Package 'Compositionalcln'

September 22, 2026

Type Package

Title Modelling Compositional Data with Zero Values

Version 1.0

Date 2026-09-12

Author Michail Tsagris [aut, cre]

Maintainer Michail Tsagris <mtsagris@uoc.gr>

Depends R (>= 4.0)

Imports graphics, grDevices, mziln, rangen, Rfast, stats

Suggests Compositional, Rfast2

Description Modelling structural zeros in compositional data using a conditional logistic normal model as described by Aitchison (1986), where MLE (Maximum Likelihood Estimation) is performed via the EM (Expectation-Maximization) algorithm. The relevant paper is Alzeley and Tsagris (2026) <doi:10.48550/arXiv.2608.29954>.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2026-09-22 07:10:16 UTC

Contents

Compositionalcln-package

Modelling Compositional Data with Zero Values

Description

Modelling Compositional Data with Zero Values.

Details

Package:	Compositionalcln
Type:	Package
Version:	1.0
Date:	2026-09-12

Maintainers

Michail Tsagris <mtsagris@uoc.gr>.

Author(s)

Michail Tsagris <mtsagris@uoc.gr>

References

Alzeley O. and Tsagris M. (2026). Modelling compositional data with structural zero values https://arxiv.org/pdf/2608.29954

Aitchison J. (1986). The statistical analysis of compositional data.

boo.clnreg

Bootstrap for the conditional logistic normal regression

Description

Bootstrap for the conditional logistic normal regression.

Usage

```
boot.clnreg(y, x, tol = 1e-6, maxit = 500, R = 1000)
```

Arguments

y	A matrix with the compositional data (dependent variable). The number of observations (vectors) with no zero values should be more than the columns of the predictor variables. Otherwise, the initial values will not be calculated.
x	The predictor variable(s), they can be either continnuous or categorical or both.
tol	The tolerance value to terminate the EM algorithm.
maxit	The maximum number of iterations allowed for the EM algorithm.
R	The number of bootstrap resamples to perform.

Details

Bootstrap to estimate the covariance matrix of the coefficients of the conditional logistic normal regression is performed.

Value

A list including:

beta	The bootstrapped beta coefficients.
sigma	The bootstrap estimate of the covariance matrix of the regression parameters.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Alzeley O. and Tsagris M. (2026). Modelling compositional data with structural zero values https://arxiv.org/pdf/2608.29954

Aitchison J. (1986). The statistical analysis of compositional data, pages 272–274.

See Also

cln.reg

Examples

```
x <- as.vector(iris[, 4])
y <- as.matrix(iris[, 1:3])
y <- y / rowSums(y)
ind <- sample(150, 10)
for ( k in ind )  y[k, sample(3, 1)] <- 0
y <- y / rowSums(y)
mod <- cln.reg(y, x)
```

`cln.contour` *Contour plot of the conditional logistic normal distribution in S^2*

Description

Contour plot of the conditional logistic normal distribution in S^2.

Usage

```
cln.contour(m, S, n = 100, y = NULL, cont.line = FALSE)
```

Arguments

`m`	A value with the concentration parameter.
`S`	A vector with the mean vector (3 values) in the simplex.
`n`	The number of grid points to consider over which the density is calculated.
`y`	This is either NULL (no data) or contains a 3 column matrix with compositional data.
`cont.line`	Do you want the contour lines to appear? If yes, set this TRUE.

Details

The user can plot only the contour lines of a zero adjusted Dirichlet distribution with som given parameters, or can also add the relevant data should he/she wish to.

Value

A ternary diagram with the points and the conditional multivariate normal contour lines.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Alzeley O. and Tsagris M. (2026). Modelling compositional data with structural zero values https://arxiv.org/pdf/2608.29954

Aitchison J. (1986). The statistical analysis of compositional data, pages 272–274.

See Also

`dcln`

Examples

```
y <- as.matrix(iris[, 1:3])
y <- y / rowSums(y)
ind <- sample(150, 10)
for ( k in ind )  y[k, sample(3, 1)] <- 0
y <- y / rowSums(y)
mod <- cln.mle(y)

cln.contour( m = mod$m, S = mod$S )
```

`cln.mle` *Maximum likelihood estimation of the conditional logistic normal model*

Description

Maximum likelihood estimation of the conditional logistic normal model.

Usage

```
cln.mle(y, tol = 1e-6, maxit = 500)
```

Arguments

`y`	A numerical matrix with compositional data with zero values.
`tol`	The tolerance value to terminate the EM algorithm.
`maxit`	The maximum number of iterations allowed for the EM algorithm.

Details

The function fits the conditional multivariate normal distribution described in Aitchison (1986), pages

Value

A list including:

`m`	The mean vector in the Euclidean space.
`mesi`	The mean vector in the simplex space.
`S`	The estimated covariance matrix.
`loglik`	The log-likelhiood value.
`patterns`	A matrix with the patterns of zeros, where the value of 0 indicates the presence of a zero, and the last column contains the percentage of occurrence each pattern. This is useful for random values simulation.
`iters`	The number of iterations required by the EM algorithm.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Alzeley O. and Tsagris M. (2026). Modelling compositional data with structural zero values https://arxiv.org/pdf/2608.29954

Aitchison J. (1986). The statistical analysis of compositional data, pages 272–274.

See Also

cln.reg

Examples

```
y <- as.matrix(iris[, 1:3])
y <- y / rowSums(y)
ind <- sample(150, 10)
for ( k in ind )  y[k, sample(3, 1)] <- 0
y <- y / rowSums(y)
cln.mle(y)
```

---

cln.reg *Conditional logistic normal regression*

Description

Conditional logistic normal regression.

Usage

```
cln.reg(y, x, tol = 1e-6, maxit = 500, xnew = NULL)
```

Arguments

y	A matrix with the compositional data (dependent variable). The number of observations (vectors) with no zero values should be more than the columns of the predictor variables. Otherwise, the initial values will not be calculated.
x	The predictor variable(s), they can be either continnuous or categorical or both.
tol	The tolerance value to terminate the EM algorithm.
maxit	The maximum number of iterations allowed for the EM algorithm.
xnew	If you have new data use it, otherwise leave it NULL.

Details

The conditional logistic normal regression is being fitted. The likelihood conists of two components. The contributions of the non zero compositional values and the contributions of the compositional vectors with at least one zero value. The second component may have many different sub-categories, one for each pattern of zeros.

Value

A list including:

`patterns`	A matrix with the patterns of zeros, where the value of 0 indicates the presence of a zero, and the last column contains the percentage of occurrence each pattern. This is useful for random values simulation.
`iters`	The number of iterations required by the EM algorithm.
`beta`	The beta coefficients.
`loglik`	The value of the log-likelihood.
`est`	The fitted or the predicted values (if xnew is not NULL).

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Alzeley O. and Tsagris M. (2026). Modelling compositional data with structural zero values https://arxiv.org/pdf/2608.29954

Aitchison J. (1986). The statistical analysis of compositional data, pages 272–274.

See Also

`cln.mle`

Examples

```
x <- as.vector(iris[, 4])
y <- as.matrix(iris[, 1:3])
y <- y / rowSums(y)
ind <- sample(150, 10)
for ( k in ind )  y[k, sample(3, 1)] <- 0
y <- y / rowSums(y)
mod <- cln.reg(y, x)
```

dcln *Density of the conditional logistic normal model*

Description

Density of the conditional logistic normal model.

Usage

```
dcln(y, m, S, logged = FALSE)
```

Arguments

y	A vector or a matrix with compositional data with zero values.
m	The mean vector in R^{D-1}.
S	The covariance matrix in R^{D-1}.
logged	If you want the log of the density set this TRUE, otherwise leave it FALSE.

Details

The function computes the density values of the conditional multivariate normal model.

Value

The density values at the given compositional data x.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Alzeley O. and Tsagris M. (2026). Modelling compositional data with structural zero values https://arxiv.org/pdf/2608.29954

Aitchison J. (1986). The statistical analysis of compositional data, pages 272–274.

See Also

cln.mle, cln.contour

Examples

```
x <- as.vector(iris[, 4])
y <- as.matrix(iris[, 1:3])
y <- y / rowSums(y)
ind <- sample(150, 10)
for ( k in ind )  y[k, sample(3, 1)] <- 0
y <- y / rowSums(y)
mod <- cln.mle(y)
f <- dcln(y, mod$m, mod$S)
```

`Ternary diagram` *Ternary diagram*

Description

Ternary diagram.

Usage

```
ternary(y, dg = FALSE, hg = FALSE, m = NULL, colour = NULL)
```

Arguments

`y`	A matrix with the compositional data.
`dg`	Do you want diagonal grid lines to appear? If yes, set this TRUE.
`hg`	Do you want horizontal grid lines to appear? If yes, set this TRUE.
`m`	If you know the estimated mean pass it here, otherwise leave it NULL. It will appear with the symbol +.
`colour`	If you want the points to appear in different colour put a vector with the colour numbers or colours.

Details

There are two ways to create a ternary graph. We used here that one where each edge is equal to 1 and it is what Aitchison (1986) uses. For every given point, the sum of the distances from the edges is equal to 1. Horizontal and or diagonal grid lines, and the mean can appear. Zeros in the data appear with red x.

Value

The ternary plot. Additionally, horizontal or diagonal grid lines can appear as well.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Aitchison, J. (1983). Principal component analysis of compositional data. Biometrika 70(1): 57–65.

Aitchison J. (1986). The statistical analysis of compositional data. Chapman & Hall.

See Also

cln.contour

Examples

```
y <- as.matrix(iris[, 1:3])
y <- y / rowSums(y)
ind <- sample(150, 10)
for ( k in ind )  y[k, sample(3, 1)] <- 0
y <- y / rowSums(y)
mod <- cln.mle(y)
ternary(y, hg = TRUE, dg = TRUE, m = mod$mesi)
```

Index