The software now prompts at the command line asking whether to continue to find the best-fit triangle.
 - e. Press Y to continue.
 - f. A figure should come up with a blue line indicating the convex hull of the data, and a red line indicating the best-fit triangle. In addition, the vertices of the triangle (suggesting archetypes) are given in the command line.
 - g. Close the figure window to continue.
 - h. The program now asks whether to perform a statistical analysis of the triangularity of the data. Press Y to proceed.
 - i. Enter the number of repetitions the software should perform. This results in the p-value that can be obtained. Note also that a significant number of repetitions, say 10,000, might result in a lengthy runtime.
 - j. % of repetitions with t-ratio lower than original data is returned at the command line, and program exists.

Command line options

Pareto.exe <input data file name> [-a <num of repetitions>]

Note : < > - means should be replaced by the relevant input. [] - means optional input.

Simple command line example that performs 10,000 repetitions:

```
'pareto.exe ..\finches.txt -a 10000'
```

Input file format

Input data file format should be ".csv". Each line is a data point (species) with measurements (e.g. wing length, beak length) separated by a comma.

If there are two dimensions to each datapoint, the software continues to find the convex hull and minimum area triangle.

If there are more than two dimensions (i.e. more than two measurements per species), the software first performs principal component analysis (PCA), and then uses the first two components for analysis. When the scales of the different measurements are different (e.g. wing length measured in cm, and beak length in mm), it is advisable that the user will first calculate the z-score of the data, and then will use the z-score as input to the software (the z-score is calculated by subtracting the mean value for a measurement (e.g. mean beak length across species), and then dividing by the standard deviation of that measurement). Using the z-score allows a 'fair' comparison of the different measured traits. The software will then return the archetypes in the two-dimensional PCA coordinates, and in addition, and archetypes in the original data's coordinates.

Provided data-sets

The zip file includes three datasets:

1. Measurements of beak and body sizes of Darwin's finches - 'Finches.csv'.
File includes log of data from tables A1-A5 in (2).
2. Measurements of A. Sexdens ants – 'Ants.csv'.
Data points digitized from figure 16 in (3).
3. Measurements of Microbats' wings – 'Bats.csv'.
Data from table 1 in (4).

Future versions will include:

1. Analysis of a general number of archetypes from 2 and above.
2. Additional statistics and data normalization procedures.
3. Improved graphical interface.
4. Executable options for operating systems other than Windows.

References and Notes

1. Software uses quickhull code from:
[http://en.literateprograms.org/Quickhull_\(Python,_arrays\)?action=history&offset=20091103134026](http://en.literateprograms.org/Quickhull_(Python,_arrays)?action=history&offset=20091103134026)
(Copyright (c) 2011 the authors listed at the URL, and/or the authors of referenced articles or incorporated external code)
2. P. R. Grant, I. Abbott, D. Schluter, R. L. Curry, L. K. Abbott, Variation in the size and shape of Darwin's finches, *Biological Journal of the Linnean Society* **25**, 1–39 (1985).
3. E. O. Wilson, Caste and division of labor in leaf-cutter ants (Hymenoptera: Formicidae: Atta), *Behav Ecol Sociobiol* **7**, 143–156 (1980).
4. U. M. Norberg, J. M. V. Rayner, Ecological Morphology and Flight in Bats (Mammalia; Chiroptera): Wing Adaptations, Flight Performance, Foraging Strategy and Echolocation, *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **316**, 335–427 (1987).