Correspondence analysis

Description

A simple function for correspondence analysis, following the algorithm described in Section 9.4 of Legendre and Legendre (1998).

Usage

```
CA(Y, use.svd=TRUE)
# Write total inertia, eigenvalues, relative eigenvalues, cumulative rel. eigenvalues to R window:
x   or   print(x, ...)
# Draw a biplot
biplot(x, xax=1, yax=2, scaling=1, aspect=1, cex=1, color.sites="black",
color.sp="red", ...)
```

Arguments

`Y`	Data matrix
`use.svd`	TRUE: the decomposition is done by svd (default). FALSE: the decomposition is done by eigen. The signs of the coefficients along any one axis may differ between the two methods.
`x`	Name of the output object of function CA().
`xax, yax`	Axes that will be used to draw the biplot. Default: `xax=1`, `yax=2`.
`scaling`	scaling = {1, 2, 3, 4} to obtain biplot with the corresponding scaling type. See details.
`aspect`	Aspect ratio of the plot; `aspect=1` makes the scales the same along the two axes. Use `aspect=NA` to remove the effect of parameter `aspect` in the biplot.
`cex`	A numerical value giving the amount by which plotting text and symbols should be magnified relative to the default, `cex=1`.
`color`	Color of the site and species symbols and labels in the biplots. Defaults: `color.sites="black"`, `color.sp="red"`.
`...`	Other parameters passed to the `print` or `biplot` functions.

Details

Correspondence analysis (CA) of a table of frequencies producing scaling 1, 2, 3, or 4 biplots. All variables must be frequency-like. Negative values are not allowed in CA.

Scaling type 1 biplot uses V for species and F for sites (notation as in Legendre and Legendre 1998, Section 9.4). The sites are at the centroids (barycentres) of the species. This projection preserves the chi-square distance among the sites

Scaling type 2 biplot uses Fhat for species and Vhat for sites. The species are at the centroids (barycentres) of the sites. This projection preserves the chi-square distance among the species.

Scaling type 3 biplot is a compromise between scalings 1 and 2. This scaling, called “symmetric” in CANOCO, does not preserve the chi-square distances among the species nor among the sites.

Scaling type 4 biplot: use this scaling when analyzing a contingency table, where the rows and columns are equivalent in nature. In this hybrid scaling, the positions of the rows of the original data table are represented as in scaling 1 whereas the positions of the columns are as in scaling 2. In this scaling, the relative positions of the row and column symbols along each axis of the plot are the same as in scaling 3. The axes in scaling 4 are compressed compared to the corresponding axes in scaling 3 because the eigenvalues are always smaller than 1 in CA. The compression is not isotropic, however, because the eigenvalues differ among axes.

Algorithmic notes – The data matrix is transformed into matrix Qbar of the contributions to chi-square, following equation 9.32 of Legendre and Legendre (1998). Then the matrix (t(Qbar) %*% Qbar) is decomposed by svd (default) or by eigen.

Value

Function CA returns a list containing the following results and matrices:

<u>x$general</u>

$inertia	Total inertia in matrix Qbar.
$values	CA eigenvalues.
$rel.values	Relative eigenvalues.
$cum.rel	Cumulative sum of the relative eigenvalues.

<u>x$scaling1</u>

$species, $sites	Matrices required to produce the scaling=1 biplot: V for species, F for sites, following the notation of Legendre and Legendre (1998, Section 11.2).

<u>x$scaling2</u>

$species, $sites	Matrices required to produce the scaling=2 biplot: Fhat for species, Vhat for sites.

<u>x$scaling3</u>

$species, $sites	Matrices required to produce the scaling=3 biplot: spec3 for species, site3 for sites.

<u>x$scaling4</u>

$species, $sites	Matrices required to produce the scaling=4 biplot: Fhat for species, F for sites.

x$other

$U, $Uhat	Matrices produced by SVD, used to compute the matrices needed for the biplots.
$site.names	Vector of site names.
$sp.names	Vector of species names.
$Qbar	CA is the result of the eigen-analysis of matrix Qbar. See Legendre and Legendre (1998, Section 11.2).

References

Aart, P. J. M. (van der) and N. Smeenk-Enserink. 1975. Correlations between distributions of hunting spiders (Lycosidae, Ctenidae) and environmental characteristics in a dune area. *Neth. J. Zool.* 25: 1-45.

Legendre, P. and Legendre, L. 1998. *Numerical Ecolog*y. 2nd English ed. Elsevier, Amsterdam.

Author

Pierre Legendre, Département de Sciences Biologiques, Université de Montréal.

Examples

```
# Example: data from Table 9.11 of Legendre and Legendre (1998)
table = matrix(c(10,10,15,10,15,5,20,10,5),3,3)
rownames(table) = c("Site1", "Site2", "Site3")
colnames(table) = c("Sp.1", "Sp.2", "Sp.3")
res = CA(table)
res                              # Print out a summary of the results
biplot(res) or biplot.CA(res)    # Draw a biplot (default: scaling=1)
summary(res)                     # Print the structure of the output object
res$other$U                      # Print out the matrix of eigenvectors of t(Qbar) %*% Qbar
res$scaling1$species             # Print out the matrix of species scores used in scaling 1

# Example: the spider data of Aart and Smeenk-Enserink (1975), available in library mvpart.
# The spider data frame has 28 rows and 18 columns. The first 12 columns are abundances of
different species of spiders and the next 6 are environmental data.
library(mvpart)
data(spider)                     # Note: this data file does not contain site names
res = CA(spider[,1:12])
res                              # Print out a summary of the results
biplot(res, scaling=2, color.sites="blue")  # Draw a biplot (scaling=2) with axes 1 and 2 (default)
biplot(res, xax=1, yax=3)        # Draw a biplot (scaling=1) with axes 1 and 3

# Draw two biplots side by side in a graphic window
par(mfrow=c(1,2))
biplot(res, scaling=1)
biplot(res, scaling=2)
```